



UNIVERSITAT DE  
BARCELONA

## Characterization of the molecular mechanisms used by JmjC-lysine demethylases to regulate gene expression

Marta Vicioso Mantis

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

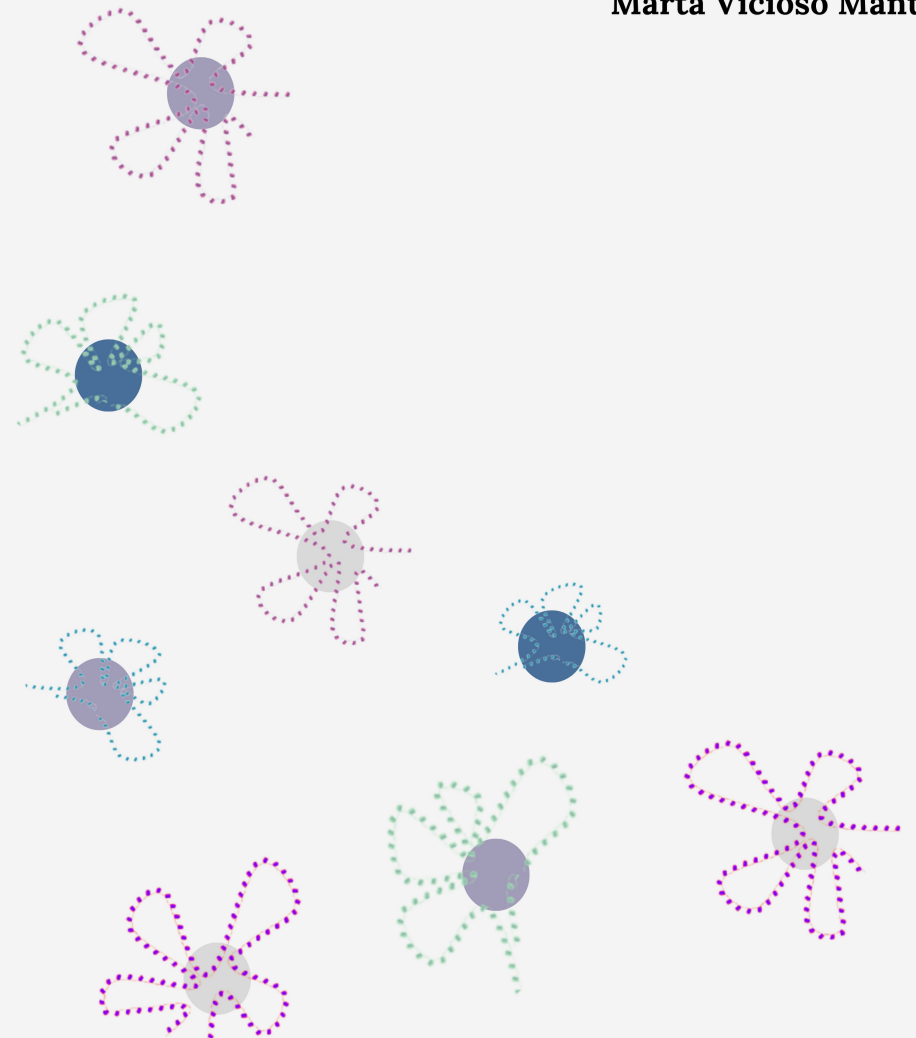
**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



2022

# CHARACTERIZATION OF THE MOLECULAR MECHANISMS USED BY JMJC-LYSINE DEMETHYLASES TO REGULATE GENE EXPRESSION

Doctoral Thesis  
Marta Vicioso Mantis



Doctoral program in Genetics  
Faculty of Biology, University of Barcelona



Research performed at the  
Institute of Molecular Biology of Barcelona, CSIC

***Characterization of the molecular  
mechanisms used by JmjC-lysine  
demethylases to regulate gene  
expression***

*A thesis submitted by Marta Vicioso Mantis to obtain the  
doctoral degree by the University of Barcelona*

**Author:**  
Marta Vicioso  
Mantis

**Director:**  
Dr María A.  
Martínez Balbás

**Tutor:**  
Dr. Pedro  
Martínez Serra

Barcelona 2022



## Acknowledgements

Estos inolvidables años de duro trabajo que han marcado mi vida para siempre no tienen cabida en mi mente sin todas esas personas que me han acompañado a lo largo del camino. Sin duda alguna, lo mínimo que merecen es una mención especial para agradecerles todo lo que, conscientes o no de ello, me han aportado durante todo este tiempo.

Debo empezar dando las gracias a Marian, la mejor directora de tesis que se puede imaginar. Con su puerta del despacho siempre abierta y su taza de café lista para discutir sobre ciencia, animarme cuando parece que está todo perdido y hacerme ver lo que a veces se olvida, para arreglar el mundo e incluso contar confidencias. Gracias por enseñarme tanto, no concibo un mejor lugar para hacer la tesis que tu laboratorio. Aprovecho para continuar agradeciendo a mis compis de labo, tanto los presentes como los pasados. Empezando por los que ya marcharon, pero igualmente me acompañaron en este proceso, una mención muy especial a Raquel, mi mentora en ciencia y en nuestra JMJD3. Te debo prácticamente todo lo que sé del laboratorio, gracias por tu dedicación y paciencia, eres una científica admirable. Gracias a Stella, por todos los momentos compartidos en el labo y fuera de él. Se me llena la boca pudiendo decir que estuve en tu boda, que cada vez que vienes a Barcelona nos vemos, y que nuestra amistad continúa a pesar de la distancia hasta Atenas. En esta línea también agradecer a Simo; sin duda eres la persona que más me ha apoyado dentro del labo, compartiendo tantos momentos las dos solas. Gracias por cada risa, cada palabra de ánimo, cada abrazo, cada comida al sol y cada café con helado. Y, sobre todo, gracias por seguir compartiendo tantas cenas, vermouths y confidencias. Eres una amiga inmejorable. Por su puesto, debo agradecer a mis compis actuales: a Samuel, por esa vitalidad y alegría con la que anima a cualquiera, por todos los detalles que tienes conmigo, por haberme sostenido en mis momentos más duros en el labo, y por todas las aventuras que vivimos tanto dentro como fuera de él (¡Rosalía nos espera!). A Marta, desde que has llegado al labo te has convertido en una amiga increíble, siempre tan dulce e inocente. Haces que todo sea mucho más llevadero. A Mónica, la máster más divertida que ha pasado por el labo, siempre dispuesta a hacer lo que haga falta, tanto a nivel de experimentos como social. En general, agradecer a todas las chicas con las que he coincidido estos años: Claudia, Elia, María, gracias por tener siempre una sonrisa en el momento adecuado. Por supuesto, también agradecer a

todos los demás compis que no son del labo de Marian, empezando por las chicas de Albert. Mónica, Nuria y Jaione, gracias por cada comida, cada charla, cada after, cada lab meeting. Sois increíbles. Gracias a Dani, que, aunque esté lejos y nos veamos poco, siempre está presente, desde que ambos llegamos al Hélix como jóvenes estudiantes. Gracias a Carles, Albert, Guille, Alicia e Inma por las discusiones científicas y las risas en los lab meeting. Gracias a nuestros vecinos, los labos de Elisa, de Murielle y de Mariona. Gracias especial a Jose, por todas las conversaciones, por aportar calma, por las comidas, los cafés y cervezas, y por nuestro amor a la navidad. Gracias al ser de luz que es Mar, a Pilar y Paula por contagiarme su vitalidad, a Alex por estar siempre ahí para todo. Otro gracias especial a Mariajo, por tantísimo apoyo, por su inagotable optimismo y su maravillosa risa, todos sabemos que nuestros labos no serían lo mismo sin ti. Y gracias a los que ya no están pero que tanto aportaron: Elena, Lucía, Susana, Isa. Acabando con los amigos de la segunda planta del Hélix, gracias a María y a Laura, ambas sois unas maravillosas compañeras de risas. No menos importante son los agradecimientos a los amigos de la tercera planta, especialmente a Marcos y Mónica, con los que tanto he compartido a lo largo de estos años (e incluso desde antes de empezar este camino), gorroneos y frutos secos incluidos. Por supuesto, también mencionar a Alexis, Blanca y Silvia. Todos habéis aportado brillo a mis años de doctorado. Para acabar con el IBMB, no puedo dejar de agradecer a Elena Rebollo por estar siempre ahí y por dedicarme tantas de tus valiosas horas. Gran parte de este trabajo no habría sido posible sin ti.

Toca ahora mencionar a mis amigos del alma, tanto los andaluces como los que están por aquí. No puedo empezar por otra que por mi Rubi, mi hermana elegida. Contigo sobran las palabras, pero debo recordar que este camino empezó hace 10 años a tu lado, cuando tuvimos que enfrentarnos a la vida real solas por primera vez. Gracias por crecer juntas, gracias por conocerme mejor que yo misma, gracias por ser una prolongación de mí, gracias por todo. Continuar con mi Yola, mi confidente, mi seriéfila preferida, mi compi de aventuras y de llantos. Gracias por aparecer en mi vida y no marcharte. Gracias a Patri, por sacar siempre un ratito para vernos cuando voy a casa, por todos esos momentos inolvidables compartidos desde los 15 años. A Ana, con la que llevo compartida una vida entera. Está claro que nuestros caminos están enlazados, gracias por darme tanto. Gracias a Bei, cómo puede ser que una persona tan pequeñita llene tanto. Gracias por convertirme en mi inseparable. Una mención muy especial a mi otro hermano elegido, mi Pepito,

que tanto me entiende y conoce, con el que parece que voy a la par en la vida; ojalá sea una señal de que siempre vamos a estar juntos. Por supuesto, agradecer también a Pablo, por acogerme tan cariñosamente desde el primer instante. Continúo con todos los amigos que, de una forma u otra, me ha traído la vida y que se han quedado en ella: a los compis de viajes Jaime y Rafa, tantísimos años de amistad hablan por sí solos; a Ramón, por haberse convertido en mi día a día, y por ser de las voces que más he escuchado en los últimos años a pesar de la distancia; a Paloma, por compartir cada etapa de la vida desde la uni y progresar juntas; a Gore, porque lo que l'Ovella Negra une, no se separa; a los que me trajó Sevilla: Inés, Luz, Diana, Isa, Pedro, Rocío, Manolo, los compis del labo de Sebastián. A mis ex compis de piso, especialmente a Leo, Musa y Alberto. A los amigos que me ha traído Barcelona, Joan, Aleix, Pau, Luisa, Borges, y especialmente a Edu, por interesarse siempre por mi trabajo y mis avances. Gracias a los vecis Raquel, Nacho, Julian y Nico, porque la pandemia tuvo que traer algo bueno a nuestras vidas. Por último, un gracias muy especial a mi querida Pauli, la que se ha convertido en mi inseparable en Barcelona. Por todos esos momentos de desahogo, de cotilleos, de risas, esas Moritz y esos brunchs. Nos queda mucho por vivir aún.

Ni que decir tiene que si he llegado hasta este día ha sido gracias a mi familia. Ellos son el motor de mi vida, a los que les debo todo. Gracias, mamá y papá, por haberme inculcado los valores que me definen, por hacerme ser quien soy, por haberme dado tanto sin pedirme jamás nada a cambio. De verdad que no podría imaginar unos padres mejores. Esta tesis es para vosotros, como todo lo que he hecho y todo lo que haga en mi vida. Gracias a mi peque, mi fiel compañera de vida que sé que siempre va a estar a mi lado, aunque paradójicamente haga tanto que no compartimos espacio vital. Contigo me sobran las palabras, eres hermana, mejor amiga, y mi mayor debilidad a un mismo tiempo. Un gracias muy muy especial a mis segundos padres, mis abuelos. Gracias a mis abuelas Mercedes y Juani, por estar orgullosas de mi sin ni siquiera saber muy bien lo que hago. Gracias, abuelo Manolo, porque aunque ya no estés, siempre me transmitiste tu apoyo, y tú sí que entendías un poquito más lo que hago. Este trabajo también va dedicado a ti. Muchas gracias a todos mis tíos, especialmente a mis padrinos, Esther y Kiko, pero también a Cinta y Sandra, a Carlos, a Mamé, a Jose, Migue, Marian y Rebeca. A mis primos, tanto los peques Carlitos, Chedes y Álvaro, como los mayores, sobre todo a María, mi predecesora en todo, incluido el ser doctora.

