



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

“Regulación transcripcional del gen *CHD5* por CTCF y KDM4A vía la modificación postraduccional H3K36me en un modelo de cáncer de mama.”

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. LISSANIA XIMENA GUERRA CALDERAS

TUTOR PRINCIPAL:

DR. ERNESTO SOTO REYES SOLÍS

Universidad Autónoma Metropolitana- Unidad Cuajimalpa

COMITÉ TUTOR:

DR. ALEJANDRO ZENTELLA DEHESA

Instituto de Investigaciones Biomédicas- UNAM

DR. ÁNGEL ALFONSO ZARAIN HERZBERG

Facultad de Medicina- UNAM

Ciudad de México, Agosto, 2020.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

AGRADECIMIENTOS INSTITUCIONALES.

El presente trabajo fue realizado en el laboratorio de Biología Celular en la Universidad Autónoma Metropolitana unidad Cuajimalpa, y en el Laboratorio de Virus y Cáncer, a cargo del Dr. Alejandro Manuel García Carranca, de la Unidad de Investigación Biomédica en Cáncer del Instituto de Investigaciones Biomédicas y el Instituto Nacional de Cancerología, bajo la asesoría del Dr. Ernesto Soto Reyes Solís y el Dr. Rodrigo González Barrios de la Parra. Tesis apoyada por el programa de Fondo Sectorial de Investigación en Salud y Seguridad Social (FOSISS, Grant No. 0261181), Fondo CB-SEP-CONACyT (284748) y el Fondo Desarrollo Tecnológico e Innovación COVID-19 (312021); y UAM-PTC-704.

Los estudios de doctorado de la M. en C. Lissania Ximena Guerra Calderas fueron apoyados por una beca de doctorado, otorgada por el programa nacional de becas del Consejo Nacional de Ciencia y Tecnología con número de CVU 588391. La estancia a la que asistió Lissania Ximena Guerra Calderas fue apoyada por el Programa de Apoyo a los Estudios del Posgrado (PAEP).

El comité tutor que asesoró el desarrollo de esta tesis estuvo formado por:

Dr. Ernesto Soto Reyes Solís Universidad Autónoma Metropolitana- Unidad Cuajimalpa
Dr. Alejandro Zentella Dehesa Instituto de Investigaciones Biomédicas- UNAM
Dr. Ángel Alfonso Zarain Herzberg Facultad de Medicina- UNAM

El jurado de examen estuvo integrado por:

Presidente: Dr. Alejandro Manuel García Carrancá
Vocal: Dra. Lorena Aguilar Arnal
Vocal: Dra. Ma. de la Paz Sánchez Jimenez
Vocal: Dr. Carlos Fabián Flores Jasso
Secretaria: Dra. Martha Verónica Vázquez Laslop

ÍNDICE.

1.	ÍNDICE DE FIGURAS.	6
2.	ÍNDICE DE TABLAS.	7
3.	ÍNDICE DE ABREVIATURAS.	8
4.	RESUMEN.	11
5.	ABSTRACT.	12
6.	INTRODUCCIÓN.	13
6.1	LA ESTRUCTURA DE LA CROMATINA Y LA REGULACIÓN TRANSCRIPCIONAL.	13
6.2	LAS MODIFICACIONES POSTRADUCCIONALES DE LAS HISTONAS.	13
6.3	LAS DESMETILASAS DE HISTONAS.	15
6.4	ESTRUCTURA PROTEICA Y MECANISMO DE DESMETILACIÓN DE LA DESMETILASA DE HISTONAS KDM4A.	16
6.5	FUNCIONES DE LA DESMETILASA DE HISTONAS KDM4A.	18
6.6	EL PAPEL DE KDM4A EN EL DESARROLLO DE CÁNCER.	19
6.6.1	Cáncer de pulmón.	20
6.6.2	Cáncer de mama.	20
6.7	EL GEN <i>CHD5</i> .	22
6.8	PARTICIPACIÓN DE H3K36me3 EN LA REGULACIÓN DE LA TRANSCRIPCIÓN.	24
6.9	REGULACIÓN DE LAS DESMETILASAS DE HISTONAS.	25
6.10	EL FACTOR NUCLEAR CTCF.	27
6.10.1	FUNCIONES DE CTCF.	28
6.10.2	PROTEÍNAS ASOCIADAS A CTCF.	28
7	ANTECEDENTES DIRECTOS.	30
7.1	<i>KDM4A se sobreexpresa en las líneas celulares MCF7, MDA-MB-231 Y HeLa.</i>	30
7.2	Presencia de CTCF, KDM4A y las modificaciones postraduccionales H3K36me3 y H3K36me2 en el primer intrón de <i>CHD5</i> .	32
8	PREGUNTA DE INVESTIGACIÓN.	35
9	HIPÓTESIS.	35
10	OBJETIVO GENERAL.	35
10.1	Objetivos particulares.	35
11	RESULTADOS.	36
11.1	Análisis de metilación del DNA en el promotor de <i>CHD5</i> .	36
11.2	El complejo CTCF-KDM4A se localiza en el primer intrón de <i>CHD5</i> en las líneas celulares MCF7 y HeLa.	37
11.3	El Knockdown de KDM4A o CTCF por medio siRNA y shRNA está relacionado con el incremento de la expresión de <i>CHD5</i> .	40

11.4	El Knockout de KDM4A en células MCF7 restablece la marca de histonas H3K36me3 en el primer intrón de <i>CHD5</i> y reactiva su expresión.....	42
11.5	El Knockout de KDM4A en células MCF7 afecta los niveles globales de H3K36me3 y H3K36me2.	44
11.6	Análisis de la distribución de las proteínas CTCF y KDM4A y el complejo CTCF-KDM4A en diversos genes.....	46
12	DISCUSIÓN.	49
13	CONCLUSIONES.	54
14	PERSPECTIVAS.	54
15	<i>Identificación y caracterización de fármacos moduladores de la actividad enzimática de la desmetilasa de histonas KDM4A.</i>	55
15.1	OBJETIVO GENERAL.	56
15.1.1	<i>Objetivos particulares.</i>	56
15.2	RESULTADOS.	56
15.2.1	Los compuestos PKF118-310, FS21, FS22 y FS23 inhiben a la enzima KDM4A.	56
15.2.2	<i>Los compuestos PKF118-310, FS21, FS22 y FS23 se unen directamente a la enzima KDM4A in vitro.</i>	57
15.2.3	<i>El compuesto FS23 se une directamente a KDM4A in vivo.</i>	58
15.2.4	<i>La inducción con los inhibidores afecta la abundancia de las marcas de las histonas H3K9me3, H3K9me2 y H3K9me1.</i>	59
15.2.5	<i>La inducción con el epifármaco PKF118-310 de las células KDM4A^{KO} y WT afecta los niveles globales de H3K9me3, H3K9me2, H3K9me1, H3K36me3 y H3K36me2.</i>	
	63	
15.3	DISCUSIÓN.	65
15.4	CONCLUSIONES.	67
15.5	PERSPECTIVAS.	67
16	INFORMACIÓN SUPLEMENTARIA.	67
16.1	Validación de los ensayos de ChIP contra KDM4A, CTCF y H3K36me3.....	67
16.2	Validación del ensayo de ChIP-Re-ChIP contra KDM4A y CTCF.	71
16.3	Citometría de flujo y selección celular.	73
17	MÉTODOS.	74
17.1	EXTRACCIÓN DE RNA.	74
17.2	REACCIÓN DE TRANSCRIPTASA REVERSA PARA OBTENER EL cDNA Y PCR CUANTITATIVA.	76
17.3	INMUNOFLUORESCENCIAS (IF).	78
17.4	ENSAYO DE INMUNOPRECIPITACIÓN DE LA CROMATINA (ChIP) E INMUNOPRECIPITACIÓN SECUENCIAL DE LA CROMATINA (ChIP-Re-ChIP).	80
17.5	COINMUNOPRECIPITACIÓN.	88
17.6	CONVERSION POR BISULFITO DE SODIO Y MS-PCR.	89

17.7	TRANSFECCIÓN DE siRNAs y shRNAs.	92
17.8	TRANSFECCIÓN DE PLÁSMIDOS CRISPR/CAS.	92
17.9	CITOMETRÍA DE FLUJO Y SELECCIÓN CELULAR.....	93
17.10	EXTRACCIÓN ÁCIDA DE HISTONAS.	93
17.11	ENSAYOS DE WESTERN BLOT.	94
17.12	<i>ENSAYO ENZIMÁTICO.</i>	<i>99</i>
17.13	<i>ENSAYO DE LABEL-FREE.</i>	<i>99</i>
17.14	<i>ENSAYO DE CAMBIO TÉRMICO CELULAR (CETSA).....</i>	<i>100</i>
18	ANEXOS.	101
18.1	OLIGONUCLEÓTIDOS SINTÉTICOS EMPLEADOS.....	101
18.1.1	<i>Análisis de expresión.</i>	<i>101</i>
18.1.2	Ensayos de MS-PCR e inmunoprecipitación de la Cromatina (ChIP).....	102
18.2	ANTICUERPOS UTILIZADOS.	103
18.2.1	Anticuerpos primarios.....	103
18.2.2	Anticuerpos secundarios.....	104
18.3	LÍNEAS CELULARES.	105
19	REFERENCIAS.....	106
20	ARTÍCULOS PUBLICADOS.	111
21	ARTÍCULOS EN PROCESO.	226

1. ÍNDICE DE FIGURAS.

Figura 1: “Writers”, “readers” y “erasers” de las MPT de las histonas.....	14
Figura 2: Estructura y mecanismo de desmetilación de la enzima KDM4A.....	17
Figura 3: Papel de KDM4A en la regulación de <i>CHD5</i>	24
Figura 4: Modelos de reclutamiento de las desmetilasas de histonas.	27
Figura 5: Localización de CTCF a lo largo de <i>CHD5</i>	30
Figura 6: <i>KDM4A</i> se sobreexpresa en las líneas celulares MCF7, MDA-MB-231 y HeLa. 31	
Figura 7: La presencia de KDM4A y CTCF en el primer intrón de <i>CHD5</i> se relaciona con la disminución de H3K36me3 y H3K36me2 en esta región en las líneas celulares MCF7, MDA-MB-231 y HeLa.....	34
Figura 8: La disminución de la expresión de <i>CHD5</i> , no esta relacionada con la metilación del DNA en su promotor en las líneas celulares HeLa y MCF7.	37
Figura 9: El complejo CTCF-KDM4A se localiza en el primer intrón de <i>CHD5</i> en las líneas celulares MCF7 y HeLa.....	39
Figura 10: La expresión de <i>CHD5</i> aumenta después del Knockdown de CTCF y KDM4A en las líneas celulares MCF7 y HeLa.	41
Figura 11: El Knockout de KDM4A promueve el restablecimiento de la MPT H3K36me3 en el primer intrón de <i>CHD5</i> así como el aumento en su expresión.....	44
Figura 12: El Knockout de KDM4A afecta los niveles globales de H3K36me2 y H3K36me3.	46
Figura 13: Calidad de las secuencias.	47
Figura 14: Distribución de CTCF, KDM4A y el complejo proteico CTCF-KDM4A en distintos genes por medio de CHIP-seq y CHIP-Re-ChIP-seq.	49
Figura 15: Modelo de represión transcripcional de <i>CHD5</i> mediado por el complejo proteico CTCF-KDM4A.....	53
Figura 16: Los compuestos PKF118-310, FS21, FS22 y FS23 actúan como inhibidores de la enzima KDM4A.....	57
Figura 17: Los compuestos PKF118-310, FS21, FS22 y FS23 se unen a la enzima KDM4A in vitro.	58
Figura 18: El compuesto FS23 se une directamente a KDM4A.	59
Figura 19: Las modificaciones postraduccionales de las histonas H3K9me3, H3K9me2, H3K9me1, se ven afectadas después del tratamiento con los inhibidores en la línea celular HCT-116.....	61
Figura 20: Evaluación de las modificaciones postraduccionales de las histonas H3K9me3, H3K9me2, H3K9me1 después del tratamiento con los inhibidores en las líneas celulares MCF7 y PANC-1.....	63
Figura 21: La inducción con el epifármaco PKF118-310 afecta los niveles globales de H3K36me3 en las células WT y KDM4A ^{KO}	64
Figura 22: Estructura de los inhibidores empleados.	65
Figura 23: Controles de las inmunoprecipitaciones de KDM4A, CTCF y H3K36me3.....	70
Figura 24: Controles del CHIP-Re-ChIP contra KDM4A y CTCF.	73
Figura 25: Citometría de flujo y selección celular.....	73
Figura 26: Gel representativo de la evaluación de la cromatina sonicada.	82

2. ÍNDICE DE TABLAS.

Tabla 1: Estructura proteica de las desmetilasas de histonas.....	15
Tabla 2: Participación de la desmetilasa de histonas KDM4A en diversos tipos de cáncer.....	19
Tabla 3: Análisis estadístico de los ensayos ChIP-seq y ChIP-Re-ChIP-seq (ChIP-seq).....	47
Tabla 4: Disoluciones usadas para determinar la integridad del RNA.....	75
Tabla 5: Componentes usados para cada reacción de retrotranscripción.....	77
Tabla 6: Componentes usados para cada reacción de qPCR en el análisis de expresión.....	78
Tabla 7: Disoluciones empleadas para los ensayos de inmunofluorescencia.....	78
Tabla 8: Disoluciones empleadas para los ensayos de ChIP y ChIP-Re-ChIP.....	80
Tabla 9: Componentes usados para cada reacción de qPCR en el análisis de los ensayos de ChIP.....	85
Tabla 10: Componentes usados para cada reacción de qPCR en el análisis de los ensayos de ChIP-Re-ChIP.....	88
Tabla 11: Componentes del buffer de lisis para llevar a cabo los ensayos de inmunoprecipitación.....	89
Tabla 12: Disoluciones empleadas en los ensayos de Western Blot.....	94
Tabla 13: Cantidades y concentraciones para preparar de la curva estándar de albúmina.....	96
Tabla 14: Oligonucleótidos empleados para llevar a cabo el análisis de expresión.....	101
Tabla 15: Oligonucleótidos empleados para analizar la metilación y evaluar los ensayos de ChIP y ChIP-Re-ChIP.....	102
Tabla 16: Anticuerpos primarios empleados.....	103
Tabla 17: Anticuerpos secundarios empleados.....	104
Tabla 18: Líneas celulares usadas en este estudio.....	105

3. ÍNDICE DE ABREVIATURAS.

5hmC: 5-hidroximetil-citosina.

5mC: 5-metil-citocina.

AOL-C: región carboxilo-terminal que contiene el dominio amino-oxidas-like.

AOL-N: región amino-terminal que contiene el dominio amino-oxidas-like.

ARHI: Aplasia Ras Homolog member I.

ARID: AT-Rich Interaction Domain.

ASCL2: Achaete-Scute Complex Homolog 2.

BAX: BCL2 Associated X Protein.

C5HC2 : dedo de Zinc de tipo C5HC2.

CCND1: Cyclin D1.

CENP-E: Centromere Protein E.

CETSA: Cellular Thermal Shift Assay.

CHD5: Chromodomain Helicase DNA-binding protein 5.

CHD8: Chromodomain Helicase DNA-binding protein 8.

ChIP-Re-ChIP: Inmunoprecipitación secuencial de la cromatina.

ChIP: Inmunoprecipitación de la cromatina.

CIITA: Class II Major Histocompatibility Complex Transactivator.

CP190: Centrosome-associated Zinc Finger Protein CP190.

CpG: Citosina-fosfato-Guanina.

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeat

CTCF: CCCTC-Binding Factor.

CTD: Carboxi-Terminal Domain.

CXXC-ZF: dedo de Zinc de tipo CXXC.

DAPI: 4',6-diamino-2-fenilindol.

DNA: Ácido Desoxirribonucleico.

ENCODE: Encyclopedia Of DNA Elements.

ER: Estrogen Receptor.

FAD: Flavín Adenín Dinucleótido.

GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase.

GFAP: Glial Fibrillary Acid Protein.

H1: Histona 1.

H2A: Histona 2A.

H2A.Z: Variante Z de la histona H2A.

H2B: Histona 2B.

H3: Histona 3.

H3K27me3: monometilación de la lisina 27 de la histona 3.

H3K36me1: monometilación de la lisina 36 de la histona 3.

H3K36me2: dimetilación de la lisina 36 de la histona 3.

H3K36me3: trimetilación de la lisina 36 de la histona 3.

H3K9me1: monometilación de la lisina 9 de la histona 3.

H3K9me2: dimetilación de la lisina 9 de la histona 3.

H3K9me3: trimetilación de la lisina 9 de la histona 3.

H4: Histona 4.

HAT: Histone Acetyl Transferase.

HCP: High CpG Promoter.

HDAC: Histone Deacetylase.

HP1 γ : Heterocromatin Protein 1 gamma.

hTERT: human Telomerase Reverse Transcriptase.

IF: Inmunofluorescencia.

IGF2: Insulin-like growth factor 2.

JmjC: Jumonji C Domain.

JmjN: Jumonji N Domain.

KDM1A: Lysine Demethylase 1 A.

KDM2: Lysine Demethylase 2.

KDM4A: Lysine Demethylase 4A.

KDM5B: Lysine Demethylase 5B.

LBP: Lipopolysaccharide-Binding Protein.

LCP: Low CpG Promoter.

LRR: Leucine-Rich Repeat.

LSD1: Lysine Specific Demethylase 1.

MDM2: Mouse Double Minute 2.

MPT: Modificaciones Postraduccionales.

mRNA: RNA Mensajero.

N-CoR: Nuclear Receptor Corepressor 1

NeP1: Negative Protein 1.

NSCLC: Non- Small Cell Lung Carcinoma.

NuRD: Nucleosome Remodeling and Deacetylation.

PARP1: Poly [ADP-ribose] polymerase 1.
Pb: pares de bases.
PCR: Polymerase Chain Reaction.
PHD: Plant Homeodomain.
PSA: Persulfato de Amonio.
qPCR: Quantitative Polymerase Chain Reaction.
Rad51: DNA repair protein RAD51 homolog 1.
RFX: Regulatory Factor X.
RNA: Ácido Riibunucléico.
RT-qPCR: Reverse Transcription Quantitative Polymerase Chain Reaction.
SETD2: SET Domain-Containing Protein 2.
shRNA: Short Hairpin RNA.
SIN3A: Histone Deacetylase Complex Subunit Sin3a.
siRNA: Small Interfering RNA.
SirT1: Sirtuin 1.
SMAD3: Mothers Against Decapentaplegic Homolog 3.
SUZ12: Suppressor of Zeste 12.
SWIRM: SWI3p, Rsc8p, and Moira.
TCA: Tricarboxylic Acid Cycle.
TCGA: The Cancer Genome Atlas.
TFAP4: Transcription Factor AP-4.
Topo II: DNA Topoisomerase II.
TPR: Tetratricopeptide Repeat.
WB: Western Blot.
WEE1: Wee1-Like Protein Kinase.
YB-1: Y-Box Binding Protein 1.
YY1: Ying Yang 1.

4. RESUMEN.

La desmetilasa histonas KDM4A remueve los grupos metilo de las modificaciones postraduccionales de las histonas, H3K9me3 y H3K36me3, las cuales son marcas epigenéticas asociadas con el silenciamiento génico y la elongación de la transcripción por la RNA polimerasa II, respectivamente. Diversos estudios reportan que *KDM4A* se sobreexpresa en diferentes tipos de cáncer y afecta la expresión de múltiples genes blanco, entre ellos, *CHD5*. *KDM4A* se localiza en el primer intrón de *CHD5*, y su disociación aumenta la expresión de *CHD5*. Ensayos *in vitro* mostraron que la desmetilación mediada por KDM4A aumenta en presencia de CTCF, lo que sugiere que CTCF podría aumentar su actividad enzimática *in vivo*, sin embargo, el mecanismo específico por el cual CTCF y KDM4A están implicadas en la regulación de *CHD5* es poco conocido. En este trabajo mostramos que CTCF y KDM4A forman un complejo proteico que se recluta en el primer intrón de *CHD5*. Esto correlacionó con una disminución en las marcas de histonas H3K36me3 y H3K36me2, a su vez, esta disminución, se asoció con la reducción de los niveles de mRNA de *CHD5*. El "knockdown" de *CTCF* o *KDM4A* mediado por siRNAs, desencadenó la reactivación de la expresión de *CHD5*, lo cual sugiere que ambas proteínas están involucradas en la regulación negativa de este gen. Localmente, el "knockout" de *KDM4A* mediado por CRISPR/Cas9 restauró la expresión de *CHD5* así como las marcas de histonas H3K36me3 y H3K36me2 en el primer intrón de *CHD5*. Globalmente, el "knockout" de *KDM4A* también afectó los niveles totales de H3K36me3 y H3K36me2. Finalmente, al realizar ensayos ChIP-seq y ChIP-Re-ChIP-seq, encontramos que CTCF y KDM4A coexisten en nueve sitios a lo largo de *CHD5*.

Nuestros hallazgos proponen un nuevo mecanismo de represión epigenética en el cuerpo del gen que no implica el silenciamiento del promotor.

5. ABSTRACT.

The histone demethylase KDM4A is involved in H3K9me3 and H3K36me3 demethylation, which are epigenetic modifications related to gene silencing and transcription elongation, respectively. *KDM4A* is abnormally expressed in cancer, affecting the expression of multiple target genes, such as *CHD5*. The KDM4A enzyme localizes at the first intron of *CHD5*, and the dissociation of this protein increases *CHD5* expression. Additionally, *in vitro* assays showed that KDM4A-mediated demethylation is enhanced by the presence of CTCF, suggesting that CTCF could increase its enzymatic activity *in vivo*, however, the specific mechanism by which CTCF and KDM4A might be involved in the *CHD5* regulation is poorly understood. In this work, we show that CTCF and KDM4A form a protein complex, which is recruited to the first intron of *CHD5*, a fact that is correlated with a decrease in H3K36me3 and H3K36me2 histone marks and transcriptional downregulation. The depletion of *CTCF* or *KDM4A* mediated by siRNAs, triggered the reactivation of *CHD5* expression, suggesting that both proteins are involved in the negative regulation of this gene. Locally, the knockout of *KDM4A* mediated by CRISPR/Cas9 restored *CHD5* expression and H3K36me3, and H3K36me2 histone marks. Globally, the knockout of KDM4A also affected the total levels of H3K36me3 and H3K36me2. Finally, by performing ChIP-seq and ChIP-Re-ChIP-seq assays, we found that CTCF and KDM4A coexist in nine sites along *CHD5*.

Our findings propose a novel epigenetic mechanism of repression at the gene body that does not involve promoter silencing.

6. INTRODUCCIÓN.

6.1 LA ESTRUCTURA DE LA CROMATINA Y LA REGULACIÓN TRANSCRIPCIONAL.

El DNA de eucariontes está empacado en un complejo formado por proteínas y RNA conocido como cromatina, la unidad fundamental de este complejo es el nucleosoma, el cual se encuentra conformado por 165 pares de bases (pb) enrolladas alrededor de un octámero de histonas (H2A, H2B, H3 y H4, un par de cada una) (revisado en Felsenfeld y Groudine, 2003; Travers et al., 2012). La cromatina se puede estructurar en una fibra más compacta de 30 nm de diámetro conocida como solenoide, el cual, se estabiliza por la unión de la histona H1. A este nivel se forma la heterocromatina (revisado en Luger y Hansen, 2005).

La organización del genoma en cromatina tiene repercusiones directas en la actividad transcripcional, siendo necesaria la reestructuración de la cromatina con el fin de expresar genes específicos en un tiempo y espacio bien definido. Las proteínas responsables de llevar a cabo el remodelaje de la cromatina contribuyen a la realización de diversos procesos epigenéticos. (revisado en Kim y Kaang, 2017). Las modificaciones en la arquitectura de la cromatina son reguladas por distintos mecanismos epigenéticos, entre ellos, la metilación del DNA, ciertos RNAs no codificantes, las variantes de histonas y las modificaciones postraduccionales de las histonas (MPT).

6.2 LAS MODIFICACIONES POSTRADUCCIONALES DE LAS HISTONAS.

Las MPT de las histonas funcionan como señales altamente específicas para la regulación transcripcional (revisado en Felsenfeld y Groudine, 2003; Li et al., 2007). Los nucleosomas y las histonas que los conforman son blancos de múltiples modificaciones epigenéticas, las cuales al ser reconocidas por diversos complejos proteicos regulan la relajación o compactación de la cromatina. Entre los cambios bioquímicos que presentan las histonas se encuentran: la ADP-ribosilación, la metilación de lisinas y argininas, la fosforilación de serinas y treoninas, la ubiquitinación, la SUMOilación y la acetilación de lisinas, entre otras (revisado en Rando, 2012).

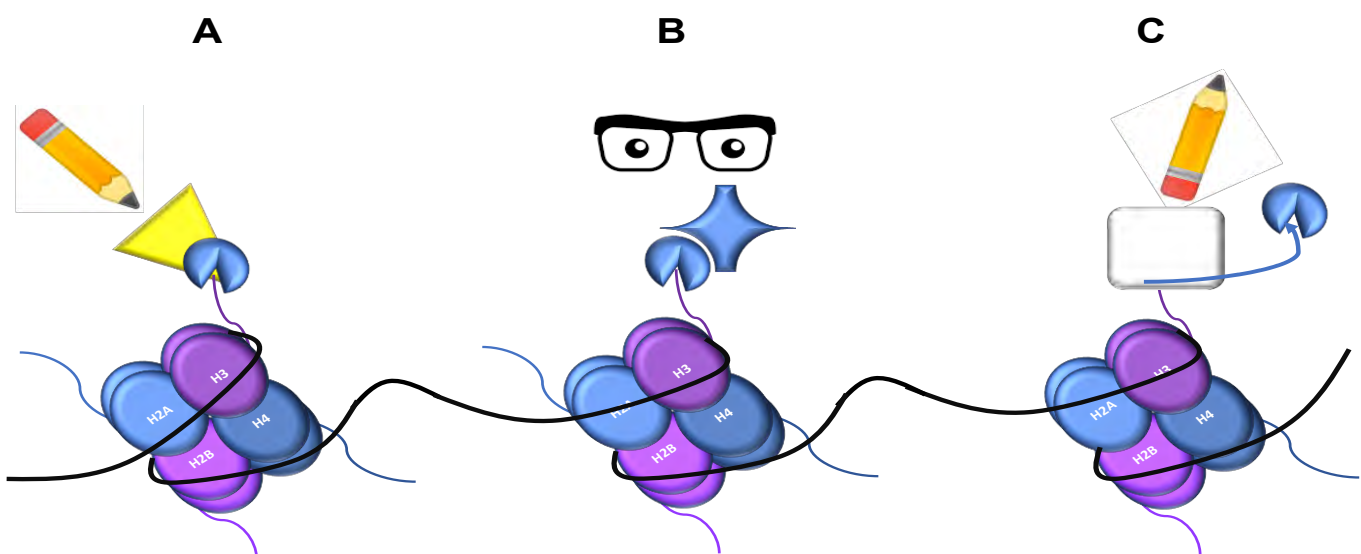
Una modificación extensamente estudiada es la acetilación de histonas, la cual es un proceso dinámico controlado por la actividad enzimática de acetiltransferasas (HAT, "Histone acetyltransferases") y desacetilasas (HDAC, "Histone Deacetylases"); estas modificaciones provocan la neutralización de cargas en las histonas lo cual resulta en la relajación de la cromatina (revisado en Bannister y Kouzarides, 2011; Peserico y Simone, 2011).

Se ha sugerido que la unión de algunas proteínas lectoras es influenciada por las modificaciones epigenéticas adyacentes, un ejemplo de este fenómeno es la interferencia de la fosforilación de la serina 10 de la histona tres, debido al proceso de metilación de la lisina nueve de la histona tres (Rea et al., 2000).

Otra modificación ampliamente estudiada, es la metilación de lisinas de las histonas, esta modificación epigenética puede ser reconocida por al menos 4 dominios proteicos: **i)** cromodominio, **ii)** dominios tipo TUDOR, **iii)** dominios PHD y **iv)** dominios WD40; por lo que las proteínas que contienen estos dominios son reclutadas de manera específica por las lisinas metiladas. Particularmente, las lisinas pueden presentar distinto grado de metilación, mono-, di- o trimetilación (revisado en Bannister et al., 2002).

Los complejos modificadores que introducen las MPT en las histonas son conocidos como escritores (“writers”), las proteínas que contienen dominios especializados que identifican e interpretan estas modificaciones se conocen como lectoras (“readers”) y las enzimas encargadas eliminar las modificaciones se llaman borradores (“erasers”) (revisado en Biswas y Rao, 2018).

Anteriormente, se pensaba que la metilación de histonas constituía un mecanismo de modificación bioquímica estable e irreversible ya que la vida media de las histonas canónicas



de las cuales se enrollan ~ 147 pb de DNA. Las regiones amino-terminales de las proteínas histonas sobresalen del núcleo octamérico y son los sitios de modificaciones covalentes como la acetilación, metilación, fosforilación y ubiquitinación entre otras, (circulo azul). (A) La ganancia de modificaciones covalentes es catalizada por enzimas modificadoras de histonas, “writers”. (B) Los “readers” reconocen modificaciones específicas y, al hacerlo, ayudan al ensamblaje de complejos de remodelación de cromatina entre otras cosas, y (C) los “erasers” catalizan la eliminación de modificaciones covalentes.

6.3 LAS DESMETILASAS DE HISTONAS.

Paik y colaboradores (1964), reportaron una enzima que podía desmetilar lisinas mono- y dimetiladas (Kim et al., 1964). Años más tarde el mismo grupo de investigación purificó parcialmente una proteína que podía desmetilar residuos de histonas (Paik y Kim, 1973, 1974). No obstante, la identidad molecular de esta enzima permaneció poco clara durante varias décadas.

Fue hasta el 2004 cuando el grupo de Yang Shi identificó y caracterizó la primera desmetilasa de histonas, LSD1 (“Lysine Specific Demethylase 1”) (Shi et al. 2004), después se renombró como KDM1A (“Lysine Demethylase 1A”) (revisado en Allis et al., 2007). Esta enzima remueve el grupo metilo de las lisinas cuatro y nueve de la histona tres (H3K4me2, H3K4me1, H3K9me2 y H3K9me1) mediante una reacción de oxidación, lo cual sugiere que KDM1A es relevante en el dinamismo de las MPT y por lo tanto en la transcripción génica (Metzger et al., 2005; Shi et al., 2004).

Actualmente, la desmetilación de las lisinas que se encuentran en las histonas es llevada a cabo por dos familias de enzimas, las amino oxidasas y las oxigenasas, los dominios proteicos de cada familia se enlistan en la Tabla 1. La primera familia realiza la desmetilación de residuos mono y dimetilados, utilizando FAD como aceptor de electrones (Anand and Marmorstein, 2007; Shi et al., 2004). La segunda familia puede desmetilar residuos mono, di y trimetilados, donde las oxigenasas utilizan como co-sustratos O₂ y α- cetoglutarato (Anand and Marmorstein, 2007; Whetstine et al., 2006).

Tabla 1: Estructura proteica de las desmetilasas de histonas.

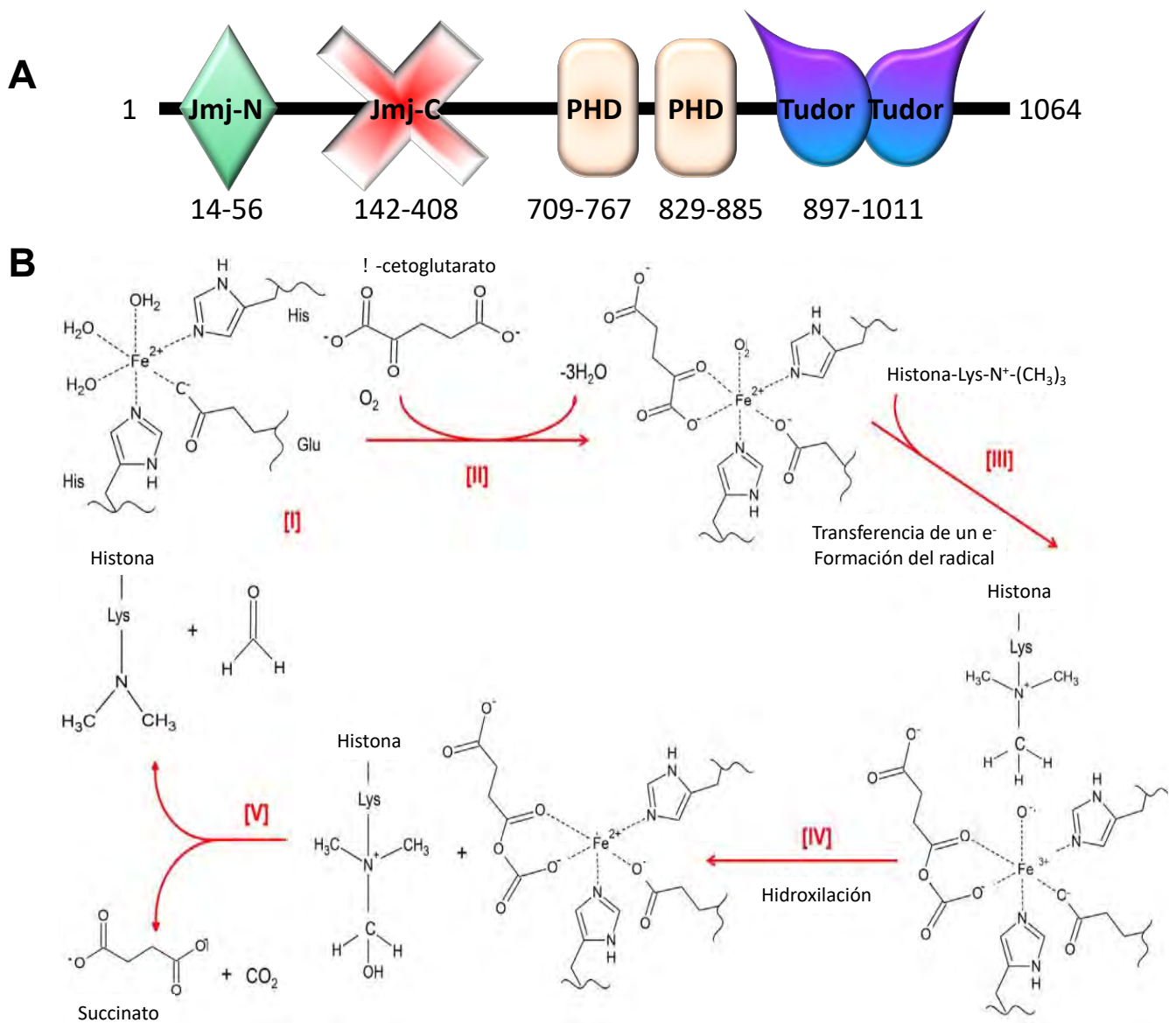
Representación esquemática del las distintas desmetilasas de histonas conocidas hasta la fecha agrupadas por familia y subfamilia, se enfatiza en los dominios proteicos que las conforman. AOL-N, región amino terminal que contiene el dominio amino-oxidasas-like; AOL-C, región carboxilo terminal que contiene el dominio amino-oxidasas-like; N-term, región amino terminal; SWIRM, SWI3p, Rsc8p, and Moira; ARID, AT-Rich Interaction Domain; C5HC2-ZF, dedo de Zinc de tipo C5HC2; CXXC-ZF, CXXC, dedo de Zinc de tipo CXXC; FBox, dominio con una caja F; JmjC, Jumonji C Domain; JmjN, Jumonji N Domain; LRR, Leucine-Rich Repeat; TPR, Tetratricopeptide Repeat; Tudor, dominio tipo Tudor; PHD, Plant Homeodomain.

6.4 ESTRUCTURA PROTEICA Y MECANISMO DE DESMETILACIÓN DE LA DESMETILASA DE HISTONAS KDM4A.

La desmetilasa de histonas KDM4A, pertenece a la familia de las oxigenasas y a la subfamilia KDM4 (anteriormente JMJD2); fue la primera enzima reportada que desmetila residuos trimetilados. KDM4A contiene un dominio Jmj-C y un dominio Jmj-N, que constituyen su sitio catalítico compuesto, así como dos dominios PHD y dos dominios híbridos tipo TUDOR (Figura 2A) (Couture et al., 2007; Patel y Wang, 2013). La función de los dominios PHD de KDM4A aún no se ha elucidado (revisado en Berry y Janknecht, 2013), en contraste con los dominios tipo PHD que están contenidos en otras proteínas. Los dominios híbridos tipo TUDOR forman una estructura bilobulada, donde cada lóbulo asemeja un dominio TUDOR canónico. Estos dominios se estructuran por el intercambio de las cadenas $\beta 3$ y $\beta 4$ y debido a esto, el segundo dominio posee un mayor potencial electronegativo comparado con el primero (Chen et al., 2006; Whetstine et al., 2006). Los dominios TUDOR de KDM4A reconocen, se unen pero no desmetilan a H3K4me3 y H4K20me2/3, debido a esto, se propone que KDM4A puede reconocer diversas regiones en la cromatina (Lee et al., 2008). Estudios *in vitro* revelaron que KDM4A desmetila a los residuos H3K9me3, H3K9me2 y H3K36me3 (Klose et al., 2006), sin embargo *in vivo* KDM4A no puede desmetilar residuos dimetilados y monometilados y presenta mayor afinidad hacia H3K9me3 que hacia H3K36me3 (Whetstine et al., 2006). La especificidad de KDM4A es la capacidad de diferenciar sustratos que tienen características semejantes. La capacidad de KDM4A de reconocer dos regiones con secuencias distintas de aminoácidos, se explica debido a que los residuos N-terminales de H3K9me3 y H3K36me3 comparten una conformación estereoquímica similar y ambos péptidos se unen en la misma dirección dentro del sitio de unión al sustrato en la desmetilasa, permitiendo que la trimetil-lisina se deposite en el sitio catalítico, el cual contiene un átomo de Fe que es fundamental para la actividad catalítica de la enzima (Klose et al., 2006; Yamane et al., 2006).

El mecanismo de reacción propuesto es muy similar al que llevan a cabo las ferrohidroxilasas dependientes de α -cetoglutarato (Hausinger, 2004), el cual consta de cinco pasos que a continuación se describen (Figura 2B): El átomo de Fe activo, se encuentra en un estado de oxidación +dos y está coordinado por dos residuos de histidina, uno de glutamato y tres moléculas de agua [I]. Inicialmente, el α -cetoglutarato y el oxígeno son coordinados en el ion Fe^{2+} , lo que conlleva el desplazamiento de las moléculas de agua [II], después ocurre la transferencia de un solo electrón desde el Fe^{2+} hacia el oxígeno, generándose un radical peróxido. Este radical ataca al α -cetoglutarato y genera un anhídrido mixto que está unido al

Fe^{3+} -radical hidroxilo [III]; entonces el radical hidroxilo, altamente reactivo, es capaz de activar el enlace C-H del grupo metilo que se encuentra en la trimetil-lisina mediante la sustracción de un protón y la transferencia del grupo OH^- al átomo de carbono del grupo metilo formándose una hidroximetil-lisina [IV]. La reacción de desmetilación procede con la pérdida espontánea de formaldehído de la hidroximetil-lisina, debido a que el grupo carbonilo es un buen grupo saliente [V]. Dado que la transferencia de OH^- deja una brecha en la esfera de coordinación del Fe^{2+} , se provoca la ruptura del anhídrido mixto que se disocia en succinato y CO_2 como subproductos, y la unión de tres moléculas de agua al Fe^{2+} regenera el sitio catalítico original (Figura 2B) (Hausinger, 2004).



142-308 a.a, dos dominios PHD (en café) los cuales abarcan del 709 al 767 a.a y del 829 al 885 a.a, y dos dominios híbridos tipo TUDOR (en morado), los cuales ocupan los 897-1011 a.a respectivamente. (B) El mecanismo de reacción propuesto es muy similar al que llevan a cabo las ferro-hidroxilasas dependientes de α -cetoglutarato, se puede dividir en cinco pasos, primeramente el átomo de Fe se encuentra coordinado por dos residuos de histidina, uno de glutamina y tres moléculas de agua, posteriormente se activa el Fe mediante la coordinación del α -cetoglutarato y oxígeno molecular (II), a continuación se transfiere un electrón del Fe^{2+} al oxígeno molecular y se forma un radical (III), después, la trimetil-lisina es atacada por este radical lo cual, resulta en la hidroxilación de un grupo metilo (IV), después, se pierde espontáneamente el grupo carbonilo (V), finalmente se regenera el sitio catalítico de la enzima mediante la coordinación de tres moléculas de agua al ion Fe^{2+} (Hausinger *et al.* 2004). Las estructuras fueron elaboradas con el programa ChemSketch.Ink de ACDLABS FREEWARE (revisado en Guerra-Calderas *et al.*, 2015).

Debido a la actividad catalítica de KDM4A, su estructura y capacidad de reconocer distintas modificaciones epigenéticas en la cromatina, se han asociado diversas funciones a esta proteína, algunas de las cuales se describen a continuación.

6.5 FUNCIONES DE LA DESMETILASA DE HISTONAS KDM4A.

En *Caenorhabditis elegans* (*C. elegans*), KDM4A se involucra en la disminución de H3K36me3 en el cromosoma X (revisado en Guerra-Calderas *et al.*, 2015). KDM4A también está implicada en la represión génica de *ASCL2* (“*Achaete-Scute Complex Homolog 2*”), ya que actúa como co-factor del co-represor N-CoR, para lo cual requiere de su actividad de desmetilasa (revisado en Klose y Zhang, 2007; Zhang *et al.*, 2005). Así mismo, KDM4A también se asocia con desacetilasas de histonas y la proteína de retinoblastoma, pRb, para reprimir los promotores regulados por E2F, pero en este caso el papel de KDM4A como desmetilasa no se ha estudiado (Gray *et al.*, 2005). Durante el desarrollo, se observó que KDM4A coactiva a *NANOG* en células troncales (Loh *et al.*, 2007) y la isoforma Δ N-KDM4A, la cual carece de la región amino terminal, lleva a cabo la desmetilación de H3K9me3 del promotor de *Myog*, durante la diferenciación del músculo esquelético de mioblastos a miotubos (Verrier *et al.*, 2011), estos resultados proponen que KDM4A junto con sus isoformas pueden tener una función importante en la regulación de la expresión de distintos genes.

La sobreexpresión de *KDM4A* en células humanas, se asocia con un relajamiento de la estructura de la cromatina y favorece un paso más rápido por la fase S. Por el contrario, mutaciones en el gen que codifica para la proteína ortóloga en *C. elegans* conduce al aumento en el tiempo de replicación y del daño al DNA e induce la muerte celular programada dependiente de p53 (Black *et al.*, 2010; Whetstone *et al.*, 2006). Los niveles de KDM4A son dependientes del ciclo celular, además reportes indican que KDM4A antagoniza con las funciones de HP1 γ (“Heterocromatin Protein 1 gamma”). Estos resultados sugieren un modelo en el

KDM4A regula la replicación del DNA ya que controla la accesibilidad de la cromatina (Black et al., 2010).

KDM4A tiene un papel relevante durante el desarrollo, sin embargo múltiples estudios reportan que la expresión aberrante de *KDM4A* se relaciona con insuficiencia cardíaca, hipertrofia cardíaca, la progresión de infecciones virales, así como con trastornos como la alopecia areata y con el desarrollo de diversos tipos de cáncer (revisado en Labbé et al., 2013).

6.6 EL PAPEL DE KDM4A EN EL DESARROLLO DE CÁNCER.

El patrón de expresión de *KDM4A* está alterado en múltiples tipos de cáncer los cuales se resumen en la Tabla 2. Particularmente profundizaremos en el papel de KDM4A en cáncer de pulmón y de mama.

Tabla 2: Participación de la desmetilasa de histonas KDM4A en diversos tipos de cáncer.

Las flechas hacia arriba significan que hay un aumento en los niveles de esta enzima a nivel proteico. Las flechas hacia abajo significan que hay una disminución de los niveles de esta desmetilasa a nivel proteína.

Tipo de cáncer	Abundancia de KDM4A	Parámetro clínico asociado	Referencias
Próstata	↑	Promueve el inicio del cáncer de próstata.	(Kim et al., 2012a, 2016; Li et al., 2018; Shin y Janknecht, 2007).
Colorrectal	↑	Promueve la proliferación de las células de cáncer de colon.	(Kim et al., 2012b).
Carcinoma de células escamosas de cabeza y cuello	↑	Poca supervivencia global.	(Ding et al., 2013; Jin et al., 2017).
Vejiga	↑	Participa en etapas tempranas de la carcinogénesis del cáncer de vejiga.	(Kogure et al., 2013)
Vejiga	↓	La pérdida de KDM4A correlaciona con pobre pronóstico.	(Kauffman et al., 2011)
Pulmón	↑	La sobreexpresión de <i>KDM4A</i> en NSCLC se considera un evento temprano en la carcinogénesis de pulmón.	(Kogure et al., 2013; Mallette y Richard, 2012)
Mama	↑	Progresión tumoral.	(Berry et al., 2012; Li et al., 2014; Metzger et al., 2017)

6.6.1 Cáncer de pulmón.

Dos reportes independientes observaron la presencia nuclear de KDM4A en carcinomas pulmonares y carcinomas de células no pequeñas NSCLC (“Non- Small Cell Lung Carcinoma”), mientras que no se detectó KDM4A en tejido pulmonar histológicamente normal. En estos trabajos no se estableció la relación entre la abundancia de KDM4A y el pronóstico de los pacientes (Malette y Richard 2012; Kogure et al. 2013). El grupo de Hamamoto, no encontró diferencias significativas de la expresión de KDM4A en tumores de distinto grado, lo cual puede sugerir que la sobreexpresión de *KDM4A* en NSCLC pudiera considerarse un evento temprano en la carcinogénesis de este tipo de tumores (Kogure et al. 2013).

Malette y colaboradores (2012) observaron que *KDM4A* se sobreexpresa en líneas celulares neoplásicas de pulmón de modelos murinos y también células derivadas de humano. Entre las células que sobreexpresan a *KDM4A* está la línea celular A549 y cuando en estas células se abate a *KDM4A*, se induce la senescencia (Malette y Richard 2012). Por medio de ensayos de inmunoprecipitación de la cromatina y microarreglos (técnica conocida como ChIP-on-ChIP), se identificó que *CHD5* es un blanco génico de KDM4A (Malette y Richard, 2012). CHD5 se involucra en la ruta de ubiquitinación de p53. En este sentido, se ha reportado que CHD5 regula de manera positiva a p19^{Arf}, el cuál codifica para una proteína que secuestra a Mdm2 (Murine doble minute 2), la función de Mdm2 es ubiquitinar a p53 para su posterior degradación. Por lo que, la deficiencia de p19^{Arf} y de CHD5 compromete a p53 y sus funciones cómo supresor de tumores e inductor de senescencia (Bagchi et al., 2007; Serrano et al., 1997). Se observó que KDM4A se localiza en la región situada a +741 pb con respecto al TSS *CHD5*. El aumento de los niveles de KDM4A conduce a una reducción del transcrito y la proteína de CHD5, por el contrario, el abatimiento de KDM4A, aumenta los niveles de CHD5, Los resultados obtenidos por este grupo de investigación proponen que el aumento de los niveles de KDM4A en cáncer pulmonar puede cooperar con otros factores en promover la transformación celular, bloquea la senescencia dependiente de p53 vía *CHD5* (Malette y Richard 2012).

6.6.2 Cáncer de mama.

La sobreexpresión de *KDM4A* en cáncer de mama de tipo lobulillar, ductal y en tumores triple negativo se relaciona con la pérdida de H3K9me3, la cual en condiciones normales se encuentra enriquecida en la región pericentromérica. Este fenómeno contribuye al desarrollo de aneuploidías e inestabilidad cromosómica en tumores sólidos y, por tanto, a la progresión tumoral (Slee et al. 2012).

Se observaron diferencias significativas en la abundancia de KDM4A en tejido neoplásico comparado con tejido no neoplásico de mama, donde estas diferencias se asocian con parámetros patológicos y clínicos (Patani *et al.* 2011).

El abatimiento de KDM4A mediante la transfección de siRNAs en líneas celulares derivadas de cáncer de mama, reduce la proliferación, invasión y migración (Li *et al.* 2011; 2012). No obstante, el papel de KDM4A en el desarrollo de cáncer de mama, permanece poco claro. Un estudio encontró que KDM4A funciona como un co-activador del ER (“Estrogen Receptor”) debido a que forman el complejo, KDM4A-ER. Este estudio encontró que la sobreexpresión de *KDM4A* deriva en un incremento de la transcripción dependiente de estrógenos, mientras que su abatimiento, disminuye la transcripción de los genes blanco del ER (Berry *et al.* 2012). La disminución de KDM4A redujo la expresión del gen que codifica para la ciclina D1, *CCND1*, gen que se sobreexpresa en tumores mamarios y cuya transcripción se induce por estradiol en células de cáncer de mama cuya transcripción se induce por estradiol en células de cáncer de mama (Altucci *et al.* 1996; Cicatiello *et al.* 2004). Por otra parte, la expresión de *JUN*, la cual es independiente de señalización hormonal, también se ve disminuida después del abatimiento de KDM4A. *JUN* tiene funciones importantes en tejidos cancerosos, su sobreexpresión estimula la invasión, la migración y la formación de tumores y su inactivación causa la interrupción del ciclo celular (Berry *et al.* 2012). En conjunto, estos datos sugieren que KDM4A puede regular el crecimiento y proliferación de las células de cáncer de mama al promover la expresión de al menos dos oncogenes, *JUN* y *CCND1* (Berry *et al.* 2012). Otro reporte encontró correlación entre la expresión de *KDM4A* y los niveles de *ARHI* (“*Aplasia Ras Homolog member 1*”) y *ER*, tanto a nivel mRNA como proteína. Así mismo, la abundancia de KDM4A fue mayor en muestras de carcinoma ductal infiltrante comparadas con muestras de fibroadenoma. Los resultados proponen a KDM4A como un posible blanco diagnóstico y terapéutico en el cáncer de mama humano (Li *et al.*, 2013). Aunado a este estudio se suma otro llevado a cabo en el 2014, el cual encontró que *KDM4A* se encuentra sobreexpresado en cáncer de mama y se correlacionó positivamente con la progresión del tumor. El abatimiento de *KDM4A* incrementó la expresión de *ARHI*, mientras que la sobreexpresión de *KDM4A* disminuyó la expresión de *ARHI* tanto a nivel proteína como mRNA. Además, los factores de transcripción E2F y las desacetilasas histonas están involucradas en la represión transcripcional de *ARHI* por KDM4A. También se encontró que el comportamiento agresivo dependiente de *KDM4A* en el cáncer de mama pudo revertirse mediante la re-expresión de *ARHI in vitro* e *in vivo* (Li *et al.*, 2014).

KDM4A coactiva genes dependientes e independientes de la señalización hormonal y regula el crecimiento celular ya que controla la expresión de al menos dos oncogenes *CCND1* y *C-JUN*. La participación de KDM4A en el cáncer de mama puede ser más compleja de lo que

originalmente se creía. La literatura sugiere que KDM4A puede participar en el diagnóstico del cáncer y como un blanco terapéutico.

6.7 EL GEN *CHD5*.

CHD5 está localizado en el cromosoma 1 en el locus *1p36.31*, codifica para una enzima, CHD5 (“Chromodomain Helicase DNA-binding protein 5”) que pertenece a la familia de las helicasas (Thompson et al., 2003). CHD5 está conformada por dos dominios PHD, dos cromodominios, un dominio con actividad de helicasa y un dominio de unión a DNA, el cual se encuentra conservado dicha familia de enzimas (Thompson et al., 2003). CHD5 es el quinto miembro de una familia de proteínas remodeladoras de la cromatina. Algunas evidencias sugieren que CHD5 puede formar un complejo con NuRD (“Nucleosome Remodeling and Deacetylation”), el cual regula la transcripción de genes específicos como *WEE1* (“*Wee1-Like Protein Kinase*”), en células derivadas de neuroblastoma. *CHD5* se expresa en el sistema nervioso, principalmente en cerebro y cerebelo, así como en los testículos. *CHD5* codifica para una proteína que funciona como supresora de tumores, ya que controla la proliferación, apoptosis y senescencia celular por la vía de $p19^{Arf}/p53$ (Bagchi et al., 2007; Serrano et al., 1997), CHD5 regula positivamente a $p19^{Arf}$, el cual codifica para una proteína que secuestra a MDM2 (“Murine Doble Minute 2”), cuya función es ubiquitinar a p53 para su posterior degradación. Por esta razón la deficiencia de $p19^{Arf}$ y CHD5 compromete a p53 y sus funciones como supresor de tumores e inductor de senescencia (Bagchi et al., 2007; Serrano et al., 1997).

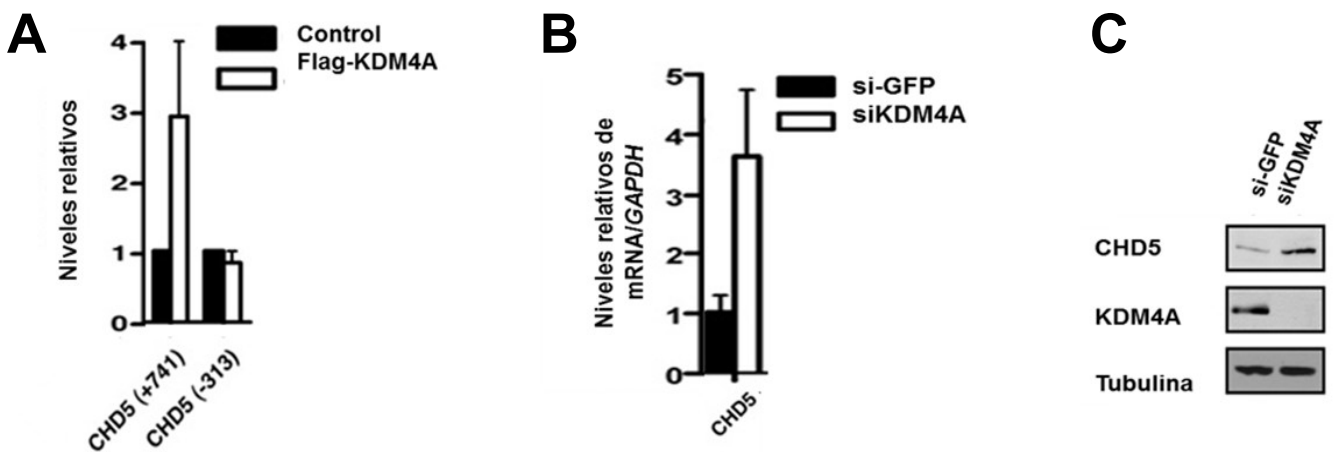
El promotor de *CHD5* se encuentra embebido en una isla CpG de alta densidad (HCP, “High CpG Promoter”), la cual contiene aproximadamente 180 CpGs. Los promotores inmersos en HCP se inactivan a través de la hipermetilación del DNA y el enriquecimiento de las modificaciones postraduccionales asociadas con represión como H3K9me3 y H3K27me3. Tal es el caso del *miR-125b1* (Soto-Reyes et al., 2012), mientras que los promotores embebidos en islas de baja densidad CpG (LCP, “Low CpG Promoter”) no son inactivados por la hipermetilación del DNA sino por las modificaciones postraduccionales de las histonas (Weber et al., 2007).

CHD5 se inactiva en varios tipos de cáncer, entre ellos el cáncer de mama tanto ductal, lobulillar y triple negativo (Mulero-Navarro y Esteller, 2008). Hasta ahora se han identificado dos mecanismos de inactivación de *CHD5*: el primero consiste en un proceso genético basado en la delección del gen en neuroblastomas (Bello et al., 1995; Law et al., 2005) y el segundo mecanismo consiste en la hipermetilación del DNA del promotor, la cual se presenta en gliomas y cáncer colorrectal (Fatemi et al., 2014; Mulero-Navarro y Esteller, 2008). En el cáncer de

mama, la hipermetilación del DNA del promotor de esta helicasa es específica de cierto tipo de tumores (Mulero-Navarro y Esteller, 2008).

La menor abundancia de CHD5 en tumores mamarios en comparación con tejido mamario histológicamente normal, se relaciona con el desarrollo y progresión del cáncer de mama (Mulero-Navarro y Esteller, 2008). Este fenómeno también se observa en las líneas celulares derivadas de cáncer de mama, MDA-MB-231 y MCF7, las cuales no presentan la delección del gen y solo se detectó la hipermetilación del promotor en células MDA-MB-231. Sin embargo, cuando estas células se tratan con 5-aza-2-deoxicitidina, fármaco que inhibe la metilación del DNA, no se observa la reactivación completa de la expresión de *CHD5* (Mulero-Navarro y Esteller, 2008), lo que sugiere que la presencia de otros mecanismos epigenéticos de inactivación de *CHD5* que no dependen exclusivamente del inicio de la transcripción.

Por medio de ensayos de inmunoprecipitación de la cromatina y microarreglos (técnica conocida como ChIP-on-ChIP), se identificó que *CHD5* es un blanco de KDM4A (Mallette y Richard, 2012). En este estudio se observó que KDM4A se localiza en la región situada a +741 pb con respecto al TSS, región que se encuentra dentro del primer intrón de este gen (Figura 3A) (Mallette y Richard, 2012). El aumento de KDM4A conduce a una reducción del transcrito y la proteína de CHD5, mientras que, el abatimiento de KDM4A, aumenta los niveles de CHD5 (Figura 3A, 3B y 3C) y desencadena la senescencia en la línea celular de cáncer de pulmón A549 (Fig. 3B, 3C, 3D, 3E) (Mallette y Richard, 2012).



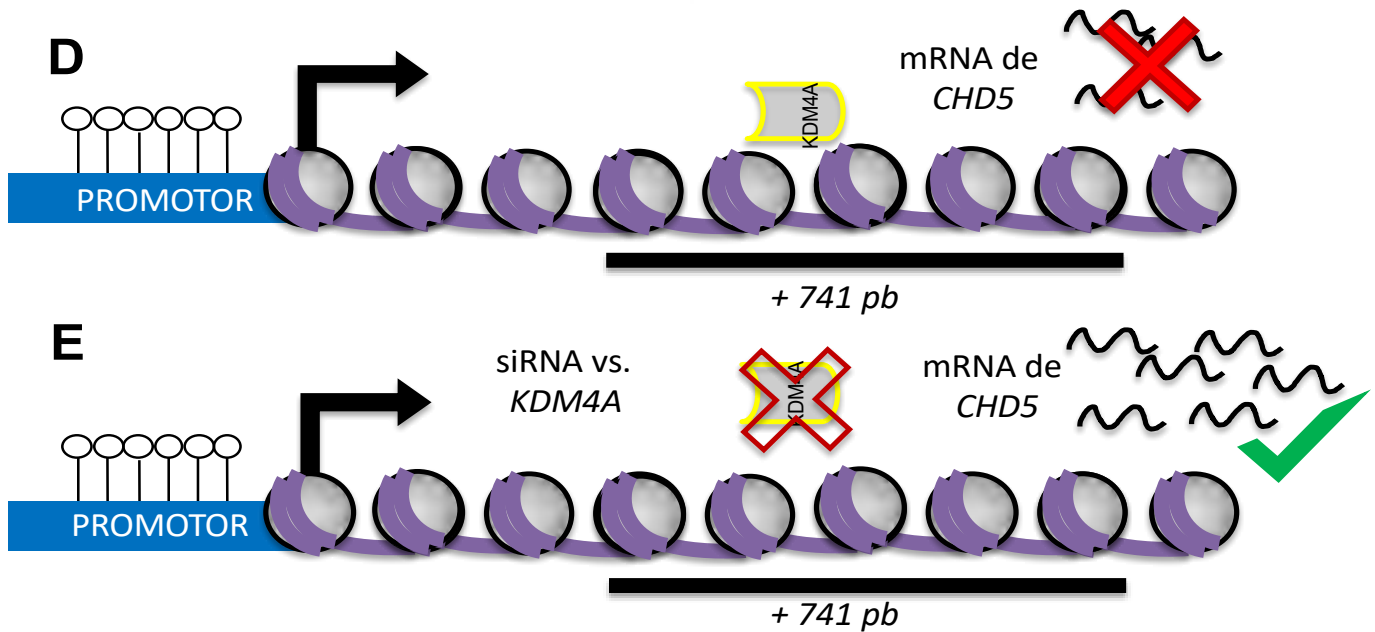


Figura 3: Papel de KDM4A en la regulación de *CHD5*.

(A) Inmunoprecipitación de la cromatina evaluada por qPCR cuantitativo (qPCR) donde se observa que KDM4A se localiza a la región +741 de *CHD5*. (B) Evaluación de la expresión de *CHD5* por qPCR en la cual se demuestra que el abatimiento de KDM4A conlleva a un aumento de la expresión de este gen. (C) Evaluación de la abundancia de *CHD5* y KDM4A por análisis de tipo Western se muestra la presencia de *CHD5* y KDM4A, post-transfección de si-GFP y siKDM4A (modificado de Mallette y Richard, 2012). (D) Esquema de la unión y efecto de KDM4A sobre *CHD5*. KDM4A tiene un efecto negativo en la transcripción de *CHD5*, y (E) Esquema de la unión y efecto de KDM4A sobre *CHD5*, después del abatimiento de KDM4A mediante un RNA de interferencia, se observa que la expresión de *CHD5* aumenta.

Estos hallazgos sugieren que la regulación negativa de *CHD5* está dada por mecanismos diferentes a la metilación del DNA en el promotor y a la delección del gen.

6.8 PARTICIPACIÓN DE H3K36me3 EN LA REGULACIÓN DE LA TRANSCRIPCIÓN.

Un mecanismo alternativo de regulación negativa de *CHD5*, diferente a la metilación del DNA en el promotor, puede estar mediado por la pérdida de la marca histónica H3K36me3, la cual, se establece por la enzima Set2. Esta modificación postraduccional es importante ya que se encuentra asociada con el proceso de elongación de la transcripción por RNA polimerasa II (revisado en Li et al., 2007; Vavouri y Lehner, 2012). La presencia de H3K36me3 en el cuerpo de los genes correlaciona con el aumento en la procesividad y los niveles de transcripción de la RNA polimerasa II. Por el contrario, cuando la H3K36me2 está presente en el cuerpo de los genes se relaciona con la disminución de procesividad de la RNA polimerasa II y de los niveles de expresión (Fong et al., 2017). Tal es el caso de *GFAP* (“*Glial Fibrillary Acid Protein*”)

(Cascante et al., 2014). Por otra parte, se reportó que en pez cebra la marca H3K36me3 esta enriquecida en el extremo 3' de los genes activamente transcritos, aunque, también se encontró esta modificación en los promotores de genes silenciados que regulan el proceso de espermatogénesis (Wu et al., 2011). Otros estudios realizados en levadura mostraron que Set2, enzima responsable de las tres formas de metilación de H3K36, se encuentra acoplada al proceso de elongación de la transcripción. Set2 se asocia la RNA polimerasa II (revisado en Li et al., 2007; Vavouri y Lehner, 2012). Esta interacción, se encuentra mediada por los residuos fosforilados, en específico la serina 2 del CTD ("Carboxy-Terminal Domain") de la subunidad grande de la RNA polimerasa II, RBP1, la cual es característica del proceso de elongación de la transcripción por la RNA polimerasa II. Se ha sugerido que la proteína SET2 humana también se pueden unir a la RNA polimerasa II y metilar a H3K36 (revisado en Wagner y Carpenter, 2012).

Estudios recientes mostraron que KDM4A desmetila a H3K36me3 y provoca la pérdida de reclutamiento de la RNA polimerasa II en regiones exónicas de *GFAP*. Al controlar la elongación de la transcripción por la RNA polimerasa II, KDM4A regula la diferenciación de los astrocitos (Cascante et al., 2014). El nivel de metilación de H3K36 puede regular el inicio y la elongación de la transcripción de *LBP* ("*Lipopolysaccharide-Binding Protein*") que es un regulador negativo de la diferenciación de adipocitos y ayuda a la diferenciación a osteoblastos. Estos hallazgos sugieren que la H3K36me3 mediada por SETD2 podría regular el destino celular de las células troncales estromales mesenquimales *in vitro* e *in vivo* (Wang et al., 2018).

Con base en lo anterior, el estudio de la regulación de las desmetilasas de histonas ha cobrado suma importancia para el entendimiento del control transcripcional y epigenético en condiciones de normalidad y eventos neoplásicos.

6.9 REGULACIÓN DE LAS DESMETILASAS DE HISTONAS.

La demostración de la actividad de las desmetilasas de histonas *in vitro* sugiere que la simple asociación de las enzimas activas con su sustrato es suficiente para que la reacción de desmetilación se lleve a cabo, por esta razón, la función de estas enzimas se debe controlar por mecanismos regulatorios *in vivo* con el fin de prevenir procesos aberrantes de desmetilación (revisado en Lan et al., 2008).

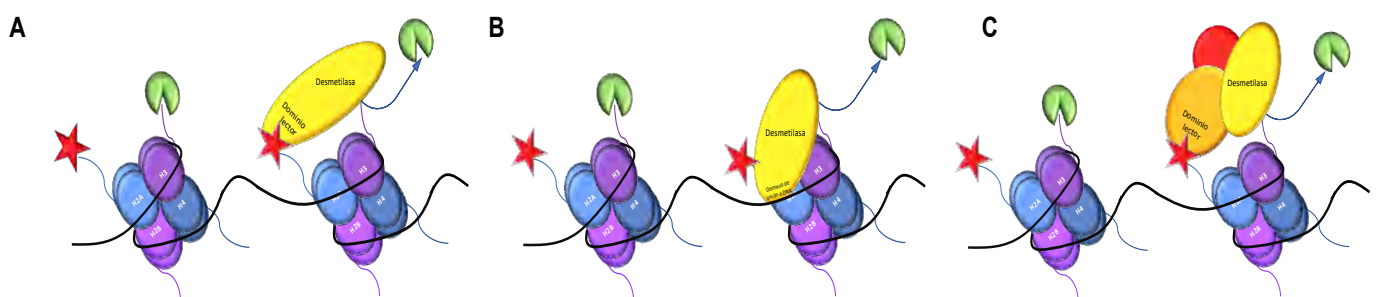
La regulación de la expresión de los genes que codifican para las desmetilasas es importante para modular su actividad biológica. Los patrones de expresión de estas enzimas son específicos de tejido (revisado en Lan et al., 2008). Otra manera de regular a las desmetilasas, es mediante el control de su actividad catalítica.

Shi y Lan (2008) plantean que la actividad de las desmetilasas se controla por sus modificaciones postraduccionales, las interacciones con factores de unión a DNA y con otras enzimas remodeladoras de la cromatina y el reconocimiento del ambiente cromatínico, por ejemplo, modificaciones específicas en las histonas pueden ser distinguidas por dominios contenidos en las desmetilasas (revisado en Lan et al., 2008). También, los microRNAs pueden regular a las desmetilasas. Por ejemplo un estudio encontró por medio de ensayos de luciferasa, que *KDM4A* es un gen blanco del miR-10a y demostró mediante ensayos de PCR cuantitativa y Western blot que los niveles de KDM4A disminuyen en respuesta al aumento del miR-10a en células derivadas de cáncer de próstata (Mu et al., 2018).

Evidencia reciente sugiere que los niveles cambiantes de los metabolitos pueden regular la actividad enzimática de las proteínas modificadoras de la cromatina, lo que muestra una interrelación entre el metabolismo celular y la epigenética (revisado en Tran et al., 2017).

Muchas enzimas modificadoras de la cromatina requieren un cofactor específico que se encuentra involucrado en el metabolismo, por ejemplo, las desmetilasas de histonas requieren O_2 , FAD y α -cetoglutarato, el cual es un intermediario del ciclo de los ácidos tricarbóxicos (TCA). La disponibilidad de un metabolito específico puede oscilar en respuesta a afectaciones en el metabolismo causadas por variaciones genéticas o como resultado del dinamismo del microambiente. El α -cetoglutarato es crucial para la desmetilación de histonas y se sabe que altos niveles de succinato o fumarato dentro de la célula antagonizan la actividad de las desmetilasas de histonas. La modulación de los niveles de metabolitos proporciona un nivel adicional de control para estas enzimas (revisado en Tran et al., 2017).

Se ha sugerido que los mecanismos de reclutamiento de las desmetilasas tienen un papel crítico en la regulación de estas enzimas, no obstante, aún no se entiende claramente cómo es que las desmetilasas son dirigidas hacia sitios específicos en la cromatina para evitar la desmetilación aberrante (revisado en Lan et al., 2008). Se han propuesto varios modelos de reclutamiento los cuales se describen en la Figura 4.



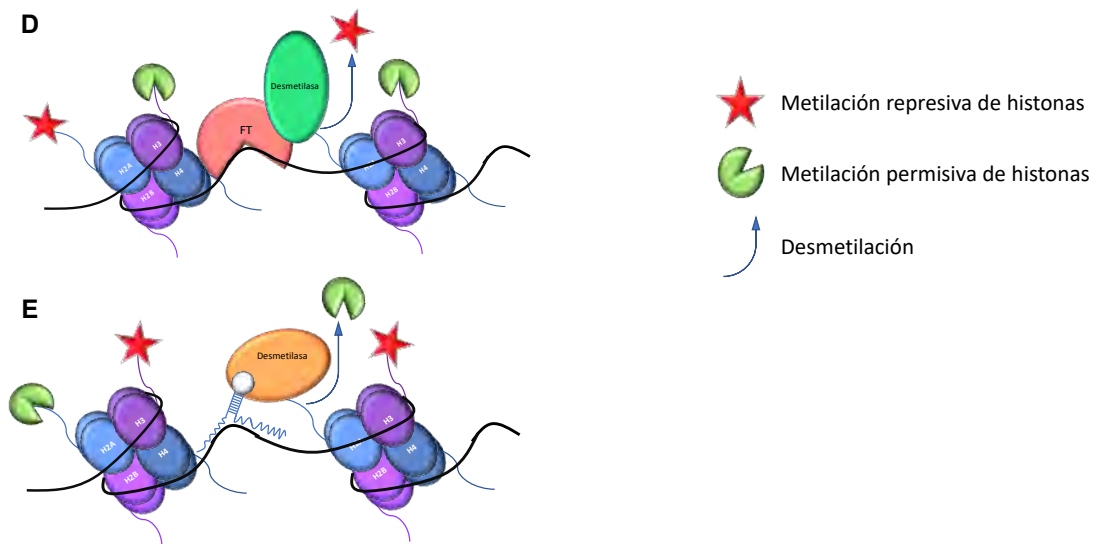


Figura 4: Modelos de reclutamiento de las desmetilasas de histonas.

(A) Algunas desmetilasas de histonas poseen dominios lectores los cuales se unen y leen las modificaciones de histonas que se encuentran ampliamente en todo el genoma. Estas interacciones funcionan para dirigir a las desmetilasas de histonas a la cromatina y regular su actividad. (B) Algunas desmetilasas de histonas interaccionan con la cromatina a través de la unión directa al DNA, tal es el caso de la familia KDM2 que contiene un dominio CXXC. Esta familia es capaz de unirse a las islas CpG. (C) Sin embargo, generalmente las desmetilasas se encuentran en grandes complejos multiproteicos los cuales contienen otras proteínas que se unen a la cromatina. (D) Las desmetilasas están dirigidas a sitios específicos en el genoma a través de la interacción con factores de transcripción (FT) o mediante RNAs no codificantes (ncRNAs) (E).

Se ha propuesto que los factores transcripcionales pueden tener una función importante en mediar el reclutamiento de distintos componentes epigenéticos como las desmetilasas de histonas. Uno de los factores que se ha comenzado a estudiar es el factor nuclear CTCF (“CCCTC-binding factor”) que puede interaccionar con la desmetilasa de histonas KDM5B y aumentar su actividad en líneas celulares derivadas de cáncer de mama (Yamamoto et al., 2014).

6.10 EL FACTOR NUCLEAR CTCF.

CTCF, es una proteína multifuncional con 11 dedos de zinc (ZF) que está conservada en diversos organismos (Lobanenkov et al., 1990). Consiste en 727 aminoácidos y contiene un dominio de unión al DNA que está compuesto por 11 ZF, flanqueados por 10 aminoácidos en la región carboxilo-terminal y 267 aminoácidos en la región amino-terminal (Vostrov et al., 2002). CTCF se caracterizó inicialmente como un regulador transcripcional, conocido como NeP1 (“Negative Protein 1”) (Burcin et al., 1997).

Este factor se une a un total de 77,811 sitios distintos en todos los linajes celulares (Wang et al., 2012), el genoma humano posee entre 15,000 y 40,000 sitios de unión a CTCF según el tipo de célula, la unión de CTCF se inhibe por la metilación del DNA (Bell y Felsenfeld,

2000; Soto-Reyes et al., 2012). La capacidad de CTCF de unirse a varias secuencias se debe al uso diferencial de sus ZF. Debido a esta habilidad peculiar, se ha denominado a CTCF como una "proteína multivalente" (Filippova et al., 1996).

6.10.1 FUNCIONES DE CTCF.

CTCF funciona como represor, activador, bloquea el contacto entre los promotores y los potenciadores al controlar su actividad y cuando se une a su secuencia blanco funge como un elemento frontera. También actúa como barrera ya que detiene la propagación de estructuras heterocromáticas (Cuddapah et al., 2009) y define los límites entre el DNA activo y heterocromático. CTCF controla la estructura tridimensional de la cromatina ya que permite la formación de bucles y ancla el DNA a estructuras celulares como la lámina nuclear (Guelen et al., 2008). CTCF media las interacciones cromatínicas a larga distancia (revisado en Phillips y Corces, 2009) y regula los genes improntados *H19* e *IGF2* ("*Insulin-like Growth Factor 2*") (Bell y Felsenfeld, 2000; Hark et al., 2000). CTCF también afecta el corte y empalme del mRNA, al unirse río abajo en los exones alternativos ya que facilita la pausa de la RNA polimerasa II, lo que permite el contexto necesario para la conformación del spliceosoma (Shukla et al., 2011).

CTCF tiene múltiples funciones que parecen no tener conexión. Esta multifuncionalidad puede residir en las interacciones de CTCF con innumerables proteínas.

6.10.2 PROTEÍNAS ASOCIADAS A CTCF.

Actualmente aún no está claro cuales son los factores que determinan qué proteína se une a CTCF para cada ocasión, pero se pueden considerar numerosas posibilidades. En primer lugar, CTCF emplea sus 11 ZF de manera combinatoria (Ohlsson et al., 2001) para reconocer y unirse a una variedad de secuencias de DNA. El uso diferencial de un subgrupo ZF para la unión al DNA podría crear, a partir de los ZF restantes, plataformas específicas para la interacción con múltiples proteínas. Otro mecanismo por el cual se puede controlar la elección de la proteína con la que se une CTCF son las diversas modificaciones postraduccionales de la proteína y / o de CTCF, que podrían usarse en diferentes condiciones celulares. Se ha descrito al menos un ejemplo en el que las modificaciones postraduccionales de CTCF alteran su interacción con una proteína asociada, en este caso, la RNA polimerasa II (Chernukhin et al., 2007).

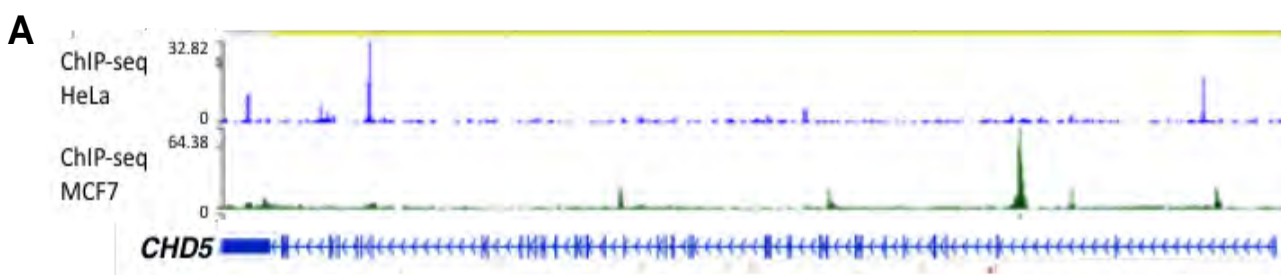
Diversos estudios demostraron que CTCF puede formar homodímeros (Yusufzai et al., 2004), lo que causa que el DNA unido forme bucles (Hou et al., 2008), la formación de los homodímeros depende de los ZF, mientras que la interacción con otras proteínas depende de los ZF restantes y de las regiones amino- y carboxilo-terminales (revisado en Caiafa y

Zlatanova, 2009). CTCF puede interactuar con proteínas que se unen al DNA como YY1, Kaiso, YB-1, CIITA, RFX, y TFAP4, proteínas diversas como Lamina A/C, Rad51, SMAD3, CP190, CENP-E y la RNA polimerasa II, proteínas multifuncionales como PARP1, Nucleofosmina, Topo II, Vigilina, ER y Nucleolina y proteínas asociadas a cromatina como son cohesina, la desacetilasa de histonas SIN3A, las histonas H2A y H2A.Z, la metil-transferasa de histona SUZ12, la helicasa CHD8 y el correpresor nuclear N-CoR (revisado en Zlatanova y Caiafa, 2009).

En células madre embrionarias la unión de la desmetilasa de DNA, TET1, promueve la conversión de 5mC (5-metil-citocina) a 5hmC (5-hidroximetil-citosina) y permite la unión de CTCF. Esto sugiere que existe una posible interacción entre las desmetilasas de DNA y CTCF (Teif et al., 2014). Además, la unión de CTCF a una región heterocromatínica promueve la rápida desmetilación de H3K27me3, lo que propone que CTCF puede reclutar *in vivo* desmetilasas de histonas (Weth et al., 2014). CTCF inmunoprecipita con KDM5B en líneas celulares derivadas de cáncer de mama luminal, y hay múltiples sitios en la cromatina donde KDM5B y CTCF coexisten (Yamamoto et al., 2014).

El 29% de los sitios de unión a CTCF se localizan en intrones, mientras que solo el 8% de esos sitios están en promotores (Wang et al., 2012). Ya que KDM4A está localizada en el primer intrón de *CHD5*. Es probable que CTCF pueda mediar el reclutamiento de KDM4A a la región localizada en el primer intrón de *CHD5*. Además, estudios *in vitro* mostraron que la capacidad de desmetilación de KDM4A se incrementa en presencia de CTCF hasta en un 60% (Jeong et al., 2011), lo cual apoya la hipótesis que *in vivo* pudiera existir alguna interacción entre CTCF y KDM4A. Ya que CTCF puede interactuar con diversos remodeladores de la cromatina, resulta sumamente interesante considerar a CTCF como un posible mediador del reclutamiento de las desmetilasas de histonas, como KDM4A, a sus sitios blanco.

Se emplearon las bases de datos del proyecto ENCODE (“Encyclopedia Of DNA Elements”) y se encontró la presencia de CTCF en al menos cinco sitios a lo largo del cuerpo de *CHD5*, en distintas líneas celulares (Figura 5A).



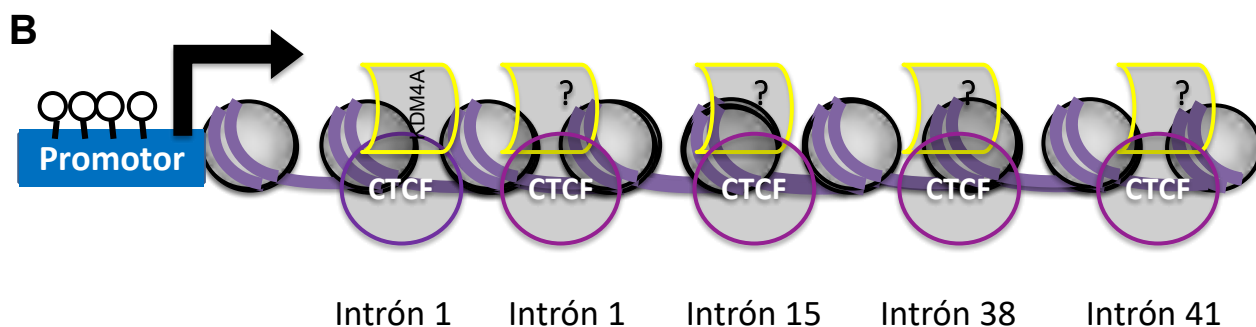


Figura 5: Localización de CTCF a lo largo de *CHD5*.

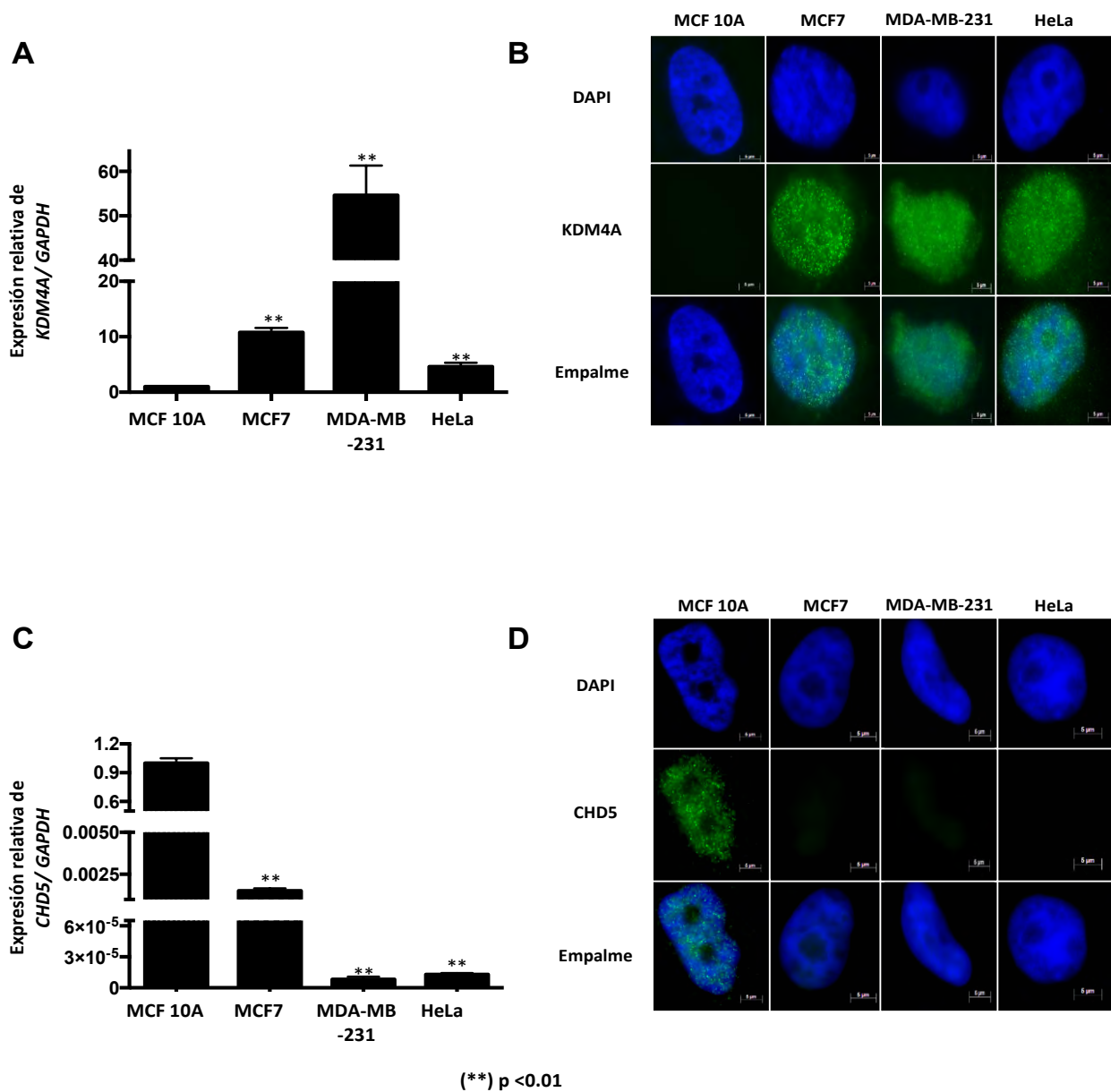
(A) Se emplearon las bases de datos del proyecto ENCODE, se encontró la localización de CTCF en al menos cinco sitios a lo largo del cuerpo de *CHD5*, tanto en la línea celular HeLa con un valor máximo de 32.82 y en la línea celular MCF7 con un valor máximo de 64.38. (B) Localización de CTCF a lo largo de *CHD5*, CTCF se localiza en el intrón 1, el intrón 15, el intrón 38 y el intrón 41 de *CHD5*.

Estos antecedentes sugieren que la unión de KDM4A y CTCF a lo largo del cuerpo de *CHD5*, puede afectar la transcripción de dicho gen, no obstante, aún no se sabe cuál es el mecanismo por el cual KDM4A y CTCF controlan la transcripción del gen *CHD5* (Figura 5B).

7 ANTECEDENTES DIRECTOS.

7.1 *KDM4A* se sobreexpresa en las líneas celulares MCF7, MDA-MB-231 Y HeLa.

Como parte de mis resultados de maestría y con el objetivo de caracterizar los modelos celulares propuestos, se evaluaron los niveles de expresión del RNA mensajero de *KDM4A* y *CHD5* por RT-qPCR, así como la presencia y localización de las proteínas *CHD5* y *KDM4A* en líneas celulares neoplásicas (MCF7, MDA-MB-231 Y HeLa) y una no neoplásica (MCF 10A) por medio de ensayos de inmunofluorescencia (IF). Los resultados mostraron que *KDM4A* se sobreexpresa en las líneas celulares neoplásicas en comparación con una línea celular no neoplásica (Figuras 6A y 6B). En contraste, la expresión de *CHD5* se abate en las líneas celulares neoplásicas, cuando se comparan con MCF 10A. Así mismo, solo detectamos a *CHD5* en el núcleo de la línea celular MCF 10A mientras que no se detectó en las líneas celulares neoplásicas (Figuras 6C y 6D). Estos datos sugieren una relación entre la presencia de *KDM4A* y la disminución de *CHD5*.



=7,
ión

FIGURE 1

de KDM4A en los modelos celulares se evaluaron por medio de microscopía de epifluorescencia con el uso de anticuerpos primarios y posteriormente anticuerpos secundarios acoplados a un fluoróforo. (C) Análisis de expresión de *CHD5* humano por medio de RT-qPCR en las líneas celulares MCF 10A, MCF7, MDA-MB-231 y HeLa. Los datos se normalizaron contra la expresión de *GAPDH*. (D) La presencia y localización de CHD5 en las líneas celulares se evaluaron por medio de microscopía de epifluorescencia (100 aumentos, se usó el microscopio de Epifluorescencia Axiolmager.A2 marca Zeiss). Los resultados muestran la media de tres experimentos independientes. Los datos de expresión se normalizaron con respecto a la línea celular MCF 10A a la cual asignamos un valor relativo de 1. (**) p < 0.01 con respecto a la línea celular MCF 10A, como prueba estadística se utilizó *t*-student.

7.2 Presencia de CTCF, KDM4A y las modificaciones postraduccionales H3K36me3 y H3K36me2 en el primer intrón de *CHD5*.

En 2012, Mallette y cols. demostraron que KDM4A se encuentra en el primer intrón de *CHD5* en la línea celular U2OS por ensayos de inmunoprecipitación de cromatina acoplada a microarreglo (ChIP-on chip) (Mallette y Richard, 2012). Sin embargo, ya que KDM4A no tiene dominio de unión a DNA, aún no está claro qué proteínas puedan reclutarla a esta región. Uno de los candidatos potenciales es CTCF. Con el objetivo de evaluar la localización de KDM4A en el primer intrón de *CHD5* en nuestros modelos celulares, realizamos ensayos de inmunoprecipitación de la cromatina (ChIP) en las líneas celulares MCF 10A, MCF7, MDA-MB-231 y HeLa. KDM4A solo se encontró en esta región en las líneas celulares MCF7, MDA-MB-231 y HeLa y no se detectó en la línea celular no neoplásica MCF 10A (Figura 7A y Figura 7C).

Con el fin determinar si CTCF se localizaba en el primer intrón de *CHD5*, realizamos ensayos de ChIP en las células MCF 10A, MCF7, MDA-MB-231 y HeLa.

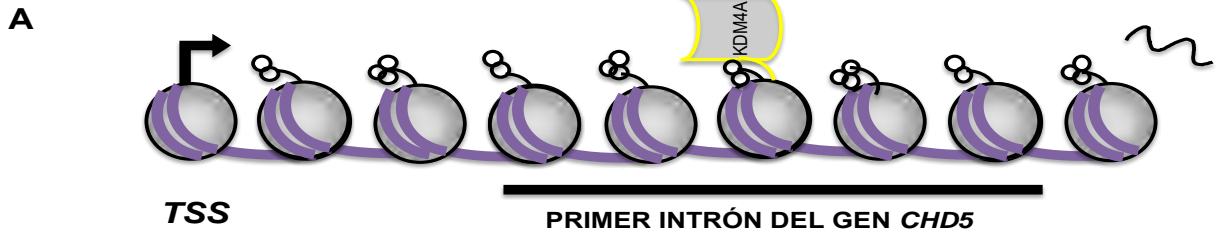
Nuestros resultados muestran que CTCF se encuentra en el primer intrón de *CHD5* en todas las líneas celulares evaluadas (Figura 7B y Figura 7C), incluida la línea celular MCF 10A. Esta es la misma región donde se demostró que KDM4A estaba presente en las líneas celulares neoplásicas (Figura 7A).

La disminución de KDM4A incrementa los niveles de mRNA y proteína de *CHD5* (Mallette y Richard, 2012). Sin embargo, el mecanismo por el cual KDM4A regula negativamente la transcripción de *CHD5* sigue poco claro.

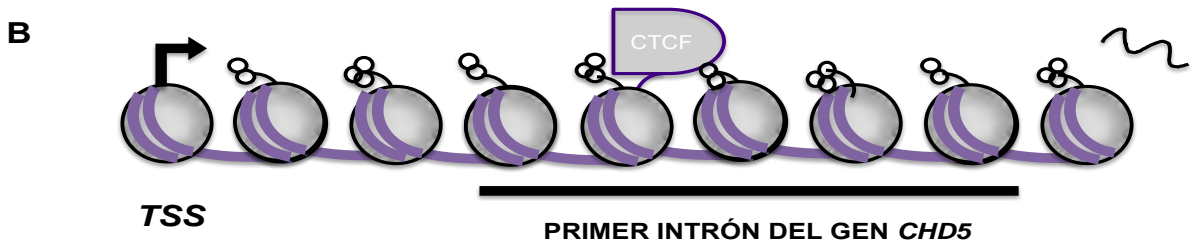
Una marca epigenética relevante para la elongación de la transcripción es H3K36me3. Esta MPT está principalmente enriquecida en los cuerpos de los genes o en las regiones estructurales, donde la disminución en su forma trimetilada está asociada con el silenciamiento génico (revisado en Li et al., 2007). En algunos genes, este silenciamiento no está relacionado con la inactivación del promotor del gen. Ya que KDM4A es capaz de eliminar esta marca de histonas, especulamos que la desmetilación de H3K36me3 podría desempeñar un papel importante en la regulación negativa de *CHD5*. Por lo tanto, para determinar el impacto de la presencia de CTCF y KDM4A en las marcas de histonas que se relacionan con la elongación de la transcripción, analizamos la abundancia de H3K36me3 y H3K36me2 en el primer intrón de *CHD5* mediante el ensayo ChIP (Figuras 7A, 7B y 7C). Cuando comparamos el enriquecimiento de las marcas H3K36me3 y H3K36me2 presentes en la región del primer intrón con la línea celular no neoplásica MCF 10A, encontramos que la presencia de CTCF y KDM4A se relaciona con una disminución de estas marcas epigenéticas en las líneas celulares

tumorales (Figuras 7A y 7B). Estos resultados sugieren que la presencia de CTCF y KDM4A podría alterar las marcas epigenéticas que se relacionan con la elongación transcripcional y, por lo tanto, afectar la transcripción de dichos genes (Figuras 7A, 7B y 7C).

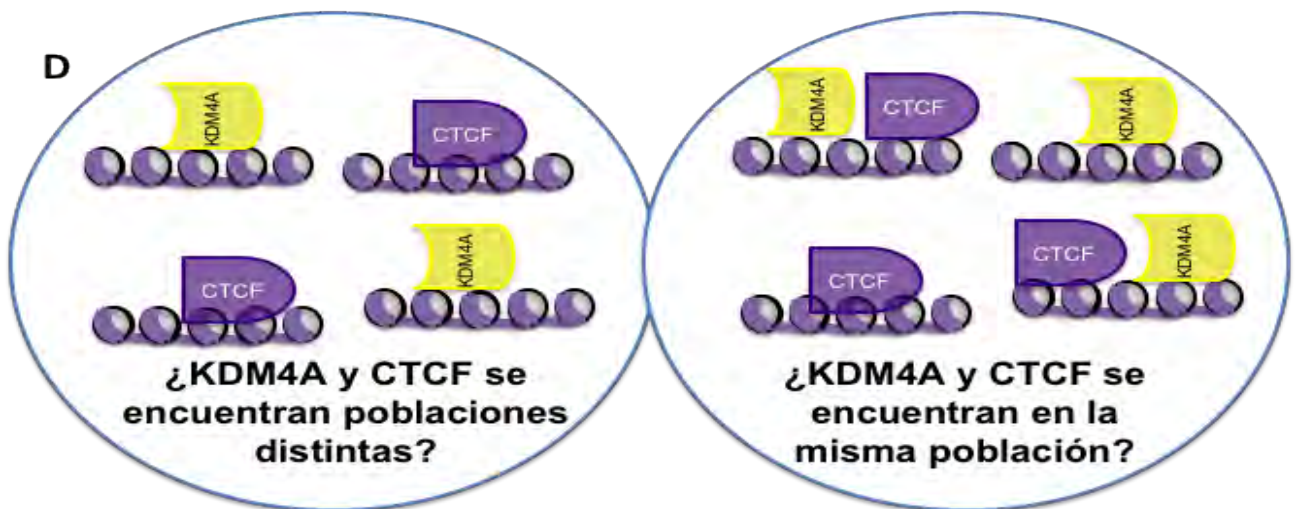
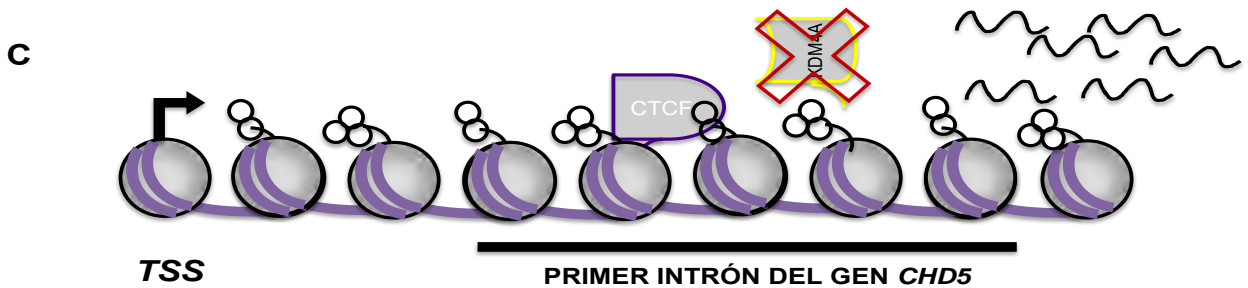
MCF7, MDA-MB-231 y HeLa, líneas celulares neoplásicas.



MCF7, MDA-MB-231 y HeLa, líneas celulares neoplásicas.



MCF 10A, línea celular no neoplásica.



H3K36me3

Disminución de H3K36me3

H3K36me2

Disminución de H3K36me2

mRNA de *CHD5*

Figura 7: La presencia de KDM4A y CTCF en el primer intrón de *CHD5* se relaciona con la disminución de H3K36me3 y H3K36me2 en esta región en las líneas celulares MCF7, MDA-MB-231 y HeLa.

(A) Esquema de los resultados del ChIP de KDM4A y posterior evaluación del primer intrón de *CHD5* en las líneas celulares MCF7, MDA-MB-231 y HeLa. (B) Esquema de los resultados del ChIP de CTCF y posterior evaluación del primer intrón de *CHD5* en las líneas celulares MCF7, MDA-MB-231 y HeLa. (C) Esquema de los resultados del ChIP con anticuerpos contra KDM4A y CTCF en la línea celular MCF 10A. (A, B, C) Esquema de los resultados del ChIP con anticuerpos contra H3K36me3 y H3K36me2 y posterior evaluación del primer intrón de *CHD5* en las líneas celulares MCF 10A, MCF7, MDA-MB-231 y HeLa. Los resultados muestran la media de tres réplicas biológicas independientes.

Las preguntas que quedan abiertas son saber si CTCF y KDM4A pueden coexistir o no en la misma región en la misma población (Figura 5D), así como determinar el efecto del abatimiento tanto de KDM4A o CTCF sobre las MPT H3K36me2, H3K36me3 y la expresión de *CHD5*.

8 PREGUNTA DE INVESTIGACIÓN.

¿Cuál es la participación de CTCF, KDM4A y H3K36me en la regulación de la expresión del gen *CHD5*?

9 HIPÓTESIS.

La localización de CTCF y KDM4A a lo largo del cuerpo de *CHD5* disminuirá su transcripción mediante la desmetilación de H3K36me3.

10 OBJETIVO GENERAL.

1. Determinar la participación de CTCF y KDM4A y las modificaciones de las histonas H3K36me3 y H3K36me2 en la regulación transcripcional de *CHD5* en un modelo de cáncer de mama.

10.1 Objetivos particulares.

1. Evaluar el estado de metilación del promotor de *CHD5*.
2. Determinar la co-existencia de KDM4A y CTCF en primer intrón de *CHD5*.
3. Determinar la existencia de un complejo entre KDM4A y CTCF.
4. Evaluar el efecto del "knockdown" de KDM4A y CTCF en la expresión de *CHD5*.
5. Determinar el efecto que tiene el "knockout" de KDM4A en la expresión de *CHD5* así como en la presencia de H3K36me3, H3K36me2 en el primer intrón de *CHD5*.
6. Evaluar la abundancia de las modificaciones postraduccionales de las histonas H3K36me3, H3K36me2, H3K9me3, H3K9me2 y H3K9me1 después del "knockout" de KDM4A.
7. Determinar la localización genómica de CTCF, KDM4A y del complejo CTCF-KDM4A.

11 RESULTADOS.

11.1 Análisis de metilación del DNA en el promotor de *CHD5*.

Algunos autores han reportado que la metilación del DNA en el promotor de *CHD5* puede alterar su expresión en varios tipos de cáncer y en líneas celulares neoplásicas (Du et al., 2016; Fatemi et al., 2014; Mulero-Navarro y Esteller, 2008; Wang et al., 2009; Zhao et al., 2012). Por lo tanto, con el objetivo de conocer el estado de metilación del promotor de *CHD5*, se realizó un análisis de la metilación del DNA a lo largo del *locus* de este gen en 743 pacientes con cáncer de mama y 98 muestras no neoplásicas obtenidas de la base de datos de The Cancer Genome Atlas (TCGA) (Illumina Human Methylation 450 K) a través del servicio web TCGA Wanderer (Díez-Villanueva et al., 2015). Este panel mide los niveles de metilación de 485,000 sitios CpG distribuidos a lo largo del genoma, de los cuales 63 CpG se encuentran dentro de la región genética de *CHD5* (Figura 8A); de estos sitios, solamente ocho CpG están localizados dentro del promotor del gen, los 55 sitios restantes se distribuyen a lo largo del cuerpo del gen. En el cuerpo del gen, se encontró que 34 CpG están metiladas (con un valor Beta ≥ 0.6 , que se considera como una región metilada) en el 50% de los pacientes, 20 de estos 34 sitios se encuentran metilados en el 80% de los pacientes. Sin embargo, al evaluar los niveles medios de metilación de los ocho sitios CpG ubicados dentro de la región promotora de *CHD5* (parte resaltada de la figura con un rectángulo en la Figura 8A) (Ensembl versión 75), observamos que solo uno de los 743 pacientes muestra metilación en promotor. Si el valor Beta es ≥ 0.6 , indica que la región analizada está metilada. Los valores Beta se calculan tomando las intensidades de lectura del fluorosforo asignado a DNA metilado (M) y el asignado al DNA no metilado (U), son la estimación del nivel de metilación utilizando la relación de intensidades entre sitios metilados y no metilados. Los valores β están entre cero y uno con cero sin metilar y uno completamente metilado. Si el valor Beta es ≥ 0.6 , indica que la región analizada está metilada (Figura 8B).

