UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Búsqueda de epifactores como posibles blancos terapéuticos en cáncer

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctora en Ciencias

PRESENTA:

M. en C. Aylin del Moral Morales

TUTOR PRINCIPAL
Dr. Ernesto Soto Reyes Solís
Universidad Autónoma Metropolitana - Unidad Cuajimalpa

MIEMBROS DEL COMITÉ TUTOR
Dr. Jan Baumbauch
Universidad de Hamburgo

Dr. José Luis Medina Franco
Universidad Nacional Autónoma de México - Facultad de Química

Ciudad de México, octubre, 2024

Agradecimientos

Esta tesis se realizó en la Universidad Autónoma Metropolitana, Unidad Cuajimalpa bajo la dirección del Dr. Ernesto Soto Reyes Solís

A los miembros del comité tutor, Dr. Jan Baumbauch, Dr. José Luis Medina Franco y Dr. Marcelino Arciniega Castro por su asesoría en este proyecto.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por otorgar la beca número 894530 para estudios de posgrado.

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) por el apoyo económico otorgado para la asistencia al Congreso Nacional de Transducción de Señales.

Al Deutscher Akademischer Austauschdienst (DAAD) de Alemania por otorgar la beca número 91833882 para doctorados supervisados binacionalmente.

Al Dr. Jan Baumbach y la Dra. Maya Toph por su asesoría y apoyo durante mi estancia en sus grupos de trabajo.

Este trabajo fue financiado por el Consejo Nacional de Ciencia y Tecnología (CONACyT) a través del Fondo CB-SEP-CONACyT 284748 y PRODEP (511/2023-3066-1599). Todos los análisis computacionales se realizaron en el servidor esrs-epigenetics de la Universidad Autónoma Metropolitana (UAM) - Cuajimalpa, el cual fue financiado por CONACyT mediante el Apoyo para Proyectos de Investigación Científica, Desarrollo Tecnológico e Innovación en Salud ante la Contingencia por COVID-19, (00312021). También se recibió apoyo del Departamento de Ciencias Naturales (DCN) de la División de Ciencias Naturales e Ingeniería (DCNI) de la UAM-Cuajimalpa a través del Proyecto Divisional número 47301026.

Índice general

Glosario de abreviaturas

BET: Bromodomain and Extra-Terminal (Proteínas bromodominio y extraterminal)

COCONUT: Colección abierta de productos naturales (por sus siglas en Inglés).

BORIS: Hermano del regulador de sitios improntados (por sus siglas en Inglés).

CBP: CREB-Binding Protein (Proteína de unión a CREB).

CTCF: Factor de unión a CCCTC (por sus siglas en Inglés).

EGCG: galato de epigalocatequina

DNA: Ácido desoxirribonucleico.

DNMT: DNA Metiltransferasa.

DNMTi: Inhibidores de DNA Metiltransferasas.

FDA: Administración de Alimentos y Medicamentos de Estados Unidos.

FAD: Flavina adenina dinucleótido.

HAT: Histona Acetiltransferasa.

HDAC: Histona Desacetilasa.

HGSC: cáncer de ovario epitelial de alto grado (por sus siglas en Inglés).

H3K4me3: Trimetilación en la lisina 4 de la histona H3.

H3K9me3: Trimetilación en la lisina 9 de la histona H3.

H3K27: Residuo de lisina 27 de la histona H3.

IP: inmunoprecipitación

JmjC: Dominio catalítico Jumonji C.

JmjN: Dominio Jumonji N.

KDM: Desmetilasa de lisinas.

KDMi: Inhibidores de Desmetilasas de lisinas.

KMT: Metiltransferasa de lisinas.

lncRNA: RNA largo no codificante.

PAE: Error Predictivo de Alineación (por sus siglas en Inglés).

PHD: Plant Homeodomain (Dominio de unión a histonas modificado).

PIC: Complejo de preiniciación de la transcripción

pLDDT: Test de diferencias en distancias locales (por sus siglas en Inglés).

PRC2: Complejo represor Polycomb 2.

RMSD: desviación cuadrática media (por sus siglas en Inglés).

RMSF: Fluctuación cuadrática media (por sus siglas en Inglés).

RNA: Ácido ribonucleico.

TBP: Proteína de unión a TATA (por sus siglas en Inglés).

TAD: Dominio asociado topológicamente.

ZF: Dedo de zinc (por sus siglas en Inglés).

1. Resumen

La cromatina, un complejo dinámico de DNA, RNA y proteínas, es esencial para la regulación de la expresión génica. Los epifactores, proteínas que modifican el estado de la cromatina, desempeñan un papel crucial en este proceso. Las alteraciones en la función de los epifactores son una de las firmas moleculares del cáncer, por lo que las terapias epigenéticas, están emergiendo como opciones de tratamiento contra el cáncer. Sin embargo, la mayoría se centra en unas pocas enzimas, como las metilasas de DNA y las desacetilasas de histonas, mientras que otras, como las desmetilasas de histonas KDM4 y la proteína BORIS, han sido menos estudiadas debido a su similitud entre sí o con proteínas de genes parálogos como CTCF. Empleando una combinación de análisis transcriptómicos, modelado molecular y análisis de acoplamiento molecular, se propusieron moléculas que podrían actuar como inhibidores específicos de las proteínas KDM4 y BORIS. El análisis transcripcional permitió correlacionar la expresión de las KDM4 con firmas moleculares características de diversos tipos de cáncer, sugiriendo su implicación en procesos oncogénicos. Mediante técnicas de acoplamiento molecular, se diseñaron y evaluaron *in silico* moléculas capaces de unirse selectivamente a cada miembro de la subfamilia KDM4. Paralelamente, se abordó el desafío de desarrollar inhibidores específicos para BORIS, a pesar de su similitud estructural con CTCF. Utilizando modelado por homología, se identificaron sitios de unión únicos en BORIS, lo que permitió proponer moléculas capaces de interferir con su interacción con el DNA. Además, se demostró la interacción de BORIS y CTCF con TBP, sugiriendo una nueva estrategia terapéutica basada en la inhibición de estas interacciones proteína-proteína.

2. Abstract

Chromatin, a dynamic DNA, RNA, and protein complex, is essential for regulating gene expression. Epifactors, proteins that modify chromatin, play a crucial role in this process. Alterations in epifactor function is one of the molecular signatures of cancer, so epigenetic therapies are emerging as cancer treatment options. However, most focus on a few enzymes, such as DNA methylases and histone deacetylases; while others, such as histone demethylases KDM4 and the BORIS protein, have been less studied due to their similarity to each other or to proteins of paralogous genes such as CTCF. Using a combination of transcriptomic analysis, molecular modeling and molecular docking analysis, molecules that could act as specific inhibitors of KDM4 and BORIS proteins were proposed. Transcriptional analysis allowed correlating KDM4 expression with molecular signatures characteristic of various types of cancer, suggesting their involvement in oncogenic processes. Using molecular docking techniques, molecules capable of selectively binding to each member of the KDM4 subfamily were designed and evaluated *in silico*. In parallel, the challenge of developing specific inhibitors for BORIS, despite its structural similarity to CTCF, was addressed. Using homology modeling, non-conserved binding sites were identified in BORIS, which allowed us to propose molecules capable of interfering with its interaction with DNA. In addition, the interaction of BORIS and CTCF with TBP was demonstrated, suggesting a new therapeutic strategy based on the inhibition of these protein-protein interactions.

3. Introducción

3.1 Epigenética

El término *epigenética* fue acuñado originalmente por Conrad Waddington en la década de 1950 para referirse a los cambios en el fenotipo celular que son hereditarios e independientes de las alteraciones en la secuencia de DNA (Berger et al., 2009; Noble, 2015). En la actualidad, la epigenética se describe como la ciencia que estudia los cambios en la expresión génica que no implican modificaciones en la secuencia de DNA y que son heredables (Berdasco and Esteller, 2019). Este término se utiliza generalmente para describir eventos relacionados con la estructura de la cromatina, que incluyen la modificación covalente del DNA y las histonas, así como la actividad de los RNA largos no codificantes (lncRNA) (Dawson and Kouzarides, 2012).

La cromatina es el complejo macromolecular de DNA e histonas que proporciona un andamiaje para el empaquetamiento del material genético dentro de la célula. El DNA se compacta dentro del núcleo gracias a su interacción con las histonas, que tienen carga positiva por ser ricas en residuos de lisina y arginina. Los eucariotas poseen cinco tipos de histonas: H1, H2A, H2B, H3 y H4. Las histonas H2A, H2B, H3 y H4 forman un octámero compuesto por dos dímeros H2A-H2B y un tetrámero H3-H4, alrededor del cual se envuelve el DNA para formar el nucleosoma, que es la unidad funcional de la cromatina. La histona H1 se conoce como "linker" ya que forma bucles de cromatina entre nucleosomas, lo que facilita la compactación del material genético (Felsenfeld and Groudine, 2003; Turner, 2005).

El estudio de los nucleosomas ha demostrado que todos sus componentes pueden modificarse covalentemente, lo que altera la organización y la función de la cromatina. En términos generales, la cromatina se divide en dos tipos principales: la heterocromatina, que está compactada y asociada con genes inactivos, y la eucromatina, que está relativamente relajada y contiene la mayoría de los genes activos. A su vez, la heterocromatina se divide en facultativa y constitutiva. La primera contiene genes cuya expresión está reprimida, pero que pueden reactivarse en las circunstancias adecuadas, mientras que la segunda es más compacta y tiene principalmente funciones estructurales (Liu et al., 2020a).

El estado de la cromatina es dinámico y puede regularse mediante la modificación covalente de los extremos aminoterminales de las histonas que sobresalen del nucleosoma y

son accesibles a las enzimas que las modifican químicamente a través de un sistema de complejos escritores, lectores y borradores (Felsenfeld and Groudine, 2003). Estas modificaciones conforman una especie de código que funciona en conjunto con la secuencia de DNA para establecer y estabilizar patrones de expresión génica (Esteller, 2011). Las modificaciones de histonas más estudiadas son la acetilación, la metilación, la fosforilación, la ubiquitinación y la SUMoilación, las cuales se encuentran en residuos específicos y pueden tener efectos opuestos en la transcripción génica (Bannister and Kouzarides, 2011). Por ejemplo, la acetilación de la lisina 9 de la histona H3 (H3K9ac) y la metilación de la lisina 36 de la histona H3 (H3K36me3) están asociadas con la activación transcripcional, mientras que la metilación de la lisina 9 de la histona H3 (H3K9me3) y la metilación de la lisina 27 de la histona H3 (H3K27me3) están asociadas con la represión transcripcional (Bannister and Kouzarides, 2011).

Además de las modificaciones de DNA e histonas, la organización tridimensional de la cromatina también es controlada por proteínas arquitectónicas, cuyas funciones involucran el mantenimiento de dominios topológicamente asociados (TADs), la definición de los límites entre la eucromatina y la heterocromatina y el anclaje de la cromatina a la lámina nuclear (Fiorito et al., 2016; Kaczmarczyk et al., 2022; Kentepozidou et al., 2020; Ong and Corces, 2014).

Es indudable que los diferentes niveles de regulación epigenética involucran la coordinación de diversas proteínas como las DNA metiltransferasas (DNMTs), histonas metiltransferasas de lisinas (KMTs), entre otras, a las cuales se han denominado epifactores. Un epifactor es una proteína o una molécula que participa en la regulación epigenética (Medvedeva et al., 2015). Los epifactores incluyen una variedad de componentes que influyen en la estructura de la cromatina y la modificación covalente del DNA, RNA y proteínas. Estos abarcan:

- Histonas y variantes de histonas
- Protaminas
- Chaperonas de histonas
- Modificadores de histonas
- Lectores de modificaciones de histonas
- Remodeladores de cromatina
- Modificadores de DNA y RNA
- Lectores de modificaciones de DNA y RNA
- Cofactores proteicos que forman complejos con los epifactores

Además, la definición se ha ampliado recientemente para incluir lncRNAs (RNAs largos no codificantes) que regulan la expresión génica a nivel transcripcional o la expresión de otros factores epigenéticos a cualquier nivel (Marakulina et al., 2022).

3.2 Alteraciones epigenéticas en cáncer

Debido a la complejidad de los procesos epigenéticos y a su papel fundamental para el establecimiento de patrones transcripcionales, no es de extrañarse que su alteración sea una de las firmas moleculares del cáncer. El genoma de las células cancerosas se reprograma no solo por consecuencia de mutaciones oncogénicas, sino también por cambios en el epigenoma, lo que beneficia la adquisición de características clave del cáncer, tales como la proliferación descontrolada, metástasis, la resistencia a la terapia y la evasión del sistema inmune (Hanahan, 2022). Por ejemplo, uno de los fenómenos más estudiados en cáncer es la alteración en los patrones de metilación del DNA. Esta es una modificación usualmente asociada a promotores de genes transcripcionalmente reprimidos y a heterocromatina (Ehrlich, 2002). En cáncer, generalmente ocurre una pérdida de metilación global lo que ocasiona descompactación y activación transcripcional de oncogenes, acompañado de hipermetilación local de algunos promotores, usualmente en genes supresores de tumores (Ehrlich, 2002; Besselink et al., 2023).

Más complejo es el entendimiento de las modificaciones postraduccionales de histonas y el papel que juegan en el cáncer. Las modificaciones postraduccionales de histonas regulan la transcripción por medio de patrones dinámicos, que permiten cambios rápidos y reversibles en la expresión génica. En cáncer, usualmente hay una desregulación en la expresión de varias enzimas fundamentales para los complejos escritores, borradores y lectores, lo que ocasiona cambios en los patrones de modificaciones postraduccionales de histonas y, por ende, el encendido o apagado de genes (Noberini et al., 2018; Lee and Kim, 2022). Un claro ejemplo de este fenómeno son las acetiltransferasas de histona (HAT), enzimas que catalizan la acetilación de residuos de lisina, lo cual provoca una disminución en la fuerza de interacción proteína-DNA de las histonas y por lo tanto la relajación de la cromatina, haciéndola accesible a factores de transcripción. El desarrollo del cáncer va asociado a cambios en los patrones de acetilación de histonas. La sobreexpresión y el aumento de la actividad de las HDAC se han identificado como motores del desarrollo tumoral y la metástasis, al grado en que las acetiltransferasas *p300* y *CBP* se consideran oncogenes (Yang et al., 2022).

Por otra parte, la metilación de histonas es un proceso complejo y altamente regulado. Cada una de las enzimas escritoras y borradoras son específicas para cierto residuo de lisina o arginina. En particular, las lisina metiltransferasas (KMT) y las lisina demetilasas (KDMs) pueden activar o reprimir la transcripción de genes dependiendo del sitio de metilación. Por ejemplo, la trimetilación de la lisina 4 en la histona H3 (H3K4me3) generalmente está asociada con la activación génica, mientras que la trimetilación de la lisina 9 en la histona H3 (H3K9me3) está asociada con represión transcripcional (Noberini et al., 2018; Lee and

Kim, 2022). En este sentido, se ha descrito la sobeexpresion de varias KDMs en cáncer, por ejemplo, KDM4A está sobreexpresada y a veces amplificada en varios tipos de neoplasias como leucemia, pulmón, próstata, cáncer colorrectal y cáncer de mama (Guerra-Calderas et al., 2015; Metzger et al., 2017). KDM4B promueve la carcinogénesis en cáncer de mama positivo para receptores de estrógeno (Yang et al., 2010) y se ha asociado con mal pronóstico en cáncer de próstata (Sha et al., 2020) y osteosarcoma (Liu et al., 2020b).

De igual importancia es la disrupción de los dominios topológicamente asociados (TADs), un fenómeno frecuentemente observado en cáncer que contribuye a la descompactación de la cromatina y al establecimiento de patrones transcripcionales alterados (Hnisz et al., 2016; Akdemir et al., 2020). La estructura tridimensional de la cromatina es establecida y mantenida por proteínas con función arquitectónica, destacando entre ellas CTCF, conocido como el regulador maestro de la cromatina (Kentepozidou et al., 2020). *CTCF* es un gen esencial de expresión ubicua cuyas funciones involucran el mantenimiento de TADs, el anclaje de la cromatina a la lámina nuclear y la regulación del contacto enhancer-promotor. En cáncer, se ha reportado que estos dominios se desregulan por una competencia entre CTCF y BORIS, una proteína paráloga de CTCF que se expresa principalmente en el testículo y en algunas neoplasias (Liu et al., 2019). Aunque BORIS tiene el mismo dominio de unión al DNA que CTCF, no puede interactuar con proteínas clave para la formación de asas de cromatina mediada por CTCF, como las cohesinas, que son cruciales para el mantenimiento de los TADs (de Necochea-Campion et al., 2011; Hore et al., 2008). Este hallazgo sugiere que CTCF y BORIS tienen funciones opuestas en la regulación de la estructura de la cromatina y, por ende, en la expresión génica, lo que podría tener implicaciones importantes en la biología del cáncer (Pugacheva et al., 2015; Nishana et al., 2020). Un resumen de los procesos anteriormente descritos puede observarse en la **Figura 3.1**.

3.2.1 Fármacos inhibidores de epifactores

Debido a la importancia de los procesos epigenéticos, las terapias farmacológicas cuyos blancos son epifactores son una opción terapéutica emergente para el tratamiento del cáncer. Actualmente existen algunos fármacos con blancos epigenéticos que han sido aprobados por la FDA para el tratamiento de distintas neoplasias, una recopilación de estas moléculas se meustra en la **Tabla 3.1**. La mayoría de estos compuestos tienen como blanco a las enzimas metilasas de DNA (DNMTs), como es el caso de la azacitidina y la decitabina, que son análogos nucleotídicos de la citidina. La azacitidina se incorpora en el DNA y el RNA, inhibiendo las DNMTs y provocando hipometilación, lo que resulta en la desrepresión de genes supresores de tumores y otros reguladores del ciclo celular. La decitabina, otro análogo de la citidina, actúa de manera similar pero se incorpora sólo en el DNA, inhibiendo también a las DNMTs y promoviendo la hipometilación del DNA. Ambos fármacos han mostrado buenos resultados en la supervivencia de pacientes con leucemia mieloide (Ye et al., 2024; Kagan et al., 2023). Además de estos compuestos, hay otros dirigidos a las DNMTs que se encuentran en pruebas clínicas para el tratamiento de la leucemia mieloide y otros tipos de cáncer, sin embargo, resalta que los compuestos

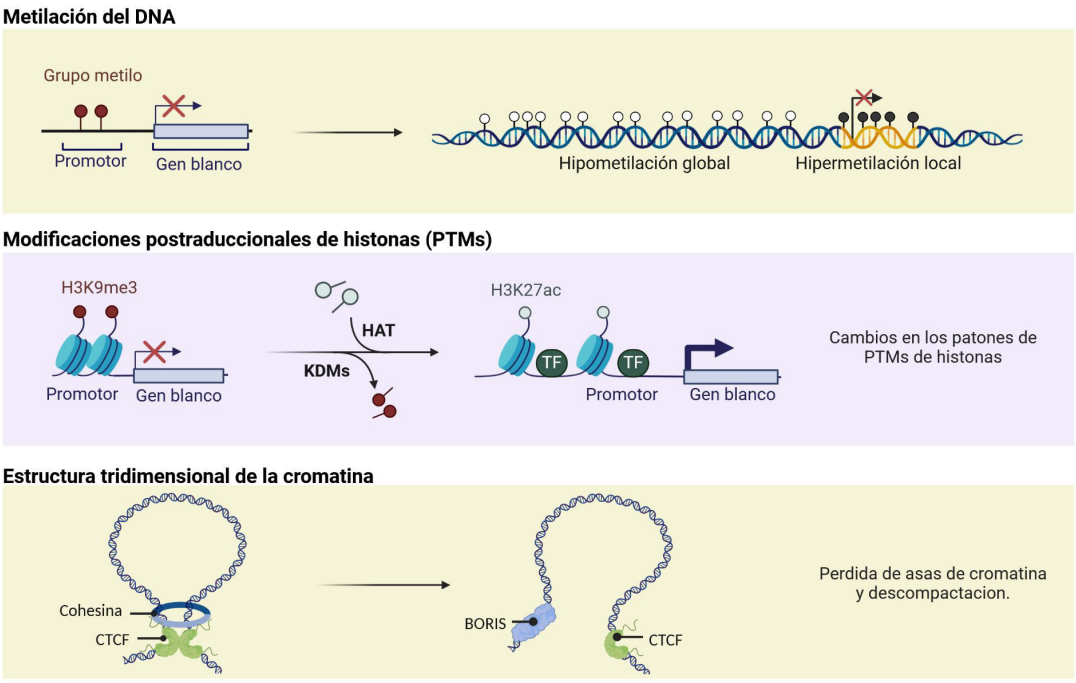


Figura 3.1: Principales procesos epigenéticos alterados en cáncer. DNMTs: Metiltransferasas de DNA, HAT: Acetiltransferasas de histona, KDMs: Desmetilasas de lisina.

disponibles van dirigidos contra todos los miembros de la familia DNMT, siendo nanaomycin el único compuesto específico para un solo miembro, DNMT3B (Georgina S. Daher-Reyes and Yee, 2019; Kuck et al., 2010).

Para las modificaciones postraduccionales de histonas (PTMs), existen varios inhibidores aprobados que se dirigen contra las desacetilasas de histonas (HDACs), tales como el vorinostat y la romidepsina, usados en el tratamiento del linfoma de células T (Wawruszak et al., 2021; Bertino and Otterson, 2011). Otros compuestos, como la tubacina y el ricolinostat, se encuentran en pruebas clínicas y han mostrado buenos resultados en diversos tipos de cáncer, incluyendo gliomas, cáncer de próstata y carcinoma hepatocelular (Biersack et al., 2022a,b). Otro grupo importante de epifármacos son los inhibidores de las proteínas BET, que se unen a histonas acetiladas por medio de sus bromodominios y a su vez regulan la unión de factores de transcripción al DNA. Hay varios compuestos dirigidos contra estas proteínas que se encuentran en pruebas clínicas y han mostrado actividad antitumoral y antiinflamatoria (Wang et al., 2023). La mayoría de los inhibidores han mostrado resultados prometedores en las pruebas clínicas con bajo porcentaje de resistencia a la terapia, sin embargo, los compuestos no son específicos para un sólo blanco, si no para varias proteínas BET por la similitud en sus bromodominios.

Llama la atención que nuevamente la mayoría de los fármacos aprobados o en pruebas clínicas están dirigidos a HDACs, DNMTs y las proteínas BET, mientras que pocas moléculas se usan como inhibidores de las enzimas metiltransferasas (KMT) y demetilasas (KDM). Entre las moléculas aprobadas se incluye al tazemetostat, que es un inhibidor de

Tabla 3.1: Fármacos aprobados por la FDA o en pruebas clínicas para su aprobación que tienen como blanco epifactores. DNMTs: Metiltransferasas de DNA, HDAC: Desacetilasas de histona, KDMs: Desmetilasas de lisina, KMTs: Metilasas de lisina, BET: Proteínas bromodominio y extraterminal i: todas las enzimas pertenecientes a ese grupo.

Blanco	**Fármaco**	**Estatus**	**Blanco específico**	**Referencia**
DNMT	Azacitidina	Aprobado	DNMTi	Kagan et al. (2023)
	Decitabina	Aprobado	DNMTi	Ye et al. (2024)
	SgI-110	En pruebas clínicas	DNMTi	Georgina S. Daher-Reyes and Yee (2019)
	CP-4200	En pruebas clínicas	DNMTi	Hummel-Eisenbeiss et al. (2013)
	Nanaomycin	En pruebas clínicas	DNMT3B	Kuck et al. (2010)
HDAC	Vorinostat	Aprobado	HDACi	Wawruszak et al. (2021)
	Romidepsina	Aprobado	HDACi	Bertino and Otterson (2011)
	Belinostat	Aprobado	HDACi	Lee et al. (2015)
	Parabinostat	Aprobado	HDACi	Laubach et al. (2015)
	Givinostat	Aprobado	HDACi	Ganai (2016)
	Trichostatin A	En pruebas clínicas	HDACi	Mogal and Abdulkadir (2006)
	Tubacin	En pruebas clínicas	HDAC6	Lu et al. (2017)
	MC1568	En pruebas clínicas	HDAC Clase II	He et al. (2022)
	MC1575	En pruebas clínicas	HDAC Clase II	Venza et al. (2013)
	Ricolinostat	En pruebas clínicas	HDAC6	Cosenza et al. (2017)
	AR-42	En pruebas clínicas	HDACi	Valencia et al. (2016)
BET	Jq1	En pruebas clínicas	BRDi	Miller et al. (2019)
	OTx015	En pruebas clínicas	BRD2, BRD3, y BRD4	Berthon et al. (2016)
	MS436	En pruebas clínicas	BRDi	Cipriano et al. (2022)
	ABBV-075	En pruebas clínicas	BRDi	Piha-Paul et al. (2019)
	I-BET151	En pruebas clínicas	BRDi	Zhao et al. (2013)
	NHWD-870	En pruebas clínicas	BRD4	Zhang et al. (2018)
KDM	Iadademstat	En pruebas clínicas	LSD1	Cuyàs et al. (2020)
KMT	Tazemetostat	Aprobado	EZH2	Hoy (2020)
	UNC0638	En pruebas clínicas	EHMT2	Mabe et al. (2020)
	Pinometostat	En pruebas clínicas	DOT1L	Menghrajani et al. (2019)
	EPZ004777	En pruebas clínicas	DOT1L	Premkumar and Shankar (2022)
	BIX-01294	En pruebas clínicas	EHMT2	Xu et al. (2021)
	GSK126	En pruebas clínicas	EZH2	Zeng et al. (2017)

EZH2, el núcleo catalítico del complejo represor de Polycomb 2 (PRC2) y que cataliza la metilación de H3K27, una marca de represión transcripcional. El tazemetostat está aprobado por la FDA para el tratamiento del sarcoma epitelioide y es específico para EZH2, aunque ha mostrado baja actividad contra su homólogo EZH1 (Hoy, 2020). Otros inhibidores, como UNC0638, pinometostat, BIX-01294 y GSK126, se encuentran en pruebas clínicas y han mostrado resultados prometedores tanto en cáncer como en otras enfermedades. En este caso, estas moléculas son específicas para sus blancos, y van dirigidas contra EZH2, DOT1L (cataliza la metilación de H3K9) y EHMT2 (cataliza la metilación de H3K79) (Menghrajani et al., 2019; Xu et al., 2021; Zeng et al., 2017).

En cuanto a las lisina desmetilasas (KDM), sólo hay un inhibidor aprobado, el iadademstat, que actúa sobre KDM1A y se encuentra en fase II de estudios clínicos para el tratamiento de leucemia mieloide aguda (Salamero et al., 2020), aunque también ha mostrado eficacia en la reducción del crecimiento de tumores xenográficos de cáncer de mama en ratones (Cuyàs et al., 2020). Sin embargo, actualmente se conocen más de 20 diferentes desmetilasas de histonas, cada una con diferentes afinidades por los residuos de lisina metilados en las histonas (Sterling et al., 2021), por lo que la búsqueda de moléculas inhibitorias para estas proteínas es un área de oportunidad. Más notable aún es la ausencia de fármacos para al restablecimiento de los dominios topológicamente asociados (TADs), a pesar de que éstos están frecuentemente alterados en múltiples neoplasias. En este sentido, una proteína que destaca como potencial blanco terapéutico es BORIS, un parálogo de CTCF (Klenova et al., 2002). BORIS es un prometedor blanco terapéutico para cáncer de ovario debido a que en tejido no neoplásico se expresa sólo en testículo, pero en diversos cánceres se ha reportado sobreexpresión de BORIS, especialmente en cáncer de ovario (Liu et al., 2019). A pesar de esto, el alto porcentaje de similitud que comparte con CTCF ha dificultado enormemente el desarrollo de terapias dirigidas a la inhibición de BORIS (Cheever et al., 2009).

Dado que los procesos epigenéticos están finamente regulados, es deseable tener inhibidores farmacológicos específicos para reducir los posibles efectos secundarios. La mayoría de los compuestos tiene como blanco a varias proteínas de la misma familia, con sólo unos cuantos ejemplos de compuestos específicos para un sólo blanco. Sin embargo, los procesos epigenéticos usualmente involucran proteínas con altos porcentajes de identidad entre sí, por lo que es comprensible que sea difícil encontrar fármacos específicos para un solo epifactor. Como ejemplo de ello, se encuentran las enzimas demetilasas de lisina KDM4. Las proteínas KDM4 son prometedores blancos terapéuticos para el cáncer debido a su implicación en procesos epigenéticos tales como la desmetilación de las marcas H3K36 y H3K9, asociadas con la transcripción activa y la heterocromatina (represión transcripcional), respectivamente y su asociación con diferentes tipos de cáncer (Katoh and Katoh, 2004; Shin and Janknecht, 2007b; Labbé et al., 2013). A pesar de su potencial como blancos terapéuticos, encontrar inhibidores específicos para las enzimas KDM4 ha sido difícil debido a la similitud entre sus dominios catalíticos; hasta el momento se han identificado algunos fármacos con actividad inhibitoria de KDM4A, sin embargo, todos ellos tienen como blanco a los otros miembros de la subfamilia también (Lee et al., 2020). Por lo anterior, la disponibilidad de inhibidores específicos para cada enzima es crucial

debido a las variaciones en la expresión de diferentes miembros de la subfamilia KDM4 en distintos tipos de cáncer (del Moral-Morales et al., 2022).

3.3 Las desmetilasas de lisina KDM4

La metilación de histonas es una modificación postraduccional relacionada con múltiples funciones biológicas. La metilación ocurre principalmente en los residuos de arginina (R) y lisina (K). Las argininas pueden ser mono- o dimetiladas mientras que los residuos de lisina pueden ser mono-, di- o trimetilados; estas modificaciones están asociadas con la activación o represión transcripcional, dependiendo del residuo en el que se encuentren. Por ejemplo, la di y trimetilación en H3K4 está relacionada con una mayor expresión génica, mientras que la trimetilación en H3K9 y H3K27 está asociada con la represión transcripcional (Millán-Zambrano et al., 2022; Blanc and Richard, 2017). Debido a que la metilación de histonas es una modificación covalente, inicialmente se asumió que era estable e irreversible. Sin embargo, en 2004, se caracterizó la primera desmetilasa de lisina de histona (KDM) y, desde entonces, se han descrito más de 20 enzimas que pueden remover esta modificación postraduccional (Shi et al., 2004; Guerra-Calderas et al., 2015). Estas enzimas actúan como activadores o represores transcripcionales, dependiendo del residuo de lisina que tengan como blanco y del contexto de la cromatina (Katoh and Katoh, 2004; Shin and Janknecht, 2007b).

Las desmetilasas de lisina de histonas (KDM) se clasifican en dos grandes familias según su mecanismo de acción: las similares a la amino-oxidasa y las oxigenasas (Sterling et al., 2021). La primera familia, más reducida, incluye a KDM1A y KDM1B, ambas dependientes de FAD y capaces de eliminar grupos metilo mono y dimetilados de las histonas. Sin embargo, su mecanismo catalítico les impide desmetilar lisinas trimetiladas (Anand and Marmorstein, 2007; Yang et al., 2006).

La familia de las oxigenasas, por su parte, es mucho más diversa y abarca a más de 20 enzimas que son dependientes de α-cetoglutarato. Dentro de esta familia, destaca la subfamilia KDM4, compuesta por cinco miembros (KDM4A a E) que desempeñan un papel crucial en diversos procesos cancerígenos. Estas enzimas se especializan en remover grupos metilo de las lisinas H3K9 y H3K36, marcando así regiones de la cromatina asociadas a la heterocromatina y a la transcripcion activa, respectivamente. Es importante destacar que las KDM4 muestran una mayor afinidad por la lisina H3K9 trimetilada (H3K9me3) en comparación con la dimetilada (H3K9me2), y prefieren desmetilar H3K9 en lugar de H3K36 (Labbé et al., 2013; Meng et al., 2018). Los valores de las constantes de Michaelis para cada miembro de la subfamilia KDM4 se detallan en el Apéndice A, lo que permite una comparación cuantitativa de sus actividades enzimáticas.

Todos los miembros de esta subfamilia poseen un dominio Jumonji N y un Jumonji C (JmjC), siendo este último el dominio catalítico. KDM4A, B y C poseen además un doble dominio PHD y un doble dominio Tudor, los cuales reconocen modificaciones

postraduccionales de histonas adyacentes al sitio de desmetilación, por lo que se cree que son responsables de guiar a las KDMs hacia sus sustratos (**Figura 3.2a**). A pesar de estas diferencias, los dominios catalíticos de las proteínas KDM4 presentan una alta identidad de secuencia, oscilando entre un 74 a 93 % de identidad en su dominio catalítico (**Figura 3.2b**). Estas proteínas encierran un ion Fe^{2+} y un Zn^{2+} en su sitio catalítico (**Figura 3.2c**), además de que utilizan α-cetoglutarato como cosustrato (Klose et al., 2006; Meng et al., 2018).

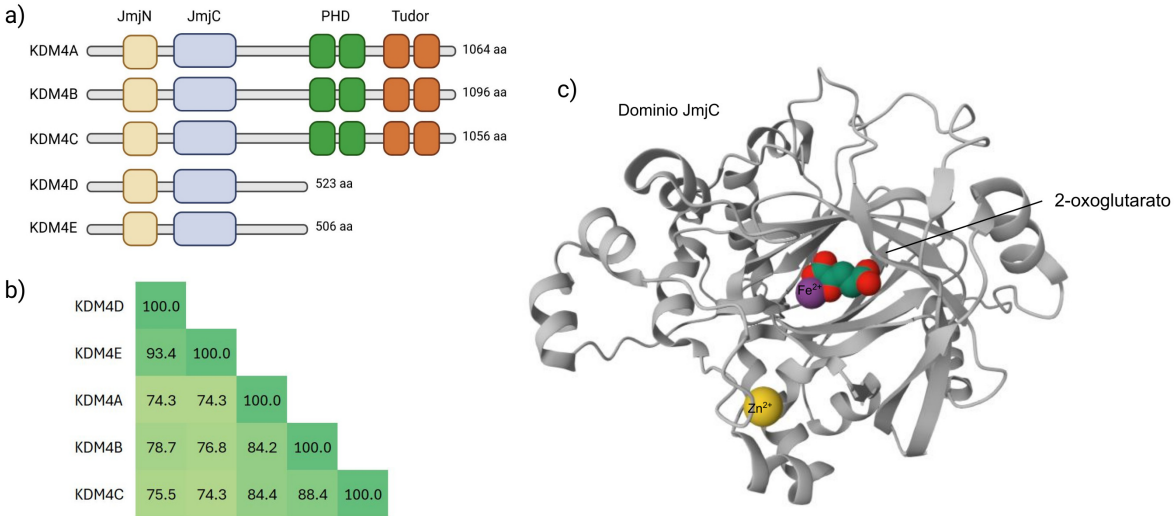


Figura 3.2: Estructura de las enzimas pertenecientes a la subfamilia KDM4. a) Distribción de dominios en los miembros de la subfamilia KDM4. JmjN: jumonji N. JmjC: jumonji C. PHD: Plant homeodomain. **b)** Porcentaje de Identidad entre los dominios JmjC de los miembros de la familia KDM4

Al ser importantes reguladores de la transcripción, las enzimas KDM4 se han visto involucradas en diferentes procesos celulares tales como la diferenciación neuronal (Zhu et al., 2024) y la neurogénesis asociada con ansiedad y depresión en modelos murinos. (Maitra et al., 2020). En este sentido, se ha observado que un aumento en la expresion de KDM4D induce una diferenciación neuronal a través de la inducción de la expresión de genes como *ID2* y *SOX2*. A su vez, las enzimas KDM4s desempeñan roles cruciales en diversos procesos biológicos como la espermatogénesis (Iwamori et al., 2011), además de regular la transcripción génica en general por su papel de desmetilasas de H3K9me3 (Guerra-Calderas et al., 2015). Por otro lado, hasta hace poco, KDM4E se consideraba un pseudogén debido a sus bajos niveles de expresión; sin embargo, informes recientes indican que codifica una enzima activa involucrada en la desmetilación de H3K9me3 y que está involucrada en el proceso de desarrollo embrionario (Liu et al., 2018).

Las proteínas KDM4 han emergido como blancos terapéuticos prometedores en el contexto oncológico debido a su papel en la proliferación y progresión tumoral (Rotili and Mai, 2011). Estudios han demostrado que la sobreexpresión y actividad de los diferentes miembros de esta familia de demetilasas de histonas está asociada con diversos tipos de cáncer. KDM4A se ha vinculado a la proliferación en leucemia mieloide aguda, cáncer de mama, próstata y pulmón (Massett et al., 2021; Metzger et al., 2017; Mu et al., 2019; Guerra-Calderas et al., 2015). KDM4B promueve la proliferación y el crecimiento tumoral en cáncer de mama positivo a estrógenos, y se asocia con un mal pronóstico en cáncer

gástrico resistente a la castración y osteosarcoma (Yang et al., 2010; Kawazu et al., 2011; Sha et al., 2020; Liu et al., 2020b). KDM4C fomenta la proliferación en mieloma múltiple y glioblastoma, además de ser crucial para el mantenimiento de células troncales cancerosas en carcinoma de células escamosas (Lv and Liu, 2021; Lee et al., 2021; Yuan et al., 2016b). KDM4D contribuye a la carcinogénesis en cáncer de próstata resistente a la castración, actúa como represor de la expresión de p53 en cáncer colorectal y promueve la proliferación de células troncales cancerosas en cáncer de hígado Shin and Janknecht (2007a); Li et al. (2020a); Deng et al. (2021). La **Figura 3.3** y la **Tabla 3.2** muestra un resumen de las principales funciones de las enzimas KDM4 tanto en tejido no neoplásico como en cáncer.

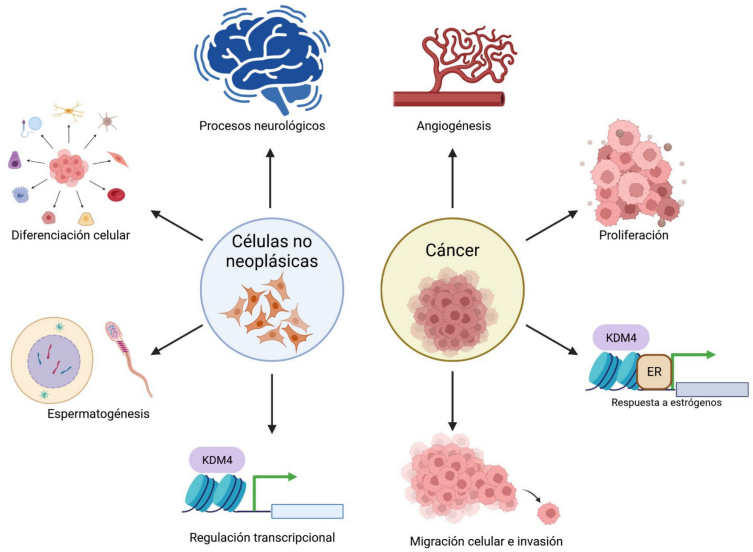


Figura 3.3: Principales procesos en los cuales participan las enzimas de la subfamilia KDM4, tanto en tejido no neoplásico como en cáncer.

3.3.1 Inhibidores de KDM4

A pesar de que las proteínas KDM4 son prometedores blancos para la terapia contra el cáncer, actualmente no hay inhibidores aprobados para su uso clínico. Existen varios compuestos con actividad de inhibición de KDM4 que se usan en investigación (ver Lee et al. (2020) para una revisión completa). Sin embargo, todos ellos apuntan a más de un miembro de la familia debido a la similitud de sus dominios catalíticos.

Si bien hasta el momento la FDA no ha aprobado ningún medicamento que tenga como blanco a los miembros de la subfamilia KDM4, existe un grupo de moléculas que se han catalogado como inhibidores de estas enzimas y que se usan en investigación. Estos compuestos se dividen en 4 grupos de acuerdo a su mecanismo de acción, los cuales se describirán brevemente a continuación (Lee et al., 2020; Baby et al., 2021).

Disruptores de cofactores metálicos

Tabla 3.2: Principales estudios sobre el papel de la subfamilia KDM4 en cáncer.

Enzima	Tipo de cáncer	Papel	Referencias
KDM4A	Leucemia mieloide aguda	Sobreexpresión y aumento en proliferación	Guerra-Calderas et al. (2015) Massett et al. (2021)
	Cáncer de mama	Sobreexpresión y aumento en proliferación	Metzger et al. (2017)
	Cáncer de próstata	Sobreexpresión y aumento en proliferación	Mu et al. (2019) Shin and Janknecht (2007a)
	Cáncer de pulmón	Sobreexpresión	Guerra-Calderas et al. (2015)
KDM4B	Cáncer de mama positivo a estrógenos	Promueve proliferación y crecimiento tumoral	Yang et al. (2010) Kawazu et al. (2011)
	Cáncer gástrico resistente a la castración	Mal pronóstico	Sha et al. (2020)
	Osteosarcoma	Mal pronóstico	Liu et al. (2020b)
KDM4C	Mieloma múltiple	Promueve proliferación	Lv and Liu (2021)
	Glioblastoma	Promueve proliferación	Lee et al. (2021)
	Carcinoma de células escamosas	Mantenimiento de células troncales cancerosas	Yuan et al. (2016b)
KDM4D	Cáncer de próstata resistente a la castración	Contribuye a la carcinogénesis	Shin and Janknecht (2007a)
	Cáncer colorectal	Actúa como represor de p53	Li et al. (2020a)
	Cáncer de hígado	Promueve la proliferación de células troncales cancerosas	Deng et al. (2021)

En esta categoría se encuentran dos de los compuestos más ampliamente usados como inhibidores de KDMs, el disulfiram y el ebselen, los cuáles inhiben KDM4A a través de la disrupción del ion Zn^{2+} en su sitio catalítico. Estas moléculas parecen ser específicas para la subfamilia KDM4 debido a que el resto de desmetilasas no tienen Zn^{2+} en su sitio catalítico (Rotili and Mai, 2011). Sin embargo, estos fármacos pueden actuar sobre otras proteínas que contengan zinc, lo que reduce su especificidad. El ebselen es conocido por sus propiedades antioxidantes, por lo que se ha usado en pruebas clínicas para varias enfermedades tales como asma, diabetes y contra parásitos gastrointestinales (Wang et al., 2020). Por otra parte, el disulfiram es un fármaco aprobado por FDA para el tratamiento del alcoholismo, sin embargo, recientemente se ha reportado que también posee actividad antitumoral, aunque al ser un quelante de metales, se une inespecíficamente a varias proteínas y iones metálicos libres en el ambiente extracelular (Meraz-Torres et al., 2020).

Análogos de 2-oxoglutarato

Otros inhibidores conocidos de KDM4 son análogos de 2-oxoglutarato, estas moléculas actúan como inhibidores competitivos al unirse al Fe^{2+} del sitio activo y desplazando al α-cetoglutarato requerido para la actividad de las desmetilasas de la familia JMJC, sin embargo, dado que el α-cetoglutarato es un cofactor para varias otras enzimas incluidas la HDACs y las TETs, estas moléculas tienen baja especificidad (Baby et al., 2021). Ejemplo de esto es el fármaco vorinostat, que actualmente se encuentra aprobado por la FDA para el tratamiento de leucemia mieloide y cuyo principal blanco son las HDACs (Wawruszak et al., 2021).

Inhibidores competitivos del sustrato de histonas

En esta categoría se encuentran las moléculas que comparten similitud con los residuos de lisina trimetilados que son blanco de la subfamilia KDM4. Estas moléculas son péptidos modificados que compiten con el sustrato de histona metilado. En este sentido, los compuestos de esta sección no sólo van dirigidos contra el dominio JmjC de las demetilasas, si no que algunos se unen a los dominios Tudor, inhibiendo la actividad de las demetilasas (Upadhyay et al., 2018). Los compuestos que conjuntan análogos del sustrato de histona junto análogos de α-cetoglutarato son los que han demostrado mejor capacidad para discriminar entre diferentes miembros de la subfamilia KDM4 (Woon et al., 2012).

Inhibidores independientes de sustrato o cofactor

Dado que el mecanismo de catálisis en los miembros de la subfamilia KDM4 es compartido entre ellas y con otro tipo de enzimas, los fármacos dirigidos al sitio catalítico tienen mayores probabilidades de ser inespecíficos (Lee et al., 2020), por lo tanto, una de las estrategias que se pueden abordar es el uso de moléculas que puedan unirse a la superficie externa de las proteínas, este es el caso del péptido 22, el cuál fue diseñado para unirse a la superficie de dimerización de KDM4C y que demostró actividad inhibitoria no competitiva (Leurs et al., 2014).

La **figura 3.4** muestra la clasificación y estructuras de los fármacos descritos en esta sección. En resumen, la mayoría de los inhibidores descritos tienen como blanco a todos o varios miembros de la subfamilia KDM4. Debido a que hay tipos de cáncer que muestran desregulación de solo un miembro de la familia KDM4 (Sterling et al., 2021), es importante lograr inhibidores específicos y efectivos para cada enzima. Además, la mayoría de los inhibidores de KDM4 reportados hasta la fecha solo han mostrado actividad *in vitro* (Chin and Han, 2015), por lo tanto, todavía falta la validación de fármacos que puedan ser utilizados en la terapia contra el cáncer.

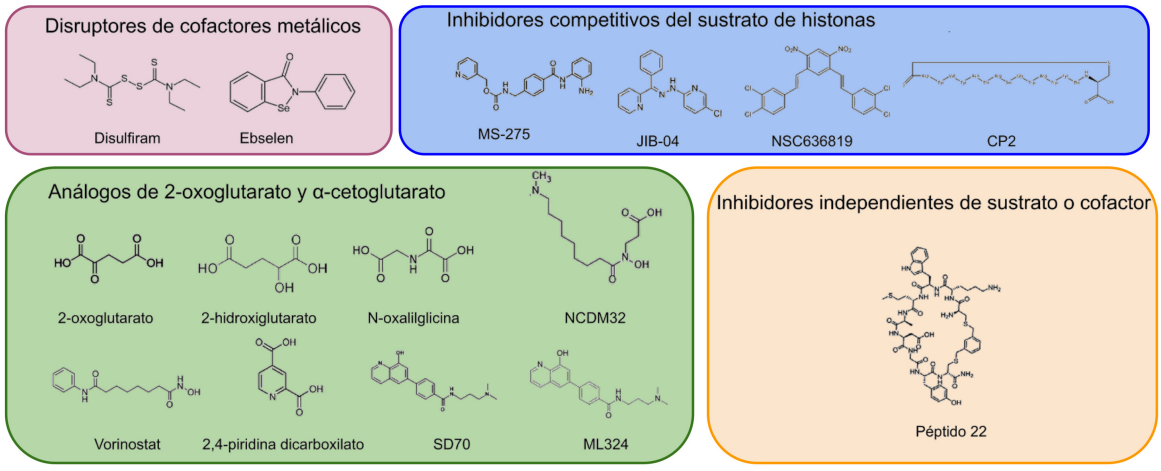


Figura 3.4: Moléculas que han demostrado actividad inhibitoria de las enzimas de la subfamilia KDM4.

3.4 Las proteínas BORIS y CTCF

CTCF es una proteína de unión al DNA que tiene decenas de miles de sitios de unión en el genoma, algunos de los cuales están conservados entre especies y tejidos (Schmidt et al., 2012). Las funciones principales de CTCF incluyen mantener los dominios asociados topológicamente (TADs), actuar como una barrera para la propagación de estructuras heterocromáticas y definir los límites entre la eucromatina y la heterocromatina; por esta razón, CTCF ha sido denominada una proteína arquitectónica (Fiorito et al., 2016; Kaczmarczyk et al., 2022; Kentepozidou et al., 2020; Ong and Corces, 2014). CTCF también regula el anclaje del DNA a estructuras celulares como la lámina nuclear (Fiorito et al., 2016; Kaczmarczyk et al., 2022), actúa como un aislante proteico controlando las interacciones entre enhancers y promotores (Ren et al., 2017), y puede funcionar como una proteína estructural para factores de transcripción (Schwalie et al., 2013; Chernukhin et al., 2000; Defossez et al., 2005) y factores epigenéticos (Wei et al., 2020). Basado en la localización de CTCF en otros sitios genómicos, también se ha demostrado que está involucrado en procesos como el splicing alternativo al pausar el recorrido de la RNA Polimerasa II (RNAP II), favoreciendo el ensamblaje del spliceosoma (Shukla et al., 2011). CTCF también interactúa con lncRNAs, lo cual es importante para la regulación transcripcional de genes como Xist, un lncRNA responsable de la inactivación del cromosoma X. Por esta razón, CTCF ha sido considerada una proteína muy versátil, similar a una navaja suiza. Un resumen de sus funciones se muestra en la **Figura 3.5**.

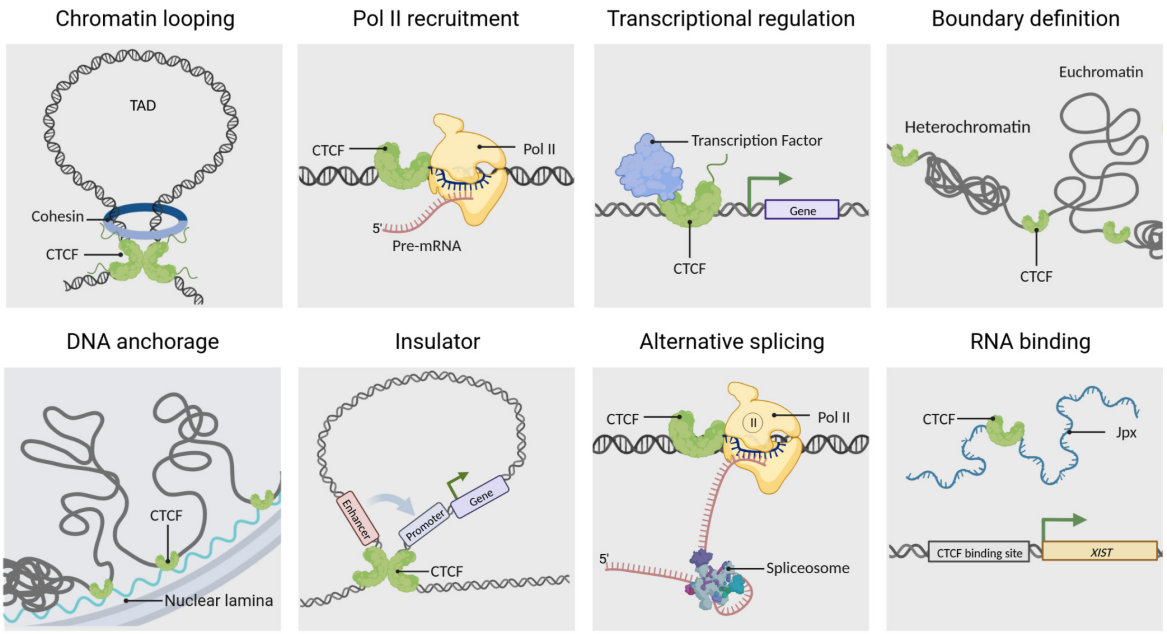


Figura 3.5: El factor arquitectónico CTCF. Resumen del amplio rango de mecanismos de acción de CTCF, tales como: formación de bucles de cromatina, reclutamiento de la RNA Polimerasa II (Pol II), regulación transcripcional, definición de límites entre eucromatina y heterocromatina, anclaje de DNA, aislante, splicing alternativo, unión a RNA, entre otros. TAD: dominio asociado topológicamente. Imagen de Del Moral-Morales et al. (2023).

Por otro lado, *CTCFL*, el gen que codifica la proteína conocida como Hermano del Regulador de Sitios Improntados (BORIS por sus siglas en inglés), es un parálogo de *CTCF*. Se cree que *CTCFL* surgió de un evento de duplicación en algún momento antes de la evolución de los mamíferos (Loukinov et al., 2002), sin embargo, su origen exacto sigue siendo incierto. A diferencia de *CTCF* que es un gen esencial y de expresión ubicua, *CTCFL* sólo se expresa en testículo bajo circunstancias normales, sin embargo, este gen ha ganado notoriedad en los últimos años porque se ha detectado en varias neoplasias, lo que sugiere un papel importante de esta proteína en el cáncer (Klenova et al., 2002; Liu et al., 2019).

Tanto BORIS como CTCF son proteínas de unión al DNA y están compuestas por tres dominios: el carboxi-terminal (C-terminal), el de unión al DNA y el amino-terminal (N-terminal). El dominio de unión al DNA comprende once dedos de zinc (ZF), con un porcentaje de identidad cercano al 80 %, mientras que la conservación entre CTCF y BORIS en los dominios N y C-terminal es inferior al 15 % (**Figura 3.6**). La similitud entre los dominios de unión al DNA de BORIS y CTCF sugiere que podrían unirse a los mismos sitios en el genoma (Klenova et al., 2002). Sin embargo, se ha demostrado que solo una parte de sus genes objetivo es compartida entre las dos proteínas (Nishana et al., 2020); además, CTCF se une a regiones intrónicas e intergénicas, mientras que BORIS se une preferentemente a promotores (Sleutels et al., 2012). Varios estudios han encontrado que BORIS y CTCF tienen consecuencias opuestas sobre la expresión génica (Vatolin et al., 2005; Nguyen et al., 2008; Hong et al., 2005); mientras que BORIS es un activador transcripcional que promueve la proliferación celular (Gaykalova et al., 2012; Smith et al., 2009), CTCF es un represor que inhibe la progresión del ciclo celular (Debaugny and Skok, 2020).

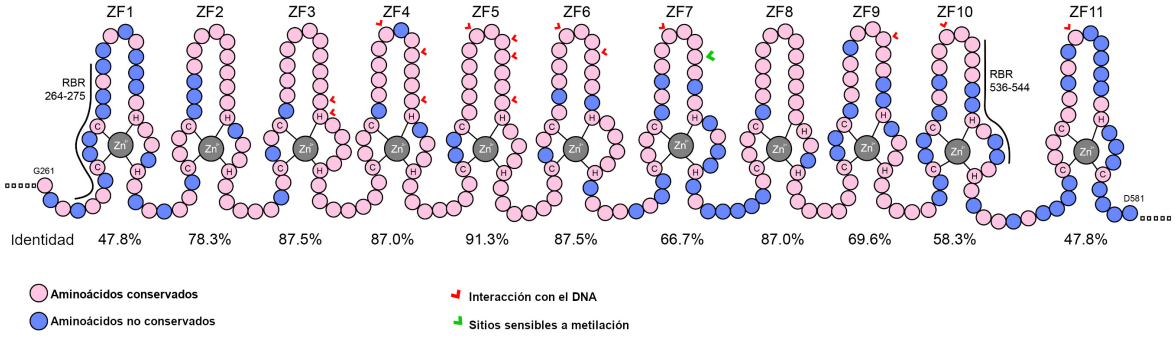


Figura 3.6: BORIS es un parálogo de CTCF. Representación gráfica del dominio de unión al DNA de CTCF. Cada círculo representa un aminoácido, y el color indica si está conservado en BORIS. Las flechas corresponden a los aminoácidos que forman puentes de hidrógeno con el DNA según (Yin et al., 2017). RBD significa RNA-binding domain (dominio de unión al RNA).

Debido al bajo grado de conservación entre los dominios terminales de BORIS y CTCF se ha sugerido que estos sirven como anclaje para distintos cofactores, lo que podría ser responsable de las diferencias entre los efectos de BORIS y CTCF sobre la expresión génica (de Necochea-Campion et al., 2011; Hore et al., 2008). Una de las principales funciones de CTCF es la retención de cohesina para establecer bucles de cromatina (Li et al., 2020b); se ha descrito que la interacción CTCF-cohesina está mediada por el dominio N-terminal y los

dos primeros dedos de zinc. La sustitución de esta región por la perteneciente a BORIS resulta en la pérdida de los anillos de cohesina en los sitios de unión a CTCF y por subsecuente, una pérdida de los bucles de cromatina establecidos (Pugacheva et al., 2020), apoyando aún más las funciones divergentes de estas dos proteínas y sugiriendo que BORIS puede competir por algunos de los sitios de unión a CTCF causando desregulación en la estructura 3D de la cromatina y un cambio en los patrones de expresión génica de las células (Nishana et al., 2020).

Pocos estudios han explorado el papel fisiológico de BORIS; sin embargo, los datos sugieren que la expresión de BORIS en tejidos no neoplásicos está restringida a los espermatocitos, donde podría estar implicado en el proceso de maduración de las células germinales masculinas a través del mantenimiento de los sitios de unión a CTCF durante el proceso de metilación *de novo* del DNA (Monk et al., 2008). También se ha observado que la expresión de BORIS y CTCF es mutuamente excluyente durante el proceso de espermatogénesis (Loukinov et al., 2002), el knock-down de CTCF induce la expresión de BORIS, y CTCF se une directamente al promotor de BORIS (Kholmanskikh et al., 2008) lo que en conjunto sugiere un bucle de regulación entre ambas proteínas.

3.4.1 BORIS como blanco terapéutico en cáncer

BORIS, un parálogo de CTCF, ha sido identificado como un marcador de diversos tipos de cáncer, incluyendo osteosarcoma, mama, pulmón y ovario (Ulaner et al., 2003; D'Arcy et al., 2008; Hong et al., 2005; Woloszynska-Read et al., 2007). Si bien su patrón de expresión tumoral es bien conocido, los mecanismos moleculares relacionados a su función oncogénica aún no están completamente esclarecidos.

Evidencias experimentales sugieren que BORIS desempeña un papel crucial en la regulación de la expresión génica en células cancerosas. Estudios han demostrado que BORIS modula la expresión de genes clave implicados en la tumorigénesis, como el receptor de andrógenos (Salgado-Albarrán et al., 2019), la telomerasa *hTERT* (Renaud et al., 2011), y antígenos cáncer-testículo como *MAGE-A1* y *NY-ESO-1* (Vatolin et al., 2005; Hong et al., 2005). Además, se ha observado que la sobreexpresión de BORIS promueve la migración, proliferación e invasión celular, mientras que su inhibición induce apoptosis (Dougherty et al., 2008; Zhang et al., 2017).

Sin embargo, el papel de BORIS en la proliferación celular parece ser más complejo de lo que inicialmente se pensaba. Algunos estudios han reportado una correlación positiva entre la expresión de BORIS y la proliferación celular (Gaykalova et al., 2012), mientras que otros han observado una reducción en la proliferación tras la inducción de BORIS (Tiffen et al., 2013; Rosa-Garrido et al., 2012). Estos resultados sugieren que el efecto de BORIS en la proliferación celular podría depender de diversos factores, como el tipo de célula, el contexto tumoral y los niveles de expresión de BORIS. Además, se ha propuesto que la regulación de la expresión de genes relacionados con la invasión, como *TGFB1*, podría ser un mecanismo adicional por el cual BORIS contribuye a la progresión tumoral (Janssen et al., 2020).

Por otra parte, BORIS ha ganado notoriedad en cáncer de ovario. Esto es debido a que, en tejido no neoplásico, la expresión de BORIS es indetectable en la mayoría de los tejidos, excepto en los testículos (Sleutels et al., 2012); sin embargo, se sobreexpresa en varias neoplasias, especialmente en el cáncer de ovario epitelial de alto grado (HGSC), donde se cree que desempeña un papel esencial en la progresión tumoral (Woloszynska-Read et al., 2007). El cáncer de ovario es la neoplasia ginecológica más agresiva y una de las principales causas de muerte en mujeres posmenopáusicas (Palmirotta et al., 2017). Dado que este tipo de tumores son en su mayoría asintomáticos, el diagnóstico es complicado y, por lo general, se detectan cuando la enfermedad ya se encuentra en un estadio avanzado (estadio III o IV), lo que hace que la terapia estándar no tenga éxito. Por lo tanto, la búsqueda de nuevas estrategias terapéuticas es imperativa para lograr una mejoría de las pacientes con cáncer de ovario.

Lo anterior ha llevado a pensar que una terapia dirigida contra esta proteína tendría un impacto significativo en el cáncer de ovario, por esta razón, BORIS es un antígeno tumoral de alta prioridad de acuerdo con el Instituto Nacional del Cáncer (NCI, Estados Unidos) (Cheever et al., 2009). Se ha observado que BORIS se relaciona positivamente con el estadio de la enfermedad (Woloszynska-Read et al., 2011; Gong et al., 2019), y con la malignidad en tumores HGSC (Salgado-Albarrán et al., 2019). Además, los altos niveles de BORIS se asocian con una baja supervivencia libre de recurrencia y una corta supervivencia global en pacientes con cáncer de ovario (Hillman et al., 2019; Zhang et al., 2022a).

Estudios previos han intentado desarrollar una vacuna contra los dominios N y C-terminal de la proteína para probar su eficacia contra el crecimiento de tumores mamarios implantados en ratones. En estos estudios se observó que la inmunización disminuía el crecimiento tumoral y aumentaba la supervivencia de los ratones inmunizados sin afectar al resto de tejidos somáticos, lo que sugiere que BORIS es una prometedora diana terapéutica en cáncer (Ghochikyan et al., 2007; Mkrtichyan et al., 2008, 2011). Sin embargo, no se ha explorado la posibilidad de inhibir la actividad de BORIS mediante métodos farmacológicos específicos. En este sentido, una de las principales limitantes es el porcentaje de similitud que comparte con CTCF, ya que al ser esta una proteína esencial, su inhibición eventualmente causaría muerte celular (Klenova et al., 2002). Aunado a esto, BORIS tiene 17 isoformas, las cuales cuentan con diferentes combinaciones de dedos de zinc, lo que añade una capa más de complejidad al problema de inhibición farmacológica de esta proteína.

3.4.2 Las isoformas de BORIS

La expresión de BORIS está controlada por tres promotores diferentes: A, B y C, los cuales son regulados por la unión a CTCF y la metilación del DNA. Hay una isla CpG en el promotor C y dos sitios de unión a CTCF, uno en el promotor A y otro en el promotor B (**Figura 3.7**). En testículos, CTCF no está unido y la isla CpG no está metilada, por lo que todos los promotores tienen actividad transcripcional. En el resto de tejidos, la isla CpG está metilada y CTCF está unido a sus dos sitios de unión; en consecuencia, la expresión de

BORIS está reprimida. En cáncer, existen diferentes combinaciones de promotores activos. Renaud et al. (2007) dividieron los tumores que expresan BORIS en dos grupos: B y A/C. Los tumores del grupo B expresan BORIS sólo a partir del promotor B, mostrando desmetilación de la isla CpG y unión de CTCF a los promotores A y C. Por otro lado, el grupo A/C presenta metilación de la isla CpG y unión de CTCF al promotor B (Renaud et al., 2007).

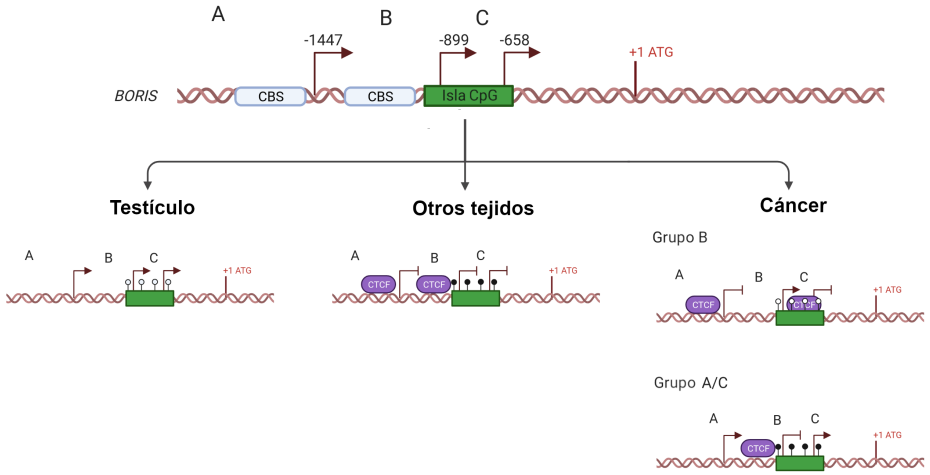


Figura 3.7: La expresión de BORIS está controlada por tres promotores. Representación gráfica de los promotores de BORIS. CBS significa CTCF Binding Site (sitio de unión a CTCF). El color del círculo representa el estado metilado de la isla CpG, blanco significa no metilado y negro metilado. Adaptado de Renaud et al. (2007).

Hasta ahora hay 23 isoformas diferentes descritas para BORIS. El grupo de Lobanenkov las ha dividido en tres familias según el promotor que regula su transcripción (A, B o C) y en seis subfamilias (sf1-6) según su extremo 3'-UTR (véase **Apéndice B**). Debido a sus similitudes, ha sido imposible evaluar cada isoforma individualmente; sin embargo, es posible medir los niveles de expresión de cada subfamilia mediante RT-qPCR. Las seis familias están presentes en testículo, mientras que en cáncer se detectaron dos patrones principales, el primero consiste en una alta expresión de sf1 únicamente, y el segundo se caracteriza por una alta expresión de todas las isoformas excluyendo la subfamilia sf5. Ninguno de estos grupos se asoció a un tipo de cáncer específico (Pugacheva et al., 2010). En el cáncer de ovario, se ha detectado la expresión de las subfamilias sf2, sf3, sf4 y sf6; siendo sf2 y sf6 las más abundantes (Link et al., 2013). Desafortunadamente, se desconoce el impacto biológico del patrón de expresión de las isoformas de BORIS.

Las 23 isoformas codifican para 17 proteínas diferentes (**Figura 3.8**). BORIS B0 es la proteína canónica y contiene los dominios N-terminal, Zinc Finger y C-terminal. La mayoría de las isoformas tienen un dominio C-terminal más corto que la proteína canónica. Existen diferentes combinaciones de dedos de zinc entre las isoformas, sin embargo, cabe mencionar que la isoforma C6 es la única que tiene un ZF híbrido (mitad ZF4, mitad ZF9). Como ya se ha mencionado, la funcionalidad de estas proteínas no ha sido explorada, sin embargo, se ha planteado la hipótesis de que las diferentes combinaciones de ZF y dominios terminales podrían permitir a BORIS unirse a un amplio conjunto de motivos de DNA.

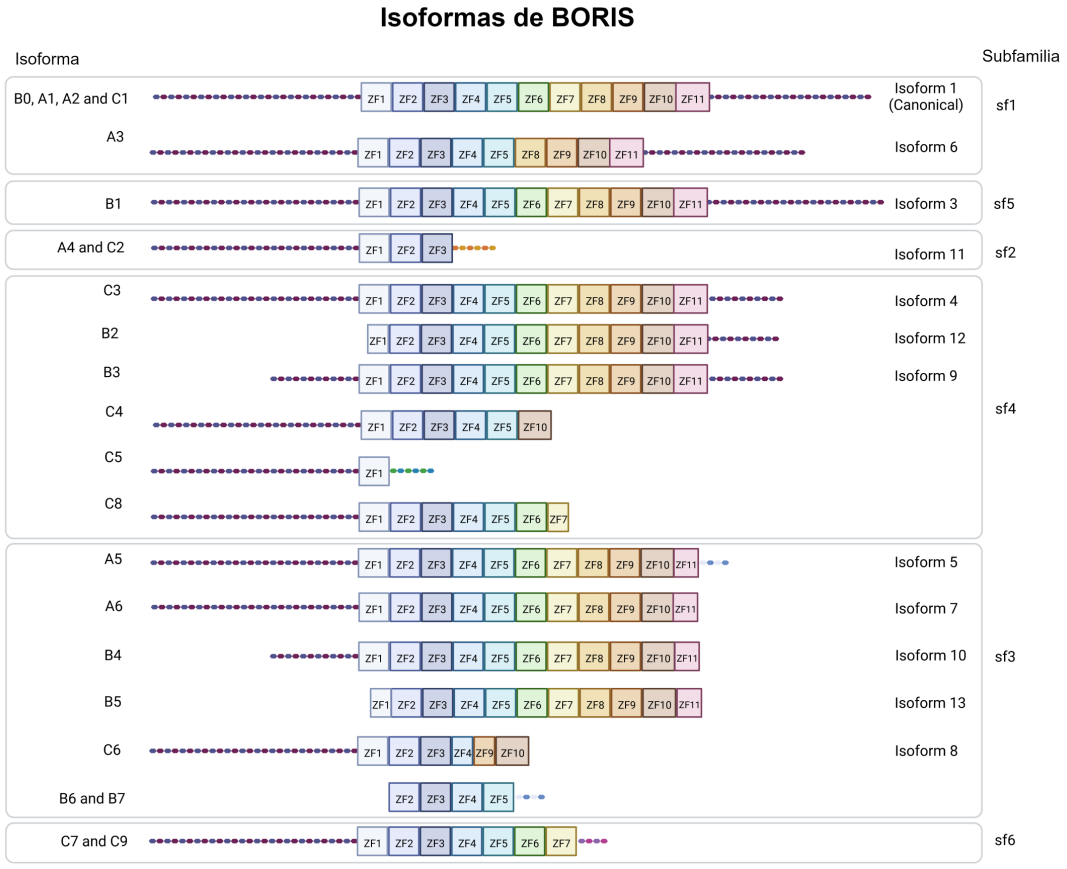


Figura 3.8: Representación esquemática de las 17 isoformas BORIS. El nombre del transcrito asignado por citePugacheva2015 se muestra a la izquierda. El número de isoforma de la derecha corresponde al registro disponible en la base de datos de genes de NCBI (NCBI gene, 2025).

4. Planteamiento del problema

Los procesos epigenéticos que participan en cáncer involucran diversas proteínas que comparten altos porcentajes de identidad entre sí. Esto ha dificultado el desarrollo de inhibidores farmacológicos específicos que puedan ser llevados a la clínica. En particular, no existen estudios enfocados en encontrar inhibidores específicos para las desmetilasas de histonas o las proteínas involucradas en la organización tridimensional de la cromatina.

5. Hipótesis

Mediante enfoques bioinformáticos y de biología estructural, se podrán identificar moléculas candidatas a inhibidores farmacológicos específicos para los miembros de la subfamilia KDM4 y para BORIS, con mínima interferencia sobre sus proteínas homólogas.

6. Objetivos

6.1 Objetivo general

Identificar compuestos con potencial de unirse selectivamente al sitio activo de las proteínas KDM4 o al dominio de unión al DNA de BORIS, minimizando efectos sobre proteínas homólogas.

6.2 Objetivos específicos

1. Identificar las desmetilasas de lisinas de histonas (KDMs) que son relevantes en distintos tipos de cánceres empleando la base de datos de TCGA.
2. Identificar las firmas moleculares que favorecen la interacción de ligandos con el sitio catalítico de las KDM4.
3. Seleccionar compuestos que tengan alta probabilidad de unirse al sitio catalítico de cada uno de los miembros de la subfamilia KDM4 mediante acoplamiento molecular.
4. Proponer moléculas que puedan actuar como agentes terapéuticos desde una perspectiva transcriptómica, integrando los datos obtenidos mediante acoplamiento molecular.
5. Explorar las diferencias estructurales entre CTCF y BORIS para buscar sitios que puedan ser blanco de inhibidores específicos de BORIS.
6. Proponer compuestos que puedan inhibir la actividad de BORIS sin afectar a CTCF.

7. Metodología

7.1 Análisis de datos transcriptómicos de cáncer

7.1.1 Obtención de datos de expresión génica

Con el objetivo de evaluar los niveles de expresión génica de los miembros de la subfamilia KDM4 en muestras de tejido no neoplásico y tumorales, se descargaron las bases de datos GTEx y TCGA de Xena Browser (Goldman et al., 2020). Solo se tuvieron en cuenta para el análisis los tipos de cáncer con su contraparte de tejido no neoplásico disponible. La cantidad de muestras analizadas por cada tipo de cáncer se muestra en la **Figura 7.1** Los conteos en RSEM (RNA-Seq by Expectation Maximization) se utilizaron como entrada para el análisis de expresión diferencial (TcgaTargetGtex_gene_expected_count). RSEM es una estrategia para la normalización del número de lecturas obtenidas mediante RNAseq (Li and Dewey, 2011). Para convertir de RSEM a cuentas normalizadas se usó la siguiente fórmula:

$$Cuentas\ normalizadas = log_2(RSEM + 1)$$

7.1.2 Análisis de supervivencia

Para evaluar la asociación entre la expresión de los genes de la subfamilia KDM4 y la supervivencia global de los pacientes se utilizó la información clínica depositada en TCGA y el paquete Survival v3.2 a 11 para R (Therneau et al., 2000). Para la regresión de COX (COXPH), la asociación se consideró significativa si el valor p <0,05. La estimación COXPH <0 se etiquetó como ”buen pronósticoz la estimación COXPH >0 como ”mal pronóstico”. Posteriormente se dividieron los pacientes en cuartiles de acuerdo con sus niveles de expresión de cada KDM. Se realizaron el gráfico de Kaplan Meier y la prueba de rango logarítmico utilizando pacientes con expresión de KDM <Q1 (Bajo-KDM) y pacientes con expresión de KDM >Q3 (High-KDM). La diferencia en la supervivencia global entre los grupos se consideró significativa con un valor p <0,05.

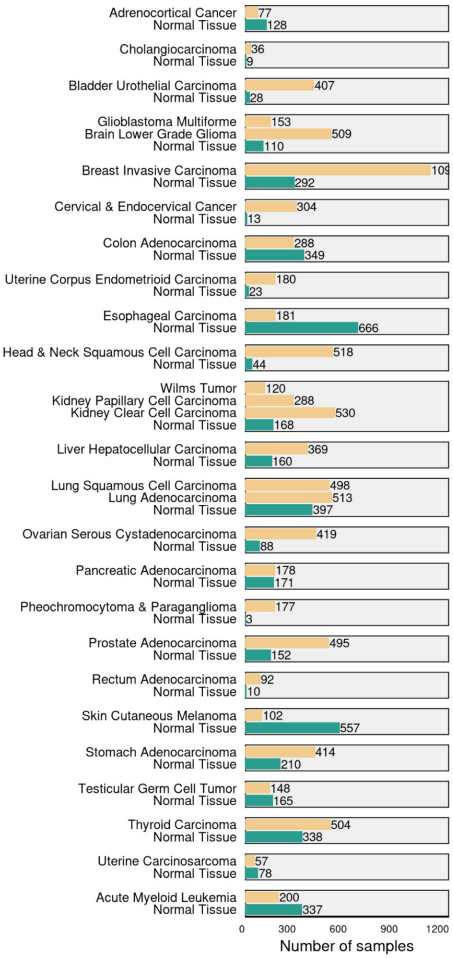


Figura 7.1: Muestras usadas para el análisis transcriptómico. Las muestras fueron obtenidas de las bases de datos TCGA y GTEx. El color verde denota a las muestras de tejido no neoplásico (Normal Tissue), el color amarillo a las muestras de tejido neoplásico.

7.1.3 Análisis de expresión diferencial

Con el objetivo de conocer las diferencias en la expresión de los diferentes miembros de la subfamilia KDM4 en los distintos tipos de cáncer, se utilizó el paquete DESeq2 v1.32.0 (Love et al., 2014) para comparar el perfil transcriptómico de muestras neoplásicas contra los controles no neoplásicos. También se realizó un análisis de expresión diferencial dentro de cada tipo de cáncer específico comparando dos grupos: pacientes con expresión de KDM >Q3 (Alto-KDM) frente a pacientes con expresión de KDM <Q1 (Bajo-KDM). Los genes con abs(log2FoldChange) >log2(1,5) y padj <0,05 se seleccionaron como genes expresados diferencialmente (DEG). Para este último análisis se eligieron las muestras etiquetadas como de ”mal pronóstico”por la prueba COXPH o aquellas en las que altos niveles de KDM mostraron una menor supervivencia en el gráfico de Kaplan Meier en comparación con las muestras de bajo KDM.

7.1.4 Análisis de enriquecimiento de DEGs

Para describir los procesos biológicos en los que podrían estar participando los DEGs, se realizó un análisis de enriquecimiento con gProfiler2 (Kolberg et al., 2020) y la base de datos Hallmark Gene Set Collection de la Molecular Signatures Database (MSigDB) (Liberzon et al., 2015). Este dataset se eligió debido a que tiene sets de genes no redundantes y que representan procesos biológicos primordiales. Se usó *g:SCS* como método de corrección por prueba múltiple (Reimand et al., 2007) y se consideró significativo los enriquecimientos con un valor p <0,05. Como fondo se utilizó la lista de todos los genes expresados en las muestras.

7.2 Cribado virtual de potenciales ligandos para la subfamilia KDM4

7.2.1 Bases de datos de ligandos

Para la identificación de posibles ligandos de las proteínas de la subfamilia KDM4, se usaron tres bases de datos: DrugBank (Wishart et al., 2018), la biblioteca de fármacos en Fase I y aprobados por la FDA (obtenida de www.selleckchem.com) y COCONUT (Sorokina et al., 2021a). Las bases de datos contenían 9,131, 3034 y 406,747 compuestos, respectivamente. Las bibliotecas se filtraron utilizando el algoritmo *FILTER* de OpenEye (OpenEye Scientific Software, disponible en www.eyesopen.com); los filtros aplicados se pueden encontrar en el **Apéndice C**. El estado de ionización de cada molécula se estableció mediante el algoritmo *FIXPKA* de OpenEye. Las cargas se calcularon con la herramienta *molcharge* de OpenEye y el método AM1-BCC (Jakalian et al., 2002). Se generaron diez conformadores de baja energía para cada molécula con el algoritmo *OMEGA*.

7.2.2 Preparación de las dianas farmacológicas

Las estructuras cristalográficas de los sitios activos de KDM4 se descargaron del Protein Data Bank (PDB, www.rcsb.org). Los números de acceso y las referencias de todos los modelos utilizados están disponibles en la **Tabla 7.1**. En los casos requeridos, las porciones faltantes de las moléculas se modelaron con SWISS-MODEL(Waterhouse et al., 2018). Las estructuras se prepararon para el acoplamiento con el programa *SPRUCE* incluido en la distribución OEDocking de OpenEye.

En el caso de KDM4D, se usaron nueve estructuras diferentes porque al alinear estructuralmente los modelos en PDB se notaron diferencias significativas (mayores a $2\,\text{Å}$), por lo que se analizaron por separado.

Tabla 7.1: Estructuras cristalográficas usadas en el análisis de acoplamiento molecular.

Enzima	PDB ID	Referencia
KDM4A	5F32	Le Bihan et al. (2016)
KDM4B	4LXL	Wang et al. (2014)
KDM4C	2XML	Yue et al. (2010)
KDM4D	4HON	Krishnan and Trievel (2012b)
	4HOO	Krishnan and Trievel (2012a)
	5F5C	Krojer et al. (2015)
	5FP4	Chung (2016b)
	5FP7	Chung (2016c)
	5FP8	Chung (2016d)
	5FPA	Chung (2016e)
	5FPB	Chung (2016a)
	6H10	Malecki et al. (2020)
KDM4E	2W2I	Yue et al. (2009)

7.2.3 Acoplamiento molecular (*Molecular Docking*)

Las estructuras de KDM4 se ajustaron mediante alineamiento estructural usando PyMol (DeLano et al., 2002) para mantener la misma orientación del sitio activo. El residuo conservado GLU190 se definió como punto de anclaje para el espacio de búsqueda. Se utilizó el campo de fuerza *Amber ff94* para el cálculo de las cargas parciales de la proteína. Se implementaron dos sistemas para comprender la influencia del metal sobre el sitio activo de las proteínas KDM4, HOLO y APO. La forma HOLO incluyó los cofactores Zn^{2+} en el sitio activo. Para el sistema APO, se eliminó cualquier cofactor metálico para el análisis. Cada base de datos química curada se acopló al receptor rígido en forma APO y HOLO con el programa FRED de OpenEye Scientific (OpenEye Scientific Software, 2021). Se utilizó la función de puntuación Chemgauss4, y las 100 moléculas con mayor puntuación en cada caso se consideraron posibles aciertos.

7.2.4 Análisis de similitud de flexóforos entre compuestos

El análisis de similitud de farmacóforo entre compuestos se realizó con DataWarrior (Sander et al., 2015) utilizando el descriptor *Flexophore*. Dos compuestos se consideraron similares si su relación de similitud superaba un umbral del 95 %. Para cada proteína y base de datos evaluada, se analizaron los 100 compuestos con mayor puntuación. Como se

utilizaron nueve estructuras PDB diferentes para KDM4D, se eligió una muestra aleatoria de 100 compuestos de cada base de datos.

7.2.5 Simulaciones de dinámica molecular y cálculos de energía de enlace absoluta

Los métodos de acoplamiento molecular son herramientas eficaces para el cribado de grandes bases de datos; sin embargo, su principal limitación son las estimaciones imprecisas de la energía de unión. Para estimar la energía de unión absoluta (ΔG_{PBSA}) de los ligandos se utilizó el método de mecánica molecular del área superficial de Poisson-Boltzmann (MM-PBSA). Dado que este enfoque requiere una gran cantidad de recursos computacionales, realizamos el cálculo de ΔG_{PBSA} sólo para un subconjunto de moléculas de acuerdo con los siguientes criterios:

a) Moléculas que se unen sólo a una KDM4.

b) Moléculas que se unen a alguna o a todas las KDM4 significativamente sobreexpresadas en un tipo de cáncer.

c) Las 10 mejores moléculas para cada KDM4 según su puntuación FRED/Chemgauss4, independientemente de la base de datos.

Brevemente, cada complejo proteína-ligando fue sometido a 20 ns de simulaciones de dinámica molecular utilizando GROMACS 5.1.15 (Abraham et al., 2015). Los archivos se procesaron con pdb2gmx, utilizando *AMBER99SB* como campo de fuerza y *TIP3P* como modelo de agua. Debido a la dificultad de simular los estados de coordinación del Zn^{2+}, todas las simulaciones se realizaron utilizando la forma APO. Se conservó la carga parcial AM1BCC obtenida con *MOLCHARGE* para cada ligando. Los parámetros de Van der Waals y topología de los ligandos se generaron con ACPYPE estableciendo *GAFF* como campo de fuerza (Sousa da Silva and Vranken, 2012). Los complejos se encerraron en una caja dodecaédrica con una distancia mínima caja-soluto de 1,0 nm, y la celda se llenó con agua. Cada sistema se equilibró utilizando las condiciones descritas previamente por Kumari et al. (2014). Tras el equilibrado, se llevó a cabo una corrida de producción de 20 ns. El ΔG_{PBSA} se calculó con la función *g_mmpbsa* de GROMACS.

7.3 Red de interacción fármaco-proteína-enfermedad

Para la red se tomaron los 100 fármacos mejor calificados de la base de datos DrugBank para cada KDM4 en el análisis de acoplamiento molecular. La herramienta NeDRex (Sadegh et al., 2021) se usó para recuperar las proteínas diana de cada fármaco utilizando las interacciones proteína-fármaco integradas en el paquete. Finalmente, se seleccionaron

los genes regulados al alza identificados en las comparaciones entre Alto KDM y Bajo KDM.

La red de interacción proteína-proteína utilizada como referencia se obtuvo de IID versión 2021-04 (Kotlyar et al., 2019), sólo se utilizaron los bordes validados experimentalmente (.^{ex}p", .^{ex}p;ortho", .^{ex}p;ortho;pred.^{o} .^{ex}p;pred"). Las redes se ensamblaron con KeyPathwayMiner (K=3 y L=0) (Alcaraz et al., 2020). Sólo se utilizaron como entrada las proteínas reguladas por los 100 mejores compuestos de la base de datos DrugBank para cada KDM. Los miembros de la subfamilia KDM4 expresados diferencialmente en cada tumor se definieron como nodos positivos. Las redes de interacción proteína-fármaco y proteína-proteína se fusionaron y editaron utilizando Cytoscape 3.8.2 (Shannon et al., 2003).

7.4 Modelado tridimensional de BORIS y CTCF

Hasta la fecha, no hay datos cristalográficos disponibles para BORIS, por lo tanto, es necesario generar un modelo 3D que más tarde se utilizará para el análisis de acoplamiento molecular. Se utilizó Modeller versión 10.1 (Šali and Blundell, 1993) para la creación de modelos de homología tridimensionales (3D) del dominio de dedos de zinc de BORIS, así como para fusionar los diferentes fragmentos cristalizados de CTFC en un modelo. Para BORIS la región modelada abarcaba desde la PRO284 hasta CYS568 (ZF2-ZF11) y para CTCF la región utilizada fue desde PRO293 hasta CYS577 (también ZF2-ZF11). En ambos casos, se utilizaron como plantillas las estructuras de rayos X de CTCF 5T0U (Hashimoto et al., 2017), 5YEF, y 5YEL (Yin et al., 2017) disponibles en Protein Data Bank (PDB) (**Figura 7.2**).

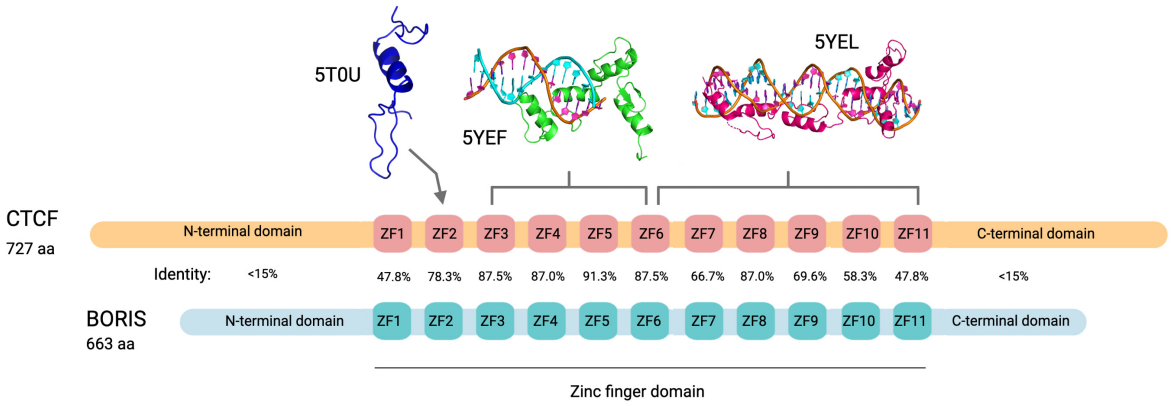


Figura 7.2: Representación esquemática de las estructuras utilizadas para modelar CTCF y BORIS.

Los nucleótidos en las cadenas de DNA de 5YEF y 5YEL se definieron como ligandos en el archivo de alineación (.), y los iones Zn^{2+} se declararon como *z*. Los contactos entre los aminoácidos y el Zn^{2+} se especificaron como restricciones especiales y se fijaron en una media de $2{,}3\,\text{Å}$ con una desviación estándar de 0,3. Se generaron mil modelos para cada

proteína y se seleccionó el que tenía la puntuación más baja de Energía Proteica Optimizada Discreta (DOPE por sus siglas en inglés) como el mejor (Eswar et al., 2006). LA secuencia de comandos y los archivos de alineación utilizados para generar los modelos se pueden encontrar en el **Apéndice D**.

La calidad estereoquímica de las mejores estructuras proporcionadas por Modeller se evaluó con los algoritmos ERRAT (Colovos and Yeates, 1993), PROCHECK (Laskowski et al., 2012), y PROVE (Pontius et al., 1996), del Structural Analysis and Verification Server (SAVES v6.0, https://saves.mbi.ucla.edu/). Se utilizó PyMOL Version 2.4.2 para la visualización y el análisis de desviación cuadrática media (RMSD por sus siglas en inglés) de las estructuras predichas.

Dado que Modeller no puede generar cadenas de DNA, sólo copiar las que están presentes en los templados, se construyeron dos nuevas cadenas de DNA complementarias con el servidor w3DNA (http://web.x3dna.org/) utilizando como plantillas las cadenas de DNA de 5YEF y 5YEL, esto permitió tener una doble hélice continua en lugar de los dos fragmentos obtenidos en el modelo original (Li et al., 2019b). La secuencia de la nueva molécula de DNA es la siguiente:

```
FW: 5'-TTGCAGTACCACATTTAACCAGCAGAGGGCGAGTCGACT-3'

RV: 3'-AACGTCATGGTGTAAATTGGTCGTCTCCCGCTCAGCTGA-5'
```

Para las isoformas BORIS, las secuencias de proteínas se extrajeron de (Link et al., 2013). Los archivos *.fasta* se utilizaron como entrada para el algoritmo AlphaFold2 (Jumper et al., 2021). Se seleccionaron los parámetros CASP14 y el modelo 0 se consideró el mejor en todos los casos.

Se utilizó la versión 10.1 de Modeller (Šali and Blundell, 1993) para construir el modelo del promotor BORIS-hTERT. Los modelos utilizados fueron: 5T0U (Hashimoto et al., 2017) para los residuos 293 a 345, 5KKQ (Hashimoto et al., 2017) para los residuos 317 a 461, y 5YEL (Yin et al., 2017) para los residuos 437 a 577, los dominios terminales (residuos 1-209 y 545 a 663) se modelaron utilizando AlphaFold2 (Jumper et al., 2021). Se construyó una cadena de DNA con el servidor w3DNA utilizando la secuencia del sitio de unión a CTCF en el promotor *hTERT*, que también se ha reportado como sitio de unión a BORIS (Renaud et al., 2007). AlphaFold2 también se utilizó para modelar dímeros CTCF-CTCF, CTCF-BORIS y BORIS-BORIS (Evans et al., 2021).

La secuencia de DNA utilizada para el modelo BORIS-*hTERT* es:

```
FW: 5'-CGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGG-3'

RV: 5'-CCTCGCGGTAGTGGCTGCGCAGCAGGGAGCGCACGGCTCG-3'
```

7.5 Simulaciones de Dinámica Molecular

Las estructuras obtenidas con MODELLER fueron refinadas y sometidas a minimización energética mediante GROMACS versión 2020.4 (Abraham et al., 2015). Previamente a las simulaciones, los residuos de cisteína que se unen a los iones Zn^{2+} fueron cambiados a su estado desprotonado (CYN) ya que se ha demostrado que las cisteínas cargadas negativamente son importantes para la correcta interacción con el Zn^{2+} y la estabilidad global del dedo de zinc (Nguyen et al., 2020). Los residuos de histidina que interactúan con el Zn^{2+} en cada dedo de zinc se declararon como HSD para que coincidieran con el estado protonado anteriormente descrito para las histidinas que se coordinan con iones Zn^{2+} (Li et al., 2008).

Los archivos se procesaron con *pdb2gmx* utilizando el campo de fuerza Charmm36-feb2021 (Huang et al., 2016) y tip3p como modelo de agua. El complejo se encerró en una caja dodecaédrica con una distancia mínima caja-soluto de 1,0 nm y la celda se llenó con agua y NaCl a una concentración total de 150 nM además de suficientes iones Na^+ para neutralizar la carga neta del complejo. En total, hay 373 residuos (58,3 kDa) en cada complejo, 285 son residuos de aminoácidos, 39 son residuos de DNA y 10 son iones de Zn^{2+}.

7.5.1 Minimización de energía

Para eliminar los contactos potencialmente de alta energía y los átomos solapados, los sistemas se sometieron a dos pasos de minimización de energía, una primera fase de minimización de energía se llevó a cabo con el algoritmo de descenso más pronunciado seguido de una segunda de gradientes conjugados. Los archivos *mdp* utilizados para cada ejecución se pueden encontrar en **Apéndice E.1** y **Apéndice E.2**.