Sabes que a ti también te debo el ser como soy hoy, por enseñarme tanto cuando las vacaciones significaban ver de nuevo a la que era casi una hermana más. Tampoco puede faltar un gracias a David, por todo su apoyo al ser el miembro de la familia más cercano a mi mundo, y a Pablo, porque, aunque ahora casi no nos veamos, de pequeños fuimos inseparables, y eso no se olvida. Por último, agradecer a mi familia política. Gracias a Pilar y Pedro, por acogerme desde el primer momento como a una hija. También gracias a Pedro y Tere, por ser los mejores cuñados posibles, y al pequeñín de la casa, Martinxito, por darme tantísimas alegrías sin siquiera saberlo. A Isabel, Manolo, Celia y Enrique, gracias por hacerme una más de la familia.

Para finalizar, debo agradecer todo mi presente a mi compañero vital, a la persona que me aguanta cada día desde que me levanto hasta que me acuesto, al que se come todos mis quebraderos de cabeza, escucha todos mis dramas y me da el empujón que necesito para poder seguir avanzando. Porque nadie sabe tan bien lo que han supuesto estos años para mí, de principio a final, momentos buenos y malos incluidos, y nadie ha estado tan presente en cada uno de ellos como él. Gracias a ti, Javi, por tener siempre una mano tendida. Por ser mi salvavidas, mi hogar, mi refugio, y mi mayor apoyo. Esto también es para ti.



## Abstract

En mamíferos, el neurodesarrollo está dirigido por la cooperación entre señales de desarrollo, factores de transcripción y reguladores epigénéticos. Su coordinación activa elementos reguladores que determinan los programas de expresión génica específicos de cada tipo celular. Los *enhancers* son regiones reguladoras del genoma que activan espacio/temporalmente a sus genes diana en respuesta a señales del desarrollo. Por tanto, constituyen la mayor unidad reguladora durante el desarrollo. Está claramente establecido que la cromatina se reorganiza para formar contactos entre *enhancers*, originando *clusters de enhancers* que activan promotores, dirigiendo así la expresión génica. Lo que actualmente se debate es cómo se establecen estas agrupaciones regulatorias.

Para investigar esta cuestión, nuestro trabajo se centró en examinar la regulación de la expresión génica dirigida por la vía de señalización de TGF $\beta$  y su proteína cooperadora JMJD3, cuyo papel regulador en *enhancers* ha sido demostrado en células madre neurales. Por tanto, elegimos como locus de estudio el de un gen cuya expresión depende del eje JMJD3-TGF $\beta$ , el gen *Chst8*. En concreto, estudiamos la formación de un *cluster de enhancers* en respuesta a TGF $\beta$ , y de forma dependiente de JMJD3, que regula la transcripción de dicho locus. JMJD3 contiene una región intrínsecamente desordenada especialmente larga, dominio que se ha demostrado implicado en procesos de separación de fases. Así pues, en esta tesis hemos demostrado que JMJD3 participa en dichos procesos de forma dependiente de su región desordenada. Además, mostramos la existencia de una correlación entre esta capacidad de JMJD3 y su activación transcripcional del locus de *Chst8*.

Además de JMJD3, otras desmetilasas con dominios JmjC (JMJC-KDMs) están implicadas en la regulación de procesos transcripcionales. Considerando nuestros hallazgos para JMJD3 y los recientemente

descritos condensados transcripcionales, decidimos investigar la universalidad del mecanismo de separación de fases para las JMJC-KDMs. Así, descubrimos que estas proteínas están enriquecidas en regiones desordenadas, y que ejemplos representativos de distintas familias (KDM2A, KDM4B y PHF2) forman condensados que correlacionan con sus características transcripcionales. Por tanto, nuestros resultados sugieren que la separación de fases constituye otro mecanismo, además del catalítico, usado comúnmente por las JMJC-KDMs para promover sus funciones biológicas.

# List of contents

<b>List of abbreviations</b>	1
<b>Introduction</b>	5
<b>1. Chromatin</b>	5
1.1. Epigenetic modifications of chromatin	6
1.1.1. DNA methylation	6
1.1.2. Histones post-translational modifications	7
1.1.2.1. Histone methylation	9
1.1.2.2. Histone demethylation	11
1.1.2.2.1. JmJC containing KDMs: KDM6 subfamily and JMJD3	12
1.2. Regulatory regions on DNA	16
1.2.1. Gene promoters	17
1.2.2. Enhancers	17
1.2.2.1. Superenhancers	20
1.2.3. Silencers and insulators	21
1.3. Three-dimensional structure of chromatin	22
1.3.1. Chromosome territories	24
1.3.2. Chromosomal compartments	24
1.3.3. Topologically associating domains	25
1.3.4. Loop domains	26
1.4. Regulatory proteins of chromatin	27
1.4.1. Transcription factors	27
1.4.2. Transcriptional coregulators	28
1.4.3. RNA-polymerase II	29
<b>2. The cortex development</b>	30
2.1. Mouse corticogenesis and its regulating molecular pathways	30
2.2. Neural stem cells as an <i>in vitro</i> model of study	32
<b>3. The TGF<math>\beta</math> pathway</b>	33
3.1. The signaling cascade of TGF $\beta$	33
3.2. Functions of the TGF $\beta$ pathway	35

3.3. JMJD3 and the TGF $\beta$ pathway in corticogenesis	36
<b>4. Biomolecular condensates</b>	<b>38</b>
4.1. Formation of biomolecular condensates	39
4.1.1. Liquid-liquid phase separation	40
4.1.1.1. Intrinsically Disordered Regions	42
4.1.2. Liquid-gel phase separation	43
4.1.3. Polymer-polymer phase separation	43
4.2. Physical properties of biomolecular condensates	44
4.3. Composition of biomolecular condensates	45
4.4. Regulation of biomolecular condensates formation through phase separation	45
4.5. Physiological roles of biomolecular condensates	46
4.5.1. Nuclear biomolecular condensates	46
4.5.1.1. Biomolecular condensates and chromatin structure	48
4.5.1.2. Biomolecular condensates and transcriptional regulation	50
4.5.1.2.1. Transcriptional condensates at typical enhancers and super-enhancers	51
4.5.1.3. Biomolecular condensates and post-transcriptional regulation	53
4.5.2. Other biomolecular condensates (cytoplasm, mitochondria and chloroplasts)	53
4.6. Dysregulation of biomolecular condensates	53
<b>Objectives</b>	<b>55</b>
<b>Materials and methods</b>	<b>57</b>
<b>1. Materials</b>	<b>57</b>
1.1. Plasmids	57
1.2. Antibodies	58
1.3. Primers	59
<b>2. Experimental methods</b>	<b>62</b>
2.1. Cell culture	62
2.1.1. Cell culture growth and maintenance	62

2.1.1.1. Mouse neural stem cells (NSCs)	62
2.1.1.2. HEK293T and NIH3T3	63
2.1.2. Cells freezing and thawing	64
2.1.3. Genetic manipulation of growing cells	65
2.1.3.1. Calcium phosphate transfection	65
2.1.3.2. Lentiviral transduction	65
2.1.4. Cell treatments	66
2.1.4.1. Doxycycline cell induction	66
2.1.4.2. Transforming Growth Factor $\beta$ (TGF $\beta$ ) treatment	66
2.1.4.3. GSK-J4 HCl inhibitor	66
2.1.4.4. 1,6-Hexanediol treatment for live imaging cells	67
2.2. Molecular biology procedures	67
2.2.1. Nucleic acids-related	67
2.2.1.1. Genomic DNA extraction	67
2.2.1.2. Phenol chloroform extraction and ethanol precipitation	68
2.2.1.3. RNA extraction and DNase treatment	68
2.2.1.4. Retrotranscription of RNA into cDNA	69
2.2.1.5. qPCR	69
2.2.1.6. PCR	70
2.2.1.7. 4C-seq assay	70
2.2.1.8. Cloning of plasmidic DNA	72
2.2.1.8.1. Luciferase constructs	72
2.2.1.8.2. Overexpression constructs	74
2.2.1.9. Amplification of plasmidic DNA: mini and maxi preparations of DNA	75
2.2.1.10. Electrophoresis in agarose gels	76
2.2.1.11. Luciferase reporter assay	76
2.2.2. Protein-related	77
2.2.2.1. Total protein extraction and quantification	77
2.2.2.2. SDS-Page electrophoresis	78
2.2.2.3. Western Blot	78
2.2.2.4. Chromatin immunoprecipitation (ChIP)	79
2.2.2.5. Droplet assays in nuclear extracts	80

2.2.2.6. Indirect immunofluorescence	81
2.2.2.7. Immunofluorescence with DNA FISH	82
2.2.2.8. Live imaging	83
2.2.2.9. Fluorescence recovery after photobleaching (FRAP) in live cells	83
2.2.2.10. Quantification of droplets liquid-like features	84
2.2.2.11. Focus calling (immunofluorescence, 1,6-Hexanediol live-imaging)	85
<b>3. Bioinformatic methods</b>	85
3.1. 4C-seq computational analysis	85
3.2. Gene expression omnibus (GEO) accessions	85
3.3. Capture obtaining	86
3.4. Protein sequence analysis and predictions	86
3.5. Amino acid composition of conserved IDRs within KDM families	87
<b>4. Statistical analysis</b>	87
4.1. Sample size	87
4.2. Graphics and measures of dispersion (mean and standard error of the mean)	88
4.3. Student's t-test	88
4.4. Wilcoxon-Mann-Whitney test	88
<b>Results</b>	89
<b>1. Characterization of JMJD3-mediated transcriptional regulation at enhancers during early neurogenesis</b>	89
1.1. TGF $\beta$ promotes chromatin reorganization around the <i>Chst8</i> locus	89
1.2. TGF $\beta$ induces the formation of the <i>Chst8</i> enhancer cluster	95
1.3. TGF $\beta$ -promoted <i>Chst8</i> enhancer cluster depends on JMJD3	98
1.4. TGF $\beta$ and JMJD3 drive chromatin reorganization	102
1.5. JMJD3's role mediating contacts seems to be independent of its demethylase activity	109
1.6. JMJD3 is a highly disordered protein with potential to phase separate	111

1.7. JMJD3 undergoes liquid-liquid phase separation <i>in vitro</i> and <i>in vivo</i>	116
1.8. JMJD3 IDR is necessary for condensate formation	125
1.9. JMJD3 demethylase activity is not necessary for condensate formation	128
1.10. JMJD3 promotes <i>Chst8</i> enhancer cluster assembly and gene transcription	132
<b>2. Analysis of the contribution of JmjC-lysine demethylases common domains to gene expression</b>	137
2.1. JMJC-KDMs are highly disordered proteins	137
2.2. JMJC-KDMs are predicted to phase separate	140
2.3. KDM2A, KDM4B and PHF2 sequence features prime them to phase separate	142
2.4. KDM2A, KDM4B and PHF2 undergo liquid-liquid phase separation <i>in vitro</i> and <i>in vivo</i>	145
2.5. KDM2A, KDM4B and PHF2 condensates are related to transcription	150
<b>Discussion</b>	155
<b>1. Regarding chromatin three-dimensional structure and transcription regulation</b>	155
1.1. About the TGF $\beta$ -driven assembly of enhancer clusters	155
1.2. About the contribution of JMJD3 to the enhancer cluster establishment	157
1.3. About the interplay between three-dimensional genome structure and function	159
<b>2. Regarding phase separation and transcription regulation</b>	160
2.1. About JMJC-KDM IDRs characterization	160
2.2. About JMJC-KDM phase separation functional outcomes regarding transcription	163
2.3. About JMJC-KDMs phase separation and disease	166
<b>3. Regarding the interplay between phase separation and genome structure to promote transcription</b>	167

<b>Conclusions</b>	171
<b>Bibliography</b>	173
<b>Appendix</b>	191



## List of abbreviations

ALS: amyotrophic lateral sclerosis  
ATP: adenosine triphosphate  
BLAST: basic local alignment search tool  
Bp: base pair  
BMP: bone morphogenetic proteins  
BSA: bovin serum albumin  
ChIP: chromatin immunoprecipitation  
ChIP-seq: chromatin immunoprecipitation sequencing  
CNS: central nervous system  
cNT: chicken neural tube  
CTD: carboxy-terminal domain  
DNA: deoxyribonucleic acid  
DNMT: DNA methyltransferase  
dNTP: deoxynucleotide triphosphate  
EGF: epidermal growth factor  
EMT: epithelial to mesenchymal transition  
eRNA: enhancer RNA  
ESC: embryonic stem cell  
FAD: flavin adenine dinucleotide  
FBS: fetal bovine serum  
FGF: fibroblast growth factor  
FISH: fluorescence in situ hybridization  
FRAP: fluorescence recovery after photobleaching  
FTD: frontotemporal dementia  
GEO: gene expression omnibus  
GTF: general transcription factors  
H: histone  
HAT: histone acetyl transferase  
HD: hexanediol  
HDAC: histone deacetylase