Dado el estado de metilación de *CHD5* en pacientes con cáncer de mama encontrado en la base de datos de TCGA, el siguiente objetivo fue caracterizar el estado de metilación del DNA en el promotor de *CHD5*, en las líneas celulares, para ello, se realizó un ensayo de PCR sensible a la metilación (MS-PCR) en la isla CpG que se observó previamente no estaba metilada en 742 pacientes (Figuras 8B y 8C). Descubrimos que la metilación del DNA en el promotor de *CHD5* está ausente en la mayoría de las líneas celulares, con la excepción de MDA-MB-231, esto se podría deber a un aumento en la abundancia de las metiltransferasas de DNA (Figura 8D); donde Mulero-Navarro y Esteller (Mulero-Navarro y Esteller, 2008) informaron

un hallazgo similar. Como control positivo de metilación del ensayo, utilizamos un DNA metilado *in vitro* (IVD) (Figura 8D).

Los resultados de la MS-PCR y el análisis de la base de datos de TCGA, sugieren que la metilación del DNA en el promotor de *CHD5* no es un mecanismo común involucrado en la represión de este gen. Por lo tanto, nos centramos en otro mecanismo epigenético que es independiente de la metilación del DNA, como la desmetilasa de histonas KDM4A.

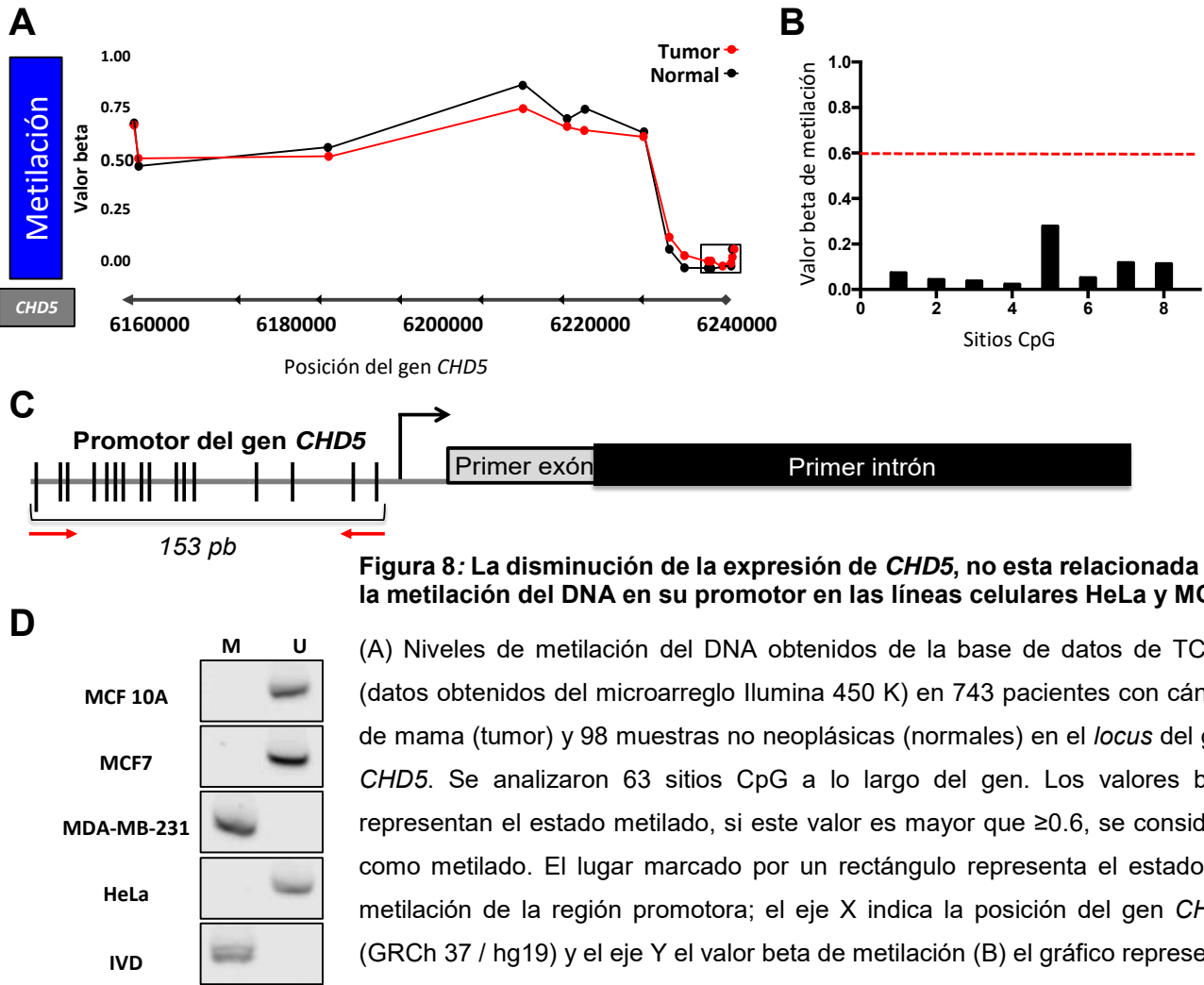


Figura 8: La disminución de la expresión de *CHD5*, no está relacionada con la metilación del DNA en su promotor en las líneas celulares HeLa y MCF7.

(A) Niveles de metilación del DNA obtenidos de la base de datos de TCGA (datos obtenidos del microarreglo Illumina 450 K) en 743 pacientes con cáncer de mama (tumor) y 98 muestras no neoplásicas (normales) en el *locus* del gen *CHD5*. Se analizaron 63 sitios CpG a lo largo del gen. Los valores beta representan el estado metilado, si este valor es mayor que ≥ 0.6 , se considera como metilado. El lugar marcado por un rectángulo representa el estado de metilación de la región promotora; el eje X indica la posición del gen *CHD5* (GRCh 37 / hg19) y el eje Y el valor beta de metilación (B) el gráfico representa los 8 sitios CpG analizados de la región promotora de 743 pacientes. La línea de puntos indica el umbral de metilación ($> 0.6 =$ metilado). (C) Esquema del gen *CHD5* y de la región analizada por MS-PCR, cada línea vertical representa un dinuceótido CpG. (D) Análisis de metilación del DNA del promotor *CHD5* en todas las líneas celulares por MS-PCR, U= no metilado, M= metilado, se usaron los oligonucleótidos indicados en la Tabla 15 que corresponden el amplicón que se indica en la figura 8C.

11.2 El complejo CTCF-KDM4A se localiza en el primer intrón de *CHD5* en las líneas celulares MCF7 y HeLa.

Como parte de los antecedentes directos de mi proyecto de maestría, encontramos que KDM4A y CTCF estaban localizadas en el primer intrón de *CHD5* por ensayos de ChIP, sin embargo una de las limitaciones del ensayo de ChIP es que no puede determinar la

coexistencia entre dos proteínas en una misma población, es por esto que se decidió realizar ensayos de ChIP-Re-ChIP, con el objetivo de demostrar la coexistencia de CTCF y KDM4A en el primer intrón de *CHD5* en una misma población celular (Figuras 9A y 9B), también conocido como inmunoprecipitación de la cromatina secuencial, en las líneas celulares MCF7 y HeLa. Nos centramos en estas líneas celulares ya que presentan bajos niveles de transcrito de *CHD5*, y además no poseen metilación del DNA en el promotor de *CHD5*. La primera inmunoprecipitación se realizó con cada uno de los anticuerpos (KDM4A o CTCF), y la segunda inmunoprecipitación con un segundo anticuerpo diferente al primero (CTCF-KDM4A o CTCF-KDM4A). Empleamos como control negativo de la técnica, el anticuerpo de interés seguido por IgGs o las IgGs seguidas por el anticuerpo de interés. Como control positivo para el ChIP-Re-ChIP de KDM4A, analizamos la región -1922 pb del TSS de *ASCL2* (Ver información suplementaria, figura 24A). Para CTCF, empleamos la región promotora de *WRAP53* (Ver información suplementaria, figura 24B). Como control negativo para ambas proteínas, evaluamos el exón 27 del *RB* (Ver información suplementaria, figura 24C). Los resultados de ChIP-Re-ChIP mostraron la coexistencia de CTCF y KDM4A en el primer intrón de *CHD5* tanto en células MCF7 como en HeLa (Figura 9A). Con los resultados de ChIP y de ChIP-Re-ChIP, evaluamos el porcentaje de co-ocupancia para ambos experimentos, CTCF-KDM4A o KDM4A-CTCF y también para los experimentos de IgG, con resultados negativos representados con un valor de 0. El cálculo del % de co-ocupancia se llevó a cabo como lo reportó Geisberg en el 2005 (Geisberg y Struhl, 2005) (Figura 9B). Estos resultados, junto con el análisis de expresión mostrado en la figura 6 sugieren que la mayor co-ocupancia de CTCF y KDM4A se asocia con menor expresión de *CHD5*.

Los resultados del ChIP-Re-ChIP mostraron que CTCF y KDM4A coexisten en el primer intrón del gen *CHD5* en la misma población de células MCF7 y HeLa (Figura 9A) y sugieren que CTCF y KDM4A pueden formar un complejo.

Con el propósito de demostrar la posible formación de un complejo entre CTCF y KDM4A, se realizó un ensayo de co-inmunoprecipitación en células HeLa. Como un primer enfoque se inmunoprecipitó contra CTCF y reveló con CTCF (Figura 9C). Posteriormente, usando las proteínas obtenidas de la inmunoprecipitación de CTCF, se reveló utilizando un anticuerpo contra KDM4A (Figura 9D). Nuestros resultados muestran que CTCF y KDM4A están formando un complejo en células HeLa (Figuras 9C y 9D).

Además, estos resultados sugieren que el complejo CTCF-KDM4A puede estar involucrado en la represión transcripcional de *CHD5*. Ya que en la línea celular MCF 10A dónde solamente se encuentra CTCF y está localizado en el primer intrón de *CHD5*, no se observa una baja expresión de *CHD5*, contrario a las líneas celulares MCF7 y HeLa donde el complejo

de KDM4A y CTCF está localizado en el primer intrón, se observa una menor expresión de *CHD5* en comparación con las células MCF 10A (Figuras 6, 7 y 9).

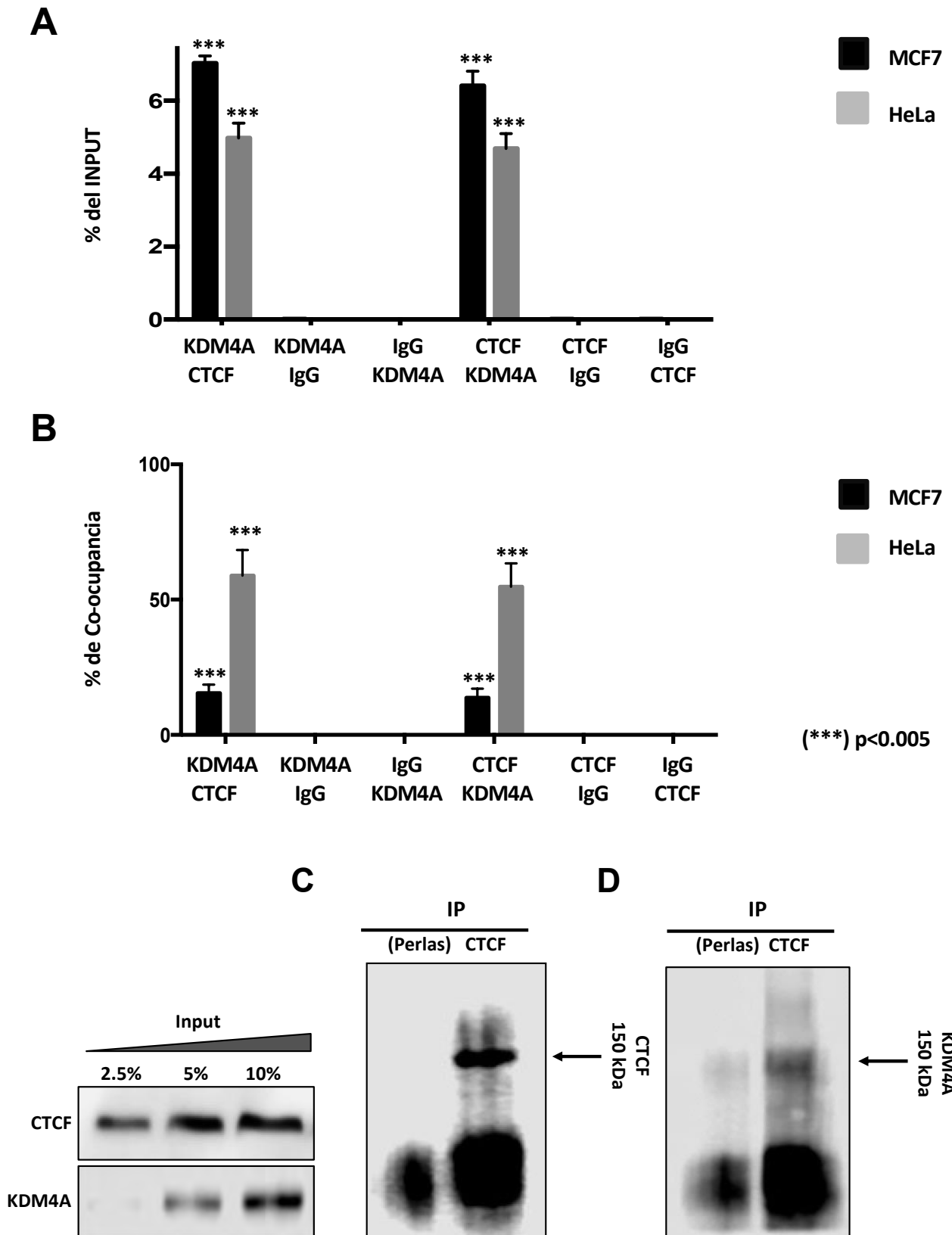


Figura 9: El complejo CTCF-KDM4A se localiza en el primer intrón de *CHD5* en las líneas celulares MCF7 y HeLa.

(A) Ensayo de ChIP/re-ChIP empleando los anticuerpos indicados en la primera fila y posteriormente se inmunoprecipitó usando los anticuerpos indicados en la segunda fila en las líneas celulares MCF7 y HeLa. (B) El análisis de co-ocupancia se realizó de acuerdo con Geisberg y Struhl (Geisberg y Struhl, 2005). Para el análisis de co-ocupancia, se utilizaron los datos del experimento de ChIP-Re-ChIP y de ChIP (CTCF-KDM4A o KDM4A-CTCF). Además, se evaluó la co-ocupancia de los experimentos con IgG. Se muestra el promedio de 3 réplicas biológicas. Los resultados están representados en % de co-ocupancia. Las diferencias estadísticas se determinaron utilizando la prueba de la t de Student, (***) $p < 0,005$ en comparación con las IgG. (C – D) Se realizó un ensayo de co-inmunoprecipitación con CTCF y se reveló un anticuepro anti-CTCF (150 kDa) (C). Con las proteínas obtenidas de la IP de CTCF se reveló con un anticuerpo anti-KDM4A (150 kDa) (D). A la izquierda, el INPUT se evaluó contra CTCF y KDM4A en cantidades crecientes de proteína (2.5, 5 y 10%).

11.3 El Knockdown de KDM4A o CTCF por medio siRNA y shRNA está relacionado con el incremento de la expresión de *CHD5*.

Con el objetivo de determinar la participación de KDM4A en la represión transcripcional del gen *CHD5*, se transfectaron células MCF7 y HeLa con el conjunto de 5 diferentes siRNAs contra KDM4A, 72 horas después de la transfección se realizaron análisis de expresión para los genes *KDM4A* y *CHD5* por medio de RT-qPCR. Los resultados muestran que después de la transfección hay una disminución de los niveles de mRNA de *KDM4A* en un 60% en MCF7 y en 40% en HeLa (Figura 10A). Adicionalmente, los niveles de mRNA de *CHD5* aumentaron en ambas líneas celulares (Figura 10B), en MCF7, se puede observar que la expresión de *CHD5* es mayor que en las células MCF 10A, y en las células HeLa los niveles de *CHD5* son comparables con la línea celular no neoplásica MCF 10A (Figura 10B).

Ya que CTCF y KDM4A forman un complejo, nos propusimos averiguar cuál es el papel de CTCF en la represión transcripcional del gen *CHD5*, con este objetivo, se transfretaron células MCF7 y HeLa con un shRNA contra CTCF, el cual fue amablemente donado por Nakao (Ishihara et al., 2006); 48 horas después de la transfección se realizaron análisis de expresión de *CTCF* y *CHD5* por medio de RT-qPCR (Figuras 10C y 10D). Los resultados revelaron que después de la transfección hay una disminución en los niveles de mRNA de *CTCF* en un 50% en MCF7 y en 40% en HeLa (Figura 10C). La disminución en la expresión de *CTCF*, se relacionó con el aumento de la expresión de *CHD5* en ambas líneas celulares, donde los niveles de mRNA de *CHD5* son comparables con la línea celular no neoplásica MCF 10A (Figura 10D). Conjuntando los resultados, sugieren que la presencia de CTCF y KDM4A en el primer intrón del gen *CHD5* actúa como represor transcripcional de este gen.

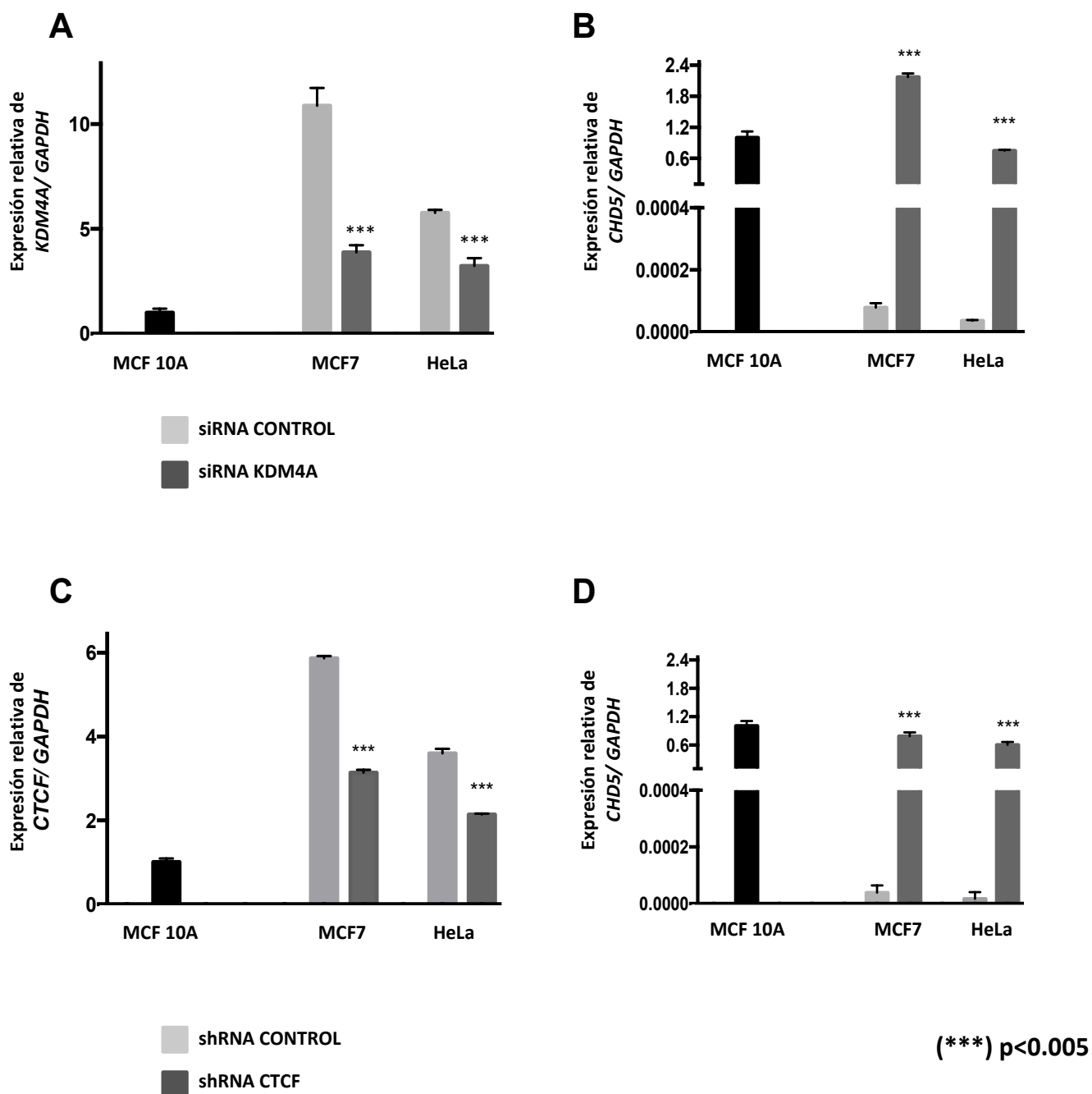


Figura 10: La expresión de *CHD5* aumenta después del Knockdown de CTCF y KDM4A en las líneas celulares MCF7 y HeLa.

Células MCF7 y HeLa fueron transfectadas transitoriamente con siRNAs en contra de KDM4A. Evaluación de la expresión de *KDM4A* (A) y *CHD5* (B) en células MCF7 y HeLa después de la transfección transitoria de un siRNA en contra de KDM4A. Análisis de expresión de *CTCF* (C) y *CHD5* (D) en células MCF7 y HeLa después de la transfección transitoria de un shRNA contra CTCF. Los datos fueron normalizados usando a *GAPDH*. Los resultados muestran la media de tres réplicas biológicas independientes. Los datos de expresión se normalizaron con respecto a la línea celular MCF 10A a la cual asignamos un valor relativo de uno. (***) p < 0.005 con respecto a la línea celular transfectada con los controles, como prueba estadística se utilizó una t de Student.

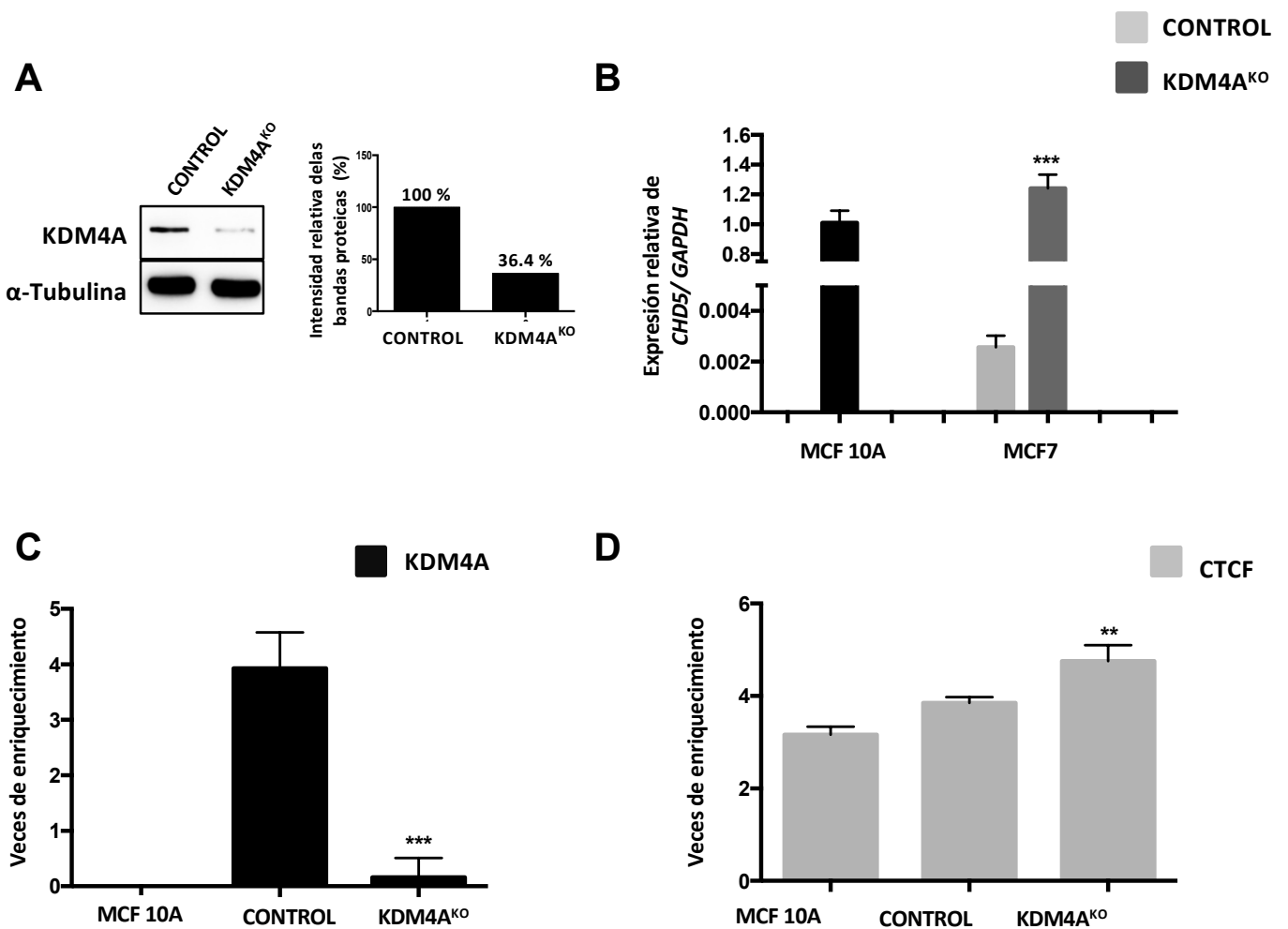
11.4 El Knockout de KDM4A en células MCF7 restablece la marca de histonas H3K36me3 en el primer intrón de *CHD5* y reactiva su expresión.

Con el fin de validar con otro modelo que KDM4A regula negativamente a *CHD5*, se realizó un modelo Knockout (KDM4A^{KO}) usando el sistema CRISPR/Cas9 KO (Santa Cruz, sc-404599 y sc-404599-HDR). Este sistema empleó tres RNAs guía gRNA que se dirigieron al exón 3 y 8 del gen KDM4A (Ver información suplementaria, figura 25A). Posteriormente, se seleccionaron las células mediante el tratamiento con Puromicina y con el objetivo de enriquecer aún más nuestra población KDM4A^{KO} se seleccionaron las células que presentaban una fluorescencia más alta por medio de citometría de flujo y FACS (Información suplementaria, figura 25B). Como control, empleamos un plásmido de gRNA no dirigido (Mock) (Santa Cruz, sc-418922).

Subsecuentemente, se evaluó mediante Western Blot la abundancia de la proteína KDM4A en células MCF7 Mock y KDM4A^{KO}, donde se observó una reducción del 63,6% de KDM4A en células KDM4A^{KO} (Figura 11A). También se realizó el análisis por medio de RT-qPCR de la expresión de *CHD5* en células MCF 10A, MCF7 Control y KDM4A^{KO}, se puede observar una reactivación significativa de la expresión de *CHD5* en células KDM4A^{KO}, con niveles similares a los observados en MCF 10A (Figura 11B). Para evaluar si esta reactivación se encuentra relacionada con la pérdida de KDM4A del primer intrón de *CHD5*, realizamos un análisis de ChIP contra KDM4A. Nuestros resultados muestran una pérdida significativa de KDM4A en KDM4A^{KO} en comparación con las células Control (Figura 11C). Con respecto a nuestros resultados anteriores que sugieren que el complejo CTCF-KDM4A regula *CHD5*, se evaluó si la localización de CTCF podría verse afectado por la pérdida de KDM4A en el primer intrón de *CHD5*. Nuestros resultados muestran que la unión a CTCF es independiente de la presencia de KDM4A, lo que sugiere que CTCF puede actuar como represor cuando está en un complejo con KDM4A (Figura 11D). Ya que en la línea celular MCF 10A dónde solamente se encuentra CTCF y está localizado en el primer intrón de *CHD5*, no se observa una baja expresión de *CHD5*, así mismo en la línea celular MCF7^{KO}, KDM4A se disocia y CTCF permanece en el primer intrón de *CHD5* y la expresión de este gen aumenta tras el KO de KDM4A, contrario a las líneas celulares MCF7 y HeLa donde el complejo de KDM4A y CTCF está localizado en el primer intrón, se observa una menor expresión de *CHD5* comparado con la expresión de este gen en las células MCF 10A (Figuras 6, 7, 9 y 11). Debido a los resultados obtenidos, se intentó establecer un modelo de CTCF^{KO}, sin embargo, estas células no eran viables, por lo que el enfoque experimental no fue posible. La generación de un modelo CTCF^{KO}

ha sido un desafío experimental para diferentes grupos de investigación. Particularmente, la alteración en la abundancia de CTCF afecta la proliferación celular e incluso puede ser causal de un fenotipo letal en modelos murinos como ya ha sido reportado anteriormente (González-Buendía et al., 2014; Moore et al., 2012; Splinter et al., 2006).

Una de las preguntas centrales que queríamos abordar es si la pérdida de KDM4A podría restaurar el patrón H3K36me3 en el primer intrón de *CHD5*. Por lo tanto, realizamos un análisis CHIP contra H3K36me2 y H3K36me3 en las células MCF 10A, MCF7 Control y KDM4A^{KO}. Observamos el restablecimiento de la modificación en células KDM4A^{KO}, y también no se observó un cambio significativo en la modificación H3K36me2 (Figura 11E). Esto sugiere que la pérdida de desmetilasa KDM4A permite la reincorporación de H3K36me3 en el primer intrón de *CHD5*, favoreciendo la reactivación de la expresión génica.



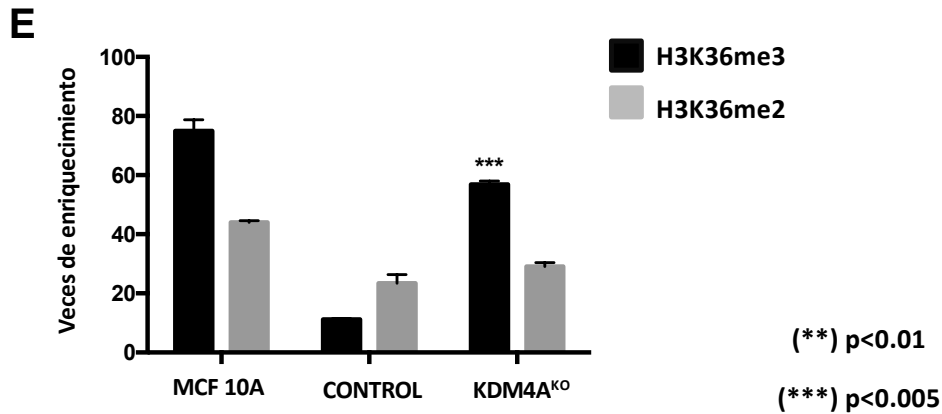


Figura 11: El Knockout de KDM4A promueve el restablecimiento de la MPT H3K36me3 en el primer intrón de *CHD5* así como el aumento en su expresión.

(A) Caracterización de la abundancia de la proteína KDM4A mediante inmunoblots en células MCF7 transfectadas con un plásmido control (Control) o transfectadas con plásmidos KDM4A^{KO} CRISPR/Cas9 y HDR KDM4A. La cuantificación de la intensidad relativa de las bandas proteicas muestra una disminución del 63.6% de la abundancia de KDM4A en las células KO. (B) Análisis de expresión de *CHD5* en células Control y KDM4A^{KO}. Los datos fueron normalizados con *GAPDH* como control interno en dos experimentos independientes, se emplearon las células MCF 10A como control. (C, D, E) Evaluación por medio de qPCR del primer intrón de *CHD5* del DNA obtenido de los ensayos de inmunoprecipitación de la cromatina se utilizaron anticuerpos anti-KDM4A (C), anti-CTCF (D), anti-H3K36me3 y anti-H3K36me2 (E) en las líneas celulares MCF 10A, MCF7 Control y KDM4A^{KO}. Como control negativo se empleó el anticuerpo irrelevante IgG incluido en el OneDay ChIP kit (Diagenode, NJ, USA, Kch-onedIP-180). (**) p < 0.01 (***) p < 0.005 comparado con las células MCF7 Control. Como prueba estadística se empleó una *t* de Student. Los resultados muestran el promedio de tres réplicas biológicas.

11.5 El Knockout de KDM4A en células MCF7 afecta los niveles globales de H3K36me3 y H3K36me2.

Con el objetivo de investigar si los cambios en la abundancia de las histonas se podían observar a nivel global, se realizaron extracción de proteína total y extracción ácida de las histonas del pase 5 de la línea celular MCF7 KDM4A^{KO} y la línea celular MCF7 Wild Type (WT). Posterior a la extracción de este material se evaluó la abundancia de la proteína KDM4A empleando distintas cantidades de proteína total (Figura 12A) y también se evaluó la abundancia de las marcas de histonas H3K9me3, H3K9me2, H3K36me3 y H3K36me2, se utilizaron estas MPT de las histonas ya que son los blancos enzimáticos de KDM4A, esta enzima desmetila H3K9me3 y produce H3K9me2 y también remueve el grupo metilo de H3K36me3 y produce H3K36me2. Adicionalmente se usó H3K9me1 como control negativo. (Figura 12C). Los resultados mostraron que hay una reducción del 78% en las células MCF7

KDM4A^{KO} comparando con células MCF7 WT (Figura 12B). Así mismo, de acuerdo con nuestros datos anteriores, donde se observó un aumento de H3K36me3 y una disminución de H3K36me2 en el primer intrón del gen *CHD5*, el análisis de la abundancia relativa de las modificaciones H3K9me3, H3K9me2 y H3K9me1 (Figura 12D), arrojo que dichas marcas no cambian en las células KDM4A^{KO} con respecto a las células WT, por el contrario se observó una disminución de la H3K36me2 y un aumento en la H3K36me3 en las células KDM4A^{KO} con respecto a las células WT. Los resultados del análisis de los ensayos de ChIP y de WB indican que los cambios en las MPT son tanto locales como globales.

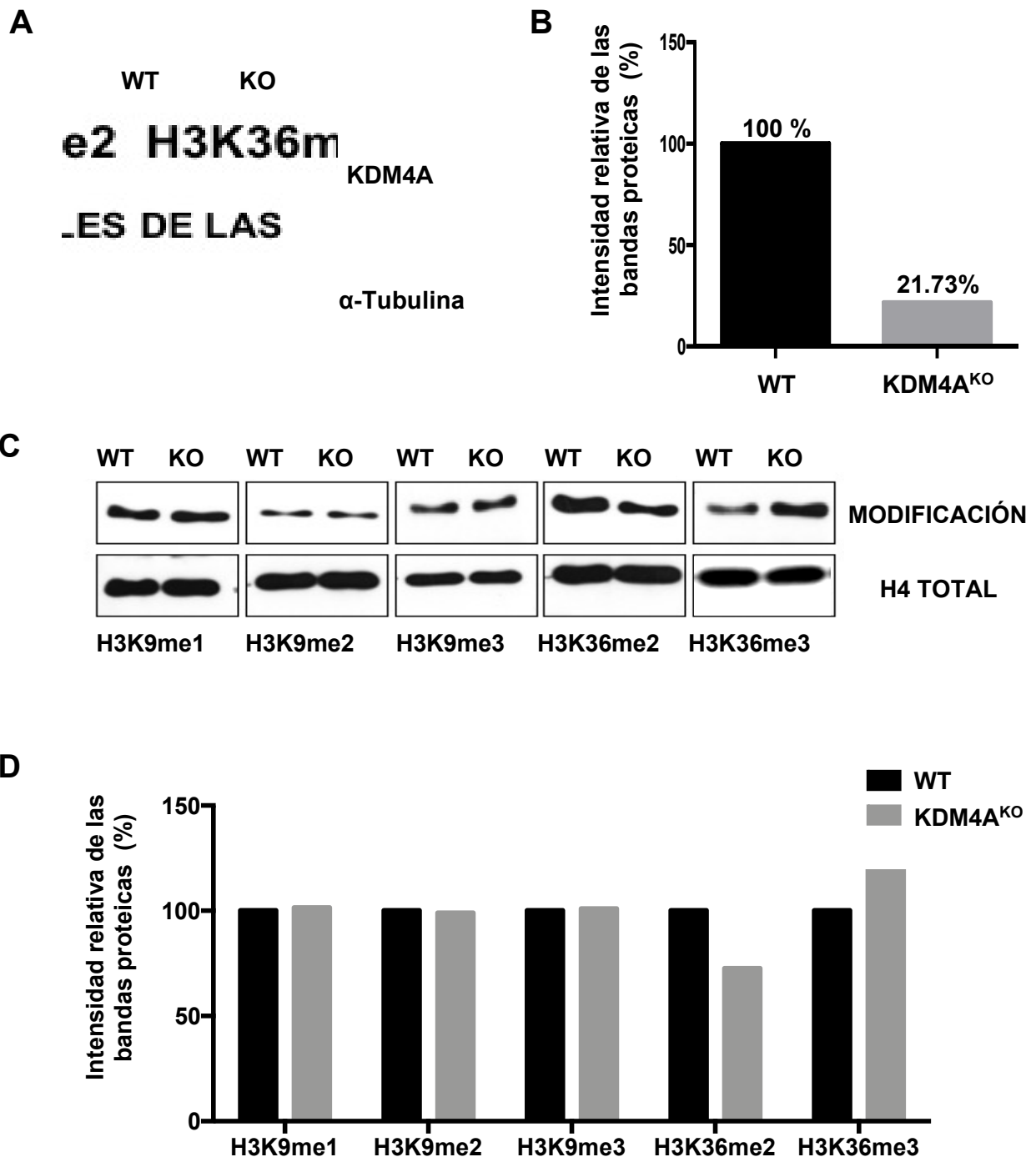


Figura 12: El Knockout de KDM4A afecta los niveles globales de H3K36me2 y H3K36me3.

(A) Caracterización de la abundancia de la proteína KDM4A mediante inmunoblots, se emplearon 60µg cantidades de proteína total, en células MCF7 transfectadas con plásmidos KDM4A^{KO} CRISPR/Cas9 y HDR KDM4A y células MCF7 WT (B). La cuantificación de la intensidad relativa de las bandas proteicas muestra una disminución del 78.26% de la abundancia de KDM4A en las células KO comparando con las células WT. (C) Caracterización de la abundancia de las marcas de histonas H3K9me1, H3K9me2, H3K9me3, H3K36me2 y H3K36me3 mediante inmunoblots, se cargaron 8µg de extracto histónico para cada condición y se normalizó con la abundancia de la histona H4 total. (D) La cuantificación de la intensidad relativa de las bandas proteicas muestra que no hay un cambio en las marcas H3K9me1, H3K9me2 y H3K9me3, la marca H3K36me2 disminuye 27.273% y la marca H3K36me3 aumenta 37.73% en las células KDM4A^{KO} comparado con las células WT. Se muestran los resultados de una réplica biológica.

11.6 Análisis de la distribución de las proteínas CTCF y KDM4A y el complejo CTCF-KDM4A en diversos genes.

Con el objetivo de conocer la localización a nivel genómico del factor nuclear CTCF, la desmetilasa de histonas KDM4A y el complejo CTCF-KDM4A, se realizaron ensayos de ChIP-seq y ChIP-Re-ChIP-seq. De manera general, el primer camino para analizar los de datos de secuenciación masiva provenientes de un ensayo de ChIP-seq, es un análisis de calidad, para esto se usan herramientas como MultiQC, las cuales generan informes estadísticos del desempeño de los ensayos de secuenciación masiva a partir de diferentes parámetros bioinformáticos (Ewels et al., 2016). La Tabla 13 muestra los resultados estadísticos generales obtenidos con la herramienta MultiQC a partir de los datos de la secuenciación masiva de los ensayos de ChIP-seq y ChIP-Re-ChIP-seq en células MCF7, en la Tabla 3 se muestran el número total de secuencias encontradas (M Seqs), el total de lecturas mapeadas al genoma humano (M Reads Mapped) y el porcentaje de secuencias mapeadas que representa dicha cantidad de lecturas (%Mapped). Debido a que algunos fragmentos de DNA no fueron identificados durante el proceso de secuenciación, solo las secuencias más enriquecidas de cada muestra fueron mapeadas al genoma. Dependiendo del comportamiento de la proteína bajo estudio, se recomienda cumplir con una cierta cantidad de lecturas. En el estudio de moléculas que se distribuyen ampliamente en el genoma, como las marcas de histonas, se recomienda tener más de 20 millones de lecturas (M Reads Mapped). En el caso de los factores de transcripción; es decir, proteínas que tienen sitios de unión discretos en el genoma, se recomienda contar con más de 10 de millones de lecturas. En la Tabla 3, se observa que la mayoría de los ensayos de ChIP-seq cubren con este requerimiento. Además, los altos porcentajes de secuencias mapeadas (%Mapped) revelan una baja cantidad de artefactos o

contaminación de las muestras. Así mismo, con ayuda de la herramienta MultiQC se puede visualizar el rango de valores de calidad de todas las bases para cada posición en una secuencia, mediante un histograma de calidad, el cual muestra que todas las secuencias analizadas se encuentran en un rango óptimo de calidad (Figura 13).

Tabla 3: Análisis estadístico de los ensayos ChIP-seq y ChIP-Re-ChIP-seq (ChIP-seq).

M reads Mapped, Millones de lecturas mapeadas; % Mapped, Porcentaje de lecturas mapeadas; M seqs, Millones de secuencias de cada muestra.

	Nombre de la muestra	M Reads Mapped	% Mapped	M Seqs
CTCF	ChIP de CTCF R1.	2.7	63.4%	4.3
	ChIP de CTCF R2.	17.8	71.5%	24.9
KDM4A	ChIP de KDM4A R1.	18.6	75.2%	24.7
	ChIP de KDM4A R2.	15.1	81.2%	18.6
	ChIP de KDM4A R3.	14.3	72.1%	19.8
CTCF-KDM4A	ChIP-Re-ChIP CTCF-KDM4A R1.	12.0	60.1%	20.0
	ChIP-Re-ChIP CTCF-KDM4A R2.	14.7	74.6%	19.7
	ChIP-Re-ChIP CTCF-KDM4A R3.	16.9	72.6%	23.3
KDM4A-CTCF	ChIP-Re-ChIP KDM4A-CTCF R1.	2.0	80.0%	2.5
	ChIP-Re-ChIP KDM4A-CTCF R2.	1.2	74.1%	1.7
	ChIP-Re-ChIP KDM4A-CTCF R3.	22.4	72.5%	31.0

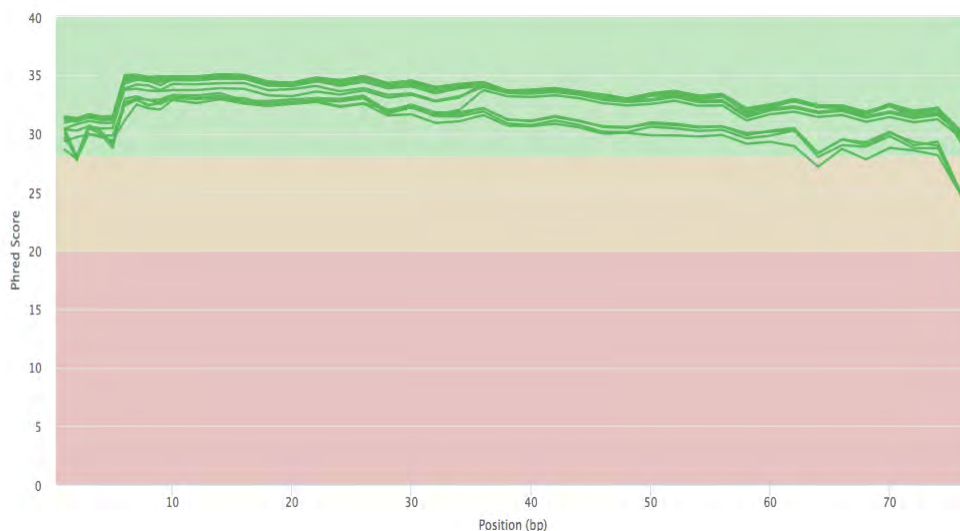


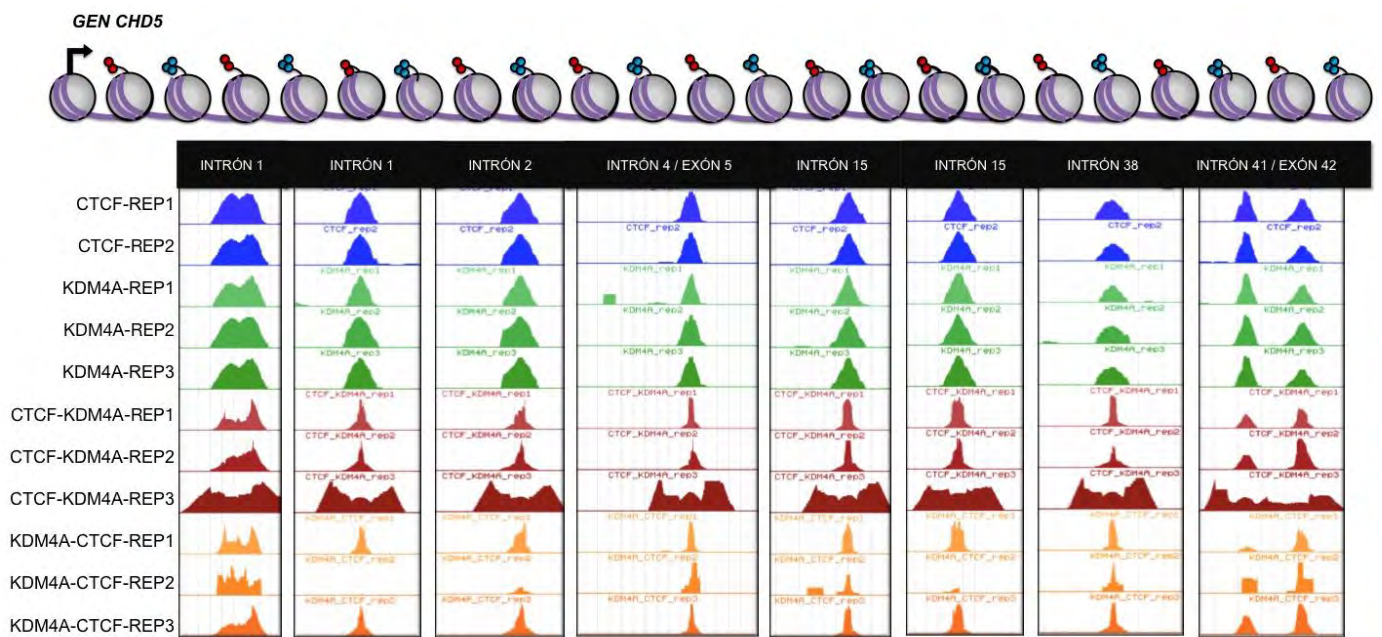
Figura 13: Calidad de las secuencias.

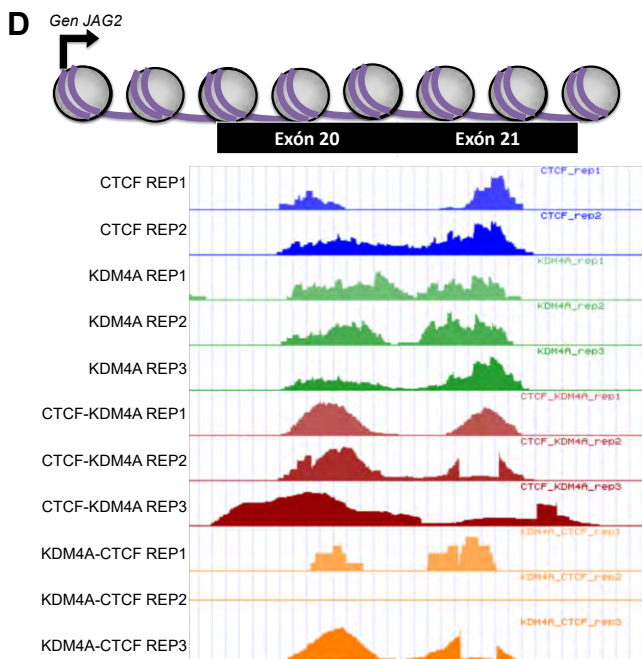
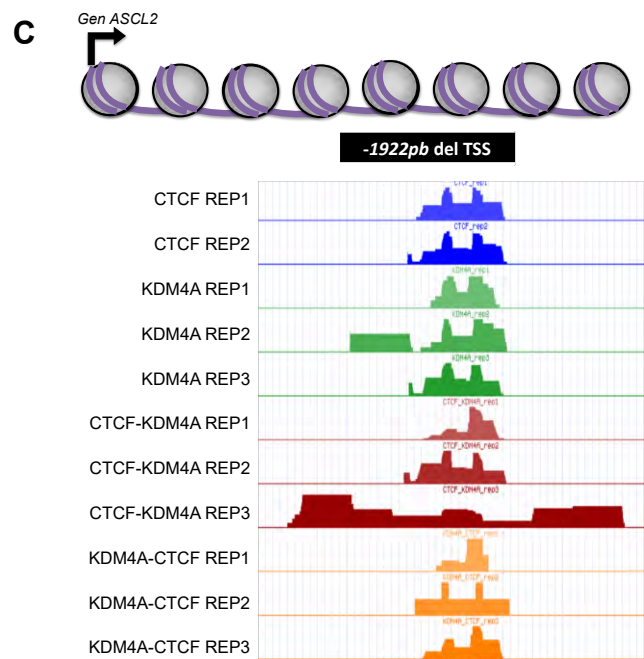
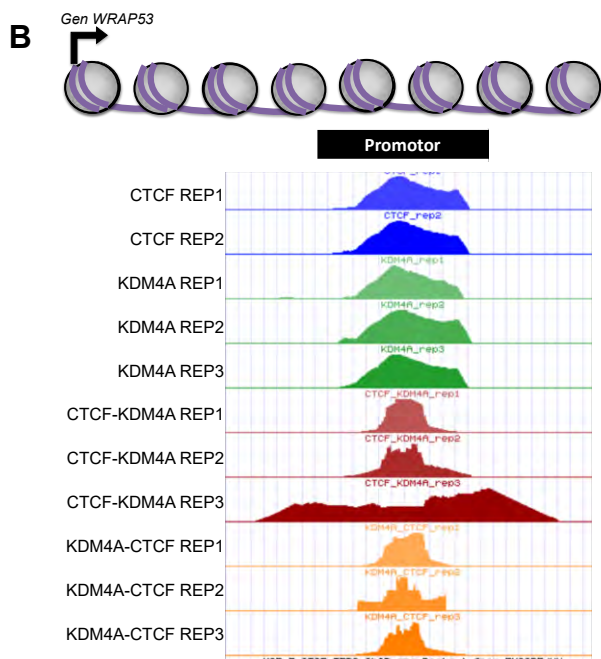
El eje “y” muestra los puntajes de calidad. Cuanto mayor sea el puntaje, mejor será llamada base. El fondo del gráfico divide al eje Y en anotaciones de muy buena calidad (verde), anotaciones de calidad razonable (naranja) y anotaciones el de mala calidad (rojo). MultiQC.

Después de alinear las muestras y llamar los picos con las herramientas bioinformáticas BWA y MACS2, se visualizó la señal de enriquecimiento de las proteínas de interés en cada

muestra con ayuda del visualizador UCSC genome browser. De esta manera, fue posible corroborar la presencia de las proteínas KDM4A, CTCF y su complejo en el primer intrón de *CHD5*, como fue anteriormente reportado e identificar otros siete sitios intrónicos y exónicos dentro del mismo gen que eran desconocidos (Figura 14A). En el visualizador las réplicas con más lecturas tienen picos de enriquecimiento más grandes. El análisis bioinformático nos mostró que en casi la mayoría de los casos las señales de CTCF y KDM4A se traslapan, y esto coincide con las señales encontradas por el ChIP-Re-ChIP-seq (Figura 14). Además de encontrar a KDM4A y CTCF y el complejo CTCF-KDM4A en el gen *CHD5* se encontró a estas proteínas en la región promotora del gen *WRAP53*, la regio situada a -1922 pb del TSS de *ASCL2* y los exones 20 y 21 del gen *JAG2*. Estos datos abren la posibilidad de realizar un análisis masivo de ChIP-seq evaluando las modificaciones que son afectadas por KDM4A en las células KDM4A^{KO} y posteriormente comparar los datos obtenidos con las células control.

A





corre
y 3 r
ChIP.

Fig el
cor. proye proteico CTCF-KDM4A en distintos genes
por medio de ChIP-seq y ChIP-Re-ChIP-seq.

Se muestra la localización del factor nuclear CTCF, la desmetilasa de histonas KDM4A y el complejo proteico CTCF-KDM4A (A) a lo largo de *CHD5* (B) en la región promotora de *WRAP53* (C) en el sitio localizado a -1922pb del TSS de *ASCL2* y (D) en los exones 20 y 21 de *JAG2*. En la columna de la izquierda indica el anticuerpo que se usó y la réplica, correspondiendo a las primeras dos filas (picos azules) a los ensayos de ChIP-seq contra CTCF, réplicas 1 y 2 respectivamente. Las filas 3, 4 y 5 (picos verdes) corresponden a los ensayos de ChIP-seq contra KDM4A, réplicas 1, 2 y 3 respectivamente. Las filas 6, 7 y 8 (picos rojos) e se inmunoprecipitó CTCF y luego KDM4A, réplicas 1, 2 y 3 respectivamente. Los picos naranjas corresponden a los ensayos de ChIP-Re-ChIP-seq, réplicas 1, 2 y 3 respectivamente.

12 DISCUSIÓN.

Las alteraciones epigenéticas son una característica común encontrada en los procesos cancerígenos (revisado en Jovanovic et al., 2010; Rea et al., 2000). Principalmente, los componentes epigenéticos clave, que incluyen metilasas y desmetilasas como KDM4A, así como las proteínas arquitectónicas tales como CTCF, están desregulados (revisado en Berry y

Janknecht, 2013; Soto-Reyes et al., 2012). Varios estudios han reportado que los niveles de KDM4A aumentan en tejidos de cáncer de mama. Esta desmetilasa elimina el grupo metilo de H3K9me3 y H3K36me3, la primera modificación se asocia con heterocromatina y la represión transcripcional (Rea et al., 2000), mientras que la última modificación está enriquecida en los cuerpos de los genes que son transcripcionalmente activos y se asocia con el reclutamiento de la RNA polimerasa II. y el proceso de elongación transcripcional (Cascante et al., 2014; Li et al., 2007; Wang et al., 2018). Por lo tanto, la alteración de H3K36me3 podría afectar la transcripción de los genes sin alterar sus promotores, lo que sugiere un nuevo mecanismo de desregulación genética no asociado con las regiones reguladoras.

CHD5 es un gen que codifica una enzima que pertenece a la familia de las helicasas (Chromodomain DNA binding protein 5) (Thompson et al., 2003). La proteína CHD5 puede funcionar como un supresor de tumores regulando la apoptosis y la senescencia celular, ya que participa en la vía p19Arf / p53 al interactuar con MDM2 (Bagchi et al., 2007; Tao y Levine, 1999). Debido a que esta interacción conduce a la atenuación de la degradación de p53 mediada por MDM2 (Fujita et al., 2008), CHD5 y p19ARF ayudan a estabilizar p53. Además, CHD5 inhibe el crecimiento clonogénico *in vitro*, así como el crecimiento tumoral por medio de xenoinjertos, lo que sugiere que su inactivación puede estar involucrada en el desarrollo del cáncer (Du et al., 2016; Mulero-Navarro y Esteller, 2008). Algunos estudios han sugerido que *CHD5* puede inactivarse por procesos genéticos (Fujita et al., 2008) o epigenéticos, pero estos informes se enfocaron principalmente en su represión por la metilación del DNA en su promotor (Du et al., 2016; Mokarram et al., 2009; Mulero-Navarro y Esteller, 2008; Zhao et al., 2012). El análisis de los conjuntos de datos de TCGA muestra que la región promotora de *CHD5* no está metilada en pacientes con cáncer de mama, lo que sugiere que otro mecanismo epigenético podría estar involucrado en la represión de este gen. En este aspecto, existe evidencia que sugiere que la alteración en el promotor *CHD5* no es el principal mecanismo de represión de este gen (Malette y Richard, 2012).

Anteriormente, se reportó que KDM4A se localiza en el primer intrón de *CHD5* y la reducción en KDM4A conduce a un aumento de la expresión de *CHD5* en las células U2OS; esto propone que KDM4A regula negativamente a *CHD5* (Malette y Richard, 2012). Sin embargo, dado que KDM4A no se encontró en el promotor de *CHD5*, el mecanismo de regulación negativa dependiente de KDM4A hasta ahora queda poco claro.

Se sabe que la sobreexpresión de KDM4A está asociada con la proliferación celular y el mal pronóstico en varios cánceres (Berry et al., 2012; Guerra-Calderas et al., 2015; Jin et al., 2017, 2017; Kauffman et al., 2011; Kim et al., 2016). Comprender los mecanismos moleculares

subyacentes a los efectos de KDM4A puede tener un impacto terapéutico en el futuro y sus implicaciones en el cáncer es un tema importante para futuras investigaciones clínicas (Chang et al., 2011; Duan et al., 2015; Franci et al., 2017; Metzger et al., 2017). Nuestros hallazgos revelan que KDM4A funciona como un represor del supresor de tumores *CHD5* al afectar las marcas epigenéticas que se asocian con la elongación transcripcional y no al regular el promotor del gen. Este fenómeno se ha reportado en otros modelos celulares, donde KDM4A / C alteran específicamente la modificación H3K36me3 (Cascante et al., 2014). Este fenómeno también se asocia con la pérdida del reclutamiento de la RNA polimerasa II en las regiones transcritas del gen *GFAP* (Cascante et al., 2014). Nuestros resultados sugieren un nuevo mecanismo de represión de *CHD5*, donde la disminución de H3K36me3 en el cuerpo de *CHD5* puede conducir a la represión transcripcional. Una hipótesis es que este fenómeno se produce debido a la falta de fosforilación de la segunda serina en el dominio carboxi terminal (CTD) de la RNA polimerasa II, lo que resulta en el enriquecimiento de H3K36me2 y una disminución la elongación transcripcional, o debido a un aumento en las marcas represivas de histonas (Fong et al., 2017).

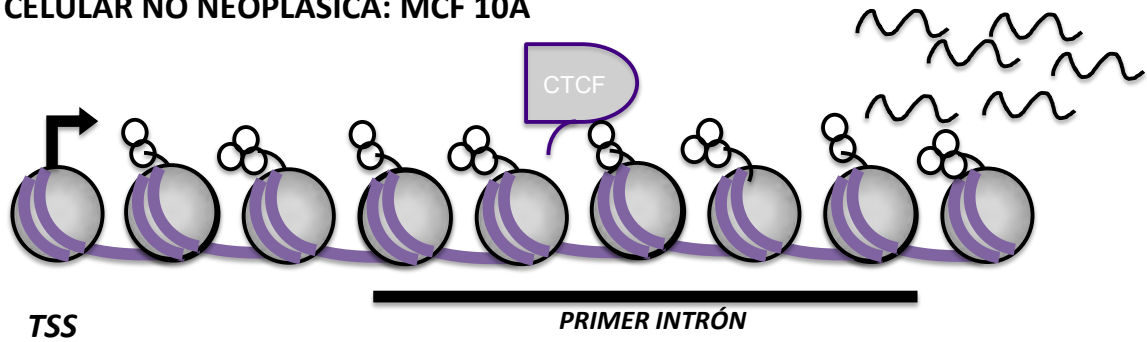
Ensayos *in vitro* han reportado que la presencia de CTCF aumenta la frecuencia de desmetilación de KDM4A hasta en un 60%, lo que sugiere que CTCF tiene un papel en la función de desmetilación de KDM4A (Jeong et al., 2011). Otro estudio proporcionó apoyo adicional para estos datos ya que demostraron que CTCF puede interactuar con la desmetilasa de histonas KDM5B y aumentar su actividad de desmetilación en líneas celulares de cáncer de mama (Yamamoto et al., 2014).

Existen pocos estudios donde se ha reportado que CTCF puede actuar como un represor transcripcional. Se han propuesto varios mecanismos, uno de ellos es la asociación con el co-represor SIN3A y desacetilasas de histonas HDACs (Lutz et al., 2000). Por otro lado, *c-MYC*, *Bax*, *hTERT*, son ejemplos de genes que están regulados de manera negativa por la unión de CTCF-P (CTCF- fosforilado) a regiones localizadas río abajo del TSS. Esta unión impide el reconocimiento de diversos factores de transcripción (Filippova et al., 1996; Méndez-Catalá et al., 2013; Renaud et al., 2005). También, hay evidencias de que la represión de los genes ribosomales, depende, en parte, de la poli-ADP-ribosilación de CTCF, modificación postraduccional que ayuda en a la translocación de CTCF en el nucléolo y se ha sugerido que CTCF-poli-ADP-ribosilado reprime la transcripción del rDNA (revisado en Caiafa y Zlatanova, 2009; Torrano et al., 2006). Otros estudios han demostrado que CTCF juega un papel relevante en la inactivación del cromosoma X, ya que se une al promotor del RNA no codificante *Jpx*, la transcripción de *Jpx* es indispensable para la transcripción de *Xist* el cual es necesario para

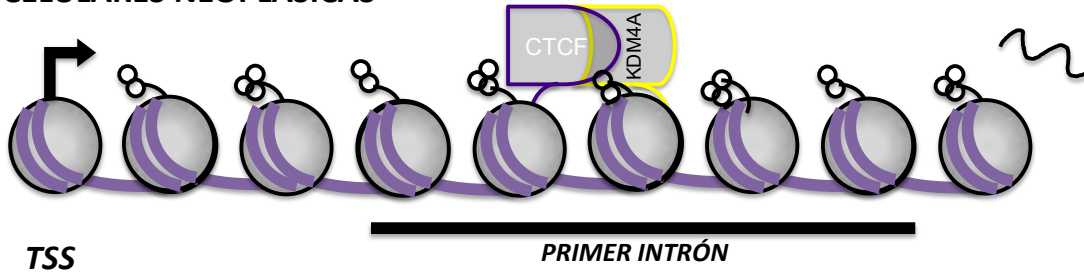
llevar a cabo la inactivación del cromosoma X en hembras (Sun et al., 2013). Curiosamente, observamos un complejo de proteínas formado por CTCF-KDM4A, que se encuentra en el primer intrón de *CHD5*. Cuando evaluamos la co-ocupancia de KDM4A y CTCF, demostramos que la línea celular HeLa exhibe un mayor porcentaje de co-ocupancia en comparación con la línea celular MCF7. Nuestros resultados sugieren que KDM4A actúa como un represor transcripcional cuando está en complejo con CTCF. La pérdida de KDM4A en el primer intrón de *CHD5* restaura la marca de histonas H3K36me3 y aumenta la expresión de *CHD5*. Así mismo la pérdida de KDM4A afecta los niveles globales de H3K36me3 y H3K36me2, estos resultados proponen que los cambios en las MPT de las histonas son tanto a nivel local como global, adicionalmente ensayos de ChIP-seq y ChIP-Re-ChIP-seq revelaron que KDM4A y CTCF y el complejo CTCF-KDM4A, se encuentran no solo en el primer intrón de *CHD5*, también se encontraron en ocho sitios más a lo largo de *CHD5*, en la región promotora de *WRAP53*, en la región situada a -1922 pb del TSS de *ASCL2* y los exones 20 y 21 de *JAG2*. Lo cual propone que el complejo formado por CTCF-KDM4A puede regular otros genes.

Por lo tanto, proponemos un nuevo mecanismo de represión transcripcional mediado por el complejo formado por KDM4A y CTCF, este complejo altera la marca de elongación de la transcripción H3K36me3 en el primer intrón de *CHD5*, y la pérdida de KDM4A en este intrón reestablece los niveles de H3K36me3 así como la expresión de *CHD5*, es importante destacar que este mecanismo nuevo de regulación no implica el silenciamiento del promotor de *CHD5* (Figura 15). Hasta la fecha, se desconoce si este complejo está relacionado con la represión de otros genes, y cuáles podrían ser las implicaciones de este complejo en distintas enfermedades como el cáncer. Otra pregunta que queda abierta es saber si el complejo aquí reportado está formado por otras proteínas, así como saber si CTCF y KDM4A tienen una interacción física directa.

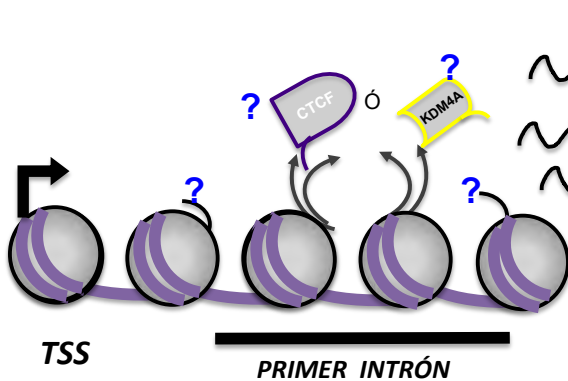
LÍNEA CELULAR NO NEOPLÁSICA: MCF 10A



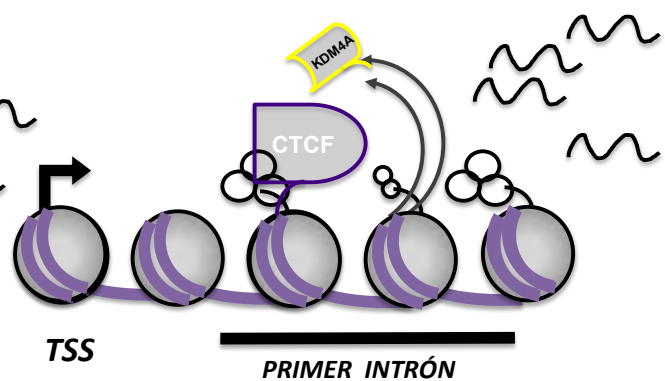
LÍNEAS CELULARES NEOPLÁSICAS



**LÍNEAS CELULARES NEOPLÁSICAS
KDM4A ó CTCF KD**



**MCF7
KDM4A^{KO}**



- H3K36me3 H3K36me3 (disminución)
- H3K36me2 H3K36me2 (disminución)

mRNA de *CHD5*

Figura 15: Modelo de represión transcripcional de *CHD5* mediado por el complejo proteico CTCF-KDM4A.

El complejo proteico CTCF-KDM4A se recluta al primer intrón de *CHD5* y promueve la desmetilación de la histona H3K36me. En células no neoplásicas, CTCF se localiza en el primer intrón de *CHD5* y H3K36me3 / 2 se encuentran enriquecidas. Estos eventos correlacionan con la expresión de *CHD5*. Por el contrario, en células neoplásicas, el complejo proteico CTCF-KDM4A promueve la desmetilación de H3K36me3 / 2 esto conduce a la represión génica. El abatimiento de CTCF o KDM4A knockdown (KD) reactiva la expresión de *CHD5*. La pérdida de KDM4A en células KDM4A^{KO} permite el restablecimiento de la marca de histonas H3K36me3 en el primer intrón y a la reactivación de la expresión génica de *CHD5*.

13 CONCLUSIONES.

- La presencia de CTCF y KDM4A en el primer intrón de *CHD5* está asociada con la disminución de H3K36me3 y H3K36me2.
- CTCF y KDM4A forman un complejo.
- El complejo CTCF-KDM4A regula negativamente la expresión de *CHD5*.
- El mecanismo de represión mediado por el complejo CTCF-KDM4A es independiente de la metilación del DNA del promotor de *CHD5*.
- La disminución de CTCF o KDM4A aumenta los niveles de expresión de *CHD5*.
- El “knockout” de KDM4A restablece las marcas de histonas H3K36me3 y H3K36me2 en el primer intrón de *CHD5* y permite la reactivación de la expresión de este gen.
- El “knockout” de KDM4A afecta los niveles globales de las modificaciones postraduccionales de histonas H3K36me2 y H3K36me3.
- El complejo CTCF-KDM4A se encontró en nueve sitios a lo largo de *CHD5*, en la región promotora de *WRAP53*, en la región situada a -1922 pb del TSS de *ASCL2* y en los exones 20 y 21 de *JAG2*.

14 PERSPECTIVAS.

KDM4A y CTCF juegan un papel dual en la regulación transcripcional. Conocer cuáles son las condiciones para que estas proteínas puedan activar o silenciar un gen, sería muy interesante. Con este objetivo, se evaluará la expresión global de genes mediante ensayos de RNA-seq, y se estudiará la abundancia de H3K36me3, H3K36me2, H3K9me3, H3K9me2 y la RNA polimerasa II, en células control y deficientes para KDM4A o CTCF.

Así mismo, para correlacionar la co-ocupancia de KDM4A y CTCF en distintos eventos biológicos y en las etapas de la transcripción como: el silenciamiento, activación, elongación o en la generación de variantes por corte y empalme alternativo de los genes, se integrarán distintos datos obtenidos de análisis masivos como ChIP, Chip-Re-ChIP y RNA-seq.

15 Identificación y caracterización de fármacos moduladores de la actividad enzimática de la desmetilasa de histonas KDM4A.

Resultados obtenidos durante la estancia de investigación en el laboratorio de la Dra. Lucia Altucci, localizado en la Università degli Studi della Campania Luigi Vanvitelli, Nápoles, Italia. (Fecha de la estancia: Diciembre 2017-Junio 2018).

Como se mencionó anteriormente, las desmetilasas de histonas se involucran en el inicio y la progresión del cáncer (Morera et al., 2016; Shi et al., 2004; Song et al., 2016). Distintos estudios reportan que la desregulación de las desmetilasas de histonas correlaciona con: 1) el incremento en la expresión de los oncogenes, 2) disminución de la expresión de los genes supresores de tumor, 3) alteración de la estabilidad cromosómica, y 4) interacción con los receptores hormonales. La familia KDM4 se sobreexpresa en varios tumores incluyendo cáncer de mama, próstata y linfomas (revisado en Berry y Janknecht, 2013). La desmetilasa de histonas KDM4A participa la iniciación, promoción y progresión tumoral. La sobreexpresión de esta enzima correlaciona con el puntaje de Gleason en tumores de próstata (Kim et al., 2016), también KDM4A regula genes supresores de tumores como *ARH1* y *CHD5*, y activa oncogenes como *CCND1* y *C-JUN* (Berry et al., 2012; Li et al., 2013; Mallette y Richard, 2012).

Por los antecedentes mencionados, se están comenzando a considerar a las enzimas desmetilasas de histonas como posibles blancos terapéuticos, y actualmente se buscan moléculas pequeñas o epifármacos que sean capaces de regular la actividad de estas enzimas, entre estas enzimas destaca KDM4A, la cual es una de las principales proteínas que se utilizan en el descubrimiento de epifármacos.

El desarrollo de los inhibidores para las desmetilasas de histonas se inició después de la identificación de las enzimas. El descubrimiento de los primeros fármacos se centró en productos naturales e imitaciones de α -cetoglutarato. Se ha reportado que SAHA (un inhibidor de HDAC) también mostró actividad inhibidora leve. Se han informado varios quimiotipos diferentes de inhibidores. La mayoría de los inhibidores son imitadores de α -cetoglutarato con un motivo que es capaz de quelar al metal que puede unirse al ion Fe (II) en el núcleo catalítico. Muchos inhibidores naturales, como el catecol, también contienen motivos de unión a metales. Algunos otros tipos de inhibidores (incluidos los inhibidores de eyección de iones Zn (II), inhibidores basados en péptidos, etc.) Hasta ahora, los inhibidores competitivos de α -cetoglutarato siguen siendo la estrategia principal para las enzimas KDM (revisado en Lin et al., 2018). En el caso de los compuestos aquí usados tienen la propiedad química de funcionar

como agentes quelantes, ya que KDM4A contiene un ion Fe^{2+} en su sitio activo, es probable que estos compuestos ejerzan su actividad inhibitoria mediante la quelación el ion Fe^{2+} en el sitio activo de la enzima.

En este aspecto, la línea de investigación de la Dra. Lucia Altucci se enfoca en nuevas estrategias terapéuticas contra el cáncer, el descubrimiento y desarrollo de epifármacos, así como la identificación de las alteraciones del epigenoma en procesos neoplásicos. Por ello anterior, esta estancia me permitió conocer el empleo de los distintos inhibidores de las desmetilasas de histonas, entre estos inhibidores se destaca el PKF118-310 o toxoflavina el cual causa muerte celular ya que entre otros efectos inhibe la expresión de la survivina y la actividad de KDM4A.

Actualmente las enzimas desmetilasas representan un potencial importante como nuevos blancos terapéuticos, por lo cual parte de los intereses de la estancia es el evaluar el papel de los distintos inhibidores propuestos por el grupo de la Dra. Altucci, como son el PKF118-310, FS21, FS22 y el FS23 para conocer su impacto en distintos modelos celulares (Franci et al., 2017).

15.1 OBJETIVO GENERAL.

Identificar y caracterizar la participación de los compuestos PKF118-310, FS21, FS22 y FS23 como inhibidores de KDM4A en distintos modelos celulares derivados de cáncer.

15.1.1 Objetivos particulares.

1. Evaluar la capacidad inhibitoria de los compuestos.
2. Determinar la unión de los distintos compuestos a la enzima KDM4A *in vitro*.
3. Evaluar la unión de los inhibidores con KDM4A en distintas líneas celulares *in vivo*.
4. Estudiar el efecto de los inhibidores PKF118-310, FS21, FS22, FS23 en las marcas de histonas que son blanco de KDM4A.

15.2 RESULTADOS.

15.2.1 Los compuestos PKF118-310, FS21, FS22 y FS23 inhiben a la enzima KDM4A.

Con el objetivo de corroborar la capacidad inhibitoria de los compuestos, se realizó un ensayo enzimático *in vitro* para KDM4A. Comparado con el vehículo (DMSO), los compuestos PKF118-310, cos8, cos 28 y cos49 (después renombrados FS21, FS22 y FS23) inhibieron la actividad de KDM4A, PKF118-310 inhibió a KDM4A en un 70%, y los compuestos FS21, FS22 y FS23 inhibieron en 55%, 40% y 45% respectivamente (Figura 16).

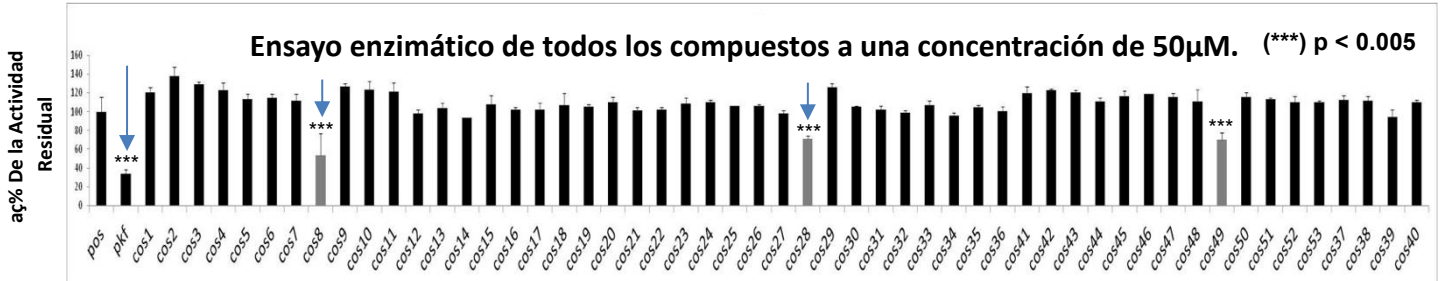


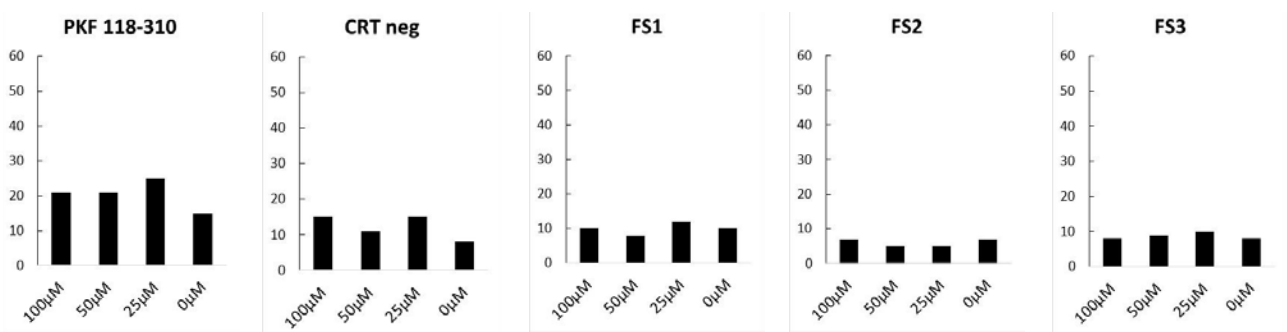
Figura 16: Los compuestos PKF118-310, FS21, FS22 y FS23 actúan como inhibidores de la enzima KDM4A.

Se muestra la actividad residual de la enzima KDM4A, después de ser incubada con todos los compuestos a una concentración de 50µM, se puede observar que el compuesto PKF118-310 inhibe la actividad de KDM4A en un 70%, FS21 tiene una acción del 55%, FS22 inhibe en un 40% y por último FS23 inhibió a KDM4A hasta en un 45%. Como prueba estadística se empleó *t* de Student. Se muestran los resultados de tres ensayos independientes.

15.2.2 Los compuestos PKF118-310, FS21, FS22 y FS23 se unen directamente a la enzima KDM4A *in vitro*.

Con el objetivo de evaluar si los compuestos PKF118-310, FS21, FS22 y FS23 eran capaces de unirse directamente a la enzima KDM4A en un modelo *in vitro*, se llevó a cabo el ensayo de Label-Free, este ensayo permite medir de manera indirecta si un compuesto se une directamente a una proteína o no (Figura 17). Se puede observar que la unión es dependiente de la dosis de los compuestos FS21, FS22 y FS23, comparando con el control negativo. Sin embargo para el compuesto PKF118-310, no se observó este comportamiento, esto se puede deber a que este compuesto podría inhibir de manera diferente a la enzima, tal vez por un metabolito secundario.

a



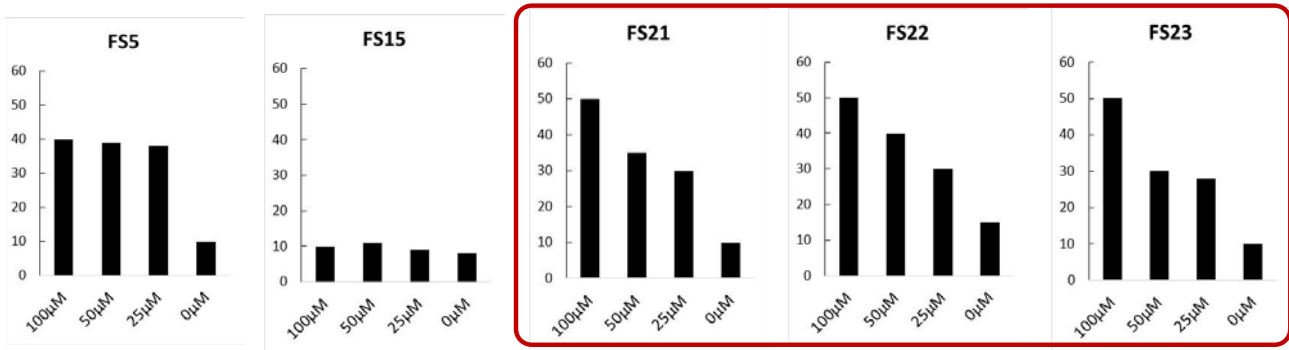


Figura 17: Los compuestos PKF118-310, FS21, FS22 y FS23 se unen a la enzima KDM4A in vitro.

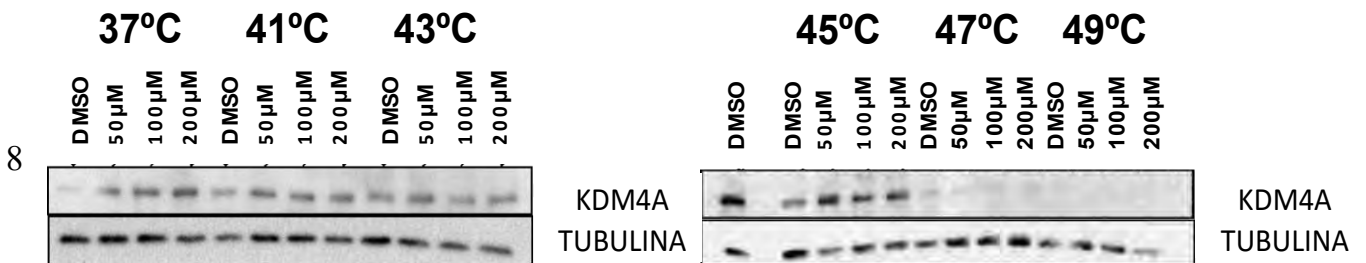
Figura 2: Los compuestos PKF118-310, FS21, FS22 y FS23 se unen a la enzima KDM4A in vitro. (A) Se muestra que la unión de los compuestos FS21, FS22 y FS23 es dependiente de la dosis ya que a mayor concentración de los inhibidores la señal aumenta. Para el caso de PKF118-310, no se observa este comportamiento ya que a mayor concentración de los inhibidores la señal aumenta. Para el caso de PKF118-310, no se observa este comportamiento, quizá la acción inhibitoria de este compuesto sea mediante un metabolito secundario.

15.2.3 El compuesto FS23 se une directamente a KDM4A in vivo.

Con el objetivo de evaluar la unión directa entre el compuesto FS23, fármaco que teníamos disponible en ese momento, y KDM4A, se llevaron a cabo ensayos de cambio térmico celular, CESTA (Cellular Thermal Shift Assay). Este ensayo permite retardar o prevenir la

degradación térmica de la proteína blanco como resultado de su estabilización por unión del KDM4A, se llevaron a cabo ensayos de cambio térmico celular, (cellular thermal shift assay CESTA). Este ensayo permite retardar o prevenir la degradación térmica de la proteína blanco como resultado de su estabilización por unión del compuesto inhibidor FS23 (Figura 18A). Específicamente, observamos la estabilización de esta enzima en

las temperaturas 43°C y 45°C (Figura 18B). Estos resultados confirman la interacción directa del compuesto FS23 con la enzima KDM4A. Nuestros hallazgos muestran la estabilización de la enzima a una mayor temperatura en comparación con la temperatura ambiente en respuesta a la presencia del inhibidor FS23 (Fig. 3A). Específicamente, observamos la estabilización de esta enzima en las temperaturas 43°C y 45°C (Fig. 3B). Estos



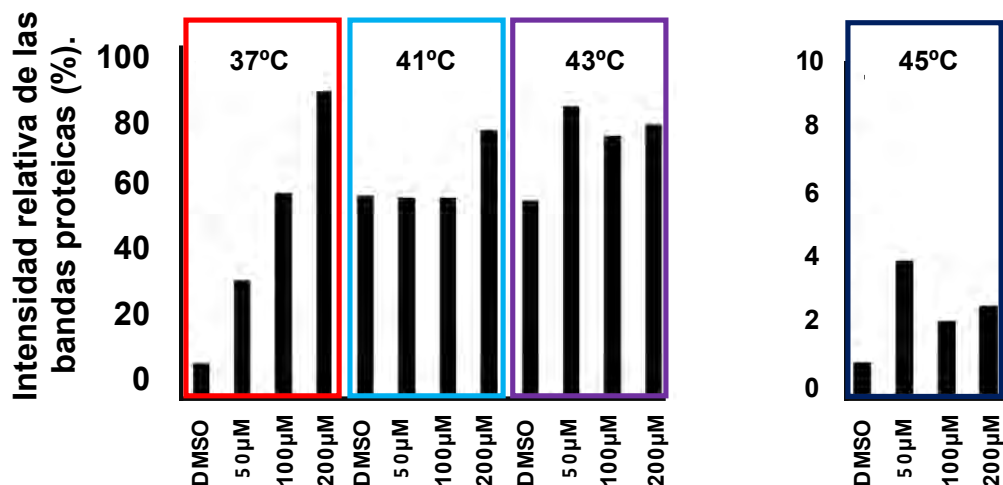
B

Figura 18: El compuesto FS23 se une directamente a KDM4A.

(A) Evaluación del ensayo de cambio térmico celular por medio de Western Blot a tres distintas concentraciones. (B) Cuantificación de la intensidad relativa de las bandas proteicas, muestra que hay una mayor estabilización de KDM4A con la concentración 50 μM, en las temperaturas 43°C y 45°C en comparación con el vehículo. La cuantificación de la abundancia relativa de KDM4A se realizó usando como proteína control a la TUBULINA. Se muestran los resultados de una réplica biológica.

15.2.4 La inducción con los inhibidores afecta la abundancia de las marcas de las histonas H3K9me3, H3K9me2 y H3K9me1.

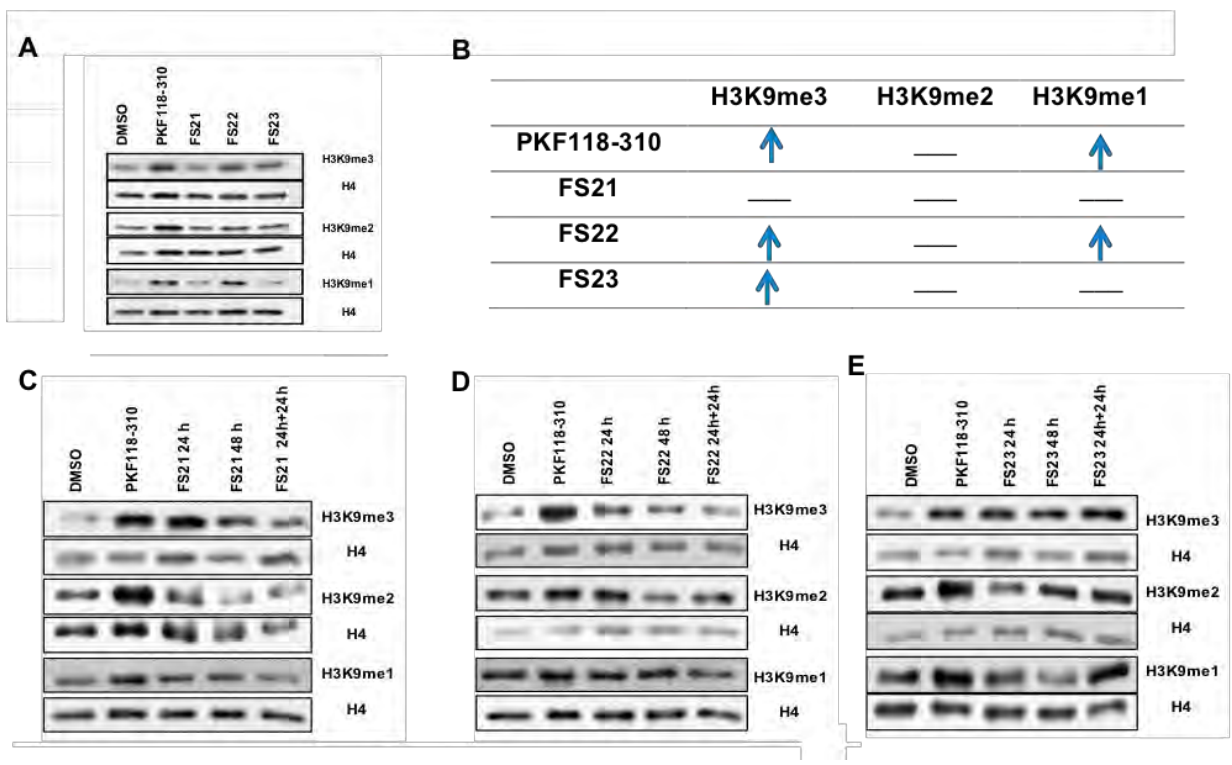
Con el objetivo de investigar cuales son los efectos de los inhibidores en las modificaciones postraduccionales de las histonas. Se estimularon las células con los distintos compuestos y posteriormente se extrajeron las histonas, después se evaluaron los niveles de las marcas H3K9me3, H3K9me2 y H3K9me, por medio de Western Blot. Como primera aproximación, se probó la inducción de los compuestos PKF118-310, FS21, FS22 y FS23 en las células de cáncer de colon HCT-116, durante 24 horas con una concentración 50 μM (Figura 19A). Los resultados revelaron que después del tratamiento con PKF118-310, FS22 y FS23, la abundancia de la modificación H3K9me3 aumentaba en comparación con las muestras que se trataron con el vehículo (Figura 19B).

Posteriormente, con el fin de probar estos inhibidores en otro contexto celular, se probó la inducción de los compuestos FS21, FS22 y FS23 en las células de cáncer de colon HCT-116, durante 24h, 48h y 24h+24h (compuesto añadido nuevamente 24 h después de la estimulación inicial) con una concentración de 50 μM, como control positivo se empleó el

compuesto PKF118-310 a una concentración de 10 μ M por 24 horas (Figuras 19C, 19D, 19E). Se encontró que la modificación H3K9me3 aumentaba en las células tratadas con los inhibidores FS21 y FS23 tanto en 24h, 48h y 24h+24h todos los tiempos probados, sin embargo en el caso del inhibidor FS22 no se encontraron alteraciones en ninguna de las marcas, esto podría ser debido a que este compuesto presenta poca permeabilidad en comparación con los otros dos o puede ser que el efecto de este inhibidor sea de carácter local y no tan global como el efecto que tienen PKF118-310, FS21 y FS23 (Figura 19F).

Las concentraciones de 50 μ M para los compuestos FS21, FS22 y FS23 y de 10 μ M para el compuesto PKF118-310, se determinaron previamente en el laboratorio de la Dra. Altucci como las concentraciones inhibitorias medias o IC₅₀.

También se encontró que la modificación H3K9me1 aumentaba tras la estimulación con los compuestos FS21 y FS23 en el tiempo 24h+24h, esta marca no es sustrato para KDM4A, esto se podría explicar debido a que estos compuestos son capaces de inhibir a otras enzimas diferentes a KDM4A (Figura 19F).



F		H3K9me3	H3K9me2	H3K9me1
	FS21 24h	↑	—	—
	FS21 48h	↑	—	—
	FS21 24+24h	↑	—	↑
	FS22 24h	—	—	—
	FS22 48h	—	—	—
	FS22 24+24h	—	—	—
	FS23 24h	↑	—	—
	FS23 48h	↑	—	—
	FS23 24+24h	↑	—	↑

Figura 19: Las modificaciones postraduccionales de las histonas H3K9me3, H3K9me2, H3K9me1, se ven afectadas después del tratamiento con los inhibidores en la línea celular HCT-116.

Evaluación de la abundancia de las modificaciones postraduccionales H3K9me3, H3K9me2, H3K9me1 mediante Inmuno Blots. (B) Tabla que resume el comportamiento de las marcas de histonas después del tratamiento con los inhibidores, se observa que tanto la H3K9me3 y H3K9me1 aumentan después de la inducción con los compuestos PKF-118-310 y FS22. En el caso de la inducción con el compuesto FS21 no se observó ningún cambio en ninguna de las MPT de las histonas, por último, se observó un aumento de H3K9me3 tras la inducción con el compuesto FS23. (C, D, F) Evaluación de la abundancia de las modificaciones postraduccionales H3K9me3, H3K9me2, H3K9me1 mediante Westen Blots, después del tratamiento con 50,µM de los compuestos FS21 (C), FS22 (D) y FS23 (E) por 24h, 48h y 24h+24h (adición del compuesto 24 horas después del estímulo inicial), como control se empleó el compuesto PKF118-310 a una concentración de 10µM por 24 horas. (F) Tabla que resume el comportamiento de las marcas de histonas después del tratamiento con los inhibidores. Se observa que la MPT H3K9me3 aumenta después de la inducción con los compuestos FS21 y FS23 por 24 horas, por 48 horas y con una segunda inducción por 24 horas (24+24 horas). Así mismo para la MPT H3K9me1 se observó un aumentó con los compuestos FS21 y FS23 por 24 horas y con una segunda inducción por 24 horas (24+24 horas). La cuantificación de la abundancia relativa de las MPTs se realizó usando como proteína control a la histona H4 total. Se muestran los resultados de una réplica biológica.

Con el objetivo de evaluar si existía alguna modificación en la abundancia de las marcas de histonas en otros contextos celulares así como en modelos de diferentes tipos de cáncer, se procedió a inducir las líneas celulares MCF7 y PANC- 1 con 50µM de los inhibidores FS21, FS22 y FS23 durante 24h, 48h y 24h+24h, es decir se administra por segunda vez el fármaco 24 horas después de la primera administración, posteriormente se extrajeron las histonas y se evaluaron los niveles de las modificaciones H3K9me3, H3K9me2 y H3K9me1 mediante Inmunoblots (Figura 20A y 20B). Se emplearon las líneas celulares MCF7 y PANC-1. Los resultados revelaron que en el caso de la línea celular MCF7 (Figuras 20A, 20C y 20D) las modificaciones H3K9me2 y H3K9me1 no presentan algún cambio aparente en comparación

mediante inmunoblots. (B) Tabla que resume el comportamiento de las marcas después del tratamiento con los inhibidores. (C, D, F) Evaluación de la abundancia de las marcas después del tratamiento con los inhibidores. (C, D, F) Evaluación de la abundancia de las marcas postraduccionales H3K9me3, H3K9me2, H3K9me1 mediante Western Blots, después del tratamiento con 50mM de los compuestos FS21 (C), FS22 (D) y FS23 (E) por 24h, 48h y 24h+24h. (F) Evaluación de la abundancia de las marcas postraduccionales H3K9me3, H3K9me2, H3K9me1 mediante Western Blots, después del tratamiento con 50mM de los compuestos FS21 (C), FS22 (D) y FS23 (E) por 24h, 48h y 24h+24h. (F) Tabla que resume el comportamiento de las marcas postraduccionales H3K9me3, H3K9me2, H3K9me1 mediante Western Blots, después del tratamiento con los inhibidores. (F) Tabla que resume el comportamiento de las marcas postraduccionales H3K9me3, H3K9me2, H3K9me1 mediante Western Blots, después del tratamiento con los inhibidores.

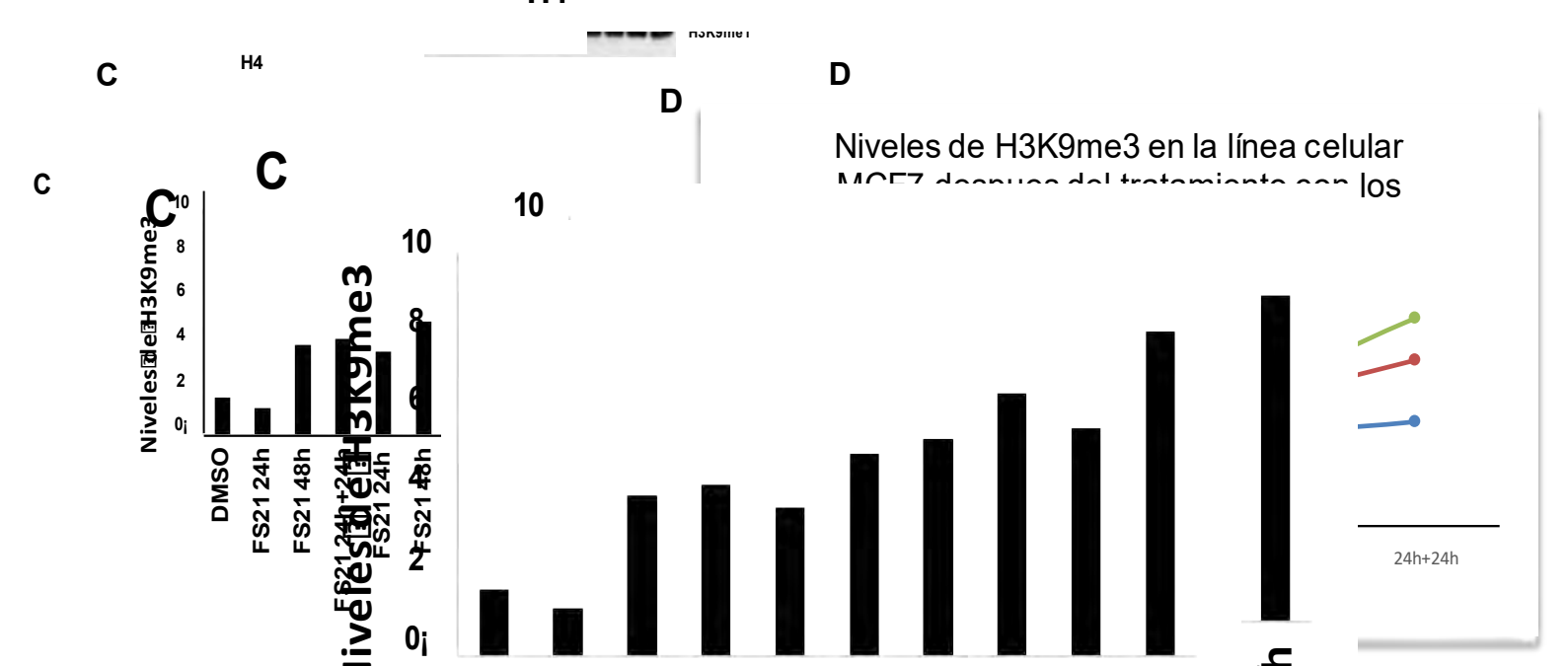
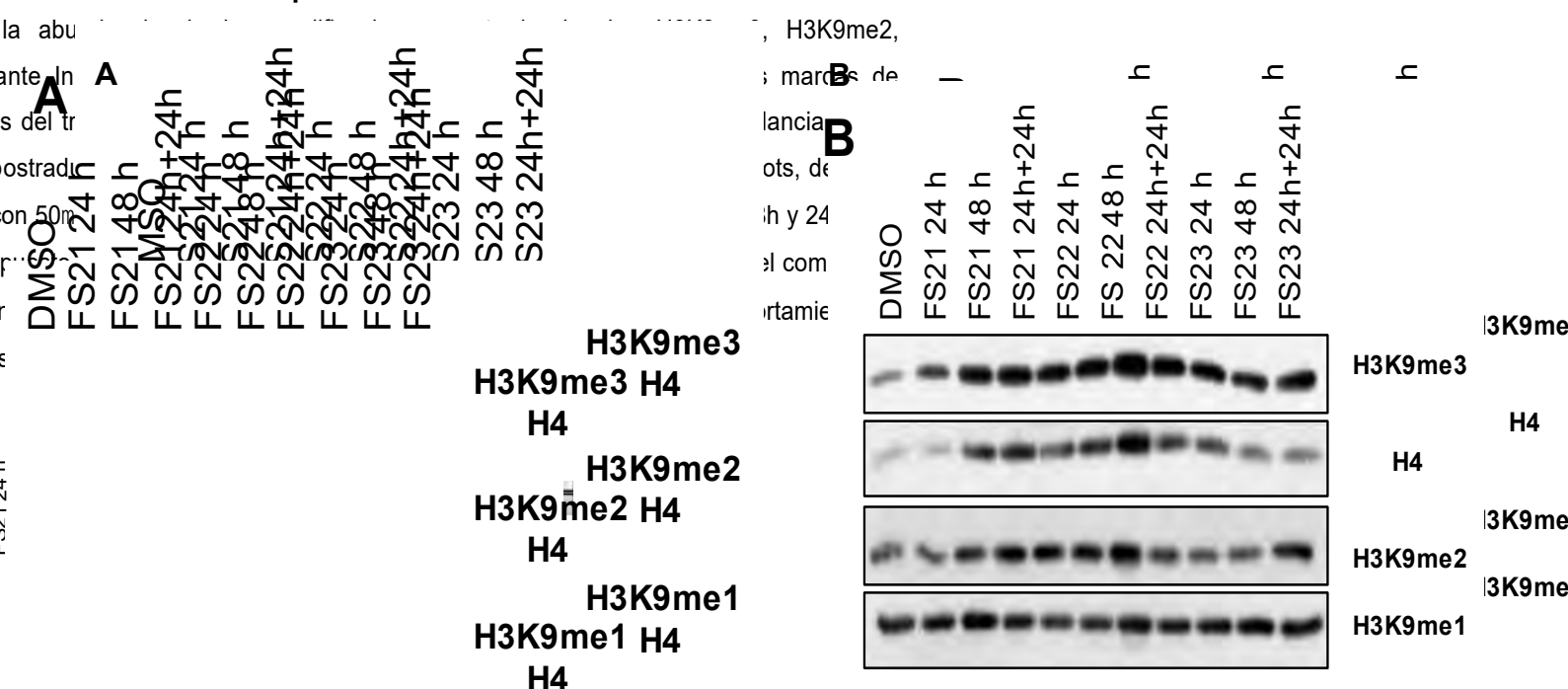


Figura 2. Efecto de los inhibidores de histonas en las líneas celulares MCF7 (A) y en la línea celular PANC-1 (C). (A) Western Blots de la marca H3K9me3, H3K9me2, H3K9me1 y H4 en la línea celular MCF7. (B) Western Blots de la marca H3K9me3, H3K9me2, H3K9me1 y H4 en la línea celular PANC-1. (C) Cuantificación de la intensidad relativa de las bandas proteicas de la marca H3K9me3 en la línea celular MCF7. (D) Gráfico que muestra el comportamiento de la marca H3K9me3 en la línea celular PANC-1.