A continuación, cada sistema se sometió a un equilibro de temperatura mediante el método *V-rescale*. Se utilizaron dos termostatos, uno para el complejo (Proteína, DNA y Zn^{2+}) y el otro para el agua, Na^+, y Cl^- iones. Se aumentó la temperatura poco a poco, iniciando en 50 K y aumentando 50 K cada 100 ps hasta alcanzar los 300 K. Un ejemplo del archivo *mdp* para cada ejecución se puede encontrar en **Appendix E.3**. El equilibrio de temperatura fue seguido por el equilibrio de presión durante 100 ps utilizando el termostato anterior y el método Berendsen para mantener la presión a 1 bar. El archivo *mdp* utilizado se encuentra en **Apéndice E.4**. Para ambos equilibrios, se aplicaron restricciones a las cadenas de proteína, Zn^{2+} y DNA. Tras el equilibrio, se llevó a cabo una simulación de producción de 500 ns utilizando el método *V-rescale* para el acoplamiento de temperatura y el barostato Parrinello-Rahman en ausencia de restricciones. Tanto las interacciones electrostáticas como las de van der Waals se limitaron a 1,2 nm y se aplicaron condiciones periódicas en las tres dimensiones. Todos los enlaces de hidrógeno se restringieron con el algoritmo LINCS. El archivo *mdp* utilizado para configurar las simulaciones se puede encontrar en **Appendix E.5**.

7.5.2 Análisis de las simulaciones de dinámica molecular

Los cálculos de la desviación cuadrática media (RMSD) y la fluctuación cuadrática media (RMSF) se realizaron utilizando el paquete Bio3D versión 2.4-1 para R (Grant et al., 2021), sólo se consideraron los átomos Cα de la proteína y los átomos O4' del DNA. Los enlaces de hidrógeno entre la proteína y el DNA se analizaron con la herramienta gromacs hbond, utilizando los parámetros por defecto (distancia entre los átomos donador y aceptor $3{,}5\,\text{Å}$ y ángulo entre los átomos implicados en el enlace de hidrógeno <30°). La estructura secundaria local se analizó con *gmx do_ssp*. La visualización de la estructura 3D se realizó con Chimera versión 1.15.

7.6 Análisis de acoplamiento molecular para BORIS y CTCF

Se emplearon tres algoritmos distintos para realizar simulaciones de acoplamiento o *docking* molecular de la base de datos de DrugBank contra las proteínas objetivo. Se usaron los algoritmos de docking PLANTS (Gorgulla et al., 2021) y AutoDock Vina (Eberhardt et al., 2021), mientras que como herramienta de *docking* ciego se usó DiffDock (Corso et al., 2022) acoplado al sistema de puntuación gnina (McNutt et al., 2021). Los parámetros y comandos utilizados se muestran en el **Apéndice F**. En el caso de los algoritmos dirigidos, la caja de búsqueda se centró en los dos sitios no conservados de BORIS identificados mediante el análisis de modelado tridimensional.

Cada algoritmo generó 10 diferentes poses para cada ligando, las cuales fueron clasificadas de los mejores a los peores puntajes. Para consolidar los resultados, se calculó un ranking consenso para la pose de mayor puntuación producida por cada herramienta, utilizando el método de "exponential consensus ranking". Se estableció un valor de corte predeterminado de 0.0015, designando a los ligandos con puntajes por encima de este umbral como posibles ligandos de las proteínas. Para refinar la selección, se excluyeron los ligandos que demostraron capacidad de unión tanto a CTCF como a BORIS, independientemente del sitio. Las moléculas restantes se propusieron como potenciales inhibidores de BORIS.

7.7 Vectores de expresión para BORIS y CTCF

Con el objetivo de sobreexpresar las proteínas BORIS y CTCF, se transfectaron células HEK-293 y SKOV3 con los siguientes vectores:

- **BORIS-GFP**
  CTCFL/BORIS cDNA ORF Clone, Human, C-GFPSpark® tag vector (cat. HG15751-ACG, SinoBiological)

Apéndice G.1

- **CTCF-GFP**
  CTCF cDNA ORF Clone, Human and C-GFPSpark® tag vector (cat. HG15017-ACG, SinoBiological)
  Apéndice G.2

- **CTCF-HA**
  CTCF cDNA ORF Clone, Human and C-HA tag vector (cat. HG15017-CY, SinoBiological)
  Apéndice G.3

- **BORIS-HA**
  BORIS cDNA ORF Clone, Human and C-HA tag vector (cat. HG15751-CY, SinoBiological)
  Apéndice G.4

Los plásmidos mencionados se transformaron en *E. coli* competente, se mezcló $0,1$µg de cada plásmido con 50µL de bacterias competentes. La mezcla se incubó en hielo durante 30 minutos y luego se sometió a un choque térmico de 42°C durante 30 segundos. Se añadieron 450µL de medio LB y las bacterias se incubaron a 37°C durante una hora. Por último, se extendió el medio en una placa de agar LB con kanamicina 50 µg/mL y se incubó toda la noche a 37°C. Se inoculó una colonia de cada placa en 200 mL de medio LB con el antibiótico correspondiente y se incubó toda la noche a 37°C y 250 rpm. Tras la incubación, se extrajeron los plásmidos con el kit QIAGEN® Plasmid Midi (cat. 12143) siguiendo las instrucciones del fabricante.

7.7.1 Cultivo celular y transfecciones

La línea celular HEK-293 derivada de riñón fetal humano y la línea SKOV3 se cultivaron en medio DMEM suplementado con un 10 % de suero bovino fetal a 37°C y un 5 % de CO_2. Todos los plásmidos se transfectaron con el reactivo de transfección Xfect (cat. 631317 Takara Bio, CA, EE.UU.) siguiendo las instrucciones del proveedor. Se sembraron 2x10^5 células y se transfectaron con $7,5$ µg de plásmido. Para las transfecciones transitorias, el medio de transfección se incubó durante la noche y las células se cosecharon 48 horas después de la transfección.

7.7.2 Western Blot

Tras los tratamientos, se obtuvo proteína total por lisis y se cuantificó utilizando el kit DC™ Protein Assay (cat. 5000112, Bio-Rad) 40µL de proteína total se separaron por electroforesis en un gel de SDS-poliacrilamida al 10 %. Las proteínas se transfirieron a una

membrana de PVDF a 50 V en cámara húmeda (4°C, toda la noche). Las membranas se bloquearon durante una hora a temperatura ambiente con leche baja en grasa al 5 %. Los anticuerpos primarios se incubaron como sigue: α-GFP 1:1000 (5246S, Cell Signaling, EE.UU.), α-BORIS 1:200 (cat. HPA001472, Sigma), anti-CTCF 1:750 (cat.N17, Santa Cruz Biotechnology) y anti-GAPDH (cat. sc-47724, Santa Cruz Biotechnology). El anticuerpo anti-GFP se incubó durante una hora (temperatura ambiente) en TBS tween 0,05 % y 5 % de leche baja en grasa. Los anticuerpos anti-BORIS y anti-CTCF se incubaron a 4°C, durante toda la noche en TBS tween 0,05 % y 0,5 % de leche baja en grasa. Los anticuerpos secundarios conjugados con peroxidasa de rábano utilizados fueron: anti-conejo de cabra (sc-2004, Santa Cruz Biotechnology, EE.UU.) y anti-cabra de burro (sc-2056, Santa Cruz Biotechnology, EE.UU.). Ambos se incubaron a temperatura ambiente durante 1 hora. Para detectar los complejos inmunes se utilizó el reactivo Super Signal West Femto Maximum Sensitivity Substrate (cat. 34096, Thermo Scientific). El análisis densitométrico se realizó con el programa Image J v.1.8.0.

7.7.3 Inmunofluorescencia

Se sembraron $2x10^5$ células SKOV3 en placas de 6 pocillos y se fijaron con formaldehído al 4 % durante 10 min. A continuación, las células se permeabilizaron con Triton X-100 al 0,1 % durante 5 minutos, seguido de una incubación de una hora con 1 ml de glicina 100 mM. A continuación, las células se incubaron durante 1 h a temperatura ambiente con el anticuerpo primario correspondiente y se lavaron con PBS. Por último, las células se incubaron con IgG anti-conejo de cabra Alexa Fluor 647 (21246, Santa Cruz Biotechnology, EE.UU.) a temperatura ambiente durante 60 min y se lavaron con PBS. Los núcleos se tiñeron con medio de montaje DAPI (Polysciences Inc., EE.UU.). Por último, las células se cubrieron con un cubreobjetos y se visualizaron en un microscopio Olympus Bx43. Los anticuerpos utilizados se muestran en la **Tabla 7.2**.

Tabla 7.2: Anticuerpos usados para Inmunofluorescencia.

Anticuerpo	Cantidad	Marca	Catálogo
BORIS (rabbit)	1:200	Sigma Aldrich	HPA001472
CTCF (rabbit)	1:200	Cell Signaling	3418S
anti-conejo de cabra Alexa Fluor 647	1:1000	Santa Cruz	21246

7.8 Interacciones proteína-proteína de BORIS y CTCF

En primer lugar, se realizó una búsqueda bibliográfica para encontrar las proteínas con evidencia experimental de interacción con BORIS o CTCF. Posteriormente, se modelaron

todas las posibles interacciones utilizando AlphaFold v2.3.0 (Evans et al., 2021), empleando el modelo de multímero. Para evaluar la confianza de las predicciones, se utilizó la puntuación DockQ con un umbral de 0,23, tal como se sugiere en el documento original (Bryant et al., 2022).

7.8.1 Inmunoprecipitaciones

Para la co-inmunoprecipitación (Co-IP) de BORIS-HA y CTCF-HA, se utilizó el kit Pierce™ Classic Magnetic IP/Co-IP siguiendo las instrucciones del fabricante. Se emplearon 500 µg de proteína total para cada reacción, obtenida de la transfección de células HEK-293 como se describió previamente. Los anticuerpos utilizados y cantidades se muestran en la **Tabla 7.3**.

Tabla 7.3: Anticuerpos usados para la inmunoprecipitación.

Anticuerpo	Cantidad	Marca	Catálogo
HA (mouse)	2 ug	Invitrogen	32-6700
IgG (mouse)	2 ug	Merk	12-371

Para el análisis por Western Blot, las muestras fueron separadas en geles NuPAGE™ Bis-Tris Mini Protein Gels al 4-12 % con un espesor de 1,5 mm. Se cargaron 30 µl de la muestra obtenida mediante IP mezclada con 10 µl de buffer de carga. Las muestras se calentaron a 72 °C durante 10 minutos antes de la carga. La transferencia se realizó durante 7 minutos a 20 V utilizando el sistema iBlot 2 de Thermo Fisher. Las membranas se bloquearon en TBS 1 % Tween con 5 % de leche libre de grasa. La membrana se incubó con el anticuerpo primario durante la noche a 4 °C, seguido de una incubación de 1 hora a temperatura ambiente con el anticuerpo secundario. En ambos casos los anticuerpos se diluyeron en TBS 1 % Tween con 5 % de leche libre de grasa. La detección se llevó a cabo utilizando el sustrato SuperSignal™ West Femto Maximum Sensitivity y el sistema de captura de imágenes ChemiDoc MP. Los anticuerpos y diluciones usadas se muestran en la **Tabla 7.4**.

Tabla 7.4: Anticuerpos usados para Western Blot.

Anticuerpo	Cantidad	Marca	Catálogo
BORIS (rabbit)	1:1000	Sigma Aldrich	HPA001472
CTCF (rabbit)	1:1000	Cell Signaling	3418S
TBP (rabbit)	1:1000	Cell Signaling	44059S
a-Rabbit	1:5000	Cell Signaling	7074P2

8. Resultados

8.1 Los miembros de la subfamilia KDM4 son prometedores blancos terapéuticos para compuestos naturales

8.1.1 La expresión de la subfamilia KDM4 tiene un mal pronóstico en el cáncer

Para abordar el papel de la subfamilia KDM4 en el cáncer, se realizó un análisis de expresión en un amplio conjunto de muestras tumorales y no neoplásicas disponibles públicamente. La comparación entre las muestras tumorales frente al tejido no neoplásico mostró que existen distintas combinaciones de KDM4s diferencialmente expresadas para cada tipo de tumor. KDM4D y KDM4A son los genes más notables, ya que están principalmente sobreexpresados, mientras que KDM4B y C suelen estar subexpresados en comparación con el tejido no neoplásico (**Figura 8.1a**). Para caracterizar la importancia clínica de los miembros de la subfamilia KDM4, se realizaron dos análisis de supervivencia. La sobreexpresión de KDM4A indica un mal pronóstico para varios tipos de cáncer, como carcinoma endometrioide del cuerpo uterino, carcinoma hepatocelular de hígado, cáncer adrenocortical, glioma de bajo grado y carcinoma uterino. La sobreexpresión de KDM4B es un mal pronóstico para el cáncer adrenocortical y el carcinoma tiroideo. La expresión de KDM4C se relaciona con un mal pronóstico para el adenocarcinoma de recto y el feocromocitoma y paraganglioma. KDM4D se relaciona con un mal pronóstico para el adenocarcinoma de pulmón, el cáncer adrenocortical y el carcinoma hepatocelular de hígado (**Figura 8.1b**). Finalmente, dado que la expresión de KDM4E es baja en la mayoría de las muestras evaluadas, no se consideró para el resto de los análisis.

A continuación, se estudió el papel de las proteínas KDM4 en los 12 tipos de cáncer donde la desregulación de algún miembro de la subfamilia es de mal pronóstico. Se realizó un análisis de expresión diferencial comparando sólo muestras tumorales clasificadas como de alta expresión de KDM (cuarto cuartil) usando como control aquellas con baja expresión de KDM (primer cuartil). La **Figura 8.2a** muestra el log2FC de cada KDM en las comparaciones hechas, así como el número de genes diferencialmente expresados (DEGs, $|\log2FC| > 1$ y p-adj < 0.05) en cada condición. El análisis de enriquecimiento de los DEGs contra la base de datos GSEA Hallmarks (una colección de procesos biológicos bien definidos y no redundantes) mostró que, en cáncer, los genes regulados por la subfamilia KDM4 están involucrados en

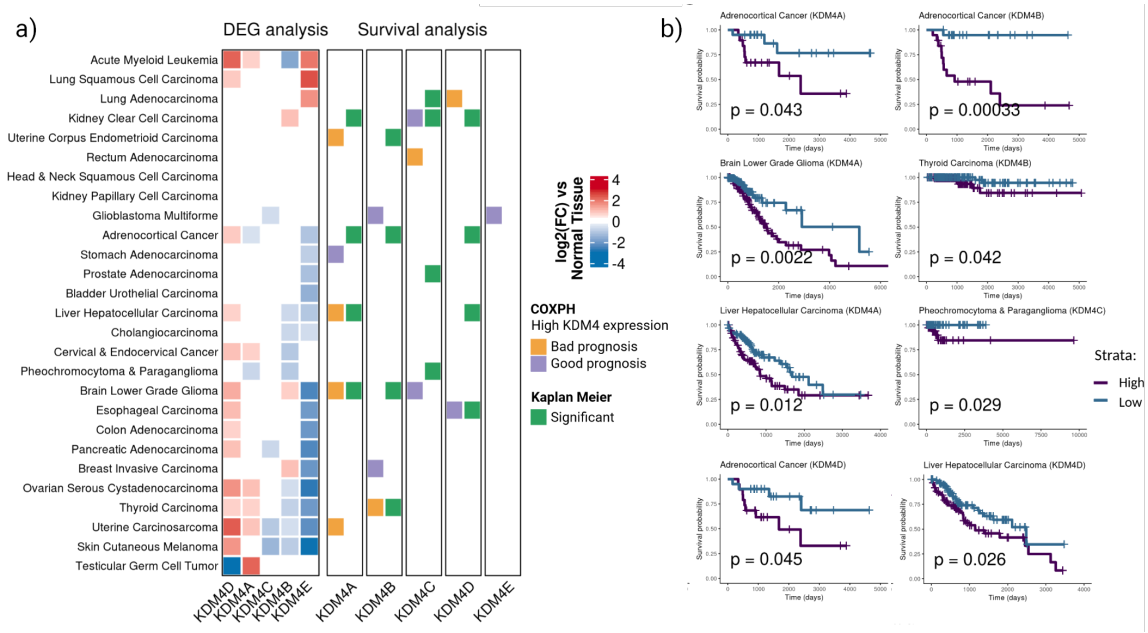


Figura 8.1: Expresión de la subfamilia KDM4 en el cáncer. a) Análisis de expresión génica y supervivencia para cada proteína KDM4. El primer panel muestra el análisis de expresión diferencial de las muestras tumorales frente al tejido no neoplásico. El segundo panel muestra el análisis de supervivencia de CoxPH y Kaplan Meier como columnas adyacentes para cada proteína KDM4. Para el análisis de CoxPH (primera columna), el color de la casilla indica si los niveles altos de KDM4 son de mal o buen pronóstico (valor de p <0.05). Para el análisis de Kaplan-Meier (segunda columna), las muestras tumorales se dividieron en dos grupos según su expresión de KDM: bajo y alto. Las casillas blancas representan asociación no significativa (valor de p >0.05). **b)** Curvas de Kaplan Meier significativas de la proteína KDM4 sobreexpresada en el tipo de cáncer donde se encontró una relación de mal pronóstico con un valor de p <0.05. El color de la línea representa la expresión de KDM4: morado para pacientes con alta expresion y azul para baja.

procesos como la señalización de TNFα por NFκB, la respuesta al interferón-gamma, la respuesta inflamatoria, el punto de control G2M y la vía p53 (**Figura 8.2b**). En general, nuestros datos sugieren que las proteínas de la subfamilia KDM4 son blancos relevantes para buscar inhibidores específicos que podrían ser beneficiosos en el tratamiento de neoplasias y que cada neoplasia tiene un diferente patrón de expresión de las proteínas KDM4.

8.1.2 Compuestos Naturales como Inhibidores Potenciales de la Subfamilia KDM4

Dado que se observó que las KDM4 están desreguladas en diferentes neoplasias y que su expresión está relacionada con diversos procesos asociados con el cáncer, a continuación, se utilizó docking molecular usando el algoritmo FRED de la plataforma OpenEye para buscar potenciales compuestos inhibidores. Se probaron un total de 418,912 compuestos provenientes de tres bases de datos diferentes (DrugBank, FDA y COCONUT) contra los sitios activos de cada KDM4; en total se usaron 13 estructuras cristalográficas disponibles en PDB. Previo al análisis de acoplamiento molecular, se superpusieron las diferentes estructuras PDB disponibles para cada una de las proteínas KDM4 y no se encontraron diferencias significativas en los sitios catalíticos entre ellas; sólo KDM4D mostró

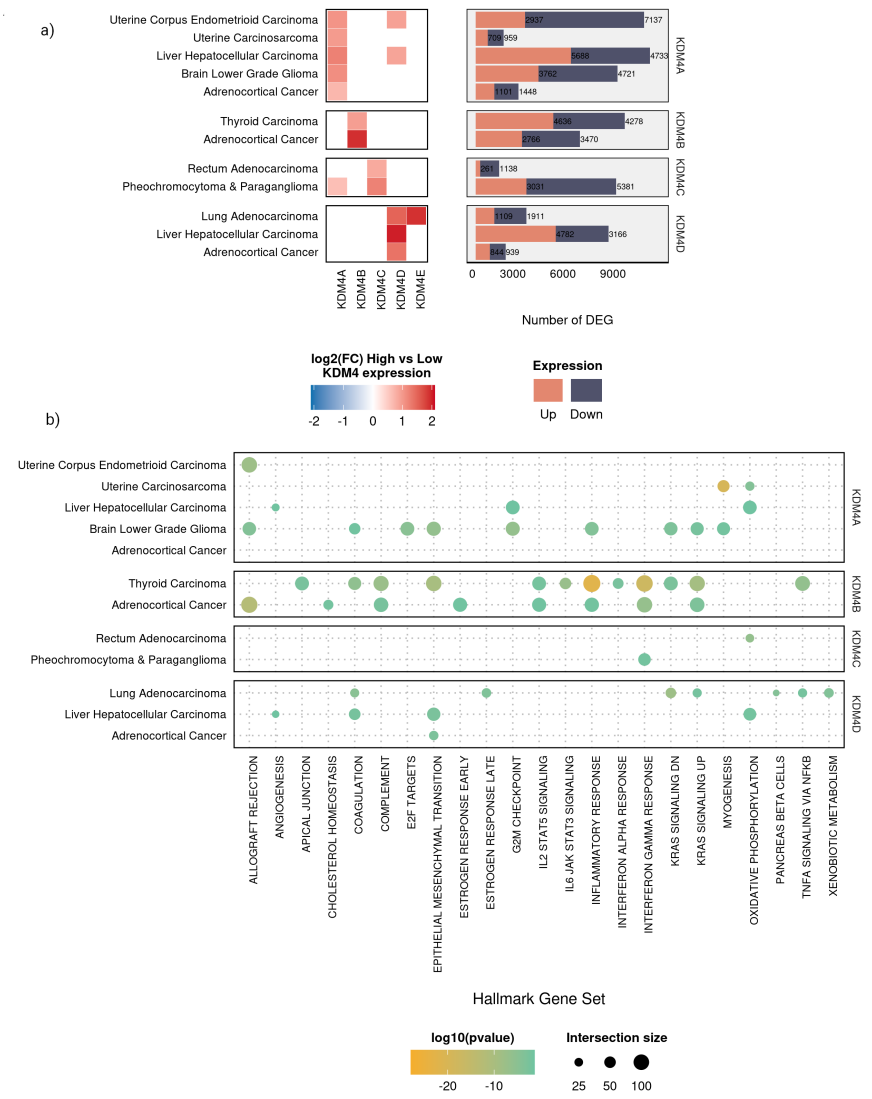


Figura 8.2: Principales firmas moleculares enriquecidas en los 12 tipos de tumores donde la desregulación de la subfamilia KDM4 es de mal pronóstico. a) El análisis de expresión diferencial se realizó comparando muestras con alto y bajo nivel de expresión de KDM4. La intensidad del color indica el log2(FC). El panel derecho representa el número de genes expresados diferencialmente (DEG) para cada comparación. **b)** El gráfico muestra los procesos moleculares enriquecidos en genes diferencialmente expresados (DEG) para cada comparación. La intensidad del color representa el valor p y el tamaño de la intersección.

variaciones estructurales importantes entre los distintos modelos disponibles en PDB, principalmente en los bucles que rodean la entrada del sitio activo. Por este motivo, se utilizó una única estructura para las KDMs 4A, B, C y E, mientras que se mantuvieron las 9 de KDM4D para tener una muestra representativa de sus diferentes conformaciones. La fluctuación cuadrática media (RMSF) de las estructuras de KDM4 utilizadas en este trabajo indica que, en general, la conformación del sitio catalítico es similar entre los distintos KDM4, aunque existe un pico alrededor del residuo 150 (los números de aminoácidos son relativos a KDM4A) que pertenece a la región del bucle exterior (**Figura 8.3a**).

Para el análisis de acoplamiento, se prepararon tanto las formas HOLO (con metales) como APO (sin metales) de todas las estructuras disponibles. En total se usaron 26 estructuras para el acoplamiento FRED. Se seleccionaron los 100 mejores ligandos de acuerdo con el puntaje FRED/Chemgauss4 obtenido, registrando un total de 7,800 interacciones proteína-ligando. La **Figura 8.3b** muestra la distribución de puntajes de esos 7800 compuestos y el número de estructuras diferentes que podrían ser blanco de cada ligando. El puntaje FRED/Chemgauss4 está relacionado con la energía de unión del complejo proteína-ligando; es decir, valores más negativos representan interacciones más fuertes y sugieren que una molécula tiene un mayor potencial de unión (**Figura 8.3b**). El número de blancos indica la cantidad de miembros de la subfamilia KDM4 a la que se puede unir un ligando. Valores altos indican alta promiscuidad de los ligandos, mientras que los valores cercanos a cero sugieren que la unión del ligando es específica para una enzima, lo cual es deseable para el diseño de fármacos. También observamos que los mejores puntajes se lograron con el sistema HOLO en comparación con el sistema APO, lo que sugiere que los ligandos pueden proporcionar grupos funcionales que actúan como agentes quelantes que forman enlaces de coordinación con el metal divalente en el sitio activo de la forma HOLO de KDM4.

Para todas las enzimas KDM4, los compuestos de la base de datos COCONUT tuvieron mejores puntajes, mientras que la mayoría de los valores de FDA y DrugBank tienden a ubicarse cerca de puntajes menos favorables (entre -16 y -11), aunque tienen una proporción notable de ligandos con puntajes mayores a -10 (**Figura 8.3c**).

8.1.3 Los grupos fenoles y azúcares son clave para el diseño de inhibidores potenciales de KDM4

A continuación, se exploraron las similitudes estructurales entre los compuestos predichos para unirse con alta afinidad a los miembros de la subfamilia KDM4; tales hallazgos podrían ser importantes para entender las firmas moleculares involucradas en las interacciones proteína-ligando y para el desarrollo futuro de inhibidores de la subfamilia KDM4. Para abordar esta pregunta, evaluamos la estructura de los 100 mejores ligandos para cada enzima (un total de 1,500 moléculas) utilizando un mapa de similitud de flexóforos. Nuestro análisis recuperó 15 grupos de compuestos con similitudes estructurales (**Figura 8.4a,b**), numerados arbitrariamente. Un compuesto representativo de cada grupo se muestra en la Figura (**Figura 8.4c**). Es notable que los dos grupos centrales (números 14 y 15) contienen principalmente compuestos de COCONUT. Estos son también los grupos con mejor puntuación Fred/Chemgauss4 y mayor densidad de nodos; contienen moléculas compuestas por 3-4 anillos de combinaciones de grupos fenol o piranosa unidos por enlaces glucosídicos que aumentan la flexibilidad de las moléculas. En general, las moléculas con menos de 3 anillos (así como las moléculas lineales), tienen una puntuación más baja. Lo anterior indica que los azúcares y moléculas aromáticas favorecen la interacción con el sitio de unión de las proteínas KDM4. También observamos que los grupos OH y O- son esenciales para la interacción entre el ligando y KDM4 al realizar enlaces de coordinación con sus cofactores metálicos, como Zn^{2+}, Ni^{2+} o Fe^{2+}; por lo tanto, en el diseño de fármacos,

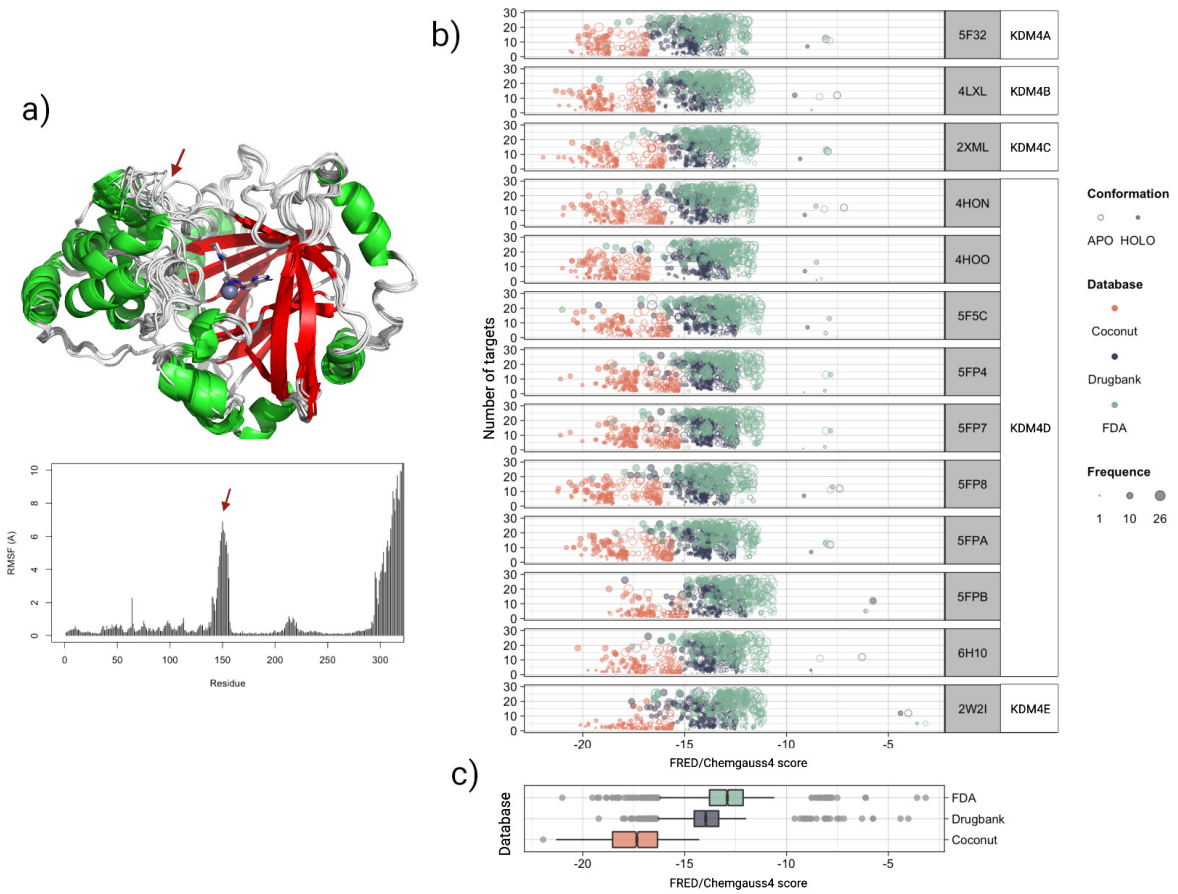


Figura 8.3: Docking molecular contra la subfamilia KDM4. a) Alineación estructural tridimensional para las estructuras de la subfamilia KDM4 utilizadas en este trabajo. El panel inferior muestra los valores de RMSD por residuo para todas las estructuras. **c)** Distribución de la puntuación FRED/Chemgauss4 para cada una de las tres bases de datos evaluadas. Los puntos atípicos se muestran en gris.

la inclusión de azúcares y fenoles representa una ventaja para lograr un inhibidor competitivo.

En la literatura existen moléculas validadas experimentalmente como inhibidores de la subfamilia KDM4; por lo tanto, para el análisis de flexóforos, se incluyeron 16 compuestos citados por Baby et al. (2021) cuyo IC50 es de rango micro a nanomolar (nodos en negrita de la (**Figura 8.4a**)). Se observó que la mayoría de estas moléculas permanecían como nodos aislados cuyos flexóforos no comparten similitudes con los compuestos de las bases de datos COCONUT, DrugBank y FDA.

8.1.4 El sitio activo de KDM4 está estabilizado por interacciones de apilamiento Pi y favorece la unión de ligandos flavonoides.

Para predecir la estabilidad de los complejos proteína-ligando previamente predichos, se llevó a cabo simulaciones de dinámica molecular utilizando el método MM-PBSA para

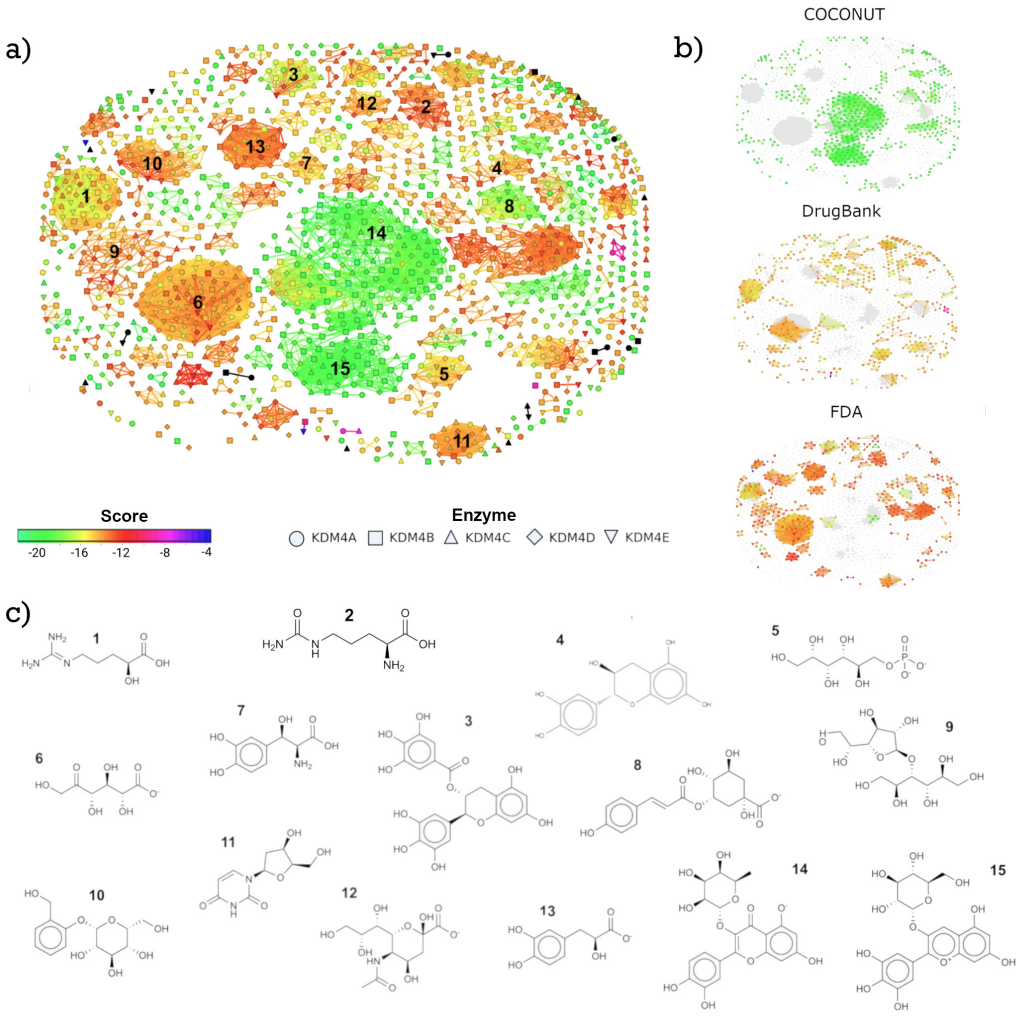


Figura 8.4: Mapa de similitud de flexóforos para COCONUT, DrugBank y FDA. a) Mapa de similitud de flexóforos. Cada nodo representa un compuesto, el color del nodo representa su puntuación FRED/Chemgauss4. La forma del nodo indica a qué KDM4 se une el ligando. Los bordes de la red indican una relación de al menos un 95 % de similitud de flexóforos entre pares de compuestos (vecinos). Los puntos negros representan los compuestos de Baby et al. (2021). **b)** Distribución de nodos para los compuestos pertenecientes a cada una de las bases de datos evaluadas (COCONUT, DrugBank y FDA). **c)** Estructura 2D para un compuesto representativo de cada uno de los clusters elegidos. Los clusters se seleccionaron en función de su tamaño y número de aristas.

calcular la energía de unión absoluta (ΔG_{PBSA}) de un subconjunto representativo de moléculas (20 de FDA, 16 de DrugBank y 25 de COCONUT). Dado que la simulación de dinámica molecular con sistemas HOLO presenta desafíos computacionales, se optó por simular únicamente los sistemas APO. Un complejo proteína-ligando exitoso implica que la interacción predicha por el algoritmo de acoplamiento es reproducible a través de simulaciones de dinámica molecular, sugiriendo una alta probabilidad de que ese ligando actúe como inhibidor de KDM4. Si los valores de ΔG_{PBSA} son negativos, se considera al ligando propuesto como potencial inhibidor. Dado que se observó una tasa de éxito superior al 60 % en todos los compuestos evaluados (**Figura 8.5**a), consideramos confiables las predicciones obtenidas por el algoritmo FRED.

Dado que el complejo KDM4A-CNP0371131 presentaba la puntuación FRED/Chemgauss4 más negativa de todos los complejos proteína-ligando evaluados, se eligió como ejemplo representativo de los cambios conformacionales observados durante las simulaciones de dinámica molecular. Este complejo presentó la puntuación FRED/Chemgauss4 más negativa. Los valores de RMSF por residuo mostraron que las áreas de bucles alrededor de la cavidad de KDM4A son las más flexibles, y que el ligando vibra dentro del sitio activo de la proteína (**Figura 8.5**b). El área de la cavidad de KDM4A es de $753{,}5\ \text{Å}^2$, su volumen es de $824\ \text{Å}^3$, y tiene un potencial electrostático exclusivamente negativo (**Figura 8.5**c). Debido a su tamaño, el sitio de unión de KDM4A podría acomodar moléculas dos veces el tamaño de las estructuras 14 o 15 en la **Figura 8.4**c. Lo anterior sugiere que solo la mitad de la cavidad está ocupada por el ligando, dejando la otra mitad a los cofactores metálicos y al solvente. Por tanto, el tamaño de la molécula no es una limitación para el diseño de un inhibidor competitivo, sino que son los grupos funcionales que coordinan los cofactores metálicos y las interacciones con los residuos del sitio catalítico los que determinan la especificidad de la unión ligando-receptor.

Un rasgo notable del sitio de unión de KDM4A es la presencia de varios aminoácidos aromáticos (Y, F, W y H), que no solo estabilizan el sitio de unión, sino que también contribuyen a la interacción proteína-ligando mediante interacciones de apilamiento pi con otros grupos aromáticos. Los residuos que interactúan con mayor frecuencia con los ligandos son I71, Q84, N86, Y132, A134, D135, G170, V171, Y175, Y177, F185, H188, E190, D191, S196, N198, W208, L241 y S288. La **figura 8.5d** muestra la unión del ligando CNP0371131 a KDM4A, siendo el complejo con mejor puntuación. H188 destaca por establecer dos enlaces de coordinación con los cofactores metálicos (Zn^{2+},Ni^{2+} o Fe^{2+}). También se observó que, aunque los residuos E y D en el sitio catalítico no interactúan directamente con los ligandos, contribuyen al microambiente negativo de la cavidad. Por ejemplo, la epigalocatequina galato (EGCG), una molécula del grupo número 3, establece un enlace de coordinación con el metal de KDM4A a través del grupo flavonoide, mientras que el catecol secundario se desvía en la dirección opuesta debido a las fuerzas repulsivas entre ellos. Además, los compuestos de los grupos 14 y 15 muestran una orientación favorable de los grupos OH del azúcar en el flavonoide, lo que permite la formación de 2 a 3 enlaces de coordinación con el metal; en este caso, aunque aumenta el número de enlaces de coordinación, estos no son proporcionados por el catecol sino por los azúcares. Esto se debe a la menor disponibilidad de electrones en el oxígeno del OH del catecol secundario en comparación con los azúcares, lo que ocasiona una mayor capacidad para formar enlaces de coordinación con el Zn^{2+} en el sitio activo de KDM4A.

Para estudiar la distribución de la energía ΔG_{PBSA} en la proteína, se calculó la contribución de la energía de unión por residuo en el complejo KDM4A-CNP0371131. Observamos que la energía de unión está impulsada principalmente por interacciones electrostáticas de largo alcance y se distribuye a lo largo de todos los residuos, no solo en los presentes en la cavidad (**Figura 8.5**e). En general, las fuerzas atractivas (valores negativos) compensan las fuerzas de repulsión (valores positivos), y aunque se encuentran algunos picos de repulsión (como en el residuo 180), estos se compensan con otras interacciones estabilizadoras (como los residuos 78, 172, 228, 298 y 300), lo que conduce a

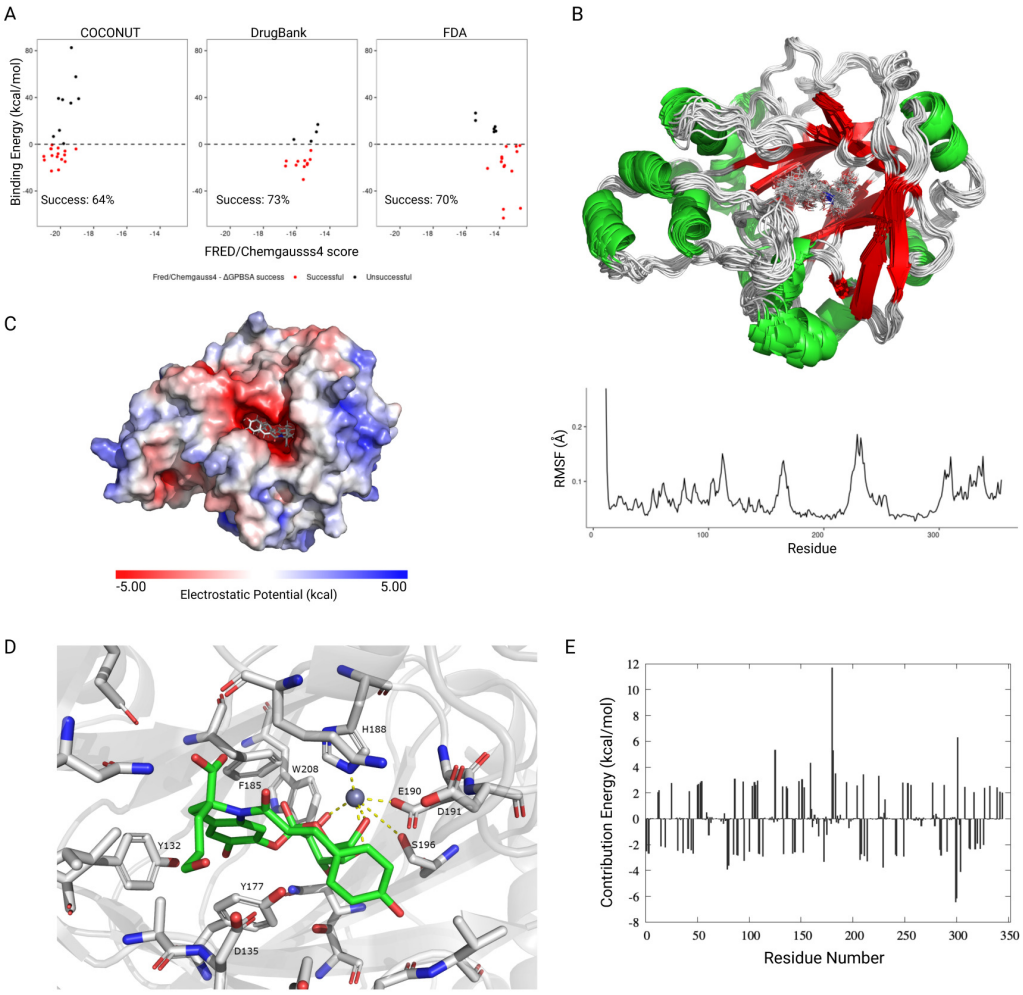


Figura 8.5: Simulaciones de dinámica molecular y cálculo de la energía de enlace absoluta. a) Relación FRED/Chemgauss4 vs ΔG_{PBSA} para los ligandos con mejor puntuación de cada base de datos. Los ligandos que mostraron una energía de unión favorable (menor a 0 kcal/mol) y una puntuación FRED/Chemgauss4 negativa se consideraron exitosos (puntos rojos). El porcentaje de éxito representa la proporción de moléculas que cumplieron los criterios establecidos para cada base de datos. **b)** Panel superior: Representación gráfica (20 fotogramas) de la simulación de dinámica molecular para KDM4A (PDB ID: 5F32) en complejo con el ligando CNP0371131 de COCONUT. Panel inferior: Valor RMSF para cada residuo. **c)** Potencial electrostático para el complejo KDM4A-CNP0371131. **d)** Representación gráfica de la molécula CNP0371131 (verde) unida al sitio catalítico de KDM4A. Los números de residuo corresponden a la estructura PDB 5F32. **e)** Contribución media por residuo de la energía libre de unión MM-PBSA para el complejo KDM4A-CNP0371131.

una energía ΔG_{PBSA} favorable en general. Esto indica que el complejo es estable, por lo que el ligando tiene probabilidades de mostrar actividad inhibidora de KDM4A *in vitro*. La energía de unión se logra gracias a la contribución de la energía de interacción intrínseca favorable (ΔE_{MM}) y la energía de interacción no polar ($\Delta E_{nopolar}$), mientras que una penalización desfavorable es aplicada por la energía de interacción polar (ΔE_{polar}), principalmente debido al efecto de solvatación tanto del ligando como del sitio activo. Estos resultados proporcionan información sobre las interacciones moleculares entre el sitio catalítico de KDM4A y las moléculas pequeñas, lo que podría ayudar en el diseño actual y

futuro de pequeños inhibidores. La **Tabla 8.1** enumera las principales moléculas obtenidas de las bases de datos COCONUT, DrugBank y FDA.

Tabla 8.1: Lista de las principales moléculas con potencial actividad inhibidora de las proteínas de la subfamilia KDM4 determinada mediante acoplamiento molecular utilizando las bases de datos COCONUT, DrugBank y FDA.

Blanco	**Base de datos**	**Ligando**
KDM4A	COCONUT	CNP0058667, CNP0150788, CNP0216191, CNP0002425, CNP0371131, Pulchellidin 3-Glucoside (CNP0359043), CNP0223133, CNP0258703 (Epigallocatechin gallate)
	DrugBank	6-O-capryloylsucrose, Zanamivir, Acteoside, DB04211, DB03249, DB07719, DB12116
	FDA	Glucosamine, Glucosamine sulfate, Doripenem, Neohesperidin, Sulisobenzone, Verbascoside, Wedelolactone, Epigallocatechin gallate
KDM4B	COCONUT	CNP0322725, CNP0216191, CNP0098686, CNP0316754, CNP0107391, CNP0239128, Crispine D (CNP0119105)
	DrugBank	Carba-glucotropaeolin, Ascorbyl glucoside, Zanamivir, Iodo-Willardiine, beta-D-arabinofuranose 5-phosphate, DB03250,DB02488
	FDA	Methazolamide, Sulisobenzone, Baricitinib, Lanraplenib, Pentostatin
KDM4C	COCONUT	CNP0187735, CNP0417860, CNP0226084, CNP0298305, CNP0289146, CNP0350449, CNP0106665
	DrugBank	Peramivir, DB03717, Edotecarin, 3'-Uridine Monophosphate
	FDA	Cynarin, Quercitrin, Chlorogenic acid, (-)-Epigallocatechin gallate, Hyperoside, Gastrodin, Polydatin
KDM4D	COCONUT	6-C-Glucosylorobol (CNP0299696), CNP0002425, CNP0362352, CNP0243580, CNP0216191, Isovolubilin (CNP0151675), CNP0397301
	DrugBank	6-O-Capryloylsucrose, Balanol, 10-hydroxycamptothecin, DB07102, 2'-Deoxycytidine-5'-Monophosphate, Cidofovir, Levoglucose
	FDA	Glucosamine, Glucosamine Sulfate, Oleuropein, Sulpiride, Sulisobenzone, Levosulpiride (Levogastrol), Hydroxycamptothecin
KDM4E	COCONUT	CNP0131606, CNP0186792, CNP0125603, 4-hydroxy-2-ketoarginine (CNP0433705), CNP0295348, Quercetin 5-Glucuronide (CNP0081446), CNP0249133
	DrugBank	Azacitidine, Meglumine, Balanol, Levoglucose, Ascorbic acid, L-Xylulose 5-Phosphate, 5-phospho-D-arabinohydroxamic acid
	FDA	Glucosamine, Glucosamine Sulfate, Minoxidil Sulphate, Sulfamonomethoxine, Sulpiride, Xylitol, Orotic Acid (6-Carboxyuracil)

8.1.5 Los inhibidores de la subfamilia KDM4 podrían tener un impacto significativo en la terapia del cáncer.

Dado que nuestros datos muestran que hay algunos tipos de cáncer en los que están involucradas más de una KDM, es posible que un fármaco que tenga como blanco a todas

las proteínas KDM4 significativas en una neoplasia podría ser altamente efectivo como terapia. Para integrar esta información, se construyó una red proteína-fármaco-enfermedad que contiene las cinco KDM4 y los siete principales ligandos en cada una de las tres bases de datos evaluadas. Observamos que de acuerdo al patrón de expresión de cada KDM4, se puede encontrar un conjunto de fármacos diferente para cada neoplasia (**Figura 8.6)**; por ejemplo, KDM4E y D son importantes para el adenocarcinoma de pulmón, por lo tanto, sulpirida y balanol son fármacos aprobados por la FDA que podrían considerarse para el tratamiento de ese cáncer. Además, KDM4A, B y D son relevantes para el cáncer adrenocortical y el carcinoma tiroideo; por lo tanto, los compuestos COCONUT CNP0002425, CNP0299696 y CNP216191 son candidatos destacados para el tratamiento de esas neoplasias. En cuanto al complejo KDM4A-CNP0371131 (que tuvo la puntuación más alta de FRED/Chemgauss4), observamos que CNP0371131 es exclusivo para KDM4A; por lo tanto, podría usarse como tratamiento para cánceres donde solo KDM4A está desregulado, como el carcinoma endometrioide del cuerpo uterino.

De forma similar podemos utilizar los blancos conocidos de las moléculas ya aprobadas por la FDA para inhibir más de un gen desregulado en cáncer. Un ejemplo detallado de la utilidad de este análisis es la red construida para el carcinoma endometrioide del cuerpo uterino (**Figura 8.7**), que muestra que KDM4D y KDM4A están sobreexpresados y ambos interactúan con DNMT1 (una DNA metiltransferasa); nuestro análisis de acoplamiento muestra que hay 5 compuestos de COCONUT capaces de unirse a KDM4D y 5 compuestos del DrugBank que se dirigen a KDM4A. Sin embargo, los compuestos del DrugBank tienen como blanco a otras proteínas de la red además de las KDM4; por ejemplo, el DB07602 se dirige a KDM4A y EGFR; mientras que la Azacitidina inhibe tanto a KDM4A como a DNMT1, lo que sugiere que la Azacitidina podría modular proteínas esenciales implicadas en la regulación negativa de la metilación de la histona H3K9 (como muestra la sombra azul claro de la **Figura 8.7**). Este mismo enfoque para interpretar los resultados puede aplicarse para las otras redes generadas específicamente de acuerdo con los perfiles de expresión caracterizados en este trabajo.

8.2 Búsqueda de inhibidores específicos para la proteína BORIS

8.2.1 El modelado por homología de BORIS y CTCF produce estructuras confiables

Además de los miembros de la subfamilia KDM4, otra proteína que se ha posicionado como un potencial blanco terapéutico es BORIS, un parálogo de CTCF. A pesar del potencial que tiene BORIS como blanco terapéutico, las estrategias usadas para las proteínas de la subfamilia KDM4 no pueden aplicarse a esta ya que, a diferencia del caso previamente estudiado, la proteína BORIS no cuenta con estructuras cristalográficas disponibles, por lo tanto antes de implementar el cribado virtual, es necesario predecir su estructura tridimensional. Como aproximación inicial, se generaron con Modeller mil modelos de homología 3D para los dedos de zinc 2 al 11 (ZF2-11) de BORIS utilizando

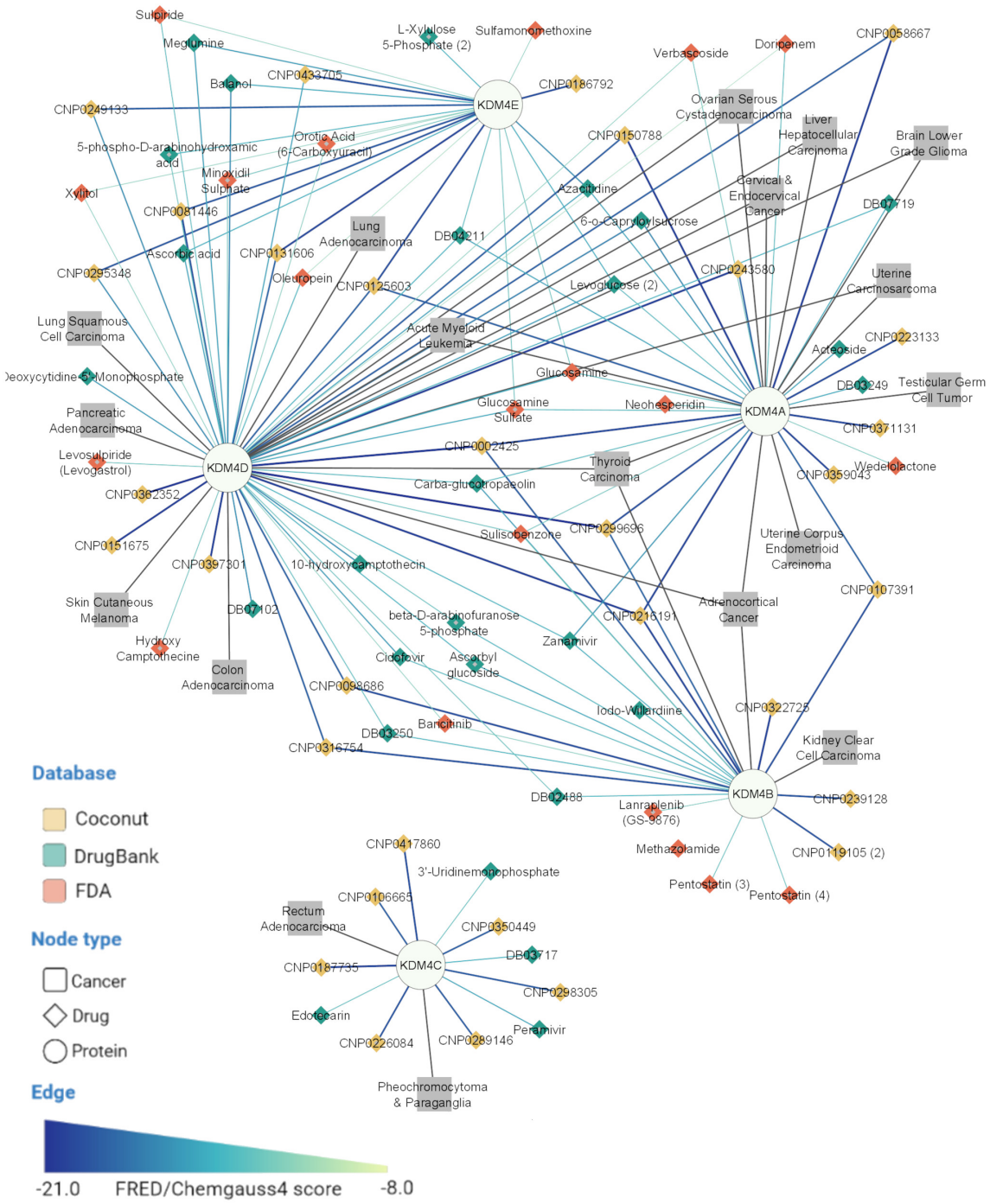


Figura 8.6: Principales inhibidores potenciales de la subfamilia KDM4. La red representa la relación Fármaco-Proteína y Enfermedad-Proteína entre los miembros de la subfamilia KDM4. El ancho del borde Fármaco-Proteína y la intensidad del color representan la puntuación FRED/Chemgauss4. Para nombres largos de compuestos, solo se proporciona el ID de la base de datos.

como templados las estructuras cristalográficas disponibles para CTCF. El modelo número 639 fue seleccionado como el mejor debido a que mostró la puntuación DOPE más baja

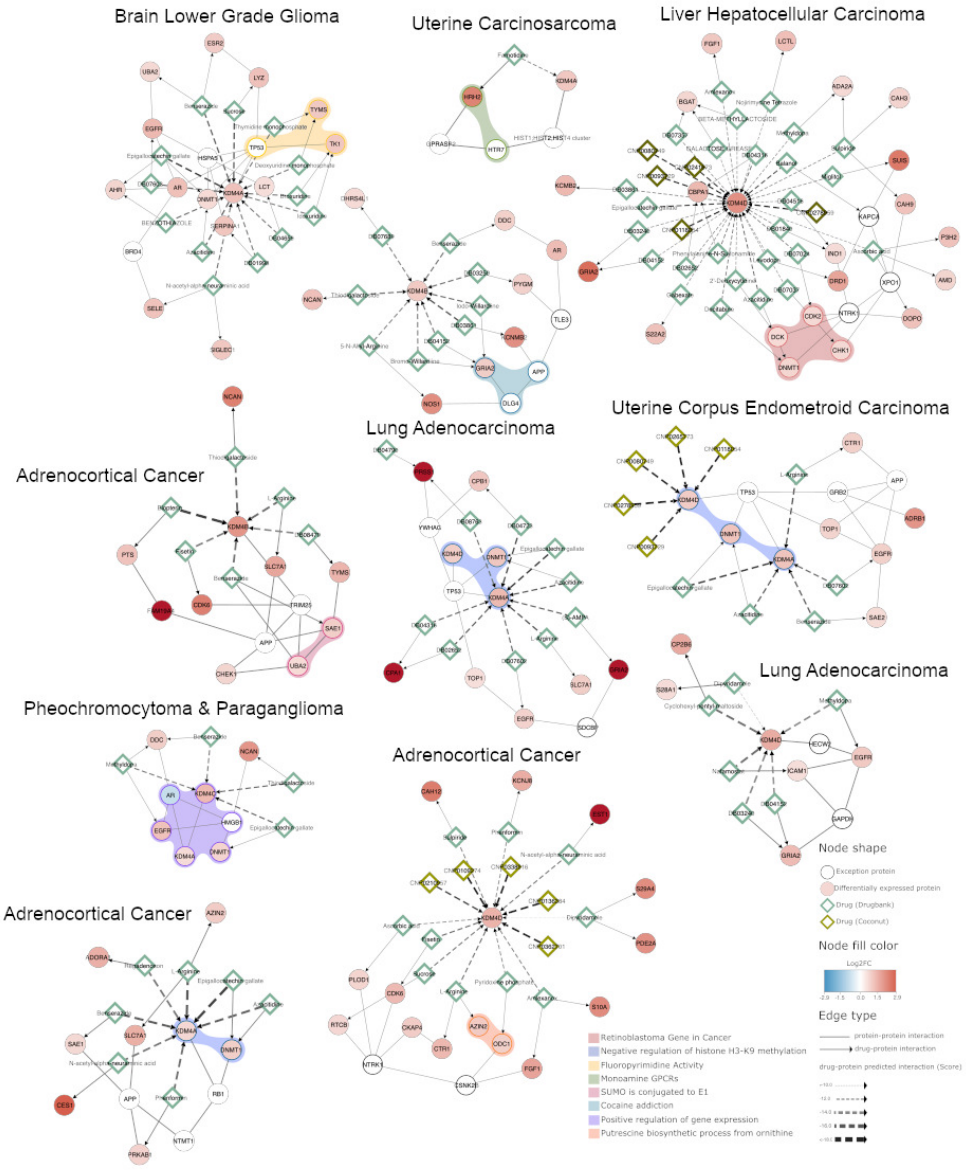


Figura 8.7: Análisis de redes de interacción para los inhibidores potenciales de KDM4 en diferentes tipos de cáncer.

(-18995,75). La puntuación DOPE está directamente relacionada con la energía libre de un complejo y, por tanto, el modelo con el valor más bajo se considera el más cercano al estado nativo de la proteína (Shen and Sali, 2006). Dado que Modeller no puede modelar cadenas de DNA, sólo copiar las disponibles en las plantillas, Los dos segmentos de DNA en el modelo final fueron sustituidos por un nuevo segmento creado con w3DNA para conseguir una única molécula continua, la secuencia de esta molécula corresponde al sitio de unión consenso de CTCF (**Figure 8.8a, b**). El diagrama de Ramachandran para el modelo 639 muestra que el 90,3 % de los aminoácidos se encuentran en las regiones más favorables del diagrama y no hay residuos en las regiones no permitidas (**Figure 8.8c**).

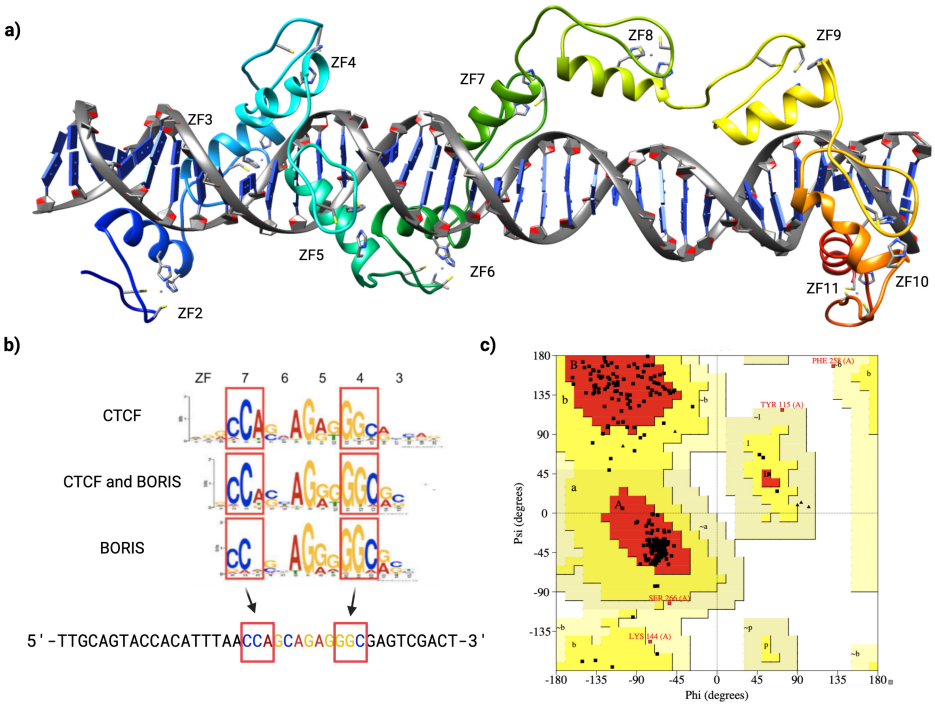


Figura 8.8: Modelado 3D por homología para los dedos de zinc 2 a 11 de BORIS. a) Modelo creado con Modeller después del reemplazo de las cadenas de DNA. **b)** Motivos consenso para los sitios de unión únicos y compartidos de CTCF y BORIS, la imagen es tomada de (Nishana et al., 2020). La secuencia de DNA utilizada para todos los modelos en este trabajo se muestra en la parte inferior del panel. **c)** Gráfico de Ramachandran para el modelo en el panel **a**. Residuos en las regiones más favorecidas [A,B,L] = 90.2 %, residuos en regiones adicionalmente permitidas [a,b,l,p] = 8.3 %, residuos en regiones generosamente permitidas [a, b, l, p] = 1.5 % y residuos en regiones no permitidas = 0.0 %.

Se siguió un proceso similar para crear un modelo de CTCF basado en los fragmentos disponibles en PDB para esta proteína, los resultados se muestran en la **Figura 7.2**. El modelo de CTCF con el puntaje DOPE más bajo fue el número 724. Los fragmentos de DNA incluidos originalmente por Modeller fueron reemplazados por la misma molécula utilizada para el modelo de BORIS (**Figura 8.9a**). El gráfico de Ramachandran para este modelo muestra que el 90.2 % de los residuos están en las regiones más favorecidas del gráfico, mientras que el 0.4 % se encuentran en las regiones no permitidas (**Figura 8.9b**).

El puntaje DOPE por residuo muestra que, en general, la energía para cada residuo es similar a la de los modelos utilizados como plantillas, lo que indica que ambos modelos son de buena calidad (**Figura 8.9c**) y que la alineación es correcta (Eswar et al., 2006). Sin embargo, el Factor de Calidad Global para ambos modelos está por debajo de lo esperado (**Tabla 8.2**). Este factor se calcula mediante el algoritmo ERRAT e indica la calidad de las interacciones atómicas no enlazantes en una estructura proteica dada; se espera que un buen modelo tenga un factor de calidad global superior al 90 % (Colovos and Yeates, 1993). El algoritmo de Evaluación de Volumen de Proteínas (PROVE) calcula la desviación Z estadística para los volúmenes de los átomos en los modelos comparados con los átomos de estructuras de PDB con alta resolución y refinadas (Pontius et al., 1996). El porcentaje indica el número de valores atípicos; se espera que un buen modelo tenga menos del 1 %, aunque porcentajes entre el 1 y

el 5 % es aceptable. Para nuestros modelos, el puntaje PROVE indica que el modelo de CTCF es aceptable, mientras que los valores atípicos en el modelo de BORIS representan el 6.7 % de los átomos, lo cual indica que existe traslape en los átomos del modelo (**Tabla 8.2**).

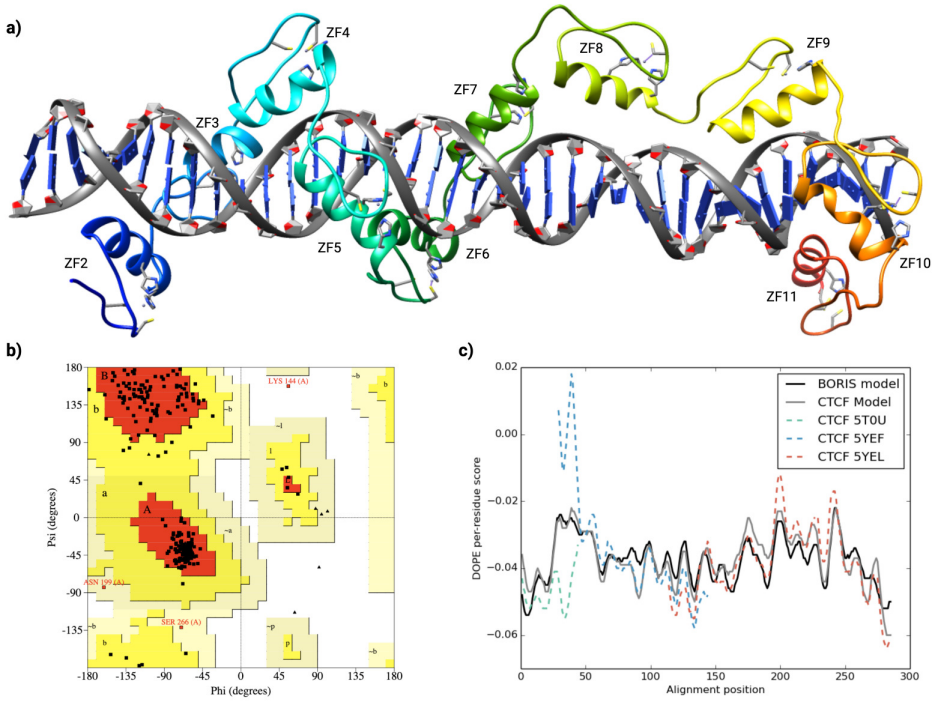


Figura 8.9: Modelo 3D para CTCF. **a)** Modelo de los dedos de zinc 2 a 11 de CTCF creado con Modeller después del reemplazo de las cadenas de DNA. **b)** Gráfico de Ramachandran para el modelo en el panel **a**. Estadísticas del gráfico: residuos en las regiones más favorecidas [A,B,L] = 90.2 %, residuos en regiones adicionalmente permitidas [a,b,l,p] = 8.3 %, residuos en regiones generosamente permitidas [a, b, l, p] = 1.5 % y residuos en regiones no permitidas = 0.0 %. **c)** Puntaje DOPE por residuo para los modelos de BORIS y CTCF comparados con las plantillas utilizadas.

La alineación estructural entre los modelos de CTCF y BORIS, así como los templados utilizados para crear ambos modelos, se muestra en la **Figura 8.10a**. Se observó un RMSD inferior a 2,0 Å para todas las comparaciones (**Figura 8.10b**), lo que indica que las diferencias entre los modelos y las plantillas no son significativas. El RMSF entre los modelos de BORIS y CTCF muestra que, en general, las estructuras son similares; sin embargo, se observan algunas diferencias importantes en los primeros residuos de cada dedo de zinc, estas regiones corresponden a los segmentos de los modelos con bucles, por lo que podrían tener mayor flexibilidad (**Figura 8.10c**).

8.2.2 Las simulaciones de Dinámica Molecular muestran que los modelos propuestos son estables

Para mejorar los modelos, estos fueron sometidos a minimización de energía, equilibración de temperatura y presión con gromacs. Después de estos pasos, se observó

Tabla 8.2: Valores de puntaje DOPE, Factor de Calidad Global, parámetros de PROCHECK y puntajes de PROVE para cada modelo obtenido.

Model	BORIS	CTCF
Factor de Calidad Global(ERRAT)	36	53.65
PROVE	6.7 %	4.9 %
PROCHECK		
Residuos en las regiones más favorecidas	90.2 %	90.2 %
Residuos en regiones adicionalmente permitidas	8.3 %	8.7 %
Residuos en regiones generosamente permitidas	1.5 %	0.8 %
Residuos en regiones no permitidas	0.0 %	0.4 %

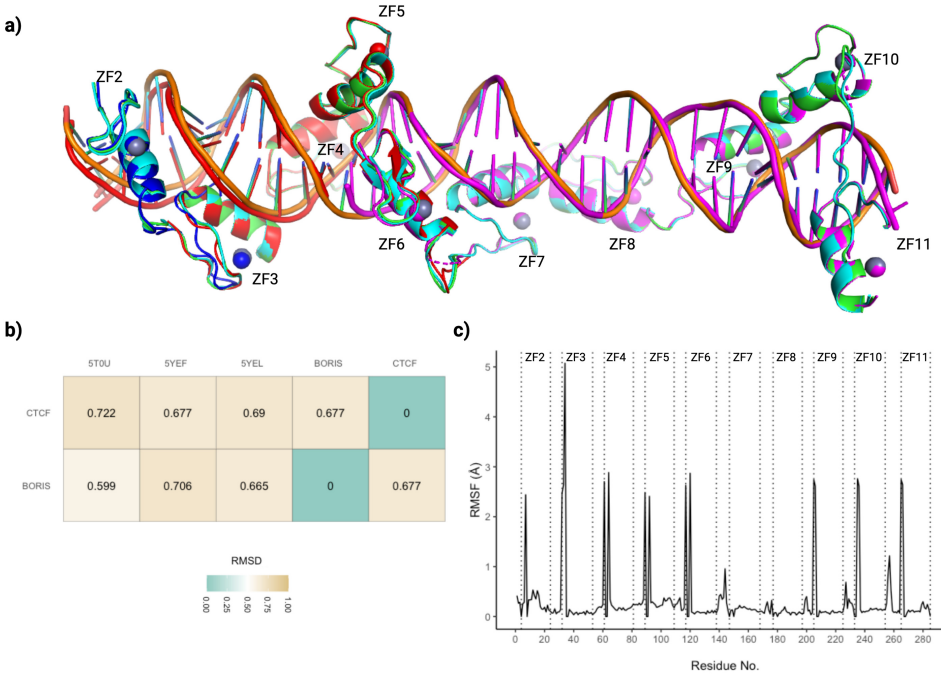


Figura 8.10: Alineación estructural de los modelos para CTCF y BORIS. **a)** Alineación de las plantillas de CTCF y los modelos creados. El modelo de CTCF se muestra en cian, BORIS en verde, el templado 5T0U en azul oscuro, 5YEF en rojo y 5YEL en magenta. **b)** RMSD entre los modelos y las plantillas utilizadas. **c)** RMSF para los modelos de CTCF vs BORIS. Las líneas discontinuas marcan el inicio y el final de cada dedo de zinc. El número de residuo es relativo al inicio de los fragmentos modelados.

una mejora considerable en ambos modelos, tal como se muestra en la **Tabla 8.3**. Para BORIS, el Factor de Calidad Global aumentó hasta 99.05 y los valores atípicos en PROVE disminuyeron al 3.5 %; sin embargo, hubo una ligera disminución en el porcentaje de residuos en las regiones más favorables del gráfico de Ramachandran en comparación con la conformación inicial.

Modelo	BORIS	CTCF
Factor de Calidad Global(ERRAT)	99.05	79.41
PROVE	3.5 %	7.1 %
PROCHECK		
Residuos en las regiones más favorecidas	87.2 %	84.8 %
Residuos en regiones adicionalmente permitidas	11.3 %	12.9 %
Residuos en regiones generosamente permitidas	1.1 %	1.9 %
Residuos en regiones no permitidas	0.4 %	0.4 %

Tabla 8.3: Valores de la puntuación DOPE, el Factor de Calidad Global, los parámetros PROCHECK y las puntuaciones PROVE de cada modelo obtenido.

El enlace coordinado entre los residuos de Zn^{2+}, CYS e HIS contribuye a la estabilidad conformacional de los dedos de zinc y a la interacción con el DNA. El estado de protonación de los residuos CYS y HIS en los dedos de zinc es crucial para su interacción con el metal, ya que los residuos cargados negativamente son necesarios para la estabilización del metal en su sitio de unión (Godwin et al., 2017; Li et al., 2008). Para las simulaciones, los residuos CYS se fijaron en su estado desprotonado (CYN) y los HIS se protonaron en el átomo ND1 (HSD) como se muestra en la **Figura 8.11a**. Esta configuración permitió mantener el Zn^{2+} en su lugar sin necesidad de aplicar restricciones posicionales. La **Figura 8.11b** muestra el primer y último fotograma de la simulación de dinámica molecular para el modelo de BORIS. Se observa que el Zn^{2+} permaneció en contacto con las histidinas y cisteínas en lugar de saltar fuera de su sitio de unión.

Para evaluar las fluctuaciones estructurales de los modelos, se calculó el RMSD entre los diferentes puntos de la simulación y el modelo inicial. BORIS y CTCF fueron simulados en presencia y en ausencia de la cadena de DNA (**Figura 8.12a**). Es notable que el complejo CTCF+DNA no mostró fluctuaciones importantes en sus valores de RMSD a lo largo del tiempo mientras que se observó el comportamiento opuesto para la proteína CTCF sola, donde el RMSD fluctúa durante todo el tiempo simulado. El mismo comportamiento se observa en el RMSF y el radio de giro (Rg), lo que en general indica que el modelo de CTCF es estable cuando está unido al DNA, pero no cuando está solo (**Figura 8.12b,c**). Por otro lado, el complejo BORIS+DNA presenta algunas fluctuaciones importantes en su RMSD, RMSF y Rg en comparación con CTCF+DNA, mientras que BORIS solo se comporta de forma similar a CTCF, lo que indica que la conformación 3D de BORIS también depende del DNA.

Se observó que las cadenas de DNA parecen seguir la misma tendencia que la proteína. Las cadenas en el modelo de BORIS muestran un ligero aumento en el RMSD alrededor de

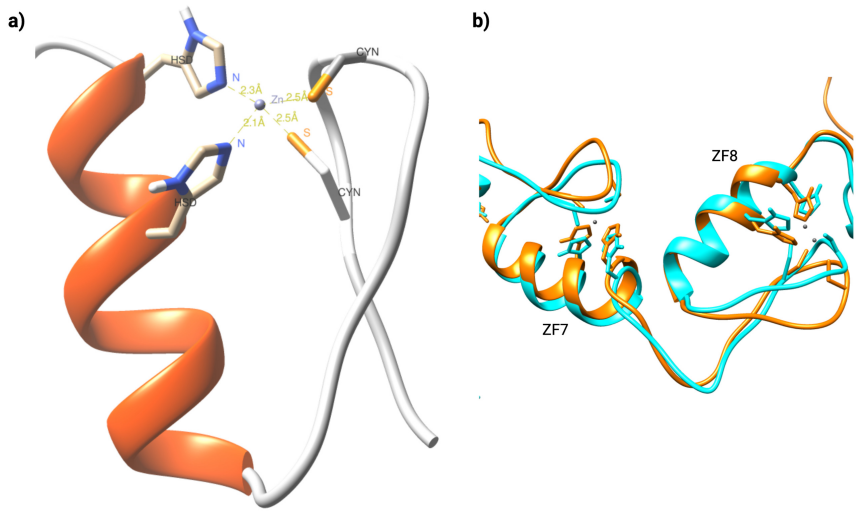


Figura 8.11: Configuración de los dedos de zinc en los modelos. a) Se muestra el ZF2 del modelo de BORIS para ilustrar la conformación estructural de los dedos de zinc después de que los residuos HIS y CYS que interactúan con el ion Zn^{2+} fueron declarados como HSD y CYN, respectivamente. **b)** Comparación entre el primer (cian) y último (naranja) fotograma de la simulación de dinámica molecular de BORIS, se muestran los ZF7 y 8.

150 ns, mientras que las cadenas de DNA pertenecientes al modelo de CTCF permanecen estables durante toda la simulación (**Figura 8.13a**). Para el modelo de BORIS, las moléculas de DNA tienen un poco más de flexibilidad en comparación con el complejo CTCF+DNA, esto es evidenciado por los valos de RMSD más altos para el modelo de BORIS que para el de CTCF (**Figura 8.13b**).

Debido a que las interacciones DNA-Proteína están mayoritariamente mediadas por puentes de hidrógeno, se calculó el número de enlaces de hidrógeno para evaluar la interacción entre BORIS y el DNA. La media de enlaces de hidrógeno para BORIS es de 63,7 y para CTCF es de 58,3, lo que indica una diferencia en la interacción de ambas proteínas con el DNA. En general, la interacción del DNA para los modelos BORIS y CTCF es estable (**Figura 8.14**). El DNA utilizado para ambos modelos CTCF y BORIS corresponde al sitio de unión consenso de CTCF y aunque se sabe que BORIS comparte sitios con CTCF, hay otros que son exclusivos para BORIS y que tienen un motivo de unión ligeramente diferente.

Finalmente, la estructura secundaria para ambas proteínas cuando se simulan con o sin DNA se muestra en la **Figura 8.15**. En general, se observa que las estructuras se conservan cuando la proteína está unida al DNA pero no cuando está separada. La estructura secundaria de ambas proteínas se mantiene a través de las simulaciones y corresponde a la estructura conocida para los dedos de zinc, es decir, dos láminas β y una hélice α **Figure 8.15**a. Sin embargo, existen algunas diferencias entre CTCF y BORIS, por ejemplo, el tamaño de la región conectiva entre los ZF5 y ZF6 es mayor en CTCF que en BORIS.

Figura 8.12: Análisis de Simulaciones de Dinámica Molecular. **a)** El RMSD se calculó utilizando solo los átomos Cα en la proteína. Se muestra el comportamiento a lo largo del tiempo y la distribución general. **b)** RMSF de los átomos Cα del esqueleto de BORIS y CTCF. Cada dedo de zinc está marcado con un cuadro gris. **c)** Radio de giro para las proteínas a lo largo de las simulaciones.

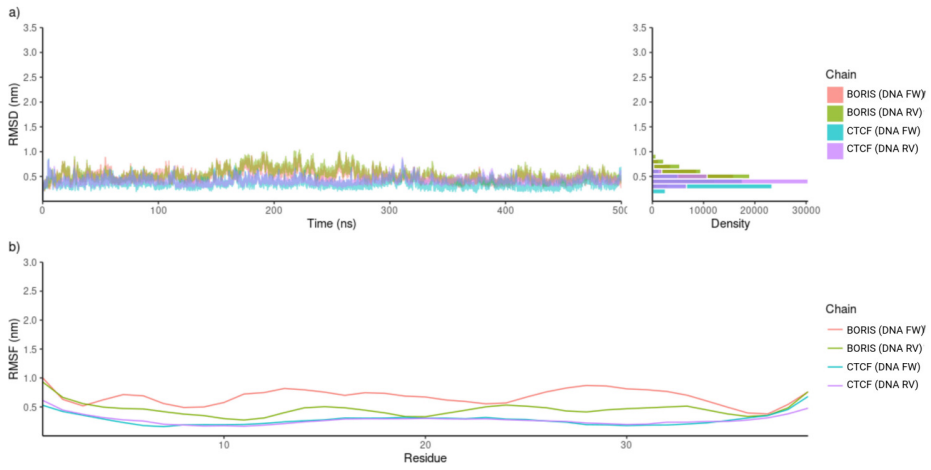


Figura 8.13: Análisis de las cadenas de DNA a través de las simulaciones. a) Gráfico de RMSD para los átomos de O_4' correspondientes al esqueleto de DNA de BORIS y CTCF. La distribución de los RMSDs a lo largo de la simulación se muestra a la derecha de los gráficos. **b)** Gráfico de RMSF para el esqueleto de DNA (átomos de O4), ambos modelos se comparan.

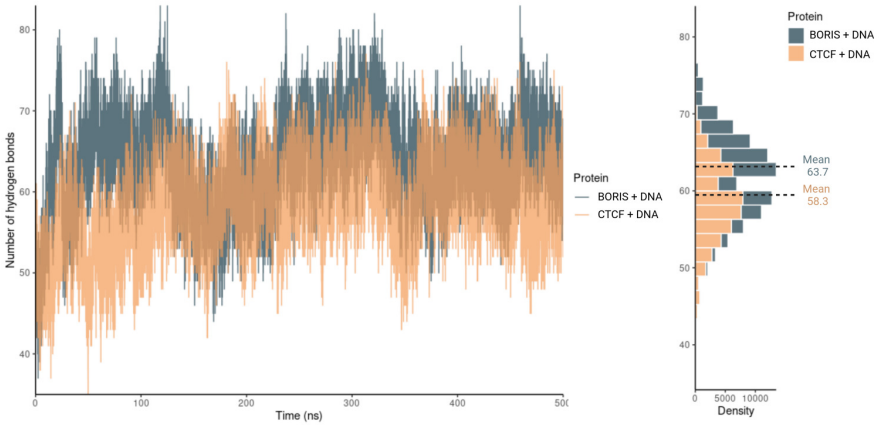


Figura 8.14: Análisis de los enlaces de hidrógeno entre la proteína y el DNA a través de las simulaciones. El panel derecho muestra la distribución del número de enlaces de hidrógeno. La línea discontinua corresponde a la mediana, para BORIS es 63,7 y para CTCF es 58,3.

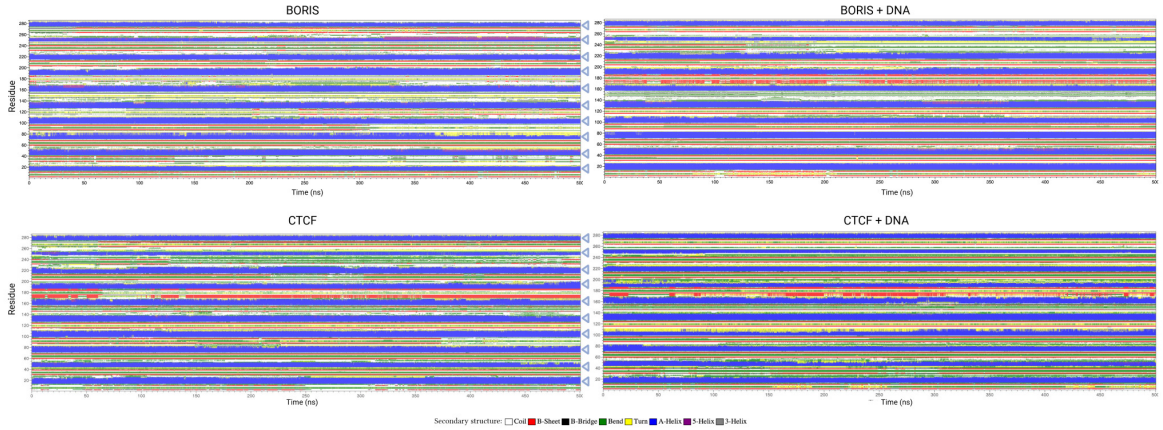


Figura 8.15: Análisis de la estructura secundaria de BORIS y CTCF. Estructura secundaria a través de la simulación de BORIS y CTCF, los residuos están codificados por colores de la siguiente manera: azul, α-hélice; gris, 3-hélice; violeta, 5-hélice; rojo, β-hoja; negro, β-puente; amarillo, giro con enlace de hidrógeno; verde, loop; y blanco, bobina. Las flechas azules indican cada uno de los dedos de zinc.

8.2.3 Obtener un modelo completo para BORIS es complicado debido a las regiones intrínsecamente desordenadas

Debido a que no se dispone de un modelo 3D completo de BORIS y CTCF, se utilizó AlphaFold2 para modelar las isoformas canónicas de CTCF y BORIS, así como cada dominio individualmente. Los resultados se muestran en la **Figura 8.16**. Los residuos están coloreados según la puntuación de confianza por residuo (pLDDT). Se observa que la predicción para la estructura del dominio de unión al DNA tiene una confianza alta, mientras que los dominios terminales tienen puntuaciones de confianza muy bajas, como era de esperar para regiones intrínsecamente desordenadas. Es importante mencionar que, debido a las características del algoritmo AlphaFold, a los modelos les faltan los iones Zn^{2+} y las moléculas de DNA. Al comparar con el modelo creado con Modeller, se observa que hay algunas diferencias en la disposición de los dedos de zinc entre las predicciones de AlphaFold y Modeller, muy probablemente debido a la ausencia de una molécula de DNA en el modelo de AlphaFold (**Figure 8.16**).

Dado que BORIS tiene 17 isoformas reportadas (Link et al., 2013), se utilizó AlphaFold para modelar cada una de ellas. Los resultados se muestran en el **Appendix B.2**. La isoforma B0 es la secuencia canónica. La principal diferencia entre las isoformas es la longitud de los dominios terminales. Las isoformas B4 y B5 carecen de dominios terminales. La isoforma C5 carece de 10 de los 11 dedos de zinc del dominio de unión al DNA. La isoforma C6 tiene un dedo de zinc híbrido formado por la combinación de los ZF4 y ZF9. En general, la confianza de las predicciones es buena en el dominio de unión al DNA, pero baja para los dominios terminales, lo que sugiere que son regiones intrínsecamente desordenadas. Es importante mencionar que, debido a las características del algoritmo AlphaFold, a los modelos les faltan los iones Zn^{2+} y el DNA. Se requieren más análisis para mejorar las predicciones estructurales.

Con el objetivo de evaluar el comportamiento de los dominios terminales, se generó un

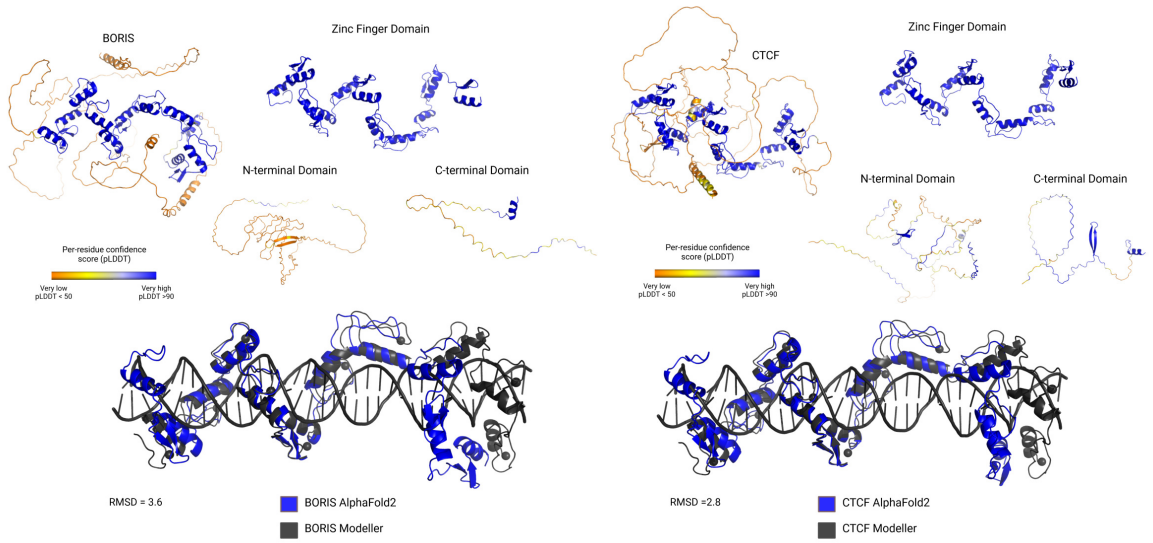


Figura 8.16: Modelado 3D de BORIS y CTCF con AlphaFold. Los residuos están coloreados según la puntuación de confianza por residuo (pLDDT). El panel inferior muestra la alineación entre el modelador (negro) y las predicciones AlphaFold (azul)

modelo completo para BORIS con Modeller. Como templados se usó el modelo previamente generado para los dedos de zinc 2 a 11, mientras que para los extremos terminales se usaron los modelos generados con AlphaFold 2. En cuanto a la cadena de DNA, se modeló como se describió anteriormente utilizando la secuencia del promotor de *hTERT*. Se ha demostrado que BORIS induce la expresión del gen *hTERT* en líneas celulares de cáncer de ovario y testículo, además de que CTCF también puede unirse a su promotor y reprimir su expresión; por lo tanto, se sugiere que BORIS compite con CTCF e induce la expresión de *hTERT* (Renaud et al., 2011). Dado que Renaud et al. validaron el sitio de unión para CTCF y BORIS en el promotor de hTERT, se usaron estos datos para construir una cadena de DNA con la secuencia de esa región (**Figura 8.17a**).

La cadena de DNA del promotor hTERT se insertó en el modelo y la calidad global se evaluó mediante el servidor SAVES v6.0. El gráfico de Ramachandran muestra que el 97,8 % de los residuos pertenecen a las regiones permitidas (**Figura 8.17b**). Además, todos los aminoácidos atípicos pertenecen a los dominios N-terminal y C-terminal.

Para evaluar la estabilidad de nuestro modelo, el complejo proteína-DNA fue sometido a una simulación de dinámica molecular durante 1 µs. Es notable que los valores de RMSD permanecen estables en el tiempo (**Figura 8.18a**). En el caso del RMSF, se observa que el dominio de dedos de zinc permanece estable, mientras que el N-terminal y el C-terminal son más flexibles, como cabría esperar de una región desordenada (**Figure 8.18b**). También se observaron variaciones en el radio de giro (Rg), lo que podría deberse a la flexibilidad de los dominios terminales (**Figure 8.18c**). En general, nuestros datos indican que las regiones terminales de BORIS no se mantienen estables durante la simulación y esto afecta a la estabilidad global de nuestro modelo.

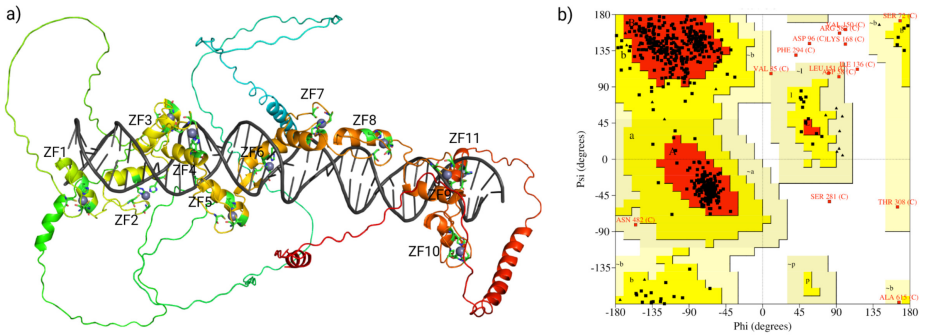


Figura 8.17: Modelo 3D para BORIS acoplado al promotor de *hTERT*. a) Modelo de la proteína completa BORIS creado con Modeller. **b)** Gráfico de Ramachandran para el modelo en el panel **a**. Estadísticas del gráfico: residuos en las regiones más favorecidas [A,B,L] = 88.1 %, residuos en regiones adicionalmente permitidas [a,b,l,p] = 9.7 %, residuos en regiones generosamente permitidas [a, b, l, p] = 1.1 % y residuos en regiones no permitidas = 1.1 %.