HKMT: histone lysine methyltransferase  
HP1: heterochromatin protein 1  
HRMT: histone arginine methyltransferase  
HSC: hematopoietic stem cell  
IDR: intrinsically disordered region  
IF: immunofluorescence  
I-SMAD: inhibitory SMAD  
JmJC: Jumonji-C domain containing  
Kb: kilobase  
KDM: lysine demethylase  
KO: knockout  
LAD: lamina-associated domain  
LCR: low-complexity region  
LGPS: liquid-gel phase separation  
LLPS: liquid-liquid phase separation  
LMC: loop maintenance complex  
LSD: lysine-specific histone demethylase  
Mb: megabase  
MBD: methyl-CpG binding domain  
MeCP2: methyl-CpG binding protein 2  
EGFP: enhanced green fluorescence protein  
mEGFP: monomeric EGFP  
mRNA: messenger RNA  
mRNP: messenger ribonucleoprotein  
MSC: mesenchymal stem cell  
MZ: marginal zone  
ncRNA: non-coding RNA  
NE: neuroepithelial progenitors  
NELF: negative elongation factor  
NSC: neural stem cell  
PBS: phosphate buffered saline  
PcG: polycomb group of proteins  
PCR: polymerase chain reaction

PIC: pre-initiation complex  
PLB: passive lysis buffer  
PML: promyelocytic leukemia protein  
PPPS: polymer-polymer phase separation  
PolyA: polyadenylic  
PTM: post-translational modification  
RG: radial glial cells  
RNA: ribonucleic acid  
RNAPII: RNA polymerase II  
RNP: ribonucleoprotein  
ROI: region of interest  
RPM: reads per million  
RT: retrotranscription  
R-SMAD: receptor SMAD  
SBE: SMAD binding elements  
SDS: sodium dodecyl sulphate  
SEM: standard error of the mean  
SG: stress granule  
Shh: Sonic Hedgehog  
shRNA: short hairpin RNA  
SVZ: subventricular zone  
TAD: topologically associating domain  
TBE: Tris-borate-EDTA  
TBP: TATA-box binding protein  
TET: ten eleven translocations  
TF: transcription factor  
TFBS: transcription factor binding site  
TGF $\beta$ : transforming growth factor beta  
TGFBR: TGF $\beta$  receptor  
tRNA: transfer RNA  
TSS: transcription start site  
VZ: ventricular zone  
WT: wild type

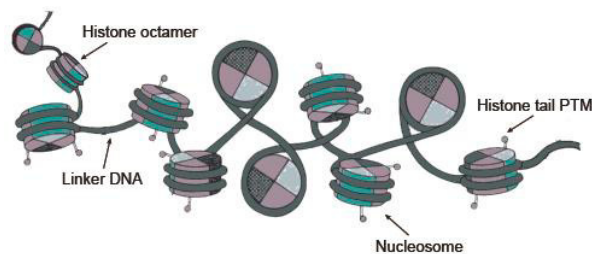


# Introduction

The embryonic development depends on the interrelation of signaling pathways, transcription factors (TFs) and epigenetic regulators, which are coordinated in a spatial-temporal manner to promote the specific gene expression program of each cell type (1). Along the present work we explored the crosstalk between the TGF $\beta$  pathway, lineage specific TFs and chromatin modifying enzymes to regulate gene expression during neurogenesis in neural stem cells (NSCs), focusing our efforts on the role of the histone demethylase JMJD3 on these events. Also, considering the discoveries made for JMJD3 along this thesis, we decided to broaden our study to deepen into the common intrinsic structural features that govern the JmjC-containing demethylases transcriptional regulation.

## 1. Chromatin

Chromatin is the polymer in which DNA is packaged within the cell nucleus. Its basic repeating structural unit is the nucleosome, which is formed by 147 base pairs (bps) of DNA wrapped around an octamer of the four core histones (H) (two copies of each histone H3, H4, H2A and H2B) and by linker DNA bound by histone H1. This last protein holds the core together and further condenses chromatin (Figure I1) (2).



**Figure I1.** Representation of chromatin polymer, which is formed by histone octamers and linker DNA with H1. Adapted from (3).

Thanks to microscopic observations, eukaryotic chromatin has been classically classified into two groups: the gene-rich, open, and

transcriptionally active euchromatin, that can be found at the center of the nucleus, and the gene-poor, compacted and transcriptionally silent heterochromatin, which is positioned at the nuclear periphery. Furthermore, heterochromatin can be subclassified into constitutive heterochromatin, which is conserved among cell types, and facultative heterochromatin, which is not conserved (4).

Functionally, chromatin acts as a platform to coordinate transcriptional events that take place specifically at euchromatic regions, both at the three-dimensional (3D) and the linear levels. This will be further reviewed in the following subsections of this introduction.

## **1.1. Epigenetic modifications of chromatin**

Chromatin is decorated by modifications, including DNA methylation and histone post-translational modifications, which impact its topological organization and its accessibility (5). The main modifications that chromatin is submitted to will be reviewed in the following subsections.

### **1.1.1. DNA methylation**

DNA methylation is an epigenetic modification that has been traditionally related to repression. It consists in the addition of methyl groups to the GC-rich DNA regions CpG islands (6). These CpG islands are located at the 5' ends of many genes, so that when they become methylated during development their associated promoter is silenced. The way in which this modification regulates transcriptional repression is both through the direct hindrance that the methyl group imposes for other proteins to bind DNA, and through the recruitment of factors that induce repression, such as methyl-CpG binding domain proteins (MBD1, 2 and 3) and the methyl-CpG binding protein 2 (MeCP2) (7). Apart from stabilizing repression of promoters, DNA methylation also regulates enhancer and insulator regions,

constituting a fundamental mechanism for a proper embryonic development (8).

There are many enzymes implied in the addition of methyl groups to DNA, and they can act both to maintain or to increase the global levels of methylation. The DNA methyltransferase 1 (DNMT1) enzyme maintains the methylation pattern between cell generations through the addition of the mark to the CpG islands that contain a methyl group in the parental strand (9). On the other hand, DNMT3A and DNMT3B catalyze de novo methylation mainly in embryonic cells (10)

DNA can be demethylated too, both in an active or in a passive way. While active DNA demethylation involves the removal of the methyl group through ten eleven translocation (TET) enzymes (11), passive demethylation implies the loss of the mark through successive rounds of replication without its active addition. Furthermore, there are many factors that promote DNA demethylation, including DNA cytosine deaminases, DNA glycosylases, DNA repair factors and DNA methyltransferases (12).

### **1.1.2. Histones post-translational modifications**

Histone proteins that conform chromatin are also subjected to the post-translational addition or removal of different chemical groups. The extensive combinatorial repertoire of modifications that histones can present, which give them the potential to regulate many chromatin-templated functions, constitute the so-called “histone-code” (13). These histone proteins, which are highly conserved among organism, contain two common domains: the fold or globular domain, which mediates the formation of H2A–H2B and H3–H4 dimers, and the unstructured, flexible N-terminal tail, which protrudes from the nucleosome core and interacts with DNA (14). Although residues found in both domains can be post-translationally modified (15), in this thesis I will focus on the N-terminal tail post-translational modifications (PTMs) as they are the most relevant for

transcriptional regulation (as well as for replication, recombination and DNA repair). In particular, I will review PTMs occurring on lysines of histone H3 N-terminal tails, although there are over 60 different residues on histone tails that are exposed to PTMs (2). These PTMs are deposited and/or removed by chromatin-modifying enzymes recruited through regulatory sequences, transcription factors and coactivators – the so-called “writer” and “eraser” proteins, respectively. They induce chemical changes on the histones they decorate, affecting their interaction with DNA and hence the compaction state of chromatin (2,16). Furthermore, they serve as targets for chromatin acting proteins – “readers” - such as remodelers, potentiating the regulatory capacity of chromatin (17). The most well-characterized modifications include acetylation, methylation, phosphorylation, ubiquitylation, and SUMOylation. They all exert different functions on chromatin, as seen in Table I1.

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	K-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	K-me1 K-me2 K-me3	Transcription, Repair
Methylation (arginines)	R-me1 R-me2a R-me2s	Transcription
Phosphorylation	S-ph T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription
Deimination	R > Cit	Transcription
Proline Isomerization	P-cis > P-trans	Transcription

**Table I1.** Main types of histone modifications, including the affected residues and the regulated biological functions. Adapted from (2).

Historically, acetylation is the most studied PTM on histone tails and both histone acetyl transferases (HATs) and histone deacetylases (HDACs) have been extensively analyzed (18). HATs utilize the acetyl-CoA as a cofactor to catalyze the addition of an acetyl group to the  $\epsilon$ -amino group of the lysine (Table I1). This reaction weakens the interaction between histones and DNA, so HATs usually work as coactivators and HDACs as corepressors (19,20) Among the main target residues of histone H3 for these enzymes we found K9, K14, K18, K23 and K27. They correlate with active transcription, as they are found at transcription start sites (TSS) and



enhancers. Due to its location at enhancers, the acetylation of H3K27 by the well-studied HAT p300/CBP is of particular interest for this thesis (21,22).

During the past decade, the focus has shifted more towards the study of histone methylation. Taking this into account, as well as the relevance that this PTM has for the present work, in the following subsection I will deepen into it. However, there are other PTMs apart from acetylation and methylation that play relevant roles in transcriptional regulation, such as phosphorylation, which is deposited by kinase proteins upon extracellular signals, DNA damage and mitosis and define chromatin structure (23); ubiquitylation, which consists in the deposition of ubiquitin to lysine residues and contribute to histone crosstalk and DNA repair processes (24); or sumoylation, which prevents the addition of ubiquitin to lysines (17).

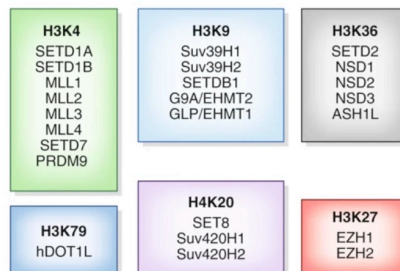
#### **1.1.2.1. Histone methylation**

Histone methylation consists in the addition of one, two or three –CH<sub>3</sub> groups to lysines, or one or two to arginines, and is performed by histone lysine methyltransferases (HKMTs) or arginine methyltransferases (HRMT) respectively. Despite not changing the electrical charge of the amino acid, this reaction has functional consequences (Table I1) (25). Although all histones can be methylated on one or more residues, some sites demonstrate certain prevalence to the targeting of HKMTs, including K4, K9, K27, K36 and K79 on histone H3 and K20 on histone H4. These residues can be mono-, di- and tri-methylated, in a way in which up to 40-80% of mammalian genomes are dimethylated at H3K9, H3K27, H3K36 or H4K20, while mono- and trimethylation is less abundant (26,27).

The association of methylation and gene regulation depends on the target residue and the degree of methylation at this site. Accordingly, H3K4me<sub>3</sub> and H3K36me<sub>3</sub> correlate with active transcription, while H3K27me<sub>2/3</sub>, H3K9me<sub>2/3</sub> and H4K20me<sub>2/3</sub> are associated with repression. The location

of these marks differs along the genome, so that through the combination of PTMs they establish an epigenetic pattern that distinguishes chromatin regions. This way, within active regions H3K4me1/2 is typically found at enhancers, H3K4me2/3 at promoters or around the TSS, and H3K36me3 at gene bodies (28). On inactive regions, H3K27me3 is enriched at promoters of silent genes whereas H3K9 and H4K20 methylation marks are relatively homogeneously distributed (29). The presence of these marks is not restricted to the mentioned regions; in fact, H3K27me3 and H3K4me3 marks coexist on the so-called “bivalent” promoters, which maintain genes “poised” for their subsequent activation. “Bivalent” regions are found on developmental genes in stem cell states, as they provide them with the capability to be rapidly activated or repressed upon developmental cues (30).

Regarding HKMTs, although there are many (31), the ones that are more related to our work are EZH2/1 and MLL, that form part of the PRC2 and MLL complexes, respectively (Figure I2). PRC2 is part of the Polycomb group of proteins (PcG), which mediate chromatin compaction and transcriptional repression; particularly, PRC2 complexes repress cell identity and developmental genes through the di- and trimethylation of the H3K27 residue (32). On the other hand, MLL complexes regulate gene activation through the methylation of the H3K4 residues. The antagonistic effect of both complexes is fundamental to ensure the correct expression of developmental genes during cellular identity determination (33).



**Figure I2:** Histone KATs found in humans, grouped by their substrate specificity. Adapted from (34).

### 1.1.2.2. Histone demethylation

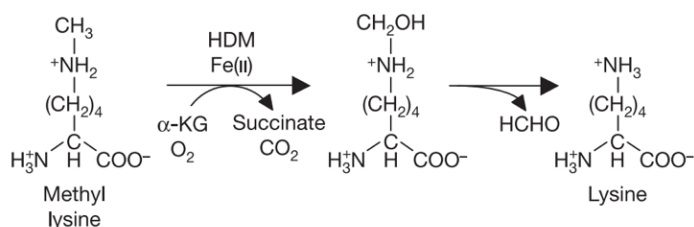
During many years histone demethylation was thought to be a passive process that resulted from histone exchange or dilution during replication. However, nowadays many histone lysine demethylases (KDMs) “eraser” proteins have been identified (reviewed in (35)), which target most of the methylated residues - H3K4, H3K9, H3K27, H3K36 and H4K20 (Table I2).