Figura 20: Evaluación de las modificaciones postraduccionales de las histonas H3K9me3, H3K9me2, H3K9me1 después del tratamiento con los inhibidores en las líneas celulares MCF7 y PANC-1.

(A y B) Evaluación de la abundancia de las modificaciones postraduccionales H3K9me3, H3K9me2, H3K9me1 mediante Inmuno Blots en la línea celular (A) MCF7 y en la línea celular (B) PANC- 1. (C) Cuantificación de la intensidad relativa de las bandas proteicas de la marca H3K9me3 en la línea celular MCF7. (D) Grafica que muestra el comportamiento de la marca H3K9me3 en la línea celular MCF7 a través del tiempo y dependiendo del compuesto empleado. Se muestran los resultados de una réplica biológica.

15.2.5 La inducción con el epifármaco PKF118-310 de las células KDM4A^{KO} y WT afecta los niveles globales de H3K9me3, H3K9me2, H3K9me1, H3K36me3 y H3K36me2.

Uno de los objetivos que formaban parte de mi estancia realizada con la Dra. Lucia Altucci, era aprender acerca de los llamados epifármacos. Los epifármacos son moléculas pequeñas que son capaces de regular la actividad enzimática de proteínas que se encargan de remodelar la cromatina, entre estas proteínas destaca KDM4A, la cual es una de las principales enzimas utilizadas en el descubrimiento de epifármacos.

Con el objetivo de evaluar cual era el efecto del epifármaco PKF118-310, fármaco que teníamos disponible en ese momento, sobre las MPT H3K9me3, H3K9me2, H3K9me1, H3K36me3 y H3K36me2 en otros contextos celulares diferentes a los usados anteriormente. Se indujeron células MCF7 WT y MCF7 KDM4A^{KO} con 0.8 μ M de PKF118-310 por 24 horas así como el vehículo, DMSO. Posterior al tiempo de inducción se realizó una extracción ácida de histonas, y se evaluó la abundancia de las marcas de histonas H3K9me3, H3K9me2, H3K9me1, H3K36me3 y H3K36me2 por medio de inmunoblots (Figura 21A). El análisis de la abundancia relativa de las modificaciones H3K9me1, H3K9me2, H3K9me3 H3K36me2 (Figura 21B) arrojó que no hay cambios en la abundancia de estas MPT. Para el caso de H3K36me3 se observó que esta marca aumenta en las células WT tratadas con el fármaco, las células KDM4A^{KO} tratadas con DMSO y con el fármaco respectivamente en las células tanto WT como KDM4A^{KO} tratadas con el fármaco PKF118-310 con respecto a las células WT tratadas con el vehículo (Figura 21B).

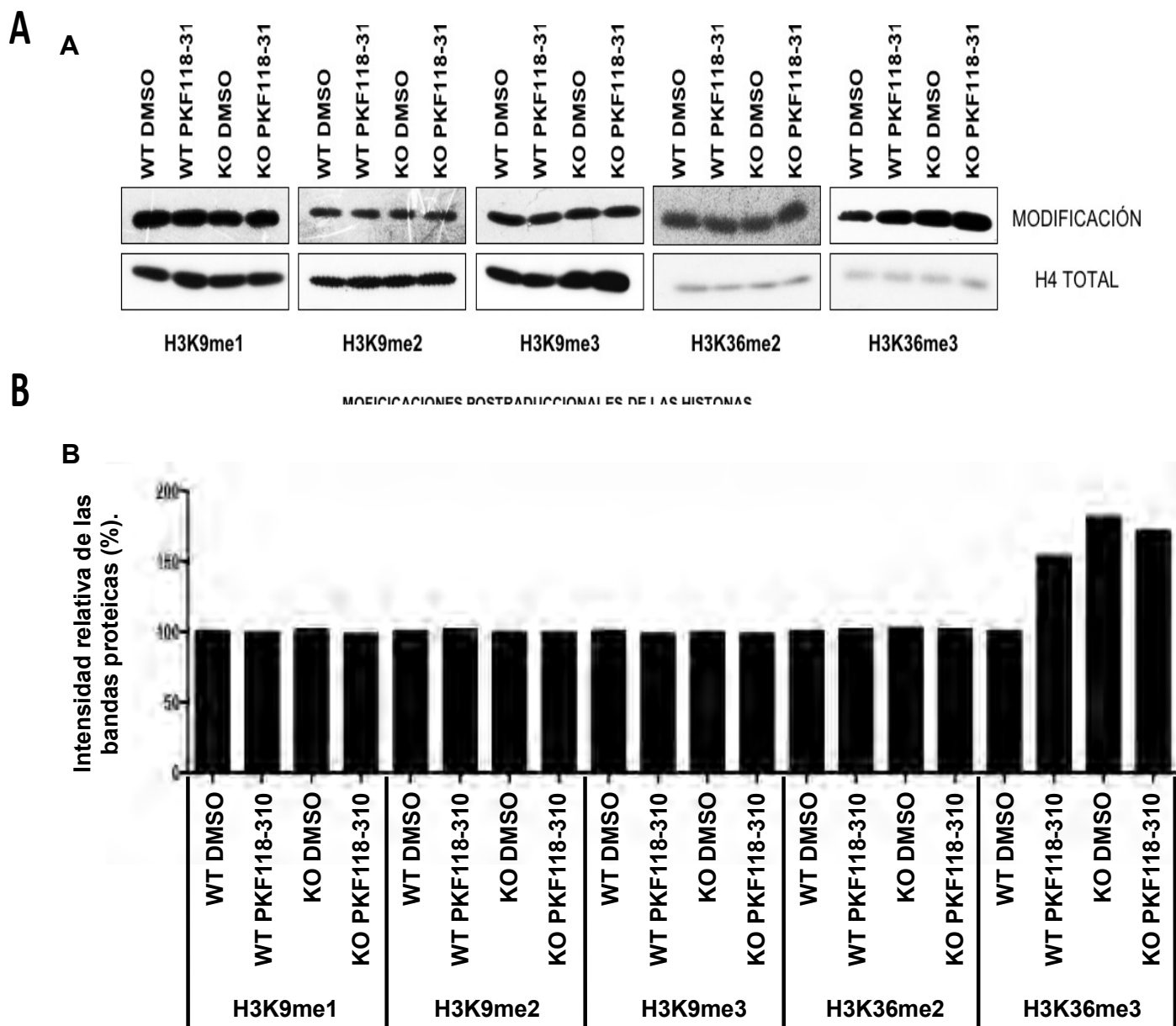


Figura 21: La inducción con el epifármaco PKF118-310 afecta los niveles globales de H3K36me3 en las células WT y KDM4A^{KO}.

(A) Caracterización de la abundancia de las marcas de histonas H3K9me1, H3K9me2, H3K9me3, H3K36me2 y H3K36me3 mediante inmunoblots, se cargaron 8µg de extracto histónico para cada condición y se normalizó con la abundancia de la histona H4 total. (B) La cuantificación de la intensidad relativa de las bandas proteicas muestra que las marcas H3K9me1, H3K9me2, H3K9me3 y H3K36me3 no presentan cambios significativos en las células WT tratadas con el fármaco, las células KDM4A^{KO} tratadas con DMSO y con el fármaco respectivamente. Finalmente, se observó que la modificación de histonas H3K36me3 aumenta en 53%, 81% y 71% para las células WT tratadas con el fármaco, las células KDM4A^{KO} tratadas con DMSO y con el fármaco respectivamente. Se muestran los resultados de una réplica biológica.

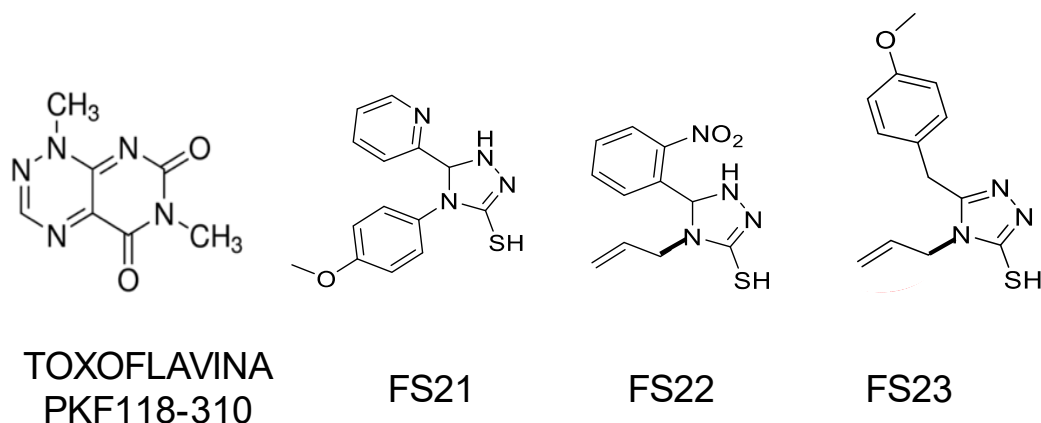


Figura 22: Estructura de los inhibidores empleados.

Esquema de la estructura molecular de los inhibidores usados en este estudio. Es probable que la presencia de grupos aromáticos y cíclicos sea la responsable de la acción inhibitoria.

15.3 DISCUSIÓN.

Las proteínas que reconocen, establecen y remueven las modificaciones postraduccionales de las histonas, son las mediadoras en condiciones fisiológicas y de enfermedad. El poder modular la actividad de estas proteínas es el objetivo de los llamados epifármacos. Dentro de los primeros epifármacos descritos se encuentran a los inhibidores de las desacetilasas de histonas (HDAC) los cuales ya están uso clínico (revisado en Conte et al., 2018). Por el contrario, el entendimiento molecular de las proteínas que se encargan de regular la desmetilación de las lisinas que se encuentran en las histonas, se ha visto rezagado (Franci et al., 2017). Actualmente se buscan moléculas pequeñas que sean capaces de regular la actividad de las enzimas desmetilasas de histonas, entre estas enzimas destaca KDM4A, la cual es una de las principales proteínas que se utilizan en el descubrimiento de epifármacos (Franci et al., 2017; Hoffmann et al., 2012). Por lo anterior uno de los objetivos de este proyecto fue identificar y caracterizar la participación de nuevos compuestos como inhibidores de KDM4A.

Como parte de los antecedentes directos, en el laboratorio de la Dra. Altucci se llevó a cabo un screening *in silico*, a partir de este estudio se seleccionaron varios candidatos posibles para inhibir a KDM4A entre estos se encontró PKF118-310 o toxoflavina, así como los compuestos denominados FS21, FS22 y FS23, cuyas estructuras se presentan en la figura 22. Es probable que la acción inhibitoria de estos compuestos sea mediante la quelación el ion Fe^{2+} que se encuentra en el sitio activo de la enzima.

Los compuestos PKF118-310, FS21, FS22 y FS23 revelaron que son inhibidores de la enzima KDM4A mediante ensayos enzimáticos, además mediante el ensayo de Label-Free, se determinó que estos compuestos FS21, FS22 y FS23 se unen directamente a la enzima KDM4A *in vitro*, no se observó el mismo comportamiento para el compuesto PKF118-310, podría ser que la acción de este compuesto sea mediante algún metabolito secundario.

Los resultados obtenidos del ensayo de CETSA, revelaron la estabilización de KDM4A en las temperaturas 43°C y 45°C. Estos resultados confirman la interacción directa entre proteína blanco KDM4A y el compuesto FS23.

También nos enfocamos en el efecto de los inhibidores sobre la abundancia de las modificaciones de las histonas blanco de KDM4A, es decir las marcas H3K9me3 y H3K9me2. Los resultados muestran que hay un incremento en la abundancia de H3K9me3 en las líneas celulares MCF7 y HCT-116. Así mismo no se encontraron diferencias después de tratar la línea celular PANC-1 con todos los compuestos, podría ser que estos compuestos no sean permeables en esta línea celular, otra opción es que el efecto de los inhibidores en este contexto celular sea de carácter local y no global como el efecto que tienen en las células MCF7 y HCT- 116.

La desregulación de las desmetilasas de histonas en diversos tipos de cáncer ha sido reportada en múltiples estudios (Berry et al., 2012; Duan et al., 2015; Franci et al., 2017, 2017; Guerra-Calderas et al., 2015; Li et al., 2013; Mallette y Richard, 2012), debido a su estructura, interacciones y actividad pleiotrópica, el papel de estas enzimas y de KDM4A en el desarrollo del cáncer puede ser más complejo de lo que originalmente se creía. Se requieren más estudios para aclarar como los compuestos aquí presentados afectan a KDM4A y a su vez afectan el desarrollo del cáncer.

Otro desafío es diseñar un fármaco que sea selectivo para una subfamilia de desmetilasas, lo cual podría lograrse vinculando el medicamento al menos a tres dominios de las enzimas; sin embargo, el medicamento diseñado podría ser demasiado grande para penetrar la membrana celular. También podría diseñarse un inhibidor alostérico que modifique la conformación del sitio catalítico de la enzima sin unirse a el enlace a este. Un estudio computacional identificó sitios alostéricos probables que podrían ser utilizados para este propósito (Campagna-Slater et al., 2010). Sin embargo, se requiere más investigación para identificar y caracterizar experimentalmente a estos sitios.

15.4 CONCLUSIONES.

- Los compuestos PKF118-310, FS21, FS22 y FS23 son inhibidores de la enzima KDM4A.
- Los compuestos PKF118-310, FS21, FS22 y FS23 se unen *in vitro* a la enzima KDM4A.
- El compuesto FS23 se une directamente a KDM4A *in vivo*.
- Las modificaciones de histonas H3K9me3, H3K9me1 se ven afectadas después de estimular las células MCF7 y HCT-116 con los inhibidores, FS21, FS22 y FS23.
- No se observaron cambios aparentes en las modificaciones H3K9me3, H3K9me2 y H3K9me1 después de la estimulación de las células PANC-1 después de la inducción con los inhibidores FS21, FS22 y FS23.
- Los compuestos FS21, FS22 y FS23 afectan los niveles globales de la marca H3K9me3 en las células MCF7.
- La inducción con el epifármaco PKF118-310 afecta los niveles globales de H3K36me3 en las líneas celulares MCF7^{WT} y MCF7^{KDM4AKO}.

15.5 PERSPECTIVAS.

Evaluar el efecto de los inhibidores PKF118-310, FS21, FS22 y FS23 sobre la expresión génica.

Determinar el impacto de los inhibidores PKF118-310, FS21, FS22 y FS23 en el reclutamiento de la desmetilasa de histonas KDM4A en distintos blancos génicos.

16 INFORMACIÓN SUPLEMENTARIA.

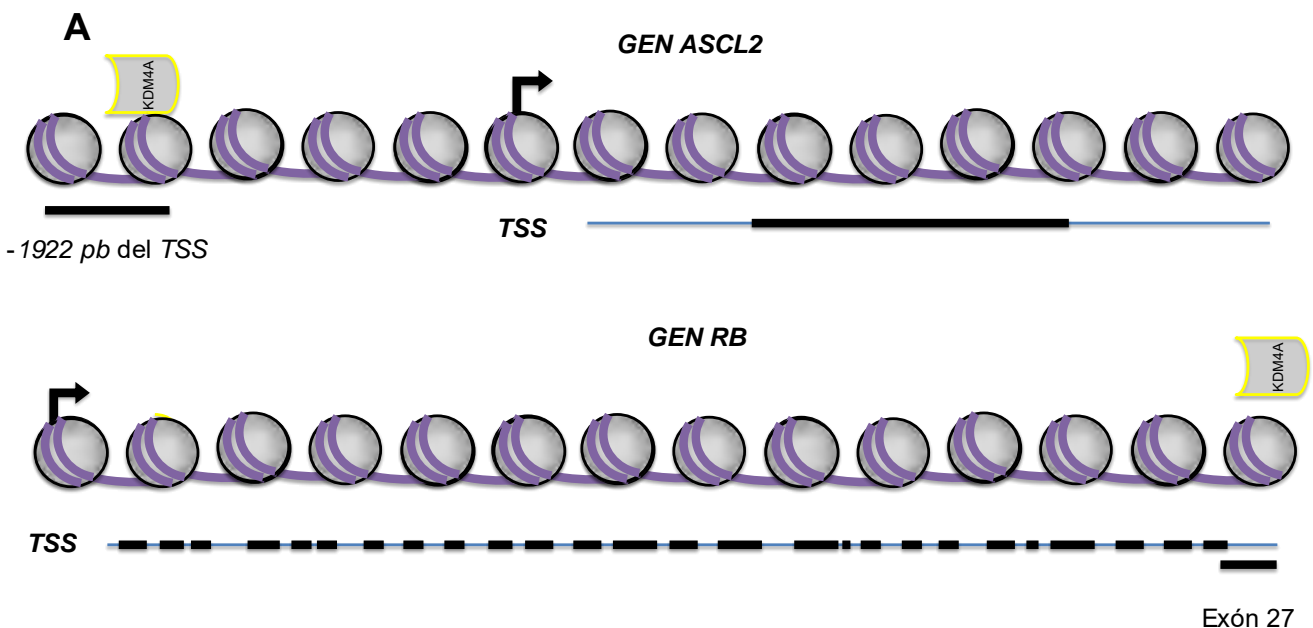
16.1 Validación de los ensayos de ChIP contra KDM4A, CTCF y H3K36me3.

Con el objetivo de validar los ensayos de ChIP para KDM4A se diseñaron, estandarizaron y evaluaron por PCR en tiempo real, los oligonucleótidos correspondientes a la región -1922 pb del TSS de *ASCL2* y los oligonucleótidos que amplifican el exón 27 de *RB*, que se emplearon como control positivo y negativo respectivamente. Los oligonucleótidos control para KDM4A fueron diseñados con base a los reportados por Mallette en el 2012. Las evaluaciones por PCR cuantitativo de las regiones control de los ChIPs de KDM4A, para las líneas celulares

neoplásicas, mostraron el comportamiento esperado, es decir hay más enriquecimiento en los controles positivos que en los negativos. Esto indica que la inmunoprecipitación es confiable. Sin embargo, en la línea celular MCF 10A el enriquecimiento de KDM4A en la región correspondiente al control negativo es mayor que el control positivo, pero ambas evaluaciones están por debajo del valor de las IgGs, que es uno, esto coincide con que KDM4A no está presente en la línea celular MCF 10A (Figura 23A).

Para los ChIPs contra CTCF, se diseñaron, estandarizaron y evaluaron por PCR en tiempo real, como control positivo el promotor de *WRAP53*, así como el control negativo el exón 27 de *RB*. Los qPCR, de los controles de las inmunoprecipitaciones CTCF mostraron mayor enriquecimiento en los controles positivos que en los negativos, esto indica que la inmunoprecipitación de este anticuerpo en esta línea celular es confiable (Figura 23B).

Para los ChIPs contra H3K36me3, se diseñaron, estandarizaron y evaluaron por PCR en tiempo real, los oligonucleótidos correspondientes a una región del segundo intrón de *GAPDH*, región en la cual H3K36me3 se encuentra enriquecida, esta región se usó como control positivo para H3K36me3. Como control negativo se empleó una región del tercer exón de *MYOG*, región en la cual H3K36me3 no se encuentra enriquecida (Figura 23C). El análisis de H3K36me3 se realizó con las bases de datos del proyecto ENCODE. Los qPCR, de los ChIPs contra H3K36me3 mostraron mayor enriquecimiento en los controles positivos que en los negativos, esto indica que la inmunoprecipitación de este anticuerpo en todas las líneas celulares es confiable (Figura 23C).



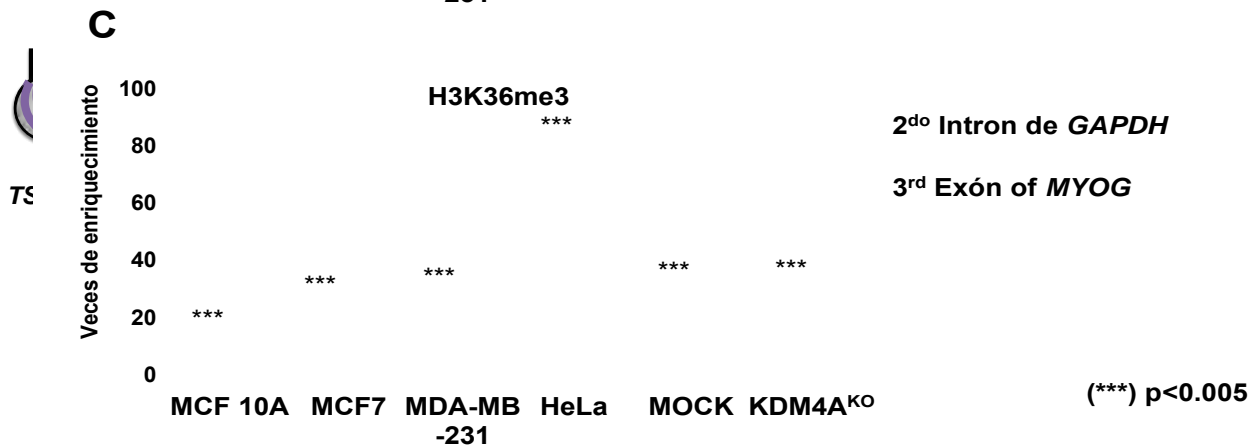
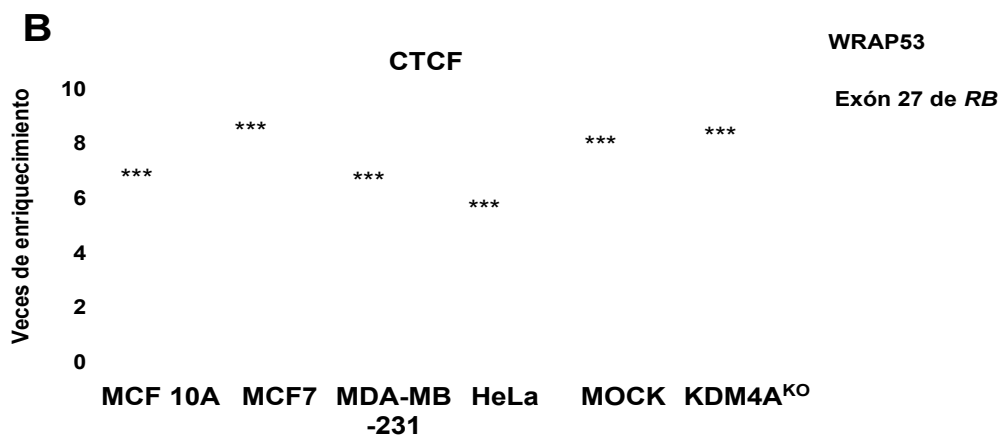
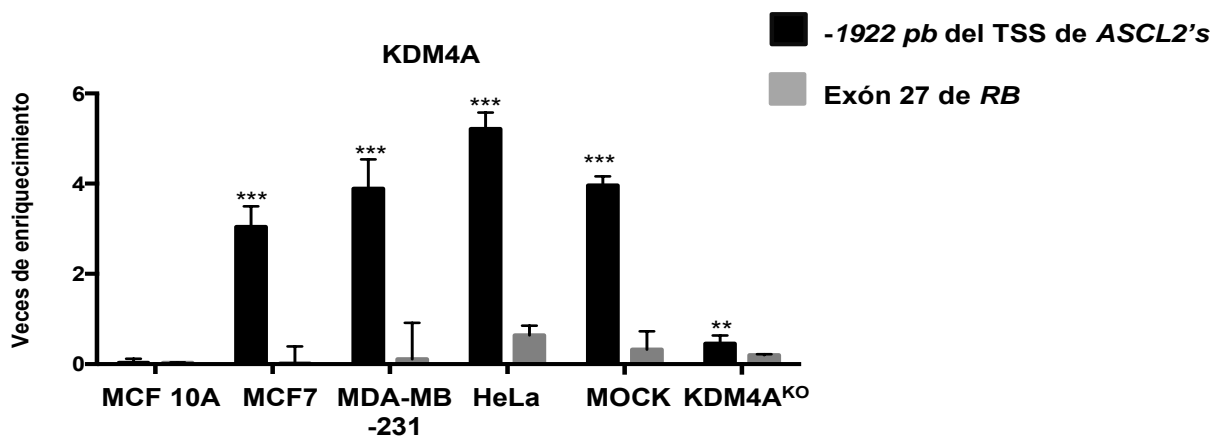
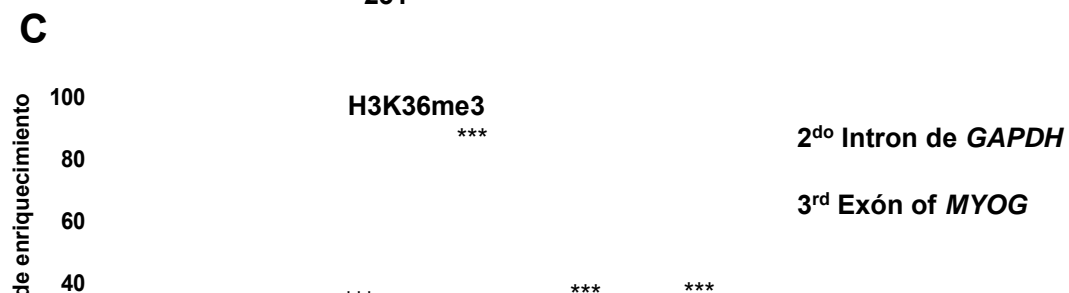
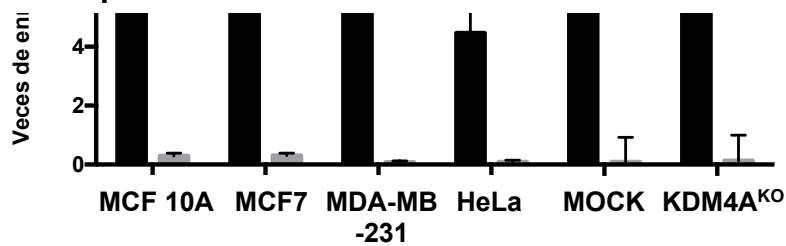
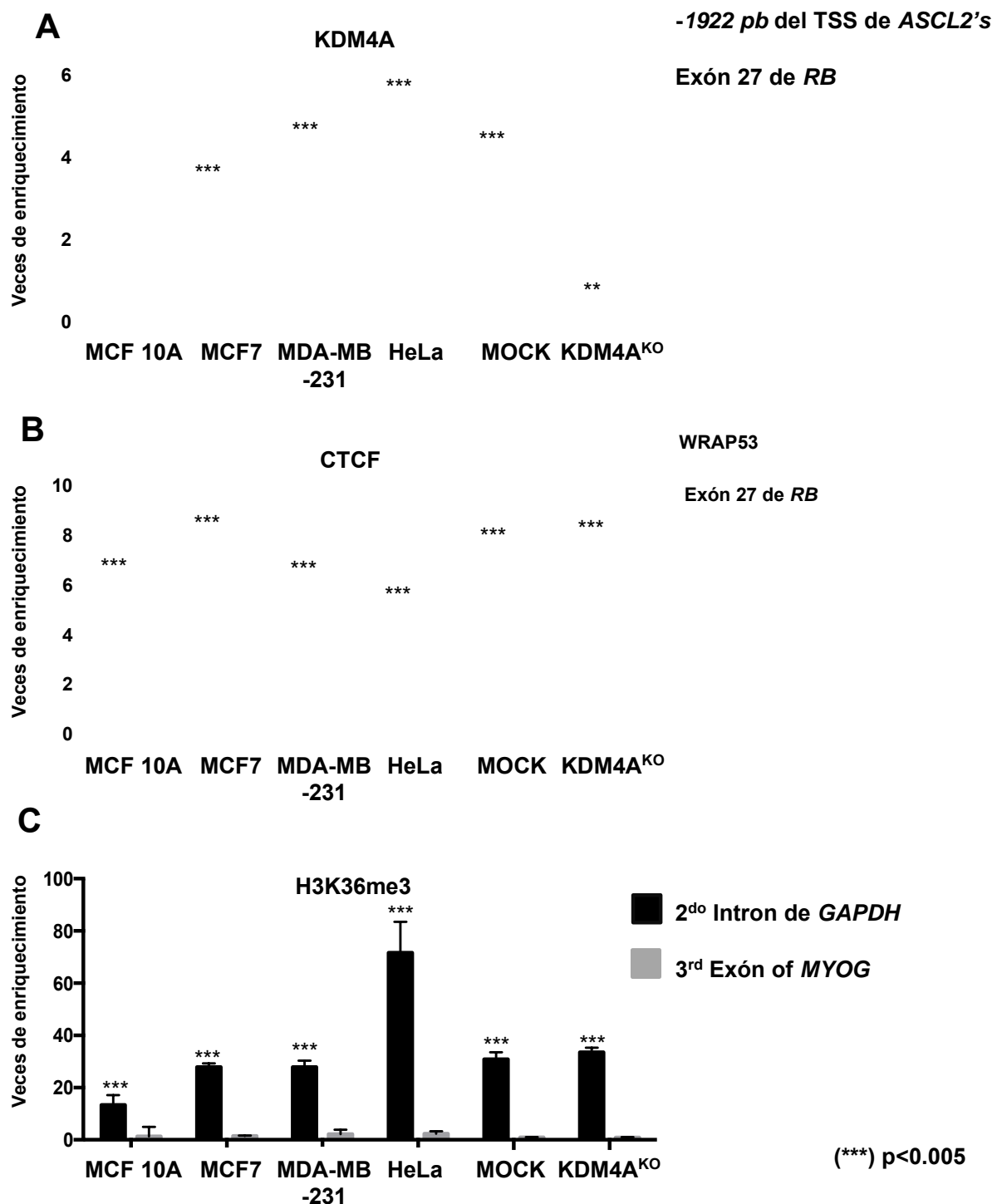


Figura suplementaria 2





Figura

Figura suplementaria 2

(A) Esquemas de los genes evaluados para controlar la inmunoprecipitación de KDM4A, evaluado por qPCR en células MCF 10A, MCF7, MDA-MB-231 y HeLa. Como control positivo para el enriquecimiento de KDM4A, utilizamos la región ubicada a -1922 pb del TSS de *ASCL2*. Como control negativo, empleamos el exón 27 de *RB*. (B) Esquemas de los genes evaluados para controlar la Inmunoprecipitación de CTCF. Análisis de los ChIPs contra CTCF evaluados por qPCR en MCF 10A, MCF7, MDA-MB-231 y células HeLa como un control positivo para el enriquecimiento de CTCF, se utilizó la región promotora de *WRAP53* y como control negativo, empleamos el exón 27 de *RB*. (C) Esquemas de los genes evaluados para controlar la Inmunoprecipitación de H3K36me3, análisis de ChIP contra H3K36me3, evaluado por qPCR en MCF 10A, MCF7, MDA-MB-231 y células HeLa. Como control positivo para H3K36me3, utilizamos el segundo intrón de *GAPDH*

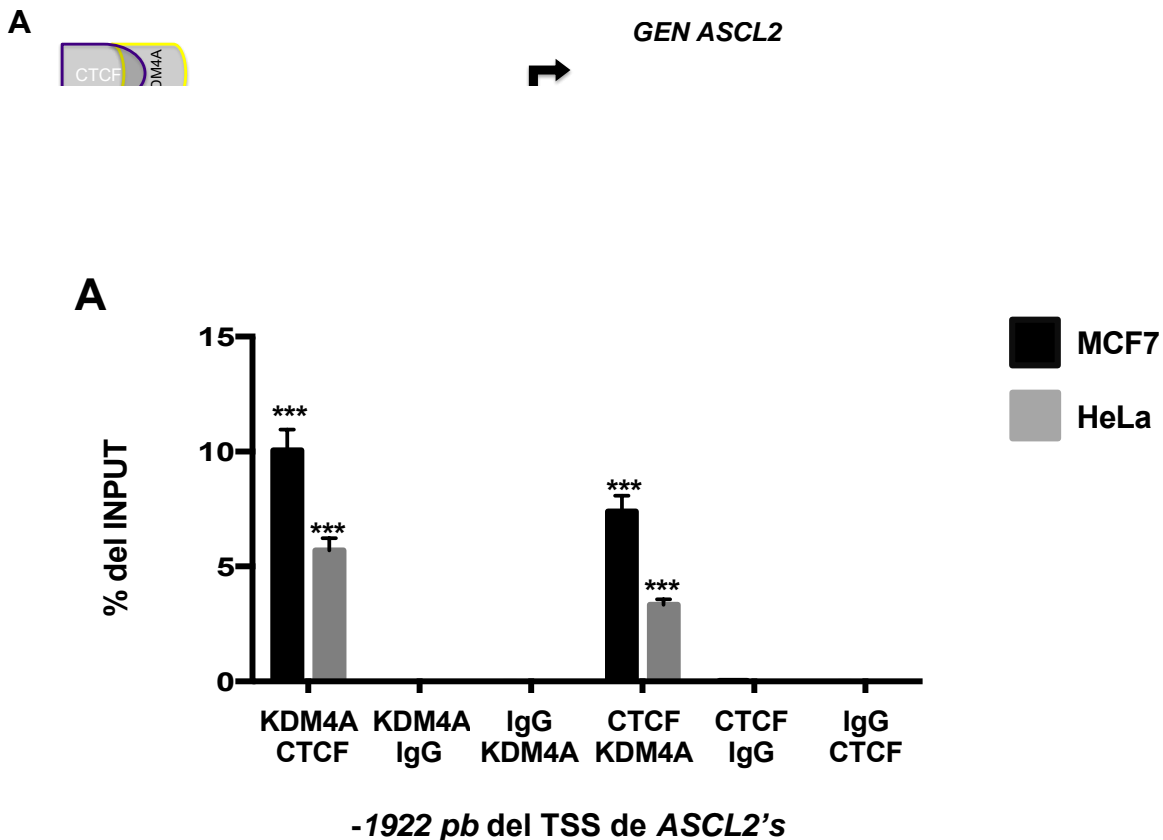
(gris) y, como control negativo, utilizamos el tercer exón de *MYOG*. Usamos el anticuerpo IgG incluido en el kit OneDay CHIP (Diagenode, NJ, EE. UU., Kch-onedIP-180), como control negativo, (**) $p < 0.01$ y (***) $p < 0.005$ en comparación con la inmunoprecipitación de el IgG. Las diferencias estadísticas se determinaron con la prueba t de Student.

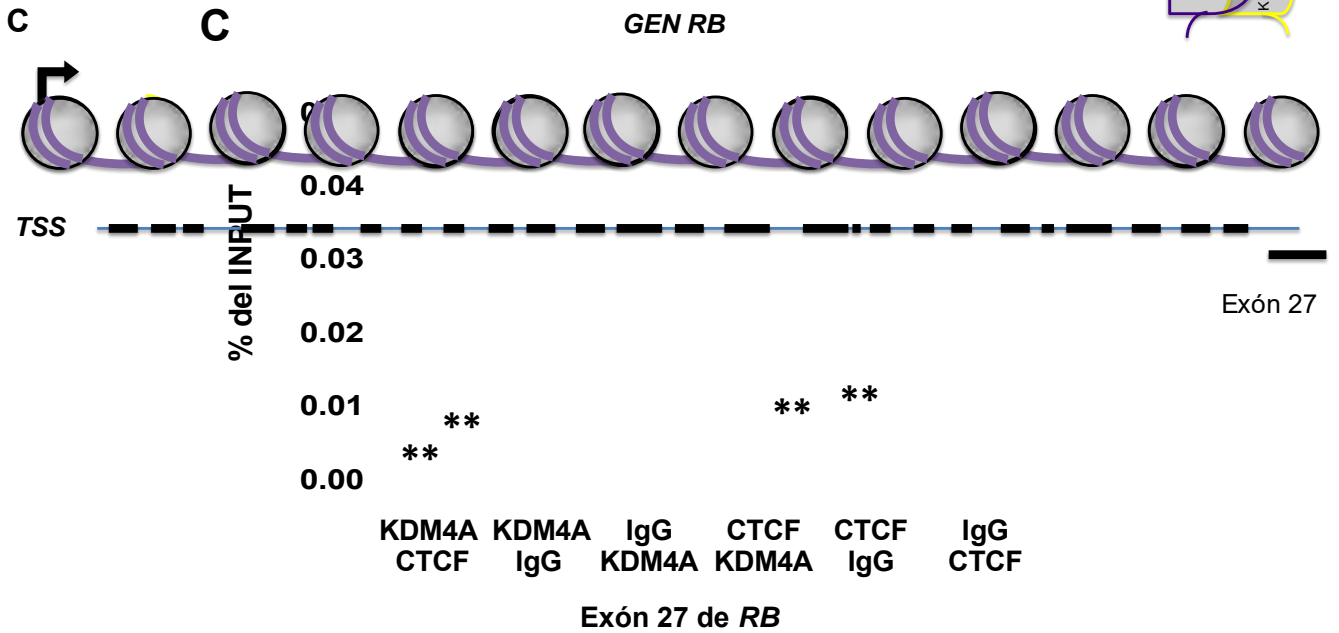
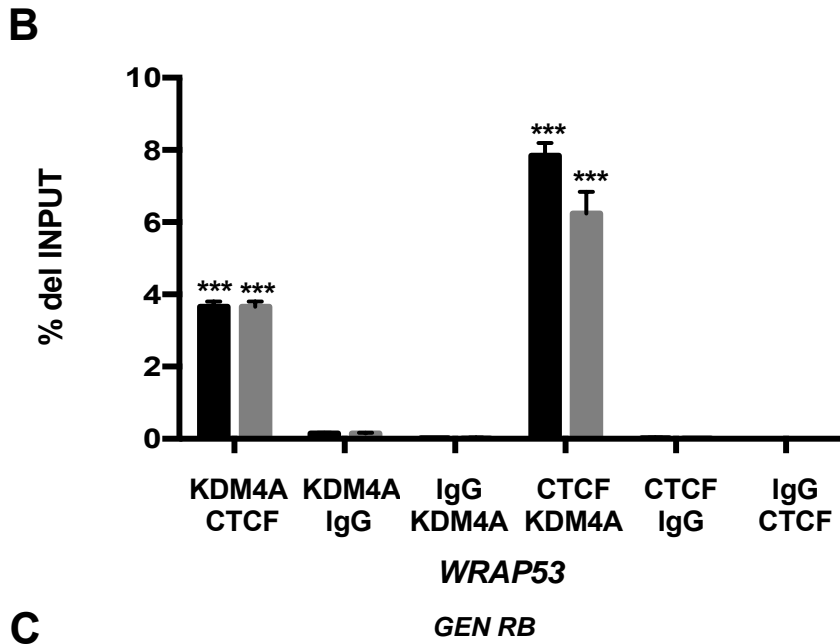
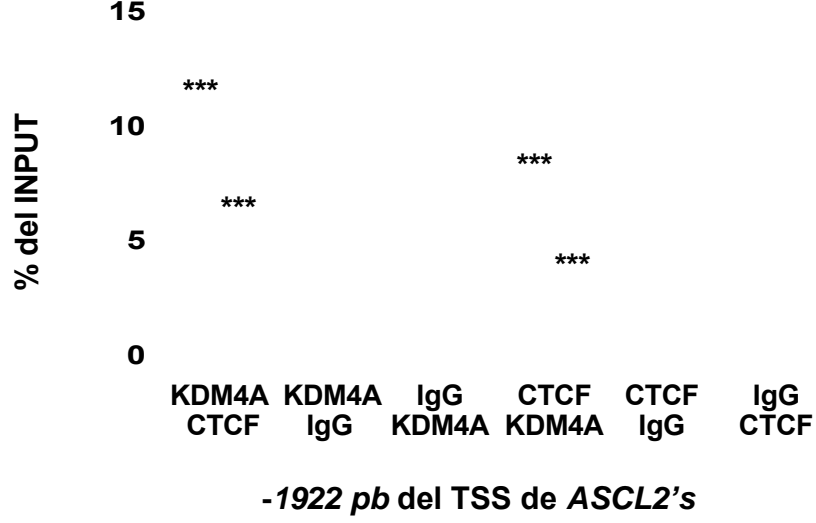
16.2 Validación del ensayo de ChIP-Re-ChIP contra KDM4A y CTCF.

Con el objetivo de validar los ensayos de ChIP-Re-ChIP, para KDM4A y CTCF, se evaluaron las inmunoprecipitaciones secuenciales en las regiones anteriormente reportadas como control positivo, para KDM4A, -1922 pb del TSS de *ASCL2* (Figura 24A), y para CTCF, *WRAP53* (Figura 24B), así como a las regiones negativas antes reportadas, para KDM4A y para CTCF, el exón 27 de *RB* (Figura 24C).

Los qPCR, de los controles de las inmunoprecipitaciones secuenciales contra KDM4A y CTCF mostraron mayor enriquecimiento en los controles positivos que en los negativos, esto nos indica que la inmunoprecipitación de este anticuerpo en esta línea celular es confiable.

De manera inesperada observamos que KDM4A coexiste con CTCF en *WRAP53*, y que CTCF también se encuentra en la región a -1922 pb del TSS de *ASCL2*, región en la cual ya se ha reportado la presencia de KDM4A, esto nos indica que KDM4A y CTCF pueden coexistir en regiones diferentes al primer intrón de *CHD5*, lo cual indica que este fenómeno no es tan localizado.





(**) p<0.01 (***) p<0.005

Figura 24: Controles del CHIP-Re-CHIP contra KDM4A y CTCF.

(A) Esquema de la región de *ASCL2* que se usó para controlar los ensayos de CHIP/re-CHIP contra KDM4A y CTCF en las líneas celulares MCF7 y HeLa. CHIP-Re-CHIPs evaluados por qPCR en la región a -1922 pb del TSS de *ASCL2* (control positivo). (B) esquema de la región de *WRAP53* que se usó para controlar la inmunoprecipitación de la cromatina secuencial contra KDM4A y CTCF. CHIP-Re-CHIPs evaluados por qPCR en la región promotora de *WRAP53* (control positivo) (C) Esquema de la región de *RB* que se usó como control negativo, el exón 27. CHIP-Re-CHIPs evaluados en células MCF7 y HeLa. Las primeras IP se realizaron con los anticuerpos mostrados en la primera fila y fueron seguidos por las IP descritas en la segunda fila. (*) $p < 0.05$, (**) $p < 0.01$ y (***) $p < 0.005$ en comparación con la inmunoprecipitación de la IgG. Las diferencias estadísticas se determinaron mediante el uso de la prueba t de Student.

16.3 Citometría de flujo y selección celular.

Con el fin de validar con otro modelo que KDM4A puede regular negativamente a *CHD5*, se realizó un modelo Knockout (KDM4A^{KO}) con el sistema CRISPR / Cas9 KO (Santa Cruz, sc-404599 y sc-404599-HDR). Este sistema empleó tres RNAs guía gRNA que se dirigieron al exón tres y ocho de *KDM4A* (Figura 25A). Posteriormente, se seleccionaron las células mediante el tratamiento con puromicina y con el objetivo de enriquecer aún más nuestra población KDM4A^{KO} se seleccionaron las células que presentaban una fluorescencia más alta por medio de citometría de flujo y FACS (Figura 25B). Como control, empleamos un plásmido de gRNA no dirigido (Mock) (Santa Cruz, sc-418922).

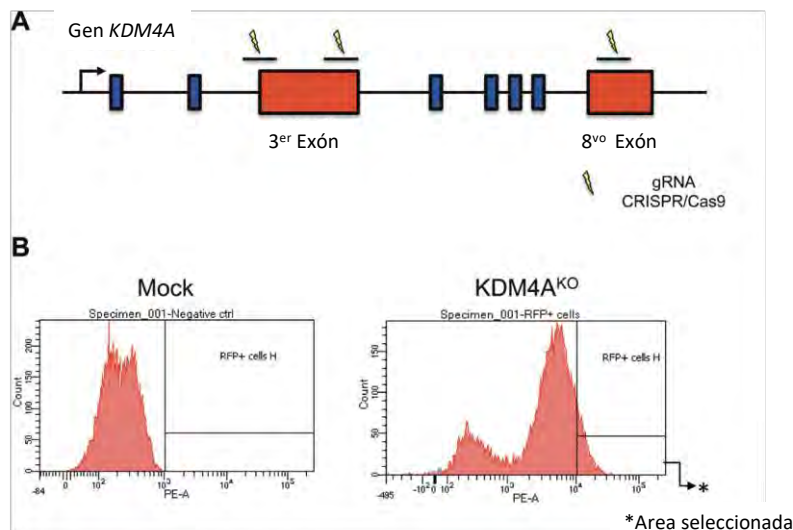


Figura 25: Citometría de flujo y selección celular.

(A) Se generaron células KDM4A^{KO} mediante el sistema CRISPR / Cas9 KO. Este sistema empleó tres gRNAs que se dirigen al exón tres y al exón ocho para el gen *KDM4A*. Como control negativo se usó un plásmido de gRNA no dirigido. (B) Las células KDM4A^{KO} se seleccionaron mediante puromicina y se determinó la fluorescencia

de RFP en células control y KDM4A^{KO}. Las células KDM4A^{KO} con las células fluorescentes más alta se seleccionaron para realizar los experimentos posteriores. (* se define el área seleccionada).

17 MÉTODOS.

17.1 EXTRACCIÓN DE RNA.

Fundamento teórico de la técnica:

El Trizol es una solución monofásica de fenol e isocianato de guanidina, que durante la homogenización o lisis celular mantiene la integridad del RNA y disuelve los componentes celulares. La posterior adición de cloroformo y subsecuente centrifugación separa la muestra en tres fases una acuosa, en la fase de en medio se encuentran las proteínas y en la fase inferior se encuentran los solventes orgánicos. La fase acuosa contiene el RNA el cual puede ser precipitado utilizando alcohol isopropílico.

1. Retirar el medio de cultivo, lisar las células directamente en el frasco de cultivo mediante la adición de 3 mL de TRIZOL para cajas de 75 cm², y mezclar varias veces con una pipeta, con el fin de que las células se despeguen en su totalidad.
2. Trasladar el TRIZOL con células a un tubo Falcon ® de 15 mL, incubarlas a temperatura ambiente (15°C-30°C) por 5 minutos.
3. Agregar 0.2 mL de cloroformo por cada mL de TRIZOL agregado, agitar vigorosamente los tubos con la mano durante 15 segundos. Lo anterior se realiza con el fin de limpiar el RNA.
4. Centrifugar las muestras a 3,200 x g durante 15 minutos, a una temperatura de 2-8°C. Después de la centrifugación la mezcla se divide en tres fases: superior: fase acuosa; media: fase fenol-cloroformo; inferior: fase roja. El RNA se encuentra solamente en la fase acuosa.
5. Transferir, cuidadosamente la fase acuosa a un tubo Eppendorf etiquetado, perturbar la fase de fenol-cloroformo, de hacerlo centrifugar de nuevo.
6. Agregar 0.5 mL de alcohol isopropílico por cada mL utilizado de TRIZOL. Lo anterior ayuda a precipitar el RNA.
7. Incubar por 10 minutos a Temperatura ambiente y centrifugar a no más de 25,910 x g durante 15 minutos, a una temperatura de 2°C a 8°C. Después de esta centrifugación es posible que se vea el pellet de RNA, si no sucede esto dejar un pequeño remanente de líquido en la parte inferior del tubo.

8. Retirar el sobrenadante y lavar el pellet con 1 mL de ET-OH al 80% en agua libre de RNAsas, mezclar vigorosamente, (con vortex), centrifugar la muestra durante 5 minutos a 16,580 x g con una temperatura de 2°C a 8°C.
9. Dejar secar el botón de RNA a temperatura ambiente, y resuspenderlo con 20 µL de agua DPEC, dependiendo del tamaño del pellet, incubar 10 minutos a 60°C.
10. Determinar la concentración de la muestra en el Nanodrop (A260/280).
11. Cuantificar la pureza e integridad del RNA mediante un gel de agarosa al 1.2 %.
12. Ajustar la concentración de RNA a 1µg/µL.
13. Conservar la muestra a -70°C.

Para más información revisar (Rio et al., 2010).

Gel para determinar la integridad del RNA.

Disoluciones:

Tabla 4: Disoluciones usadas para determinar la integridad del RNA.

FA gel buffer 10x	Buffer de corrida
200 mM MOPS (libre de ácidos).	100 mL de FA gel buffer 10x.
50 mM de acetato de sodio.	20 mL de formaldehído al 37%.
10 mM de EDTA.	880 mL de agua libre de RNAsa.
pH: 7.0 (ajustar con NaOH).	
Buffer de carga	Gel FA 1.2%
16 µL de azul de bromofenol acuoso.	1.2 g de agarosa.
80 µL de EDTA 500mM, pH: 8.	10 mL de FA gel buffer 10x.
720 µL de formaldehído al 37%.	Aforar a 100 mL con agua libre de RNAsa.
2 mL de glicerol 100%.	Después de calentar para fundir la agarosa, agregar 1.8 mL de formaldehído al 37 %.
3084 µL de formamida.	
4 mL de FA gel buffer 10x.	
Aforar a 10 mL con agua libre de RNAsa.	Agregar 1 µL de gel red por cada 25mL de gel al 1.2 %.

Preparación de la muestra.

1. Cargar aproximadamente 200 ng de RNA por pozo.
2. Adicionar 2 μ L de buffer de carga.
3. Llevar todas las muestras al mismo volumen.
4. Incubar a 65°C durante 5 minutos.
5. Correr el gel FA durante 45 minutos a 95 volts.
6. Si se presenta DNA, tratar con DNAsa, agregar 1 unidad de enzima por cada microgramo de RNA, incubar a 37°C durante una hora, para inactivar la enzima, agregar 2 μ L de EDTA 50mM e incubar 20 minutos a 65°C.

Para más información revisar (Aranda et al., 2012).

17.2 REACCIÓN DE TRANSCRIPTASA REVERSA PARA OBTENER EL cDNA Y PCR CUANTITATIVA.

Fundamento teórico de las técnicas:

El DNA complementario (cDNA) es sintetizado por la transcriptasa reversa, la cual utiliza como molde al RNA extraído. Los cebadores utilizados pueden ser oligo de T, el cual se alinea con el extremo 3' de la cadena de RNA que tiene un oligo de poli-adenina; o hexámeros random, los cuales se alinean a secuencias aleatorias del RNA; durante la síntesis se añaden deoxinucleótidos trifosfatados (dNTPs), que la transcriptasa reversa incorpora al naciente cDNA. El proceso de síntesis ocurre en condiciones controladas de temperatura, pH y concentración de sales.

En la PCR cuantitativa, qPCR, el DNA amplificado es detectado al mismo tiempo que la reacción procede, es decir en tiempo real. Para detectar el DNA amplificado se utilizan moléculas fluorescentes que se unen a la doble cadena de DNA, en este caso se usó SYBR Green el cual se excita con luz azul (λ :488nm) y emite luz verde (λ :522nm). Después de cada ciclo los niveles de fluorescencia son medidos con el lector del termociclador para tiempo real, SYBR Green solo fluoresce cuando se encuentra unido a DNA de doble cadena.

El análisis de expresión se llevó a cabo con el método de doble delta CT ($\Delta\Delta$ CT), para esto se tomó como control interno a GAPDH, y como calibrador a la línea celular no neoplásica MCF 10A. El método de $\Delta\Delta$ CT consiste en calcular:

- La diferencial de los CTs: el gen analizado menos el control interno en nuestro caso: $CT_{KDM4A} - CT_{GAPDH}$, esto se realiza para todas las líneas celulares analizadas.

- Posteriormente, se calcula el diferencial de las líneas celulares neoplásicas menos la no neoplásica (calibrador): $(CT_{KDM4A}-CT_{GAPDH})_{MCF7}-CT_{KDM4A}-CT_{GAPDH})_{MCF 10A}$.
- Por último se utiliza esta ecuación: $2^{(CT_{KDM4A}-CT_{GAPDH})_{MCF7}-CT_{KDM4A}-CT_{GAPDH})_{MCF 10A}}$.

El CT mide el número de ciclos en el que la señal fluorescente es mayor que un umbral definido durante la fase logarítmica de amplificación, ciclos umbral (por sus siglas en inglés cycle threshold).

1. Por cada reacción de retrotranscripción agregar:

Tabla 5: Componentes usados para cada reacción de retrotranscripción.

Componente	Volumen	Concentración final
MgCl ₂	4 µL	5mM
10X Buffer II para PCR	2 µL	1X
dATP	2 µL	1 mM
dCTP	2 µL	1 mM
dTTP	2 µL	1 mM
dGTP	2 µL	1 mM
Inhibidores de RNAsa	1 µL	1 U/µL
Transcriptasa reversa	1 µL	2.5 U/µL
Oligo d(T) ₁₆	1 µL	2.5 µM
RNA	1 µL	≤1 µg
H ₂ O	2 µL	
Volumen final	20 µL	

Para la obtención del cDNA se utilizó el Kit GeneAmp® RNA PCR KIT No. De catálogo: N808-0017.

2. Programar el termociclador a 42°C durante 30 minutos (temperatura óptima de la enzima), 99°C durante 5 minutos (inactivación de la enzima), 5°C durante 5 minutos más.

3. Cuantificar el cDNA con ayuda del nanodrop y diluirlo para que el cDNA de todas las muestras se encuentre a la misma concentración, 20 ng/μL.
4. Por cada reacción de qPCR agregar lo siguiente:

Tabla 6: Componentes usados para cada reacción de qPCR en el análisis de expresión.

Componente	Volumen
SYBR-GREEN	10 μL
OLIGO FWD	1 μL
OLIGO RVS	1 μL
cDNA	5 μL
H2O	3 μL

Para el análisis de expresión se empleó el método de doble delta CT (dΔCT) (Livak and Schmittgen, 2001).

17.3 INMUNOFLUORESCENCIAS (IF).

Fundamento teórico de la técnica:

En esta técnica los anticuerpos primarios reconocen y se unen específicamente a sus antígenos (proteínas blanco). El anticuerpo secundario que se encuentra acoplado a un fluoróforo reconoce al anticuerpo primario y se une a él. Para detectar el fluoróforo se utiliza un microscopio de fluorescencia, el cual excita a los fluoróforos y detecta la luz que estos emiten a determinada longitud de onda, por ejemplo, el fluoróforo Alexa 488 absorbe luz a una $\lambda=495\text{nm}$ y emite luz con una $\lambda=519\text{nm}$.

El 4',6-diamino-2-fenilindol (DAPI) se une a regiones ricas en Adenina y Timina en el DNA, la longitud de onda de absorción es de $\lambda=358\text{nm}$ y de emisión es de $\lambda=461\text{nm}$.

Disoluciones:

Tabla 7: Disoluciones empleadas para los ensayos de inmunofluorescencia.

PBS	Tritón
-----	--------

8 g de NaCl.	2.5 mL de tritón 100-X.
0.2 g de KCl.	Aforar a 50 mL con PBS (solución a 5 %).
1.44 g de Na ₂ HPO ₄ .	
0.24 g de KH ₂ PPO ₄ .	Tomar 1 mL y llevarlo a 50 mL con PBS (solución al 0.1%).
Ajustar pH a 7.4.	
Aforar a 1 L.	

Glicina

Formaldehído 1 %

0.375 g de glicina.	370 µL de formaldehído al 37 %.
Aforar a 50 mL con PBS.	

1. Sembrar aproximadamente 100000 células en cada pozo, con un cubreobjetos previamente limpio y radiado, se dejan el cultivo de 12 a 24 horas, dependiendo de la línea celular.
2. Agregar 1 mL de formaldehído al 1 % o al 4 % a cada pozo e incubar 10 minutos.
3. Retirar el formaldehído y realizar 3 lavados con PBS reposando por 5 minutos entre cada lavado con agitación.
4. Adicionar 1 mL de tritón, incubar durante 5 minutos.
5. Retirar el tritón y realizar 3 lavados con PBS 1 mL por 5 minutos, con agitación, cuidar que no se sequen las laminillas.
6. Agregar 1 mL de glicina 100 mM y se incuba durante 1 hora a temperatura ambiente.
7. Eliminar la glicina y adicionar 25 µL de anticuerpo primario, cortar parafilm y colocarlo una vez que ya se haya agregado el anticuerpo primario con el fin de que el anticuerpo y las células se encuentren en contacto y las laminillas no se sequen, incubar durante 1 hora a temperatura ambiente, o toda la noche a 4°C.
8. Lavar de 3 a 5 veces con PBS durante 3 minutos.
9. Agregar 25 µL del anticuerpo secundario e incubar por una hora a temperatura ambiente, en cámara húmeda y aislado de la luz.
10. Realizar tres lavados con PBS durante 3 minutos.
11. Previamente al montaje se limpian con alcohol, los portaobjetos.
12. Posteriormente, se agregan 15 µL de medio de montaje, Vectashield con DAPI diluido con 6.5µL con Vectashield sin DAPI (), (Vector Labs.) a cada inmunofluorescencia, se

coloca el cubreobjetos procurando no dejar burbujas, se retira el excedente del medio de montaje.

13. Se sellan las muestras montadas con barniz de uñas. Las muestras se pueden conservar a 4°C tapadas con aluminio.

Para más información revisar (Donaldson, 2015).

17.4 ENSAYO DE INMUNOPRECIPITACIÓN DE LA CROMATINA (ChIP) E INMUNOPRECIPITACIÓN SECUENCIAL DE LA CROMATINA (ChIP-Re-ChIP).

Fundamento teórico de la técnica:

Durante el entrecruzamiento el formaldehído forma un enlace covalente (-CH₂-) entre los grupos amino de las cadenas de DNA y los grupos amino de los aminoácidos que se encuentran en las histonas; los enlaces formados por el formaldehído son fácilmente reversibles. La reacción de entrecruzamiento se detiene mediante la adición de glicina, esto evita que se formen enlaces donde no hay interacción. Después de que las células sean lisadas, es necesario romper la cromatina en fragmentos de aproximadamente 500-700pb, esto se realiza con el fin de minimizar las falsas interacciones. La inmunoselección se lleva a cabo ya que los anticuerpos se unen específicamente a las proteínas blanco. Posteriormente los complejos cromatina-anticuerpo son inmunoprecipitados con perlas de proteína A, la cual se acopla al anticuerpo. Después de lavar para quitar los agregados no específicos, se revierte el entrecruzamiento con calor y se eliminan las proteínas con proteinasa K. Finalmente se purifica el DNA y las inmunoprecipitaciones son evaluadas por una PCR cuantitativa. La inmunoprecipitación secuencial de la cromatina utiliza DTT con el fin de desacoplar los complejos cromatina-anticuerpo-perla, la posterior centrifugación descarta los complejos anticuerpo-perla y el sobrenadante se trata con el segundo anticuerpo.

Disoluciones:

Tabla 8: Disoluciones empleadas para los ensayos de ChIP y ChIP-Re-ChIP.

Buffer de entrecruzamiento	Buffer de lisis
Formaldehído al 11 %. NaCl 100 mM. EGTA 0.5 mM.	Tris-HCl pH 8.0 50 mM. NaCl 150 mM. Inhibidor de proteasas (Complete tablets Mini, EDTA-free. Roche) 2X.

HEPES 50 mM con pH 8.0.

NP40 1%.

Glicina 2.5 M

PBS 1x

Pesar 9.38 g de glicina.

8 g de NaCl.

Aforar a 50 mL con agua bidestilada.

0.2 g de KCl.

1.44 g de Na₂HPO₄.

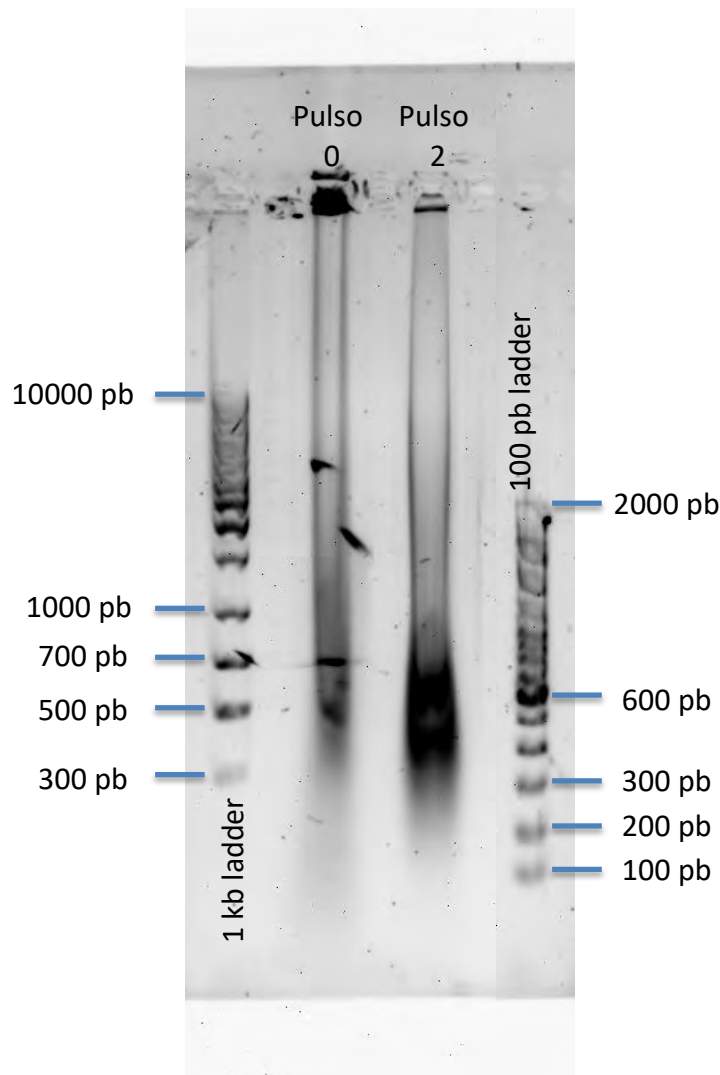
0.24 g de KH₂PPO₄.

Ajustar pH a 7.4.

Aforar a 1 L.

1. **CÉLULAS:** Cultivar las células en 3 cajas de 150 mm a una confluencia del 80 %, checar el volumen del medio, tiene que ser 18 mL.
2. **ENTRECruzamiento O “CROSSLINKING”:** Añadir 2 mL de buffer “crosslinking. Mix” obtener una concentración final de formaldehído al 1 %. Agitar las células por 10 minutos.
3. **DETENER EL “CROSSLINKING”:** Añadir 1 mL de glicina 2.5 M para obtener una concentración final de 125 mM. Mantener las células en agitación constante durante 5 minutos.
4. **LAVADO DE LAS CÉLULAS CON PBS:** Retirar el medio lavar las células 3 veces con PBS frío (4°C).
5. **LISIS:** Remover el PBS y se adicionar 1 mL de Buffer de Lisis con inhibidores a cada una de las cajas, e incuba a 4°C de 5-10 minutos. Transcurrido el tiempo cosechar las células con un scraper y transferir la cromatina a un tubo Falcon de 50 mL.
6. **SONICACIÓN:** Colocar el tubo que contiene la cromatina en un vaso con hielo para que se mantenga frío, evitando generar mucha espuma. (PRESTAR ATENCIÓN AL PROGRAMA DEL SONICADOR). . El programa de sonicación es de 30 segundos, pulso 2.0 - 0.5, amplitud de 37% en un sonicador Ultrasonic Processor (marca GENEQ, modelo GEX500, número de catálogo SOVC505-00). La cromatina se puede guardar a 4°C. La cromatina sonicada se puede conservar durante 3 a 4 meses con un coctel apropiado de inhibidores, y a 4°C.
7. **EVALUACIÓN DE LA CROMATINA:** Determinar el grado de fragmentación de la cromatina, el tamaño adecuado para las Inmunoprecipitaciones es de 500-700 pb. Tomar

una alícuota de 100 μ L y agregar 5 μ L de RNasa A e incubar a 37°C durante una hora, posteriormente agregar 5 μ L de proteínasa K e incubar a 65°C, por lo menos 6 horas. Subsecuentemente realizar una extracción fenol/cloroformo (llevar la muestra a un volumen final de 400 μ l con agua) y se precipita el DNA (añadir 1 volumen de ET-OH al 100 % y 0.66



usar la misma cantidad de cromatina para cada ml, para esta determinación se procede a determinar la concentración de las cromatinas deseadas por el método de Lowry.

- a. **Preparación de los reactivos:** Agregar 20 μ L de la solución S por cada mL de la solución A.

- b. Curva estándar de albúmina.** Partir de un stock de 1 mg/mL, realizar la curva estándar con las siguientes cantidades de albúmina, 0 mg, 1 mg, 5 mg, 10 mg, 15 mg, 20 mg.
- c. Cargar la placa de ELISA:** Con el fin de anular el ruido causado por los detergentes se utiliza la solución S, se agrega 20 µL de solución S por cada mL de reactivo A. Se emplearon las soluciones del kit DC™ Protein Assay Kit I con número de catálogo: 5000111 de la marca Biorad. Por cada muestra agregar 165 µL del reactivo B de Lowry, 20 µL del reactivo A, de 2 µL a 5 µL de la muestra, ajustar el volumen final con agua a 200 µL. Para los volúmenes de la curva estándar se pueden adicionar de 1 µL a 15 µL.
- 9. INMUNOSELECCIÓN:** Preparar el ChIP buffer 1 x, adicionar 100 mL de ChIP Buffer a 400 mL de agua desionizada, almacenar el ChIP buffer 1 x a 4°C. Preparar 5 mL de ChIP Buffer 1x con 25 µL de inhibidores de proteasas. Diluir la cromatina con esta última mezcla de modo que el volumen de la cromatina no exceda de 28 %, por inmunoprecipitación. Repartir la cromatina en tubos de 1.5 mL.
- a. Input:** Tomar una alícuota del mismo volumen que las inmunoprecipitaciones, esta alícuota se procesará más adelante.
- b. Anticuerpos:** Añadir los anticuerpos deseados (ver tabla 16), agitar vigorosamente con el vortex por 5 segundos a potencia media. Incubar durante 1 hora a 4°C en movimiento.
- c. Perlas:** Resuspender perfectamente las perlas antes de cada uso, tomar una alícuota de 840 µL para 20 IP, transferir estos 840 µL a un tubo de 15 mL agregar 10.5 mL de ChIP buffer 1x, lavar las perlas mediante la inversión del tubo. Centrifugar a 500 x g durante 3 minutos a 4°C. Repetir esta operación una vez más. Alicuotar 500 µL las perlas en tubos nuevos de 1.5 mL, centrifugar a 500 x g durante 2 minutos a 4°C, checar que todos los tubos tengan la misma cantidad de perlas. Quitar el sobrenadante, sin perturbar el pellet de perlas, mantener en hielo las perlas alicuotadas.
- 10. Inmunoprecipitación:** Cuando termina el tiempo de incubación de la inmunoselección centrifugar los tubos que contienen los complejos anticuerpo-cromatina, a 14,000 x g, durante 10 minutos a 4°C. Este paso se realiza con el fin de eliminar agregados inespecíficos, los complejos anticuerpo-cromatina se encuentran en el sobrenadante. Transferir el sobrenadante al tubo correspondiente que contiene las perlas, tener cuidado de no perturbar el pellet ya que este contiene los agregados inespecíficos. Incubar durante 1 hora a 4°C en movimiento.
- a. Input:** Agregar 1 mL de Et-OH al 100% a la alícuota tomada en el paso anterior. Mezclar de 2 a 3 veces por inversión, incubar durante 10 minutos en hielo, posteriormente centrifugar a 10,000 x g durante 10 minutos a 4°C. retirar el sobrenadante sin perturbar el

pellet, lavar el pellet con 1 mL de Et-OH al 75 %, deshacer el pellet con ayuda de la pipeta y centrifugar a 10,000 x g durante 10 minutos a 4°C. Mantener el pellet a temperatura ambiente.

b. Aislamiento de las inmunoprecipitaciones: Una vez terminado el tiempo de incubación agregar 1 mL de ChIP buffer 1x a los tubos de 1.5 que contienen los complejos perlas-anticuerpo-cromatina, mezclar los tubos por inversión 2 veces, centrifugar los tubos a 500 x g, durante 2 minutos, a 4°C. Retirar el sobrenadante sin perturbar el pellet y agregar 1 mL más de ChIP buffer 1x, resuspender el pellet y transferirlo a un tubo Falcon de 15mL el cual contiene 12 mL de ChIP buffer 1 x frío, incubar durante 5 minutos en hielo y centrifugar a 500 x g durante 3 minutos a 4°C. Retirar 12mL de ChIP buffer 1x, resuspender el mililitro restante y trasladarlo a un tubo nuevo de 1.5 mL, centrifugar a 500 x g durante 2 minutos a 4°C, tirar el sobrenadante sin perturbar el pellet de perlas.

11. Purificación del DNA: a partir de este paso ya no trabajar en hielo, calentar agua a punto de ebullición.

a. Input: agregar 100 µL de agua para PCR y resuspender el pellet, incubar a temperatura ambiente.

b. Inmunoprecipitaciones: sacar la suspensión de purificación del DNA, es importante que durante el uso de este reactivo la suspensión sea uniforme, hay que moverlo constantemente para conservar esta uniformidad. Añadir 100 µL de suspensión purificadora a cada inmunoprecipitación. Incubar a temperatura ambiente durante 1 minuto.

c. Input: Añadir 100 µL de suspensión purificadora del DNA e incubar a temperatura ambiente durante 1 minuto.

d. Input e inmunoprecipitaciones: mezclar por inversión, los tubos de los dos pasos anteriores, cerrar los tubos con pinzas.

e. Revertir el entrecruzamiento: colocar los tubos en agua hirviendo durante 10 minutos. Terminado este tiempo agregar 1 µL de Proteínasa K, incubar en el termomixer a 55°C durante por lo menos 30 minutos. Terminado el tiempo de incubación cerrar los tubos con pinzas e incubarlos durante 10 minutos en agua hirviendo. Y centrifugar a 14,000 x g durante 1 minutos 4°C

f. Recuperación del DNA: sin perturbar el pellet, transferir 70 µL del sobrenadante a otro tubo de 1.5 mL, añadir 130 µL de agua para PCR al pellet, mezclar vigorosamente con ayuda del vortex y centrifugar a 14,000 x g, durante 1 minuto a 4°C. Sin perturbar el pellet, tomar 130 µL del sobrenadante y trasladarlo al tubo de 1.5 mL donde se encuentran los 70 µL previos, el volumen final de cada muestra es de 200 µL, conservar las muestras a -20°C.

La suspensión purificadora del DNA que se añadió en los pasos 11b y 11c ayuda a recuperar el DNA de las muestras, por ello no es necesario extraer con etanol y sales.

12. Evaluación del CHIP por PCR en tiempo real y análisis de los datos.

a. Por cada reacción de qPCR añadir:

Tabla 9: Componentes usados para cada reacción de qPCR en el análisis de los ensayos de CHIP.

Componente	Volumen
SYBR-GREEN	10 µL
OLIGO FWD	1.5 µL
OLIGO RVS	1.5 µL
Inmunoprecipitación	5 µL
H ₂ O	2 µL
Volumen final	20 µL

b. Validación de los oligonucleótidos: se realiza una curva estándar con diluciones del input 0.01 %, 0.1 %, 1 %, 10 %, para hacer las diluciones sucesivas se toman 100 µL del tubo del input 100% y se diluyen con 900 µL de agua para PCR, de esta dilución (10% del input) se toman nuevamente 100 µL y se diluyen con 900 µL de agua para PCR, de esta dilución (1% del input) se toman nuevamente 100 µL y se diluyen con 900 µL de agua para PCR. Para obtener las diluciones 0.1% y 0.01% se repiten los pasos dos veces más.

Posteriormente graficar CT *versus* log (input %), con el uso de la pendiente de esta gráfica y la siguiente ecuación se obtiene la eficiencia de amplificación (EA), la cual debe de ser cercana a 2. (Ver gráfica 1).

$$\text{Ecuación 1} > \quad EA = 10^{(-1/m)}$$

Dónde:

m: pendiente de la recta calculada por las diferentes concentraciones del input.

EA: eficiencia de amplificación.

c. % Input: este valor indica la eficiencia de inmunoprecipitación con respecto al material inicial y se calcula con la siguiente ecuación:

Ecuación 2 >
$$\%Input = EA^{(Ct_{input} - Ct_{ChIP})}(FD)(100)$$

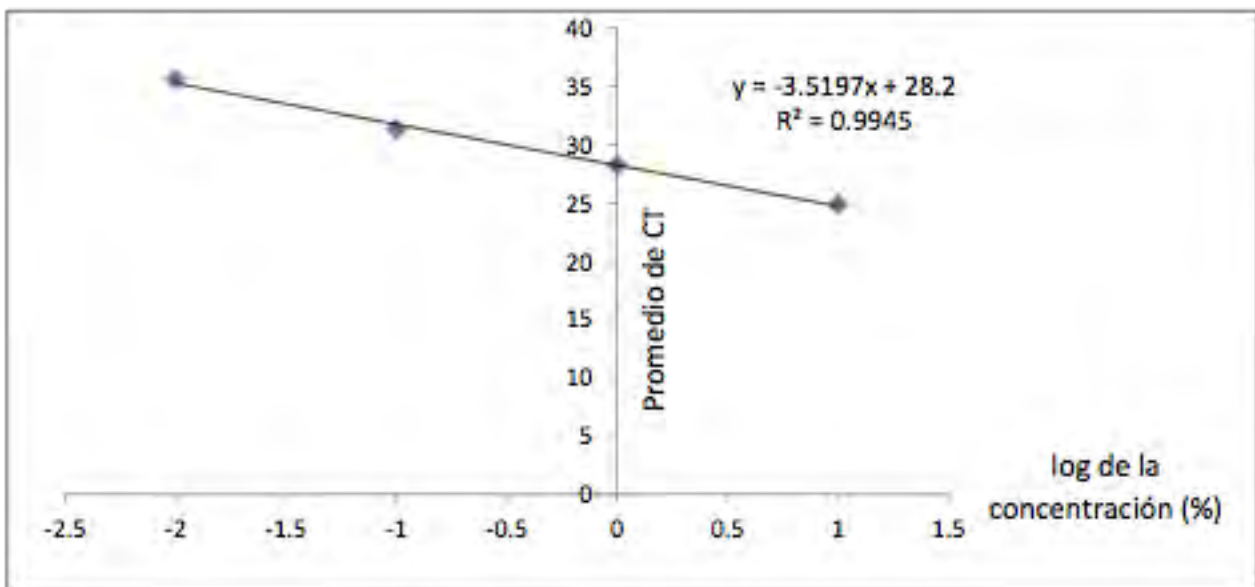
Dónde:

EA: eficiencia de amplificación.

FD: factor de dilución del input con el cual se comparó las inmunoprecipitaciones.

d. Veces de enriquecimiento: este parámetro compara el % del input de las Inmunoprecipitaciones con el % del input de el anticuerpo inespecífico, IgG.

Ecuación 3 >
$$\text{veces de enriquecimiento} = \frac{\% \text{ input loci específico}}{\% \text{ input loci inespecífico}}$$



Gráfica 1: Ejemplo de curva estándar de INPUT. Las concentraciones utilizadas del INPUT de HeLa fueron, 10 %, 1 %, 0.1 % y 0.01 %.

Nota: Se empleó el kit **One Day ChIP kit**, Cat. No. kch-oneDIP-060 / kch-oneDIP-180, de la marca Diagenode, así mismo, las ecuaciones aquí planteadas se extrajeron del manual de dicho Kit, para más información visitar la página: <http://www.diagenode.com/en/applications/chromatin-immunoprecipitation.php>,

Para el ChIP-Re-ChIP se siguen los pasos del ensayo de ChIP hasta los lavados de los complejos cromatina-anticuerpo-perla.

- a. **Lavado de las inmunoprecipitaciones:** Una vez terminado el tiempo de incubación agregar 1 mL de ChIP buffer 1x a los tubos de 1.5 que contienen los complejos perlas-anticuerpo-cromatina. Dejar 5 minutos rotando a 4°C, centrifugar los tubos a 2750 rpm, durante 2 minutos, a 4°C, retirar el sobrenadante sin perturbar el pellet. Repetir el lavado dos veces.

- b. Elución con DTT:** con el objetivo de remover el complejo perla-anticuerpo del antígeno de interés, se incuban las inmunoprecipitaciones con DTT 10 mM MÁXIMO 75 μ L de buffer durante 30 min a 37°C, centrifugar los tubos a 2750 rpm, durante 2 minutos, a 4°C, recuperar el sobrenadante y se lleva con ChIP buffer a 1.5mL.
- c. Incubación con el segundo anticuerpo.** Incubar los sobrenadantes del paso anterior con los anticuerpos y dejar rotando durante toda la noche a 4°C.
- 2. Segunda inmunoprecipitación.** Cuando termina el tiempo de incubación de la inmunoselección centrifugar los tubos que contienen los complejos anticuerpo-cromatina, a 12000 rpm, durante 10 minutos a 4°C. Este paso se realiza con el fin de eliminar agregados inespecíficos, los complejos anticuerpo-cromatina se encuentran en el sobrenadante. Transferir el sobrenadante al tubo correspondiente que contiene las perlas, tener cuidado de no perturbar el pellet ya que este contiene los agregados inespecíficos. Incubar durante 3 horas a 4°C en movimiento.
- 3. Aislamiento de las inmunoprecipitaciones:** Una vez terminado el tiempo de incubación agregar 1 mL de ChIP buffer 1x a los tubos de 1.5 que contienen los complejos perlas-anticuerpo-cromatina. Dejar 5 minutos rotando a 4°C centrifugar los tubos a 2750rpm, durante 2 minutos, a 4°C, retirar el sobrenadante sin perturbar el pellet. Repetir el lavado dos veces.
- 4. Purificación del DNA:** a partir de este paso ya no trabajar en hielo, calentar agua a punto de ebullición.
- a. Input:** agregar 100 μ L de agua para PCR y resuspender el pellet, obtenido en el paso 10b del ChIP, e incubar a temperatura ambiente.
- b. Inmunoprecipitaciones:** sacar la suspensión de purificación del DNA, es importante que durante el uso de este reactivo la suspensión sea uniforme, hay que moverlo constantemente para conservar esta uniformidad. Añadir 100 μ L de suspensión purificadora a cada inmunoprecipitación. Incubar a temperatura ambiente durante 1 minuto.
- c. Input:** Añadir 100 μ L de suspensión purificadora a. Incubar a temperatura ambiente durante 1 minuto.
- d. Input e inmunoprecipitaciones:** mezclar por inversión, los tubos de los dos pasos anteriores, cerrar los tubos con pinzas.
- e. Revertir el entrecruzamiento:** colocar los tubos en agua hirviendo durante 10 minutos. Terminado este tiempo agregar 1 μ L de Proteínasa K, incubar en el termomixer a 55°C durante por lo menos 30 minutos. Terminado el tiempo de incubación cerrar los tubos con pinzas e

incubarlos durante 10 minutos en agua hirviendo. Y centrifugar a 12000rpm durante 1 minuto a 4°C.

- f. **Recuperación del DNA:** sin perturbar el pellet, transferir 70 µL del sobrenadante a otro tubo de 1.5 mL, añadir 130 µL de agua para PCR al pellet, mezclar vigorosamente con ayuda del vortex y centrifugar a 12000 rpm, durante 1 minuto a 4°C. Sin perturbar el pellet, tomar 130 µL del sobrenadante y trasladarlo al tubo de 1.5 mL donde se encuentran los 70 µL previos, el volumen final de cada muestra es de 200 µL, conservar las muestras a -20°C.

5. PCR en tiempo real y análisis de los datos.

- a. **Por cada reacción de qPCR añadir:**

Tabla 10: Componentes usados para cada reacción de qPCR en el análisis de los ensayos de ChIP-Re-ChIP.

Componente	Volumen
SYBR-GREEN	10 µL
OLIGO FWD	1.5 µL
OLIGO RVS	1.5 µL
c-DNA	5 µL
H ₂ O	2 µL

- b. **Validación de los oligonucleótidos:** se realizó la validación de los oligonucleótidos como se menciona en el paso 12b del protocolo del ChIP.

- c. **% Input:** con el valor de la EA obtenido en el paso anterior se calculo el % del input como en el paso 12C del ChIP.

Los resultados de ChIP-Re-ChIP se muestran en porcentaje de input, ya que hay dos controles negativos y no es posible calcular las veces de enriquecimiento.

El cálculo del % de co-ocupancia se realizó como lo reportaron Geisberg y Struhl (Geisberg and Struhl, 2005), % de co-ocupancia = $100 (AB - A) / (A \times B - A)$, donde A y B representan el valor de cada ChIP, y AB el valor del ensayo de ChIP-Re-ChIP en las líneas celulares MCF7 y HeLa.

Para más información revisar (Beischlag et al., 2018; Truax y Greer, 2012).

17.5 COINMUNOPRECIPITACIÓN.

Fundamento teórico de la técnica:

La inmunoprecipitación (IP) es una técnica que consiste en precipitar el antígeno de la proteína en una solución utilizando un anticuerpo que se une específicamente a esa proteína

en particular. Después de que se acopla el anticuerpo al antígeno se precipita usando perlas especiales que se unen a la fracción cristalizable del anticuerpo. Posterior a la IP, los complejos se evalúan mediante western blot.

Tabla 11: Componentes del buffer de lisis para llevar a cabo los ensayos de inmunoprecipitación.

IP LISIS BUFFER	
Stocks	Concentraciones finales
Tris-HCl pH 8.0 1 M	50 mM
NaCl 2 M	150 mM
Inhibidor de proteasas (Complete tablets Mini, EDTA-free. Roche)	2X
NP40 10%	1%

1. Extraer proteína de las líneas celulares de interés, se cultivan a una confluencia aproximadamente del 80% en cajas p60, previo a la extracción de proteína se lavan 2 veces con PBS friso durante 5 minutos en agitación moderada, posteriormente se lisan las células agregando 850 µl de IP Buffer frío, lisar por 10 min a 4°C obtener las células con ayuda de un scrapper.
2. Cosechar en un tubo eppendorf de 1.5mL y centrifugar a 13,000 rpm 10 min a 4° C.
3. Recuperar el sobrenadante y pasarlo a un tubo Eppendorf de 1.5 ml frío.
4. Las proteínas fueron incubadas con 2µg del anticuerpo anti-CTCF (Santa Cruz Biotech, sc-5916) o sin anticuerpo (-) y con 25µL de perlas magnéticas acopladas a proteína A/G (Pierce, 88802) a 4°C aproximadamente 16 h.
5. Las perlas fueron recuperadas con una gradilla magnética y lavadas 5 veces con IP-lisis buffer durante 20 minutos en rotación a 4°C.
6. Finalmente, las proteínas se eluyeron mediante la ebullición en buffer Laemmli 1x y se evaluaron mediante Western Blot usando anticuerpos contra CTCF (Santa Cruz Biotech, sc-5916) y KDM4A (Cell Signaling, 2A # 5328). Se evaluaron al menos tres repeticiones biológicas independientes.

Para más información revisar (Tang y Takahashi, 2018).

17.6 CONVERSION POR BISULFITO DE SODIO Y MS-PCR.

Fundamento teórico de la técnica:

La química de la conversión del DNA por bisulfito de sodio consta de tres etapas: (1) sulfonación: La adición de bisulfito al enlace doble 5-6 de la citosina (2) La desaminación: desaminación hidrolítica de las resultantes citosina-bisulfito de derivados para dar un derivado de uracilo-bisulfito (3) Desulfonación alcalina: La eliminación de la sulfonato grupo por un tratamiento alcalino, para dar el uracilo. Bisulfito de preferencia desamina citosina a uracilo en el DNA de cadena sencilla. La amplificación por PCR, el uracilo se amplifica como timina mientras que el 5-MEC residuos permanecen como citosinas, lo que permite metilado CpGs deben distinguirse de los CpGs metilado por la presencia de una citosina "C" frente a la timina "T" de residuos durante la secuencia.

1. Preparación del reactivo de conversión (CT). El reactivo CT suministrado en este kit (D5006 EZ DNA Methylation-Gold Kit) es una mezcla sólida y debe ser preparado antes del primer uso.

1.1. Agregar 900 μ L de agua, 300 μ L de buffer de dilución M y 50 μ L de buffer de disolución M a un tubo de reactivo CT.

1.2. Mezclar a temperatura ambiente con agitación frecuente durante 10 minutos. Nota: Es normal ver trazas de reactivo no disuelto en el Reactivo de CT. Cada tubo de reactivo CT está diseñado para 10 reacciones de conversión.

1.3. Almacenamiento: el reactivo CT debe usarse inmediatamente después de la preparación. Si no se utiliza de inmediato, la solución de reactivo CT se puede almacenar cubierto de la luz, una semana a 4 ° C, o hasta un mes a -20 ° C. La solución debe calentarse a 37 ° C, luego agitarse en vórtex antes de usarla.

2. Preparación del buffer de lavado M

2.1. Añadir 24 ml de etanol al 100% al buffer de lavado M-6 de 6 ml (D5005) o 96 ml de etanol al 100% hasta el concentrado de 24 ml de tampón de lavado M (D5006) antes de usar.

3. Agregar 130 μ L del reactivo de conversión CT a 20 μ L de el DNA previamente purificado en un tubo para PCR. Si el volumen de la muestra de DNA es menor a 20 μ L completar con agua. Mezclar la muestra moviendo el tubo o pipeteando la muestra hacia arriba y hacia abajo, luego centrifugar.

4. Colocar el tubo de muestra en un termociclador y con el siguiente programa:

- 4.1.1. 98 ° C durante 10 minutos
 - 4.1.2. 64 ° C durante 2.5 horas
 - 4.1.3. Almacenamiento a 4 ° C hasta 20 horas.
5. Agregar 600 µL de M-Binding Buffer a una columna Zymo-Spin™ IC y coloque el columna en un tubo de recogida proporcionado.
 6. Cargar la muestra del Paso 4 en la columna IC de Zymo-Spin™ que contiene el Mbinding Buffer. Cerrar la tapa y mezclar invirtiendo la columna varias veces.
 7. Centrifugar a velocidad máxima (> 10,000 x g) durante 30 segundos. Deseche el sobrenadante.
 8. Agregar 100 µL de M-Wash Buffer a la columna. Centrifugar a toda velocidad durante 30 segundos.
 9. Agregar 200 µL de tampón de desulfonación M a la columna y deje reposar en la habitación Temperatura (20-30 ° C) durante 15-20 minutos. Después de la incubación, centrifugar al máximo. Velocidad durante 30 segundos.
 10. Agregar 200 µL de M-Wash Buffer a la columna. Centrifugar a toda velocidad durante 30 segundos. Agregue otros 200 µL de M-Wash Buffer y centrifugar por 30 adicionales segundos.
 11. Colocar la columna en un tubo de microcentrífuga de 1.5 mL. Añadir 10 µL de M-Elution Buffer directamente a la matriz de la columna. Centrifugar durante 30 segundos a toda velocidad para eluir el DNA.
 12. El DNA está listo para el análisis inmediato o se puede almacenar a -20 ° C o menos para su posterior uso. Para almacenamiento a largo plazo, conservar a -70 ° C o menos. Recomendamos usar 1-4 µl de DNA eluido para cada PCR, sin embargo, se pueden usar hasta 10 µl si es necesario.
 13. El volumen de elución puede ser ≤10 µL dependiendo de los requisitos de los experimentos, pequeños volúmenes de elución producirán DNA más concentrado.

Para más información visitar:

<https://files.zymoresearch.com/protocols/ d5005 d5006 ez dna methylation-goldga o kit.pdf>.

17.7 TRANSFECCIÓN DE siRNAs y shRNAs.

1. Sembrar 300000 células en una caja de 6 pozos, 24 h antes de la transfección de los shRNAs y siRNAs.
2. Se usó el reactivo de transfección X-fect (PT5003-2, de la marca Clontech) para transfectar 1 µg del siRNA contra KDM4A o el shRNA contra CTCF, también fueron transfectados 1 µg de los siRNA (siRNA contra KDM4A Dharmacon, E-004292-00-0010) y (siRNA non-targeting o control negativo, Dharmacon, D-001910-01-05) y shRNA (pCT1 el cual abate a CTCF y el vector vacío o control negativo pSilencer, los cuales fueron donados por Ko Ishihara).
3. Diluir 1 µg de los siRNA o shRNA en 100 µL de buffer de reacción Xfect. Posteriormente, agregar 1 µL del polímero Xfect e incubar durante 15 min.
4. Distribuir los 100 µl de solución de complejo de nanopartículas por goteo a las células sembradas.
5. 48 h después de la transfección extraer RNA y analizar la expresión de los genes como se menciona en el punto 13.1.

17.8 TRANSFECCIÓN DE PLÁSMIDOS CRISPR/CAS.

- 1 Sembrar 300000 células en una caja de 6 pozos, 24 h antes de la transfección del plásmido.
- 2 Se usó el reactivo de transfección X-fect (PT5003-2 de la marca Clontech) para transfectar 1 µg del plásmido KO CRMPR / Cas9 KDM4A (sc-404599) y 1 µg del plásmido HDR (sc-404599-HDR), también se transfectaron los plásmidos control.
- 3 Diluir 1 µg de los plásmidos KDM4A KO y HDR en 100 µL de buffer de reacción Xfect. Posteriormente, agregar 2 µL del polímero Xfect y se incubaron durante 15 min.
- 4 Distribuir los 100 µl de solución de complejo de nanopartículas por goteo a las células sembradas.

- 5 48 h después de la transfección, evaluar la presencia de GFP y RFP por microscopía de epifluorescencia (Carl Zeiss, AXIO Imager D2).
- 6 Seleccionar las células transfectadas con medio suplementado con puromicina (3 μg / mL), cambiando el medio cada 24 h durante al menos 5 días.

17.9 CITOMETRÍA DE FLUJO Y SELECCIÓN CELULAR.

1. Las células KDM4A^{KO} o Mock se resuspendieron a una concentración de un millón de células por mililitro de medio en DMEM / F-12, que contenía FBS al 10% y antibiótico/antimicótico 1X.^j
2. Filtrar las células a través de un filtro de 70 μm y posteriormente a través de un filtro de 40 μm .
3. Se clasificaron las células resuspendidas y filtradas en un citómetro de flujo para clasificación de células FACSAria III (BD Biosciences, San Jose, CA).
4. Antes de la selección, se empleó MCF7 WT para acotar el tamaño de las células y las medidas de autofluorescencia.
5. Se determinó que las células control no mostraban fluorescencia de RFP, mientras que la población KO que era positiva para RFP, solo se seleccionaron las células con la fluorescencia más alta. (Figura 25B suplementaria).
6. Se seleccionaron 2.61×10^5 células en medio DMEM / F12 que contenía FBS al 10% y antibiótico/antimicóticos 2X y se sembraron en una placa de cultivo de células p60.
7. Esta última población KDM4A^{KO}, que exhibía alta expresión de RFP +, y las células control se utilizaron para los experimentos posteriores.

17.10 EXTRACCIÓN ÁCIDA DE HISTONAS.

Fundamento teórico de la técnica:

La extracción ácida de histonas se basa en el carácter básico de estas proteínas, esta característica les permite ser solubles en soluciones con pH ácido. Primero se lava con PBS para retirar el medio y debris celular. Posteriormente se lisa con el buffer TEB, este buffer contiene tritón el cual es un detergente y permite lisar las células, también contiene PMSF, este

compuesto se usa para inhibir a las proteasas. Una vez que se termina el tiempo de incubación en hielo, se usa el scraper para cosechar las células lisadas, subsecuentemente se centrifuga, y se desecha el sobrenadante, es importante destacar que las histonas no son solubles en buffer TEB. El pellet contiene los núcleos con las histonas y se lava con buffer TEB, y se vuelve a centrifugar, posteriormente el botón de núcleos se resuspende en HCl 0.2 N, y se deja rotando toda la noche a 4°C, esto permite que las histonas sean extraídas del botón de núcleos. Finalmente se centrifuga para retirar el debris nuclear que queda en el fondo del tubo. Se determina la concentración de los extractos histónicos por el método de Bradford.

1. Para células en monocapa, lavar dos veces con PBS helado. El PBS y las disoluciones amortiguadoras posteriores se pueden complementar con butirato de sodio 5 mM para retener los niveles de acetilación de las histonas.
2. Agregar 200µL, por cada caja p100, de buffer de extracción con tritón, (TEB: PBS que contiene Tritón X 100 al 0,5% (v / v), fluoruro de fenilmetilsulfonilo 2 mM (PMSF).
3. Lisar las células en hielo durante 10 minutos con agitación suave.
4. Obtener las células con ayuda de un scraper.
5. Centrifugar a 2000 RPM durante 10 min a 4 ° C. Retirar y desechar el sobrenadante.
6. Lavar los núcleos a la mitad del volumen de TEB y centrifugar como el paso anterior.
7. Vuelva a suspender el sedimento en HCl 0.2 N, calcular el volumen del ácido en función al botón obtenido.
8. Dejar rotando a 4 ° C durante toda la noche.
9. Centrifugar a 2000 RPM durante 10 min a 4 ° C. Guarde el sobrenadante (que contiene las histonas).
10. Determine la concentración de proteína utilizando el ensayo de Bradford.
11. Almacenar las alícuotas a -20 ° C.
12. Para correr las histonas en un gel SDS-PAGE, neutralice el HCl con tris al 2 M pH:8, nota, cargar 8µg de proteína por cada pozo.

Para más información revisar (Shechter et al., 2007).

17.11 ENSAYOS DE WESTERN BLOT.

Disoluciones:

Tabla 12: Disoluciones empleadas en los ensayos de Western Blot.

Buffer de lisis	Buffer de corrida	Resolving buffer	Spacer buffer	TBS10X

Tris-HCl pH 8.0 50 mM NaCl 150 mM Inhibidor de proteasas (Complete tablets Mini,EDTA-free. Roche) 2X NP40 1%	TRIS-BASE 3g Gly 14.4g SDS (10%) 10mL Aforar a 1L	HCl 1M 48mL Tris-Base 36.3g pH:8.9 (HCl) Aforar a 100mL	Tris-Base 5.98g pH: 6.7 (HCl) Aforar a 100mL	NaCl 80g KCl 12g Tris-base 30g pH:7.4 Aforar a 1L		
Buffer de carga	Buffer de transferencia	PSA	TBS-Tween-leche (1x-0.1%-5%)			
Spacer buffer 5mL SDS (10%) 1mL Glycerol 4mL Azul de BF (5%) 0.4mL Geles desnat. β -mercapto. Concentración final: 5% DTT 1M 1/5 volúmenes.	TRIS-BASE 3g Gly 14.4g Met-OH 200mL Aforar a 1L	PSA 0.1g Aforar a 1mL	TBS 10X 50mL Leche en polvo 25g Tween 20 500uL			
Geles de poliacrilamida	3%	7.5%	10%	12.5%	15%	30%
Resolving buffer	-----	1.25mL	1.25mL	1.25mL	1.25mL	1.25mL
Spacer buffer	1.25mL	-----	-----	-----	-----	-----
Acrilamida (40%)	0.75mL	1.875mL	2.5mL	3mL	3.75mL	7.5mL
SDS (10%)	100uL	100uL	100uL	100uL	100uL	100uL
H ₂ O	7.8mL	6.67mL	6.04mL	5.54mL	4.8mL	9.6mL
PSA (10%)	100uL	100uL	100uL	100uL	100uL	100uL
TEMED	10uL	10uL	10uL	10uL	10uL	10uL

1. Preparación de los extractos proteicos.

- a. Lavar las células con PBS, agitar a lo mínimo durante 5 minutos, retirar el PBS cuidadosamente, repetir este paso una vez mas.
- b. IMPORTATNE, retirar TODO el PBS en el segundo lavado.

- c. Agregar de 100 μ L a 200 μ L de buffer de lisis con inhibidores, asegurarse que el volumen agregado cubra la totalidad de la caja.
- d. Incubar a 4°C durante 10 min, posteriormente, scrappear y recolectar las proteínas en un tubo eppendorf etiquetado y con ayuda de una pipeta de 1mL.
- e. Centrifugar 14000rpm (centrifuga refrigerada) (OJO! Poner a enfriar en el tiempo de incubación), por 10min a 4°C.
- f. Dividir los extractos en alícuotas de 50 μ L-100 μ L, y congelarlos a -70°C.

2. Cuantificación de los extractos proteicos mediante BCA:

- a. Diluir 1:50 del reactivo B con el reactivo A, esta es la disolución de trabajo.
- b. Preparar la curva estándar de albúmina por triplicado, como se muestra a continuación:

Tabla 13: Cantidades y concentraciones para preparar de la curva estándar de albúmina.

μ L de Albúmina (2mg/mL)	μ L de buffer de lisis	μ g de Albúmina en la curva
0	25	0
2.5	22.5	5
5	20	10
10	15	20
20	5	40
25	0	50

- d. Añadir a cada pozo 200 μ L de la dilución de trabajo.
- e. Para las muestras adicionar 25 μ L.
- f. Incubar por 30 minutos protegido de la luz.
- g. Emplear el lector de ELISA con el filtro o la absorbancia 562 nm.

3. Elaborar los geles de poliacrilamida.

- a. Dependiendo del peso de la proteína será el porcentaje de los geles utilizados, en este caso, para KDM4A y CTCF, los respectivos pesos son 150KDa 150KDa, utilizaremos 3% y 7.5%.