Tabla 8.4: Valores de puntuación DOPE, Factor de Calidad Global, los parámetros PROCHECK, y las puntuaciones PROVE para BORIS acoplado al promotor hTERT.

Model	Antes de la minimización	Después de la minimización
Factor de Calidad General (ERRAT)	51.66	78.12
PROVE	7.4 %	6.8 %
PROCHECK		
Residuos en las regiones más favorecidas	88.1 %	87.2 %
Residuos en las regiones adicionalmente favorecidas	9.7 %	9.7 %
Residuos en las regiones generosamente favorecidas	1.1 %	2.1 %
Residuos en las regiones no permitidas	1.1 %	1.1 %

Por otra parte, las cadenas de DNA permanecieron estables durante toda la simulación. Tanto los valores RMSD (**Figura 8.19**a) como RMSF (**Figura 8.19**b) mostraron poca variación. La estructura secundaria para ambas proteínas cuando se simula con o sin DNA se muestra en **Figure 8.15**. También evaluamos el número de enlaces de hidrógeno entre BORIS y el promotor *hTERT*. La media de enlaces de hidrógeno es de 82, que es considerablemente mayor que los enlaces de hidrógeno reportados para nuestros modelos anteriores. En general, la interacción proteína-DNA es estable en nuestro modelo (**Figura 8.20a**), lo que se ve reflejado en los bajos valores de RMSF reportados para el dominio de dedos de zinc.

8.2.4 Las condiciones de simulación no recapitulan la unión metilosensible de CTCF al DNA

Para evaluar la capacidad de nuestras condiciones de simulación actuales para recrear las interacciones DNA-proteína, realizamos una serie de simulaciones de control utilizando

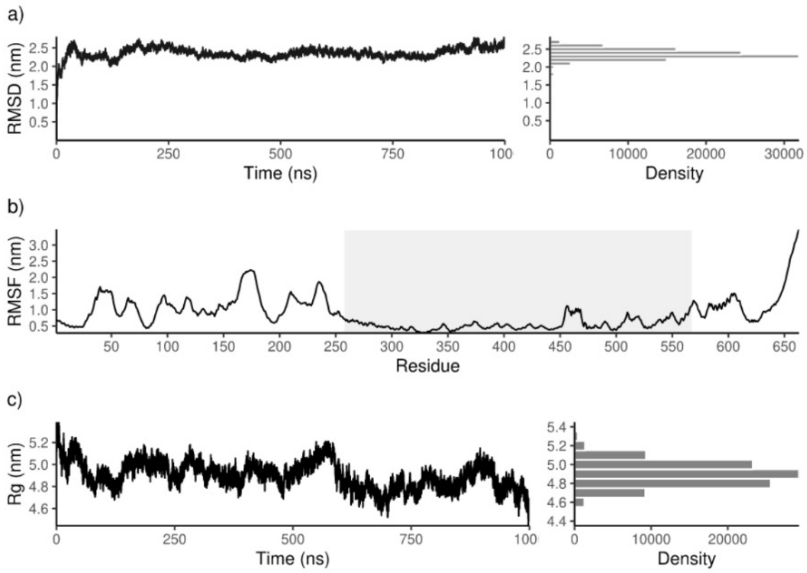


Figura 8.18: Análisis de simulaciones de dinámica molecular. a) El RMSD se calculó utilizando sólo los átomos Cα de la proteína. Se muestra el comportamiento en el tiempo y la distribución global. **b)** RMSF de los átomos Cα de BORIS. El dominio de dedos de zinc está marcado con un recuadro gris. **c)** Radio de giro de la proteína junto con las simulaciones.

las estructuras cristalográficas de CTCF 5T0U y 5T00 como referencia. Estas estructuras representan los dedos de zinc 3-7 de CTCF unidos al DNA (5T0U) y al DNA metilado (5T00). Los modelos cristalográficos se eligieron para eliminar el error asociado al proceso de modelado por homología. Se realizaron simulaciones para varios escenarios, incluyendo CTCF solo (sin DNA), CTCF con DNA, CTCF con DNA metilado y una secuencia de DNA aleatoria como control para evaluar el comportamiento de la proteína. Además, se realizaron simulaciones para la región ZF3-7 modelo por homología BORIS con secuencias de DNA normales y aleatorias. Las simulaciones revelaron que el CTCF no unido es intrínsecamente inestable. Sin embargo, curiosamente, el complejo CTCF-DNA metilado no mostró cambios significativos en su desviación cuadrática media (RMSD) en comparación con el complejo CTCF + DNA, lo que sugiere que el complejo se mantuvo estable durante toda la simulación, contrariamente a informes experimentales anteriores. Por otro lado, cuando se introdujo una secuencia de DNA aleatoria, CTCF se desprendió del DNA como era de esperar. Sorprendentemente, en el caso de BORIS, no hubo diferencias perceptibles entre las simulaciones que utilizaban secuencias de DNA normales y aleatorias. Este hallazgo inesperado contrasta con el comportamiento observado en CTCF (**Figura 8.21**).

8.2.5 Inhibir a BORIS sin inhibir a CTCF

Para identificar inhibidores específicos de BORIS, realizamos análisis de acoplamiento utilizando los modelos construidos para CTCF y BORIS. En primer lugar, se realizó una búsqueda de cavidades para encontrar las superficies con menor conservación en BORIS con respecto a CTCF. La superficie molecular de BORIS se marcó por colores en función

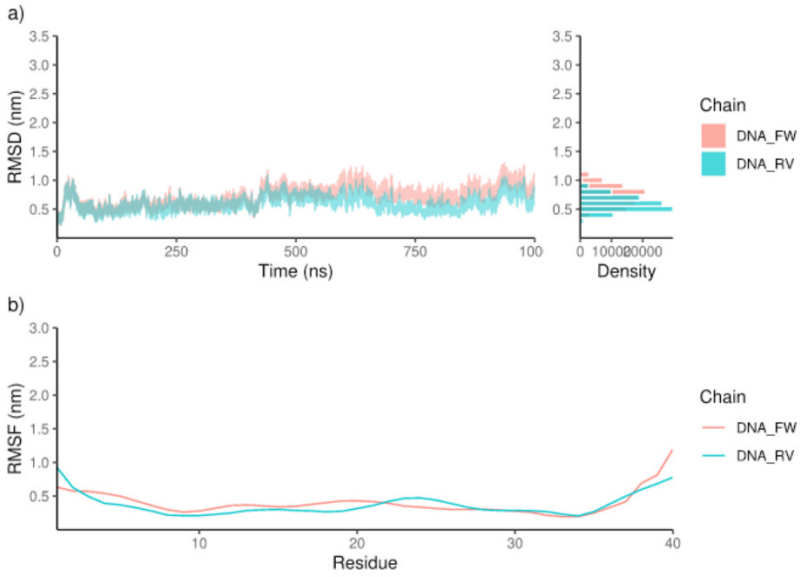


Figura 8.19: Análisis del promotor *hTERT* a través de la simulación. a) RMSD de la molécula de DNA (átomos O4). La distribución de los RMSDs a través de la simulación se muestra a la derecha de los gráficos. **b)** Gráfico RMSF para la columna vertebral del DNA (átomos O4).

de la conservación de aminoácidos en comparación con CTCF (**Figura 8.22**). En particular, este análisis desveló dos sitios distintos en la superficie de interacción proteína-DNA de BORIS que carecen de conservación con CTCF (**Figure 8.22**). Estos sitios no conservados son particularmente interesantes como potenciales dianas farmacológicas, dada su mayor probabilidad de unirse a ligandos específicos. Para probar esta hipótesis, sometimos la biblioteca DrugBank a simulaciones de acoplamiento contra estos sitios identificados tanto en BORIS como en CTCF. Este enfoque tiene como objetivo identificar moléculas que puedan unirse a BORIS sin afectar a CTCF, ofreciendo así información valiosa para el desarrollo de inhibidores dirigidos.

Se emplearon tres programas de software distintos para realizar simulaciones de acoplamiento de la biblioteca de ligandos frente a las proteínas diana. Posteriormente, se generaron múltiples poses para cada ligando, que se clasificaron de mejor a peor puntuación. Para consolidar los resultados, se calculó una clasificación consenso para la pose con mejor puntuación producida por cada herramienta, utilizando el método exponencial de clasificación consenso. Se estableció un valor de corte predeterminado de 0,0015, designando los ligandos con una puntuación superior a este umbral como posibles ligandos de las proteínas diana. Para refinar la selección, se excluyeron los ligandos que demostraban capacidad de unión tanto a CTCF como a BORIS, independientemente del sitio. Como medida de control, se investigaron fármacos adicionales anotados por su afinidad de unión a proteínas de dedos de zinc. Curiosamente, las clasificaciones de estos compuestos no superaron a las observadas para los demás compuestos de la biblioteca, como se muestra en la **Figura 8.23**. No se observó una diferencia significativa entre los puntajes obtenidos por los ligandos conocidos de otras proteínas de dedos de zinc y el resto de ligandos evaluados. En total, se identificaron 159 inhibidores específicos para el sitio 1

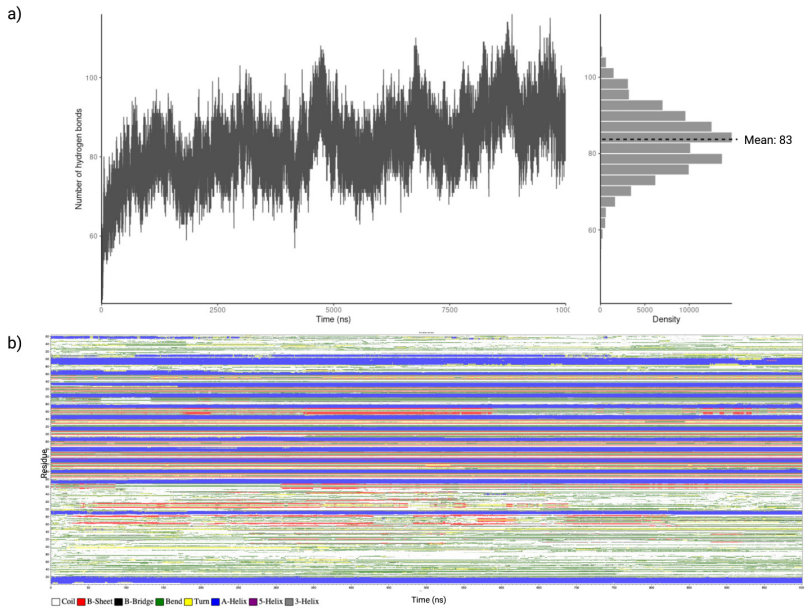


Figura 8.20: Análisis de la unión de BORIS al promotor de *hTERT*. a) Número de enlaces de hidrógeno entre la proteína y el DNA a través de las simulaciones. El panel derecho muestra la distribución del número de puentes de hidrógeno. La línea discontinua corresponde a la media. b) Análisis de la estructura secundaria de BORIS y CTCF. El color de los residuos es como sigue: azul, α-hélice; gris, 3-hélice; violeta, 5-hélice; rojo, β-plegada; negro, β-bridge; amarillo, giro con enlace de hidrógeno; verde, flexión; y blanco, coil.

de CTCF, 136 para el sitio 2 de CTCF, 169 para el sitio 1 de BORIS y 133 para el sitio 2 de BORIS. Los 10 ligandos principales para cada sitio en BORIS se muestran en la **Tabla 8.5**.

Para el Sitio 1, los compuestos principales fueron el patidegib, un inhibidor de la vía de hedgehog empleado en el tratamiento de carcinoma basocelular (Jimeno et al., 2013), el lorlatinib, aprobado por la FDA para el cáncer de pulmón (Waqar and Morgensztern, 2018), y el elacridar, utilizado como adyuvante en el tratamiento de tumores del sistema nervioso central (Agarwal et al., 2011). En el caso del Sitio 2, los compuestos mejor posicionados fueron el venlafaxine, un inhibidor de la recaptura de serotonina, el deprocin, una droga con efectos alucinógenos reportados (Awuchi et al., 2023), y el ulimorelin, un inductor de la motilidad gastrointestinal (James et al., 2020).

De la lista de potenciales ligandos de BORIS se excluyeron todos los compuestos que rankearon alto para CTCF y que mostraron potencial de unión a otras proteínas con dominios de dedos de zinc, esto con el objetivo de disminuir las probabilidades de inhibir blancos secundarios. También se eliminaron los compuestos que puntuaron alto para ambos sitios de unión de BORIS, bajo el supuesto de que al unirse con alta afinidad a ambos sitios es más probable que sean moléculas inespecíficas que podrían también inhibir a otras proteínas con dedos de zinc.

Es notable que tanto para el Sitio 1 como para el 2 se obtuvieron principalmente compuestos aromáticos de alto peso molecular. Esto es deseable debido a que lo que se busca es inhibir la unión de BORIS al DNA, por lo tanto, un compuesto de alto peso

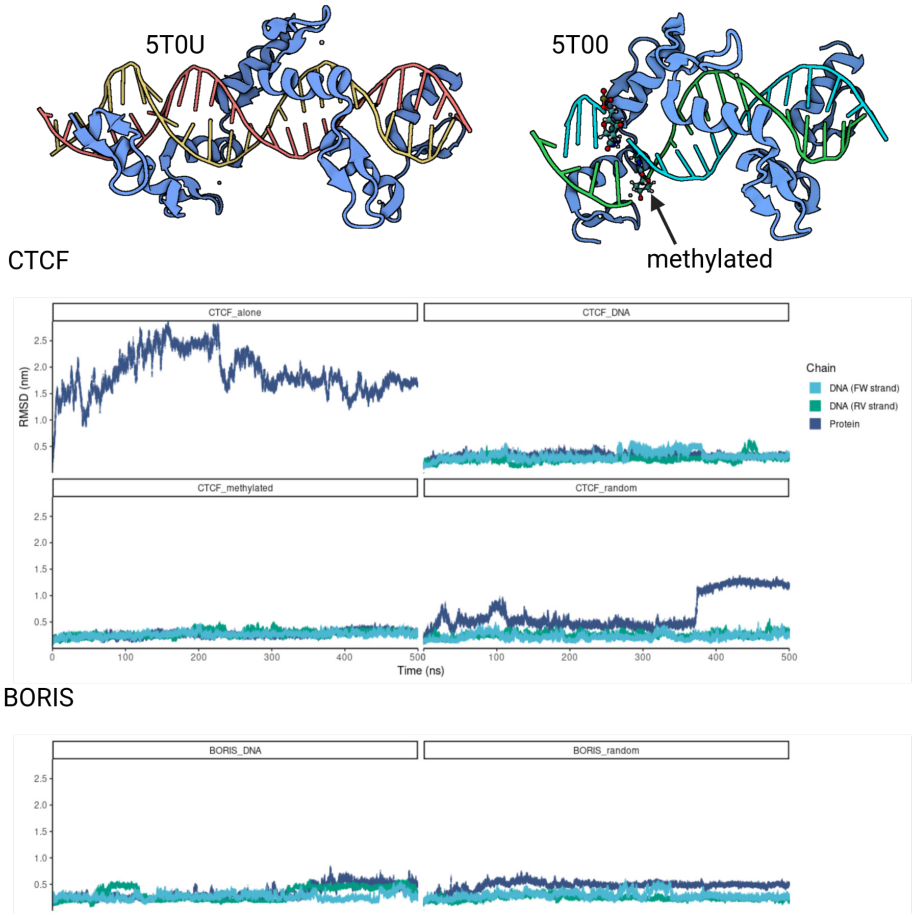


Figura 8.21: Simulaciones de dinámica molecular para los complejos de CTCF y BORIS con el DNA. El panel superior muestra las estructuras de rayos X utilizadas para las simulaciones de CTCF. 5T0U es CTCF unido a su sitio de unión habitual. 5T00 es CTCF unido a una cadena de DNA metilada (la misma secuencia que en 5T0U). El panel inferior muestra los gráficos RMSD de las proteínas y las cadenas de DNA a lo largo de la simulación. Los complejos evaluados son CTCF solo, CTCF + DNA (5T0U), CTCF + DNA metilado (5T00), CTCF + secuencia aleatoria de DNA (secuencia scramble de 5TOU), BORIS + DNA (modelo de homología) y BORIS + DNA aleatorio.

molecular podría ser lo más adecuado para interferir con las funciones de BORIS.

8.3 Las interacciones proteína-proteína en BORIS podrían ser una buena fuente de inhibidores selectivos.

Como se ha mostrado anteriormente, el modelado de las regiones intrínsecamente desordenadas de BORIS no es posible utilizando las herramientas disponibles actualmente, sin embargo, es posible que dichos dominios adquieran una estructura secundaria local al interactuar con otras proteínas. Por lo tanto, el siguiente paso que se tomó fue el modelado de heterodímeros de proteínas con AlphaFold2. La lista de proteínas cuya interacción con

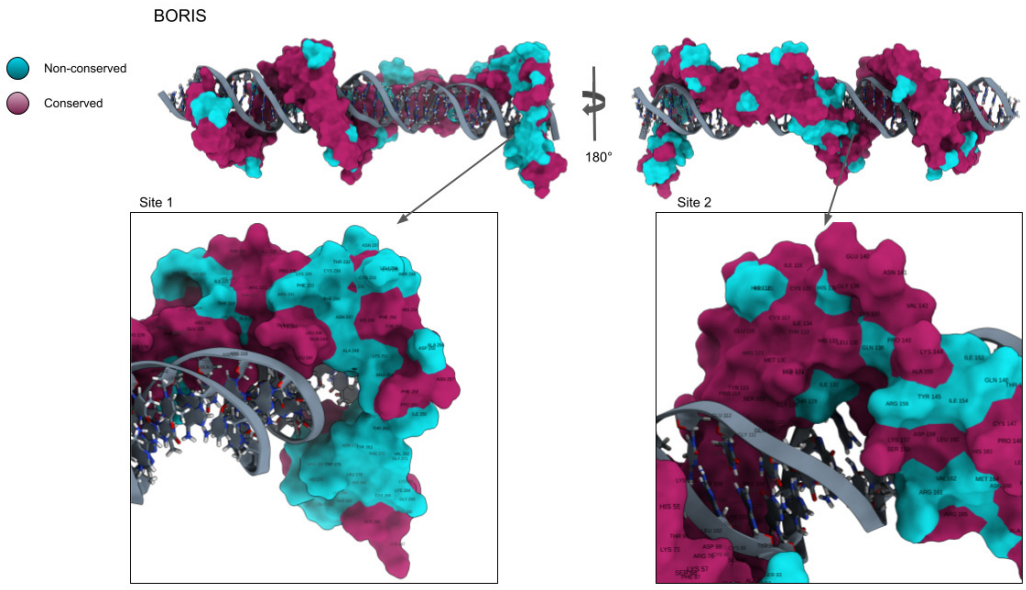


Figura 8.22: Superficie electrónica de BORIS. La superficie está coloreada según su conservación de secuencia respecto a CTCF. Los dos sitios de unión no conservados están ampliados.

BORIS se ha comprobado experimentalmente fue obtenida mediante una búsqueda bibliográfica y se presenta en la **Tabla 8.6**.

Cada pareja se modeló utilizando la información disponible; si se conocía el dominio específico de la interacción, la secuencia de BORIS se truncó para contener solo la región de interacción. En caso contrario, se utilizó como entrada la secuencia completa de la proteína. Para evaluar la confianza de los modelos, se calificaron con la puntuación DockQ (Mirabello and Wallner, 2024). El artículo original establece un umbral de 0,23 para considerar que un modelo es confiable. Entre los modelos creados, sólo el par CTCF-BORIS superó este umbral. Sin embargo, la siguiente mejor predicción fue la interacción entre el dominio C-terminal de BORIS y la proteína de unión a TATA (TBP), con una puntuación DockQ de 0,197. Aunque no superó el umbral, esta puntuación fue superior a las demás. Para mejorar el modelo, se refinó iterativamente la secuencia reteniendo sólo la porción del dominio C-terminal que mostraba interacción (**Figure 8.24**). Con cada iteración, la confianza en el modelo fue mejorando. El modelo final mostró una alta puntuación de DockQ, un bajo error de alineación predicho (PAE) y una alta confianza por residuo (pLDDT). Los resultados muestran que el dominio C-terminal de BORIS interactúa con el dominio de unión al DNA de TBP y que el modelo generado con AlphaFold es confiable. En la **Figura 8.25** se observa que el dominio de unión al DNA de TBP interactúa con una hélice formada por el extremo carboxilo del dominio C-terminal de BORIS.

Debido al alto porcentaje de identidad entre BORIS y CTCF y a la ausencia de informes previos en la literatura sobre una interacción entre CTCF y TBP, se modeló el complejo CTCF-TBP utilizando AlphaFold y el dominio C-terminal de CTCF. Sin embargo, los niveles

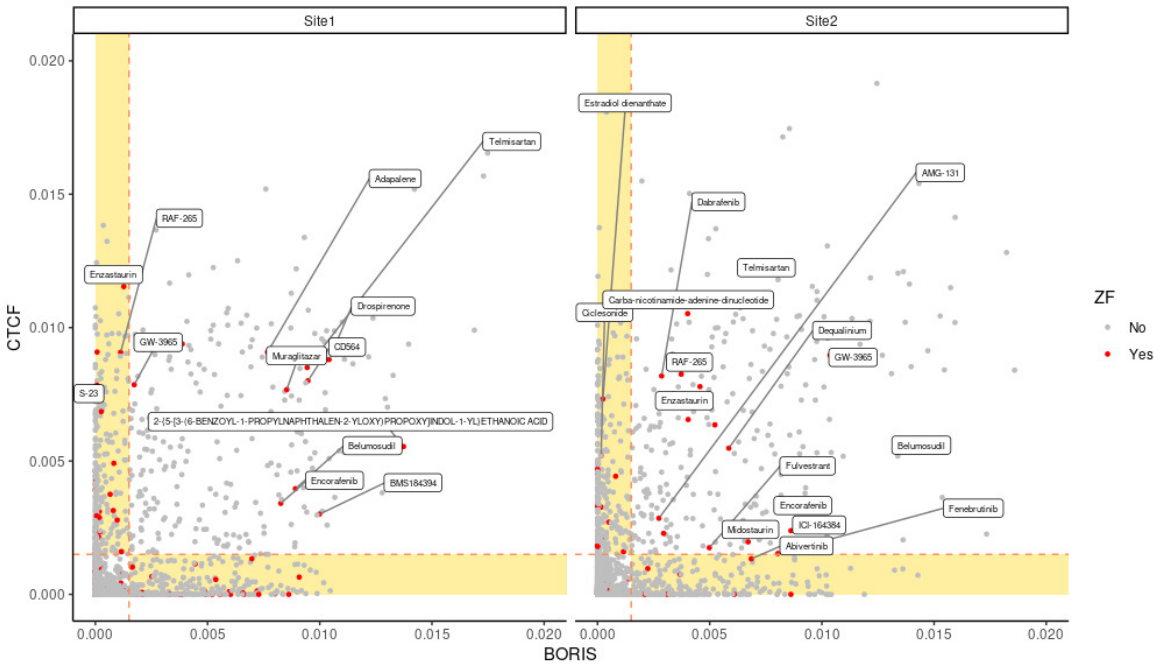


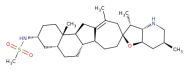
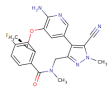
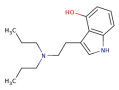
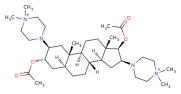
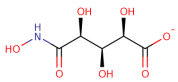
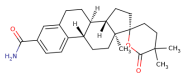
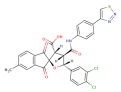
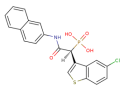
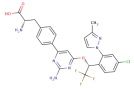
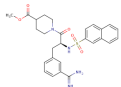
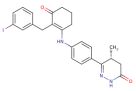
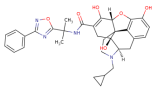
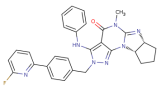
Figura 8.23: Ranking exponencial de consenso para los ligandos de DrugBank acoplados contra CTCF y BORIS. Los recuadros amarillos marcan la localización de los ligandos considerados como específicos para cada una de las proteínas. Los puntos rojos indican fármacos que tienen como blanco proteínas de dedos de zinc.

de confianza de estos modelos no fueron tan altos como los obtenidos para BORIS, y el PAE para las predicciones de CTCF fue menor (**Figure 8.26**). Se siguió el mismo proceso de refinamiento que para el complejo BORIS-TBP y en este caso se encontró que solamente uno de los modelos tiene un puntaje DockQ mayor al umbral establecido y se observa que los valores de PAE son más altos que para los modelos de BORIS, lo que indica una menor confianza del algoritmo en la predicción.

Al comparar las secuencias de las regiones C-terminal de CTCF y BORIS, observamos una similitud del 25 % (**Figura 8.27a**). Sin embargo, la hélice identificada como la región de interacción con TBP en BORIS no se conserva en CTCF. Esto es evidente a partir de la predicción de regiones desordenadas (DISOPRED) y de estructura secundaria en **Figura 8.27b**. Basándonos en nuestros datos, se puede hipotetizar que TBP tiene el potencial de unirse a BORIS pero no a CTCF. Sin embargo, para validar esta hipótesis, se realizaron experimentos de co-inmunoprecipitación (IP).

Para la validación de las interacciones entre TBP, BORIS y CTCF, se eligieron como modelo experimental las células HEK-293 ya que carecen de expresión endógena de BORIS (**Figura 8.28a**). Se transfectaron en las células vectores de sobreexpresión de BORIS y CTCF, ambos marcados con hemaglutinina (HA). Tras la transfección, se realizó la extracción de proteínas e inmunoprecipitación con un anticuerpo contra HA para posteriormente llevar a cabo un Western blot para BORIS, CTCF y TBP. La vinculina sirvió como control negativo debido a su presencia en las uniones de adhesión focal (Bays and DeMali, 2017), situadas

Tabla 8.5: Top 10 ligandos para BORIS. Se muestran los 10 ligandos con la puntuación más alta para ambos sitios de unión en BORIS. La lista excluye moléculas que muestran potencial de unión a CTCF o que se ha reportado que son inhibidores de otras proteínas con dominios de dedos de zinc.

Sitio 1			Sitio 2		
DB12655	Patidegib		DB00285	Venlafaxine	
DB12130	Lorlatinib		DB13990	Deprocin	
DB04881	Elacridar		DB12128	Ulimorelin	
DB01338	Pipecuronium		DB03734	Xylarohydroxamate	
DB07700	DB07700		DB04330	Bilh 434	
DB07680	DB07680		DB12148	Menatetrenone	
DB02009	L-756423		DB00677	Isoflurophate	
DB14218	Telotristat		DB01745	DB01745	
DB01640	DB01640		DB16266	Olodanrigan	
DB11691	Naldemedine		DB15039	ITI-214	

fuera del núcleo, lo que hace muy improbable cualquier interacción con BORIS o CTCF. Ninguna de las dos proteínas mostró interacción con vinculina, además como control del anticuerpo contra HA se utilizó uno dirigido a IgG, que no mostró señal para ninguna de las proteínas evaluadas. Los resultados demostraron la recuperación exitosa de TBP con la IP de proteínas marcadas con HA, indicando la formación de complejos TBP-BORIS y TBP-CTCF (**Figura 8.28**).

Tabla 8.6: Interacciones proteína-proteína conocidas de BORIS. Las proteínas que fueron validadas experimentalmente para interactuar con CTCF también están etiquetadas en rojo.

Tipos de proteínas	Proteína	Función del complejo	Evidencia experimental	Referencias
Proteínas asociadas a la cromatina	PRMT7	Metilación de arginina para controlar la impronta.	Inmunoprecipitación.	(Jelinic et al., 2006)
	CTCF	Función desconocida en la espermatogénesis.	Ensayo de ligación de proximidad in situ. Inmunoprecipitación	(Rivero-Hinojosa et al., 2021), (Pugacheva et al., 2015)
	BAG6	Activación transcripcional de c-myc y BRCA1.	Ensayo de dos híbridos en levadura	(Nguyen et al., 2008)
	SET1A			
	POGZ	Desconocida	Ensayo de dos híbridos en levadura	(Nguyen et al., 2008)
	SRCAP			
Factores de transcripción	TBP	Activación transcripcional de MAGE-A1.	Ensayo de pull down.	(Schwarzenbach et al., 2014)
	SP1	Activación transcripcional de NY-ESO-1.	Inmunoprecipitación. Ensayo de pull down.	(Kang et al., 2007), (Schwarzenbach et al., 2014)
	ELF2	Desconocida	Ensayo de dos híbridos en levadura	(Nguyen et al., 2008)
	HCFC2, HCFC1			
	MGA			
	TLK2			
	NFAT5			
	ZNF518			
	ATF7			
	MKL2			
Proteínas de unión al DNA	Ku70	Reparación de daño en el DNA.	Inmunoprecipitación	(Zhang et al., 2022b)
	UBF	Regulación transcripcional del DNA ribosomal.	Inmunoprecipitación.	(van de Nobelen et al., 2010)
	CHD8	Desconocida	Ensayo de dos híbridos en levadura	(Nguyen et al., 2008)
Proteínas de señalización	CSTA			
	FHL2			

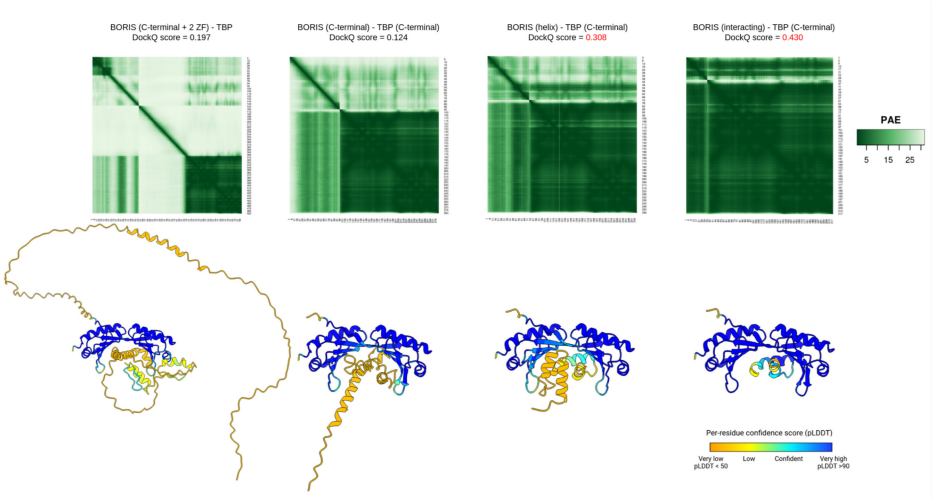


Figura 8.24: Modelos de AlphaFold2 para los complejos BORIS-TBP. El panel superior muestra el Error de Alineamiento Predicho (PAE), y el panel inferior muestra los modelos coloreados según la confianza por residuo (pLDDT).

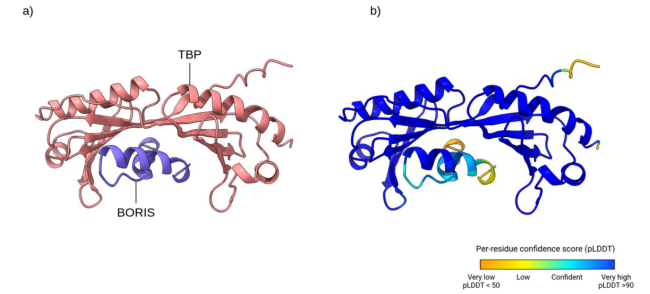


Figura 8.25: Predicción de AlphaFold2 para el complejo BORIS-TBP. a) El modelo está coloreado según la cadena. El rosa corresponde a TBP y el violeta a BORIS. **b)** Confianza por residuo (pLDDT) del modelo predicho para el complejo BORIS-TBP.

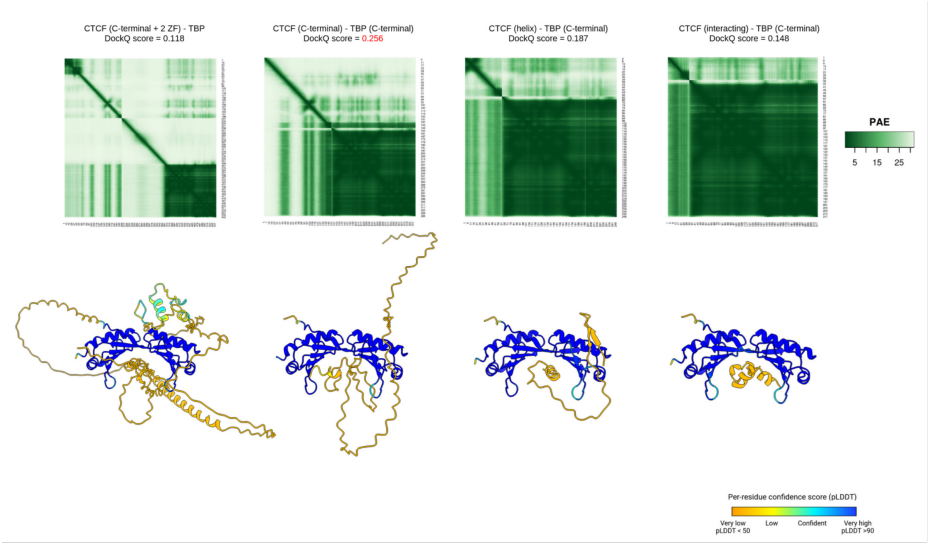


Figura 8.26: Predicción de AlphaFold2 para el complejo CTCF-TBP. El panel superior muestra el Predicted Aligned Error (PAE), el panel inferior muestra los modelos coloreados por la confianza por residuo (pLDDT).

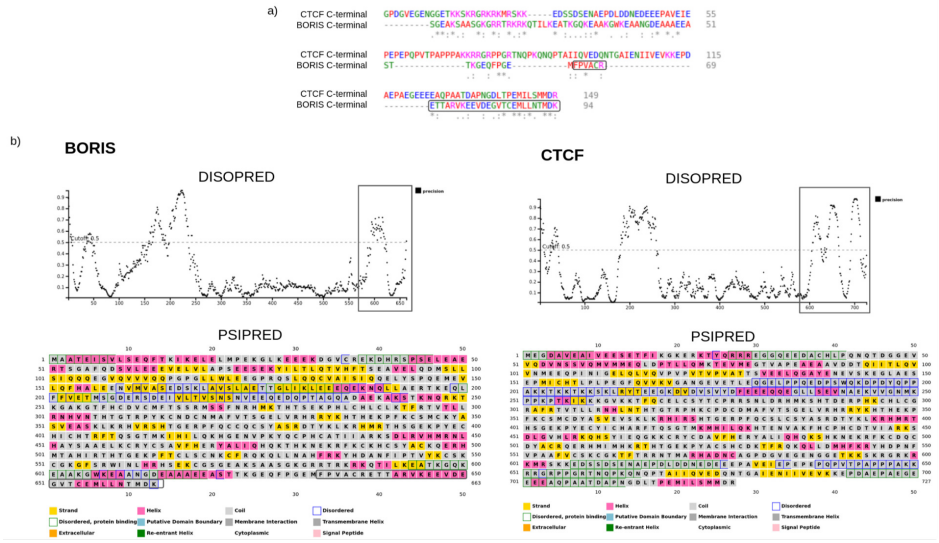


Figura 8.27: Comparación de secuencias entre los dominios C-terminales de BORIS y CTCF. a) Alineamiento de secuencias para los extremos carboxilo terminal de CTCF y BORIS. Los aminoácidos están coloreados según sus propiedades fisicoquímicas. La hélice de BORIS que se supone interactúa con TBP está marcada con un cuadro negro. **b)** Predicción de regiones desordenadas (DISOPRED) y predicción de estructura secundaria (PSIPRED) calculadas con el servidor http://bioinf.cs.ucl.ac.uk/.

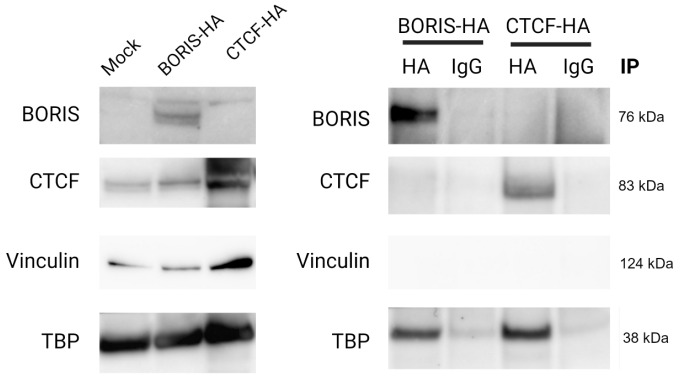


Figura 8.28: Análisis de inmunoprecipitación de lisados de células HEK-293 transfectadas con proteína BORIS o CTCF marcada con HA. Tras la inmunoprecipitación con un anticuerpo HA-específico, se utilizó Western blot para detectar BORIS, CTCF, TBP y Vinculina. El panel izquierdo muestra las muestras de entrada y el derecho las muestras después de la IP. Se presenta un blot representativo de 3 réplicas biológicas independientes.

9. Discusión

La alteración de procesos epigenéticos es una de las firmas moleculares que caracterizan al cáncer (Hanahan, 2022). Dichos procesos involucran principalmente cambios en los patrones de metilación del DNA, en el establecimiento de modificaciones postraduccionales de histonas y los patrones de expresión génica, además de desregulación de la estructura tridimensional de la cromatina, en específico de los dominios topológicamente asociados (Besselink et al., 2023; Lee and Kim, 2022; Akdemir et al., 2020). Por lo tanto, las proteínas involucradas en estos procesos son un blanco de interés para el desarrollo de estrategias terapéuticas enfocadas a contrarrestar la progresión del cáncer. En este sentido, la mayoría de los esfuerzos se han enfocado en caracterizar moléculas inhibitorias de las DNMTs, mientras que otros procesos como la desmetilación de histonas y la regulación de asas de cromatina aún carecen de fármacos enfocados a los epifactores que los regulan (Yu et al., 2024; Lee and Kim, 2022). Dicho fenómeno puede ser debido a que varios elementos epigenéticos son regulados por proteínas que comparten altos porcentajes de identidad entre sí, lo que dificulta la especificidad de las terapias y aumenta las probabilidades de obtener moléculas con efectos secundarios indeseables.

En el presente trabajo se evaluaron dos grupos de epifactores que se han clasificado como potenciales blancos terapéuticos en cáncer. El primero es la subfamilia de desmetilasas KDM4, las cuales están implicadas en el desarrollo del cáncer por su capacidad para alterar el estado de la cromatina e influir en la expresión génica (García et al., 2016). El segundo grupo son las proteínas de unión al DNA CTCF y BORIS, las cuales regulan de forma dinámica la formación de asas de cromatina y regiones topológicamente asociadas (Pugacheva et al., 2020). Ambos casos son interesantes desde el punto de vista farmacológico e involucran proteínas homólogas, en este trabajo se evaluaron por separado con el propósito de proponer nuevas alternativas terapéuticas.

La expresión de los miembros de la subfamilia KDM4 (KDM4A, B, C, D y E) está estrechamente regulada en tejidos no neoplásicos pero a menudo desregulada en varias neoplasias (Guerra-Calderas et al., 2015; Chen et al., 2020; Wu et al., 2021). En este estudio, utilizando amplios conjuntos de datos de RNAseq de tumores y tejidos no neoplásicos, se demostró que las proteínas KDM4 son relevantes para la mayoría de tumores. Mediante análisis transcriptómicos, se observó que las proteínas KDM4, cuando se sobreexpresan, desencadenan cambios de expresión que son distintos en cada tipo de cáncer y que tienen diferentes consecuencias en la supervivencia de los pacientes. Se mostró

también que los procesos influenciados por las KDM4s involucran varias firmas moleculares del cáncer, incluyendo angiogénesis, respuesta inmune e inflamatoria, ciclo celular y transición epitelio-mesénquima (Hanahan, 2022), lo que concuerda con reportes previos que sugieren que estas proteínas tienen un papel importante para la enfermedad (Katoh and Katoh, 2004; Shin and Janknecht, 2007c; Labbé et al., 2013).

Nuestros resultados sugieren que los miembros de la subfamilia KDM4 son dianas farmacológicas prometedoras para el desarrollo de alternativas terapéuticas en distintos tipos de cáncer. Es notable que cada una de las KDM4 tiene diferente impacto en la supervivencia de los pacientes y en los patrones de expresión que inducen. Este hecho sugiere que las terapias dirigidas contra las enzimas KDM4 deben ser específicas, a fin de acoplarse a las necesidades de cada individuo. Sin embargo, un reto al que hay que enfrentarse durante el desarrollo de inhibidores específicos es la semejanza de los dominios catalíticos de todos los miembros de la familia JMJC. Dado que estas proteínas comparten mecanismo catalítico, sus sitios activos tienen un alto porcentaje de identidad, lo que complica la búsqueda de ligandos dirigidos a una sola enzima (Markolovic et al., 2016). El establecimiento de un mecanismo de unión específico para un inhibidor competitivo de una o algunas proteínas KDM4 es importante para un adecuado control de la terapia. De hecho, se sabe que la mayoría de los inhibidores de KDM4 descritos se dirigen también a otras demetilasas de histona, lo que limita su uso como agentes terapéuticos en pacientes con cáncer (Chin and Han, 2015; Baby et al., 2021).

Para la búsqueda de nuevas moléculas con actividad inhibitoria de KDM4, se usaron distintos repositorios de moléculas pequeñas, entre ellas una colección de compuestos naturales (COCONUT), la cual es una fuente de diversas moléculas con actividad biológica. Además, se incluyeron las bases de datos DrugBank y FDA, las cuales han sido ampliamente utilizadas en estudios de reposicionamiento de fármacos (Masoudi-Sobhanzadeh et al., 2020; Sorokina et al., 2021a). Comparadas con los compuestos de la FDA y DrugBank, las moléculas de la base de datos COCONUT obtuvieron una mejor puntuación en los análisis de acoplamiento molecular contra la subfamilia KDM4, lo que sugiere que los compuestos naturales podrían ser una rica fuente de terapias contra el cáncer (Pushpakom et al., 2019). Sin embargo, es importante tener en cuenta que no se hizo una corrección de la calificación FRED/Chemgauss para contrarrestar el efecto del tamaño de las moléculas, por lo que también es de esperarse que los compuestos de COCONUT tengan una puntuación más alta, ya que es sabido que las calificaciones de acoplamiento molecular toman en cuenta el número de interacciones proteína-ligando, por lo que es indudable que moléculas de mayor peso molecular tendrán a su vez mayor número de interacciones y por subsecuente, mejor puntuación (Li et al., 2019a). También se observó que la cavidad del sitio activo de estas enzimas es de gran tamaño, por lo que puede acomodar ligandos de alto peso molecular, tales como las presentes en la base de datos COCONUT. Es importante destacar que los compuestos naturales se han utilizado durante siglos para tratar una amplia gama de enfermedades, incluyendo el cáncer (Fang et al., 2020; Gómez-Cansino et al., 2017; Gutiérrez-Rebolledo et al., 2017).

Entre las moléculas propuestas por este trabajo destacan los catecoles y flavonoides, que mostraron mayores puntajes en los análisis de acoplamiento molecular. Estos ligandos han sido propuestos como inhibidores de las KDM4 debido a su elevado contenido de grupos OH, que les permite actuar como reguladores de radicales libres (Baby et al., 2021). Estos grupos funcionales favorecen la interacción con los cofactores metálicos en el sitio activo de las proteínas KDM4, los cuales participan en el mecanismo de transferencia de electrones durante la desmetilación de lisinas (Warshakoon et al., 2006). Los grupos OH de los catecoles (como flavonoides o fenoles) pueden formar enlaces de coordinación con dichos metales, compitiendo con los sustratos naturales de la proteína y aumentando la afinidad ligando-proteína (Xu, 2013). Varios de estos compuestos interfieren en procesos epigenéticos; por ejemplo, los flavonoides son compuestos presentes en la frambuesa negra (y en muchas otras plantas) que inhiben la actividad de la DNA metiltransferasa 1 (DNMT1) potenciando la expresión de genes supresores de tumores (Wang et al., 2013). Sin embargo, aunque hay trabajos previos sobre moléculas naturales que podrían interferir con la actividad de la subfamilia KDM4, hasta ahora no se conocen inhibidores naturales directos (Guillade et al., 2018). Es por esto que en nuestra búsqueda de inhibidores específicos de la subfamilia KDM4 utilizamos la base de datos COCONUT, que reúne 406,744 productos naturales de más de 50 bases de datos diferentes, donde casi la mitad de los compuestos proceden principalmente de plantas, hongos, bacterias y, en menor medida, de origen animal o marino (Capecchi and Reymond, 2021; Sorokina et al., 2021b). La mayoría de estos compuestos han sido utilizados como medicina tradicional en China, India (Ayurveda), Japón (Kampo), Corea, México, entre otros países (Gutiérrez-Rebolledo et al., 2017; Yuan et al., 2016a) y proceden de Asia, África, Brasil y México (Sorokina et al., 2021b).

Un ejemplo de las moléculas identificadas en este estudio es el galato de epigalocatequina (EGCG), un metabolito secundario del té (*Camellia sinensis*) que contiene un grupo catecol; este ha sido ampliamente estudiado por su efecto sobre la metilación del DNA y actualmente es usado como agente terapéutico para diabetes e hipertensión (Fang et al., 2003; Choi et al., 2009). En este estudio, el EGCG mostró una interacción favorable con KDM4, sugiriendo su potencial como inhibidor. Además, se ha reportado que el EGCG puede quelar metales divalentes como el zinc y se ha propuesto en ensayos clínicos como coadyuvante en varios procesos terapéuticos (Shirakami and Shimizu, 2018). Otro caso similar es el del pirogalol, que contiene tres grupos OH en lugar de dos y ha sido investigado por sus efectos citotóxicos en líneas celulares de cáncer de pulmón, mostrando su potencial terapéutico (Yang et al., 2009).

Además de los enlaces de coordinación, otras fuerzas no covalentes como puentes salinos, interacciones hidrofóbicas, enlaces de hidrógeno y apilamiento π influyen en la unión proteína-ligando (Churchill and Wetmore, 2009; Houser et al., 2020; Wilson et al., 2014). Las interacciones π entre anillos aromáticos son clave en la formación del complejo con KDM4, ya que sus residuos aromáticos en los sitios activos favorecen esta interacción. Moléculas con anillos aromáticos tienen mayor afinidad, mientras que las que carecen de estas interacciones muestran menor eficacia. Inhibidores con anillos aromáticos y grupos amina han demostrado interactuar bien con los cofactores metálicos de KDM4 y

proporcionan una oportunidad para el desarrollo de una nueva generación de inhibidores de novo para la subfamilia KDM4 (Małecki et al., 2019).

Dado que la especificidad es difícil de alcanzar para los inhibidores de las KDM, es posible usar este hecho en beneficio de la terapia al buscar ligandos que puedan dirigirse a todas las KDM4 relevantes para una neoplasia específica. Como se muestra en este trabajo, las moléculas identificadas podrían tener un efecto terapéutico amplificado al modular, no sólo las funciones de KDM4, sino procesos celulares completos, modificando la actividad de proteínas implicadas en las mismas vías. Este mecanismo de acción ha sido propuesto para otras enfermedades y dianas proteicas con anterioridad (Cheng et al., 2019); sin embargo, el estudio de los inhibidores de KDM4 sigue siendo abordado sin tener en cuenta el contexto molecular necesario para su correcta función (Chin and Han, 2015; Baby et al., 2021). En este sentido, nuestros datos sugieren que los ligandos propuestos podrían ser utilizados como terapias multiblanco, aprovechando su actividad inespecífica contra otras KDMs e inclusive contra proteínas cuya expresión es inducida por las KDM4. Esto abre una nueva ventana de oportunidades para el uso de inhibidores de la subfamilia KDM4 y contribuye a la búsqueda de terapias personalizadas contra el cáncer.

Por ejemplo, las redes construidas para el feocromocitoma y el paraganglioma muestran que cuando hay sobreexpresión de KDM4A, también hay sobreexpresión de proteínas implicadas en la regulación positiva de la expresión génica, las cuales están estrechamente relacionadas con la regulación negativa de la función de metilación de H3K9 encontrada en las redes de cáncer corticosuprarrenal, adenocarcinoma de pulmón y carcinoma endometrioide del cuerpo uterino. Además, encontramos que algunos procesos relacionados a la sobreexpresión de las KDM4 están relacionados con GPCRs monoaminérgicos o con vías de adicción a la cocaína en carcinomas tiroideos y uterinos, esto es altamente relevante dado el hecho de que proteínas implicadas en estos procesos metabólicos han demostrado previamente estar afectadas en algunos cánceres; como linfoma, próstata, cáncer de pulmón y algunos cánceres cerebrales (Shih, 2018; Rybaczyk et al., 2008). Ciertos fármacos que mostraron alta afinidad por KDM4A son actualmente usados como inhibidores de otras proteínas con función epigenética, tal es el caso de la azacitidina y el EGCG que son ligandos de DNMT1 también. Por lo tanto, estos resultados sugieren que dirigirse a varias proteínas KDM4 a la vez también puede ser un enfoque terapéutico prometedor ya que los fármacos dirigidos a ellas pueden modular también a las proteínas de los procesos celulares que contribuyen al fenotipo neoplásico.

Por otra parte, BORIS y CTCF son otro caso de proteínas que comparten altos niveles de identidad entre sí y cuya función es de interés en cáncer. A diferencia del caso de las enzimas KDM4, BORIS y CTCF tienen funciones y patrones de expresión totalmente opuestos, en el caso de CTCF es una proteína de expresión ubicua que es esencial para el mantenimiento de la estructura de la cromatina, mientras que BORIS está presente sólo en testículo y en algunos tipos de cáncer, principalmente de ovario (Klenova et al., 2002; Del Moral-Morales et al., 2023). BORIS es un prometedor blanco terapéutico debido a que se ha visto que su abatimiento interfiere con la proliferación y crecimiento de células tumorales, además de que sus patrones de expresión lo hacen ideal para el tratamiento del cáncer de ovario (Mkrtichyan

et al., 2011; Salgado-Albarrán et al., 2019). Sin embargo, hay tres principales problemas que han limitado el surgimiento de nuevas terapias dirigidas a esta proteína, las cuáles se discutirán a continuación junto con las estrategias que se tomaron para sobreponerse a ellas.

En primer lugar, no se cuenta con una estructura cristalográfica para BORIS. Uno de los requisitos principales para los estudios de acoplamiento molecular y de cribado virtual es la disponibilidad de una estructura determinada experimentalmente para la proteína blanco (Li et al., 2019a). Esta limitante logró ser superada mediante la creación de modelos por homología, usando como plantillas las estructuras disponibles para CTCF. Mediante este abordaje se logró obtener modelos tridimensionales confiables de los dominios de unión al DNA tanto de BORIS como de CTCF. Las simulaciones de dinámica molecular corroboraron la estabilidad conformacional de estos modelos al interactuar con secuencias de DNA consenso, pero revelaron limitaciones en la predicción de fenómenos como la metilación del DNA, que se sabe inhibe la unión de CTCF (Renda et al., 2007; Hashimoto et al., 2017). Estos resultados sugieren que las interacciones electrostáticas en la interfaz proteína-DNA, frecuentemente sobreestimadas en las simulaciones (Yoo et al., 2020), podrían desempeñar un papel crucial en la especificidad de unión. Las simulaciones dependen de la precisión de los modelos de interacción molecular (campos de fuerza) por lo que la correcta recapitulación de los detalles atómicos de la interacción está sujeta a la precisión del campo de fuerza usado (Tucker et al., 2022).

Una segunda limitación en la búsqueda de un inhibidor específico para BORIS radica en la elevada similitud de su secuencia de aminoácidos con la proteína CTCF, particularmente en los dominios responsables de la unión al DNA. Como se ha mencionado, BORIS y CTCF comparten aproximadamente un 80 % de identidad en el dominio de dedos de zinc, lo que conlleva diversas implicaciones funcionales. Entre ellas destaca la competencia de BORIS por los sitios de unión de CTCF, resultando en la disrupción de las estructuras de cromatina (Pugacheva et al., 2015). Los análisis estructurales comparativos de ambas proteínas revelan conservación de todos los residuos de aminoácidos que interactúan directamente con el DNA, lo que dificulta considerablemente el diseño de moléculas capaces de inhibir selectivamente la unión de BORIS sin afectar a CTCF (Marshall et al., 2014). No obstante, se identificaron dos regiones con diferencias estructurales significativas, lo que sugiere su potencial como blancos terapéuticos.

Mediante estudios de acoplamiento molecular (docking), se ha llevado a cabo una búsqueda de compuestos candidatos en la base de datos DrugBank, seleccionada por su contenido de moléculas biológicamente activas aprobadas para uso clínico. Esta estrategia computacional nos permite explorar un amplio espacio químico de moléculas biológicamente activas, agilizando así la identificación de potenciales inhibidores. Entre los compuestos identificados, destacan moléculas aprobadas clínicamente para el tratamiento de tumores sólidos, como patidegib (Jimeno et al., 2013), lorlatinib y elacridar (Agarwal et al., 2011). La mayoría de estos compuestos presentan características estructurales comunes, tales como alto peso molecular y presencia de anillos aromáticos. Nuestro objetivo es diseñar moléculas capaces de interferir en la interacción proteína-DNA mediada por BORIS, con el fin de desplazar a la proteína de su sitio de unión y restaurar la función

normal de la cromatina. Al excluir compuestos con afinidad por CTCF, aseguramos que la inhibición de BORIS no afecte negativamente otros procesos reguladores.

La inhibición de las superficies de interacción DNA-Proteína mediante moléculas pequeñas se considera difícil de lograr debido a que generalmente los factores transcripcionales carecen de bolsillos que puedan acomodar ligandos, además de que los dominios de unión al DNA generalmente están conservados entre diferentes factores de transcripción que pertenecen a la misma familia, tal es el caso de BORIS y CTCF, como ya se ha expuesto, o el de los receptores andrógenos, estrógenos y progesterona (Bushweller, 2019; Radaeva et al., 2021). Sin embargo, el uso de métodos computacionales ha permitido sobreponerse a los retos que implica la identificación de compuestos dirigidos a la interacción DNA-Proteína, tal es el caso del compuesto VPC-14449, el cual inhibe la interacción del Receptor de Andrógenos con el DNA sin afectar a ninguno de los otros receptores de hormonas esteroides, además de que este mecanismo de acción evita la resistencia que eventualmente desarrollan los tumores dependientes de andrógenos (Li et al., 2014; Dalal et al., 2017).

Además del receptor de andrógenos, el uso de estrategias computacionales ha facilitado la identificación de moléculas que interfieren con la interacción DNA-proteína de otros factores de transcripción tales como STAT3 (Buettner et al., 2011), PAX2 (Grimley et al., 2017) y FOXM1 (Gormally et al., 2014). Todas estas moléculas tienen en común que son de alto peso molecular y contienen varios anillos aromáticos, además de que han mostrado actividad *in vitro* con una alta selectividad a pesar de que las proteínas contra las que van dirigidas tienen homólogos cercanos (Radaeva et al., 2021).

Finalmente, el tercer desafío para el diseño de inhibidores de BORIS radica en la naturaleza intrínsecamente desordenada de sus extremos amino y carboxilo terminales. La falta de estructura secundaria y terciaria definida en estas regiones dificulta significativamente tanto el modelado molecular como la determinación por medio de métodos experimentales y, por consiguiente, la identificación de sitios de unión para potenciales inhibidores. Esta limitación resulta particularmente desafortunada dado que estas regiones presentan una menor identidad de secuencia con CTCF, lo que sugiere un potencial para el desarrollo de inhibidores selectivos.

Debido a la baja identidad de estos dominios, se ha hipotetizado que las consecuencias funcionales opuestas de la unión de BORIS y CTCF en sus regiones blanco son debido al reclutamiento de diferentes cofactores (Del Moral-Morales et al., 2023). Dado que se ha reportado que las regiones intrínsecamente desordenadas pueden adquirir una estructura secundaria temporal al formar complejos proteicos con otras proteínas (Chakrabarti and Chakravarty, 2022), se modelaron las estructuras de las interacciones proteína-proteína de cada par descrito en la literatura usando AlphaFold, el cual ha demostrado que puede generar estas estructuras con alta precisión cuando el valor de PAE es bajo (Zhu et al., 2023). Se observó que de todos los pares modelados, únicamente se obtuvo el par CTCF-BORIS con alta confianza; el resto de estos modelos no pasaron el umbral de DockQ score para considerarlos confiables. El heterodímero CTCF-BORIS está ampliamente

reportado en la literatura y, de acuerdo con nuestros modelos y con los datos experimentales, la interacción se da entre los dominios de unión al DNA (Pugacheva et al., 2015), no entre los extremos terminales, por lo que no sirve para el propósito de la búsqueda de un inhibidor que se dirija a los extremos amino terminal de BORIS.

En consecuencia, se seleccionó como la interacción más probable aquella que involucra a BORIS y la proteína de unión a cajas TATA (TBP). Los modelos predictivos generados por AlphaFold sugieren que el dominio carboxilo terminal de BORIS puede establecer un contacto directo con el dominio de unión al DNA de TBP mediante una α-hélice parcialmente conservada en CTCF. Aunque los modelos del dímero CTCF-TBP presentan menor confianza, la predicción de una interacción estable entre BORIS y TBP resulta coherente con la función activadora de BORIS. Al formar parte del complejo de preiniciación de la transcripción (PIC), TBP desempeña un papel crucial en la regulación de la expresión génica Patel et al. (2018); Majello et al. (1998). La unión de BORIS a TBP podría facilitar el reclutamiento del PIC a los promotores de los genes diana, desencadenando así la activación transcripcional (Hong et al., 2005; Vatolin et al., 2005). Este mecanismo de activación transcripcional es congruente con estudios previos que han demostrado la capacidad de diversas proteínas para modular la actividad de TBP a través de interacciones con su surco de unión al DNA. En estos casos, regiones intrínsecamente desordenadas adquieren estructura secundaria al unirse a TBP, favoreciendo la regulación de la transcripción (Mal et al., 2007; Anandapadamanaban et al., 2013). Los resultados obtenidos sugieren que BORIS podría emplear una estrategia similar para activar la transcripción de sus genes diana.

Experimentos de co-inmunoprecipitación han evidenciado la capacidad de TBP para interactuar tanto con CTCF como con BORIS. No obstante, estudios más recientes han revelado una localización diferencial de TBP en los sitios de unión de estas proteínas. Mientras que TBP se colocaliza con BORIS, su presencia en los sitios de unión de CTCF resulta escasa. Estos hallazgos sugieren que BORIS podría desempeñarse como un factor de transcripción pionero, facilitando la apertura de la cromatina y el reclutamiento de otros factores de transcripción, incluyendo a TBP. Esta función de BORIS es coherente con la capacidad de TBP de formar parte del PIC, por lo cual nuestros resultados aportan nuevas evidencias sobre el mecanismo molecular mediante el cual BORIS regula la expresión génica (Pugacheva et al., 2024).

En resumen, los datos presentados sugieren un modelo de interacción entre BORIS y TBP que ofrece una nueva perspectiva sobre los mecanismos moleculares implicados en la regulación transcripcional mediada por BORIS. La identificación de la región de interacción entre BORIS y TBP, junto con su baja identidad con CTCF, plantea una prometedora oportunidad para el desarrollo de terapias dirigidas específicamente a BORIS. Al inhibir su interacción con TBP, se podría modular la actividad transcripcional de genes clave involucrados en procesos patológicos, sin afectar significativamente a la función de CTCF. Sin embargo, se requieren estudios adicionales para validar este modelo y explorar el potencial terapéutico de esta estrategia.

10. Conclusiones

A través del análisis de datos transcriptómicos, se identificaron las lisinas de histonas de la subfamilia KDM4 como blancos terapéuticos prometedores en diversos tipos de cáncer. Estudios *in silico* revelaron que los flavonoides poseen una afinidad preferencial por el sitio catalítico de estas enzimas. Mediante técnicas de acoplamiento molecular, se diseñaron y evaluaron moléculas candidatas a fármacos capaces de inhibir la actividad de las KDM4. La integración de los datos transcriptómicos y de acoplamiento molecular permitió construir una red de interacción fármaco-ligando-proteína, lo cual constituye una herramienta invaluable para la selección de compuestos candidatos para su uso en la medicina personalizada. Además, mediante el empleo de herramientas de modelado molecular, se exploraron las posibilidades de desarrollar inhibidores selectivos de BORIS. Los resultados obtenidos permitieron identificar regiones específicas de la proteína como blancos potenciales para el diseño de fármacos. Además, se validó experimentalmente la interacción de BORIS con TBP, lo cual amplía el panorama para la búsqueda de moléculas inhibitorias que no tengan como blanco a CTCF.

Este estudio ha evaluado el potencial de las proteínas KDM4 y BORIS como nuevas dianas terapéuticas en el tratamiento del cáncer. A través de un enfoque interdisciplinario que combina análisis de datos ómicos y modelado molecular, hemos identificado moléculas con potencial de inhibir selectivamente la actividad de estas enzimas. Nuestros hallazgos no solo amplían nuestro conocimiento sobre los mecanismos moleculares subyacentes a la tumorigénesis, sino que también abren nuevas vías para el desarrollo de terapias eficaces y personalizadas contra una amplia gama de tipos de cáncer. Los resultados obtenidos en este trabajo constituyen un punto de partida sólido para futuras investigaciones encaminadas a la evaluación preclínica y clínica de estos compuestos, así como a la exploración de nuevos mecanismos moleculares regulados por KDM4 y BORIS.

11. Perspectivas

Este trabajo sienta las bases para explorar nuevas estrategias terapéuticas contra el cáncer, centradas en la inhibición de las enzimas KDM4 y BORIS. Las perspectivas futuras de esta investigación se pueden enmarcar en los siguientes ejes:

Desarrollo de nuevas estrategias terapéuticas:

- La identificación de moléculas con afinidad por los sitios catalíticos de KDM4 y los dominios específicos de BORIS abre la puerta al diseño racional de fármacos más potentes y selectivos.
- Los compuestos candidatos identificados en este estudio deberán ser evaluados en modelos preclínicos para determinar su eficacia y seguridad en la inhibición del crecimiento tumoral y la inducción de apoptosis. El uso de repositorios de fármacos ya aprobados por la FDA permite reducir el tiempo de pruebas preclínicas, así como el costo de la investigación necesaria para llevar las terapias a los pacientes.
- La exploración de sinergia entre los diferentes blancos de un fármaco, podría potenciar los efectos terapéuticos de los inhibidores de KDM4 y BORIS, tal como se sugiere en las redes construidas en este trabajo. Además, la combinación de la información genómica del paciente con los datos obtenidos en este estudio permitirá seleccionar los tratamientos más adecuados para cada individuo.

Comprensión de los mecanismos moleculares de la tumorigénesis:

- El estudio de las interacciones moleculares entre TBP, BORIS y otras proteínas involucradas en la regulación de la expresión génica permitirá comprender mejor los mecanismos moleculares subyacentes a la tumorigénesis.
- La generación de modelos con alteraciones en la expresión de KDM4 y BORIS permitirá estudiar el papel de estas enzimas en la iniciación y progresión del cáncer, así como evaluar la eficacia de las terapias dirigidas.
- El análisis de los perfiles de expresión génica en células tratadas con inhibidores de KDM4 y BORIS podría contribuir a comprender los mecanismos mediante los cuales estas proteínas regulan la expresión génica.

En resumen, este trabajo abre nuevas perspectivas de investigación en el campo de la farmacología, con un potencial impacto significativo en el desarrollo de nuevas terapias contra el cáncer.

12. Bibliografía

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25.

Agarwal, S., Sane, R., Oberoi, R., Ohlfest, J. R., and Elmquist, W. F. (2011). Delivery of molecularly targeted therapy to malignant glioma, a disease of the whole brain. *Expert reviews in molecular medicine*, 13:e17.

Akdemir, K. C., Le, V. T., Chandran, S., Li, Y., Verhaak, R. G., Beroukhim, R., Campbell, P. J., Chin, L., Dixon, J. R., Futreal, P. A., et al. (2020). Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nature genetics*, 52(3):294–305.

Alcaraz, N., Hartebrodt, A., and List, M. (2020). De novo pathway enrichment with KeyPathwayMiner. *Methods Mol. Biol.*, 2074:181–199.

Anand, R. and Marmorstein, R. (2007). Structure and mechanism of lysine-specific demethylase enzymes. *Journal of Biological Chemistry*, 282(49):35425–35429.

Anandapadamanaban, M., Andresen, C., Helander, S., Ohyama, Y., Siponen, M. I., Lundström, P., Kokubo, T., Ikura, M., Moche, M., and Sunnerhagen, M. (2013). High-resolution structure of tbp with taf1 reveals anchoring patterns in transcriptional regulation. *Nature structural & molecular biology*, 20(8):1008–1014.

Awuchi, C. G., Aja, M. P., Mitaki, N. B., Morya, S., Amagwula, I. O., Echeta, C. K., and Igwe, V. S. (2023). New psychoactive substances: Major groups, laboratory testing challenges, public health concerns, and community-based solutions. *Journal of Chemistry*, 2023(1):5852315.

Baby, S., Gurukkala Valapil, D., and Shankaraiah, N. (2021). Unravelling KDM4 histone demethylase inhibitors for cancer therapy. *Drug Discov. Today*, 26(8):1841–1856.

Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research*, 21(3):381–395.

Bays, J. L. and DeMali, K. A. (2017). Vinculin in cell–cell and cell–matrix adhesions. *Cellular and Molecular Life Sciences*, 74:2999–3009.

Berdasco, M. and Esteller, M. (2019). Clinical epigenetics: seizing opportunities for translation. *Nature Reviews Genetics*, 20(2):109–127.

Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes & development*, 23(7):781–783.

Berthon, C., Raffoux, E., Thomas, X., Vey, N., Gomez-Roca, C., Yee, K., Taussig, D. C., Rezai, K., Roumier, C., Herait, P., et al. (2016). Bromodomain inhibitor otx015 in patients with acute leukaemia: a dose-escalation, phase 1 study. *The Lancet Haematology*, 3(4):e186–e195.

Bertino, E. M. and Otterson, G. A. (2011). Romidepsin: a novel histone deacetylase inhibitor for cancer. *Expert Opinion on Investigational Drugs*, 20(8):1151–1158. PMID: 21699444.

Besselink, N., Keijer, J., Vermeulen, C., Boymans, S., de Ridder, J., van Hoeck, A., Cuppen, E., and Kuijk, E. (2023). The genome-wide mutational consequences of dna hypomethylation. *Scientific Reports*, 13(1):6874.

Biersack, B., Nitzsche, B., and Höpfner, M. (2022a). Hdac inhibitors with potential to overcome drug resistance in castration-resistant prostate cancer. *Cancer Drug Resistance*, 5(1):64.

Biersack, B., Nitzsche, B., and Höpfner, M. (2022b). Immunomodulatory properties of hdac6 inhibitors in cancer diseases: New chances for sophisticated drug design and treatment optimization. *Seminars in cell & developmental biology*.

Blanc, R. S. and Richard, S. (2017). Arginine methylation: The coming of age. *Mol. Cell*, 65(1):8–24.

Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using alphafold2. *Nature communications*, 13(1):1265.

Buettner, R., Corzano, R., Rashid, R., Lin, J., Senthil, M., Hedvat, M., Schroeder, A., Mao, A., Herrmann, A., Yim, J., et al. (2011). Alkylation of cysteine 468 in stat3 defines a novel site for therapeutic development. *ACS chemical biology*, 6(5):432–443.

Bushweller, J. H. (2019). Targeting transcription factors in cancer—from undruggable to reality. *Nature Reviews Cancer*, 19(11):611–624.

Capecchi, A. and Reymond, J.-L. (2021). Classifying natural products from plants, fungi or bacteria using the COCONUT database and machine learning. *J. Cheminform.*, 13(1):82.

Chakrabarti, P. and Chakravarty, D. (2022). Intrinsically disordered proteins/regions and insight into their biomolecular interactions. *Biophysical Chemistry*, 283:106769.

Cheever, M. A., Allison, J. P., Ferris, A. S., Finn, O. J., Hastings, B. M., Hecht, T. T., Mellman, I., Prindiville, S. A., Viner, J. L., Weiner, L. M., et al. (2009). The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. *Clinical cancer research*, 15(17):5323–5337.

Chen, G.-Q., Ye, P., Ling, R.-S., Zeng, F., Zhu, X.-S., Chen, L., Huang, Y., Xu, L., and Xie, X.-Y. (2020). Histone demethylase KDM4C is required for ovarian cancer stem cell maintenance. *Stem Cells Int.*, 2020:8860185.

Cheng, F., Kovács, I. A., and Barabási, A.-L. (2019). Network-based prediction of drug combinations. *Nat. Commun.*, 10(1):1197.

Chernukhin, I. V., Shamsuddin, S., Robinson, A. F., Carne, A. F., Paul, A., El-Kady, A. I., Lobanenkov, V. V., and Klenova, E. M. (2000). Physical and functional interaction between two pluripotent proteins, the y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J. Biol. Chem.*, 275(38).

Chin, Y.-W. and Han, S.-Y. (2015). KDM4 histone demethylase inhibitors for anti-cancer agents: a patent review. *Expert Opin. Ther. Pat.*, 25(2):135–144.

Choi, K.-C., Jung, M. G., Lee, Y.-H., Yoon, J. C., Kwon, S. H., Kang, H.-B., Kim, M.-J., Cha, J.-H., Kim, Y. J., Jun, W. J., Lee, J. M., and Yoon, H.-G. (2009). Epigallocatechin-3-gallate, a histone acetyltransferase inhibitor, inhibits EBV-induced B lymphocyte transformation via suppression of RelA acetylation. *Cancer Res.*, 69(2):583–592.

Chung, C. (2016a). Crystal structure of human KDM4D in complex with 2-1h-pyrazol-4-yloxy- 3H,4H-pyrido-3,4-d-pyrimidin-4-one.

Chung, C. (2016b). Crystal structure of human KDM4D in complex with 3-(4-phenylbutanamido)pyridine-4-carboxylic acid.

Chung, C. (2016c). Crystal structure of human KDM4D in complex with 3-4-methylthiophen-2- yl methylaminopyridine-4-carboxylic acid.

Chung, C. (2016d). Crystal structure of human KDM4D in complex with 3-4-methylthiophen-2- ylmethylaminopyridine-4-carboxylic acid.

Chung, C. (2016e). Crystal structure of human KDM4D in complex with 3H,4H-pyrido-3,4-d- pyrimidin-4-one.

Churchill, C. D. M. and Wetmore, S. D. (2009). Noncovalent interactions involving histidine: the effect of charge on pi-pi stacking and t-shaped interactions with the DNA nucleobases. *J. Phys. Chem. B*, 113(49):16046–16058.

Cipriano, A., Milite, C., Feoli, A., Viviano, M., Pepe, G., Campiglia, P., Sarno, G., Picaud, S., Imaide, S., Makukhin, N., et al. (2022). Discovery of benzo [d] imidazole-6-sulfonamides as bromodomain and extra-terminal domain (bet) inhibitors with selectivity for the first bromodomain. *ChemMedChem*, 17(20):e202200343.

Colovos, C. and Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, 2:1511–1519.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. (2022). Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*.

Cosenza, M., Civallero, M., Marcheselli, L., Sacchi, S., and Pozzi, S. (2017). Ricolinostat, a selective hdac6 inhibitor, shows anti-lymphoma cell activity alone and in combination with bendamustine. *Apoptosis*, 22:827–840.

Cuyàs, E., Gumuzio, J., Verdura, S., Brunet, J., Bosch-Barrera, J., Martin-Castillo, B., Alarcón, T., Encinar, J. A., Martin, Á. G., and Menendez, J. A. (2020). The lsd1 inhibitor iadademstat (ory-1001) targets sox2-driven breast cancer stem cells: a potential epigenetic therapy in luminal-b and her2-positive breast cancer subtypes. *Aging (Albany NY)*, 12(6):4794.

Dalal, K., Che, M., Que, N. S., Sharma, A., Yang, R., Lallous, N., Borgmann, H., Ozistanbullu, D., Tse, R., Ban, F., et al. (2017). Bypassing drug resistance mechanisms of prostate cancer with small molecules that target androgen receptor–chromatin interactions. *Molecular cancer therapeutics*, 16(10):2281–2291.

D'Arcy, V., Pore, N., Docquier, F., Abdullaev, Z. K., Chernukhin, I., Kita, G. X., Rai, S., Smart, M., Farrar, D., Pack, S., Lobanenkov, V., and Klenova, E. (2008). BORIS, a paralogue of the transcription factor, CTCF, is aberrantly expressed in breast tumours. *Br. J. Cancer*, 98(3):571–579.

Dawson, M. A. and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell*, 150(1):12–27.

de Necochea-Campion, R., Ghochikyan, A., Josephs, S. F., Zacharias, S., Woods, E., Karimi-Busheri, F., Alexandrescu, D. T., Chen, C. S., Agadjanyan, M. G., and Carrier, E. (2011). Expression of the epigenetic factor boris (ctcfl) in the human genome. *Journal of Translational Medicine*, 9:213.

Debaugny, R. E. and Skok, J. A. (2020). CTCF and CTCFL in cancer. *Curr. Opin. Genet. Dev.*, 61:44–52.

Defossez, P. A., Kelly, K. F., Filion, G. J., 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(52).

del Moral-Morales, A., Salgado-Albarrán, M., Ortiz-Gutiérrez, E., Pérez-Hernández, G., and Soto-Reyes, E. (2022). Transcriptomic and drug discovery analyses reveal natural compounds targeting the kdm4 subfamily as promising adjuvant treatments in cancer. *Frontiers in Genetics*, 13:860924.

Del Moral-Morales, A., Salgado-Albarrán, M., Sánchez-Pérez, Y., Wenke, N. K., Baumbach, J., and Soto-Reyes, E. (2023). CTCF and its Multi-Partner network for chromatin regulation. *Cells*, 12(10).

DeLano, W. L. et al. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*, 40(1):82–92.

Deng, Y., Li, M., Zhuo, M., Guo, P., Chen, Q., Mo, P., Li, W., and Yu, C. (2021). Histone demethylase jmjd2d promotes the self-renewal of liver cancer stem-like cells by enhancing epcam and sox9 expression. *Journal of Biological Chemistry*, 296.

Dougherty, C. J., Ichim, T. E., Liu, L., Reznik, G., Min, W. P., Ghochikyan, A., Agadjanyan, M. G., and Reznik, B. N. (2008). Selective apoptosis of breast cancer cells by siRNA targeting of BORIS. *Biochem. Biophys. Res. Commun.*, 370(1):109–112.

Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021). Autodock vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of chemical information and modeling*, 61(8):3891–3898.

Ehrlich, M. (2002). Dna methylation in cancer: too much, but also too little. *Oncogene*, 21(35):5400–5413.

Esteller, M. (2011). Epigenetic changes in cancer. *F1000 biology reports*, 3(9).

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, 15.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. (2021). Protein complex prediction with alphafold-multimer. *bioRxiv*.

Fang, M. Z., Wang, Y., Ai, N., Hou, Z., Sun, Y., Lu, H., Welsh, W., and Yang, C. S. (2003). Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. *Cancer Res.*, 63(22):7563–7570.

Fang, Y., Yang, C., Yu, Z., Li, X., Mu, Q., Liao, G., and Yu, B. (2020). Natural products as LSD1 inhibitors for cancer therapy. *Acta Pharm Sin B*.

Felsenfeld, G. and Groudine, M. (2003). Controlling the double helix. *Nature*, 421(6921):448–453.

Fiorito, E., Sharma, Y., Gilfillan, S., Wang, S., Singh, S. K., Satheesh, S. V., Katika, M. R., Urbanucci, A., Thiede, B., Mills, I. G., and Hurtado, A. (2016). CTCF modulates estrogen receptor function through specific chromatin and nuclear matrix interactions. *Nucleic Acids Res.*, 44(22):10588–10602.

Ganai, S. A. (2016). Histone deacetylase inhibitor givinostat: the small-molecule with promising activity against therapeutically challenging haematological malignancies. *Journal of Chemotherapy*, 28(4):247–254.

García, M. A., R., F., and Martínez-Balbás, M. A. (2016). Lysine demethylases: Structure, function, and disfunction. In Binda, O. and Fernandez-Zapico, M. E., editors, *Chromatin Signaling and Diseases*, pages 179–194. Academic Press, Boston.

Gaykalova, D., Vatapalli, R., Glazer, C. A., Bhan, S., Shao, C., Sidransky, D., Ha, P. K., and Califano, J. A. (2012). Dose-Dependent Activation of Putative Oncogene SBSN by BORIS. *PLoS One*, 7(7):e40389.

Georgina S. Daher-Reyes, B. M. M. and Yee, K. W. L. (2019). Guadecitabine (sgi-110): an investigational drug for the treatment of myelodysplastic syndrome and acute myeloid leukemia. *Expert Opinion on Investigational Drugs*, 28(10):835–849.

Ghochikyan, A., Mkrtichyan, M., Loukinov, D., Mamikonyan, G., Pack, S. D., Movsesyan, N., Ichim, T. E., Cribbs, D. H., Lobanenkov, V. V., and Agadjanyan, M. G. (2007). Elicitation of T Cell Responses to Histologically Unrelated Tumors by Immunization with the Novel Cancer-Testis Antigen, Brother of the Regulator of Imprinted Sites. *J. Immunol.*, 178(1):566–573.

Godwin, R. C., Melvin, R. L., Gmeiner, W. H., and Salsbury, F. R. (2017). Binding site configurations probe the structure and dynamics of the zinc finger of nemo (nf-κb essential modulator). *Biochemistry*, 56:623–633.

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A. N., et al. (2020). Visualizing and interpreting cancer genomics data via the xena platform. *Nature biotechnology*, 38(6):675–678.

Gómez-Cansino, R., Guzmán-Gutiérrez, S. L., Campos-Lara, M. G., Espitia-Pinzón, C. I., and Reyes-Chilpa, R. (2017). Natural compounds from mexican medicinal plants as potential drug leads for Anti-Tuberculosis drugs. *An. Acad. Bras. Cienc.*, 89(1):31–43.

Gong, M., Yan, C., Jiang, Y., Meng, H., Feng, M., and Cheng, W. (2019). Genome-wide bioinformatics analysis reveals CTCFL is upregulated in high-grade epithelial ovarian cancer. *Oncol. Lett.*, 18(4):4030–4039.

Gorgulla, C., Çınaroğlu, S. S., Fischer, P. D., Fackeldey, K., Wagner, G., and Arthanari, H. (2021). Virtualflow ants—ultra-large virtual screenings with artificial intelligence driven docking algorithm based on ant colony optimization. *International Journal of Molecular Sciences*, 22(11):5807.

Gormally, M. V., Dexheimer, T. S., Marsico, G., Sanders, D. A., Lowe, C., Matak-Vinković, D., Michael, S., Jadhav, A., Rai, G., Maloney, D. J., et al. (2014). Suppression of the foxm1 transcriptional programme via novel small molecule inhibition. *Nature communications*, 5(1):5165.

Grant, B. J., Skjærven, L., and Yao, X. Q. (2021). The bio3d packages for structural bioinformatics. *Protein Science*, 30:20–30.

Grimley, E., Liao, C., Ranghini, E. J., Nikolovska-Coleska, Z., and Dressler, G. R. (2017). Inhibition of pax2 transcription activation with a small molecule that targets the dna binding domain. *ACS chemical biology*, 12(3):724–734.

Guerra-Calderas, L., González-Barrios, R., Herrera, L. A., Cantú de León, D., and Soto-Reyes, E. (2015). The role of the histone demethylase KDM4A in cancer. *Cancer Genet.*, 208(5):215–224.

Guillade, L., Sarno, F., Tarhonskaya, H., Nebbioso, A., Alvarez, S., Kawamura, A., Schofield, C. J., Altucci, L., and de Lera, Á. R. (2018). Synthesis and biological evaluation of tripartin, a putative KDM4 natural product inhibitor, and 1-dichloromethylinden-1-ol analogues. *ChemMedChem*, 13(18):1949–1956.

Gutiérrez-Rebolledo, G. A., Drier-Jonas, S., and Jiménez-Arellanes, M. A. (2017). Natural compounds and extracts from mexican medicinal plants with anti-leishmaniasis activity: An update. *Asian Pac. J. Trop. Med.*, 10(12):1105–1110.

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1):31–46.

Hashimoto, H., Wang, D., Horton, J. R., Zhang, X., Corces, V. G., and Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol. Cell*, 66(5):711–720.e3.

He, X., Sun, T., Zhang, P., Xia, Z., Gao, C., Ren, H., and Ji, D. (2022). Selective inhibition of histone deacetylase class iia with mc1568 ameliorates podocyte injury. *Frontiers in Medicine*, 9:848938.

Hillman, J. C., Pugacheva, E. M., Barger, C. J., Sribenja, S., Rosario, S., Albahrani, M., Truskinovsky, A. M., Stablewski, A., Liu, S., Loukinov, D. I., Zentner, G. E., Lobanenkov, V. V., Karpf, A. R., and Higgins, M. J. (2019). BORIS expression in ovarian cancer precursor cells alters the CTCF cistrome and enhances invasiveness through GALNT14. *Mol. Cancer Res.*, 17(10):2051–2062.

Hillringhaus, L., Yue, W. W., Rose, N. R., Ng, S. S., Gileadi, C., Loenarz, C., Bello, S. H., Bray, J. E., Schofield, C. J., and Oppermann, U. (2011). Structural and evolutionary basis for the dual substrate selectivity of human kdm4 histone demethylase family. *Journal of Biological Chemistry*, 286(48):41616–41625.

Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., Goldmann, J., Lajoie, B. R., Fan, Z. P., Sigova, A. A., Reddy, J., Borges-Rivera, D., Lee, T. I., Jaenisch, R., Porteus, M. H., Dekker, J., and Young, R. A. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351(6280):1454–1458.

Hong, J. A., Kang, Y., Abdullaev, Z., Flanagan, P. T., Pack, S. D., Fischette, M. R., Adnani, M. T., Loukinov, D. I., Vatolin, S., Risinger, J. I., Custer, M., Chen, G. A., Zhao, M., Nguyen, D. M., Barrett, J. C., Lobanenkov, V. V., and Schrump, D. S. (2005). Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. *Cancer Res.*, 65(17):7763–7774.

Hore, T. A., Deakin, J. E., and Marshall Graves, J. A. (2008). The Evolution of Epigenetic Regulators CTCF and BORIS/CTCFL in Amniotes. *PLoS Genet.*, 4(8):e1000169.

Houser, J., Kozmon, S., Mishra, D., Hammerová, Z., Wimmerová, M., and Koča, J. (2020). The CH–π interaction in Protein–Carbohydrate binding: Bioinformatics and in vitro quantification.

Hoy, S. M. (2020). Tazemetostat: first approval. *Drugs*, 80(5):513–521.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., Groot, B. L. D., Grubmüller, H., and MacKerell, A. D. (2016). Charmm36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14:71–73.

Hummel-Eisenbeiss, J., Hascher, A., Hals, P.-A., Sandvold, M. L., Müller-Tidow, C., Lyko, F., and Rius, M. (2013). The role of human equilibrative nucleoside transporter 1 on the cellular transport of the dna methyltransferase inhibitors 5-azacytidine and cp-4200 in human leukemia cells. *Molecular Pharmacology*, 84(3):438–450.

Iwamori, N., Zhao, M., Meistrich, M. L., and Matzuk, M. M. (2011). The testis-enriched histone demethylase, kdm4d, regulates methylation of histone h3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility. *Biology of reproduction*, 84(6):1225–1234.

Jakalian, A., Jack, D. B., and Bayly, C. I. (2002). Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation. *Journal of computational chemistry*, 23(16):1623–1641.

James, J., Mair, S., Doll, W., Sandefer, E., Wurtman, D., Maurer, A., Deane, A. M., and Harris, M. S. (2020). The effects of ulimorelin, a ghrelin agonist, on liquid gastric emptying and colonic transit in humans. *Neurogastroenterology & Motility*, 32(3):e13784.

Janssen, S. M., Moscona, R., Elchebly, M., Papadakis, A. I., Redpath, M., Wang, H., Rubin, E., van Kempen, L. C., and Spatz, A. (2020). BORIS/CTCFL promotes a switch from a proliferative towards an invasive phenotype in melanoma cells. *Cell Death Discov.*, 6(1):1.

Jelinic, P., Stehle, J.-C., and Shaw, P. (2006). The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. *PLoS Biol.*, 4(11):e355.

Jimeno, A., Weiss, G. J., Miller Jr, W. H., Gettinger, S., Eigl, B. J., Chang, A. L. S., Dunbar, J., Devens, S., Faia, K., Skliris, G., et al. (2013). Phase i study of the hedgehog pathway inhibitor ipi-926 in adult patients with solid tumors. *Clinical cancer research*, 19(10):2766–2774.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589.

Kaczmarczyk, L. S., Levi, N., Segal, T., Salmon-Divon, M., and Gerlitz, G. (2022). CTCF supports preferentially short lamina-associated domains. *Chromosome Res.*, 30(1):123–136.

Kagan, A. B., Garrison, D. A., Anders, N. M., Webster, J., Baker, S. D., Yegnasubramanian, S., and Rudek, M. A. (2023). Dna methyltransferase inhibitor exposure–response: Challenges and opportunities. *Clinical and Translational Science*, 16(8):1309–1322.

Kang, Y., Hong, J. A., Chen, G. A., Nguyen, D. M., and Schrump, D. S. (2007). Dynamic transcriptional regulatory complexes including BORIS, CTCF and sp1 modulate NY-ESO-1 expression in lung cancer cells. *Oncogene*, 26(30):4394–4403.

Katoh, M. and Katoh, M. (2004). Identification and characterization of jmjd2 family genes in silico. *International journal of oncology*, 24(6):1623–1628.

Kawazu, M., Saso, K., Tong, K. I., McQuire, T., Goto, K., Son, D.-O., Wakeham, A., Miyagishi, M., Mak, T. W., and Okada, H. (2011). Histone demethylase jmjd2b functions as a co-factor of estrogen receptor in breast cancer proliferation and mammary gland development. *PloS one*, 6(3):e17830.

Kentepozidou, E., Aitken, S. J., Feig, C., Stefflova, K., Ibarra-Soria, X., Odom, D. T., Roller, M., and Flicek, P. (2020). Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.*, 21(1):1–19.

Kholmanskikh, O., Loriot, A., Brasseur, F., De Plaen, E., and De Smet, C. (2008). Expression of BORIS in melanoma: Lack of association with MAGE-A1 activation. *Int. J. Cancer*, 122(4):777–784.

Klenova, E. M., Morse, H. C., Ohlsson, R., and Lobanenkov, V. V. (2002). The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Semin. Cancer Biol.*, 12(5):399–414.

Klose, R. J., Kallin, E. M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.*, 7(9):715–727.

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2–an r package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler. *F1000Research*, 9.

Kotlyar, M., Pastrello, C., Malik, Z., and Jurisica, I. (2019). IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.*, 47(D1):D581–D589.

Krishnan, S. and Trievel, R. C. (2012a). Crystal structure of human JMJD2D/KDM4D apoenzyme.

Krishnan, S. and Trievel, R. C. (2012b). Crystal structure of human JMJD2D/KDM4D in complex with an H3K9me3 peptide and 2-oxoglutarate.

Krojer, T., Vollmar, M., Crawley, L., Bradley, A. R., Szykowska, A., Ruda, G. F., Yang, H., Burgess-Brown, N., Brennan, P., Bountra, C., Arrowsmith, C. H., Edwards, A., Oppermann, U., von Delft, F., and Structural Genomics Consortium (SGC) (2015). Crystal structure of human JMJD2D complexed with KDOPP7.

Kuck, D., Caulfield, T., Lyko, F., and Medina-Franco, J. L. (2010). Nanaomycin a selectively inhibits dnmt3b and reactivates silenced tumor suppressor genes in human cancer cells. *Molecular cancer therapeutics*, 9(11):3015–3023.

Kumari, R., Kumar, R., Open Source Drug Discovery Consortium, and Lynn, A. (2014). g_mmpbsa–a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.*, 54(7):1951–1962.

Labbé, R. M., Holowatyj, A., and Yang, Z.-Q. (2013). Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential. *Am. J. Transl. Res.*, 6(1):1–15.

Laskowski, R. A., Macarthur, M. W., and Thornton, J. M. (2012). Chapter 21.4. procheck: validation of protein-structure coordinates. *nternational Tables for Crystallography*, F:684–687.

Laubach, J. P., Moreau, P., San-Miguel, J. F., and Richardson, P. G. (2015). Panobinostat for the treatment of multiple myeloma. *Clinical cancer research*, 21(21):4767–4773.

Le Bihan, Y.-V., Dempster, S., Westwood, I. M., and van Montfort, R. L. M. (2016). Crystal structure of human KDM4A in complex with compound 40.

Lee, D. H., Kim, G. W., Jeon, Y. H., Yoo, J., Lee, S. W., and Kwon, S. H. (2020). Advances in histone demethylase kdm4 as cancer therapeutic targets. *The FASEB Journal*, 34(3):3461–3484.

Lee, D. H., Kim, G. W., Yoo, J., Lee, S. W., Jeon, Y. H., Kim, S. Y., Kang, H. G., Kim, D.-H., Chun, K.-H., Choi, J., et al. (2021). Histone demethylase kdm4c controls tumorigenesis of glioblastoma by epigenetically regulating p53 and c-myc. *Cell Death & Disease*, 12(1):89.

Lee, H.-Z., Kwitkowski, V. E., Del Valle, P. L., Ricci, M. S., Saber, H., Habtemariam, B. A., Bullock, J., Bloomquist, E., Li Shen, Y., Chen, X.-H., et al. (2015). Fda approval: belinostat for the treatment of patients with relapsed or refractory peripheral t-cell lymphoma. *Clinical Cancer Research*, 21(12):2666–2670.

Lee, J. E. and Kim, M.-Y. (2022). Cancer epigenetics: Past, present and future. *Semin. Cancer Biol.*, 83:4–14.

Leurs, U., Lohse, B., Rand, K. D., Ming, S., Riise, E. S., Cole, P. A., Kristensen, J. L., and Clausen, R. P. (2014). Substrate-and cofactor-independent inhibition of histone demethylase kdm4c. *ACS chemical biology*, 9(9):2131–2138.

Li, B. and Dewey, C. N. (2011). Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC bioinformatics*, 12:1–16.

Li, H., Ban, F., Dalal, K., Leblanc, E., Frewin, K., Ma, D., Adomat, H., Rennie, P. S., and Cherkasov, A. (2014). Discovery of small-molecule inhibitors selectively targeting the dna-binding domain of the human androgen receptor. *Journal of medicinal chemistry*, 57(15):6458–6467.

Li, J., Fu, A., and Zhang, L. (2019a). An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdisciplinary Sciences: Computational Life Sciences*, 11:320–328.

Li, M., Deng, Y., Zhuo, M., Zhou, H., Kong, X., Xia, X., Su, Z., Chen, Q., Guo, P., Mo, P., et al. (2020a). Demethylase-independent function of jmjd2d as a novel antagonist of p53 to promote liver cancer initiation and progression. *Theranostics*, 10(19):8863.

Li, S., Olson, W. K., and Lu, X. J. (2019b). Web 3dna 2.0 for the analysis, visualization, and modeling of 3d nucleic acid structures. *Nucleic Acids Research*, 47:W26–W34.