FAMILY	HDM	OTHER NAMES	HISTONE SUBSTRATE	REFERENCES
LSD	KDM1A	KDM1, KIAA0601, LSD1, AOF2	H3K4me2/me1, H3K9me2/me1	Shi et al. 2004, Metzger et al. 2005
	KDM1B	LSD2, AOF1, C6orf193	H3K4me2/me1, H3K9me2/me1	Ciccione et al. 2009
JMJC	KDM2A	JHDM1A, KIAA1004, CXXC8, FBL7, FBXL11	H3K36me2/me1	Tsukada et al. 2006
	KDM2B	JHDM1B, CXXC2, FBL10, FBXL10, PCCX2	H3K4me3, H3K36me2	Frescas et al. 2007, He et al. 2008
	KDM3A	JHDM2A, JMJD1, JMJD1A, KIAA0742, TSGA	H3K9me2/me1	Yamane et al. 2006
	KDM3B	JHDM2B, JMJD1B, KIAA1082, C5orf7	H3K9me2/me1	Kim et al. 2012
	JMJD1C	JHDM2C, KIAA1380, TRIF8	H3K9me2/me1	Kim et al. 2010
	KDM4A	JHDM3A, JMJD2A, KIAA0677	H3K9me3/me2, H3K36me3/me2, H1.4K26me3	Whetstone et al. 2006, Klose et al. 2006, Trojer et al. 2009
	KDM4B	JHDM3B, JMJD2B, KIAA0876	H3K9me3/me2, H3K36me3/me2, H1.4K26me3	Whetstone et al. 2006, Fodor et al. 2006, Trojer et al. 2009
	KDM4C	JHDM3C, JMJD2C, KIAA0780, GAS1	H3K9me3/me2, H3K36me3/me2, H1.4K26me3	Whetstone et al. 2006, Cloos et al. 2006, Trojer et al. 2009
	KDM4D	JHDM3D, JMJD2D, FLJ10251	H3K9me3/me2, H1.4K26me3/me2	Whetstone et al. 2006, Trojer et al. 2009
	KDM5A	JARID1A, RBP2	H3K4me3/me2	Christensen et al. 2007, Klose et al. 2007
	KDM5B	JARID1B, PLU1	H3K4me3/me2	Christensen et al. 2007
	KDM5C	JARID1C, SMCX	H3K4me3/me2	Christensen et al. 2007, Iwase et al. 2007
	KDM5D	JARID1D, SMCY	H3K4me3/me2	Lee et al. 2007a
	KDM6A	UTX	H3K27me3/me2	Agger et al. 2007, Lan et al. Al 2007
	KDM6B	JMJD3, KIAA0346	H3K27me3/me2	Agger et al. 2007, Lan et al. Al 2007
	UTY			
	KDM7A	JHDM1D, KIAA1718, KDM7	H3K9me2/me1, H3K27me2/me1	Tsukada et al. 2010, Hbrton et al. 2010, Huang et al. 2010b
PHF8	JHDM1F, KIAA1111, ZNF422	H3K9me2/me1, H4K20me1	Loenarz et al. 2010, Liu et al. 2010, Qi et al. 2010	
PHF2	JHDM1E, KIAA0662	H3K9me2/me1, H4K20me3	Wen et al. 2010, Baba et al. 2011, Stender et al. 2012	

**Table I2:** Histone demethylases families, subfamilies, alternative names and target substrates. Adapted from (35).

According to their catalytic mechanism of action, two evolutionarily conserved families of KDMs exist: the lysine-specific histone demethylase (LSD) family and the JmJC (Jumonji-C domain containing proteins) family (36,37). The LSD family uses a flavine adenine dinucleotide (FAD)-dependent amine oxidation reaction to catalyze the demethylation (38–40). There are only two enzymes within this family, which catalyze the removal of H3K4me1/2 and H3K9me1/2: LSD1, the first discovered KDM, and LSD2

(41). On the other hand, the JmjC family coordinates Fe (II) to catalyze the removal of methylated lysines using alpha-ketoglurate as a cofactor (Figure 13). Through this reaction, the methyl group is converted to a hydroxymethyl group that is finally released as formaldehyde. These are the KDMs we have studied along this thesis, so I will further review them in the following subsection.



**Figure 13:** JmjC-domain containing KDM lysine demethylation reaction. Adapted from (42).

Recently, other two enzymes have been assigned as H4K20 demethylases that require Fe (II) and alpha-ketoglurate as cofactors, as well as the ubiquitin-associated domain: the well-known DNA repair proteins hHR23A and hHR23B (43,44).

### 1.1.2.3. JmjC containing KDMs: KDM6 subfamily and JMJD3

The JmjC-containing KDM family encompasses 24 JmjC-domain containing KDMs, which are evolutionarily conserved from yeast to human. They play major roles in many physiological and pathological processes, such as gene expression (45), embryonic stem cell renewal (46), cellular differentiation (47), X-linked mental retardation (48), and cancer (49).

Apart from the JmjC catalytic domain, their function is determined by the combination of other conserved structural domains such as the PHD, Tudor, CXXC, FBOX, ARID, LRR and JmjN. According to their structural similarities and their specificity for histones, they can be classified into

seven subfamilies (Table I2). Due to its relevance for the present work, I will focus the attention on the KDM6 subfamily (36).

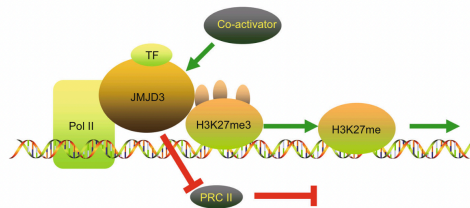
The KDM6 subfamily of JmjC KDMs comprise three members: KDM6A (UTX), KDM6B (JMJD3) and KDM6C (UTY), which present structural differences. UTX and UTY contain a JmjC catalytic domain and several tetratricopeptide repeats that mediate protein-protein interactions, while the only defined domain in JMJD3 is the catalytic one (50–52) (Figure I4). KDM6 substrates are the repressive histone marks H3K27me2/3, which are deposited by the PRC2 complex and are fundamental during development.



**Figure I4.** Phylogenetic tree that shows the three members of the KDM6 subfamily of KDMs. JmjC: Jumonji-C domain; TPR: tetratricopeptide repeats. Adapted from (53).

Along the present work we have studied specifically JMJD3. Human JMJD3 encodes a 1643 amino acid protein, while murine *Jmjd3* consists of 1641 residues; they are highly conserved proteins, sharing more than 90% of identity. This enzyme is ubiquitously expressed and is involved in many different physiological and pathological processes according to the context, such as proliferation, senescence, apoptosis, cell cycle regulation, differentiation, tumorigenesis, neurodegeneration or inflammation (54,55). Indeed, JMJD3 is regulated in a context-specific manner, so that in normal conditions it is expressed in low levels, and it is specifically induced upon developmental, immune, and oncogenic stress-related stimuli. This way, it is upregulated by the MAPK, Wnt, NF- $\kappa$ B, bone morphogenetic protein (BMP), TGF $\beta$ , IL-4-STAT6, and T-bet signaling pathways. Then, it is recruited to the chromatin through the interaction with transcription factors (TFs), where it regulates gene expression (51,56,57).

During the present work we specifically studied the participation of JMJD3 in development. It is worth noting the protein's role on this process as it targets the repressive marks H3K27me2/3, which are fundamental to determine cell fate (32,35). This way, JMJD3 induces gene activation through the demethylation of the repressive marks at promoters, the release of the PRC2 complexes, and the promotion of transcription elongation (Figure I5). However, its regulatory capacity relies not only on the demethylase activity, but also on demethylase-independent mechanisms. Indeed, JMJD3 can behave as a transcription factor that interacts with coactivators in gene promoters and enhancers to activate transcription, as observed in inflammation, reprogramming and differentiation processes (58,59) (Figure I5). These two mechanisms have been particularly observed during neurogenesis promotion, which will be further commented in the subsection 3.3. of the thesis.



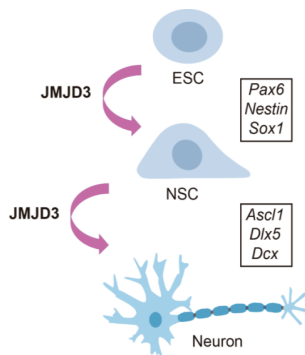
**Figure I5.** Model depicting the demethylase-dependent and independent mechanisms used by JMJD3 to regulate gene expression. Adapted from (60).

In mice, JMJD3 knockout (KO) experiments have shed different results depending on the followed strategy. The homozygous deletion of its catalytic domain caused perinatal death due to premature alveolar development (61,62), while mice lacking 50% of JMJD3 died peri- or neonatally due to the malfunction of the interneurons that comprise the Pre-Böttinger complex, which generates the respiratory rhythm (63). The total deletion of JMJD3 caused embryonic lethality at the E6.5 stage (64).

Regarding embryonic stem cells (ESCs) biology, JMJD3 is not necessary to maintain their pluripotent state, but it is to generate the three germinal

layers. Upon differentiation signals, JMJD3 promotes meso-, endo- or ectodermal transcriptional programs through the activation of lineage specific regulators (56,64). With respect to the mesoderm lineage, JMJD3 is involved in cardiomyogenic lineage determination and differentiation through its cooperation with the insulin gene enhancer-binding protein 1 (ISL1) and the Wnt signaling pathway. Concerning endoderm differentiation, JMJD3 has been demonstrated to interact with the TGF $\beta$  pathway to demethylate NODAL promoter. Finally, JMJD3 promotes ectoderm differentiation through the regulation of Pax6, Nestin and Sox1, three key markers of neurogenesis, and the enhancement of neural commitment (58).

JMJD3 plays major roles in multipotent stem cells. Indeed, it has been reported that JMJD3 catalytic activity is important in the osteogenic commitment of mesenchymal stem cells (MSCs) and in the self-renewal capacity of hematopoietic stem cells (HSCs) (58). Moreover, JMJD3 is necessary for the catalytic-dependent induction of neuronal transcriptional programs in neural stem cells (NSCs) (47), and for the retinoic acid-induced differentiation to neurons, which results from the de-repression of the neurogenic pioneer transcription factor *Ascl1* (65) (Figure 16). It also increases the expression of the *CDKN2a* locus to stabilize the tumor suppressor p53, which has a key function in mouse neurogenesis (58).



**Figure 16.** The KDM JMJD3 participates in the commitment and differentiation to the neuronal lineage. Adapted from (35).

The role of JMJD3 within this neurodevelopmental context has been extensively studied, both through *in vivo* experiments in the chicken neural tube (cNT) and taking advantage of the *in vitro* model of NSCs. This way, it has been shown that JMJD3 and the developmental pathway of TGF $\beta$  crosstalk to promote neuronal differentiation (66,67). In the cNT, the BMP signaling pathway was also shown to induce JMJD3, which activates the negative regulator of the BMP pathway *Noggin*. Hence, it establishes an inhibitory feedback loop that controls the appearance of dorsal interneurons (68). These results will be further reviewed in the subsection 3.3., as they were the basis for the present work.

Regarding its implication in neural disorders, JMJD3 has been shown to exert dual functions. On the one hand, it is found to be a protective agent in some neurodegenerative disorders, such as Parkinson's disease, by promoting a microglia anti-inflammatory response and mediating the differentiation of midbrain dopaminergic neurons (69–71). A similar role was detected in Alzheimer's disease, where JMJD3 stabilizes p63 nuclear accumulation, so that the tumor suppressor antagonizes neuronal apoptosis (72,73). On the other hand, the demethylase was demonstrated to promote neuropathic pain and neuroinflammation, and hence neurodegenerative disorders progression, through the cooperation with NF- $\kappa$ B, STAT, TGF- $\beta$ /SMAD3, and T-bet signaling (74–76). Furthermore, genetic variants in JMJD3 gene were associated with dysmorphic facial features and neurodevelopmental delays, including speech and motor delays, and some degree of intellectual disability (77).

## **1.2. Regulatory regions on DNA**

The genome contains many cis- and trans-regulatory regions, that basically are clusters of binding motifs that recruit TFs in an orchestrated spatial-temporal manner to control gene expression programs (78). The different regulatory regions found in the genome will be reviewed along this section.