- b. Verter primero el gel de mayor porcentaje, y agregar *i*-pr, para evitar irregularidades en el la parte superior del gel. Esperar a que polimerice.
- c. Retirar con cuidado el *i*-pr, hacer dos lavados con H₂O, verter el gel de menor concentración de poliacrilamida, y colocar el peine para cargar.
- d. En el tiempo de polimerización del segundo gel, preparar las muestras, de 15-50ug de muestra, agregar el volumen necesario de buffer Laemmli 5x con β-mercaptoetanol, y calentar en agua hirviendo durante 10 min. Nota: después de la desnaturalización, no colocar las proteínas en hielo, ya que se tornan viscosas.
- e. Colocar los geles en la cámara de electroforesis vertical, y llenar con buffer de corrida, cargar primero el marcador de peso molecular y posteriormente las muestras. Correr a 100mV durante una hora.

4. Transferencia húmeda:

- a. Activar la membrana de PVDF, previamente cortada, mediante su incubación en Me-OH, al activar la membrana esta se torna más hidrofílica, terminado el tiempo de activación, realizar dos lavados con agua de 5min cada uno.
- b. Durante el tiempo de corrida, preparar la cámara de transferencia húmeda, colocar papel 3M previamente mojado con el buffer de transferencia y retirar las posibles burbujas mediante el uso de una pipeta de plástico.
- c. Colocar en forma sucesiva, esponja, papel 3M, el gel, la membrana, y otro papel 3M, cuidando de retirar las burbujas después de la membrana y del papel 3M con ayuda de la pipeta.
- d. Transferir con una diferencia de potencial de 50Volts durante 15 horas.

5. Anticuerpo primario.

- a. Durante el tiempo de transferencia, preparar la solución de bloqueo.
- b. Posterior al tiempo de transferencia, recortar la membrana del tamaño del gel, teñirla con rojo de Ponceau, para corroborar la presencia de las proteínas. Lavar la membrana dos veces con, TBS-TWEEN.
- c. Agregar la solución de bloqueo incubar en movimiento durante 1 hora.

- d. Durante la hora de bloqueo, preparar los anticuerpos en la solución de bloqueo, o según las especificaciones del anticuerpo a utilizar.
 - e. Finalizada la hora de bloqueo, realizar dos lavados con TBS-TWEEN-LECHE, y un tercero con TBS1x, de 5 minutos cada uno, durante el tiempo de lavados elaborar las bolsas de incubación, para el anticuerpo primario usando el sellador. Después de los lavados, colocar las membranas en las bolsas y agregar los anticuerpos primarios correspondientes.
 - f. Evitar la formación de burbujas en las membranas, incubar toda la noche a 4°C.
6. Anticuerpo secundario.
- a. Realizar 3 lavados con TBS-TWEEN de 5 minutos cada uno.
 - b. Adicionar el anticuerpo secundario en solución de bloqueo una hora a temperatura ambiente.
7. Revelado de las membranas.
- a. Terminado el tiempo de incubación, realizar 3 lavados con TBS-TWEEN de 5 minutos cada uno y un último con TBS.
 - b. Al cuarto oscuro llevar:
 - i. Placas.
 - ii. Kasette.
 - iii. Revelador, fijador.
 - iv. Contenedores.
 - v. Sanitas.
 - vi. Luminol, se prepara 1:1 y se cubre de la luz.
 - vii. Parafilm.
 - viii. Cinta adhesiva.
 - ix. Pipeta de 1mL y puntas.

Para más información visitar: <https://www.abcam.com/protocols/transfer-and-staining-of-proteins-in-western-blot>

17.12 ENSAYO ENZIMÁTICO.

Fundamento teórico de la técnica:

Con el objetivo de corroborar la capacidad inhibitoria de los compuestos, se realizó un ensayo enzimático in vitro para KDM4A. Particularmente, se sabe que la enzima KDM4A actúa sobre los sustratos H3K9me3, H3K9me2 y H3K36me3 con la producción de formaldehído. La combinación de formaldehído, amoníaco y acetoacetanilida produce un compuesto fluorescente que reacciona a una longitud de onda de excitación de 370 nm y una longitud de onda de emisión de 470 nm. Cuando un compuesto es un posible inhibidor, la actividad de KDM4A se bloquea y el compuesto fluorescente final disminuye en comparación con el control. En este estudio se utilizó el péptido H3K9me3. Se incubaron los distintos inhibidores junto con enzima GST-KDM4A recombinante, cofactores y sustrato, y se registró la actividad enzimática.

1. Incubar por 30 minutos a 37°C 50µM de los compuestos con, 13.5 µL buffer, 6 µL del susstrato y 4.5 µL enzima recombinante humana, en una placa de 96 pozos negra. Se usó el kit de KDM4A Inhibitor Enzymatic Assay Kit (número de catalogo: 2016-001, de la marca Epi-C)
2. Agregar 24 µL de solución reveladora 1 y 6 µL de solución reveladora 2 en cada pozo, incubar por 30 minutos a temperatura ambiente.
3. Leer la fluorescencia con un lector TECAN M-200 a una longitud de onda de excitación de 370 nm y de emisión de 470 nm. El experimento se realizó por triplicado. Para más información revisar: (Sarno et al., 2018).

17.13 ENSAYO DE LABEL-FREE.

Fundamento teórico de la técnica:

Con el objetivo de determinar si los compuestos que inhiben a KDM4A se unen a esta enzima, se llevo a cabo el ensayo de Label-Free. Este ensayo se utiliza para medir la unión de un compuesto con una enzima in vitro. Primero se necesita obtener la enzima pura, ya sea por

cromatografía o por algún otro método, posteriormente se inmoviliza la enzima a una superficie, después se añade el compuesto y finalmente se evalúa la señal.

- 1 Transformar bacterias competentes con el plásmido PGEX-4T-1-KDM4A.
 - 2 Cultivar una colonia bacteriana en medio LB líquido que contiene ampicilina 100 µg / ml en una incubadora con agitación hasta que densidad óptica llegue en un rango entre 0.6 y 0.8.
 - 3 Inducir la producción de KDM4A con 200 µM de isopropil-β-D-1-tiogalactopiranoside durante 7 h. Cosechar las bacterias por centrifugación.
 - 4 Lisar las bacterias con un sonicador (Bioruptor, Diagenode) en buffer de lisis con 1 mM de ditioneitol (DTT), 0.5 mM de PMSF y cóctel inhibidor de proteasas.
 - 5 Purificar la enzima con una columna GSTrap 4b (28401745, GE Healthcare Life Sciences) y dializar la enzima purificada en una solución de agua (NaCl 100 mM, DTT 1 mM, acetato de sodio 50 mM, pH 6,0).
 - 6 Diluir la enzima en solución de acetato de sodio (20 mM, pH 6,0) para obtener una concentración final de 150 µg / mL.
 - 7 Colocar 15 µL de esta solución en cada pocillos de una microplaca óptica de alto rendimiento.
 - 8 Centrifugar la placa a 800 rpm durante 1 minuto e incubar durante toda la noche a 4° C.
 - 9 Al día siguiente, lavar la enzima inmovilizada cuatro veces con 25 µl de PBS y centrifugar la placa a 800 rpm durante 1 minuto después de cada lavado.
 - 10 Incubar la placa a temperatura ambiente durante 3 h, y durante los últimos 30 min colocarla en el instrumento EnSpire (PerkinElmer) con el fin de equilibrar antes del ensayo de unión.
 - 11 Leer la placa, esta es la lectura basal.
 - 12 Agregar 15 µL de los compuestos en diferentes concentraciones 100µM, 50µM, 25µM (en PBS con DMSO al 0.1%) y la misma cantidad del vehículo.
 - 13 Introducir la placa dentro del instrumento EnSpire y leer la placa.
- Para más información revisar: (Sarno et al., 2018).

17.14 ENSAYO DE CAMBIO TÉRMICO CELULAR (CETSA).

Fundamento teórico de la técnica:

Con el fin de evaluar la unión de los inhibidores con KDM4A en la línea celular HCT-116, se realizaron ensayos de cambio térmico celular después del tratamiento con los inhibidores y se evaluó la unión de estos con KDM4A por medio de Inmuno Blot. El ensayo de cambio térmico celular (CETSA) se basa en el principio biofísico de la estabilización térmica inducida por la unión de un compuesto a una proteína blanco. El ensayo implica el tratamiento de células con un compuesto de interés, en este caso FS23, calentamiento para desnaturalizar y precipitar proteínas, lisis celular y la separación de los desechos y agregados celulares de la fracción de proteína soluble. Las proteínas no unidas se desnaturalizan y precipitan a temperaturas elevadas, las proteínas unidas al fármaco permanecen en solución.

1. Cosechar las células en este caso, HCT-116, y lavar con PBS, después del tratamiento con FS23 (50 μ M, 100 μ M y 200 μ M) y un volumen igual de DMSO, como control, durante 1 h.
2. Resuspender las muestras con 1.5 mL de PBS, dividir en alicuotas de 100 μ L.
3. Calentar las muestras a diferentes temperaturas por 3 minutos, incubar a 4°C por 3 minutos mas.
4. Añadir 100 μ L de buffer de lisis e incubar durante 15 minutos en hielo.
5. Centrifugar las muestras a 13000 rpm por 30 minutos a 4°C.
6. Determinar la concentración del sobrenadante por medio del ensayo de Bradford, se utilizo el kit de la marca Bio-Rad Protein Assay Kit I, número de catalogo 5000001.
7. Del extracto de proteína total cargar 20 μ g en un gel de SDS-PAGE al 10%.
8. Trasferir el gel a una membrana de nitrocelulosa.
9. Realizar el Western Blot.

Para más información revisar: (Martinez Molina et al., 2013).

18 ANEXOS.

18.1 OLIGONUCLEÓTIDOS SINTÉTICOS EMPLEADOS.

18.1.1 Análisis de expresión.

Tabla 14: Oligonucleótidos empleados para llevar a cabo el análisis de expresión.

Nombre y función	Oligonucleótidos	Secuencia en DNA (5' – 3')	Tm (°C)	Tamaño esperado en cDNA	Tamaño esperado en DNA genómico

Expresión <i>KDM4A</i>	<i>KDM4A</i> FWD (Exón 20)	CGGCCAAGTCTATGGAGCC	62	177 pb	577 pb
	<i>KDM4A</i> RVS (Exón 22)	TCATTGAAGCGCATGTCTGAG	62		
Expresión <i>GAPDH</i>	<i>GAPDH</i> FWD (Exón 5)	TGCACCACCAACTGCTTAGC	62	95pb	95pb
	<i>GAPDH</i> RVS (Exón 5)	GGCTGGACTGTGGTCATGAG	62		
Expresión <i>CHD5</i>	<i>CHD5</i> FWD (Exón 42)	TCGAGACTTCCCTGTGTTGC	62	127 pb	933 pb
	<i>CHD5</i> RVS (Exón 44)	CTTTTTGTCCCAAGGTGGCG	62		
Expresión <i>CTCF</i>	<i>CTCF</i> FWD (Exón 5-6)	GAGAAGCCATTCAAGTGTTCAT	66	85 pb	399 pb
	<i>CTCF</i> FWD (Exón 20)	CTCCAGTATGAGAGCGAATGTGA	66		

18.1.2 Ensayos de MS-PCR e inmunoprecipitación de la Cromatina (ChIP).

Tabla 15: Oligonucleótidos empleados para analizar la metilación y evaluar los ensayos de ChIP y ChIP-Re-ChIP.

Oligonucleótidos	Secuencia en DNA (5' – 3')	Tm (°C)	Tamaño esperado en DNA genómico
PROM <i>CHD5</i> M FWD	TTGTGCGTTGTGATCGTC	51	153 pb
PROM <i>CHD5</i> M RVS	AAACGTCAAACCCGTAACC	51	
PROM <i>CHD5</i> UN FWD	TTTTTGTGTGTTGTGATTGTT	51	153 pb
PROM <i>CHD5</i> UN RVS	AAACATCAAACCCATAACCAAA	51	
<i>CHD5 +741pb</i> FWD	TCCAAGCACTTTACCCG	53	236 pb
<i>CHD5 +741pb</i> RVS	AAGAACTGTCCCGCAAGG	53	

-1922pb ASCL2 FWD	CGAAGACCGGGTAGAGAGC	53	80 pb
-1922pb ASCL2 RVS	TCTGGGTGTGGCTGCATAC	53	
Promotor WRAP53 FWD	GACAGGTCTGAAGCCTG	53	230 pb
Promotor WRAP53 RVS	CGGGACGTGAAAGGTTAG	53	
Exón 27 RB FWD	CTAACACTGGCATGTTCAAAGC	53	163 pb
Exón 27 RB RVS	GGTGTAGGGGAGGGG	53	

18.2 ANTICUERPOS UTILIZADOS.

18.2.1 Anticuerpos primarios.

Tabla 16: Anticuerpos primarios empleados.

	Anti-KDM4A	Anti-CTCF	Anti-CHD5	Anti-H3K36me3
Origen	Ratón	Conejo	Conejo	Conejo
Marca	Abcam	Millipore	Santa Cruz	Diagenode
Número de catálogo	ab105953	07-729	sc-68389	C15410058
Tipo	Monoclonal	Policlonal	Policlonal	Policlonal
Dilución IF	1:50	1:100	1:50	1:200
ChIP(μL)	5	2	----	2
	Anti-H3K36me2	Anti-H3K9me3	Anti-H3K9me2	IgG
Origen	Conejo	Conejo	Ratón	Conejo
Marca	Diagenode	Diagenode	Abcam	Diagenode
Número de	CS-127-100	CS-056-100	ab-1220	kch-oneDIP-180

catálogo

Tipo	Policlonal	Policlonal	Monoclonal	Policlonal
Dilución IF	1:100	1:100	1:50	
ChIP(μL)	2	1	3.7	2
	Anti-KDM4A	Anti-CTCF	Anti-CTCF	Anti-KDM4A
Origen	Conejo	Cabra	Ratón	Conejo
Marca	Cell signaling	Santa Cruz BT	Santa Cruz BT	Diagenode
Número de catálogo	C37E5	SC-5916	SC-398149	pAb-126-050
Tipo	Monoclonal	Policlonal	Monoclonal	Policlonal
Dilución IF	N/A	N/A	1:100	1:50
WB	1:1000	1:750	N/A	N/A

18.2.2 Anticuerpos secundarios.

Tabla 17: Anticuerpos secundarios empleados.

	Anti-Conejo	Anti-Ratón	Anti-Conejo	Anti-Cabra
Origen	Cabra	Cabra	Cabra	Conejo
Marca	Invitrogen	Invitrogen	Santa Cruz	Santa Cruz
Número de catálogo	A32731	A32723	sc-2030	sc-2020

Tipo	Alexa	Alexa	HRP	Policlonal
Dilución IF	1:200	1:200	N/A	N/A
Dilución WB	N/A	N/A	1:5000	1:5000

18.3 LÍNEAS CELULARES.

Las distintas líneas celulares humanas utilizadas en el presente estudio fueron cultivadas con las condiciones que a continuación se describen.

Tabla 18: Líneas celulares usadas en este estudio.

Línea celular	Origen	Medio de cultivo
MCF 10A	Glándula mamaria, enfermedad fibroquística.	DMEM/F12 (3:1) adicionado con 10% de suero fetal bovino, 2 mM de glutamina, 1x de amp/strepto, 10 ng/ml de EGFrh (factor de crecimiento epidérmico humano recombinante; Invitrogen), 120 mU/ml de insulina recombinante humana y 1 µg/ml de hidrocortisona.
MCF7	Efusión pleural, metástasis	RPMI con 10% SFB y 1%.
MDA-MB-231	Efusión pleural, metástasis	RPMI con 10% SFB.
HeLa	Adenocarcinoma de Cérvix.	DMEM High Glucose, con 10% SFB.
HCT-116	Efusión pleural, metástasis.	RPMI con 10% SFB y 1% de amp/strepto.
PANC-1	Carcinoma epitelial de páncreas.	DMEM, con 10% SFB y 1% de amp/strepto.

Todas las líneas celulares se mantuvieron a 37°C con 5% de CO₂. Sin antibiótico.

19 REFERENCIAS.

- Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., Helin, K. (2007). UTX y JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731–734.
- Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhhattar, R., et al. (2007). New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633–636.
- Anand, R., y Marmorstein, R. (2007). Structure and mechanism of lysine-specific demethylase enzymes. *J. Biol. Chem.* **282**, 35425–35429.
- Aulmann, S., Bläker, H., Penzel, R., Rieker, R.J., Otto, H.F., y Sinn, H.P. (2003). CTCF gene mutations in invasive ductal breast cancer. *Breast Cancer Res. Treat.* **80**, 347–352.
- Bagchi, A., Papazoglu, C., Wu, Y., Capurso, D., Brodt, M., Francis, D., Bredel, M., Vogel, H., y Mills, A.A. (2007). CHD5 is a tumor suppressor at human 1p36. *Cell* **128**, 459–475.
- Bannister, A.J., Schneider, R., y Kouzarides, T. (2002). Histone methylation: dynamic or static? *Cell* **109**, 801–806.
- Bao, L., Zhou, M., y Cui, Y. (2008). CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators. *Nucleic Acids Res.* **36**, D83–87.
- Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27**, 1139–1152.
- Bell, A.C., y Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485.
- Berry, W.L., y Janknecht, R. (2013). KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. *Cancer Res.* **73**, 2936–2942.
- Berry, W.L., Shin, S., Lightfoot, S.A., y Janknecht, R. (2012). Oncogenic features of the JMJD2A histone demethylase in breast cancer. *Int. J. Oncol.* **41**, 1701–1706.
- Biswas, S., y Rao, C.M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur. J. Pharmacol.* **837**, 8–24.
- Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., Filippova, G.N., Lobanenkova, V.V., y Renkawitz, R. (1997). Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. *Mol. Cell. Biol.* **17**, 1281–1288.
- Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., y Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **99 Suppl 4**, 16433–16437.
- Butcher, D.T., y Rodenhiser, D.I. (2007). Epigenetic inactivation of BRCA1 is associated with aberrant expression of CTCF y DNA methyltransferase (DNMT3B) in some sporadic breast tumours. *Eur. J. Cancer* **43**, 210–219.
- Butcher, D.T., Mancini-DiNardo, D.N., Archer, T.K., y Rodenhiser, D.I. (2004). DNA binding sites for putative methylation boundaries in the unmethylated region of the BRCA1 promoter. *Int. J. Cancer* **111**, 669–678.
- Byvoet, P., Shepherd, G.R., Hardin, J.M., y Noland, B.J. (1972). The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch. Biochem. Biophys.* **148**, 558–567.
- Caiafa, P., y Zlatanova, J. (2009). CCCTC-binding factor meets poly(ADP-ribose) polymerase-1. *J. Cell. Physiol.* **219**, 265–270.
- Cascante, A., Klum, S., Biswas, M., Antolin-Fontes, B., Barnabé-Heider, F., y Hermanson, O. (2014). Gene-specific methylation control of H3K9 and H3K36 on neurotrophic BDNF versus astroglial GFAP genes by KDM4A/C regulates neural stem cell differentiation. *J. Mol. Biol.* **426**, 3467–3477.
- Chang, K.-H., King, O.N.F., Tumber, A., Woon, E.C.Y., Heightman, T.D., McDonough, M.A., Schofield, C.J., y Rose, N.R. (2011). Inhibition of histone demethylases by 4-carboxy-2,2'-bipyridyl compounds. *ChemMedChem* **6**, 759–764.
- Chen, Z., Wang, X., Liu, R., Chen, L., Yi, J., Qi, B., Shuang, Z., Liu, M., Li, X., Li, S., et al. (2016). KDM4B-mediated epigenetic silencing of miRNA-615-5p augments RAB24 to facilitate malignancy of hepatoma cells. *Oncotarget*.
- Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergström, R., Kwon, Y.-W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., et al. (2007). CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol. Cell. Biol.* **27**, 1631–1648.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., y Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840.
- Cloos, P.A.C., Christensen, J., Agger, K., y Helin, K. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **22**, 1115–1140.
- Cohen, I., Poręba, E., Kamieniarz, K., y Schneider, R. (2011). Histone Modifiers in Cancer. *Genes Cancer* **2**, 631–647.
- Couture, J.-F., Collazo, E., Ortiz-Tello, P.A., Brunzelle, J.S., y Trievel, R.C. (2007). Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14**, 689–695.
- Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., y Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32.
- Defossez, P.-A., Kelly, K.F., Fillion, G.J.P., Pérez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C.L., Daniel, J.M., and Gilson, E. (2005). The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J. Biol. Chem.* **280**, 43017–43023.
- Díez-Villanueva, A., Mallona, I., y Peinado, M.A. (2015). Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. *Epigenetics Chromatin* **8**, 22.
- Ding, X., Pan, H., Li, J., Zhong, Q., Chen, X., Dry, S.M., y Wang, C.-Y. (2013). Epigenetic activation of AP1 promotes squamous cell carcinoma metastasis. *Sci. Signal.* **6**, ra28.1-13, S0-15.

Donohoe, M.E., Zhang, L.-F., Xu, N., Shi, Y., y Lee, J.T. (2007). Identification of a Ctfc cofactor, Yy1, for the X chromosome binary switch. *Mol. Cell* 25, 43–56.

Du, Z., Li, L., Huang, X., Jin, J., Huang, S., Zhang, Q., y Tao, Q. (2016). The epigenetic modifier CHD5 functions as a novel tumor suppressor for renal cell carcinoma and is predominantly inactivated by promoter CpG methylation. *Oncotarget* 7, 21618–21630.

Duan, L., Rai, G., Roggero, C., Zhang, Q.-J., Wei, Q., Ma, S.H., Zhou, Y., Santoyo, J., Martinez, E.D., Xiao, G., et al. (2015). KDM4/JMJD2 Histone Demethylase Inhibitors Block Prostate Tumor Growth by Suppressing the Expression of AR and BMYB-Regulated Genes. *Chem. Biol.* 22, 1185–1196.

Ewels, P., Magnusson, M., Lundin, S., y Käller, M. (2016). MultiQC: summarize analysis results for multiple tools y samples in a single report. *Bioinforma. Oxf. Engl.* 32, 3047–3048.

Fatemi, M., Paul, T.A., Brodeur, G.M., Shokrani, B., Brim, H., y Ashktorab, H. (2014). Epigenetic silencing of CHD5, a novel tumor-suppressor gene, occurs in early colorectal cancer stages. *Cancer* 120, 172–180.

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

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., y Lobanenkov, V.V. (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.

Fong, N., Saldi, T., Sheridan, R.M., Cortazar, M.A., y Bentley, D.L. (2017). RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. *Mol. Cell* 66, 546–557.e3.

Franci, G., Sarno, F., Nebbioso, A., y Altucci, L. (2017). Identification and characterization of PKF118-310 as a KDM4A inhibitor. *Epigenetics* 12, 198–205.

Fu, Y., Sinha, M., Peterson, C.L., y Weng, Z. (2008). The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. *PLoS Genet.* 4.

Fujita, T., Igarashi, J., Okawa, E.R., Gotoh, T., Manne, J., Kolla, V., Kim, J., Zhao, H., Pawel, B.R., London, W.B., et al. (2008). CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. *J. Natl. Cancer Inst.* 100, 940–949.

Geisberg, J.V., y Struhl, K. (2005). Analysis of protein co-occupancy by quantitative sequential chromatin immunoprecipitation. *Curr. Protoc. Mol. Biol. Chapter 21*, Unit 21.8.

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

Gray, S.G., Iglesias, A.H., Lizcano, F., Villanueva, R., Camelo, S., Jingu, H., Teh, B.T., Koibuchi, N., Chin, W.W., Kokkotou, E., et al. (2005). Functional characterization of JMJD2A, a histone deacetylase- and retinoblastoma-binding protein. *J. Biol. Chem.* 280, 28507–28518.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951.

Guerra-Calderas, L., González-Barrios, R., Herrera, L.A., Cantú de León, D., y Soto-Reyes, E. (2015). The role of the histone demethylase KDM4A in cancer. *Cancer Genet.* 208, 215–224.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., y Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486–489.

Hausinger, R.P. (2004). Feil/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39, 21–68.

Hou, C., Zhao, H., Tanimoto, K., y Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20398–20403.

Huang, J., y Berger, S.L. (2008). The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.* 18, 152–158.

Ishihara, K., Oshimura, M., y Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell* 23, 733–742.

Jeong, Y.S., Park, J.S., Ko, Y., y Kang, Y.-K. (2011). JHDM3A module as an effector molecule in guide-directed modification of target chromatin. *J. Biol. Chem.* 286, 4461–4470.

Jin, X., Xu, H., Wu, X., Li, T., Li, J., Zhou, Y., Dan, H., Jiang, L., Zeng, X., Ji, P., et al. (2017). KDM4A as a prognostic marker of oral squamous cell carcinoma: Evidence from tissue microarray studies in a multicenter cohort. *Oncotarget* 8, 80348–80357.

Jovanovic, J., Rønneberg, J.A., Tost, J., y Kristensen, V. (2010). The epigenetics of breast cancer. *Mol. Oncol.* 4, 242–254.

Kauffman, E.C., Robinson, B.D., Downes, M.J., Powell, L.G., Lee, M.M., Scherr, D.S., Gudas, L.J., y Mongan, N.P. (2011). Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol. Carcinog.* 50, 931–944.

Kelly, W.G., Schaner, C.E., Dernburg, A.F., Lee, M.-H., Kim, S.K., Villeneuve, A.M., y Reinke, V. (2002). X-chromosome silencing in the germline of *C. elegans*. *Dev. Camb. Engl.* 129, 479–492.

Kim, S., y Kaang, B.-K. (2017). Epigenetic regulation and chromatin remodeling in learning and memory. *Exp. Mol. Med.* 49, e281.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., y Bedford, M.T. (2006a). Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* 7, 397–403.

Kim, S., Benoiton, L., y Paik, W.K. (1964). EPSILON-ALKYLLYSINASE. PURIFICATION AND PROPERTIES OF THE ENZYME. *J. Biol. Chem.* 239, 3790–3796.

Kim, S.C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., et al. (2006b). Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* 23, 607–618.

Kim, T.-D., Oh, S., Shin, S., y Janknecht, R. (2012a). Regulation of tumor suppressor p53 and HCT116 cell physiology by histone demethylase JMJD2/KDM4D. *PLoS One* 7, e34618.

Kim, T.-D., Shin, S., Berry, W.L., Oh, S., and Janknecht, R. (2012b). The JMJD2A demethylase regulates apoptosis and proliferation in colon cancer cells. *J. Cell. Biochem.* *113*, 1368–1376.

Kim, T.-D., Jin, F., Shin, S., Oh, S., Lightfoot, S.A., Grande, J.P., Johnson, A.J., van Deursen, J.M., Wren, J.D., and Janknecht, R. (2016). Histone demethylase JMJD2A drives prostate tumorigenesis through transcription factor ETV1. *J. Clin. Invest.* *126*, 706–720.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkova, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* *128*, 1231–1245.

Klose, R.J., and Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. *Nat. Rev. Mol. Cell Biol.* *8*, 307–318.

Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHD3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* *442*, 312–316.

Kogure, M., Takawa, M., Cho, H.-S., Toyokawa, G., Hayashi, K., Tsunoda, T., Kobayashi, T., Daigo, Y., Sugiyama, M., Atomi, Y., et al. (2013). Dereglulation of the histone demethylase JMJD2A is involved in human carcinogenesis through regulation of the G(1)/S transition. *Cancer Lett.* *336*, 76–84.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* *128*, 693–705.

Kwon, S.Y., Xiao, H., Wu, C., and Badenhorn, P. (2009). Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities. *PLoS Genet.* *5*, e1000574.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* *410*, 116–120.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstone, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., et al. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* *449*, 689–694.

Lan, F., Nottke, A.C., and Shi, Y. (2008). Mechanisms involved in the regulation of histone lysine demethylases. *Curr. Opin. Cell Biol.* *20*, 316–325.

Lee, B.-K., and Iyer, V.R. (2012). Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. *J. Biol. Chem.* *287*, 30906–30913.

Lee, J., Thompson, J.R., Botuyan, M.V., and Mer, G. (2008). Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat. Struct. Mol. Biol.* *15*, 109–111.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* *437*, 432–435.

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

Li, B.-X., Li, J., Luo, C.-L., Zhang, M.-C., Li, H., Li, L.-L., Xu, H.-F., Shen, Y.-W., Xue, A.-M., and Zhao, Z.-Q. (2013). Expression of JMJD2A in infiltrating duct carcinoma was markedly higher than fibroadenoma, and associated with expression of ARHI, p53 and ER in infiltrating duct carcinoma. *Indian J. Exp. Biol.* *51*, 208–217.

Li, L.-L., Xue, A.-M., Li, B.-X., Shen, Y.-W., Li, Y.-H., Luo, C.-L., Zhang, M.-C., Jiang, J.-Q., Xu, Z.-D., Xie, J.-H., et al. (2014). JMJD2A contributes to breast cancer progression through transcriptional repression of the tumor suppressor ARHI. *Breast Cancer Res. BCR* *16*, R56.

Li, T., Hu, J.-F., Qiu, X., Ling, J., Chen, H., Wang, S., Hou, A., Vu, T.H., and Hoffman, A.R. (2008). CTCF Regulates Allelic Expression of Igf2 by Orchestrating a Promoter-Polycomb Repressive Complex 2 Intrachromosomal Loop. *Mol. Cell. Biol.* *28*, 6473–6482.

Li, X., Moon, G., Shin, S., Zhang, B., and Janknecht, R. (2018). Cooperation between ETS variant 2 and Jumonji domain-containing 2 histone demethylases. *Mol. Med. Rep.* *17*, 5518–5527.

Liu, Z., Scannell, D.R., Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* *146*, 720–731.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* *25*, 402–408.

Lobanenkova, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (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.

Loh, Y.-H., Zhang, W., Chen, X., George, J., and Ng, H.-H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* *21*, 2545–2557.

Luger, K., and Hansen, J.C. (2005). Nucleosome and chromatin fiber dynamics. *Curr. Opin. Struct. Biol.* *15*, 188–196.

Lutz, M., Burke, L.J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkova, V., et al. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* *28*, 1707–1713.

Lutz, M., Burke, L.J., LeFevre, P., Myers, F.A., Thorne, A.W., Crane-Robinson, C., Bonifer, C., Filippova, G.N., Lobanenkova, V., and Renkawitz, R. (2003). Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor. *EMBO J.* *22*, 1579–1587.

Mallete, F.A., and Richard, S. (2012). JMJD2A promotes cellular transformation by blocking cellular senescence through transcriptional repression of the tumor suppressor CHD5. *Cell Rep.* *2*, 1233–1243.

Méndez-Catalá, C.F., Gretton, S., Vostrov, A., Pugacheva, E., Farrar, D., Ito, Y., Docquier, F., Kita, G.-X., Murrell, A., Lobanenkova, V., et al. (2013). A novel mechanism for CTCF in the epigenetic regulation of Bax in breast cancer cells. *Neoplasia N. Y. N* *15*, 898–912.

Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H.F.M., Günther, T., Buettner, R., and Schüle, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* *437*, 436–439.

Metzger, E., Stepputtis, S.S., Strietz, J., Preca, B.-T., Urban, S., Willmann, D., Allen, A., Zenk, F., Iovino, N., Bronsert, P., et al. (2017). KDM4 Inhibition Targets Breast Cancer Stem-like Cells. *Cancer Res.* *77*, 5900–5912.

Mokarram, P., Kumar, K., Brim, H., Naghibalhossaini, F., Saberi-firoozi, M., Nouraie, M., Green, R., Lee, E., Smoot, D.T., y Ashktorab, H. (2009). Distinct high-profile methylated genes in colorectal cancer. *PLoS One* 4, e7012.

Moore, J.M., Rabaia, N.A., Smith, L.E., Fagerlie, S., Gurley, K., Loukinov, D., Disteche, C.M., Collins, S.J., Kemp, C.J., Lobanekov, V.V., et al. (2012). Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. *PLoS One* 7, e34915.

Mu, H., Xiang, L., Li, S., Rao, D., Wang, S., y Yu, K. (2018). MiR-10a functions as a tumor suppressor in prostate cancer via targeting KDM4A. *J. Cell. Biochem.*

Mulero-Navarro, S., y Esteller, M. (2008). Chromatin remodeling factor CHD5 is silenced by promoter CpG island hypermethylation in human cancer. *Epigenetics* 3, 210–215.

Nakahashi, H., Kwon, K.-R.K., 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.

Neault, M., Mallette, F.A., y Richard, S. (2016). miR-137 Modulates a Tumor Suppressor Network-Inducing Senescence in Pancreatic Cancer Cells. *Cell Rep.* 14, 1966–1978.

Ohlsson, R., Renkawitz, R., y Lobanekov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet. TIG* 17, 520–527.

Paik, W.K., y Kim, S. (1973). Enzymatic demethylation of calf thymus histones. *Biochem. Biophys. Res. Commun.* 51, 781–788.

Paik, W.K., y Kim, S. (1974). Epsilon-alkyllysine. New assay method, purification, and biological significance. *Arch. Biochem. Biophys.* 165, 369–378.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433.

Patel, D.J., y Wang, Z. (2013). Readout of epigenetic modifications. *Annu. Rev. Biochem.* 82, 81–118.

Peserico, A., y Simone, C. (2011). Physical and functional HAT/HDAC interplay regulates protein acetylation balance. *J. Biomed. Biotechnol.* 2011, 371832.

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

Plasschaert, R.N., Vigneau, S., Tempera, I., Gupta, R., Maksimoska, J., Everett, L., Davuluri, R., Mamorstein, R., Lieberman, P.M., Schultz, D., et al. (2014). CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic Acids Res.* 42, 774–789.

Rando, O.J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr. Opin. Genet. Dev.* 22, 148–155.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406, 593–599.

Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., y Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6883–6888.

Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S., et al. (2000). A global profile of germline gene expression in *C. elegans*. *Mol. Cell* 6, 605–616.

Renaud, S., Loukinov, D., Bosman, F.T., Lobanekov, V., y Benhattar, J. (2005). CTCF binds the proximal exonic region of hTERT y inhibits its transcription. *Nucleic Acids Res.* 33, 6850–6860.

Reuben, M., y Lin, R. (2002). Germline X chromosomes exhibit contrasting patterns of histone H3 methylation in *Caenorhabditis elegans*. *Dev. Biol.* 245, 71–82.

Ross-Innes, C.S., Brown, G.D., y Carroll, J.S. (2011). A co-ordinated interaction between CTCF and ER in breast cancer cells. *BMC Genomics* 12, 593.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., y Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 105, 8309–8314.

Schwalie, P.C., Ward, M.C., Cain, C.E., Faure, A.J., Gilad, Y., Odom, D.T., y Flicek, P. (2013). Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome Biol.* 14, R148.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., y Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* 88, 593–602.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstone, J.R., Cole, P.A., Casero, R.A., y Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941–953.

Shin, S., y Janknecht, R. (2007). Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D. *Biochem. Biophys. Res. Commun.* 359, 742–746.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79.

Soto-Reyes, E., González-Barrios, R., Cisneros-Soberanis, F., Herrera-Goepfert, R., Pérez, V., Cantú, D., Prada, D., Castro, C., Recillas-Targa, F., y Herrera, L.A. (2012). Disruption of CTCF at the miR-125b1 locus in gynecological cancers. *BMC Cancer* 12, 40.

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.

Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S., y Lieberman, P.M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* 27, 654–666.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., y Lee, J.T. (2013). Jpx RNA activates Xist by evicting CTCF. *Cell* 153, 1537–1551.

Tao, W., y Levine, A.J. (1999). P19(ARF) stabilizes p53 by blocking nucleo-cytoplasmic shuttling of Mdm2. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6937–6941.

Teif, V.B., Beshnova, D.A., Vainshtein, Y., Marth, C., Malm, J.-P., Höfer, T., y Rippe, K. (2014). Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res.* **24**, 1285–1295.

Thompson, P.M., Gotoh, T., Kok, M., White, P.S., y Brodeur, G.M. (2003). CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. *Oncogene* **22**, 1002–1011.

Torrano, V., Navascués, J., Docquier, F., Zhang, R., Burke, L.J., Chernukhin, I., Farrar, D., León, J., Berciano, M.T., Renkawitz, R., et al. (2006). Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribose)ylation-dependent mechanism. *J. Cell Sci.* **119**, 1746–1759.

Travers, A.A., Vaillant, C., Arneodo, A., y Muskhelishvili, G. (2012). DNA structure, nucleosome placement and chromatin remodelling: a perspective. *Biochem. Soc. Trans.* **40**, 335–340.

Vavouri, T., y Lehner, B. (2012). Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol.* **13**, R110.

Verrier, L., Escaffit, F., Chailleux, C., Trouche, D., y Vandromme, M. (2011). A new isoform of the histone demethylase JMJD2A/KDM4A is required for skeletal muscle differentiation. *PLoS Genet.* **7**, e1001390.

Vostrov, A.A., Taheny, M.J., y Quitschke, W.W. (2002). A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter. *J. Biol. Chem.* **277**, 1619–1627.

Wagner, E.J., y Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126.

Wallace, J.A., y Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* **17**, 400–407.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688.

Wang, L., Niu, N., Li, L., Shao, R., Ouyang, H., y Zou, W. (2018). H3K36 trimethylation mediated by SETD2 regulates the fate of bone marrow mesenchymal stem cells. *PLoS Biol.* **16**, e2006522.

Wang, X., Lau, K.K.K., So, L.K.Y., y Lam, Y.W. (2009). CHD5 is down-regulated through promoter hypermethylation in gastric cancer. *J. Biomed. Sci.* **16**, 95.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801.

Weth, O., Paprotka, C., Günther, K., Schulte, A., Baierl, M., Leers, J., Galjart, N., y Renkawitz, R. (2014). CTCF induces histone variant incorporation, erases the H3K27me3 histone mark and opens chromatin. *Nucleic Acids Res.* **42**, 11941–11951.

Whetstine, J.R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., et al. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467–481.

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

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., y Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7145–7150.

Yamamoto, S., Wu, Z., Russnes, H.G., Takagi, S., Peluffo, G., Vaske, C., Zhao, X., Moen Volla, H.K., Maruyama, R., Ekram, M.B., et al. (2014). JARID1B is a luminal lineage-driving oncogene in breast cancer. *Cancer Cell* **25**, 762–777.

Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., y Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**, 483–495.

Yang, X.-J. (2004). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* **32**, 959–976.

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

Zampieri, M., Guastafierro, T., Calabrese, R., Ciccarone, F., Bacalini, M.G., Reale, A., Perilli, M., Passananti, C., y Caiafa, P. (2012). ADP-ribose polymers localized on Ctf-Parp1-Dnmt1 complex prevent methylation of Ctf target sites. *Biochem. J.* **441**, 645–652.

Zhang, D., Yoon, H.-G., y Wong, J. (2005). JMJD2A is a novel N-CoR-interacting protein y is involved in repression of the human transcription factor achaete scute-like homologue 2 (ASCL2/Hash2). *Mol. Cell. Biol.* **25**, 6404–6414.

Zhao, R., Yan, Q., Lv, J., Huang, H., Zheng, W., Zhang, B., y Ma, W. (2012). CHD5, a tumor suppressor that is epigenetically silenced in lung cancer. *Lung Cancer Amst. Neth.* **76**, 324–331.

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

20 ARTÍCULOS PUBLICADOS.

- **PUBLICACIÓN 1:** The role of the histone demethylase KDM4A in cáncer.
- **AUTORES:** Guerra -Calderas L, González-Barrios R, Herrera LA, Cantú de León D & Soto-Reyes E.
- **REVISTA:** Cancer Genetics. **VOLÚMEN:** 208. **PÁGINAS:** 215–224. **AÑO:** 2015.

REVIEW ARTICLE

The role of the histone demethylase KDM4A in cancer

Lissania Guerra-Calderas¹, Rodrigo González-Barrios¹, Luis A. Herrera, David Cantú de León, Ernesto Soto-Reyes*

Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología-Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

Histone posttranslational modifications are important components of epigenetic regulation. One extensively studied modification is the methylation of lysine residues. These modifications were thought to be irreversible. However, several proteins with histone lysine demethylase functions have been discovered and characterized. Among these proteins, KDM4A is the first histone lysine demethylase shown to demethylate trimethylated residues. This enzyme plays an important role in gene expression, cellular differentiation, and animal development. Recently, it has also been shown to be involved in cancer. In this review, we focus on describing the structure, mechanisms, and function of KDM4A. We primarily discuss the role of KDM4A in cancer development and the importance of KDM4A as a potential therapeutic target.

Keywords KDM4A, histone demethylase, cancer, Jmjd2A, therapeutic target

© 2015 Elsevier Inc. All rights reserved.

Eukaryotic DNA is packed in a complex composed of RNA and proteins known as chromatin. The fundamental unit of this complex is the nucleosome, which is composed of 147 bp of DNA wrapped in 1.67 superhelical turns around the octameric histone core (composed of one pair each of histones H2A, H2B, H3, and H4). The DNA structure in chromatin leads to a five- to 10-fold DNA compaction (1). These compact structures negatively affect gene expression (2) and are modulated by several mechanisms, including histone posttranslational modifications such as the methylation of lysine and arginine residues, acetylation, the phosphorylation of serine and threonine, ADP-ribosylation, and the ubiquitination and SUMOylation of lysines. These modifications occur mainly at the histone N-terminal tail and promote either chromatin relaxation or compaction into a heterochromatin structure, affecting chromatin architecture and therefore gene transcription (3). One of the most-studied histone modifications is acetylation, which is controlled by acetyl transferases and deacetylases, suggesting that acetylation is a dynamic histone mark (4).

Lysine methylation is another prominently studied covalent histone modification. This histone mark can be recognized by at least four protein motifs: the chromodomain, the plant homeodomain zinc finger PHD, the Tudor domain, and the WDR40-repeat domain (5e7). Proteins that contain these motifs are recruited by specific methylated lysines. However, the mechanism becomes more complex because lysine residues can be mono-, di-, or trimethylated, and the binding affinity of a protein for a particular modification might be affected by an adjacent modification (8,9). Histone lysine and arginine methylation were believed to be stable and irreversible modifications (10). However, approximately 30 enzymes capable of removing this covalent modification have been discovered to date. The search for histone demethylases began in the 1960s, when an enzyme that could remove a methyl group from mono- and dimethylated lysine residues was reported (11). Years later, the same research group partially purified a protein that had histone demethylase function (12,13); nevertheless, the molecular identity of this enzyme was not fully known for several decades. Not until 2004 was the first histone demethylated lysine-specific demethylase 1 (LSD1), later renamed lysine (K) demethylase 1 (KDM1) identified and characterized (14). This enzyme can remove the methyl groups from lysines 4 and 9 of the histone (H3K4me2/1 and H3K9me2/1, respectively), suggesting that this protein plays a role in the dynamic structure of chromatin and transcription (15,16). KDM1 belongs to the oxidase family that includes enzymes that can demethylate

Received April 14, 2014; received in revised form October 20, 2014; accepted November 5, 2014.

* Corresponding author.

E-mail address: ctcf@ciencias.unam.mx

¹ Both authors contributed equally to this work and both are considered first authors.

2210-7762/\$ - see front matter © 2015 Elsevier Inc. All rights reserved.
<http://dx.doi.org/10.1016/j.cancergen.2014.11.001>

mono- and dimethylated residues using flavin adenine dinucleotide (FAD) as an electron acceptor (15,17). The oxygenase family, also known as the Fe(II) oxygenases, can demethylate mono-, di-, and trimethylated residues; this type of enzyme uses diatomic oxygen and α -oxoglutarate as cosubstrates (18–21). Lysine (K)-specific demethylase 4A (KDM4A, also known as JMJD2A, JHDM3A, and KIA0677) is categorized as a member of the Fe(II) oxygenase family.

In this review, we focus on the structure and function of the KDM4A protein, its role in cancer development, and the importance of this enzyme as a therapeutic target. For further review of the KDM4 family, see Shi et al., Whetstone et al., and Berry et al. (20–22).

KDM4A protein structure and enzymology

The *KDM4A* gene is a member of the Jumonji domain 2 (*JMJD2*) family and encodes a protein that contains JmjC and JmjN domains that form a composite active site, two PHD-type zinc finger domains, and two hybrid Tudor domains that form a bilobal structure, with each lobe resembling a normal Tudor domain (Figure 1A) (24,25). The function of the PHD fingers of KDM4A is not yet clear (22), in contrast to the functions of the PHD fingers present in other proteins, such as those in the NURF complex, which are known to bind to the H3K4me3 histone mark (26). The hybrid Tudor domains are formed by the exchange of the β 3 and β 4 chains; therefore, the electrostatic potential of the second Tudor domain is more negative than that of the first domain (21,27,28). Because of the folding of the hybrid Tudor domains of KDM4A, the side chain of H3K4me3 is inserted into the aromatic cage pocket of one Tudor domain, whereas the side chains of the other Tudor domain form intermolecular contacts; these domains also bind H4K20me3 peptides but in the opposite direction (28–32). Additionally, in vitro assays have demonstrated that this enzyme can demethylate di- and trimethylated residues at lysines 9 and 36 of histone 3 (H3K9me3/2 and H3K36me3/2, respectively), but this enzyme cannot demethylate monomethylated residues; in vivo however, KDM4A demethylates only trimethylated residues (18). KDM4A also has a higher affinity for H3K9me3 than for H3K36me3 (21,27,28). In particular, the H3K9me3 mark is associated with heterochromatic regions and transcriptional repression (8). Although H3K36me3 is associated with transcriptional repression in some models, it is primarily involved in transcription elongation by the RNA polymerase II, transcription initiation, alternative splicing, and DNA repair and recombination. For further review, see Pradeepa et al. (33).

Interestingly, the unusual KDM4A specificity for two regions with different sequences can be explained because the interplay between the enzyme and the histone peptides is governed by weak interactions such as hydrogen bonds and van der Waals interactions and by interactions with substrate backbone peptides (18,34). In addition, the N-terminal residues of H3K9me3 and H3K36me3 share a similar β -chain conformation, and the peptides bind in the same direction within the substrate-binding site (18,34). Thus, the trimethyl lysine is deposited in the catalytic site, which has an Fe(II) ion that is essential for the catalytic activity of the enzyme (18,34).

The proposed reaction mechanism of KDM4A is very similar to that of other Fe(II)-containing and α -ketoglutarate-dependent hydroxylases (Figure 1B). This process involves five general steps (35). (1) First, the active unbound Fe(II) ion is in a +2 oxidation state and is coordinated by two histidine residues, one glutamate residue, and three molecules of water. (2) Second, the α -ketoglutarate and diatomic oxygen are coordinated to the iron center, displacing the water molecules. (3) Third, a single electron transfer occurs from the Fe(II) ion to the oxygen molecule, leading to the formation of a peroxide radical that attacks the α -ketoglutarate and yields a mixed anhydride that is attached to the Fe³⁺-hydroxyl radical. (4) Fourth, this highly reactive radical activates the carbon-hydrogen bond of the methyl group located on the methyl lysine by removing a proton and transferring the hydroxyl group to the carbon atom of the methyl group, leading to hydroxymethyl lysine formation. (5) Finally, the demethylation reaction proceeds with the spontaneous loss of formaldehyde from the hydroxymethyl lysine because the carbonyl is a good leaving group. Due to the hydroxyl group transfer, which leaves a gap in the coordination sphere of the Fe²⁺, the mixed anhydride dissociates, producing succinate and carbon dioxide as byproducts. The union of three water molecules with the Fe²⁺ regenerates the original catalytic site (35).

In vitro studies have described the kinetic parameters of the KDM4A catalytic site (cKDM4A) (Figure 1C) (23); the k_{cat}/K_M (k_{cat} as the catalytic constant and K_M as the Michaelis constant) values represent how fast the enzyme reacts with the substrate once it encounters the substrate, where the values are proportional to the catalytic efficiency. The k_{cat}/K_M values of the dimethylated and trimethylated peptide ($2.4 \times 10^{-3} (\mu\text{mol/L})^{-1} \text{min}^{-1}$ and $3.0 \times 10^{-2} (\mu\text{mol/L})^{-1} \text{min}^{-1}$, respectively) show that cKDM4A has a stronger preference for the trimethylated substrate than the dimethylated substrate. Furthermore, a comparison of the k_{cat}/K_M values for a modified nucleosome and an analogue trimethylated peptide that has an aminoethylcysteine but behaves in a way similar to that of natural lysine residues suggests that the catalytic site of KDM4A predominantly recognizes the residues immediately surrounding the H3K9 and not additional structures on the nucleosome (Figure 1C). These data suggest that the catalytic site of KDM4A acts in a distributive manner and that the recognition of other chromatin features or modifications by the double Tudor or PHD domains of the entire demethylase may result in a tighter association, additional interactions, and an increase in demethylase activity (23).

Such interactions and protein-protein cross talk may play an important role in the regulation of KDM4A activity and processivity. In vivo, in the presence of chromatin histone marks or protein partners, the entire KDM4A may demethylate in a processive manner, and this regulation of KDM4A has significant implications on the specific output of KDM4 proteins in a context-dependent manner (31,36–39). Additionally, the demethylation activity toward H3K9me3 is influenced by other posttranslational modifications on the same peptide. Further studies of these cross-talk interactions at the peptide level are needed to obtain a more accurate understanding of the dynamics of epigenetic marks (40). Due to its catalytic activity, interactions, particular structure, and recognition ability, several functions have been attributed to KDM4A. Below, we describe some functions of KDM4A.

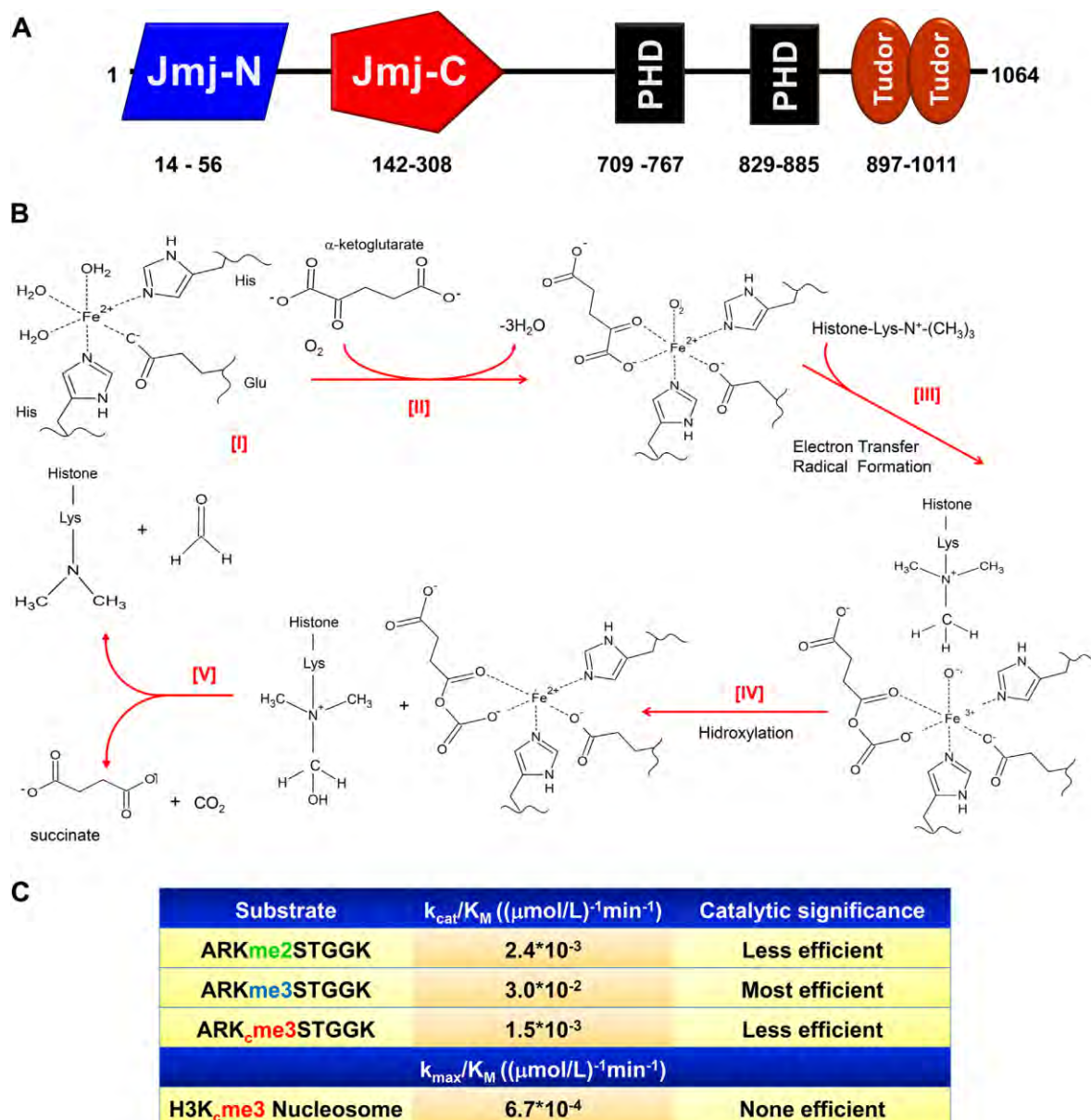


Figure 1 The KDM4A demethylase protein structure and its biochemical mechanisms. (A) The KDM4A protein architecture consists of one JmjN domain and one JmjC domain, which contain an Fe(II) ion, two PHD domains, and two hybrid Tudor domains in the catalytic site. (B) The proposed reaction mechanism of KDM4A is very similar to that of other Fe(II)-containing and α -ketoglutarate-dependent hydroxylases. More details can be found in the text. The structures were created using ACD/ChemSketch from ACD/Labs Freeware, Toronto, Canada. (C) Kinetic parameters (k_{cat}/K_M) of KDM4A catalytic site and the catalytic significance of the values; ARK_{me2}STGGK and ARK_{me3}STGGK correspond to the H3K9me2 and H3K9me3 peptides, respectively; trimethyllysine analogue peptide and recombinant homogeneous H3K9me3 nucleosomes correspond to ARK_cme3STGGK and H3K_cme3 (23).

KDM4A functions

Through KDM4A activity, H3K9me3 demethylation promotes an open chromatin state, contributing to the transcription activation of promoter regions (Figure 2A) (48). Regarding the functional impact of KDM4A-mediated demethylation of H3K36, the outcome is less clear.

Notably, H3K36 and H3K27 methylation are antagonistic histone marks, because nucleosomes that are methylated at H3K27 inhibit the enzymatic methylation of H3K36 and vice versa (49). Whereas H3K27 methylation is a characteristic repressive histone mark that is associated with the Polycomb group, H3K36me3 histone modification has been implicated in other processes that affect euchromatin functions. For

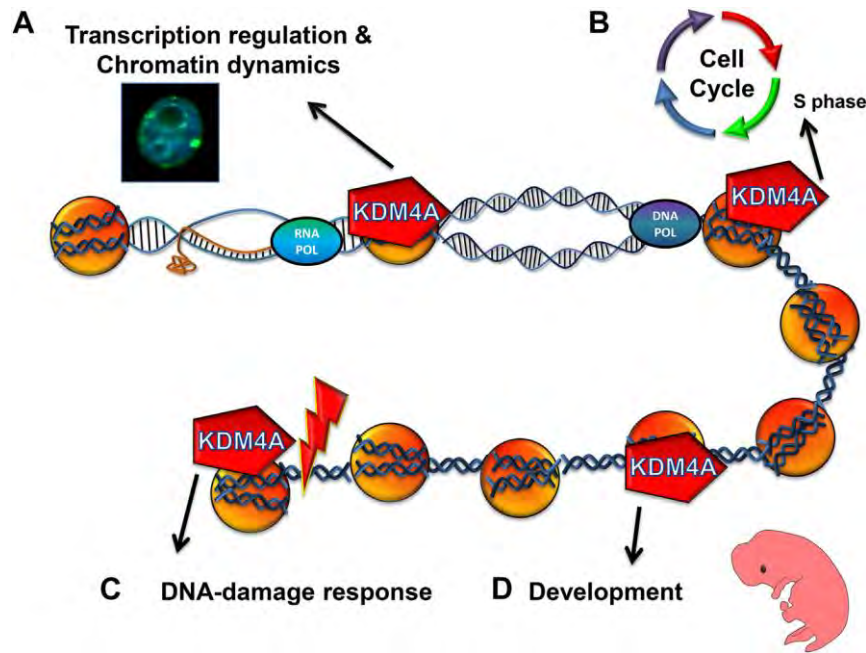


Figure 2 Biological implications of KDM4A. (A) KDM4A can act as a transcription repressor or activator and its chromatin dynamics, through its interactions with N-CoR, pRB, HDACs, and histones 3 and 4 (8,18,36,41–44). (B) In the cell cycle, KDM4A accelerates replication during S phase (45). (C) KDM4A associated with the DNA damage response avoids the recruitment of 53BP1 (46), the lightning bolt symbolizes DNA damage. (D) During development, KDM4A leads to the reduction of H3K36me3 in the X chromosome and is also involved in the activation of the *Myog* gene (47,43).

example, H3K36me3 is recognized by proteins that antagonize exon definition, affecting alternative splicing (50,51).

The role of H3K36me3 in transcription is controversial. Whereas some investigators have demonstrated that H3K36me3 couples with RNA polymerase II (RNA Pol II) in transcription elongation (52), others have shown that H3K36me3 is recognized by DNMT3A, promoting DNA methylation at nearby DNA regions and suggesting a DNMT3A-mediated gene repression link to H3K36me3 (53).

KDM4A is implicated in replication timing and genomic stability (45). *KDM4A* overexpression in human cells increases chromatin accessibility, coinciding with accelerated replication during S phase (Figure 2B). In contrast, a mutation in the *Caenorhabditis elegans* orthologue leads to increased replication timing and DNA damage, which induces p53-dependent apoptosis. KDM4A abundance is cell cycle-dependent, and this protein antagonizes the function of heterochromatin protein 1 gamma (HP1 γ) (45).

Additionally, KDM4A is involved in the DNA damage response through the tandem Tudor domains of KDM4A and KDM4B that bind to the preexisting methylated residues of H4. After DNA damage, KDM4A/B is ubiquitinated by RNF8 and RNF168 and degraded by the proteasome, allowing the binding of 53BP1 to H4K20me2. Furthermore, *KDM4A* overexpression abrogates 53BP1 recruitment to DNA damage sites, suggesting a possible role of KDM4A in the DNA damage response (Figure 2C) (46).

In *C. elegans*, KDM4A appears to be involved in H3K36me3 reduction on the X chromosome, suggesting that this protein has a relevant role in germ cell development (Figure 2A) (47). In addition, in HeLa cells, KDM4A is associated with the repression of the achaete-scute complex

homologue 2 (*ASCL2*) gene by acting as a cofactor of the nuclear receptor corepressor (N-CoR); this function requires its demethylase activity (18,41). KDM4A also interacts with histone deacetylases (HDACs) and retinoblastoma protein (pRb); this partnership is involved in the repression of E2F-dependent promoters (Figure 2A). However, the role of this protein as a demethylase has not been studied in this context (42). Remarkably, H3K9me3 demethylation of the myogenin (myogenic factor 4, or *Myog*) gene promoter during skeletal muscle differentiation of myoblasts into myotubes is performed by a Δ N-KDM4A isoform (Figure 2D) (43). These data suggest that functional KDM4A isoforms might also play a major role in the regulation of gene expression.

Genes repressed or activated by *Drosophila melanogaster* KDM4A (dKDM4A) were reported to not require KDM4A catalytic activity for expression; nevertheless, activated genes require its demethylase activity for proper expression (44). These findings suggest that some of these genes are indirect dKDM4A targets and that epigenetic regulation can be either dependent on or independent of the demethylase catalytic activity (44).

One study demonstrated that some dKDM4A-regulated genes were near one another, suggesting that genes controlled by this enzyme may require a common chromatin environment (44).

Interestingly, chromatin immunoprecipitation (ChIP) assays showed that the H3K36me3 levels were very low in both wild types and dKDM4A mutants (P/DdKDM4A) and that no detectable amounts of H3K9me3 were present in either wild types or mutants. These results indicate that many of the dKDM4A-controlled genes are modulated not by these histone marks but by other dKDM4A-dependent functions. A

Table 1 Summary of the role of *KDM4A* in several cancer types

Cancer type	<i>KDM4A</i> expression ^a	Associated with	References
Prostate	↑	AR, PSA	(61)
Colorectal	↑	p53, p21	(62)
Lung	↑	CHD5, Ras, ADAM12, CXL5	(63,64)
Breast	↑	ER, C-Jun, Cyclin D1	(65)
Squamous cell carcinoma, lymph node metastases	↑	FOS1, JUN	(66)
Bladder	↓	Patient survival, new biomarker for patient risk stratification	(67)
Bladder	↑	Involved in an early stage of human bladder carcinogenesis	(63)

^a Arrows represent overexpression (↑), and underexpression (↓).

likely explanation for the changes found in mutants may be that they are due to the interaction of *KDM4A* with other proteins, such as pRb and N-CoR (Figure 2A) (41–44).

Histone modifications are important in chromatin structure. Global and local chromatin architecture alterations are common findings in tumors (54,55). The expression pattern of *KDM4A* has been suggested to be altered in several cancer types that involve such chromatin modifications. Here, we summarize some aspects of the role of *KDM4A* in cancer development.

The role of *KDM4A* in cancer development

KDM4A in genomic stability

Chromosomal instability and copy number alterations are common features in cancer (56); nevertheless, there is little information regarding the molecular mechanisms that demonstrate how copy number variations (CNVs) are involved in the timing of tumor progression. However, recent studies have demonstrated that the overexpression of *KDM4A* leads to focalized copy gains at the 1q12, 1q21, and Xq13.1 loci (57).

The 1q12 and 1q21 regions harbor several putative oncogenes (58,59) and are often amplified in multiple myeloma and lung cancer. In addition, primary tumors of different cancer types that exhibit *KDM4A* overexpression also have increased copy gains at 1q12, 1q21, and Xq13.1; however, surprisingly, it has been suggested that the *KDM4A*-mediated copy gain does not cause chromosome instability (57). The 1q12 loci copy gain also appears to occur in one cell cycle and is not stably inherited by daughter cells. Moreover, the sites with amplified copy numbers are re-replicated and have increased occupancy by DNA polymerase, *KDM4A* and MCM (57). The 1q12/1q21 copy gains may be associated with drug resistance in multiple myeloma and ovarian cancer cell lines (60).

These results suggest that 1q12 copy gains are not incorporated into the genome but exist as extrachromosomal DNA. Additionally, these results establish how copy number changes originate during tumorigenesis and provide evidence showing that the overexpression of specific chromatin modulators promotes these events (57).

KDM4A has been reported to be deregulated in several cancer types, such as prostate, bladder, colorectal, squamous cell carcinoma, and lung and breast cancers.

Prostate, bladder, and colorectal cancer

The overexpression of *KDM4A* has been observed in prostate cancer (Table 1), with *KDM4A* functioning as a coactivator of the androgen receptor (AR) under simulated conditions of low AR ligand levels. *KDM4A* appears to activate the basal transcription of prostate-specific antigen (PSA) (61).

These results could impact patients undergoing androgen ablation. *KDM4A* overexpression has been suggested to be involved in prostate tumors that become refractory to androgen ablation therapy (61).

In contrast, *KDM4A* levels were shown to be reduced in bladder cancer (Table 1). *KDM4A* and AR are absent in primary and advanced bladder tumors, suggesting that these proteins are involved in neither initiation nor tumor progression; however, these proteins might be involved in delaying the onset of carcinogenesis. The physiological significance of the AR and *KDM4A* losses in androgen signaling remains to be determined (67). Furthermore, there is an association between decreased levels of *KDM4A* and patient smoking, with *KDM4A* greater presence being associated with patient survival. Protein loss is correlated with a particularly aggressive bladder disease and poor prognosis in bladder cancer patients. These data suggest evidence of a possible new biomarker for patient risk stratification (67).

Although a recent study demonstrated that the expression levels of *KDM4A* are upregulated in bladder cancer tissue compared with normal bladder tissue, no significant differences among different tumor grades have been found. This finding suggests that the elevated expression of *KDM4A* could be involved in an early stage of human bladder carcinogenesis (63).

As in prostate cancer, *KDM4A* overexpression has been observed in colorectal tumors (Table 1). ChIP assays have shown that *KDM4A* and p53 are recruited to the *p21* gene promoter after Adriamycin-induced DNA damage. *KDM4A* reduction leads to increases in p53, p21, and the proapoptotic protein PUMA, thus inducing apoptosis in the HCT116 cell line model (62). Interestingly, *KDM4A* knockdown results in reduced cell proliferation, whereas *KDM4A* overexpression correlates with cell proliferation under low serum concentration conditions (61). Therefore, *KDM4A* overexpression could be advantageous in tumors, as they are often surrounded by stroma and extracellular matrix that limit the diffusion of growth factors, which resembles a low serum

environment. Thus, researchers have proposed that KDM4A promotes cell proliferation and survival in colon cancer (59) and that KDM4A inhibition may sensitize cells to chemotherapeutics such as Adriamycin (62).

Head and neck squamous cell carcinoma

The transcription factor activating protein 1 (AP-1) plays a critical role in metastasis and tumor growth. AP-1 is composed of two proteins, c-Jun and c-Fos. AP-1 activation can be partially mediated by the transcriptional activation of *JUN* and *FOS*. In addition, *JUN* and *FOS* undergo positive feedback with the recruitment of AP-1 to their own gene promoters, thereby enhancing AP-1 activation (68–71). When H3K9me3 is enriched in this region, AP-1 cannot be recruited. The demethylation of this histone mark, mediated by KDM4A, can promote the gene activation of *JUN* and *FOSL1* (66).

Furthermore, the abundance of this enzyme correlates with the abundance of JUN and FOSL1, increasing the activity of AP-1 in human squamous cell carcinoma tissues (66). Remarkably, *KDM4A* is overexpressed in lymph node metastases (66) and squamous cell carcinoma tissue compared with expression levels in normal tissues (Table 1) (63). These data suggest that KDM4A could be involved in squamous cell carcinoma invasion and metastasis of the head and neck (66).

Lung cancer

KDM4A is overexpressed in mouse and human lung cancer cell lines (Table 1). The depletion of KDM4A in the human lung cancer cell line A549, which bears an activated K-Ras allele, triggers senescence. Therefore, *KDM4A* could function as an oncogene that represents a target for Ras-expressing tumors (64). Additionally, KDM4A appears to be involved in the regulation of the tumor suppressor gene chromodomain helicase DNA binding protein 5 (*CHD5*). *CHD5* targets p19ARF, which is involved in the p53 ubiquitination pathway (72). Thus, *KDM4A* overexpression may cooperate with Ras in the transformation of primary cells by blocking cellular p53-dependent senescence via *CHD5* in lung adenocarcinomas (64).

Furthermore, the nuclear presence of KDM4A in neoplastic tissues such as lung carcinomas and non-small cell lung carcinomas (NSCLC) was detected, unlike in normal lung tissue (63,64). Surprisingly, no association between *KDM4A* expression and prognosis was observed. The above data suggest that *KDM4A* overexpression may be an early event in NSCLC carcinogenesis (63).

New candidate genes that appear to be upregulated by KDM4A through the demethylation of H3K9me3 have recently been reported in the A549 cell line, including three cancer-related genes, *ADAM12*, *CXCL5*, and *JAG1* (63). *ADAM12* is overexpressed in several types of human carcinomas (73–75) and enhances tumor cell growth by the proteolytic shedding of EGFR ligands (76). The *CXCL5* gene may be implicated in the promotion of tumor growth, progression, and angiogenesis (77). *JAG1* encodes a ligand involved in the Notch intracellular pathway and angiogenesis (78) and has also been implicated in enhancing cell

proliferation by activating the canonical Notch signaling pathway (79).

Taken together, these results suggest that KDM4A may have a dual role in lung carcinogenesis by downregulating the tumor suppressor gene *CHD5* (64) and by activating tumor growth- and cell proliferation-related genes (63).

Breast cancer

In triple-negative breast tumors, the overexpression of *KDM4B* and *KDM4A* has been observed to correlate with the loss of H3K9me3, which is normally enriched in the pericentromeric region. This phenomenon may contribute to the development of aneuploidy and chromosomal instability in solid tumors and thus to tumor progression (80). However, other factors may cause increased KDM4A expression and promote chromosomal instability due to the loss of H3K9me3 in pericentromeric regions, such as the downregulation of the expression of the methyltransferase SUV39H1/2 (81) or of complexes that help correct chromosome segregation and tumor suppression, such as pRb, SWI/SNF, and mSds3 (82–85).

A study comparing breast cancer tissue and normal breast tissue found significant differences in several proteins that modify histones, including KDM4A (Table 1). These differences were associated with pathological and clinical parameters. However, further studies are required to determine the biological and clinical significance of this altered expression for each histone-modifier gene and for the different expression profile combinations (86). Moreover, the depletion of KDM4A by siRNA in breast cancer cell lines suppresses tumor proliferation, invasion, and migration (87,88).

Similarly, KDM4A has been proposed as an estrogen receptor coactivator (ER α) that forms a KDM4A-ER α complex, by which the overexpression of *KDM4A* increases estrogen-dependent transcription. Meanwhile, KDM4A depletion causes a transcriptional decrease of ER α target genes such as *CCND1*, which is overexpressed in breast cancer (89). Another protein that is also decreased after *KDM4A* downregulation is c-Jun. The inactivation of c-Jun causes cell cycle arrest. This protein, which is regulated by ER α , has important functions in cancer tissues, and its overexpression stimulates the invasion, migration, and formation of tumors (90).

Taken together, these associations suggest that KDM4A can coactivate both hormone signaling-dependent and signaling-independent genes and that it may regulate cell growth by influencing the expression of at least two oncogenes, *CCND1* and *c-Jun* (65).

Recently, *KDM4A* has been revealed to have a higher expression in infiltrating duct carcinoma than in benign lesions in situ at the mRNA and protein levels (91). The same study showed a negative correlation between the expression levels of *KDM4A* and ADP-ribosylarginine hydrolase 1 (ARH1). In contrast, the expressions of *KDM4A*, *p53*, and *ER* were positively correlated. Although the exact mechanism of KDM4A's involvement in human breast cancer is not yet clear, these results suggest that KDM4A has a role in the diagnosis of cancer and as a possible therapeutic target (91).

The ability of KDM4A to activate or repress transcription may be dictated by chromatin structure, the presence or

absence of other transcriptional regulators, stressful stimuli such as hormonal stimulation, and transcription factor recruitment (64,65). These data suggest that the reduction or inhibition of KDM4A may be beneficial for the treatment of different cancer types.

KDM4A as a potential therapeutic target

Given our understanding of the structures of demethylases, their catalytic reaction mechanisms, the selectivity of their methylation marks, and the implications of these marks in cancer, a great interest in developing inhibitors of these demethylases has emerged.

Many histone demethylase inhibitors have been described; these inhibitors can be classified into five groups: iron chelators, α -ketoglutarate analogues, catalytic site inhibitors, prodrugs, and zinc chelators (Table 2) (92–104). However, the lack of research on the selectivity and specificity of histone demethylases and thus the deficiency of our knowledge regarding undesirable targets have prevented these inhibitors from progressing toward clinical research. Therefore, the use of such inhibitors remains in the preclinical phase (105).

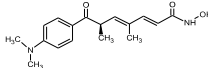
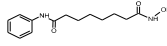
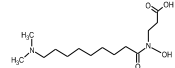

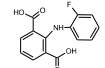

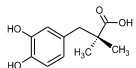
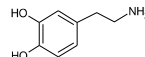
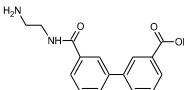
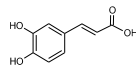
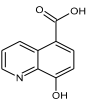
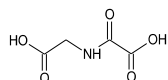
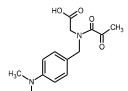
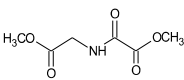
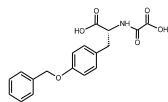
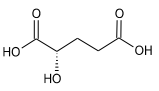
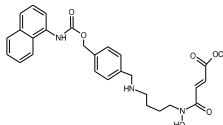
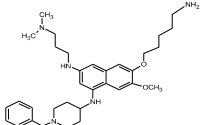
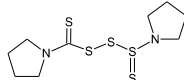
These reported inhibitors include N-oxalylglycine (NOG) and its derivatives, which are analogues of the cosubstrate α -ketoglutarate. They inhibit KDM4A and other members of

the KDM4 family via competition. In addition, another analogue of α -ketoglutarate is the oncometabolite 2-hydroxyglutarate, which also inhibits KDM4 enzymes by competition but is a weak antagonist of α -ketoglutarate (92–94). However, these chemicals are not very specific, because they target different α -ketoglutarate-dependent enzymes (92–94).

Another molecule that has inhibitory action is pyrimidine 2,4-dicarboxylic acid, whose mechanism of action is based on electrostatic interactions with a lysine residue within the active site; nevertheless, little is known about its uses in therapy (105). Additionally, hydroxamic acid and its derivatives, such as trichostatin A, most likely function as iron chelators to inhibit the catalytic activity of JmjC domain-containing demethylases such as KDM4A (65). Moreover, the main use of these compounds is against HDACs (106); therefore, the probable side effects make the future development of this class of compounds unpromising (105).

The structures of KDM4A revealed a Cys-His Zn(II) binding site that is close to the substrate binding spot, which bioinformatic analyses indicated was not present in any other histone demethylase subfamily. Therefore, an alternative method to inhibit the KDM4 family (95) would be to use compounds that chelate Zn(II) ions. One derivative of disulfiram is a potent KDM4A inhibitor; it changes the methyl lysine-binding site by the chelation of Zn(II) ions (95). This strategy may have potential for the development of selective inhibitors for those Jumonji protein subtypes containing a

Table 2 Classification of JumonjiC domain-containing demethylase inhibitors, according to the inhibition mechanism

Iron chelators					
 Trichostatin [94].	 Suberoylanilide hydroxamic [94].	 Internal hydroxamic acid [98].	 Pyridine 2,4-dicarboxylic acid [97].	 PDCA derivative [97].	 β -lapachone [99].
 [S]-[-]-carbidopa [99].	 Dopamine [99].	 8-hydroxyquinoline-5-carboxylic [96].	 3,4-dihydroxy cinnamic acid [101].	 2,2'-bipyridine derivative [100].	
α -ketoglutarate analogs					
 N-oxalylglycine [93].	 NOG-derivative [93].	 Dimethyl N-oxalylglycine [DMOG] [93].	 N-oxalyl D-tyrosine derivative [102].	 [S]-2-hydroxyglutaric acid [2-HG] [92].	
Prodrug		Binds to the active site		Zinc chelators	
 Methylstat [103].		 E67 [104].		 Disulfiram derivative [95].	

More details can be found in the text. The structures were created using ACD/ChemSketch from ACD/Labs Freeware, Toronto, Canada.

structural Zn(II) ion (KDM4) (96). Different chelants that may be involved in KDM4 inhibition are 8-hydroxyquinoline and its derivatives; 8-hydroxyquinoline chelates the Fe(II) ion with a bidentate structure and executes its inhibitory action via a carboxylic acid motif positioned toward the active site (97). Interestingly, there is a new compound that performs its inhibitory action by similarly chelating the Fe(II) ion and binding to the cosubstrate cleft (107). This compound consists of a peptide and an α -ketoglutarate analogue that are connected by a disulfide bridge. Although studies have revealed potent and partially high selectivity of this compound, its peptide nature may be an obstacle to the further development of these compounds into potential drugs due to cell permeability, intracellular stability, and other pharmacokinetic parameters (105).

The search for novel histone demethylase inhibitors provides a starting point for the development of new therapies with selective agents against aberrant epigenetic phenomena. However, further research is still needed for such therapies to become a reality.

Conclusions and final remarks

KDM4A has a dual role as an epigenetic transcriptional regulator. Knowing which conditions are required for it to activate or suppress a gene would be extremely interesting. Due to its structure, interactions, and pleiotropic activity, the role of KDM4A in cancer development may be more complex than originally believed. Further research is required to clarify how KDM4A affects cancer development and to create a comprehensive overview of the functions performed by this protein in various cancer types.

Another challenge is designing a drug that is selective for a subset of demethylases. This selectivity could potentially be achieved by linking the drug to at least three protein domains; however, the designed drug would be too large to penetrate the cellular membrane. In addition, an allosteric inhibitor that changes the conformation of the catalytic site of the enzyme without binding to this site could be designed. A recent computational screen identified putative allosteric sites that could be used for this purpose (108). However, more research is required to further experimentally identify and characterize these sites.

Acknowledgments

This work was supported by the Consejo Nacional de Ciencia y Tecnología (grant number 83959 and 182997) and the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, Universidad Nacional Autónoma de México (grant number IN213311). L. Guerra-Calderas was supported by an undergraduate fellowship from Básica SEP-CONACyT (19071).

References

- Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. *Nature* 2003;423:145–150.
- Luger K, Hansen JC. Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol* 2005;15:188–196.
- Berger SL. The complex language of chromatin regulation during transcription. *Nature* 2007;447:407–412.
- Kouzarides T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* 2000;19:1176–1179.
- Bannister AJ, Schneider R, Kouzarides T. Histone methylation: dynamic or static? *Cell* 2002;109:801–806.
- Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 2005;6:838–849.
- Li H, Ilin S, Wang W, et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 2006;442:91–95.
- Lachner M, O'Carroll D, Rea S, et al. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 2001;410:116–120.
- Rea S, Eisenhaber F, O'Carroll D, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000;406:593–599.
- Byvoet P, Shepherd GR, Hardin JM, et al. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch Biochem Biophys* 1972;148:558–567.
- Kim S, Benoiton L, Paik WK. Epsilon-alkyllysine. Purification and properties of the enzyme. *J Biol Chem* 1964;239:3790–3796.
- Paik WK, Kim S. Enzymatic demethylation of calf thymus histones. *Biochem Biophys Res Commun* 1973;51:781–788.
- Paik WK, Kim S. Epsilon-alkyllysine. New assay method, purification, and biological significance. *Arch Biochem Biophys* 1974;165:369–378.
- Allis CD, Berger SL, Cote J, et al. New nomenclature for chromatin-modifying enzymes. *Cell* 2007;131:633–636.
- Shi Y, Lan F, Matson C, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 2004;119:941–953.
- Metzger E, Wissmann M, Yin N, et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* 2005;437:436–439.
- Anand R, Marmorstein R. Structure and mechanism of lysine-specific demethylase enzymes. *J Biol Chem* 2007;282:35425–35429.
- Klose RJ, Yamane K, Bae Y, et al. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* 2006;42:312–316.
- Schneider J, Shilatifard A. Histone demethylation by hydroxylation: chemistry in action. *ACS Chem Biol* 2006;1:75–81.
- Shi Y, Whetstone JR. Dynamic regulation of histone lysine methylation by demethylases. *Mol Cell* 2007;25:1–14.
- Whetstone JR, Nottke A, Lan F, et al. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* 2006;125:467–481.
- Berry WL, Janknecht R. KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. *Cancer Res* 2013;73:2936–2942.
- Shiau C, Trnka MJ, Bozicevic A, et al. Reconstitution of nucleosome demethylation and catalytic properties of a Jumonji histone demethylase. *Chem Biol* 2013;20:494–499.
- Couture JF, Collazo E, Ortiz-Tello PA, et al. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Bio* 2007;8:689–695.
- Ng SS, Kavanagh KL, McDonough MA, et al. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* 2007;448:87–91.
- Kwon SY, Xiao H, Wu C, et al. Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities. *PLoS Genet* 2009;5:e1000574.

- **PUBLICACIÓN 2:** CTCF-KDM4A complex correlates with histone modifications that negatively regulate CHD5 gene expression in cancer cell lines. The role of the histone demethylase KDM4A in cáncer.
- **AUTORES:** Guerra-Calderas L, González-Barrios R, Patiño CC, Alcaraz N, Salgado-Albarrán M, León DC de, Hernández CC, Sánchez-Pérez Y, Maldonado-Martínez HA, Rosa-Velazquez IAD Ia, Vargas-Romero F, Herrera LA, García-Carrancá A, Soto-Reyes E, Guerra-Calderas L, González-Barrios R, Patiño CC, Alcaraz N, Salgado-Albarrán M, de León DC, Hernández CC, Sánchez-Pérez Y, Maldonado-Martínez HA, De la Rosa-Velazquez IA, Vargas-Romero F, Herrera LA, García-Carrancá A & Soto-Reyes E.
- **REVISTA:** Oncotarget. **VOLÚMEN:** 9. **PÁGINAS:** 17028–17042. **AÑO:** 2018.

CTCF-KDM4A complex correlates with histone modifications that negatively regulate *CHD5* gene expression in cancer cell lines

Lissania Guerra-Calderas¹, Rodrigo González-Barrios¹, Carlos César Patiño¹, Nicolás Alcaraz⁴, Marisol Salgado-Albarrán¹, David Cantú de León³, Clementina Castro Hernández^{1,2}, Yesennia Sánchez-Pérez¹, Héctor Aquiles Maldonado-Martínez⁵, Inti A. De la Rosa-Velazquez⁵, Fernanda Vargas-Romero⁷, Luis A. Herrera^{1,2}, Alejandro García-Carrancá^{1,2} and Ernesto Soto-Reyes¹

¹Cancer Biomedical Research Unit, Instituto Nacional de Cancerología (INCan) Mexico City, Mexico

²Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

³Clinical Research, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico

⁴The Bioinformatics Centre, Section for RNA and Computational Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark

⁵Department of Surgical Pathology, Instituto Nacional de Cancerología, Mexico City, Mexico

⁶Genomics Lab, Universidad Nacional Autónoma de México, Red de Apoyo a la Investigación-CIC and Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán" Mexico City, Mexico

⁷Instituto de Fisiología Celular-Neurociencias, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

Correspondence to: Ernesto Soto-Reyes, email: ctcf@ciencias.unam.mx

Keywords: KDM4A; CTCF; histone demethylation; H3K36me; CHD5

Received: August 10, 2017

Accepted: February 26, 2018

Published:

Copyright: Guerra-Calderas et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Histone demethylase KDM4A is involved in H3K9me3 and H3K36me3 demethylation, which are epigenetic modifications associated with gene silencing and RNA Polymerase II elongation, respectively. *KDM4A* is abnormally expressed in cancer, affecting the expression of multiple targets, such as the *CHD5* gene. This enzyme localizes at the first intron of *CHD5*, and the dissociation of KDM4A increases gene expression. *In vitro* assays showed that KDM4A-mediated demethylation is enhanced in the presence of CTCF, suggesting that CTCF could increase its enzymatic activity *in vivo*, however the specific mechanism by which CTCF and *KDM4A* might be involved in the *CHD5* gene repression is poorly understood. Here, we show that CTCF and KDM4A form a protein complex, which is recruited into the first intron of *CHD5*. This is related to a decrease in H3K36me3/2 histone marks and is associated with its transcriptional downregulation. Depletion of CTCF or KDM4A by siRNA, triggered the reactivation of *CHD5* expression, suggesting that both proteins are involved in the negative regulation of this gene. Furthermore, the knockout of *KDM4A* restored the *CHD5* expression and H3K36me3 and H3K36me2 histone marks. Such mechanism acts independently of *CHD5* promoter DNA methylation. Our findings support a novel mechanism of epigenetic repression at the gene body that does not involve promoter silencing.

INTRODUCTION

Gene regulation in eukaryotes is driven in part by chromatin architecture, where histone post-translational modifications play a major role in this process [1]. In particular, the methylation of lysine residues in histones is involved in transcriptional activation and repression, depending on specific lysines and the degree of methylation. For example, H3K4me3 and H3K36me3 are associated with transcriptional activation, while H3K9me3 and H3K27me3 are related with transcriptional repression [2].

Although, it was long thought that lysine methylation was a stable and irreversible process, recent reports have found approximately 25 enzymes capable of removing the methyl groups of lysines in histones. These enzymes are grouped into two families depending on their chemical mechanism of demethylation, the oxidases and the oxygenases [3]. The majority of histone demethylases belong to the second family, including lysine (K)-specific demethylase 4A (KDM4A). KDM4A actively removes the methyl groups from H3K36me3 to produce H3K36me2 [3]. In particular, H3K36me3 is enriched in genes that are transcriptionally active and is associated with recruitment of RNA polymerase II and transcriptional elongation, loss of H3K36me3 leads to transcriptional repression [4].

KDM4A is overexpressed in several types of cancer, including breast cancer [5]. One of the target genes of KDM4A is chromodomain helicase DNA binding protein 5 gene (*CHD5*). *CHD5* was identified as a tumor suppressor gene, and it has been reported deregulated in glioma, colon, lung, ovarian, prostate and breast cancers. Thus, based on its likely involvement as a tumor suppressor gene (TSG) in neuroblastomas, gliomas, and many common adult neoplasms, *CHD5* may play an important developmental role in many other tissues besides the nervous system and testis [6]. Particularly, this gene is involved in cell proliferation, apoptosis and senescence by regulating p19^{Arf}, modulating p53 activity [6]. KDM4A has been reported to negatively regulate *CHD5* by its recruitment to the first intron [7]. Neither the mechanism by which KDM4A negatively regulates *CHD5* nor the mechanism by which KDM4A is recruited to this target site are known. Furthermore, *in vitro* assays have shown that the demethylation frequency of KDM4A increases up to 80% in the presence of the architectural protein CTCF [8], suggesting that CTCF may play a major role in the activity of KDM4A which has not been addressed until now. Hence, the aim of this study was to elucidate the mechanism underlying the role of CTCF and KDM4A on histone modifications and in the downregulation of *CHD5*.

RESULTS

***KDM4A* is highly expressed in MCF7, MDA-MB-231 and HeLa cell lines**

As a first approach, we evaluated the expression of *KDM4A* in four different cell lines using RT-qPCR. We observed that *KDM4A* was highly expressed in MCF7 and MDA-MB-231 cell lines compared to the expression levels of the non-tumorigenic epithelial breast cell line MCF 10A (Figure 1A). Previously, *KDM4A* has been reported to be highly expressed in HeLa cells [9], hence we used this cell line as a positive control. Immunofluorescence assays show that KDM4A is located mainly at the nucleus in the neoplastic cell lines (Figure 1B), but it is not detected in the non-tumorigenic breast cell line MCF 10A (Figure 1B). We also observed *CHD5*, which has been reported to be regulated by KDM4A and highly expressed in the MCF 10A cells compared with MCF7, MDA-MB-231 and HeLa cells (Figure 1C) [7]. Additionally, *CHD5* is only detected in the MCF 10A cell line, where *KDM4A* is not present (Figure 1B and 1D). When looking into breast cancer cell line expression data available at the Cancer Cell Line Encyclopedia we found that 83.34% (50/60) of these cell lines show high expression of *KDM4A*, while not expressing *CHD5*. In this regard, MCF7 and MDA-MB-231 cell lines exhibit the same behavior that we observed previously in our results (Figure 1 and Supplementary Figure 1A) [10]. In contrast to what is observed in cell lines, we did not find a significant correlation between *KDM4A* and *CHD5* expression in breast cancer patients (Supplementary Figure 1B) from The Cancer Genome Atlas (TCGA). We argue that this could be due to the heterogeneity of the tumor tissue or tumor subtypes.

DNA methylation at the *CHD5* gene promoter is not the main mechanism of epigenetic silencing in the neoplastic cell lines

Some authors have reported that DNA methylation at *CHD5* gene promoter can alter the expression of this gene in several cancers and neoplastic cell lines [11, 12]. Thus, we analyzed the methylation status along the *CHD5* gene locus of 743 breast cancer patients and 98 normal samples obtained from TCGA (Illumina Human Methylation 450 K) through the TCGA wanderer web service [13]. This panel measures the methylation levels of 485,000 CpG sites distributed along the genome, of which 63 CpGs fall within the *CHD5* gene region (Figure 2A); of these sites, 8 CpGs are located within the gene promoter, the remaining 55 sites are distributed along the gene body. At the gene body, 34 CpGs are found to be methylated (having Beta-value ≥ 0.6 , which is considered as a methylated region) in 50% of the patients, and 20 of these 34 sites that are present at the gene body are methylated

in 80% of the patients. Nevertheless, when evaluating the mean methylation levels of the 8 CpG sites located within the gene promoter region (Highlighted part of the figure with a rectangle in Figure 2A) (Ensembl version 75), we observed that only 1 out of the 743 patients shows promoter methylation, where the CpG methylation Beta value is less than 0.6, indicating that *CHD5* gene promoter is considered as not methylated (Figure 2B). In order to determine if the absence of methylation in the *CHD5* promoter was restricted only to breast cancer, we also looked into the methylation status in other neoplasms such as Low-Grade Gliomas or Glioblastomas where we also did not find methylation at the promoter region (Supplementary Figure 1C). Hence, these datasets point out that DNA methylation at the promoter region is not related with *CHD5* gene silencing, suggesting that there may be other mechanisms related to its repression in breast cancer (Figure 2A).

Given the methylation status of *CHD5* gene in breast cancer patients found in TCGA, we aimed to characterize DNA methylation status at the *CHD5* promoter. We carried out a methylation sensitive-PCR assay (MS-PCR) at the CpG island which we observed to be unmethylated in 742 patients (Figure 2B and 2C). We found DNA methylation at the *CHD5* promoter to be absent in most of the cell lines, with the exception of MDA-MB-231 (Figure 2D);

a similar finding was previously reported by Mulero-Navarro and Esteller [12]. As a positive methylation control of the assay we used an *in vitro*-methylated DNA (IVD) (Figure 2D).

Results from the MS-PCR reinforce the observation of the methylation status in the TCGA patients, where DNA methylation at the *CHD5* gene promoter is not a common mechanism involved in repression of *CHD5*. Therefore, we focused on another epigenetic mechanism that is independent of DNA methylation, such as the histone demethylase KDM4A.

The localization of KDM4A at the *CHD5* first intron correlates with the decrease of H3K36me3 and H3K36me2 in neoplastic cell lines

In 2012, Mallette and colleagues demonstrated by chromatin immunoprecipitation (ChIP) assays that KDM4A is located at *CHD5* first intron in the U2OS cell line, and that the depletion of KDM4A increased *CHD5* mRNA and protein levels [7]. Nevertheless, the mechanism by which KDM4A negatively regulates transcription of the *CHD5* gene remained unclear.

One epigenetic mark relevant to transcriptional elongation is H3K36me3. This histone mark is mainly enriched in gene bodies, where a decrease in its

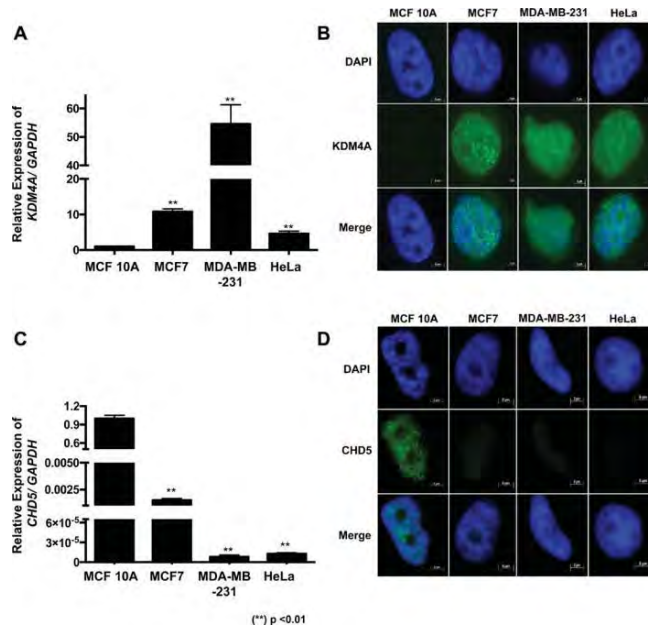


Figure 1: KDM4A overexpression correlates with CHD5 decrease in neoplastic cell lines. (A) Expression profile of the human *KDM4A* gene in MCF 10A, MCF7, MDA-MB-231 and HeLa cell lines obtained by RT-qPCR. The data were normalized against GAPDH expression in three independent experiments. (B) The presence and localization of KDM4A in MCF 10A, MCF7, MDA-MB-231 and HeLa cells were assessed by immunofluorescence assay. (C) Expression profile of *CHD5* gene in the MCF 10A, MCF7, MDA-MB-231 and HeLa cell lines obtained by RT-qPCR. The data were normalized against GAPDH expression in three independent experiments. (D) The presence and localization of *CHD5* in MCF 10A, MCF7, MDA-MB-231 and HeLa cells were assessed by immunofluorescence assay. The DNA was stained with DAPI. (**) $p < 0.01$ compared with the MCF 10A cell line. Statistical differences were determined using Student's *t* test.

trimethylated form is associated with gene silencing. In some genes, such silencing is not related to inactivation of the gene's promoter. Since KDM4A is capable of removing this histone mark, we speculated that demethylation of H3K36me3 could play a role in the downregulation of *CHD5* gene expression. To assess our hypothesis, we performed a ChIP assay to determine whether KDM4A could be found at the *CHD5* first intron in our cell lines. KDM4A was present at this region in the MCF7, MDA-MB-231 and HeLa cell lines; and was not detected in the non-neoplastic cell line MCF 10A (Figure 2E). One of the best-characterized gene targets of KDM4A

is the region located -1922 bp upstream from the TSS of *ASCL2* [14]. Therefore, we used this region as a positive control of ChIP assay to confirm the presence of KDM4A in all the cell lines, and the 27th exon of the *RB* gene as negative control (Supplementary Figure 2A). To determine the impact of the presence of KDM4A on histone marks related to transcriptional elongation, we analyzed the abundance of H3K36me3 and H3K36me2 at the *CHD5* first intron by ChIP assay (Figure 2F). As a positive control for the H3K36me3 histone modification, we used the ENCODE database to identify a region that is enriched with this histone mark in different cell lines; based on the

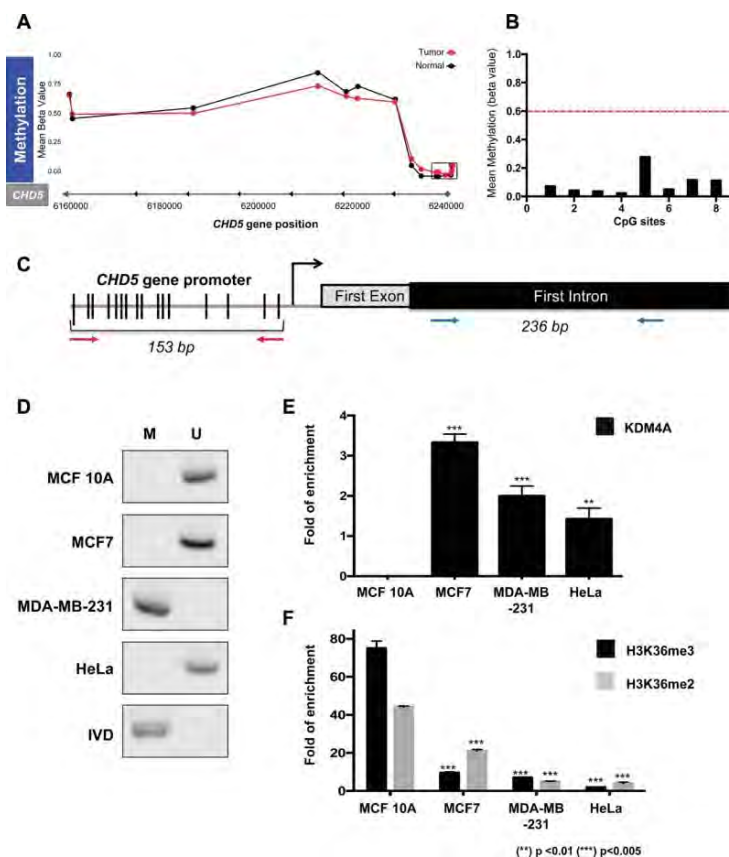


Figure 2: *CHD5* repression is associated to histone demethylation by KDM4A at the first intron and not to promoter DNA methylation. (A) TCGA DNA methylation levels (Illumina 450 K data) in 743 breast cancer patients (tumor) and 98 non-neoplastic samples (normal) in *CHD5* gene locus. 63 CpG sites were analyzed along the gene. Cutoff ≥ 0.6 Beta-values represents methylated status. The locus marked by a rectangle represents the methylation status of the promoter region; X axes represents the *CHD5* gene position (GRCh 37/hg19) (B) the graphic represents the 8 CpGs sites analyzed of the promoter region of 743 patients. The dot line is the threshold of DNA methylation ($>0.6 =$ Methylated). (C) Schematic representation of the *CHD5* gene that includes the promoter region and the *CHD5* first intron region analyzed by MS-PCR (153 bp PCR product) and ChIP assays (236 bp PCR product) respectively. The red arrows represent the primers employed for MS-PCR and blue arrows for ChIP qPCR (D) Promoter DNA methylation status was assessed by MS-PCR in MCF 10A, MCF7, MDA-MB-231 and HeLa cells. DNA from lymphocytes was methylated *in vitro* by SssI methyltransferase and used as a methylated DNA positive control (IVD). M represents methylated, and U represents non-methylated. (E and F) qPCR evaluation of the products obtained from the ChIP assay of the *CHD5* first intron, precipitated with anti-KDM4A (D), anti-H3K36me3 and anti-H3K36me2 (E) antibodies in MCF 10A, MCF7, MDA-MB-231 and HeLa cells. As a negative control, we used the IgG antibody included in the OneDay ChIP kit (Diagenode, NJ, USA, Kch-onedIP-180). (**) $p < 0.01$ and (***) $p < 0.005$ compared with the MCF 10A cell line. Statistical differences were determined using Student's *t* test.

results of this analysis, we decided to use the second intron of the *GAPDH* gene. As a negative control, we employed the third exon of the silenced gene *MYOG* (Supplementary Figure 2B). When we compared the enrichment of the methyl marks present in the intron 1 region to the non-neoplastic cell line MCF10A we found that the presence of KDM4A was associated with a decrease in these epigenetic marks in the tumor cell lines (Figure 2F). These results suggest that the presence of KDM4A could alter epigenetic marks related to transcriptional elongation and thus affect gene transcription (Figure 2E and 2F).

CTCF and KDM4A coexist at the *CHD5* first intron in neoplastic cell lines

Given that CTCF plays a major role in the demethylation function of KDM4A [8], we decided to characterize its expression in our cellular model. By RT-qPCR, we observed that CTCF was overexpressed in the MCF7, MDA-MB-231 and HeLa cell lines when compared to the MCF 10A cells (Figure 3A). In addition, CTCF was located in the nucleus of all the cell lines evaluated (Figure 3B).

To determine if CTCF could participate in KDM4A's demethylation activity, we decided to evaluate by ChIP assay the presence of CTCF in the first intron of *CHD5*. As a negative and positive controls, we used the 27th exon of *RB* gene and *WRAP53* promoter region, respectively (Supplementary Figure 2C). Our results show that CTCF is found in this region in all cell lines evaluated (Figure 3C and 3D), including MCF 10A. This is the same region where KDM4A was shown to be present in the neoplastic cell lines (Figure 2C). Since MDA-MB-231 exhibits promoter methylation, we decided to focus only in MCF7 and HeLa, where *CHD5* is repressed even though its promoter region is not methylated. Thus, to determine the coexistence of CTCF and KDM4A at the same genomic region, we performed a ChIP/re-ChIP experiment in the MCF7 and HeLa cell lines (Figure 4A). A first immunoprecipitation was performed with each of the antibodies (KDM4A or CTCF), and a subsequent immunoprecipitation was performed with a second antibody (KDM4A-CTCF or CTCF-KDM4A). As a negative control assay, we used the antibody of interest followed by IgGs or the IgGs followed by the antibody of interest. As a positive control for KDM4A recruitment, we analyzed the region -1922 bp from the *ASCL2* TSS (Supplementary Figure 3A). For CTCF, we employed the *WRAP53* promoter region as a positive control (Supplementary Figure 3B). As a negative control for KDM4A and CTCF, we evaluated the 27th exon of the *RB* gene (Supplementary Figure 3C). The ChIP/ReChIP results showed a co-existence of CTCF and KDM4A at the first intron of *CHD5* both in MCF7 and HeLa cells (Figure 4A). Using the ChIP and ChIP/ReChIP results we evaluated the percentage of co-occupancy in MCF7 and

HeLa cell lines (Figure 4B). These results suggest that the higher co-occupancy of KDM4A and CTCF is associated with an increase in *CHD5* repression. Also, these results imply a possible interaction between CTCF and KDM4A.