Li, W., Zhang, J., Wang, J., and Wang, W. (2008). Metal-coupled folding of cys2his2 zinc-finger. *Journal of the American Chemical Society*, 130:892–900.

Li, Y., Haarhuis, J. H., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M. S., Willems, L., Teunissen, H., Muir, K. W., de Wit, E., Rowland, B. D., and Panne, D. (2020b). The structural basis for cohesin–CTCF-anchored loops. *Nature*, 578(7795):472–476.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6):417–425.

Link, P. A., Zhang, W., Odunsi, K., and Karpf, A. R. (2013). BORIS/CTCFL mRNA isoform expression and epigenetic regulation in epithelial ovarian cancer. *Cancer Immun.*, 13(1).

Liu, F., Wu, D., and Wang, X. (2019). Roles of CTCF in conformation and functions of chromosome. *Seminars in cell & developmental biology*, 90:168–173.

Liu, J., Ali, M., and Zhou, Q. (2020a). Establishment and evolution of heterochromatin. *Annals of the New York Academy of Sciences*.

Liu, X., Wang, Y., Gao, Y., Su, J., Zhang, J., Xing, X., Zhou, C., Yao, K., An, Q., and Zhang, Y. (2018). H3k9 demethylase kdm4e is an epigenetic regulator for bovine embryonic development and a defective factor for nuclear reprogramming. *Development*, 145(4):dev158261.

Liu, X., Zhang, Q., Zhao, Y., Xun, J., Wu, H., and Feng, H. (2020b). Association of jmjd2b and hypoxia-inducible factor 1 expressions with poor prognosis in osteosarcoma. *Analytical Cellular Pathology*, 2020(1):2563208.

Loukinov, D. I., Pugacheva, E., Vatolin, S., Pack, S. D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A. A., Cui, H., Niemitz, E. L., Rasko, J. E., Docquier, F. M., Kistler, M., Breen, J. J., Zhuang, Z., Quitschke, W. W., Renkawitz, R., Klenova, E. M., Feinberg, A. P., Ohlsson, R., Morse, H. C., and Lobanenkov, V. V. (2002). 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. U. S. A.*, 99(10):6806–6811.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15:1–21.

Lu, C.-Y., Chang, Y.-C., Hua, C.-H., Chuang, C., Huang, S.-H., Kung, S.-H., Hour, M.-J., and Lin, C.-W. (2017). Tubacin, an hdac6 selective inhibitor, reduces the replication of the japanese encephalitis virus via the decrease of viral rna synthesis. *International Journal of Molecular Sciences*, 18(5):954.

Lv, M. and Liu, Q. (2021). Jmjd2c triggers the growth of multiple myeloma cells via activation of β-catenin. *Oncology Reports*, 45(3):1162–1170.

Mabe, N. W., Garcia, N. M. G., Wolery, S. E., Newcomb, R., Meingasner, R. C., Vilona, B. A., Lupo, R., Lin, C.-C., Chi, J.-T., and Alvarez, J. V. (2020). G9a promotes breast cancer recurrence through repression of a pro-inflammatory program. *Cell reports*, 33(5).

Maitra, S., Khandelwal, N., Kootar, S., Sant, P., Pathak, S. S., Reddy, S., K, A. P., Murty, U. S., Chakravarty, S., and Kumar, A. (2020). Histone lysine demethylase jmjd2d/kdm4d and family members mediate effects of chronic social defeat stress on mouse hippocampal neurogenesis and mood disorders. *Brain Sciences*, 10(11):833.

Majello, B., Napolitano, G., De Luca, P., and Lania, L. (1998). Recruitment of human tbp selectively activates rna polymerase ii tata-dependent promoters. *Journal of Biological Chemistry*, 273(26):16509–16516.

Mal, T. K., Takahata, S., Ki, S., Zheng, L., Kokubo, T., and Ikura, M. (2007). Functional silencing of tata-binding protein (tbp) by a covalent linkage of the n-terminal domain of tbp-associated factor 1. *Journal of Biological Chemistry*, 282(30):22228–22238.

Małecki, P. H., Rüger, N., Roatsch, M., Krylova, O., Link, A., Jung, M., Heinemann, U., and Weiss, M. S. (2019). Structure-Based screening of tetrazolylhydrazide inhibitors versus KDM4 histone demethylases. *ChemMedChem*, 14(21):1828–1839.

Malecki, P. H., Weiss, M. S., Heinemann, U., and Link, A. (2020). Crystal structure of KDM4D with tetrazolylhydrazide ligand NR073.

Marakulina, D., Vorontsov, I. E., Kulakovskiy, I. V., Lennartsson, A., Drabløs, F., and Medvedeva, Y. (2022). EpiFactors 2022: expansion and enhancement of a curated database of human epigenetic factors and complexes. *Nucleic Acids Research*, 51(D1):D564–D570.

Markolovic, S., Leissing, T. M., Chowdhury, R., Wilkins, S. E., Lu, X., and Schofield, C. J. (2016). Structure-function relationships of human JmjC oxygenases-demethylases versus hydroxylases. *Curr. Opin. Struct. Biol.*, 41:62–72.

Marshall, A. D., Bailey, C. G., and Rasko, J. E. (2014). CTCF and BORIS in genome regulation and cancer.

Masoudi-Sobhanzadeh, Y., Omidi, Y., Amanlou, M., and Masoudi-Nejad, A. (2020). Drug databases and their contributions to drug repurposing. *Genomics*, 112(2):1087–1095.

Massett, M. E., Monaghan, L., Patterson, S., Mannion, N., Bunschoten, R. P., Hoose, A., Marmiroli, S., Liskamp, R. M., Jørgensen, H. G., Vetrie, D., et al. (2021). A kdm4a-paf1-mediated epigenomic network is essential for acute myeloid leukemia cell self-renewal and survival. *Cell death & disease*, 12(6):573.

McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J., and Koes, D. R. (2021). Gnina 1.0: molecular docking with deep learning. *Journal of cheminformatics*, 13(1):43.

Medvedeva, Y. A., Lennartsson, A., Ehsani, R., Kulakovskiy, I. V., Vorontsov, I. E., Panahandeh, P., Khimulya, G., Kasukawa, T., Consortium, T. F., and Drabløs, F. (2015). EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database*, 2015:bav067.

Meng, Y., Li, H., Liu, C., Zheng, L., and Shen, B. (2018). Jumonji domain-containing protein family: the functions beyond lysine demethylation. *Journal of Molecular Cell Biology*, 10(4):371–373.

Menghrajani, K., Cai, S. F., Devlin, S. M., Armstrong, S. A., Piekarz, R., Rudek, M., Tallman, M. S., and Stein, E. M. (2019). A phase ib/ii study of the histone methyltransferase inhibitor pinometostat in combination with azacitidine in patients with 11q23-rearranged acute myeloid leukemia. *Blood*, 134:2655.

Meraz-Torres, F., Plöger, S., Garbe, C., Niessner, H., and Sinnberg, T. (2020). Disulfiram as a therapeutic agent for metastatic malignant melanoma—old myth or new logos? *Cancers*, 12(12):3538.

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 research*, 77(21):5900–5912.

Millán-Zambrano, G., Burton, A., Bannister, A. J., and Schneider, R. (2022). Histone post-translational modifications - cause and consequence of genome function. *Nat. Rev. Genet.*, 23(9):563–580.

Miller, A. L., Fehling, S. C., Garcia, P. L., Gamblin, T. L., Council, L. N., van Waardenburg, R. C., Yang, E. S., Bradner, J. E., and Yoon, K. J. (2019). The bet inhibitor jq1 attenuates double-strand break repair and sensitizes models of pancreatic ductal adenocarcinoma to parp inhibitors. *EBioMedicine*, 44:419–430.

Mirabello, C. and Wallner, B. (2024). Dockq v2: Improved automatic quality measure for protein multimers, nucleic acids, and small molecules. *bioRxiv*.

Mkrtichyan, M., Ghochikyan, A., Davtyan, H., Movsesyan, N., Loukinov, D., Lobanenkov, V., Cribbs, D. H., Laust, A. K., Nelson, E. L., and Agadjanyan, M. G. (2011). Cancer-testis

antigen, BORIS based vaccine delivered by dendritic cells is extremely effective against a very aggressive and highly metastatic mouse mammary carcinoma. *Cell. Immunol.*, 270(2):188–197.

Mkrtichyan, M., Ghochikyan, A., Loukinov, D., Davtyan, H., Ichim, T. E., Cribbs, D. H., Lobanenkov, V. V., and Agadjanyan, M. G. (2008). DNA, but not protein vaccine based on mutated BORIS antigen significantly inhibits tumor growth and prolongs the survival of mice. *Gene Ther.*, 15(1):61–64.

Mogal, A. and Abdulkadir, S. A. (2006). Effects of histone deacetylase inhibitor (hdaci); trichostatin-a (tsa) on the expression of housekeeping genes. *Molecular and cellular probes*, 20(2):81–86.

Monk, M., Hitchins, M., and Hawes, S. (2008). Differential expression of the embryo/cancer gene ECSA(DPPA2), the cancer/testis gene BORIS and the pluripotency structural gene OCT4, in human preimplantation development. *Mol. Hum. Reprod.*, 14(6):347–355.

Mu, H., Xiang, L., Li, S., Rao, D., Wang, S., and Yu, K. (2019). Mir-10a functions as a tumor suppressor in prostate cancer via targeting kdm4a. *Journal of Cellular Biochemistry*, 120(4):4987–4997.

NCBI gene, N. C. f. B. I. (2025). Ctcfl ccctc-binding factor like [homo sapiens (human)] https://www.ncbi.nlm.nih.gov/gene/140690.

Nguyen, L. H., Tran, T. T., Truong, L. T. N., Mai, H. H., and Nguyen, T. T. (2020). Overcharging of the zinc ion in the structure of the zinc-finger protein is needed for dna binding stability. *Biochemistry*, 59:1378–1390.

Nguyen, P., Cui, H., Bisht, K. S., Sun, L., Patel, K., Lee, R. S., Kugoh, H., Oshimura, M., Feinberg, A. P., and Gius, D. (2008). CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region. *Cancer Res.*, 68(14):5546–5551.

Nishana, M., Ha, C., Rodriguez-Hernaez, J., Ranjbaran, A., Chio, E., Nora, E. P., Badri, S. B., Kloetgen, A., Bruneau, B. G., Tsirigos, A., and Skok, J. A. (2020). Defining the relative and combined contribution of CTCF and CTCFL to genomic regulation. *Genome Biol.*, 21(1).

Noberini, R., Osti, D., Miccolo, C., Richichi, C., Lupia, M., Corleone, G., Hong, S.-P., Colombo, P., Pollo, B., Fornasari, L., et al. (2018). Extensive and systematic rewiring of histone post-translational modifications in cancer model systems. *Nucleic acids research*, 46(8):3817–3832.

Noble, D. (2015). Conrad Waddington and the origin of epigenetics. *Journal of Experimental Biology*, 218(6):816–818.

Ong, C.-T. and Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.*, 15(4):234–246.

OpenEye Scientific Software (2021). Cheminformatics software. https://www.eyesopen.com. Accessed: 2021-12-8.

Palmirotta, R., Silvestris, E., D'Oronzo, S., Cardascia, A., and Silvestris, F. (2017). Ovarian cancer: Novel molecular aspects for clinical assessment. *Crit. Rev. Oncol. Hematol.*, 117:12–29.

Patel, A. B., Louder, R. K., Greber, B. J., Grünberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., and Nogales, E. (2018). Structure of human tfiid and mechanism of tbp loading onto promoter dna. *Science*, 362(6421):eaau8872.

Piha-Paul, S. A., Sachdev, J. C., Barve, M., LoRusso, P., Szmulewitz, R., Patel, S. P., Lara Jr, P. N., Chen, X., Hu, B., Freise, K. J., et al. (2019). First-in-human study of mivebresib (abbv-075), an oral pan-inhibitor of bromodomain and extra terminal proteins, in patients with relapsed/refractory solid tumors. *Clinical Cancer Research*, 25(21):6309–6319.

Pontius, J., Richelle, J., and Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. *Journal of Molecular Biology*, 264:121–136.

Premkumar, K. and Shankar, B. S. (2022). Identification of epz004777 and fg2216 as inhibitors of tgf-β1 induced treg cells by screening a library of epigenetic compounds. *Life Sciences*, 301:120643.

Pugacheva, E. M., Bhatt, D. N., Rivero-Hinojosa, S., Tajmul, M., Fedida, L., Price, E., Ji, Y., Loukinov, D., Strunnikov, A. V., Ren, B., et al. (2024). Boris/ctcfl epigenetically reprograms clustered ctcf binding sites into alternative transcriptional start sites. *Genome Biology*, 25(1):40.

Pugacheva, E. M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A. L., Strunnikov, A. V., Zentner, G. E., Ren, B., and Lobanenkov, V. V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci. U. S. A.*, 117(4):2020–2031.

Pugacheva, E. M., Rivero-Hinojosa, S., Espinoza, C. A., Méndez-Catalá, C. F., Kang, S., Suzuki, T., Kosaka-Suzuki, N., Robinson, S., Nagarajan, V., Ye, Z., Boukaba, A., Rasko, J. E. J., Strunnikov, A. V., Loukinov, D., Ren, B., and Lobanenkov, V. V. (2015). Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. *Genome Biol.*, 16(1):161.

Pugacheva, E. M., Suzuki, T., Pack, S. D., Kosaka-Suzuki, N., Yoon, J., Vostrov, A. A., Barsov, E., Strunnikov, A. V., Morse, H. C., Loukinov, D., and Lobanenkov, V. (2010). The Structural Complexity of the Human BORIS Gene in Gametogenesis and Cancer. *PLoS One*, 5(11):e13872.

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., and Pirmohamed, M. (2019). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.*, 18(1):41–58.

Radaeva, M., Ton, A.-T., Hsing, M., Ban, F., and Cherkasov, A. (2021). Drugging the 'undruggable'. therapeutic targeting of protein–dna interactions with the use of computer-aided drug discovery methods. *Drug discovery today*, 26(11):2660–2679.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research*, 35(suppl_2):W193–W200.

Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D. R., and Zhao, K. (2017). CTCF-Mediated Enhancer-Promoter interaction is a critical regulator of Cell-to-Cell variation of gene expression. *Mol. Cell*, 67(6):1049–1058.e6.

Renaud, S., Loukinov, D., Alberti, L., Vostrov, A., Kwon, Y. W., Bosman, F. T., Lobanenkov, V., and Benhattar, J. (2011). BORIS/CTCFL-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. *Nucleic Acids Res.*, 39(3):862–873.

Renaud, S., Pugacheva, E. M., Delgado, M. D., Braunschweig, R., Abdullaev, Z., Loukinov, D., Benhattar, J., and Lobanenkov, V. (2007). Expression of the CTCF-paralogous cancer-testis gene, brother of the regulator of imprinted sites (BORIS), is regulated by three alternative promoters modulated by CpG methylation and by CTCF and p53 transcription factors. *Nucleic Acids Res.*, 35(21):7372–7388.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P. V. (2007). Critical dna binding interactions of the insulator protein ctcf: A small number of zinc fingers mediate strong binding, and a single finger-dna interaction controls binding at imprinted loci. *Journal of Biological Chemistry*, 282:33336–33345.

Rivero-Hinojosa, S., Pugacheva, E. M., Kang, S., Méndez-Catalá, C. F., Kovalchuk, A. L., Strunnikov, A. V., Loukinov, D., Lee, J. T., and Lobanenkov, V. V. (2021). The combined action of CTCF and its testis-specific paralog BORIS is essential for spermatogenesis. *Nat. Commun.*, 12(1):3846.

Rosa-Garrido, M., Ceballos, L., Alonso-Lecue, P., Abraira, C., Delgado, M. D., and Gandarillas, A. (2012). A Cell Cycle Role for the Epigenetic Factor CTCF-L/BORIS. *PLoS One*, 7(6):e39371.

Rotili, D. and Mai, A. (2011). Targeting histone demethylases: a new avenue for the fight against cancer. *Genes & cancer*, 2(6):663–679.

Rybaczyk, L. A., Bashaw, M. J., Pathak, D. R., and Huang, K. (2008). An indicator of cancer: downregulation of monoamine oxidase-a in multiple organs and species. *BMC Genomics*, 9:134.

Sadegh, S., Skelton, J., Anastasi, E., Bernett, J., Blumenthal, D. B., Galindez, G., Salgado-Albarrán, M., Lazareva, O., Flanagan, K., Cockell, S., Nogales, C., Casas, A. I., Schmidt, H. H. H. W., Baumbach, J., Wipat, A., and Kacprowski, T. (2021). Network medicine for disease module identification and drug repurposing with the NeDRex platform. *Nat. Commun.*, 12(1):6848.

Salamero, O., Montesinos, P., Willekens, C., Pérez-Simón, J. A., Pigneux, A., Récher, C., Popat, R., Carpio, C., Molinero, C., Mascaró, C., et al. (2020). First-in-human phase i study of iadademstat (ory-1001): a first-in-class lysine-specific histone demethylase 1a inhibitor, in relapsed or refractory acute myeloid leukemia. *Journal of Clinical Oncology*, 38(36):4260–4273.

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, L. A., Baumbach, J., and Soto-Reyes, E. (2019). The epigenetic factor BORIS (CTCFL) controls the androgen receptor regulatory network in ovarian cancer. *Oncogenesis*, 8(8):1–12.

Sander, T., Freyss, J., von Korff, M., and Rufener, C. (2015). DataWarrior: an open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.*, 55(2):460–473.

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G. D., Marshall, A., Flicek, P., and Odom, D. T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, 148(1-2):335–348.

Schwalie, P. C., Ward, M. C., Cain, C. E., Faure, A. J., Gilad, Y., Odom, D. T., and Flicek, P. (2013). Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome Biol.*, 14(12):1–15.

Schwarzenbach, H., Eichelser, C., Steinbach, B., Tadewaldt, J., Pantel, K., Lobanenkov, V., and Loukinov, D. (2014). Differential regulation of MAGE-A1 promoter activity by BORIS and sp1, both interacting with the TATA binding protein. *BMC Cancer*, 14(1):1–15.

Sha, J., Han, Q., Chi, C., Zhu, Y., Pan, J., Dong, B., Huang, Y., Xia, W., and Xue, W. (2020). Upregulated kdm4b promotes prostate cancer cell proliferation by activating autophagy. *Journal of cellular physiology*, 235(3):2129–2138.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11):2498–2504.

Shen, M. and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15:2507–2524.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, 119(7):941–953.

Shih, J. C. (2018). Monoamine oxidase isoenzymes: genes, functions and targets for behavior and cancer therapy. *J. Neural Transm.*, 125(11):1553–1566.

Shin, S. and Janknecht, R. (2007a). Activation of androgen receptor by histone demethylases jmjd2a and jmjd2d. *Biochemical and biophysical research communications*, 359(3):742–746.

Shin, S. and Janknecht, R. (2007b). Diversity within the jmjd2 histone demethylase family. *Biochemical and biophysical research communications*, 353(4):973–977.

Shin, S. and Janknecht, R. (2007c). Diversity within the JMJD2 histone demethylase family. *Biochem. Biophys. Res. Commun.*, 353(4):973–977.

Shirakami, Y. and Shimizu, M. (2018). Possible mechanisms of green tea and its constituents against cancer. *Molecules*, 23(9).

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(7371):74–79.

Sleutels, F., Soochit, W., Bartkuhn, M., Heath, H., Dienstbach, S., Bergmaier, P., Franke, V., Rosa-Garrido, M., Van De Nobelen, S., Caesar, L., Van Der Reijden, M., Bryne, J. C., Van Ijcken, W., Grootegoed, J. A., Delgado, M. D., Lenhard, B., Renkawitz, R., Grosveld, F., and Galjart, N. (2012). The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics and Chromatin*, 5(1):8.

Smith, I. M., Glazer, C. A., Mithani, S. K., Ochs, M. F., Sun, W., Bhan, S., Vostrov, A., Abdullaev, Z., Lobanenkov, V., Gray, A., Liu, C., Chang, S. S., Ostrow, K. L., Westra, W. H., Begum, S., Dhara, M., and Califano, J. (2009). Coordinated Activation of Candidate Proto-Oncogenes and Cancer Testes Antigens via Promoter Demethylation in Head and Neck Cancer and Lung Cancer. *PLoS One*, 4(3):e4961.

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., and Steinbeck, C. (2021a). Coconut online: collection of open natural products database. *Journal of Cheminformatics*, 13(1):2.

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., and Steinbeck, C. (2021b). COCONUT online: Collection of open natural products database. *J. Cheminform.*, 13(1):2.

Sousa da Silva, A. W. and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon parser interface. *BMC Res. Notes*, 5(1):1–8.

Sterling, J., Menezes, S. V., Abbassi, R. H., and Munoz, L. (2021). Histone lysine demethylases and their functions in cancer. *International Journal of Cancer*, 148(10):2375–2388.

Therneau, T. M., Grambsch, P. M., Therneau, T. M., and Grambsch, P. M. (2000). *The cox model*. Springer.

Tiffen, J. C., Bailey, C. G., Marshall, A. D., Metierre, C., Feng, Y., Wang, Q., Watson, S. L., Holst, J., and Rasko, J. E. (2013). The cancer-testis antigen BORIS phenocopies the tumor suppressor CTCF in normal and neoplastic cells. *Int. J. Cancer*, 133(7):1603–1613.

Tucker, M. R., Piana, S., Tan, D., LeVine, M. V., and Shaw, D. E. (2022). Development of force field parameters for the simulation of single-and double-stranded dna molecules and dna–protein complexes. *The Journal of Physical Chemistry B*, 126(24):4442–4457.

Turner, B. M. (2005). Reading signals on the nucleosome with a new nomenclature for modified histones. *Nature structural & molecular biology*, 12(2):110–112.

Ulaner, G. A., Vu, T. H., Li, T., Hu, J. F., Yao, X. M., Yang, Y., Gorlick, R., Meyers, P., Healey, J., Ladanyi, M., and Hoffman, A. R. (2003). Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. *Hum. Mol. Genet.*, 12(5):535–549.

Upadhyay, A. K., Judge, R. A., Li, L., Pithawalla, R., Simanis, J., Bodelle, P. M., Marin, V. L., Henry, R. F., Petros, A. M., and Sun, C. (2018). Targeting lysine specific demethylase 4a (kdm4a) tandem tudor domain–a fragment based approach. *Bioorganic & Medicinal Chemistry Letters*, 28(10):1708–1713.

Valencia, H., Newton, H. B., Hade, E., Sborov, D. W., Cavaliere, R., Poi, M., Phelps, M., Wang, J., Coss, C. C., Khountham, S., et al. (2016). A phase 1 study of ar-42 in patients with advanced solid tumors, including nervous system tumors.

van de Nobelen, S., Rosa-Garrido, M., Leers, J., Heath, H., Soochit, W., Joosen, L., Jonkers, I., Demmers, J., van der Reijden, M., Torrano, V., Grosveld, F., Delgado, M. D., Renkawitz, R., Galjart, N., and Sleutels, F. (2010). CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenetics Chromatin*, 3(1):1–21.

Vatolin, S., Abdullaev, Z., Pack, S. D., Flanagan, P. T., Custer, M., Loukinov, D. I., Pugacheva, E., Hong, J. A., Morse, H., Schrump, D. S., Risinger, J. I., Barrett, J. C., and Lobanenkov, V. V. (2005). Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes. *Cancer Res.*, 65(17):7751–7762.

Venza, I., Visalli, M., Oteri, R., Cucinotta, M., Teti, D., and Venza, M. (2013). Class ii-specific histone deacetylase inhibitors mc 1568 and mc 1575 suppress il-8 expression in human melanoma cells. *Pigment cell & melanoma research*, 26(2):193–204.

Wang, J., Wang, P., Dong, C., Zhao, Y., Zhou, J., Yuan, C., and Zou, L. (2020). Mechanisms of ebselen as a therapeutic and its pharmacology applications. *Future medicinal chemistry*, 12(23):2141–2160.

Wang, L.-S., Kuo, C.-T., Cho, S.-J., Seguin, C., Siddiqui, J., Stoner, K., Weng, Y.-I., Huang, T. H.-M., Tichelaar, J., Yearsley, M., Stoner, G. D., and Huang, Y.-W. (2013). Black raspberry-derived anthocyanins demethylate tumor suppressor genes through the inhibition of DNMT1 and DNMT3B in colon cancer cells. *Nutr. Cancer*, 65(1):118–125.

Wang, W.-C., Chu, C.-H., and Chen, C.-C. (2014). Crystal structure of JMJD2B complexed with pyridine-2,4-dicarboxylic acid and H3K9me3.

Wang, Z.-Q., Zhang, Z.-C., Wu, Y.-Y., Pi, Y.-N., Lou, S.-H., Liu, T.-B., Lou, G., and Yang, C. (2023). Bromodomain and extraterminal (bet) proteins: biological functions, diseases, and targeted therapy. *Signal transduction and targeted therapy*, 8(1):420.

Waqar, S. N. and Morgensztern, D. (2018). Lorlatinib: a new-generation drug for alk-positive nsclc. *The Lancet Oncology*, 19(12):1555–1557.

Warshakoon, N. C., Wu, S., Boyer, A., Kawamoto, R., Sheville, J., Renock, S., Xu, K., Pokross, M., Zhou, S., Winter, C., Walter, R., Mekel, M., and Evdokimov, A. G. (2006). Structure-based design, synthesis, and SAR evaluation of a new series of 8-hydroxyquinolines as HIF-1alpha prolyl hydroxylase inhibitors. *Bioorg. Med. Chem. Lett.*, 16(21):5517–5522.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., et al. (2018). Swiss-model: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1):W296–W303.

Wawruszak, A., Borkiewicz, L., Okon, E., Kukula-Koch, W., Afshan, S., and Halasa, M. (2021). Vorinostat (saha) and breast cancer: an overview. *Cancers*, 13(18):4700.

Wei, L., Liu, Q., Huang, Y., Liu, Z., Zhao, R., Li, B., Zhang, J., Sun, C., Gao, B., Ding, X., Yu, X., He, J., Sun, A., and Qin, Y. (2020). Knockdown of CTCF reduces the binding of EZH2 and affects the methylation of the SOCS3 promoter in hepatocellular carcinoma. *Int. J. Biochem. Cell Biol.*, 120:105685.

Wilson, K. A., Kellie, J. L., and Wetmore, S. D. (2014). DNA-protein π-interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Res.*, 42(10):6726–6741.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research*, 46(D1):D1074–D1082.

Woloszynska-Read, A., James, S. R., Link, P. A., Yu, J., Odunsi, K., and Karpf, A. R. (2007). DNA methylation-dependent regulation of BORIS/CTCFL expression in ovarian cancer. *Cancer Immun.*, 7.

Woloszynska-Read, A., Zhang, W., Yu, J., Link, P. A., Mhawech-Fauceglia, P., Collamat, G., Akers, S. N., Ostler, K. R., Godley, L. A., Odunsi, K., and Karpf, A. R. (2011). 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(8):2170–2180.

Woon, E. C., Tumber, A., Kawamura, A., Hillringhaus, L., Ge, W., Rose, N. R., Ma, J. H., Chan, M. C., Walport, L. J., Che, K. H., et al. (2012). Linking of 2-oxoglutarate and substrate binding sites enables potent and highly selective inhibition of jmjc histone demethylases. *Angewandte Chemie*, 124(7):1663–1666.

Wu, W., Cao, X., and Mo, L. (2021). Overexpression of KDM4D promotes acute myeloid leukemia cell development by activating MCL-1. *Am. J. Transl. Res.*, 13(4):2308–2319.

Xu, L., Gao, X., Yang, P., Sang, W., Jiao, J., Niu, M., Liu, M., Qin, Y., Yan, D., Song, X., et al. (2021). Ehmt2 inhibitor bix-01294 induces endoplasmic reticulum stress mediated apoptosis and autophagy in diffuse large b-cell lymphoma cells. *Journal of Cancer*, 12(4):1011.

Xu, Z. (2013). Mechanics of metal-catecholate complexes: the roles of coordination state and metal types. *Sci. Rep.*, 3:2914.

Yang, C.-J., Wang, C.-S., Hung, J.-Y., Huang, H.-W., Chia, Y.-C., Wang, P.-H., Weng, C.-F., and Huang, M.-S. (2009). Pyrogallol induces G2-M arrest in human lung cancer cells and inhibits tumor growth in an animal model. *Lung Cancer*, 66(2):162–168.

Yang, J., Jubb, A. M., Pike, L., Buffa, F. M., Turley, H., Baban, D., Leek, R., Gatter, K. C., Ragoussis, J., and Harris, A. L. (2010). The histone demethylase jmjd2b is regulated by estrogen receptor α and hypoxia, and is a key mediator of estrogen induced growth. *Cancer research*, 70(16):6456–6466.

Yang, M., Gocke, C. B., Luo, X., Borek, D., Tomchick, D. R., Machius, M., Otwinowski, Z., and Yu, H. (2006). Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell*, 23(3):377–387.

Yang, Y., Zhang, M., and Wang, Y. (2022). The roles of histone modifications in tumorigenesis and associated inhibitors in cancer therapy. *Journal of the National Cancer Center*, 2(4):277–290.

Ye, C., Jiang, N., Zheng, J., Zhang, S., Zhang, J., and Zhou, J. (2024). Epigenetic therapy: Research progress of decitabine in the treatment of solid tumors. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1879(2):189066.

Yin, M., Wang, J., Wang, M., Li, X., Zhang, M., Wu, Q., and Wang, Y. (2017). Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.*, 27(11):1365–1377.

Yoo, J., Winogradoff, D., and Aksimentiev, A. (2020). Molecular dynamics simulations of dna–dna and dna–protein interactions. *Current Opinion in Structural Biology*, 64:88–96.

Yu, X., Zhao, H., Wang, R., Chen, Y., Ouyang, X., Li, W., Sun, Y., and Peng, A. (2024). Cancer epigenetics: from laboratory studies and clinical trials to precision medicine. *Cell Death Discov*, 10(1):28.

Yuan, H., Ma, Q., Ye, L., and Piao, G. (2016a). The traditional medicine and modern medicine from natural products. *Molecules*, 21(5).

Yuan, X., Kong, J., Ma, Z., Li, N., Jia, R., Liu, Y., Zhou, F., Zhan, Q., Liu, G., and Gao, S. (2016b). Kdm4c, a h3k9me3 histone demethylase, is involved in the maintenance of human escc-initiating cells by epigenetically enhancing sox2 expression. *Neoplasia*, 18(10):594–609.

Yue, W. W., Gileadi, C., Krojer, T., Pike, A. C. W., von Delft, F., Ng, S., Carpenter, L., Arrowsmith, C., Weigelt, J., Edwards, A., Bountra, C., and Oppermann, U. (2010). Crystal structure of human JMJD2C catalytic domain.

Yue, W. W., Ng, S., Shafqat, N., Ugochukwu, E., McDonough, M., Pike, A. C. W., Filippakopoulos, P., von Delft, F., Arrowsmith, C., Weigelt, J., Edwards, A., Bountra, C., Schofield, C., and Oppermann, U. (2009). Crystal structure of the human 2-oxoglutarate oxygenase LOC390245.

Zeng, D., Liu, M., and Pan, J. (2017). Blocking ezh2 methylation transferase activity by gsk126 decreases stem cell-like myeloma cells. *Oncotarget*, 8(2):3396.

Zhang, Y., Fang, M., Li, S., Xu, H., Ren, J., Tu, L., Zuo, B., Yao, W., and Liang, G. (2022a). Btapep-tat peptide inhibits adp-ribosylation of boris to induce dna damage in cancer. *Molecular Cancer*, 21(1):158.

Zhang, Y., Fang, M., Li, S., Xu, H., Ren, J., Tu, L., Zuo, B., Yao, W., and Liang, G. (2022b). BTApep-TAT peptide inhibits ADP-ribosylation of BORIS to induce DNA damage in cancer. *Mol. Cancer*, 21(1):1–17.

Zhang, Y., Fang, M., Song, Y., Ren, J., Fang, J., and Wang, X. (2017). Brother of Regulator of Imprinted Sites (BORIS) suppresses apoptosis in colorectal cancer. *Sci. Rep.*, 7(1):1–12.

Zhang, Y., Wu, F., and Yang, N. (2018). Nhwd-870, a novel bet family bromodomain inhibitor targeting brd4, proved to be effective and promising for treatment of small cell lung cancer. *Cancer Research*, 78(13_Supplement):1668–1668.

Zhao, Y., Yang, C.-Y., and Wang, S. (2013). The making of i-bet762, a bet bromodomain inhibitor now in clinical development. *Journal of medicinal chemistry*, 56(19):7498–7500.

Zhu, K., Zhang, H., Luan, Y., Hu, B., Shen, T., Ma, B., Zhang, Z., and Zheng, X. (2024). Kdm4c promotes mouse hippocampal neural stem cell proliferation through modulating apoe expression. *The FASEB Journal*, 38(5):e23511.

Zhu, W., Shenoy, A., Kundrotas, P., and Elofsson, A. (2023). Evaluation of alphafold-multimer prediction on multi-chain protein complexes. *Bioinformatics*, 39(7):btad424.

Šali, A. and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234:779–815.

Apéndices

A. Constantes de Michaelis para los miembros de la subfamilia KDM4

Tabla A.1: Valores de Km para cada miembro de la subfamilia KDM4. Datos tomados de Hillringhaus et al. (2011)

	H3K9me2	**H3K9me3**
KDM4A	73 ± 26 μM	45 ± 7 μM
KDM4B	50 ± 11 μM	31 ± 10 μM
KDM4C	66 ± 14 μM	48 ± 9 μM
KDM4D	74 ± 6 μM	37 ± 9 μM
KDM4E	25 ± 3 μM	23 ± 4 μM

B. Isoformas de BORIS

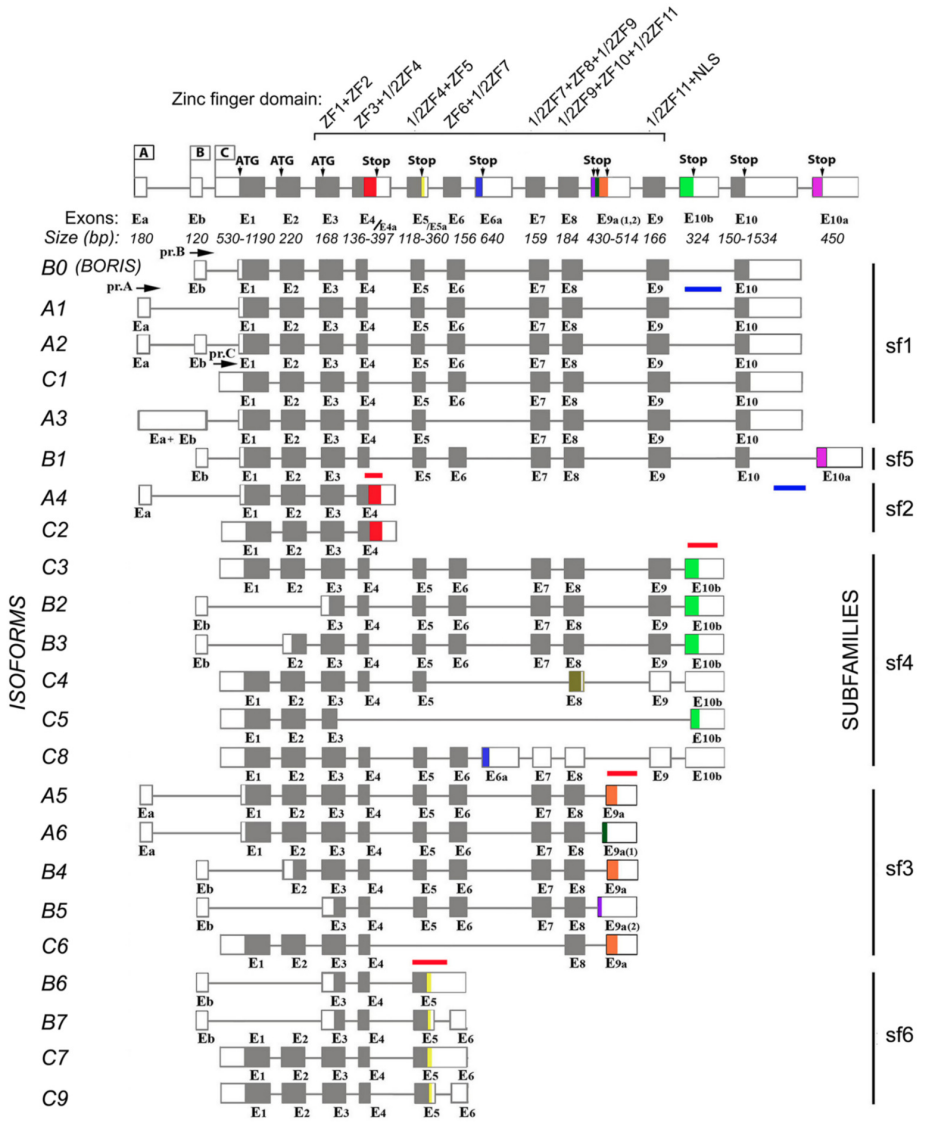


Figura B.1: Clasificación de las isoformas de BORIS. El gen *CTCFL* está representado en la parte superior con sus tres sitios de transcripción diferentes. Los dieciséis exones se representan como recuadros, los exones codificantes se denotan en gris. Imagen tomada de Pugacheva et al. (2010).

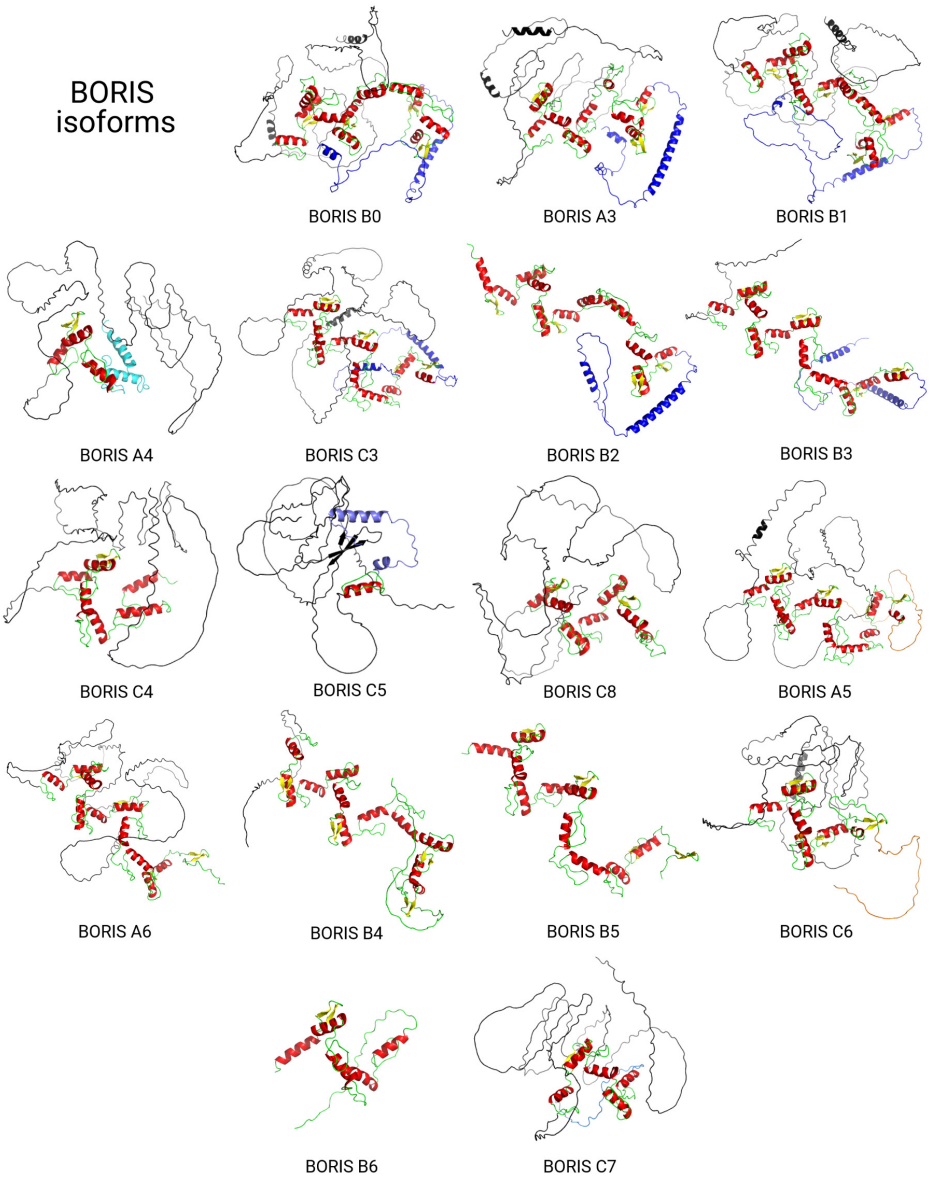


Figura B.2: Modelado 3D de las isoformas de BORIS. Las 17 isoformas de BORIS descritas por Link et al. (2013) se modelaron con AlphaFold. El dominio N-terminal se muestra en negro, el C-terminal en azul, y los dedos de zinc en rojo y amarillo.

C. Parámetros para el algoritmo *FILTER* de OpenEye.

```
MAX_COUNT_FORMAL_CRG 4 "Maximum number of formal charges"
MAX_XLOGP 1.0 "Maximum XLogP"
MIN_SOLUBILITY soluble "Minimum solubility"
MAX_LIPINSKI 3 "Maximum number of Lipinski violations"
PSA_USE_SandP false "Count S and P as polar atoms"
MIN_2D_PSA 100.0 "Minimum 2-Dimensional (SMILES) Polar Surface Area"
MAX_2D_PSA 250.0 "Maximum 2-Dimensional (SMILES) Polar Surface Area"
RULE 1 acid
RULE 1 acid_chloride
RULE 1 acid_halide
RULE 1 alkyl_phosphate
RULE 1 alkylating_agent
RULE 1 amide
RULE 1 amine
RULE 1 amino_acid
RULE 1 carboxylic_acid
RULE 1 cation_C_Cl_I_P_or_S
RULE 1 disulfide
RULE 1 hydroxylamine
RULE 1 nitro
RULE 1 N_P_S_Halides
RULE 1 organometallic
RULE 1 phosphoric_acid
RULE 1 phosphoric_ester
RULE 1 sulfinimine
RULE 1 sulfinylthio
RULE 1 sulfonamide
RULE 1 thiocarbonyl
RULE 1 thioester
RULE 1 thiol
RULE 1 thiourea
```

D. Archivos de entrada para Modeller

```
from modeller import *
from modeller.automodel import *

class MyModel(AutoModel):
    def special_restraints(self, aln):
        rsr = self.restraints
        for ids in (('SG:4:A', 'ZN:286:B'),
                    ('SG:7:A', 'ZN:286:B'),
                    ('NE2:20:A', 'ZN:286:B'),
                    ('NE2:24:A', 'ZN:286:B'),
                    ('SG:32:A', 'ZN:287:B'),
                    ('SG:35:A', 'ZN:287:B'),
                    ('NE2:48:A', 'ZN:287:B'),
                    ('NE2:53:A', 'ZN:287:B'),
                    ('SG:61:A', 'ZN:288:B'),
                    ('SG:64:A', 'ZN:288:B'),
                    ('NE2:77:A', 'ZN:288:B'),
                    ('NE2:81:A', 'ZN:288:B'),
                    ('SG:89:A', 'ZN:289:B'),
                    ('SG:92:A', 'ZN:289:B'),
                    ('NE2:105:A', 'ZN:289:B'),
                    ('NE2:109:A', 'ZN:289:B'),
                    ('SG:117:A', 'ZN:290:B'),
                    ('SG:120:A', 'ZN:290:B'),
                    ('NE2:133:A', 'ZN:290:B'),
                    ('NE2:138:A', 'ZN:290:B'),
                    ('SG:147:A', 'ZN:291:B'),
                    ('SG:150:A', 'ZN:291:B'),
                    ('NE2:163:A', 'ZN:291:B'),
                    ('NE2:168:A', 'ZN:291:B'),
                    ('SG:177:A', 'ZN:292:B'),
                    ('SG:180:A', 'ZN:292:B'),
                    ('NE2:193:A', 'ZN:292:B'),
                    ('NE2:197:A', 'ZN:292:B'),
                    ('SG:205:A', 'ZN:293:B'),
                    ('SG:208:A', 'ZN:293:B'),
                    ('NE2:221:A', 'ZN:293:B'),
```

```
                    ('NE2:225:A', 'ZN:293:B'),
                    ('SG:233:A', 'ZN:294:B'),
                    ('SG:236:A', 'ZN:294:B'),
                    ('NE2:249:A', 'ZN:294:B'),
                    ('NE2:254:A', 'ZN:294:B'),
                    ('SG:265:A', 'ZN:295:B'),
                    ('SG:268:A', 'ZN:295:B'),
                    ('NE2:281:A', 'ZN:295:B'),
                    ('SG:285:A', 'ZN:295:B')

):
            atoms = [self.atoms[i] for i in ids]
            rsr.add(forms.UpperBound(group=physical.upper_distance,
                                     feature=features.Distance(*atoms),
                                     mean=2.3, stdev=0.3))

env = Environ()
env.io.hetatm = True
a = MyModel(env, alnfile='BORIS_CTCF.ali',
               knowns=('5T0U_A','5YEL', '5YEF'),
        sequence='BORIS_ZF',
        assess_methods=(assess.DOPE,
                                #soap_protein_od.Scorer(),
                                assess.GA341))
a.starting_model = 1
a.ending_model = 1000
a.make()
```

Alignment file for BORIS in .pir format (BORIS_CTCF.ali)

```
>P1;BORIS_ZF
sequence:BORIS_ZF::::human:BORIS:::
PHLCHLCLKTFRTVTLLRNHVNTHTGTRPYKCNDCNMAFVTSGELVRHRRYKHTHEKPFK
CSMCKYASVEASKLKRHVRSHTGERPFQCCQCSYASRDTYKLKRHMRTHSGEKPYECHIC
HTRFTQSGTMKIHILQKHGENVPKYQCPHCATIIARKSDLRVHMRNLHAYSAAELKCRYC
SAVFHERYALIQHQKTHKNEKRFKCKHCSYACKQERHMTAHIRTHTGEKPFTCLSCNKCF
RQKQLLNAHFRKYHDANFIPTVYKCSKCGKGFSRWINLHRHSEKC/zzzzzzzzzz/...
...................../......................../.........
..../.............../
*
>P1;5T0U_A
structureX:Temp_CTCF5t0u.pdb:293:A:336::::3.20:0.231
THKCHLCGRAFRTVTLLRNHLNTHTGTRPHKCPDCDMAFVTSGE----------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------------------------/zz--------/---
----------------------/-------------------------/---------
----/--------------/
*
```

```
>P1;5YEL
structureX:Temp_CTCF5YEL.pdb:406:A:577::::2.96:0.269
------------------------------------------------------------
-----------------------------------------------------PYECYIC
HARFTQSGTMKMHILQKH---VAKFHCPHCDTVIARKSDLGVHLRKQHSYIEQGKKCRYC
DAVFHERYALIQHQKSHKNEKRFKCDQCDYASRQERHMIMHKRTHTG-KPYACSHCDKTF
RQKQLLDMHFKRYHDP--VPAAFVCSKCGKTFTRRNTMARHDNAC/----zzzzzz/...
....................../......................../---------
----/--------------/
*
>P1;5YEF
structureX:Temp_CTCF5yef.pdb:321:A:436::::2.96:0.269
---------------------------PHKCPDCDMAFVTSGELVRHRRYKHTHEKPFK
CSMCDYASVEVSKLKRHIRSHTGERPFQCSLCSYASRDTYKLKRHMRTHSGEKPYECYIC
HARFTQSGTMKMHILQKHTENVAK-----------------------------------
------------------------------------------------------------
----------------------------------------------/-zzz------/---
-----------------------/-------------------------/.........
..../.............../
*
```

Alignment file for CTCF in .pir format (CTCF_CTCF.ali)

```
>P1;CTCF_ZF
sequence:CTCF_ZF:293::557:human:CTCF:::
PHKCHLCGRAFRTVTLLRNHLNTHTGTRPHKCPDCDMAFVTSGELVRHRRYKHTHEKPFK
CSMCDYASVEVSKLKRHIRSHTGERPFQCSLCSYASRDTYKLKRHMRTHSGEKPYECYIC
HARFTQSGTMKMHILQKHTENVAKFHCPHCDTVIARKSDLGVHLRKQHSYIEQGKKCRYC
DAVFHERYALIQHQKSHKNEKRFKCDQCDYACRQERHMIMHKRTHTGEKPYACSHCDKTF
RQKQLLDMHFKRYHDPNFVPAAFVCSKCGKTFTRRNTMARHDNAC/zzzzzzzzzz/...
....................../......................../.........
..../.............../
*
>P1;5T0U_A
structureX:Temp_CTCF5t0u.pdb:293:A:336::::3.20:0.231
THKCHLCGRAFRTVTLLRNHLNTHTGTRPHKCPDCDMAFVTSGE----------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
----------------------------------------------/zz--------/---
-----------------------/-------------------------/---------
----/--------------/
*
>P1;5YEL
structureX:Temp_CTCF5YEL.pdb:406:A:577::::2.96:0.269
------------------------------------------------------------
-----------------------------------------------------PYECYIC
HARFTQSGTMKMHILQKH---VAKFHCPHCDTVIARKSDLGVHLRKQHSYIEQGKKCRYC
DAVFHERYALIQHQKSHKNEKRFKCDQCDYASRQERHMIMHKRTHTG-KPYACSHCDKTF
RQKQLLDMHFKRYHDP--VPAAFVCSKCGKTFTRRNTMARHDNAC/----zzzzzz/...
....................../......................../---------
----/--------------/
```

```
*
>P1;5YEF
structureX:Temp_CTCF5yef.pdb:321:A:436::::2.96:0.269
----------------------------PHKCPDCDMAFVTSGELVRHRRYKHTHEKPFK
CSMCDYASVEVSKLKRHIRSHTGERPFQCSLCSYASRDTYKLKRHMRTHSGEKPYECYIC
HARFTQSGTMKMHILQKHTENVAK-----------------------------------
------------------------------------------------------------
---------------------------------------/-zzz------/---
-----------------------/-------------------------/.........
..../................/
*
```

E. Parámetros usados en las simulaciones de Dinámica Molecular

E.1 Steepest descent minimization

```
title       = Minimization

integrator      = steep
emtol           = 100.0
emstep          = 0.01
nsteps          = 50000

nstlist         = 10
cutoff-scheme   = Verlet
ns_type         = grid
rlist           = 1.2
coulombtype     = PME
rcoulomb        = 1.2
vdwtype         = cutoff
vdw-modifier    = force-switch
rvdw-switch     = 1.0
rvdw            = 1.2
pbc             = xyz
DispCorr        = no
```

E.2 Conjugate gradients minimization

```
title       = Minimization


integrator      = cg
emtol           = 100.0
emstep          = 0.02
nsteps          = 50000
nstcgsteep      = 100

nstlist         = 10
```

```
cutoff-scheme   = Verlet
ns_type         = grid
rlist           = 1.2
coulombtype     = PME
rcoulomb        = 1.2
vdwtype         = cutoff
vdw-modifier    = force-switch
rvdw-switch     = 1.0
rvdw            = 1.2
pbc             = xyz
DispCorr        = no
```

E.3 NVT equilibration

```
define                  = -DPOSRES
; Run parameters
integrator              = md
nsteps                  = 50000
dt                      = 0.002
; Output control
nstenergy               = 500
nstlog                  = 500
nstxout-compressed      = 500
; Bond parameters
continuation            = no
constraint_algorithm    = lincs
constraints             = h-bonds
lincs_iter              = 1
lincs_order             = 4
; Neighbor searching and vdW
cutoff-scheme           = Verlet
ns_type                 = grid
nstlist                 = 20
rlist                   = 1.2
vdwtype                 = cutoff
vdw-modifier            = force-switch
rvdw-switch             = 1.0
rvdw                    = 1.2
; Electrostatics
coulombtype             = PME
rcoulomb                = 1.2
pme_order               = 4
fourierspacing          = 0.16
; Temperature coupling
tcoupl                  = V-rescale
tc-grps                 = Protein_ZN2_DNA Water_NA_CL
tau_t                   = 0.1   0.1
ref_t                   = 300   300
; Pressure coupling
pcoupl                  = no
; Periodic boundary conditions
```

```
pbc                     = xyz


DispCorr                = no
; Velocity generation
gen_vel                 = yes
gen_temp                = 300
gen_seed                = -1
```

E.4 NPT equilibration

```
define                  = -DPOSRES

; Run parameters
integrator              = md
nsteps                  = 50000
dt                      = 0.002
; Output control
nstenergy               = 500
nstlog                  = 500
nstxout-compressed      = 500
; Bond parameters
continuation            = yes
constraint_algorithm    = lincs
constraints             = h-bonds
lincs_iter              = 1
lincs_order             = 4
; Neighbor searching and vdW
cutoff-scheme           = Verlet
ns_type                 = grid
nstlist                 = 20
rlist                   = 1.2
vdwtype                 = cutoff
vdw-modifier            = force-switch
rvdw-switch             = 1.0
rvdw                    = 1.2
; Electrostatics
coulombtype             = PME
rcoulomb                = 1.2
pme_order               = 4
fourierspacing          = 0.16
; Temperature coupling
tcoupl                  = V-rescale
tc-grps                 = Protein_ZN2_DNA Water_NA_CL
tau_t                   = 0.1   0.1
ref_t                   = 300   300
; Pressure coupling
pcoupl                  = Berendsen
pcoupltype              = isotropic
tau_p                   = 2.0
ref_p                   = 1.0
```

```
compressibility         = 4.5e-5
refcoord_scaling        = com
; Periodic boundary conditions
pbc                     = xyz

DispCorr                = no
; Velocity generation
gen_vel                 = no
```

E.5 Molecular dynamics: Production run

```
; Run parameters
integrator              = md
nsteps                  = 25000000
dt                      = 0.002
; Output control
nstenergy               = 5000
nstlog                  = 5000
nstxout-compressed      = 5000
; Bond parameters
continuation            = yes
constraint_algorithm    = lincs
constraints             = h-bonds
lincs_iter              = 1
lincs_order             = 4
; Neighbor searching and vdW
cutoff-scheme           = Verlet
ns_type                 = grid
nstlist                 = 20
rlist                   = 1.2
vdwtype                 = cutoff
vdw-modifier            = force-switch
rvdw-switch             = 1.0
rvdw                    = 1.2
; Electrostatics
coulombtype             = PME
rcoulomb                = 1.2
pme_order               = 4
fourierspacing          = 0.16
; Temperature coupling
tcoupl                  = V-rescale
tc-grps                 = Protein_ZN2_DNA Water_NA_CL
tau_t                   = 0.1   0.1
ref_t                   = 300   300
; Pressure coupling
pcoupl                  = Parrinello-Rahman
pcoupltype              = isotropic
tau_p                   = 2.0
ref_p                   = 1.0
compressibility         = 4.5e-5
; Periodic boundary conditions
```

```
41 pbc                      = xyz
42
43 DispCorr                 = no
44 ; Velocity generation
45 gen_vel                  = no
```

F. Secuencia de comandos utilizados para el análisis de acoplamiento molecular de BORIS y CTFC

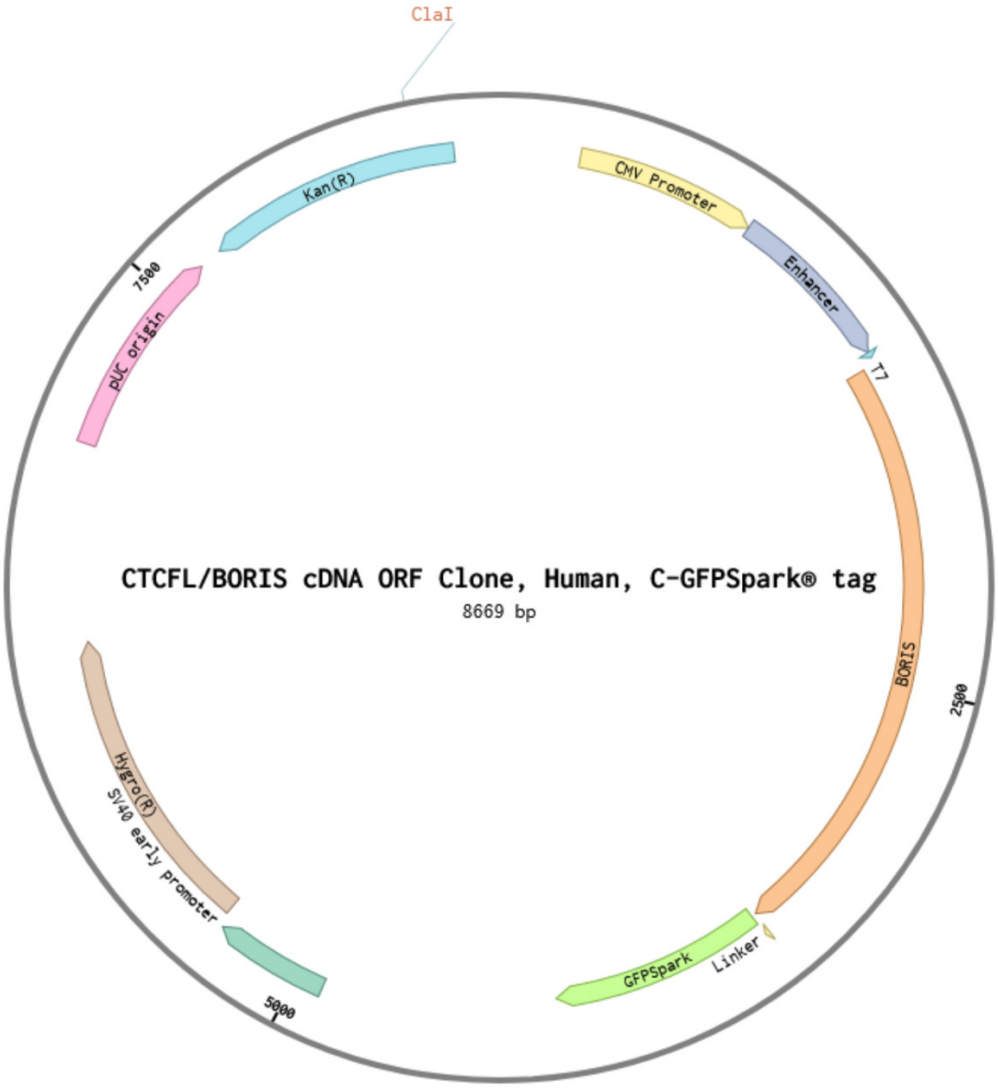
```
    #! /bin/bash
workdir=/DATA/Aylin/BORIS_CTCF_Docking/
ligLib=/DATA/LigandsDB/DrugBank/mol2/
cpus=2

proteins="BORIS CTCF"
databases="DrugBank"
sites="Site1 Site2"

for i in $databases
do
  for x in $proteins
    do
      echo "----------DiffDock--------"
      if [ ! -d "/$workdir/$x/$i/DiffDockResults" ]; then
      mkdir /$workdir/$x/$i/
      bash /DATA/Aylin/Scripts/Docking/diffdock_batch.sh \
      --ligandDir /DATA/LigandsDB/$i/sdf/ \
      --protein $workdir/$x\_clean.pdb \
      --splits 1 \
      --outdir $workdir/$x/$i/
      else
      echo "/$workdir/$x/$i/DiffDockResults already exists. Skipping"
      fi

      mkdir $workdir/$x/$i/Site1/
      mkdir $workdir/$x/$i/Site2/

      mkdir /$workdir/$x/$i/DiffDock_sdf
      for f in $workdir/$x/$i/DiffDockResults/*/*.sdf
        do
          b=`basename $f .sdf`
          dir=$(cat $f | head -n 1 | awk '{print $2}')
          obabel -i sdf $f -o sdf -O /$workdir/$x/$i/DiffDock_sdf/$dir\_$b.sdf
        done
      echo "---------- Vina ---------"
```

```
        echo "SITE1"

        bash /DATA/Aylin/Scripts/Docking/vina.sh \
        --ligand /DATA/LigandsDB/$i \
        --outDir $workdir/$x/$i/Site1/ \
        --split 20 \
        --config /DATA/Aylin/BORIS_CTCF_Docking/config_$x\_S1.txt

        echo "SITE2"

        bash /DATA/Aylin/Scripts/Docking/vina.sh \
        --ligand /DATA/LigandsDB/$i/ \
        --outDir $workdir/$x/$i/Site2/ \
        --split 20 \
        --config /DATA/Aylin/BORIS_CTCF_Docking/config_$x\_S2.txt

        echo "----------- PLANTS ---------"
        echo "SITE 1"

        if [ ! -d "/$workdir/$x/$i/Site1/PLANTS" ]; then
        mkdir $workdir/$x/$i/Site1/PLANTS/
        bash /DATA/Aylin/Scripts/Docking/parallelPLANTS.sh \
        $workdir/$x/$x\_clean.mol2 \
        /DATA/LigandsDB/$i/PLANTS.mol2 \
        $cpus \
        $workdir/bindingsite_S1.def \
        $workdir/$x/$i/Site1/PLANTS/
        else
        echo "/$workdir/$x/$i/Site1/PLANTS already exists. Skipping"
        fi

        echo "SITE 2"
        if [ ! -d "/$workdir/$x/$i/Site2/PLANTS" ]; then
        mkdir $workdir/$x/$i/Site2/PLANTS/
        bash /DATA/Aylin/Scripts/Docking/parallelPLANTS.sh \
        $workdir/$x/$x\_clean.mol2 \
        /DATA/LigandsDB/$i/PLANTS.mol2 \
        $cpus \
        $workdir/bindingsite_S2.def \
        $workdir/$x/$i/Site2/PLANTS/
        else
        echo "/$workdir/$x/$i/Site2/PLANTS already exists. Skipping"
        fi
        for site in $sites
        do
        find $workdir/$x/$i/$site/PLANTS/results -name "* *" -type f | rename 's/ /_/g'
        rm $workdir/$x/$i/$site/names.txt
        touch $workdir/$x/$i/$site/names.txt
        #Convert Vina output to mol2
        #
        echo "Spliting Vina results and converting to .mol2"
        mkdir $workdir/$x/$i/$site/Vina_mol2
        for f in $workdir/$x/$i/$site/VinaResults/*.pdbqt
        do
```

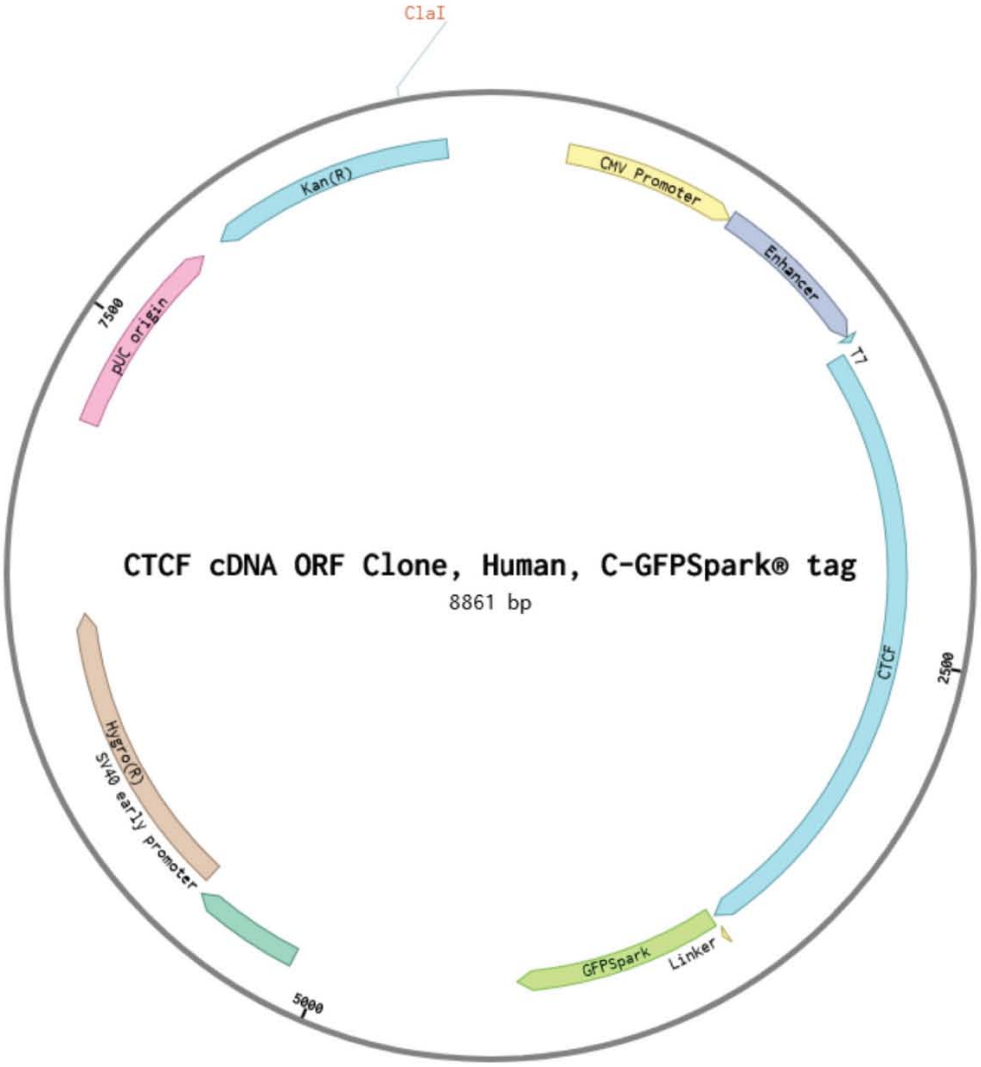
```
91            b=`basename $f .pdbqt`
92            ligand=$(cat $ligLib/$b.mol2 | sed -n '2 p' | awk '{print $2}')
93            echo "$ligand   $b" >> $workdir/$x/$i/$site/names.txt
94            obabel -i pdbqt $f -o mol2 -O $workdir/$x/$i/$site/Vina_mol2/$ligand\_pose.mol
95        done
96
97    done
98 done
```

G. Mapas de plásmidos

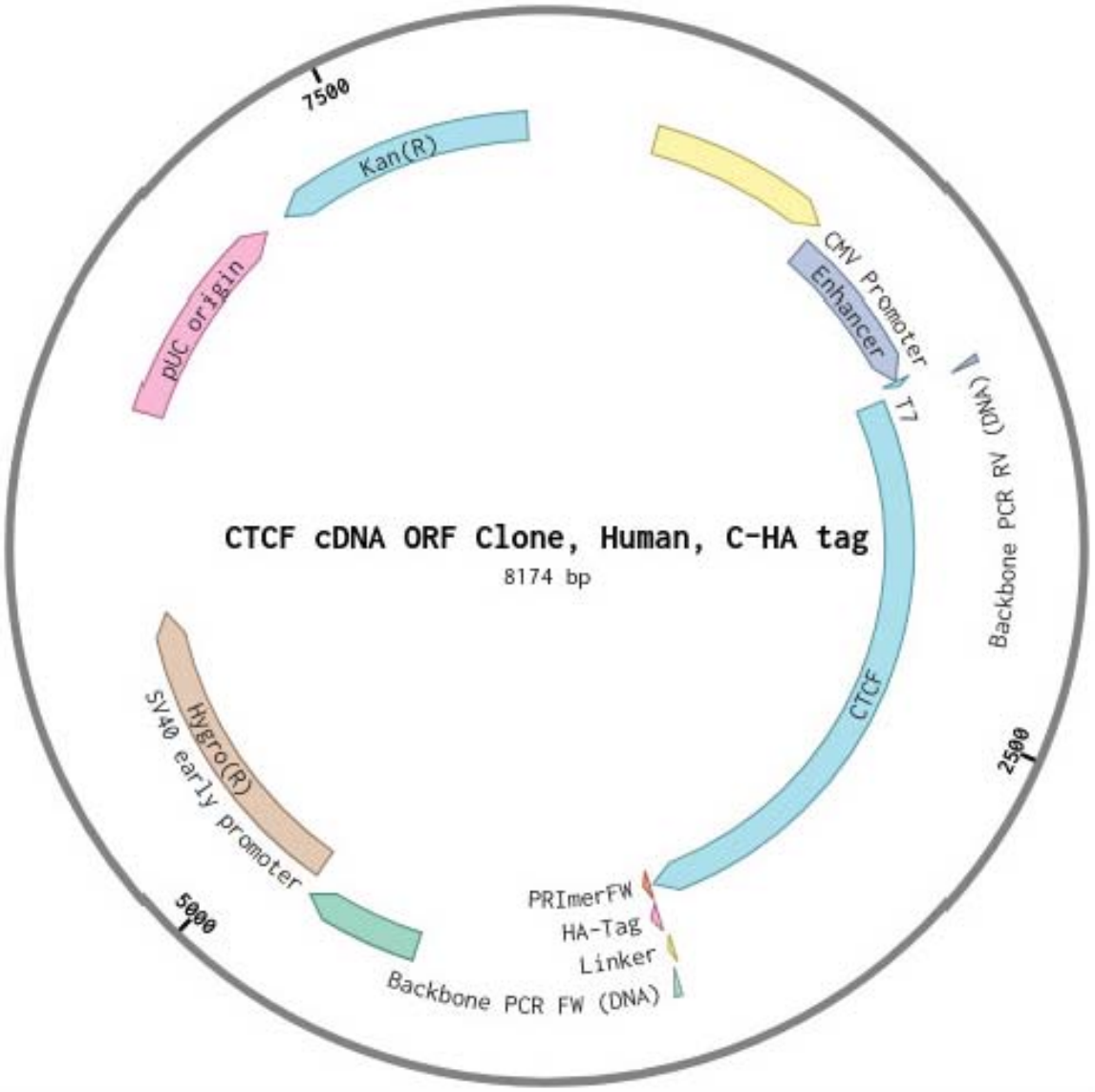
G.1 CTCFL/BORIS cDNA ORF Clone, Human, C-GFPSpark® tag



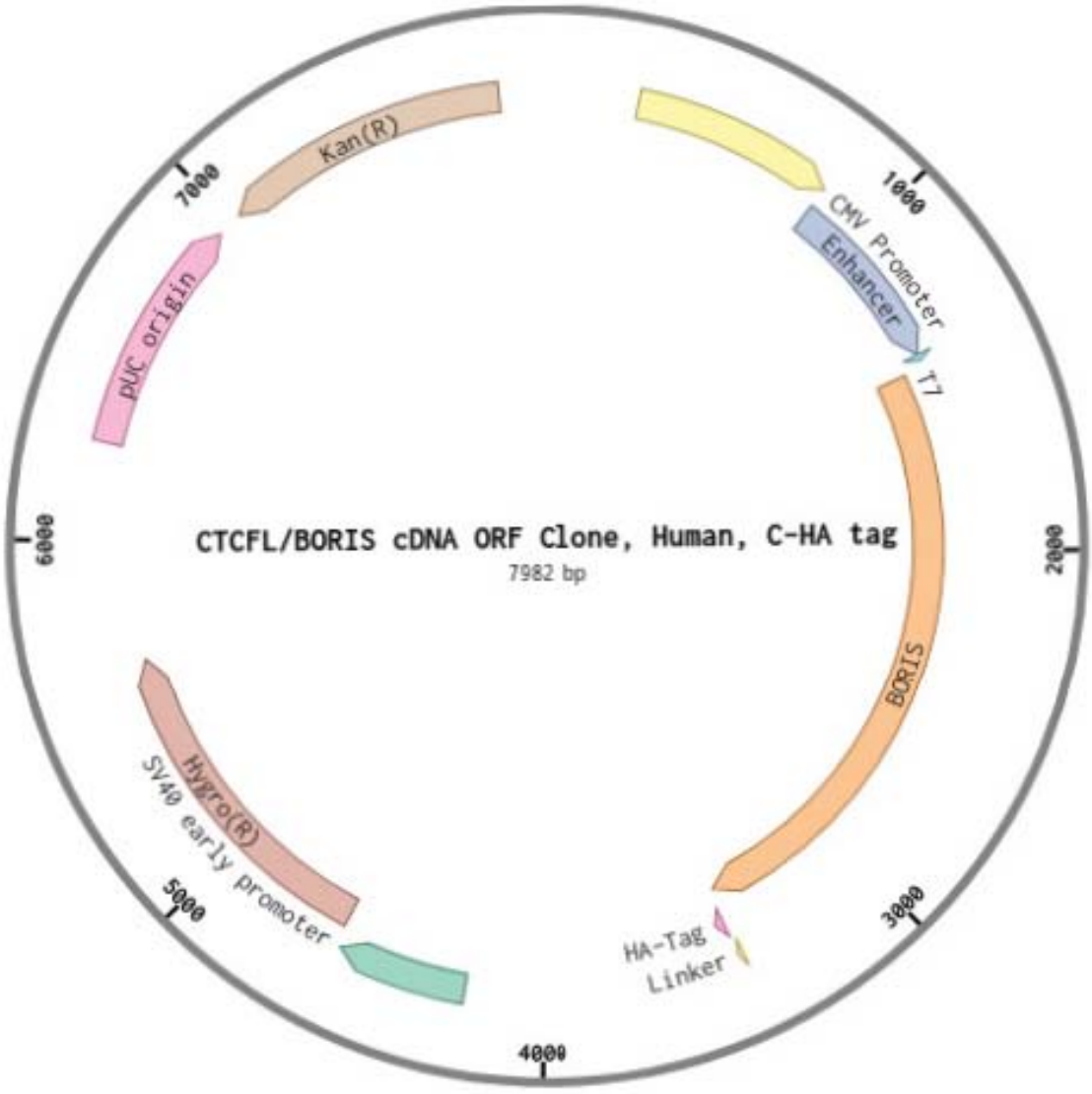
G.2 CTCF cDNA ORF Clone, Human, C-GFPSpark® tag



G.3 CTCF cDNA ORF Clone, Human, HA tag



G.4 BORIS cDNA ORF Clone, Human, HA tag



H. Artículos publicados

npj | Systems Biology and Applications

www.nature.com/npjsba

ARTICLE OPEN

Check for updates

Comparative transcriptome analysis reveals key epigenetic targets in SARS-CoV-2 infection

Marisol Salgado-Albarrán[1,2,7], Erick I. Navarro-Delgado[3,7], Aylin Del Moral-Morales[1,7], Nicolas Alcaraz[4], Jan Baumbach[5,6], Rodrigo González-Barrios[3]✉ and Ernesto Soto-Reyes[1]✉

COVID-19 is an infection caused by SARS-CoV-2 (Severe Acute Respiratory Syndrome coronavirus 2), which has caused a global outbreak. Current research efforts are focused on the understanding of the molecular mechanisms involved in SARS-CoV-2 infection in order to propose drug-based therapeutic options. Transcriptional changes due to epigenetic regulation are key host cell responses to viral infection and have been studied in SARS-CoV and MERS-CoV; however, such changes are not fully described for SARS-CoV-2. In this study, we analyzed multiple transcriptomes obtained from cell lines infected with MERS-CoV, SARS-CoV, and SARS-CoV-2, and from COVID-19 patient-derived samples. Using integrative analyses of gene co-expression networks and de-novo pathway enrichment, we characterize different gene modules and protein pathways enriched with Transcription Factors or Epifactors relevant for SARS-CoV-2 infection. We identified EP300, MOV10, RELA, and TRIM25 as top candidates, and more than 60 additional proteins involved in the epigenetic response during viral infection that has therapeutic potential. Our results show that targeting the epigenetic machinery could be a feasible alternative to treat COVID-19.

npj Systems Biology and Applications (2021)7:21 ; https://doi.org/10.1038/s41540-021-00181-x

INTRODUCTION

The coronavirus family (CoV) are non-segmented, positive-sense, and enveloped RNA viruses that have been identified as the cause of multiple enteric and respiratory diseases in both animals and humans[1]. Three major CoV strains of this family have caused recent human pandemics: the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2002–2003[2], severe acute respiratory syndrome coronavirus 1 (SARS-CoV) in 2012, and SARS-CoV-2 in 2020[3]. The most recent one was identified in Wuhan, China by the end of 2019 and is the etiological origin of atypical pneumonia known as Coronavirus Disease 2019 (COVID-19), which has caused a global outbreak and is one of the top sixth public health emergencies of international concern[4] with 98,089,877 confirmed cases and 2,100,404 deaths as of January 2021, leading to the biggest CoV pandemic in modern times[5].

By being intracellular pathogens, viruses' infection strategy requires the continuous subordination and exploitation of cellular transcriptional machinery and metabolism in order to ensure its own expansion. To do so, the host genome expression must be used and, to be successful, this will depend on chromatin dynamics and transcription regulation, which are principally ruled by epigenetic mechanisms, such as DNA methylation, histone post-translational modifications (HPTM), and transcription factors (TFs)[6]. During a viral infection, it has been reported that epigenetic and transcriptional changes occur for both sides: the infected cell promotes an antiviral environmental response, leading to the induction of pathways to survive, while the virus switches off the expression of critical anti-viral host cell genes[7,8].

Several studies have reported the importance of epigenetic modifications in viral infections. In the influenza virus, specific gene promoter DNA methylation[9], decreased H3K4me3 (a hallmark of active chromatin)[10], histone acetylation in H3 and H4 histones, and increased levels of H4K20me2 and unmodified H3K36 and H4K79 have been reported[11]. Interestingly, these HPTMs do not always trigger the same mechanisms and lead to similar phenotypes; for example, depletion of H3K79me2, an epigenetic mark that is usually increased upon viral infections due to an upregulation of DOT1L, results in impaired viral growth in human cytomegalovirus infection[12], while enhancing the replication in influenza virus[11]. However, these mechanisms usually lead to host transcriptional inactivation, which contributes to the altered cellular transcription produced by viral infections.

Regarding CoVs, few experimental studies have been conducted to unravel the epigenetic proteins and marks involved in their infection and pathogenesis in MERS-CoV and SARS-CoV, being especially scarce in SARS-CoV-2 due to its recent appearance. For MERS-CoV and SARS-CoV, different outcomes have been reported, such as the mechanisms used to control the interferon-stimulated genes, which involve H3K27 methylation in MERS-CoV but not in SARS-CoV[13], and the ones used to down-regulate antigen-presenting molecules, which involves DNA methylation in MERS-CoV and not in SARS-CoV[9]. These studies show that epigenetic mechanisms are highly important in the host gene expression control carried out by the virus and that, despite the phylogenetic closeness, these mechanisms can be very different between strains, highlighting the need to understand the epigenetic processes that play a role in SARS-CoV-2 infection.

Integrative computational methods are promising approaches used to generate research hypotheses, generate consensus regulatory networks and describe deregulated processes in SARS-CoV-2 infection[14,15]. Nevertheless, they have overlooked key epigenetic and TFs that underlie the infected phenotype.

[1]Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana-Cuajimalpa (UAM-C), Mexico City, Mexico. [2]Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany. [3]Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología, Mexico City, Mexico. [4]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [5]Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany. [6]Computational BioMedicine Lab, Institute of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark. [7]These authors contributed equally: Marisol Salgado-Albarrán, Erick I. Navarro-Delgado, Aylin Del Moral-Morales ✉email: rodrigop@ciencias.unam.mx; esotoreyes@cua.uam.mx

Since drugs that target the epigenetic landscape of diseased cells have shown great potential and have proved to be game-changing as complementary treatments of complex diseases, such as cancer[16], the identification of these key epigenetic proteins and TFs become highly important in our current context, where popular regimen candidates for treating COVID-19, such as Remdesivir, Hydroxychloroquine, Lopinavir, and Interferon have shown to have little or no effect on reducing mortality of hospitalized COVID-19 individuals[17].

In this work, we gathered publicly available RNA-seq data from SARS-CoV-2, SARS-CoV, and MERS-CoV infected cell lines and patient samples and performed differential expression analyses together with a weighted gene co-expression network analysis to identify unique and shared central epigenetic players in SARS-CoV-2, SARS-CoV, and MERS-CoV. Candidate genes were further prioritized by integrating differentially expressed genes (DEGs), enrichment tests, gene-coexpression network, and viral–host protein–protein interaction network analysis to propose potential key epigenetic proteins involved in SARS-CoV-2 infection. Finally, we identified currently approved drugs that target key epigenetic drivers of SARS-CoV-2 infection, and thus they are potential new therapeutic approaches for COVID-19.

RESULTS

SARS-CoV-2, SARS-CoV, and MERS-CoV induce different transcriptional and epigenetic responses during infection in pulmonary cell lines

In order to identify the genes that change their expression in pulmonary cell lines (Calu-3, MRC-5, A549, and NHBE) due to infection of Coronaviruses such as MERS-CoV, SARS-CoV, or SARS-CoV-2, differential expression analysis was performed in RNA-seq data (Supplementary Table 1).

As a first approach, we evaluated the overlapping DEGs identified for each virus regardless of the cell type and in common among viruses (Supplementary Table 2). For MERS-CoV and SARS-CoV, the overlap among all cell conditions was considered. For SARS-CoV-2, the overlap among 3 out of the 4 cell conditions was used, since the NHBE cell line showed a small number of DEGs most likely because these cells are derived from normal bronchial epithelial cells[18,19] (Supplementary Fig. 1a). We observed that the majority of the virus-associated genes are unique for each virus and a small proportion is shared among them. Specifically, only three genes were differentially expressed during infection in cell lines regardless of the virus evaluated (Fig. 1a). Furthermore, GO enrichment analysis (Fig. 1b) shows that the top ten enriched GO terms are different for each virus, except "cellular response to lipopolysaccharide", shared between SARS-CoV-2 and SARS-CoV; however, the three viruses share terms related to immune response processes (Supplementary Fig. 1b). The latter shows that, despite their phylogenetic relationship, the main changes in gene expression driven by MERS-CoV, SARS-CoV-2, and SARS-CoV infection are divergent at both levels: at the DEGs and the cellular processes, suggesting that each virus uses specific molecular strategies during infection.

Subsequently, we inspected the DEGs with epigenetic or transcriptional regulatory function present among viruses, hereinafter referred to as epigenes. A comparative analysis of the DEGs among the three viruses revealed that only *INO80D*, a regulatory component of the chromatin remodeling INO80 complex, is shared among them. MERS-CoV and SARS-CoV only share the histone deacetylase *HDAC9*; while MERS-CoV and SARS-CoV-2 share *DUSP1*, *KDM6B*, *CHD2*, and *GADD45A*. Between SARS-CoV and SARS-CoV-2, we found *PBK MYSM1*, *ZNF711*, and *PCGF5* (Fig. 1c, Supplementary Fig. 1c). In addition, given that TFs are also key elements in gene remodeling and regulation, we evaluated the ones differentially expressed across viruses and none of them was affected in all conditions. However, MERS-CoV and SARS-CoV share *ZNF484* and *CEBPD*; SARS-CoV and SARS-CoV-2 share *ZEB1*, *ZEBTB20*, *NR4A1*, and *FOXN2*, and between MERS-CoV and SARS-CoV-2 15 shared TFs were found, including *RELB*, *JUN*, *FOSB*, *E2F8*, among others (Fig. 1c, Supplementary Fig. 1d).

Furthermore, the analysis showed that the differentially expressed epifactors belong to a wide range of functional categories, such as histone writers, histone readers, histone erases, Polycomb group proteins, chromatin remodeling, DNA modifications, among others (Fig. 1d). In addition, regarding differentially expressed TFs, cell lines infected with SARS-CoV-2 show differential expression of TFs of the STAT (mediators of the cellular response to cytokine) and IRF (interferon-regulatory factor) family, which are not differentially expressed in MERS-CoV and SARS-CoV (Fig. 1e). We noted that most of the TFs that are differentially expressed and shared between two or more of the Coronaviruses infected cells are members of the Znf TF family (ZNF436, 448, 543, 597, 773, XSCAN12, ZEB1, ZDTB20, KLF10, and HIVEP1) bHLH family (MXD1 and MXD4), involved in CCAAT/Enhancer Binding Protein (C/EBP) (DDIT3 and CEPPD), NF-κB complex (RELB), AP-1 complex (FOSB, JUN), ETS family (SPDEF) and E2F TF, among others (Supplementary Fig. 1c).

Since the repeated elements contained in the genome and their expression can also be regulated by epigenetic components, we evaluated the changes in gene expression of repeat elements after viral infection (Supplementary Table 3). We found that 47, 22, and 319 repeat elements are differentially expressed in SARS-CoV-2, SARS-CoV, and MERS-CoV infected cell lines, respectively. In SARS-CoV-2, the repeat elements belong predominantly to the long interspersed nuclear elements (LINE; 13 elements), long terminal repeat (LTR; 17 elements), and DNA repeat elements (17 elements) families. Similarly, for SARS-CoV and MERS-CoV we found LINE (5 and 59 elements), LTR (10 and 179 elements), and DNA repeat (5 and 65 elements). Interestingly, short interspersed nuclear elements (SINE) elements are not differentially expressed (only three elements found in MERS-CoV) and few Satellite elements were identified (3, 1, and 10 in SARS-CoV-2, SARS-CoV, and MERS-CoV, respectively) The elements L1MA4:L1:LINE, L1PA8A:L1:LINE and LTR54:ERV1:LTR are shared among all viruses. Notably, the L1 or LINE-1 elements are the only autonomous transposons that remain active in the human genome and are mainly repressed by epigenetic mechanisms such as HPTMs (H3K9me3)[20]. The latter, along with the fact that SARS-CoV-2 infected cells overexpress the histone demethylases KDM7A and KDM6B that target the heterochromatin histone marks such as H3K9me and H3K27me[21] (Supplementary Fig. 1c), suggest that SARS-CoV-2 infection could promote an open chromatin conformation, thus affecting the transcriptional expression and the derepression of the repeated sequences.

SARS-CoV-2 transcriptional effect in COVID-19 patient-derived samples

Afterward, we evaluated the transcriptional response in patient samples infected with SARS-CoV-2 to assess their resemblance to the previously observed results in cell lines. For this purpose, we obtained datasets from samples of bronchoalveolar lavage fluid (BALF) and lung. From the differential expression analysis, we found 389 DEGs shared among both samples (Fig. 2a). In this geneset, we identified 28 epigenes whose fold change direction was consistent in most of the cases (Fig. 2b); GO enrichment analysis shows that most of the DEGs were related to the immune response to viral infection such as leukocyte mediated immunity and humoral immune response (Fig. 2c). Furthermore, they were involved in a wide range of epigenetic processes, such as histone modification, chromatin remodeling, DNA modification, and TF (Fig. 2d). Following, we evaluated the similarity of these results

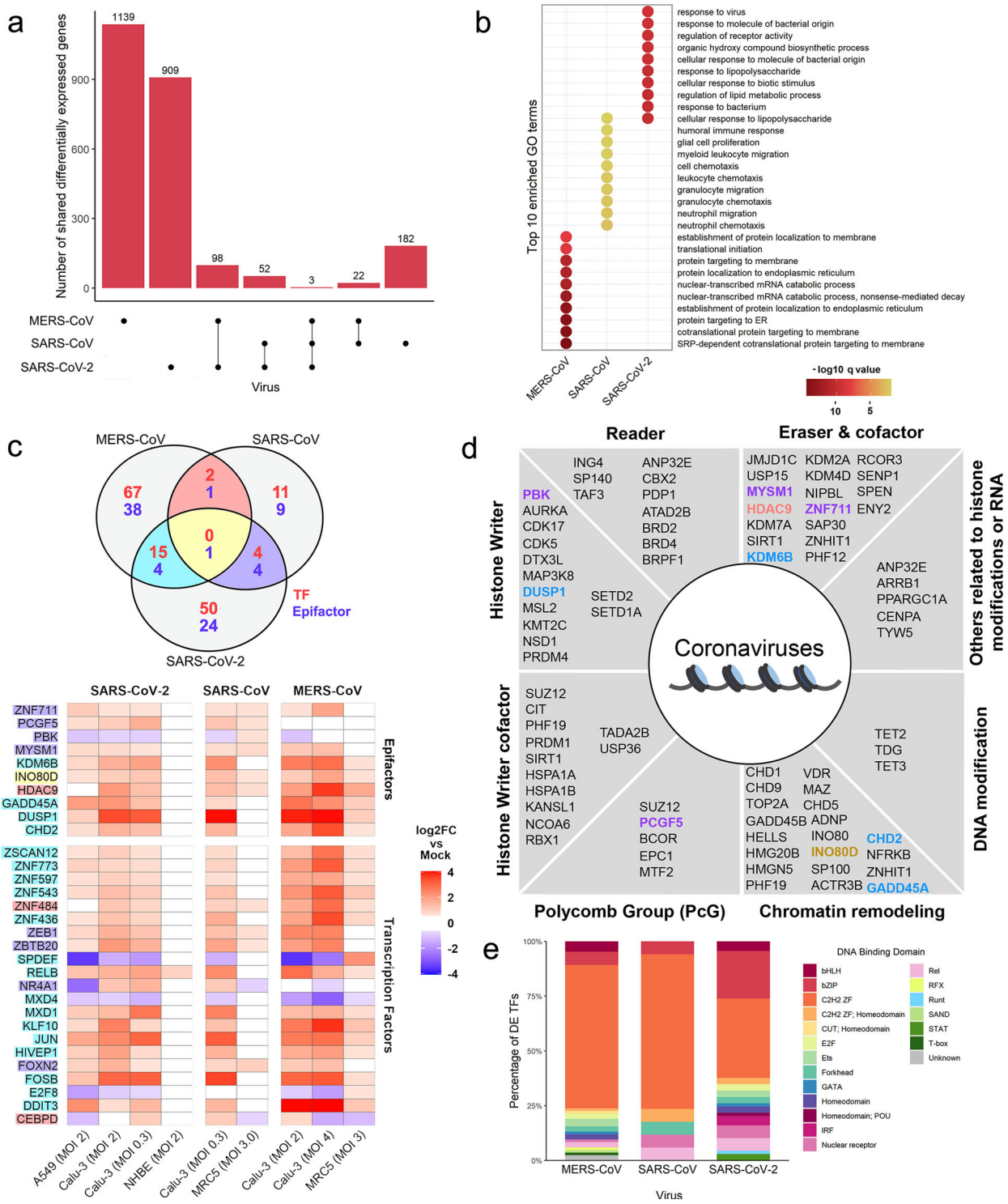


Fig. 1 Differential expression analysis of coronavirus-infected cell lines. a Intersection size of the DEGs common to each viral infection represented as single dots (virus-associated gene sets) and the size of their intersections with the other sets (multiple vertical dots). **b** Top ten simplified enriched Gene Ontology terms of biological process in the virus-associated gene sets ordered by *q*-value. **c** Shared differentially expressed epigenes between virus-associated gene sets; text color corresponds to the gene classification as either TF (red) or epifactor (blue) (upper panel). Log2 fold change of shared differentially expressed epifactors in each cell line are also shown as a heatmap (lower panel); blank color represents non-significant differential expression, text highlight corresponds to the intersections shown in the Venn diagram. **d** Functional classification of the identified epifactors; text color corresponds to the intersection color of subsection (**c**). **e** Characterization of the DNA-binding domain (DBDs) of human transcription factors (TFs) altered by the viral infection of coronaviruses.