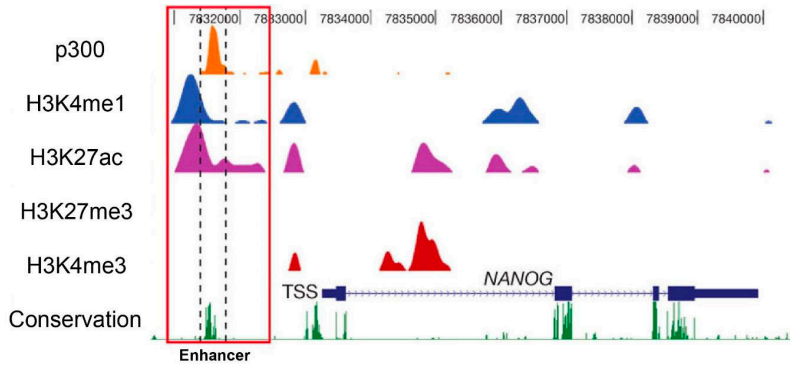


### **1.2.1. Gene promoters**

Promoters are cis-regulatory regions that control the transcriptional output rate through the recruitment of the transcription initiation complex. They provide great regulatory versatility by differing in their accessibility and architecture (79). Core promoters are the regions that contain the TSS, and sometimes a TATA-box, which constitute the platforms to which the pre-initiation complex (PIC) binds to initiate transcription. This way, a competent PIC contains TATA-box binding proteins (TBP), the general transcription factors (GTFs), that include TFIIB, TFIIF, TFIIE and TFIIH, and the RNA polymerase II (RNAPII) (80). Apart from PIC recruitment, the RNAPII pause-release constitutes another transcriptional regulatory step from promoters. After the transcription of 50 nucleotides by RNAPII, the enzyme pauses until receiving cellular release signals that trigger elongation (81).

### **1.2.2. Enhancers**

Enhancers are defined as cis-regulatory regions that can induce gene activation from a target promoter in an orientation and distance-independent manner. They typically consist of hundreds of bps that contain multiple binding sites for TFs, which are distally located from promoters (82). They also display a specific pattern of chromatin features that allow their identification, consisting of histone marks, cofactors, and chromatin accessibility data. This way, enhancers show high levels of H3K4me1 and absence or low levels of H3K4me3. Active enhancers also contain H3K27ac, together with H3K27me3 for the inactive “poised” enhancers. Usually, both classes of enhancers are also bound by the HAT p300 (Figure 17) (83). With respect to chromatin accessibility, enhancer regions are characterized by low nucleosome density and clusters of TF binding sites (TFBS) (84).



**Figure 17.** Capture that shows the chromatin pattern characteristic of active enhancers. Adapted from (83).

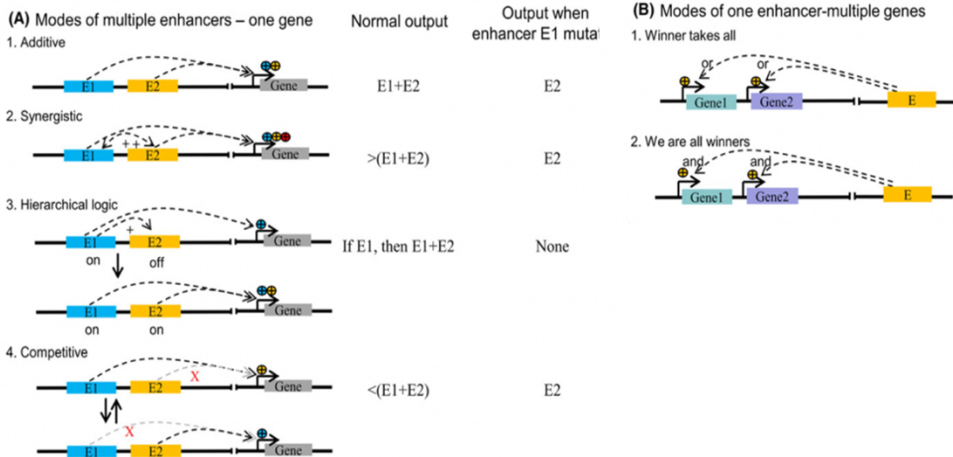
Enhancers constitute the major regulatory unit during development, when they act in a cell-type specific manner to achieve a precise spatial-temporal control of gene expression. In this regard, enhancers are able to regulate the levels of gene expression according to the particular cellular context through their own modulation, both by signaling cascades and the binding of specific TFs (1). Then, to promote gene transcription they contact promoters through the establishment of chromatin loops. Although enhancer-promoter interaction is fundamental for gene transcription, not much is clearly known regarding the processes that mediate it. For example, it is not known how the specificity between enhancer and promoter is established. Classically, enhancers have been assigned to their nearby promoters just as a matter of linear proximity, but it is currently known that it is not always the case that enhancers regulate their nearest promoters (85). Nowadays, with the development of chromatin conformation capture techniques enhancer-promoter contacts are starting to be studied in more detail, but in-depth knowledge in the field is still missing.

Apart from driving gene transcription, enhancers themselves have been reported to be transcribed by the RNAPII into nuclear unspliced bidirectional RNA species, named enhancer RNAs (eRNAs) (86,87).

Although they are expressed at low levels, nowadays it is accepted that these non-coding RNAs (ncRNA) are functional molecules necessary for the enhancer activity, the expression of nearby genes, and the establishment of enhancer-promoter loops. Indeed, they facilitate DNA decompaction, the stabilization of TF binding and the recruitment of cofactors, and the release of the negative elongation factor (NELF) from promoters (82,88). Classically, eRNAs have been defined as short, non-polyadenylated ncRNAs that arise by divergent transcription; however, recently they have also been shown to be longer, contain a polyA tail, and generate unidirectionally (89). Anyway, despite its current controversial nature, eRNA has been determined as a mean to assess enhancer activation.

Enhancer regions have been shown to cooperate to achieve a higher transcriptional activity than the one promoted by a single, typical enhancer. This way, they can form enhancer clusters that work in a coordinated manner to induce gene transcription. Taking this into account, several models have arisen to explain the way in which enhancers regulate gene expression (Figure I8). They have been proposed to act in an additive manner, so that transcription is determined by the additive effect of multiple enhancers; in a synergistic manner, when multiple enhancers produce higher levels of transcription than their sum individually; in a hierarchical manner, in which one enhancer could induce transcription at basal levels while activating other close enhancers to potentiate gene expression; and in a redundancy manner, so that more than one enhancer control the same gene and compete, this way ensuring gene expression even though an enhancer is lost. Indeed, enhancer redundancy is a remarkable feature of mammalian genomes. All these mechanisms correspond to the activity that results from the above-mentioned enhancer clusters and are summarized as “multiple enhancers-one target gene” (Figure I8A). Also, there are models that depict the opposite situation, where individual typical enhancers regulate various genes. These are the so-called “one enhancer-multiple genes” models (Figure I8B), which basically comprehend the

“winner takes it all” model, when one target gene is activated in each cell, and the “we are all winners” model, when many genes are activated in all cells, but not at maximum levels (85).



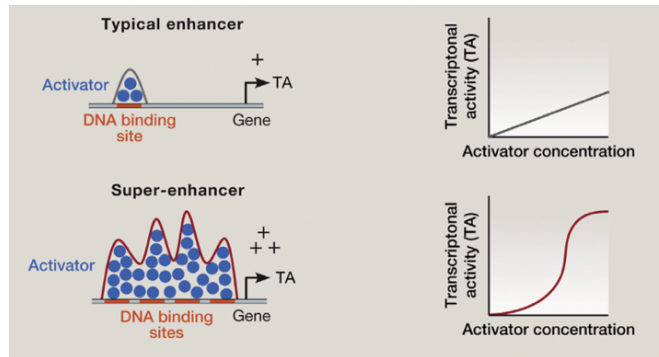
**Figure 18.** Models of enhancer-driven transcriptional activation, including the “multiple enhancers-one target gene” and the “one enhancer-multiple genes” models. Adapted from (85).

Currently, superenhancers (SEs) are the most well-studied enhancer clusters, as they play a major role during development by controlling genes that determine cell identity and cell fate. They will be further analyzed in the following subsection.

### 1.2.2.1. Superenhancers

SEs are constituted by several clusters of enhancers that act together to promote a high expression of cell-type-specific genes to maintain cell identity and determine cell fate. Indeed, SEs induce activation in a much higher manner than typical enhancers do (90,91) (Figure I9). To achieve such a level of transcription, they accumulate large amounts of interacting factors, up to 10-fold the density of the same components found at typical enhancers. These interacting components include lineage specific TFs, co-

factors (such as the Mediator complex), chromatin regulators, RNAPII, and ncRNAs (such as eRNAs) (Figure I9) (82,91). They are also characterized by a higher enrichment than typical enhancers in H3K4me1, H3K27ac, p300 (88). Despite the commented features that define SEs, due to the low resolution of current methods to study them, both their definition and composition are very ambiguous.



**Figure I9.** Comparison between SEs and typical enhancers features and potential to induce transcriptional activity (92).

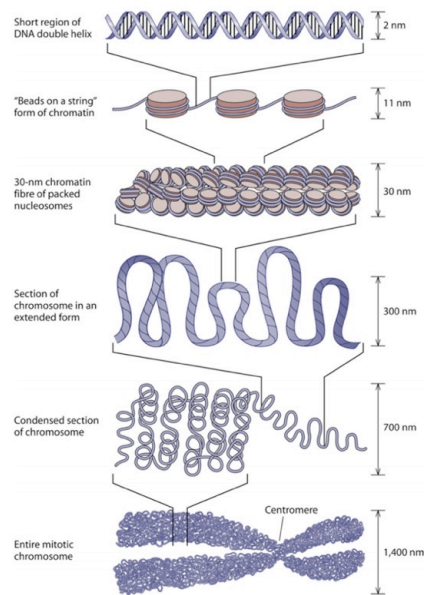
SEs are very sensitive to the alteration of some of their components, and they arise from a single nucleation event. These facts highlight the importance of cooperation for their formation and function. Apart from acting together, the cluster of enhancers that form the SE physically contact one another and with the promoter of their target genes (82). Not only they are vulnerable to the alteration of their components, but also to their constituents' enhancers, demonstrating they are interdependent (82,93).

### 1.2.3. Silencers and insulators

Silencers are regulatory regions very similar to enhancers in that they operate in an orientation and distance-independent manner, but instead of activating, they hinder transcription, as they target repressor proteins (94). For its part, insulators delimit gene expression within genomic boundaries (95).

### 1.3. Three-dimensional structure of chromatin

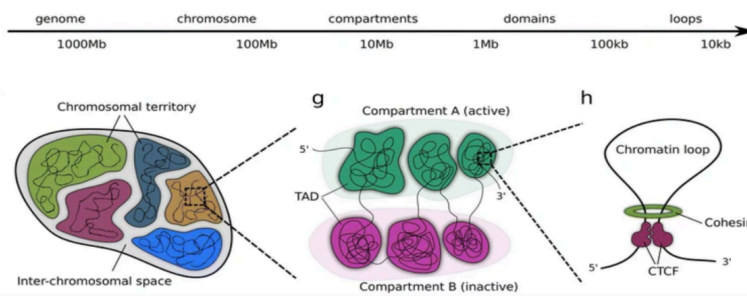
The DNA found within eukaryotic cells is around 2 meters long, so it needs to be packaged into chromatin to fit the 5-10  $\mu\text{m}$  of diameter of the nucleus (96). The first level of compaction is the “beads on a string”, a flexible 10 nm fiber formed by repeating nucleosomes, which shorten the DNA length up to seven times (Figure I10). Traditionally, the 30 nm chromatin fiber formed by the coiling of nucleosomes and linker DNA is considered the next level of compaction. However, the contradictory fluid-like model views the chromatin as a dynamic structure based on the irregular 10 nm fiber (97,98). Anyway, if the traditional view is kept, the 30 nm fiber continues coiling to form higher order stages of condensed and compacted chromatin, which end up in the metaphase chromosome (99,100) (Figure I10).



**Figure I10.** Depiction of the different levels of chromatin compaction, from DNA double helix to mitotic chromosomes. Adapted from (101).

During interphase, the compacted chromatin fiber is organized in the three-dimensional space into structures of growing complexity. This way, we can

distinguish chromosome territories, topologically associating domains (TADs), chromosome compartments and loop domains (102) (Figure I11). The significance of genome 3D-organization cannot be dismissed, as there is an intimate relationship between genome structure and function that is not fully understood yet. Nowadays, there is a lot of controversy on whether the three-dimensional structure of the genome is a cause or a consequence of function, but what does seem clear is that genome architecture plays major regulatory roles. Given this significance, albeit the many unanswered essential questions and the current methodological limitations to study them, in the following subsections I will deepen into each one of these levels of 3D structure.