CTCF and KDM4A form a protein complex in neoplastic cell lines

In order to demonstrate the physical interaction between CTCF and KDM4A a co-immunoprecipitation assay in HeLa cells was performed. This was carried out by an immunoprecipitation against CTCF and revealed with a CTCF antibody (Figure 4C). Subsequently, the proteins obtained from the CTCF Immunoprecipitation (IP) were used in an independent experiment and were revealed against KDM4A (Figure 4D). Our data shows a detectable interaction between endogenous CTCF and endogenous KDM4A in HeLa cells (Figure 4C and 4D). Our results demonstrate a novel protein complex formed by CTCF and KDM4A, which may be localized at the first intron of the *CHD5* gene (Figure 4).

***KDM4A* and *CTCF* siRNA knock down is associated with the reactivation of *CHD5* expression in neoplastic cell lines**

To determine the participation of KDM4A in the repression of *CHD5*, HeLa and MCF7 cells were transfected with siRNAs against *KDM4A*. At 72 hours, post-transfection with the siRNA, expression analyses of the *KDM4A* and *CHD5* genes were performed by RT-qPCR. The results revealed that *KDM4A* mRNA decreased after transfection (Figure 5A). The decrease of KDM4A in the MCF7 and HeLa cell lines induced the reactivation of *CHD5* mRNA, even above the basal expression of MCF 10A (Figure 5B). Because CTCF and KDM4A can potentially form a protein complex, we further investigated which was the participation of CTCF in the repression of *CHD5*. Therefore, we transiently transfected a small hairpin RNA expression vector against CTCF (pCT1) in MCF7 and HeLa cells (Figure 5C). Our results show that diminishing of CTCF leads to a reactivation of *CHD5* expression similar to MCF 10A (Figure 5D). Taken together, our results suggest that the presence of KDM4A and CTCF at the first intron of *CHD5* acts as repressors of *CHD5* expression in neoplastic cells.

***KDM4A* knockout (KO) in MCF7 reestablish the H3K36me3 histone mark at the first intron of *CHD5* and reactivates gene expression**

In order to further validate that KDM4A is negatively regulating *CHD5* we establish a Knockout model (*KDM4A*_{KO}) using CRISPR/Cas9 KO system (Santa Cruz, sc-404599 and sc-404599-HDR). This system employed three gRNAs that target exon 3 and

8 of the *KDM4A* gene (Supplementary Figure 4A). We selected cells by puromycin treatment and further enrich our *KDM4A*_{KO} by FACS cell sorting selecting the highest fluorescent cells (Supplementary Figure 4B). As control, we employed a non-targeting gRNA plasmid (Mock) (Santa Cruz, sc-418922).

We evaluated by Western Blot the protein expression of *KDM4A* in MCF7 Mock and *KDM4A*_{KO} cells, where a 63.6% reduction of *KDM4A* in *KDM4A*_{KO} cells is observed (Figure 6A). We also performed RT-

qPCR analysis of *CHD5* expression in MCF10A, MCF7 Mock and *KDM4A*_{KO} cells. Here we observe a significant reactivation of the *CHD5* expression in *KDM4A*_{KO} cells, with levels similar to the observed in MCF10A (Figure 6B). In order to evaluate if such reactivation is related to the loss of *KDM4A* of the *CHD5* first intron, we performed a ChIP analysis of *KDM4A*. Our results show a significant loss of *KDM4A* in *KDM4A*_{KO} compared to Mock cells (Figure 6C). Regarding our previous results that suggest that *KDM4A*-CTCF complex regulates

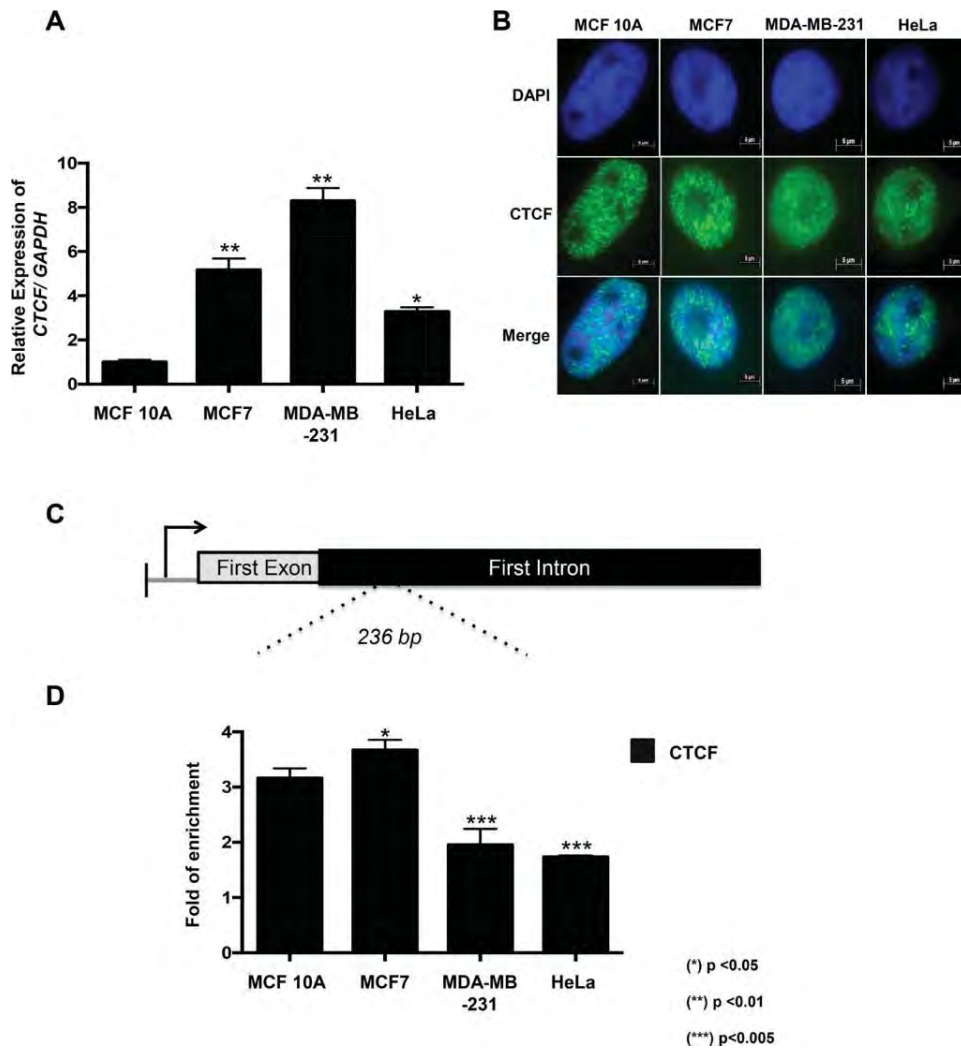


Figure 3: CTCF is overexpressed in neoplastic cell lines and is recruited to *CHD5* first intron. (A) Expression profile of the human *CTCF* gene in the MCF 10A, MCF7, MDA -MB-231 and HeLa cell lines was obtained with RT-qPCR. The data were normalized against GAPDH expression in three independent experiments. (*) $p < 0.05$ and (**) $p < 0.01$ compared to the MCF 10A cell line. Statistical differences were determined using Student's *t* test. (B) The presence and localization of CTCF in MCF 10A, MCF7, MDA-MB-231 and HeLa cells were assessed by immunofluorescence assay. (C) Schematic representation of the *CHD5* gene that includes the promoter region and the *CHD5* first intron region (236 bp PCR product). (D) qPCR analysis of the *CHD5* first intron was performed on the DNA obtained from anti-CTCF ChIP assays in MCF 10A, MCF7, MDA-MB-231 and HeLa cells. As a negative control, we used the IgG antibody included in the OneDay ChIP kit (Diagenode, NJ, USA, Kch-onedIP-180). The data was evaluated by qPCR at *CHD5* first intron and the data is expressed in fold of enrichment over IgG immunoprecipitation. Statistical differences were determined using Student's *t* test. (*) $p < 0.05$ and (**) $p < 0.01$ compared to the MCF 10A cell line.

CHD5, we evaluated if CTCF could be affected by the loss of KDM4A at the *CHD5* first intron. Our results show that CTCF binding is independent of KDM4A presence, suggesting that CTCF may act as repressor when it is in a complex with KDM4A (Figure 6D). Because of the obtained results, we attempted to establish a CTCF^{ko} model, however these cells were not viable so the experimental approach was not possible. The generation

of a CTCF^{ko} model has been an experimental challenge for different research groups. Particularly, alteration in the abundance of CTCF affects cell proliferation and can even be causal of a lethal phenotype in murine models [15–17].

One of the central questions we wanted to address is whether the loss of KDM4A could restore the H3K36me3 pattern at the *CHD5* first intron. Therefore, we performed a ChIP analysis of H3K36me2/3 in MCF10A, MCF7

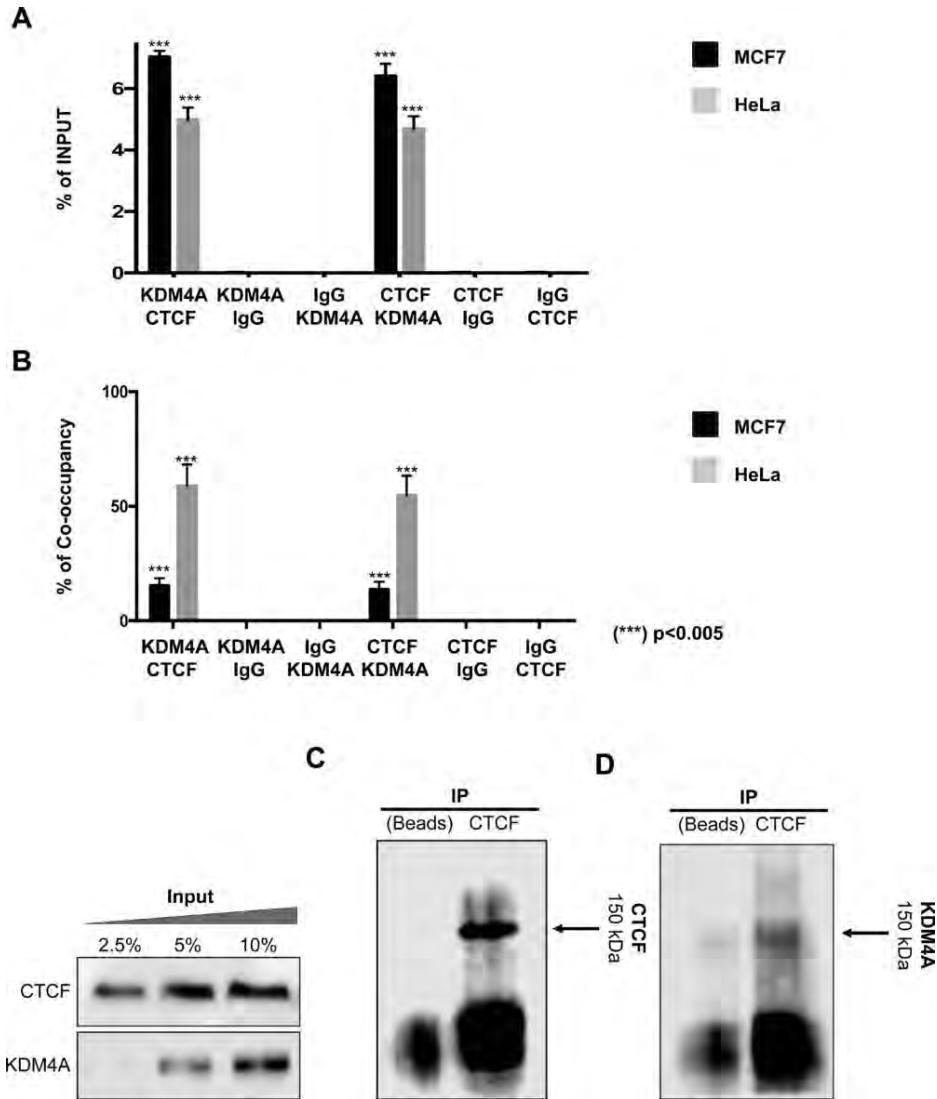


Figure 4: CTCF-KDM4A complex is located at the *CHD5* first intron in MCF7 and HeLa cell lines. (A) ChIP/re-ChIP assays were performed using the antibodies shown in the first row and subsequently immunoprecipitated by the antibodies described at the second row in the MCF7 and HeLa cells. The data was evaluated by qPCR at *CHD5* first intron and the results are represented as % of input. Statistical differences were determined using Student's *t* test, (***) $p < 0.005$ compared with IgGs. (B) Co-occupancy analysis was performed in accordance to Geisberg and Struhl [38]. For occupancy analysis, the ChIP-Re-ChIP data from both experiment data (CTCF-KDM4A or KDM4A-CTCF) were used. Also, the co-occupancy of IgG experiments was evaluated. The results are represented in % of co-occupancy. Statistical differences were determined using Student's *t* test, (***) $p < 0.005$ compared with IgGs. (C–D) A Co-immunoprecipitation assay was performed against CTCF and revealed with CTCF (150 kDa) (C). Using the proteins obtained from the CTCF IP we revealed employing a KDM4A antibody (150 kDa) (D). To the left, the input material was evaluated against CTCF and KDM4A in increasing amounts of protein (2.5, 5 and 10 %).

Mock and KDM4A^{ko} cells. We found a H3K36me3 recovery in KDM4A^{ko} cells, which does not affect the H3K36me2 (Figure 6E). This suggests that the loss of KDM4A demethylase allows the reincorporation of H3K36me3 at the first intron of *CHD5*, favoring the reactivation of the gene expression.

DISCUSSION

Epigenetic alterations are a common feature of cancer processes [18, 19]. Mainly, key epigenetic components, which include methylases and demethylases such as KDM4A as well as architectural proteins like CTCF, are deregulated [5]. Several studies have reported that KDM4A is highly expressed in breast cancer

tissues. This demethylase removes the methyl group of H3K9me3 and H3K36me3, with the former related to heterochromatin and the repression of transcription [20], while the latter is enriched in the bodies of genes that are transcriptionally active and is associated with the recruitment of RNA polymerase II and the process of transcriptional elongation [4]. Hence, H3K36me3 alteration could affect gene transcription without disturbing the gene promoters, suggesting a novel mechanism of gene dysregulation not associated with regulatory regions.

CHD5 is a gene that encodes an enzyme which belongs to the helicase family (chromodomain helicase DNA-binding protein 5) [21]. The CHD5 protein can function as a tumor suppressor by regulating apoptosis

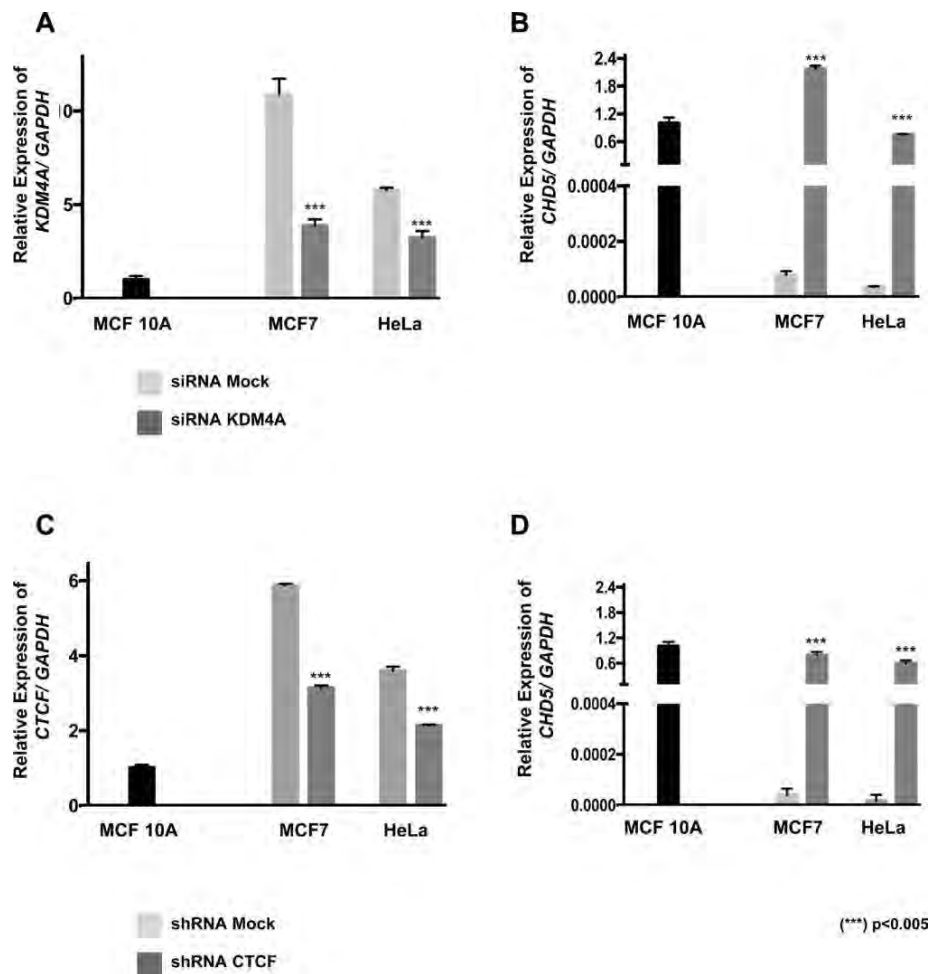


Figure 5: The *CHD5* expression is reactivated by CTCF and KDM4A knockdown in MCF7 and HeLa cells. Analysis of *KDM4A* (A) and *CHD5* (B) expression in MCF7 and HeLa cell lines following *KDM4A* siRNA transfection. Analysis of *CTCF* (C) and *CHD5* (D) expression in MCF7 and HeLa cell lines following *CTCF* shRNA transfection. Data were normalized against *GAPDH* expression in three independent experiments using MCF 10A cells as the normal expression control. siRNA mock and shRNA mock transfected cells were used as negative controls. Statistical differences were determined using Student's *t* test compared with mock-transfected cells. (***) $p < 0.005$.

and cellular senescence, and is involved in the p19^{Arf}/p53 pathway by interacting with MDM2 [22, 23]. Because this interaction leads to the attenuation of MDM2-mediated p53 degradation [24], CHD5 and p19^{ARF} help to stabilize p53. In addition, CHD5 inhibits clonogenic growth *in vitro* as well as tumor xenograft growth, suggesting that its inactivation may be involved in cancer development [11]. Some studies have suggested that CHD5 can be inactivated by genetic [25] or epigenetic processes, but these reports focused mainly on its repression by DNA

promoter methylation [11, 12, 26–28]. Analysis of TCGA datasets show that the CHD5 promoter region is not methylated in breast cancer patients, which suggests that another epigenetic mechanism could be involved in gene repression. In this regard there is evidence that suggests that alteration at the CHD5 promoter is not the major mechanism of repression of this gene [12].

Previously, it was reported that KDM4A localizes to the CHD5 first intron and the reduction in KDM4A leads to an increase of CHD5 expression in U2O2 cells; this

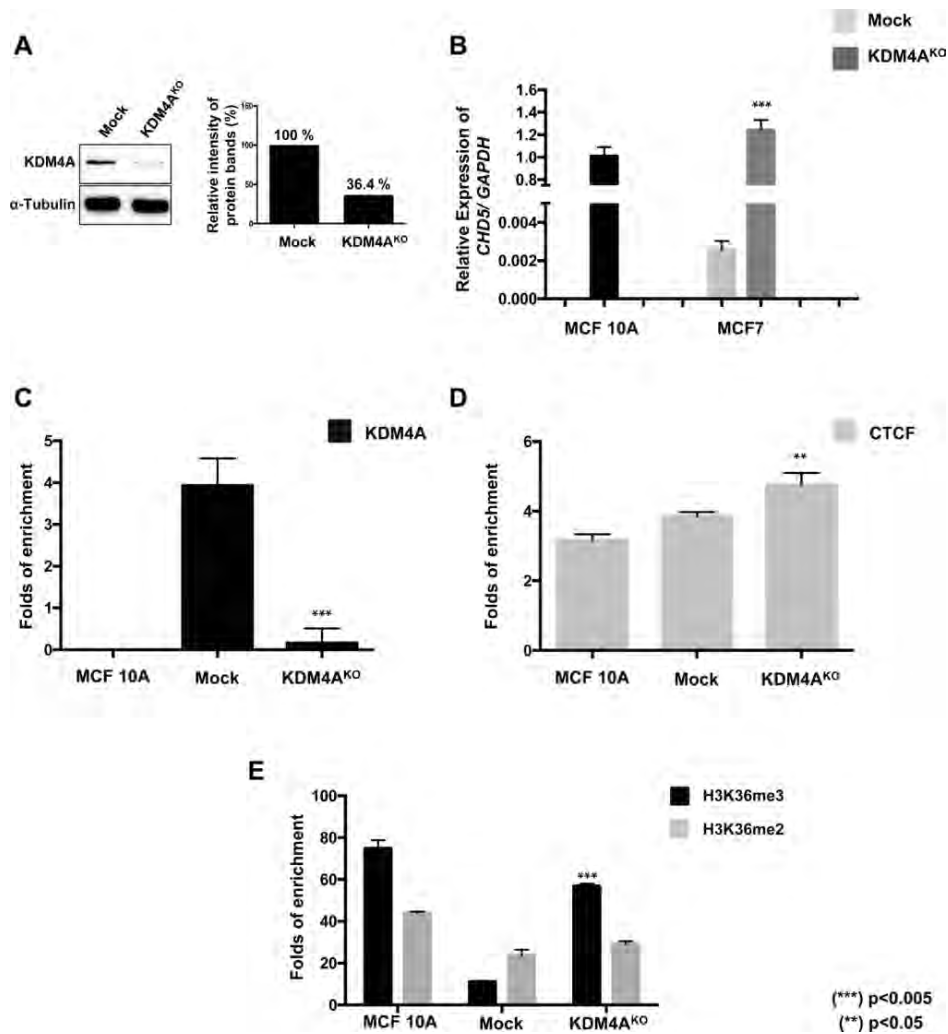


Figure 6: The KDM4A knockout promotes the reestablishment of the H3K36me3 histone mark at the first intron and the reactivation of the expression of CHD5 gene. (A) Characterization of KDM4A protein abundance by immunoblots in MCF7 cells transfected with a non-targeting gRNA plasmid (Mock) or KDM4A^{KO} CRISPR/Cas9 and HDR plasmids. The quantitation of the relative intensity of the protein bands showed a decrease of 63.6% of KDM4A in KO cells. (B) CHD5 expression analysis in Mock and KDM4A^{KO} cells. Data were normalized against GAPDH expression in two independent experiments using MCF 10A cells as the normal expression control. (C, D, E) qPCR evaluation of the CHD5 first intron from DNA obtained from the ChIP assay using anti-KDM4A (C), anti-CTCF (D) and anti-H3K36me3 and anti-H3K36me2 (E) antibodies in MCF 10A cells, Mock and KDM4A^{KO} cells. As a negative control, we used the IgG antibody included in the OneDay ChIP kit (Diagenode, NJ, USA, Kch-onedIP-180). (**) $p < 0.01$ and (***) $p < 0.005$ compared with the MCF7 Mock cells.

indicates that KDM4A could be associated with *CHD5* repression [7]. However, since KDM4A was not found at the promoter of *CHD5*, the mechanism of how KDM4A downregulates *CHD5* remained unclear.

The overexpression of *KDM4A* is known to be associated with cell proliferation and poor prognosis in several cancers [29, 30]. Identifying how KDM4A inhibits gene expression has a therapeutic impact on cancer in the future; therefore, understanding the molecular mechanisms underlying the effects of KDM4A and their implications in cancer are an important topic for future clinical research [31]. Our findings show that KDM4A functions as a repressor of the *CHD5* TSG by affecting epigenetic marks associated with elongation and not by regulating the gene promoter. This phenomenon has been reported in other

cellular models, where KDM4A/C specifically alters H3K36me3 [32]. The phenomenon is also associated with the loss of RNA polymerase II recruitment in transcribed regions of the *GFAP* gene [32]. Our results suggest a novel mechanism of *CHD5* gene repression, where the decrease of H3K36me3/2 at the gene body could lead to transcriptional repression. One hypothesis is that this phenomenon occurs due to lack of phosphorylation of the second serine in the carboxy terminal domain of RNA polymerase II, which results in the enrichment of H3K36me2 and a decrease of transcriptional elongation, or due to an increase in repressive histone marks.

In vitro assays have reported that the presence of CTCF increases the demethylation frequency of KDM4A by up to 80%, suggesting that CTCF has a role in the

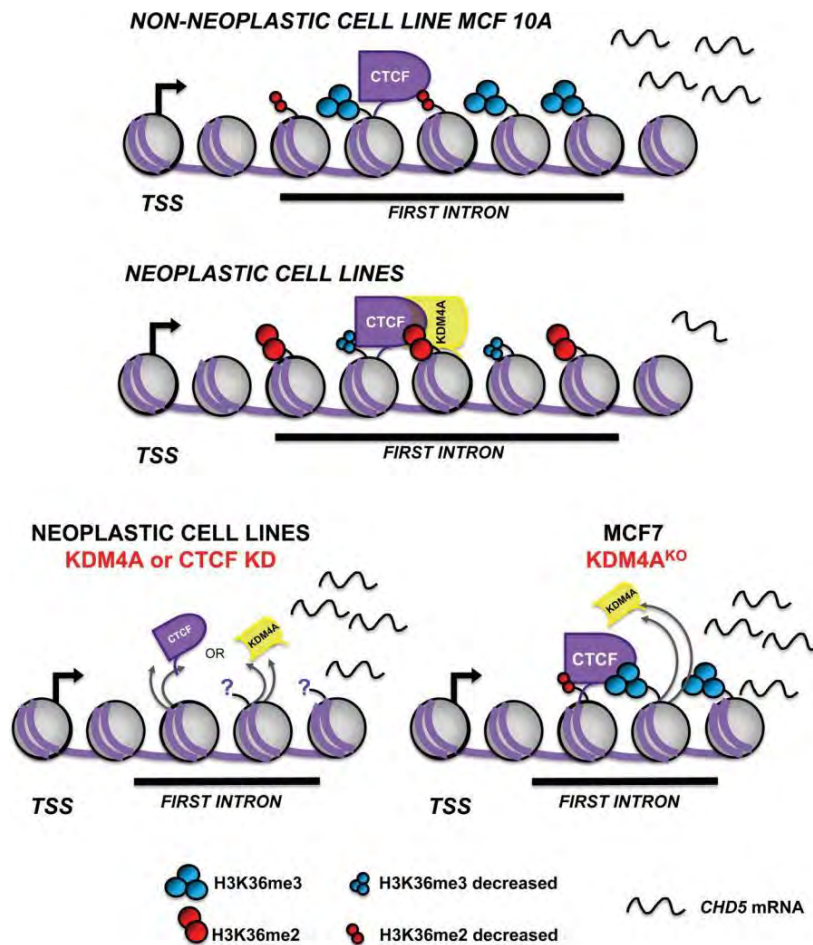


Figure 7: Schematic model of *CHD5* transcriptional repression mediated by CTCF-KDM4A protein complex. CTCF-KDM4A protein complex is recruited to the first intron of the *CHD5* gene and promotes demethylation of histone H3K36me. In non-neoplastic cells, CTCF is located at the first intron of *CHD5*, and H3K36me3/2 are enriched. These events correlate with *CHD5* expression. In contrast, in the neoplastic cells, CTCF-KDM4A protein complex promotes the demethylation of H3K36me3/2 and leads to gene repression. CTCF or KDM4A knockdown (KD) reactivates *CHD5* gene expression. The loss of KDM4A in KDM4A^{KO} cells leads to the reestablishment of the H3K36me3 histone mark at the first intron and the reactivation of *CHD5* gene expression.

demethylation function of KDM4A [8]. Additional support for these datasets was provided by another study that demonstrated that CTCF can interact with the KDM5B histone demethylase and increase its demethylation activity in breast cancer cell lines [33]. CTCF has been reported to act occasionally as a transcriptional repressor of genes, such as *c-MYC*, *Bax*, *Xist* and *hTERT*, by interacting with SIN3A and recruiting HDACs or by preventing the binding of transcription factors that affect expression [34–38]. Interestingly, we observed a protein complex formed by CTCF-KDM4A, which is found at the first intron of *CHD5*. When we evaluated the co-occupancy of KDM4A and CTCF, we showed that the HeLa cell line exhibits a higher percentage of co-occupancy in comparison with MCF7 cell line. Our results suggest that KDMA4 acts as a transcriptional repressor when it is in complex with CTCF. The loss of KDM4A at *CHD5* first intron restores H3K36me3 histone mark and recovers *CHD5* gene expression. Therefore, we propose a novel mechanism of transcriptional repression mediated by KDM4A and CTCF (Figure 7). To date it is unknown if this complex is related with the repression of other genes, and what could be the implications of this complex in diseases such as cancer. Further studies are needed to understand the biological meaning of this new regulatory mechanism.

MATERIALS AND METHODS

Cell culture

MCF 10A cells were cultured in 1 part DMEM-Dulbecco's Modified Eagle Medium (GIBCO, 11965-084) to 1 part of Dulbecco's Modified Eagle Medium/Ham's F-12 Nutrient Mixture (DMEM/F-12, GIBCO, 11320-033) supplemented with 10% fetal bovine serum (GIBCO, 10500056), 2 mM L-Glutamine (GIBCO, 25030081), 10 ng/ml EGFRh (Invitrogen), 120 mU/ml insulin and 1 µg/ml hydrocortisone (SIGMA). MCF7 and MDA-MB-231 cells were cultured in Dulbecco's Modified Eagle Medium/Ham's F-12 Nutrient Mixture (DMEM/F-12, GIBCO, 11320-033) supplemented with 10% fetal bovine serum. HeLa cells were cultured in DMEM high glucose (GIBCO, 11965-084) supplemented with 10% fetal bovine serum. All cell lines were cultured at 37° C in a 5% CO₂ incubator.

Expression analysis

Total RNA was extracted using TRIzol (Invitrogen, 15596018) according to the manufacturer's specifications. 2 µg of total RNA were reverse transcribed in a final volume of 40 µL using the Kit GeneAmp® RNA PCR KIT (Applied Biosystems, N8080143) as described by the manufacturer's protocol. Gene expression levels for *KDM4A*, *CTCF*, and *CHD5* were determined using the

primers listed in Supplementary Table 1; *GAPDH* was used as an internal control. The qPCRs were performed using Thermo Maxima SYBR Green/ROX 1 PCR Master Mix (Thermo Scientific, K0222) with a StepOnePlus Real-Time PCR System (Applied Biosystems). All reactions were run in triplicate, and the average C_t values were used for quantification. The plots show the mean of three biological replicates. The analysis of the relative quantification of target genes was performed using the $\Delta\Delta C_t$ method as described by Livak [39].

Immunofluorescence assays

Cells were cultured in 22 × 22 mm coverslips at least 18 h before the immunofluorescence staining was performed. The cells were fixed in 2% formaldehyde for 10 min and then washed three times with 1x PBS for 5 min each. Subsequently, cells were permeated with 2% Triton X-100 in 1× PBS for 20 min and then washed three times with 1× PBS for 5 min. Non-specific antigens were blocked by incubating the cells with 1% fetal bovine serum in 1× PBS for 40 min at room temperature. Then, the cells were incubated with the primary antibodies diluted in blocking solution for 60 min at 37° C. The coverslips were washed three times with 2% Triton X-100 in 1× PBS for 3 min; in between these washes, the cells were quickly rinsed with 1× PBS. Afterwards, the coverslips were incubated with the secondary antibodies diluted in blocking solution for 60 min at room temperature in the dark. The cells were washed three times with 2% Triton X-100 in 1× PBS for 3 min; in between these washes, the cells were quickly rinsed with 1× PBS. Finally, the coverslips were mounted on a previously cleaned slide with 10 µL-15 µL mounting medium with DAPI (Vector Labs, H-1200). To prevent drying and movement under the microscope, the coverslips were sealed with nail polish and then stored in the dark at 4° C. For all experiments, at least 100 cells from three coverslips were analyzed. The antibodies used are listed in Supplementary Table 2. The cells were observed using a Zeiss Axio Imager A2 epifluorescence microscope (Carl Zeiss), and the images were analyzed using AxioVision 4.8 software (Carl Zeiss). The concentrations and quantities of antibodies were chosen based on the manufacturer's specifications.

CHD5 Promoter methylation analysis by MS-PCR

DNA was obtained from the cell lines by phenol/chloroform extraction. 500 ng of genomic DNA were modified using the EZ DNA methylation Gold kit (ZYMO, D5006). The MS-PCR assay was performed with DNA treated with sodium bisulfite. The primers for MS-PCR were designed using Methyl Primer Express software and are listed in Supplementary Table 3. As a positive control, 1 µg of DNA from lymphocytes of a

healthy donor was methylated *in vitro* (IVD) for 8 h using SssI methyltransferase (NEB, M226S).

Chromatin immunoprecipitation (ChIP) and ChIP/re-ChIP assays

Cells were cultured until 80% confluence, and then, chromatin was extracted in accordance with the protocol of the OneDay ChIP kit (Diagenode, Kch-onedIP-180). ChIP assays were performed following the manufacturer's instructions. For all experiments, at least two chromatin preparations were analyzed. As a negative control, we used an IgG antibody included in the kit. The antibodies used are listed in Supplementary Table 4.

The ChIP/re-ChIP assays were performed following the method previously described by [40]. In brief, cells were treated according to the first steps in the ChIP assay and then incubated at 37° C in 10 mM DTT in 1X ChIP buffer for 30 min. Eluents were then diluted at 1.5 mL with ChIP buffer and incubated with the indicated second antibody overnight. The following day, protein A agarose beads were added to the solution, which was then incubated for 3 h at 4° C. The DNA-protein-antibody complexes were washed three times with 1X ChIP buffer. Finally, the DNA-protein complexes were treated with proteinase K overnight, and to break the crosslinked complexes, the samples were boiled for 10 min. The DNA was extracted as suggested by the OneDay ChIP kit protocol, and qPCR was performed with the specific primers listed in Supplementary Table 5.

The obtained results represent experiments from four separate amplifications that were used to calculate the standard deviation. qPCRs were done in triplicate using fast optical 96-well qPCR plates. Then, the oligonucleotides were amplified in triplicate by a fast optical 96-well qPCR plate (Applied Biosystems). The qPCR was performed using Thermo Maxima SYBR Green/ROX 1 PCR Master Mix (Thermo Scientific, K0222) with a StepOnePlus Real-Time PCR System (Applied Biosystems). We used the concentration of antibodies indicated by the manufacturer's specifications.

ChIP and ChIP-Re-ChIP data analysis

The oligonucleotides were validated with a standard curve performed with Input serial dilutions. The amplification efficiency (AE) value was calculated as $AE = 10^{(-1/slope)}$. The percentage of the input was calculated as $\% \text{ input} = AE^{(Ct_{input} - Ct_{ChIP})} \times Fd$ ($Fd = \text{factor Dilution}$) $\times 100$, using 10% of the input value as reference. Afterwards, to calculate the fold of enrichment of the immunoprecipitated proteins we used the following equation $\text{fold of enrichment} = \frac{\% \text{ input (ip)}}{\% \text{ input (IgG)}}$ as described in the OneDay ChIP (Diagenode) manufacturer's manual. For the ChIP/ Re-ChIP analysis, we calculated the % of the input using the 10% of the input as reference and compared the data obtained from IgG.

Co-occupancy data analysis

To determine the co-occupancy of CTCF and KDM4A at the first intron of *CHD5* gene, we used the fold of enrichment over the background for each individual ChIP. The percentage of co-occupancy, was calculated according to Geisberg and Struhl [41]: $\% \text{ co-occupancy} = 100 \frac{(AB-A)}{(A \times B-A)}$, where A and B represent the IP of each experiment, and AB the ChIP-Re-ChIP assay. The occupancy was determined in the ChIP-Re-ChIP data for both experiments (CTCF-KDM4A or KDM4A-CTCF), as well as the co-occupancy in the IgG experiments, with negative results plotted with a value of 0.

CTCF and KDM4A knockdown

HeLa and MCF7 cells were transiently transfected using Xfect transfection reagent (Clontech, 631317) following the manufacturer's specifications, using 2.5 µg of a small hairpin RNA expression vector against CTCF (pCT1) kindly provided by Ko Ishihara (Institute of Molecular Embryology and Genetics, Kumamoto University, Japan) [42]. As a mock control, we employed the empty vector from pSilencer-3.1-H1 puro (Ambion).

For KDM4A knockdown, siRNA transfections were performed using KDM4A SMART pool siRNAs (Dharmacon, E-004292-00-0010) and non-targeting siRNA (Dharmacon, D-001910-01-05). HeLa and MCF7 cells were seeded at 3×10^4 cells/well and 6×10^4 cells/well, respectively, in 12-well plates. 24 h later, the cells were transfected with ACCELL siRNA Delivery Media (Dharmacon, B-005000-500) over 72 h according to the manufacturer's protocol. The results were obtained from three separate biological replicates. RNA and cDNA were obtained as previously described.

Co-immunoprecipitation of CTCF and KDM4A (Co-IP)

Extracts from HeLa cells were prepared with IP lysis buffer containing 50 mM Tris-HCl (pH 8.0), NaCl 150 mM and 1% of NP40 supplemented with 2× complete protease inhibitor cocktail (Sigma-Aldrich). The cell lysate was cleared by centrifugation at 13,000 rpm for 10 minutes at 4° C. The proteins were incubated with 2 µg of anti-CTCF (Santa Cruz Biotech, sc-5916) or without antibody (using beads) and the complex were precipitated employing 25 µL of Protein A/G magnetic beads (Pierce, 88802) and incubated at 4° C approximately 16 h. The beads were recovered with a magnetic stand, and washed five times for 20 minutes with IP lysis buffer. Finally, proteins were eluted by boiling in 1× Laemmli buffer and evaluated by Western Blot using antibodies against CTCF (Santa Cruz Biotech, sc-5916) and KDM4A (Cell Signaling, JMJD2A #5328) as two independent experiments. At least three independent biological replicates were evaluated.

CRISPR/Cas9 KO and HDR plasmids transfection

We used the Xfect Transfection Reagent (PT5003-2) to transfect 1 µg of the CRISPR/Cas9 KDM4A KO Plasmid (sc-404599) and 1 µg of the HDR Plasmid (sc-404599-HDR). In brief, 3×10^5 cells were seeded in a 6-well chamber, 24 h before plasmid transfection. We diluted 1 µg of the KDM4A KO and HDR plasmids onto 100 µL of Xfect Reaction Buffer. Afterwards, we added 2 µL of the Xfect Polymer and incubated for 15 min. Finally, we distributed the entire 100 µL of nanoparticle complex solution dropwise to the cell culture medium. 48 h after transfection we evaluate the GFP and RFP expression by epifluorescence microscopy (Carl Zeiss, AXIO Imager D2). We selected the transfected cells with media supplemented with Puromycin (3 µg/mL), changing the media every 24 h for at least 5 days.

Flow cytometry and cell sorting

KDM4A_{KO} or Mock cells were resuspended at a concentration of 1×10^6 cells/mL in DMEM/F-12, containing 10% FBS and 1X antibiotic-antimycotic. First, cells were filtered through a 70 µm cell strainer and subsequently through a 40 µm cell strainer and sorted on a FACSAria III Cell Sorting Flow Cytometer (BD Biosciences, San Jose, CA). Prior to sorting, MCF7 WT was used for cell size and autofluorescence measures. It was determined that the Mock cells did not show RFP fluorescence, while the KO population that was positive for RFP, only cells with the highest fluorescence were sorted (Supplementary Figure 4B). 2.61×10^5 cells were sorted into DMEM/F12 medium containing 10% FBS, and 2X antibiotic-antimycotic and were seeded in a p60 cell culture plate. The KDM4A_{KO} population, that exhibit RFP⁺ high expression, and Mock sorted cells were used for the subsequent experiments.

Abbreviations

KDM4A: Lysine specific demethylase 4; CTCF: CCCTC-Binding Factor; CHD5: chromodomain helicase DNA binding protein 5; qPCR: quantitative PCR; ChIP: Chromatin immunoprecipitation; Co-IP: co-immunoprecipitation; MS-PCR; Methylation-specific PCR; H3K36me3: trimethylation of histone H3 at lysine 36; H3K36me2: demethylation of histone H3 at lysine 36; TSG: tumor suppressor gene.

Author contributions

LGC, RGB and ESR designed experimental strategy, analyzed the results, and drafted the manuscript; CCP, performed the Co-IP and Western Blot assays, MSA carried out the siRNA assays and drafted the manuscript;

NA performed the bioinformatics analysis; FVR and CCH participated in the DNA methylation analysis; DCL, HAMM, LAH and AGC analyzed and discussed the results; YSP and IARV, participated in the manuscript discussion. All authors have contributed to seen and approved the manuscript.

ACKNOWLEDGMENTS

L. Guerra-Calderas is a doctoral student in the Programa de Doctorado en Ciencias Bioquímicas, UNAM, and received a fellowship from CONACyT (588391). We thank Dr. Guillermo Juárez Vega for his technical assistance with cell sorting.

CONFLICTS OF INTEREST

The authors declare no conflicts of interests.

FUNDING

This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT) by the Fondo Sectorial de Investigación en Salud y Seguridad Social (FOSISS, grant number 0261181), CONACyT 182997 and CB-CONACyT 284748. This research project had support from the National Cancer Institute of Mexico (INCan).

REFERENCES

1. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007; 128:693–705. <https://doi.org/10.1016/j.cell.2007.02.005>.
2. Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*. 2005; 6:838–49. <https://doi.org/10.1038/nrm1761>.
3. Guerra-Calderas L, González-Barrios R, Herrera LA, Cantú de León D, Soto-Reyes E. The role of the histone demethylase KDM4A in cancer. *Cancer Genet*. 2015; 208:215–24. <https://doi.org/10.1016/j.cancergen.2014.11.001>.
4. Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. Psp1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet*. 2012; 8:e1002717. <https://doi.org/10.1371/journal.pgen.1002717>.
5. Berry WL, Shin S, Lightfoot SA, Janknecht R. Oncogenic features of the JMJD2A histone demethylase in breast cancer. *Int J Oncol*. 2012; 41:1701–6. <https://doi.org/10.3892/ijo.2012.1618>.
6. Kolla V, Zhuang T, Higashi M, Naraparaju K, Brodeur GM. Role of CHD5 in human cancers: 10 years later. *Cancer Res*. 2014; 74:652–8. <https://doi.org/10.1158/0008-5472.CAN-13-3056>.
7. Mallette FA, Richard S. JMJD2A promotes cellular transformation by blocking cellular senescence through

- **PUBLICACIÓN 3:** Angiotensinogen rs5050 germline genetic variant as potential biomarker of poor prognosis in astrocytoma.
- **AUTORES:** Perdomo-Pantoja A, Mejía-Pérez SI, Reynoso-Noverón N, Gómez-Flores-Ramos L, Soto-Reyes E, Sánchez-Correa TE, Guerra-Calderas L, Castro-Hernandez C, Vidal-Millán S, Sánchez-Corona J, Taja-Chayeb L, Gutiérrez O, Cacho-Díaz B, Alvarez-Gomez RM, Gómez-Amador JL, Ostrosky-Wegman P, Corona T, Herrera-Montalvo LA & Wegman-Ostrosky.
- **REVISTA:** PLoS ONE. **VOLÚMEN:** 13. **PÁGINAS:** e0206590. **AÑO:** 2018.

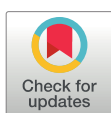
RESEARCH ARTICLE

Angiotensinogen rs5050 germline genetic variant as potential biomarker of poor prognosis in astrocytoma

Alexander Perdomo-Pantoja^{1,2}, Sonia Iliana Mejía-Pérez², Nancy Reynoso-Noverón³, Lilliana Gómez-Flores-Ramos³, Ernesto Soto-Reyes³, Thalía Estefanía Sánchez-Correa², Lissania Guerra-Calderas³, Clementina Castro-Hernandez⁴, Silvia Vidal-Millán³, José Sánchez-Corona⁵, Lucía Taja-Chayeb³, Olga Gutiérrez³, Bernardo Cacho-Díaz⁶, Rosa María Alvarez-Gomez³, Juan Luis Gómez-Amador², Patricia Ostrosky-Wegman⁷, Teresa Corona⁸, Luis Alonso Herrera-Montalvo^{3,4}, Talia Wegman-Ostrosky^{3*}

1 Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, United States of America, **2** Departamento de Neurocirugía, Instituto Nacional de Neurología y Neurocirugía, "Manuel Velasco Suarez", Mexico City, Mexico, **3** Dirección de Investigación, Instituto Nacional de Cancerología, Mexico City, Mexico, **4** Unidad de Investigación Biomédica en Cáncer, Instituto de Investigaciones Biomédicas, UNAM-INCAN, Mexico City, Mexico, **5** Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, Mexico, **6** Unidad de Neuro-oncología, Instituto Nacional de Cancerología, Mexico City, Mexico, **7** Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico, **8** Laboratorio Clínico de Enfermedades Neurodegenerativas, Instituto Nacional de Neurología y Neurocirugía, "Manuel Velasco Suarez", Mexico City, Mexico

* taliaw@gmail.com



 OPEN ACCESS

Citation: Perdomo-Pantoja A, Mejía-Pérez SI, Reynoso-Noverón N, Gómez-Flores-Ramos L, Soto-Reyes E, Sánchez-Correa TE, et al. (2018) Angiotensinogen rs5050 germline genetic variant as potential biomarker of poor prognosis in astrocytoma. PLoS ONE 13(11): e0206590. <https://doi.org/10.1371/journal.pone.0206590>

Editor: Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

Received: July 23, 2018

Accepted: October 16, 2018

Published: November 1, 2018

Copyright: © 2018 Perdomo-Pantoja et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT) (México) (Salud-2013-01-202720).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Introduction

Renin-angiotensin system (RAS) in brain cancer represents a scarcely explored field in neuro-oncology. Recently, some pre- and clinical studies have reported that RAS components play a relevant role in the development and behavior of gliomas. The angiotensinogen (AGT) rs5050 genetic variant has been identified as a crucial regulator of the transcription of AGT mRNA, which makes it a logical and promising target of research. The aim of this study was to determine the relationship between the AGT rs5050 genetic variant in blood with prognosis in astrocytoma.

Methods

A prospective pilot study was performed on forty-eight astrocytoma patients, who received the standard-of-care treatment. Blood samples were taken prior to surgery and DNA was sequenced using Ion Torrent next-generation sequencing and analyzed by Ion Reporter software. Descriptive, bivariate, multivariate, and survival analyses were performed using SPSS v21, STATA 12 and GraphPad Prism 7.

Results

Median follow-up was 41 months (range 1–48). Survival analysis showed a significant difference between the rs5050 genotypes ($p = .05$). We found lower survival rates in individuals

with the GG-genotype of rs5050 *AGT* compared to patients with the TT- and TG-genotype (2 months vs. 11.5 months, respectively [$p = .01$]). In bivariate and multivariate analyses, GG-genotype was negatively associated with survival.

Conclusions

In patients with astrocytoma, *AGT* rs5050 GG-genotype was associated with poor prognosis. We propose this germline genetic variant as a complementary biomarker, which can be detected practically and safely in blood samples or saliva.

Introduction

Gliomas are the most common intrinsic primary tumors of the central nervous system (CNS) in the adult population worldwide, representing approximately 27% of all CNS tumors and 80% of CNS malignant tumors in the United States [1, 2]. Astrocytomas, oligodendrogliomas, and ependymomas are three different glial cell-derived types of gliomas, astrocytomas being the most frequent [2]. Glioblastoma (GBM), the most malignant astrocytic tumor, is considered an incurable disease with a mean survival of 15 months for patients treated with the standard-of-care [3, 4].

The World Health Organization (WHO) grading system belongs to a set of clinical criteria aimed to predict treatment response and outcomes [5]. The WHO classification of tumors of the CNS has become much more accurate with the use of molecular markers, making them an integral part of deciding how to treat gliomas, reducing the interobserver variability and distinguishing new types and variants of tumors [6–8]. Although several biomarkers for diagnosis, risk, and prognosis have been studied, most of them require tumor tissue to be detected [3, 9].

In recent years, biochemical pathways involved in diverse mechanisms, such as the Renin-Angiotensin System (RAS) in blood-pressure control, are now being considered for playing a significant role in carcinogenesis [10]. The RAS, besides its well-known systemic regulation of the circulatory homeostasis, has a local or paracrine function [11]. Local expression of the RAS has been described in multiple tissues, such as liver, kidneys, or pancreas [12]; and also in cancer tissues, such as breast cancer [13], colorectal cancer [14], and renal cell carcinoma [15]. The expression of RAS components have been linked to the hallmarks of cancer [10, 12], and some of those components have been found to be upregulated in some cancer types, including GBM [16–18]. RAS demonstrated involvement in sustaining proliferative signaling, evading growth suppressors, resisting apoptosis, inducing angiogenesis, deregulating cellular energetics, as well as in inflammation, cellular migration, invasion and metastasis [10, 11, 19].

One of the essential RAS components is the human angiotensinogen (*AGT*) gene. The genetic variant rs5050 is a thymidine to guanosine substitution at nucleotide -58 of the 5'UTR of the gene *AGT*. In *in vitro* studies rs5050 has been confirmed to have a functional effect on promoter activity [20]. This has been consistent with studies that have found a correlation between this genetic variant, including haplotypes that contain it, and differences in blood levels of *AGT* [21, 22]. Additionally, the *AGT* rs5050 has been correlated with an increased risk of developing gastric cancer [23].

Recently, the discovery of RAS peptides and receptors in GBM [17] has urged the planning of clinical studies to elucidate the role of this new concept of the RAS in brain cancer [24–27] [28]. A better-characterized analysis of RAS in gliomas has been already described previously

[18]. The present study aimed to determine the relationship between the *AGT* rs5050 germline genetic variant with prognosis in astrocytoma.

Methods

Source of data

This prospective analytical study and its informed consent were approved by the Institutional Review Board of the National Institute of Neurology and Neurosurgery, Mexico City, Mexico, before recruitment of patients. TRIPOD reporting guideline was implemented [29]. ([S1 TRIPOD Checklist](#)).

Participants

A cohort of adult patients of both sexes, newly diagnosed with primary astrocytoma via histopathology, without prior treatment, were included after signing consent form. Patients with other glial cell-type tumors, prior treatments, or insufficient/degraded DNA samples, were excluded. The included patients underwent surgery for therapeutic and/or diagnostic purpose between 2013 and 2015. After surgery, complementary treatment with standard radiotherapy and chemotherapy were administered to high-grade gliomas and cases with progressive grade II glioma.

Outcomes

Demographic and clinical information was obtained from medical records by one of the researchers in a blinded fashion. Long-term survival was considered ≥ 3 years. Patients were grouped according to age into one of four groups for analysis purpose. The performance status was assessed within 1-week before surgery by the Karnofsky performance status (KPS) [30], which is an 11-level scale with scores ranging from normal activity (100) to death (0). A cutoff point of KPS ≥ 70 was used for analysis. The histological grading was taken from the pathology report, which was based on 2007 WHO classification. The extent of resection (EOR) was evaluated using the postoperative T1-weighted MRI scan with contrast and classified as gross total (100%), subtotal ($>90\%$) or partial (70–90%) resection. The major outcomes were risk and survival. Survival was defined as the lapse of time from when surgery is performed to the patient's death or last clinic visit.

Predictors

A sample of 5 mL of blood was taken from each patient before the surgery. Samples were unidentified and labeled using a coding system for internal control, and then, submitted to the laboratory. For DNA genomic extraction, Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) was used. As part of a more extensive study, Ion Torrent NGS (Thermo Fisher Scientific, Waltham, MA, USA) was used for sequencing the genetic variant rs5050. Customized Ion AmpliSeq panel was designed using Ion AmpliSeq designer software. Libraries were constructed using Ion AmpliSeq Library Kit v2.0 according to the manufacturer's instructions. The library was labeled with an individual adapter given in the Ion Xpress Barcode Adapters Kit. Sequencing was done using Ion 316 chip. Protocols were run on the NGS Ion OneTouch 2 System and the Ion OneTouch ES Instrument according to the user manual. All barcoded specimens were sequenced on the Ion View OT2 Kit. Ion Reporter software was used to perform primary to tertiary analysis, including optimized signal processing, base calling, sequence alignment, and variant analysis. In 12.5% of the samples,

orthogonal verification was performed and further verified by conventional Sanger sequencing with 100% validation.

Sample size and missing data

Sample size was calculated as described by Schoenfeld [31] to determine the minimum sample size for statistical purposes (S1 Table). Complete-case analysis was performed with no missing data identified.

Statistical analysis methods

In order to compare the allelic and genotypic frequencies of astrocytoma patients in this study, a control group was taken from the 1000 Genomes Browser Phase 3 version 3.7 [32][33][34], where the proportion of the rs5050 *AGT* allele and genotype frequencies were obtained. Hardy-Weinberg equilibrium (HWE) testing was performed on this control group. The wild-type allele and genotype were defined as the most common in the population. We performed descriptive statistics using means, medians, percentages, and maximum and minimum values. The bivariate analysis used contingency tables, and chi-square test and odds ratio were calculated within a 95% confidence interval, and p-value was calculated by Fisher's exact test. Median follow-up time was calculated using the Schemper and Smith method. The effect of each measured factor on time to death was identified using Kaplan Meier curves, Cox regression models, and Log-rank tests and were used to determine differences in survival function between subgroups. A p-value < .05 was considered statistically significant. All analyses were performed using SPSS v21 (IBM, Armonk, North Castle, NY), STATA 12 (StataCorp LLC, College Station, TX) and GraphPad Prism 7 (GraphPad Software, Inc., La Jolla, CA).

Results

Participants

Forty-eight astrocytoma patients were identified (50% males, 50% females, mean age 49.1, range 22–79 years). The most common tumor location was in the left frontal and temporal lobes, with 16.7% of patients in each location. 68.7% of cases corresponded to high-grade gliomas. Resection of >90% was achieved in 43.7% of patients, and partial resection in 56.3%. Out of the 24 (50%) patients who received chemotherapy, 11 (45.8%) could cover the cost of temozolomide, while the remaining patients received treatment regimes with carboplatin, vincristine, chloroquine, cisplatin, and/or carmustine. Postoperative KPS was ≥ 70 in 79.2% of patients, and <70 in 20.8%. Additional demographic data is shown in Table 1.

Model development

AGT rs5050 gene genetic variant. The rs5050 control group consisted of 2,504 genotypes counted for the worldwide population, including 170 genotypes for the Mexican-ancestry population, which were in HWE after calculations. According to the 1000 Genomes Browser, the distribution of the rs5050 *AGT* alleles and genotypes in the worldwide population is as follows: T-allele was present in 82.4% while G-allele in 17.6%; and the TT-genotype was in 67.9%, TG- in 29% and GG- in 3.1%. In the Mexican-ancestry population in Los Angeles, California, the frequencies were: T-allele in 79.8% and G-allele in 20.2%; and the TT-genotype was in 62.7%, TG- in 34.3% and GG- in 3.0%. From these data, the wild-type and the risk allele and genotype were defined. Then, the proportion of the rs5050 frequencies in our cohort was identified. Out of the 48 patients, the T-allele was present in 77.1% while G-allele in 22.9%; the genotypes corresponded to TT-, TG- and GG-genotypes in 58.3%, 37.5%, and 4.1%, respectively. When

Table 1. Demographical and clinical characteristics.

n = 48	n (%)
Demographic Variables	
Age*	49 ± 14.1
Sex	M: 24 (50), F: 24 (50)
Morbidity	
Hypertension	9 (18.7)
Diabetes Mellitus type 2	8 (16.6)
Smoking	15 (31.2)
Alcohol Abuse	5 (10.4)
Previous Neoplasm	4 (8.3)
Socio-economic Status	
Very-Low income	28 (58.3)
Low income	17 (35.4)
Medium income	2 (4.2)
High income	1 (2.1)
Clinical Variables	
Age* at Diagnosis	
Grade II	40
Grade III	46.6
Grade IV	54.5
Tumor Location	
Frontal	16 (33.3)
Temporal	14 (29.1)
Parietal	9 (18.8)
Occipital	2 (4.2)
Others	7 (14.6)
Initial Symptom	
Headache	14 (29.1)
Motor deficit	9 (18.8)
Generalized Seizures	10 (20.8)
Cognitive Functions	8 (16.7)
Partial Seizures	2 (4.2)
Others	5 (10.4)
Karnofsky Performance Status	
≥70	38 (79.2)
<70	10 (20.8)
WHO Grade	
Grade II	15 (31.3)
Grade III	5 (10.4)
Grade IV	28 (58.3)
Extent of Resection	
>90%	21 (43.8)
<90%	27 (56.2)

* = years. M = male, F = female.

<https://doi.org/10.1371/journal.pone.0206590.t001>

comparing genotypic and allelic frequencies, we identified that the allelic and genotypic frequencies of our cohort were similar compared to the Mexican-ancestry population.

Table 2. AGT rs5050 allelic and genotypic frequencies (%).

	AGT rs5050 Allelic and Genotypic Frequencies (%)		
Allelic Frequencies (n)	T	G	
Present study (96)	77.1	22.9	
MXL (134)*	79.8	20.2	
Global (5,008)*	82.4	17.6	
Genotypic Frequencies (n)	TT	TG	GG
Present study (48)	58.3	37.5	4.1
MXL (67)*	62.7	34.3	3.0
Global (2,504)*	67.9	29	3.1

* 1000 Genomes Browser, version 3.7 (last updated: March 1, 2018)

MXL = Mexican-ancestry in Los Angeles, California.

<https://doi.org/10.1371/journal.pone.0206590.t002>

Additionally, we observed a slightly higher prevalence of the G-allele and GG- and TG-genotypes in our cases in comparison with the worldwide population, which was not statistically significant. Nevertheless, statistical analysis did not reveal AGT rs5050 to be a significant risk factor ($p = .48$ and $p = .34$) (Table 2).

Model specification

Survival analysis. Median follow-up was 41 months (range 1–48). The mean survival was 14.8 (range 1–50) months. Long-term survival was reported in five (10.4%) patients. Out of the 40 deceased patients, 17 (42.5%) patients died at ≤ 0 –6 months, 10 (25%) at > 6 to ≥ 12 months, 8 (20%) at > 12 to ≤ 24 months, and 5 (12.5%) at ≥ 24 months.

The patients were grouped by AGT rs5050 genotype, and the survival times were analyzed to obtain the Kaplan-Meier curves. The survival analysis was performed with the rs5050 genotypes (TT, TG, and GG) separately, and with every genotype against the other two genotypes grouped. Survival analysis showed a trend towards significance when genotypes were studied separately ($p = .05$) (Fig 1A). When comparing each genotype against the two remaining genotypes in Kaplan-Meier survival curves, we found lower survival rates among individuals with the GG-genotype of AGT rs5050 (2 months in the GG-genotype carriers vs. 11.5 months in the group of TT- plus TG-genotype patients [$p = .01$]) (Fig 1B).

Bivariate analysis showed a statistically significant difference between the survival status comparing GG- vs. TG-, and GG- vs. TT-genotypes ($p = .011$ and $p = .016$, respectively). Table 3 presents the crude hazard ratio (HR) for survival, by cohort, at the end of the follow-up period. Clinical variables analyses, such as histological grade, KPS, and EOR, among others, were also conducted. Clinical markers of younger age ($p = .02$) (S1A Fig), lower histological grade ($p = .003$) (S1B Fig) and higher KPS ($p = .04$) (S1C Fig) were also related to longer survival.

Model performance

In the multivariate Cox proportional HR regression analysis, the model was adjusted for clinical covariables that were considered relevant for survival. After adjusting for gender, KPS, EOR, chemotherapy, radiotherapy, and WHO grade, AGT rs5050 genotype remained an independent risk factor for survival with an adjusted HR of 1.000, 0.009 (95% C.I. 0.00–0.09, $p = .000$) and 0.02 (95% C.I. 0.00–0.17, $p = .000$), for GG-, TG- and TT-genotypes, respectively. Other covariables that were significant risk factors in this multivariate model for survival were KPS ($p = .025$), WHO grade ($p = .001$), and non-chemotherapy ($p = 0.000$) after adjusting.

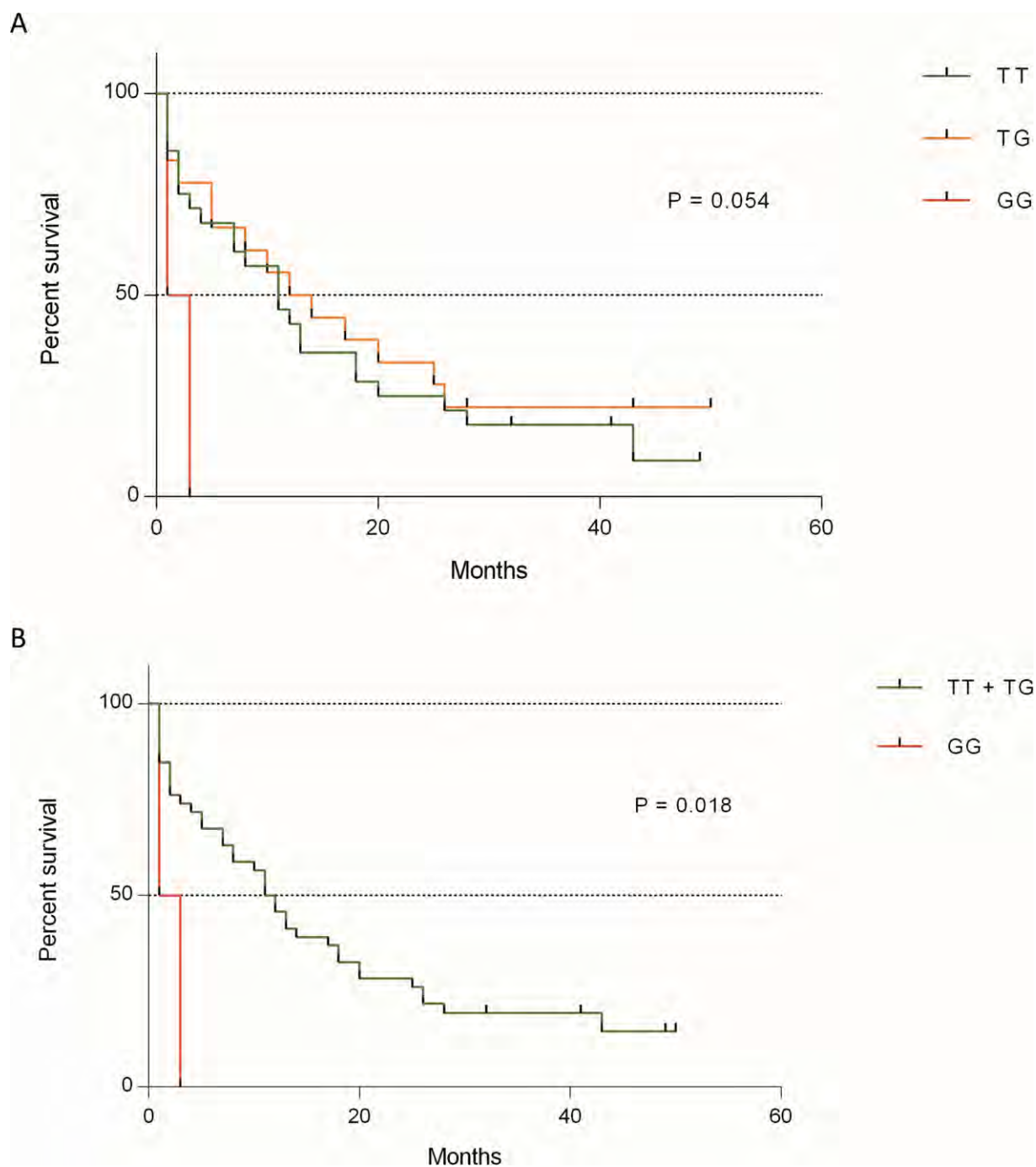


Fig 1. Kaplan-Meier survival estimates. (A) Survival analysis of the *AGT* rs5050 genotypes (TT in green, TG in orange, and GG in red) separately, showing the shorter survival in GG-genotype patients ($p = .05$). (B) Comparing GG-genotype against TT + TG-genotypes of *AGT* rs5050, patients harboring GG-genotype (in red) exhibited lower survival rates compared to TT + TG (in green) genotypes patients (2 vs. 11.5 months [$p = .01$]).

<https://doi.org/10.1371/journal.pone.0206590.g001>

Table 3. Crude hazard ratios for survival.

Cohort	HR	95% CI	P value
Age			
	1.032	1.01–1.06	0.010
Gender			
Female	1.000	-	-
Male	1.514	0.81–2.83	0.193
AGT rs5050 Genotype**			
GG	1.000	-	-
TG	0.130	0.03–0.62	0.011
TT	0.152	0.03–0.70	0.016
Karnofsky Performance Status**			
	0.976	0.96–0.99	0.003
Extent of Resection			
>90%	1.000	-	-
<90%	0.718	0.39–1.34	0.296
Radiotherapy			
No	1.000	-	-
Yes	0.890	0.48–1.66	0.718
Chemotherapy**			
No	1.000	-	-
Yes	0.858	0.45–1.63	0.639
WHO Grade**			
II	1.000	-	-
III	2.015	0.22–18.26	0.533
IV	4.590	0.62–34.09	0.136
Smoking			
No	1.000	-	-
Yes	1.195	0.61–2.33	0.604
Alcoholism			
No	1.000	-	-
Yes	1.130	0.40–3.19	0.817
Familial cancer history			
No	1.000	-	-
Yes	2.090	1.08–4.07	0.030

** Study cohorts with *p* values < .05 in adjusted hazard ratios calculated using the Cox proportional hazard regression analysis.

<https://doi.org/10.1371/journal.pone.0206590.t003>

Discussion

In this study, the statistical analysis revealed the GG-genotype of *AGT* rs5050 to be a significantly poor prognosis factor. With a median follow-up of 41 months, GG-genotype carriers of the *AGT* rs5050 presented a statistically significant shorter survival than the other two genotypes (2 vs. 11 months, *p* = .01). The blood-detected GG-genotype reported this significant correlation with poor prognosis demonstrating value not just as an independent variable but also in the context of other variables. It would give the clinician an additional tool, which combined with other factors, provides a basis for the pursuit of a personalized prognosis for these patients.

Limitations

Even though the size of the population was relatively small, the number of cases was sufficient for statistical purposes to obtain valid results and to define the direction of further studies. Another limitation of this study is the fact that was performed in a single tertiary referral center, which is entirely dedicated to diseases of the nervous system. Our institution is one of the main hospitals of the Department of Public Health in Mexico, which receives patients from all throughout the central region of the country. The population of this study was composed of Mexicans only. Recently, a US-based study using the zip code tabulation areas codes as a proxy for socioeconomic status (SES), showed that SES did not affect prognosis in patients newly diagnosed GBM [35]. For this reason, the SES of the patients included in this cohort, where almost all of them were classified as very-low and low income, was not considered a bias for survival. However, it is necessary to mention that most of these patients had to cover the cost of their chemotherapy drugs, hindering the treatment with more standardized regimes among the patients. Finally, a reasonable limitation is that O⁶-methylguanine DNA methyltransferase (MGMT) and Isocitrate Dehydrogenase (IDH) mutation were not consistently investigated, as a result of lack of financial resources, and as the patients were recruited before the updated WHO classification was published [36]. Therefore, an attractive opportunity to examine in upcoming studies is the full molecular characterization of the tumor tissue, to analyze the interaction between the expression of *AGT* rs5050 in blood and these biomarkers currently used.

Interpretation

AGT rs5050 gene and angiotensinogen. The *AGT* gene and its genetic variants have been previously studied in the context of the pathogenic mechanisms of essential hypertension, particularly the impact of plasma AGT levels on blood pressure. *In vitro* studies inferred that the transcriptional cis-element of the *AGT* gene regulates blood pressure by managing plasmatic levels of AGT [37]. Yanai *et al.* identified the 5'UTR of the human *AGT* gene as a regulator for the transcription of AGT mRNA [38]. Ishigami *et al.* suggested that rs5050 might influence the level of transcription of AGT mRNA in humans, and therefore modify the plasma AGT concentration. However, they reported a weak correlation between the *AGT* rs5050 and plasmatic AGT levels in a multiple regression analysis, and no differences in plasma AGT concentration among the three genotypes in the analysis of variance [39]. Fourteen years later and using a more sensitive and specific quantification system for human AGT, Balam-Ortiz *et al.* found differences in AGT plasma levels between the genotypes of rs5050, with the GG-genotype associated with the lowest levels. Their results of a bivariate analysis (TT = 25.3±8.3 µg/mL; TG = 22.1±7.1 µg/mL; GG = 19.4±4.8 µg/mL) were statistically significant (TT vs. TG, $p = .03$; TT vs. GG, $p = .05$; TT vs. TG+GG, $p = .008$). Their regression analysis confirmed these differences of the plasmatic AGT levels in the H2 (which contains rs5050) and H8 (which contains -58 wildtype) haplotypes. The H2 haplotype was associated with the lowest plasma AGT levels (-5.1 µg/mL [95% C.I. -8.6 to -1.6], $p = .004$), while the H8 haplotype was linked to the highest plasma AGT levels (6.5 µg/mL [95% C.I. 2.5–10.6], $p = .001$) [22].

AGT rs5050 and gliomagenesis. The mechanism through which the *AGT* gene influences cancer behavior might stem from genetic variants, bioactive peptides, enzymes, and receptors that have been recently summed to the RAS network [10, 19, 40]. Angiotensin peptides act principally via the AngII receptor type 1 (AT1R) and type 2 (AT2R), and secondarily through the Mas receptor [19, 40]. AT1R and AT2R are receptors with pleiotropic actions with opposing effects. When stimulated, AT1R favors cellular proliferation and angiogenesis while AT2R

has antiproliferative attributes [10]. Growing data demonstrate that AT1R is present in several types of neoplasms and that its expression is correlated with tumor growth and a more aggressive disease [12, 41–43]. AT1R stimulates diverse intracellular signaling pathways, leads upregulation of transforming growth factor beta, and induces vascular endothelial growth factor (VEGF) [44–46]. The function of AT2R in cancer is less known. While AT2R has been mainly described as an antiproliferative and proapoptotic mediator, proproliferative and angiogenic effects *in vivo* have also been mentioned in conflicting reports, such as in an AT2R knockout mouse model that showed that the inhibition of AT2R hinders tumor growth by reducing VEGF expression [47].

The promoter activity of *AGT* rs5050 regulates the beginning of the pathway, decreasing the transcription of *AGT* and its concentration in the plasma. Using well-established *in vitro* and *in vivo* models, C  lerier *et al.* demonstrated an antiangiogenic effect of *AGT* [48]. As they concluded, these opposite effects of *AGT*, showing an antiangiogenic property as a serpin and a proangiogenic activity as the precursor of AngII, might depend on local conditions that define which of the effects prevails [48] (Fig 2). One of the advantages of the *AGT* rs5050 is that is a germline genetic variant, and thus it can be found in the majority of the patient's cells, including leukocytes from blood and even in normal tissue neighboring neoplasms, unlike the somatic mutations that are found heterogeneously in tumor tissue only [49]. With the knowledge that the GG-genotype is related to decreased plasmatic *AGT* levels, and that *AGT* owns a physiological antiangiogenic activity, we infer that the GG-genotype might contribute to a proangiogenic tumor environment, and therefore, to more aggressive behavior and worse outcomes.

Implications. To date, just a few clinical studies have explored the feasibility of a relation between RAS components and gliomas [18]. Arrieta *et al.* described the potential of the AT1R as prognosis biomarker [50]. Expression of both receptors, AT1R and AT2R were analyzed in tumor tissue from astrocytoma patients, and a higher prevalence of the two receptors was found in high-grade astrocytomas. AT1R and AT2R were associated with higher cellular proliferation and angiogenesis. AT1R-positive tumors were related to a lower survival rate compared to those which were AT1R-negative [50]. In another study, Lian *et al.* suggested the insertion/deletion (I/D) genetic variant of the *ACE* gene as a risk biomarker for glioma, and performed a case-control study in a Chinese population, which showed that glioma patients had a significantly higher prevalence of the *ACE* DD-genotype detected in blood [51]. The *AGT* rs5050 is so far the only RAS component that is simultaneously related to prognosis and identified by a blood test, which would be both valuable and practical features for a biomarker in a clinical setting. Recent studies have indirectly supported this notion of a relationship between the RAS and prognosis in gliomas, by analyzing the impact of the use of AngII receptor 1 blockers (ARBs) in survival: Carpentier *et al.* described the use of ARBs as a significant prognostic factor for both progression-free survival (PFS) and OS in GBM patients treated with the standard-of-care [25]; and, Levin *et al.* reported an OS benefit offered by ARBs in recurrent GBM patients treated with low-dose Bevacizumab (BVZ) [46]. Recently, Urup *et al.* proposed the *AGT* gene as a predictive biomarker of BVZ response [52]. In that retrospective study, the low expression of *AGT* in the tumor tissue was associated with a prolonged PFS and OS in recurrent GBM patients treated with BVZ [52]. As mentioned, the GG-genotype of rs5050 has been correlated with lower *AGT* expression. C  lerier *et al.* [48] reported that *AGT* could exhibit an antiangiogenic effect as a serpin. Hence, the downregulation of *AGT* might decrease the serpine-mediated antiangiogenic activity of *AGT*, and consequently, contribute to a proangiogenic tumor microenvironment. This effect might cause the tumor to become more susceptible particularly to anti-angiogenic therapy. This hypothetical mechanism could explain, at least in part, the increased survival seen by Urup *et al.* [52] in those recurrent GBM

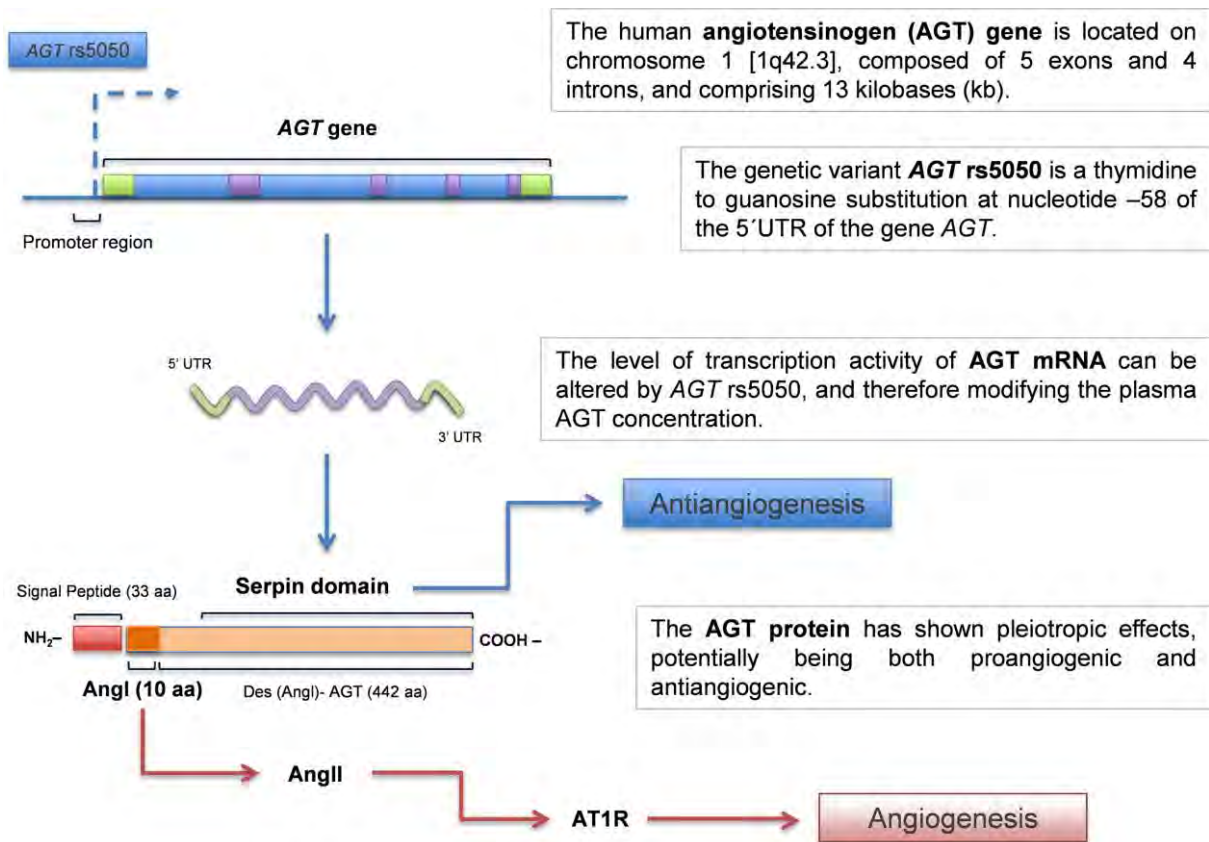


Fig 2. Hypothetical dual mechanism of the AGT rs5050 genetic variant. The 5' upstream core promoter region of the human AGT gene, where rs5050 is identified, has been recognized as an authentic regulator for the transcription of AGT mRNA [38]. Differences in the AGT plasma levels were found between the genotypes of rs5050, with the GG-genotype associated with the lowest levels [22]. AGT expresses opposite effects, showing an antiangiogenic property, such as some Serpins family members, and a proangiogenic activity as the precursor of AngII. It might depend upon local conditions that define which of the effects dominates [48].

<https://doi.org/10.1371/journal.pone.0206590.g002>

patients with low expression of AGT when they were treated with BVZ, and also might be the reason for the poor prognosis that was seen in the patients harboring GG-genotype (and not treated with BVZ) in our study. In light of the evidence that the GG-genotype of rs5050 is linked to the lower AGT expression, and that the lower AGT expression, in turn, is related to longer survival particularly in these GBM subjects with anti-VEGF treatment, we might hypothesize for future studies a plausible role of AGT rs5050 as a potential biomarker of BVZ response.

Conclusions

The AGT rs5050 germline genetic variant is proposed as a complementary biomarker to predict survival, which is detected by a blood or saliva test, safely, less invasive, and before surgery. The GG-genotype of AGT rs5050 was related to poor prognosis in our cohort, becoming the first study that analyzes and identifies a blood RAS component as a prognosis factor. These results encourage further and broader investigation to endorse this finding and to validate our

conclusions in other populations. Future necessary studies regarding *AGT* rs5050 will include its consideration as a treatment-response biomarker or as a druggable target.

Supporting information

S1 TRIPOD Checklist.

(DOCX)

S1 Fig. Kaplan-Meier survival estimates. Clinical variables analyses, such as histological grade, KPS, and EOR, among others, were also conducted. (A) Younger age ($p = .02$), (B) lower histological grade ($p = .003$), and (C) higher KPS ($p = .04$) were related to longer survival.

(TIF)

S1 Table. Parameters used in sample size calculation.

(DOCX)

Acknowledgments

We thank Patricia Rosas, Laura Marqu ez and Patricia de la Torre for their invaluable help in this project. A sincere thank you to Joshua Casaos for his diligent proofreading of this manuscript.

Author Contributions

Conceptualization: Alexander Perdomo-Pantoja, Sonia Iliana Mej a-P erez, Nancy Reynoso-Nover n, Silvia Vidal-Mill n, Jos  Sanchez-Corona, Talia Wegman-Ostrosky.

Data curation: Alexander Perdomo-Pantoja, Sonia Iliana Mej a-P erez, Nancy Reynoso-Nover n, Thal a Estefania S nchez-Correa, Talia Wegman-Ostrosky.

Formal analysis: Alexander Perdomo-Pantoja, Nancy Reynoso-Nover n, Liliana G mez-Flores-Ramos, Lissania Guerra-Calderas, Clementina Castro-Hernandez, Silvia Vidal-Mill n, Lucia Taja-Chayeb, Luis Alonso Herrera-Montalvo.

Funding acquisition: Talia Wegman-Ostrosky.

Investigation: Alexander Perdomo-Pantoja, Sonia Iliana Mej a-P erez, Ernesto Soto-Reyes, Thal a Estefania S nchez-Correa, Olga Guti rrez, Rosa Maria Alvarez-Gomez, Talia Wegman-Ostrosky.

Methodology: Jos  Sanchez-Corona, Bernardo Cacho-Diaz, Talia Wegman-Ostrosky.

Project administration: Talia Wegman-Ostrosky.

Resources: Juan Luis G mez-Amador, Teresa Corona, Talia Wegman-Ostrosky.

Supervision: Talia Wegman-Ostrosky.

Validation: Nancy Reynoso-Nover n, Patricia Ostrosky-Wegman.

Writing – original draft: Alexander Perdomo-Pantoja, Liliana G mez-Flores-Ramos, Talia Wegman-Ostrosky.

Writing – review & editing: Sonia Iliana Mej a-P erez, Nancy Reynoso-Nover n, Liliana G mez-Flores-Ramos, Ernesto Soto-Reyes, Thal a Estefania S nchez-Correa, Lissania Guerra-Calderas, Clementina Castro-Hernandez, Silvia Vidal-Mill n, Jos  Sanchez-Corona, Lucia Taja-Chayeb, Olga Guti rrez, Bernardo Cacho-Diaz, Rosa Maria Alvarez-

Gomez, Juan Luis Gómez-Amador, Patricia Ostrosky-Wegman, Teresa Corona, Luis Alonso Herrera-Montalvo.

References

1. Xavier-Magalhaes A, Nandhabalan M, Jones C, Costa BM. Molecular prognostic factors in glioblastoma: state of the art and future challenges. *CNS Oncol*. 2013; 2(6):495–510. <https://doi.org/10.2217/cns.13.48> PMID: 25054820.
2. Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008–2012. *Neuro-oncol*. 2015; 17 Suppl 4:iv1–iv62. <https://doi.org/10.1093/neuonc/nov189> PMID: 26511214; PubMed Central PMCID: PMC4185005.
3. Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, et al. Epidemiologic and molecular prognostic review of glioblastoma. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2014; 23(10):1985–96. <https://doi.org/10.1158/1055-9965.EPI-14-0275> PMID: 25053711; PubMed Central PMCID: PMC4185005.
4. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *The Lancet Oncology*. 2009; 10(5):459–66. [https://doi.org/10.1016/S1470-2045\(09\)70025-7](https://doi.org/10.1016/S1470-2045(09)70025-7) PMID: 19269895.
5. Delgado-Lopez PD, Corrales-Garcia EM. Survival in glioblastoma: a review on the impact of treatment modalities. *Clin Transl Oncol*. 2016. <https://doi.org/10.1007/s12094-016-1497-x> PMID: 26960561.
6. Aldape K, Zadeh G, Mansouri S, Reifenberger G, von Deimling A. Glioblastoma: pathology, molecular mechanisms and markers. *Acta neuropathologica*. 2015; 129(6):829–48. <https://doi.org/10.1007/s00401-015-1432-1> PMID: 25943888.
7. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell*. 2006; 9(3):157–73. <https://doi.org/10.1016/j.ccr.2006.02.019> PMID: 16530701.
8. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010; 17(1):98–110. <https://doi.org/10.1016/j.ccr.2009.12.020> PMID: 20129251; PubMed Central PMCID: PMC2818769.
9. Preusser M. Neuro-oncology: a step towards clinical blood biomarkers of glioblastoma. *Nature reviews Neurology*. 2014; 10(12):681–2. <https://doi.org/10.1038/nrneuro.2014.208> PMID: 25366110.
10. Wegman-Ostrosky T, Soto-Reyes E, Vidal-Millan S, Sanchez-Corona J. The renin-angiotensin system meets the hallmarks of cancer. *Journal of the renin-angiotensin-aldosterone system: JRAAS*. 2015; 16(2):227–33. <https://doi.org/10.1177/1470320313496858> PMID: 23934336.
11. George AJ, Thomas WG, Hannan RD. The renin-angiotensin system and cancer: old dog, new tricks. *Nat Rev Cancer*. 2010; 10(11):745–59. <https://doi.org/10.1038/nrc2945> PMID: 20966920.
12. Ager EI, Neo J, Christophi C. The renin-angiotensin system and malignancy. *Carcinogenesis*. 2008; 29(9):1675–84. <https://doi.org/10.1093/carcin/bgn171> PMID: 18632755.
13. Gonzalez-Zuloeta Ladd AM, Arias Vasquez A, Siemes C, Yazdanpanah M, Coebergh JW, Hofman A, et al. Differential roles of Angiotensinogen and Angiotensin Receptor type 1 polymorphisms in breast cancer risk. *Breast cancer research and treatment*. 2007; 101(3):299–304. <https://doi.org/10.1007/s10549-006-9290-0> PMID: 16823505.
14. Vasku A, Vokurka J, Bienertova-Vasku J. Obesity-related genes variability in Czech patients with sporadic colorectal cancer: preliminary results. *International journal of colorectal disease*. 2009; 24(3):289–94. <https://doi.org/10.1007/s00384-008-0553-6> PMID: 18704460.
15. Andreotti G, Boffetta P, Rosenberg PS, Berndt SI, Karami S, Menashe I, et al. Variants in blood pressure genes and the risk of renal cell carcinoma. *Carcinogenesis*. 2010; 31(4):614–20. <https://doi.org/10.1093/carcin/bgp321> PMID: 20047954; PubMed Central PMCID: PMC2847086.
16. Friis S, Sorensen HT, Mellemkjaer L, McLaughlin JK, Nielsen GL, Blot WJ, et al. Angiotensin-converting enzyme inhibitors and the risk of cancer: a population-based cohort study in Denmark. *Cancer*. 2001; 92(9):2462–70. PMID: 11745304.
17. Juillerat-Jeanneret L, Celerier J, Chapuis Bernasconi C, Nguyen G, Wostl W, Maerki HP, et al. Renin and angiotensinogen expression and functions in growth and apoptosis of human glioblastoma. *British*

- **PUBLICACIÓN 4:** Epigenetic Data and Disease, Capitulo del libro: **Analysing Network Data in Biology and Medicine.**
- **AUTORES:** Rodrigo González-Barrios, Marisol Salgado-Albarrán, Nicolás Alcaraz, Cristian Arriaga-Canon, Lissania Guerra-Calderas, Laura Contreras-Espinosa, and Ernesto Soto-Reyes
- **EDITORIAL:** Cambridge University Press. **AÑO:** 2019.

2 Epigenetic Data and Disease

Rodrigo González-Barrios, Marisol Salgado-Albarrán, Nicolás Alcaraz, Cristian Arriaga-Canon, Lissania Guerra-Calderas, Laura Contreras-Espinosa, and Ernesto Soto-Reyes

2.1 Background

Genetic material carries the information for every process necessary for life. Environmental exposure to mutagens can alter genetic information throughout lifetime and together with genetic predisposition could generate diseases in an organism. However, genetic information is not the only mechanism underlying the trans-generational transmission and environment influence on human variation and disease. The relatively new field of epigenetics has given a new perspective to the origin of diseases. The term “epigenetics” was first defined by Waddington in 1939 as “The random interactions between genes and their products, which result in a phenotype” [1]. Nowadays, epigenetics is mostly defined as heritable changes that regulate gene expression and chromosome architecture independent of any DNA sequence. In recent years, it has been emphasized that epigenetic components dictate and coordinate gene expression. The human body has more than 100 distinct cell types, which have essentially the same genome but contain a unique epigenome that serves to instruct specific gene expression programs present within each cell type. Epigenetic modifications are highly dynamic and can be altered throughout aging and environmental exposure. There are many types of epigenetic modifications, which include DNA methylation and posttranslational histone modifications such as methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation. Also, non-coding RNAs play a key role in the regulation of epigenetic processes. Altogether, epigenetic mechanisms are essential to many cellular functions, and when dysregulated, major adverse health and behavioral effects occur.

Epigenetic aberrations could explain the origin and prevalence of certain diseases. The most studied relationship between epigenetics and disease has been cancer [2], but aberrant epigenetic patterns go beyond oncology, stretching to a variety of biomedical fields including imprinting, metabolism, neurology, immunology, development, cardiovascular disease, etc. [2, 3, 4, 5, 6]. Researchers worldwide understand the importance of epigenetic mechanisms in disease biology, leading to the origin of many techniques to study different levels of epigenetic information, regarding DNA

methylation, histone post-translational modification, chromatin associated proteins, chromatin remodelers, and non-coding RNAs. With the introduction of whole genome sequencing and epigenomics, the knowledge of epigenetic mechanisms and its implications in cellular processes rapidly increased and became one of the most flourishing areas in biology and medicine, leading to great scale international efforts to characterize the epigenome from different samples and diseases. However, this field is still in its infancy and further research is still needed to understand the epigenetic machinery and its biological function. In this chapter, we will review different mechanisms from a bioinformatics perspective, experimental approaches and computational methods to study epigenetic data (see Box 2.1 for glossary of terms).

2.2 DNA Methylation and its Role in Genome Regulation

Among the most widely studied epigenetic mechanisms is DNA methylation, a covalent modification that mainly occurs at the fifth carbon position of the cytosine, thus forming the 5-methylcytosine (5mC). The molecular machinery related to 5mC can be divided into three components that can establish, read, and remove this mark: (1) The three active DNA methyltransferases (DNMT1, DNMT3a and 3b) which catalyze and maintain cytosine methylation patterns; (2) the methyl binding proteins (MBPs), that read this mark and carry out different effector mechanisms like gene silencing; and (3) the DNA demethylases (TET1, 2 and 3) that actively “erase” 5mC through a series of oxidations and glycosylations (see Table 2.1). In normal cells, genome wide studies demonstrated that 70 to 80% of the CpG sites are methylated, predominantly in repetitive genomic regions, including satellite repeats, viral sequences, LINEs, and SINEs (long and short interspersed transposable elements) and gene bodies [7, 8, 9, 10].

Paradoxically, the regions of the genome that contains most of CpG dinucleotides are usually not methylated. These regions are known as “CpG islands” (CGI) and were first described in 1987 by Gardiner and Frommer. They are short genomic regions (around 200–1000bp) that are defined by an increase in cytosine and guanine greater than 50%. CGI are particularly associated with gene promoters and regulatory regions, where approximately 70% of annotated gene promoters are associated with CGI, making this the most common promoter type element in the vertebrate genome [11, 12]. Virtually all housekeeping genes, as well as a proportion of tissue specific genes and developmental regulator genes are associated with CGI [13, 14]. There is evidence that a large class of CGIs are located remotely from annotated transcription start sites (TSS), moreover, some show evidence for promoter function, although an increasing number of exceptions are being identified suggesting other functions related to this CGIs [15].

Methylation at CGI promoters is associated with transcriptional repression by interfering with transcription factor binding; nevertheless, repression seems to occur largely indirectly, via recruitment of MBPs that induce chromatin remodeling. However, the strength of repression mediated by 5mC depends on the local concentration of CpGs within the promoter. Through DNA methylation mapping by massive

Box 2.1: Glossary

Chromatin conformation capture (5C, HiC): A technique used to profile all chromatin interactions in specific regions of the genome by the hybridization of a mixture of DNA primers to chromosome conformation capture (3C) templates followed by high-throughput sequencing.

Cis-acting: Regions of non-coding DNA, which regulate transcription within the same chromosome.

CpG Island: Genomic regions with a minimum of 200 bp, with a G+C content greater than 50% and observed/ expected CpG ratio above 60%.

Enhancer: A cis-acting regulatory sequence that markedly increases expression of a neighboring gene. Enhancers are typically capable of operating over considerable distances (sometimes ~50 kb) upstream or downstream of the gene and in either orientations.

Epigenomics: Is the systematic analysis of the global state of gene expression modulated by epigenetic processes such as DNA methylation, posttranslational modifications of histones non-coding RNA and the organization of chromatin inside the nucleus.

Euchromatin: Less densely packed or open chromatin that is often associated with active transcription.

Global hypomethylation: Loss of DNA methylation across the genome that commonly occurs in cancer cells.

Heterochromatin: Tightly packed form of chromatin that lack high number of genes and is commonly constituted by repetitive sequences in the genome, which is associated with inactive transcription and serves as a structural element of the chromosome.

Hi-C contact matrix: Matrix which displays all chromatin interactions found within a genomic range. First, the genome is partitioned into bins of fixed size. Then, a contact matrix is generated, where every cell corresponds to the frequency of contacts between the associated pair of loci. The frequency or number of interactions is then converted into color signal.

Local hypermethylation: Gain of methylation that occurs at specific regulatory regions that alters the normal state of transcription in diseases like cancer.

paralleled sequencing it was reported that there are three classes of promoters based on CpG ratio, CG content, and length of the CpG rich region: High CpG promoters, intermediate CpG promoters and low CpG promoters (HCP, ICP, and LCP respectively) [16]. HCP and most ICP remain largely hypomethylated and transcriptionally active, these types of promoters are associated with housekeeping and some tissue specific genes [16]. In contrast, CGI with LCP are predominantly methylated, and such hypermethylation does not affects gene expression, which indicates that repression

Table 2.1: DNA modifications, function, and writer enzymes

DNA Modification	Name	Function
5mC	5-methylcytosine	Repression, alternative expression
5hmC	5-hydroxymethylcytosine	Transcription regulation
5fmC	5-formylcytosine	Unknown
5cmC	5-carboxylcytosine	Unknown
Human enzymes	Gene name	Function
DNMT1	DNA methyltransferase 1	DNA methylation maintenance
DNMT2	DNA methyltransferase 2	Low DNA activity, methylates tRNA
DNMT3A	DNA methyltransferase 3A	De novo methylation
DNMT3B	DNA methyltransferase 3B	De novo methylation
DNMT3L	DNA methyltransferase 3L	Cofactor of De novo methyltransferases activity
TET1	Ten-eleven-translocation protein 1	DNA oxydase/ demethylation
TET2	Ten-eleven-translocation protein 2	DNA oxydase/ demethylation
TET3	Ten-eleven-translocation protein 3	DNA oxydase/ demethylation

by DNA methylation requires high 5mC density, and such density of 5mC is also necessary for the repressive function of MBPs.

DNA methylation is a major epigenetic mechanism that determines cellular outcome and response to environmental signals, by dynamically regulating gene expression and chromatin formation. Due to its importance in development and cell differentiation, specific patterns of 5mC are established among cell types. This pattern of methylation is not static and can be altered in response to environmental stressors. Specific cell patterns are maintained by molecular mechanisms that keep “epigenetic memory” and maintain the cellular identity across its lifetime. However, altered DNA methylation is frequently observed in diseases, like cancer, compared with corresponding healthy cells.

Comprehensively studying the profiles of different healthy individuals and tissue types enables us to estimate the variance of a particular CpG site or of regions such as promoters. Reference data sets are now being created in consortia such as Blueprint, the International Human Epigenome Consortium (IHEC) and Roadmap using high-resolution technologies, which will be discussed later. Focusing on normal tissue types, these joint efforts aim freely to release reference datasets of integrated epigenomic profiles of stem cells, as well as developmental and somatic tissue types to the research community. Concordantly, filtering for *loci* that are unstable in DNA methylation between individuals excludes unsuitable CpG sites before biomarker candidate selection approaches. Systematic screening of reference data sets obtained from different individuals will enable us to identify and exclude variable CpG sites and regions, facilitating future biomarker selections.

2.2.1 DNA Demethylation and its Role in Genomic Profiles

Until recently, it was believed that DNA methylation was an irreversible epigenetic mark because it is a covalent modification. Now, it is known that DNA methylation is a dynamic process because different processes can remove it. DNA demethylation can occur passively by the reduction or absence in the enzymatic activity of the DNMTs, or actively, mediated by enzymes with the ability to remove the methyl group from the cytosines [17].

In 2009, it was discovered the first enzyme capable of erasing the 5mC, by an oxidation process [18]. These enzymes belong to the family of dioxygenases known as TET (ten-eleven-translocation proteins), composed of three members (TET1, TET2, and TET3) [19]. TET enzymes have the ability to convert de 5mC to 5-hydroxymethylcytosine (5hmC), mediating an active DNA demethylation process. Also, the 5hmC can be further catalyzed to 5-formylcytosine (5fC), presenting the highest level of oxidation with 5-carboxylcytosine (5caC). 5caC can be efficiently removed by thymine DNA glycosylase (TDG)-mediated bases excision repair.

The distribution of 5hmC depends on multiple factors such as cell type, cell differentiation and response to the environment. These questions have been addressed from a genome wide analysis perspective. An example of the importance of these enzymes in diseases is TET2. TET2 is one of the most frequently mutated genes in hematopoietic malignancies as an early event in cancer [20].

2.2.2 Different Experimental Strategies for the DNA Methylation Analysis

DNA methylation has been actively studied for the past four decades, using different approaches according to the available technologies (Figure 2.1). Initially global 5mC levels were studied with different approaches like: High-performance liquid chromatography (HPLC) or luminometric methylation assay (LUMA) [21, 22]. However, these methods only screen general levels of 5mC regardless of its genomic patterns. Other methods use methyl-sensitive restriction endonucleases; nevertheless, these analyses were restricted to the sequence of the enzymes and show digestion biases. Bisulfite conversion promptly became a gold standard technique. With this method, methylation research acquired the possibility to do aimed studies for specific sites by methyl specific polymerase chain reaction (MS-PCR), or to search for methylation patterns from specific regions to whole genome mapping by combining sequencing to bisulfite converted sequences (see Section 2.3).

Alternatively, region specific or global 5mC status can be assessed by immunoprecipitation of methylated sequences (MeDIP). This method is an adaptation of the chromatin immunoprecipitation (ChIP) protocol for DNA and uses an antibody against 5-methylcytosine. However, some biases come from MeDIP method, because it cannot establish CpG methylation patterns due to the resolution based on the size of immunoprecipitated DNA fragments (~200 to 500 bp), and the amount of enrichment of methylated DNA depends on the abundance of CpGs in a given sequence and the capacity of the antibody to detect such levels.

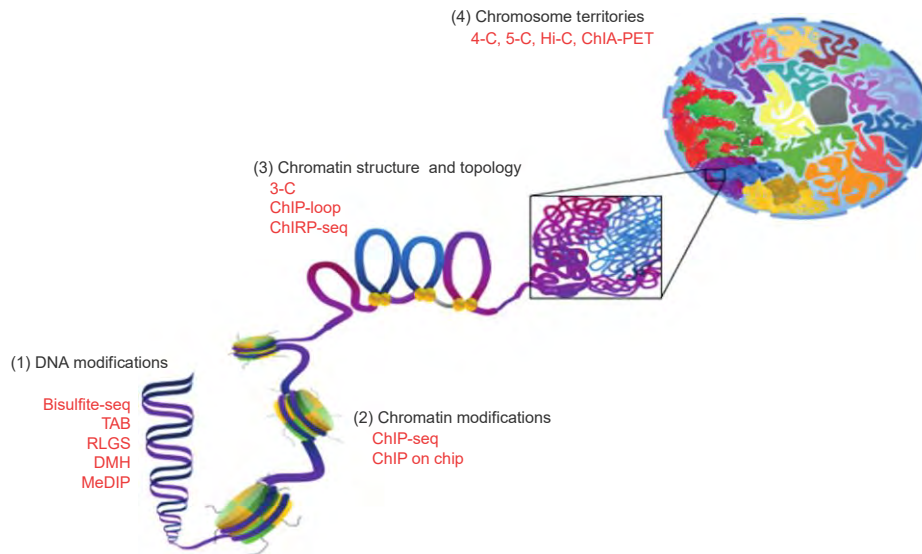


Figure 2.1: Overview of epigenetic modifications and different methods of analysis (1) Analysis of DNA covalent modifications, that includes and restriction analysis (RLGS, etc.), bisulfite sequencing and modifications for 5hmC analysis (TAB seq), and methylation immunoprecipitation (MeDIP) approach oxidation levels. (2) Chromatin modifications, which include histone post-translational modifications and other chromatin binding complexes. (3) Chromatin structure and topology, massive profile analysis of one vs one interactions of proteins associated to chromatin or ncRNAs that show 3D architecture of the chromatin regions and determine its transcriptional or structural function, methods commonly used for this are 3C, ChIP-loop and ChIP-seq. (4) Chromosome territories.