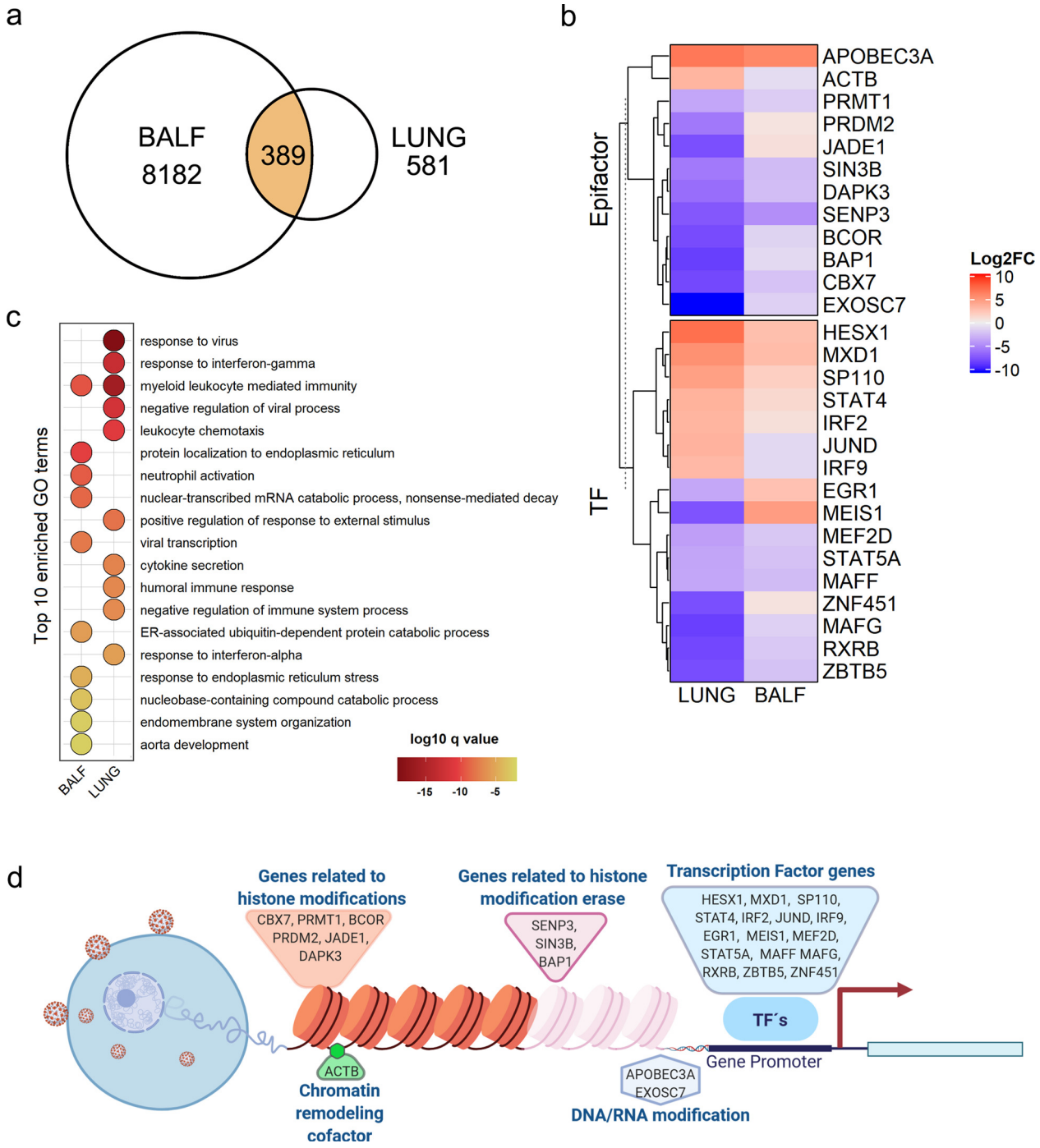


Fig. 2 Differential expression analysis of COVID-19 patient samples. a Number of shared differentially expressed genes between the samples. **b** Log2 fold change of shared differentially expressed epigenes in patients' samples. **c** Top ten simplified Gene Ontology enriched terms belonging to the biological process sub-ontology; ordered by *q*-value. **d** Epigenetic processes associated with the shared differentially expressed epigenes between patient samples. Created with BioRender.com.

with the data observed for SARS-CoV-2 in infected cell lines by comparing the overlap between the virus-associated genes with the DEGs present in the patient's samples (hereinafter referred as patient-DEGs, 389 genes). We found 46, 10, and 22 genes in common with SARS-CoV-2, SARS-CoV, and MERS-CoV respectively. In particular, for SARS-CoV-2 infected cell lines and patients, 5 TFs (*STAT5A, MAFF*, *IRF9*, *MXD1*, and *STAT4)* and no epifactors were identified. The shared DEG between samples of SARS-CoV-2 infected patients and MERS-CoV and SARS-CoV infected cell lines, which were found at a lesser extent, is likely to be non-specific viral-responding immune genes. Finally, we found that 804 and 20 repeat elements are differentially expressed in BALF and LUNG samples, respectively, being LTR elements the most differentially expressed in both samples (Supplementary Table 3). Collectively, these results show that the gene expression changes promoted by SARS-CoV-2 infection in patients are similar in respiratory tract samples, where immune response processes are the main ones affected.

SARS-CoV-2 and MERS-CoV infection induce different transcriptional fold changes in shared gene co-expression modules, which recapitulate the expression profiles in COVID-19 patient-derived samples

So far, our analyses showed that cell lines and patient samples infected with SARS-CoV-2 exhibited DEGs related to immunological processes, which has been previously described by Blanco-Melo et al.[22] and is congruent with our results. However, differential expression analysis often overlooks the subtle differences in several genes that all together can be responsible for major changes in global transcriptional regulation. Weighted gene co-expression network analysis overcomes this limitation by studying the expression of thousands of genes in the same analysis[23]. Thus, we expanded our previous results by including a co-expression analysis to identify gene modules associated with each viral infection and genes that play central roles within them.

We constructed the co-expression network with the log2 fold changes of each sample compared to its controls. After identifying the modules, we calculated the correlation between each module and the different traits, where we found that out of the 24 total modules identified, 13 were significantly correlated to the infection of any of the three viruses (Supplementary Fig. 2; Supplementary Table 4). Specific modules are associated with MERS-CoV (module 1), SARS-CoV (module 7), and SARS-CoV-2 (modules 9 and 10) (Fig. 3); shared modules were also identified. Notably, more than half of the modules (7 out of 13) are significantly associated with both SARS-CoV-2 and MERS-CoV; contrary to SARS-CoV-2 and SARS-CoV, which have only one module jointly associated. Remarkably, SARS-CoV-2 and MERS-CoV share a higher number of modules that are significantly associated with each virus. Even though the transcriptional profile is not the same, as it can be seen by the opposing response, the same genes from the shared modules are involved in both infections. Therefore, those infections share more players relevant to the infection than SARS-CoV-2 and SARS-CoV, as it would be expected.

Further analysis with GO enrichment analysis (Supplementary Fig. 3), shows that genes in modules 11 and 12 are involved in the host cell response to viral infection, while modules 4, 5, 7, 8, 10, and 13 were associated with intracellular processes used by the viruses during the infection, such as DNA replication, translation, ribosome biogenesis, and protein folding. Interestingly, module 6 was found related to epigenetic processes, particularly, transcriptional activation of promoters. These results show that, in addition to the immunological processes identified in our previous differential expression analysis, the transcriptional response to SARS-CoV-2 and MERS-CoV infection contain genes that also participate in RNA translation, DNA replication, and epigenetic regulation.

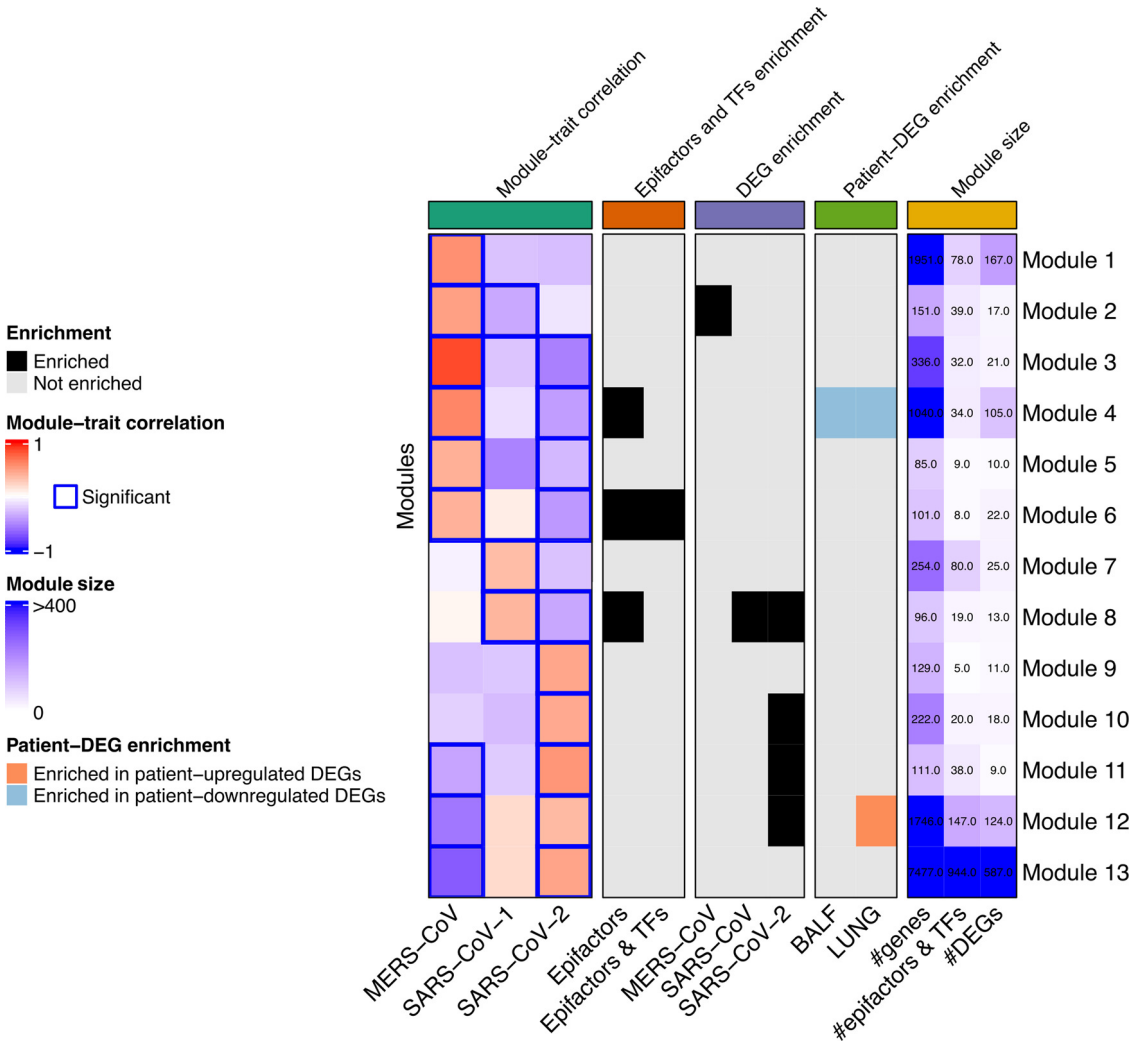


Fig. 3 Relevant modules for coronavirus infection. Summary of the analyses used to identify relevant modules for each infection. From left to right, grids show the module-trait correlation, the enrichment of epigenes, the enrichment of DEGs found in cell lines, enrichment of DEGs found in patients' samples, and information of the module size.

Following, in order to determine the modules that might be more relevant to SARS-CoV, MERS-CoV, or SARS-CoV-2 infection in terms of epigenetic regulation, we conducted enrichment analyses to identify an over-representation of epigenes, virus-associated gene sets, DEGs found in patients (Fig. 3).

When integrating these analyses, we found module 1 relevant for MERS-CoV and module 7 for SARS-CoV infection since they are exclusive for these viruses according to the co-expression analysis (Fig. 3). To determine the most important epigenes for each of these modules, we evaluated the eigengene-based connectivity (Module Membership, MM) to find hub genes and the gene significance (GS) of each gene. Subsequently, we prioritized them by identifying drugs targeting them. After looking for candidate drugs that targeted these central players, we found approved medication for CDK7 and PCNA for SARS-CoV and NCOA1, NR1H2, PRKAB2, CLOCK, KDM1B, and ATF2 (Supplementary Fig. 4).

Regarding SARS-CoV-2, module 9 was uniquely associated with it, but also other modules stood out. First, module 4 had a consistent behavior between cell lines and patients, since we found it negatively correlated to SARS-CoV-2 while being enriched in downregulated genes in BALF and LUNG samples; in addition, it was enriched in epifactors, suggesting an important role in viral-related epigenetic modifications carried out by these genes. A similar phenomenon is observed for module 12, which was positively correlated with SARS-CoV-2 infection and enriched with upregulated genes in patients and with SARS-CoV-2 DEGs. In addition, modules 10 and 11 were positively correlated to SARS-CoV-2 and enriched in SARS-CoV-2 DEGs, with module 11 being also enriched with patient's DEGs. Module 6 was negatively associated with SARS-CoV-2 in the co-expression network and enriched with epigenes, and module 8 was enriched with upregulated genes in PBMC and negatively associated with SARS-CoV-2 in the co-expression network while being enriched with epifactors and SARS-CoV-2-DEGs (Fig. 3). Finally, the enrichment of TFs targets in each module was evaluated to identify the ones that could explain the co-expression patterns of the genes within the module. With this analysis, it was found that Module 4 is enriched in the target genes of the transcriptional factors MTA1, MORC2, and RBM34 that belong to the same module (Supplementary Table 5), being MTA1 and RBMC34 differentially expressed in BALF samples. It is worth mentioning that in most of these modules epigenes showed a higher MM than the rest of the genes (Module 1: $W = 164{,}249$, $p < 0.05$, Module 2: $W = 1497$, $p < 0.05$, Module 4: $W = 51{,}768$, $p < 0.05$, Module 6: $W = 1359$, $p < 0.05$, Module 7: $W = 3741$, $p < 0.05$, Module 12: $W = 113{,}820$, $p < 0.05$, Module 13: $W = 2{,}182{,}049$, $p < 0.05$), evincing their central role within their modules.

Collectively, these results show that the transcriptional response to the infection of SARS-CoV-2 and MERS-CoV involve a higher similarity regarding gene modules but with a different extent of transcriptional change in host cells during infection, which extends our previous observations in the differential expression analysis. Therefore, the same genes in the shared modules play a potential role in both infections, despite presenting a different transcriptional behavior. Importantly, the virus-correlated co-expression modules either recapitulate the changes in gene expression observed in different COVID-19 patient sample types or are enriched with epifactors, and also contain genes involved in several biological processes related to viral infection, suggesting that the data obtained in the cell lines could recapitulate what was found in infected patients

Protein–protein interaction network analysis provides additional therapeutic alternatives and new targets for drug development for COVID-19

To prioritize epigenes that play a key function in each co-expression module relevant for SARS-CoV-2 (modules 4, 6, 8–12), we examined them at the protein–protein interaction (PPI) level in the context of SARS-CoV-2 infection. We constructed a PPI network containing all experimentally validated human protein interactions[24] and the reported virus–host protein interactions from Gordon et al. and Stukalov et al.[25,26]. Using the virus–human PPI network, we performed de novo pathway enrichment analysis with KeyPathwayMiner[27] to extract the largest network using a selection of epigenes as input, while also taking into account the SARS-CoV-2-DEGs and patient–DEGs previously identified. The selection of epigenes for each module was based on their shortest path length with viral proteins, their expression correlation with viral genes, and their MM.

All genes contained in the networks identified (Fig. 4) provide insights about the molecular machinery involved in SARS-CoV-2 infection, since the genes are either differentially expressed in infected cell lines or patients, or they are hub-epigenes in the co-expression analysis.

For module 4, the network obtained contains mainly epifactors. Notably, DNMT1 directly interacts with the viral protein ORF8 and with TRIM28, to which it is also highly co-expressed. Other relevant epigenes in the networks are SIRT6 (highly co-expressed and interactor of TRIM28), SENP3, MTA1 (a TF whose targets are also enriched in module 4; Supplementary Table V), and BAP1 (differentially expressed in patients). Furthermore, MEPCE, an snRNA methyl phosphate capping enzyme, is differentially expressed in patients and interacts with viral protein NSP8. Module 6 contains BRD4, which directly interacts with E viral protein and is highly co-expressed with EP300, a histone acetyltransferase. Another relevant epigene is SETD1B (related to the trimethylation of H3K4[3], a unique epigenetic histone mark related to transcriptional activation), which interacts with TRIM28, present in module 4[28]. For module 8, notable epigenes are CENPF, differentially expressed in SARS-CoV and SARS-CoV-2 and directly interacting with NSP13; and TOP2A, differentially expressed in SARS-CoV infected cell lines and patients. EP300, an exception node in module 9, interacts with several TFs such as NR2F2, HOXB9, NR3C2, and SOX9. In module 10, RELA (also known as nuclear factor NF-κB p65 subunit) and MOV10 are exception nodes that interact with the TFs SMAD3, ZNF277, and UBE2D3. Also, viral proteins ORF7B and ORF3 interact with FXYD2, STEAP1B, and TMEM156, which are differentially expressed in SARS-CoV-2 infected cell lines. For module 11, IRF7 (interferon regulatory factor 7) and STAT5A interact with EP300. Module 12, also contains epigenes of interest, such as MOV10 (Putative helicase MOV-10) that interacts with the N protein, TRIM25, RELA, and TLE1, which has a direct interaction with viral protein NSP13. Further genes which are classified as hub-epigenes and are also differentially expressed in cell lines or patients are FOS, CEBPD, NR4A1, PRDM1, PCGF5, ZNF652, IRF2, and ZEB2.

Briefly, we identified relevant TFs known to participate in Coronavirus infection and support the veracity of our results, such as TFs from the STAT family (STAT1, STAT2, and STAT5A), interferon regulatory factors (IRF7 and IRF2), cytokines (CCL3, CCL4, and IL1B), and FOS and JUND, members of the AP-1 complex[29]. However, we also identify important genes that appear to be drivers of SARS-CoV-2 infection; such as the epifactors MOV10 and EP300 and the TF RELA, since they are exception connectors (genes that do not belong to the specific module, but are important in the protein pathway found) in more than one module (modules 6 and 9–11), and belong to modules enriched in genes that participate in histone H3–K4 methylation and in the response to interferon-gamma. EP300 is a histone acetyltransferase that was also identified in SARS-CoV-2 infected cell lines[14]. Additionally, we found that MOV10, a putative helicase, also participates in SARS-CoV-2 infection. The TF RELA has been increasingly recognized as a crucial modulator of the response to SARS-CoV-2 infection[14,30] and is part of the NF-κB complex, along with RELB[31], which is differentially expressed in MERS-CoV and

SARS-CoV-2 infected cell lines (Supplementary Fig. 1D). TRIM25 is a ubiquitin ligase required for the production of INF-1 and is inhibited by the Nucleocapsid of SARS-CoV[32]. Finally, TRIM28 (also known as KAP1) has been shown to interfere with viral integration into host genome[33] and represses the expression of repeat elements of the LINE family, in particular L1NA4[21], which was previously identified as differentially expressed in cell lines infected with the three Coronaviruses.

Afterward, we evaluated whether the proteins in the networks had annotated drugs targeting them. We found drugs for 69 out of the total 260 proteins, being PLAU, RELA, NEK6, NR1H4, PTGS2, PRKDC, ESR1, NR3C2, TTK, TOP2A, ADRB2, HDAC4, TRIM25, STK10, RPS6KA5, and EP300 the ones with the most drugs identified (more than 20). A total of 799 drugs were found, where Erlotinib, Imatinib, Lapatinib, Sunitinib, S-adenosyl-L-homocysteine (SAH), Quercetin, Tandutinib, RAF-265, Pictilisib, Neratinib, and Fedratinib

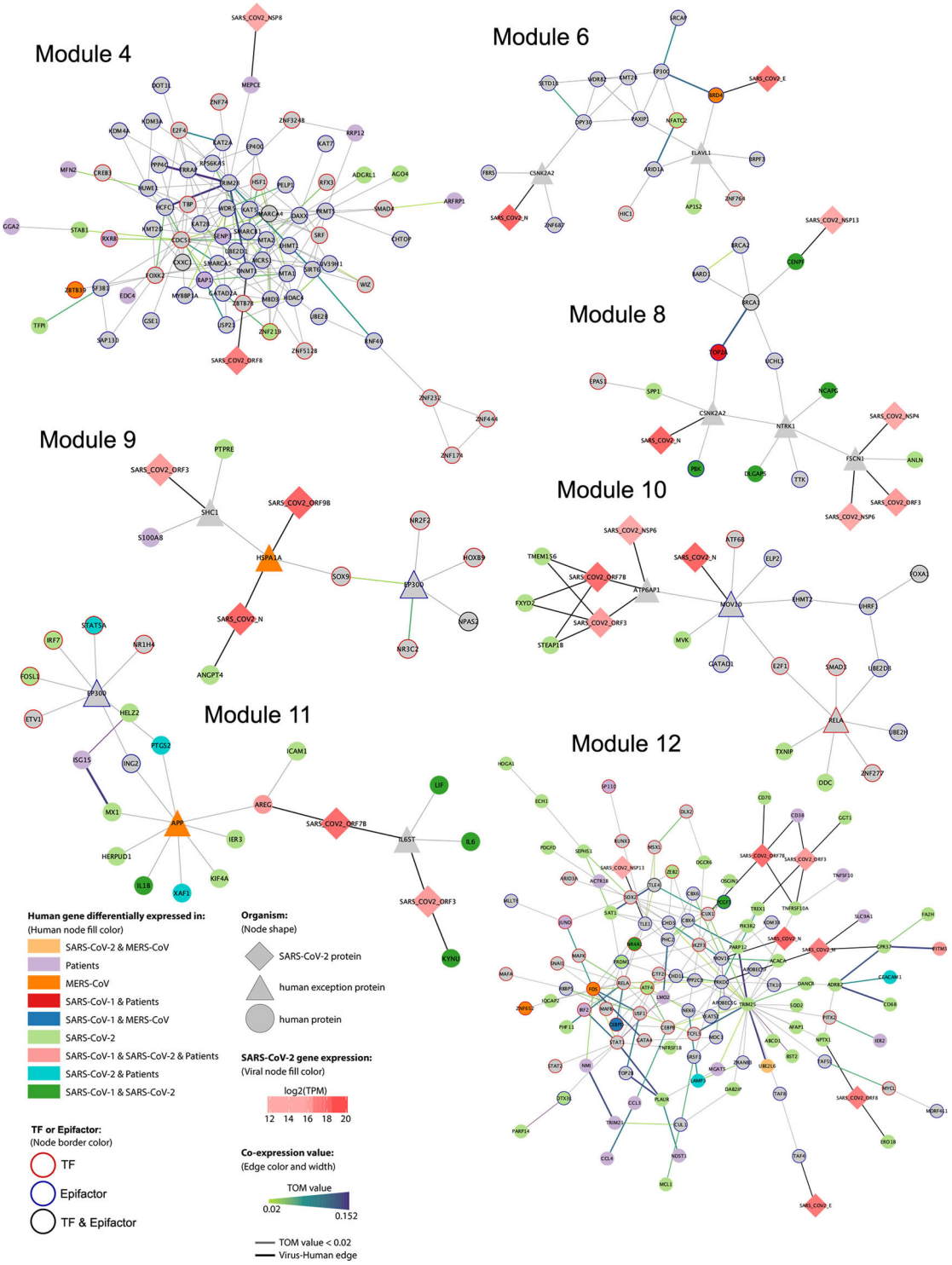


Fig. 4 Protein-protein interactions network containing SARS-CoV-2-DEGs, patient-DEGs, or selected epigenes for modules 4, 6, 8–12. Nodes and edges represent proteins and the interaction between them, respectively. The node and edge color's meaning is indicated in the annotation panel.

Table 1. Drugs targeting candidate epigenes from modules 4, 6, 8, and 10–12.

Target protein	Drug name	Module	Function
TOP2A	Genistein, fluorouracil, intoplicine, enoxacin, sparfloxacin, amrubicin, etoposide, epirubicin, ciprofloxacin, myricetin, mitoxantrone, trovafloxacin, RTA 744, daunorubicin, norfloxacin, finafloxacin, dexrazoxane, 13-deoxydoxorubicin, idarubicin, lomefloxacin, lucanthone, pefloxacin, valrubicin, amsacrine, levofloxacin, doxorubicin, declopramide, annamycin, banoxantrone, ZEN-012, podofilox, aldoxorubicin, teniposide, moxifloxacin, SP1049C, amonafide, dactinomycin, fleroxacin, becatecarin, ofloxacin, and elsamitrucin	Module 8	Epifactor (chromatin remodeling)
BRD4	Fedratinib, panobinostat, romidepsin, birabresib, alprazolam, vorinostat, volasertib, alobresib, belinostat, and apabetalone	Module 6	Epifactor (histone modification read)
EP300	Curcumin	Module 6	Epifactor (histone modification write)
DNMT1	S-adenosyl-L-homocysteine, procainamide, palifosfamide, cefalotin, decitabine, azacitidine, flucytosine, epigallocatechin gallate, and hydralazine	Module 4	Epifactor (DNA methylation)
SENP3	Methylphenidate	Module 4	Epifactor (histone modification erase, histone modification write cofactor)
SIRT6	7-[4-(Dimethylamino)Phenyl]-N-Hydroxy-4,6-Dimethyl-7-Oxo-2,4-Heptadienamide	Module 4	Epifactor (histone modification erase)
FOS	Pseudoephedrine and nadroparin	Module 12	TF
RELA	SC-236, betulinic acid, bortezomib, dimethyl fumarate, PHENYL-5-(1H-PYRAZOL-3-YL)-1,3-thiazole, indoprofen	Module 12	TF
STAT1	Epigallocatechin gallate	Module 12	TF
STAT5A	AZD-1480	Module 11	TF
SMAD3	Ellagic acid	Module 10	TF

are the drugs with more targets (more than 5; Supplementary Table 6). Relevant epigenes that have associated medication are shown in Table 1.

Most notably, RELA is targeted by SC-236, bortezomib, indoprofen (an anti-inflammatory) and betulinic acid, whose derivatives show anti-HIV activity[34,35]. EP300 is targeted by curcumin, a molecule with anti-inflammatory properties[36]. The latter proposes RELA and EP300 as new potential drug target candidates for SARS-CoV-2 infection, not only because they participate in immune-related processes, but also because they belong to the cellular epigenetic machinery used by the virus during infection. Furthermore, self-evident immune-related targets STAT5A, STAT1, and FOXN2 are also good candidates for treatment. Finally, the proteins MOV10, TRIM25, and TRIM28 do not have associated drugs, thus they are good candidates for drug development, as well as other relevant epigenes shown in Table 2.

Together, network analysis at the protein level allowed the identification of several epigenes that are part of the molecular machinery used by the virus during infection (Fig. 5). Epigenes that participate in the immune response through different mechanisms (response to interferon or NF-kB complex) are among the main genes identified and are evident drug target candidates for COVID-19 because they already have associated drugs targeting them (such as STAT5A and STAT1). Furthermore, new candidate druggable epigenes were also identified, notable examples are EP300 and RELA, which are targeted by drugs with anti-inflammatory or antiviral properties; and TRIM25, TRIM28, and MOV10, which are good candidates for drug development.

Table 2. Candidate epigenes for drug development in modules 4, 6, 8, and 10–12.

Target protein	Module	Function
MOV10	Module 12	Epifactor (chromatin remodeling)
MTA1	Module 4	Epifactor (chromatin remodeling cofactor)
TLE1	Module 12	Epifactor (chromatin remodeling, histone modification cofactor)
TAF4	Module 12	Epifactor (histone chaperone)
BAP1	Module 4	Epifactor (histone modification erase, polycomb group (PcG) protein)
TRIM28	Module 4	Epifactor (histone modification read)
PRDM1	Module 12	Epifactor (histone modification write cofactor)
SETD1B	Module 6	Epifactor (histone modification write)
UBE2D3	Module 10	Epifactor (histone modification write)
PCGF5	Module 12	Epifactor (polycomb group (PcG) protein)
STAT2	Module 12	TF
GMEB2	Module 6	TF
ZNF277	Module 10	TF
IRF7	Module 11	TF
CEBPD	Module 12	TF
NR4A1	Module 12	TF
ZNF652	Module 12	TF
IRF2	Module 12	TF
ZEB2	Module 12	TF
SP110	Module 12	TF
CUX1	Module 12	TF
TRIM25	Module 12	E3 ubiquitin ligase

DISCUSSION

Cells are in constant adaptation with their environment, in fact they can sense and respond to different stimuli by changing their transcriptional patterns. This cellular plasticity allows cells to adapt almost immediately to insults, including virus infections[37].

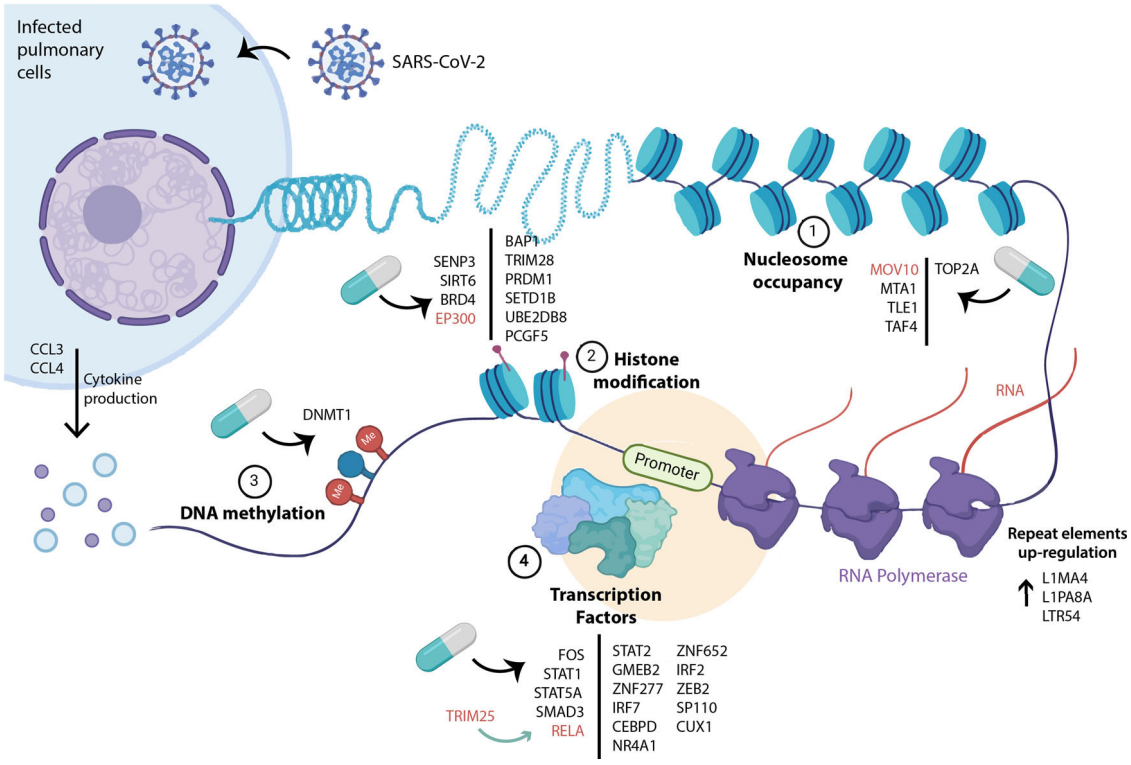


Fig. 5 Relevant epigenes in SARS-CoV-2 infection with therapeutic potential. Epigenetic targets are indicated in different processes such as nucleosome occupancy (1), histone modification (2), DNA methylation (3), and also TFs (4). Top gene candidate targets are highlighted in red. Created with BioRender.com.

Epigenetic proteins and TFs are one of the main elements involved in the transcriptional response of cells during viral infection. These elements can be used as protein targets for drug identification and treatment. In this work, we aimed to identify key TFs and proteins involved in the epigenetic response to viral infection of SARS-CoV-2, SARS-CoV, and MERS-CoV by integrating co-expression and de novo pathway enrichment analyses. Therefore, our study focused on the infection part of COVID-19, which is relevant mostly during the early stages of the disease, in contrast to the immune pathologies seen in the later ones.

One of our main findings is that the transcriptional response (regarding DEGs and significantly co-expressed modules) induced by SARS-CoV-2 and MERS-CoV involves a higher similarity regarding gene players and biological processes than SARS-CoV-2 and SARS-CoV, despite presenting a different transcriptional behavior. However, it is interesting to notice that regarding the transcriptional trend of the modules (i.e., correlation sign), SARS-CoV-2 and SARS-CoV behave more similarly despite many modules not being significantly associated with SARS-CoV. Nevertheless, unique modules, patterns and DEG were found in each CoV. Despite they belong to the coronavirus family, each one has unique characteristics that could influence its pathogenicity and virulence. This finding agrees with a recent study that has found specific biological process deregulations in SARS-CoV-2 infected cell lines, which are not found in other CoVs[15]. In addition, different transcriptional change patterns have been observed between MERS-CoV and SARS-CoV during the infection; these changes are not recapitulated by phylogenetic relationships since, in some groups of genes, MERS-CoV-infected transcriptional behavior appears to be more similar to the more remotely related influenza H5N1 virus infection[13].

Furthermore, the contrasting transcriptional response induced by the infection of SARS-CoV-2 and MERS-CoV in several modules suggests that genes in those modules participate in both viral infections but with a different mechanism, which leads to distinct pathways of infection that could explain the dissimilar phenotypes observed in both diseases. Divergent fold change trends, such as the ones described in this study, have been previously described in MERS-CoV and SARS-CoV infections to limit the host type I interferon (IFN-I) response, where predominant active and repressive epigenetic marks in involved genes are the opposite between both CoVs[13]. In our study, we present a list of epigenes and biological processes whose fold change trend is the opposite between MERS-CoV and SARS-CoV-2; further investigation on them could shed light on the mechanisms responsible for the differences in pathogenesis and outcome of both viral infections.

We further identify at the protein interaction level, that several TFs take part mainly in the immunological response to viral infection. One example is NF-κB, whose p65 subunit (also known as *RELA*) is a central part in the protein interaction network for SARS-CoV-2. NF-κB induces the expression of several pro-inflammatory cytokines, including IL-6, CCL2, and CCL3[38], which had been found in high levels in COVID19 patients[39]. On the other hand, TRIM25, a ubiquitinase, is essential for the activation of NF-κB and the production of IL-6[40]. TRIM25 is overexpressed in cell lines infected with SARS-CoV-2 but not in those with MERS-CoV, which furthermore suggests that NF-κB could be a medullary part of the host immune response against SARS-CoV-2. The previous observation is reinforced by the fact that it was observed that RELA directly interacts with histone acetyltransferase EP300, and both proteins interact with various components of the AP-1 complex such as FOS, JUND, and FOSL1. AP-1. EP300 and NF-κB regulate chromatin accessibility in the proximal promoter region of IL-6 and CCL2, both pro-inflammatory cytokines[41,42]. The p300/CBP complex is one of the best-characterized cofactors of NF-κB and specifically binds RELA and acetylates it along with the surrounding histones[31]. It is known that adults older than 65 years have higher NF-κB levels compared to younger adults[43] and some authors had suggested that this may be one reason older adults are more susceptible to develop the severe form of COVID-19[44].

According to our results, SARS-CoV-2 infection modifies the expression of several TFs of the interferon regulatory factor (IRF) and STAT families, which are primarily involved in the immune response against pathogens. STAT1 and STAT2 are key elements of the signaling induced by type I interferons, these proteins form a dimer upon interferon-mediated phosphorylation and, together with IRF9, form the complex ISGF3 that activates the transcription of interferon-stimulated genes[45]. Our results also showed that IRF9 is upregulated in cell lines infected with SARS-CoV-2; however, module 12's interactome showed that STAT2 and STAT1 interact with IRF2. IRF2 is a negative regulator of IFNα and its inhibition causes an increase in the antiviral response induced by IFNα[46]. This fact further suggests an impairment of interferon type I stimulated genes activation, as previously described as a hallmark of SAR-CoV-2 infection[47]. On the other hand, IRF1 and IRF7 were also upregulated in SARS-CoV-2 infected cell lines. IRF7 is a key TF for IFNα expression, and it has been previously identified as a hub gene for SARS-CoV-2 infection together with IFR9 and STAT1[48]. It is also interesting that IRF7 loss of function mutations was associated with severe COVID-19 patients[49] and with the development of life-threatening influenza in children[50] which suggests that inhibition of IRF7 activity is crucial for SARS-CoV-2 pathology.

Viruses have been reported to use epigenetic machinery to take advantage of the cell and hijack its regulatory capacity for their own benefit[13]. The epigenetic machinery can be affected by coronaviruses in this same sense, and this can happen either by promoting alterations in the epigenetic code, such as DNA methylation and post-translational modifications of histones, or directly by promoting the dysregulation of enzymes and other proteins associated with the epigenome.

We found that among the deregulated epifactors with histone acetylation function are HDAC9 and SIRT1 enzymes. In this sense, it has recently been reported that the SIRT1 protein (a class 3 HDAC) was positively regulated in the lung of patients with severe COVID-19 comorbidities[51]. Likewise, another work demonstrated that under conditions of cellular energy stress, SIRT1 can epigenetically regulate the ACE2 receptor[52]. Also, it has been observed that treatment with non-steroidal anti-inflammatory drugs can inhibit SIRT1 activity, which in turn could affect ACE2 expression[53]. Accordingly, it has been postulated that in some diseases where the epigenetic dysregulation is implicit (such as lupus) the entrance of SARS-CoV-2 into the host cells may be facilitated[54].

Interestingly, the enzymes HAT1, HDAC2, and KDM5B have been reported to also potentially regulate ACE2 in human lungs. KDM5B has gained interest, because it is associated with other viral infections such as the hepatitis B virus[55], and potentially with SARS-CoV-2[51]. Remarkably, in breast cancer cells, it has been shown that inhibition of this enzyme triggers a robust interferon response that results in resistance to infection by DNA and RNA viruses[56]. In this regard, we observed several deregulated KDMs in the different coronavirus infections, in which KDM6B stands out by being deregulated in both MERS-CoV and SARS-CoV2 infection. KDM6B is a specific demethylase of H3K27me3, which acts as a repressive histone mark. Although it remains to be fully studied, it is associated with the regulation of a wide range of genes involved in inflammatory agents, development, cancer, viral infection response, senescence and is an important host response against environmental, cellular stress[57]. Therefore, adding to the above, it is suggested that demethylases, such as KDM6B, are potential epigenes that are affected during SARS-CoV-2 infection and can be presented as potential targets for the treatment of COVID-19. However, this should be further studied.

Several epigenes previously involved in response to viral infections stood out in our protein interaction analysis, such as BRD4, TOP2A, and TRIM28. Bromodomain protein 4 (BRD4) is a histone acetylation reader and writer that plays an important role in DNA replication, transcription, and DNA repair[58]. This epigene is critical for the maintenance of the higher-order chromatin structure since its inhibition leads to chromatin decondensation and fragmentation, and it also can stimulate innate antiviral immunity[58]. BRD4 complexes with RELA and CDK9 and are functionally required for effective activation of NF-kB-dependent immediate-early cytokine genes in response to viral patterns. In this sense, our results show a PPI with EP300, which involves the p300/CBP complex, one of the best-characterized cofactors of NF-kB and binds specifically to RELA[59], validating the possible importance of this system in infection with SARS-CoV2. Examples like this suggest that the virus, through these epigenetic remodelers, promotes chromatin remodeling that could lead to opening, both at the local and global level. Accordingly, an indicator of global changes is the increased expression of transcripts from repeated sequences such as LINE1. If this is so, then the virus is manipulating the chromatin aperture to promote the expression of genes that support its invasion. In this regard, other work has suggested the importance of LINE1 elements. Where these types of repetitive elements are very relevant in gene regulation, especially when these elements are in proximity to neighboring genes since they could alter their expression. Therefore, the dysregulation of repeated elements such as LINE1 could indirectly change the cellular transcriptome[60].

Furthermore, we find epigenes that interact with the viral proteins directly or very closely. This connection suggests a virus-promoted modulation to affect the epigenome of the host cell's interactome. Which reinforces the idea that the virus strategy is partly to take advantage of the epigenetic machinery. In general, our data suggest that the SARS-CoV-2 infection deregulates the epigenetic master machinery of the host cell. One of the points that should be taken into consideration in the future is that if this epigenetic machinery is not re-established after disease courses it could generate other diseases such as cancer in the long term. This is based on the fact that many of the genes that we found in our study have been proposed as epigenetic hallmarks in various neoplasms.

Our last key finding is the identification of driver epigenetic proteins and TFs involved in SARS-CoV-2 infection that can be targeted by existing drugs. We identified SAH targeting several epigenetic components of the host response to SARS-CoV-2 infection. SAH is the product of the chemical reaction performed by methyltransferases using nucleic acids or proteins as substrates and has been previously suggested as a potential treatment for viral infections such as ZIKA, MERS-CoV, and SARS-CoV[61–64] due to its inhibitory activity of the viral RNA cap 2′-O-methyltransferase, formed by the NSP16-NSP10 complex[65,66]. Furthermore, given the interaction between DNMT1 and ORF8 at the protein level, SAH could potentially work against SARS-CoV-2 infection, not only by inhibiting the methyltransferase activity of NSP16–NSP10 but also by directly modulating the activity of the key host proteins involved in the transcriptional response to infection or by interfering with the interactions observed between ORF8 and DNMT1.

Furthermore, as anticipated, many proteins with epigenetic functions involved in SARS-CoV-2 infection have kinase activity and can be targeted by kinase inhibitors. One important example is imatinib, which we identified as a potential drug for SARS-CoV-2 and SARS-CoV, and is currently undergoing clinical trials to evaluate its efficacy in COVID-19 patients (NCT04394416, NCT04422678, NCT04346147, and NCT04357613; www.clinicaltrials.gov). Similarly, we found quercetin targeting several epifactors with kinase activity. Quercetin is a plant-derived compound with anti-inflammatory and antiviral effects[67,68] that has been evaluated in clinical trials as a dietary supplement or prophylaxis for COVID-19 (NCT04578158, NCT04377789, and NCT0446813). Even though some independent studies show no clear evidence of its effectiveness, preliminary data shows that it

could be effective to decrease the frequency and duration of respiratory tract infections[69–71]. It is worth mentioning that these drugs are being tested in clinical trials based on their described inhibitory activity of enzymes related to the activation of immune response and inflammation, such as growth receptors[72]. The latter, together with our results, suggests that drugs targeting epigenetic mechanisms could be also effective to treat SARS-CoV-2 by modulating their kinase activity.

Finally, we also identified Bortezomib and betulinic acid associated with RELA. Bortezomib is a proteasome inhibitor that has been proposed as COVID-19 therapy given its capacity to inhibit (although only marginally) the papain-like protease (NSP3) of SARS-CoV, which also has deubiquitinase activity[73–75]. Likewise, betulinic acid has been proposed as a target of NSP3 in SARS-CoV-2[76].

Together, we have supporting evidence that current drug-based therapies to treat COVID-19 also target the transcriptional response to infection by the modulation of the epigenetic proteins identified in this study. Furthermore, we provide additional new potential drug targets and drug candidates which could be effective and whose potential use has not been exploited yet. These results provide comprehensive evidence that epigenetic therapy could aid in restoring the transcriptional changes observed during infection. By using epigenetic drugs, a therapeutic effect can be achieved due to their systemic effects, which can be advantageous to treat a disease that targets different tissues and cellular mechanisms, as observed in COVID-19.

In this study, we used a blend of bioinformatic approaches to comparatively analyze transcriptomic data from SARS-CoV-2, SARS-CoV, and MERS-CoV infected pulmonary cell lines and COVID-19 patient-derived samples. In particular, we focused on the epigenetic processes and transcriptional factors, since these have been widely proposed as the master regulators of the expression of most genes. We found that the transcriptional response to the infection of SARS-CoV-2 and MERS-CoV is more similar to that observed for SARS-CoV regarding shared significantly associated gene modules; however, the transcriptional change elicited by MERS-CoV and SARS-CoV seems to be opposite. At the same time, we identified specific altered modules in the response to infection with SARS-CoV2 that could serve as a guide for the proposal of different therapeutic strategies based on epigenetic therapy. Thus, our results add a piece to the puzzle of the strategies used by the different coronaviruses to manipulate the gene regulation capacity of the cell. Although the pathways are differential between them, the virus's objective is to take advantage of the TFs and various chromatin remodelers to avoid being detected and prevail in the invasion. This is a very fine strategy that the virus uses and it has been poorly studied in both its biological importance and its future therapeutic application. This could open a new window of opportunities for treatment and thus close the chapter on this pandemic disease.

METHODS

Data processing and differential expression analysis

Raw sequencing data was trimmed with Trimmomatic version 0.39[77] using the parameters ILLUMINACLIP 2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36; and the quality of reads was evaluated with FastQC version 0.11.9[78]. Technical replicates (when existing) were merged and each biological replicate was aligned to the GRCh38 v33 human genome with STAR version 2.7.3[79] using the mapping parameters suggested in Jin et al.[80]: (—outFilterMultimapNmax 100—winAnchorMultimapNmax 100). To estimate the abundance of the transcripts accounting for coding and non-coding genes as well as repetitive elements, we used TETranscripts version 2.1.4[80] with the multi-mode. Raw count tables were used for differential expression analysis using DESeq2[81]. DEGs were identified with a p adj. < 0.05 and abs($\log_2$ fold change) > $\log_2$(1.5).

Viral transcripts quantification

The viral transcriptome was constructed with the 11 gene sequences reported in the SARS-CoV-2 genome (NCBI Reference Sequence NC_045512.2). Viral transcript expression was quantified in each trimmed RNA-seq file of SARS-CoV-2 infected samples with Salmon v 1.3.0[82].

Virus and patient DEGs

Virus-associated gene sets were obtained with the intersection of DEGs identified in all the cell lines infected with the corresponding virus, except for SARS-CoV-2. For SARS-CoV, the intersection between the cell lines infected consisted of 182 genes (SARS-CoV-DEGs); for MERS-CoV, the intersection was 1139 genes (MERS-CoV-DEGs); and for SARS-CoV-2, the intersection between at least 3 out of the 4 cell lines was used instead, and consisted in 909 genes (SARS-CoV-2-DEGs) (Supplementary Table 2). The patient-associated gene set was obtained with the shared DEGs in lung and BALF conditions (389 genes, patient-DEGs) (Supplementary Table 2).

Epigenes catalog

To build the Epigenes catalog, 4 different databases were used: EpiFactors[83], Histome[84], dbEM[85], and the manually curated TF list from Lambert et al.[86]. TFs' functional annotation was taken from Lambert et al.[86]. The final list consisted of 2161 genes (776 epifactors, 1348 TFs, and 41 categorized as both TF and epifactor).

Co-expression analysis

Count matrices of the analyzed cell lines were filtered to remove low-expressed genes using the function filterByExpr from edgeR[87] while accounting for the treatment (i.e., virus infection) and cell type in the filtering design. Following, normalization of gene counts was performed with vst function from DESeq2[81] (treatment and cell type of each sample were included in the design matrix and accounted for these effects with the blind argument). The gene co-expression network was built with the $\log_2$ fold changes ($\log_2$FC) of each biological sample compared with the controls of the same biological condition by applying the formula (1).

$$\log_2 FC_i = \log_2(SC_i/ACC_i) \quad (1)$$

where SC and ACC correspond to the normalized counts of gene *i* in the infected and controls samples, respectively. The resulting matrix containing the log2FoldChanges per sample was used to construct the weighted gene co-expression network with the WGCNA package[23]. A soft threshold of 9 was used to construct the network and modules were identified with a minimum size of 20. Modules whose expression was similar were merged using a dissimilarity threshold of 0.25, resulting in a total of 24 modules. Finally, the module-eigengene Pearson correlation of each module with the viruses was tested.

Enrichment analysis

Gene Ontology (GO) enrichment analyses were performed using clusterProfiler[88] in virus-associated and patient gene sets. For the differential expression analyses of infected cell lines, the enrichment of GO terms in DEGs was tested using the expressed genes on each particular comparison as background. For the co-expression network, the enrichment of GO terms was tested in each module using the genes of the full network as background.

Epigenes, virus-associated DEGs, and TF-target enrichment analyses were performed with gProfiler2[89] using a custom gmt file or the TRANSFAC database included in the package for TF-target enrichment. The correction method used was g:SCS and an adjusted *p* value significance threshold of 0.05. As background, all the genes annotated in the co-expression network were used for epigenes and TF-target enrichment and the expressed genes in each virus for virus-associated DEG enrichment.

Co-expression module selection

SARS-CoV-2 modules were selected from the co-expression analysis based on whether they were uniquely and significantly associated with SARS-COV-2 in the co-expression analysis. If they were not uniquely associated with SARS-CoV-2, the modules enriched with at least one dataset (DEG, patient–DEG, or Epigenes) were selected. Based on these criteria, modules 4, 6, and 8–12 were selected. MERS-CoV and SARS-CoV modules were selected on whether they were uniquely associated with each specific virus in the co-expression analysis. Module 1 was selected for MERS-CoV and

module 7 for SARS-CoV. SARS-CoV-2 selected modules were further analyzed, as described in the following sections.

Virus–host network construction

Virus–human interactions were obtained from Gordon et al.[25] and Stukalov et al.[26]. The human PPI network was obtained from IID version 2018-11[24] using only the experimentally validated interactions ("exp", "exp;ortho", "exp;ortho;pred", or "exp;pred"). After homogenizing the viral protein nomenclature, the three sources of interactions were merged to create the entire virus-human PPI, followed by the removal of duplicated edges and self-loops. The final integrated network contained 30 viral nodes, 17,524 human nodes and 329,054 edges. The mapping of viral transcript counts to viral proteins in the PPI was based on the reference sequence annotation (NCBI Reference Sequence NC_045512.2) and the data provided in Supplementary Data from Gordon et al.[25].

Epigene selection

For co-expression modules 4, 6, 8, and 10–12, relevant epigenes were selected based on whether they satisfied at least one of the following criteria: (1) its shortest path length with viral proteins, (2) the correlation value between its expression and the expression of viral proteins, and (3) its MM value, a measure of the correlation between a gene expression profile and the module eigengene, which is highly related to the intramodular connectivity, and GS the correlation of a gene with an external trait (viral infection)[90].

(1) The shortest path length was calculated between all pairs of viral proteins and human proteins in the PPI network with the igraph package version 1.0.0[91]. The retained epigenes were the ones whose shortest path length with at least one viral protein was less than 3.
(2) Pearson's correlation coefficient was computed between the count values of viral transcripts and count values of epigenes in infected cell lines. Epigenes with p value < 0.05 and abs(correlation_estimate) > 0.5 with at least one viral transcript were selected.
(3) Epigenes with abs(MM) > 0.8 in the corresponding module of the co-expression network were retained.

For modules 1 and 7, epigenes with abs(MM) > 0.8 and abs(GS) > 0.3 were selected.

De novo pathway enrichment

De novo pathway enrichment analysis for co-expression modules 4, 6, 8, 10–12 was performed with KeyPathwayMiner[27], the built virus–human PPI network, the full list of viral proteins as positive nodes and a customized input indicator matrix for each module containing as active genes those which belonged to any of the following categories: (1) it was a SARS-CoV-2-DEG, (2) it was a patient-DEG, or (3) it was an epigene selected as described above. The parameters used for all the analyses were the Greedy search algorithm, INES search strategy, remove border exception nodes, $L = 0$, and $K = 0$ for modules 4 and 12, $K = 2$ for module 6, and $K = 3$ for modules 8–11.

Drug identification

All approved and non-approved drugs targeting the genes/proteins contained in each network were obtained with CoVex[92] by mapping the gene names to uniprot IDs, using the closeness centrality algorithm and the following parameters: result size = 50,000, disabled hub penalty, disabled max degree, include indirect drugs = FALSE and include non-approved drugs = TRUE. The latter parameters ensure the retrieval of all drugs associated with the input genes. A total of 265 out of 277 genes mapped to the CoVex database.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Raw RNA-seq data was obtained from the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) of the National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine, and the Genome Sequence Archive in BIG Data Center (bigd.big.ac.cn/), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (Supplementary Table 1).

Received: 9 November 2020; Accepted: 8 March 2021;
Published online: 24 May 2021

REFERENCES

1. Payne, S. Chapter 17—family coronaviridae. in *Viruses* (ed Payne, S.) 149–158 (Academic Press, 2017).
2. Memish, Z. A., Perlman, S., Van Kerkhove, M. D. & Zumla, A. Middle east respiratory syndrome. *Lancet* **395**, 1063–1077 (2020).
3. Hui, D. S. C. & Zumla, A. Severe acute respiratory syndrome: historical, epidemiologic, and clinical features. *Infect. Dis. Clin. North Am.* **33**, 869–889 (2019).
4. Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J. & Hsueh, P.-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents* **55**, 105924 (2020).
5. World Health Organization. Coronavirus disease (COVID-19) pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (2020).
6. Youssef, N., Budd, A. & Bielawski, J. P. Introduction to genome biology and diversity. *Methods Mol. Biol.* **1910**, 3–31 (2019).
7. Marazzi, I. et al. Suppression of the antiviral response by an influenza histone mimic. *Nature* **483**, 428–433 (2012).
8. Flanagan, J. M. Host epigenetic modifications by oncogenic viruses. *Br. J. Cancer* **96**, 183–188 (2007).
9. Menachery, V. D. et al. MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc. Natl Acad. Sci. USA* **115**, E1012–E1021 (2018).
10. Marcos-Villar, L., Pazo, A. & Nieto, A. Influenza virus and chromatin: role of the CHD1 chromatin remodeler in the virus life cycle. *J. Virol.* **90**, 3694–3707 (2016).
11. Marcos-Villar, L. et al. Epigenetic control of influenza virus: role of H3K79 methylation in interferon-induced antiviral response. *Sci. Rep.* **8**, 1230 (2018).
12. O'Connor, C. M., DiMaggio, P. A. Jr, Shenk, T. & Garcia, B. A. Quantitative proteomic discovery of dynamic epigenome changes that control human cytomegalovirus (HCMV) infection. *Mol. Cell. Proteom.* **13**, 2399–2410 (2014).
13. Menachery, V. D. et al. Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses. *MBio* **5**, e01174–14 (2014).
14. Ochsner, S. A., Pillich, R. T. & McKenna, N. J. Consensus transcriptional regulatory networks of coronavirus-infected human cells. *Sci. Data* **7**, 314 (2020).
15. Singh, K. et al. Network analysis and transcriptome profiling identify autophagic and mitochondrial dysfunctions in SARS-CoV-2 Infection. *bioRxiv*. https://doi.org/10.1101/2020.05.13.092536 (2020).
16. Ganesan, A., Arimondo, P. B., Rots, M. G., Jeronimo, C. & Berdasco, M. The timeline of epigenetic drug discovery: from reality to dreams. *Clin. Epigenetics* **11**, 174 (2019).
17. WHO Solidarity Trial Consortium. Repurposed antiviral drugs for Covid-19—interim WHO solidarity trial results. *N. Engl. J. Med.* **384**, 497–511 (2021).
18. Rayner, R. E., Makena, P., Prasad, G. L. & Cormet-Boyaka, E. Optimization of normal human bronchial epithelial (NHBE) cell 3D cultures for in vitro lung model studies. *Sci. Rep.* **9**, 500 (2019).
19. Davis, A. S. et al. Validation of normal human bronchial epithelial cells as a model for influenza A infections in human distal trachea. *J. Histochem. Cytochem.* **63**, 312–328 (2015).
20. Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 (2014).
21. Castro-Diaz, N. et al. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
22. Blanco-Melo, D. et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036–1045.e9 (2020).
23. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
24. Kotlyar, M., Pastrello, C., Malik, Z. & Jurisica, I. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.* **47**, D581–D589 (2019).
25. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
26. Stukalov, A. et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* https://doi.org/10.1038/s41586-021-03493-4 (2021).
27. Alcaraz, N. et al. Robust de novo pathway enrichment with KeyPathwayMiner 5. *F1000Research* **5**, 1531 (2016).

28. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. 3rd SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).
29. Hess, J., Angel, P. & Schorpp-Kistner, M. AP-1 subunits: quarrel and harmony among siblings. *J. Cell Sci.* **117**, 5965–5973 (2004).
30. Fagone, P. et al. Transcriptional landscape of SARS-CoV-2 infection dismantles pathogenic pathways activated by the virus, proposes unique sex-specific differences and predicts tailored therapeutic strategies. *Autoimmun. Rev.* **19**, 102571 (2020).
31. Bhatt, D. & Ghosh, S. Regulation of the NF-κB-mediated transcription of inflammatory genes. *Front. Immunol.* **5**, 71 (2014).
32. Hu, Y. et al. The severe acute respiratory syndrome coronavirus nucleocapsid inhibits type I interferon production by interfering with TRIM25-mediated RIG-I ubiquitination. *J. Virol.* **91**, e02143–16 (2017).
33. Fehervari, Z. Putting a KAP on infection. *Nat. Immunol.* **12**, 816 (2011).
34. Pavlova, N. I., Savinova, O. V., Nikolaeva, S. N., Boreko, E. I. & Flekhter, O. B. Antiviral activity of betulin, betulinic and betulonic acids against some enveloped and non-enveloped viruses. *Fitoterapia* **74**, 489–492 (2003).
35. Aiken, C. & Chen, C. H. Betulinic acid derivatives as HIV-1 antivirals. *Trends Mol. Med.* **11**, 31–36 (2005).
36. Gupta, S. C., Patchva, S. & Aggarwal, B. B. Therapeutic roles of curcumin: lessons learned from clinical trials. *AAPS J.* **15**, 195–218 (2013).
37. Bollati, V. & Baccarelli, A. Environmental epigenetics. *Heredity* **105**, 105–112 (2010).
38. Beacon, T. H., Su, R.-C., Lakowski, T. M., Delcuve, G. P. & Davie, J. R. SARS-CoV-2 multifaceted interaction with the human host. Part II: Innate immunity response, immunopathology, and epigenetics. *IUBMB Life* **72**, 2331–2354 (2020).
39. Chu, H. et al. Comparative replication and immune activation profiles of SARS-CoV-2 and SARS-CoV in human lungs: an ex vivo study with implications for the pathogenesis of COVID-19. *Clin. Infect. Dis.* **71**, 1400–1409 (2020).
40. Liu, Y. et al. TRIM25 promotes TNF-α–induced NF-κB activation through potentiating the K63-linked ubiquitination of TRAF2. *J. Immunol.* **204**, 1499–1507 (2020).
41. Wolter, S. et al. c-Jun controls histone modifications, NF-κB recruitment, and RNA Polymerase II Function To Activate the ccl2 Gene. *Mol. Cell. Biol.* **28**, 4407–4423 (2008).
42. Ndlovu, M. N. et al. Hyperactivated NF-{kappa}B and AP-1 transcription factors promote highly accessible chromatin and constitutive transcription across the interleukin-6 gene promoter in metastatic breast cancer cells. *Mol. Cell. Biol.* **29**, 5488–5504 (2009).
43. Bektas, A. et al. Age-associated changes in basal NF-κB function in human CD4+ T lymphocytes via dysregulation of PI3 kinase. *Aging* **6**, 957–974 (2014).
44. Do, L. A. H., Anderson, J., Mulholland, E. K. & Licciardi, P. V. Can data from paediatric cohorts solve the COVID-19 puzzle? *PLoS Pathog.* **16**, e1008798 (2020).
45. Martinez-Moczygemba, M., Gutch, M. J., French, D. L. & Reich, N. C. Distinct STAT structure promotes interaction of STAT2 with the p48 subunit of the interferon-α-stimulated transcription factor ISGF3. *J. Biol. Chem.* **272**, 20070–20076 (1997).
46. Robichon, K. et al. Identification of interleukin1β as an amplifier of interferon alpha-induced antiviral responses. *PLoS Pathog.* **16**, e1008461 (2020).
47. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
48. Prasad, K. et al. Targeting hub genes and pathways of innate immune response in COVID-19: a network biology perspective. *Int. J. Biol. Macromol.* **163**, 1–8 (2020).
49. Zhang, Q. et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).
50. Ciancanelli, M. J. et al. Infectious disease. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science* **348**, 448–453 (2015).
51. Pinto, B. G. G. et al. ACE2 expression is increased in the lungs of patients with comorbidities associated with severe COVID-19. *J. Infect. Dis.* **222**, 556–563 (2020).
52. Clarke, N. E., Belyaev, N. D., Lambert, D. W. & Turner, A. J. Epigenetic regulation of angiotensin-converting enzyme 2 (ACE2) by SIRT1 under conditions of cell energy stress. *Clin. Sci.* **126**, 507–516 (2014).
53. Dell'Omo, G. et al. Inhibition of SIRT1 deacetylase and p53 activation uncouples the anti-inflammatory and chemopreventive actions of NSAIDs. *Br. J. Cancer* **120**, 537–546 (2019).
54. Sawalha, A. H., Zhao, M., Coit, P. & Lu, Q. Epigenetic dysregulation of ACE2 and interferon-regulated genes might suggest increased COVID-19 susceptibility and severity in lupus patients. *Clin. Immunol.* **215**, 108410 (2020).
55. Wang, X. et al. Hepatitis B virus X protein induces hepatic stem cell-like features in hepatocellular carcinoma by activating KDM5B. *World J. Gastroenterol.* **23**, 3252–3261 (2017).
56. Wu, L. et al. KDM5 histone demethylases repress immune response via suppression of STING. *PLoS Biol.* **16**, e2006134 (2018).
57. Zhang, X., Liu, L., Yuan, X., Wei, Y. & Wei, X. JMJD3 in the regulation of human diseases. *Protein Cell* **10**, 864–882 (2019).
58. Wang, J. et al. BRD4 inhibition exerts anti-viral activity through DNA damage-dependent innate immune responses. *PLoS Pathog.* **16**, e1008429 (2020).
59. Tian, B. et al. BRD4 couples NF-κB/RelA with airway inflammation and the IRF-RIG-I amplification loop in respiratory syncytial virus infection. *J. Virol.* **91**, e00007–17 (2017).
60. Faulkner, G. J. et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563–571 (2009).
61. Bray, M., Driscoll, J. & Huggins, J. W. Treatment of lethal Ebola virus infection in mice with a single dose of an S-adenosyl-L-homocysteine hydrolase inhibitor. *Antivir. Res.* **45**, 135–147 (2000).
62. Coutard, B. et al. Zika virus methyltransferase: structure and functions for drug design perspectives. *J. Virol.* **91**, e02202–16 (2017).
63. Aouadi, W. et al. Binding of the methyl donor S-adenosyl-l-methionine to middle east respiratory syndrome coronavirus 2′-O-methyltransferase nsp16 promotes recruitment of the allosteric activator nsp10. *J. Virol.* **91**, e02217–16 (2017).
64. Li, G. & De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* **19**, 149–150 (2020).
65. Mahalapbutr, P., Kongtaworn, N. & Rungrotmongkol, T. Structural insight into the recognition of S-adenosyl-L-homocysteine and sinefungin in SARS-CoV-2 Nsp16/Nsp10 RNA cap 2′-O-methyltransferase. *Comput. Struct. Biotechnol. J.* **18**, 2757–2765 (2020).
66. Krafcikova, P., Silhan, J., Nencka, R. & Boura, E. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat. Commun.* **11**, 3717 (2020).
67. Uchide, N. & Toyoda, H. Antioxidant therapy as a potential approach to severe influenza-associated complications. *Molecules* **16**, 2032–2052 (2011).
68. Nair, M. P. N. et al. The flavonoid, quercetin, differentially regulates Th-1 (IFNγ) and Th-2 (IL4) cytokine gene expression by normal peripheral blood mononuclear cells. *Biochim. Biophys. Acta* **1593**, 29–36 (2002).
69. Karunakaran, K. B., Balakrishnan, N. & Ganapathiraju, M. Potentially repurposable drugs for COVID-19 identified from SARS-CoV-2 host protein interactome. Res Sq. https://doi.org/10.21203/rs.3.rs-30363/v1 (2020).
70. Colunga Biancatelli, R. M. L., Berrill, M., Catravas, J. D. & Marik, P. E. Quercetin and vitamin C: an experimental, synergistic therapy for the prevention and treatment of SARS-CoV-2 related disease (COVID-19). *Front. Immunol.* **11**, 1451 (2020).
71. Aucoin, M. et al. The effect of quercetin on the prevention or treatment of COVID-19 and other respiratory tract infections in humans: a rapid review. *Adv. Integr. Med.* **7**, 247–251 (2020).
72. Luo, W. et al. Targeting JAK-STAT signaling to control cytokine release syndrome in COVID-19. *Trends Pharmacol. Sci.* **41**, 531–543 (2020).
73. Lindner, H. A. et al. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J. Virol.* **79**, 15199–15208 (2005).
74. Schneider, M. et al. Severe acute respiratory syndrome coronavirus replication is severely impaired by MG132 due to proteasome-independent inhibition of M-calpain. *J. Virol.* **86**, 10112–10122 (2012).
75. Longhitano, L. et al. Proteasome inhibitors as a possible therapy for SARS-CoV-2. *Int. J. Mol. Sci.* **21**, 3622 (2020).
76. Chojnacka, K., Witek-Krowiak, A., Skrzypczak, D., Mikula, K. & Młynarz, P. Phytochemicals containing biologically active polyphenols as an effective agent against Covid-19-inducing coronavirus. *J. Funct. Foods* **73**, 104146 (2020).
77. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
78. Andrews, S. et al. FastQC: a quality control tool for high throughput sequence data. (2010).
79. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
80. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).
81. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
82. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
83. Medvedeva, Y. A. et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database* **2015**, bav067 (2015).
84. Khare, S. P. et al. HIstome—a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.* **40**, D337–D342 (2011).
85. Singh Nanda, J., Kumar, R. & Raghava, G. P. S. dbEM: A database of epigenetic modifiers curated from cancerous and normal genomes. *Sci. Rep.* **6**, 19340 (2016).
86. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).

87. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
88. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).
89. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
90. Horvath, S. & Langfelder, P. Tutorials for the WGCNA package for R: WGCNA Background and glossary. https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/ (2011).
91. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. *InterJ. Complex Syst.* **1695**, 1–9 (2006).
92. Sadegh, S. et al. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. *Nat. Commun.* **11**, 3518 (2020).

ACKNOWLEDGEMENTS

This work was supported by Apoyo para proyectos de investigación científica, desarrollo tecnológico e innovación en salud ante la contingencia por COVID-19, CONACyT [00312021 to ESR], Fondo CB-SEP-CONACyT [284748 to ESR], and DSA-SEP (PRODEP) [47310681, id-250690 to ESR]. M.S.A. and A.D.M.M. are doctoral students in the "Programa de Doctorado en Ciencias Bioquímicas, UNAM" and received a fellowship funding from CONACYT (MSA CVU659273 and ADMM CVU894530). M.S.A. was awarded by the German Academic Exchange Service, DAAD (ref. 91693321). N.A. would like to acknowledge the Independent Research Fund Denmark (6108-00038B). E.S.R. was supported by the Departamento de Ciencias Naturales, UAM-Cuajimalpa. J.B.'s work was supported by his VILLUM Young Investigator Grant no. 13154. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 777111. This publication reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

AUTHOR CONTRIBUTIONS

M.S.A., E.I.N.D., and A.D.M.M. equally contributed to the data collection, bioinformatic analyses, and paper writing. N.A. and J.B. provided critical feedback and helped to improve the paper. R.G.B. and E.S.R. were in charge of overall direction, planning, and supervision. M.S.A., E.I.N.D., and A.D.M.M. contributed equally to this work.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41540-021-00181-x.

Correspondence and requests for materials should be addressed to R.G.-B. or E.S.-R.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fgene.2022.860924

Transcriptomic and Drug Discovery Analyses Reveal Natural Compounds Targeting the KDM4 Subfamily as Promising Adjuvant Treatments in Cancer

Aylin del Moral-Morales[1†], *Marisol Salgado-Albarrán*[1,2†], *Elizabeth Ortiz-Gutiérrez*[1‡], *Gerardo Pérez-Hernández*[1*] *and Ernesto Soto-Reyes*[1*]

[1]*Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana-Cuajimalpa (UAM-C), Mexico City, Mexico,* [2]*Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany*

KDM4 proteins are a subfamily of histone demethylases that target the trimethylation of lysines 9 and 36 of histone H3, which are associated with transcriptional repression and elongation respectively. Their deregulation in cancer may lead to chromatin structure alteration and transcriptional defects that could promote malignancy. Despite that KDM4 proteins are promising drug targets in cancer therapy, only a few drugs have been described as inhibitors of these enzymes, while studies on natural compounds as possible inhibitors are still needed. Natural compounds are a major source of biologically active substances and many are known to target epigenetic processes such as DNA methylation and histone deacetylation, making them a rich source for the discovery of new histone demethylase inhibitors. Here, using transcriptomic analyses we determined that the KDM4 family is deregulated and associated with a poor prognosis in multiple neoplastic tissues. Also, by molecular docking and molecular dynamics approaches, we screened the COCONUT database to search for inhibitors of natural origin compared to FDA-approved drugs and DrugBank databases. We found that molecules from natural products presented the best scores in the FRED docking analysis. Molecules with sugars, aromatic rings, and the presence of OH or O- groups favor the interaction with the active site of KDM4 subfamily proteins. Finally, we integrated a protein-protein interaction network to correlate data from transcriptomic analysis and docking screenings to propose FDA-approved drugs that could be used as multitarget therapies or in combination with the potential natural inhibitors of KDM4 enzymes. This study highlights the relevance of the KDM4 family in cancer and proposes natural compounds that could be used as potential therapies.

Keywords: epigenetics (chromatin remodeling), KDM4 inhibitor, cancer, natural compounds, drug discovery, structural biology

OPEN ACCESS

Edited by:
César López-Camarillo, Universidad Autónoma de la Ciudad de México, Mexico

Reviewed by:
Christo Christov, Michigan Technological University, United States
Zeng-Quan Yang, Wayne State University, United States

***Correspondence:**
Ernesto Soto-Reyes
esotoreyes@cua.uam.mx
Gerardo Pérez-Hernández
gperezh@cua.uam.mx

‡Present address:
Elizabeth Ortiz-Gutiérrez, Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico City, Mexico.

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *24 January 2022*
Accepted: *07 March 2022*
Published: *11 April 2022*

Citation:
del Moral-Morales A, Salgado-Albarrán M, Ortiz-Gutiérrez E, Pérez-Hernández G and Soto-Reyes E (2022) Transcriptomic and Drug Discovery Analyses Reveal Natural Compounds Targeting the KDM4 Subfamily as Promising Adjuvant Treatments in Cancer. Front. Genet. 13:860924. doi: 10.3389/fgene.2022.860924

INTRODUCTION

Histone methylation is the addition of methyl groups to the arginine (R) and lysine (K) residues on histone tails (Portela and Esteller, 2010). The methylation and demethylation of the different lysines in each histone tail allow a dynamic regulation of the chromatin state (Jambhekar et al., 2019) that affects transcription depending on the residue and the number of methyl groups added (lysines can be mono, di, and trimethylated). Histone lysine methylation marks are regulated by two sets of enzymes: histone lysine methyltransferases and histone lysine demethylases (KDMs) (García et al., 2016). KDMs can be divided into two families according to their mechanisms of action (Sterling et al., 2020). The lysine-specific demethylases (LSD) family is characterized by its catalytic site, which requires an available pair of electrons in the nitrogen atom from the lysine that is going to be demethylated; thus, they can only remove mono and dimethyl groups (Stavropoulos et al., 2006). On the other hand, the Jumonji-C domain-containing (JMJC) family is dependent on Fe^{2+} and 2-oxoglutarate and does not require an available pair of electrons for its catalytic activity, which is why it can target mono, di, and trimethylated lysines (Tsukada et al., 2006a); for further reaction mechanism details, see (Guerra-Calderas et al., 2015); (Ramanan et al., 2020), and (Cortopassi et al., 2015). Besides, *in vitro* studies have detected arginine demethylase activities for KDM4A and KDM4E (Walport et al., 2016), even though molecular dynamics simulations combined with quantum mechanical and molecular mechanical calculations suggest that KDM4E demethylase activity is more efficient when an arginine residue is the substrate rather than a lysine residue (Ramanan et al., 2021). KDMs are also divided into eight subfamilies (KDM1-8) according to the similarity of their catalytic domain and their substrate specificity (Sterling et al., 2020).

The KDM4 subfamily is part of the JMJC group. It is composed of five functional members (KDM4A-E) that mainly target the trimethylation of H3K36 and H3K9, which are associated with active transcription and heterochromatin (transcriptional repression), respectively (Katoh and Katoh 2004; Shin and Janknecht 2007b; Labbé et al., 2013; Zhao and Garcia 2015). The KDM4 proteins are of great interest as drug targets due to their oncogenic potential (Rotili and Mai, 2011; Agger et al., 2019). For instance, KDM4A is overexpressed and sometimes amplified in several neoplasms such as leukemia, lung, prostate, colorectal, and breast cancer (Guerra-Calderas et al., 2015). It has also been reported that the inhibition or downregulation of KDM4A causes a decrease in the proliferation of acute myeloid leukemia (Massett et al., 2021), breast cancer (Metzger et al., 2017), and prostate cancer (Mu et al., 2019). KDM4B promotes carcinogenesis in estrogen receptor-positive breast cancer (Yang et al., 2010; Kawazu et al., 2011) and has also been associated with poor outcomes in gastric cancer (Wu et al., 2019), castration-resistant prostate cancer (Sha et al., 2020) and osteosarcoma (Liu et al., 2020). KDM4C promotes malignancy in multiple neoplasms, such as multiple myeloma (Lv and Liu, 2021), glioblastoma (Lee et al., 2021), and squamous cell carcinoma (Labbé et al., 2013).

Only a few studies have explored KDM4D and E's role in cancer; these proteins are shorter than KDM4A-C because they lack the C-terminal PHD and Tudor domains, required for histone recognition and binding (Labbé et al., 2013). In non-neoplastic tissues, KDM4D is mainly expressed in the testis (Iwamori et al., 2011); a few reports suggested that it contributes to the establishment of androgen-independent prostate cancer (Shin and Janknecht, 2007a), acts as a repressor of *p53* in colorectal cancer (Li et al., 2020) and promotes liver cancer progression (Deng et al., 2021). On the other hand, until recently, KDM4E was considered a pseudogene due to its low expression levels; however, recent reports point out that it encodes an active enzyme involved in H3K9me3 demethylation (Hillringhaus et al., 2011; Liu et al., 2018); nevertheless, KDM4E's role in cancer has not been explored yet.

Although the KDM4 proteins are promising targets for cancer therapy, currently there are few reports of KDM4 small-molecule inhibitors [see (Lee et al., 2020) for a comprehensive review]. Nevertheless, all of them target more than one family member due to the similarity of their catalytic domains. For example, disulfiram and ebselen are metal cofactor disruptors that inhibit KDM4A through the obstruction of the Zn^{2+} ion at its catalytic site; however, those drugs target other zinc-binding proteins as well, including other KDMs (Rotili and Mai, 2011). Other known KDM4A inhibitors are 2-oxoglutarate analogs, these molecules act as competitive inhibitors but, since 2-oxoglutarate is a cofactor for several other enzymes including all the JMJC family, these molecules have low specificity (Baby et al., 2021). Because there are cancer types that show dysregulation of only one family member (Sterling et al., 2020), it is important to achieve specific and effective inhibitors for each enzyme. Moreover, most of the KDM4 inhibitors reported to date have only shown *in vitro* activity (Chin and Han, 2015), consequently, there is still a lack of validated drugs that could be used in cancer therapy.

Natural compounds have always been a major source of biologically active substances, and many are known for their effect on epigenetic processes such as DNA methylation, histone marks and lncRNAs (Yang et al., 2018). The KDM enzymes are no exception; for example, several natural products like resveratrol, curcumin and melatonin have been reported as inhibitors of the LSD1 enzymes (Fang et al., 2020). Tripartin, a compound produced by a bacteria found in dung beetles, is the only natural inhibitor reported for the KDM4 subfamily (Kim et al., 2013). However, another study showed that tripartin and its analogs increased H3K9me3 levels but did not directly interact with KDM4 proteins, suggesting that their mechanisms of action could involve other enzymes (Guillade et al., 2018).

Currently, drug repurposing allows the use of medications, previously indicated to certain diseases, as new therapeutic alternatives for other diseases by identifying the protein targets of these drugs. It is cost-effective and has been reinforced by computational approaches such as molecular docking (Pushpakom et al., 2019). In this work, we evaluated the KDM4 subfamily's role in cancer and searched for natural and previously FDA-approved compounds that could potentially inhibit the KDM4 proteins. Our work highlights the value of the KDM4 subfamily as therapeutic targets and, using a combination of transcriptomic and structural

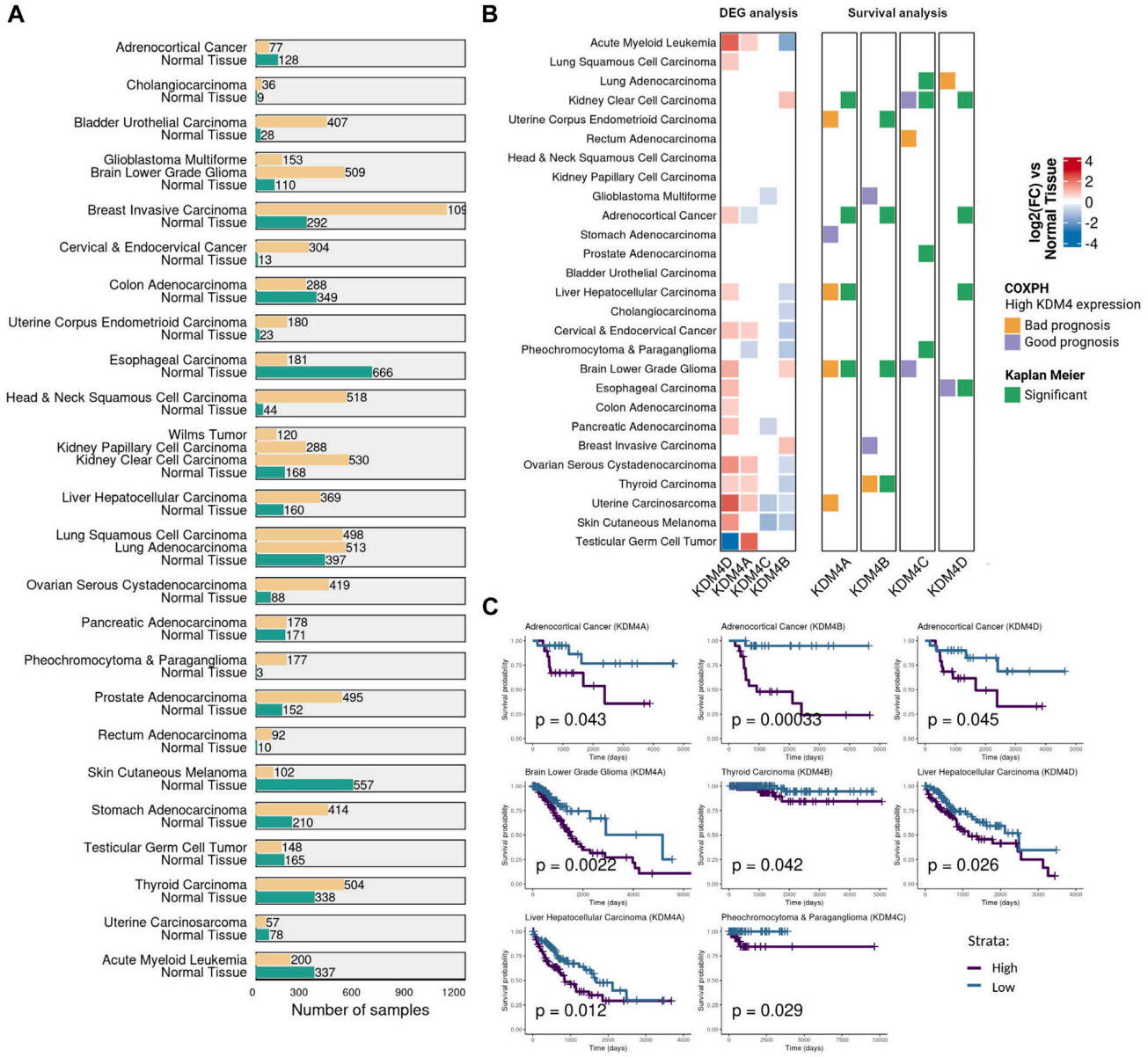


FIGURE 1 | KDM4 family expression in cancer. **(A)** The number of samples used for the transcriptomic and survival analysis. Samples were obtained from TCGA, TARGET and GTEx databases. **(B)** Gene expression and survival analysis for each KDM4 protein. The first panel shows the differential expression analysis of the tumor samples vs. the non-neoplastic tissue. The second panel shows the CoxPH and Kaplan Meier survival analysis as adjacent columns for each KDM4 protein. For CoxPH analysis (first column), the tile color indicates if high levels of the KDM4 are of bad or good prognosis (p value <0.05). For the Kaplan-Meier analysis (second column), tumor samples were divided into two groups according to their KDM expression: Low-KDM and patients with High-KDM (p value <0.05). White tiles represent non significant association. **(C)** Significant Kaplan Meier curves of the KDM4 protein overexpressed in the cancer type where only a bad prognosis relationship was found with p value <0.05.

biology approaches, we provide a set of compounds with high inhibitory and clinical potential in cancer.

MATERIALS AND METHODS

Gene Expression Datasets

Survival information and gene expression levels of non-neoplastic and tumor samples (the "TCGA TARGET GTEx" dataset) were downloaded from Xena Browser (Vivian et al., 2017; Goldman et al., 2020). RSEM expected counts (TcgaTargetGtex_gene_expected_count) were used as input for differential expression analysis. Only cancer types with associated normal tissue available were considered for analysis.

Survival Analysis

Event and time-to-event information was used to evaluate the association between expression of the KDM4 subfamily genes and

Overall Survival of patients using the Survival v3.2-11 package (Grambsch and Therneau, 2000). For COX Proportional Hazards, association was considered significant if p value <0.05. COXPH estimate <0 was labeled as "good prognosis" and COXPH estimate >0 as "bad prognosis". Kaplan Meier plot and Log Rank Test were performed using patients with KDM expression < Q1 (Low-KDM) and patients with KDM expression > Q3 (High-KDM). Difference in overall survival between groups was considered significant with p value <0.05.

Differential Expression Analysis

Normalized RSEM expected counts from Xena Browser were converted to RSEM expected counts (RSEM expected counts = 2^(normalized RSEM expected counts) -1) and used as input for DESeq2 v1.32.0 to compare neoplastic vs. non-neoplastic samples (Love et al., 2014). Differential expression analysis was also performed within a specific cancer type by comparing two groups: patients with KDM expression > Q3 (High-KDM) versus patients with KDM expression < Q1 (Low-KDM). Genes with abs (log2FoldChange) > log2 (1.5) and padj <0.05 were selected as Differentially Expressed Genes (DEGs). For this last analysis samples were chosen if they were labeled as "bad prognosis" by the COXPH test or if they had a significant Log-rank test between the groups used for the Kaplan Meier plot where the High-KDM group had a lower survival expectancy than the Low-KDM group.

DEGs Enrichment Analysis

Enrichment analyses for DEGs were performed with gProfiler2 (Kolberg et al., 2020) using the Hallmark Gene Set Collection gmt file from the Molecular Signatures Database (MSigDB) (Liberzon et al., 2015). The correction method used was g:SCS and an adjusted p value significance threshold of 0.05. All the genes expressed in each sample were used as background.

Ligand Libraries Preparation

Virtual ligand screening studies were performed against three databases: DrugBank (Wishart et al., 2018), the FDA-approved and passed phase I drug library (obtained from www.selleckchem.com), and COCONUT (Sorokina et al., 2021). The databases contained 9131, 3034, and 406,747 compounds, respectively. The libraries were filtered using OpenEye's FILTER algorithm (OpenEye Scientific Software, 2021); the filters applied can be found in **Supplementary File S1**. The ionization state was established through OpenEye's FIXPKA algorithm. Charges were calculated with OpenEye's *molcharge* tool and the AM1-BCC method (Jakalian et al., 2002). Ten low-energy conformers were generated for each molecule with the OMEGA algorithm.

Target Preparation

The crystallographic structures for KDM4 active sites were downloaded from Protein Data Bank (PDB). The accession numbers and references for all the models used are available in **Table 1**. The missing portions of the molecules were modeled with SWISS-MODEL (Waterhouse et al., 2018). The structures (using the active site of KDM4 as target) were prepared for docking with the SPRUCE program included in OpenEye's OEDocking distribution.

TABLE 1 | Structures used for the molecular docking screenings.

Enzime	PDB accession	References
KDM4A	5F32	Bavetsias et al. (2016)
KDM4B	4LXL	Chu et al. (2014)
KDM4C	2XML	Hillringhaus et al. (2011)
KDM4D	4HON	Krishnan and Trievel (2013)
KDM4E	4HOO	Krishnan and Trievel (2013)
	5F5C	Bavetsias et al. (2016)
	5FP4	Westaway et al. (2016)
	5FP7	
	5FP8	
	5FPA	
	5FPB	
	6H10	Małecki et al. (2019)
	2W2I	Hillringhaus et al. (2011)

Molecular Docking

The KDM4 structures were fitted by structural alignment to maintain the same active site orientation. Different reports have established that KDM4 is active in the presence of Fe^{2+} and Zn^{2+} as cofactors in the active site; crystallographic reports have also shown the presence of Ni^{2+} as a cofactor, with no significant differences in KDM conformation. Since the development of a competitive inhibitor must consider the effect of the metal in the active site, in this work, Zn^{2+} was considered as the representative metal for the functional activity of KDM4. For the docking study, the representation of coordination bonds between metals and the active site is not necessary, so we consider this as a good representation of the electrostatic potentials for the metal in the force field used. The conserved residue GLU190 of the active site was defined as the anchor point for the search docking box. Amber ff94 force field was used for protein and Zn^{2+} partial charges calculation.

Two systems were implemented to understand the metal influence over KDM4 proteins' active site, HOLO and APO. The HOLO form included the Zn^{2+} cofactors in the active site of each KDM4. For the APO system, any metal cofactor was removed for the analysis. Each cured chemical database was docked to the receptor in APO as GLU190 residue as reference of binding site, while for the HOLO form we used GLU190 and Zn^{2+} as reference for the FRED program from OpenEye Scientific software. The Chemgauss4 scoring function was used, and the top 100 scoring molecules for each case were considered possible hits.

Flexophore Similarity Analysis Between Compounds

The 3D-pharmacophore similarity analysis between compounds was performed with DataWarrior (Sander et al., 2015) using the *Flexophore* descriptor. Two compounds were considered similar if their similarity relationship surpassed a threshold of 95%. For each protein and database evaluated, the top 100 scoring compounds were

analyzed. As nine different PDB structures were used for KDM4D, a random sample of 100 compounds from each database were chosen.

Molecular Dynamics Simulations and Absolute Binding Energy Calculations

Molecular docking methods are efficient tools for large database screening; however, their main limitation is inaccurate binding energy estimations. The molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method was used to estimate the absolute binding energy (ΔG_{PBSA}) of the ligands. Since this approach requires a large amount of computational resources, we performed the ΔG_{PBSA} calculation for only a subset of molecules according to the following criteria:

a) Molecules that bind only to one KDM4
b) Molecules that bind to some or all the significantly overexpressed KDM4s in a cancer type.
c) The best 10 molecules for each KDM4 according to their FRED/Chemgauss4 score, regardless of the source database.

Briefly, each protein-ligand complex was subjected to 20 ns of molecular dynamics simulations using GROMACS 5.1.15 (Abraham et al., 2015). The files were processed by pdb2gmx, setting AMBER99SB as the force field and TIP3P as the water model. Due to the difficulty of simulating Zn coordination states, all the simulations were performed using the APO form. Partial charge AM1BCC obtained with MOLCHARGE for each ligand was conserved. The van der Waals and topology parameters of the ligands were generated with ACPYPE setting GAFF as the force field (Silva and Vranken, 2012). The complexes were enclosed into a dodecahedral box with a minimum box-solute distance of 1.0 nm, and the cell was filled with water. Each system was equilibrated using the conditions previously described by Kumari et al. (2014). After equilibration, a 20 ns production run was carried out. The ΔG_{PBSA} was calculated with GROMACS *g_mmpbsa* (Kumari et al., 2014).

Network Analysis

The network analysis was performed with a selection of input genes which were selected as follows:

- The starting point were the top 100 drugs from the DrugBank database for each KDM as evaluated through the molecular docking analysis, which were used to retrieve their target proteins using the protein-drug interactions integrated in NeDRex (Sadegh et al., 2021).
- Only upregulated genes identified in the comparisons between High-KDM and Low-KDM were selected.

The protein-protein interaction network used as reference was obtained from IID version 2021-04 (Kotlyar et al., 2019), only the experimentally validated edges ("exp", "exp;ortho", "exp;ortho; pred" or "exp;pred") were used. The networks were assembled with KeyPathwayMiner (K = 3 and L = 0) (Alcaraz et al., 2020). Only the upregulated proteins targeted by the top 100 compounds from the DrugBank database for each KDM were used as input. The differentially expressed KDM4s in each tumor were defined as positive nodes. Protein-drug and protein-protein interaction networks were merged and edited using Cytoscape 3.8.2 (Shannon et al., 2003).

RESULTS

KDM4 Subfamily Expression is of Bad Prognosis in Cancer

To address the KDM4 subfamily's role in cancer, we carried out gene expression analysis on a broad set of publicly available tumor and non-neoplastic tissue samples (**Figure 1A**). The differential expression analysis of the tumor samples vs. the non-neoplastic tissue showed that KDM4A-D subfamily members are deregulated in several tumors and there are several combinations of differentially expressed KDM4s for each tumor type. KDM4D and KDM4A are the most notable genes since they are mostly overexpressed while KDM4B and C are usually downregulated compared to non-neoplastic tissue (**Figure 1B**). To further characterize the clinical significance of the KDM4 subfamily members, two survival analyses were conducted; CoxPH and Kaplan Meier (**Figure 1B**). For the last one, samples were divided into two groups according to their KDM expression: low (first quartile) and high (fourth quartile, **Figure 1C**). KDM4A overexpression indicates a bad prognosis for Uterine Corpus Endometrioid Carcinoma, Liver Hepatocellular Carcinoma, Adrenocortical Cancer, Brain Lower Grade Glioma, and Uterine Carcinoma. KDM4B overexpression is a bad prognosis for Adrenocortical Cancer and Thyroid Carcinoma. KDM4C expression is related to a bad prognosis for Rectum Adenocarcinoma and Pheochromocytoma and Paraganglioma. KDM4D is related to a bad prognosis for Lung Adenocarcinoma, Adrenocortical Cancer, and Liver Hepatocellular Carcinoma. Finally, since KDM4E expression is low in most of the samples evaluated, we do not report the differential expression analysis or the survival analysis for this gene.