**Figure I11.** Genome is organized in the three-dimensional space into chromosome territories, compartments A and B, TADs and chromatin loop structures. Adapted from (102).

It is worth noting that the current knowledge regarding the three-dimensional structure of chromatin comes from two complementary type of assays: imaging techniques and chromosome conformation capture (3C) methods. On the one hand, microscopy-related assays, such as RNA and DNA fluorescence in situ hybridization (FISH), live-cell imaging, and electron microscopy, uncover major principles of genome organization and correlate nuclear location with transcriptional output. On the other hand, 3C techniques are based on the digestion and religation of crosslinked chromatin in cells, followed by the quantification of ligated fragments. This way, it is possible to detect DNA contacts frequencies, both in a locus-

specific (3C) or in a high-throughput sequencing manner (4C, 5C, Hi-C, and other derived methods). Recent advances in both superresolution microscopy and loci tracking in live-cells, and single-cell 3C technologies, represent promising opportunities to explore the connections between chromosome structure and nuclear functioning (103).

### **1.3.1. Chromosome territories**

Nowadays it is known that the chromosome location at the nucleus is not random. Each chromosome occupies discrete regions within the nuclear space, so that gene-rich chromosomes are most likely located at the center of the nucleus while gene-poor are at the periphery, and they even cluster in a species and cell-type specific manner (Figure I11). These specific positions have been named as chromosomal territories, which can be observed both by fluorescence techniques and chromosome capture methods (104). Chromosome territories show limited intermingling between them. Thus, the functional consequences of these locations cannot be dismissed; chromosome territories will determine gene expression programs, and consequently chromosome mislocalizations give rise to many pathological processes (105).

### **1.3.2. Chromosomal compartments**

Hi-C experiments have shown plaid-like contacts maps constituted by alternating blocks of enriched and depleted interaction frequencies. This suggests that chromatin is further divided into spatial clusters of tens to hundreds megabases (Mb) size with a similar epigenetic and transcriptional state (104,106). These clusters are named compartments, and according to its functionality two compartments can be defined: A and B compartments (107) (Figure I11). While A compartment corresponds to active regions - gene-rich, accessible, transcriptionally active chromatin -, B compartment correlates to gene-poor regions and late replication. The latter can be further subdivided into constitutive heterochromatin, located



close to nuclear periphery lamina-associated domains (LADs) and nucleoli, and facultative heterochromatin, positioned within the nuclear interior (108). It is worth noting that loci contained within a compartment preferentially contact between them.

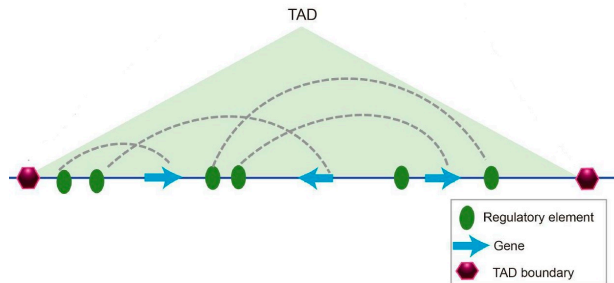
Compartments are not conserved but change across cell types and during development (107,109). Regarding their formation, it is believed that it is promoted by homotypic chromatin interactions that can segregate distal loci into a determined compartment. Apparently, these mechanisms would be specific for each compartment type, although currently they are not fully understood (110).

### **1.3.3. Topologically associating domains**

TADs are insulated self-interacting domains on the scale of few kb to Mb that seem to be intrinsic to mammalian genomes, as they are very stably conserved across species and cell types. They were defined by means of chromatin conformation capture techniques, specifically the high-throughput version Hi-C (111) (Figure I12). Despite their conservation, TADs have been demonstrated to be dynamic structures that disappear during mitosis to be later reorganized during G1 (112).

Functionally, genes contained within TADs are coregulated (111), and enhancer-promoter loops predominantly take place within them (113), so TADs have been proposed as the basic regulatory unit of genome folding. Indeed, TAD borders or boundaries have been shown to act as insulators (Figure I12). They also constitute fundamental regulatory regions found between TADs, which show enrichment in binding motifs for the insulator protein CTCF and, to a lesser extent, housekeeping genes, active chromatin marks such as H3K4me3 and H3K36me3, retrotransposons, and transfer RNAs (tRNAs) (111). Therefore, through their insulation capacity, TADs might represent fundamental regulatory units that promote and delimit specific enhancer-promoter interactions, preventing ectopic

interactions to occur (114). This way, they are used as indicative of high frequency interactions between loci within a domain and low frequency contacts between loci in different domains (115).



**Figure I12.** Schematic representation of a TAD, including the contacts established within them and its main components. Adapted from (116).

These three-dimensional chromatin structures are formed and maintained by the loop maintenance complex of proteins (LMC), which include CTCF and the cohesin complex (SMC1, SMC3, RAD21 and SA1/2 proteins). Both protein complexes have been found in TAD borders; particularly, CTCF proteins are placed in a convergent orientation within borders (111,115). Taking this into account, the current model of TADs appearance is the loop extrusion mechanism, which sustains that cohesin is loaded to enhancers and promoters to progressively extrude larger loops, until encountering convergent CTCF proteins at TAD boundaries (117) (Figure I11). This mechanism seems to be a completely different and opposing manner of chromatin organization than the homotypic chromatin interactions underlying chromosome compartments formation (110).

#### 1.3.4. Loop domains

Intra-TADs genomic structures are also observed through targeted 3C-based approaches, which provide lower coverage by higher resolution. This way, it has been observed that chromatin further establishes contact or loop domains, which are long-range interactions that bring together two distant

regions of the genome, typically enhancers and promoters, to increase gene transcription (106) (Figure I11). These loops have a median length of 185 kb, although they can span from few kb to more than 100 Mb, and they are more variable and cell-type specific as compared to TADs (118). Indeed, loop domains are also very dynamic, disassembling during mitosis and reassembling again on G1 (112).

There are various mechanisms beneath loop appearance. Although cohesin-dependent loop extrusion and CTCF-cohesin interactions mechanisms were considered the prototype ones for loop formation, as for TADs, it is becoming apparent that they only involve a limited number of enhancer-promoter contacts. Alternatively, other mechanisms seem to be more prevalent, such as enhancer-promoter dimerization processes driven by YY1 or LDB1 proteins, or the formation of transcriptional condensates through phase separation within which enhancer-promoter contacts are facilitated (82,110). The phase separation phenomena will be further commented on the subsection 4.5.1 of this introduction.

## **1.4. Regulatory proteins of chromatin**

### **1.4.1. Transcription factors**

TFs are proteins that bind both enhancer and promoter regulatory regions to increase or decrease gene transcription and protein synthesis, and hence they modulate cell functions. At enhancers, activator TFs promote enhancer-promoter loop formation; at promoters, they induce the formation of the transcription initiation complex through the recruitment of the GTFs and Mediator proteins, which then recruit the RNAPII (119,120). In addition to the DNA sequence itself, other factors contribute to the binding of TFs to cis-regulatory regions. For instance, it is known that the chromatin state, the presence of cofactors, the need of TF cooperative binding, and TFs post-translational modifications will regulate their DNA-binding capacity (121).

Nowadays many TFs have been identified, both ubiquitous and cell specific, and they are classified into distinct families according to their structural characteristics. It is worth noting that TFs can form homodimers or heterodimers that regulate their transcriptional activity at different sites from their specific target, allowing crosstalk between distinct signaling pathways. In any case, they translate environmental signals into changes in gene transcription, so their role over cell biology determination is remarkable (119).

Sometimes DNA is not accessible for TFs due to chromatin compaction and the binding of repressor complexes, which hinder the recognition of their TFBS. Pioneer TFs are proteins that can overcome this situation, binding closed chromatin and opening it to allow transcription of silent regions. These TFs are of major importance during cell lineage determination (122). One such pioneer TF involved in neurogenesis is the pro-neural factor ASCL1, which participates in the proliferation and differentiation of neural progenitor cells to neurons. Indeed, in NSCs ASCL1 is able to bind its target sites in close chromatin, promoting chromatin loosening and gene expression (123). Together with ASCL1, the effector of the TGF $\beta$  pathway SMAD3 will be especially relevant for the present work. Both TFs will be further described in the subsection 3.3. of this introduction.

#### **1.4.2. Transcriptional coregulators**

The set of proteins that modulate transcription without directly binding to DNA are known as transcriptional coregulators. After the specific binding of TFs to DNA, genes become marked for activation or repression through coactivator or corepressor proteins, respectively. As this thesis has been focused on transcription activation, we will further comment on transcriptional coactivators.

Coactivators promote gene expression after docking to TFs both through the direct enzymatic modification of chromatin (primary coactivators) or through the recruitment of other modifying proteins (secondary coactivators). In fact, coactivators can directly regulate gene expression by being the primary target of signaling pathways. Given this, three classes of coactivators can be distinguished: the ones that open chromatin to increase its accessibility for other factors, such as p300 and CBP HATs; the ones that recruit RNAPII and the transcriptional machinery, such as the Mediator complex; and the ones that unwind DNA to allow gene transcription, such as the SWI/SNF complexes (124).

Especially relevant for the context of this thesis is the coactivator complex formed by the histone demethylase JMJD3 and the ATP-dependent chromatin remodeler CHD8, which is highly related to autism spectrum disorder (125). This protein binds H3K4me2 and H3K4me3 histone marks and disrupts or promotes the winding of nucleosomes and DNA (126). The coactivator complex will be described with more detail in the subsection 3.3. of this thesis.

#### **1.4.3. RNA-polymerase II**

The ultimate goal of the so far reviewed proteins is to promote transcription, which is conducted by the RNA-polymerases (RNAP) complexes that attach to the template DNA and produce complementary RNA. Although in eukaryotes there exist three RNA-polymerases, the one that transcribes mRNA and eRNA is the RNAPII (127). In humans, RNAPII constitutes a complex formed by 12 subunits named RBPs, of which RBP1 is the enzymatic core. From this RBP1 subunit emerges the carboxy-terminal domain (CTD) – 52 tandem repeats of the consensus sequence YSPTSPS in humans – that is fundamental for the control of transcriptional and co-transcriptional processes, such as chromatin modification or splicing (128). Lately, the CTD has been related to phase separated transcriptional

condensates; this will be further commented on the section 4.5.1.2 of the present introduction.

## **2. The cortex development**

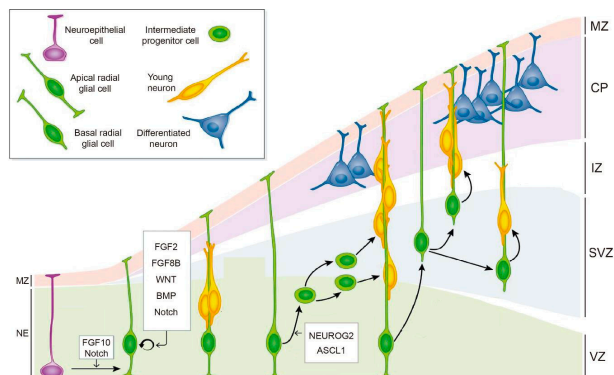
In this section of the introduction, I will review the essential aspects of the neural cortex development to properly introduce the *in vitro* model of mouse NSCs, which I used along the present work.

Intellectual ability is believed to depend on the mammalian brain size, which in turn is the result of the cerebral cortex size. Indeed, the cortex is the brain region to which cognitive function, consciousness and sensory perception are attributed. Human cortex size and folding is notorious; although it is conserved in all vertebrates, it is thought to provide the uniqueness to the human species (129). It consists of six layers of different neuronal subtypes and glial cells, which conform a really well-organized structure where each area provides specific motor, sensory and cognitive functions (130). It is worth noting that alterations in the proliferation or survival of neural progenitors translate into abnormal brain size, either in megacephaly, microcephaly, or dysplasia, which translates in pathological outcomes.

### **2.1. Mouse corticogenesis and its regulating molecular pathways**

In mice, corticogenesis begins at day 11 (E11). At this embryonic stage, cortical progenitors populate the dorsal telencephalon proliferative zones: the ventricular and the subventricular zones (VZ and SVZ). These progenitors, which initially form a polarized neuroepithelium, are submitted to successive rounds of maturation to restrict their fate and generate all cortical cell types. In mammals, this process is triggered by intrinsic and extrinsic signals that regulate the specific spatial-temporal neuronal generation (131).