Also, the introduction of microarrays technology for methylation also became of importance to quickly establish epigenetic patterns for specific regions, like CpG islands located in promoter regions of selected genes. By combining this technology with epigenetic methods like sodium bisulfite, this method has helped the discovery of different roles of DNA methylation in the genome and its translation to applied approaches in biomedical research.

From the discovery of DNA structure to the publication of the Human Genome Project, massive genome sequencing approaches have advanced in a very important way. This was possible thanks to the discovery of next generation sequencing (NGS) methods. This technology is significantly cheaper, faster, more accurate and more reliable than the ones used in the past, thus allowing the development of clinical approaches (see Box 2.2) These approaches have also allowed the development of interdisciplinary projects such as The International Genome Sample Resource (IGSR; www.internationalgenome.org). This consortium represents the largest open collection of human variation data [23]. Most of current epigenomic data come from NGS, however, some data is still obtained by microarrays (ChIP on chip) design for specific regions, like gene regulators or CGI (Figure 2.2). Taken together, the application of different approaches to study DNA methylation has provided tools for researchers to understand its implication in many biological functions at a

Box 2.2: Scope and limitations of genomic experimental strategies

DNA microarrays: Developed in the 1980s, this genomic strategy is based on the hybridization of fluorophore labeled DNA to a solid surface. The number of DNA molecules bound to the surface can be determined by the intensity of the signal emitted by the fluorophore. The limitations of this technique are the number of probes printed in the array.

NanoString: Similar to microarrays or quantitative PCR, this technology is based on fluorophore-labeled probes that target a gene of interest. It provides high resolution, less than one copy per cell, and fidelity. The limitation of this assay is the low number of gene targets (around 800, which are below microarrays).

Short-read NGS: This methodology relies upon sequencing by ligation (SBL) or sequencing by synthesis (SBS). The SBL approach is based on a fluorophore-labeled probe that is hybridized to a DNA sample and ligated to an adjacent oligonucleotide, which together are used for image capture. The SBS identifies nucleotide addition by employing a polymerase and a signal such as a fluorophore or a change in ionic concentration. These two experimental strategies are based on solid surface DNA amplification.

Sequencing by synthesis (CRT): CRT assay employs a terminator molecule similar to those used in Sanger sequencing, where the ribose 3'-OH group is blocked, preventing elongation. The fluorophore and the blocking group can be removed before starting a new cycle.

Sequencing by synthesis (SNA, 454 Ion Torrent): This technology is based on the incorporation of a dNTP, which works as a single signal, into an elongating strand. In this technique, it is not necessary to employ blocking dNTPs. The 454 pyrosequencing was the first NGS technology.

Single-molecule long-read sequencing (PacBio and ONT): One of the most used platforms for long-read sequencing is the single-molecule real-time (SMRT) sequencing method developed by Pacific Biosciences (PacBio). This equipment employs a flow cell containing thousands of individual wells where the polymerases travel along the DNA template.

genome-wide level and obtain new insights on the mechanisms underlying genome regulation, and the implication of changes in methylation patterns in diseases.

2.2.3 Processing and Analysis Methods and Tools for DNA Methylation Data from Bisulfite Based Assays

2.2.3.1 Bisulfite Conversion

Bisulfite conversion is one of the most accepted methods to determine the methylation state in a genomic region, using different experimental strategies such as bisulfite

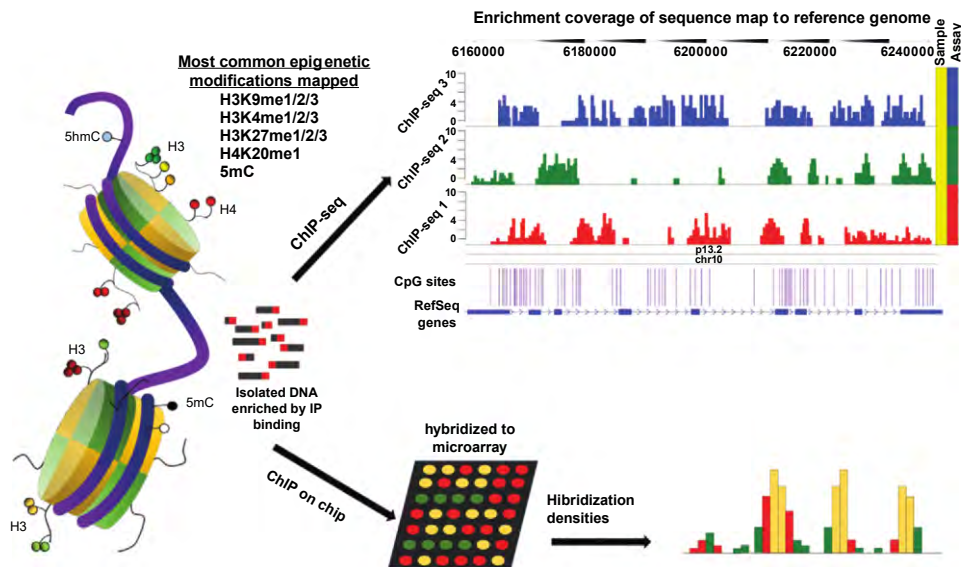


Figure 2.2: Epigenetic mapping performed by NGS and microarrays. Enrichment coverage example taken from WashU EpiGenome Browser viewer.

sequencing (BS-seq) and methylation sensitive PCR (MS-PCR). Moreover, bisulfite conversion is unable to differentiate 5mC from 5hmC, because both protect cytosine from its conversion to uracil [24]. New methods have been developed to distinguish and profile 5mC and 5hmC, using specific antibodies (MeDIP) or TET assisted bisulfite sequencing (TAB-seq).

In general, bisulfite conversion method generates a chemical conversion of unmethylated cytosines (C) to thymines (T). Therefore, the analysis aims for counting and mapping the number of C to T conversions and quantifying the proportion of methylation per base. We can carry this out by identifying C to T conversions in the aligned reads and dividing the number of Cs by the sum of Ts and Cs for each cytosine in the genome. However, bisulfite treatment has a 95% conversion efficiency rate, and base-calling quality is not constant and could change within the same read and between sequencing runs. Hence, bisulfite libraries are susceptible to errors and biases that could generate miscalled bases that could be counted as C–T conversions erroneously. To avoid this, attention must be made to initial quality control, trimming, and suitable alignment of bisulfite libraries.

Performing quality control of the data is important to avoid mis-mapping events and incorrect methylation calls. Therefore, it is important to check the base quality, which represents the level of confidence in the base calls, in order to avoid miscalled bases that can be counted as C–T conversion erroneously. Such basic checks can be performed via fastQC software (www.bioinformatics.babraham.ac.uk/projects/fastqc/). In addition, evaluating, and reducing potential sequence contamination or adapter contamination from the libraries is a common first approach. Base composition and CG-content plots will be useful for this purpose. It is common for BS-seq experiments performed in mammals, to observe an average of cytosine content around 1–2% of the

Table 2.2: Alignment methods for bisulfite based experiments

Aligner	Webpage	Reference
MethylCoder	https://github.com/brentp/methylcode .	[26]
BS-seeker2	http://pellegrini.mcdb.ucla.edu/BS_Seeker2/	[27]
Bison	http://sourceforge.net/projects/dna-bison/	[28]
BSMAP	https://sourceforge.net/projects/bsmapper/	[29]
VALiBS	https://github.com/wwwyxder/valibs	[30]

complete sequence length. Also, the CG content observed in BS-seq libraries peaks around 30%. This will certainly change among different tissues and cell types or species with different methylation rates. However, the rate of C and CG content helps to identify adapter contamination by observing spikes increase the occurrence of C at the later cycles, or an increase of the CG profile to more than 40%. Such errors can be fixed by trimming the sequence file. This can be achieved using trimming programs such as Trim Galore (www.bioinformatics.babraham.ac.uk/projects/trim-galore/).

After pre-alignment quality control and processing is done, the next step is the alignment. For this purpose, the BS-seq methods rely on modification of known short read alignment methods, and the conversion of the genome that is studied to an *in silico* bisulfite converted sequence. For example Bismark employs Bowtie (or Bowtie 2) and *in silico* C–T conversion of reads and genomes [25], some methods that use this strategy are cited in Table 2.2. After completing the alignment and methylation calling, there is no need for further quality control, and analysis of methylation proportions and differential methylation regions (DMRs) analysis can start. We will discuss differential methylation analysis later in the chapter.

2.2.3.2 Methylation Microarrays

As an alternative to the bisulfite-based mass sequencing, microarrays emerge as another promising method for the understanding of DNA methylation in cell biology and diseases. This experimental approach has decreased in costs and increased in coverage. The simplicity of the generation of specific panels for the analysis of discrete regions of the genome is also an advantage (for further discussion of microarray methods and applications see Chapter 1). There are different types of array-based assessments of methylation, but here we will focus on Illumina’s array assays from Illumina 27K and 450K platforms (see Box 2.3).

Illumina’s microarray technology has been adapted for the study of DNA methylation from sequences converted by bisulfite. Where the types of beads are quantified as the average signal representing methylated and non-methylated alleles, which are computed as β -values. The β -value is defined as the ratio of the methylated probe intensity to the overall intensity that summarized methylated and unmethylated probe intensities. Following the notation used by Illumina methylation assay [12], the β -value for an interrogated CpG site is defined as:

$$\beta = \frac{\text{Max}(M, 0)}{\text{Max}(M, 0) + (U, 0) + 100} \quad (2.1)$$

Box 2.3: Microarrays platforms and different probe types

Illumina methylation 27K platform contains 27,578 CpG loci that targets 14,000 genes. It covers around 2 CpG in regions within 1Kb upstream to 0.5 kb downstream of genes TSS. Biologically is a rather weak coverage of methylation of promoter regions. Although this platform was phased out in 2010, many datasets are held in public databases like Gene Expression Omnibus (GEO) that can be used for analysis and data mining. Human methylation 450K platform contains approximately 485,577 CPGs sites covering 99% RefSeq genes and 96% of CpG islands outside coding regions. These assays are configured with conjugated oligonucleotides in beads to measure specific target sequences. The Illumina 450 chip has two probe designs Infinium I and II. Infinium I targets each CpG with two beads labeled by the same dye. These probes contain from 0 to 10 CpG sites, designed either methylated or unmethylated to match the bisulfite converted sequence at the target sites. The type II probes employ one bead type labeled with two different dye colors for unmethylated (red) and methylated (green) CpGs. The Illumina 27K only has Infinium I probes, while the Illumina 450K platform contains both types of probes.

Here U and M represent the intensities measured in a specific region by the unmethylated and methylated probes, respectively. Any negative value will be reset to zero, in order to avoid negative values after background adjustment. A constant offset (100) is added to the denominator to regularize the β -value when both U and M probe intensities are low. A β -value statistic results in a number from 0 to 1, or 1 to 100%. Where a β -value of zero represents a site in which all copies were completely unmethylated, and a β -value of 1 indicates a CpG site that is fully methylated in all copies. Illumina has developed their Genome-Studio software [31], for basic data analysis and other purposes.

Quality control is essential to evaluate the data quality of the samples and to avoid miscalled and false positive errors. For this reason, Illumina's array includes several control probes to determine the data quality. In order to analyze the quality and to detect samples that behave poorly in the arrays, diagnostic plots of the control probes are generated using different tools, such as the Genome-Studio software [31], or the HumMethQCReport R-package [32]. Also, other packages available in R and Bioconductor, which are based on the use of P values can be employed for quality control measures; such as IMA, the pre-processing and analysis pipeline, Minfi [31, 33, 34, 35], available at <http://bioconductor.org/packages/2.12/bioc/vignettes/minfi/inst/doc/minfi.pdf> and www.bioconductor.org/packages/2.12/bioc/html/methylumi.html.

Note: Bisulfite conversion may present incomplete conversion, where some unmethylated Cs are not converted to Ts. Causing false positive results. For mammals and other species non-CpG methylation are almost not present in somatic tissues, hence, we can

calculate the conversion rate by using the percentage of non-CpG methylation. The closer the conversion rate is to 100%, the higher the quality of the experiment. Typical values for a good quality of bisulfite conversion will be higher than 99.5%. This will not apply in organisms such as plants, which commonly present non-CpG methylation [36, 37].

Microarray Normalization and Batch Effect Correction The purpose of normalization is to remove sources of experimental artefacts, random noise, and systematic technical biases introduced during the assay; in addition to biases between samples of different batches [38]. There are two different types of normalization: Between array and within array normalization. The first type removes technical artefacts between samples on different arrays; the second type corrects intensity related dye biases [39]. There are many normalization and data processing algorithms available for 450K arrays (Table 2.3). The peak adjustment is initially proposed to correct probe I and II bias. Therefore, many researchers use the *M*-value, which uses methylation *M*-value and rescales design II probe data to the peak positions of design I probes:

$$M = \log_2 \frac{\text{Max}(M, 0)}{\text{Max}(U, 0)}. \quad (2.2)$$

It is a normalization method within a sample, which is implemented in R package IMA. A normalized *M*-value near zero signifies a semimethylated locus, a positive *M*-value indicates more CpGs are methylated than unmethylated, whereas a negative *M*-value represents the opposite interpretation. However, because DNA methylation is not distributed in a balanced manner in the genome, a bias is generated in the distribution of the log-ratio of methylation. This bias is dependent on the levels of methylation found in a particular tissue. Such imbalance is created by the non-random distribution of CpG sites throughout the genome. For example CGI, which are the

Table 2.3: 450k array data processing algorithms/ methods

Method/Package	Utility	Reference
Genome Studio	Raw data processing, background control.	[31]
Subset quantile normalization	Probe I and II bias correction, color bias adjustment across sample normalization.	[33]
Subset quantile within normalization	Probe I and II bias correction, color bias adjustment.	[40]
Methylumi/ lumi	Raw data processing, background control, color bias adjustment across sample normalization.	[41]
Beta mixture quantile dilation	Probe I and II bias correction.	[42]
Illumina methylation analyser	Probe I and II bias correction, across sample normalization.	[34]

densest CpG regions are normally not methylated, whereas the opposite behavior is observed in the less dense CpG regions in human cells. It is also important to note that this is common behavior of normal cells. Likewise, this behavior in DNA methylation is reversed in diseases such as cancer. Because of this and other reasons, there is still a lack of consensus regarding which is the optimal approach for normalization of methylation data.

Differential Methylation Region Analysis Once processing of the data has been performed, the most common goal of DNA methylation profiling is to find differentially methylated CpGs (DMC) between two groups of samples. For this purpose, the average β -values from microarray and methylation ratio from BS-seq data are equivalent, thus the same statistic can be applied. An advantage of methylation sequencing against microarrays is the direct count of methylated and unmethylated cytosines, which is much more accurate than the signal intensity of the probes. This is because DNA methylation occurs in discrete regions of the genome, called differentially methylated regions (DMR), where such DMR clusters have associated specific functions in the regulation of the genome. The search for DMR helps us to understand the biological significance of the DNA methylation. For this reason, several algorithms that identify DMR have been developed for BS-seq and microarrays data. In summary, these methods need two steps: First they identify DMCs through Fisher or chi square test. Afterwards, these methods put together the CpGs with the similar statistics into a DMR according to the defined threshold of distance and DMC statistics. Some algorithms such as BSmooth conduct a smoothing step before DMC detection, which smooth out outlier CpGs and utilize CpGs with a low coverage. Some of the methods employed for DMC and DMR analysis are summarized in Table 2.4.

The study of DNA methylation profiles is a growing field in the understanding of diseases, whose implications is both in the understanding of genomic and cell regulation, as well as in the development of diseases. The methylation profiles serve to recognize specific characteristics of the disease as well as possible new treatment or biomarkers that can contribute to modern medicine. Many statistical methods can be used for DMC or DMR detection, however, they have performed very differently. More tests need to be developed and deployed for clinical use and better understanding of DNA methylation in a genome scale.

Table 2.4: Examples of DMC and DMR detection methods

Method/Package	Usage	Algorithm	Reference
Methylsig	DMC/ DMR	Beta binominal	[43]
MOABS	DMC/ DMR	Beta-binominal hierarchical model	[44]
Radmeth	DMC/ DMR	Beta-binominal	[45]
methylKit	DMR	Logistic regression	[46]
Bump hunting	DMR	Linear regression	[47]
BSmooth	DMR	Smooth t test	[48]
Biseq	DMR	Smooth/ beta regression	[49]

2.3 The Post-Translational Modifications of Histones

DNA is packaged in a complex with proteins and RNA known as chromatin. The fundamental unit of this complex is the nucleosome, that is composed of 165 base pairs (bp) rolled around an octamer of histones (H2A, H2B, H3, and H4, a couple from each) [50, 51]. Chromatin can be arranged in a more compact structure, that has a diameter of 30 nm, which is known as solenoid. This structure is stabilized by the union of the histone H1. At this level the formation of heterochromatin alters the expression of multiple genes [52].

The modifications in the chromatin architecture are regulated by different epigenetic mechanisms, including DNA methylation, non-coding RNA, ATP-dependent complexes, histone variants, and histone post-translational modifications. This latter happens mainly in the histone amino-terminal region and has distinct effects on chromatin state depending on the modification and the amino acid modified. They can induce relaxation and/or compaction of the chromatin, affecting the accessibility to the DNA sequence. Among the biochemical changes in histones are: ADP-ribosylation, lysine and arginine methylation, serine and threonine phosphorylation, ubiquitination, SUMOylation and lysine acetylation [50, 53, 54, 55]. These post-translational modifications act as a code, known as the “histone code,” that tells the genome of the cell which genes to repress or activate, whether promoters or enhancers are active or not, or whether a sequence is a structural region or parasitic sequence that must be compacted. Among the most studied activation marks are H3K4me3, H3K9ac and H3K27ac, H3K36me3. These are recognized by different elements associated with the genetic transcription. Moreover, histone marks such as H3K9me3 and H3K27me3 are known to recruit protein complexes associated with genetic repression and chromatin compaction [50, 53, 54, 55]. The most used technique to make profiles and mapping of how histones are found in the genome is the chromatin immunoprecipitation assay (ChIP). In the next subsection we will describe this method and how the analysis is performed and which computational tools are available.

2.3.1 Experimental Evaluation of Post-Translational Modifications of Histones

To evaluate post-translational modifications among the entire genome, the preferred method is chromatin immunoprecipitation (ChIP). A classic ChIP assay evaluates the presence of a protein or histone modification in a specific region of the genome. It starts from a sheared cross-linked chromatin (the complex of proteins bound to the DNA sequence) incubated with an antibody (Ab) that binds to a specific histone modification, so the Ab will only recognize its target epitope. Then, the Ab–chromatin complex is isolated from the rest of the chromatin and the DNA contained in the Ab–chromatin complex is purified [56]. Finally, using specific PCR primers, the target region is amplified by PCR [57].

Currently, ChIP experiments are coupled to next generation sequencing technologies (ChIP-seq) to evaluate the presence of a post-translational modification, not only in a specific gene, but along the entire genome. DNA fragments from ChIP-seq

are sequenced as reads, which are then mapped onto the reference genome, and the genomic regions that are significantly enriched for ChIP reads, compared with input reads, are detected as peaks [58, 59].

ChIP-seq experiments can be used to evaluate several targets, from DNA-bound proteins, post-translational modifications of histones, among others. Depending on the type of target, the bioinformatics approach will vary. Usually, there are three modes of protein-DNA interactions:

- **Sharp mode:** Certain proteins bind to specific sequences in the genome and produce a highly localized signal; for instance, transcription factors and cofactors.
- **Broad mode:** This mode of interaction with DNA is characterized by proteins associated with large genomic domains; for example, histone modifications involved in elongation of transcription or heterochromatin.
- **Mixed mode:** Includes proteins that with differential distribution along the genome, from sharp to broad mode depending on the genomic region, one example is RNA Pol II.

Depending on the mode of interaction between a protein and the DNA, some considerations are needed in the experimental design. For proteins displaying sharp signals, 10–14 millions of reads (M) are recommended; meanwhile, for broad signals a higher number of reads is suggested (20–40 M) [60].

Also, controls are important for a well-designed study. Each ChIP condition should have an input sample as control. The input sample consists of a sheared cross-linked chromatin obtained in the same conditions as the immunoprecipitated sample, but it is not incubated with any Ab. This control serves to correct biases during the data analysis.

2.3.2 ChIP-seq Data Analysis

A ChIP-seq experiment produces millions of reads. Once the raw data is obtained, an appropriate analysis is needed. Here, we describe the basic steps to analyze a ChIP-seq experiment [60, 61, 62, 63].

1. **Quality metrics:** Raw data obtained from sequencing platforms (FASTQ files) is evaluated to discard low quality reads. This step also includes trimming of barcodes and low-quality nucleotides. The preferred tool to evaluate read quality is FastQC [61, 64].
2. **Reads mapping:** Once the quality of the reads is evaluated, they are aligned to a reference genome using mapping tools, producing a BAM file. This step usually allows for a small number of mismatches in the alignment [65].
3. **Background evaluation:** Given that a ChIP-seq assay uses an Ab to recognize a specific protein–DNA complex (the signal), it can also bind non-specifically to a region (noise). Most of a ChIP-seq library is noise rather than signal (80–90% of reads). Thus, a method to establish a threshold to better identify signal peaks from background is needed [60].
4. **Peak calling:** A peak is a region of the genome where multiple reads have mapped. Peak calling tools have been developed to better identify true peaks.

Table 2.5: List of peak calling tools to analyse ChIP-seq data

Name	Webpage	Reference
MACS	https://github.com/taoliu/MACS/tree/macsv1	[66]
SICER	http://home.gwu.edu/approximatelywpeng/Software.htm	[67]
PeakSeq	https://github.com/gersteinlab/PeakSeq	[68]

Table 2.6: List of differential binding analysis tools for ChIP-seq data

Name	Webpage	Reference
ChIPComp	https://bioconductor.org/packages/release/bioc/html/ChIPComp.html	[69]
ChIPDiff	http://cmb.gis.a-star.edu.sg/ChIPSeq/paperChIPDiff.htm	[70]
DBChIP	http://bioconductor.org/packages/release/bioc/html/DBChIP.html	[71]
DESeq	http://bioconductor.org/packages/release/bioc/html/DESeq.html	[72]
diffReps	https://github.com/shenlab-sinai/diffreps	[73]
EdgeR	http://bioconductor.org/packages/release/bioc/html/edgeR.html	[74]
GFOLD	https://bitbucket.org/feeldead/gfold	[75]
JAMM	https://github.com/mahmoudibrahim/JAMM	[76]
MANorm	http://bcb.dfc.harvard.edu/~gcyuan/MANorm/MANorm.htm	[77]
ODIN	www.regulatory-genomics.org/odin-2/download-installation/	[78]
PePr	https://github.com/shawnzhangyx/PePr	[79]
RSEG	http://smithlabresearch.org/software/rseg/	[80]
THOR	www.regulatory-genomics.org/thor-2/download-installation/	[81]

There is no better method for peak calling and multiple tools have been developed. Some of these tools are shown in Table 2.5

5. **Differential binding analysis:** After obtaining peak sets, the identification of regions with differential protein binding patterns between conditions is important. For this end, normalization is performed [61]. Some of the differential binding analysis tools are listed in Table 2.6.
6. **Integrative analyses:** Further analysis can be performed by integrating ChIP-seq data to other type of experiments such as RNA-seq, Hi-C, and DNA methylation. Also, ChIP-seq data can be used to identify protein binding sites of transcription factors [61].

In summary, transcriptional regulation of a gene can be influenced by post-translational modifications of histones. ChIP-seq is an antibody-based technique to evaluate the presence of a protein or a histone modification along the entire genome. ChIP-seq experiments produce millions of sequences, named “reads,” which are further

analyzed to obtain enrichment peaks and which can be used to identify differential binding of proteins among conditions using different bioinformatics tools.

The ChIP-seq has opened a new window of opportunities in biomedicine. This allows a global perspective to search for novel regulators of transcription, which, viewed in a locus specific manner, would have been a challenge to predict. There are many examples of the contribution of this method to the knowledge of diseases, allowing the proposal of possible new targets for therapies for multiple diseases. Using the ChIP-seq approach, in solid tumors derived from breast cancer patients it was shown the differential estrogen receptor binding and H3K27me3 histone mark has been associated with poor clinical outcome [82].

2.4 Higher Order Chromatin Organization

Each cell in the human body contains the same DNA sequence. Different cell types in an organism are the result of different expression profiles that are presented in each of them. Where the position of these genes within the nucleus is will affect their transcription capacity. Therefore, one of the current efforts in epigenetics is to know how the tridimensionality of the nucleus influences gene regulation. In this regard, the eukaryotic genome is organized in a hierarchical and three-dimensional fashion. On the first level exist double stranded DNA, which is then packaged with histones to form nucleosomes, then nucleosome contacts form fibers or clutches, which can form dynamic long distance loops [83]. Some of them are established by architectural and regulatory proteins to give rise to structural landmarks, named domains [84, 85, 86]. Chromatin domains are comprised of clusters of genes with similar patterns of expression. For example, actively transcribed or inactive domains. The interaction among domains with similar characteristics form compartments, and the fusion of compartments in the same chromosome form chromosome territories [87].

In this context, the observation that chromosomes contain chromatin loops, led to the idea that these are separated by regulatory sequences. In this hypothesis, domains have well-defined borders marked by specific characteristics. These boundaries, also named insulators, represent one of the components that contribute to chromatin domain formation and maintenance of a specific configuration [88, 89].

Another component involved in the establishment of three-dimensional organization are enhancers. Enhancers are control sequences, that interact with promoters to regulate gene expression, they are usually located far from the promoter [90]. Transcription factors bind to enhancers or promoters to establish chromatin loops, which allow these distant regulatory sequences to be brought into close spatial proximity with its target gene [91]. It should be noted, that the activity of insulators or enhancers depends on the cell type and stage of development; meaning that cells from an adult kidney may have different patterns of enhancer/ insulator activity than brain cells from embryos.

Some experimental methods have been developed to study enhancer/ promoter interactions and the three-dimensional conformation of chromatin. The aim of this section is to show some of the most common experimental techniques used to study

higher-order chromatin organization and the bioinformatics methods to analyze these data.

2.4.1 Technologies to Study Chromatin Conformation

The study of long-range chromatin interactions has increased after the development of the chromosome conformation capture (3C) technique [92]. The development of 3C-derived techniques made possible a genome-wide study of chromatin conformation. These methods include 4C, 5C, chromatin interaction analysis with paired-end-tag sequencing (ChIA-PET), and Hi-C (genome-wide 3C).

The 3C and 4C techniques require choosing one locus and evaluating an interaction with another locus (3C) or genome wide (4C). The 5C method probes multiple loci genome wide, while Hi-C allows a genome-wide analysis of chromatin interactions. In sum, 3C can be seen as a one-vs-one experiment, 4C as one-vs-all, 5C as many-vs-many and Hi-C as all-vs-all. All these approaches produce a DNA fragment composed of the two interacting sequences, depending on its tri-dimensional localization within the nucleus. (See Box 2.3.)

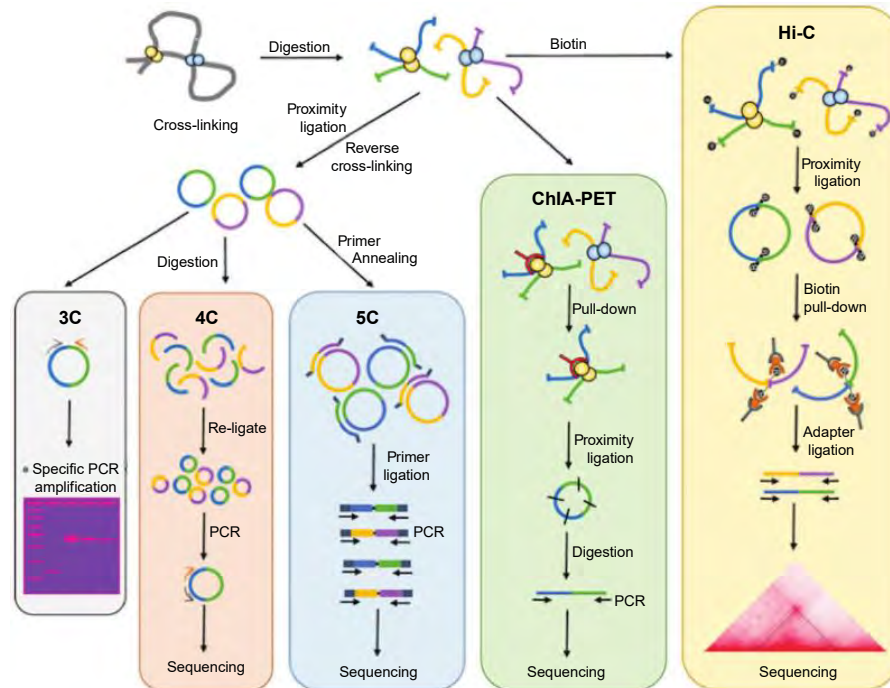
2.4.1.1 *The 3C, 4C, 5C, and ChIA-PET Technologies*

In a 3C experiment, the template is a pool of DNA fragments that reflect the interaction between two genomic loci and which can be detected by PCR using primer pair combinations [93, 94]. The abundance of each DNA fragment is a measurement of the frequency with which the two loci interact [95]. Thus, the 3C method can be used to capture and quantify only one physical interaction between a gene and a distant element, both in cis and in trans. On the contrary, a 4C experiment evaluates the DNA contacts made across the genome by a given genomic site of interest [96]. The 4C technique has been applied to demonstrate that individual gene loci can be engaged in many long-range DNA contacts with loci elsewhere on the same chromosome and on other chromosomes [97, 98]. The 5C method can be used to generate dense interaction maps that cover most or all potential interactions between all fragments of any genomic region. Dense interaction maps can provide a global overview of the conformation of a given genomic region [95]. Finally, ChIA-PET is a combination of 3C and ChIP that allows the genome-wide identification of potential interacting loci bound by a given protein [99].

2.4.1.2 *The Hi-C Technology*

A Hi-C experiment produces hundreds of millions of reads; these reads are analyzed and generate genome-wide maps showing interactions between genomic loci. A Hi-C experiment consists of these steps: (1) Cross-linking cells, (2) digestion of the DNA with a restriction enzyme, (3) mark the restriction sites with biotin, (4) ligation of the cross-linked fragments, (5) shearing the resulting DNA and pulling down the fragments with biotin, and (6) sequencing the pulled down fragments using paired-end reads. This procedure produces a genome-wide sequencing library that provides a proxy for measuring the three-dimensional distances among all possible locus pairs in the genome [100].

Box 2.4: 3C-based approaches to study chromatin conformation



Detecting DNA fragments that preferentially interact together based on their proximity in the 3D space are the basis of all chromosome conformation capture (3C) technologies, including Hi-C (a high-throughput derivative of 3C). The first step of most 3C-based methods involves cross-linking of cells to capture chromatin organization, followed by fragmentation of the chromatin by digestion with a restriction enzyme or by sonication. The digestion is pieced back together by proximity-ligation of adjacent DNA ends. After reverse cross-linking, different approaches can be used to identify the chromatin interactions. During 3C, 4C, and 5C, the ends are simply re-ligated to generate circular ligation products. For Hi-C, restriction ends are first marked with biotin to identify sites that were cut and re-ligated. During 3C, primers are designed upstream of specific cut sites to measure the frequency of given junctions by (q)PCR one at a time. In 4C, the 3C library is further digested with an enzyme that cuts more frequently and fragments are circularized by ligation. Primers designed against a region of interest are then used to simultaneously amplify all other fragments contacting it, and the products are sequenced. In the 5C approach, primer sequences overlapping restriction fragment ends are ligated only when the two ends are immediately adjacent, then products are amplified and sequenced. In the Hi-C method the restriction fragment ends are labeled using biotin, ligated products are enriched using streptavidin pull-down after and interactions are interrogated in a genome-wide all-versus-all manner. This method could potentially be combined with an enrichment step, either using sequence-specific probes or with an antibody against a protein of interest (ChIA-PET).

2.4.2 Bioinformatic Methods of Hi-C Analysis

A Hi-C experiment generates a huge amount of information, which needs to be analyzed to identify those chromatin interactions that have a functional impact on the cell from those “random” interactions, also known as non-functional. Thus, the bioinformatics methods available to analyze Hi-C data are mainly focused in the normalization and statistical analysis. Nevertheless, in the next subsection we briefly review the steps required in a typical Hi-C experiment.

2.4.3 Mapping and Filtering

The first step in the analysis of Hi-C data is the alignment of reads to a reference genome. There are four methods for alignment of reads obtained from a Hi-C experiment: (1) Pre-truncation, (2) iterative mapping, (3) allow split alignments, and (4) split if not mapped. For a more detailed explanation of the alignment methods see reference [101].

After the mapping step, a filtration of reads is needed. Usually, filters applied are: (1) Mismatches, (2) read quality, and (3) uniqueness of reads. Then, reads that passed the previous filters are classified to identify the “informative” ones. These filters are: (1) Strand filters and (2) distance filters [101, 102].

2.4.4 Normalization

The input data for normalization is a raw contact matrix. There are three normalization methods [101]: (1) Explicit-factor correction, (2) matrix balancing, and (3) joint correction.

- **Explicit-factor correction:** These methods need an a priori knowledge of bias factors (GC content, mappability, and fragment length) and corrects by modeling the probability of observing an interaction between two sequences [101].
- **Matrix balancing:** These methods correct for all bias causing factors without modeling them. They assume that in case of no bias, then loci would produce equal number of reads [101].
- **Joint correction:** These methods correct considering the one-dimensional distance between the two sequences, given that regions adjacent to each other in one dimension cannot be far way in the three-dimensional space, and this does not necessarily mean they form a functional loop [101].

Some of the tools that can be employed for Hi-C data normalization are shown in Table 2.7.

2.4.5 Statistical Analysis

Once biases are eliminated during normalization, the identification of statistically significant interactions between genomic regions is important. The number of interactions between two regions depends on the distance between them; there is random looping that needs to be evaluated when assigning statistical significance to the contacts. Many tools are available for Hi-C data analysis; some of them are enlisted in

Table 2.7: List of Hi-C data normalization tools

Name	Programming language	Website	Reference
Hi-Corrector	ANSI C	http:// zhoulab.usc.edu/ Hi- Corrector/	[103]
Hi-Five	Phyton	http:// taylorlab.org/ software/ hifive/	[104]
HiCNorm	R	www.people.fas.harvard.edu/ junliu/ HiCNorm/	[105]

Table 2.8. The methods developed to assess the significance of an interaction in a Hi-C experiment are described below [101].

2.4.6 Visualization of Hi-C Data

There are generally three ways of visualizing Hi-C data: (1) Square heat maps, (2) circular plots, and (3) genome browsers. To visualize large-scale data, such as complete chromosomes or the genome, a square heat map or a circular plot are the selected options. On the contrary, to evaluate local data corresponding to a specific region of the genome, a genome browser is the appropriate option [101, 102, 117, 118]. Browsers of Hi-C data are summarized in Table 2.9.

Square Heat Maps and Circular Plots Contact matrices are square heat maps where the contact count is transformed into colour. Every Hi-C heat map is a diagonal in the middle, showing adjacent loci and they can represent the entire genome or only one chromosome. Circular plots can be used to visualize intra-chromosomal interactions, where the contacts are show by arcs connecting distal loci [118].

Genome Browsers Genome browsers are useful for the inspection of a small region of the genome. Usually, the DNA sequence is shown horizontally and the Hi-C data is added in parallel to the DNA sequence. They allow the visualization of Hi-C data by local arc tracks, similar to circular plots [118].

2.4.7 Topological Associated Domain Identification from Hi-C Data

One of the purposes of epigenetics is the identification of regulatory domains, which have been characterized by specific histone marks and transcription factors. Topological associated domains (TADs) are densely interacting regions that can be identified in a contact map as interacting squares. TADs are of importance in the tri-dimensional organization of chromatin, thus many methods have been developed to study TADs from Hi-C experiments (Table 2.10).

The importance of data generated by experiments that capture chromatin interaction in the clinics is a growing topic in the field of epigenetics. One of the applications of techniques such as Hi-C is the prediction of the emergence of chromosomal

Table 2.8: List of Hi-C data analysis tools

Name	Programming language	Website	Reference
diffHic	R	https:// bioconductor.org/ packages/ release/ bioc/ html/ diffHic.html	[106]
Fit-Hi-C	R	https:// bioconductor.org/ packages/ release/ bioc/ html/ FitHiC.html	[107]
Gothic	R	https:// bioconductor.org/ packages/ release/ bioc/ html/ GOTHiC.html	[108]
HiC-Inspector*	R	https:// github.com/ HiC-inspector/ HiC-inspector	[109]
HiC-Pro*	Phyton, R	http:// github.com/ nservant/ HiC-Pro	[110]
HiCdat*	R	https:// github.com/ MWSchmid/ HiCdat	[111]
Hiclib*	Phyton	https:// bitbucket.org/ mirnylab/ hiclib	[112]
HiCUP*	Perl, R	www.bioinformatics.babraham.ac.uk/ projects/ hicup/	[113]
HIPPIE*	Phyton	wanglab.pcbi.upenn.edu/ hippie	[114]
HOMER	Perl	http:// homer.ucsd.edu/ homer/ interactions/	[115]
Juicer*	Java	https:// github.com/ theaidenlab/ juicer	[116]

*Tool that provides the complete workflow to analyze Hi-C data.

Table 2.9: List of genome browsers to visualize Hi-C data

Name	Webpage	Reference
3D Genome Browser	http:// promoter.bx.psu.edu/ hi-c/	[119]
Epigenome Browser	http:// epigenomegateway.wustl.edu/ browser/	[120]
HiBrowse	http:// hyperbrowser.uio.no/ 3d	[121]
Juicebox	www.aidenlab.org/ juicebox/	[122]
My5C	http:// my5c.umassmed.edu/ welcome/ welcome.php	[123]

Table 2.10: List of TAD identification tools from Hi-C experiments

Name	Website	Reference
DI-HMM	https:// github.com/ gcyuan/ diHMM	[70]
HiCseg	https:// cran.r-project.org/ web/ packages/ HiCseg/ index.html	[124]
TADbit	https:// github.com/ 3DGenomes/ TADbit	[125]
TADtree	http:// compbio.cs.brown.edu/ projects/ tadtree/	[126]

translocations. It is currently believed that for a translocation to occur, there must be proximity between two chromosomes at the chromosomal territory level, this is known as the “proximity effect.” Proximity effects have been described in analysis of cancer-causing translocations. An example widely studied is the one involving ABL and BCR genes. Genesis of the BCR–ABL hybrid gene appears to be a fundamental step in the development of chronic myeloid leukemia. These genes have been found in close proximity in normal hematopoietic cells and its Hi-C contact frequency is significantly higher compared with other loci in the same nucleus [127, 128, 129, 130, 131]. Suggesting the predisposition to generate relapses in these patients. Similar results were found in Burkitt’s lymphoma. In this case, the gene MYC (8q24) translocate in ~90% of the cases with IGH (14q21), and less frequently with IGL (22q11) and IGK (2p11). In such patients, locus proximity correlated to the observed incidence of translocation [132].

2.5 Long Non-Coding RNAs, Novel Molecular Regulators

Non-coding RNAs (ncRNAs) have been known for decades, with classic examples such as the ribosomal, transfer, small nucleolar RNAs, among others [133]. With the development of genome-wide sequencing and other high throughput technologies the presence of thousands of novel transcripts have been revealed and reported by the ENCODE consortium [133]. ncRNAs are now suggested to be functional products and not transcription noise. Of these novel transcripts, many are considered to be long non-coding RNA (lncRNA). However, we still do not know the function of the great majority of these transcripts. Although RNA polymerase II can generate spurious transcripts by non-specific binding, current evidence has shown that these ncRNAs are involved in many biological processes with specific expression patterns and regulation

[134, 135, 136, 137]. For this reason it is very important to elucidate the functions of this kind of non-coding transcripts in regulation of gene expression and diseases.

In general, lncRNAs are defined as transcripts longer than 200 base pairs, lacking open reading frames (ORF). The majority of lncRNAs have in addition the structure called CAP at the 5' end, the nonetheless, some exceptions have been observed. Likewise, as mRNAs, lncRNAs can be polyadenylated, suffer splicing, and may have several isoforms [138]. Expression analysis based RNA-seq performed on cell lines and different human tissues have suggested a higher amount of lncRNAs in the human genome than protein-coding genes [139]. The main features of lncRNAs are: (1) that they are synthesized in low scale, (2) they are located in both nucleus and cytoplasm, and (3) they are expressed in a tissue specific manner [137]. Depending on their genomic location and the sense in which they are expressed, lncRNAs can be divided into intergenic, intragenic, bidirectional, enhancers, as well as sense and antisense.

Experimental evidence has demonstrated that lncRNAs may be involved in different biological processes such as DNA methylation [140], genome imprinting [141], nuclear compartmentalization [142], gene to gene interactions [143], and regulation of chromatin structure [144], among others [145].

The first long non-coding RNA was described in 1988. It was identified as a transcript that was expressed during liver development in mice and it does not generate a protein product in vivo [146]. Actually, a high number of lncRNA have been identified through chromatin modifications associated with active promoters such as H3K4me3, and epigenetic modifications associated with transcription elongation as H3K36me3 [138]. Interestingly, many of these RNAs are transcribed relatively close to protein coding genes, occurring mostly in an antisense manner [147]. Moreover, genomic context does not necessarily explain the function or evolution of these genes [138, 148, 149]. Besides, in an attempt to understand lncRNAs and their cellular functions, they have been classified in four archetypes of molecular mechanisms: Signal, decoy, guide, and scaffold [150]. Despite what it is known about lncRNAs, there is a great debate on how to classify them [151, 145]

In summary, our knowledge of lncRNA is very limited and we are still unaware of the function of the great majority of these transcripts, therefore, it has increasingly become an exciting field of study in biology. Yet, their implications in biological mechanisms are leading to a high research interest which is additionally enhanced by the discovery of its possible implications in many diseases, such as cancer and maybe future therapeutic targets [152].

2.5.1 The Implications of lncRNAs in Precision Medicine

Conventional treatments are designed to treat sickness in terms of the average response of patients, assuming that one average approach fits all individuals. However, what is successful for some might not work for other patients. In order to improve the effectiveness of treatments, an emerging approach for disease treatment and prevention has gained attention, which is currently known as precision medicine. Precision medicine is based in the knowledge of the individual variability observed

in genes, epigenetic profiles, environment, and lifestyle for each person [153]. The concept is sustained by the assumption that genetic or molecular aberrations can be observed as a cause or contributor of a disease. Therefore, genomic studies and related information can shed light on various questions associated with the health and disease of an individual. Genomic approaches such as DNA sequence variation, expression profiles, proteomics, and metabolomics have become valuable tools for precise disease management and prediction. The information from an individual genome sequence and the associated expressed biomarker are essential to establish risk checkpoints and achieve precision therapies [154].

Actually, numerous reports have demonstrated the use of genomic data and expression signatures as clinical prognostic factors in cancer and other diseases. Most of these studies are focused on coding genes and regulatory regions to identify different types of biomarkers and in some cases establish companion diagnostics that will positively impact the outcome for patients [155]. Diverse reports have observed aberrant expression patterns of lncRNAs associated with numerous diseases including cardiovascular disease, Alzheimer', fragile X syndrome, and cancer. Therefore, lncRNAs have become novel targets with potential for therapeutics, prognoses, and diagnostics of diseases [156].

In order to understand the implications and the biology of these transcripts, several computational tools have been generated that help in their study. In the next section, we will discuss some of these tools, as well as the analysis strategies that can be used.

2.5.2 Bioinformatic Tools for lncRNAs Analysis

The study of the transcriptome represents a new challenge for the modern medical area, since it breaks with the traditional analysis of only some coding genes [157]. In a conceptual manner, the transcriptome is defined as the set of total transcripts present in a cell, at a particular time in its life cycle [158, 159]. The current tools of NGS allow us to obtain information about the transcriptome of a specific sample, and in consequence we can know the expression levels of all the transcripts. This implies that the amount of data obtained is too large and requires sophisticated bioinformatics and statistical tools for processing. Despite this, advances in bioinformatics have developed various *in silico* methodologies [160, 161].

On the other hand, NGS methods such as RNA-seq, show some advantages over methodologies such as expression microarrays. In the sense that RNA-seq scans the whole transcriptome and is not limited to the detection of sequences or transcripts that are already known, therefore, this tool can be used to discover new isoforms, gene fusions, and unknown lncRNAs. Thus, these observations have shown the great potential of RNA-seq for the search of new transcripts that can be of coding and non-coding nature [162].

Once the expression information of an RNA-seq has been obtained, it is important to be clear about what you want to know or search. In this sense, there are two main ways to analyze the results obtained: (2) Work with already known annotated transcripts [163], or look for new transcripts that have not been characterized or annotated

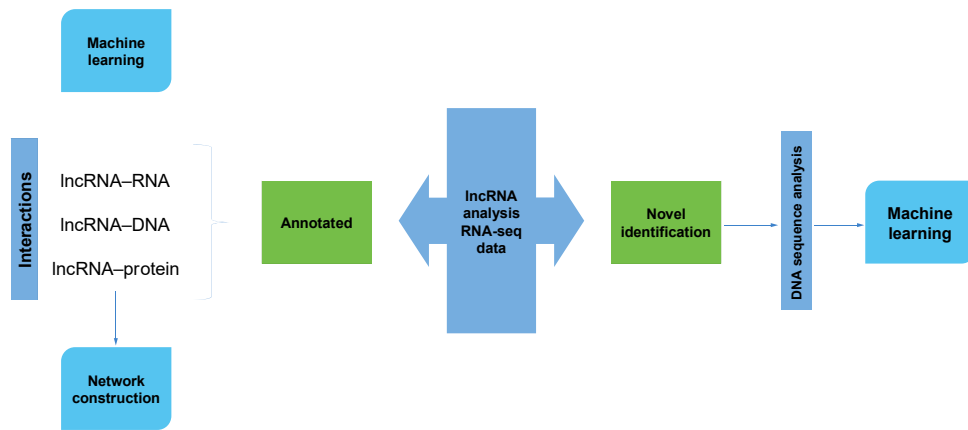


Figure 2.3: New methodological approaches for the bioinformatic analysis of lncRNAs. From RNA-seq data, it is possible to identify the expression of already annotated transcripts, in order to perform interaction analysis among some transcripts of interest, and determine the possible functions of the RNAs of interest, or their relationship with a specific cellular process.

[163, 164, 165] (Figure 2.3). In the specific case of lncRNA, they are not fully characterized, so the field of work in annotated and non-annotated transcripts is extensive [163].

2.5.3 Analysis of Annotated lncRNAs

When it is desired to work with lncRNA that have already been previously annotated, there are different strategies for the analysis of these, all bioinformatics analysis of them begins with the alignment of the transcriptome [165]. For this, it is important to know that the reference files for the alignment will depend on each source. GENCODE is a source that has alignment files for the complete genome, or only the subset belonging to the lncRNA can be used [166]. Ensembl, on the other hand, has coding and non-coding transcripts in independent files [167]. The databases also provide, in addition to these files, information of a structural, functional, or expression type (see Table 2.11).

Once the alignment of the reads has been done, the expression quantification process is carried out, those analyses can be analyzed by using fragments per kilobase of transcript per million mapped reads (FPKM), reads per kilobase of transcript per million mapped reads (RPKM), or transcripts per kilobase million (TPM). Once transcript quantification is done a differential expression analysis is desirable, because it is used to determine which will be the lncRNA of interest to study. These analyses allow us to compare the levels of expression between different conditions (normal vs disease, experimental vs control, stages of development, etc.) [178]. An example could be the determination of the level expression of a transcript under normal and cancer conditions, in order to determine which transcripts are up- or down-regulated. All of the above analysis can be carried out with packages such as R, or with the use of tools in Galaxy, depending on the user's experience [178, 179, 180].

Table 2.11: LncRNAs and databases

Database	Description	Disease-association data	Number of lncRNA	Species	Reference
LNCipedia 5.0	A database for annotated human lncRNA transcripts sequences, structures, and available literature.	No	120,353	Human (hg37/ hg19 Hg38)	[168]
NONCODEV5.0	A database for annotated ncRNA (except tRNAs and rRNAs).	Yes	548,640	Multiple (17 species)	[169]
lncRNAator	A database for annotated lncRNA sequences, expression profiles, protein interactions, and phylogenetic conservation.	No	31,725	Multiple (6 species)	[170]
TANRIC	An interactive platform for research and exploration of lncRNAs in cancer.	Yes	~13,000	Human	[171]
ChIPBase v2.0	An integrative database of transcription regulation and expression profiles of lncRNAs.	Yes	38,293	Multiple (10 species)	[172]
lnc2Cancer	A curated database for cancer-associated lncRNAs.	Yes	666	Human	[173]
lncRNADisease	A database for lncRNA and disease association.	Yes	1,564	Human	[174]
lncRNome	A database for human lncRNA transcripts.	No	~170,000	Human	[175]
lncRNADB v2.0	A database for eukaryotic lncRNA structural information, gene expression data, genomic context, and phylogenetic conservation.	No	>200	Multiple (68 species)	[176]
EVLncRNAs	An integrative and curated database for annotated and experimentally validated human lncRNAs.	Yes	1,543	Multiple (77 species)	[177]

From the results of quantification and differential expression analysis, other types of analysis can be performed, which allow us to understand relationships between the expression of the transcripts and the conditions analyzed [181]. In this sense, there are three main types of analysis: Those that relate the expression of lncRNA with a particular pathology; those that associate the levels of lncRNA with the expression of proteins; or those that associate the levels of lncRNA with the expression of other non-coding transcripts such as miRNA [182, 183, 184]. This is called correlation studies, and they are based mainly on the Pearson correlation coefficient, which is calculated with operations between matrices [184]. The result of this process is the construction of a network that allows analyzing the relationship between the species and/ or the study condition [185].

Furthermore, to carry out a deeper functional study about lncRNA functions, other types of tools capable of predicting the interaction between molecules have been developed. In this case, we can determine the possible interaction between lncRNA–ncRNA–miRNA [184, 186, 187], lncRNA–protein. [188, 189] or lncRNA–DNA [190] (Figure 2.4). Basically, the reason why those algorithms were designed is to analyze the binding sites that each of the biological molecules has, and the probability that the molecules of interest can interact with each other. These tools use for their construction, information that has been previously corroborated experimentally, or that is available in the literature (text mining), but it represents a limitation of analysis [188, 189]. However, there are reports of tools capable to thermodynamically predict some interactions between nucleic acids molecules (ncRNA [187] and DNA [190]), thus the complementary use of these tools can be a way to resolve the lack of reported information on interactions between nucleic acids.

As a consequence, all of the above criteria lead us to obtain much more complete information about the possible function of each lncRNA, or even the cellular processes in which it participates. For example, if an lncRNA is able to interact with chromatin remodeling complex, or transcription factors, among others regulatory elements, with this information we can get a general landscape of the lncRNA function. Actually, there exist two main methodologies for functional analysis of lncRNAs: Networks construction [183, 184, 188], and machine learning [191]. The algorithms designed by machine learning have taken on a lot of importance, since they provide reliable results, and they are capable of working with much more information. Currently, there are groups working on the improvement of network construction algorithms. The algorithms differ from each other by automation and training, in addition to being conditioned by the inputs accepted by each of the matrices that are used. In this sense, there is a lot of literature that describes in detail each of the algorithms and the matrices used for each of them.

By working on the transcripts already annotated, it can be clearly seen that the focus of these analyses is guided to be the functional characterization of each of them (Figure 2.5). In Epigenetics, it is important to describe the role that each lncRNAs play as epigenetic regulators and describe the pathways by which they could act. In this sense, the tools described above allow us to measure the relationship that exists between the expression of some proteins that participate in epigenetic processes, or the association with other non-coding transcripts and even begin to discern the possible mechanisms of action. Thus, the lncRNAs display a more tissue-specific expression

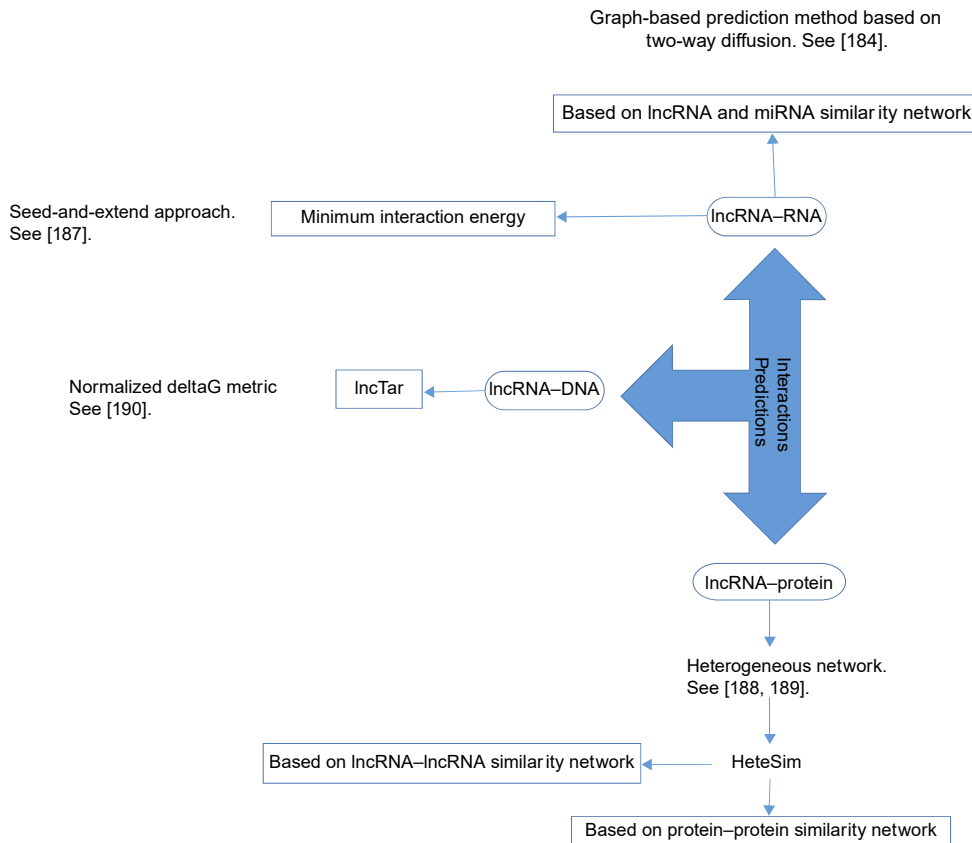


Figure 2.4: Algorithms designed to predict RNA interactions. Examples of algorithms used to predict lncRNA interactions with different types of RNAs, its can be predict interactions with DNA and proteins too. Among them, the figure shows the algorithms based on the construction of networks (HeteSim) and in the calculation of some parameters that characterize the physical interactions, such as the minimum interaction energy.

pattern even in pathological conditions or stages of development, for this reason this kind of transcripts can be useful indicators of disease status in patients this makes them attractive molecules for the development of new algorithms that allow us to develop new tools to determine which of the diseases are specific to diseases and can be used to modern therapies [192].

2.5.4 Analysis of Unannotated lncRNAs

The bioinformatics tools to discover new transcripts have faced several challenges along their development, among them are the inherent characteristics of the lncRNA, such as the little evolutionary conservation, and splicing sites that differ with the sites of canonic splicing sites of mRNAs [164, 165]. Despite this, the first tools designed based their pipeline on the analysis of mRNAs. Currently, there are tools and pipelines

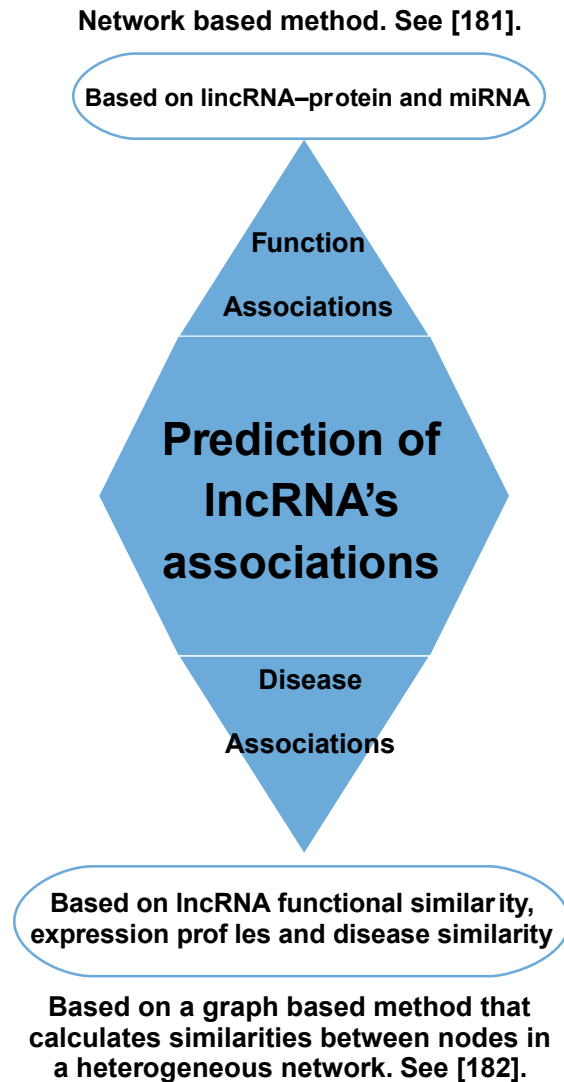


Figure 2.5: Algorithms used to predict the association of lincRNAs with diseases. Examples of algorithms used to predict the association of lincRNAs based on diseases. Some of these methodologies depend strictly on the data already reported in the literature, as is the case of the associations of lincRNAs with diseases. However, the characteristic shared by many of the pipelines is that they are based on previous networks of similarity of the lincRNAs in terms of function, structure, expression profile, or association to diseases. Another important input that these methodologies receive is the information coming from the co-expression studies.

designed for the discovery of new lincRNA transcripts, mainly based on the identification of splicing sites, genomic sequences, and de novo assemblies [164]. In this sense, the implementation of methods based on machine learning have been proven to have better results than previously designed tools [164]. lincRNAs are transcripts difficult to

characterize for their low abundance, for this reason more efforts need to be made in this area in order to get the all compendium of non-coding transcripts that form part of the human transcriptomes.

2.6 Epigenetic Databases

Here we provide a brief overview of the most relevant projects, consortia, and databases containing large-scale epigenetic datasets, which have been used to provide novel insights into a wide range of disease-focused studies.

2.6.1 Encyclopedia of DNA Elements in the Human Genome

The Encyclopedia of DNA elements (ENCODE, www.encodeproject.org) is an international collaborative effort funded by the National Human Genome Research Institute (NHGRI) in the US. Its main objective is to map and characterize all the functional and regulatory regions within the genome [193]. Though initially intended for humans, the ENCODE project has expanded to include other model organisms such as mouse [194], fruit fly [195], and worm [196].

The current version of ENCODE (as of February 2017) contains information of more than 6,500 samples which span different types of immortalized cell lines, tissues, and stem cells. Types of epigenetic information included are DNA methylation (MeDIP-seq), histone mark enrichment (ChIP-seq), open chromatin regions (DNase-seq), and non-coding transcripts (RNA-seq).

The rich datasets in ENCODE have already helped us to understand the mechanisms of disease development. For example, studies using ENCODE data have shown that 90% of variants linked to disease risk fall outside protein-coding regions in the DNA.

2.6.2 The Roadmap Epigenomics Project

The aim of the NIH Roadmap Epigenomics project (www.roadmapepigenomics.org/) is to produce and make publicly available human epigenomic data from various stem cell types and ex-vivo tissues [197]. Their vision is to provide a genome-wide landscape of epigenetic variation that can take place during normal human development and disease. Also, a major aim is to provide reference epigenomes, to which Version 9 of the Roadmap currently includes 2,804 genome-wide datasets of which 1,821 are histone modifications datasets, 360 DNase datasets, 277 DNA methylation, and 166 RNA-seq datasets. Release 9 also includes 111 reference epigenomes, each containing a set of five core histone modifications (H3K4me3, H3K4me1, H3K27me3, H3K9me3, and H3K36me3).

Analysis of the Roadmap data has led to elucidation on epigenetic mechanisms behind Alzheimer's disease and non-coding SNPs in autoimmune diseases [4, 198]. Another study revealed that the mutation density in cancer genomes can be predicted

by epigenetic features such as chromatin accessibility and modification, derived from the most likely cell of origin of the corresponding tumor [199].

2.6.3 Functional Annotation of the Mammalian Genome

The Functional Annotation of the Mammalian Genome (FANTOM, <http://fantom.gsc.riken.jp>) consortium is an international research effort that attempts to annotate and understand all the elements encompassing the complex transcriptional regulatory network of the mammalian genome. In phase 5 (FANTOM5) of the project, by using capped analysis of gene expression (CAGE) followed by next generation sequencing [200, 201], researchers of FANTOM systematically profiled the sets of transcripts, transcription factors, promoters, and enhancers active in the majority of mammalian primary cell types and a series of cancer cell lines, and tissues.

Datasets of FANTOM5 have led to important discoveries, such as that enhancers are the first to be transcribed, followed by transcription factors and finally non-transcription factor genes [202]. Also, it was shown that enhancers produce bidirectional, exosome-sensitive unspliced RNAs, which so far are the strongest predictors of enhancer activity [203].

FANTOM5 data has been used to identify new pan-cancer biomarkers by comparing the transcriptional profiles with RNA-seq datasets from The Cancer Genome Atlas [204, 205]. Enhancer RNAs were shown to be upregulated in cancer, where also promoters overlap with repetitive elements that are also upregulated in cancer.

2.6.4 BLUEPRINT Epigenome

Project BLUEPRINT is a European funded effort with the objective to obtain and make available epigenomes from all types of blood cells [206]. Samples in their data repository include primary cells from healthy patients, as well as patients suffering from blood-bases diseases such as type I diabetes and hematopoietic neoplasia and autoimmune diseases among others. Types of assays include DNA methylation (WGBS), histone modifications (ChIP-seq), chromatin accessibility (DNase-seq), and RNA expression (RNA-seq).

2.6.5 The International Human Epigenome Consortium

The International Human Epigenome Consortium (IHEC, <http://ihc-epigenomes.org/>) is an organization that coordinates the productions of reference epigenomes for multiple cell types relevant to health and disease research [206]. Their goal is to map 1,000 epigenomes by 2020. The IHEC data portal collects epigenetic information totaling 8,753 (February 2017) datasets produced by all its members, these include ENCODE, NIH Roadmap [207], the European Union BLUEPRINT Project, the Canadian Epigenetics Environment, and Health Research Consortium (CEEHRC), and the German Epigenome Programme DEEP, among others (Table 2.12).

Table 2.12: A non-exhaustive list of epigenetic focused databases

Name	Type of data	Platforms	Description
MethylomeDB	DNA methylation	Methyl-MAPS	Provides genome-wide methylation profiles for human and mouse brains.
NGSmethDB	DNA methylation	Whole genome bisulfite sequencing	Contains methylation profiles for different human cell lines, primary tissues, pathological biopsies and autopsies curated mainly from NCBI GEO [20] and the Roadmap project.
MethBase	DNA methylation	Whole genome bisulfite sequencing and reduced representation bisulfite sequencing	Contains hundreds of annotated methylomes for human, mouse and other model organisms.
4DGenome	Chromatin interactions	3C, 4C, 5C, Hi-C, Capture-C, ChIA-PET, IM-Pet	Contains thousands of chromatin interactions for human and other model organisms.
EpiFactors	Genes and gene products corresponding to epigenetic factors.	N/ A	Contains information about genes and gene products that code into epigenetic factors such as histones, histone modifiers, DNA methyltransferases, etc.
DNAMod	DNA chemical modifications	N/ A	Contains information about all types of covalent DNA modifications such as 5mC and 5-hydroxymethylcytosine and several others, found in many model species.
NONCODE	Long non-coding RNAs	N/ A	Collects and annotates information about lncRNAs from human and other model organisms. Provides information about conservation and the relationships between lncRNAs and diseases

2.7 Conclusion and Final Remarks

The epigenome is an important mechanism that regulates which parts of the genome are condensed and inactive to transcription and which are accessible and therefore transcriptionally active. Hence, epigenetic modifications are a major driver of development and determine and maintain cellular identity. The epigenome suffers different aberrations in diseases, where such aberrations have been related to many biological alterations related to the aetiology and the clinical outcome for patients. Different fundamental epigenetic mechanisms that regulate gene expression: Including DNA methylation, histone modifications, chromatin remodeling proteins, associated proteins, lncRNAs, and master regulators of the chromatin have been observed in association with diseases. Here we summarized some of the different epigenetic components and their importance in different pathologies, and present an overview of profiling methods and technologies that have rapidly matured in the last decade with examples of some of the epigenetic databases that exist.

There is currently a worldwide effort to understand and determine the precise human epigenome in healthy and diseased tissues, in order to help establish profiles that predict disease outcome in terms of patient prognosis and treatment response and also the possibility of epigenetic specific therapies, which could improve the life expectancy of patients. The comparison of epigenetic profiles and the integration of data from different studies from healthy and diseased tissue will allow the generation of predictive models for disease outcomes using epigenetic biomarkers and signatures. However, many key questions in the field remain unanswered, regarding the rules that command the epigenetic code, and the functions of the epigenetic components in the regulation process. Do we know all the players in epigenetics? It is clear that further effort in research towards completing the epigenetic map are still needed, therefore detailed epigenomic analysis are needed to create the maps in healthy and diseased tissues. The advances in this field will be of great importance in promising research of epigenetic therapies.

2.8 Exercises

2.1 How is epigenetics defined and why is it important in cellular mechanisms?

2.2 What is the importance of epigenetics in medicine?

2.3 Select the methodologies for DNA methylation analysis.

- LUMA
- ChIP
- RT-PCR
- MeDIP
- MS-PCR
- Western Blot
- Bisulfite conversion

- **PUBLICACIÓN 5:** Histamine Modulates Midbrain Dopamine Neuron Differentiation Through the Regulation of Epigenetic Marks..
- **AUTORES:** Vargas-Romero F, González-Barrios R, Guerra-Calderas L, Escobedo-Avila I, Cortés-Pérez D, López-Ornelas A, Rocha L, Soto-Reyes E & Velasco I.
- **REVISTA:** Frontiers in Cellular Neuroscience. **VOLUMEN:** 13. **PÁGINAS:** 215. **AÑO:** 2019.



Histamine Modulates Midbrain Dopamine Neuron Differentiation Through the Regulation of Epigenetic Marks

Fernanda Vargas-Romero^{1†}, Rodrigo González-Barrios^{2†}, Lissania Guerra-Calderas^{2†}, Itzel Escobedo-Avila^{1,4}, Daniel Cortés-Pérez^{1,4}, Adolfo López-Ornelas^{1,4}, Luisa Rocha⁵, Ernesto Soto-Reyes^{3*} and Iván Velasco^{1,4*}

¹Instituto de Fisiología Celular – Neurociencias, Universidad Nacional Autónoma de México, Mexico City, Mexico, ²Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología-Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico, ³Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana, Unidad Cuajimalpa, Mexico City, Mexico, ⁴Laboratorio de Reprogramación Celular, Instituto Nacional de Neurología y Neurocirugía “Manuel Velasco Suárez” – Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Mexico, ⁵Departamento de Farmacobiología, Centro de Investigación y de Estudios Avanzados (Cinvestav), Mexico City, Mexico

OPEN ACCESS

Edited by:

Tommaso Pizzorusso,
University of Florence, Italy

Reviewed by:

Maria Beatrice Passani,
University of Florence, Italy
Yuriko Iwakura,
Niigata University, Japan

*Correspondence:

Ernesto Soto-Reyes
esotoreyes@correo.cua.uam.mx
Iván Velasco
ivelasco@ifc.unam.mx

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Cellular Neurophysiology,
a section of the journal
Frontiers in Cellular Neuroscience

Received: 13 February 2019

Accepted: 29 April 2019

Published: 21 May 2019

Citation:

Vargas-Romero F,
González-Barrios R,
Guerra-Calderas L, Escobedo-Avila I,
Cortés-Pérez D, López-Ornelas A,
Rocha L, Soto-Reyes E and Velasco I
(2019) Histamine Modulates Midbrain
Dopamine Neuron Differentiation
Through the Regulation of Epigenetic
Marks. *Front. Cell. Neurosci.* 13:215.
doi: 10.3389/fncel.2019.00215

During midbrain development, dopamine neuron differentiation occurs before birth. Epigenetic processes such as DNA methylation and demethylation as well as post-translational modification of histones occur during neurogenesis. Here, we administered histamine (HA) into the brain of E12 embryos *in vivo* and observed significant lower immunoreactivity of Lmx1a+ and Tyrosine Hydroxylase (TH)+ cells, with parallel decreases in the expression of early (*Lmx1a*, *Msx1*) and late (*Th*) midbrain dopaminergic (mDA) genes. With MeDIP assays we found that HA decreases the percentage of 5-methylcytosine of *Pitx3* and *Th*, without changes in 5-hydroxymethylcytosine. Additionally, HA treatment caused a significant increase in the repressive epigenetic modifications H3K9me3 in *Pitx3* and *Th*, and also more H3K27me3 marks in *Th*. Furthermore, HA has a long-term effect on the formation of the nigrostriatal and mesolimbic/mesocortical pathways, since it causes a significant decrease in midbrain TH immunoreactivity, as well as alterations in dopaminergic neuronal fibers, and significant lower TH-positive area in the forebrain in whole-mount stainings. These findings suggest that HA diminishes dopaminergic gene transcription by altering several epigenetic components related to DNA and histone modifications, which affects mDA neuron progression during development.

Keywords: midbrain development, dopamine neuron differentiation, epigenetic modifications, neuroepigenetics, ultrasound-guided injection in utero

INTRODUCTION

Midbrain dopaminergic (mDA) neurons are essential for motor function control as well as for the regulation of reward and emotions (van Heesbeen et al., 2013). Development of mDA neurons can be divided into three stages: neural plate regionalization, midbrain cell fate determination and terminal differentiation to mDA neurons (Gale and Li, 2008). Each step presents activation of

different cascades of transcriptional factors triggered by extrinsic signals (Ang, 2006; Abeliovich and Hammond, 2007; Huang et al., 2015). During the specification stage, dopaminergic neural progenitors express specific genes such as *Lmx1a/b*, *Foxa2*, *Msx1/2*, and *Ng2* (Ang, 2006; Gale and Li, 2008; Nakatani et al., 2010; Yan et al., 2011). At this stage neural progenitors acquire mDA neuronal fate and start the last stage to become post-mitotic neurons (Yi et al., 2014). The activation of *Nurr1* and *Pitx3* allows the expression of late-differentiation genes associated to the dopaminergic neuron phenotype, as they regulate the expression of enzymes such as Tyrosine Hydroxylase (TH), the rate limiting enzyme in the production of dopamine (Gale and Li, 2008; Jacobs et al., 2009; Kadkhodaei et al., 2009). Terminally differentiated mDA neurons extend their axons to reach the *striatum* (ST) and release dopamine to regulate motor function in the basal ganglia (Hegarty et al., 2013).

Histamine (HA) is a neurotransmitter/neuromodulator that participates in the sleep/awake cycle, motor activity, cognition, feeding and energy balance in the adult brain (Panula and Nuutinen, 2013). During rat embryonic development, the cerebral concentration of HA reaches higher levels than those observed in the adult; its highest concentration is found at embryonic days (E) 14–16 (Vanhala et al., 1994; Nissinen and Panula, 1995; Molina-Hernández et al., 2012). This amine has a cell proliferation and neuronal differentiation effect on cortical neural stem/progenitor cell (NSPC) cultures, through multiple effects, including: elevation of intracellular Ca^{2+} , up-regulation of FGF receptors and increased expression of proneural genes (Molina-Hernández and Velasco, 2008; Rodríguez-Martínez et al., 2012; Molina-Hernández et al., 2013). Similar effects of HA, have been observed on adult NSPC at the subventricular zone (Bernardino et al., 2012). In addition, the *in vivo* effects of HA injected at E12 include a decrease in NSPC proliferation in the ventricular zone and also a decrease in TH staining in the midbrain, without affecting serotonergic nor GABAergic neurons; co-injection of HA with the H_1R antagonist chlorpheniramine prevented the decrease on TH induced by HA (Escobedo-Avila et al., 2014).

Recent studies describe different pathways to regulate neural stem cell differentiation via regulation of epigenetic modifications, which result in changes on gene expression. DNA demethylation has been related to the brain development control (Hahn et al., 2013; Wheldon et al., 2014). Particularly, the increase of 5-hydroxymethylcytosine (5hmC) along gene bodies has been associated with transcriptional activation during neuronal differentiation (Hahn et al., 2013; Kim et al., 2014). Furthermore, abundance of 5-methylcytosine (5mC) on gene body regions has been related to up-regulation of pre-mRNA splicing (Shukla et al., 2011; Guo et al., 2014; Yearim et al., 2015). These and other epigenetic modifications can be induced by environmental stimuli (Aguilera et al., 2010; van Heesbeen et al., 2013). For example, long-term exposure of cultured ventral midbrain (VM) NSPC to depolarizing potassium concentrations promotes mDA neuron differentiation, increasing binding of *Nurr1* to the regulatory regions, and higher transcriptional activity, of *Th* and *Dat* genes, with parallel increases in H3K4me3 and decreased H3K9me3/H3K27me3,

although the latter changes observed only in the *Th* promoter (He et al., 2011). Furthermore, VM NSPC treated with the neuropeptide urocortin increases mDA differentiation by inhibition of Histone Deacetylase 1 (HDAC1), resulting in hyperacetylation of histone H3, which allows the binding of *Nurr1* to the upstream *Th* regulatory regions (Huang et al., 2015). This suggests that post-translational histone modifications participate in the regulation of the expression of genes involved in brain development.

In this work, we aimed to evaluate the effect of HA on several epigenetic marks on intragenic regions that regulate expression of genes involved in mDA development *in utero*, as well as its long-term effects *in vivo* before birth. We found an enrichment of repressive histone marks, H3K9me3 and H3K27me3, together with a decrease of 5mC at intergenic regulatory regions. Such modifications were associated to significantly lower mRNA levels of genes related to midbrain DA neuron fate specification. Injection of HA at E12 caused a decrease in the number of mDA neurons and modified the trajectory of its axons to the ST at late stages of midbrain development.

MATERIALS AND METHODS

Ultrasound-Guided Injections

All animal procedures were approved by the Institute of Cellular Physiology's Animal Care and Use Committee and conformed to National guidelines (*Norma Oficial Mexicana NOM-062-ZOO-1999*). We used timed-pregnant Wistar rats with embryos at gestational age E12 to perform ultrasound-guided injections as previously reported (Escobedo-Avila et al., 2014). Control (vehicle-injected) or experimental (HA-injected) conditions were performed by triplicate or quadruplicate in different embryos from the same rat, and repeated with six different pregnant dams. Each group was evaluated in a single uterine horn. Rats were deeply anesthetized in a chamber and then maintained with a facemask to administer inhaled sevoflurane (5% for induction and 1% for maintenance). We performed a midline laparotomy, the uterine horns were carefully exposed and the number of embryos was recorded. To visualize the embryos, we used an MHF-1 Ultraview UltraSound System (E-Technologies) with a focal distance of 7 mm. Intraventricular injections were performed through the uterine wall using pulled microcapillaries (borosilicate glass, Sutter Instruments). The glass needles were visualized in the ultrasound imaging system and when in the lateral ventricles, the injection of 2 μ l was verified by observing the liquid going out of the needle with the aid of an automatic injector (Quintessential injector, Stoelting). Control embryos received injectable water and experimental embryos were injected with 50 μ g of HA dihydrochloride (Sigma-Aldrich), a dose previously determined to decrease TH in the MB (Escobedo-Avila et al., 2014). The osmolarity of this HA solution was 348 mOsm, and HA decreased 0.6 pH units in relation to the vehicle. We have showed (Escobedo-Avila et al., 2014) that the injection of vehicle, or HA at this concentration, did not cause morphological abnormalities in the MB, ruling out damage caused by the osmolarity or pH.

After injection, the uterus was repositioned in its physiological site. The incisions of the abdominal muscle and skin were stitched with separate sutures. Finally, we applied Buprenorphine as analgesic (0.1 mg/kg, Pisa Laboratories) and animals were monitored and left to recover. Two or 6 days after the injections, pregnant rats were euthanized by an overdose of sodium pentobarbital (Pfizer), and identified embryos were recovered for analysis. We performed the following number of successfully injected E12 embryos: vehicle: 53; HA: 59. The number of dams used was 22.

Immunohistochemistry

Embryos were removed and transferred to PBS. Subsequently fixed by immersion in 4% paraformaldehyde solution in PBS at 4°C, overnight (for E14 embryos) and for 6 days (for E18 embryos). Afterward, tissues were cryoprotected overnight in 30% (v/v) sucrose, embedded in Tissue Freezing Medium (Tissue-Tek, Zakura Finetek) and frozen at -80°C. Coronal and sagittal sections of 20 µm were obtained on a cryostat (Leica). For immunohistochemistry, slides were rehydrated in PBS. Tissue sections were treated with immuno-retriever for 30 min at 65°C and subsequently, permeabilized and blocked for 1 1/2 h at room temperature with 0.3% Triton X-100 and 5% horse serum in PBS. Slides were incubated overnight at 4°C with the following primary antibodies diluted in PBS containing 5% horse serum: rabbit polyclonal anti-TH (1:1000, Pel-freez), rabbit polyclonal anti-Lmx1a (1:1500, Millipore), mouse monoclonal anti-β-III Tubulin (1:500, Babco Covance). Then, the sections were washed three times for 5 min in PBS and incubated with the secondary antibodies Alexa-Fluor 488 anti-rabbit IgG and Alexa-Fluor 568 anti-mouse IgG (1:500; Molecular Probes) in PBS containing 5% horse serum for 2 h. Nuclei were stained with Hoechst 33258 (1 ng/mL). After three further washes in PBS, the sections were mounted in Aqua Poly/Mount (Polysciences, Inc.) Immunostainings were analyzed with an epifluorescence microscope (Nikon, Eclipse TE2000-U) and photographed with a Nikon digital camera (DMX1200 F). Some images were acquired with a FV1000 Olympus confocal microscope. Negative controls were performed in the absence of primary antibodies and showed no unspecific staining.

Immunohistochemistry of Whole-Mount Brains From Embryos

We perform immunohistochemistry of whole-mount brain as previously reported (Joyner and Wall, 2008). Embryos at E18 stage were incubated in 20% dimethyl sulfoxide (DMSO): 80% methanol (fixative solution) for 24 h at 4°C. Then, the embryos were transferred to 10% hydrogen peroxide diluted in fixative and incubated overnight at room temperature. This treatment effectively bleached the pigment of the embryo and blocked endogenous peroxidase activity. After bleaching, embryos were incubated in 100% methanol at -20°C for 4 h. For staining, bleached embryos were washed twice for 1 h with 2% skim milk and 0.5% Triton X-100 in PBS (PBSMT). Embryos were incubated overnight at 4°C with

rabbit polyclonal anti-TH (1:1000, Pel-freez) in PBSMT. After this incubation, the embryos were washed seven times in PBSMT; each wash lasted 1 h. The embryos were then incubated overnight at 4°C with affinity-purified goat anti-rabbit immunoglobulin antibody conjugated to horseradish peroxidase (1:500, Santa Cruz) in PBSMT. The seven washes were repeated. Bound peroxidase-conjugated antibody was visualized using the DAB Substrate Kit for Peroxidase (VECTOR laboratories); reactions were carried out for 10 min at room temperature. Embryos were post-fixed in 4% paraformaldehyde in PBS at 4°C overnight and dehydrated in 50, 80, and 100% methanol (1 h each). Embryos were cleared by placing them in a 1:2 mixture of benzyl alcohol:benzyl benzoate. Immunostainings were visualized with a stereomicroscope (Nikon, SMZ1500) and photographed with a Nikon digital camera (Coolpix S10). This preparation allows to visualize the nigro-striatal as well as the mesolimbic and mesocortical pathways. We measured the area of TH+ staining with ImageJ software in: (a) the midbrain (including the *substantia nigra pars compacta* and the Ventral Tegmental Area); (b) the Medial Forebrain Bundle (MFB); and (c) the forebrain (striatum/cortex, including the Nucleus Accumbens). The average of four embryos was determined per treatment. Analysis is presented as the mean ± standard error of the mean (SEM) and statistical analysis by multiple *t*-test. Since the immunostaining with positive TH was tridimensional, the area of positive label was the preferred parameter to measure.

Quantitative RT-PCR (RT-qPCR) Analysis

Total RNA was extracted with TRIzol (Invitrogen) according to the manufacturer's specifications. The mRNA levels were determined by RT-qPCR performed with cDNA generated from 2 µg of total RNA using the Kit GeneAmp RNA PCR KIT (Applied Biosystems). Primers were synthesized based on the UCSC Genome Browse database from *Rattus norvegicus* (Gibbs et al., 2004; Havlak et al., 2004). Gene expression levels were determined using the following primers (all listed in 5' to 3' orientation; Sense, S; Antisense, AS): *Lmx1a*, S: GCACGGAAGCTAGACTCAA; AS: GCTCTGCC CAGCAAAGAG; 143 bp. *Msx1*, S: CTGTTGGGGGACTCC TCAA; AS: GCCGCCTGGCTGGGGG; 119 bp. *Foxa2*, S: CAGAAAAAGGCCTGAGGTG; AS: CAGCATACTTTAACTC GCTG; 137 bp. *Nurr1*, S: GTCACAGAGACACGGG; AS: GGTAGTTGGGTTCGGTTCAA; 121 bp. *Pitx3*, S: CTCGAAGCCCTGCGCTGT; AS: GCCTTCTCCGAGTC ACTGT; 100 bp. *Th*, S: CCACTGGAGGCTGTGGTATT; AS: CCGGGTCTCTAAGTGGTGAA; 145 bp. H₁ receptor, S: CTCTACCTCCCCACTTTGCT; AS: TTCCCTTTCCC CCTCTTG; 292 bp. H₂ receptor S: TTCTTGGACTCCT GGTGCTGC; AS: CATGCCCCCTCTGGTCCC; 309 bp. *Gapdh*, S: GTGGACCTCATGGCCTACAT; AS: GGATGGAATT GTGAGGGAGA; 160 bp. *Gapdh* was used as an internal control. The qPCR reactions for 96-well plate format were performed using 50 ng of cDNA and the Thermo Maxima SYBR Green/ROX 1 PCR Master Mix (Thermo Fisher Scientific) with a StepOnePlus Real-Time PCR System (Applied Biosystems).

The fold change was calculated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Sodium Bisulfite Sequencing

Genomic DNA was isolated by phenol:chloroform:isoamyl alcohol (Sigma-Aldrich) extraction technique as reported (Clark et al., 2006). One microgram of genomic DNA was modified by sodium bisulfite using the EZ DNA methylation kit (ZYMO Research). For sequencing, modified DNA was amplified by PCR using the following primers recognizing *Th*, S: GTAGGTGTTTGTGATAGTGGATG; AS: AAAACCACTCCAACCCTAAATA, with a length of 439 bp. These primers were synthesized based on the UCSC Genome Browse database from *R. norvegicus* (Gibbs et al., 2004; Havlak et al., 2004). The PCR products were purified and cloned using the pGEM-T Easy Vector system (Promega). Eight clones were randomly chosen and sequenced individually. The percentage of modified cytosines was calculated for vehicle- and HA-treated embryos.

Methylated and Hydroxymethylated DNA Immunoprecipitation Assay (MeDIP and hMeDIP)

Genomic DNA was isolated as described above and processed by ultrasonic pulses (GENEQ, GEX500; Cole-Parmer). Initial experiments were set up to obtain fragments of an average size of 300 bp. Such conditions included 3 rounds of pulses of 30 s, with sequences of pulse-on for 5 s and pulse-off for 3 s at 35% amplitude. DNA was immunoprecipitated using a MeDIP/hMeDIP kit (Diagenode). Briefly, 1 μ g of fragmented DNA was denatured for 3 min at 95°C and incubated with 2.5 μ g of the company-validated mouse monoclonal anti-5hmC (Diagenode) or 2 μ g of mouse monoclonal anti-5mC (Diagenode). The antibody-bound DNA was recovered by immunoabsorption with anti-mouse IgG-coated magnetic beads (Diagenode) for 16 h at 4°C. The immunoprecipitated DNA was amplified by 30 cycles of PCR. Primers were designed from the sequences obtained in the UCSC Genome Browse database from *R. norvegicus* (Gibbs et al., 2004; Havlak et al., 2004). For the mDA differentiation genes, oligonucleotides were designed in exons, which represent regulatory regions for neurogenesis. Exons were chosen based on their higher percentage of CpG content or within a CpG island. Primers used are as follows: *Lmx1a*, exon 2, S: GGGTCATCTCGGATAGGTTT; AS: CAGGCACTTACTTCTCGTAG; 156 bp. *Msx1*, exon 1, S: CGAAACCCATGATCCAGGG; AS: GTCCTCCACTTTGACACCG; 112 bp. *Foxa2*, exon 3, S: CGTCATCACCATGGCCAT; AS: GCCGGTAGAAAGGGAAGAG; 96 bp. *Nurr1*, exon 2, S: CAGTCCGAGGAGATGATGC; AS: GGAAGTTGTGAAGGGAGCC; 133 bp. *Pitx3*, exon 4, S: CTTAGTCCCTGCCAGTACG; AS: GCACCCCTTTTCAGACCCT; 149 bp. *Th*, exon1, S: GAGACAGAACTCGGGACC; AS: CGGGTGACAGCATATCCTC; 145 bp. The efficiency of MeDIP or hMeDIP products of a particular genomic locus was calculated from qPCR by a fast optical 96-well qPCR reaction plate (Applied Biosystems). The qPCR reaction was performed using

Thermo Maxima SYBR Green/ROX 1 PCR Master Mix (Thermo Fisher Scientific) with a StepOnePlus Real-Time PCR System (Applied Biosystems). According to the manufacturer inset, data for each sample is reported as recovery of starting material: % MeDIP or hMeDIP/INPUT.

Chromatin Immunoprecipitation to Detect H3K9me3 and H3K27me3

Ventral midbrains were dissected and pooled from 7 embryos treated with HA or vehicle. Tissue was maintained in ice-cold PBS before crosslinking, which was carried out for 10 min, with 1% of formaldehyde. The crosslink reaction was stopped for 5 min with 125 mM glycine and the crosslinked tissue was washed three times with ice-cold PBS for 5 min. After the last wash, tissue was lysed with 500 μ L of lysis buffer. Chromatin was sonicated using an Ultrasonic Processor (GENEQ, GEX500; SOVC505-00). After sonication, DNA was extracted and its length evaluated. Quantitation of chromatin was made by the Lowry reaction and 100 μ g of material was used per IP. Immunoselection was performed according to the protocol of the OneDay ChIP kit (Diagenode, Kch-oned IP- 180), with anti-H3K9me3 or anti-H3K27me3 (Diagenode). Immunoprecipitation was performed by incubating 20 μ L of magnetic beads (16-663 | Magna ChIP™ Protein A+G Magnetic Beads), with the antibody-chromatin complexes for 3 h in a rotating wheel; then the beads-antibody-chromatin complexes were washed three times with ice-cold ChIP buffer 1 \times . Finally, the DNA was decrosslinked overnight and purified employing the MinElute Reaction Cleanup Kit (Cat No./ID: 28204) following the manufacturer's instructions. ChIP assays were performed by qPCR in a fast optical 96-well qPCR reaction plate (Applied Biosystems) with the same primers used for MeDIP/hMeDIP for *Pitx3* and *Th*. The qPCR reaction was performed using Thermo Maxima SYBR Green/ROX1 PCR Master Mix (Thermo Fisher Scientific) with a StepOnePlus Real-Time PCR System (Applied Biosystems), according to the manufacturer's instructions.

Dopamine Quantification by High-Performance Liquid Chromatography (HPLC)

E18 embryos were recovered after vehicle or HA injection at E12. The striatum was dissected and individually homogenized with a mixture of 0.1 M perchloric acid (Baker), 4 mM Na₂S₂O₃ (Sigma), 0.1 mM EDTA (Sigma). The homogenate was centrifuged (12,600 rpm, 20 min at 4°C). Subsequently, the supernatant was filtered with 0.45 μ m filters (Millex-HV) and 20 μ l were mixed with 7.5 μ L of antioxidant solution (0.1 N perchloric acid; 0.02% EDTA and 1% EtOH) and injected into the solvent stream of a HPLC system, using a reversed-phase column (C18, 3 μ m; 2.1 \times 50 mm; Atlantis, Waters) coupled to a pre-column (Nova-Pack, Waters) with a mobile phase solution containing 0.054 mM EDTA, 50 mM citric acid, and 0.1 mM octansulfonic acid, dissolved in milli-Q water and mixed with methanol in a proportion of 97:3, respectively. The pH was 2.9 and the flow rate 0.35 mL/min. Dopamine detection was performed by a

single-channel electrochemical amperometric detector (Waters model 2465) at 450 mV at a temperature of 30°C, and quantified by peak height measurements against standard solutions with known dopamine concentration, as described (Rocha et al., 2012).

Statistical Analysis

Data from four to six independent experiments are expressed as the mean \pm SEM. The differences between groups were analyzed using the unpaired Student's *t*-test comparing the HA- with the vehicle-treated rat embryos. Differences between groups were considered statistically significant when $p < 0.05$.

RESULTS

Intrauterine Administration of HA Decreases TH and Lmx1a Immunoreactivity in the VM

We performed *in utero* microinjections of HA or vehicle, into the ventricular lumen close to the MB of E12 rat embryos, allowing them to develop to E14. Because it was previously reported that intrauterine injection of HA at E12 precludes mDA differentiation at E14 (Escobedo-Avila et al., 2014), we quantified the immunoreactivity of TH and Lmx1a in vehicle- or HA-injected embryos, confirming a significant decrease in TH (Figure 1A). We also assessed the effect of HA on VM dopamine progenitors positive to Lmx1a (Nakatani et al., 2010; Yan et al., 2011). We found that HA down-regulates Lmx1a reactivity, without altering β -III Tubulin, in contrast to the vehicle-treated embryos (Figure 1B).

Administration of HA Down-Regulates the Expression of Specification and Differentiation Dopaminergic Genes

Based on the Lmx1a-positive domain, we performed dissections selecting the VM of injected embryos 2 days after surgery. Subsequently, we analyzed the expression of DA specification (*Lmx1a*, *Msx1*, and *Foxa2*) and differentiation (*Nurr1*, *Pitx3*, and *Th*) genes by RT-qPCR. After normalization with *Gapdh*, we found that HA significantly decreases the expression of *Lmx1a*, *Msx1*, and *Th* compared to vehicle-injected embryos (Figure 1C). H₁ and H₂ receptors are expressed in the VM; we did not find significant changes in these receptors' expression after HA injection (data not shown).

HA Reduces 5mC at the Exons Regions of Genes Associated With DA Differentiation

Next, we asked if HA affected epigenetic modifications associated to DNA. By bisulfite sequencing, we evaluated the DNA methylation patterns in HA-treated embryos, after MB dissections (Figure 2A). We did not observe overall changes in the percentage of modified cytosines on *Th* gene, including exon 1, on HA-treated embryos, compared to vehicle-treated

embryos, i.e., the percentages of modified bases were very similar (Figure 2B).

One limitation of bisulfite sequencing is that it does not allow discriminating between 5mC and 5hmC. Therefore, in order to evaluate both DNA modifications, we perform a MeDIP and hMeDIP, followed by qPCR in dissected VM (Figure 2A). Particularly, we focused on the intragenic regulatory regions of specification (*Lmx1a*, *Msx1*, and *Foxa2*) and differentiation (*Nurr1*, *Pitx3*, and *Th*) genes. We found that HA-injected embryos showed a significant decrease in 5mC at the exonic regions of the differentiation genes *Pitx3* and *Th*, compared to vehicle-treated embryos (Figure 2C). The 5hmC did not show significant changes on specification or differentiation genes neither in HA- nor in vehicle-treated embryos (Figure 2D).

HA Treatment Is Associated With an Increase in the H3K9me3 and H3K27me3 Histone Modifications

Due to the histone post-translational modifications may also participate in the modulation of gene expression, we decided to study two particular histone modifications the H3K9 trimethylation (H3K9me3) and H327 trimethylation (H3K27me3). In order to evaluate the effect of HA treatment in the histone modifications H3K9me3 and H3K27me3 we performed a chromatin immunoprecipitation assay (ChIP) after the HA or vehicle treatment, and then evaluate the immunoprecipitated DNA by qPCR analyzing the above-mentioned regions of *Pitx3* (Figure 2E) and *Th* (Figure 2F) genes. We found that HA-injected embryos showed a significant increase of H3K9me3 and H3K27me3, in contrast with vehicle-treated embryos (Figures 2E,F).