We next seek to evaluate the relevance of the KDM4 proteins in a selected group of cancer types, the selection considered the fact that we are interested in inhibitor molecules; thus, the tumors used for further evaluation are the ones where the overexpression is related to bad prognosis. We conducted a differential expression analysis comparing only the tumor samples ranked by KDM expression: high-KDM (fourth quartile) vs. low-KDM (first quartile). The log2 Fold Change (log2FC) of each KDM4 protein and the number of differentially expressed genes (DEGs) in each comparison are shown in **Figure 2A**. Enrichment analysis of the DEGs against the GSEA Hallmarks database showed that, in cancer, the genes regulated by the KDM4 family are involved in processes such as TNFα signaling by NFκB, interferon-gamma response, inflammatory response, G2M checkpoint, and p53 pathway (**Figure 2B**). Overall, our data suggest that KDM4 proteins are relevant targets to screen for specific inhibitors that could be beneficial in the treatment of neoplasms.

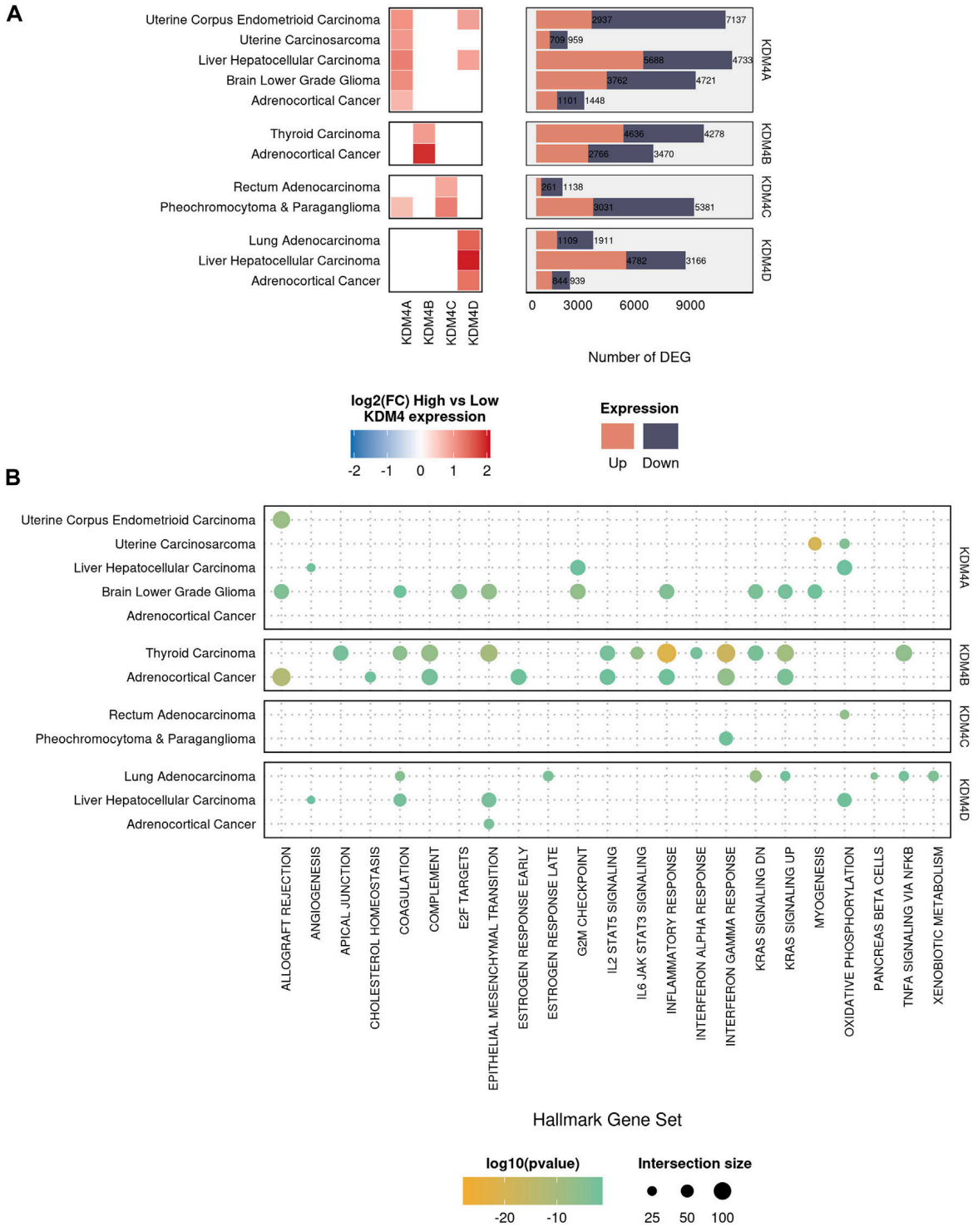


FIGURE 2 | Differential expression and enrichment analysis of the KDM4 family. **(A)** The left panel of squares represents the 12 types of tumors where the deregulation of the KDM4 subfamily is of bad prognosis. The differential expression analysis was performed comparing High-KDM vs. Low-KDM samples. Color intensity is related to the log2(FC). The right panel represents the number of differentially expressed genes (DEG) for each comparison. **(B)** Hallmarks of Cancer enrichment analysis for the DEG in each sample. Color intensity represents the pvalue and size of the intersection size.

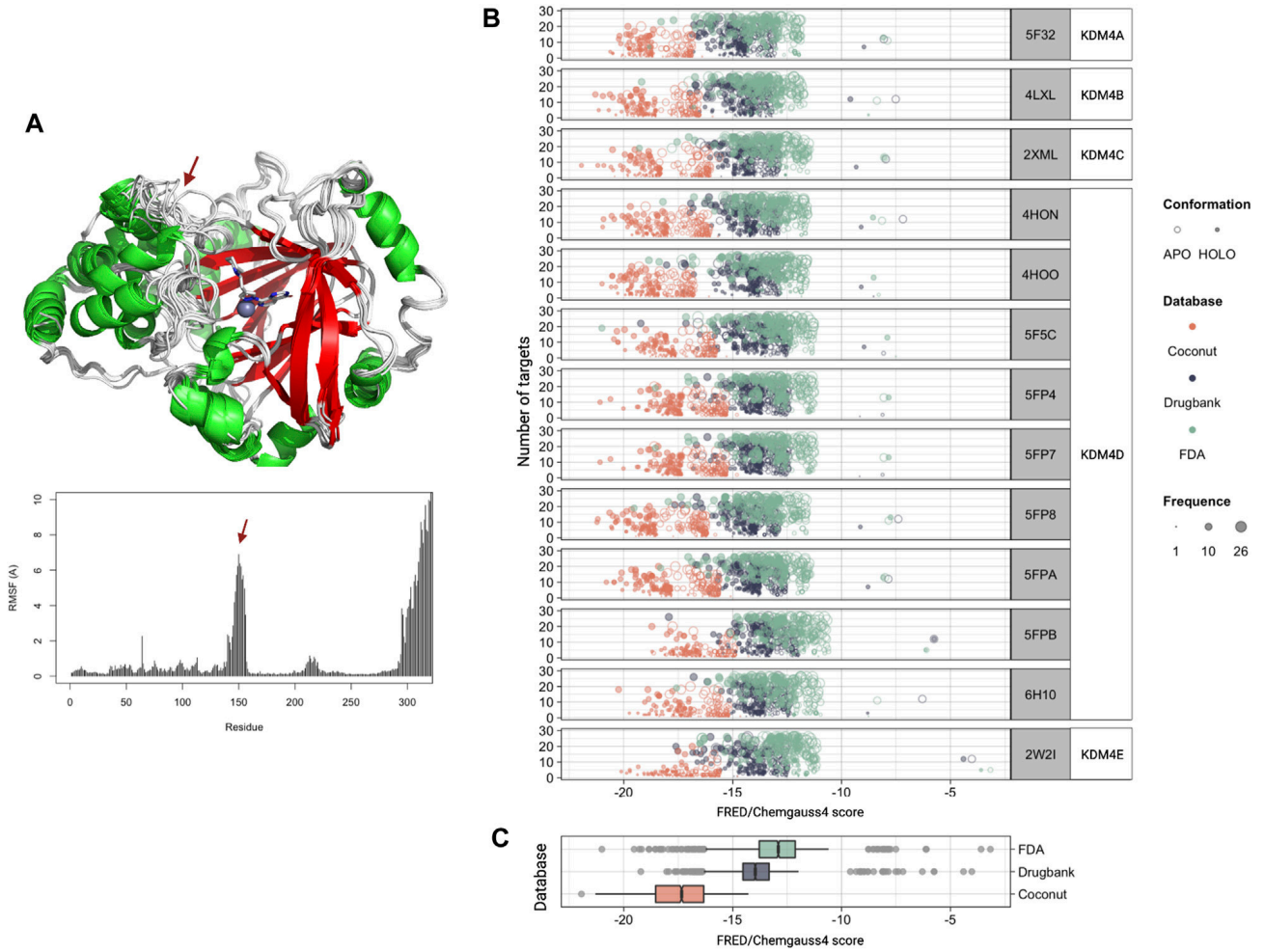


FIGURE 3 | Molecular docking against the KDM4 subfamily. **(A)** Structural 3D alignment (upper panel) and root mean square fluctuation (RMSF, lower panel) for the aminoacid residues of the KDM4 subfamily structures used in this work. The red arrow indicates the location of the residues with the highest RMSF. **(B)** FRED/Chemgauss4 score distribution for the top 100 compounds from each database (COCONUT, FDA, and DrugBank) that were predicted to bind to each of the KDM4 family members. The fill indicates the enzyme system used, APO (without metal cofactors), and HOLO (with all metallic ions). Size is proportional to each compound's number of targets according to our docking analysis. **(C)** FRED/Chemgauss4 score distribution for each of the three databases evaluated. Outlier points are shown in gray.

Natural Compounds as Promising Potential KDM4 Subfamily Inhibitors

Since we observed that KDM4 proteins are deregulated in several neoplasies and that their expression is related to several processes associated with cancer, we next used molecular docking to screen for potential inhibitory compounds. In order to explore the scaffold for inhibition specificity, we docked a total of 418,912 compounds from three different databases (DrugBank, FDA, and COCONUT) against the active sites of each KDM4; the protein targets used are disclosed in **Table 1**. Previous to the molecular docking analysis, the available PDB structures for each of the KDM4 proteins were superimposed; no significant changes in the catalytic sites were found between them. Only KDM4D showed important structural variations between the different models available in PDB, mainly in the loops surrounding the active site entrance. For this reason, a single structure was used for KDMs 4A, B, C, and E while we kept 9 for KDM4D to have a representative sample of its different conformations. The Root Mean Square Fluctuation (RMSF) for the KDM4 structures used in this work indicates that overall, the catalytic site's structure conformation is similar between the different KDM4s, although there is a peak around residue 150 (amino acid numbers are relative to KDM4A) which belongs to the outer loop region with higher mobility (**Figure 3A**).

For docking analysis, both HOLO and APO forms of 13 KDM4 structures were prepared; thus, a total of 26 structures were sampled for FRED/Chemgauss4 docking. The 100 best-scored results were selected, recording a total of 7,800 protein-ligand interactions (**Supplementary File S2**). **Figure 3B** shows the score distribution of the 7800 compounds related to the number of different structures that could be targeted by each ligand. Note that the FRED/Chemgauss4 score is related to the binding energy of the protein-

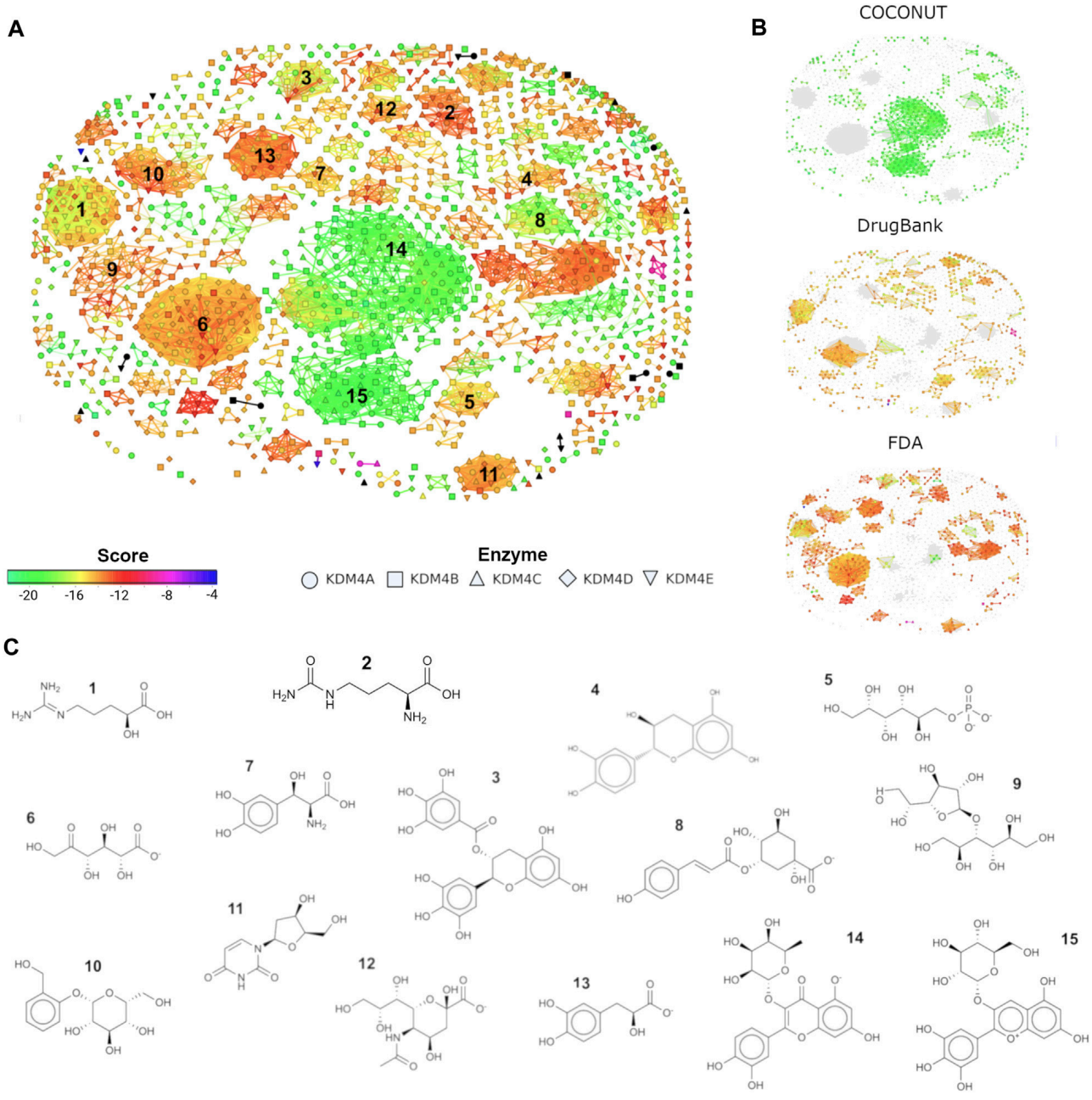


FIGURE 4 | Similarity flexophores map for COCONUT, DrugBank, and FDA top hits. **(A)** Similarity flexophores map. Each node represents a compound, the node color depicts its FRED/Chemgauss4 score. The node shape indicates which KDM4 the ligand binds. The network edges indicate a relationship of at least 95% of flexophore similarity between compound pairs (neighbors). Black dots represent the compounds from Baby et al., 2021. **(B)** Node distribution for the compounds belonging to each of the databases evaluated (COCONUT, DrugBank, and FDA). **(C)** 2D structure for a representative compound from each of the chosen clusters. Clusters were selected based on their size, and edge number.

ligand complex; thus, large negative values stand for stronger interactions and suggest that a molecule has a higher binding potential. For all the KDM4 enzymes, COCONUT compounds had the best favorable binding score, set between −21 and −15, meanwhile, most of the FDA and DrugBank values trend to locate near less favorable scores (between −16 and −11) and have a notable proportion of outlier ligands with scores greater than −10 (**Figure 3C**). As shown in **Figures 3A,B** high protein-ligand interaction count is related to high ligand promiscuity for different KDM4 proteins, whereas the values near to zero suggest that the ligand binding is specific for an enzyme, which is desirable for drug design (**Supplementary Files S3, S4**). We also observed that the best scores were achieved with the HOLO system in comparison to the APO system, suggesting that the ligands can provide functional

groups that act as chelating agents that form coordination bonds with the divalent metal in the active site of the HOLO form of KDM4.

Flexophore-Based Scaffold Suggests Phenols and Sugars as Key for the Design of Potential KDM4 Inhibitors

Next, we seek to further explore whether structural similarities exist among the compounds predicted to bind the KDM4 subfamily members; such findings could be important for understanding the molecular signatures involved in the protein-ligand interactions and for the future development of KDM4 subfamily inhibitors. To address this idea, we evaluated the chemical scaffold of the top 100 hits for each KDM4 from FDA, DrugBank, and COCONUT databases (1,500 molecules total) using a similarity flexophores map. This graphical method tests whether the conformational flexibility of a molecule plays a significant role as a potential inhibitor of proteins (von Korff et al., 2009). Usually, ligands adopt subtle conformations to achieve geometric complementarity with their targets, allowing them to reorganize the attractive and repulsive forces required during their binding. Thus, a molecule with several rotatable bonds (higher flexibility) is more likely to adapt to a binding site. Our flexophore analysis retrieved clusters with maximized edges and nodes that match similar compounds. We analyzed 15 representative clusters, arbitrarily numbered, while isolated nodes (436 out of the 1,500 compounds evaluated) represent ligands with no similar molecules according to the criteria used (**Figure 4A**). In the literature had been reported molecules experimentally validated as inhibitors of the KDM4 subfamily; thus, for the flexophore analysis, we included 16 compounds cited by (Baby et al., 2021) whose IC50 is of micro to nanomolar range (nodes in bold in **Figure 4A**). We observed that most of these molecules remained as isolated nodes whose floxophores did not share similarities with the compounds from COCONUT, DrugBank, and FDA databases.

The molecules' distribution by library is shown in **Figure 4B**. The node color represents the Fred/Chemgauss4 score, and the shape indicates to which KDM4 the compound potentially binds. It is noticeable that the two central clusters (numbers 14 and 15) contain mainly COCONUT compounds. Clusters 1, 6, 7, 11, and 12 have a mixture of DrugBank and FDA molecules; cluster 5 has mainly DrugBank compounds, and clusters 1 and 3 contain a combination of the three databases, while the remaining clusters are composed primarily of compounds from the FDA database. A representative molecule for each cluster is displayed in **Figure 4C**. Clusters with better Fred/Chemgauss4 score and highest node density like clusters 14 and 15, contain molecules composed of 3–4 rings of phenol or pyranose group combinations joined by glycosidic bonds that increase flexibility to the molecules. A similar composition was observed for the molecules from cluster 3, although this set had more members with lower score values than the ones previously mentioned, this can be due to the ketone group joining the rings instead of a glycosidic bond, and the carbonyl of ketone can influence the dipolar moment and flexibility of the molecule altering the possible pi-interactions with the receptor. In general, molecules with fewer than 3 rings (as well as linear molecules), tend to have a lower score. The former indicates that rings from sugars and aromatic molecules favor the interaction with the binding site of KDM4 proteins. We also noticed that OH and O- groups are essential for the interaction between the ligand and KDM4 to doing coordination bonds with their metal cofactors, such as Zn^{2+}, Ni^{2+} or Fe^{2+}; thus, in drug design, the inclusion of sugars and phenols represents an advantage for the achievement of a competitive inhibitor.

Active Site of KDM4 is Stabilized by Pi-Stacking Aromatic Residues and Favor Flavonoid-Carbohydrates Ligand Binding

Docking algorithms are powerful tools for the identification of potentially inhibitory molecules; however, since their main purpose is to narrow down large compound databases, the protein-ligand binding affinity calculations are often sacrificed to achieve higher calculation speeds. The scoring functions used by these algorithms have serious limitations to adequately estimate binding energies, in addition, they do not consider the conformational changes of ligands and targets. To overcome this challenge, we validated the affinity of Protein-Ligand complexes through molecular dynamics simulations. The absolute binding energy (ΔG_{PBSA}) was calculated with the MM-PBSA method for a representative subset of molecules (20 from FDA, 16 from DrugBank and 25 from COCONUT). Because performing molecular dynamics with HOLO systems represents a computational challenge (Vidossich and Magistrato, 2014), it was decided to calculate the ΔG_{PBSA} only for the APO systems.

The FRED/Chemgauss4 score vs. the calculated ΔG_{PBSA} for each ligand were compared; in both cases, a negative value means that the protein-ligand interaction is favorable; if both values were negative, the hit was considered a success. A successful Protein-Ligand complex means that the interaction predicted by the docking algorithm could be replicated through molecular dynamics simulations, thus there is a high possibility for that ligand to be a KDM4 inhibitor. Since we observed a success rate higher than 60% in all the compounds evaluated (**Figure 5A**), we considered the predictions obtained by the FRED algorithm as potential KDM4 family inhibitors.

Since the KDM4A-CNP0371131 complex had the more negative FRED/Chemgauss4 score out of all of the protein-ligand complexes evaluated, it was chosen as a representative example for the conformational changes observed during the molecular dynamics simulations. The per residue RMSF values showed that the loop areas surrounding the KDM4A cavity (residues 170, 225, and C-terminal) are the most flexible areas. It is also noticeable that the ligand is vibrating inside the protein's active site (**Figure 5B**). KDM4A's cavity area is 753.5 A^2, its volume is 824 A^3, and has an exclusively negative electrostatic potential (**Figure 5C**). Due to its size, the KDM4A binding site could fit molecules twice the size of structures 14 or 15 in **Figure 4C**, which have an area of 320 A^2 or 303 A^2,

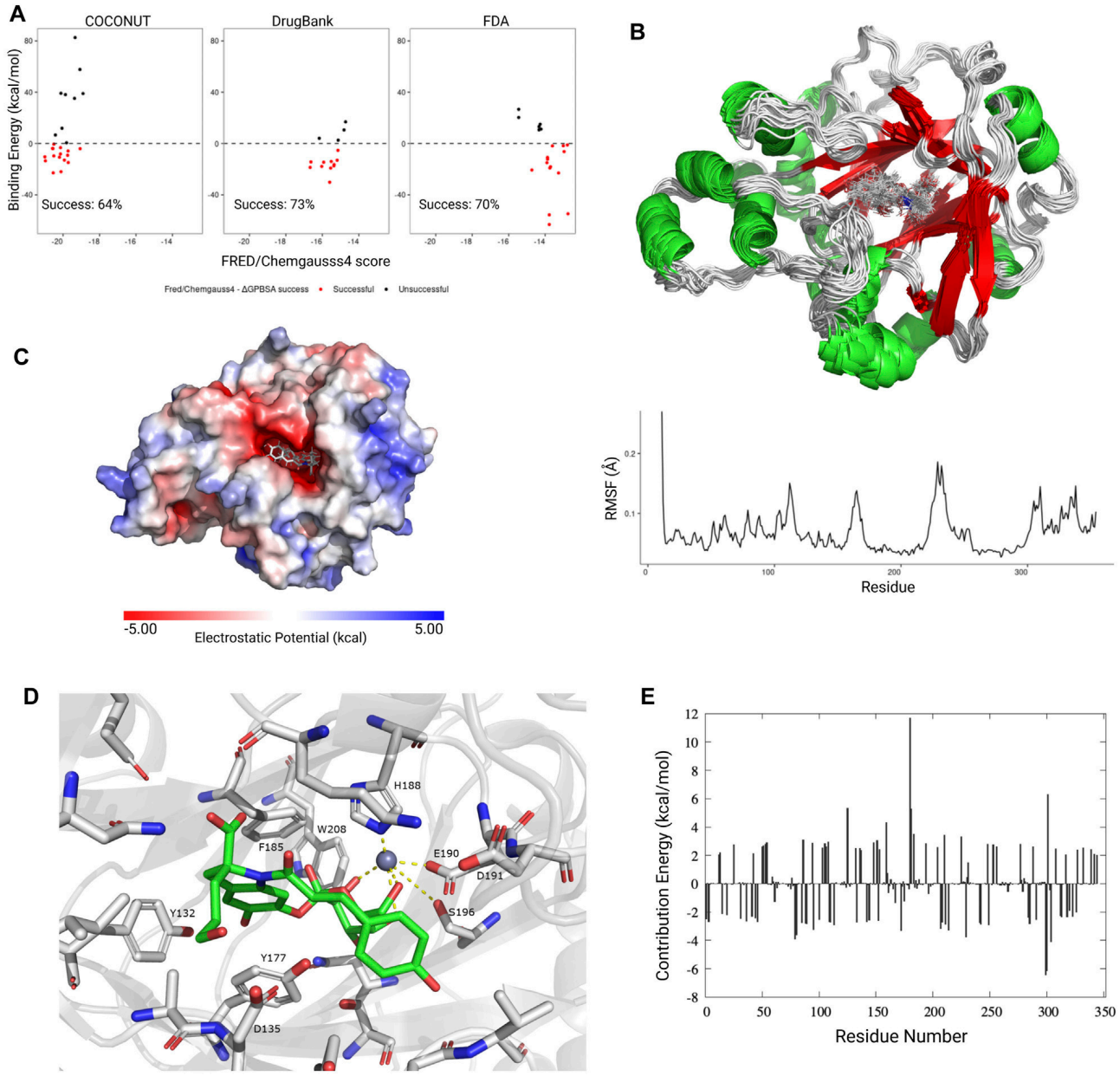


FIGURE 5 | Molecular dynamics simulations and absolute binding energy calculation. **(A)** FRED/Chemgauss4 vs ΔG_{PBSA} correlation for the best scoring ligands from each database. Ligands that showed a favorable binding energy (<0 kcal/mol) and negative FRED/Chemgauss4 score were considered as successful (red dots). The success percentage represents the proportion of successful molecules for each database. **(B)** *Upper panel*: Graphical representation (20 frames) of the molecular dynamics simulation for KDM4A (PDB ID: 5F32) in complex with the CNP0371131 ligand from COCONUT. *Lower panel:* RMSF value for each residue. **(C)** Electrostatic potential for the KDM4A-CNP0371131 complex. **(D)** Graphical representation of the CNP0371131 molecule (green) bound to KDM4A's catalytic site. The residue numbers correspond to PDB structure 5F32. **(E)** Average per residue MM-PBSA binding free energy contribution for the KDM4A-CNP0371131 complex.

respectively. The former suggests that only half of the cavity is occupied by the ligand, leaving the other half to the metallic cofactors and the solvent. Therefore, the molecule's size is not a limitation for the design of a competitive inhibitor; instead, it is the functional groups that coordinate the metallic cofactors and the interactions with the catalytic site's residues that determines the specificity of the ligand-receptor binding.

A remarkable characteristic of the KDM4A binding site is the presence of several aromatic amino acids (Y, F, W, and H) which not only stabilize the binding site but also contribute to the protein-ligand binding through pi stacking interactions with other aromatic groups. The residues that most frequently interact with the ligands are I71, Q84, N86, Y132, A134, D135, G170, V171, Y175, Y177, F185, H188, E190, D191,

TABLE 2 | List of the top molecules with potential inhibitory activity of KDM4 subfamily proteins determined with molecular docking using COCONUT, DrugBank and FDA databases.

Target	Database	Ligand
KDM4A	COCONUT	CNP0058667, CNP0150788, CNP0216191, CNP0002425, CNP0371131, Pulchellidin 3-Glucoside (CNP0359043), CNP0223133, CNP0258703 (Epigallocatechin gallate)
	DrugBank	6-O-capryloylsucrose, Zanamivir, Acteoside, DB04211, DB03249, DB07719, DB12116
	FDA	Glucosamine, Glucosamine sulfate, Doripenem, Neohesperidin, Sulisobenzone, Verbascoside Wedelolactone, Epigallocatechin gallate
KDM4B	COCONUT	CNP0322725, CNP0216191, CNP0098686, CNP0316754, CNP0107391, CNP0239128, Crispine D (CNP0119105)
	DrugBank	Carba-glucotropaeolin, Ascorbyl glucoside, Zanamivir, Iodo-Willardiine, beta-D-arabinofuranose 5-phosphate, DB03250,DB02488
	FDA	Methazolamide, Sulisobenzone, Baricitinib, Lanraplenib, Pentostatin
KDM4C	COCONUT	CNP0187735, CNP0417860, CNP0226084, CNP0298305, CNP0289146, CNP0350449, CNP0106665
	DrugBank	Peramivir, DB03717, Edotecarin, 3′-Uridine Monophosphate
	FDA	Cynarin, Quercitrin, Chlorogenic acid, (-)-Epigallocatechin gallate, Hyperoside, Gastrodin, Polydatin
KDM4D	COCONUT	6-C-Glucosylorobol (CNP0299696), CNP0002425, CNP0362352, CNP0243580, CNP0216191, Isovolubilin (CNP0151675), CNP0397301
	DrugBank	6-O-Capryloylsucrose, Balanol, 10-hydroxycamptothecin, DB07102, 2′-Deoxycytidine-5′-Monophosphate, Cidofovir, Levoglucose
	FDA	Glucosamine, Glucosamine Sulfate, Oleuropein, Sulpiride, Sulisobenzone, Levosulpiride (Levogastrol), Hydroxycamptothecin
KDM4E	COCONUT	CNP0131606, CNP0186792, CNP0125603, 4-hydroxy-2-ketoarginine (CNP0433705), CNP0295348, Quercetin 5-Glucuronide (CNP0081446), CNP0249133
	DrugBank	Azacitidine, Meglumine, Balanol, Levoglucose, Ascorbic acid, L-Xylulose 5-Phosphate, 5-phospho-D-arabinohydroxamic acid
	FDA	Glucosamine, Glucosamine Sulfate, Minoxidil Sulphate, Sulfamonomethoxine, Sulpiride, Xylitol, Orotic Acid (6-Carboxyuracil)

**For long compound names only the database ID is provided.*

S196, N198, W208, L241, and S288. **Figure 5D** shows an example for CNP0371131 binding to KDM4A, since it is the best scoring complex. H188 stands out because it establishes two coordination bonds with the metallic cofactors (Zn^{2+}, Ni^{2+}, or Fe^{2+}). It was also observed that, although the E and D residues in the catalytic site do not directly interact with the ligands, they do contribute to the overall negative microenvironment of the cavity. For example, epigallocatechin gallate (EGCG), a molecule belonging to cluster number 3, establishes one coordination bond with KDM4A metal through the flavonoid group, meanwhile, the secondary catechol bends in the opposite direction of the metal due to repulsive forces effect between them. Additionally, compounds in clusters 14 and 15 exhibit a favorable orientation of the OH groups of the sugar on the flavonoid that allows the formation of 2–3 coordination bonds with the metal (**Figure 5D**); although in this case the number of coordination bonds increases, they are not provided by catechol but by the carbohydrates. The former is due to the reduced availability of electrons in the oxygen from the OH of the secondary catechol to form coordination bonds in comparison with those of the sugars that show a higher electron availability and thus, capacity to form more coordination bonds with the Zn^{2+} at the active site of KDM4A. This is a possible explanation as to why molecules with sugars and phenol groups achieved the best FRED/Chemgauss4 scores.

To study the ΔG_{PBSA} energy distribution through the protein, we calculated the per residue binding energy contribution of the KDM4A-CNP0371131 complex. We observed that the binding energy is mainly driven by long-range electrostatic interactions and it is distributed along all the residues, not only the ones present in the cavity (**Figure 5E**). In general, the attractive forces (negative values) compensate for the repulsion forces (positive values), and although some peak repulsion forces can be found (such as the one for residue 180), these are compensated by other stabilizing interactions (such as residues 78, 172, 228, 298 and 300), leading to an overall favorable ΔG_{PBSA} energy. The former indicates that the complex is stable; thus the ligand has probabilities of showing KDM4A inhibition activity *in vitro*. The binding energy is achieved by the contribution of the favorable intrinsic interaction energy (ΔE_{MM}) and the nonpolar interaction energy ($\Delta E_{nonpolar}$), while an unfavorable penalty is applied by the polar interaction energy (ΔE_{polar}), mainly due to the solvation effect of both the ligand and the active site (**Supplementary Figure S4**). Together, these results provide an insight into the molecular interactions between the KDM4A catalytic site and small molecules, which could assist in the present and future design of small inhibitors. As an example, **Table 2** lists the top molecules obtained from COCONUT, DrugBank, and FDA databases.

KDM4 Subfamily Inhibitors are Potential Multitarget Therapies in Cancer

Since our data show that there are some cancer types where more than one KDM is involved, we suggest that a drug that targets all the significant KDM4 proteins in a neoplasm could be highly effective as a therapy. To integrate all this information (KDM4 gene expression, drug inhibitors, and transcriptomic profiles of each cancer with KDM4 overexpression), facilitate interpretation and explore the applicability of the results, we constructed a protein-drug-disease network containing the five KDM4s and the

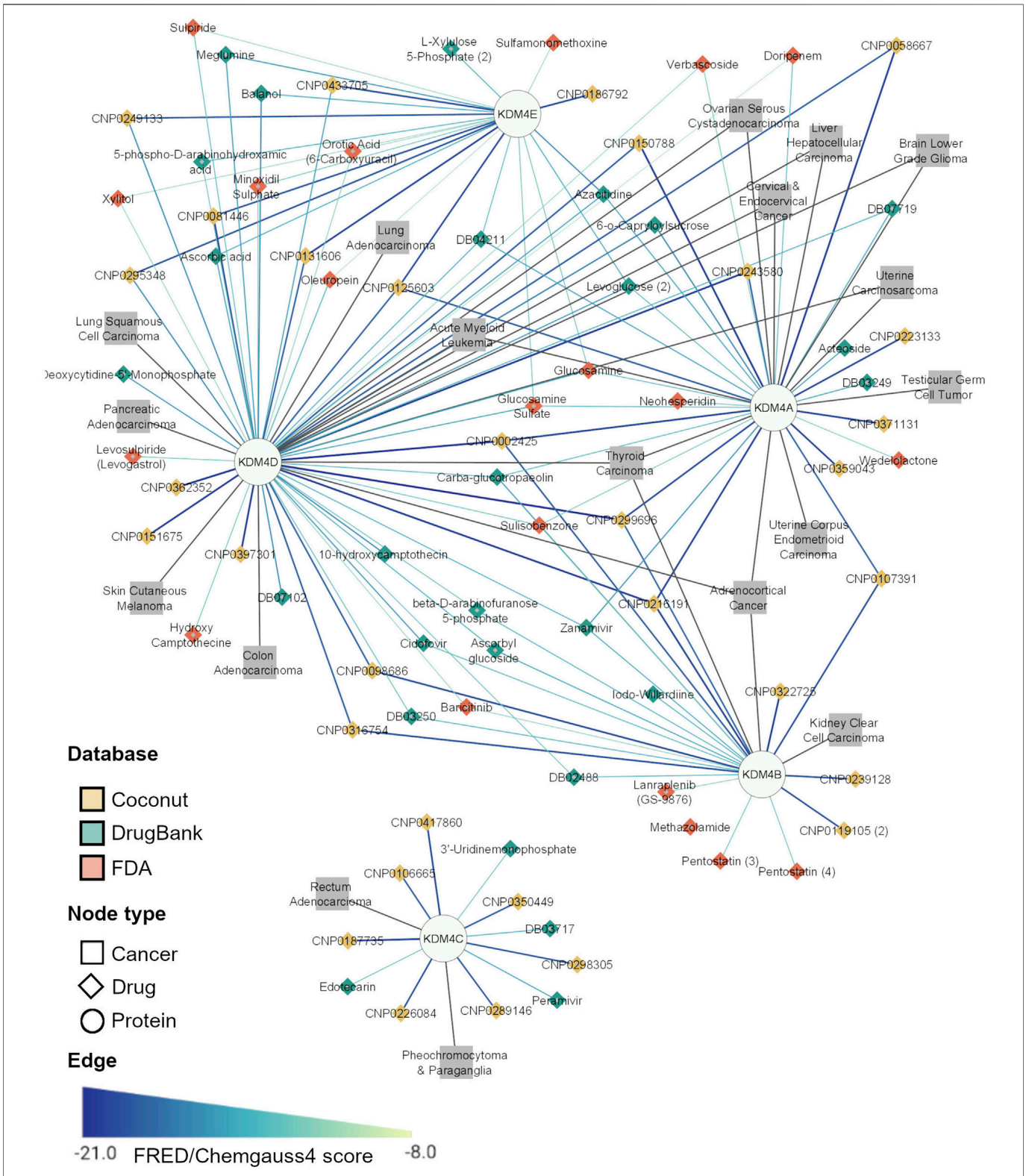


FIGURE 6 | Top potential inhibitors of the KDM4A family. The network represents the Drug-Protein and Disease-Protein relationship between the members of the KDM4 family. The Drug-Protein edge width and color intensity represents the FRED/Chemgauss4 score. For long compound names only the database ID is provided.

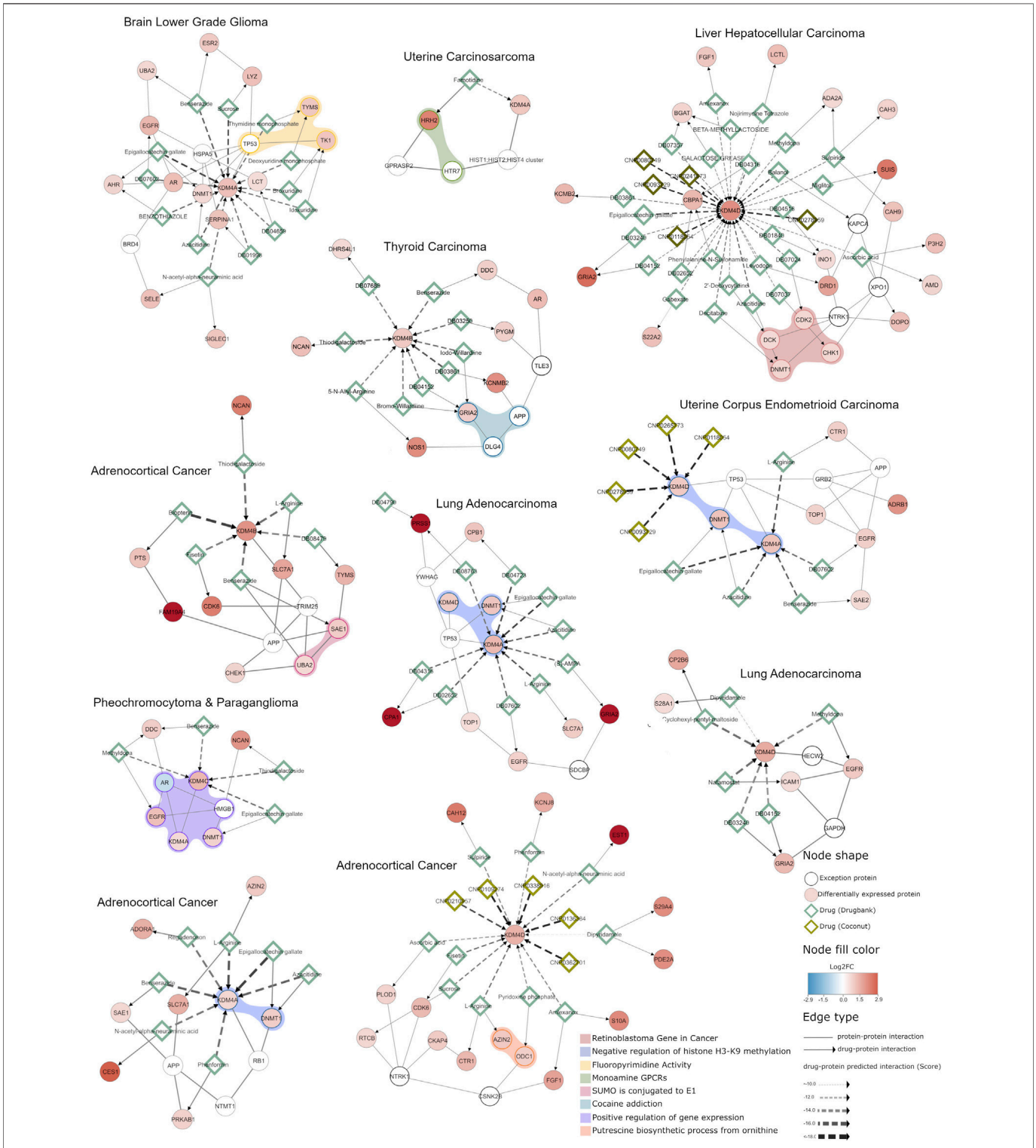


FIGURE 7 | Integrative network analysis of KDM4 potential inhibitors in different cancer types. A network enriched with differentially expressed genes obtained from each cancer type selected is shown. The circular nodes represent proteins, and the edges the interactions between them. The color of the circular nodes represents the fold change in gene expression between tumors with high and low KDM4 expression. Drugs targeting the proteins are represented by diamond nodes, where dark green is used for Drugbank drugs and light green for natural compounds (Coconut database). The Drugbank drug and protein target interactions were retrieved from curated databases (NeDRex platform), while the natural compound interactions with proteins are predicted by the *in silico* analysis performed previously. The colored shadow highlights the proteins that participate in a cellular process according to g.Profiler enrichment. Overall, the network depicts the KDM4 proteins, their protein interaction context and shared interactions with known drugs and natural compounds.

top seven hits for each KDM4 from the three databases evaluated. We also included the neoplasms related to each enzyme; a neoplasm was included if a KDM4 was overexpressed or if it was of bad prognosis in any of the two survival analyses. When integrating these data we observed that according to their KDM4 expression pattern, a different drug set for each neoplasm can be found (**Figure 6**); for example, KDM4E and D are of importance for Lung Adenocarcinoma, therefore, sulpiride and balanol are FDA-approved drugs that could be considered for the treatment of that cancer. Moreover, KDM4A, B, and D are relevant for Adrenocortical Cancer and Thyroid Carcinoma; thus the COCONUT CNP0002425, CNP0299696, and CNP216191 compounds are prominent candidates for the treatment of those neoplasms. For the therapy of Acute Myeloid Leukemia, since KDM4A, D and E are involved, the CNP0131606 is promising given the fact that it could target those three enzymes. As for the KDM4A-CNP0371131 complex (which had the highest FRED/Chemgauss4 score) we observed that CNP0371131 was exclusive for KDM4A; thus, could be used as a treatment for cancers where only KDM4A is deregulated, such as Uterine Corpus Endometrioid Carcinoma and Testicular Germ Cell Tumor.

A detailed example of the usefulness of this analysis is the network extracted for uterine corpus endometrioid carcinoma (**Figure 7**), which shows that KDM4D and KDM4A are overexpressed and both interact with DNMT1 (a DNA methyltransferase involved in gene regulation); our docking analysis shows that there are 5 COCONUT compounds able to target KDM4D and 5 DrugBank compounds targeting KDM4A. However, the DrugBank compounds target other proteins in the network in addition to KDM4; for instance, DB07602 targets KDM4A and EGFR; and Azacitidine inhibits KDM4A and DNMT1, which suggests that Azacitidine could modulate essential proteins involved in the negative regulation of histone H3K9 methylation (as depicted by the light blue shadow in **Figure 7**). This same approach to interpret the results can be applied for the other networks specifically generated according to the expression profiles shown in **Figure 2A**.

This analysis also allows us to observe that KDM4 proteins, when overexpressed, trigger expression changes that affect genes involved in various cellular processes. For example, the network detected for pheochromocytoma and paraganglioma shows that the proteins are involved in the positive regulation of gene expression, which is also closely related to the negative regulation of H3K9 methylation function found in networks adrenocortical cancer (KDM4A overexpression), lung adenocarcinoma (KDM4A overexpression) and uterine corpus endometrioid carcinoma. Furthermore, we found that some druggable processes are related to monoamine GPCRs or closely related to cocaine addiction pathways, in thyroid and uterine carcinomas (**Figure 7**), this is highly relevant given the fact that proteins involved in these metabolic processes have previously been demonstrated to be affected in some cancer; such as lymphoma, prostate, lung cancer and some brain cancers (Rybaczyk et al., 2008; Shih, 2018). Thus, these results suggest that targeting KDM4 proteins can also be a promising therapeutic approach because the drugs targeting them can potentially modulate cellular processes that contribute to the neoplastic phenotype.

DISCUSSION

Epigenetic processes play an important role in the regulation of gene transcription. The discovery of histone demethylases has contributed to understanding the dynamic process of histone marks establishment where the deregulation of these enzymes can contribute to the development of several diseases including cancer (Guerra-Calderas et al., 2015). These types of enzymes can affect the expression of multiple genes such as oncogenes, cell cycle genes and tumor suppressor genes (Sterling et al., 2020). Many of these demethylases have been involved in cancer, such as KDM1A, related to the maintenance of clonogenicity and the inhibition of differentiation (Harris et al., 2012). As well as KDM2A and KDM2B, which have K3K36me2 and H3K4me3 as their substrate, where its deregulation is associated with increased proliferation of stem cells and tumor growth and metastasis (Harris et al., 2012; Wagner et al., 2013) among other processes such as cell proliferation and drug resistance, among others. In the present work we focus on the role of KDM4 subfamily members since they have been involved in cancer development for their ability to alter the chromatin's state and influence gene expression (García et al., 2016). KDM4A, B, C, and D's expression is tightly regulated in non-neoplastic tissues but often deregulated in several neoplasias such as prostate, liver, bladder, colorectal, squamous cell carcinomas, acute myeloid leukemia, breast, lung and ovarian cancer (Guerra-Calderas et al., 2015, 2018; Lu et al., 2015; Lin et al., 2019; Chen et al., 2020; Wu et al., 2021). KDM4E's expression has only been detected in testis, however, its physiological role remains unknown (Hillringhaus et al., 2011). In this study, using large RNAseq tumor and non-neoplastic tissue datasets, we show that KDM4 proteins are relevant in different neoplasias and potential drug targets for therapy. One of the widest sources of novel biologically active molecules are natural compounds. These have been used for centuries to treat a wide range of diseases, including cancer (Gómez-Cansino et al., 2017; Gutiérrez-Rebolledo et al., 2017; Fang et al., 2020). Importantly, plenty of natural compounds are known to interfere with epigenetic processes; for example, flavonoids are compounds found in black raspberry (and many other plants) which inhibit DNA methyltransferase 1 (DNMT1) activity and enhance the expression of tumor suppressor genes (Wang et al., 2013). Nevertheless, although there are reports about natural molecules that could interfere with KDM4 subfamily activity, no direct natural inhibitors are known so far (Guillade et al., 2018).

Since our main interest is to propose natural KDM4 inhibitors, we used the COlleCtion of Open Natural Products (COCONUT), which gathers 406,744 natural products from over 50 different databases, where nearly half the compounds come mainly from plants, fungi, bacteria, and to a lesser extent, from animal or marine origins (Capecchi and Reymond, 2021; Sorokina et al., 2021). Most of these compounds (Sorokina et al., 2021) have been used as traditional medicine in China, India (Ayurveda), Japan (Kampo),

Korea, Mexico, among other countries (Yuan et al., 2016; Gutiérrez-Rebolledo et al., 2017) and come from Asia, Africa, Brazil, and Mexico (Sorokina et al., 2021). The former indicates that this database is a very powerful bioinformatic tool for natural compound screening. Also, in this work, we included the DrugBank and FDA databases, which have been the first-line source for drug repurposing. When compared against the FDA and DrugBank compounds, the molecules from the COCONUT database stood out in the molecular docking analyses against the KDM4 subfamily, which further suggests that natural compounds could be a rich source of anticancer therapies (Pushpakom et al., 2019).

However, a challenge faced during the development of specific inhibitors is the resemblance of JMJC family members' catalytic sites. Since these proteins share a catalytic mechanism, their active sites have a high resemblance, which complicates the design of ligands that could be specific for a single enzyme (Markolovic et al., 2016). The KDM4 subfamily active site consists of a TIM-barrel fold (16 beta-sheets and 15 alpha helix), which is a usual structural pattern in proteins that allows a wide assortment of functions (Romero-Romero et al., 2021). Thus, the TIM-barrel fold pattern is a challenge for drug design since ligands could bind to different proteins. In this sense, it is relevant that a specific binding mechanism with a competitive inhibitor is established for one or some KDM4 proteins. In fact, most of the KDM4 inhibitors reported are known to target other KDMs which limits their use for cancer treatment (Chin and Han, 2015; Baby et al., 2021). Furthermore, we showed that the KDM4 subfamily's expression is heterogeneous among different cancer types, which adds another layer of complexity to the search for inhibitor molecules that could favor the treatment of neoplasms where the KDM4 proteins are relevant.

Since the KDM4 subfamily is a promising therapeutic target for drug design, a wide number of synthetic and nature-inspired molecules have been explored. Among them, it has been proposed that catechol and flavonoids as structural scaffolds, these kinds of molecules have gained attention because of their high content of OH with redox capacity that can also act as free radical regulators (Baby et al., 2021). Such functional groups can favor their interaction with high electronegative residues located in the active site of KDM4 proteins, as shown in this study. Furthermore, the metallic cofactors in KDM4 active sites, assist the catalytic mechanism of electrons transfer during lysine methylation, thus OH groups can form coordination bonds that compete with KDM4's natural substrates and allow a greater affinity than the substrate itself (Warshakoon et al., 2006). It has been reported that coordination bonds between catechol-containing groups (such as flavonoids, or phenols) and KDM4 metal cofactors lead to an enhancement of interaction forces (Xu, 2013). An example of the former is the epigallocatechin gallate (EGCG), a secondary metabolite derived from the tea plant (*Camellia sinensis*) that contains catechol and whose effect has been studied in various epigenetic processes (Fang et al., 2003; Choi et al., 2009). EGCG is a compound included in FDA and COCONUT databases which showed a favorable FRED/Chemgauss4 score on its interaction with KDM4 (**Figure 4C**, compound 3), this suggests that it may be a promising inhibitor candidate for these enzymes. On the other hand, it has been reported that EGCG chelates divalent metals, including zinc, and it has been proposed in many clinical assays as an adjuvant in multiple processes (Shirakami and Shimizu, 2018). Another variant of catechol, pyrogallol, which contains 3 OH groups instead of 2, has been studied as a therapeutic agent in lung cancer cell lines showing cytotoxic effects (Yang et al., 2009).

In addition to the contribution of coordination bonds that favor specificity, there are other non-covalent binding forces that can also have an impact on specificity and binding affinity such as salt bridges, hydrophobic interactions, hydrogen bonds, and pi-stacking interactions. In particular, pi-stacking interactions among aromatic rings are an important factor in the protein-ligand complex formation; in such interactions, the geometric orientation of the rings change the dipole attraction forces among them as well as the hydrophobic and van der Waals forces rearrange (Churchill and Wetmore, 2009; Wilson et al., 2014; Houser et al., 2020). The presence of five aromatic residues and one histidine in KDM4 active sites promotes a favorable environment to design specific inhibitors (Churchill and Wetmore, 2009; Brylinski, 2018). The former is evident for linear molecules (such as the ones in clusters 1, 2, 5, and 6 in **Figure 4C**), since those obtained a lower score due to their interaction with the metal through ionic groups of amines, carboxyl, or phosphates groups and pi-stacking interactions are not present. While, molecules of clusters 3, 14, and 15 have higher scores due to the presence of aromatic rings that favor pi-stacking interaction. Similar results have been reported for KDM4 proteins and tetrazolyl hydrazide inhibitors which have an aromatic ring and amine functional groups that interact with the protein's metal cofactors (Małecki et al., 2019). Metal coordination capability of sugars coupled with flavonoids favors the physicochemical properties of the KDM4 active sites and provides an opportunity for the development of a new generation of *de novo* molecules for cancer treatment. One of the limitations of our study is that it is not supported by experimental assays, but its strength is that this work is the first step towards an experimental approach that could contribute to the treatment of different neoplasms.

Our results also show that members of the KDM4 subfamily are promising drug targets for the development of therapeutic alternatives in different types of cancer. Since specificity is hard to achieve for KDM inhibitors, we aimed to use this to our advantage searching for ligands that could target all the KDM4s relevant for a specific neoplasm without altering the others. We highlight the importance of natural compounds against KDM4 subfamily members, not only because of their high potential as inhibitors but also because these compounds could contribute to an integrative cancer treatment. As shown in this work the identified molecules could have an amplified therapeutic effect by modulating, not only KDM4 functions but entire cellular processes, by modifying the activity of proteins involved in the same pathways. This mechanism of action has been proposed for other diseases and protein targets before (Cheng et al., 2019); however, the study of KDM4 inhibitors remains approached without considering the molecular context required for their proper function (Chin and Han, 2015; Baby et al., 2021). Overall, our data suggest that natural compounds could be used as adjuvant therapies in cancer, which opens a new window of opportunities for the

search of KDM4 subfamily inhibitors and contributes to the search of novel cancer therapies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

MS-A, AM-M and EO-G contributed to the data collection and analyses. MS-A, AM-M, EO-G, GP-H and ES-R contributed to manuscript writing. GP-H and ES-R were in charge of overall direction, planning, and supervision.

FUNDING

This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT) Fondo CB-SEP-CONACyT 284748 and PROMEP 250690 to ES-R. Also, GP-H and ES-R were supported by the Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana (UAM)-Cuajimalpa. All the computational analyses were performed on the esrs-epigenetics server from Universidad Autónoma Metropolitana (UAM)-Cuajimalpa funded by CONACyT through the Apoyo para proyectos de investigación científica, desarrollo tecnológico e innovación en salud ante la contingencia por COVID-19, grant number 00312021. AM-M and MS-A are doctoral students from "Programa de maestría y doctorado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México (UNAM)" and received a PhD fellowship funding from CONACYT (CVU894530 to AM-M and CVU659273 to MS-A respectively).

ACKNOWLEDGMENTS

We thank OpenEye Scientific software for the academic license provided to the laboratory of GP-H and Dr. Elizabeth Del Moral Ramírez for her helpful contributions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.860924/full#supplementary-material

REFERENCES

Abraham, M. J., Murtol, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* 1-2, 19–25. doi:10.1016/j.softx.2015.06.001

Agger, K., Nishimura, K., Miyagi, S., Messling, J.-E., Rasmussen, K. D., and Helin, K. (2019). The KDM4/JMJD2 Histone Demethylases Are Required for Hematopoietic Stem Cell Maintenance. *Blood* 134 (14), 1154–1158. doi:10.1182/blood.2019000855

Alcaraz, N., Hartebrodt, A., and List, M. (2020). De Novo Pathway Enrichment with KeyPathwayMiner. *Methods Mol. Biol.* 2074, 181–199. doi:10.1007/978-1-4939-9873-9_14

Baby, S., Valapil, D. G., and Shankaraiah, N. (2021). Unravelling KDM4 Histone Demethylase Inhibitors for Cancer Therapy. *Drug Discov. Today* 26 (8), 1841–1856. doi:10.1016/j.drudis.2021.05.015

Bavetsias, V., Lanigan, R. M., Ruda, G. F., Atrash, B., McLaughlin, M. G., Tumber, A., et al. (2016). 8-Substituted Pyrido[3,4-D]pyrimidin-4(3h)-One Derivatives as Potent, Cell Permeable, KDM4 (JMJD2) and KDM5 (JARID1) Histone Lysine Demethylase Inhibitors. *J. Med. Chem.* 59 (4), 1388–1409. doi:10.1021/acs.jmedchem.5b01635

Brylinski, M. (2018). Aromatic Interactions at the Ligand-Protein Interface: Implications for the Development of Docking Scoring Functions. *Chem. Biol. Drug Des.* 91 (2), 380–390. doi:10.1111/cbdd.13084

Capecchi, A., and Reymond, J.-L. (2021). Classifying Natural Products from Plants, Fungi or Bacteria Using the COCONUT Database and Machine Learning. *J. Cheminform* 13 (1), 82. doi:10.1186/s13321-021-00559-3

Chen, G. Q., Ye, P., Ling, R. S., Zeng, F., Zhu, X. S., Chen, L., et al. (2020). Histone Demethylase KDM4C Is Required for Ovarian Cancer Stem Cell Maintenance. *Stem Cell Int* 2020, 8860185. doi:10.1155/2020/8860185

Cheng, F., Kova´cs, I. A., and Baraba´si, A.-L. (2019). Network-based Prediction of Drug Combinations. *Nat. Commun.* 10 (1), 1197. doi:10.1038/s41467-019-09186-x

Chin, Y.-W., and Han, S.-Y. (2015). KDM4 Histone Demethylase Inhibitors for Anti-cancer Agents: a Patent Review. *Expert Opin. Ther. patents* 25 (2), 135–144. doi:10.1517/13543776.2014.991310

Choi, K.-C., Jung, M. G., Lee, Y.-H., Yoon, J. C., Kwon, S. H., Kang, H.-B., et al. (2009). Epigallocatechin-3-Gallate, a Histone Acetyltransferase Inhibitor, Inhibits EBV-Induced B Lymphocyte Transformation via Suppression of RelA Acetylation. *Cancer Res.* 69 (2), 583–592. doi:10.1158/0008-5472.can-08-2442

Chu, C.-H., Wang, L.-Y., Hsu, K.-C., Chen, C.-C., Cheng, H.-H., Wang, S.-M., et al. (2014). KDM4B as a Target for Prostate Cancer: Structural Analysis and Selective Inhibition by a Novel Inhibitor. *J. Med. Chem.* 57 (14), 5975–5985. doi:10.1021/jm500249n

Churchill, C. D. M., and Wetmore, S. D. (2009). Noncovalent Interactions Involving Histidine: The Effect of Charge on π–π Stacking and T-Shaped Interactions with the DNA Nucleobases. *J. Phys. Chem. B* 113 (49), 16046–16058. doi:10.1021/jp907887y

Cortopassi, W. A., Simion, R., Honsby, C. E., França, T. C. C., and Paton, R. S. (2015). Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment. *Chem. Eur. J.* 21 (52), 18869. doi:10.1002/chem.201504536

Deng, Y., Li, M., Zhuo, M., Guo, P., Chen, Q., Mo, P., et al. (2021). Histone Demethylase JMJD2D Promotes the Self-Renewal of Liver Cancer Stem-like Cells by Enhancing EpCAM and Sox9 Expression. *J. Biol. Chem.* 296, 100121. doi:10.1074/jbc.ra120.015335

Fang, M. Z., Wang, Y., Ai, N., Hou, Z., Sun, Y., Lu, H., et al. (2003). Tea Polyphenol (-)-Epigallocatechin-3-Gallate Inhibits DNA Methyltransferase and Reactivates Methylation-Silenced Genes in Cancer Cell Lines. *Cancer Res.* 63 (22), 7563–7570.

Fang, Y., Yang, C., Yu, Z., Li, X., Mu, Q., Liao, G., et al. (2021). Natural Products as LSD1 Inhibitors for Cancer Therapy. *Acta Pharmaceutica Sinica B* 11, 621–631. doi:10.1016/j.apsb.2020.06.007

García, M. A. R., and Fueyo, M. A. (2016). "Lysine Demethylases," in *Chromatin Signaling and Diseases*. Editors O. Binda and M. E. Fernandez-Zapico (Boston: Academic Press), 179–194. doi:10.1016/b978-0-12-802389-1.00010-1

Gómez-cansino, R., Guzmán-gutiérrez, S. L., Campos-lara, M. G., Espitia-pinzón, C. I., and Reyes-chilpa, R. (2017). Natural Compounds from Mexican Medicinal Plants as Potential Drug Leads for Anti-tuberculosis Drugs. *Acad. Bras. Ciênc.* 89 (1), 31–43. doi:10.1590/0001-3765201720160298

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., et al. (2020). Visualizing and Interpreting Cancer Genomics Data via the Xena Platform. *Nat. Biotechnol.* 38 (6), 675–678. doi:10.1038/s41587-020-0546-8

Grambsch, P. M., and Therneau, T. M. (2000). "Modeling Survival Data: Extending the Cox Model," in *Statistics for Biology and Health [Preprint]* (Berlin/Heidelberg, Germany: Springer).

Guerra-Calderas, L., González-Barrios, R., Herrera, L. A., Cantú de León, D., and Soto-Reyes, E. (2015). The Role of the Histone Demethylase KDM4A in Cancer. *Cancer Genet.* 208 (5), 215–224. doi:10.1016/j.cancergen.2014.11.001

Guerra-Calderas, L., González-Barrios, R., Patiño, C. C., Alcaraz, N., Salgado-Albarrán, M., de León, D. C., et al. (2018). CTCF-KDM4A Complex Correlates with Histone Modifications that Negatively Regulate CHD5 Gene Expression in Cancer Cell Lines. *Oncotarget* 9 (24), 17028–17042. doi:10.18632/oncotarget.24798

Guillade, L., Sarno, F., Tarhonskaya, H., Nebbioso, A., Alvarez, S., Kawamura, A., et al. (2018). Synthesis and Biological Evaluation of Tripartin, a Putative KDM4 Natural Product Inhibitor, and 1-Dichloromethylinden-1-Ol Analogues. *ChemMedChem* 13 (18), 1949–1956. doi:10.1002/cmdc.201800377

Gutiérrez-Rebolledo, G. A., Drier-Jonas, S., and Jiménez-Arellanes, M. A. (2017). Natural Compounds and Extracts from Mexican Medicinal Plants with Anti-leishmaniasis Activity: An Update. *Asian Pac. J. Trop. Med.* 10 (12), 1105–1110. doi:10.1016/j.apjtm.2017.10.016

Harris, W. J., Huang, X., Lynch, J. T., Spencer, G. J., Hitchin, J. R., Li, Y., et al. (2012). The Histone Demethylase KDM1A Sustains the Oncogenic Potential of MLL-AF9 Leukemia Stem Cells. *Cancer cell* 21 (4), 473–487. doi:10.1016/j.ccr.2012.03.014

Hillringhaus, L., Yue, W. W., Rose, N. R., Ng, S. S., Gileadi, C., Loenarz, C., et al. (2011). Structural and Evolutionary Basis for the Dual Substrate Selectivity of Human KDM4 Histone Demethylase Family. *J. Biol. Chem.* 286 (48), 41616–41625. doi:10.1074/jbc.m111.283689

Houser, J., Kozmon, S., Mishra, D., Hammerová, Z., Wimmerová, M., and Koča, J. (2020). The CH-π Interaction in Protein-Carbohydrate Binding: Bioinformatics and *In Vitro* Quantification. *Chem. Eur. J.* 26, 10769–10780. doi:10.1002/chem.202000593

Iwamori, N., Zhao, M., Meistrich, M. L., and Matzuk, M. M. (2011). The Testis-Enriched Histone Demethylase, KDM4D, Regulates Methylation of Histone H3 Lysine 9 during Spermatogenesis in the Mouse but Is Dispensable for Fertility1. *Biol. Reprod.* 84 (6), 1225–1234. doi:10.1095/biolreprod.110.088955

Jakalian, A., Jack, D. B., and Bayly, C. I. (2002). Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* 23 (16), 1623–1641. doi:10.1002/jcc.10128

Jambhekar, A., Dhall, A., and Shi, Y. (2019). Roles and Regulation of Histone Methylation in Animal Development. *Nat. Rev. Mol. Cel Biol* 20 (10), 625–641. doi:10.1038/s41580-019-0151-1

Katoh, M., and Katoh, M. (2004). Identification and Characterization of JMJD2 Family Genes In Silico, International Journal of Oncology [Preprint]. Available at: https://www.spandidos-publications.com/ijo/24/6/1623.

Kawazu, M., Saso, K., Tong, K. I., McQuire, T., Goto, K., Son, D.-O., et al. (2011). Histone Demethylase JMJD2B Functions as a Co-factor of Estrogen Receptor in Breast Cancer Proliferation and Mammary Gland Development. *PloS one* 6 (3), e17830.

Kim, S.-H., Kwon, S. H., Park, S.-H., Lee, J. K., Bang, H.-S., Nam, S.-J., et al. (2013). Tripartin, a Histone Demethylase Inhibitor from a Bacterium Associated with a Dung Beetle Larva. *Org. Lett.* 15 (8), 1834–1837. doi:10.1021/ol4004417

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 - an R Package for Gene List Functional Enrichment Analysis and Namespace Conversion Toolset g:Profiler. *F1000Res* 9, ELIXIR-709. doi:10.12688/f1000research.24956.2

Kotlyar, M., Pastrello, C., Malik, Z., and Jurisica, I. (2019). IID 2018 Update: Context-specific Physical Protein-Protein Interactions in Human, Model Organisms and Domesticated Species. *Nucleic Acids Res.* 47, D581–D589. doi:10.1093/nar/gky1037

Krishnan, S., and Trievel, R. C. (2013). Structural and Functional Analysis of JMJD2D Reveals Molecular Basis for Site-Specific Demethylation Among JMJD2 Demethylases. *Structure* 21 (1), 98–108. doi:10.1016/j.str.2012.10.018

Kumari, R., Kumar, R., and Lynn, A. (2014). g_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* 54 (7), 1951–1962. doi:10.1021/ci500020m

Labbé, R. M., Holowatyj, A., and Yang, Z. Q. (2013). Histone Lysine Demethylase (KDM) Subfamily 4: Structures, Functions and Therapeutic Potential. *Am. J. Transl Res.* 6 (1), 1–15.

Lee, D. H., Kim, G. W., Jeon, Y. H., Yoo, J., Lee, S. W., and Kwon, S. H. (2020). Advances in Histone Demethylase KDM4 as Cancer Therapeutic Targets. *FASEB j.* 34 (3), 3461–3484. doi:10.1096/fj.201902584r

Lee, D. H., Kim, G. W., Yoo, J., Lee, S. W., Jeon, Y. H., Kim, S. Y., et al. (2021). Histone Demethylase KDM4C Controls Tumorigenesis of Glioblastoma by Epigenetically Regulating P53 and C-Myc. *Cell Death Dis* 12 (1), 89. doi:10.1038/s41419-020-03380-2

Li, M., Deng, Y., Zhuo, M., Zhou, H., Kong, X., Xia, X., et al. (2020). Demethylase-independent Function of JMJD2D as a Novel Antagonist of P53 to Promote Liver Cancer Initiation and Progression. *Theranostics* 10 (19), 8863–8879. doi:10.7150/thno.45581

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Lin, C.-Y., Wang, B.-J., Chen, B.-C., Tseng, J.-C., Jiang, S. S., Tsai, K. K., et al. (2019). Histone Demethylase KDM4C Stimulates the Proliferation of Prostate Cancer Cells via Activation of AKT and C-Myc. *Cancers* 11, 1785. doi:10.3390/cancers11111785

Liu, X., Wang, Y., Gao, Y., Su, J., Zhang, J., Xing, X., et al. (2018). H3K9 Demethylase KDM4E Is an Epigenetic Regulator for Bovine Embryonic Development and a Defective Factor for Nuclear Reprogramming. *Development* 145, dev158261. doi:10.1242/dev.158261

Liu, X., Zhang, Q., Zhao, Y., Xun, J., Wu, H., and Feng, H. (2020). Association of JMJD2B and Hypoxia-Inducible Factor 1 Expressions with Poor Prognosis in Osteosarcoma. *Anal. Cel Pathol (Amst)* 2020, 2563208. doi:10.1155/2020/2563208

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Lu, J.-W., Ho, Y.-J., Lin, L.-I., Huang, Y.-C., Yeh, K.-T., Lin, Y.-H., et al. (2015). JMJD2B as a Potential Diagnostic Immunohistochemical Marker for Hepatocellular Carcinoma: A Tissue Microarray-Based Study. *Acta Histochem.* 117 (1), 14–19. doi:10.1016/j.acthis.2014.10.002

Lv, M., and Liu, Q. (2021). JMJD2C Triggers the Growth of Multiple Myeloma Cells via Activation of β-catenin. *Oncol. Rep.* 45 (3), 1162–1170. doi:10.3892/or.2021.7934

Małecki, P. H., Rüger, N., Roatsch, M., Krylova, O., Link, A., Jung, M., et al. (2019). Structure-Based Screening of Tetrazolylhydrazide Inhibitors versus KDM4 Histone Demethylases. *ChemMedChem* 14 (21), 1828–1839. doi:10.1002/cmdc.201900441

Massett, M. E., Monaghan, L., Patterson, S., Mannion, N., Bunschoten, R. P., Hoose, A., et al. (2021). A KDM4A-PAF1-Mediated Epigenomic Network Is Essential for Acute Myeloid Leukemia Cell Self-Renewal and Survival. *Cel Death Dis* 12 (6), 573. doi:10.1038/s41419-021-03738-0

Metzger, E., Stepputtis, S. S., Strietz, J., Preca, B.-T., Urban, S., Willmann, D., et al. (2017). KDM4 Inhibition Targets Breast Cancer Stem-like Cells. *Cancer Res.* 77 (21), 5900–5912. doi:10.1158/0008-5472.can-17-1754

Mu, H., Xiang, L., Li, S., Rao, D., Wang, S., and Yu, K. (2019). MiR-10a Functions as a Tumor Suppressor in Prostate Cancer via Targeting KDM4A. *J. Cel Biochem* 120 (4), 4987–4997. doi:10.1002/jcb.27774

OpenEye Scientific Software (2021). Cheminformatics Software. Available at: https://www.eyesopen.com (Accessed December 8, 2021).

Portela, A., and Esteller, M. (2010). Epigenetic Modifications and Human Disease. *Nat. Biotechnol.* 28 (10), 1057–1068. doi:10.1038/nbt.1685

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug Repurposing: Progress, Challenges and Recommendations. *Nat. Rev. Drug Discov.* 18 (1), 41–58. doi:10.1038/nrd.2018.168

Ramanan, R., Chaturvedi, S. S., Lehnert, N., Schofield, C. J., Karabencheva-Christova, T. G., and Christov, C. Z. (2020). Catalysis by the JmjC Histone Demethylase KDM4A Integrates Substrate Dynamics, Correlated Motions and Molecular Orbital Control. *Chem. Sci.* 11 (36), 9950–9961. doi:10.1039/d0sc03713c

Ramanan, R., Waheed, S. O., Schofield, C. J., and Christov, C. Z. (2021). What Is the Catalytic Mechanism of Enzymatic Histone N-Methyl Arginine Demethylation and Can it Be Influenced by an External Electric Field? *Chem. Eur. J.* 27 (46), 11750. doi:10.1002/chem.202102158

Romero-Romero, S., Kordes, S., Michel, F., and Höcker, B. (2021). Evolution, Folding, and Design of TIM Barrels and Related Proteins. *Curr. Opin. Struct. Biol.* 68, 94–104. doi:10.1016/j.sbi.2020.12.007

Rotili, D., and Mai, A. (2011). Targeting Histone Demethylases: A New Avenue for the Fight against Cancer. *Genes & Cancer* 2 (6), 663–679. doi:10.1177/1947601911417976

Rybaczyk, L. A., Bashaw, M. J., Pathak, D. R., and Huang, K. (2008). An Indicator of Cancer: Downregulation of Monoamine Oxidase-A in Multiple Organs and Species. *BMC genomics* 9, 134. doi:10.1186/1471-2164-9-134

Sadegh, S., Skelton, J., Anastasi, E., Bernett, J., Blumenthal, D. B., Galindez, G., et al. (2021). Network Medicine for Disease Module Identification and Drug Repurposing with the NeDRex Platform. *Nat. Commun.* 12 (1), 6848. doi:10.1038/s41467-021-27138-2

Sander, T., Freyss, J., von Korff, M., and Rufener, C. (2015). DataWarrior: An Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* 55 (2), 460–473. doi:10.1021/ci500588j

Sha, J., Han, Q., Chi, C., Zhu, Y., Pan, J., Dong, B., et al. (2020). Upregulated KDM4B Promotes Prostate Cancer Cell Proliferation by Activating Autophagy. *J. Cel Physiol* 235 (3), 2129–2138. doi:10.1002/jcp.29117

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Shih, J. C. (2018). Monoamine Oxidase Isoenzymes: Genes, Functions and Targets for Behavior and Cancer Therapy. *J. Neural Transm.* 125 (11), 1553–1566. doi:10.1007/s00702-018-1927-8

Shin, S., and Janknecht, R. (2007a). Activation of Androgen Receptor by Histone Demethylases JMJD2A and JMJD2D. *Biochem. biophysical Res. Commun.* 359 (3), 742–746. doi:10.1016/j.bbrc.2007.05.179

Shin, S., and Janknecht, R. (2007b). Diversity within the JMJD2 Histone Demethylase Family. *Biochem. biophysical Res. Commun.* 353 (4), 973–977. doi:10.1016/j.bbrc.2006.12.147

Shirakami, Y., and Shimizu, M. (2018). Possible Mechanisms of Green Tea and its Constituents against Cancer. *Molecules* 23, 2284. doi:10.3390/molecules23092284

Silva, A. W. S. D., and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* 5 (1), 367–368. doi:10.1186/1756-0500-5-367

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., and Steinbeck, C. (2021). COCONUT Online: Collection of Open Natural Products Database. *J. Cheminform* 13 (1), 2. doi:10.1186/s13321-020-00478-9

Stavropoulos, P., Blobel, G., and Hoelz, A. (2006). Crystal Structure and Mechanism of Human Lysine-specific Demethylase-1. *Nat. Struct. Mol. Biol.* 13 (7), 626–632. doi:10.1038/nsmb1113

Sterling, J., Menezes, S. V., Abbassi, R. H., and Munoz, L. (2020). Histone Lysine Demethylases and Their Functions in Cancer. *Int. J. Cancer* 148, 2375–2388. doi:10.1002/ijc.33375

Tsukada, Y.-I., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., et al. (2006). Histone Demethylation by a Family of JmjC Domain-Containing Proteins. *Nature* 439 (7078), 811–816. doi:10.1038/nature04433

Vidossich, P., and Magistrato, A. (2014). QM/MM Molecular Dynamics Studies of Metal Binding Proteins. *Biomolecules* 4 (3), 616–645. doi:10.3390/biom4030616

Vivian, J., Rao, A. A., Nothaft, F. A., Ketchum, C., Armstrong, J., Novak, A., et al. (2017). Toil Enables Reproducible, Open Source, Big Biomedical Data Analyses. *Nat. Biotechnol.* 35 (4), 314–316. doi:10.1038/nbt.3772

von Korff, M., Freyss, J., and Sander, T. (2009). Comparison of Ligand- and Structure-Based Virtual Screening on the DUD Data Set. *J. Chem. Inf. Model.* 49 (2), 209–231. doi:10.1021/ci800303k

Wagner, K. W., Alam, H., Dhar, S. S., Giri, U., Li, N., Wei, Y., et al. (2013). KDM2A Promotes Lung Tumorigenesis by Epigenetically Enhancing ERK1/2 Signaling. *J. Clin. Invest.* 123 (12), 5231–5246. doi:10.1172/jci68642

Walport, L. J., Hopkinson, R. J., Chowdhury, R., Schiller, R., Ge, W., Kawamura, A., et al. (2016). Arginine Demethylation Is Catalysed by a Subset of JmjC Histone Lysine Demethylases. *Nat. Commun.* 7 (1), 11974. doi:10.1038/ncomms11974

Wang, L.-S., Kuo, C.-T., Cho, S.-J., Seguin, C., Siddiqui, J., Stoner, K., et al. (2013). Black Raspberry-Derived Anthocyanins Demethylate Tumor Suppressor Genes through the Inhibition of DNMT1 and DNMT3B in Colon Cancer Cells. *Nutr. Cancer* 65 (1), 118–125. doi:10.1080/01635581.2013.741759

Warshakoon, N. C., Wu, S., Boyer, A., Kawamoto, R., Sheville, J., Renock, S., et al. (2006). Structure-based Design, Synthesis, and SAR Evaluation of a New Series of 8-hydroxyquinolines as HIF-1α Prolyl Hydroxylase Inhibitors. *Bioorg. Med. Chem. Lett.* 16 (21), 5517–5522. doi:10.1016/j.bmcl.2006.08.040

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427

Westaway, S. M., Preston, A. G., Barker, M. D., Brown, F., Brown, J. A., Campbell, M., et al. (2016). Cell Penetrant Inhibitors of the KDM4 and KDM5 Families of Histone Lysine Demethylases. 1. 3-Amino-4-Pyridine Carboxylate Derivatives. *J. Med. Chem.* 59 (4), 1357–1369. doi:10.1021/acs.jmedchem.5b01537

Wilson, K. A., Kellie, J. L., and Wetmore, S. D. (2014). DNA-protein π-interactions in Nature: Abundance, Structure, Composition and Strength of Contacts between Aromatic Amino Acids and DNA Nucleobases or Deoxyribose Sugar. *Nucleic Acids Res.* 42 (10), 6726–6741. doi:10.1093/nar/gku269

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037

Wu, M.-C., Cheng, H.-H., Yeh, T.-S., Li, Y.-C., Chen, T.-J., Sit, W. Y., et al. (2019). KDM4B Is a Coactivator of C-Jun and Involved in Gastric Carcinogenesis. *Cel Death Dis* 10 (2), 68. doi:10.1038/s41419-019-1305-y

Wu, W., Cao, X., and Mo, L. (2021). Overexpression of KDM4D Promotes Acute Myeloid Leukemia Cell Development by Activating MCL-1. *Am. J. Transl Res.* 13 (4), 2308–2319.

Xu, Z. (2013). Mechanics of Metal-Catecholate Complexes: The Roles of Coordination State and Metal Types. *Sci. Rep.* 3, 2914. doi:10.1038/srep02914

Yang, C.-J., Wang, C.-S., Hung, J.-Y., Huang, H.-W., Chia, Y.-C., Wang, P.-H., et al. (2009). Pyrogallol Induces G2-M Arrest in Human Lung Cancer Cells and Inhibits Tumor Growth in an Animal Model. *Lung cancer* 66 (2), 162–168. doi:10.1016/j.lungcan.2009.01.016

Yang, J., Jubb, A. M., Pike, L., Buffa, F. M., Turley, H., Baban, D., et al. (2010). The Histone Demethylase JMJD2B Is Regulated by Estrogen Receptor α and Hypoxia, and Is a Key Mediator of Estrogen Induced Growth. *Cancer Res.* 70 (16), 6456–6466. doi:10.1158/0008-5472.can-10-0413

Yang, Y., Chia, Z., Gao, R., and Lei, Z. (2018). The Roles of Natural Compounds in Epigenetics. *Nat. Prod. Commun.* 13 (8), 1067–1072. doi:10.1177/1934578x1801300835

Yuan, H., Ma, Q., Ye, L., and Piao, G. (2016). The Traditional Medicine and Modern Medicine from Natural Products. *Molecules* 21, 559. doi:10.3390/molecules21050559

Zhao, Y., and Garcia, B. A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor Perspect. Biol.* 7 (9), a025064. doi:10.1101/cshperspect.a025064

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

cells

MDPI

Review

CTCF and Its Multi-Partner Network for Chromatin Regulation

Aylin Del Moral-Morales [1,2], Marisol Salgado-Albarrán [1,2], Yesennia Sánchez-Pérez [3], Nina Kerstin Wenke [2], Jan Baumbach [2,4] and Ernesto Soto-Reyes [1,*]

1 Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana-Cuajimalpa (UAM-C), Mexico City 05348, Mexico
2 Institute for Computational Systems Biology, University of Hamburg, D-22607 Hamburg, Germany
3 Subdirección de Investigación, Instituto Nacional de Cancerología, Mexico City 14080, Mexico
4 Computational BioMedicine Lab., University of Southern Denmark, DK-5230 Odense, Denmark
* Correspondence: esotoreyes@cua.uam.mx

Abstract: Architectural proteins are essential epigenetic regulators that play a critical role in organizing chromatin and controlling gene expression. CTCF (CCCTC-binding factor) is a key architectural protein responsible for maintaining the intricate 3D structure of chromatin. Because of its multivalent properties and plasticity to bind various sequences, CTCF is similar to a Swiss knife for genome organization. Despite the importance of this protein, its mechanisms of action are not fully elucidated. It has been hypothesized that its versatility is achieved through interaction with multiple partners, forming a complex network that regulates chromatin folding within the nucleus. In this review, we delve into CTCF's interactions with other molecules involved in epigenetic processes, particularly histone and DNA demethylases, as well as several long non-coding RNAs (lncRNAs) that are able to recruit CTCF. Our review highlights the importance of CTCF partners to shed light on chromatin regulation and pave the way for future exploration of the mechanisms that enable the finely-tuned role of CTCF as a master regulator of chromatin.

Keywords: CTCF; epigenetics; chromatin regulation; histone; demethylases; lncRNAs; TET; KDM; BORIS; CTCF-s

Citation: Del Moral-Morales, A.; Salgado-Albarrán, M.; Sánchez-Pérez, Y.; Wenke, N.K.; Baumbach, J.; Soto-Reyes, E. CTCF and Its Multi-Partner Network for Chromatin Regulation. *Cells* **2023**, *12*, 1357. https://doi.org/10.3390/cells12101357

Academic Editor: Ali Hamiche

Received: 31 March 2023
Revised: 5 May 2023
Accepted: 9 May 2023
Published: 10 May 2023

1. Introduction

Chromatin, a macromolecular complex of DNA, RNA, and proteins, provides a framework for the packaging of genetic material within the cell nucleus. Its organization plays a crucial role in gene expression and is regulated by a diverse array of protein complexes in response to a dynamic code of histone posttranslational modifications and DNA modifications [1]. CTCF (CCCTC-binding factor) is a crucial architectural protein believed to play a critical role in maintaining chromatin organization through its interactions with various protein complexes [2]. Among other functions, CTCF is a versatile protein known to participate in various processes related to the chromatin structure, including insulation [3], alternative splicing [4–6], transcriptional activation [7], and chromatin loop formation [8]. It is not clear how CTCF has such a dynamic range of functions; however, the response to this question may lie in the context-dependent interactions of CTCF with several protein partners.

Epigenetic complexes, which regulate histone post-translational modifications and DNA methylation, usually contain enzymes that chemically modify the amino-terminal ends of histones, forming a code that determines the chromatin state through a system of writing, reading, and erasing complexes [9–11]. The mechanisms by which epigenetic components are recruited to specific regions of the genome have not been fully understood, mainly due to the lack of DNA binding domains in most proteins with epigenetic functions [12]. This is why CTCF is a fundamental protein since it could be the bridge between many epigenetic factors and the DNA [13]. The importance of CTCF protein–protein interactions is highlighted by BORIS (Brother of the Regulator of Imprinted Sites),

the paralogous protein of CTCF. The similarity of DNA-binding domains between BORIS and CTCF suggests they share similar targets in the genome [14]; however, due to the low degree of conservation between their terminal domains, it is believed that they interact with different cofactors, which cause them to have opposite consequences in gene expression and chromatin structure [15–17].

In addition, long non-coding RNAs (lncRNAs) have been described as crucial factors in chromatin architecture [18]. Recent evidence indicates that CTCF interacts with several lncRNAs that modulate its recruitment and binding to the DNA. Depletion of CTCF RNA binding domains impairs chromatin loop formation and alters transcriptional profiles [19,20]. Moreover, lncRNAs serve as a scaffold for the interaction of CTCF with other proteins in the form of RNA bridges [21] or could even cause it to detach from its DNA binding sites [22]. Without a doubt, CTCF depends on its interactions with other proteins and nucleic acids to exert a wide range of functions. In this review, we aim to shed light on the role of CTCF partners in shaping the 3D organization and gene regulation of chromatin, specifically those with epigenetic function.

2. CTCF Is a Multifaceted Protein

Originally, CTCF was described in chickens as a protein that binds to a region upstream of the c-myc promoter. Because that binding site has three regularly spaced repetitions of the sequence CCCTC, the protein was named CCCTC-binding factor or CTCF [23]. Later, it was found that CTCF is a ubiquitously expressed and highly conserved protein in vertebrates [14,24]. CTCF consists of 727 amino acids (aa) distributed in three domains; a zinc finger DNA-binding domain flanked by the intrinsically disordered N- and C-terminal regions (Figure 1a). The DNA binding domain of CTCF has 11 zinc fingers (ZF) which allow it to interact dynamically with the DNA [25–27]. CTCF uses different combinations of its ZF to recognize and bind to a variety of DNA sequences, which is why it is considered a multivalent protein [28,29]. However, around 80% of its target sequences contain the core motif 5′-CCACCAGGTGG-3′ that is recognized by ZFs 4 to 7. Unconserved flanking sequences can be recognized by ZF 1–2 or ZF 8–11, which helps to stabilize the CTCF-DNA complex [30–32]. A peculiarity of CTCF is that ZF1 and ZF10 have an RNA binding domain (RBD) which is used to interact with several lncRNAs, providing extra anchorage points for the protein [19,31].

CTCF has tens of thousands of genomic binding sites, some of which are conserved between species and tissues [33]. CTCF actions are dependent on its binding site location; which are mainly located in intergenic regions, although they could also be present in regulatory regions such as enhancers, gene promoters, and within gene bodies [34–36]. The main functions of CTCF include maintaining topologically associated domains (TADs), acting as a barrier to the spread of heterochromatic structures, and defining the boundaries between euchromatin and heterochromatin, for this reason, CTCF has been coined as an architectural protein [37–40]. CTCF also regulates DNA anchorage to cellular structures such as the nuclear lamina [37,38], acts as a protein insulator by controlling the interactions between enhancers and promoters [41], and can function as a scaffold protein for transcription factors [42–44] and epigenetic factors [45]. Based on the location of the CTCF in other genomic sites, it has also been demonstrated to be involved in processes such as alternative splicing by pausing RNA Polymerase II (RNAP II) binding to alternative exons, thus providing the required temporal context for co-transcriptional spliceosome formation at weak upstream splice sites [4]. CTCF also interacts with lncRNAs which is important for the transcriptional regulation of genes such as Xist, a lncRNA responsible for X chromosome inactivation. For this reason, CTCF has been considered a very versatile protein similar to a swiss army knife. A summary of its functions is shown in Figure 1b.

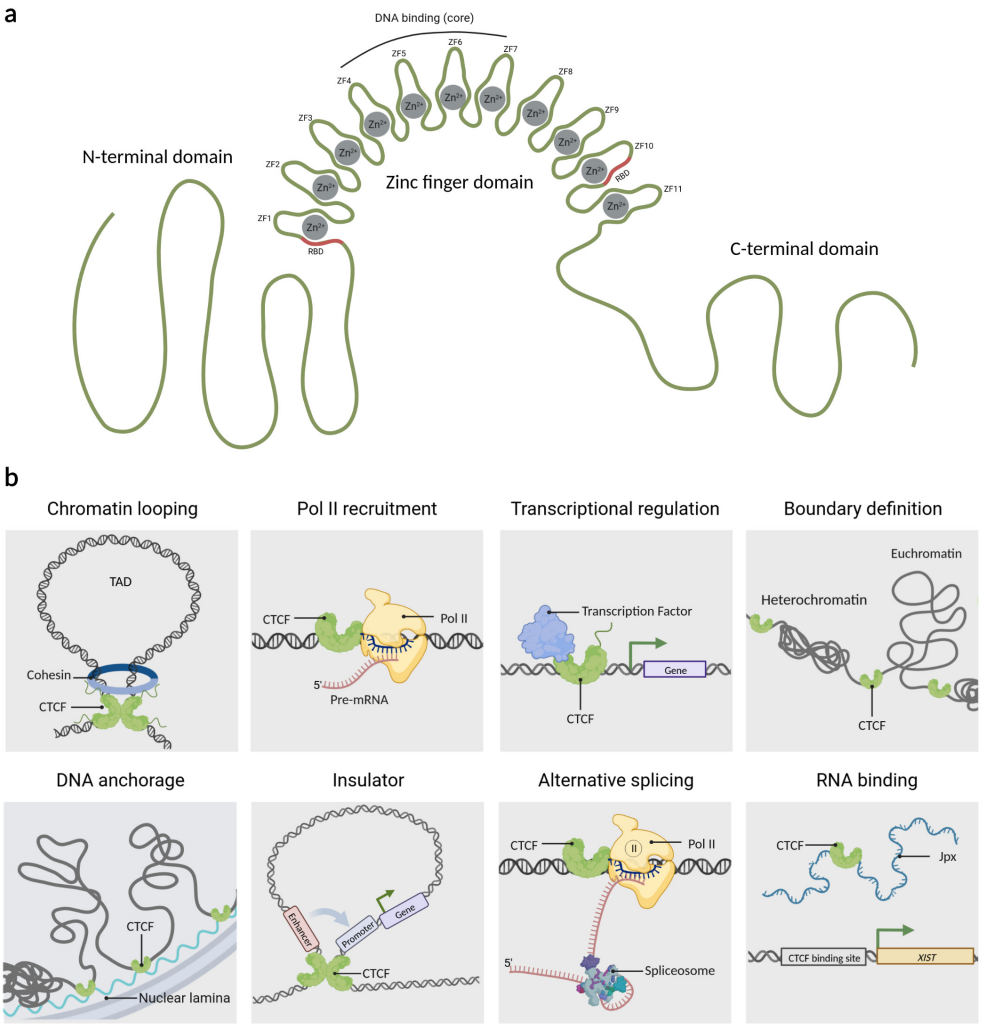


Figure 1. The architectonic factor CTCF. (**a**) CTCF is an 82-kDa protein that contains three domains: an N-terminal region, a C-terminal region, and a central domain of 11 zinc fingers. Moreover, CTCF uses the zinc finger domain cooperatively to bind to DNA. RBD:RNA binding domain, ZF: zinc finger. (**b**) Overview of the wide arrange of CTCF mechanisms of action as: Chromatin looping, RNA Polymerase II (Pol II) recruitment, transcriptional regulation, boundary definition, DNA anchorage, insulator, alternative splicing, and RNA binding, among others. TAD: topologically associated domain. Created with BioRender.com (accessed on 23 April 2023).

3. BORIS and CTCF-s Highlight the Importance of CTCF Protein–Protein Interactions

The mechanisms underlying CTCF functions are not yet fully understood, but it is probable that most of them depend on interactions with other proteins. One of the better-characterized CTCF protein–protein interactions is cohesin retention. The cohesin ring is a multi-protein complex involved in the formation of chromatin loops [46]. The mechanism of loop extrusion by cohesin involves the translocation of the complex along chromatin fibers, progressively extruding chromatin loops until it encounters a barrier that prevents further movement; such a barrier is frequently a CTCF dimer. In humans, CTCF interacts with SA1-SCC1 subunits of cohesins through its N-terminal domain, fixating the ring in place and establishing topologically associated domains [8,47,48]. Based on this mechanism of action, it has been proposed that upon binding to the DNA, the unbound ZFs and the terminal regions of CTCF might serve as a platform for interaction with other proteins. This hypothesis is supported by the discovery of a shorter isoform known as CTCF-short

(CTCF-s), which lacks the N-terminal domain and the first three zinc fingers (Figure 2a). Because they share the core DNA binding domain, CTCF-s competes for the canonical CTCF binding sites and interferes with CTCF–cohesin interactions, causing a disruption in the long-range connection between enhancers and promoters (Figure 2b). Overexpression of CTCF-s leads to increased cell apoptosis in HeLa-S3 cells, but the physiological role of this isoform and its impact on CTCF interactions with other proteins remain uncertain [49].