The most relevant neurodevelopmental process for this work is neurogenesis, which begins when neuroepithelial progenitors (NE), that serve as neural stem cells, start proliferating to amplify the pool of cortical progenitors. NE give rise to the main subtype of cortical progenitors, the radial glial cells (RG), which will finally generate all subtypes of neurons that compose the cortex. For this, RG will divide asymmetrically, giving rise to another identical cell and a more differentiated intermediate progenitor cell (132). Then, more progenitors will go through symmetric terminal divisions to generate two differentiating daughter cells, so that the expansion of the progenitors progressively slows and stops. Neurons are generated in the deeper part of the developing brain, the VZ. Next, they migrate to the surface so that they generate the six-layered cortical structure. The last step in neurogenesis is the neuron translocation to the marginal zone (MZ), where it will mature to develop axons and dendrites (133) (Figure I13).



**Figure I13.** Representation of the developing mice cortex, including the types of progenitor cells and the signaling and intrinsic factors that regulate the process. Adapted from (134).

As mentioned above, the differentiation of cortical progenitors to neurons depends upon signaling pathways and intrinsic factors, including proneural TFs, cell polarity and mitotic molecules. These factors will determine the proliferative or differentiating behavior of progenitors (132). Regarding signaling pathways, it is worth noting the major roles of the fibroblast growth

factor (FGF), which promotes proliferation of NSCs *in vitro* and *in vivo* (135); the Notch pathway, whose role is fundamental for the maintenance of NSCs self-renewal capacity (136); the Sonic Hedgehog (Shh) pathway, which controls cell cycle of RG (137); or the canonical Wnt pathway, which induce proliferation of VZ neural precursors and neuron fate (138). In addition, the BMP and TGF $\beta$  pathways participate in forebrain development. Specially interesting for this thesis are the functions of the TGF $\beta$  signaling, which promotes cell cycle exit of VZ progenitors (139) and neuronal cell fate in cortical and hippocampal progenitors (140). On the other hand, intrinsic factors operate mainly in a cell-autonomous manner. Essential intrinsic factors are the proneural TFs, including ASCL1, neurogenins (NEUROG1, 2 and 3), and neurogenic and oligodendrocyte differentiation proteins (NEUROD1, 2 4 and 6 and OLIG1, 2 and 3, respectively) (141) (Figure I13). These TFs are both necessary and sufficient to trigger neuronal differentiation programs. In fact, their expression is restricted to certain cell populations of the central nervous system (CNS) (142). Furthermore, corticogenesis is also conditioned by the cell polarity and the different modes of cell division (132,143) (Figure I13).

## **2.2. Neural stem cells as an *in vitro* model of study**

In the experimental work of this thesis, NSCs from the cortex of mice in the embryonic stage of 12.5 have been used as a model of study. NSCs are multipotent cells in the nervous system that can self-renew, proliferate in an unlimited manner, and differentiate into neurons, astrocytes, and oligodendrocytes. It is worth noting that this *in vitro* model of study displays both morphology and markers similar to the ones observed in the radial glia in the mouse embryo, such as BLBP, RC2, GLAST, PAX6 and NESTIN (144).

NSCs exist in the developing cortex and in specific regions of mammalian adult CNS. *In vivo*, they occupy a microenvironment called niche that supplies the factors required for stem cell self-renewal and differentiation.



The lack of this environment makes it necessary to add mitogens such as epidermal growth factor (EGF) and FGF to the adherent *in vitro* system, forcing the retention of their multipotency and reducing their differentiating capacity. However, the addition of mitogens might alter the transcription programs of these cells (145). Other limitations of *in vitro* NSC cultures are the loss of cells positional identity and the restricted capacity to differentiate into neuronal subtypes (146).

Due to the heterogeneity of events that take place during vertebrate neurogenesis, the study of transcriptional regulation is very difficult in the developing embryo. The use of adherent cultures of NSCs gives us the advantage of avoiding such heterogeneity of the *in vivo* systems, despite the limitations that this model imposes.

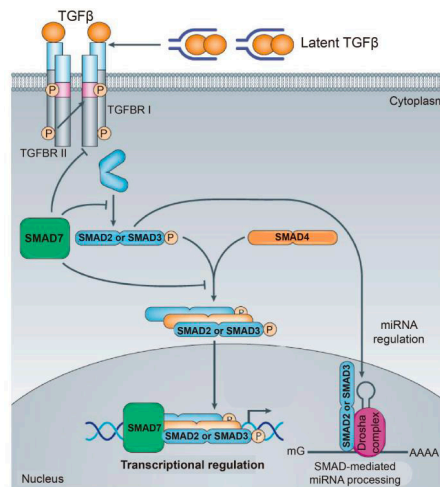
### **3. The TGF $\beta$ pathway**

The TGF $\beta$  cascade will be reviewed in this section of the manuscript, as it has served as a mechanism to commit NSCs to the neuronal lineage. Thus, the addition of TGF $\beta$  provided us with a model to study the interplay between signaling events, epigenetic coactivators, and TFs to trigger specific neurodevelopmental transcriptional programs.

#### **3.1. The signaling cascade of TGF $\beta$**

The TGF $\beta$  superfamily is composed by over 30 members, including Activins, Nodals, BMPs, TGF $\beta$ s and Growth and Differentiation Factors, which are ubiquitously expressed in embryonic and adult tissues. All these ligands participate in specific spatial-temporal intercellular communication events. Thus, their misregulation leads to pathological processes (147). Although all the mentioned ligands trigger a universal signaling cascade, for the sake of the present work we will focus on the TGF $\beta$  pathway triggered by the cytokine TGF $\beta$ 1 (from here on, TGF $\beta$ ).

The TGF $\beta$  cytokine has the ability to bind to a serine/threonine kinase receptor, named TGFBR type II, that forms a dimer with TGFBR I. Once bound, TGF $\beta$  increases the affinity of type II receptor for type I, so that the latter becomes phosphorylated by the kinase domain of type II receptor. This signal is propagated to the TFs that govern the pathway, the cytoplasmic SMAD2 and 3. Thus, the proteins get phosphorylated too, and this modification allows them to form a complex with SMAD4 to enter the nucleus, bind the SMAD binding elements (SBE) and regulate specific transcriptional programs (148) (Figure I14).



**Figure I14.** Graphic depiction of the TGF $\beta$  signaling pathway and its effector proteins, the TFs SMAD. Adapted from (149).

Regarding the effector proteins of the pathway, it is worth noting that there are three types of SMAD proteins according to their role in the cascade: the receptor SMADs (R-SMADs, SMAD2 and 3), which are substrates of the TGFBRs; the Co-SMAD (SMAD4), which contributes to the signaling through binding to the R-SMADs; and the inhibitory SMAD (I-SMAD, SMAD7), which inhibits the cascade through different mechanisms (150) (Figure I14). These TFs are formed by two globular domains - MH1, which binds the SBE, and MH2, which mediates R-SMADs interactions with

TGFBRs, Co-SMADs, cofactors and epigenetic regulators –, located in the N-terminal and the C-terminal part of the protein, respectively, and a linker region that stabilizes the proteins and is submitted to PTMs (151).

As briefly mentioned in the previous paragraph, SMAD proteins interact with cofactors and epigenetic regulators in a specific cell-context manner. This cooperation is necessary for their proper targeting to the regions they regulate, as SBEs sequences are very degenerated and can be commonly found along the genome. Although they can cooperate with many factors, for this work it is especially interesting to review the well-studied coordination between SMAD3 and the KDM JMJD3, which has been reported for endodermal differentiation, epithelial to mesenchymal transition (EMT) and neurodevelopmental processes (66,68,152–154). Given the context in which this project was developed, the latter will be detailed in the subsection 3.3.

### **3.2. Functions of the TGF $\beta$ pathway**

The TGF $\beta$  signaling pathway is implicated in many fundamental cellular processes in vertebrates. Indeed, some of its most relevant functions include the regulation of cell homeostasis, apoptosis, proliferation (exerting a cytostatic role), EMT, regeneration and differentiation (including myogenesis, morphogenesis, osteoblast differentiation or neurodevelopment). This regulatory versatility is enabled by the cell-context cooperativity capability of SMAD proteins, as previously mentioned (155).

Due to the many physiological roles the pathway exerts, its malfunctioning entails major pathological outcomes. It has been mainly related to cancerous processes, where the signaling plays a role as both a tumor suppressor in pre-malignant cells, due to its cytostatic and apoptotic capacity, and as a tumor progressor in malignant cells, because of alterations in SMAD activities (156). Moreover, it also induces cancer progression by regulating EMT processes together with the KDM JMJD3,

which has been shown to promote metastasis through the demethylation of the EMT regulator SNAI1 (153). Apart from cancer, the pathway has also been related to inflammatory diseases such as arthritis, nephritis, myocarditis or Chron's disease (157).

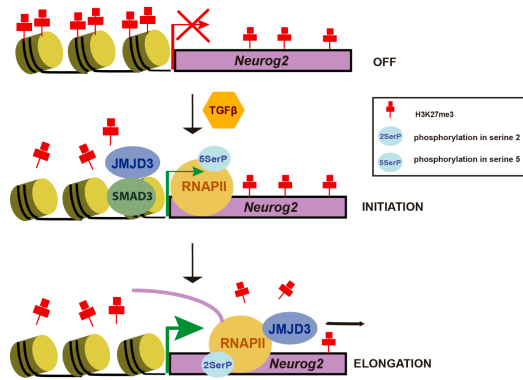
### **3.3. JMJD3 and the TGF $\beta$ pathway in corticogenesis**

Among the many processes in which the TGF $\beta$  pathway is involved, in the lab of Dr Martínez Balbás we are interested in its regulation of the neurogenic transcriptional program. Specifically, we have studied the crosstalk between the pathway and the epigenetic regulator JMJD3 in a corticogenesis context, as both are known to regulate neurodevelopment (158,159). This was investigated *in vivo*, in the cNT, and *in vitro*, in NSCs, where the crosstalk was analyzed at promoters and enhancers regulatory regions.

On the one hand, experiments performed in the cNT demonstrated that TGF $\beta$  promotes neuronal differentiation in a JMJD3-dependent manner *in vivo*. Furthermore, the demethylation of H3K27me3 seemed to play a role in the process (66).

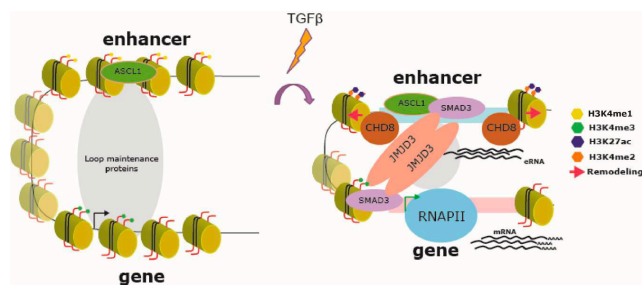
On the other hand, experiments done in NSCs demonstrated that JMJD3 and SMAD3 interact through the MH2 globular domain upon activation of the pathway. Indeed, both proteins widely co-localized at promoters, intragenic and intergenic regions along the genome. This way, it was demonstrated that SMAD3 targets JMJD3 to neural promoters, where the KDM removes H3K27me3 and activates neural genes (66). Moreover, the intragenic regions occupied by JMJD3 were also bound by RNAPII-Ser2p. Indeed, it was shown that the SMAD3-recruited JMJD3 also interacts with the elongating form of the RNAPII and travels along gene bodies, inducing their demethylation, and hence, their transcription (Figure 115). The progression of the RNAPII-Ser2p through the intragenic regions in response to TGF $\beta$  was prevented by JMJD3 depletion. The other way

around, inhibition of the transcription elongation hampered JMJD3 recruitment to intragenic regions (152).



**Figure I15.** Picture showing the crosstalk between the TGF $\beta$  pathway and the epigenetic regulator JMJD3 to regulate the neurogenic program at promoters. Adapted from (35).

Interestingly, JMJD3 binds neuronal enhancers in response to TGF $\beta$  together with pioneer TFs, such as ASCL1, and cofactors, such as the remodeler protein CHD8 (Figure I16). This way, the interaction of the signaling pathway and the epigenetic factor promotes the opening of chromatin, and so the proper activation of the neurogenic program. Contrary to what was previously observed at promoters, at enhancers the role of JMJD3 was independent of its catalytic activity (67).



**Figure I16.** Depiction of the TGF $\beta$  and JMJD3-induced activation complex formed at neural enhancers to promote the neurogenic program. Adapted from (160).

Considering these results, the study of how the interplay between JMJD3 and TGF $\beta$  activates enhancers in a KDM catalytic-independent manner is the starting point of my project.

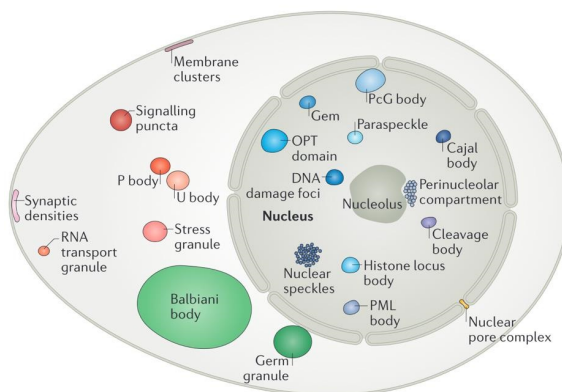
#### **4. Biomolecular condensates**

Given their major role both biologically and for the present work, in this last section of the introduction I will introduce biomolecular condensates.