HA Injection Has a Long-Term Effect During the Formation of Midbrain Dopaminergic Pathways

We set out to investigate if the effect of HA on mDA 2-days after its injection has a long-term consequence, affecting the formation of nigrostriatal pathway. *In utero* injections of E12 embryos with either HA or vehicle were made, and the embryos were collected after 6 days, when normal development of the nigrostriatal pathway has reached the striatum. TH staining in sagittal sections or whole-mount immunohistochemistry was analyzed in E18 brains. A single injection of HA at early stages was enough to alter the formation of the nigrostriatal pathway; in vehicle-injected embryos, the MFB is apparently normal, whereas after HA administration a decrease in this bundle is observed and aberrant axons innervating the *thalamus* appeared (solid rectangles in Figure 3A). To comprehensively analyze the mesencephalic dopaminergic pathways, TH immunohistochemistry was performed in whole-brain mounts. Image analysis was used to quantify the indicated areas positive for TH within the dotted rectangles shown in Figure 3B. HA caused a significant reduction in the TH-positive area in the midbrain and the striatum/nucleus accumbens/cortex when compared to vehicle-injected embryos (Figure 3C). Also, the quantification of dopamine by HPLC assay in the striatal area

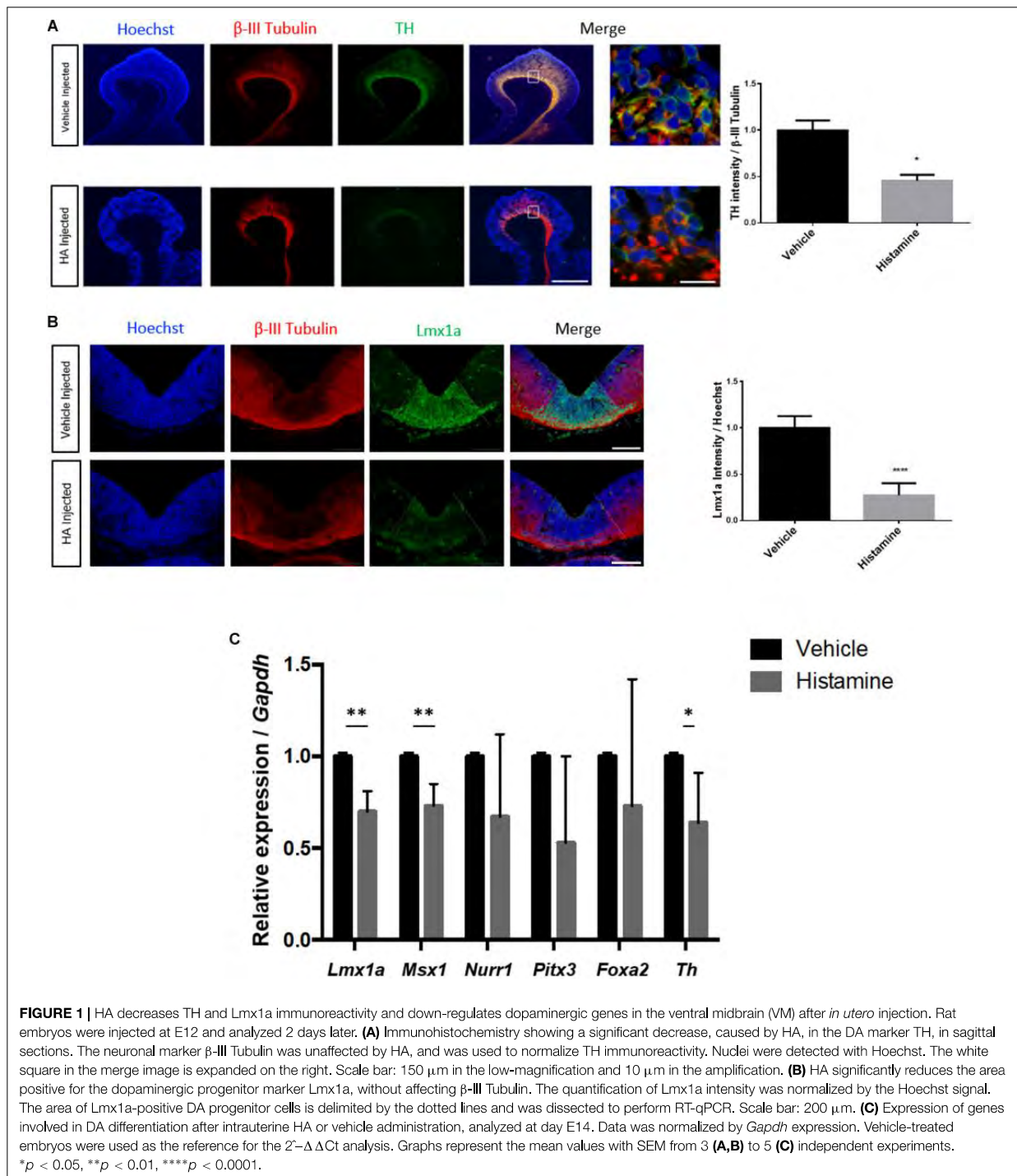
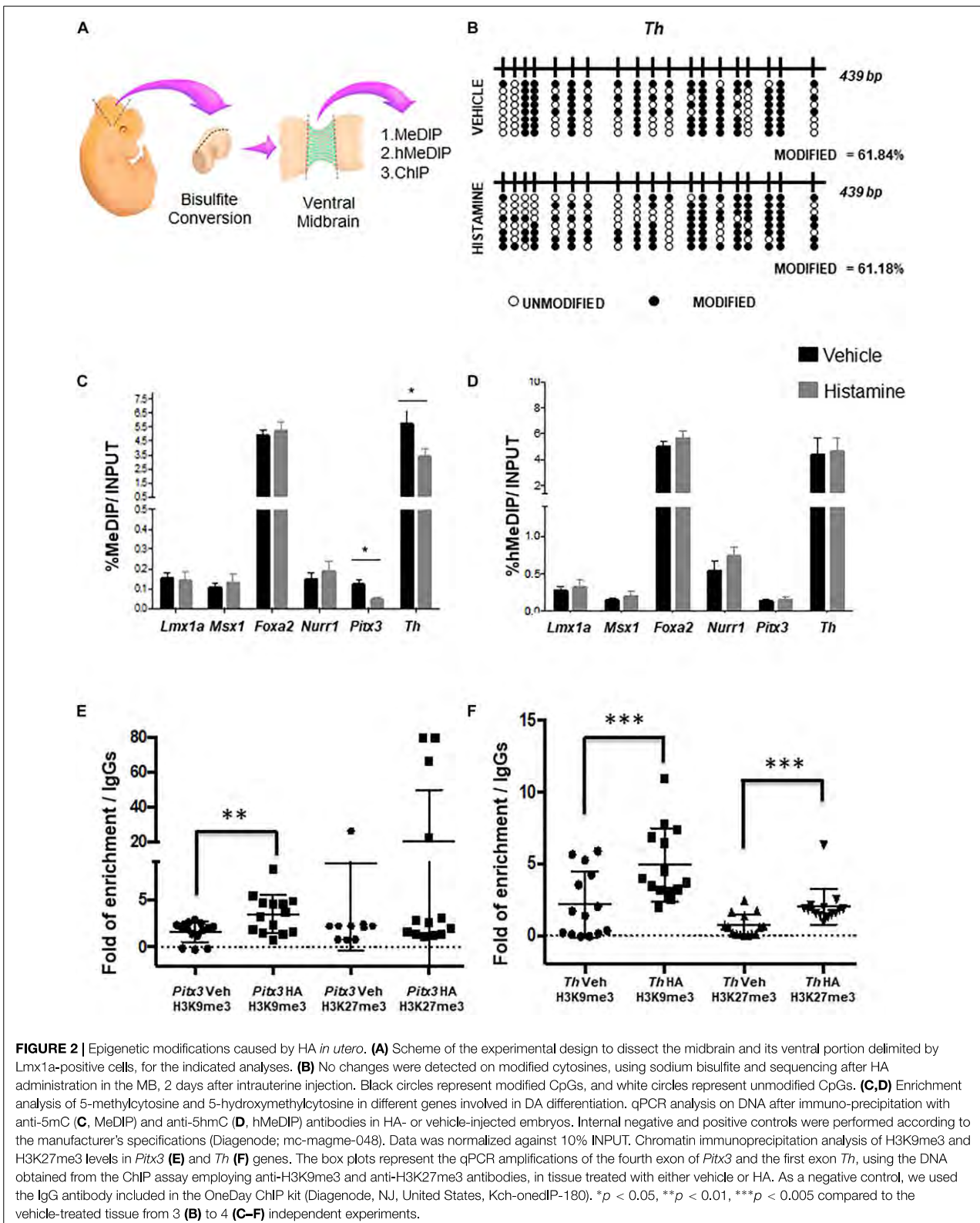


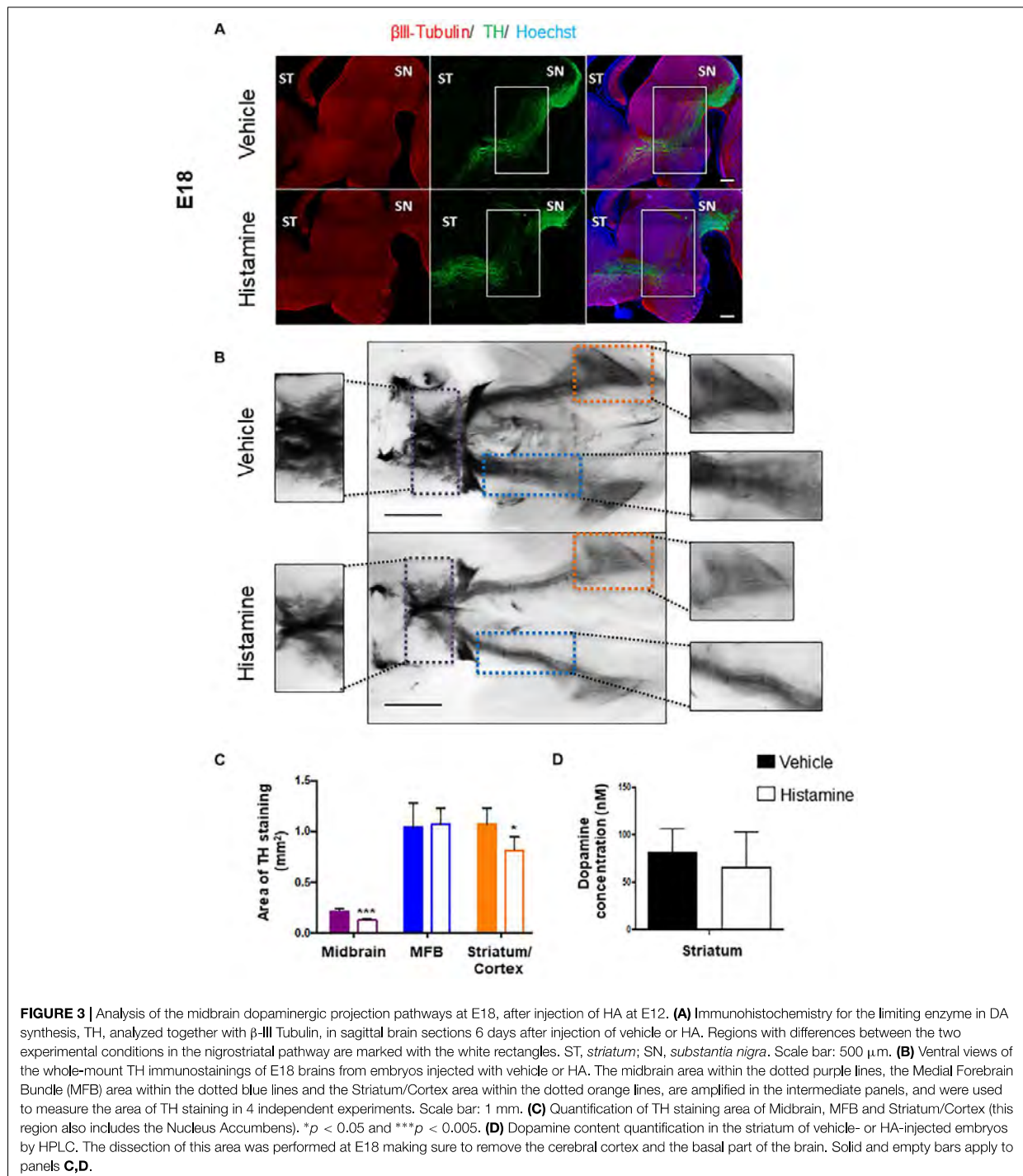
FIGURE 1 | HA decreases TH and Lmx1a immunoreactivity and down-regulates dopaminergic genes in the ventral midbrain (VM) after *in utero* injection. Rat embryos were injected at E12 and analyzed 2 days later. **(A)** Immunohistochemistry showing a significant decrease, caused by HA, in the DA marker TH, in sagittal sections. The neuronal marker β -III Tubulin was unaffected by HA, and was used to normalize TH immunoreactivity. Nuclei were detected with Hoechst. The white square in the merge image is expanded on the right. Scale bar: 150 μ m in the low-magnification and 10 μ m in the amplification. **(B)** HA significantly reduces the area positive for the dopaminergic progenitor marker Lmx1a, without affecting β -III Tubulin. The quantification of Lmx1a intensity was normalized by the Hoechst signal. The area of Lmx1a-positive DA progenitor cells is delimited by the dotted lines and was dissected to perform RT-qPCR. Scale bar: 200 μ m. **(C)** Expression of genes involved in DA differentiation after intrauterine HA or vehicle administration, analyzed at day E14. Data was normalized by *Gapdh* expression. Vehicle-treated embryos were used as the reference for the $2^{-\Delta\Delta Ct}$ analysis. Graphs represent the mean values with SEM from 3 **(A,B)** to 5 **(C)** independent experiments. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$.

shows a 19% non-significant reduction in HA injected animal compared to vehicle (Figure 3D). Together our data suggest that HA injection leads to a decrease in TH staining in the forebrain, which might result in lower dopamine release.

DISCUSSION

Brain development is initiated by formation of defined patterns in the neural ectoderm. As development proceeds, stem





cell differentiation to neurons is defined by the action of morphogenic gradients and transcription factors, expressed in a temporal and sequential order (Gale and Li, 2008). In cultured cortical NSPC, HA has a neuronal-promoting action,

involving up-regulation of proneural genes (Molina-Hernández and Velasco, 2008; Rodríguez-Martínez et al., 2012). However, HA administration *in vivo* diminishes proliferation of midbrain NSPC, down-regulating *Th* mRNA levels and the content of

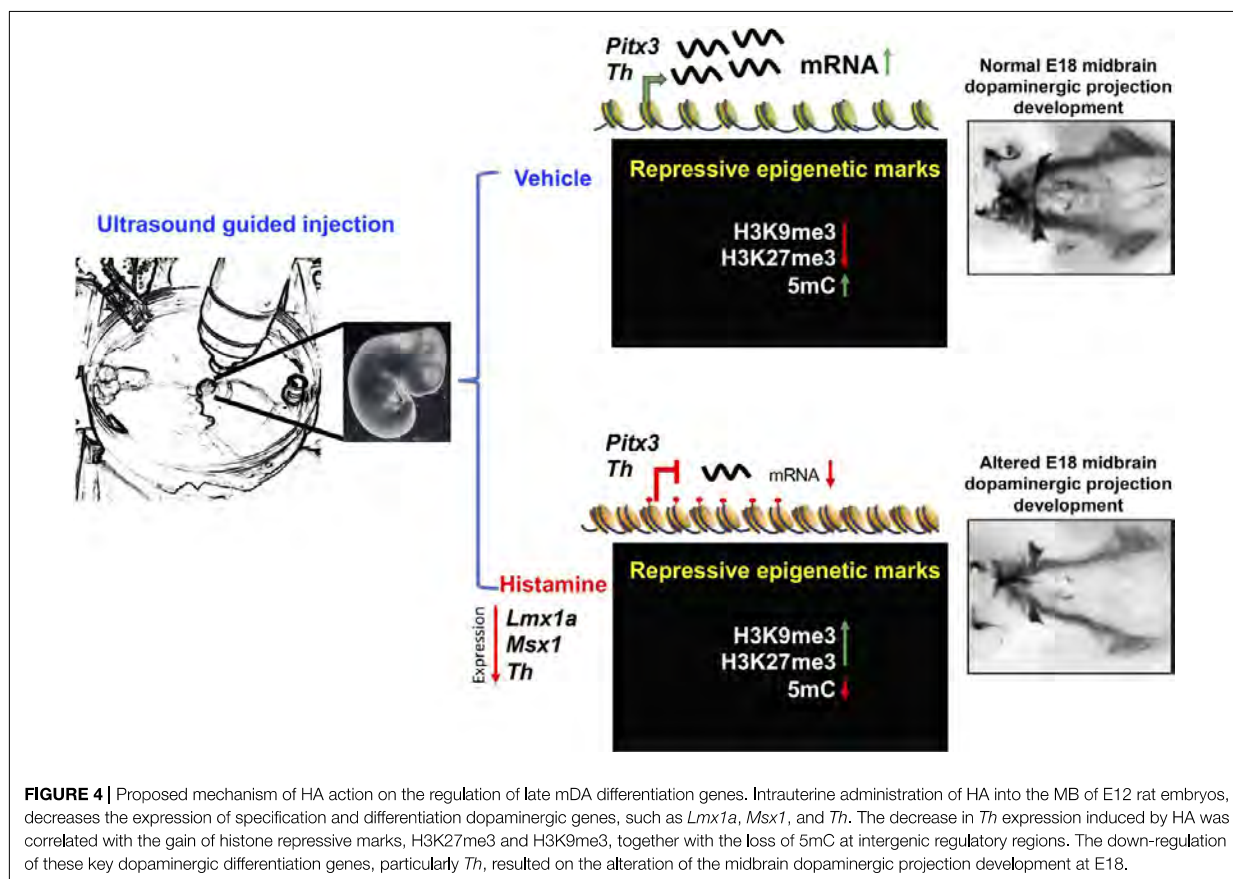


FIGURE 4 | Proposed mechanism of HA action on the regulation of late mDA differentiation genes. Intrauterine administration of HA into the MB of E12 rat embryos, decreases the expression of specification and differentiation dopaminergic genes, such as *Lmx1a*, *Msx1*, and *Th*. The decrease in *Th* expression induced by HA was correlated with the gain of histone repressive marks, H3K27me3 and H3K9me3, together with the loss of 5mC at intergenic regulatory regions. The down-regulation of these key dopaminergic differentiation genes, particularly *Th*, resulted on the alteration of the midbrain dopaminergic projection development at E18.

TH, *Lmx1a*, and *Lmx1b* assessed by western blot, 2 days after injection (Escobedo-Avila et al., 2014), but the mechanisms for these actions are just emerging (Molina-Hernández et al., 2013). In the current work, we administered embryos with either HA or vehicle, on a stage previous to the major production of DA neurons (Götz and Huttner, 2005), by ultrasound guided-injections. Injection of HA diminishes the presence of TH as well as the transcription factor *Lmx1a* in the VM.

Recently, studies about the influence of environment on epigenetics and its importance with regards to the expression of transcription factors during cell commitment have emerged. Regulation of the deposition of 5hmC in intragenic regions during progression of embryonic stem cells to NSPC is crucial for the development of mDA neurons (Kim et al., 2014). In dopaminergic neurons, methylation in one specific CpG site on the first exon of human *TH* gene down-regulates its transcription, allowing the binding of repressive transcription factors such as KAISO and NcoR and recruitment of HDACs, which might act as an additional modulator of *Th* expression (Arányi et al., 2005; van Heesbeen et al., 2013). We found that HA injection significantly decreased expression of early-differentiation genes (*Lmx1a* and *Msx1*) as well as in *Th*, a late canonical mDA differentiation gene.

Presence of 5mC and 5hmC within the exonic regions have been associated with up-regulation of gene expression, through

different mechanisms (Hahn et al., 2013; Guo et al., 2014; Kim et al., 2014). Our findings reveal that injection of HA at early stages significantly diminish the percentage of enrichment of 5mC, but only on the differentiation genes *Pitx3* and *Th*. Although DNA methylation has been related to transcriptional repression, in recent years, some reports showed that 5mC within the gene body, specifically in exons, is associated to regulation of pre-mRNA splicing (Lev Maor et al., 2015). Other studies associated the gain of 5hmC at intergenic regions with transcriptional activation of neuronal differentiation genes (Hahn et al., 2013; Guo et al., 2014). We did not observe changes on 5hmC levels, indicating that HA induced gene repression independent of 5hmC, and suggesting that other epigenetic mechanisms participate in mDA gene regulation. Interestingly, many genes associated with neuronal differentiation can be negatively regulated by decreasing 5mC and a gain in histone repressive modifications, such as H3K27me3, at their regulatory regions (Hahn et al., 2013).

During mDA differentiation, epigenetic changes occur in cultured VM NSCP: Nurr1-controlled dopaminergic gene expression was precluded by the interaction of Nurr1 with CoREST and HDAC1, a complex that caused deacetylation of these promoters. Forced expression of *Foxa2* allowed the formation of Nurr1-*Foxa2* complexes with concomitant

acetylation of H3, associated to open chromatin, in *Th* and *Dat* promoters (Yi et al., 2014). *In vitro* expansion of VM NSCP caused decreased dopaminergic differentiation, which was associated with hypoacetylation on H3 and H4 and increased H3K27me3 in the *Foxa2* promoter, as well as globally. Inhibition of HDAC activity up-regulates *Foxa2* transcripts, caused H3 acetylation, decreased methylation of H3K27 and restored dopaminergic induction (Bang et al., 2015). We observed significant increases in the repressive marks H3K27me3 and H3K9me3 in *Pitx3* and *Th* genes, although significant down-regulation was observed only in *Th*. Similar observations were recently reported by Hong et al. (2019), where they manipulated *in vitro* the levels of H3K27me3 in the promoter region of *Th*: treatment with EPZ005687, an inhibitor of histone methylase EZH2 reduces the levels of H3K27me3 and leads to an increase in TH levels; conversely, addition of GSK-J1, which affects the demethylase activity of Jmjd3, augmented H3K27me3 and decreased TH protein levels (Hong et al., 2019). Our data points that not only the H3K27me3 is involved in the regulation of *Th*, but also H3K9me3 participate.

We studied the long-term consequence of deregulation of such genes, in the formation of the MFB, and found that injection of HA caused disrupted and miss-oriented dopaminergic fibers in sagittal sections. Further, a significant reduction in TH-positive staining was found in the MB and the striatum/cortex after HA treatment, in whole-brain mounts. However, we did not observe a complete deterioration of the nigro-striatal, mesolimbic and mesocortical pathways, probably due to the plasticity of the developing brain; accordingly, dopamine level in the striatum were not reduced by HA. In newborn rats injected with 6-hydroxydopamine to degenerate mDA neurons, there is an increase in serotonergic projections (Snyder et al., 1986) and lesioned animals were hyperactive (Miller et al., 1981; Zhang et al., 2001; Avale et al., 2004). It will be of interest to analyze if embryos injected with HA at E12 show changes in locomotor activity at postnatal stages.

The precise mechanisms that link HA and the epigenetic changes reported here deserve further investigation. In cultured VM-NSPC, it has been shown that membrane depolarization and calcium influx is correlated with higher acetylation of Histone 3, trimethylation of K4 of H3 with reductions in the H3K9me3 and H3K27me3 marks (He et al., 2011). HA triggers intracellular calcium increases after receptor activation in VM-NSPC, through production of IP3 and Diacylglycerol by Phospholipase C, but the effects reported here are in the opposite direction: addition of the transcriptional repressive H3K9me3 and H3K27me3; additional signaling transduction mechanism to explore include Phospholipase A2/arachidonic acid production and nitric oxide production/cGMP increases (Molina-Hernández et al., 2012). On the other hand, the participation of the transient population of histaminergic neurons of the raphe nucleus or in the hypothalamus (Vanhala et al., 1994), and the activation of microglia after HA administration, as observed in adult mice (Rocha et al., 2016), cannot be ruled out in our experiments; therefore,

it will be interesting to determine their roles in future experiments by ablating HA-producing neurons or microglial cells, respectively. All together, the data reported here describe molecular effects of HA on dopaminergic midbrain development. The reduction in expression of dopaminergic genes correlates with gain of the repressive modifications H3K9me3 and H3K27me3 and the loss of DNA methylation at intergenic regions in *Th*. Moreover, HA induced long-term alterations on mDA neurons, as shown in **Figure 4**. Our results suggest that HA promotes an epigenetic change related to significant reductions in the expression of important dopaminergic genes. Such alterations result in anatomical dysregulation and decreased TH staining in the dopaminergic neurons targets.

In conclusion, this study contributes to the understanding of the molecular actions of HA on mDA development, as well as chromatin structure regulation mediated by extracellular stimuli, during NPSC commitment in the midbrain; such knowledge may contribute to establish different roles of HA during brain development.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was carried out in accordance with Mexican guidelines and following recommendations of Instituto de Fisiología Celular, UNAM Animal Care and Use Committee, which approved this protocol.

AUTHOR CONTRIBUTIONS

FV-R wrote the manuscript, analyzed the data and performed the majority of experiments, including epigenetics, whole mount stainings and dopamine quantification. LG-C and RG-B performed RT-qPCR and ChIP assays, contributed to the design of experiments and discussion of results. IE-A helped to perform intrauterine injections, data analysis and participated on discussion of results. DC-P contributed to set up techniques, performed results analysis and discussed results. AL-O and LR participated in dopamine quantification by HPLC. IV and ES-R contributed to the experimental design and supervision, reviewed the manuscript, and obtained funding. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México (PAPIIT IN213716 and IN213719), Consejo Nacional de Ciencia y Tecnología (CONACyT, CB09/131281), and Red Temática Células Troncales y Medicina

Regenerativa to IV. ES-R was supported by grants from Fondo Sectorial de Investigación en Salud y Seguridad Social (FOSISS, 0261181) and CONACyT (182997 and 284748). Fernanda Vargas-Romero and Lissania Guerra-Calderas received a graduate fellowship from CONACyT (662044 and 588391, respectively).

REFERENCES

- Abeliovich, A., and Hammond, R. (2007). Midbrain dopamine neuron differentiation: factors and fates. *Dev. Biol.* 304, 447–454. doi: 10.1016/j.ydbio.2007.01.032
- Aguilera, O., Fernández, A. F., Muñoz, A., and Fraga, M. F. (2010). Epigenetics and environment: a complex relationship. *J. Appl. Physiol.* 109, 243–251. doi: 10.1152/jappphysiol.00068.2010
- Ang, S.-L. (2006). Transcriptional control of midbrain dopaminergic neuron development. *Development* 133, 3499–3506. doi: 10.1242/dev.02501
- Arányi, T., Faucheux, B. A., Khalifallah, O., Vodjdani, G., Biguet, N. F., Mallet, J., et al. (2005). The tissue-specific methylation of the human tyrosine hydroxylase gene reveals new regulatory elements in the first exon. *J. Neurochem.* 94, 129–139. doi: 10.1111/j.1471-4159.2005.03173.x
- Avale, M. E., Nemirovsky, S. I., Raisman-Vozari, R., and Rubinstein, M. (2004). Elevated serotonin is involved in hyperactivity but not in the paradoxical effect of amphetamine in mice neonatally lesioned with 6-hydroxydopamine. *J. Neurosci. Res.* 78, 289–296. doi: 10.1002/jnr.20245
- Bang, S. Y., Kwon, S. H., Yi, S. H., Yi, S. A., Park, E. K., Lee, J. C., et al. (2015). Epigenetic activation of the *Foxa2* gene is required for maintaining the potential of neural precursor cells to differentiate into dopaminergic neurons after expansion. *Stem Cells Dev.* 24, 520–533. doi: 10.1089/scd.2014.0218
- Bernardino, L., Eiriz, M. F., Santos, T., Xapelli, S., Grade, S., Rosa, A. L., et al. (2012). Histamine stimulates neurogenesis in the rodent subventricular zone. *Stem Cells* 30, 773–784. doi: 10.1002/stem.1042
- Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L., and Frommer, M. (2006). DNA methylation: bisulphite modification and analysis. *Nat. Protoc.* 1, 2353–2364. doi: 10.1038/nprot.2006.324
- Escobedo-Avila, I., Vargas-Romero, F., Molina-Hernández, A., López-González, R., Cortés, D., De Carlos, J. A., et al. (2014). Histamine impairs midbrain dopaminergic development in vivo by activating histamine type 1 receptors. *Mol. Brain* 7:58. doi: 10.1186/s13041-014-0058-x
- Gale, E., and Li, M. (2008). Midbrain dopaminergic neuron fate specification: of mice and embryonic stem cells. *Mol. Brain* 1, 1–10. doi: 10.1186/1756-6606-1-8
- Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., et al. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493–521. doi: 10.1038/nature02426
- Götz, M., and Huttner, W. B. (2005). The cell biology of neurogenesis. *Nat. Rev. Mol. Cell Biol.* 6, 777–788.
- Guo, J. U., Szulwach, K. E., Su, Y., Li, Y., Yao, B., Xu, Z., et al. (2014). Genome-wide antagonism between 5-hydroxymethylcytosine and DNA methylation in the adult mouse brain. *Front. Biol.* 9:66–74. doi: 10.1007/s11515-014-1295-1
- Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., et al. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Rep.* 3, 291–300. doi: 10.1016/j.celrep.2013.01.011
- Havlak, P., Chen, R., Durbin, K. J., Egan, A., Ren, Y., Song, X.-Z., et al. (2004). The atlas genome assembly system. *Genome Res.* 14, 721–732. doi: 10.1101/gr.2264004
- He, X.-B., Yi, S.-H., Rhee, Y.-H., Kim, H., Han, Y.-M., Lee, S.-H., et al. (2011). Prolonged membrane depolarization enhances midbrain dopamine neuron differentiation via epigenetic histone modifications. *Stem Cells* 29, 1861–1873. doi: 10.1002/stem.739
- Hegarty, S. V., Sullivan, A. M., and O’Keeffe, G. W. (2013). Midbrain dopaminergic neurons: a review of the molecular circuitry that regulates their development. *Dev. Biol.* 379, 123–138. doi: 10.1016/j.ydbio.2013.04.014
- Hong, F., Zhao, M., Zhang, L., and Feng, L. (2019). Inhibition of *Ezh2* in vitro and the decline of *Ezh2* in developing midbrain promotes dopaminergic neurons differentiation through modifying H3K27me3. *Stem Cells Dev.* doi: 10.1089/scd.2018.0258 [Epub ahead of print].
- Huang, H.-Y., Chiu, T.-L., Chang, H.-F., Hsu, H.-R., Pang, C.-Y., Liew, H.-K., et al. (2015). Epigenetic regulation contributes to urocortin-enhanced midbrain dopaminergic neuron differentiation. *Stem Cells* 33, 1601–1617. doi: 10.1002/stem.1949
- Jacobs, F. M. J., van Erp, S., van der Linden, A. J. A., von Oerthel, L., Burbach, J. P. H., and Smidt, M. P. (2009). *Pitx3* potentiates *Nurr1* in dopamine neuron terminal differentiation through release of SMRT-mediated repression. *Development* 136, 531–540. doi: 10.1242/dev.029769
- Joyner, A., and Wall, N. (2008). Immunohistochemistry of whole-mount mouse embryos. *Cold Spring Harb. Protoc.* 3, 1–5.
- Kadkhodaei, B., Ito, T., Joodmardi, E., Mattsson, B., Rouillard, C., Carta, M., et al. (2009). *Nurr1* is required for maintenance of maturing and adult midbrain dopamine neurons. *J. Neurosci.* 29, 15923–15932. doi: 10.1523/JNEUROSCI.3910-09.2009
- Kim, M., Park, Y.-K., Kang, T.-W., Lee, S.-H., Rhee, Y.-H., Park, J.-L., et al. (2014). Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. *Hum. Mol. Genet.* 23, 657–667. doi: 10.1093/hmg/ddt453
- Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends Genet.* 31, 274–280. doi: 10.1016/j.tig.2015.03.002
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Miller, F. E., Heffner, T. G., Kotake, C., and Seiden, L. S. (1981). Magnitude and duration of hyperactivity following neonatal 6-hydroxydopamine is related to the extent of brain dopamine depletion. *Brain Res.* 229, 123–132. doi: 10.1016/0006-8993(81)90750-2
- Molina-Hernández, A., Díaz, N. F., and Arias-Montaño, J.-A. (2012). Histamine in brain development. *J. Neurochem.* 122, 872–882. doi: 10.1111/j.1471-4159.2012.07863.x
- Molina-Hernández, A., Rodríguez-Martínez, G., Escobedo-Ávila, I., and Velasco, I. (2013). Histamine up-regulates fibroblast growth factor receptor 1 and increases *FOXP2* neurons in cultured neural precursors by histamine type 1 receptor activation: conceivable role of histamine in neurogenesis during cortical development in vivo. *Neural Devel.* 8:4. doi: 10.1186/1749-8104-8-4
- Molina-Hernández, A., and Velasco, I. (2008). Histamine induces neural stem cell proliferation and neuronal differentiation by activation of distinct histamine receptors. *J. Neurochem.* 106, 706–717. doi: 10.1111/j.1471-4159.2008.05424.x
- Nakatani, T., Kumai, M., Mizuhara, E., Minaki, Y., and Ono, Y. (2010). *Lmx1a* and *Lmx1b* cooperate with *Foxa2* to coordinate the specification of dopaminergic neurons and control of floor plate cell differentiation in the developing mesencephalon. *Dev. Biol.* 339, 101–113. doi: 10.1016/j.ydbio.2009.12.017
- Nissinen, M. J., and Panula, P. (1995). Developmental patterns of histamine-like immunoreactivity in the mouse. *J. Histochem. Cytochem.* 43, 211–227. doi: 10.1177/43.2.7822777
- Panula, P., and Nuutinen, S. (2013). The histaminergic network in the brain: basic organization and role in disease. *Nat. Rev. Neurosci.* 14, 472–487. doi: 10.1038/nrn3526
- Rocha, L., Alonso-Vanegas, M., Villeda-Hernández, J., Mújica, M., Cisneros-Franco, J. M., López-Gómez, M., et al. (2012). Dopamine abnormalities in the neocortex of patients with temporal lobe epilepsy. *Neurobiol. Dis.* 45, 499–507. doi: 10.1016/j.nbd.2011.09.006
- Rocha, S. M., Saraiva, T., Cristóvão, A. C., Ferreira, R., Santos, T., Esteves, M., et al. (2016). Histamine induces microglia activation and dopaminergic neuronal

ACKNOWLEDGMENTS

We thank Dr. Alfredo Varela Echavarría, Dr. Luis A. Herrera Montalvo, and Jorge Landgrave-Gómez M.Sc., for continuous discussion and advice. We also thank the technical support of Francia Carmona B.Sc.

- **PUBLICACIÓN 6:** The epigenetic factor BORIS (CTCF) controls the androgen receptor regulatory network in ovarian cancer.
- **AUTORES:** Salgado-Albarrán M, González-Barrios R, Guerra-Calderas L, Alcaraz N, Estefanía Sánchez-Correa T, Castro-Hernández C, Sánchez-Pérez Y, Aréchaga-Ocampo E, García-Carrancá A, Cantú de León D, Herrera LA, Baumbach J & Soto-Reyes E (2019) 8, 41.an-Ostrosky.
- **REVISTA:** PLoS ONE. **VOLÚMEN:** 13. **PÁGINAS:** e0206590. **AÑO:** 2019.

Salgado-Albarrán et al. *Oncogenesis* (2019)8:41
<https://doi.org/10.1038/s41389-019-0150-2>

Oncogenesis

ARTICLE

Open Access

The epigenetic factor BORIS (CTCF) controls the androgen receptor regulatory network in ovarian cancer

Marisol Salgado-Albarrán^{1,2}, Rodrigo González-Barrios³, Lissania Guerra-Calderas¹, Nicolás Alcaraz⁴, Thalía Estefanía Sánchez-Correa⁵, Clementina Castro-Hernández³, Yesennia Sánchez-Pérez³, Elena Aréchaga-Ocampo¹, Alejandro García-Carrancá³, David Cantú de León³, Luis A. Herrera³, Jan Baumbach² and Ernesto Soto-Reyes¹

Abstract

The identification of prognostic biomarkers is a priority for patients suffering from high-grade serous ovarian cancer (SOC), which accounts for >70% of ovarian cancer (OC) deaths. Meanwhile, borderline ovarian cancer (BOC) is a low malignancy tumor and usually patients undergo surgery with low probabilities of recurrence. However, SOC remains the most lethal neoplasm due to the lack of biomarkers for early diagnosis and prognosis. In this regard, BORIS (CTCF), a CTCF paralog, is a promising cancer biomarker that is overexpressed and controls transcription in several cancer types, mainly in OC. Studies suggest that BORIS has an important function in OC by altering gene expression, but the effect and extent to which BORIS influences transcription in OC from a genome-wide perspective is unclear. Here, we sought to identify BORIS target genes in an OC cell line (OVCAR3) with potential biomarker use in OC tumor samples. To achieve this, we performed in vitro knockout and knockdown experiments of BORIS in OVCAR3 cell line followed by expression microarrays and bioinformatics network enrichment analysis to identify relevant BORIS target genes. In addition, ex vivo expression data analysis of 373 ovarian cancer patients were evaluated to identify the expression patterns of BORIS target genes. In vitro, we uncovered 130 differentially expressed genes and obtained the BORIS-associated regulatory network, in which the androgen receptor (AR) acts as a major transcription factor. Also, *FN1*, *FAM129A*, and *CD97* genes, which are related to chemoresistance and metastases in OC, were identified. In SOC patients, we observed that malignancy is associated with high levels of BORIS expression while BOC patients show lower levels. Our study suggests that BORIS acts as a main regulator, and has the potential to be used as a prognostic biomarker and to yield novel drug targets among the genes BORIS controls in SOC patients.

Introduction

The most frequent epithelial ovarian cancer (OC) type is serous ovarian cancer (SOC), which accounts for about 90% of the OC cases. On the other hand, borderline ovarian cancer (BOC), which shows a more favorable outcome, accounts for an estimated 15–20% of all ovarian

neoplasms. However, SOC is the most lethal gynecological neoplasm due to the lack of early diagnosis, prognosis, treatment, and response biomarkers¹. Approximately, 70% of the cases are diagnosed in late stages, where the disease has already disseminated and the survival rate is low^{1–3}. Thus, the search for new cancer biomarkers is one of the main goals in OC research.

In this regard, one protein that has gained interest by its potential use as a biomarker is the Brother Of the Regulator of Imprinted Sites (BORIS)^{4–7}. BORIS is a transcriptional factor coded by the *CTCF* gene, a paralog of *CTCF*. BORIS shows a very specific expression pattern;

Correspondence: Ernesto Soto-Reyes (esotoreyes@correo.cua.uam.mx)

¹Natural Sciences Department, Universidad Autónoma Metropolitana-Cuajimalpa (UAM-C), Mexico City 05300, Mexico

²Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany
 Full list of author information is available at the end of the article.

© The Author(s) 2019



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Oncogenesis

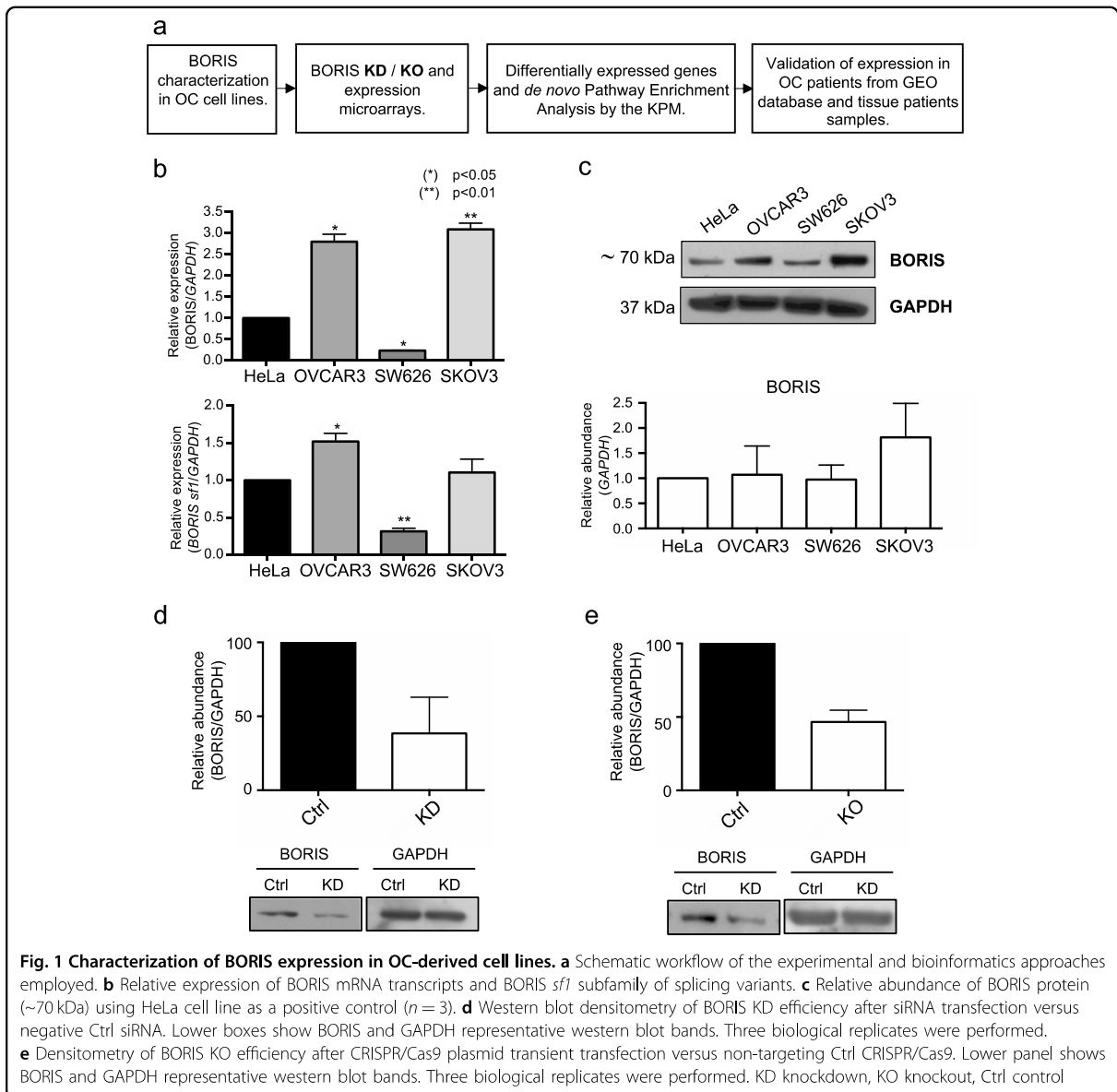


Fig. 1 Characterization of BORIS expression in OC-derived cell lines. **a** Schematic workflow of the experimental and bioinformatics approaches employed. **b** Relative expression of BORIS mRNA transcripts and BORIS *sfl* subfamily of splicing variants. **c** Relative abundance of BORIS protein (~70 kDa) using HeLa cell line as a positive control ($n = 3$). **d** Western blot densitometry of BORIS KD efficiency after siRNA transfection versus negative Ctrl siRNA. Lower boxes show BORIS and GAPDH representative western blot bands. Three biological replicates were performed. **e** Densitometry of BORIS KO efficiency after CRISPR/Cas9 plasmid transient transfection versus non-targeting Ctrl CRISPR/Cas9. Lower panel shows BORIS and GAPDH representative western blot bands. Three biological replicates were performed. KD knockdown, KO knockout, Ctrl control

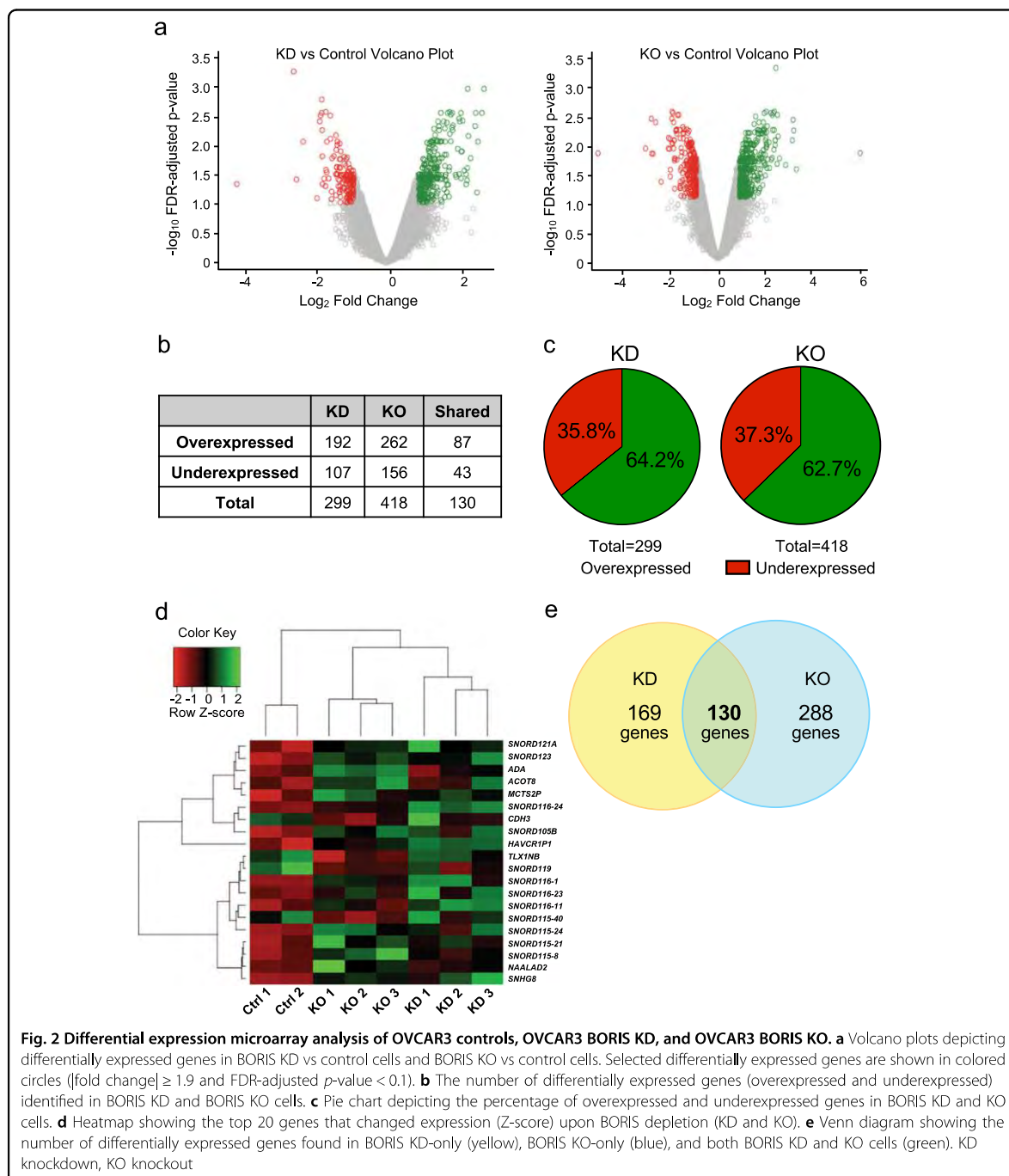
for example, it is highly expressed in testicular germ cells, and has very low levels in other somatic tissues^{8,9}. Remarkably, BORIS expression is reactivated in several neoplasms, such as lung, breast, prostate, and OC,^{5,10–12} where it participates in different cellular processes, such as cell proliferation and apoptosis^{13,14}. Given the above, BORIS was proposed by the NCI as one of the most promising cancer antigens¹⁵.

Furthermore, BORIS shows high expression levels in OC patients and is strongly associated with poor prognosis, which suggests that BORIS, or the molecular network it controls, have a role in OC progression^{16,17}. To address this question, great efforts have been made

to understand the regulatory function of BORIS in gene expression; for example, in OC cells, BORIS can act as a transcriptional activator of the *hTERT* gene, known for its important contribution in cell immortalization¹⁸. However, the function of BORIS in OC is still not completely understood. There are no genome-wide unbiased studies of BORIS activity to understand the effect and the extent to which this protein influences transcription and oncogenic processes in OC. Thus, the identification and understanding of the regulatory function of BORIS could help to predict novel drug targets or to use it as a prognostic biomarker in OC patients.

Hence, in this study we aimed to investigate the regulatory role of BORIS in OC from a genome-wide perspective using gene expression microarrays and bioinformatics analyses in OC cell lines and patients' samples. In vitro, we identified that BORIS negatively

regulates the androgen receptor (*AR*) gene, as well as fibronectin 1 (*FNI*), family with sequence similarity 129-member A (*FAM129A*) and CD97 antigen (*CD97*) genes, which are commonly deregulated in OC patients and associated with poor prognosis. In addition, we evaluated



the expression patterns of *CTCF* (BORIS) and their related targets genes on publicly available OC patient data sets, which were subsequently validated in an independent set of samples obtained from SOC, BOC patients, and non neoplastic fresh tissue.

Results

Characterization of BORIS expression in OC cell lines and BORIS-deficient cells

Given that *CTCF* (BORIS) has been found to be mainly deregulated in OC samples compared with other types of neoplasms (Supplementary Fig. S1), the aim of our study was to explore the participation of BORIS in gene regulation in OC. This work was addressed by two different experimental approaches: in vitro using OC cell lines, molecular biology techniques, and bioinformatics tools; and an ex vivo approach to evaluate the clinical relevance of BORIS in samples from OC patients, publicly available in GEO database and in fresh tissue samples from OC (Fig. 1a).

As a first approach, we evaluated the transcript and protein levels of BORIS in a set of three OC-derived cell lines (OVCAR3, SKOV3, and SW626), to select the most suitable cell model to carry out the following experiments (Fig. 1b). The expression of BORIS had been previously reported in HeLa, therefore we used this cell line as a positive control¹⁹.

We evaluated all BORIS transcripts reported using oligonucleotides located in a region shared by all its splicing variants. Also, a second pair of primers was used to amplify only the BORIS subfamily 1 of splicing variants (*sfl*), because it encodes for the canonic protein²⁰. We observed that OVCAR3 and SKOV3 cell lines have higher levels of BORIS expression compared with HeLa. In addition, the expression levels of *sfl* is higher in OVCAR3 cells than in HeLa and the other OC-derived cells (Fig. 1b). In addition, we evaluated by western blot the abundance of BORIS in the different cell lines compared to HeLa (Fig. 1c). We did not observe a significant difference in BORIS among cell lines.

Considering that we were interested in evaluating the effect of the absence of BORIS expression and its association with OC, we chose OVCAR3 for subsequent experiments.

To evaluate the expression profiles related to BORIS, we obtained two cellular models with decreased levels: one by siRNAs transfection (KD), and the second with the CRISPR/Cas9 system (KO). Then, we compared the endogenous protein levels (controls) with the KD and KO cells. The efficiencies of the decrease in KD cells were 40–90% (Fig. 1d), while the KO cells exhibit 50% decrease (Fig. 1e).

Expression profile analysis in the KD and KO cells

The first goal of this work was to identify novel genes regulated by BORIS in OC; thus, the expression profiles of the KD and KO cells versus controls were obtained with

expression microarrays. Differentially expressed genes (DEG) in KD/controls and KO/controls were selected with FDR p -value < 0.1 and |fold change| > 1.9 (Fig. 2a). We observed that the number of DEG for the KD cells was 299 (192 overexpressed and 107 underexpressed; Supplementary Table 1) and 418 (262 overexpressed and 156 underexpressed; Supplementary Table 2) in the KO cells. GO-term enrichment analysis results was performed for both conditions (Supplementary Figs. S2, S3). In addition, we detected that 130 genes were consistent in both cellular models, where 87 genes were overexpressed and 43 genes underexpressed (Fig. 2b). In general, we identified that the decrease of BORIS in both cellular models (KD and KO) corresponds to an overexpression of 63% and a decrease of 36% of DEG, suggesting that BORIS may act mainly as a transcriptional repressor (Fig. 2c). Surprisingly, we observed that a group of noncoding genes (SNORDs) showed major expression changes after KD and KO (Fig. 2d).

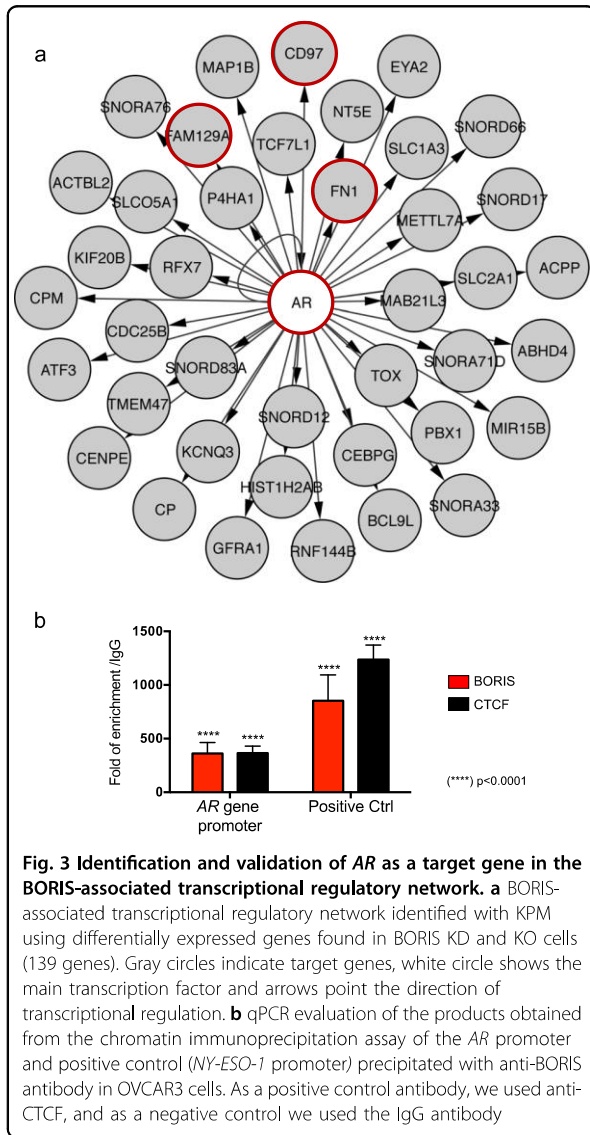
With the DEG from the KD and KO cells, we select only the genes found in both experimental strategies. This was done to exclude those genes that could be a result of the experimental technique, and not because of BORIS decrease. We identified 130 DEG in both cellular models and the following studies were based on these genes (Fig. 2b, e).

Identification of BORIS-associated network in the KD and KO cells

To identify a BORIS-associated regulatory network that provides us of potential novel BORIS targets in OC, we used the 130 DEG found in the KD and KO cells, and performed a *de novo* pathway enrichment analysis with KeyPathwayMiner (KPM)²¹. Given a biological network and set of expression studies, KPM extracts subnetworks enriched with differentially expressed genes (DEG). This method allowed us to discover previously uncharacterized regulatory networks by extracting a BORIS-associated network from HTRIdb²², a large experimentally validated human gene regulatory interaction database.

Our results show that the *AR* gene changes its expression upon BORIS depletion and is the main transcription factor that regulates the network (Fig. 3a). In addition, genes previously related to OC, such as *FNI*, *FAM129A*, and *CD97*, were identified as *AR* targets. Nevertheless, it was unclear whether BORIS regulates *AR* directly or indirectly. Thus, we performed a chromatin immunoprecipitation assay (ChIP) with an anti-BORIS antibody in the *AR* promoter. We observed that BORIS is located at the *AR* gene promoter (Fig. 3b).

Finally, to validate the BORIS-associated network, we proceeded to evaluate the expression changes by qRT-PCR in the KD and KO cells. Our results show that, indeed, *AR*, *FNI*, *FAM129A*, and *CD97* genes change their expression levels after BORIS decrease (Figs. 4a, b). In



accordance with our findings, expression data from The Cancer Cell Line Encyclopedia show that ~82% of the OC-derived cell lines reproduce the same phenomenon for *FN1*, *FAM129A*, and *CD97* genes (Supplementary Fig. S4). The latter proposes that BORIS acts as a transcriptional repressor of *AR*, *FN1*, *FAM129A*, and *CD97* genes, not only in OVCAR3 but also in a larger set of OC-derived cell lines, suggesting that this phenomenon might be related to OC biology.

Expression analysis of CTCFL (BORIS), AR, FN1, and FAM129A and CD97 genes in SOC, BOC, and non neoplastic samples

Previously, we identified that high levels of *AR*, *FN1*, *FAM129A*, and *CD97* transcripts are associated to a

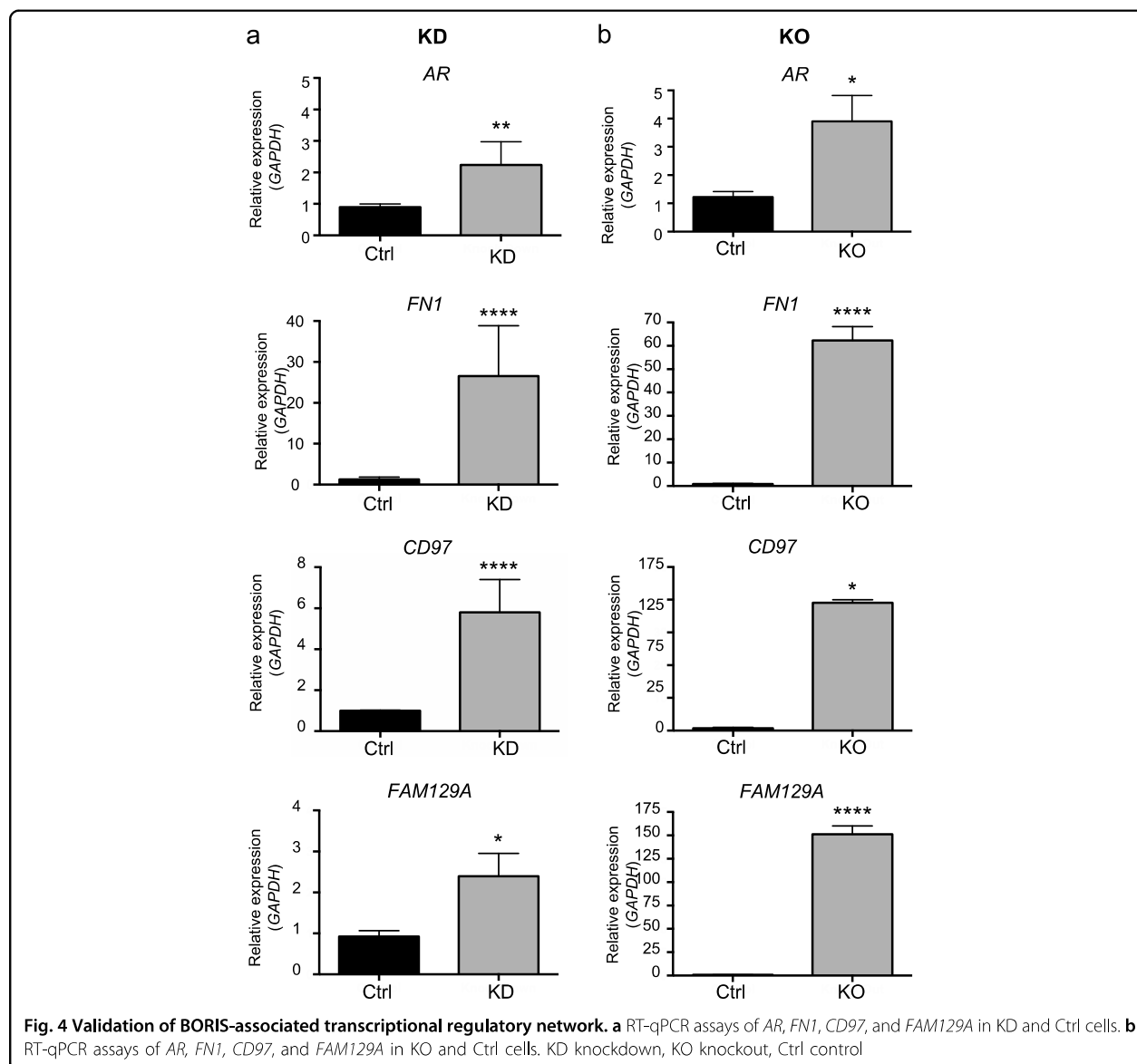
decrease in BORIS expression in OC cell lines. Thus, to determine if this expression pattern is similar in OC samples, we analyzed 343 expression microarrays from the GEO database (43 non neoplastic, 288 SOC, and 12 BOC samples).

We extracted the expression values of *BORIS*, *AR*, *FN1*, *FAM129A*, and *CD97* genes and plotted heatmaps for non neoplastic, SOC, and BOC samples (Fig. 5a). Our data reveals a wide heterogeneity among normal and cancer samples. First, we observed two main dendrogram branches in SOC samples: BORIS positive (red) and BORIS negative (green); in contrast to non neoplastic and BOC samples, which show low levels of BORIS. An important subset of SOC patients exhibits high levels of BORIS (42–47%; Fig. 5a). Furthermore, in this group of patients, we can also find different expression combinations with *AR*, *FN1*, *FAM129A*, and *CD97*; however, an important group of samples with high levels of BORIS also show low levels of *AR* (28% of samples) along with low levels of *FN1* (13%; Fig. 5b), in agreement with our in vitro findings, suggesting that both genes might be playing an important role in SOC.

Furthermore, due to the incidence of SOC patients, we aimed to classify this type of OC into clusters based on the expression patterns of these genes (Fig. 5a, c). We identified a clear expression pattern in clusters number 2 and 6, which correspond to the phenomenon previously characterized in vitro in OVCAR3 cell line (Fig. 3), where the presence of BORIS is associated with low levels of *AR*, *FN1*, *FAM129A*, and *CD97* (Cluster 2), while the absence of BORIS has the opposite behavior (Cluster 6; Supplementary Figs. S5, S6).

In addition, we evaluated *BORIS*, *AR*, *FN1*, *CD97*, and *FAM129A* in tissue from patients' samples to assess their clinical value. For instance, we wondered if the expression of these genes could be associated with a clinical feature. Thus, we obtained the relative expression levels of these genes in 11 SOC, 10 BOC, and 8 non neoplastic patients of a second cohort. We tested the results against clinical variables, and we identified that BORIS and *AR* genes are not associated with CA-125 expression, tumor size or FIGO stage (Table 1). However, we found that BORIS was associated with malignancy; for instance, SOC patients show significant higher levels of BORIS compared with BOC and non neoplastic samples ($p = 0.017$), in agreement with the previous finding in the large cohort from GEO database (Table 1). The latter proposes BORIS as a relevant deregulated gene in OC, particularly in SOC, where a subset of patients show an inverse association between BORIS and *AR* expression, and is associated with tumor malignancy.

In summary, our results show that BORIS binds to *AR* gene promoter and acts as a transcriptional repressor of *AR*. In addition, the decrease of BORIS levels is associated

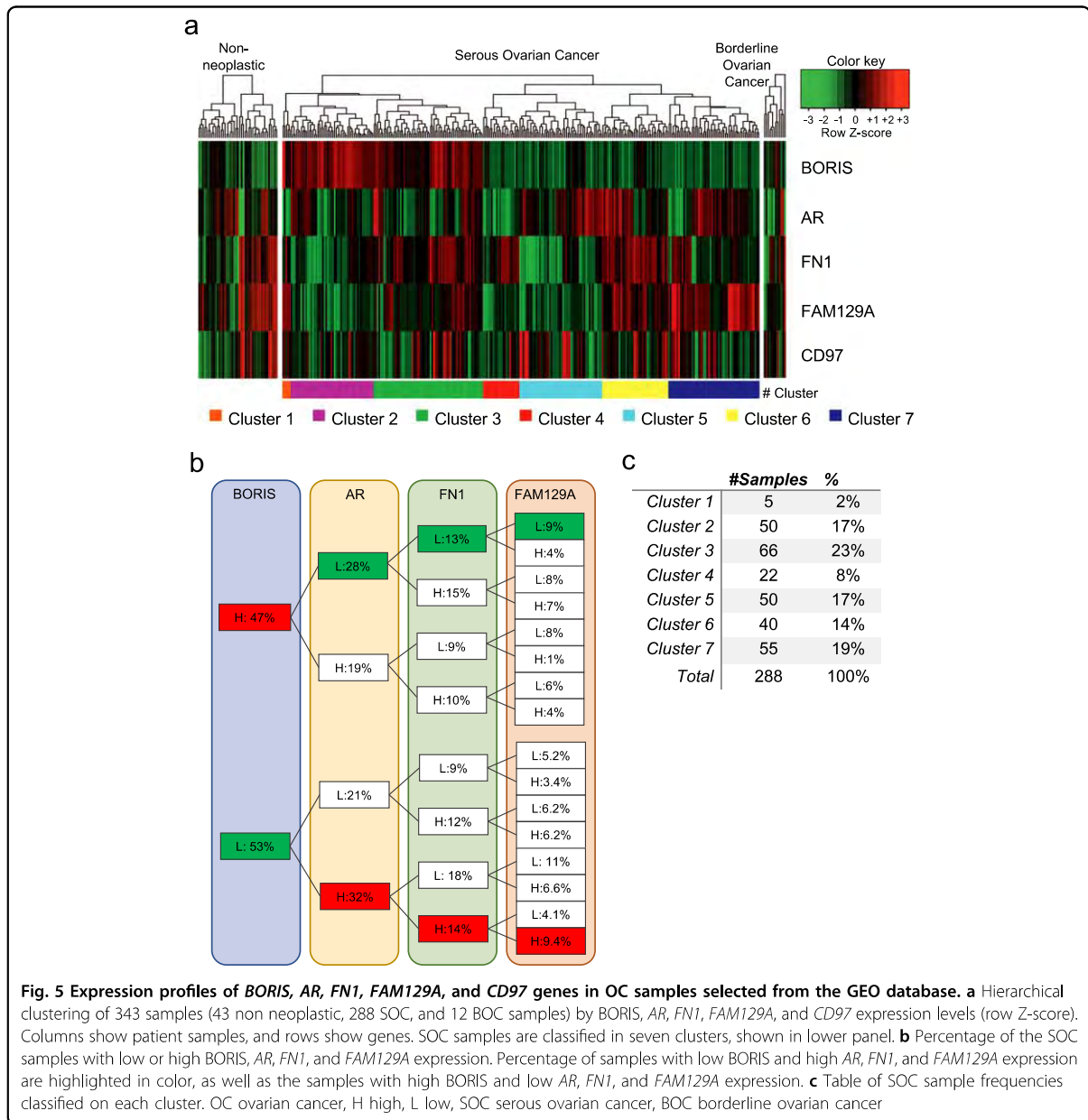


with an increase in *FN1*, *FAM129A*, and *CD97* gene expression (Fig. 6a). Furthermore, in OC patients, BORIS is overexpressed in 47% of SOC samples, and is down-regulated in borderline and non neoplastic samples, thus it is significantly associated with malignancy. Our data also indicate that a fraction of SOC samples with overexpression of BORIS (47%) can show either underexpression (28%) or overexpression (19%) of *AR*, indicating that in OC samples, the relationship between BORIS and *AR* found in OVCAR3 cell line might not necessarily happen in all OC samples (Fig. 6b). Nevertheless, we still ignore whether the different combinations of BORIS and *AR* expression levels are associated with a specific outcome, and if their joint use as prognostic biomarkers is beneficial for SOC patients. In this regard,

further studies are needed to address if BORIS and *AR* can be used together as prognostic biomarkers in OC.

Discussion

SOC is as one of the most lethal gynecologic malignancies worldwide, and we lack effective biomarkers to improve the outcome²³. Currently, only CA-125 and HE4 antigens are used in the medical practice, but they lack sensitivity²⁴. Thus, the identification of novel tumor biomarkers that could benefit patients' outcome is a priority in OC research. One of the proteins that has gained interest, given its potential use as a cancer biomarker, is BORIS^{5,6}. This protein was proposed as a priority biomarker by the National Cancer Institute, because its deregulation is related to oncogenic properties, and as a



biomarker could be used for its specificity and immunogenicity¹⁵. Previously, some studies have shown that the decrease of *BORIS* is associated with decreasing cell proliferation and cell viability^{6,25}. Hence, we aimed to identify genes regulated by *BORIS*, which could provide some insights in the biology of OC. An analysis of differential expression exhibits that several small nuclear and small nucleolar RNAs are the most affected genes in the absence of *BORIS*. To understand the function of *BORIS* from a systems perspective, we applied KPM to integrate our expression studies with known gene regulatory

interactions to extract novel disease pathways and potential drug targets²⁶. Enriched with the 130 DEG, we identified and experimentally validated the *AR* gene as one of the main regulators of the *BORIS*-associated regulatory network. This type of analysis may suggest that *BORIS* could have its effect on transcriptional regulation by altering the levels of *AR* through its binding to the *AR* promoter. Our data show biological and therapeutic relevance, since *AR* has been used as a biomarker of different neoplasms, such as prostate cancer, and is an important druggable target²⁷. Because the ovary is also

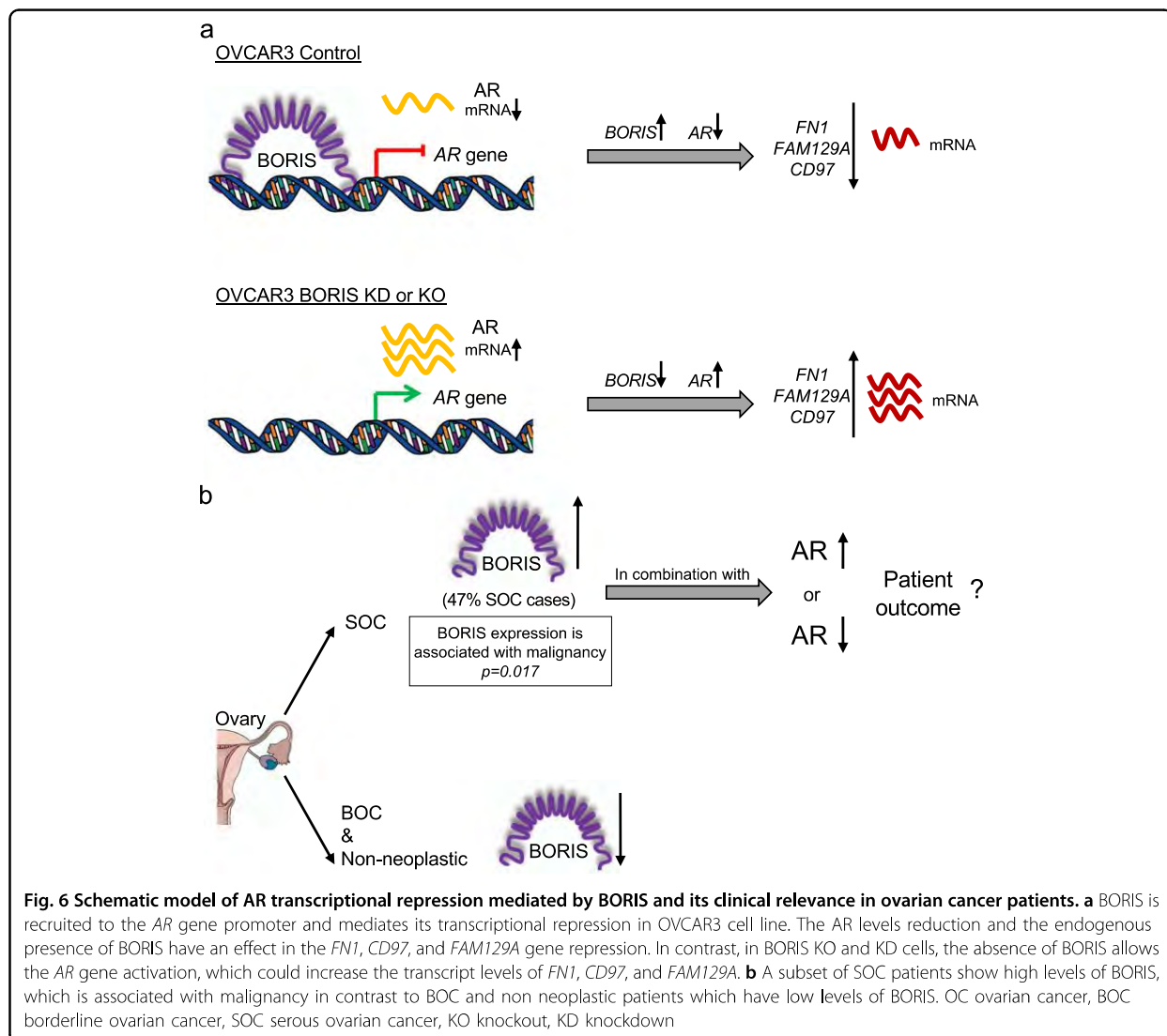
Table 1 Clinical and demographic characteristics of patients

			Expression				
			BORIS	AR	FN1	FAM129A	CD97
N =	29						
SOC	11	37.9%	p = 0.017	p = 1.000	p = 0.895	p = 0.195	p = 0.180
BOC	10	34.4%					
Non neoplastic	8	27.5%					
Age	X =	48.9 (21–81)	p = 0.793	p = 0.474	p = 0.013	p = 0.013	p = 0.483
	< 45 years	9 (31%)					
	> 45 years	20 (69%)					
Patient follow-up	38.3 months (1–80)						
Premenopause		12 (41%)	p = 0.428	p = 0.774	p = 0.876	p = 0.876	p = 0.432
Postmenopause		17 (59%)					
Number of births		2.3 (0–11)	p = 0.913	p = 0.507	p = 0.017	p = 0.017	p = 0.311
BMI	X =	26.7	p = 0.888	p = 0.663	p = 0.018	p = 0.018	p = 0.947
	SOC	27.3					
	BOC	26.5					
	Non neoplastic	26.2					
CA-125 Pre-Qx	X =	3649.1	p = 0.467	p = 0.488	p = 0.483	p = 0.586	p = 1.000
	SOC	478.2					
	BOC	9968.2					
	Non neoplastic	110.3					
Tumor size	X =	16.2	p = 0.208	p = 0.628	p = 0.812	p = 0.558	p = 0.466
	> 10 CM SOC	11 (100%)					
	> 10 CM BOC	7 (70%)					
Ascites	Yes	6 (28%)	p = 0.172	p = 0.565	p = 0.291	p = 0.457	p = 0.375
	No	15 (72%)					
Peritoneal implants	SOC	6 (54%)	p = 0.432	p = 0.183	p = 0.129	p = 0.382	p = 0.192
	BOC	2 (20%)					
FIGO	SOC	BOC					
I	0	5 (50%)	p = 0.510	p = 1.000	p = 0.523	p = 0.171	p = 0.142
II	3 (27%)	1 (10%)					
III	8 (73%)	4 (40%)					
IV	0	0					
Optimal cytoreduction	SOC	6 (54%)	p = 0.329	p = 0.236	p = 0.186	p = 0.112	p = 0.498
	BOC	100%					
Mortality	SOC	27%	p = 0.060	p = 0.420	p = 0.527	p = 0.636	p = 0.220
	BOC	0%					
	Non neoplastic	0%					

p < 0.05 is statistically significant

responsible for the synthesis of AR, it would be of great importance the search for biomarkers related to this protein^{28–30}. Interestingly, our results from KPM demonstrated that BORIS is partly responsible for the

regulation of AR gene expression, and in turn, AR establishes a gene regulatory network with genes such as FN1, CD97, and FAM129A, which are related with poor prognosis, chemoresistance, and metastasis in several cancers,



including OC^{31–33}. Previous studies are limited to the identification of possible target genes by evaluating changes in gene expression by increasing or decreasing BORIS levels;^{5,6,11,19,34} however, our *de novo* pathway enrichment analysis allowed us to perform a novel selection (based on hundreds of experimental data available in public databases) of potential genes relevant in cancer, which are closely related through AR and BORIS. In conclusion, we found that in both, BORIS KD and KO models, BORIS regulates several genes, functioning mainly as a transcriptional repressor. Remarkably, BORIS acts as a transcriptional repressor of AR gene and binds to its promoter. The latter is of great relevance given the importance of AR deregulation in the development and malignancy of many types of cancer, also due to the current use of AR gene as a prognostic biomarker and

therapy target in cancer, such as prostate cancer. In addition, our findings show that BORIS can also regulate other genes involved in OC, such as *FN1*, *FAM129A*, and *CD97*. Finally, we found a subset of SOC patients that show BORIS overexpression, which is significantly associated with malignancy.

Collectively, our *in vitro* and *ex vivo* studies confirm that BORIS has the potential to be used as a prognostic biomarker in SOC and to yield novel druggable targets among the regulatory network that BORIS controls.

Materials and methods

Cell lines culture

The cancer cell lines OVCAR3, SKOV3, SW626, and HeLa were obtained from ATCC (Manassas, VA, USA), and subcultured as described by the supplier.

Knockdown and Knockout assays

To knockdown (KD) expression of BORIS, OVCAR3 cells were transfected with a pool of small interfering RNAs (siRNAs) (Cat. A-003819-100-0005, Dharmacon, Lafayette, CO, USA) or non-targeting control siRNA (Cat. D-001910-10-10, Dharmacon) and were incubated for 48 h in Accell Delivery Media (Cat. B-005000, Dharmacon). The assays were performed by triplicate.

Knockout (KO) of BORIS expression was performed by transfection of OVCAR3 cells with 2 µg of BORIS (Cat. sc-403313, SCBT, Dallas, TX, USA) or control CRISPR/Cas9 (Cat. Sc-418922, SCBT) plasmids. The assays were performed by triplicate. Following experiments were performed 24 h post transfection.

Western blot assays

Western blot assays were performed with 30 µg of total protein. Primary and secondary antibodies used are described in Supplementary Table 3. Canonic BORIS abundance was determined by standard densitometry analysis, using ImageJ software (NIH, USA) with GAPDH as normalizing protein.

RNA extraction, cDNA, and qRT-PCR analysis

The total RNA was extracted with TRIzol Reagent (Cat. 15596026, Invitrogen, Carlsbad, CA, United States), and the integrity and quality were analyzed with TapeStation 2200. cDNA was obtained with the GeneAmp PCR Core Kit (Cat. N8080143, Thermo Scientific, Waltham, MA, USA) using oligo(dT). qRT-PCR assays were performed on StepOnePlus using SYBR Green Master Mix (Cat. 4309155, Applied Biosystems, Foster City, CA, USA). All data were normalized to *GAPDH* using the $\Delta\Delta C_t$ method. Primers used are described in Supplementary Table 4. Experiments were performed by triplicate.

Microarray analysis: hybridization and analysis

The GeneChip Human Gene 2.0 ST oligonucleotide arrays (Affymetrix, Santa Clara, CA, USA) were hybridized with two control samples of OVCAR3, three samples of OVCAR3 KD, and three samples of OVCAR3 KO cells, according to the instructions provided by the manufacturer.

Data normalization was performed with robust multi-array analysis (RMA). DEG were selected with “Limma” package³⁵ and those with |fold change| > 1.9 and FDR-adjusted *p*-value < 0.1 were selected for further downstream analysis. *De novo* pathway enrichment analysis was performed with KPM²¹ with the Greedy search algorithm and the INES search strategy. We used as input of KPM, a regulatory network constructed from 51871 interactions found in the human transcriptional regulation interaction database (HTRIdb)²². Microarrays data have been deposited at the National Center for Biotechnology

Information Gene Expression Omnibus (NCBI GEO Series Accession # GSE130163).

Chromatin immunoprecipitation assays (ChIP)

OVCAR3 cells were cultured at 80% of confluence, chromatin was extracted according to the OneDay ChIP Kit (Cat. kch-oneDIP-180, Diagenode, Denville, NJ, USA). Two independent chromatin preparations were analyzed. As a negative control, we used an IgG antibody provided by the kit. The antibodies and primers used are listed in Supplementary Tables 3 and 4.

Patients samples

The ovarian tumor samples were collected from patients undergoing surgical resections at the INCan (Mexico City, Mexico) with previous written consent and the approval of the ethical committee (approval number 015/037/ICI). We analyzed a total of 21 samples from patients diagnosed with OC (11 SOC, 10 BOC), and eight non neoplastic ovarian samples obtained from January 1st, 2014 to December 31st, 2016. Demographics, prognostic markers, and epidemiologic exposure variables were obtained from the medical record.

RNA extraction from patient samples

The tissue from patients was stored in RNAlater RNA Stabilization Reagent (Cat. AM7020, Invitrogen) at -20° C. RNA from samples was obtained according to Peña-Llopis et al.³⁶.

GEO expression microarray analysis

Gene expression data from OC patients, evaluated by GeneChip Human Genome U133A 2.0 Plus arrays, were obtained from the GEO (Accession IDs: GSE14001, GSE18520, GSE19352, GSE36668, GSE38666, GSE63885, GSE40595 and GSE26193)^{37–45}. A total of 343 samples were analyzed (43 from non neoplastic samples, 288 were from SOC samples, and 12 BOC patients). Batch effects were corrected with ComBat function from the “sva” package⁴⁶. Then, normalized expression values were obtained for BORIS, *AR*, *FNI*, *FAM129A*, and *CD97* genes. Heatmaps were generated using the “ComplexHeatmap” package⁴⁷, clustering was performed with hierarchical clustering where the optimal number of clusters was selected using the “NbClust” package⁴⁸.

Statistical analysis

The data from cell lines are shown as the mean \pm standard error. The differences between groups were analyzed with a paired Student's *t* test comparing the samples with the controls. Differences between groups were considered statistically significant when *p* < 0.05, *p* < 0.01, *p* < 0.001, and *p* < 0.0001.

For patient samples, we performed Chi-square test and Fisher's exact test using STATA software version 13.0 software (StataCorp, TX, USA), to assess the relationship between BORIS, AR, FN1, FAM129A, and CD97 with clinicopathological characteristics. $P < 0.05$ was considered statistically significant.

Acknowledgements

M. Salgado-Albarrán is a doctoral student in the "Programa de Doctorado en Ciencias Bioquímicas, UNAM", and received a fellowship from CONACYT (CVU 659273). She also was a beneficiary of the German Academic Exchange Service (DAAD grant no. 91693321). This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT) by the Fondo Sectorial de Investigación en Salud y Seguridad Social (FOSISS, grant no. 0261181), Fondo CB-SEP-CONACYT (284748), Programa para el Desarrollo Profesional Docente (PRODEP-SEP; 47310447) and UAM-PTC-704. J.B. was supported by VILLUM Young Investigator grant no. 13154, and by H2020 project no. 777111 (RepoTrial). E.S. R. and E.A.O. are supported by the Natural Science Department at UAM Cuajimalpa Unit.

Author details

¹Natural Sciences Department, Universidad Autónoma Metropolitana-Cuajimalpa (UAM-C), Mexico City 05300, Mexico. ²Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany. ³Cancer Biomedical Research Unit, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico. ⁴The Bioinformatics Centre Section for RNA and Computational Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark. ⁵Instituto Nacional de Neurología y Neurocirugía "Manuel Velasco Suárez", Mexico City, Mexico

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41389-019-0150-2>).

Received: 6 February 2019 Revised: 8 May 2019 Accepted: 1 June 2019

Published online: 12 August 2019

References

- Sopik, V., Rosen, B., Giannakeas, V. & Narod, S. A. Why have ovarian cancer mortality rates declined? Part III. Prospects for the future. *Gynecol. Oncol.* **138**, 757–761 (2015).
- Sopik, V., Iqbal, J., Rosen, B. & Narod, S. A. Why have ovarian cancer mortality rates declined? Part I. Incidence. *Gynecol. Oncol.* **138**, 741–749 (2015).
- Sopik, V., Iqbal, J., Rosen, B. & Narod, S. A. Why have ovarian cancer mortality rates declined? Part II. Case-fatality. *Gynecol. Oncol.* **138**, 750–756 (2015).
- Salmaninejad, A. et al. Cancer/testis antigens: expression, regulation, tumor invasion, and use in immunotherapy of cancers. *Immunol. Invest.* **45**, 619–640 (2016).
- Cheema, Z. et al. Expression of the cancer-testis antigen BORIS correlates with prostate cancer. *Prostate* **74**, 164–176 (2014).
- Okabayashi, K. et al. Cancer-testis antigen BORIS is a novel prognostic marker for patients with esophageal cancer. *Cancer Sci.* **103**, 1617–1624 (2012).
- Freitas, M. R. P. et al. Expression of cancer/testis antigens is correlated with improved survival in glioblastoma. *Oncotarget* **4**, 636–646 (2013).
- GTEx Consortium TGTE. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013).
- Loukinov, D. I. et al. BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl Acad. Sci. USA* **99**, 6806–6811 (2002).
- Bhan, S. et al. BORIS binding to the promoters of cancer testis antigens, MAGEA2, MAGEA3, and MAGEA4, is associated with their transcriptional activation in lung cancer. *Clin. Cancer Res.* **17**, 4267–4276 (2011).
- D'Arcy, V. et al. BORIS, a paralogue of the transcription factor, CTCF, is aberrantly expressed in breast tumours. *Br. J. Cancer* **98**, 571–579 (2008).
- Link, P. A., Zhang, W., Odunsi, K. & Karpf, A. R. BORIS/CTCF mRNA isoform expression and epigenetic regulation in epithelial ovarian cancer. *Cancer Immun.* **13**, 6 (2013).
- Dougherty, C. J. et al. Selective apoptosis of breast cancer cells by siRNA targeting of BORIS. *Biochem Biophys. Res. Commun.* **370**, 109–112 (2008).
- Soltanian, S., Dehghani, H., Matin, M. M. & Bahrami, A. R. Expression analysis of BORIS during pluripotent, differentiated, cancerous, and non-cancerous cell states. *Acta Biochim Biophys. Sin.* **46**, 647–658 (2014).
- Cheever, M. A. et al. The prioritization of cancer antigens: a National Cancer Institute pilot project for the acceleration of translational research. *Clin. Cancer Res.* **15**, 5323–5337 (2009).
- Woloszynska-Read, A. et al. DNA methylation-dependent regulation of BORIS/CTCF expression in ovarian cancer. *Cancer Immun.* **7**, 21 (2007).
- Woloszynska-Read, A. et al. Coordinated cancer germline antigen promoter and global dna hypomethylation in ovarian cancer: association with the BORIS/CTCF expression ratio and advanced stage. *Clin. Cancer Res.* **17**, 2170–2180 (2011).
- Renaud, S. et al. BORIS/CTCF-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. *Nucleic Acids Res.* **39**, 862–873 (2011).
- Alberti, L., Renaud, S., Losi, L., Leyvraz, S. & Benhattar, J. High expression of hTERT and stemness genes in BORIS/CTCF positive cells isolated from embryonic cancer cells. *PLoS One* **9**, <https://doi.org/10.1371/journal.pone.0109921> (2014).
- Pugacheva, E. M. et al. The structural complexity of the human BORIS Gene in gametogenesis and cancer. *PLoS One* **5**, e13872 (2010).
- Alcaraz, N. et al. Robust *de novo* pathway enrichment with KeyPathwayMiner 5. *F1000Research* **5**, 1531 (2016).
- Bovolenta, L. A., Acencio, M. L. & Lemke, N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genom.* **13**, 405 (2012).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30 (2018).
- Yang, W.-L., Lu, Z. & Bast, R. C. The role of biomarkers in the management of epithelial ovarian cancer. *Expert Rev. Mol. Diagn.* **17**, 577–591 (2017).
- Alberti, L., Losi, L., Leyvraz, S. & Benhattar, J. Different effects of BORIS/CTCF on stemness gene expression, sphere formation and cell survival in epithelial cancer stem cells. *PLoS One* **10**, e0132977 (2015).
- List, M. et al. KeyPathwayMinerWeb: online multi-omics network enrichment. *Nucleic Acids Res.* **44**, W98–W104 (2016).
- Pelekanou, V. & Castanas, E. Androgen control in prostate cancer. *J. Cell Biochem* **117**, 2224–2234 (2016).
- Simitsidellis, I., Saunders, P. T. K. & Gibson, D. A. Androgens and endometrium: new insights and new targets. *Mol. Cell Endocrinol.* **465**, 48–60 (2018).
- Gomora M. J. et al. Sexual steroid hormone receptors profiles in ovarian carcinoma in Mexican women. *Endocr. Connect.*; <https://doi.org/10.1530/EC-18-0158> (2018).
- Shaw, G. L. et al. The early effects of rapid androgen deprivation on human prostate cancer. *Eur. Urol.* **70**, 214–218 (2016).
- Chang, P.-Y. et al. An epigenetic signature of adhesion molecules predicts poor prognosis of ovarian cancer patients. *Oncotarget* **8**, 53432–53449 (2017).
- Wu, W. et al. Identification of proteomic and metabolic signatures associated with chemoresistance of human epithelial ovarian cancer. *Int. J. Oncol.* **49**, 1651–1665 (2016).
- Thomas, B. C. et al. Whole blood mRNA in prostate cancer reveals a four-gene androgen regulated panel. *Endocr. Relat. Cancer* **23**, 797–812 (2016).
- Chen, K. et al. BORIS, brother of the regulator of imprinted sites, is aberrantly expressed in hepatocellular carcinoma. *Genet. Test. Mol. Biomark.* **17**, 160–165 (2013).
- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
- Peña-Llopis, S. & Brugarolas, J. Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat. Protoc.* **8**, 2240–2255 (2013).

- **PUBLICACIÓN 7:** The Regulatory Roles of Non-coding RNAs in Angiogenesis and Neovascularization From an Epigenetic Perspective.
- **AUTORES:** Hernández-Romero IA, Guerra-Calderas L, Salgado-Albarrán M, Maldonado-Huerta T & Soto-Reyes E
- **REVISTA:** *Frontiers in Oncology*. **VOLÚMEN:** 9. **PÁGINAS:** 1091. **AÑO:** 2019.



The Regulatory Roles of Non-coding RNAs in Angiogenesis and Neovascularization From an Epigenetic Perspective

Itzel Alejandra Hernández-Romero[†], Lissania Guerra-Calderas[†],
Marisol Salgado-Albarrán, Tatiana Maldonado-Huerta and Ernesto Soto-Reyes*

Natural Sciences Department, Universidad Autónoma Metropolitana-Cuajimalpa, Mexico City, Mexico

OPEN ACCESS

Edited by:

Erika Ruiz-García,
National Institute of Cancerology
(INCan), Mexico

Reviewed by:

Fahd Al-Mulla,
Genatak, Kuwait
Shao-Chun Wang,
China Medical University, Taiwan

*Correspondence:

Ernesto Soto-Reyes
esotoreyes@correo.cua.uam.mx

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Molecular and Cellular Oncology,
a section of the journal
Frontiers in Oncology

Received: 20 June 2019

Accepted: 03 October 2019

Published: 24 October 2019

Citation:

Hernández-Romero IA,
Guerra-Calderas L,
Salgado-Albarrán M,
Maldonado-Huerta T and
Soto-Reyes E (2019) The Regulatory
Roles of Non-coding RNAs in
Angiogenesis and Neovascularization
From an Epigenetic Perspective.
Front. Oncol. 9:1091.
doi: 10.3389/fonc.2019.01091

Angiogenesis is a crucial process for organ morphogenesis and growth during development, and it is especially relevant during the repair of wounded tissue in adults. It is coordinated by an equilibrium of pro- and anti-angiogenic factors; nevertheless, when affected, it promotes several diseases. Lately, a growing body of evidence is indicating that non-coding RNAs (ncRNAs), such as miRNAs, circRNAs, and lncRNAs, play critical roles in angiogenesis. These ncRNAs can act *in cis* or *trans* and alter gene transcription by several mechanisms including epigenetic processes. In the following pages, we will discuss the functions of ncRNAs in the regulation of angiogenesis and neovascularization, both in normal and disease contexts, from an epigenetic perspective. Additionally, we will describe the contribution of Next-Generation Sequencing (NGS) techniques to the discovery and understanding of the role of ncRNAs in angiogenesis.

Keywords: angiogenesis, non-coding RNA, epigenetics, neovascularization, next generation sequencing, miRNAs, lncRNAs, circRNA

INTRODUCTION

In the vascular network, blood vessels act as channels for nutrients, oxygen delivery, and metabolic waste evacuation. The growth of new capillary vessels, known as angiogenesis, plays key roles in embryonic development and in tissue homeostasis and remodeling in adults, as well as in cancer initiation and progression (1, 2). The balance between pro- and anti-angiogenic factors (such as VEGF, PDGF, and TSP-1/2) coordinates angiogenesis and other neovascularization mechanisms such as intussusceptive angiogenesis, vasculogenesis, lymphangiogenesis, vessel co-option, and vasculogenic mimicry (3–5).

Over the last few decades, the study of angiogenesis has helped researchers to understand vascular physiology and its implications for several diseases. For instance, in atherosclerosis, ischemia, and retinopathy, excessive or insufficient vascular growth can affect the behavior of endothelial and smooth muscle cells (6, 7). Studies of the neovascularization processes have also provided molecular targets for the development of therapies to delay cancer progression, since it is well-known that angiogenesis is an essential process that is altered in tumors (8).

Nowadays, the study of the molecular mechanisms involved in angiogenesis is being built on different experimental approaches, such as cell migration, proliferation, and metabolic assays or histological and tri-dimensional models, that approach specific stages of angiogenesis; however, only pieces of the puzzle have been elucidated (9). With advances in high-throughput genomic

technologies such as microarrays, next-generation sequencing (NGS), and bioinformatic analyses, a genome-wide perspective of the elements involved in the angiogenic process is now being taken. Some of the newest players revealed by these approaches are non-coding RNAs (ncRNAs), which have gained relevance in the field of epigenetics (10–12). Therefore, in this review, we will describe the epigenetic regulatory functions of ncRNAs in physiological angiogenesis and vascular diseases, as well as the contribution of NGS technologies to the discovery of new roles for ncRNAs that are associated with angiogenesis.

AN OVERVIEW OF EPIGENETICS

In 1939, the term “epigenetics” was coined by Conrad Hal Waddington (13). Today, one of the most accepted definitions of the term explains that “epigenetics is the study of the heritable changes in gene expression that cannot be explained by alterations in the DNA sequence” (14). Among the epigenetic components that coordinate nucleus organization and gene transcription are DNA methylation, histone post-translational modifications (PTMs), and histone positioning, but recently, ncRNAs have been incorporated as epigenetic modifiers, because many of these can function as scaffolding elements to transport proteins with epigenetic functions (15). Each of these processes is stimulated by the signals derived from a dynamic epigenetic code that is established on the chromatin depending on the physiological and extracellular context. The writers, readers, and erasers of this code are proteins that place, recognize, or remove chemical modifications of DNA nucleotides and within the amino-terminal regions of histones. Most chromatin “writers” are methyltransferases that catalyze the transfer of methyl groups. DNA methylation occurs predominantly in regions enriched in CpG sites. The occurrence of methylation at the promoter regions of genes is associated with gene silencing. PTMs alter the regulation of gene transcription by changing the structure of chromatin depending on the particular residue that is modified (16, 17). The “readers” are proteins that recognize and associate with the epigenetic modifications, interpret them, and, in many cases, promote the assembly of protein complexes. The erasers remove the modifications and, therefore, alter signaling

Abbreviations: BDNF, Brain-Derived Neurotrophic Factor; BRG1, Brahma related gene-1; CAD, Coronary Artery Disease; circRNA, circular RNAs; DEGs, Differentially Expressed Genes; DNMT, DNA methyltransferase; EICiRNAs, Exon-intron circular RNAs; EPCs, Endothelial Progenitor Cells; EZH2, Enhancer of Zeste Homolog 2; GRO-Seq, Global run-on sequencing; HDAC, Histone deacetylase; HDL, High-density lipoprotein; HF, Heart Failure; HIF1, Hypoxia Inducible Factors 1; HUVEC, Human Umbilical Vein Endothelial Cells; IH, Infantile hemangioma; lincRNAs, intergenic lincRNAs; lncRNAs, long non-coding RNAs; LOXL2, Lysyl oxidase-like 2; MeCP2, Methyl-CpG-binding protein 2; miRNAs, microRNAs; MMP, Matrix metalloproteinase; mRNA, messenger RNA; ncRNAs, non-coding RNAs; NGS, Next-Generation Sequencing; PB-EPCs, Peripheral Blood EPCs; PRC, Polycomb Repressive Complex; pre-miRNA, precursor hairpin miRNA; REST, Repressor Element-1 Silencing Transcription; RNA-seq, RNA sequencing; SIRT1, NAD-dependent deacetylase sirtuin1; smRNA-seq, small RNA-seq; SNPs, Single Nucleotide Polymorphisms; SUZ12, Suppressor of Zeste 12 Protein Homolog 2; TF, Transcription Factor; TGF- β , Transforming Growth Factor; TSS, Transcription Start Sites; UC-EPCs, Umbilical Cord EPCs; UHRF1, E3 ubiquitin ligase with PHD and RING finger domain 1; VASH1, Angiogenesis inhibitor vasohibin 1; VEGF, Vascular Endothelial Growth Factor.

components that contribute to the regulation of gene expression. Recently, it has been reported that ncRNAs can mediate the binding of epigenetic proteins to their target sequences. Though they do not function alone as “classic” epigenetic modifiers, they play a vital role in both the recruitment and transcriptional regulation of epigenetic modifiers (18). In fact, multiple chromatin-remodeling enzymes have been shown to directly contact ncRNAs, including Enhancer of Zeste Homolog 2 (EZH2) and Suppressor of Zeste 12 Protein Homolog (SUZ12) (writer and eraser within the Polycomb repressive complex 2/PRC2, respectively), and nuclear architectural proteins like Yin Yang 1 and CTCF, among others (19–22). The incorporation of ncRNAs as epigenetic elements has opened up new fields of study in which they have been shown to regulate gene expression. In the following pages, we will provide an overview of the ncRNAs involved in angiogenesis, focusing on those involved in epigenetic processes.

MiRNAs AND THEIR EPIGENETIC TARGETS IN NEOVASCULARIZATION AND ANGIOGENIC PROCESSES

MicroRNAs (miRNAs) are short ncRNAs with a length of 19–23 nucleotides that are conserved in animals, plants, and some viruses (23–25). MiRNAs are transcribed as long pri-microRNAs (pri-miRNA) and are subsequently processed to ~70-nucleotide precursor hairpins (pre-miRNA) by the RNase Drosha (26). Pre-miRNAs are then exported to the cytoplasm and recognized by the RNase DICER, which removes the loop linking the 3' and 5' ends of the hairpin, producing a ~20-nucleotide miRNA duplex (27). Later, one of these strands is fused into the RNA Induced Silencing Complex (RISC), where both the miRNA and its messenger RNA (mRNA) target interact (28).

MiRNAs have two main functions: post-transcriptional gene regulation and RNA silencing. They act by pairing bases with a complementary sequence located in the 3'UTR region of target mRNA (29, 30). Consequently, these mRNAs are regulated by one or more mechanisms that include the inhibition of mRNA translation to proteins by ribosomes and by mRNA strand cleavage into two fragments and poly(A) tail shortening that results in mRNA disruption (29, 31). In the last 10 years, the field of miRNA biology has ignited, revealing amazing functions in angiogenesis. These miRNAs have been termed angiomiRs, and they target key angiogenesis molecular drivers, such as metalloproteinases, hypoxia inducible factor 1 (HIF1), cytokines, and growth factors, such as EGFL7, FGF11, PDGFRB, and the vascular endothelial growth factor (VEGF) family (32–34).

MiRNAs are not considered epigenetic components, but some of them are modulated by epigenetic mechanisms. This mainly affects their regulatory region through the incorporation of DNA methylation, repressive histone marks, or the loss of transcriptional factors, as has been reported for *miR-125b1* and *miR-124* (35, 36). Others, known as Epi-miRNAs, can also regulate the gene expression of epigenetic elements, DNA methyltransferases (DNMTs) (such as *miR-152*, *miR-30*, and *miR-148a/b*), histone deacetylases (HDACs) (such as *miR-140*,

miR-1, and *miR-449a*), and the Polycomb Group of genes (such as *miR-101* and *miR-26a*) (37–44), and some of them have been considered angiomiRs (39, 40). MiRNAs and their identified epigenetic targets in angiogenesis are listed in **Table 1**.

MiR-30a-3p

Transforming Growth Factor (TGF- β) is a relevant cytokine that functions in the process of vascular homeostasis and is involved in the vascular development of endothelial cells. It has been reported that the administration of TGF- β to endothelial cells leads to decreased *miR-30a-3p* expression. The absence of this microRNA results in increased levels of methyl-CpG-binding protein 2 (MeCP2), a protein associated with silencing of *SIRT1* (45). *SIRT1* is necessary for the migration of endothelial cells to occur throughout sprouting angiogenesis, and the loss of this enzyme induces abnormal angiogenesis *in vivo* (52). Conversely, increased levels of *miR-30a-3p* expression lead to the activation of *SIRT1* expression (**Figure 1A**). Further experiments revealed that MeCP2 enhanced the methylation status of the *SIRT1* promoter, probably by DNMT1 recruitment, leading to a reduction in *SIRT1* expression and endothelial angiogenic defects (53).

MiR-101

The microRNA *miR-101* acts as a tumor suppressor, promoting apoptosis and inhibiting cell proliferation, angiogenesis, invasion, and metastasis. *MiR-101* performs its regulatory functions by targeting an abundant range of epigenetic molecular effectors, such as *DNMT3A*, *EZH2*, and *HDAC9* (54, 55). In endothelial cells, high levels of VEGF are associated with the downregulation of *miR-101*, allowing an increase in *EZH2* (46). *EZH2* is associated with the formation of heterochromatin and can affect multiple target genes such as Vasohibin 1 (*VASH1*), which functions as a negative feedback modulator of angiogenesis in vascular endothelial cells (56, 57) (**Figure 1B**). The overexpression *miR-101* leads to *EZH2* repression and the activation of *VASH1* transcription. This evidence, taken together, suggests that *miR-101* is involved in multiple processes such as cellular growth attenuation, migration, and invasion mechanisms and the ability of endothelial cells to form capillary-like structures in glioblastomas (47).

MiR-20a

MiR-20a belongs to the *miR-17-92* cluster and has been linked to breast cancer cells with a high angiogenic profile. High levels of *miR-20a* are correlated with complex vascular structures and larger vessels, suggesting that *miR-20a* could be used as a potential new angiogenic target (58). Additionally, overexpression of *miR-20a* affects the mRNA stability of the lysine acetyltransferase, p300. In mouse myocardium cells, p300 is a key factor that regulates angiogenic and hypertrophic programs, influencing the expression of many related genes, such as *Hif1*, *Vegfc*, *Vegfa*, *Angpt1*, and *Egln3*. Interestingly, high p300 levels induce an increase in the expression of *miR-20a*, providing a feedback inhibition loop for p300 that prevents its pro-angiogenic effects (48).

MiR-137

MiR-137 has a tumor suppressor gene function that has been reported for several neoplasms (49, 59, 60). It was also reported that this miRNA can inhibit angiogenesis and cell proliferation by *EZH2* downregulation in glioblastomas. Overexpression of *miR-137* reduces the mRNA and protein levels of *EZH2*, while downregulation of *miR-137* is associated with poor prognosis in affected patients (49).

MiR-124

The miRNA *miR-124* is highly conserved, from nematodes to humans. Three human genes encoding *miR-124* have previously been characterized (*miR124a-1*, *miR-124a-2*, and *miR-124a-3*) and the majority have been shown to be deregulated in neoplasms (61). Also, it has been shown that expression of *miR-124* is elevated after treatment with certain drugs such as niclosamide. In this case, it is associated with the inhibition of vasculogenic mimicry formation, particularly by reducing levels of phosphorylated STAT3 (62).

Some reports propose that *miR-124* suppresses the E3 ubiquitin ligase with PHD and RING finger domain 1 (*UHRF1*) expression, a factor involved in the recruitment of epigenetic components in bladder cancer tissues. Also, *UHRF1* is known to enhance malignancy, inducing cellular proliferation, migration, and angiogenesis (63). *MiR-124* overexpression resulted in *UHRF1* suppression through the competitive binding of its 3'-UTR region. In addition, *miR-124* overexpression attenuated tumor growth and cell proliferation *in vivo* and invasion, migration, and vasculogenic mimicry *in vitro*. Further, it reduced VEGF protein levels and levels of the matrix metalloproteinases MMP-2 and MMP-9. A matrigel assay in a three-dimensional culture revealed reductions in tubular channel formation when *miR-124* was over-expressed in bladder cancer cell lines compared to the control group, suggesting that *miR-124* indirectly regulates vasculogenic mimicry in bladder cancer (44).

MiR-214

Originating from intron 14 of the *Dynamin-3* gene (*DNM3*), the primary transcript of *miR-214* produces four different miRNAs (*miR-199-3p*, *miR-199-5p*, *miR-214-3p*, and *miR-214-5p*) (64). During the endothelial differentiation of embryonic stem cells, the Brain-Derived Neurotrophic Factor (BDNF) promotes angiogenesis, *in vitro* and *in vivo*, by increasing levels of *miR-214*. The *miR-214* inhibits *EZH2* at the post-transcriptional level, leading to reductions in *EZH2* occupancy at the *NOS3* promoter (50). Also, *miR-214* controls the BDNF-mediated upregulation of neuropilin 1, VEGF-R, and Crk-associated substrate kinase (50, 65). Thus, *miR-214* is a downstream player within the BDNF signaling pathway that regulates important angiogenic targets.

MiR-200b

miR-200b is part of the miR-200 family, which is organized into two main groups according to seed sequence. The miRNAs of group A are *miR-141* and *miR-200a*, while the miRNAs in group B are *miR-200b*, *miR-200c*, and *miR-429* (66). Particularly, *miR-200b* has been indicated to have a role in the process of angiogenesis. Studies of malignant neoplasms demonstrated that

TABLE 1 | Summary of MicroRNAs and their epigenetic targets in angiogenesis and vascular disease.