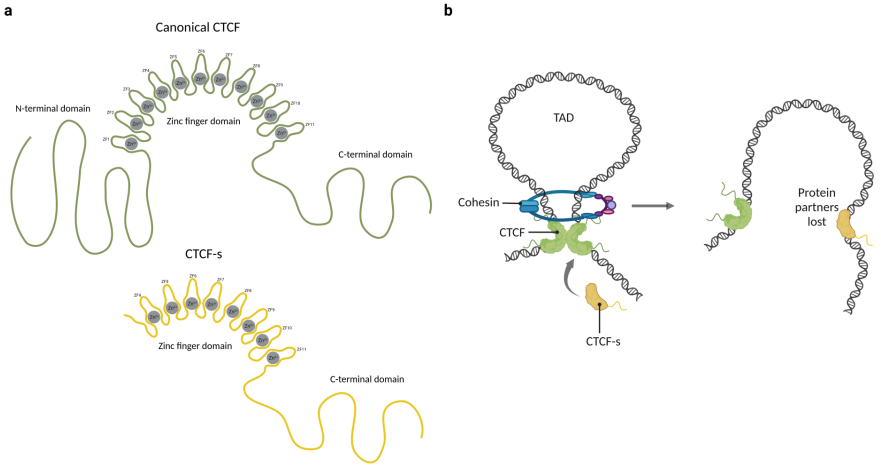


Figure 2. The participation of CTCF and CTCF-short (CTCF-s) in the formation of chromatin loops. (**a**) Representation of the domain distribution of CTCF and CTCF-short (CTCF-s). (**b**) CTCF physically binds to itself to form homodimers which promote chromatin loop formation through cohesin ring protein. CTCF-s competes with CTCF to alter the chromatin architecture and loop formation, mainly because CTCF-s is unable to interact with the cohesin ring. Created with BioRender.com (accessed on 23 April 2023).

Similarly, *CTCF* has a paralogous gene called *CTCF-Like* (*CTCFL*), which encodes the protein Brother of the Regulator of Imprinted Sites (BORIS). It is believed that *CTCFL* originated from a duplication event at some point before the evolution of mammals [14]. Unlike CTCF, BORIS is a protein that under physiological circumstances is only expressed in the testis, where it is required for spermatogenesis [50]. Nonetheless, BORIS has gained notoriety recently as a promising drug target because it is aberrantly expressed in several neoplasms and has been related to poor outcomes in cancer patients [51,52].

CTCF and BORIS share 75% of identity, mainly in their DNA binding domains, suggesting that they might compete for similar binding sites in the genome. Indeed, BORIS has been described to bind to a large subset of CTCF binding sites; however, there are a few differences in the target regions of both proteins [53]. While CTCF binds preferentially to intronic and intergenic regions, BORIS binds mainly to promoters [54,55]. Because the N- and C-terminal domains of BORIS are not conserved (Figure 3a), it has been suggested that BORIS may share binding sites with CTCF, but after binding will recruit different protein partners, interfering with the main functions of CTCF. In this regard, it has been reported that BORIS expression affects transcriptional regulation and the establishment of chromatin loops since BORIS alone is insufficient to recruit the cohesin complex, which is indispensable for CTCF-mediated chromatin loop formation [53,56].

Besides the impairment of chromatin loops, the differences between CTCF and BORIS terminal domains may affect which proteins are recruited upon binding (Figure 3b). Through a yeast two-hybrid assay, it was demonstrated that BORIS binds to a set of completely different protein partners than CTCF [57]. This explains the opposite consequences of their expression in cancer; while BORIS promotes cell proliferation and has been classified as an oncogene [58,59], CTCF is a known tumor suppressor [60]. Moreover, it has been observed that BORIS promotes the expression of some genes that are repressed by

CTCF, such as hTERT [61], NY-ESO [17], and H19 [62]. So far, CDH8 and UBF are the only proteins known to bind both CTCF and BORIS [63,64]. A summary of the currently known BORIS protein partners is displayed in Table 1. The former reinforces the importance of CTCF protein–protein interactions for the maintenance of the 3D-chromatin structure and suggests that the terminal domains of these proteins serve as scaffolds for cofactor recruitment. Together, this suggests that the cellular functions of CTCF and BORIS could be defined by their interaction with other proteins.

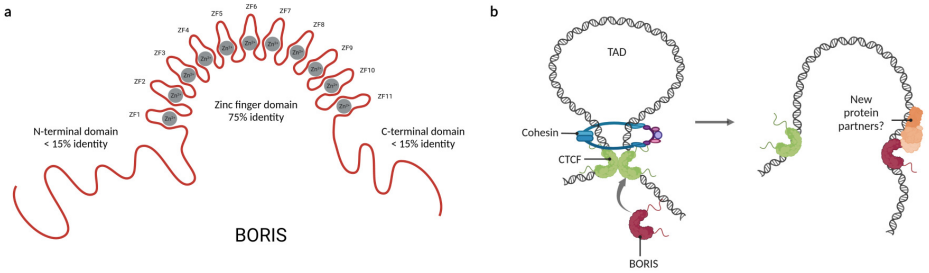


Figure 3. Features and functions of Brother of the Regulator of imprinted sites (BORIS). (**a**) Representation of the domain distribution of BORIS and their percentage of identity with CTCF. (**b**) BORIS can alter chromatin loops by a competitive mechanism with CTCF and its inability to interact with the cohesin ring. Moreover, the recruitment of new protein partners by BORIS could explain the opposite behaviors of CTCF and BORIS. Created with BioRender.com (accessed on 23 April 2023).

Table 1. Known protein–protein interactions of BORIS. Proteins that were experimentally validated to interact with CTCF as well are labeled in red.

Protein Types	Protein	Complex Function	Experimental Evidence	References
Chromatin-associated proteins	PRMT7	Arginine methylation to control imprinting.	Immunoprecipitation.	[65]
	CTCF	Unknown function in spermatogenesis.	In situ proximity ligation assay. Immunoprecipitation	[50,53]
	BAG6	Transcriptional activation of c-myc and BRCA1.	Yeast two-hybrid assay	[57]
	SET1A			
	POGZ	Unknown	Yeast two-hybrid assay	[57]
	SRCAP			
	TBP	Transcriptional activation of MAGE-A1.	Pull down assay.	[66]
	SP1	Transcriptional activation of NY-ESO-1.	Immunoprecipitation. Pull down assay.	[66,67]
Transcription factors	ELF2	Unknown	Yeast two-hybrid assay	[57]
	HCFC2, HCFC1			
	MGA			
	TLK2			
	NFAT5			
	ZNF518			
	ATF7			
	MKL2			
DNA Binding proteins	Ku70	DNA damage repair. rDNA transcriptional regulation.	Immunoprecipitation	[68]
	UBF		Immunoprecipitation.	[64]
	CHD8			
Signaling proteins	CSTA	Unknown	Yeast two-hybrid assay	[57]
	FHL2			

4. CTCF Regulates the Chromatin Structure through Interactions with Several Epigenetic Factors

The chromatin status is dynamic and can be regulated by covalent modification of the amino-terminal ends of histones that protrude from the nucleosome and are accessible to enzymes that chemically modify them through a system of writing, reading, and erasing complexes [9]. These modifications correspond to a kind of code that works in conjunction with the DNA sequence to determine the state of the chromatin and establishes and stabilizes gene expression patterns [10]. Because of CTCF's role as the master regulator of chromatin, it is highly probable that both its actions and DNA recruitment are dependent on the chromatin context. To better understand the interactions between CTCF and other proteins with epigenetic functions, we analyzed data from the literature, as well as the STRING database [69] and the Integrated Interactions Database [70] to find CTCF protein partners (Supplementary Table S1). While many of these partners are transcription factors that use CTCF as a scaffold to shape the chromatin structure [71], CTCF also interacts with other proteins that have epigenetic functions, such as DNA and histone demethylases [21,72,73]. The identification of CTCF protein partners involved in epigenetic processes may provide valuable insights into the complex regulatory mechanisms of chromatin organization and gene expression. To identify these proteins, we filtered our list of CTCF protein partners using the annotations available in the EpiFactors database [13]. The resulting CTCF epigenetic factor targets are shown in Figure 4.

Among these interactions, many of the proteins participate in the shaping of the 3D conformation of the genome such as the DNA helicases CHD7 [74], CHD8 [63] and CHD1L [75], the topoisomerases TOP2A [76] and TOP2B [77], and the components of chromatin remodeling complexes such as ARID1A [78], YY1 [79], YAF2 [42] and BPTF [71]. The former suggests that CTCF works in combination with other remodeling cofactors to establish chromatin domains.

It is also worth noticing that CTCF interacts with several members of the Polycomb group (PcG). These proteins are part of a system that regulates post-translational modifiers of histones, and their action is generally associated with the transcriptional repression of tissue-specific genes. This group has two members, the Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2). PRC2 is the complex that acts as a writer, as it is responsible for mono-, di-, and trimethylated lysine 27 of histone 3 (H3K27me3). This mark is associated with silenced gene promoters and facultative heterochromatin. H3K27me3 is recognized by PRC1 (reader) that binds to chromatin, monoubiquitinates lysine 119 of histone H2A (H2AK119ub), and prevents transcription by blocking the recruitment of RNA polymerase II [80,81]. CTCF interacts with EED and SUZ12 which are members of the PRC2 complex; a couple of studies have proposed that CTCF could guide the PRC2 complex to gene promoters that are susceptible to repression through H3K27 methylation [82,83]. Furthermore, BMI1, PCGF1, and RYBP are members of the PRC1 complex. Although the biological significance of their interaction with CTCF remains unexplored, a study shows that these proteins may regulate the organization of CTCF-mediated chromatin interactions [84].

Besides PcG proteins and chromatin remodeling factors, CTCF's relationship with proteins related to histone post-translational modifications are remarkable as well. CTCF interacts with proteins involved in the three stages of histone posttranslational modifications (writing, reading, and erasing). However, we would like to discuss further two particular cases that have not been broadly explored yet; histone and DNA demethylases.

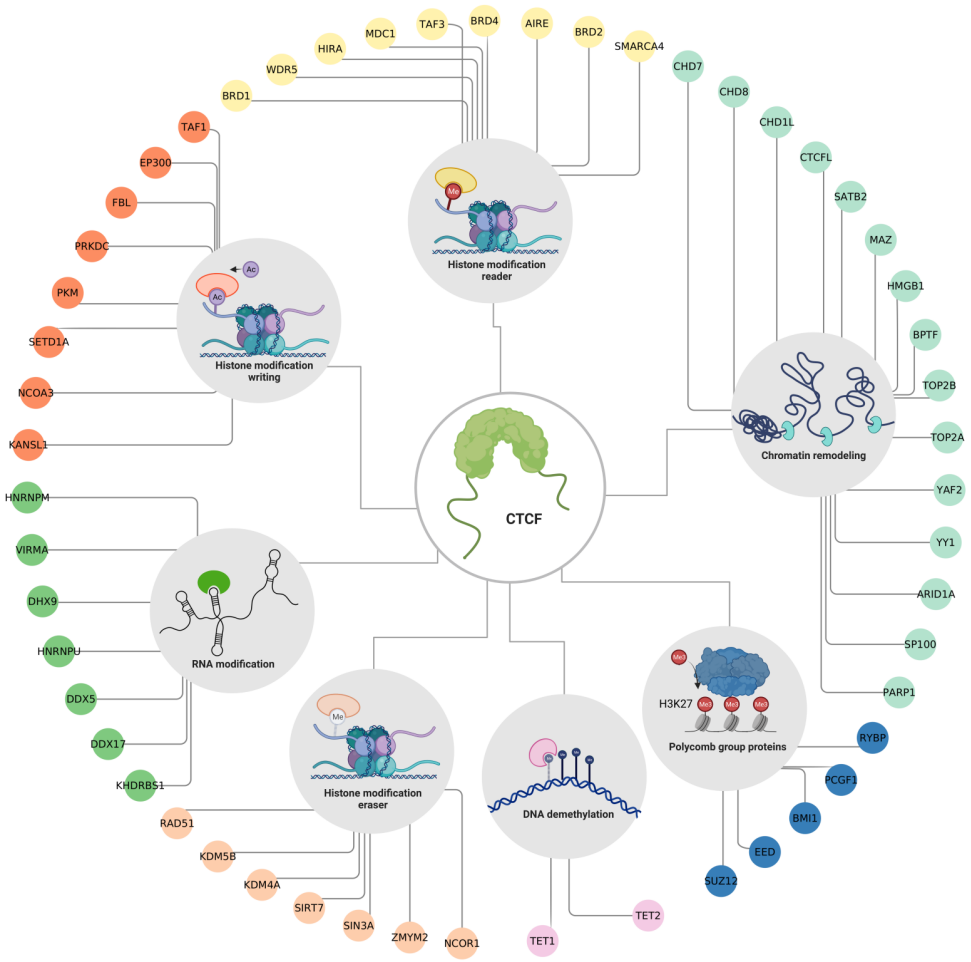


Figure 4. Epigenetic factors that interact with CTCF. The protein–protein interactions between CTCF and other proteins with epigenetic functions. Colors are according to the EpiFactor category that each protein belongs to, as follows: histone modification reader in yellow, chromatin remodeling in mint, polycomb group proteins in blue navy, DNA demethylation in pink, histone modification eraser in salmon, RNA modification in green, and histone modification writing in orange. Created with BioRender.com (accessed on 23 April 2023).

5. CTCF as a Modulator of Histone Methylation

Histone methylation is a post-translational modification related to multiple biological functions. Methylation happens mainly in arginine (R) and lysine (K) residues. Arginines can be mono- or dimethylated, and this chemical modification generally potentiates the interaction with other enzymes that modify histone tails [85]. Moreover, lysine residues can be mono-, di-, or trimethylated; these histone marks are associated with either transcriptional activation or repression, depending on the lysine residue. As an example, di- and trimethylation at H3K4 is related to enhanced gene expression, whereas trimethylation at H3K9 and H3K27 is associated with transcriptional repression [10]. Because histone methylation is a covalent modification, it was initially assumed to be stable and irreversible. However, in 2004, the first histone lysine demethylase was characterized, and since then more than 20 enzymes have been described that can remove this covalent modification [86,87].

Currently, histone lysine demethylases (KDMs) are classified into two families based on their chemical mechanism of action: the amine oxidase-like and the oxygenase enzymes [88]. The amino oxidase-like family has two members: KDM1A; the first his-

tone lysine demethylase described by Shi and colleagues in 2004; and KDM1B. These proteins have a common amine oxidase-like domain and are FAD-dependent [89]. KDM1 enzymes can remove mono- and dimethyl groups but cannot demethylate trimethylated lysines, due to their FAD-dependent catalytic mechanism [90]. The oxygenase family is the largest one, with more than 20 JmjC (Jumonji) domain-containing enzymes. These proteins enclose a Fe^{2+} ion in their catalytic domain and use α-ketoglutarate as a co-substrate [91]. This family is also divided into seven subfamilies (KDM2-8) according to the similarity of their catalytic domain and their substrate specificity [88].

In vitro studies have demonstrated that the simple binding of these enzymes to their substrates is sufficient for the demethylation reaction, suggesting that their recruitment must be tightly controlled in order to prevent aberrant demethylation [92–94]. It is not yet clear how the demethylases are directed to specific sites in the chromatin, especially since they lack DNA binding domains. One possible explanation could be that certain transcriptional factors and other chromatin-binding proteins might be responsible for the recruitment of these epigenetic components. KDMs activity could be regulated by protein–protein interactions allowing a dynamical interaction with the chromatin by taking advantage of the "reader" domains present in their binding partners [95,96]. Moreover, it has been suggested that the chromatin environment provides certain selectivity to demethylases since it controls the accessibility of these proteins to their target sites [97]. In addition, it is known that several transcription factors recruit histone demethylases upon binding to their target genes to promote a change in the chromatin state [98–100]. However, KDMs' relationship with CTCF remains partially unexplored.

Until now, few studies have demonstrated the association between CTCF and histone demethylases; in fact, only two KDM partners have been found. The first was reported in 2014 by Yamamoto et al., who found via co-immunoprecipitation that CTCF formed a complex with the H3K4me3 and H3K4me2 specific demethylase KDM5B. Moreover, when conducting ChIP-seq assays, they discovered that KDM5B sites overlap with those of CTCF in most mammary cancer cell lines, and this overlapping phenomenon correlates with a lower H3K4me3 signal compared to those non-overlapping sites (Figure 5a). The role of the KDM5B-CTCF complex is not clear, but the authors suggest that CTCF takes part in a finely tuned regulation of basal/stem cell genes, such as *ACTG2*, *APOE*, *CTGF*, *FN1*, and *TGFβ2*, among others. The perturbation of these transcriptional changes could promote breast cancer progression [73,101].

Another CTCF histone demethylase partner is KDM4A. The first clue that CTCF could be a KDM4A partner was reported in 2011 by Kang's group, who performed transfection and immunofluorescence assays and observed that the demethylation frequency of KDM4A was enhanced by the presence of CTCF [102]. This study opened the window to another report in 2018, where co-immunoprecipitation was used to demonstrate that CTCF and KDM4A form a protein complex. Furthermore, it was shown by ChIP-qPCR and ChIP-Re/ChIP-qPCR that CTCF and KDM4A coexist in the first intron of *CHD5*, the promoter of *WRAP53*, and the region located at −1922 bp of the *ASCL2* transcription starting site. The coexistence of CTCF and KDM4A correlates with the reduction of H3K36me3/2 histone modifications at the first intron of *CHD5* and is associated with its transcriptional down-regulation (Figure 5b). Moreover, CTCF or KDM4A depletion mediated by siRNAs leads to the *CHD5* reactivation expression, proposing that both proteins are involved in the negative regulation of this gene. The knockout of KDM4A by CRISPR/Cas9 restored the expression of *CHD5* and H3K36me3 and H3K36me2 histone marks, without disturbing the CTCF localization [72]. Nevertheless, it is currently unknown whether this complex is related to a genome-wide repression or activation and if CTCF might also be one of the key proteins driving the specificity of KDM4A.

To the best of our knowledge, there are no studies evaluating the association between CTCF and other histone demethylases. Nevertheless, ChIP-seq studies demonstrate some overlap between KDM5A, KDM5C, KDM1A, and CTCF, suggesting that CTCF could

be involved in their regulation; however, further studies are required to determine the participation of CTCF in the modulation of these enzymes.

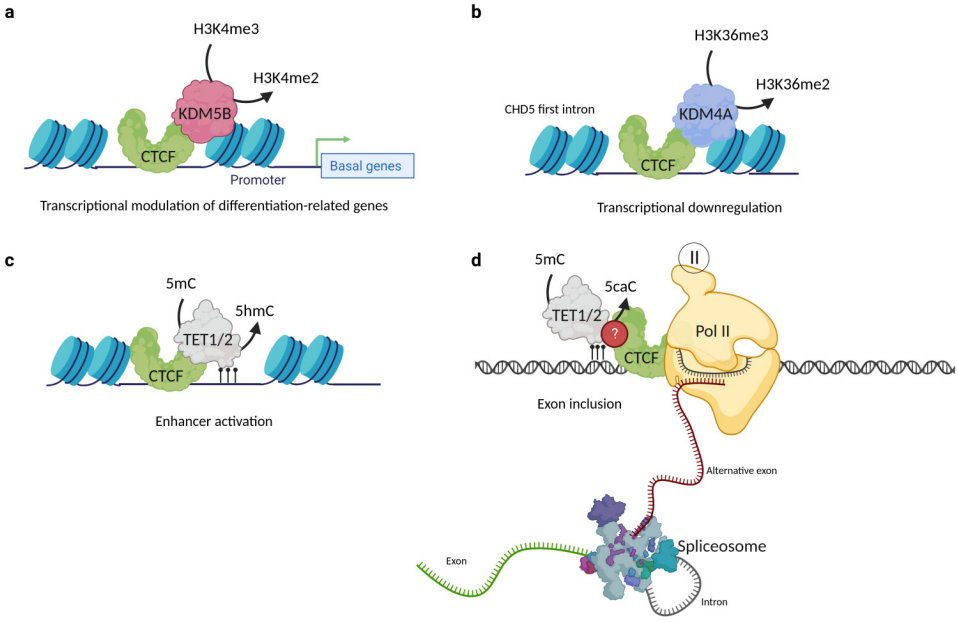


Figure 5. CTCF interactions with histone and DNA demethylases. (**a**) CTCF interacts with KDM5B and regulates the transcription rate of basal/stem cell genes in luminal breast cancer lines. (**b**) The interaction of CTCF with KDM4A is involved in the down-regulation of CHD5 gene expression in MCF7 cells. (**c**) CTCF interaction with TET1 and TET2 proteins is involved in enhancer activation. (**d**) CTCF can also interact with 5caC, which leads to RNA pol II pausing and alternative exon inclusion of the CD45+ gene. CTCF and TET protein–protein interaction is possible but remains uncharacterized for this mechanism. Created with BioRender.com (accessed on 23 April 2023).

6. CTCF and the TET Enzymes

DNA methylation is an epigenetic process involving a methyl group transfer to the C5 position of the cytosine to form 5-methylcytosine (5mC). DNA methylation has several functions; although it is generally associated with transcriptional repression; it is also involved in other vital processes, such as genomic imprinting, X chromosome inactivation, and retrotransposon element suppression [103,104]. Similarly to histones, DNA can be demethylated; this process can be accomplished either passively, by simply not methylating the new DNA strand after replication, or actively, by a replication-independent process that involves the ten-eleven translocation (TET) enzymes [105].

The first evidence of the enzyme-mediated DNA demethylation was observed in 2007, with the identification of the Trypanosoma cruzi enzymes JBP1 and JBP2 that are responsible for gene silencing through the hydroxylation and glycosylation of a thymine methyl group (known as J Base). This discovery pointed toward the existence of "eraser" proteins that are in charge of removing DNA methylation [106]. Shortly after, in 2009, when looking for mammalian homologs of the trypanosome thymidine hydroxylases, the three human ten eleven translocation (TET) proteins, TET1, TET2, and TET3 were identified [107]. Nevertheless, the TET proteins were not at a central stage until they were found to oxidize 5mC to 5-hydroxymethyl-cytosine (5hmC) as part of the DNA demethylation mechanism [108,109]. Subsequent reports revealed that TET proteins further oxidize 5hmC to 5-formyl-cytosine (5fC) and 5-carboxyl-cytosine (5caC), both of which are removed through the Base Excision Repair (BER) pathway, thereby completing the demethylation process [108,110].

Because DNA methylation is an epigenetic marker that is essential for correct cellular function and organism development [111,112], TET proteins must be subjected to finely controlled regulatory mechanisms. These enzymes have fundamental roles in epigenetic reprogramming, embryogenesis, development, and tumorigenesis, and it is well-known that their inactivation contributes to the local DNA hypermethylation observed in cancer [113,114]. Apart from catalytic activity regulation, TET1 and TET3 are more likely recruited to their genomic target sites through the direct binding of their respective CXXC domains to the DNA [115]. In vitro binding assays and in vivo chromatin immunoprecipitation assays confirm that these domains can bind CpG-rich oligonucleotides with a slight preference for unmethylated versus methylated substrates [116–118]. In contrast, TET2 does not have any obvious DNA-binding domains, and it is therefore potentially recruited through the direct binding of DNA-targeting partners [119]. In fact, it has been demonstrated that the TET2 protein binds tissue-specific transcription factors such as the early B cell factor 1 (EBF1) [120] and WT1 [121,122]. The dynamic expression of DNA-binding factors and their interactions with TET2 can likely concede the tissue-specific and temporal modulation of TET activity on a limited set of genomic loci [123]. Furthermore, interaction with several binding partners is likely to alter the genomic location and stability of TET proteins [124].

Since TET enzymes form protein complexes with other epigenetic components to modify gene transcription, the interaction of these proteins with CTCF is of particular interest. It is known that synchronized fluctuations of DNA methylation, demethylation, nucleosome positioning, and CTCF chromatin binding have an important role in establishing cell-type-specific chromatin states during differentiation. Loss of CTCF in regions such as the boundaries of chromatin loops, promoters, and TADs can be associated with the spread of DNA methylation and demethylation, and can be linked to the down-regulation of adjacent genes. A hierarchical interaction between cytosine modifications, nucleosome positioning, and DNA sequences controls CTCF binding and regulates gene expression [125,126].

It has been proposed that CTCF binding to low methylated regions could mediate local DNA demethylation through TET recruitment [127]. The first evidence was an oscillating 5hmC pattern observed around the binding sites of CTCF in mouse embryonic stem cells, which suggests that accessibility and 5hmC deposition could be related to CTCF binding [128]. The genomic co-localization of CTCF, TET1, TET2, and 5hmc was probed by co-immunoprecipitation assays on 3T3-L1 and HEK293T cell lines and correlated with enhancer activation on differentiated cells through the facilitation of the hydroxymethylation of DNA [129]. This concludes that CTCF directly interacts with the TET enzymes and promotes the DNA hydroxymethylation of enhancers driving adipocyte differentiation (Figure 5c). Nevertheless, the relationship between CTCF and TET demethylases is not only relevant to cell differentiation processes, since a study in 2016 revealed that dynamic TET1 and TET2-catalyzed DNA oxidation stimulates CTCF-dependent alternative splicing in human lymphocytes. This study found that CTCF directly interacts with 5caC in vitro and that this mark was strongly associated with alternative exon inclusion [6]. Moreover, a study demonstrated that 5caC could reinforce CTCF binding to the DNA (Figure 5d). These findings suggest that the TET mediated-induction of 5caC is a potential way to regulate CTCF binding and further reinforces the idea that there is a close relationship between CTCF and the TET proteins [130]. More studies are needed to better describe the exact functions that DNA oxidation plays in transcriptional regulatory events; additional explorations will be required to define the way in which CTCF binding is associated with 5caC in vivo.

Taken together, the above information suggests that CTCF could interact, directly or indirectly, with histone and DNA demethylases; it is still unknown whether these complexes are related to repression, activation, or other transcriptional processes.

7. Long Non-Coding RNAs as Non-Protein Partners of CTCF

Long non-coding RNAs (lncRNAs) have emerged as important regulators of chromatin structure and gene expression. They act as scaffolds, guides, or decoys that recruit chromatin modifiers to specific genomic regions, mediate higher-order chromatin organization, and influence gene expression [131]. LncRNAs have been demonstrated to play a critical role in the formation and maintenance of chromatin domains, such as TADs. In this context, lncRNAs have been found to interact with chromatin-associated proteins, CTCF for instance, to modulate their function and impact on chromatin structure and gene regulation [19,132]. Recently, lncRNAs have been identified as key regulators of CTCF [20]. CTCF interacts with RNA through the RNA-binding domains in ZF1 and ZF10. Some studies have even reported a consensus sequence for RNAs that bind to CTCF, and it has been suggested that it could have around 5000 potential RNA partners in the genome [21,133].

LncRNAs contribute to the functions of CTCF by recruiting it to specific genomic sites, modulating chromatin loops, and regulating the formation of TADs. One of the most studied cases is CTCF-mediated Xist transcriptional repression. Xist is a lncRNA involved in X chromosome inactivation. CTCF represses Xist expression by binding to its promoter; however, Jpx is a lncRNA that binds to CTCF and removes it from the *Xist* promoter, allowing its expression and subsequent X chromosome inactivation [19,134]. Recently, it was found that Jpx can also compete for CTCF binding sites in the DNA, altering the loop formation and the overall conformation of the chromatin [22]. The interplay between CTCF and other RNA-binding proteins is also important for the maintenance of TADs. As shown in Figure 4, CTCF interacts with several RNA-binding proteins. Among them, DDX5 is an RNA helicase involved in many steps of RNA-related processes, such as alternative splicing, miRNA biogenesis, and RNA unwinding [135]. It has been described that both DDX5 and the lncRNA steroid receptor RNA activator (SRA) interact with the CTCF-cohesin complex and stabilize it. Such an interaction is required for the insulation activity of CTCF [136].

Several other lncRNAs have been identified to interact with CTCF and modulate its function. HOTTIP, for instance, can recruit CTCF to specific genomic regions and promote TAD formation [137]. Similarly, GATA6-AS1 contributes to TAD formation by forming an RNA-DNA triplex and interacting with CTCF [138]. LncRNAs also regulate gene expression through the recruitment or detachment of CTCF [139–141]. PACERR recruits CTCF and p300 to promoter regions to activate gene transcription through histone acetylation [142]. LncRNAs have also been associated to increase protein stability; for instance, the lncRNA ELDR inhibits CTCF degradation by the proteasome, increasing protein levels without modifying transcript levels. Table 2 shows known interactions between CTCF and lncRNAs, along with the putative function of the complexes.

Overall, lncRNAs represent an exciting new area of research in the field of chromatin biology and gene regulation. The interaction between lncRNAs and CTCF offers a new level of complexity to the already intricate network of molecular interactions that govern gene expression and chromatin architecture.

Table 2. Long noncoding RNAs (lncRNAs) that are known to directly interact with CTCF.

lncRNA	Function	References
HOTTIP	CTCF recruitment and TAD formation	[137,140]
PACERR	Recruits CTCF and p300 to promoter regions.	[142]
JPX	Jpx binds to CTCF consensus regions causing a shift in chromatin loops. It is also involved in X chromosome inactivation.	[22,139]
DLGAP1-AS2	Reduced binding of CTCF to target genes.	[140]
GATA6-AS1	May contribute to TAD formation. Forms an RNA-DNA triplex.	[138]
ELDR	Inhibits CTCF degradation by the proteasome.	[143]
SH3PXD2A-AS1	Recruits CTCF to inhibit the expression of target genes.	[141]
CDKN2B-AS1	Recruits CTCF and EZH2 to silence target genes.	[144]
LINC00346	Prevents CTCF binding to the *c-Myc* promoter	[145]
H19	Mediates the interaction between CTCF and Vigilin to regulate IGF2 imprinting.	[146]
Firre	Anchorage of the X chromosome to the nucleolus.	[147]
CCAT1-L	Modulates chromatin loops.	[148]
SRA	Estabilizes CTCF-cohesin complex.	[136]

8. Conclusions and Final Remarks

CTCF is a nuclear factor that is involved in several chromatin-related processes, including transcriptional regulation, three-dimensional chromatin topology, and epigenetics. Part of its relevance lies in its versatility, as shown in this review, CTCF relies on a broad network of protein and RNA partners to achieve its different tasks. The role of the protein partners is clear upon comparison with CTCF–s and BORIS. In the first case, the lack of the N-terminal domain leads to the loss of the most studied CTCF interacting partners, the cohesin complex. The second case is more complex, since BORIS binding to the DNA has completely opposite consequences than CTCF, besides sharing a high degree of identity at their DNA binding domains. Most of this could be explained by their interactions with different protein partners through their unconserved terminal domains. There is current research going on in this regard, and without a doubt, the study of the interplay between CTCF and BORIS in cancer will help to understand CTCF's role in chromatin organization and other epigenetic processes.

Chromatin is finely organized inside the nucleus through a complex system that has not yet been elucidated. As mentioned previously in this review, many of the epigenetic factors that help to establish and maintain chromatin structure lack DNA binding domains; thus, it has been hypothesized that their action should rely on other proteins. CTCF is capable of binding to thousands of sites in the genome, and due to the flexibility of its DNA binding domain, it is considered a multivalent protein. In this review, we demonstrate that CTCF interacts with a wide array of epigenetic factors which suggests that it could serve as a scaffold for the assembly of different protein complexes. Nevertheless, the logistics involved in partner election, the impact of each complex, and the crosstalk between different partners is an exciting point of view that is worth further study.

Since epigenetic markers such as histone and DNA methylation are highly dependent on the chromatin context, the interaction between CTCF and different components of the epigenetic complexes is interesting. So far only a few protein–protein interactions between CTCF and other epigenetic factors have been fully characterized, and in most cases, the studies have been conducted on a specific gene or promoter; thus, genomic scale experiments could be helpful to identify the overall impact and localization of the complexes.

LncRNAs add another layer of complexity to the CTCF-mediated chromatin regulation. Currently, the role of most of these complexes in biological processes remains unknown. However, several studies hint towards the existence of a broad CTCF-RNA interaction network. The role of some of these complexes has been discussed here; among them, Jpx is the most remarkable example due to its ability to detach CTCF from its binding sites. Further studies will help to understand the role of CTCF-lncRNA interactions.

Because of CTCF's versatility, it could likely function as a scaffold for many of the epigenetic complexes required for a proper genomic organization. Without a doubt, there are still undiscovered mechanisms for CTCF; the study of this protein could aid to understand the complex mechanisms that regulate chromatin organization and gene expression.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cells12101357/s1, Table S1: Protein-Protein interactions for CTCF.

Author Contributions: Conceptualization, A.D.M.-M. and E.S.-R.; investigation and data curation M.S.-A., Y.S.-P. and A.D.M.-M.; original draft preparation, A.D.M.-M. and M.S.-A.; review and editing, E.S.-R. and N.K.W.; supervision, E.S.-R. and J.B.; funding acquisition, E.S.-R. and J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT) through the Fondo CB-SEP-CONACyT (284748, PROMED 250690) and by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grants 01ZX1910D and 01ZX2210D). Soto-Reyes, E. was supported by the Natural Science Department at UAM Cuajimalpa (DCNI-07-243-23). Del Moral-Morales, A. is a doctoral student from Programa de Maestría y Doctorado en Ciencias Bioquímicas, UNAM, and received a fellowship from CONACyT (CVU894530). Del Moral-Morales, A. was also a beneficiary of the German Academic

Exchange Service (DAAD Grant No. 91833882). Baumbach, J. was partially funded by his VILLUM Young Investigator Grant nr.13154.

Acknowledgments: We greatly acknowledge the support and collaboration with the Center for Data and Computing in Natural Sciences (CDCS).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Morrison, O.; Thakur, J. Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. *Int. J. Mol. Sci.* **2021**, *22*, 6922. [CrossRef] [PubMed]
2. Holwerda, S.J.B.; de Laat, W. CTCF: The protein, the binding partners, the binding sites and their chromatin loops. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2013**, *368*, 20120369. [CrossRef] [PubMed]
3. Bell, A.C.; West, A.G.; Felsenfeld, G. The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell* **1999**, *98*, 387–396. [CrossRef] [PubMed]
4. Shukla, S.; Kavak, E.; Gregory, M.; Imashimizu, M.; Shutinoski, B.; Kashlev, M.; Oberdoerffer, P.; Sandberg, R.; Oberdoerffer, S. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **2011**, *479*, 74–79. [CrossRef] [PubMed]
5. Ruiz-Velasco, M.; Kumar, M.; Lai, M.C.; Bhat, P.; Solis-Pinson, A.B.; Reyes, A.; Kleinsorg, S.; Noh, K.M.; Gibson, T.J.; Zaugg, J.B. CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. *Cell Syst.* **2017**, *5*, 628–637.e6. [CrossRef] [PubMed]
6. Marina, R.J.; Sturgill, D.; Bailly, M.A.; Thenoz, M.; Varma, G.; Prigge, M.F.; Nanan, K.K.; Shukla, S.; Haque, N.; Oberdoerffer, S. TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *EMBO J.* **2016**, *35*, 335–355. [CrossRef]
7. Vostrov, A.A.; Taheny, M.J.; Quitschke, W.W. 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.* **2002**, *277*, 1619–1627. [CrossRef]
8. Pugacheva, E.M.; Kubo, N.; Loukinov, D.; Tajmul, M.; Kang, S.; Kovalchuk, A.L.; Strunnikov, A.V.; Zentner, G.E.; Ren, B.; Lobanenkov, V.V. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 2020–2031. [CrossRef]
9. Felsenfeld, G.; Groudine, M. Controlling the double helix. *Nature* **2003**, *421*, 448–453. [CrossRef]
10. Millán-Zambrano, G.; Burton, A.; Bannister, A.J.; Schneider, R. Histone post-translational modifications—Cause and consequence of genome function. *Nat. Rev. Genet.* **2022**, *23*, 563–580. [CrossRef]
11. 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. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **2000**, *406*, 593–599. [CrossRef] [PubMed]
12. Medvedeva, Y.A.; Lennartsson, A.; Ehsani, R.; Kulakovskiy, I.V.; Vorontsov, I.E.; Panahandeh, P.; Khimulya, G.; Kasukawa, T.; FANTOM Consortium; Drabløs, F. EpiFactors: A comprehensive database of human epigenetic factors and complexes. *Database* **2015**, *2015*, bav067. [CrossRef] [PubMed]
13. Marakulina, D.; Vorontsov, I.E.; Kulakovskiy, I.V.; Lennartsson, A.; Drabløs, F.; Medvedeva, Y.A. EpiFactors 2022: Expansion and enhancement of a curated database of human epigenetic factors and complexes. *Nucleic Acids Res.* **2023**, *51*, D564–D570. [CrossRef] [PubMed]
14. Hore, T.A.; Deakin, J.E.; Marshall Graves, J.A. The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. *PLoS Genet.* **2008**, *4*, e1000169. [CrossRef] [PubMed]
15. Vatolin, S.; Abdullaev, Z.; Pack, S.D.; Flanagan, P.T.; Custer, M.; Loukinov, D.I.; Pugacheva, E.; Hong, J.A.; Morse, H., 3rd; Schrump, D.S.; et al. Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes. *Cancer Res.* **2005**, *65*, 7751–7762. [CrossRef]
16. Nguyen, P.; Cui, H.; Bisht, K.S.; Sun, L.; Patel, K.; Lee, R.S.; Kugoh, H.; Oshimura, M.; Feinberg, A.P.; Gius, D. CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region. *Cancer Res.* **2008**, *68*, 5546–5551. [CrossRef]
17. Hong, J.A.; Kang, Y.; Abdullaev, Z.; Flanagan, P.T.; Pack, S.D.; Fischette, M.R.; Adnani, M.T.; Loukinov, D.I.; Vatolin, S.; Risinger, J.I.; et al. Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. *Cancer Res.* **2005**, *65*, 7763–7774. [CrossRef]
18. Guh, C.Y.; Hsieh, Y.H.; Chu, H.P. Functions and properties of nuclear lncRNAs-from systematically mapping the interactomes of lncRNAs. *J. Biomed. Sci.* **2020**, *27*, 44. [CrossRef]
19. Saldaña-Meyer, R.; Rodriguez-Hernaez, J.; Escobar, T.; Nishana, M.; Jácome-López, K.; Nora, E.P.; Bruneau, B.G.; Tsirigos, A.; Furlan-Magaril, M.; Skok, J.; et al. RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol. Cell* **2019**, *76*, 412–422.e5. [CrossRef]
20. Hansen, A.S.; Hsieh, T.H.S.; Cattoglio, C.; Pustova, I.; Saldaña-Meyer, R.; Reinberg, D.; Darzacq, X.; Tjian, R. Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol. Cell* **2019**, *76*, 395–411.e13. [CrossRef]

21. Gavrilov, A.A.; Sultanov, R.I.; Magnitov, M.D.; Galitsyna, A.A.; Dashinimaev, E.B.; Lieberman Aiden, E.; Razin, S.V. RedChIP identifies noncoding RNAs associated with genomic sites occupied by Polycomb and CTCF proteins. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116222119. [CrossRef] [PubMed]
22. Oh, H.J.; Aguilar, R.; Kesner, B.; Lee, H.G.; Kriz, A.J.; Chu, H.P.; Lee, J.T. Jpx RNA regulates CTCF anchor site selection and formation of chromosome loops. *Cell* **2021**, *184*, 6157–6173.e24. [CrossRef] [PubMed]
23. Lobanenkov, V.V.; Nicolas, R.H.; Adler, V.V.; Paterson, H.; Klenova, E.M.; Polotskaja, A.V.; Goodwin, G.H. 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* **1990**, *5*, 1743–1753. [PubMed]
24. Heger, P.; Marin, B.; Bartkuhn, M.; Schierenberg, E.; Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17507–17512. [CrossRef] [PubMed]
25. Bonchuk, A.; Kamalyan, S.; Mariasina, S.; Boyko, K.; Popov, V.; Maksimenko, O.; Georgiev, P. N-terminal domain of the architectural protein CTCF has similar structural organization and ability to self-association in bilaterian organisms. *Sci. Rep.* **2020**, *10*, 2677. [CrossRef]
26. Martinez, S.R.; Miranda, J.J.L. CTCF terminal segments are unstructured. *Protein Sci.* **2010**, *19*, 1110. [CrossRef]
27. Maksimenko, O.G.; Fursenko, D.V.; Belova, E.V.; Georgiev, P.G. CTCF As an Example of DNA-Binding Transcription Factors Containing Clusters of C2H2-Type Zinc Fingers. *Acta Nat.* **2021**, *13*, 31. [CrossRef]
28. Ohlsson, R.; Renkawitz, R.; Lobanenkov, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **2001**, *17*, 520–527. [CrossRef]
29. Nakahashi, H.; Kieffer Kwon, K.R.; Resch, W.; Vian, L.; Dose, M.; Stavreva, D.; Hakim, O.; Pruett, N.; Nelson, S.; Yamane, A.; et al. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* **2013**, *3*, 1678–1689. [CrossRef]
30. Yin, M.; Wang, J.; Wang, M.; Li, X.; Zhang, M.; Wu, Q.; Wang, Y. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.* **2017**, *27*, 1365–1377. [CrossRef]
31. Hashimoto, H.; Wang, D.; Horton, J.R.; Zhang, X.; Corces, V.G.; Cheng, X. Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol. Cell* **2017**, *66*, 711–720.e3. [CrossRef] [PubMed]
32. Xu, D.; Ma, R.; Zhang, J.; Liu, Z.; Wu, B.; Peng, J.; Zhai, Y.; Gong, Q.; Shi, Y.; Wu, J.; et al. Dynamic Nature of CTCF Tandem 11 Zinc Fingers in Multivalent Recognition of DNA As Revealed by NMR Spectroscopy. *J. Phys. Chem. Lett.* **2018**, *9*, 4020–4028. [CrossRef] [PubMed]
33. Schmidt, D.; Schwalie, P.C.; Wilson, M.D.; Ballester, B.; Gonçalves, A.; Kutter, C.; Brown, G.D.; Marshall, A.; Flicek, P.; Odom, D.T. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **2012**, *148*, 335–348. [CrossRef] [PubMed]
34. Kim, T.H.; Abdullaev, Z.K.; Smith, A.D.; Ching, K.A.; Loukinov, D.I.; Green, R.D.; Zhang, M.Q.; Lobanenkov, V.V.; Ren, B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **2007**, *128*, 1231–1245. [CrossRef] [PubMed]
35. Lee, B.K.; Iyer, V.R. Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. *J. Biol. Chem.* **2012**, *287*, 30906–30913. [CrossRef] [PubMed]
36. Shen, Y.; Yue, F.; McCleary, D.F.; Ye, Z.; Edsall, L.; Kuan, S.; Wagner, U.; Dixon, J.; Lee, L.; Lobanenkov, V.V.; et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* **2012**, *488*, 116–120. [CrossRef]
37. Fiorito, E.; Sharma, Y.; Gilfillan, S.; Wang, S.; Singh, S.K.; Satheesh, S.V.; Katika, M.R.; Urbanucci, A.; Thiede, B.; Mills, I.G.; et al. CTCF modulates Estrogen Receptor function through specific chromatin and nuclear matrix interactions. *Nucleic Acids Res.* **2016**, *44*, 10588–10602. [CrossRef]
38. Kaczmarczyk, L.S.; Levi, N.; Segal, T.; Salmon-Divon, M.; Gerlitz, G. CTCF supports preferentially short lamina-associated domains. *Chromosome Res.* **2022**, *30*, 123–136. [CrossRef]
39. Kentepozidou, E.; Aitken, S.J.; Feig, C.; Stefflova, K.; Ibarra-Soria, X.; Odom, D.T.; Roller, M.; Flicek, P. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* **2020**, *21*, 5. [CrossRef]
40. Ong, C.T.; Corces, V.G. CTCF: An architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **2014**, *15*, 234–246. [CrossRef]
41. Ren, G.; Jin, W.; Cui, K.; Rodrigez, J.; Hu, G.; Zhang, Z.; Larson, D.R.; Zhao, K. CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Mol. Cell* **2017**, *67*, 1049–1058.e6. [CrossRef] [PubMed]
42. Schwalie, P.C.; Ward, M.C.; Cain, C.E.; Faure, A.J.; Gilad, Y.; Odom, D.T.; Flicek, P. Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome Biol.* **2013**, *14*, R148. [CrossRef] [PubMed]
43. Chernukhin, I.V.; Shamsuddin, S.; Robinson, A.F.; Carne, A.F.; Paul, A.; El-Kady, A.I.; Lobanenkov, V.V.; Klenova, E.M. Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J. Biol. Chem.* **2000**, *275*, 29915–29921. [CrossRef] [PubMed]
44. Defossez, P.A.; Kelly, K.F.; Filion, G.J.; Pérez-Torrado, R.; Magdinier, F.; Menoni, H.; Nordgaard, C.L.; Daniel, J.M.; Gilson, E. The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J. Biol. Chem.* **2005**, *280*, 43017–43023. [CrossRef]

45. Wei, L.; Liu, Q.; Huang, Y.; Liu, Z.; Zhao, R.; Li, B.; Zhang, J.; Sun, C.; Gao, B.; Ding, X.; et al. Knockdown of CTCF reduces the binding of EZH2 and affects the methylation of the SOCS3 promoter in hepatocellular carcinoma. *Int. J. Biochem. Cell Biol.* **2020**, *120*, 105685. [CrossRef] [PubMed]
46. Gabriele, M.; Brandão, H.B.; Grosse-Holz, S.; Jha, A.; Dailey, G.M.; Cattoglio, C.; Hsieh, T.H.S.; Mirny, L.; Zechner, C.; Hansen, A.S. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **2022**, *376*, 496–501. [CrossRef]
47. Li, Y.; Haarhuis, J.H.I.; Sedeño Cacciatore, Á.; Oldenkamp, R.; van Ruiten, M.S.; Willems, L.; Teunissen, H.; Muir, K.W.; de Wit, E.; Rowland, B.D.; et al. The structural basis for cohesin–CTCF-anchored loops. *Nature* **2020**, *578*, 472–476. [CrossRef]
48. Wutz, G.; Várnai, C.; Nagasaka, K.; Cisneros, D.A.; Stocsits, R.R.; Tang, W.; Schoenfelder, S.; Jessberger, G.; Muhar, M.; Hossain, M.J.; et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **2017**, *36*, 3573–3599. [CrossRef]
49. Li, J.; Huang, K.; Hu, G.; Babarinde, I.A.; Li, Y.; Dong, X.; Chen, Y.S.; Shang, L.; Guo, W.; Wang, J.; et al. An alternative CTCF isoform antagonizes canonical CTCF occupancy and changes chromatin architecture to promote apoptosis. *Nat. Commun.* **2019**, *10*, 1535. [CrossRef]
50. Rivero-Hinojosa, S.; Pugacheva, E.M.; Kang, S.; Méndez-Catalá, C.F.; Kovalchuk, A.L.; Strunnikov, A.V.; Loukinov, D.; Lee, J.T.; Lobanenkov, V.V. The combined action of CTCF and its testis-specific paralog BORIS is essential for spermatogenesis. *Nat. Commun.* **2021**, *12*, 3846. [CrossRef]
51. Loukinov, D.I.; Pugacheva, E.; Vatolin, S.; Pack, S.D.; Moon, H.; Chernukhin, I.; Mannan, P.; Larsson, E.; Kanduri, C.; Vostrov, A.A.; 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* **2002**, *99*, 6806–6811. [CrossRef] [PubMed]
52. Janssen, S.M.; Moscona, R.; Elchebly, M.; Papadakis, A.I.; Redpath, M.; Wang, H.; Rubin, E.; van Kempen, L.C.; Spatz, A. BORIS/CTCFL promotes a switch from a proliferative towards an invasive phenotype in melanoma cells. *Cell Death Discov.* **2020**, *6*, 1. [CrossRef] [PubMed]
53. Pugacheva, E.M.; Rivero-Hinojosa, S.; Espinoza, C.A.; Méndez-Catalá, C.F.; Kang, S.; Suzuki, T.; Kosaka-Suzuki, N.; Robinson, S.; Nagarajan, V.; Ye, Z.; et al. Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. *Genome Biol.* **2015**, *16*, 161. [CrossRef]
54. Sleutels, F.; Soochit, W.; Bartkuhn, M.; Heath, H.; Dienstbach, S.; Bergmaier, P.; Franke, V.; Rosa-Garrido, M.; van de Nobelen, S.; Caesar, L.; et al. The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics Chromatin* **2012**, *5*, 8. [CrossRef]
55. Bergmaier, P.; Weth, O.; Dienstbach, S.; Boettger, T.; Galjart, N.; Mernberger, M.; Bartkuhn, M.; Renkawitz, R. Choice of binding sites for CTCFL compared to CTCF is driven by chromatin and by sequence preference. *Nucleic Acids Res.* **2018**, *46*, 7097–7107. [CrossRef]
56. Nishana, M.; Ha, C.; Rodriguez-Hernaez, J.; Ranjbaran, A.; Chio, E.; Nora, E.P.; Badri, S.B.; Kloetgen, A.; Bruneau, B.G.; Tsirigos, A.; et al. Defining the relative and combined contribution of CTCF and CTCFL to genomic regulation. *Genome Biol.* **2020**, *21*, 108. [CrossRef] [PubMed]
57. Nguyen, P.; Bar-Sela, G.; Sun, L.; Bisht, K.S.; Cui, H.; Kohn, E.; Feinberg, A.P.; Gius, D. BAT3 and SET1A form a complex with CTCFL/BORIS to modulate H3K4 histone dimethylation and gene expression. *Mol. Cell. Biol.* **2008**, *28*, 6720–6729. [CrossRef]
58. Gaykalova, D.; Vatapalli, R.; Glazer, C.A.; Bhan, S.; Shao, C.; Sidransky, D.; Ha, P.K.; Califano, J.A. Dose-dependent activation of putative oncogene SBSN by BORIS. *PLoS ONE* **2012**, *7*, e40389. [CrossRef]
59. Rosa-Garrido, M.; Ceballos, L.; Alonso-Lecue, P.; Abraira, C.; Delgado, M.D.; Gandarillas, A. A cell cycle role for the epigenetic factor CTCF-L/BORIS. *PLoS ONE* **2012**, *7*, e39371. [CrossRef]
60. Kemp, C.J.; Moore, J.M.; Moser, R.; Bernard, B.; Teater, M.; Smith, L.E.; Rabaia, N.A.; Gurley, K.E.; Guinney, J.; Busch, S.E.; et al. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep.* **2014**, *7*, 1020–1029. [CrossRef]
61. Renaud, S.; Loukinov, D.; Alberti, L.; Vostrov, A.; Kwon, Y.W.; Bosman, F.T.; Lobanenkov, V.; Benhattar, J. BORIS/CTCFL-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. *Nucleic Acids Res.* **2011**, *39*, 862–873. [CrossRef] [PubMed]
62. Ulaner, G.A.; Vu, T.H.; Li, T.; Hu, J.F.; Yao, X.M.; Yang, Y.; Gorlick, R.; Meyers, P.; Healey, J.; Ladanyi, M.; et al. Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. *Hum. Mol. Genet.* **2003**, *12*, 535–549. [CrossRef] [PubMed]
63. Ishihara, K.; Oshimura, M.; Nakao, M. CTCF-Dependent Chromatin Insulator Is Linked to Epigenetic Remodeling. *Mol. Cell* **2006**, *23*, 733–742. [CrossRef] [PubMed]
64. van de Nobelen, S.; Rosa-Garrido, M.; Leers, J.; Heath, H.; Soochit, W.; Joosen, L.; Jonkers, I.; Demmers, J.; van der Reijden, M.; Torrano, V.; et al. CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenetics Chromatin* **2010**, *3*, 19. [CrossRef]
65. Jelinic, P.; Stehle, J.C.; Shaw, P. The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. *PLoS Biol.* **2006**, *4*, e355. [CrossRef]
66. Schwarzenbach, H.; Eichelser, C.; Steinbach, B.; Tadewaldt, J.; Pantel, K.; Lobanenkov, V.; Loukinov, D. Differential regulation of MAGE-A1 promoter activity by BORIS and Sp1, both interacting with the TATA binding protein. *BMC Cancer* **2014**, *14*, 796. [CrossRef]

67. Kang, Y.; Hong, J.A.; Chen, G.A.; Nguyen, D.M.; Schrump, D.S. Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells. *Oncogene* **2007**, *26*, 4394–4403. [CrossRef]
68. Zhang, Y.; Fang, M.; Li, S.; Xu, H.; Ren, J.; Tu, L.; Zuo, B.; Yao, W.; Liang, G. BTApep-TAT peptide inhibits ADP-ribosylation of BORIS to induce DNA damage in cancer. *Mol. Cancer* **2022**, *21*, 158. [CrossRef]
69. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [CrossRef]
70. Kotlyar, M.; Pastrello, C.; Ahmed, Z.; Chee, J.; Varyova, Z.; Jurisica, I. IID 2021: Towards context-specific protein interaction analyses by increased coverage, enhanced annotation and enrichment analysis. *Nucleic Acids Res.* **2022**, *50*, D640–D647. [CrossRef]
71. Sun, X.; Zhang, J.; Cao, C. CTCF and Its Partners: Shaper of 3D Genome during Development. *Genes* **2022**, *13*, 1383. [CrossRef] [PubMed]
72. Guerra-Calderas, L.; González-Barrios, R.; Patiño, C.C.; Alcaraz, N.; Salgado-Albarrán, M.; de León, D.C.; Hernández, C.C.; Sánchez-Pérez, Y.; Maldonado-Martínez, H.A.; De la Rosa-Velazquez, I.A.; et al. CTCF-KDM4A complex correlates with histone modifications that negatively regulate gene expression in cancer cell lines. *Oncotarget* **2018**, *9*, 17028–17042. [CrossRef]
73. Yamamoto, S.; Wu, Z.; Russnes, H.G.; Takagi, S.; Peluffo, G.; Vaske, C.; Zhao, X.; Moen Vollan, H.K.; Maruyama, R.; Ekram, M.B.; et al. JARID1B Is a Luminal Lineage-Driving Oncogene in Breast Cancer. *Cancer Cell* **2014**, *25*, 762–777. [CrossRef]
74. Allen, M.D.; Religa, T.L.; Freund, S.M.V.; Bycroft, M. Solution structure of the BRK domains from CHD7. *J. Mol. Biol.* **2007**, *371*, 1135–1140. [CrossRef] [PubMed]
75. Liu, Q.; Yang, B.; Xie, X.; Wei, L.; Liu, W.; Yang, W.; Ge, Y.; Zhu, Q.; Zhang, J.; Jiang, L.; et al. Vigilin interacts with CCCTC-binding factor (CTCF) and is involved in CTCF-dependent regulation of the imprinted genes Igf2 and H19. *FEBS J.* **2014**, *281*, 2713–2725. [CrossRef]
76. Yusufzai, T.M.; Tagami, H.; Nakatani, Y.; Felsenfeld, G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **2004**, *13*, 291–298. [CrossRef]
77. Witcher, M.; Emerson, B.M. Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol. Cell* **2009**, *34*, 271–284. [CrossRef]
78. Marino, M.M.; Rega, C.; Russo, R.; Valletta, M.; Gentile, M.T.; Esposito, S.; Baglivo, I.; De Feis, I.; Angelini, C.; Xiao, T.; et al. Interactome mapping defines BRG1, a component of the SWI/SNF chromatin remodeling complex, as a new partner of the transcriptional regulator CTCF. *J. Biol. Chem.* **2019**, *294*, 861–873. [CrossRef] [PubMed]
79. Donohoe, M.E.; Zhang, L.F.; Xu, N.; Shi, Y.; Lee, J.T. Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol. Cell* **2007**, *25*, 43–56. [CrossRef]
80. Wiles, E.T.; Selker, E.U. H3K27 methylation: A promiscuous repressive chromatin mark. *Curr. Opin. Genet. Dev.* **2017**, *43*, 31–37. [CrossRef]
81. Margueron, R.; Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **2011**, *469*, 343–349. [CrossRef] [PubMed]
82. Zhang, H.; Niu, B.; Hu, J.F.; Ge, S.; Wang, H.; Li, T.; Ling, J.; Steelman, B.N.; Qian, G.; Hoffman, A.R. Interruption of intrachromosomal looping by CCCTC binding factor decoy proteins abrogates genomic imprinting of human insulin-like growth factor II. *J. Cell Biol.* **2011**, *193*, 475–487. [CrossRef]
83. Li, T.; Hu, J.F.; Qiu, X.; Ling, J.; Chen, H.; Wang, S.; Hou, A.; Vu, T.H.; Hoffman, A.R. CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell. Biol.* **2008**, *28*, 6473–6482. [CrossRef] [PubMed]
84. Wei, C.; Jia, L.; Huang, X.; Tan, J.; Wang, M.; Niu, J.; Hou, Y.; Sun, J.; Zeng, P.; Wang, J.; et al. CTCF organizes inter-A compartment interactions through RYBP-dependent phase separation. *Cell Res.* **2022**, *32*, 744–760. [CrossRef]
85. Blanc, R.S.; Richard, S. Arginine Methylation: The Coming of Age. *Mol. Cell* **2017**, *65*, 8–24. [CrossRef] [PubMed]
86. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J.R.; Cole, P.A.; Casero, R.A.; Shi, Y. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **2004**, *119*, 941–953. [CrossRef]
87. Guerra-Calderas, L.; González-Barrios, R.; Herrera, L.A.; Cantú de León, D.; Soto-Reyes, E. The role of the histone demethylase KDM4A in cancer. *Cancer Genet.* **2015**, *208*, 215–224. [CrossRef]
88. Sterling, J.; Menezes, S.V.; Abbassi, R.H.; Munoz, L. Histone lysine demethylases and their functions in cancer. *Int. J. Cancer* **2020**, *148*, 2375–2388. [CrossRef]
89. Anand, R.; Marmorstein, R. Structure and Mechanism of Lysine-specific Demethylase Enzymes. *J. Biol. Chem.* **2007**, *282*, 35425–35429. [CrossRef]
90. Yang, M.; Gocke, C.B.; Luo, X.; Borek, D.; Tomchick, D.R.; Machius, M.; Otwinowski, Z.; Yu, H. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell* **2006**, *23*, 377–387. [CrossRef]
91. Klose, R.J.; Kallin, E.M.; Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **2006**, *7*, 715–727. [CrossRef] [PubMed]
92. Couture, J.F.; Collazo, E.; Ortiz-Tello, P.A.; Brunzelle, J.S.; Trievel, R.C. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **2007**, *14*, 689–695. [CrossRef]
93. Ng, S.S.; Kavanagh, K.L.; McDonough, M.A.; Butler, D.; Pilka, E.S.; Lienard, B.M.R.; Bray, J.E.; Savitsky, P.; Gileadi, O.; von Delft, F.; et al. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **2007**, *448*, 87–91. [CrossRef] [PubMed]

94. Lan, F.; Nottke, A.C.; Shi, Y. Mechanisms involved in the regulation of histone lysine demethylases. *Curr. Opin. Cell Biol.* **2008**, *20*, 316. [CrossRef]
95. Shiau, C.; Trnka, M.J.; Bozicevic, A.; Ortiz Torres, I.; Al-Sady, B.; Burlingame, A.L.; Narlikar, G.J.; Fujimori, D.G. Reconstitution of nucleosome demethylation and catalytic properties of a Jumonji histone demethylase. *Chem. Biol.* **2013**, *20*, 494–499. [CrossRef]
96. Ruthenburg, A.J.; Allis, C.D.; Wysocka, J. Methylation of lysine 4 on histone H3: Intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **2007**, *25*, 15–30. [CrossRef]
97. Lan, F.; Zaratiegui, M.; Villén, J.; Vaughn, M.W.; Verdel, A.; Huarte, M.; Shi, Y.; Gygi, S.P.; Moazed, D.; Martienssen, R.A.; et al. *S. pombe* LSD1 homologs regulate heterochromatin propagation and euchromatic gene transcription. *Mol. Cell* **2007**, *26*, 89–101. [CrossRef]
98. Metzger, E.; Wissmann, M.; Yin, N.; Müller, J.M.; Schneider, R.; Peters, A.H.F.M.; Günther, T.; Buettner, R.; Schüle, R. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **2005**, *437*, 436–439. [CrossRef] [PubMed]
99. Garcia-Bassets, I.; Kwon, Y.S.; Telese, F.; Prefontaine, G.G.; Hutt, K.R.; Cheng, C.S.; Ju, B.G.; Ohgi, K.A.; Wang, J.; Escoubet-Lozach, L.; et al. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* **2007**, *128*, 505–518. [CrossRef] [PubMed]
100. Wang, J.; Scully, K.; Zhu, X.; Cai, L.; Zhang, J.; Prefontaine, G.G.; Krones, A.; Ohgi, K.A.; Zhu, P.; Garcia-Bassets, I.; et al. Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* **2007**, *446*, 882–887. [CrossRef]
101. Kristensen, L.H.; Nielsen, A.L.; Helgstrand, C.; Lees, M.; Cloos, P.; Kastrup, J.S.; Helin, K.; Olsen, L.; Gajhede, M. Studies of H3K4me3 demethylation by KDM5B/Jarid1B/PLU1 reveals strong substrate recognition in vitro and identifies 2,4-pyridine-dicarboxylic acid as an in vitro and in cell inhibitor. *FEBS J.* **2012**, *279*, 1905–1914. [CrossRef]
102. Jeong, Y.S.; Park, J.S.; Ko, Y.; Kang, Y.K. JHDM3A module as an effector molecule in guide-directed modification of target chromatin. *J. Biol. Chem.* **2011**, *286*, 4461–4470. [CrossRef]
103. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **2002**, *16*, 6–21. [CrossRef]
104. Moore, L.D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2012**, *38*, 23–38. [CrossRef]
105. Wu, X.; Zhang, Y. TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nat. Rev. Genet.* **2017**, *18*, 517–534. [CrossRef] [PubMed]
106. Yu, Z.; Genest, P.A.; ter Riet, B.; Sweeney, K.; DiPaolo, C.; Kieft, R.; Christodoulou, E.; Perrakis, A.; Simmons, J.M.; Hausinger, R.P.; et al. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res.* **2007**, *35*, 2107–2115. [CrossRef] [PubMed]
107. Tahiliani, M.; Koh, K.P.; Shen, Y.; Pastor, W.A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L.M.; Liu, D.R.; Aravind, L.; et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009**, *324*, 930–935. [CrossRef]
108. Ito, S.; Shen, L.; Dai, Q.; Wu, S.C.; Collins, L.B.; Swenberg, J.A.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300–1303. [CrossRef]
109. Ito, S.; D'Alessio, A.C.; Taranova, O.V.; Hong, K.; Sowers, L.C.; Zhang, Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **2010**, *466*, 1129–1133. [CrossRef]
110. He, Y.F.; Li, B.Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **2011**, *333*, 1303–1307. [CrossRef] [PubMed]
111. Li, E.; Bestor, T.H.; Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **1992**, *69*, 915–926. [CrossRef] [PubMed]
112. Okano, M.; Bell, D.W.; Haber, D.A.; Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **1999**, *99*, 247–257. [CrossRef] [PubMed]
113. Li, D.; Guo, B.; Wu, H.; Tan, L.; Lu, Q. TET Family of Dioxygenases: Crucial Roles and Underlying Mechanisms. *Cytogenet. Genome Res.* **2015**, *146*, 171–180. [CrossRef]
114. Williams, K.; Christensen, J.; Helin, K. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep.* **2011**, *13*, 28–35. [CrossRef]
115. Zhang, H.; Zhang, X.; Clark, E.; Mulcahey, M.; Huang, S.; Shi, Y.G. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **2010**, *20*, 1390–1393. [CrossRef]
116. Wu, H.; D'Alessio, A.C.; Ito, S.; Xia, K.; Wang, Z.; Cui, K.; Zhao, K.; Sun, Y.E.; Zhang, Y. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **2011**, *473*, 389–393. [CrossRef]
117. Xu, Y.; Wu, F.; Tan, L.; Kong, L.; Xiong, L.; Deng, J.; Barbera, A.J.; Zheng, L.; Zhang, H.; Huang, S.; et al. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Mol. Cell* **2011**, *42*, 451–464. [CrossRef]
118. Xu, Y.; Xu, C.; Kato, A.; Tempel, W.; Abreu, J.G.; Bian, C.; Hu, Y.; Hu, D.; Zhao, B.; Cerovina, T.; et al. Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for Xenopus Eye and Neural Development. *Cell* **2012**, *151*, 1200–1213. [CrossRef] [PubMed]
119. Ko, M.; An, J.; Bandukwala, H.S.; Chavez, L.; Aijö, T.; Pastor, W.A.; Segal, M.F.; Li, H.; Koh, K.P.; Lähdesmäki, H.; et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **2013**, *497*, 122–126. [CrossRef]

120. Guilhamon, P.; Eskandarpour, M.; Halai, D.; Wilson, G.A.; Feber, A.; Teschendorff, A.E.; Gomez, V.; Hergovich, A.; Tirabosco, R.; Fernanda Amary, M.; et al. Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nat. Commun.* **2013**, *4*, 2166. [CrossRef]
121. Rampal, R.; Alkalin, A.; Madzo, J.; Vasanthakumar, A.; Pronier, E.; Patel, J.; Li, Y.; Ahn, J.; Abdel-Wahab, O.; Shih, A.; et al. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell Rep.* **2014**, *9*, 1841–1855. [CrossRef]
122. Wang, Y.; Xiao, M.; Chen, X.; Chen, L.; Xu, Y.; Lv, L.; Wang, P.; Yang, H.; Ma, S.; Lin, H.; et al. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Mol. Cell* **2015**, *57*, 662–673. [CrossRef]
123. Rasmussen, K.D.; Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* **2016**, *30*, 733–750. [CrossRef]
124. Wang, Y.; Zhang, Y. Regulation of TET protein stability by calpains. *Cell Rep.* **2014**, *6*, 278–284. [CrossRef] [PubMed]
125. Teif, V.B.; Beshnova, D.A.; Vainshtein, Y.; Marth, C.; Mallm, J.P.; Höfer, T.; Rippe, K. Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res.* **2014**, *24*, 1285–1295. [CrossRef]
126. Stadler, M.B.; Murr, R.; Burger, L.; Ivanek, R.; Lienert, F.; Schöler, A.; van Nimwegen, E.; Wirbelauer, C.; Oakeley, E.J.; Gaidatzis, D.; et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **2011**, *480*, 490–495. [CrossRef]
127. Feldmann, A.; Ivanek, R.; Murr, R.; Gaidatzis, D.; Burger, L.; Schübeler, D. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* **2013**, *9*, e1003994. [CrossRef] [PubMed]
128. Sun, Z.; Terragni, J.; Borgaro, J.G.; Liu, Y.; Yu, L.; Guan, S.; Wang, H.; Sun, D.; Cheng, X.; Zhu, Z.; et al. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep.* **2013**, *3*, 567–576. [CrossRef] [PubMed]
129. Dubois-Chevalier, J.; Oger, F.; Dehondt, H.; Firmin, F.F.; Gheeraert, C.; Staels, B.; Lefebvre, P.; Eeckhoute, J. A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res.* **2014**, *42*, 10943–10959. [CrossRef]
130. Nanan, K.K.; Sturgill, D.M.; Prigge, M.F.; Thenoz, M.; Dillman, A.A.; Mandler, M.D.; Oberdoerffer, S. TET-Catalyzed 5-Carboxylcytosine Promotes CTCF Binding to Suboptimal Sequences Genome-wide. *iScience* **2019**, *19*, 326–339. [CrossRef] [PubMed]
131. Statello, L.; Guo, C.J.; Chen, L.L.; Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *22*, 96–118. [CrossRef]
132. Long, Y.; Wang, X.; Youmans, D.T.; Cech, T.R. How do lncRNAs regulate transcription? *Sci. Adv.* **2017**, *3*, eaao2110. [CrossRef]
133. Kuang, S.; Wang, L. Identification and analysis of consensus RNA motifs binding to the genome regulator CTCF. *NAR Genom. Bioinform.* **2020**, *2*, lqaa031. [CrossRef] [PubMed]
134. Sun, S.; Del Rosario, B.C.; Szanto, A.; Ogawa, Y.; Jeon, Y.; Lee, J.T. Jpx RNA activates Xist by evicting CTCF. *Cell* **2013**, *153*, 1537–1551. [CrossRef] [PubMed]
135. Xu, K.; Sun, S.; Yan, M.; Cui, J.; Yang, Y.; Li, W.; Huang, X.; Dou, L.; Chen, B.; Tang, W.; et al. DDX5 and DDX17—Multifaceted proteins in the regulation of tumorigenesis and tumor progression. *Front. Oncol.* **2022**, *12*, 943032. [CrossRef] [PubMed]
136. Yao, H.; Brick, K.; Evrard, Y.; Xiao, T.; Camerini-Otero, R.D.; Felsenfeld, G. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* **2010**, *24*, 2543–2555. [CrossRef]
137. Wu, Q.; Liu, J.; Deng, J.; Chen, Y. Long non-coding RNA HOTTIP induces inflammation in asthma by promoting EFNA3 transcription by CCCTC-binding factor. *Am. J. Transl. Res.* **2022**, *14*, 8903–8917.
138. Soibam, B. Association between Triplex-Forming Sites of Cardiac Long Noncoding RNA and Chromatin Organization. *Noncoding RNA* **2022**, *8*, 41. [CrossRef]
139. Karner, H.; Webb, C.H.; Carmona, S.; Liu, Y.; Lin, B.; Erhard, M.; Chan, D.; Baldi, P.; Spitale, R.C.; Sun, S. Functional Conservation of LncRNA JPX Despite Sequence and Structural Divergence. *J. Mol. Biol.* **2020**, *432*, 283–300. [CrossRef]
140. Ren, C.; Han, H.; Pan, J.; Chang, Q.; Wang, W.; Guo, X.; Bian, J. DLGAP1-AS2 promotes human colorectal cancer progression through trans-activation of Myc. *Mamm. Genome* **2022**, *33*, 672–683. [CrossRef]
141. Chen, Q.; Jiang, S.; Liu, H.; Gao, Y.; Yang, X.; Ren, Z.; Gao, Y.; Xiao, L.; Hu, H.; Yu, Y.; et al. Association of lncRNA SH3PXD2A-AS1 with preeclampsia and its function in invasion and migration of placental trophoblast cells. *Cell Death Dis.* **2020**, *11*, 583. [CrossRef]
142. Liu, Y.; Wang, X.; Zhu, Y.; Cao, Y.; Wang, L.; Li, F.; Zhang, Y.; Li, Y.; Zhang, Z.; Luo, J.; et al. The CTCF/LncRNA-PACERR complex recruits E1A binding protein p300 to induce pro-tumour macrophages in pancreatic ductal adenocarcinoma via directly regulating PTGS2 expression. *Clin. Transl. Med.* **2022**, *12*, e654. [CrossRef] [PubMed]
143. Sur, S.; Steele, R.; Ko, B.C.B.; Zhang, J.; Ray, R.B. Long noncoding RNA ELDR promotes cell cycle progression in normal oral keratinocytes through induction of a CTCF-FOXM1-AURKA signaling axis. *J. Biol. Chem.* **2022**, *298*, 101895. [CrossRef]
144. Ou, M.; Li, X.; Zhao, S.; Cui, S.; Tu, J. Long non-coding RNA CDKN2B-AS1 contributes to atherosclerotic plaque formation by forming RNA-DNA triplex in the CDKN2B promoter. *EBioMedicine* **2020**, *55*, 102694. [CrossRef] [PubMed]
145. Peng, W.X.; He, R.Z.; Zhang, Z.; Yang, L.; Mo, Y.Y. LINC00346 promotes pancreatic cancer progression through the CTCF-mediated Myc transcription. *Oncogene* **2019**, *38*, 6770–6780. [CrossRef] [PubMed]
146. Yu, X.; Liu, Q.; He, J.; Huang, Y.; Jiang, L.; Xie, X.; Liu, J.; Chen, L.; Wei, L.; Qin, Y. Vigilin interacts with CTCF and is involved in the maintenance of imprinting of IGF2 through a novel RNA-mediated mechanism. *Int. J. Biol. Macromol.* **2018**, *108*, 515–522. [CrossRef]

147. Yang, F.; Deng, X.; Ma, W.; Berletch, J.B.; Rabaia, N.; Wei, G.; Moore, J.M.; Filippova, G.N.; Xu, J.; Liu, Y.; et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **2015**, *16*, 52. [CrossRef]
148. Xiang, J.F.; Yin, Q.F.; Chen, T.; Zhang, Y.; Zhang, X.O.; Wu, Z.; Zhang, S.; Wang, H.B.; Ge, J.; Lu, X.; et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **2014**, *24*, 513–531. [CrossRef]

Wiley
Scientifica
Volume 2024, Article ID 5460694, 23 pages
https://doi.org/10.1155/2024/5460694

WILEY

Research Article

Key Proteins for Regeneration in *A. mexicanum*

Transcriptomic Insights From Aged and Juvenile Limbs

Aylin Del Moral-Morales,[1] **Cynthia Sámano**,[1] **José Antonio Ocampo-Cervantes**,[2] **Maya Topf**,[3,4,5] **Jan Baumbach**,[6,7] **Jossephlyn Hernández**,[1] **Karla Torres-Arciga**,[8] **Rodrigo González-Barrios**,[8] **and Ernesto Soto-Reyes**[1]

[1]*Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana-Cuajimalpa (UAM-C), Mexico City, Mexico*
[2]*Centro de Investigaciones Biológicas y Acuícolas de Cuemanco (CIBAC), Universidad Autónoma Metropolitana-Xochimilco (UAM-X), Mexico City, Mexico*
[3]*Centre for Structural Systems Biology (CSSB), Hamburg, Germany*
[4]*Leibniz-Institut für Virologie (LIV), Hamburg, Germany*
[5]*University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany*
[6]*Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany*
[7]*Computational BioMedicine Lab., University of Southern Denmark, Odense, Denmark*
[8]*Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología-Instituto de Investigaciones Biomédicas, UNAM, Mexico City, Mexico*

Correspondence should be addressed to Rodrigo González-Barrios; satroz@gmail.com and Ernesto Soto-Reyes; esotoreyes@cua.uam.mx

Received 7 May 2024; Accepted 10 October 2024

Academic Editor: Chandrabose Selvaraj

The axolotl, known for its remarkable regenerative abilities, is an excellent model for studying regenerative therapies. Nevertheless, the precise molecular mechanisms governing its regenerative potential remain uncertain. In this study, we collected samples from axolotls of different ages, including 8-year-old individuals and 8-month-old juveniles, obtaining their blastemas 10 days after amputation. Subsequently, we conducted a transcriptomic analysis comparing our samples to a set of previously published experiments. Our analysis unveiled a distinctive transcriptional response in the blastema, characterized by differential gene expression associated with processes such as bone and tissue remodeling, transcriptional regulation, angiogenesis, and intercellular communication. To gain deeper insights, we compared these findings with those from aged axolotls that showed no signs of regeneration 10 days after amputation. We identified four genes—*FSTL1*, *ADAMTS17*, *GPX7*, and *CTHRC1*—that showed higher expression in regenerating tissue compared to aged axolotls. Further scrutiny, including structural and homology analysis, revealed that these genes are conserved across vertebrate species. Our discoveries point to a group of proteins relevant to tissue regeneration, with their conservation in vertebrates suggesting critical roles in development. These findings also propose a novel gene set involved in axolotl regeneration, laying a promising foundation for future investigations across vertebrates.

Keywords: aging; *Ambystoma mexicanum*; blastema; tissue regeneration; transcriptomics

1. Introduction

The axolotl is a salamander of the genus *Ambystoma*, which comprises 32 species from southern Canada to central Mexico. Among the different species present in Mexico, *Ambystoma mexicanum*, an endemic species from the lake region of Mexico City, is the most studied [1, 2]. Commonly known as axolotls, *A. mexicanum* is a vertebrate amphibian with an unusual appearance. Unlike other salamanders, axolotls do not metamorphose to become adults; instead, they remain in their larval form, retaining their external gills and aquatic lifestyle throughout their entire lives [3, 4]. In addition to their popularity as pets in many parts of the world, this animal is considered an interesting model for studying tissue regeneration. Indeed, they can regenerate a wide range of injured tissues within just a few weeks, including limbs [5, 6], tail, central and peripheral nervous systems [7, 8], iris [9], bone, and muscle [10, 11].

Previous studies have shown that the axolotl regeneration process involves three phases: wound healing, blastema formation, and redevelopment [12, 13]. After an injury, the fibroblasts surrounding the wound acquire proliferative capacity and set the stage for new tissue development [14–17]. Once the wound is closed, the fibroblasts dedifferentiate into mesenchymal progenitor cells that accumulate in an epithelium-covered region called the blastema. It is in this organ where the reconstruction of the missing limb takes place. Blastema cells are thought to have a certain "cellular memory," as they always differentiate into the same cell type from which they originated, and the new limb will have the same size, shape, and orientation as the amputated one [18–20]. Because of this, the study of the underlying mechanisms and factors involved in blastema regeneration is currently of great interest. Specifically, researchers are interested in dissecting the cellular and molecular events, leading to regeneration in axolotls that could be extrapolated to other species. A previous study demonstrated that mouse growth factors can induce blastema formation in axolotls, suggesting that the machinery required for tissue regeneration may be conserved in mammals [21]. However, the molecular factors involved in the activation of such machinery remain unexplored.

One limitation of studying axolotls is the complexity of their genome; with 32 gigabase pairs of DNA distributed across 14 pairs of chromosomes, it is one of the largest genomes among vertebrates. However, approximately 70% of their genome is composed of repetitive elements, making its assembly a challenge [22]. Furthermore, the lack of a fully annotated transcriptome has hampered comprehensive studies of tissue regeneration, making it difficult to compare the results across publications. Despite these challenges, previous research has identified several candidate genes that may play a role in the regenerative process. However, the mechanistic insights into these genes and how they interact with others are not fully understood [23–25]. Fortunately, recent advances in genome assembly and transcriptome annotation are enabling detailed transcriptomic studies that will shed light on the molecular processes underlying axolotl regeneration [26].

Additionally, the impact of environmental factors, such as habitat, diet, and age, on the regenerative capacity of axolotls has not been investigated [5]. While axolotls can regenerate tissues throughout most of their entire lifespan, this phenomenon becomes less efficient with age [9, 27]. Younger individuals can regenerate tissues in a matter of days or weeks, whereas sexually mature adults may require several months to regenerate a limb [18, 28]. It is also worth noting that most experiments have been carried out on the *d*/*d* strain, a homozygous white mutant lineage established by French naturalists in the mid-19th century. Interestingly, this strain often exhibits regeneration defects due to extensive inbreeding [1, 29]. Therefore, to fully understand the genetic, transcriptomic, and environmental factors involved in tissue regeneration in *A. mexicanum*, and how aging may affect this process [30], it is crucial to study native Mexican axolotl populations from the Xochimilco Lake area.

In this study, we conducted transcriptomic analysis on the limbs of juvenile axolotls (8 months old) and the blastema formed 10 days postamputation, aiming to elucidate the molecular fingerprint of tissue regeneration. Our investigation involved a comparative analysis of our datasets with those previously documented by Bryant in 2017, with the objective of identifying conserved transcriptomic responses in blastemas, even across different axolotl strains [23]. Through a custom annotation of the axolotl proteome, we found that the differentially expressed genes (DEGs) were primarily associated with anatomical development and cell differentiation. Additionally, we collected samples from the limbs of two aged axolotls (8 years old) that did not develop a blastema after amputation. By contrasting the DEGs in blastemas with those in aged limbs, we pinpointed four genes (*FSTL1*, *ADAMTS17*, *GPX7*, and *CTHRC1*) exhibiting heightened expression in regenerating tissues but diminished expression in aged axolotls. These genes were found to be associated with anatomical structure development, as corroborated by structural and conservation analyses, indicating their high conservation across vertebrates and their pivotal roles in development, bone morphogenesis, and cartilage formation in mammals. Our results suggest a set of key axolotl genes that potentially participate in tissue regeneration specifically in juvenile axolotls. These genes could provide a basis for studying tissue regeneration in other vertebrates and hold promise for advancing the field of regenerative medicine.

2. Materials and Methods

2.1. Sample Collection. All biological samples were sourced from a captive population of Mexican axolotls (*A. mexicanum*) housed at the Unidad de Manejo Ambiental of the Centro de Investigaciones Biológicas y Acuícolas de Cuemanco (UMA-CIBAC), affiliated with the Universidad Autónoma Metropolitana, Xochimilco Campus (UAM-Xochimilco) in Mexico City. This axolotl colony originates from wild individuals captured in the Xochimilco Lake area. Situated adjacent to the lake, CIBAC provides conditions closely resembling the axolotl's natural habitat; the water in their tanks is sourced from the lake, and their diet primarily

consists of aquatic worms (*Tubifex*). All procedures were approved and supervised by veterinarians from UMA-CIBAC and UAM-Xochimilco, in accordance with the norms and regulations set by the Mexican Ministry of Environment and Natural Resources (Reference Number CEI.2023.007).

Sedation of the axolotls was achieved through immersion in a tank containing benzocaine at a concentration of 50 mg/L before the amputation process. Biological samples were meticulously collected, using a stereoscopic microscope, from the inferior limbs of five juvenile axolotls (8 months old) as well as from two aged axolotls (8 years old). Specifically, blastema tissues were acquired 10 days postamputation from juvenile axolotls, whereas aged axolotls did not display any development of blastema tissue. Five additional 8-month-old organisms, along with their corresponding blastema generated 10 days after amputation, were collected to perform experimental validation of *ADAMTS-17*. No animals were sacrificed for this study, and the axolotls were safely reintroduced to their habitat at UMA-CIBAC following the final sample collection.

2.2. RNA Extraction and Sequencing. The collected tissue was preserved in RNA*later* (Invitrogen) at 4°C for a maximum of 24 h prior to processing. RNA was extracted according to the protocol established by Peña-Llopis and Brugarolas [31]. Its quality and concentration were assessed using the High Sensitivity RNA Tapestation (Agilent Technologies). Ribosomal RNA depletion was performed using the Ribo-Zero Gold Kit (Illumina), and libraries were constructed using the SMARTerStranded V2 kit (Takara Bio). Paired-end (PE) sequencing was carried out on Illumina with a read length configuration of 150 and 20 million reads per sample.

2.3. Quantitative Reverse Transcription-Polymerase Chain Reaction (RT-qPCR). Total RNA from five samples of juvenile axolotl limbs and their blastemas, collected 10 days postamputation, and two samples from aged axolotls were extracted using TRIzol reagent (Invitrogen, Catalog No. 15596026), and RNA integrity and quality were evaluated with TapeStation 2200. cDNA was synthesized through a reverse transcription reaction following the instructions provided by the GeneAmp RNA PCR Core Kit (Applied Biosystem, Catalog No. N8080143) using oligo(dT) primers. qRT-PCR assays were performed on 7500 Real-Time PCR System (4351105, Applied Biosystems, Foster City, CA, USA) using Maxima SYBR Green/ROX qPCR Master Mix (2x) kit (Thermo Scientific, Catalog No. K0222). To assess the expression of *ADAMTS17* and *GAPDH* as the housekeeping gene control, the following oligonucleotides were used: *ADAMTS17* (FW: 5′-CTGCTCTGACCTACAAGT GC-3′, RV: 5′-TTGGCATAAAAGTCTCGGCA-3′) and *GAPDH* (FW: 5′-GCTGCCTCCTATGACGAAAT-3′, RV: 5′-TTCCTCGGTGTATCCCAGAA-3′). The relative expression of *ADAMTS17* was determined using the −△△Ct method, comparing its expression in blastema tissues of juvenile axolotls or limbs of aged axolotls to the control condition and normalized with *GAPDH*. Statistical analyses were performed using a *t*-test in GraphPad 7.

2.3.1. Transcriptomic Analysis. In addition to the experimentally obtained samples, data from two other studies were also analyzed. The accession numbers for the samples used are displayed in Table 1.

Quality assessment of the sequencing data was performed using FastQC [32] Subsequently, raw counts were aligned to the reference genome AmexG_v6.0-DD [26] using STAR v.2.7.9a [33]. The resulting transcript-aligned read counts were quantified using featureCounts for R, and the AmexT_v47 transcriptome, accessible at https://www.axolotl-omics.org/assemblies, was used for this purpose. DESeq2 v.1.32.0 [34] was employed for the differential expression analysis. The tissue obtained after amputation of the juvenile axolotl's limb was used as a control. Two separate comparisons were carried out: aged axolotl limb (8 years old) versus control and 10-day blastema from juvenile axolotls versus control. Genes were designated as DEGs if they exhibited $|\log2 \text{ (fold change)}| > 1$ and padj < 0.05. For the Bryant dataset, differential expression analysis was conducted for proximal blastema using upper arm tissue as the control and for distal blastema using hand tissue as the control.

2.3.2. Gene Annotation and Protein–Protein Interaction (PPI) Networks. All genes were named according to the "gene_name" field in the AmexT_v47 transcriptome file, accessible at https://www.axolotl-omics.org/assemblies. In cases where the "gene_name" field was either empty or equal to "N/A," genes were assigned the same name as their gene_id. Priority was given to nomenclature with the [hs] identifier over any other annotation. To distinguish genes with the same name, a numerical identifier was appended; for example, KAZAL.3 indicated the presence of four other genes with the name KAZAL in the dataset. The predicted gene ontology (GO) function and PPIs for each gene were determined using STRING [35]. Only one open reading frame per gene was used as input for STRING; if a gene had more than one transcript, the 0.1 isoform was selected. GO term enrichment analysis was performed using gProfiler2 v.0.2.1 for R [36] with the correction method g:SCS and an adjusted p value significance threshold of 0.05. Redundant terms were clustered using rrvgo v.1.4.4 for R [37]. The PPI visualizations were built with KeyPathwayMineR [38] All network visualizations were created with Cytoscape [39].

2.3.2.1. Coexpression Analysis. Our datasets, as well as the ones from Bryant and Caballero-Pérez, were used as input for WGCNA [40]. The raw counts were normalized using variance-stabilizing transformation (VST) from the DESeq2 package. Batch effect correction was performed with Combat from the sva v3.40.0 package for R [41]. The blockwiseModules function was employed to identify coexpression modules within the data. The function was run with a power of 7 and a minimum module size of 30 genes, with the TOMType set to "unsigned." The resulting coexpression modules were named with random colors. Modules were considered significant if they had a p value < 0.05 and |Module-Trait Correlation| > 0.5. GO enrichment analysis for the genes contained in each module was

Table 1: Analysis of previously published samples.

SRA accession	Sample type	Citation
SRR2885267	Bone	[23]
SRR2885268	Bone	
SRR2885269	Bone	
SRR2885270	Forearm cartilage	
SRR2885271	Forearm cartilage	
SRR2885273	Forearm cartilage	
SRR2885553	Distal blastema	
SRR2885591	Distal blastema	
SRR2885592	Elbow	
SRR2885593	Elbow	
SRR2885594	Forearm	
SRR2885595	Forearm	
SRR2885597	Hand	
SRR2885598	Hand	
SRR2885599	Hand	
SRR2885865	Proximal blastema	
SRR2885866	Proximal blastema	
SRR2885867	Skeletal muscle	
SRR2885868	Skeletal muscle	
SRR2885869	Skeletal muscle	
SRR2885870	Skeletal muscle	
SRR2885871	Upper arm	
SRR2885873	Upper arm	
SRR2885875	Upper arm	
SRR5042766	Front leg	[24]
SRR5042769	Rear leg	

performed with gProfiler2 v.0.2.1 for R[36], and the PPI network was built with KeyPathwayMineR [38] using only the genes with module membership > 0.7. Visualizations were created with Cytoscape and BioRender.com.

2.4. Homology and Structural Biology Analysis. The amino acid sequences for each transcript were obtained from the annotation file (.gtf) for the AmexT_v47 transcriptome. If more than one transcript was available, only the 0.1 isoform was modeled. The amino acid sequence of each protein was used as a query for BLASTP 2.12.0 [42, 43] against the UniProtKB and Swiss-Prot database [44]. Only the sequence with the highest identity percentage in each organism was conserved for the identity plot, and the organisms displayed were selected based on the animal model organisms described in a comprehensive review of model organisms [45]. AlphaFold2 [46] was used to predict the structures of the proteins; templates were used for all predictions, and the ConSurf [47] web server was used to color the residues according to their conservation. Functional domain annotation was performed using the Conserved Domain Database (CDD) [48] and the NCBI Conserved Domain tool [49]. Protein visualizations were created with the PyMOL Molecular Graphics System v.4.60.

3. Results

3.1. Blastemas Share a Defined Transcriptomic Profile. To explore the dynamics of gene expression during the formation of blastema in *A. mexicanum*, we conducted an experiment involving the RNA sample collection. This was achieved by amputating the right forelimb of five juvenile axolotls (8 months old) at a proximal site. These axolotls were bred and raised within CIBAC, a facility dedicated to maintaining a colony derived from wild specimens captured in Lake Xochimilco. Blastema samples were obtained 10 days after the amputation procedure, and RNA was subsequently extracted. The extracted RNA samples underwent sequencing analysis.

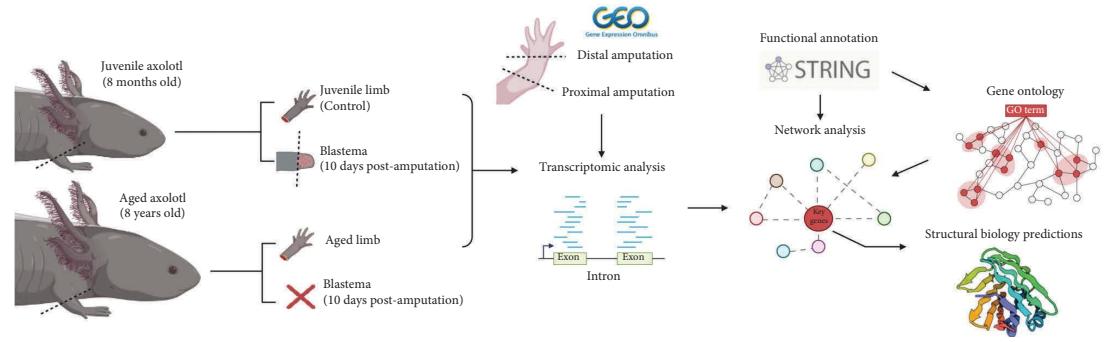
To contextualize our findings, we compared them with RNA-seq data from Bryant [23], who studied a strain of d/d axolotls (white mutant). In their study, forelimbs were amputated at two different sites (proximal and distal), and blastemas were collected 23 days postamputation. To identify the most significant changes in gene expression between the different samples, we performed differential expression analysis (Figure 1(a)).

Overall, we observed that most of the DEGs were downregulated. In our samples, we identified 667 upregulated genes and 2076 downregulated genes in the blastema compared to control tissue. In the distal blastema, we found 4809 upregulated genes and 5577 downregulated genes compared to the control, whereas in the proximal blastema, we found 6143 upregulated genes and 7671 downregulated genes (Figure 1(b)).

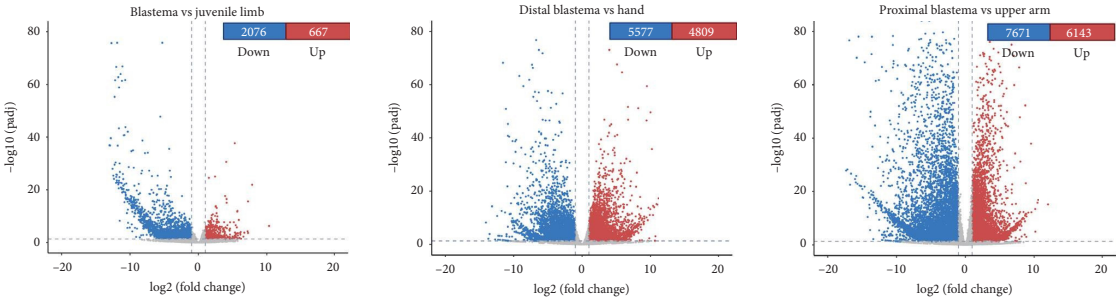
We compared the DEGs in our samples with Bryant's datasets to search for genes that were consistently regulated across experiments. Our aim was to identify expression patterns associated with regeneration regardless of strain or amputation site. We found 1277 DEGs that were common to all three datasets. Specifically, out of these common DEGs, 248 were upregulated and 886 were downregulated in all three samples (Figure 1(c), Supporting Table 1).

Since ontology annotation for *A. mexicanum* is not available, we used STRING [35] to generate a homology annotation for the coding genes in the *AmexT_v47* transcriptome. By performing a GO term enrichment analysis, we found that the DEGs in the blastema were primarily associated with biological processes related to tissue and muscle development, cytoskeleton organization, cell adhesion, extracellular matrix organization, and ossification. However, among the downregulated genes, we observed an enrichment in biological processes such as cytoskeleton organization, muscle structure development, and myofibril assembly. Conversely, the upregulated genes were enriched in terms such as anatomical structure development, negative regulation of cellular processes, and regulation of cell differentiation (Figure 1(d), Supporting Table 2).

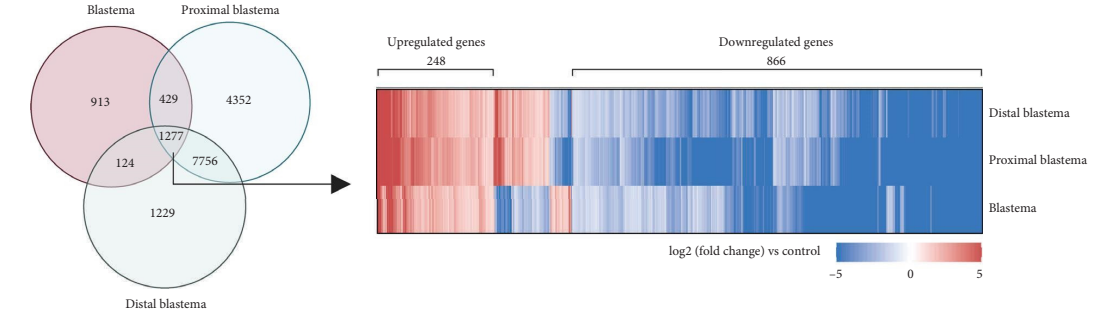
A gene–GO term network was constructed using 10 of the most representative terms from Figure 1(d) to provide a more comprehensive visualization of the genes involved in the process of regeneration. Notably, the majority of genes belonged to the term "anatomical structure development," and a significant proportion of the genes in this category showed downregulation in the blastema compared to the control. However, genes associated with the Wnt pathway (*WNT5A* and *WNT5B*) were upregulated, as were the metallopeptidases *ADAMTS17*, *ADAM8*, *MMP19*, *MMP11*, and *MMP13*, which are also associated with the



(a)

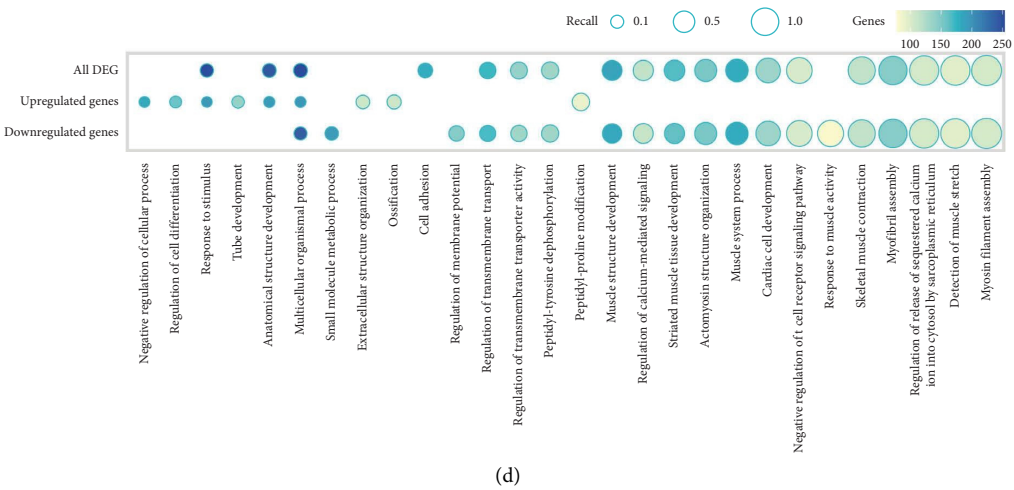


(b)



(c)

Figure 1: Continued.



(d)

Figure 1: Comparison of DEG in different blastema samples. (a) Schematic representation of tissue sampling and downstream analyses. We sampled five juvenile axolotl limbs (control) and two aged axolotl limbs (over 8 years old). Amputation was performed below the elbow (proximal amputation). Blastemas from juvenile axolotls were collected after 10 days, while aged axolotls showed no blastema at the same time point. Datasets published by Bryant et al. were also analyzed; the samples were taken from *d/d* strain axolotls at two different amputation sites: the first at hand level (distal) and the second below the elbow (proximal). Blastemas from Bryant's dataset were collected 23 days postamputation. After RNA extraction and sequencing, differential expression and network analysis were performed to find relevant genes and predict GO terms associated with regeneration. (b) Differentially expressed gene (DEG) volcano plots. The left panel shows the comparison between blastema and control limbs (juvenile axolotls). The middle panel shows the distal blastema compared to the hand. The right panel shows the proximal blastema compared to the upper arm. Red dots represent the upregulated genes (p value < 0.05 and log2 (fold change) > 1). Blue dots represent the downregulated genes (p value < 0.05 and log2 (fold change) > 1). (c) Venn diagram of the intersection between DEG for blastema vs. control, distal blastema vs. hand, and proximal blastema vs. humerus. The left panel shows the heat map for the 1277 genes in the intersection between DEG in aged axolotls and blastema. The color of the tiles represents the log2(fold change) value for each gene. Genes with the same behavior across samples are indicated by parentheses (1134 of 1277 genes). d) Top 20 gene ontology biological processes associated with all DEG in the intersection shown in Figure 1(c) (1134 genes), downregulated genes only (886 genes), or upregulated genes (248 genes). The size of the bar corresponds to the number of genes associated with the significant GO term. The color of the bars represents the log10 (p value) of the term. Created with BioRender.

"extracellular structure organization" term. Similarly, genes related to cell adhesion, myofibril assembly, and actomyosin structure organization were mostly downregulated in blastemas, with a few exceptions such as tenascin (*TNC*), fibronectin (*FN1*), and thrombospondin-2 (*THBS2*). Another group of genes, including *SMAD6*, *SMAD7*, and *BAMBI*, involved in the tumor growth factor beta (TGF-β) signaling pathway, were upregulated and associated with cell differentiation, along with *SOX4* and *NOX4*. Genes associated with epigenetic functions were also found. A notable example is the histone lysine demethylase *KDM3A*, found among the genes involved in differentiation and development. This finding suggests that tissue regeneration may be controlled by epigenetic processes. Notably, genes involved in the ossification and regulation of cell differentiation were also upregulated. Additionally, several genes previously associated with regeneration in axolotls, marked in squared boxes, were obtained from the table published by Haas and Whited [5]. Among these genes, *KAZALD1* stands out as a well-studied gene involved in anatomical development and bone regeneration in axolotls [23]. Furthermore, bone morphogenetic protein 2 (*BMP2*) and fibroblast growth factor 9 (*FGF9*), which are associated with skeletal and cartilage development, were upregulated. These proteins are particularly relevant because a previous study demonstrated their ability to induce blastema formation in axolotls [21]. In summary, our results provide a collection of genes and biological processes that are associated with blastema tissue and the regeneration process in axolotls, which is consistent with and extends the data from previous studies (Figure 2).

3.1.1. Key Genes in Regeneration Revealed by the Transcriptomic Profile of Aged Axolotls. In addition to the previously analyzed samples, we included two aged axolotls (8 years old) in our study, subjecting them to proximal forelimb amputation. Surprisingly, after a period of 10 days postamputation and a subsequent follow-up of 6 months, these aged specimens showed no signs of limb regeneration, in stark contrast to the robust regenerative response observed in juvenile axolotls. Consequently, no postamputation samples were collected from the aged cohort. When we compared the transcriptome of the aged limbs to that of the juvenile limbs, only 172 genes were differentially expressed, with most being downregulated compared to the control (Figure 3(a)). After a GO enrichment analysis, the only significant GO term associated with the DEG in aged limbs was collagen fibril organization.

To investigate relationships among the aged DEGs, we performed *de novo* pathway enrichment using KeyPathwayMineR [38]. As input, we used the DEGs in aged limbs and a PPI network constructed with STRING using

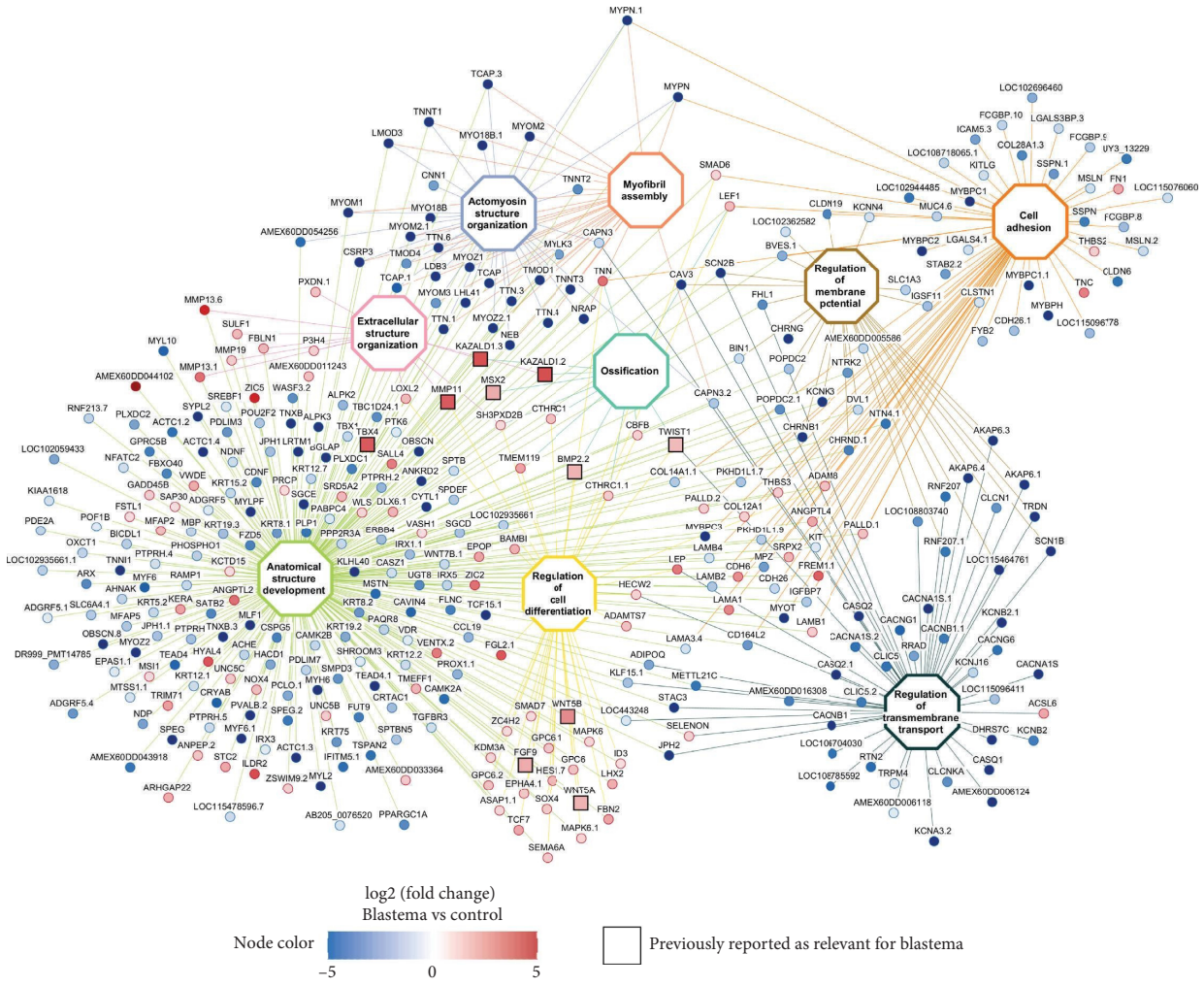


FIGURE 2: Gene ontology terms and genes associated with axolotl regeneration. Gene ontology terms enriched for the DEG in the three blastema datasets analyzed. Node color corresponds to the log2 (fold change) for the blastema vs. control dataset. GO terms are depicted in white, each with a distinct border color. Edges are colored according to the originating GO term. Squared nodes represent the genes that have been previously reported in the literature to be associated with the axolotl regenerative process. Created with BioRender.

the proteome from the *AmexT_v47* annotation file. The interaction network shows mainly the downregulation of type I (*COL1A1, COL1A2*), II (*COL2A1*), V (*COL5A1*), and XI (*COL11A1*) collagens. Some of these proteins appeared more than once because they were assigned the same name in the transcriptome annotation file, so they were marked with a numerical suffix. Additionally, several ribosomal components, such as ribosomal proteins RPS2 and RPS5, as well as signal recognition particle 9 *(SRP9*) and *SEC61G*, which are part of the complex required for protein translation at the endoplasmic reticulum, were observed (Figure 3(b)).

In an effort to identify key genes involved in the tissue regeneration process, we focused on the DEGs that showed contrasting patterns in the blastema vs. control and aged vs. control comparisons. Our goal was to identify genes that were upregulated in blastemas but downregulated in old axolotls, and vice versa, as these could provide valuable insights into the impaired tissue regeneration observed in aging axolotls. Through DEG overlap analysis, we identified 26 genes that consistently exhibited significant differential expression across all four datasets evaluated (Figure 3(c)). Among these, seven genes showed consistent differential expression specifically in blastemas compared to aged limbs. Notably, *CTHRC1.1*, *ADAMTS17*, *GPX7*, *FSTL1*, and *LOC112547415.222* were upregulated in blastemas but downregulated in aged axolotls. Conversely, *NNMT.19* and *NNMT.20* exhibited the opposite pattern, being upregulated in aged axolotls but downregulated in blastemas (Figure 3(d)). In order to experimentally validate one of the key genes involved in tissue regeneration, we analyzed the expression of the *ADAMTS17* gene using limb samples from five 8-month-old juvenile axolotls and their blastema generated 10 days after amputation, as well as samples from 8-year-old aged axolotls. Overexpression of *ADAMTS17* was observed in the blastema of the 8-month-old juvenile axolotls compared to their limb. Interestingly, the expression of this gene significantly decreases in the limbs of 8-year-old

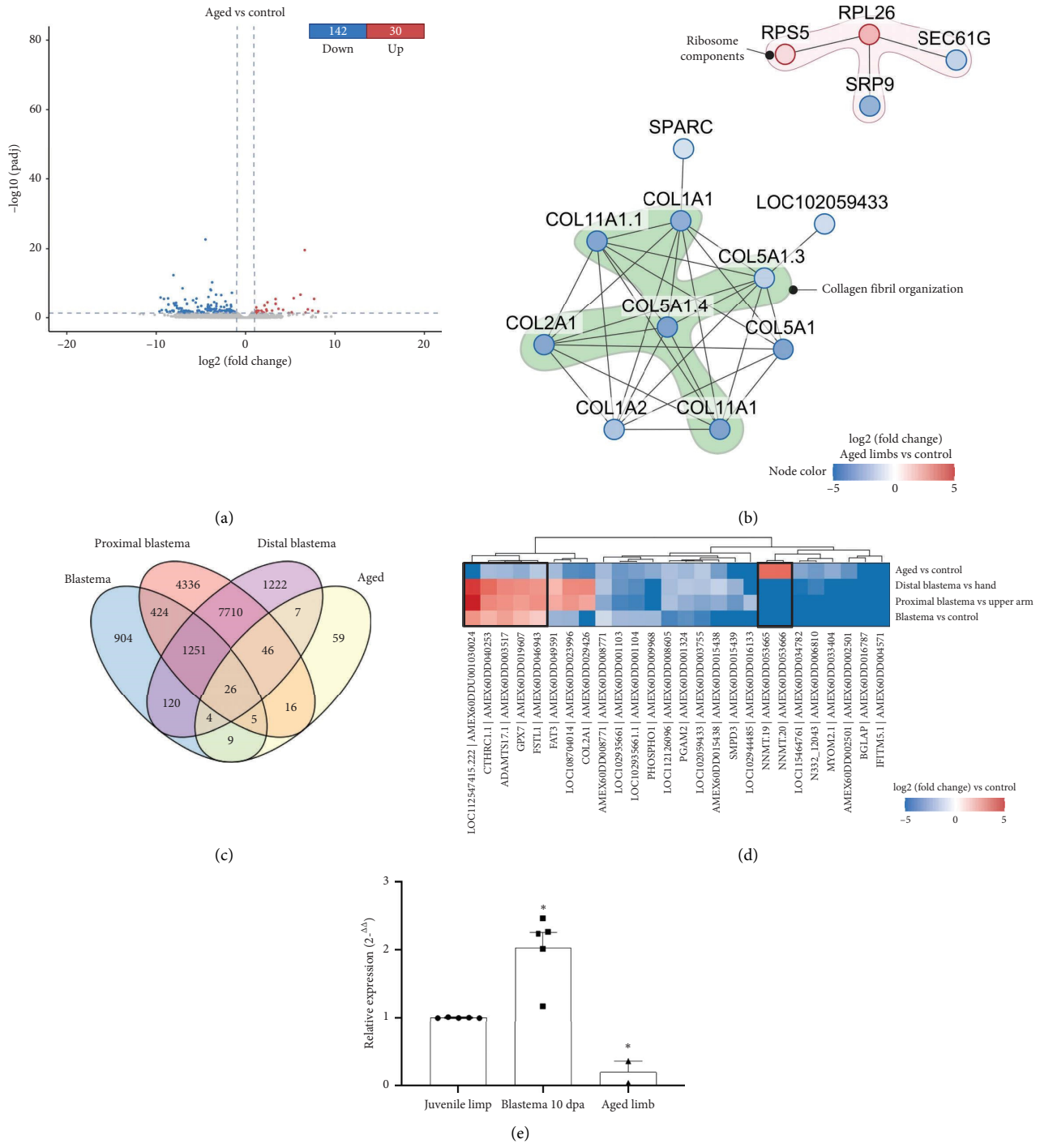


Figure 3: Differential expressed genes in aged axolotls. (a) Volcano plot of the differentially expressed genes (DEGs) obtained for the comparison between aged limbs vs. control limbs (juvenile axolotls). Red dots represent the upregulated genes (p value < 0.05 and log2 [fold change] > 1). Blue dots represent the downregulated genes (p value < 0.05 and log2 (fold change) < - 1). (b) Protein–protein interaction network for the DEGs in aged limbs. Node color represents the log2 (fold change) of aged limbs vs. control. Genes associated with the term "collagen fibril organization" are highlighted in green, while genes that are part of the ribosome are highlighted in pink. Edges represent the predicted interactions between the proteins encoded by the genes. (c) Venn diagram for the DEGs in all four datasets evaluated. The blastema DEGs correspond to the comparisons made previously in this study. (d) Heat map displaying the fold change of all DEGs in the intersection shown in (c) (26 genes). DEGs that show contrasting patterns between the blastema and aged datasets are highlighted with a black box. (e) Experimental validation of *ADAMTS17* was conducted using qRT-PCR. For this purpose, RNA was extracted from five samples of juvenile axolotl limbs and their respective blastemas, as well as from two aged axolotls. Statistical analyses were performed using a t-test. Created with BioRender.

aged axolotls compared to those from juvenile axolotls (Figure 3(e)). These genes are of particular interest because their high expression in blastemas, coupled with their downregulation in aged axolotls that have lost their regenerative capacity, suggests that they may play a critical role in the regeneration process. Therefore, we will refer to this set of genes as regeneration-related genes. Table 2 summarizes the annotation and putative function of these regeneration-related genes.

3.1.2. Coexpression Network Analysis Reveals Gene Modules Associated With Tissue Regeneration in A. mexicanum. As an additional approach, we used coexpression network analysis to examine sets of genes that might be involved in regeneration and show similar transcriptional responses to our set of regeneration-related genes. This analysis incorporated muscle, cartilage, and bone samples from the study by Bryant [23], as well as two leg samples from Caballero-Pérez [24]. The inclusion of these datasets allowed for a more comprehensive and robust understanding of the genes that are coregulated and, thus, similarly responsive, in *A. mexicanum*. The resulting network comprised 35 gene modules, of which 18 modules showed significant associations (p value < 0.05 and absolute Pearson correlation coefficient > 0.5) with the evaluated conditions (Supporting Figure 1, Supporting Table 3). Four of the significant modules showed a strong correlation with the blastema condition (yellow, magenta, dark orange, and black). The yellow and magenta modules were also significantly associated with muscle samples. Genes involved in metabolic processes and muscle structure development are enriched in the yellow module, while the magenta module showed enrichment in genes associated with the cell cycle, DNA replication, and RNA splicing. The black module is exclusively associated with blastema and was enriched in genes associated with developmental processes and multicellular organism development. Notably, the dark orange module did not show any significant GO term association but is also associated with limb samples (Figure 4(a), Supporting Figure 2).

Although none of the modules showed a strong association with the aged samples, we identified a set of 19 genes that had a significant association (absolute module–trait correlation > 0.7 and p value < 0.05) with the trait (Supporting Table 4). Among these genes, the two with the strongest correlation were *AMEX60DD027179*, a non-annotated gene encoding a 127 amino acid protein of unknown function, and *PAFAH2*, a gene encoding platelet-activating factor acetylhydrolase isoform 2. Both genes were negatively associated with aged axolotls and belonged to the turquoise module, which exhibited enrichment in genes associated with the regulation of metabolic processes, organelle organization, and chromatin organization.

To gain further insight into the biological processes associated with each module, we used KeyPathwayMiner to construct PPI networks. Figure 4(b) displays the modules that were significant for blastema and those that contain one or more of the previously identified regeneration-related genes. Network nodes are color-coded based on their gene–trait correlation for blastema, while the edges represent the predicted interactions between the proteins encoded by the genes. The colored shapes indicate the GO terms associated with the module nodes. *CTHRC1,* one of the regeneration-related genes, was found in the midnight blue module, which exhibited significance in muscle samples but had no enriched GO terms. The network revealed proteins associated with multicellular development and collagen trimer. The blue module contained three of the seven regeneration-related genes identified through differential expression analysis (*ADAMTS17*, *GPX7*, and *FSTL1*). This module displayed a strong correlation between cartilage samples and enrichment in genes associated with the regulation of developmental processes. The PPI network for this module showed enrichment of genes related to cartilage development (e.g., *BMP1* and *CHST11*) and cell differentiation (e.g., *SOX8*, *SDC2*, and *DAB2*). Furthermore, the dark orange module, which negatively correlated with blastema, showed enrichment in immune response-related proteins, with some members associated with *STAT3*. The magenta module included genes related to the cell cycle, particularly belonging to the kinesins family (*KIF*). Notably, genes related to histone modification were also observed, including the polycomb group members *EED* and *EZH2*, as well as *SUV39H1*, a histone lysine methyltransferase. Genes involved in the regulation of gene expression, including *HDAC2* (histone deacetylase), *DNMT3A* (DNA methyltransferase), and transcription factors such as *SOX4*, *MYCN*, and *SALL1*, were found in the PPI network for the black module, one of the modules closely associated with the blastema and containing the regeneration-related gene *LOC112547415*. Finally, the two *nicotinamide N-methyltransferase (NNMT)* genes were found in the yellow module, which was enriched with several proteins associated with the organization of the actin cytoskeleton, as well as various components of the cytoskeleton, such as troponin genes (*TNT*), actin (*ACTN*), myotilin (*MYOT*), and myosin (*MYL*). Interaction between *SGCE*, another protein from the cytoskeleton, and *FSTL1* was also observed in this module.

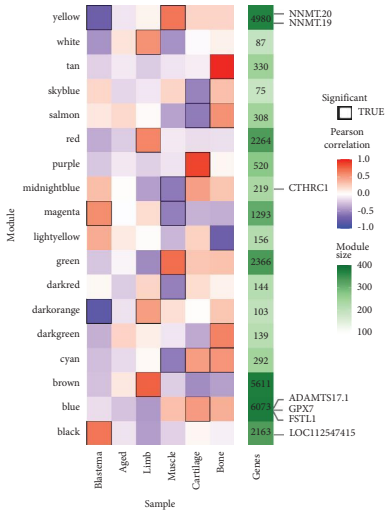
In summary, we identified gene modules associated with regeneration and genes with similar transcriptional responses to our set of regeneration-related genes. Notably, these identified modules included our proposed regeneration-related genes, as well as other genes potentially involved in regeneration. These results highlight the potential interplay between the regeneration-related genes we identified and other factors involved in axolotl limb regeneration, offering insights into the potential biological significance of the identified genes.

3.1.3. The Regeneration-Related Genes Are Conserved in Vertebrates. Through transcriptomic analysis, we identified a group of genes relevant to regeneration in *A. mexicanum*. To assess the functionality and presence of the proteins encoded by these genes in other organisms, we performed a homology search against the UniProtKB + Swiss-Prot database [44] using a selection of animals based on

Table 2: Annotation and putative function of the regeneration-related genes identified.

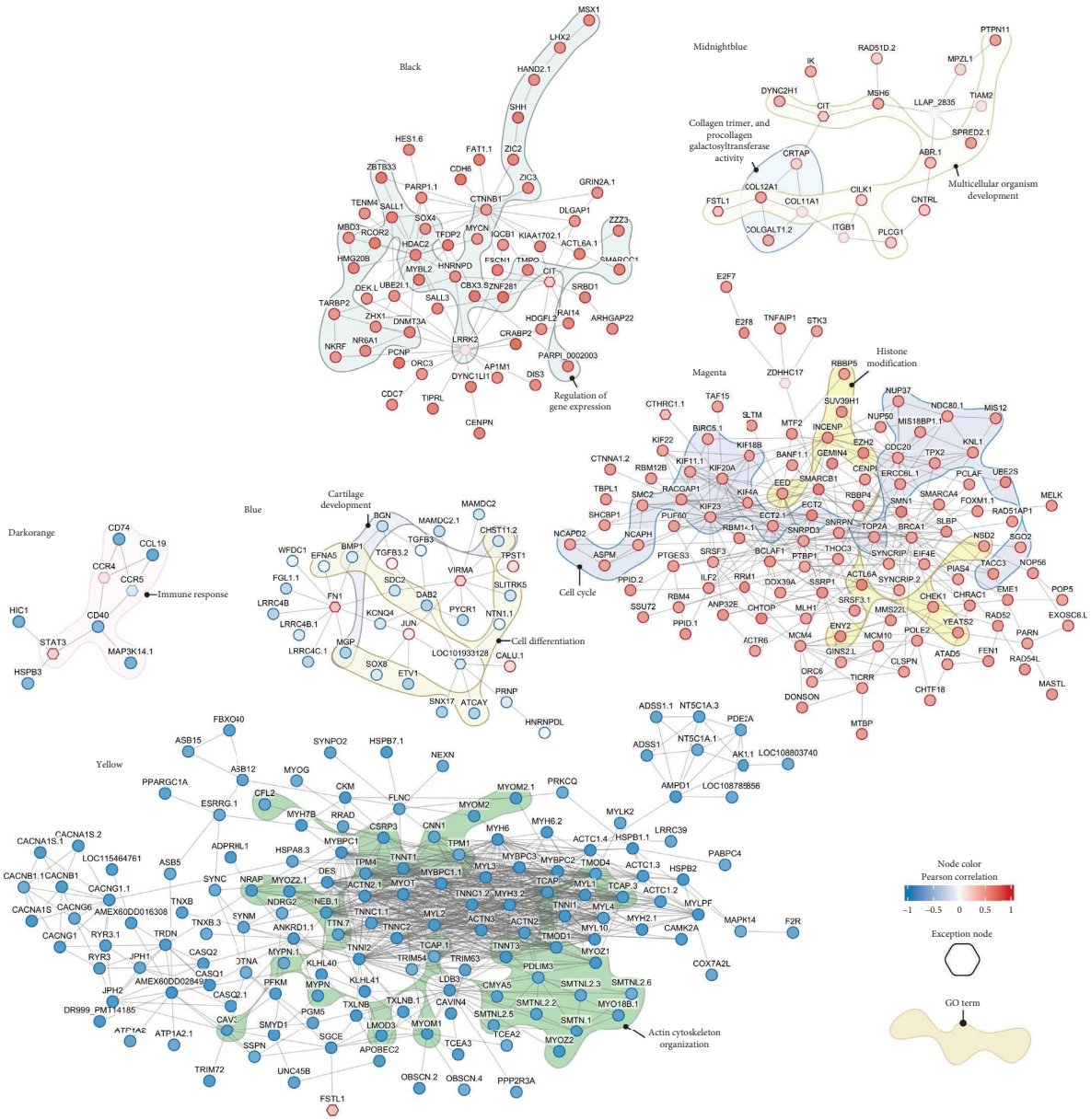
Gene ID	Gene Name	Homolog	Location		
			Chromosome	Start	End
AMEX60DD003517	*ADAMTS17*	*ADAMTS17* A disintegrin and metalloproteinase with thrombospondin motifs 17 *Terrapene carolina triunguis* NCBI Reference Sequence: XP_026504145.1	chr11p	81488489	82602596
AMEX60DD019607	*GPX7*	*GPX7* PREDICTED: glutathione peroxidase 7 *Latimeria chalumnae* NCBI Reference Sequence: XP_006006032.1	chr1q	1276741180	1277033638
AMEX60DD040253	*CTHRC1.1*	*CTHRC1* collagen triple helix repeat-containing protein 1 *Xenopus laevis* NCBI Reference Sequence: XP_018123683.1	chr5q	899669621	899687475
AMEX60DD046943	*FSTL1*	FSTL1 follistatin-like 1 *Rhinatrema bivittatum* NCBI Reference Sequence: XP_026504145.1	chr7p	70865121	71016956
AMEX60DDU001030024	*LOC112547415.222*	uncharacterized protein LOC112547415 *Pelodiscus sinensis* NCBI reference sequence: XP_025045344.1	C0173161	1	8438
AMEX60DD053665	*NNMT.19*	NNMT Indolethylamine *N*-methyltransferase-like *Microcaecilia unicolor* NCBI Reference Sequence: XP_030077372.1	chr9p	79243911	79434013
AMEX60DD053666	*NNMT.20*	NNMT Nicotinamide *N*-methyltransferase-like *Rhinatrema bivittatum* NCBI Reference Sequence: XP_029428629.1	chr9p	79380383	79403217

Note: The gene IDs correspond to the AmexT_v47 transcriptome. The homolog sequence listed is based on the closest match found in the NCBI reference database.



(a)

Figure 4: Continued.



(b)

Figure 4: Gene coexpression analysis of axolotl samples. (a) Module–trait correlation for the 18 significant coexpression modules identified. Tile color represents the module–trait Pearson correlation coefficient. Significant modules (p value < 0.05) are enclosed in a black box The green panel indicates the module size. (b) Overview of the coexpressed genes within the black, midnight blue, yellow, blue, dark orange, and magenta modules. Node color corresponds to the Pearson correlation coefficient for the blastema samples. Edges represent the putative protein–protein interactions predicted by STRING. Gene ontology terms associated with some nodes are highlighted with colored shapes. Exception nodes (genes not belonging to the module) are marked with a hexagon. Created with BioRender.

S. Hedges' comprehensive review of model organisms [45]. Figure 5 illustrates the identity between the proteins encoded by the regeneration-related genes and their closest homologs in each selected organism. The UniProt IDs for each sequence and the full set of results are given in Supporting Table 5.

It is noteworthy that all of the regeneration-related proteins in *A. mexicanum* show homology in vertebrates, although their presence in invertebrates is variable. Among the regeneration-related proteins, CTHRC1 has the highest percentage of identity in vertebrates, while the NNMT proteins have lower conservation and a lower percentage of positives, indicating that only a fraction of the target protein matches the query sequences. However, FSTL1 and ADAMTS17 also demonstrate conservation in vertebrates but not in invertebrates. GPX7 is the only protein from the regeneration-related genes that is found in all of the organisms evaluated, albeit with a noticeably lower percentage

Figure 5: Conservation of regeneration-related proteins in *A. mexicanum*. The plot shows the percentage identity of each regeneration-related *A. mexicanum* protein (query) vs. its closest homolog in other selected animals (target). Bubble color indicates the percentage identity between the query and target protein, while the bubble size is proportional to the percentage of positives. Positives refer to amino acids in the subject sequence that are either identical to or have similar chemical properties as those in the query sequence. Missing values indicate that no similar proteins were found in the subject organism. Created with BioRender.

of identity in invertebrates compared to vertebrates. Notably, the gene *LOC112547415* does not appear in the graph as it lacks an associated open reading frame, leading us to hypothesize its potential as a noncoding RNA.

Subsequently, the 3D structure of the proteins encoded by each regeneration-related gene in *A. mexicanum* was modeled using AlphaFold2 [46], while sequence conservation was assessed using the Consurf [47] server and the UniRef90 database. The complete predictions and confidence scores can be found in Supporting Figure 3. In addition, the CDD [48] was used to identify conserved domains within the proteins, with the details summarized in Table 3. Regarding the NNMT proteins, they share 97% identity with each other (Supporting Figure 4), and both have an S-adenosylmethionine-dependent methyltransferases domain (Figure 6(a)). NNMT.19 also possesses an extra N-terminal domain that does not appear to be conserved, as indicated by the "insufficient data" label from ConSurf (Figure 6(b)). However, upon comparing NNMT.19 with a previously published crystallographic structure of human NNMT [50] it became evident that both axolotl proteins possess an incomplete catalytic domain, suggesting that only half of the protein is present in axolotls (Figure 6(c)), consistent with the low percentage of positive amino acids found previously.

ADAMTS17 from *A. mexicanum* has several conserved domains, including a catalytic domain [51] with the zinc-binding HExxHxxGxxH consensus motif, characteristic of the catalytic site of the ADAMTS family of metalloproteases [52–54]. It also contains a propeptide from the reprolysin family, two cysteine-rich regions, and several TSP-1 motifs. A comparison with the crystallographic structure of human ADAMTS5 [55], an enzyme of the same family, shows that the catalytic residues of the axolotl's ADAMTS17 are present and that the sequence around them is conserved, suggesting that this protein could be catalytically active (Figure 7(a)).

On the other hand, two domains were identified for FSTL1: a Kazal-type serine protease inhibitor and an EF-hand domain. The ConSurf analysis shows that FSTL1 from *A. mexicanum* is highly conserved. Notably, a set of disulfide bonds with high conservation, together with the Kasal domain, corresponds to the follistatin-like 1 (Fstl1-FK) domain described by Li et al. [56] (Figure 7(b)).

GPX7 has only one thioredoxin-like domain; however, structural analyses revealed that the catalytic site and residues of the protein are conserved in *A. mexicanum*. A comparison between the catalytic site of *H. sapiens* GPX7 [51] and the predicted structure for *A. mexicanum* GPX7

Table 3: Domain annotation for the regeneration-related *A. mexicanum* proteins.

Name	Gene ID	Transcript ID	*E*-value	Residue range	Domain
NNMT.20	AMEX60DD053666	AMEX60DD201053666.1	1.80e- – 48	5–133	ID: cl17173 S-Adenosylmethionine-dependent methyltransferases
NNMT.19	AMEX60DD053665	AMEX60DD301053665.1	8.28e – 41	52–155	ID: cl17173 S-Adenosylmethionine-dependent methyltransferases
ADAMTS17	AMEX60DD003517	AMEX60DD201003517.1	1.18e – 24	26–163	ID: pfam01562 Reprolysin family propeptide
			3.56e – 84	217–434	ID: cd04273 Zinc-dependent metalloprotease, ADAMTS_like subgroup
			7.95e – 23	451–518	ID: cl20316 ADAM cysteine-rich domain
			7.31e – 16	531–583	ID: smart00209 TSP-1 type 1 repeat
			1.26e – 08	615–684	ID: cl41950 ADAM cysteine-rich domain
			3.06e – 07	821–873	ID: pfam19030 TSP-1 type 1 domain
			3.10e – 10	937–982	ID: pfam19030 TSP-1 type 1 domain
			2.14e – 12	991–1040	ID: pfam19030 TSP-1 type 1 domain
FSTL1	AMEX60DD046943	AMEX60DD201046943.1	3.46e – 12	30–9	ID: cd01328 Kazal-type serine protease domain
			4.27e – 73	112–225	ID: cd16233 EF-hand
GPX7	AMEX60DD019607	AMEX60DD301019607.1	5.40e – 105	46–198	ID: cl00388 Thioredoxin-like catalytic residues

Note: The domain ID is from the Conserved Domain Database (CDD).

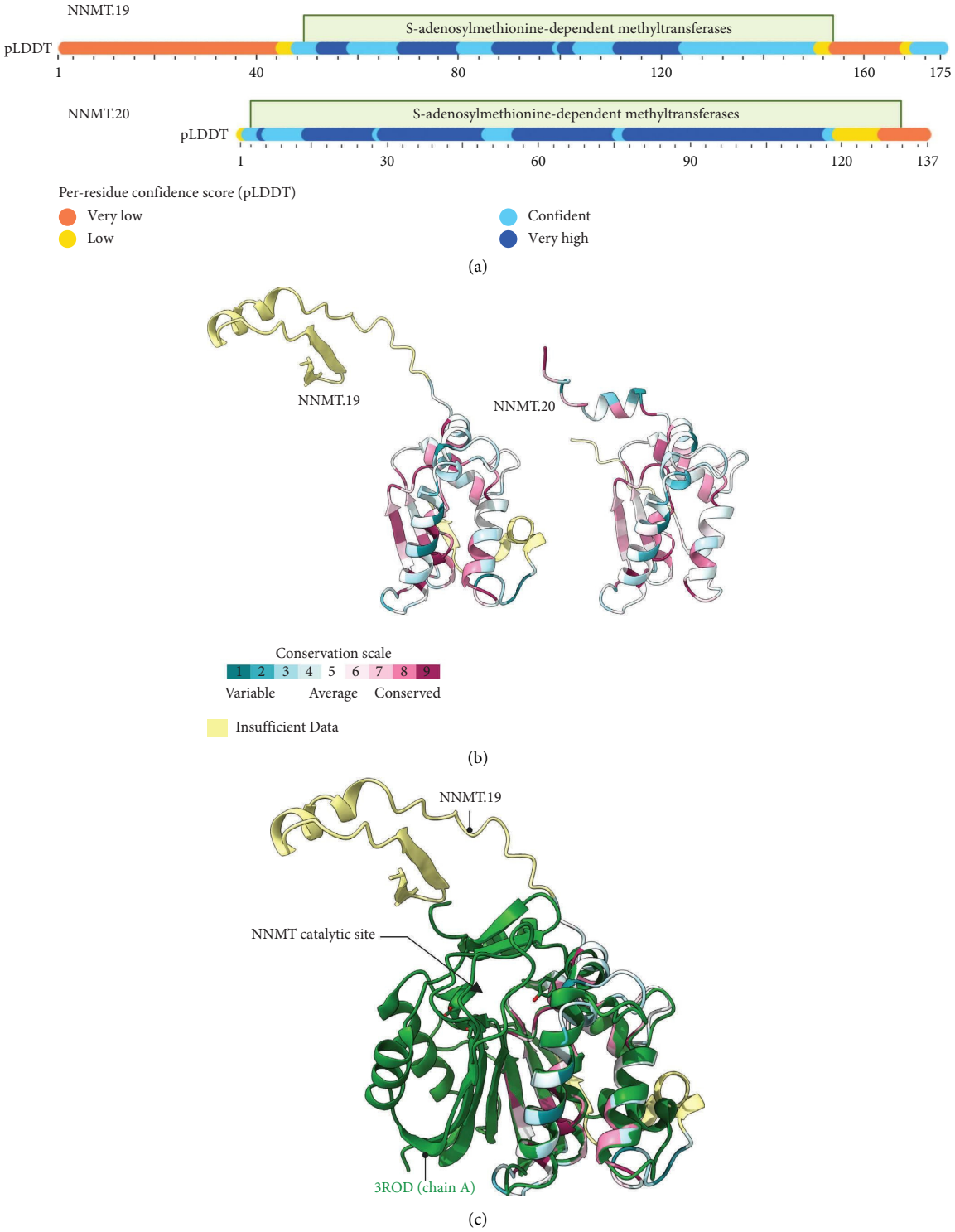


Figure 6: Sequence and conservation of *A. mexicanum* NNMT proteins. (a). Domain overview of *A. mexicanum* NNMT.19 and NNMT.20 proteins. The figures represent the sequence alignment of the proteins, with the scale corresponding to the number of amino acids in each protein. The predicted conserved domains are annotated with a green box. The per-residue confidence (pLDDT) for the AlphaFold models is represented by a bar. The complete structures can be found in Supporting Figure 3. (b) AlphaFold2 structure prediction for both NNMT proteins. The predictions were obtained using the monomer preset for AlphaFold, with templates. The structures are colored according to their sequence conservation as calculated by the ConSurf server. (c) Structural alignment between *A. mexicanum* NNMT.19 and an x-ray crystallography structure for human nicotinamide *N*-methyltransferase (NNMT) deposited in PDB (ID: 3ROD). Created with BioRender.

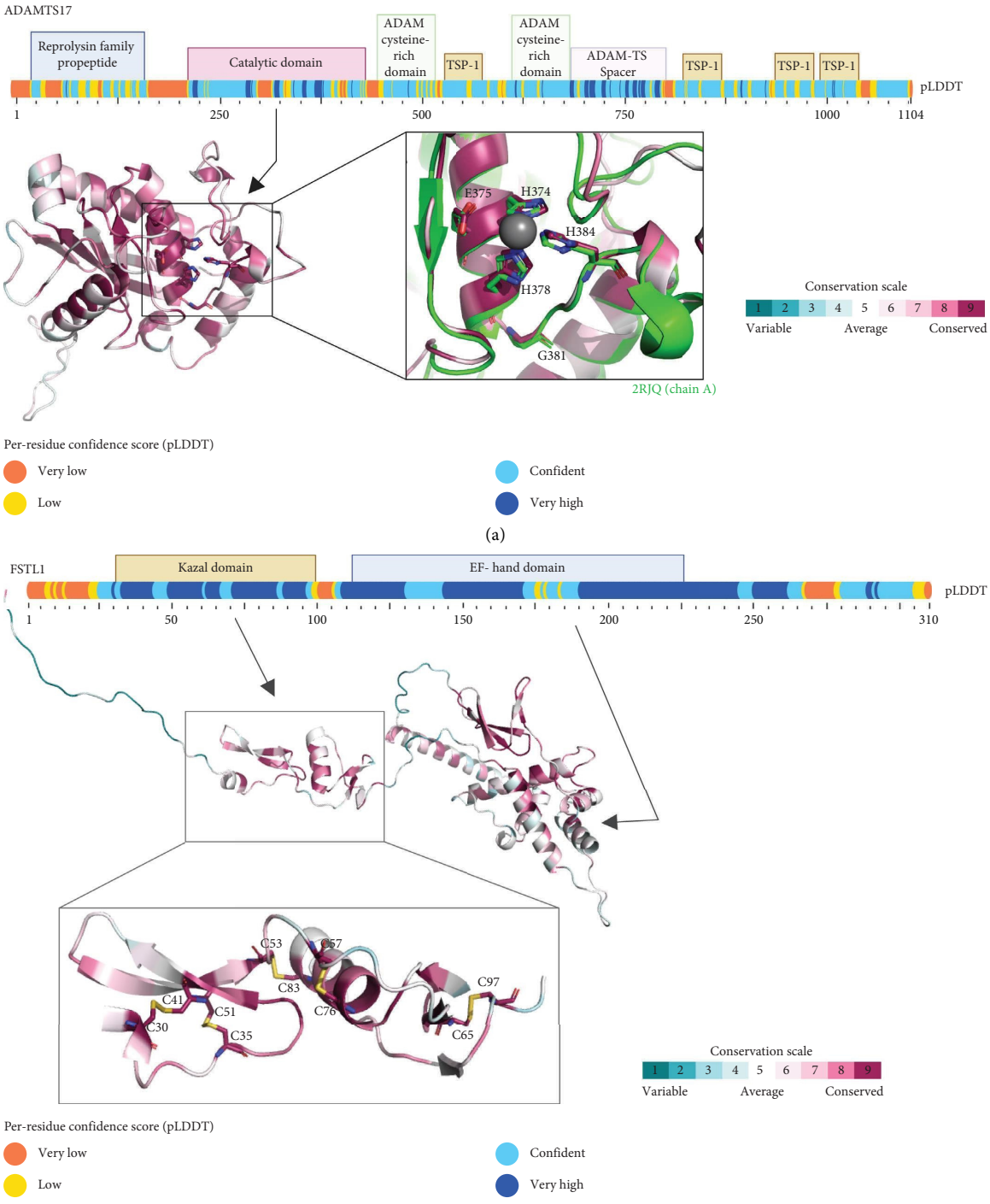


Figure 7: Continued.

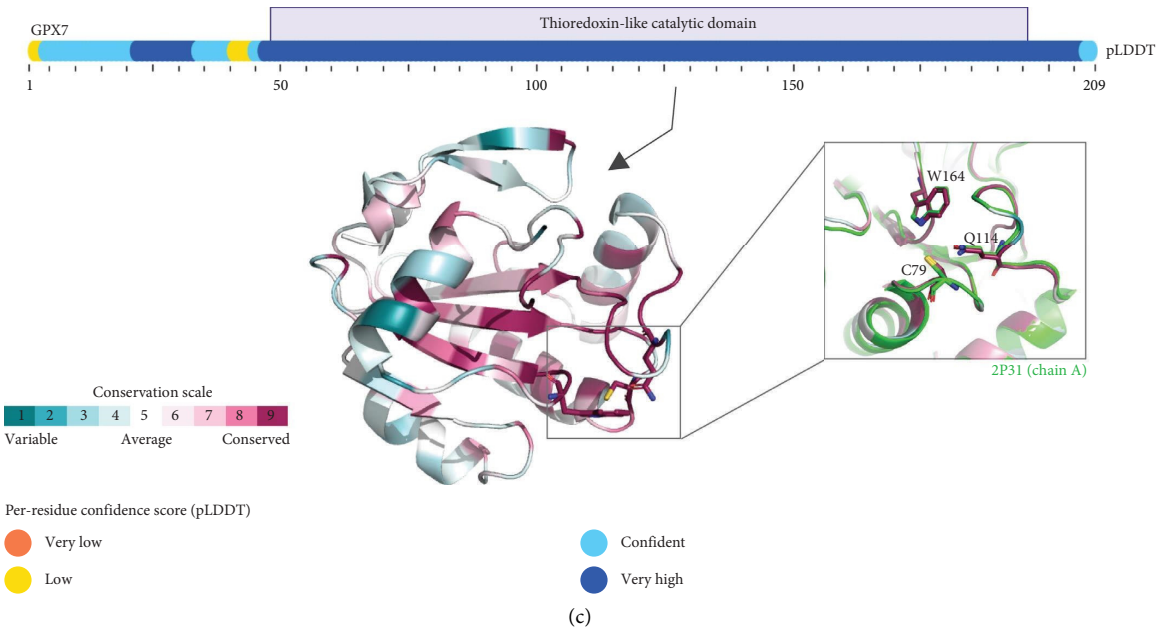


Figure 7: Conservation of ADAMTS17, FSTL1, and GPX7 proteins in *A. mexicanum*. Structures were predicted with AlphaFold2, and the conservation analysis was performed using the ConSurf server. (a) Domain overview and predicted structure of the catalytic domain of *A. mexicanum* ADAMTS17. Residues are colored according to their degree of conservation. The inset shows the structural alignment of the catalytic sites of *A.mexicanum* ADAMTS17 and *H. sapiens* ADAMTS5 (PDB: 2RJQ, green). The per-residue confidence (pLDDT) for the AlphaFold models is represented with a bar. The complete structures can be found in Supporting Figure 3. (b) Predicted domains and structure of *A. mexicanum* FSTL1. Residues are colored according to their degree of conservation. The inset shows the conservation of the cysteine bonds required for FSTL1 folding. The colored bar represents the per-residue confidence (pLDDT) for the AlphaFold model. (c) Domain overview and predicted structure of the catalytic domain of *A. mexicanum* GPX7. Residues are colored according to their degree of conservation. The inset shows the structural alignment of the catalytic sites of *A. mexicanum* GPX7 and *H. sapiens* GPX7 (PDB: 2P31, green). The colored bar represents the per-residue confidence (pLDDT) for the AlphaFold model. Created with BioRender.

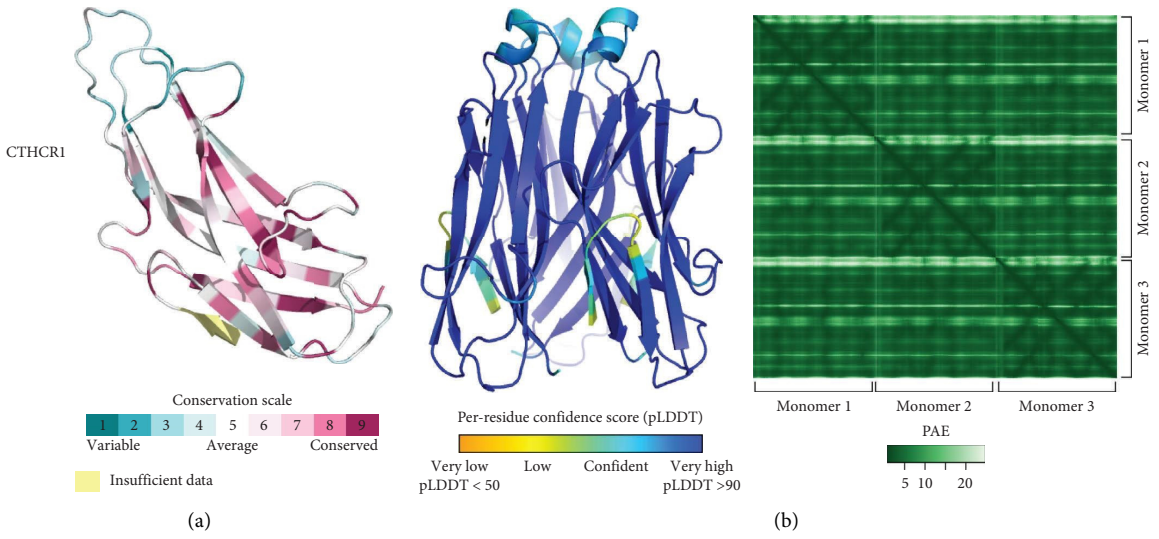


Figure 8: CTHRC1 conservation in *A. mexicanum*. (a). Complete predicted structure for *A. mexicanum* CTHRC1. The 3D structure was predicted using AlphaFold2, with residues colored according to their degree of conservation as calculated by the ConSurf server. (b) Heterotrimer 3D model prediction for CTHRC1. The prediction was obtained with AlphaFold2 multimer, with the residues colored according to the per-residue confidence score. The right panel shows the predicted aligned error for the trimer. Low values indicate the higher confidence in the residue–residue interaction. Created with BioRender.

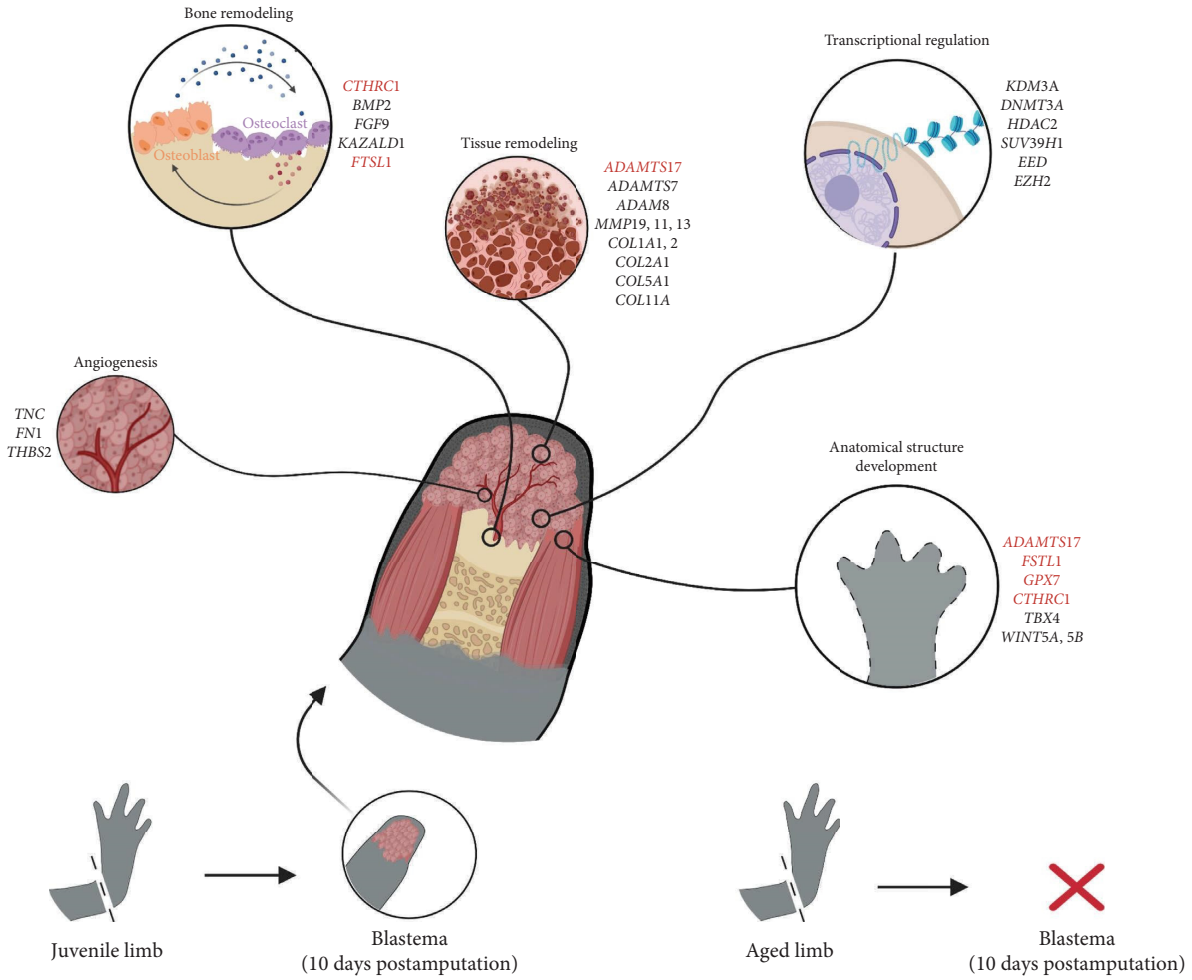


Figure 9: Graphical summary. The diagram displays key genes associated with blastema development during axolotl limb regeneration, along with the biological processes they are linked to. Genes considered vital for regeneration, highlighted in red, were identified by their absence in aged limbs that are incapable of regeneration. Created with BioRender.

shows that the residues W164, C79, and Q114, which are required for GPX7 activity, are also present in our prediction (Figure 7(c)).

Finally, no conserved domain was identified for the CTHRC1 protein; however, structural prediction and conservation analysis suggest its preservation (Figure 8). CTHRC1, also known as collagen triple helix repeat containing-1, is a secreted protein involved in osteogenesis and bone remodeling processes and typically exists as a homotrimer [57]. Using the AlphaFold multimer model, we obtained a high confidence structure for the axolotl CTHRC1 trimer. Taken together, our results suggest that the regeneration-related genes highlighted in this study encode proteins that are conserved in other vertebrates, allowing us to infer their function based on their homologs.

4. Discussion

A. mexicanum, commonly known as axolotl, is a salamander that captivates scientists due to its exceptional characteristics [58]. Unlike other amphibians, axolotls exhibit neoteny, retaining larval features while achieving sexual maturity [1]. However, their most remarkable attribute is their ability to regenerate nearly all tissues and organs, setting them apart from other vertebrates. Despite descriptive and histological studies shedding light on tissue regeneration in *A. mexicanum*, the underlying molecular mechanisms remain enigmatic [12, 14, 20]. Comprehensive studies of the molecular mechanisms involved in axolotl regeneration are of particular interest to the field of regenerative medicine, with potential implications for human therapies [30, 59]. To gain a broader understanding of the genetic basis of axolotl tissue regeneration, we conducted a transcriptomic study on a group of *A. mexicanum* from the CIBAC, a center dedicated to preserving the native axolotl population in Lake Xochimilco, Mexico City. Notably, these specimens were bred under conditions closely resembling their natural environment, reducing confounding factors that may be encountered in laboratory-raised counterparts [60, 61].

We began by examining samples from amputated limbs and the blastemas that form in response to injury. Blastema formation is triggered by signals from surrounding tissues that induce the dedifferentiation of nearby cells into a proliferative and multipotent state. This

organ is crucial for regeneration in axolotls, and, because of that, is subject to complex transcriptomic regulation [19, 62]. Employing next-generation RNA sequencing, we conducted differential expression analysis between blastemas and juvenile limbs, comparing our results to a previously published dataset to identify shared expression patterns. Additionally, we generated a predicted interactome and performed GO annotation for the axolotl proteome to further describe the putative functions of the genes identified.

Taken together, our findings suggest a global downregulation of genes associated with muscular tissue and anatomical development during limb regeneration, consistent with the cellular dedifferentiation that takes place in the blastema [63]. However, a few genes related to development, such as *TBX4* [64], *BMP2* [65], and *KAZALD1* [23], as well as wound healing, such as *SALL4* [66], were upregulated in blastemas. Furthermore, genes associated with the regulation of cell differentiation, such as *WNT5A* and *WNT5B*, exhibited increased expression. *WNT5A* and *WNT5B* play established roles in bone morphogenesis, hematopoiesis, and cartilage homeostasis [67–69]. This suggests that, while tissue-specific genes are silenced during regeneration, a select few guide the development of the new limb.

In support of the aforementioned hypothesis, our gene coexpression network analysis revealed two modules positively associated with blastema samples. These modules contained several proteins involved in histone modification and DNA methylation, including *DNMT3A*, *HDAC2*, *SUV39H1*, *KDM3A*, and the Polycomb group members *EED* and *EZH2*. Epigenetic regulation has been shown to be crucial for tissue regeneration in axolotls; the pharmacological inhibition of DNA methyltransferases causes impairments in blastema formation [70], and histone deacetylase inhibitors inhibit regeneration [71]. The Polycomb group is also known for its role in embryonic development and tissue differentiation, suggesting its likely importance in axolotl tissue regeneration due to its evolutionary conservation [72].

Intriguingly, we also observed upregulation of genes associated with extracellular matrix organization in the blastema compared to control tissue, while cell adhesion factors were predominantly downregulated. This may be linked to the extracellular matrix degradation (histolysis) that occurs at the amputation site, facilitating cell migration, differentiation, and blastema formation [73]. Among the main players in this tissue remodeling process are metalloproteases, with several, such as *MMP13*, *MMP19*, *ADAM8*, and *ADAMTS17*, found to be upregulated in the blastema [74]. Furthermore, genes involved in angiogenesis and wound healing, including tenascin (*TNC*), fibronectin (*FN1*), and thrombospondin-2 (*THBS2*), are also upregulated. Alongside the overall downregulation of myosin and actin proteins, these findings indicate the extensive histolysis and angiogenesis process occurring in the tissue surrounding the blastema [75, 76].

The significance of the extracellular matrix becomes evident in the analysis of samples collected from aged axolotls that lacked regenerative capabilities. Notably, we detected the downregulation of several collagens and ribosomal proteins in aged limbs, indicative of impaired cell–cell interactions and extracellular matrix composition, commonly associated with aging [77, 78]. Processes such as osteogenesis, proliferation, and differentiation are influenced by extracellular matrix stiffness [79–81], making the reduced collagen expression in aged organisms particularly intriguing. This leads us to hypothesize that extracellular matrix organization plays a central role in axolotl regeneration. Further investigation into the changes occurring in the extracellular matrix during blastema formation and limb regeneration in axolotls could provide valuable insights into the regulatory networks involved in this process.

A comparative analysis between aged limbs and blastemas in juvenile specimens postamputation led us to the identification of a set of genes that may be required for initiating the regeneration process. Among these, seven genes were pinpointed, including one putative lncRNA (*LOC112547415.222*) and six coding genes. A blast search against model animals demonstrated homologous counterparts of the six coding proteins in vertebrates, indicating their potential importance. In particular, four regeneration-associated genes (*FSTL1*, *ADAMTS17*, *GPX7*, and *CTHRC1*) exhibited high expression levels in regenerating tissue but were underexpressed in aged axolotls. Interestingly, *ADAMTS17* is underexpressed in 8-year-old axolotls compared to juvenile axolotl limbs. This could suggest that the molecular mechanisms required for its transcriptional activity are impaired or that an epigenetic mechanism may be involved in this process [82]. This might partially explain why these organisms were unable to regenerate their limbs. Structural and conservation analyses further highlighted that these genes encode conserved proteins in vertebrates, which, together with structural predictions, allows us to infer their potential functions.

In this regard, ADAMTS17 is an extracellular metalloprotease involved in collagen processing, extracellular matrix degradation, cartilage cleavage, development, and angiogenesis. This protein has a conserved catalytic domain [52–54], suggesting it is catalytically active. Studies in *Adamts17* knockout mice showed that it is required for proper skeletogenesis and skeletal muscle development [83]. Mutations in ADAMTS17 are also linked to the Weill-Marchesani syndrome in humans, affecting connective tissue and leading to impaired vision, short stature, and musculoskeletal anomalies [84, 85]. Additionally, ADAMTS17 mutations have been implicated in short height and glaucoma in dogs [86] and humans [87]. Nevertheless, the precise function of this protein remains enigmatic. Studies have observed that a catalytically inactive ADAMTS17 interferes with fibrillin-1 secretion, resulting in elastic fiber abnormalities and intracellular collagen accumulation in fibroblasts from patients with Weill-Marchesani syndrome [88]. Another noteworthy protein involved in cell–matrix interactions is FSTL1, a secreted glycoprotein that participates in the regulation of the TGF-β, BMP, and Wnt pathways [89, 90]. In axolotls, the identification of two domains, a Kasal and an EF-hand domain, aligns with the

architecture described in other organisms [56, 91]. FSTL1 has been found to be critical for tracheal and central nervous system development in mice, with *Fstl1*$^{-/-}$ mice exhibiting cyanotic traits due to tracheal malformations [92, 93]. It is also involved in vascularization and vascular epithelial homeostasis maintenance [94, 95]. Overall, the observed downregulation of ADAMTS17 and FSTL1 in aged axolotls compared to juvenile limbs further highlights the importance of the extracellular environment during limb regeneration.

Furthermore, CTHRC1, involved in the TGF-β pathway, functions as a mediator of osteoblast–osteoclast communication and plays a role in osteogenesis and bone remodeling [96]. In axolotls, bone resorption mediated by osteoclast is important for the adequate integration of regenerated bone [97]. CTHRC1 also promotes angiogenesis by inhibiting collagen deposition and promoting cell migration [98]. Knockout mice for this gene showed reduced bone density and arthritis [99, 100], further suggesting the importance of this protein in bone and cartilage development.

GPX7, another protein with high expression in blastema but reduced levels in aged limbs, is a peroxidase, and its catalytic domain has been found to be conserved in axolotls. GPX7 is vital for oxidative stress resistance [101, 102]. Studies have linked low GPX7 levels to increased adipogenesis and fat accumulation in mice [103]. Interestingly, a recent study has observed that the bone marrow of aged axolotls has a higher fat content than that of its younger counterpart [30], which could be associated with the reduced GPX7 expression in aged axolotl limbs. Moreover, tissue regeneration in axolotls requires the production of reactive oxygen species [104], suggesting that enzymes involved in the regulation of oxidative stress are important for this process.

Lastly, we also identified two NNMT proteins with high expression in aged axolotls but with decreased expression in the blastema. While the sequences of both proteins showed homology to other vertebrate NNMTs, we noticed that only half of the catalytic domain of the protein is present in our axolotl proteins. NNMT enzymes belong to the group of SAM-dependent methyltransferases [105], and they have been implicated in several epigenetic processes since SAM serves as the methyl donor used by DNA and histone methyltransferases. Consequently, high levels of NNMT cause a decrease in SAM availability, which leads to the inhibition of other methyltransferases [106, 107]. Although the functional status of the axolotl NNMT proteins we identified remains unclear, it is evident that these transcripts play a central role in aged axolotls, and further investigation is needed to unravel their mechanisms of action. These two transcripts were identified in the yellow coexpression module, displaying a strong negative correlation with blastema and enrichment in genes associated with metabolic processes, thereby accentuating their central role in regeneration and aged limbs. A graphical summary of our main findings is provided in Figure 9.

5. Conclusions

Our research provides a comprehensive overview of the changes in gene activity that occur during tissue regeneration in axolotls. By studying older axolotls, we discovered a group of genes that may be crucial for the regenerative abilities of *A. mexicanum*. Through an analysis of how these genes are preserved across different species, we suggest that they play critical roles in development, bone formation, and the organization of the extracellular matrix. Our results also indicate that the proteins produced by these genes exist in other vertebrates, suggesting that these genes could serve as a starting point for studying regeneration in other animals. Although our findings may not lead directly to therapeutic applications in human regenerative medicine, the insights gained from studying axolotl regeneration are invaluable to the field. Research has shown that mammalian transcription factors can trigger a regenerative response in axolotls, implying that similar regenerative mechanisms may exist in other organisms. Therefore, it is essential to identify the genes and factors responsible for initiating this process. Such knowledge has the potential to inspire innovative therapeutic strategies in human regenerative medicine in the future. In conclusion, our findings emphasize the importance of understanding the genes at work in *A. mexicanum*, providing valuable insights into the molecular processes behind tissue regeneration and the impact of aging on regenerative abilities.

Data Availability Statement

The GO term information and PPI network built and used in this work can be explored in STRING using the organism identifier STRG0034MNQ or the link https://version-11-5.string-db.org/organism/STRG0034MNQ. Additionally, the gene annotation file, ppi predicted network, an AnnotationDbi OrgDb package, and a Txdb object with the gene names and associated GO terms can be accessed at the https://github.com/aylindmm/A.mexicanum GitHub repository. The raw sequencing data were deposited in NCBI's Gene Expression Omnibus under Accession ID GSE237864.

Ethics Statement

All studies conducted with *A. mexicanum* were approved under the ethical considerations authorized by the Ethics and Research Committee of the Division of Biological and Health Sciences at the Universidad Autónoma Metropolitana, Xochimilco Campus, in accordance with the current legislation and the guidelines established by the Mexican Ministry of Environment and Natural Resources (Reference Number CEI.2023.007).

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Aylin Del Moral-Morales and Cynthia Sámano joint as first authors.

Funding

This work was funded by the Consejo Nacional de Ciencia y Tecnología (CONACyT) Fondo CB-SEP-CONACyT 284748 and PRODEP (511/2023-3066-1599) to E.S-R, and the Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCyT), Fondo Ciencia de Frontera (CF-2023-G-1558) to C.S.S. All the computational analyses were performed on the esrs-epigenetics server at Universidad Autónoma Metropolitana (UAM)-Cuajimalpa, which was funded for E.S-R. by CONACyT through the Apoyo para Proyectos de Investigación Científica, Desarrollo Tecnológico e Innovación en Salud ante la Contingencia por COVID-19, Grant Number 00312021. C.S. and E.S-R were supported by the Departamento de Ciencias Naturales (DCN), de la División de Ciencias Naturales e Ingeniería (DCNI), and de la UAM-Cuajimalpa through Divisional Project Numbers 47301025 and 47301026, respectively. C.S. and E.S-R were supported by the Red de Biotecnología y Bioingeniería for the payment of English proofreading and editing. The work of J.B. and M.T. was developed as part of the ASPIRE and SyMBoD projects and was funded by the German Federal Ministry of Education and Research (BMBF) under Grant Numbers 031L0287B and 01ZX1910D. M.T. was partly funded by the Leibniz Science Campus InterACt (from BWFGB Hamburg and the Leibniz Association). C.S., R.G.B., and E.S.-R. had equal participation as corresponding authors. J.H. is a master's student from "Posgrado en Ciencias Naturales e Ingeniería (PCNI)" at Universidad Autónoma Metropolitana, Unidad Cuajimalpa (UAM-C) and received a master's scholarship from CONAHCyT (CVU 1252067).

Acknowledgments

A.D.M.M. is a doctoral student from "Programa de maestría y doctorado en Ciencias Bioquímicas" at Universidad Nacional Autónoma de México (UNAM) and received a PhD fellowship funding from CONACyT (CVU 894530) as well as a Research Grant for Bi-nationally Supervised Doctoral Degrees/Cotutelle from Deutscher Akademischer Austauschdienst (DAAD, Personal Reference Number 91833882). J.H. is a master's student from "Posgrado en Ciencias Naturales e Ingeniería (PCNI)" at Universidad Autónoma Metropolitana, Unidad Cuajimalpa (UAM-C) and received a master's scholarship from CONAHCyT (CVU 1252067).

A former version of this manuscript is deposited on bioRxiv [108] and can be accessed via the following DOI: https://doi.org/10.1101/2023.09.07.556684.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Information. Supporting_Figures.pdf—Supporting figures 1 to 4.

Supporting Information. Supporting_Table 1.xlsx—The log2FC and other statistics associated with the DEGs in the conditions evaluated.

Supporting Information. Supporting_Table 2.xlsx—Results obtained for the enrichment analysis of the consensus DEGs in the blastema.

Supporting Information. Supporting_Table 3.xlsx—This table contains the module membership and gene–trait correlation obtained with WGCNA for the set of *A. mexicanum* samples evaluated.

Supporting Information. Supporting_Table 4.xlsx—This table contains the module membership and gene–trait correlation obtained with WGCNA for genes significantly correlated with the "aged" trait. An absolute Pearson correlation.

Supporting Information. Supporting_Table 5.xlsx—Subject sequences retrieved by Blastp for each of the "Regeneration-associated genes."

References

[1] C. Sámano, "Genomics and Epigenomics of Axolotl Regeneration," *International Journal of Developmental Biology* 65 (2021): 465–474.

[2] E. Vance, "Biology's Beloved Amphibian—the Axolotl—Is Racing towards Extinction," *Nature* 551 (2017): 286–289.

[3] T. Demircan, "A Histological Atlas of the Tissues and Organs of Neotenic and Metamorphosed Axolotl," *Acta Histochemica* 118 (2016): 746–759.

[4] J. R. Monaghan, "Gene Expression Patterns Specific to the Regenerating Limb of the Mexican Axolotl," *Biology Open* 1 (2012): 937–948.

[5] B. J. Haas and J. L. Whited, "Advances in Decoding Axolotl Limb Regeneration," *Trends in Genetics* 33 (2017): 553–565.

[6] A. Simon and E. M. Tanaka, "Limb Regeneration," *Wiley Interdisciplinary Reviews: Developmental Biology* 2 (2013): 291–300.

[7] L. McHedlishvili, "Reconstitution of the Central and Peripheral Nervous System During Salamander Tail Regeneration," *Proceedings of the National Academy of Sciences of the United States of America* 109 (2012): E2258–E2266.

[8] A. Tazaki, E. M. Tanaka, and J.-F. Fei, "Salamander Spinal Cord Regeneration: The Ultimate Positive Control in Vertebrate Spinal Cord Regeneration," *Developmental Biology* 432 (2017): 63–71.

[9] R. Suetsugu-Maki, "Lens Regeneration in Axolotl: New Evidence of Developmental Plasticity," *BMC Biology* 10 (2012): 103.

[10] C.-H. Wu, T.-Y. Huang, B.-S. Chen, L.-L. Chiou, and H.-S. Lee, "Long-Duration Muscle Dedifferentiation during Limb Regeneration in Axolotls," *PLoS One* 10 (2015): e0116068.

[11] A. Joven, A. Elewa, and A. Simon, "Model Systems for Regeneration: Salamanders," *Development* 146 (2019).

[12] D. Knapp, "Comparative Transcriptional Profiling of the Axolotl Limb Identifies a Tripartite Regeneration-Specific Gene Program," *PLoS One* 8 (2013): e61352.

[13] D. L. Stocum and J. A. Cameron, "Looking Proximally and Distally: 100 Years of Limb Regeneration and beyond," *Developmental Dynamics* 240 (2011): 943–968.
[14] S. V. Bryant, T. Endo, and D. M. Gardiner, "Vertebrate Limb Regeneration and the Origin of Limb Stem Cells," *International Journal of Developmental Biology* 46 (2004): 887–896.
[15] T. Sandoval-Guzmán, "Fundamental Differences in Dedifferentiation and Stem Cell Recruitment During Skeletal Muscle Regeneration in Two Salamander Species," *Cell Stem Cell* 14 (2014): 174–187.
[16] W. Masselink and E. M. Tanaka, "Toward Whole Tissue Imaging of Axolotl Regeneration," *Developmental Dynamics* 250 (2021): 800–806.
[17] T. Y. Lin, "Fibroblast Dedifferentiation as a Determinant of Successful Regeneration," *Developmental Cell* 56 (2021): 1541–1551.e6.
[18] C. McCusker and D. M. Gardiner, "The Axolotl Model for Regeneration and Aging Research: A Mini-Review," *Gerontology* 57 (2011): 565–571.
[19] C. McCusker, S. V. Bryant, and D. M. Gardiner, "The Axolotl Limb Blastema: Cellular and Molecular Mechanisms Driving Blastema Formation and Limb Regeneration in Tetrapods," *Regeneration (Oxf)* 2 (2015): 54–71.
[20] M. Kragl, "Cells Keep a Memory of Their Tissue Origin During Axolotl Limb Regeneration," *Nature* 460 (2009): 60–65.
[21] A. Makanae, K. Mitogawa, and A. Satoh, "Co-operative Bmp- and Fgf-Signaling Inputs Convert Skin Wound Healing to Limb Formation in Urodele Amphibians," *Developmental Biology* 396 (2014): 57–66.
[22] M. C. Keinath, "Initial Characterization of the Large Genome of the Salamander *Ambystoma mexicanum* Using Shotgun and Laser Capture Chromosome Sequencing," *Scientific Reports* 5 (2015): 16413.
[23] D. M. Bryant, "A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors," *Cell Reports* 18 (2017): 762–776.
[24] J. Caballero-Pérez, "Transcriptional Landscapes of Axolotl (*Ambystoma mexicanum*)," *Developmental Biology* 433 (2018): 227–239.
[25] S. Darnet, "Deep Evolutionary Origin of Limb and Fin Regeneration," *Proceedings of the National Academy of Sciences of the United States of America* 116 (2019): 15106–15115.
[26] S. Schloissnig, "The Giant Axolotl Genome Uncovers the Evolution, Scaling, and Transcriptional Control of Complex Gene Loci," *Proceedings of the National Academy of Sciences of the United States of America* 118 (2021).
[27] K. Sousounis, A. T. Athippozhy, S. R. Voss, and P. A. Tsonis, "Plasticity for Axolotl Lens Regeneration is Associated With Age-Related Changes in Gene Expression," *Regeneration (Oxf.)* 1 (2014): 47–57.
[28] W. A. Vieira, K. M. Wells, and C. D. McCusker, "Advancements to the Axolotl Model for Regeneration and Aging," *Gerontology* 66 (2020): 212–222.
[29] J. E. Farkas and J. R. Monaghan, "Housing and Maintenance of *Ambystoma mexicanum*, the Mexican Axolotl," *Methods in Molecular Biology* 1290 (2015): 27–46.
[30] A. Polikarpova, "The Specialist in Regeneration-The Axolotl-A Suitable Model to Study Bone Healing?" *NPJ Regen Med* 7 (2022): 35.
[31] S. Peña-Llopis and J. Brugarolas, "Simultaneous Isolation of High-Quality DNA, RNA, miRNA and Proteins From Tissues for Genomic Applications," *Nature Protocols* 8 (2013): 2240–2255.
[32] S. Andrews, "FastQC: A Quality Control Tool for High Throughput Sequence Data," (2010), http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
[33] A. Dobin, "STAR: Ultrafast Universal RNA-Seq Aligner," *Bioinformatics* 29 (2013): 15–21.
[34] M. I. Love, W. Huber, and S. Anders, "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With DESeq2," *Genome Biology* 15 (2014): 550.
[35] D. Szklarczyk, "The STRING Database in 2023: Protein-Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest," *Nucleic Acids Research* 51 (2023): D638–D646.
[36] L. Kolberg, U. Raudvere, I. Kuzmin, and J. Vilo, "gprofiler2--an R Package for Gene List Functional Enrichment Analysis and Namespace Conversion Toolset G: Profiler," *F1000Research* 9 (2020).
[37] S. Sayols, "Rrvgo: A Bioconductor Package for Interpreting Lists of Gene Ontology Terms," *microPublication Biology* 2023 (2023).
[38] K. Mechteridis, M. Lauber, J. Baumbach, and M. List, "KeyPathwayMineR: De Novo Pathway Enrichment in the R Ecosystem," *Frontiers in Genetics* 12 (2022).
[39] G. Bindea, "ClueGO: A Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks," *Bioinformatics* 25 (2009): 1091–1093.
[40] P. Langfelder and S. Horvath, "WGCNA: An R Package for Weighted Correlation Network Analysis," *BMC Bioinformatics* 9 (2008): 1–13.
[41] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods," *Biostatistics* 8 (2007): 118–127.
[42] S. F. Altschul, "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," *Nucleic Acids Research* 25 (1997): 3389–3402.
[43] A. A. Schäffer, "Improving the Accuracy of PSI-BLAST Protein Database Searches With Composition-Based Statistics and Other Refinements," *Nucleic Acids Research* 29 (2001): 2994–3005.
[44] UniProt Consortium, "UniProt: The Universal Protein Knowledgebase in 2023," *Nucleic Acids Research* 51 (2023): D523–D531.
[45] S. B. Hedges, "The Origin and Evolution of Model Organisms," *Nature Reviews Genetics* 3 (2002): 838–849.
[46] J. Jumper, "Highly Accurate Protein Structure Prediction With AlphaFold," *Nature* 596 (2021): 583–589.
[47] B. Yariv, "Using Evolutionary Data to Make Sense of Macromolecules With a "Face-Lifted" ConSurf," *Protein Science* 32 (2023): e4582.
[48] A. Marchler-Bauer, "CDD: NCBI's Conserved Domain Database," *Nucleic Acids Research* 43 (2015): D222–D226.
[49] A. Marchler-Bauer and S. H. Bryant, "CD-Search: Protein Domain Annotations on the Fly," *Nucleic Acids Research* 32 (2004): W327–W331.

[50] Y. Peng, "Structural Basis of Substrate Recognition in Human Nicotinamide *N*-Methyltransferase," *Biochemistry* 50 (2011): 7800–7808.
[51] K. L. Kavanagh, "PDB: 2P31," *Crystal Structure of Human Glutathione Peroxidase* 7 (2007).
[52] B. L. Tang, "ADAMTS: A Novel Family of Extracellular Matrix Proteases," *International Journal of Biochemistry and Cell Biology* 33 (2001): 33–44.
[53] S. Porter, I. M. Clark, L. Kevorkian, and D. R. Edwards, "The ADAMTS Metalloproteinases," *Biochemical Journal* 386 (2005): 15–27.
[54] G. C. Jones and G. P. Riley, "ADAMTS Proteinases: A Multi-Domain, Multi-Functional Family With Roles in Extracellular Matrix Turnover and Arthritis," *Arthritis Research and Therapy* 7 (2005): 160–169.
[55] L. Mosyak, "Crystal Structures of the Two Major Aggrecan Degrading Enzymes, ADAMTS4 and ADAMTS5," *Protein Science* 17 (2008): 16–21.
[56] X. Li, L. Li, Y. Chang, W. Ning, and X. Liu, "Structural and Functional Study of FK Domain of Fstl1," *Protein Science* 28 (2019): 1819–1829.
[57] L. Leclère, "Dynamic Evolution of the Cthrc1 Genes, a Newly Defined Collagen-Like Family," *Genome Biology and Evolution* 12 (2020): 3957–3970.
[58] A. Bölük, M. Yavuz, and T. Demircan, "Axolotl: A Resourceful Vertebrate Model for Regeneration and Beyond," *Developmental Dynamics* 251 (2022): 1914–1933.
[59] A. Samos, V. McGaughey, S. Rieger, and T. S. Lisse, "Reawakening GDNF's Regenerative Past in Mice and Humans," *Regenerative Therapy* 20 (2022): 78–85.
[60] S. R. Voss, M. R. Woodcock, and L. Zambrano, "A Tale of Two Axolotls," *BioScience* 65 (2015): 1134–1140.
[61] C. Reiß, "Cut and Paste: The Mexican Axolotl, Experimental Practices and the Long History of Regeneration Research in Amphibians, 1864-Present," *Frontiers in Cell and Developmental Biology* 10 (2022): 786533.
[62] S. Min and J. L. Whited, "Limb Blastema Formation: How Much Do We Know at a Genetic and Epigenetic Level?" *Journal of Biological Chemistry* 299 (2023): 102858.
[63] T. Gerber, "Single-cell Analysis Uncovers Convergence of Cell Identities During Axolotl Limb Regeneration," *Science* 362 (2018).
[64] P. Khan, B. Linkhart, and H.-G. Simon, "Different Regulation of T-Box Genes Tbx4 and Tbx5 During Limb Development and Limb Regeneration," *Developmental Biology* 250 (2002): 383–392.
[65] J.-C. Guimond, "BMP-2 Functions Independently of SHH Signaling and Triggers Cell Condensation and Apoptosis in Regenerating Axolotl Limbs," *BMC Developmental Biology* 10 (2010): 15.
[66] J. R. Erickson, "A Novel Role for SALL4 During Scar-Free Wound Healing in Axolotl," *NPJ Regenerative Medicine* 1 (2016): 16016.
[67] S. Suthon, R. S. Perkins, V. Bryja, G. A. Miranda-Carboni, and S. A. Krum, "WNT5B in Physiology and Disease," *Frontiers in Cell and Developmental Biology* 0 (2021).
[68] M. Mastelaro de Rezende, G. Zenker Justo, E. Julian Paredes-Gamero, and R. Gosens, "Wnt-5A/B Signaling in Hematopoiesis Throughout Life," *Cells* 9 (2020).
[69] F. Xuan, "Wnt/β-Catenin Signaling Contributes to Articular Cartilage Homeostasis Through Lubricin Induction in the Superficial Zone," *Arthritis Research and Therapy* 21 (2019): 247.
[70] C. Aguilar and D. M. Gardiner, "DNA Methylation Dynamics Regulate the Formation of a Regenerative Wound Epithelium During Axolotl Limb Regeneration," *PLoS One* 10 (2015): e0134791.
[71] S. R. Voss, "HDAC Inhibitor Titration of Transcription and Axolotl Tail Regeneration," *Frontiers in Cell and Developmental Biology* 9 (2021): 767377.
[72] S. J. Whitcomb, A. Basu, C. D. Allis, and E. Bernstein, "Polycomb Group Proteins: an Evolutionary Perspective," *Trends in Genetics* 23 (2007): 494–502.
[73] D. L. Stocum, "Mechanisms of Urodele Limb Regeneration," *Regeneration (Oxf)* 4 (2017): 159–200.
[74] N. Santosh, "Matrix Metalloproteinase Expression During Blastema Formation in Regeneration-Competent versus Regeneration-Deficient Amphibian Limbs," *Developmental Dynamics* 240 (2011): 1127–1141.
[75] A. M. Ritenour and R. Dickie, "Inhibition of Vascular Endothelial Growth Factor Receptor Decreases Regenerative Angiogenesis in Axolotls," *The Anatomical Record* 300 (2017): 2273–2280.
[76] J. L. Whited, J. A. Lehoczky, C. A. Austin, and C. J. Tabin, "Dynamic Expression of Two Thrombospondins During Axolotl Limb Regeneration," *Developmental Dynamics* 240 (2011): 1249–1258.
[77] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, "Hallmarks of Aging: An Expanding Universe," *Cell* 186 (2023): 243–278.
[78] M. Selman and A. Pardo, "Fibroageing: An Ageing Pathological Feature Driven by Dysregulated Extracellular Matrix-Cell Mechanobiology," *Ageing Research Reviews* 70 (2021): 101393.
[79] J. Du, "Extracellular Matrix Stiffness Dictates Wnt Expression through Integrin Pathway," *Scientific Reports* 6 (2016): 20395.
[80] Q. Zhou, "Stiffness of Nanoparticulate Mineralized Collagen Scaffolds Triggers Osteogenesis via Mechanotransduction and Canonical Wnt Signaling," *Macromolecular Bioscience* 21 (2021): e2000370.
[81] E. A. Klein, "Cell-Cycle Control by Physiological Matrix Elasticity and In Vivo Tissue Stiffening," *Current Biology* 19 (2009): 1511–1518.
[82] A. Kawaguchi, "A Chromatin Code for Limb Segment Identity in Axolotl Limb Regeneration," *Developmental Cell* 59 (2024): 2239–2253.e9.
[83] T. Oichi, "Adamts17 Is Involved in Skeletogenesis Through Modulation of BMP-Smad1/5/8 Pathway," *Cellular and Molecular Life Sciences* 76 (2019): 4795–4809.
[84] X. Yu, "Weill-Marchesani Syndrome 4 Caused by Compound Heterozygosity of a Maternal Submicroscopic Deletion and a Paternal Nonsense Variant in the Gene: A Case Report," *American Journal of Ophthalmology Case Reports* 26 (2022): 101541.
[85] P. Marzin, V. Cormier-Daire, and E. Tsilou, "Weill-Marchesani Syndrome," in *GeneReviews*, ed. M. P. Adam (Seattle, WA: University of Washington, Seattle, 2007).
[86] E. C. Jeanes, J. A. C. Oliver, S. L. Ricketts, D. J. Gould, and C. S. Mellersh, "Glaucoma-causing Mutations Are Also Reproducibly Associated with Height in Two Domestic Dog Breeds: Selection for Short Stature May Have Contributed to Increased Prevalence of Glaucoma," *Canine Genetics and Epidemiology* 6 (2019): 5.

[87] H. A. van Duyvenvoorde, "Copy Number Variants in Patients With Short Stature," *European Journal of Human Genetics* 22 (2014): 602–609.
[88] S. Z. Karoulias, "A Novel ADAMTS17 Variant that Causes Weill-Marchesani Syndrome 4 Alters Fibrillin-1 and Collagen Type I Deposition in the Extracellular Matrix," *Matrix Biology* 88 (2020): 1–18.
[89] H. O. Hambrock, "Structural Characterization of TSC-36/Flik," *Journal of Biological Chemistry* 279 (2004): 11727–11735.
[90] M. Horak, D. Fairweather, P. Kokkonen, D. Bednar, and J. Bienertova-Vasku, "Follistatin-like 1 and Its Paralogs in Heart Development and Cardiovascular Disease," *Heart Failure Reviews* 27 (2022): 2251–2265.
[91] M. Sylva, A. F. M. Moorman, and M. J. B. van den Hoff, "Follistatin-like 1 in Vertebrate Development," *Birth Defects Research Part C: Embryo Today-Reviews* 99 (2013): 61–69.
[92] Y. Geng, "Follistatin-Like 1 (Fstl1) is a Bone Morphogenetic Protein (BMP) 4 Signaling Antagonist in Controlling Mouse Lung Development," *Proceedings of the National Academy of Sciences of the United States of America* 108 (2011): 7058–7063.
[93] Y. Yang, "The Expression Pattern of Follistatin-like 1 in Mouse Central Nervous System Development," *Gene Expression Patterns* 9 (2009): 532–540.
[94] H. Jiang, "Angiocrine FSTL1 (Follistatin-Like Protein 1) Insufficiency Leads to Atrial and Venous Wall Fibrosis via SMAD3 Activation," *Arteriosclerosis, Thrombosis, and Vascular Biology* 40 (2020): 958–972.
[95] N. Ouchi, "Follistatin-Like 1, a Secreted Muscle Protein, Promotes Endothelial Cell Function and Revascularization in Ischemic Tissue Through a Nitric-Oxide Synthase-Dependent Mechanism," *Journal of Biological Chemistry* 283 (2008): 32802–32811.
[96] J.-M. Kim, C. Lin, Z. Stavre, M. B. Greenblatt, and J.-H. Shim, "Osteoblast-Osteoclast Communication and Bone Homeostasis," *Cells* 9 (2020).
[97] C. Riquelme-Guzmán, "Osteoclast-mediated Resorption Primes the Skeleton for Successful Integration during Axolotl Limb Regeneration," *Elife* 11 (2022).
[98] A. Myngbay, L. Manarbek, S. Ludbrook, and J. Kunz, "The Role of Collagen Triple Helix Repeat-Containing 1 Protein (CTHRC1) in Rheumatoid Arthritis," *International Journal of Molecular Sciences* 22 (2021).
[99] Y.-R. Jin, "Inhibition of Osteoclast Differentiation and Collagen Antibody-Induced Arthritis by CTHRC1," *Bone* 97 (2017): 153–167.
[100] S. Takeshita, "Osteoclast-Secreted CTHRC1 in the Coupling of Bone Resorption to Formation," *Journal of Clinical Investigation* 123 (2013): 3914–3924.
[101] D. Peng, "Glutathione Peroxidase 7 Protects Against Oxidative DNA Damage in Oesophageal Cells," *Gut* 61 (2012): 1250–1260.
[102] V. Bosello-Travain, "Protein Disulfide Isomerase and Glutathione are Alternative Substrates in the One Cys Catalytic Cycle of Glutathione Peroxidase 7," *Biochimica et Biophysica Acta* 1830 (2013): 3846–3857.
[103] Y.-C. Chang, "Deficiency of NPGPx, an Oxidative Stress Sensor, Leads to Obesity in Mice and Human," *EMBO Molecular Medicine* 5 (2013): 1165–1179.
[104] N. W. Al Haj Baddar, A. Chithrala, and S. R. Voss, "Amputation-Induced Reactive Oxygen Species Signaling is Required for Axolotl Tail Regeneration," *Developmental Dynamics* 248 (2019): 189–196.
[105] A.-W. Struck, M. L. Thompson, L. S. Wong, and J. Micklefield, "S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications," *ChemBioChem* 13 (2012): 2642–2655.
[106] H. Sperber, "The Metabolome Regulates the Epigenetic Landscape During Naive-To-Primed Human Embryonic Stem Cell Transition," *Nature Cell Biology* 17 (2015): 1523–1535.
[107] M. A. Eckert, "Proteomics Reveals NNMT as a Master Metabolic Regulator of Cancer-Associated Fibroblasts," *Nature* 569 (2019): 723–728.
[108] A. del Moral-Morales, "Key Proteins for Regeneration in *A. mexicanum*: Transcriptomic Insights From Aged and Juvenile Limbs," *bioRxiv* (2023): 556684, https://doi.org/10.1101/2023.09.07.556684.