Common name	Function	Mechanism	Model or disease	References
MiR-30a-3p	Required for endothelial cell migration during sprouting angiogenesis	Base-pairing with matching sequences within <i>MeCP2</i> mRNA	Human umbilical vein endothelial cells	(45)
MiR-101	Inhibits cellular proliferation, migration, invasion and attenuates formation of capillary-like structures	Base-pairing with matching sequences within <i>EZH2</i> mRNA	Human brain microvascular endothelial cells and glioblastoma	(46, 47)
MiR20-a	Inhibits angiogenic and hypertrophic programs	Base-pairing with matching sequences within <i>p300</i> mRNA	Mouse myocardium	(48)
MiR-137	Inhibits cellular proliferation and angiogenesis	Base-pairing with matching sequences within <i>EZH2</i> mRNA	Glioblastoma and xenografts of severe combined immunodeficiency mice	(49)
MiR-124	Inhibits cellular proliferation, migration, invasion and formation of capillary-like structures	Base-pairing with matching sequences within <i>UHRF1</i> mRNA	Bladder cancer	(44)
MiR-214	Promote angiogenesis and endothelial differentiation	Base-pairing with matching sequences within <i>EZH2</i> mRNA	Embryonic stem cells	(50)
MiR-200b	Inhibits the formation of capillary-like structures	Posibly base-pairing with matching sequences within <i>p300</i> mRNA	Diabetic retinopathy	(51)

miR-200b controls the epithelial to mesenchymal transition by downregulating p300 (67–70). In addition, p300 activates *HIF1*, which is a transcriptional regulator of *VEGF-A*, and triggers the development of abundant blood vessels (71–73). Since *miR-200b* negatively regulates p300, this miRNA has antiangiogenic properties (51).

In sum, these studies suggest that miRNAs have the capacity to indirectly affect epigenetic pathways in endothelial cells and influence the angiogenic response. This opens up the possibility of considering miRNAs as therapeutic targets or biomarkers, an exciting prospect since therapies for both vascular diseases and cancer are needed. In several diseases, miRNAs have proven to be excellent biomarkers as a result of their high circulating levels. Indeed, analysis of oncogenic and suppressor miRNAs that are found in primary tumors against non-neoplastic cells revealed exosome-mediated sorting mechanisms related to cancer progression (74, 75). It is unknown whether similar mechanisms could be utilized by Epi-miRNAs during the evolution of vascular diseases. Recently, the attention of the scientific community has been focused on other, widely-studied ncRNAs known as long non-coding RNAs (lncRNAs), which have master regulatory functions in angiogenesis.

LONG NON-CODING RNAs AS SCAFFOLDS FOR EPIGENETIC PARTNERS IN NEOVASCULARIZATION

lncRNAs are all ncRNAs larger than 200 nucleotides and are classified according to their proximity to protein-coding genes as intergenic, intronic, bidirectional, sense, and antisense lncRNAs. Massive analyses have revealed that lncRNAs are originated using the same mechanisms as protein-coding genes; however, contrary to protein-coding genes, lncRNAs show a preference for having

two-exon transcripts, and most of them lack any protein coding-potential. Also, lncRNAs show tissue-specific expression patterns and are predominantly located in the nucleus rather than the cytoplasm. In fact, there are several lines of evidence that suggest that lncRNAs are significantly more enriched in chromatin than miRNAs (76).

lncRNAs can indirectly modulate DNA methylation at CpG sites, which in turn regulates gene transcription. For example, Tsix recruits DNMT3a to methylate and silence the *XIST* promoter. *XIST* is an important effector involved in the inactivation of the X chromosome (77). Likewise, the lncRNA *Kcnq1ot1* recruits the *de novo* DNA demethylase DNMT1 to control the methylation status of ubiquitously imprinted genes during mouse development (78). lncRNAs can act as guides or scaffolds, facilitating interaction between several proteins, such as those that are part of chromatin-modifying complexes, causing gene activation or repression, depending on the interaction partners involved (79, 80). The polycomb repressive complexes PRC1 and PRC2, the transcriptional repressor element-1 silencing transcription factor REST, its cofactor (REST/CoREST), other epigenetic components like the mixed lineage leukemia protein and the H3K9 methyltransferase G9a, physically interact with lncRNAs (78, 80, 81). In addition, many lncRNAs such as *HOTAIR*, *Xist*, *Kcnq1ot1*, and *Breaveheart* interact with PRC2, implying that these ncRNAs play a role in recruiting this complex through its subunits (*EZH2*, *SUZ12*, *EED*, *RBBP4*, and *AEBP2*) or through a bridging protein (such as *JARID2*) to their target genes (82, 83). Likewise, the expression of many angiogenesis-related genes involved in the VEGF signaling pathway is regulated through lncRNAs (such as *H19*, *MEG3*, and *HOTAIR*), and recently, researchers discovered that some of them perform their regulatory function by influencing the expression and activity of several epigenetic modulators (20, 22). lncRNAs and their identified epigenetic targets in angiogenesis are listed in **Table 2**.

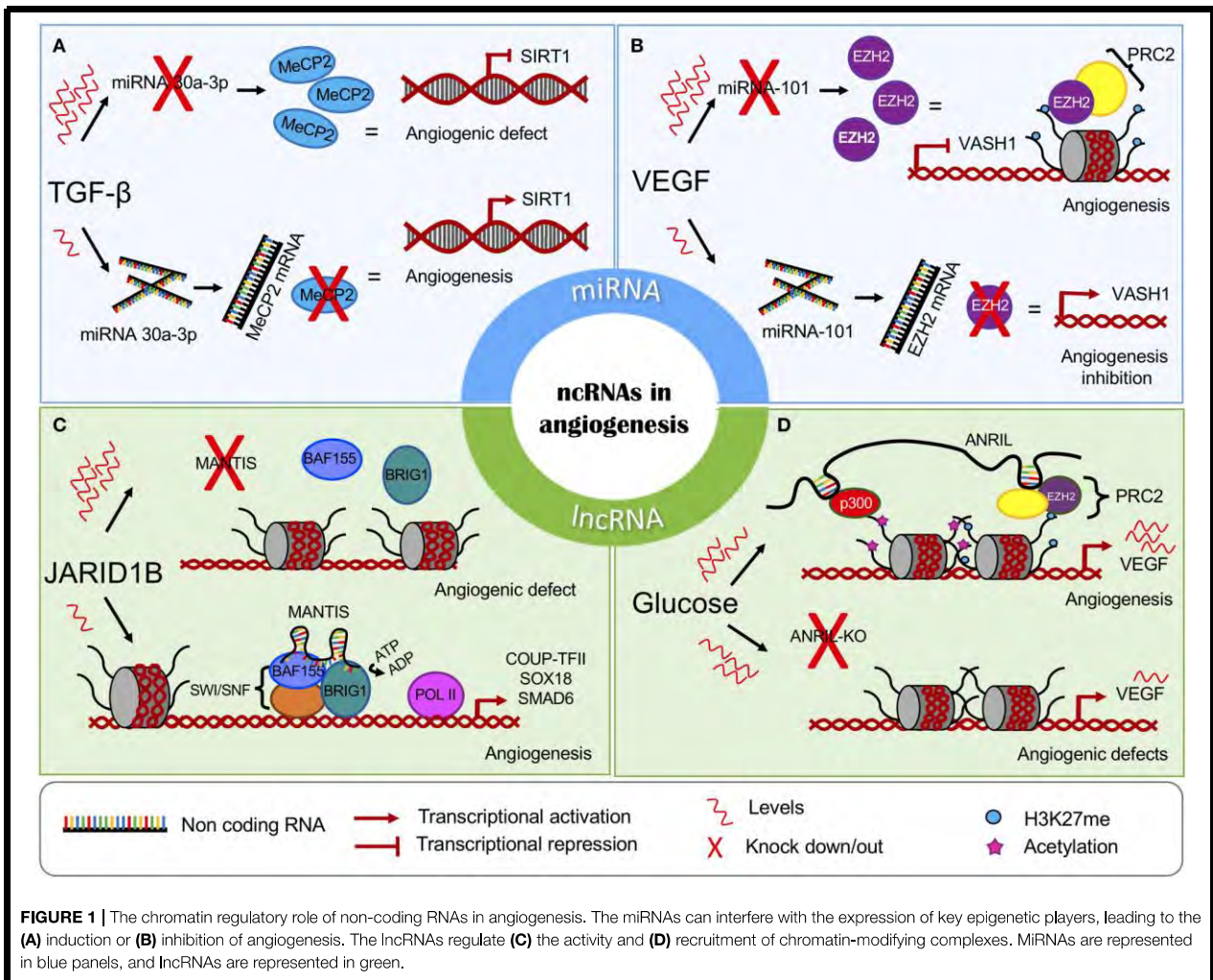


FIGURE 1 | The chromatin regulatory role of non-coding RNAs in angiogenesis. The miRNAs can interfere with the expression of key epigenetic players, leading to the (A) induction or (B) inhibition of angiogenesis. The lncRNAs regulate (C) the activity and (D) recruitment of chromatin-modifying complexes. MiRNAs are represented in blue panels, and lncRNAs are represented in green.

MANTIS

MANTIS is a recently discovered lncRNA required for endothelial cell function and proper angiogenesis. MANTIS is induced in the endothelium of glioblastoma tumors and is overexpressed during vascular regeneration in atherosclerosis regression. It alters angiogenic sprouting, tube formation, and epithelial cell migration. Loss of MANTIS expression is reported during pulmonary arterial hypertension, and its downregulation also led to the reduced expression of many angiogenesis-related mRNAs (80).

In endothelial cells, MANTIS is upregulated following the knockdown of the histone demethylase JARID1B. JARID1B loss triggers increased H3K4me3 levels at transcription start sites (TSS) of the MANTIS gene, facilitating gene expression. Interestingly, in patients with idiopathic pulmonary arterial hypertension, a disease characterized by endothelial dysfunction, MANTIS expression is downregulated, while JARID1B is upregulated (80).

Novel studies have revealed that MANTIS functions as a scaffold and regulates the activity of Brahma related gene-1 (BRG1), the catalytic subunit of the SWI/SNF chromatin remodeling complex. The MANTIS-BRG1 interaction allows for increased binding of BAF155, which is a core component of the SWI/SNF complex, enhancing BRG1 ATPase activity and chromatin relaxation at the TSS of the transcription factor COUP-TFII, which, in turn, recruits RNA Pol II binding and transcription of the pro-angiogenic genes SOX18 and SMAD6. The knockdown of MANTIS reduces BRG1 ATPase activity (80) (Figure 1C).

ANRIL

ANRIL is an antisense lncRNA from the INK4 locus. It encodes two cyclin-dependent kinase inhibitors, p15 (INK4b) and p16 (INK4a), and a protein known as ARF. All of the genes cooperate in tumor suppressor networks. When these genes are silenced, proatherosclerotic cellular mechanisms are enhanced,

TABLE 2 | Summary of lncRNAs and their epigenetic targets in angiogenesis and vascular disease.

Common name	Function	Mechanism	Model or disease	References
MANTIS	Promotes angiogenic sprouting and proper endothelial function	Interplay with BRG1 and favors ATPase activity in chromatin remodeling	Human umbilical vein endothelial cells, glioblastoma, atherosclerosis, and idiopathic pulmonary, arterial hypertension	(80)
ANRIL	Induces the formation of capillary-like structures	Recruitment of EZH2 and p300 to the VEGF promoter	Diabetic retinopathy	(84)
GATA6-AS	Promotes angiogenic sprouting	Binds to LOXL2 and regulates H3K4 trimethylation of angiogenesis- and hypoxia-related genes	Human endothelial cell-based xenograft model	(85)

such as increased adhesion and diminished apoptosis (86). In fact, ANRIL expression is correlated with the risk of some vascular diseases such as coronary atherosclerosis and carotid arteriosclerosis (87).

It has been shown that ANRIL recruits PRC2 or PRC1 to different target genes by directly interacting with their subunits EZH2, SUZ12, and CBX7 (86, 88, 89). In a diabetic retinopathy cellular model, high glucose levels upregulated ANRIL and VEGF expression. In turn, ANRIL positively regulated EZH2, EED, and p300 levels. Furthermore, ANRIL recruits EZH2 and histone acetyl-transferase p300 to the *VEGF* promoter, enhancing its expression and angiogenic effects. It was shown that ANRIL silencing prevented the formation of capillary-like structures in spite of the angiogenic influence of high glucose levels (84) (**Figure 1D**). Moreover, ANRIL silencing also promoted *miR-200b* expression, a previously described miRNA that has been shown to be involved in regulating VEGF (90).

GATA6-AS

GATA6-AS is the hypoxia-regulated long non-coding antisense transcript of *GATA6* and promotes angiogenesis by negatively regulating lysyl oxidase-like 2 (*LOXL2*). *LOXL2* catalyzes the oxidative deamination of lysines and hydroxylysines, which results in the generation of non-methylated H3K4 and gene silencing. Thus, GATA6-AS silencing leads to increased *LOXL2* activity and transcriptional repression. In the nucleus, the physical interaction between GATA6-AS and *LOXL2* positively regulates the expression of several angiogenesis- and hypoxia-related genes, such as periostin and cyclooxygenase-2. It has been shown that GATA6-AS silencing in epithelial cells significantly prevented TGF- β 2-induced endothelial to mesenchymal transition and augmented angiogenic sprouting in xenograft models *in vivo* (85).

Like epi-miRNAs, the epi-lncRNAs are excellent candidates biomarkers due to their easy collection and tissue specificity. Although there are few examples of epi-lncRNAs in angiogenesis, the implications behind these interactions provide an interesting view of the mechanisms in which lncRNAs regulate not only the recruitment but also the activity of chromatin modifiers. Another layer of complexity is added if we consider that lncRNAs have many alternative splice forms, including the non-linear, circular RNAs (circRNAs).

CIRCULAR RNAs IN NEOVASCULARIZATION

Circular RNAs (circRNA) are single-stranded RNAs that are widely conserved in all life domains and form a covalent closed loop (91). The discovery of this type of RNA has occurred fairly recently, and before their discovery, the RNAs were considered the result of errors within the process of gene transcription. These circRNAs are produced by a back-splicing process of pre-mRNA, in which a downstream splice donor is linked to an upstream acceptor (92, 93). The splice forms can circularize from exonic, intronic, or a combination of both regions (EiRNAs) (94).

In cancer-derived cell lines, it has been reported that changes in DNMTs and the hypermethylation of the CpG islands of some genes that host circRNA can induce gene silencing of both linear RNA and circRNA, suggesting an epigenetic mechanism that produces two molecular “hits” (95). Because circRNA lack 5' and 3' ends, these cannot be degraded by exonucleases. Instead, circRNA levels may be regulated by endonucleases and exosomal deportation (96). These molecules are stable, abundant and specific to certain cell types, having distinct transcriptional patterns for specific tissues and multiple isoforms in eukaryotic cells (97). CircRNAs have been linked to different biological processes, including cell proliferation, senescence, and apoptosis, among others. The study of circRNA has increased in recent years, since they have been shown to be related to both physiological and pathological processes (98). In fact, circRNAs have been proposed as potential biomarkers for neurological disorders, infectious diseases, cancer, and preeclampsia as a result of their availability in circulating body fluids (99–102).

The circRNAs have transcriptional and post-transcriptional regulatory functions. EiRNAs such as EIF3J associate with ribonucleoproteins like U1 and the Pol II at the promoters of their parental genes to enhance their own expression (94). Similar to EiRNAs, some circRNAs (such as ciANKRD52) can positively regulate their own expression through interaction with the Pol II complex (103). Other circRNAs regulate alternative splicing or serve as sponges to bind, store, or sequester miRNAs and other protein complexes containing transcription factors and RNA binding proteins (94, 104, 105). Due to the ability of circRNA to bind to miRNAs, they have been referred to as miRNA sponges (106). Despite their recent discovery, some evidence suggests that circRNAs are implicated in angiogenesis (e.g., circRNA-MYLK)

and many cardiovascular diseases, such as atherosclerosis (e.g., circR-284), myocardial infarction (e.g., ciRS-7), and coronary artery disease (CAD) (e.g., circ_0124644), among others (107, 108). However, to our knowledge, no study has shown that circRNAs have an epigenetic regulatory role in angiogenesis. Similar to the lncRNA ANRIL, a circularized and anti-sense splice variant of the INK4/ARF locus (cANRIL) has been associated with atherosclerotic vascular disease (109). Moreover, in the cytoplasm, the binding of circANRIL to the rRNA-processing machinery impairs its function and causes nucleolar fragmentation and stress signaling (110). These findings suggest that, just like their longer-sized isoform, the variant cANRIL may have a role in the epigenetic regulation of vascular disease.

The study of ncRNA has opened up a new research field, and this has been extended to the genome scale. This type of experimental approach has become common practice in both the research laboratory and at the clinical level. Therefore, along with a growing array of genomic analysis machinery, bioinformatics platforms have also been developed, thus generating a new set of tools for the study and analysis of ncRNA.

CONTRIBUTION OF NGS TECHNOLOGIES TO THE DISCOVERY OF NEW ncRNAs

In recent years, increasing quantities of data have been obtained from NGS technologies such as mass RNA sequencing (RNA-seq), small RNA-seq (smRNA-seq), and single-cell RNA-seq, among others. These technologies have revealed that the human genome encodes for more than 90,000 non-coding RNAs and that these play an important role in several diseases (111). Using publicly available genomic information, it is now possible to discover and characterize novel disease-associated ncRNAs. In the next section, we will describe some of the key discoveries that have been made thanks to NGS data, in which ncRNAs are shown to have roles in angiogenesis and neovascularization processes.

The study of the ncRNAs involved in molecular processes associated with neovascularization and angiogenesis in several diseases can be carried out by using RNA-seq approaches, especially where angiogenesis or neovascularization is one of the causes, risk factors, or consequences of the disorders. Some of the diseases studied in this manner have been ischemia stroke, CAD, hemangioma, and heart failure (HF). Furthermore, angiogenesis and neovascularization are strongly related to endothelial functioning and the transcriptional programming of endothelial progenitor cells (EPCs). Thus, the study of the molecular mechanisms involved in the regulation of EPCs is of great interest. Nevertheless, only a few studies have been conducted on human umbilical vein endothelial cells (HUVEC) or other endothelial models to understand the role of ncRNAs using NGS technologies. In this section, we will provide a compilation of some studies aiming to identify or characterize ncRNAs involved in vascular processes.

First, in 2012 Cheng et al. performed smRNA-seq on umbilical cord blood EPCs (UC-EPCs), which was known for its enrichment in EPCs, and compared the expression profiles

against EPCs derived from peripheral blood in adults (PB-EPCs) to understand the underlying mechanisms involved the functional differences between these two models. They identified specific patterns of miRNAs (miRNome) in UC-EPCs and PB-EPCs in which 54 miRNAs were overexpressed in UC-EPC and 50 miRNAs were overexpressed in PB-EPCs. For instance, UC-EPCs expressed miRNAs involved in angiogenesis such as *miR-31* and *mir-18a*, while PB-EPCs are enriched in tumor-suppressive miRNA expression such as that of *miR-10a* and *mir-26a* (112).

A study performed by Wang and colleagues in 2014 revealed that there was cooperation between VEGF and miRNAs in CAD progression. They performed smRNA-seq and identified EPC-specific miRNome that was related to angiogenic processes, which suggests that miRNAs in EPCs with a poor capacity to enhance angiogenesis might have higher levels of miRNAs targeting VEGF. Indeed, they identified anti-VEGF miRNAs such as *miR-361-5p* that were enriched in EPCs and in the plasma of patients with CAD (113).

Also, atherosclerosis appears to be one of the factors leading to CAD. In 2018, Mao and colleagues conducted a study to identify miRNAs linked with carotid atherosclerosis. They performed a differential expression analysis to identify genes that were specifically associated with either primary or advanced atherosclerotic plaque tissues. Using public databases, they predicted 23 miRNAs that targeted the differentially expressed genes, such as *miR-126*, *miR-155*, *miR-19A*, and *miR-19B*, which can play a regulating role in neovascularization and angiogenesis (114).

Furthermore, a study from Liu et al. (115) identified differentially expressed ncRNAs that were predicted to be involved in the regulation of high-density lipoprotein (HDL) metabolism, the deregulation of which is believed to be one of the main causes of CAD. To this end, they treated HUVEC cells with HDL from healthy subjects and patients with CAD and hypercholesterolemia. After RNA-seq analysis, 41 ncRNAs were identified, and researchers were able to show that the ncRNAs, along with protein-coding genes such as *DGKA* and *UBE2V1*, have critical functions in vascular cells (115).

Additionally, it is well-known that endothelial cell metabolism is sensitive to hypoxia, which is an adverse effect of atherosclerotic lesions in humans. In 2018, Moreau et al. investigated the lncRNA profiles of HUVEC cells using global run-on sequencing (GRO-Seq). GRO-seq is a sequencing method that measures active transcription, identifying newly synthesized RNA, and providing sufficient resolution to map the position and orientation of transcripts detected. This group aimed to discover changes in the expression patterns of lncRNAs in HUVEC cells exposed to hypoxia and demonstrated that hypoxia affects the transcription of ~1,800 lncRNAs. Among the most relevant lncRNAs identified were *MALAT1*, *HYMAI*, *LOC730101*, *KIAA1656*, and *LOC339803*, which were differentially expressed in human atherosclerotic lesions compared to normal vascular tissue (116).

In contrast, heart and circulatory system diseases often involve changes in vascular smooth muscle or cardiac cells. In 2018, Cheng et al. used RNA-seq to identify circRNAs in human aortic valves. They recognized 1,412 specific circRNAs, most of which

originated from exons of their host genes. Furthermore, after performing a gene ontology enrichment analysis, they found that the host genes were associated with pathways regulating aortic valve function (ECM-receptor interaction pathway, ErbB signaling pathway, and vascular smooth muscle contraction pathway) (117). In addition, Bell et al. identified novel lncRNAs in human vascular smooth muscle cells in 2014. This work expanded our knowledge of the relevance of lncRNAs in the control of smooth muscle cells. The researchers performed an RNA-seq experiment examining expression patterns in human coronary artery smooth muscle cells. Their analysis revealed 31 novel lncRNAs. They discovered and characterized a novel vascular cell-enriched lncRNA that they named *SENCR*. They performed RNA-seq after knockdown of *SENCR* and observed that expression of Myocardin and genes involved in the contraction of smooth muscle were reduced, while expression of other promigratory genes was enhanced (118). These results have enhanced our understanding of vascular cells and should be further studied in order to discern lncRNAs in vascular diseases. Finally, in 2015, Di Salvo et al. analyzed the expression profiles of cells derived from 22 human hearts from patients with Heart Failure (HF) vs. non-HF donor hearts. Initially, they discovered 2,085 lncRNAs, and subsequent analyses revealed 48 differentially expressed lncRNAs in HF patients. Among these, AP000783.2, RP11-403B2.6, and RP11-60A24.3 were identified (119).

Angiogenesis and neovascularization processes affect the prognosis of patients who have suffered from brain stroke ischemia. Thus, the identification of ncRNAs involved in these processes might be useful for their further use as drug targets or biomarkers for the disease. Therefore, Zhang et al. (120) aimed to uncover which ncRNAs have altered expression profiles after cerebrovascular dysfunction in ischemic stroke. Using bulk RNA-seq, they profiled lncRNA signatures in primary brain microvascular endothelial cells after oxygen-glucose deficiency. This approach allowed for the identification of 362 differentially expressed lncRNAs. The top three lncRNAs that were upregulated were *Snhg12*, *Malat1*, and *lnc-OGD 1006*, while the top three down-regulated lncRNAs were *281008D09Rik*, *Peg13*, and *lnc-OGD 3916* (120).

Another disease model that has been studied in order to identify ncRNAs involved in angiogenesis and neovascularization is infantile hemangioma (IH), which is a type of vascular tumor in infants. Li et al. investigated whether ncRNAs have a role in IH pathogenesis in 2018. The researchers used a bulk RNA-seq approach to examine global ncRNAs expression profiles in IH patients compared to their matched, normal-skin controls. In this study, researchers identified 256 lncRNAs and 142 miRNAs that were differentially expressed. They also found more than a thousand sponge modulators involved in miRNA-, lncRNA-, and mRNA-mediated interactions. These findings suggest the presence of an endogenous ncRNA regulatory network associated with the development of IH and other vascular diseases (121).

Overall, the studies described above have shown that NGS technologies can be very effective in identifying and

characterizing ncRNAs. This type of technology has helped researchers to understand the regulatory role of ncRNAs in angiogenic and neovascularization processes. However, studies in this field are just emerging, and additional research will be required to expand our knowledge and translated into clinical use.

CURRENT APPROACHES USED TO DISCOVER NEW ncRNAs

After the development of NGS technologies, ncRNAs have been discovered, and multiple efforts have been made to organize, collect, provide, and unify all available information regarding ncRNAs so that it can be accessed by the research community. Furthermore, new methods have developed to predict and identify novel ncRNAs. Here we present some of the cutting-edge bioinformatics approaches currently being used to study ncRNAs and give some examples of how they are used in the study of neovascularization processes (Figure 2). For a detailed explanation, see the following reference (122).

Transcriptome-wide association studies can be performed to identify expression-trait associations where ncRNAs might be involved. This method can identify single-nucleotide polymorphisms (SNPs) located in transcribed regions of ncRNA genes that can be related to a specific phenotype. A second bioinformatic approach is the use of tools for the prediction of primary, secondary, and tertiary ncRNA structures to obtain information about their potential function. This method has been used for circRNAs, smRNAs, and lncRNAs. The third approach to studying ncRNAs is the use of biological networks. These types of analyses enhance our understanding of the function of ncRNAs by integrating expression, regulatory, and protein-protein interaction networks. ncRNAs are highly connected in these networks and can influence more than one target gene in order to produce a specific phenotype. These approaches can identify disease-specific regulatory modules where ncRNAs play an important role (122).

Though the effective methods described above can be used to discover and understand the biological functions of ncRNAs, they have not been adequately exploited to reveal the roles of ncRNAs in angiogenesis or neovascularization. So far, only a few studies have used advanced bioinformatics tools for this purpose. For example, in 2018, Li et al. detected novel circRNAs related with IH using RNA-seq data. The best experimental approach for the detection of circRNAs is the use of deep sequencing of RNA treated with RNase R (which leaves a circRNA-enriched sample). The availability of tools to predict novel circRNAs from RNA-seq data is of great value, given that RNA-seq data are much more highly available (122). Thus, Li et al. used circRNAFinder, a tool able to predict circRNAs from bulk RNA-seq experiments, and identified 249 circRNA candidates differentially expressed between IH and matched normal skin samples. The circRNAs *hsa_circRNA001885* and *hsa_circRNA006612* were further investigated by this group, providing novel insights about the disease (123).

originated from exons of their host genes. Furthermore, after performing a gene ontology enrichment analysis, they found that the host genes were associated with pathways regulating aortic valve function (ECM-receptor interaction pathway, ErbB signaling pathway, and vascular smooth muscle contraction pathway) (117). In addition, Bell et al. identified novel lncRNAs in human vascular smooth muscle cells in 2014. This work expanded our knowledge of the relevance of lncRNAs in the control of smooth muscle cells. The researchers performed an RNA-seq experiment examining expression patterns in human coronary artery smooth muscle cells. Their analysis revealed 31 novel lncRNAs. They discovered and characterized a novel vascular cell-enriched lncRNA that they named *SENCR*. They performed RNA-seq after knockdown of *SENCR* and observed that expression of Myocardin and genes involved in the contraction of smooth muscle were reduced, while expression of other promigratory genes was enhanced (118). These results have enhanced our understanding of vascular cells and should be further studied in order to discern lncRNAs in vascular diseases. Finally, in 2015, Di Salvo et al. analyzed the expression profiles of cells derived from 22 human hearts from patients with Heart Failure (HF) vs. non-HF donor hearts. Initially, they discovered 2,085 lncRNAs, and subsequent analyses revealed 48 differentially expressed lncRNAs in HF patients. Among these, AP000783.2, RP11-403B2.6, and RP11-60A24.3 were identified (119).

Angiogenesis and neovascularization processes affect the prognosis of patients who have suffered from brain stroke ischemia. Thus, the identification of ncRNAs involved in these processes might be useful for their further use as drug targets or biomarkers for the disease. Therefore, Zhang et al. (120) aimed to uncover which ncRNAs have altered expression profiles after cerebrovascular dysfunction in ischemic stroke. Using bulk RNA-seq, they profiled lncRNA signatures in primary brain microvascular endothelial cells after oxygen-glucose deficiency. This approach allowed for the identification of 362 differentially expressed lncRNAs. The top three lncRNAs that were upregulated were *Snhg12*, *Malat1*, and *lnc-OGD 1006*, while the top three down-regulated lncRNAs were *281008D09Rik*, *Peg13*, and *lnc-OGD 3916* (120).

Another disease model that has been studied in order to identify ncRNAs involved in angiogenesis and neovascularization is infantile hemangioma (IH), which is a type of vascular tumor in infants. Li et al. investigated whether ncRNAs have a role in IH pathogenesis in 2018. The researchers used a bulk RNA-seq approach to examine global ncRNAs expression profiles in IH patients compared to their matched, normal-skin controls. In this study, researchers identified 256 lncRNAs and 142 miRNAs that were differentially expressed. They also found more than a thousand sponge modulators involved in miRNA-, lncRNA-, and mRNA-mediated interactions. These findings suggest the presence of an endogenous ncRNA regulatory network associated with the development of IH and other vascular diseases (121).

Overall, the studies described above have shown that NGS technologies can be very effective in identifying and

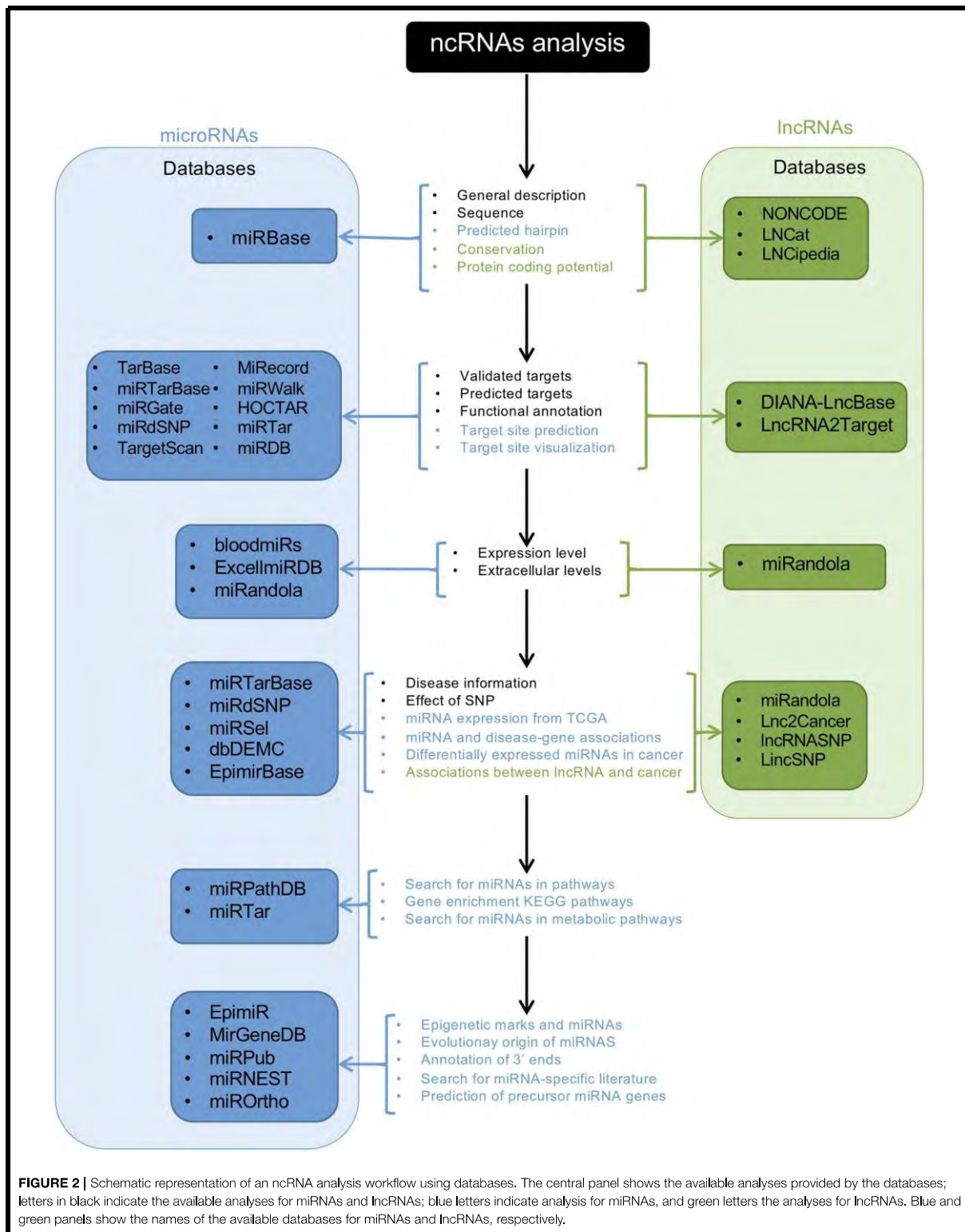
characterizing ncRNAs. This type of technology has helped researchers to understand the regulatory role of ncRNAs in angiogenic and neovascularization processes. However, studies in this field are just emerging, and additional research will be required to expand our knowledge and translated into clinical use.

CURRENT APPROACHES USED TO DISCOVER NEW ncRNAs

After the development of NGS technologies, ncRNAs have been discovered, and multiple efforts have been made to organize, collect, provide, and unify all available information regarding ncRNAs so that it can be accessed by the research community. Furthermore, new methods have developed to predict and identify novel ncRNAs. Here we present some of the cutting-edge bioinformatics approaches currently being used to study ncRNAs and give some examples of how they are used in the study of neovascularization processes (Figure 2). For a detailed explanation, see the following reference (122).

Transcriptome-wide association studies can be performed to identify expression-trait associations where ncRNAs might be involved. This method can identify single-nucleotide polymorphisms (SNPs) located in transcribed regions of ncRNA genes that can be related to a specific phenotype. A second bioinformatic approach is the use of tools for the prediction of primary, secondary, and tertiary ncRNA structures to obtain information about their potential function. This method has been used for circRNAs, smRNAs, and lncRNAs. The third approach to studying ncRNAs is the use of biological networks. These types of analyses enhance our understanding of the function of ncRNAs by integrating expression, regulatory, and protein-protein interaction networks. ncRNAs are highly connected in these networks and can influence more than one target gene in order to produce a specific phenotype. These approaches can identify disease-specific regulatory modules where ncRNAs play an important role (122).

Though the effective methods described above can be used to discover and understand the biological functions of ncRNAs, they have not been adequately exploited to reveal the roles of ncRNAs in angiogenesis or neovascularization. So far, only a few studies have used advanced bioinformatics tools for this purpose. For example, in 2018, Li et al. detected novel circRNAs related with IH using RNA-seq data. The best experimental approach for the detection of circRNAs is the use of deep sequencing of RNA treated with RNase R (which leaves a circRNA-enriched sample). The availability of tools to predict novel circRNAs from RNA-seq data is of great value, given that RNA-seq data are much more highly available (122). Thus, Li et al. used circRNAFinder, a tool able to predict circRNAs from bulk RNA-seq experiments, and identified 249 circRNA candidates differentially expressed between IH and matched normal skin samples. The circRNAs *hsa_circRNA001885* and *hsa_circRNA006612* were further investigated by this group, providing novel insights about the disease (123).



As shown previously, the development of tools used to predict and identify novel ncRNAs is invaluable. The increasing number of RNA-seq experiments and access to databases will increasingly facilitate the discovery of novel ncRNAs, and the characterization of ncRNAs will become increasingly straightforward. For instance, ANGIOGENES is a database that has been created to store information related to angiogenic processes. It depicts experimental data obtained from RNA-seq

experiments in endothelial cells. This allows for the *in-silico* detection of genes expressed in several endothelial cell types from different tissues. ANGIOGENES uses publicly-available RNA-seq experiments and identifies endothelial cell-specific ncRNAs in human, mouse, and zebrafish. The database facilitates further analyses using GO enrichment terms and is available online (124). In addition to ANGIOGENES, EndoDB is another database that retrieves information about endothelial cells from

TABLE 3 | Databases and tools for the ncRNAs study.

Database	ncRNAs	Website	Species	References
ANNOTATION RESOURCES				
TARGET RESOURCES				
TarBase	microRNAs	http://diana.imis.athena-innovation.gr/DianaTools/	Human/mouse, fruit fly, worm, and zebrafish	(130)
miRTarBase	microRNAs	http://mirtarbase.mbc.nctu.edu.tw	Human, mouse, virus	(131)
miRGate	microRNAs	http://mirgate.bioinfo.cnio.es	Human, rat, mouse	(132)
miRdSNP	microRNAs	http://mirdsnp.ccr.buffalo.edu	Human	(133)
TargetScan	microRNAs	http://www.targetscan.org/vert_72/	Human	(134)
CSmiRTar	microRNAs	http://cosbi.ee.ncku.edu.tw/CSmiRTar/	Human, mouse	(135)
MiRecords	microRNAs	http://c1.accurascience.com/miRecords/	Human, rat, mouse, fly, worm, chicken	(136)
miRsel	microRNAs	https://services.bio.ifi.lmu.de/mirsel/	Human, mouse	(137)
miRWalk	microRNAs	http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/	Human, mouse	(138)
miRPathDB	microRNAs	https://mpd.bioinf.uni-sb.de/	Human, mouse	(139)
HOCTARdb	microRNAs	http://hoctar.tigem.it/	Human	(140)
CELL TYPE SPECIFIC RESOURCES				
bloodmiRs	microRNAs	http://134.245.63.235/ikmb-tools/bloodmiRs/	Human	(145)
DISEASE-RELATED RESOURCES				
dbDEMOC	microRNAs	http://www.picb.ac.cn/dbDEMOC/	Human	(150)
EpimiRBase	microRNAs	https://www.epimirbase.eu/	Human	(151)
HMDD	microRNAs	http://www.cuilab.cn/hmdd	Human	(152)
OncomiRDB	microRNAs	http://lifeome.net/database/oncomirdb/	Human	(153)
OTHER RESOURCES				
EpimiR	microRNAs	http://210.46.85.180:8080/EpimiR/	Human	(158)
MirGeneDB	microRNAs	http://mirgenedb.org/	All	(25)
miRBaseTracker	microRNAs	http://mirbasetracker.org/	All	(127)
mirPub	microRNAs	http://www.microrna.gr/mirpub/	All	(159)
miRNEST	microRNAs	http://rhesus.amu.edu.pl/mirnest/copy/	All	(160)
miROrtho	microRNAs	http://cegg.unige.ch/mirortho	All	(161)

different platforms for several species (125). Other databases are available for the study of ncRNAs; nevertheless, these are not specialized in angiogenesis or neovascular processes. Databases and tools used for the study of ncRNAs are listed in **Table 3**.

We know that endothelial cells are heterogeneous; for instance, they function differently depending on vessel type (162). To uncover the molecular mechanisms controlling this heterogeneity, single-cell RNA sequencing analyses (scRNA-seq) have the potential to enhance our understanding of vascular biology. ScRNA-seq is currently being used to study and assess cellular heterogeneity. Particularly with respect to cancer research, this approach has proved to be valuable (163–165); nevertheless, its use in vascular research is just beginning. Recently published studies have mostly focused on protein-coding genes (166, 167). The participation of ncRNAs, along with epigenetic factors, in regulating the metabolic activities of endothelial cells from a single-cell perspective in vascular development and diseases is not yet clear.

CONCLUDING REMARKS

ncRNAs comprise a new frontier in genetic regulation that has impacts on several research areas. Undoubtedly, the study of angiogenesis and neovascularization has been enhanced through the integration of the study of ncRNAs and epigenetics. Further, ncRNAs are involved in the regulation of several angiogenic targets through epigenetic mechanisms. On the basis of this relationship, a new field of opportunity has emerged in which biomarkers and specific therapies may be identified that can improve the treatment of different vascular diseases and cancers. NGS platforms allow for the global analysis of ncRNA expression and can be used to compare different

physiological and pathological processes. Most of the pathways and mechanisms controlling the ncRNA-mediated regulation of angiogenesis remain unexplored. It is likely that new research strategies implementing an epigenetic perspective will facilitate future discoveries.

AUTHOR CONTRIBUTIONS

IH-R, LG-C, MS-A, TM-H, and ES-R wrote the manuscript. MS-A and IH-R did the artwork. TM-H and LG-C combined their information to make the tables. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT) through the Fondo Sectorial de Investigación en Salud y Seguridad Social (FOSISS, Grant No. 0261181), Fondo CB-SEP-CONACyT (284748), and UAM-PTC-704. ES-R was supported by the Natural Science Department at UAM Cuajimalpa Unit.

ACKNOWLEDGMENTS

IH-R and TM-H are masters students and LG-C and MS-A are doctoral students from Programa de Maestría y Doctorado en Ciencias Bioquímicas, UNAM and received a fellowship from CONACyT (IH-R: CVU 886138, TM-H: CVU 924685, LG-C: CVU 588391, and MS-A: CVU 659273). MS-A was also a beneficiary of the German Academic Exchange Service (DAAD Grant No. 91693321).

REFERENCES

- Carmeliet P, Jain RK. Molecular mechanisms and clinical applications of angiogenesis. *Nature*. (2011) 473:298–307. doi: 10.1038/nature10144
- Song X, Shan D, Chen J, Jing Q. miRNAs and lncRNAs in vascular injury and remodeling. *Sci China Life Sci*. (2014) 57:826–35. doi: 10.1007/s11427-014-4698-y
- Helkin A, Maier KG, Gahtan V. Thrombospondin-1,–2 and–5 have differential effects on vascular smooth muscle cell physiology. *Biochem Biophys Res Commun*. (2015) 464:1022–7. doi: 10.1016/j.bbrc.2015.07.044
- Xie H, Cui Z, Wang L, Xia Z, Hu Y, Xian L, et al. PDGF-BB secreted by preosteoclasts induces angiogenesis during coupling with osteogenesis. *Nat Med*. (2014) 20:1270–8. doi: 10.1038/nm.3668
- Xu X, Zong Y, Gao Y, Sun X, Zhao H, Luo W, Jia S. VEGF induce vasculogenic mimicry of choroidal melanoma through the PI3k Signal Pathway. *Biomed Res Int*. (2019) 2019:3909102. doi: 10.1155/2019/3909102
- Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. *Nature*. (2000) 407:249–57. doi: 10.1038/35025220
- Tímár J, Döme B, Fazekas K, Janovics A, Paku S. Angiogenesis-dependent diseases and angiogenesis therapy. *Pathol Oncol Res*. (2001) 7:85–94. doi: 10.1007/BF03032573
- Bieniasz-Krzywiec P, Martín-Pérez R, Ehling M, García-Caballero M, Pinioti S, Pretto S, et al. Podoplanin-expressing macrophages promote lymphangiogenesis and lymphoinvasion in breast cancer. *Cell Metab*. (2019) 30:1–20. doi: 10.1016/j.cmet.2019.07.015
- Nowak-Sliwinska P, Alitalo K, Allen E, Anisimov A, Aplin AC, Auerbach R, et al. Consensus guidelines for the use and interpretation of angiogenesis assays. *Angiogenesis*. (2018) 21:425–532. doi: 10.1007/s10456-018-9613-x
- Celletti FL, Waugh JM, Amabile PG, Brendolan A, Hilfiker PR, Dake MD. Vascular endothelial growth factor enhances atherosclerotic plaque progression. *Nat Med*. (2001) 7:425–9. doi: 10.1038/86490
- Li M, Liu C, Bin J, Wang Y, Chen J, Xiu J, et al. Mutant hypoxia inducible factor-1 α improves angiogenesis and tissue perfusion in ischemic rabbit skeletal muscle. *Microvasc Res*. (2011) 81:26–33. doi: 10.1016/j.mvr.2010.09.008
- Xue K, Zhao X, Zhang Z, Qiu B, Tan QSW, Ong KH, et al. Sustained delivery of anti-VEGFs from thermogel depots inhibits angiogenesis without the need for multiple injections. *Biomater Sci*. (2019). doi: 10.1039/C9BM01049A. [Epub ahead of print].
- Nicoglou A. Waddington's epigenetics or the pictorial meetings of development and genetics. *Hist Philos Life Sci*. (2018) 40:61. doi: 10.1007/s40656-018-0228-8
- Felsenfeld G. A brief history of epigenetics. *Cold Spring Harb Perspect Biol*. (2014) 6:a018200. doi: 10.1101/cshperspect.a018200
- Richard JLC, Eichhorn PJA. Deciphering the roles of lncRNAs in breast development and disease. *Oncotarget*. (2018) 9:20179–212. doi: 10.18632/oncotarget.24591
- Felsenfeld G, Groudine M. Controlling the double helix. *Nature*. (2003) 421:448–53. doi: 10.1038/nature01411
- Jenuwein T, Allis CD. Translating the histone code. *Science*. (2001) 293:1074–80. doi: 10.1126/science.1063127

- **PUBLICACIÓN 8:** Landscape of Germline Genetic Variants in *AGT*, *MGMT*, and *TP53* in Mexican Adult Patients with Astrocytoma.
- **AUTORES:** Carlos-Escalante JA, Gómez-Flores-Ramos L, Bian X, Perdomo-Pantoja A, de Andrade KC, Mejía-Pérez SI, Cacho-Díaz B, González-Barrios R, Reynoso-Noverón N, Soto-Reyes E, Sánchez-Correa TE, Guerra-Calderas L, Yan C, Chen Q, Castro-Hernández C, Vidal-Millán S, Taja-Chayeb L, Gutiérrez O, Álvarez-Gómez RM, Gómez-Amador JL, Ostrosky-Wegman P, Mohar-Betancourt A, Herrera-Montalvo LA, Corona T, Meerzaman D & Wegman-Ostrosky T
- **REVISTA:** *Frontiers in Oncology*. **VOLÚMEN:** 9. **PÁGINAS:** 1091. **AÑO:** 2020.

Cellular and Molecular Neurobiology
<https://doi.org/10.1007/s10571-020-00901-7>

ORIGINAL RESEARCH



Landscape of Germline Genetic Variants in *AGT*, *MGMT*, and *TP53* in Mexican Adult Patients with Astrocytoma

José Alberto Carlos-Escalante¹ · Liliana Gómez-Flores-Ramos² · Xiaopeng Bian³ · Alexander Perdomo-Pantoja⁴ · Kelvin César de Andrade⁵ · Sonia Iliana Mejía-Pérez^{6,7} · Bernardo Cacho-Díaz⁸ · Rodrigo González-Barrios⁹ · Nancy Reynoso-Noverón⁹ · Ernesto Soto-Reyes¹⁰ · Thalía Estefanía Sánchez-Correa⁷ · Lissania Guerra-Calderas¹⁰ · Chunhua Yan³ · Qingrong Chen³ · Clementina Castro-Hernández¹¹ · Silvia Vidal-Millán¹² · Lucía Taja-Chayeb⁹ · Olga Gutiérrez⁹ · Rosa María Álvarez-Gómez¹² · Juan Luis Gómez-Amador⁷ · Patricia Ostrosky-Wegman¹³ · Alejandro Mohar-Betancourt¹¹ · Luis Alonso Herrera-Montalvo^{11,14} · Teresa Corona¹⁵ · Daoud Meerzaman³ · Talia Wegman-Ostrosky⁹

Received: 20 March 2020 / Accepted: 6 June 2020
 © Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Astrocytoma is the most common type of primary brain tumor. The risk factors for astrocytoma are poorly understood; however, germline genetic variants account for 25% of the risk of developing gliomas. In this study, we assessed the risk of astrocytoma associated with variants in *AGT*, known by its role in angiogenesis, *TP53*, a well-known tumor suppressor and the DNA repair gene *MGMT* in a Mexican population. A case–control study was performed in 49 adult Mexican patients with grade II–IV astrocytoma. Sequencing of exons and untranslated regions of *AGT*, *MGMT*, and *TP53* from was carried in an Ion Torrent platform. Individuals with Mexican Ancestry from the 1000 Genomes Project were used as controls. Variants found in our cohort were then assessed in a The Cancer Genome Atlas astrocytoma pan-ethnic validation cohort. Variants rs1926723 located in *AGT* (OR 2.74, 1.40–5.36 95% CI), rs7896488 in *MGMT* (OR 3.43, 1.17–10.10 95% CI), and rs4968187 in *TP53* (OR 2.48, 1.26–4.88 95% CI) were significantly associated with the risk of astrocytoma after multiple-testing correction. This is the first study where the *AGT* rs1926723 variant, *TP53* rs4968187, and *MGMT* rs7896488 were found to be associated with the risk of developing an astrocytoma.

Keywords Astrocytoma · Germline genetic variant · Risk · Mexican population

Introduction

Globally, the annual age-adjusted incidence of central nervous system tumors is 4.63 per 100,000, and this incidence has been continuously increasing since 1990 (Patel et al. 2019). Gliomas are nearly a quarter of all intrinsic primary

tumors in the CNS, among which astrocytic tumors account for 75% of all reported cases. The most aggressive form of astrocytic tumors is astrocytoma grade IV (glioblastoma or GBM), and it is the most commonly diagnosed type in the United States of America (USA), with an annual age-adjusted incidence of 2.4 cases per 100,000 in Hispanic populations (Ostrom et al. 2018a). Brain cancer disproportionately contributes to cancer mortality, especially in high-risk age groups. In the USA, primary brain cancer is the first and fifth leading cause of cancer mortality in men and women, respectively, in people aged 20–39 years. Prognosis is dismal, with an overall 5-year relative survival rate is 5.6% (Siegel et al. 2016).

The risk factors for developing gliomas are still poorly understood and include environmental and genetic components. Cancer predisposition syndromes account for less than

José Alberto Carlos-Escalante and Liliana Gómez-Flores-Ramos have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10571-020-00901-7>) contains supplementary material, which is available to authorized users.

✉ Talia Wegman-Ostrosky
taliaw@gmail.com

Extended author information available on the last page of the article

Published online: 13 June 2020

Springer

5% of all adult gliomas (Ostrom et al. 2018a, b). Recent studies have estimated a 25% risk due to common genetic variations (Kinnersley et al. 2015). O-6-methylguanine-DNA methyltransferase (*MGMT*), tumor protein p53 (*TP53*), and more recently, angiotensinogen (*AGT*) are among the potential genes harboring germline variants associated with gliomas.

MGMT encodes a DNA repair enzyme (Pegg 1990). The methylation status of this gene, an epigenetic silencing mechanism, is a predictive biomarker of a favorable prognosis in temozolomide treatment in individuals with GBM (Yin et al. 2014; Zhao et al. 2018). However, little is known about the effect of germline variants in *MGMT* and the risk of glioma, but two variants (e.g., rs12917 and rs16906252) have been identified as being associated with a higher risk (Felini et al. 2007; Liu et al. 2009; Rapkins et al. 2015).

TP53 is well known for its role in cancer, and it is commonly found to be mutated in astrocytoma (Purkait et al. 2016). Rare, high-penetrance germline *TP53* variants that cause Li-Fraumeni are responsible for roughly 1% of astrocytomas (Malmer et al. 2007). Nonetheless, polymorphisms with low penetrance may also be associated with an enhanced susceptibility to astrocytoma. The *TP53* rs78378222 variant has been strongly associated with an increased risk of GBM (Wang et al. 2015).

Genetic association studies of *MGMT* and *TP53* germline variants have been performed predominantly in populations of European ancestry (Felini et al. 2007; Liu et al. 2009; Wang et al. 2016). Thus, the risk these polymorphisms may pose in populations of different ethnic background remains uncharacterized. For instance, the risk conferred by *MGMT* rs12917 variant has been reported to be higher in populations of European ancestry compared to other populations (Sheng et al. 2018). Furthermore, the ethnic composition of large glioma genomic databases, such as The Cancer Genome Atlas (TCGA) database, might not reflect the real composition of a population, and therefore, some ethnic groups may be underrepresented. The proportion of Hispanic people, in is just 2% of the TCGA GBM cohort, from which is it unknown the fraction of individual with Mexican ancestry (Spratt et al. 2016). In addition, we have previously reported that our patients tend to be younger at diagnosis than those reported in other works (Wegman-Ostrosky et al. 2016).

AGT is the central gene in the renin–angiotensin system (RAS). *AGT* encodes angiotensinogen (AGT), a protein that is cleaved by renin, forming angiotensin-I. Angiotensin-I is further cleaved to form the peptide angiotensin-II, which exerts its effects via the angiotensin-II type-1 (AT1R) and type-2 (AT2R) receptors and the Mas receptor (Fyhrquist and Saijonmaa 2008). Intact AGT and its renin-cleavage residue known as des (Ang I)AGT have also been shown to have functions independent of angiotensin-II (C el erier et al.

2002; Lu et al. 2016). RAS has been found to be involved in several hallmarks of cancer, such as proliferative signaling, evading growth suppressors, resisting cell death, inducing angiogenesis, reprogramming of energy metabolism, inflammation, cell migration, invasion, and metastasis, mainly via signaling through the AT1R (Wegman-Ostrosky et al. 2015). There is evidence suggesting the clinical utility of the *AGT* rs5050 germline genetic variant as a prognostic biomarker (Perdomo-Pantoja et al. 2018). Nevertheless, the role of *AGT* genetic variants in the risk of glioma is currently unknown.

The present study aimed to determine the association between the genetic variants of *AGT*, *MGMT*, and *TP53* with the risk of astrocytoma in a Mexican population and to compare our results with an independent, multi-ethnic cohort for validation.

Methods

Study Design

We used a discovery and a validation nested cohort study design (Fig. 1). For the discovery cohort, a single-center, case–control study following the STREGA statement reporting guidelines was conducted (Little et al. 2009) (Online Resource 1, STREGA Checklist) on Mexican patients. A pan-ethnic TCGA astrocytoma cohort was used as a validation cohort. This study and its informed consent complied with the Declaration of Helsinki, and they were approved by the Institutional Review Board of the Instituto Nacional de Neurolog a y Neurocirug a (INNN), Mexico City, Mexico, before recruitment of patients.

Participants

Our cohort comprised 49 Mexican adult patients of both genders, who were newly diagnosed with primary astrocytoma via histopathology, had not undergone prior treatment, had freely accepted to participate and signed an informed consent form. Patients with any other brain tumor, who had undergone prior treatment, or whose DNA samples were insufficient/degraded were excluded. The included patients underwent surgery for therapeutic purposes between 2013 and 2015, as previously reported (Perdomo-Pantoja et al. 2018). Information on the demographic and clinical data from our astrocytoma patients was collected via chart review in a blinded fashion. The histological grading was taken from the pathology report, which was based on the 2007 WHO classification (Louis et al. 2007). The tumor location was determined from the preoperative T1-weighted MRI scan with gadolinium contrast. The sample size was calculated as previously described (Perdomo-Pantoja et al. 2018).

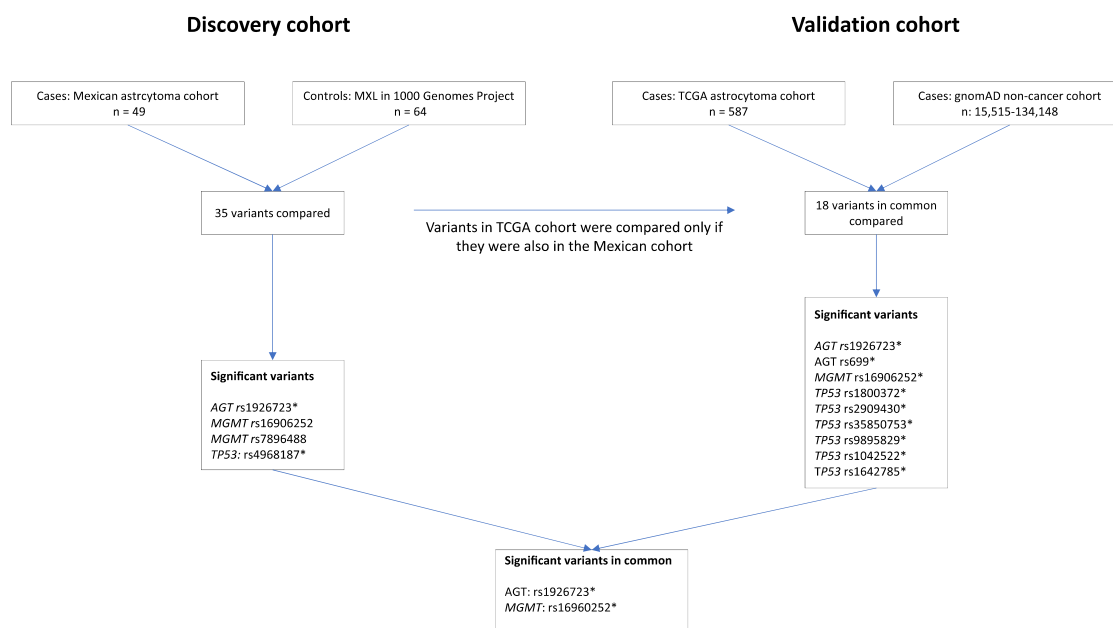


Fig. 1 Overview of the study. This work had two cohorts. The discovery cohort was comprised of Mexican individuals with the diagnosis of astrocytoma. This cohort was compared with individuals with Mexican ancestry from the 1000 Genomes Project. A total of 35 variants were evaluated. As a validation cohort, we used data from TCGA patients with astrocytoma. Individuals from a pan-ethnic, non-cancer

cohort in gnomAD were used as controls. In the TCGA cohort, we only compared the subset of variants also present in the Mexican cohort. *Significant variants after multiple-testing correction. *MXL* Mexican ancestry in Los Angeles; *TCGA* the cancer genome atlas, *gnomAD* genome aggregation database

To compare the genotypic frequencies found in the above-mentioned Mexican astrocytoma population, a population of Mexican ancestry from Los Angeles, California (MXL), was taken from the 1000 Genomes Browser Phase 3 version 3.8 (The 1000 Genomes Project Consortium 2015) and was considered to be the controls. The genotype data for the GRCh37.p13 assembly were available in the Ensembl website, release 98 (<https://grch37.ensembl.org/index.html>). No missing data were identified.

Sequencing and Data Processing

Before surgery, 5 mL of blood was drawn from each patient. The blood collection tubes were unidentified but labeled using a coding system for internal tracking and then sent to the laboratory. For the extraction of genomic DNA from the blood samples, a Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) was used. We performed a quantitative and qualitative analysis with real-time PCR with the StepOne™ system (Thermo Fisher Scientific, Waltham, MA, USA).

We used Ampliseq Designer to design 45 pairs of primers covering the exons, 5' UTR, 3' UTR and adjacent intronic

regions of *MGMT*, *TP53*, and *AGT*. Sequencing was performed with a 316-chip using the Ion Torrent PGM System. Signal processing and base calling of sequence data generated from the Ion Torrent PGM were performed with Torrent Analysis Suite version 3.4.2. The target regions were sequenced with an average coverage of 99% and a mean depth of 2946X (range 206X–9373X). High quality reads (a quality score > 20) from the FASTQ files generated from the Ion Torrent PGM Server were cut off from the adapter, barcode, and primer sequences prior to alignment with the human genome assembly 19 (Hg19) sequence.

The Ion Reporter software (Thermo Fisher Scientific, Waltham, MA, USA) (version 5.10) was used to map the sequences and to call the variants. With Integrative Genomic Viewer, we confirmed that each variant had a Phred quality score higher than 25 and a flow depth read of 30 or higher. Variant annotation was carried out with ANNOVAR (version 20180416) (Wang et al. 2010). Variant identification was carried out with dbSNP 147.

The genotypic frequencies were compared with the MXL population. The genotype proportions in the control populations were assessed for deviation from Hardy–Weinberg equilibrium (HWE) with the Haldane exact test, provided

by the Hardy–Weinberg package (version 1.6.3) in the R software environment (<https://www.r-project.org/>) (version 3.6.1) (Graffelman 2015). Multiple-testing correction was then applied to the *p* values with the Benjamini–Hochberg (BH) procedure, with a false discovery rate (FDR) of 0.20. Orthogonal verification of the identified genetic variants was conducted in 10% of the samples with 100% verification.

Statistical Analysis

Odds ratios (OR) were calculated as a measure of the association of the presence of a variant with the presence of astrocytoma, always taking the major allele as the reference. The EpiTools package (version 0.5-10) in R was the software employed to perform the calculations (Aragon 2017). Confidence intervals (CI) at 95% were calculated with the Wald normal approximation. When any value in the table was 0, small sample adjustment was applied, which consisted of a Jewell correction. Hypothesis testing was carried with the same R package. The χ^2 test was the test of choice, except when one of the expected values was less than 5, and in that case, a Fisher exact test was applied, taking the mid *p* value. The variants were compared under the genotype model (AA/AB and AA/BB, separately), dominant model, recessive model, and multiplicative model (Lewis 2002). The BH procedure was applied independently for each model, with an FDR of 0.20. As independent variants two scores were defined: a general variant score that only considered the presence of any of the variants, the maximal score would be 4. An allelic score was also defined, that is, the number of variant alleles was considered for calculating the score, with a maximum score of 8. Both variants were quantitative discrete.

Validation

We carried out the validation of the variants identified in our discovery cohort. As a validation cohort, we used whole-exome sequencing data from 587 individuals with astrocytoma in the TCGA. As this is a controlled dataset, access privilege was obtained according to the standard procedure (<https://gdc.cancer.gov/access-data/obtaining-access-controlled-data>), and the data files were downloaded as described (<https://gdc.cancer.gov/about-data/publications/PanCanAtlas-Germline-AWG>). GATK and Varscan2 were the variant callers used for generating the VCF files. The variant data for individuals with GBM and grade II and grade III astrocytoma were extracted. A more detailed account of the processing of the TCGA data can be found in the Online Resource 2.

In the TCGA cohort, we searched for the same variants found in the Mexican discovery cohort, and we worked only with these variants. The control data for this cohort were

taken from the whole non-cancer cohort from the Genomic Aggregation Database (gnomAD v2.1.1) (Karczewski et al. 2019). Odds ratios, confidence intervals, and hypothesis testing were computed in the same way as in the discovery cohort.

Results

Discovery Cohort

Forty-nine patients with astrocytoma were included in the DNA analysis after signing informed consent (51% females, 49% males, median age 50 years, range 25–79 years). The most common tumor location was the frontal lobe, (38.8%), followed by temporal lobe (32.7%), and 71.5% of the cases were high-grade astrocytomas of which GBM patients represented 59.1%. The patients were born in 14 different states of Mexico, mainly from the center of the country; 23 (46.9%) were born in Mexico City and the neighboring Estado de Mexico and Morelos. The clinical characteristics of the patients can be found in Table 1. No clinical variant was associated with the presence of the genetic variant (Online Resource).

Table 1 Clinical and demographical characteristics of Mexican patients

<i>n</i> = 49	<i>n</i> (%)
Median age	51 (IQ: 39–59)
Female	25 (51%)
Body Mass Index	25.9 (SD: 4.8)
Cigarette smoking	17 (34.7%)
Alcohol drinking	5 (10.2%)
History of cancer in relatives	18 (36.7%)
Tumor localization	
Frontal	19 (38.8%)
Temporal	12 (32.7%)
Parietal	8 (16.3%)
Insular	2 (4.1%)
Occipital	1 (2%)
Other	3 (6.1%)
WHO grade	
I	2 (4.0%)
II	12 (24.5%)
III	6 (12.2%)
IV	29 (59.3%)
Median age at diagnosis	
Grade I	31
Grade II	37 (31–57)
Grade III	50 (48–53)
Grade IV	56 (45–64)

Thirty-nine variants in the three genes of interest were identified in the cohort: 14 in *AGT* (and an additional intergenic variant, rs2067853 closely downstream), 10 in *MGMT*, and 14 in *TP53*. Out of these variants, 38 were single-nucleotide variants (SNV), and one was a deletion (rs936119541). The genotypic frequencies can be reviewed in Table 1. The variant annotations are reported Online Resource 3. Patients had a median of 11 variants in any of the three genes, with a range of 7–16.

Fourteen variants were identified in the *AGT* gene, and an additional intergenic variant (rs2067853) was registered closely downstream. Of those 14 variants, two were intronic, two occurred in the 3' UTR and 7 occurred in the 5' UTR. Two variants were nonsynonymous, and one was synonymous. The spatial distribution of the variants in *AGT* is represented in Fig. 2a. In *MGMT*, two variants were intronic,

three affected the 3' UTR, two were synonymous and three were nonsynonymous (Fig. 2b). In *TP53*, 3 variants were identified in the 3' UTR, 7 variants were intronic, two were nonsynonymous, and the remaining two were synonymous (Fig. 2c).

The data on 35 variants from 64 unrelated controls (MLX) were extracted from the 1000 Genomes Project Browser. In all instances, the major allele in the population of Mexican ancestry was the same as that in the whole 1000 Genomes population. None of the variants had a significant HWE deviation after multiple-testing correction was applied (Online Resource 4).

The odds ratios calculated under the multiplicative model are shown in Table 2. Two variants were significantly associated with the risk of astrocytoma after multiple-testing correction: rs1926723 in the *AGT* gene (OR

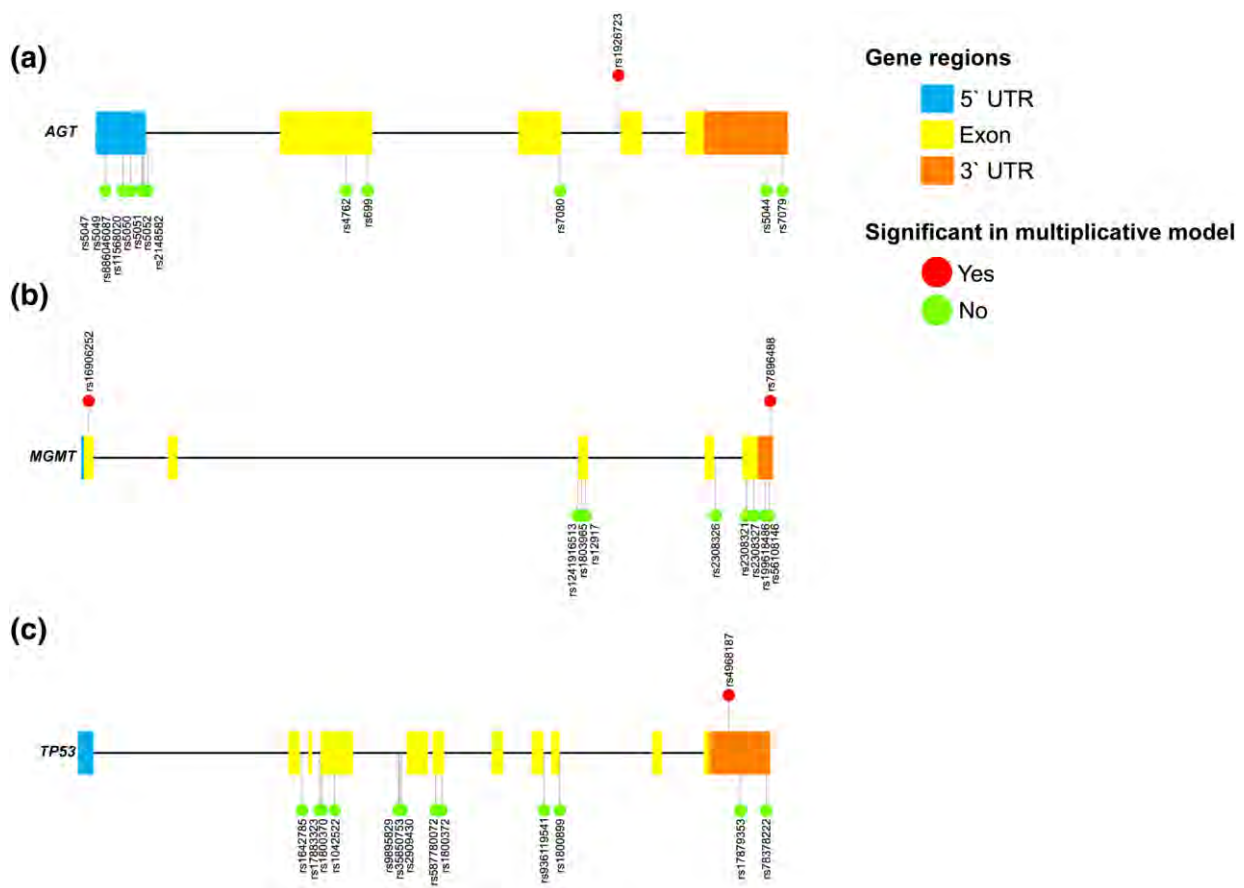


Fig. 2 Mutation diagram (lollipop) of the variants found in the Mexican population. Boxes indicate exons and lines connecting the boxes represent introns. Variants are marked in each gene. **a** In *AGT*, most of the variants were detected in the 5' UTR. Notice that the rs2067853 variant is not shown in the figure because it is an intergenic variant located downstream of *AGT*. **b** In *MGMT*, the variants we found also tended to cluster in the 3' UTR. The significant vari-

ants were at both extremes of the gene. **c** In contrast, in *TP53*, most of the variants were located in the central exons and introns. Although the *AGT* and *TP53* genes are depicted as in the forward strand, they are in the reverse strand. The figure is not to scale. Significant variants before multiple-testing correction are depicted on the upside. Non-significant variants are on the downside. UTR untranslated region

Table 2 Genotype frequencies in Mexican patients

Variant	dbSNP HGVS	Effect	Number of patients (%)		
			Homozygous reference	Heterozygous	Homozygous alternate
<i>AGT</i>					
rs2067853	g.17079C>T	Downstream variant	41 (83.7%)	8 (16.3%)	0 (0%)
rs7079	g.17006C>A	3' UTR variant	41 (83.7%)	8 (16.3%)	0 (0%)
rs5044	g.16873T>C	3' UTR variant	48 (98%)	1 (2%)	0 (0%)
rs1926723	g.15241A>G	Intronic variant	27 (55.1%)	15 (30.6%)	7 (14.3%)
rs7080	g.13650A>G	Synonymous variant	0 (0%)	1 (2%)	48 (98%)
rs699	g.9543T>C	p.Met268Thr	5 (10.2%)	15 (30.6%)	29 (59.2%)
rs4762	g.9370C>T	p.Thr207Met	36 (73.5%)	12 (24.5%)	1 (2%)
rs2148582	g.5538 T>C	Intronic variant	6 (12.2%)	12 (24.5%)	31 (63.3%)
rs5052	g.5480C>T	5' UTR variant	48 (98%)	1 (2%)	0 (0%)
rs5051	g.5465G>A	5' UTR variant	6 (12.2%)	12 (24.5%)	31 (63.3%)
rs5050	g.5451A>C	5' UTR variant	30 (61.2%)	17 (34.7%)	2 (4.1%)
rs11568020	g.5319G>A	5' UTR variant	43 (87.8%)	6 (12.2%)	0 (0%)
rs886046087	g.5318C>T	5' UTR variant	48 (98%)	1 (2%)	0 (0%)
rs5049	g.5254G>A	5' UTR variant	41 (83.7%)	8 (16.3%)	0 (0%)
rs5047	g.5086G>A	5' UTR variant	48 (98%)	1 (2%)	0 (0%)
<i>MGMT</i>					
rs16906252	g.5098C>T	Synonymous variant	48 (98%)	1 (2%)	0 (0%)
rs1241916513	g.245697T>C	Intronic variant	48 (98%)	1 (2%)	0 (0%)
rs1803965	g.245745C>T	Synonymous variant	27 (55.1%)	19 (38.8%)	3 (6.1%)
rs12917	g.245836C>T	p.Leu115phe	27 (55.1%)	19 (38.8%)	3 (6.1%)
rs2308326	g.297194C>T	Intronic variant	43 (87.8%)	5 (10.2%)	1 (2%)
rs2308321	g.304617A>G	p.Ile143Val	47 (95.9%)	2 (4.1%)	0 (0%)
rs2308327	g.304723A>G	p.Lys178Arg	47 (95.9%)	2 (4.1%)	0 (0%)
rs199618486	g.304851C>T	3' UTR variant	48 (98%)	1 (2%)	0 (0%)
rs56108146	g.304975G>A	3' UTR variant	48 (98%)	1 (2%)	0 (0%)
rs7896488	g.305062G>A	3' UTR variant	37 (75.5%)	12 (24.5%)	0 (0%)
<i>TP53</i>					
rs78378222	g.24117A>C	3' UTR variant	48 (98%)	1 (2%)	0 (0%)
rs17879353	g.23555C>A	3' UTR variant	48 (98%)	1 (2%)	0 (0%)
rs4968187	g.23427G>A	3' UTR variant	27 (55.1%)	17 (34.7%)	5 (10.2%)
rs1800899	g.19028T>C	Intronic variant	48 (98%)	1 (2%)	0 (0%)
rs936119541	g.18863del	Intronic variant	48 (98%)	1 (2%)	0 (0%)
rs1800372	g.17659A>G	Synonymous variant	48 (98%)	1 (2%)	0 (0%)
rs587780072	g.17624C>T	p.Arg202Gly	48 (98%)	1 (2%)	0 (0%)
rs2909430	g.17224G>A	Intronic variant	0 (0%)	4 (8.2%)	45 (91.8%)
rs35850753	g.17198G>A	Intronic variant	47 (95.9%)	2 (4.1%)	0 (0%)
rs9895829	g.17190T>C	Intronic variant	42 (85.7%)	7 (14.3%)	0 (0%)
rs1042522	g.16397C>G	p.Pro72His	3 (6.8%)	16 (36.4%)	25 (56.8%)
rs1800370	g.16290G>A	Synonymous variant	48 (98%)	1 (2%)	0 (0%)
rs17883323	g.16250C>A	Intronic variant	43 (87.8%)	6 (12.2%)	0 (0%)
rs1642785	g.16068C>G	Intronic variant	6 (12.2%)	22 (44.9%)	21 (42.9%)

2.74, 95% CI 1.40, 5.36; $p=0.003$) and rs4968187 in *TP53* (OR 2.48, 95% CI 1.26, 4.88, $p=0.007$). The rs4968187 variant was also significant after correction in the dominant model. Although the variants rs1690625 ($p=0.03$) and rs7896488 ($p=0.02$) in *MGMT* were nominally

significant in the multiplicative model, they were not significant after the BH procedure. None of the significant variants in the multiplicative model were associated with the clinical data of our patients (data not shown). *AGT* variants rs2148582, rs5051 (AA/AB), and rs1926723

(AA/BB) had a nominal p value < 0.05 in the genotypic model; however, they were not considered significant after multiple-testing correction. The results for all of the models and multiple-testing corrections are detailed in Online Resource 5 (Table 3).

Validation Cohort

The TCGA cohort was composed of 587 individuals: 393 had GBM, and 194 had grade II/III astrocytoma. Regarding the ethnicities of the patients, 88.1% were Caucasian, 8.3% were African American, 1.3% were Asian, and ethnicity was

Table 3 Odds ratios under the multiplicative model (allelic counts) in the Mexican cohort

Variant	Cases		Controls		OR (95% CI)	p value
	Ref	Alt	Ref	Alt		
<i>AGT</i>						
rs2067853	90	8	112	16	0.62 (0.25–1.52)	0.29 ^a
rs7079	90	8	112	16	0.62 (0.25–1.52)	0.29 ^a
rs5044	97	1	128	0	1.31 (0.16–98.11)	0.43 ^b
rs1926723	69	29	111	17	2.74 (1.40–5.36)	0.003^{a*}
rs7080	1	97	7	121	0.18 (0.02–1.47)	0.08 ^b
rs699	25	73	39	89	0.39 (0.22–0.69)	0.41 ^a
rs4762	84	14	111	17	1.09 (0.51–2.33)	0.83 ^a
rs2148582	74	24	40	88	0.71 (0.39–1.29)	0.26 ^a
rs5052	97	1	128	0	1.31 (0.16–98.11)	0.43 ^b
rs5051	74	24	40	88	0.71 (0.39–1.29)	0.26 ^a
rs5050	77	21	102	26	1.07 (0.56–2.04)	0.84 ^a
rs11568020	92	6	120	8	0.98 (0.33–2.92)	0.97 ^a
rs5049	90	8	118	10	1.05 (0.40–2.77)	0.92 ^a
rs5047	97	1	126	2	0.65 (0.06–7.27)	0.78 ^b
<i>MGMT</i>						
rs16906252	97	1	119	9	0.14 (0.02–1.09)	0.03^b
rs1803965	73	25	94	34	0.95 (0.52–1.73)	0.86 ^a
rs12917	73	25	93	35	0.91 (0.50–1.65)	0.76 ^a
rs2308326	91	7	120	8	1.15 (0.40–3.30)	0.79 ^a
rs2308321	96	2	125	3	0.87 (0.14–5.30)	0.90 ^b
rs2308327	96	2	125	3	0.87 (0.14–5.30)	0.90 ^b
rs199618486	97	1	128	0	1.31 (0.16–98.11)	0.43 ^b
rs56108146	97	1	127	1	1.31 (0.08–21.20)	0.87 ^b
rs7896488	86	12	123	5	3.43 (1.17–10.10)	0.02^a
<i>TP53</i>						
rs78378222	97	1	128	0	1.31 (0.16–98.11)	0.43 ^b
rs17879353	97	1	126	2	0.65 (0.06–7.27)	0.78 ^b
rs4968187	71	27	111	17	2.48 (1.26–4.88)	0.01^{a*}
rs1800899	97	1	127	1	1.31 (0.08–21.20)	0.87 ^b
rs1800372	97	1	127	1	1.31 (0.08–21.20)	0.87 ^b
rs2909430	4	94	8	120	0.64 (0.19–2.18)	0.47 ^a
rs35850753	96	2	128	0	2.64 (0.32–140.28)	0.19 ^b
rs9895829	91	7	118	10	0.91 (0.33–2.48)	0.85 ^a
rs1042522	22	66	40	88	0.73 (0.40–1.35)	0.32 ^a
rs1800370	97	1	125	3	0.43 (0.04–4.19)	0.52 ^b
rs17883323	92	6	118	10	0.77 (0.27–2.20)	0.62 ^a
rs1642785	34	64	41	87	1.13 (0.65–1.97)	0.67 ^a

Differences in the association of astrocytoma risk with the variants were calculated with Chi-square tests (a) or Fisher's exact tests taking the mid p value (b). The p value was considered nominally significant if $p < 0.05$ (bold). The significant variants after multiple-testing correction are indicated (*).

Ref reference allele, *Alt* alternate allele, *OR* odds ratio, *95% CI* 95% confidence interval

not reported in 2.2% of the patients. Only 11 patients were classified as Hispanic or Latino in this cohort.

Compared to the discovery cohort, 18 variants were found to be shared by our astrocytoma patients and the TCGA cohort, while 17 variants (including rs7896488 and rs496817) were not reported in the later cohort (Fig. 1). The number of individuals in the gnomAD non-cancer cohort ranged between 15,515 and 134,148, depending on the variant. The genotypic frequencies for each variant in TCGA and gnomAD are shown in Table 4.

The results of the TCGA-gnomAD comparison are shown in Table 5 for the multiplicative model. The *AGT* gene rs1926723 variant was associated with astrocytoma risk after the BH procedure in the TCGA cohort (OR 0.78, 0.64–0.95 95% CI, $p = 0.02$). Nine of the 18 variants assessed in the TCGA cohort were identified as significant before and after multiple-testing correction. Complete results are reported in Online Resource 6.

Discussion

In this study, we found 39 variants in the *AGT*, *MGMT*, and *TP53* genes in Mexican astrocytoma patients. *AGT* rs1926723, *MGMT* rs7896488, and *TP53* rs4968187 were associated with an increased astrocytoma risk before

multiple-testing correction. *MGMT* variant rs16906252 was related to decreased risk of astrocytoma. Variants rs1926723 and rs4968187 remained significant after multiple-testing correction. No patient has simultaneously the 4 variants. However, there are 2 patients that present 3 variants (rs1926723, rs7896488, and rs4968187). There were not significant co-occurrences of variants as revealed by the chi square test performed in a pairwise manner for the three sets of variants (rs1926723-rs7896488, rs1926723 /rs4968187, rs7896488, and rs4968187).

Variant rs1926723 from *AGT*, which was significant in our cohort, was also significant in TCGA cohort. Other 8 variants were significant in TCGA cohort.

AGT Gene rs1926723 Variant

We found that the rs1926723 (c.1125-13A>G rev) variant was significantly associated with the risk of astrocytoma in our cohort under the multiplicative model. This variant, located in the *AGT* gene, had an OR of 2.38, implying an increased risk of astrocytoma due to exposure to the C allele. The C allele of rs1926723 has been associated with hypertension (Padma et al. 2013), and with increased mortality in patients with coronary heart disease, and an increased incidence of type-2 diabetes, renal disease and higher body

Table 4 Genotype frequencies in TCGA astrocytoma cohort and gnomAD non-cancer cohort

Variant	Cases			Controls		
	Homozygous reference	Heterozygous	Homozygous alternate	Homozygous reference	Heterozygous	Homozygous alternate
<i>AGT</i>						
rs1926723	488 (83.1%)	95 (16.2%)	4 (0.7%)	106,772 (80.0%)	24,118 (18.1%)	2619 (2%)
rs7080	2 (0.3%)	69 (11.8%)	516 (87.9%)	574 (0.4%)	13,438 (10.0%)	119,883 (89.5%)
rs699	157 (26.7%)	280 (47.7%)	150 (25.6%)	30,518 (22.8%)	58,477 (43.6%)	45,122 (33.6%)
rs4762	442 (75.3%)	136 (23.2%)	9 (1.5%)	102,675 (76.9%)	28,575 (21.4%)	2226 (1.7%)
<i>MGMT</i>						
rs16906252	581 (99%)	4 (0.7%)	2 (0.3%)	77,639 (89.2%)	8990 (10.3%)	371 (0.4%)
rs1803965	442 (75.3%)	133 (22.7%)	12 (2%)	101,021 (75.6%)	29,865 (22.4%)	2735 (2%)
rs12917	426 (72.6%)	148 (25.2%)	13 (2.2%)	98,276 (73.6%)	32,295 (24.2%)	3020 (2.3%)
rs2308321	492 (83.8%)	85 (14.5%)	10 (1.7%)	110,050 (82.8%)	21,477 (16.2%)	1382 (1%)
rs2308327	483 (82.3%)	94 (16%)	10 (1.7%)	110,571 (82.7%)	21,758 (16.3%)	1438 (1.1%)
<i>TP53</i>						
rs1800899	574 (97.8%)	13 (2.2%)	0 (0%)	131,072 (97.7%)	3000 (2.2%)	22 (0.1%)
rs1800372	556 (94.7%)	30 (5.1%)	1 (0.2%)	130,925 (97.6%)	3194 (2.4%)	29 (0.1%)
rs2909430	196 (33.4%)	47 (8%)	344 (58.6%)	464 (3.0%)	4043 (25.9%)	11,096 (71.1%)
rs35850753	579 (98.6%)	6 (1%)	2 (0.3%)	14,985 (96.3%)	566 (3.6%)	6 (0.1%)
rs9895829	577 (98.3%)	8 (1.4%)	2 (0.3%)	13,112 (84.5%)	2288 (14.7%)	115 (0.7%)
rs1042522	46 (7.8%)	211 (35.9%)	330 (56.2%)	17,234 (12.9%)	56,052 (41.9%)	60,380 (45.2%)
rs1800370	574 (97.8%)	13 (2.2%)	0 (0%)	130,686 (97.5%)	3343 (2.5%)	38 (0.1%)
rs17883323	508 (86.5%)	77 (13.1%)	2 (0.3%)	116,694 (87.1%)	16,545 (12.4%)	718 (0.5%)
rs1642785	40 (6.8%)	220 (37.5%)	327 (55.7%)	15,635 (11.8%)	56,609 (42.6%)	60,699 (45.7%)

Table 5 Odds ratio under the multiplicative model (allelic counts) in TCGA cohort

Variant	Cases		Controls		OR (95% CI)	<i>p</i> value
	Ref	Alt	Ref	Alt		
<i>AGT</i>						
rs1926723	1071	103	237,662	29,356	0.78 (0.64–0.95)	0.02*
rs7080	73	1101	14,586	253,204	1.15 (0.91–1.46)	0.25
rs699	594	580	119,513	148,721	1.27 (1.14–1.43)	<0.001*
rs4762	1020	154	233,925	33,027	1.07 (0.90–1.27)	0.44
<i>MGMT</i>						
rs16906252	1166	8	164,268	9732	0.12 (0.06–0.23)	<0.001*
rs1803965	1017	157	231,907	35,335	1.01 (0.86–1.20)	0.88
rs12917	1000	174	228,847	38,335	1.04 (0.88–1.22)	0.64
rs2308321	1069	105	241,577	24,241	0.98 (0.80–1.20)	0.83
rs2308327	1060	114	242,900	24,634	1.06 (0.87–1.29)	0.55
<i>TP53</i>						
rs1800899	1161	13	265,144	3044	0.98 (0.56–1.69)	0.93
rs1800372	1142	32	265,044	3252	2.28 (1.60–3.25)	<0.001*
rs2909430	439	735	4971	26,235	3.15 (2.79–3.56)	<0.001*
rs35850753	1164	10	30,536	578	0.45 (0.24–0.85)	0.01*
rs9895829	1162	12	28,512	2518	0.12 (0.07–0.21)	<0.001*
rs1042522	303	871	90,520	176,812	0.68 (0.60–0.77)	<0.001*
rs1800370	1161	13	264,715	3419	0.87 (0.50–1.50)	0.61
rs17883323	1093	81	249,933	17,981	1.03 (0.82–1.29)	0.80
rs1642785	30	874	87,879	178,007	0.70 (0.61–0.79)	<0.001*

All the differences in the association of the astrocytoma risk with the variants were calculated with the χ^2 test. The *p* value was considered nominally significant if $p < 0.05$ (bold). The significant variants after multiple-testing correction are indicated (*)

Ref reference allele, *Alt* alternate allele, OR, odds ratio, 95% CI 95% confidence interval

mass index (Ellis et al. 2013). Our study is the first work to identify a cancer risk associated with rs1926723.

The rs1926723 variant is located 13 bases upstream of exon 4 (intron 3). Through in silico analysis with the Human Splicing Finder (HSF) tool, this variant was found to have a prediction variation of 19.85, suggesting the possibility of the creation of a new splice site (Padma et al. 2013). Intact AGT and its renin-cleavage residue, des(Ang I)AGT, have been shown to inhibit endothelial proliferation in vitro (Bouquet et al. 2006; C el erier et al. 2002) and to inhibit proliferation in tumor xenografts in mice in vivo through an antiangiogenic action (Bouquet et al. 2006; Vincent et al. 2009). Thus, if the rs1926723 variant indeed creates a new splice site, this could result in a frameshift in translation that leads to AGT and des(Ang I)AGT functional impairment, and, feasibly, increased angiogenesis. Functional studies of rs1926723 variants are necessary to provide a possible pathophysiological explanation for the increased risk of astrocytoma associated with this variant.

Although this variant was related to astrocytoma risk when compared to the TCGA cohort, the association was negative, with an OR of 0.78. This negative association could be a consequence of the lower frequency of this allele

in other populations compared with Latino population. In gnomAD non-cancer cohort, Latino population had the second highest alternate allele frequency (0.2117), three times as much as non-Finnish Europeans (0.07897). Thus, it could be a population-specific risk marker for the Mexican population.

TP53 Gene rs4968187 Variant

The rs4968187 (c.*485G>A) variant in *TP53* was found to be associated with an increased risk of astrocytoma in our population after multiple-testing correction, with an OR of 2.48. However, it could not be compared with TCGA because it was not reported there. This variant is located in the 3' UTR, and thus, a possible mechanism explaining the increased risk of astrocytoma could be post-transcriptional repression due to miRNAs, similar to what has been proposed for the rs78378222 variant, which is also located in the *TP53* 3' UTR and is associated with cancer predisposition in some Brazilian patients (Macedo et al. 2016). Additionally, variants in the 3' UTR of the *TP53* gene may disrupt mRNA processing, further dysregulating p53 expression (Wang et al. 2015).

MGMT Gene rs7896488 Variant

The rs7896488 variant, located in *MGMT*, was linked to astrocytoma risk in Mexican patients. This variant was not found in the TCGA cohort and thus could not be validated. Currently, there are no associations reported of rs7896488 to astrocytomas or any other disease. Located in the 3' UTR of the *MGMT* gene, rs7896488 could modify *MGMT* expression at the transcriptional or post-transcriptional levels, disrupting DNA repair. As confirmation in TCGA cohort could not be carried out for this variant, it is essential that it be addressed in other populations in future studies in order to determine if this is a specific risk variant in the Latino population.

MGMT Gene rs16906252 Variant

In the *MGMT* gene, the rs16906252 (c.66C>T) variant was negatively associated with astrocytoma risk in our cohort, and this was validated in the TCGA cohort. This variant occurs in the first exon of *MGMT* and has been shown to be related to low levels of *MGMT* promoter methylation and reduced levels of *MGMT* expression in glioblastoma, colorectal cancer, and normal tissue (Kuroiwa-Trzmielina et al. 2016; Rapkins et al. 2015). Mechanisms for rs16906252 variant-induced *MGMT* reduced expression other than promoter methylation may be of greater importance, such as disruption of transcription factor binding (Kuroiwa-Trzmielina et al. 2016).

An Australian case–control study found that rs16906252 was not related to risk in non-methylated GBM; however, in methylated GBM, these authors found that this variant increased risk (OR 2.86; 95% CI 1.57, 5.21; $p = 0.001$) (Rapkins et al. 2015). It is important to note that ethnic composition was not addressed, which is relevant, given that rs16906252 is more frequent in certain populations. In gnomAD, the T allelic frequency in non-Finnish European individuals and Ashkenazi Jewish individuals is twice as high as that in Latino and African individuals. It remains to be studied if the variant rs16906252 is related to risk in methylated GMS in Latino populations or if the low frequency of the variant in Latino populations is causing a bias in risk studies.

Additional Relevant Variants

In the *AGT* gene but only in our cohort under the dominant model, the rs2148582 and rs5051 variants were associated with the risk of astrocytoma.

Exclusively in TCGA cohort, the following variants were significantly associated to risk of astrocytoma: rs699 from *AGT*, rs1800372, and rs2909430 from *TP53*. At the other hand, variants related to less risk in TCGA were rs35850753,

rs9895829, rs1042522, and rs1642785. Only rs1042522 and rs1642785 have been previously studied in the context of glioma, conferring an increased risk in an Indian population (Jha et al. 2011).

Limitations

One of the limitations of this work was the sample size in the Mexican patient and control groups; nonetheless, this limitation was addressed with the validation TCGA cohort. Hispanic populations are underrepresented in the TCGA cohort, hindering the statistical power of any analysis employing this database, and thus, a pan-ethnic analysis was performed (Spratt et al. 2016).

In addition, patients from our cohort were not classified according to the most recent 2016 WHO classification of CNS tumors. The 2016 WHO classification is mainly a diagnostic and prognostic classification based on tumor histological grading and somatic isocitrate dehydrogenase 1 and 2 (*IDH1/2*) mutational status assessment (Louis et al. 2016). Tumor tissue could not be retrieved for this study, and, in Mexico, *IDH1/2* assessment is not routinely performed due to financial constraints in public health care. However, this is a risk study, and prognosis is not being assessed; therefore, it is unlikely that the 2016 WHO classification could change our results.

The INN is a referral institution nationwide, mainly for the central and southern states, and therefore, the demographic composition and clinical characteristics of its patients may not be representative of the country. However, our patients have not been shown to significantly differ from what is reported from the rest of the country (Wegman-Ostrosky et al. 2016).

Several variants found in our cohort could not be studied in the TCGA cohort because they were not reported in the TCGA astrocytoma cohort. Most of the absent variants were localized in the 3' UTR or 5' UTR of the genes, outside the coverage of TCGA astrocytoma data. Alternatively, some of the variants not found in TCGA may have been dropped due to quality issues. Discrepancies could have also arisen due to the use of different variant callers.

Conclusions

This is the first study to assess the genetic risk of astrocytoma in a Mexican population and the first study linking rs1926723 or any other *AGT* variant to the risk of astrocytoma. The ORs reported for rs1926723 and rs4968187 indicate a larger than two-fold increase in astrocytoma risk. This result differs from the majority of risk variants that only confer a modest (OR 1.1–1.5) increase in risk (Stadler et al. 2010). Validation studies in larger cohorts

of Latino patients are necessary. To date, few studies have obtained molecular data from astrocytoma patients from this population (Cardona et al. 2016; Ricaurte et al. 2017), and these patients are underrepresented in the TCGA. More effort should be directed towards the assembly of a genomic program of Latino astrocytoma patients that is large enough to sustain more extensive genetic association studies.

Understanding the risk factors for the emergence of astrocytoma may enable us to develop preventive approaches for at risk individuals. The impact of the RAS system on other cancers is currently under study, and some findings could be suitable for practical application. For instance, the use of angiotensin receptor blockers has been linked to a reduced risk of lung cancer (Zhang et al. 2015). Studies on the relationship between drugs targeting the RAS system and astrocytoma risk should be conducted in the future. The expression of several RAS components in astrocytoma cells has already been reported in vivo (Arrieta et al. 2008; Juillerat-Jeanneret et al. 2004), and therapeutic strategies directed at this system in patients with glioblastoma have shown some benefits (Tan et al. 2019), highlighting the importance that RAS has in astrocytoma.

The variants found in this article were assessed in peripheral blood samples, in a minimally invasive way, and they are potentially detectable in other tissues, making their future assessment more feasible.

To summarize, we identified *TP53* rs4968187, *MGMT* rs7896488, and rs16906252, and *AGT* rs1926723 as variants related to risk of astrocytoma. *MGMT* rs16906252 was validated in the TCGA cohort. *AGT* rs1926723 could be a specific risk factor in Mexican populations. This is the first study to address germline genetic variants in a Mexican population in the context of astrocytoma, taking the first step towards the clinical translation of these genetic variants as risk biomarkers that will allow us to better design preventive and early therapeutic strategies for astrocytoma.

Acknowledgements We thank Patricia Rosas, Laura Márquez, and Patricia de la Torre for their contributions in the project.

Author Contributions SIMP, APP, NRN, SVM, TWO, and POW involved in conceptualization of the study. TWO, BCD, and POW designed the study. XB, KCA, QC, SIMP, APP, TESC, TWO, JACE, JLGA, TC, and DM contributed to data acquisition and processing. ESR, RGB, POW, TESC, LSG, SVM, LTC, OG, RMAG performed data validation. JACE, LGFR, and TWO participated in statistical analysis. LGFR and TWO contributed to validation of results. All authors interpreted the results and involved in reviewing and editing. JACE and TWO involved in writing of the manuscript.

Funding This study was funded by Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico (Salud-2013-01-202720 to TWO, Scholarship number 969754 to JACE).

Data Availability The VCF files from Mexican patients can be accessed in Zenodo (<https://zenodo.org/>). Identifier: <https://doi.org/10.5281/zenodo.3737200>.

Compliance with Ethical Standards

Conflict of interest The authors do not have any competing interests to declare.

Ethical Approval This study was performed in line with principles of the Declaration of Helsinki. Approval was granted by the Research Board of the Instituto Nacional de Neurología y Neurocirugía (No 67/12).

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

- Aragon TJ (2017) Epitools: epidemiology tools. <https://cran.r-project.org/package=epitools>
- Arrieta O, Pineda-Olvera B, Guevara-Salazar P, Hernández-Pedro N, Morales-Espinosa D, Cerón-Lizarraga TL, González-De la Rosa CH, Rembao D, Segura-Pacheco B, Sotelo J (2008) Expression of AT1 and AT2 angiotensin receptors in astrocytomas is associated with poor prognosis. *Br J Cancer* 99(1):160–166. <https://doi.org/10.1038/sj.bjc.6604431>
- Bouquet C, Lamandé N, Brand M, Gasc JM, Jullienne B, Faure G, Griscelli F, Opolon P, Connault E, Perricaudet M, Corvol P (2006) Suppression of angiogenesis, tumor growth, and metastasis by adenovirus-mediated gene transfer of human angiotensinogen. *Mol Ther* 14(2):175–182. <https://doi.org/10.1016/J.YMTHE.2006.01.017>
- Cardona AF, Rojas L, Wills B, Behaine J, Jiménez E, Hakim F, Useche N, Bermúdez S, Arrieta O, Mejía JA, Ramón JF (2016) Genotyping low-grade gliomas among Hispanics. *Neuro-Oncol Pract* 3(3):164–172. <https://doi.org/10.1093/nop/npv061>
- Célérier J, Cruz A, Lamandé N, Gasc J-M, Corvol P (2002) Angiotensinogen and its cleaved derivatives inhibit angiogenesis. *Hypertension* 39(2):224–228. <https://doi.org/10.1161/hy0202.103441>
- Ellis KL, Palmer BR, Frampton CM, Troughton RW, Doughty RN, Whalley GA, Ellis CJ, Pilbrow AP, Skelton L, Yandle TG, Richards AM (2013) Genetic variation in the renin-angiotensin-aldosterone system is associated with cardiovascular risk factors and early mortality in established coronary heart disease. *J Hum Hypertens* 27(4):237–244. <https://doi.org/10.1038/jhh.2012.24>
- Felini MJ, Olshan AF, Schroeder JC, North KE, Carozza SE, Kelsey KT, Liu M, Rice T, Wiencke JK, Wrensch MR (2007) DNA repair Polymorphisms *XRCC1* and *MGMT* and risk of adult gliomas. *Neuroepidemiology* 29(1–2):55–58. <https://doi.org/10.1159/000108919>
- Fyhrquist F, Saijonmaa O (2008) Renin-angiotensin system revisited. *J Intern Med* 264(3):224–236. <https://doi.org/10.1111/j.1365-2796.2008.01981.x>
- Graffelman J (2015) Exploring diallelic genetic markers: The HardyWeinberg package. *J Stat Softw* 64(3):1–23. <https://doi.org/10.18637/jss.v064.i03>
- Jha P, Jha P, Pathak P, Chosdol K, Suri V, Sharma MC, Kumar G, Singh M, Mahapatra AK, Sarkar C (2011) TP53 polymorphisms in gliomas from Indian patients: Study of codon 72 genotype, rs1642785, rs1800370 and 16 base pair insertion in intron-3.

21 ARTÍCULOS EN PROCESO.

The role of the nuclear factors CTCF and BORIS in DNA and histone demethylation.

Lissania Guerra-Calderas¹, Itzel Alejandra Hernández-Romero¹, Marisol Salgado-Albarrán^{1,2}, Talia Wegman-Ostrosky³ and Ernesto Soto-Reyes^{1*}.

¹*Natural Sciences Department, Universidad Autónoma Metropolitana-Cuajimalpa (UAM-C), México City 05300, México.*

²*Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany*

³*Subdirección de Investigación básica, Instituto Nacional de Cancerología, Mexico City, Mexico,*

*Corresponding author: Ernesto Soto-Reyes. UAM-C, Av. Vasco de Quiroga 4871, Col. Santa Fe, Del. Cuajimalpa de Morelos. 05348, CDMX, México.

Email: esotoreyes@cua.uam.mx.

Abstract.

Architectural proteins are fundamental epigenetic regulators of the assembly of chromatin loops and chromosomal arrangements in the nucleus that affect gene expression. Growing evidence shows that these processes are controlled in part by insulator-like proteins such as CTCF and its paralogue gene *CTCF-L* (BORIS). In particular, CTCF interacts and potentially regulates multiple protein partners; such as its short isoform (CTCF-s), BORIS and histone and DNA demethylases; consequently, increasing its range of functions and, in turn, its effect in gene expression. In this review, we will summarize and discuss the functions of CTCF, CTCF-s, and BORIS in the regulation of chromatin states by the recruitment of DNA and histone demethylases. The latter is a newly described function of CTCF which has not been thoroughly discussed anywhere. The understanding of the methylation-related mechanisms that drive CTCF and BORIS functions in gene expression is of