The densely crowded cellular space needs to be organized to allow spatiotemporal control over the biochemical reactions that take place within it. Apart from classical organelles surrounded by a lipid bilayer, which acts as a physical barrier separating their content from the rest of the cell, this organization can be achieved through the formation of biomolecular condensates. These are defined as micron-scale membraneless compartments that concentrate certain biomolecules, such as proteins and nucleic acids, that are functionally related. This way, by regulating the localization of reaction components the efficiency of the process can be increased, both through their non-stoichiometric accumulation into specific compartments and through the reduction of the effective search space for proteins to find their partners. This compartmentalization property of condensates also provides the cell with a reservoir for components that can form a different temporal condensate in case of special need. All considered, it could be said that condensates share three major features: compartmentalization, selective partitioning, and concentration of the macromolecules within them (161).

The first condensate observed was the nucleolus, found in the nucleus of neuronal cells in the 1830s (162). Nowadays, many examples of biomolecular condensates in the cell can be provided, both for the cytoplasm – such as stress granules (SGs) and germ granules - and the nucleus – such as nucleoli, DNA damage foci or nuclear speckles (163,164)

(Figure I17). Independently of their localization, they comprise many of the regulatory processes that take place in the cell, such as the regulation of chromosome structure and maintenance, DNA replication, transcription, or RNA processing (165).



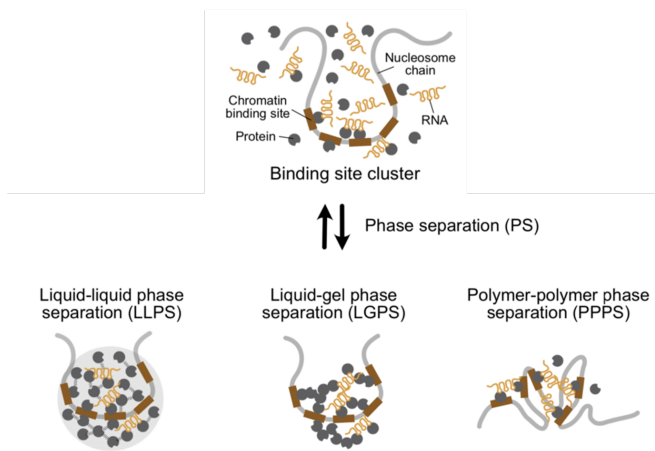
**Figure I17.** Graphic representation of the many condensates that can be found in cellular cytoplasm and nucleus (165).

In the following subsections I will review some of the most compelling aspects of biomolecular condensates, from their emergence to their physiological roles in the cell, as well as their dysfunctions.

#### 4.1. Formation of biomolecular condensates

The above referred to as membraneless compartments or condensates have been named in many ways, such as cellular or nuclear bodies, granules, speckles, aggregates, assemblages, or membrane puncta. We use the term biomolecular condensate to refer to these assemblies that concentrate biological molecules independently of their material state or the physical mechanism through which they arise. In fact, it encompasses both assemblies whose origins are unknown, and compartments formed through phase separation mechanisms (165).

During the development of the present work, we have focused on condensates that arise through phase separation, so in the following subsection I will further review this process. Basically, there are three main mechanisms involved in phase separation processes: liquid-liquid (LLPS), liquid-gel (LGPS), and polymer-polymer phase separation (PPPS) (Figure I18). These processes are not mutually exclusive; in fact, the three of them usually contribute to the formation of different phases within the cell in a way in which they are mixed or changed over time, as seen for the maturation from a liquid to a gel-like or solid state that many condensates suffer (166). It is worth noting that the three mechanisms share a common specific binding to a certain genomic region, involving protein-RNA or DNA-RNA interactions, that constitute nucleation sites (for LLPS and LGPS) or linkers between chromatin segments (for PPPS) (167). In the following subsections the three phase separation processes will be described.



**Figure I18.** Representation of the three main mechanisms of phase separation processes: liquid-liquid, liquid-gel, and polymer-polymer phase separation (167).

#### 4.1.1. Liquid-liquid phase separation

The idea that phase separation could be a common mechanism by which biomolecular condensates are formed came from the discovery of the liquid nature of P granules in germ cells of *Caenorhabditis elegans* – they fuse



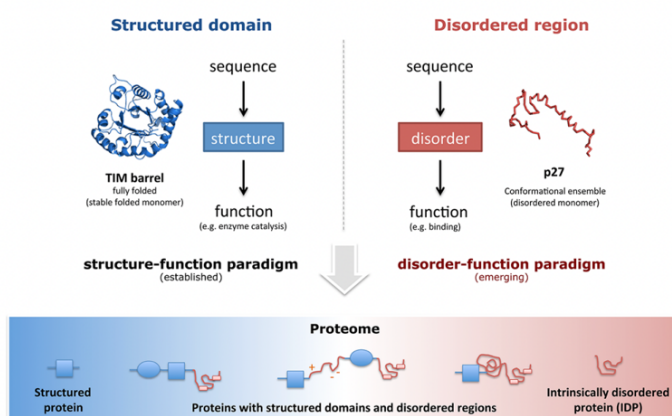
with one another, undergo fission, flow and deform under shear force, and their constituent proteins are highly mobile and exchange with the cytoplasm (168). This was further reinforced through the observation of liquid-like properties in other membranellar compartments, such as nucleoli (169) or stress granules (170).

For a molecule to phase separate it must reach its solubility limit, this is, the threshold concentration for this molecule to be miscible in solution (165). Macromolecules establish low-affinity, short-lived interactions among them and with solvent. Hence, their solubility is a product of the existing balance between macromolecules-macromolecules and macromolecules-solvent interactions. When this equilibrium is shifted toward interactions between macromolecules, they tend to phase separate from the solution in which they were miscible, resulting in two phases that coexist at equilibrium in the system: a small volume, condensed phase, and a large volume, dilute phase (171).

So essentially, the process of LLPS (Figure I17) is promoted by the establishment of low-affinity, dynamic, and multivalent interactions among macromolecules, including weak hydrophobic interactions and protein-protein/protein-nucleic acids electrostatic interactions. The valence of a molecule refers to the number of interactions that it can establish simultaneously; this way, a multivalent interaction involves a network of interactions between a single macromolecule and at least three other factors simultaneously. Multivalent molecules with phase separating behaviors include proteins with structured modular interaction domains (172), proteins with oligomerization domains (173), proteins with intrinsically disordered regions (IDRs) that provide many low-affinity adhesive elements (174), and nucleic acids such as RNA and DNA molecules. Due to the major relevance of IDRs, both biologically and for this thesis, we will further review the role of these regions in condensates formation.

#### 4.1.1.1. Intrinsically Disordered Regions

In the proteome of eukaryotic cells, most proteins contain both IDRs and structured globular domains (Figure I19). IDRs are regions that lack well-defined, fixed three-dimensional structures and present a high conformational flexibility, dynamically fluctuating from coils to collapsed globules, that allows them to promiscuously interact with many targets. Nevertheless, they are functional and actively participate in fundamental cellular functions mediated by proteins (Figure I19) (175). In fact, IDRs possess accessible sites for PTMs deposition that increase the protein functional states. Moreover, they interact with other proteins' structured domains through the exposure of short linear residues that constitute small recognition elements (176,177). It is also worth noting that IDRs present low-sequence complexity regions (LCR), meaning that they show a biased amino acid composition. Particularly, IDRs are usually enriched in glycine, serine, glutamine, proline, asparagine, phenylalanine, tyrosine and charged residues such as lysine, arginine, glutamate, and aspartate. This bias generates repetitive amino acids tracks and blocks of a certain type of amino acid, such as charged, hydrophobic or aromatic ones. In turn, these repeated sequence elements are fundamental for the phase separation role of IDRs (165).



**Figure I19.** Eukaryotic proteins contain both structured and disordered regions that participate in fundamental cellular functions (177).

All the above-mentioned features explain the IDRs condensate-promoting role, as they are prone to engage in low-affinity, dynamic, multivalent interactions. Given this, IDR-containing proteins are crucial for the regulation of signaling pathways and cellular processes such as the cell cycle, transcription, or translation (177,178). This, in turn, is reflected in their enrichment in both cytoplasmic and nuclear biomolecular condensates (165).

#### **4.1.2. Liquid-gel phase separation**

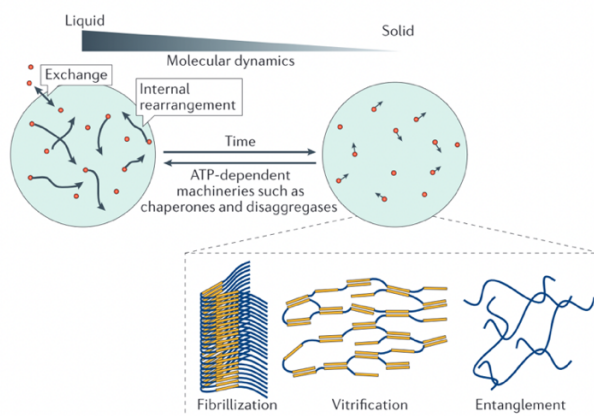
Multivalent proteins involved in phase separation can also undergo LGPS (Figure I17) or gelation processes, which consist in the formation of hydrophilic cross-linked polymer networks through noncovalent interactions between macromolecules. The resulting gel state is termed “hydrogel”, highlighting its high content in water molecules. RNA molecules are major promoters of these processes (167,179).

#### **4.1.3. Polymer-polymer phase separation**

PPPS (Figure I17) consists in the cross-linking of proteins segments that bridge to form ordered collapsed globules (instead of the liquid-like droplets formed by LLPS). PPPS is the typical phase separation mechanism through which chromatin subcompartments are assembled. In this case, bridging proteins dynamically interlink nucleosomes found in close proximity, so that they build an ordered globule or chromatin body that intends to compact chromatin fiber. Representative examples of these bridging factors include histone tails modifications and the proteins bound to them, CTCF and heterochromatin protein 1 (HP1) proteins (180–182).

## 4.2. Physical properties of biomolecular condensates

The physical nature of condensates changes over time, and these properties, in turn, will depict condensates' functions. Most biomolecular condensates possess liquid-like features, this is, they are viscous, they show a round shape due to their reduced surface tension, they are deformable by shear flows, as shown by their fusion and fission capacity, and they can exchange molecules with the soluble phase in a dynamic and rapid-manner, as assessed by fluorescence recovery after photobleaching (FRAP) experiments (179,183,184). However, liquid-like condensates formed by proteins with IDRs can undergo a process called maturation or hardening, by which they become more viscoelastic until becoming gels, glasses, or solid-like condensates (protein aggregates and amyloid fibrils, usually linked to disease) (Figure I20). This has been observed for proteins like FUS, TIA1 and hnRNPA1, or condensates such as Balbiani bodies or yeast SGs (165). The maturation process seems to be promoted by aging, mutations, cellular stress, PTMs or changes in the physicochemical features of the cell, such as in pH, osmolarity and temperature (185).



**Figure I20.** Representation of the maturation process that condensates undergo from liquid to gel or solid-like (165).

As previously mentioned, the physical properties of condensates will determine their functional state. Therefore, the cell possesses mechanisms

to fine-tune the maturation potential of IDRs. These mechanisms involve energy-dependent machinery – chaperones, disaggregases, molecular motors or ATPases - that dynamically regulate the appearance of fibers and crosslinks among condensates' polymers, so that they are promoted when the condensate needs to be static and limited when they need to be dynamic (165).

#### **4.3. Composition of biomolecular condensates**

Although biomolecular condensates share a similar shape, dynamic, and assembly manner, they show differences in their composition. Condensates contain many molecular components, typically from 10 to hundreds of proteins and RNA and DNA molecules, which will change depending on their specific function. Condensate components can be divided into two classes according to their dynamic control, so that some will be constitutive elements – scaffolds - while others will be recruited in a transient fashion – clients. (165).

#### **4.4. Regulation of biomolecular condensates formation through phase separation**

Condensate formation is an orchestrated phenomenon that occurs in the cell in a regulated manner. The assembly of phase separated condensates will be influenced by the factors that alter the solubility threshold of their key constituents. This will depend on environmental and thermodynamical parameters intrinsic to the cell, such as ionic strength, pH, and temperature, and on specific regulatory mechanisms that modify macromolecules concentration and intrinsic solubility (165). In turn, these factors will determine the readiness of macromolecules to establish the multivalent interactions that drive condensates assembly (186).

The intrinsic solubility of condensates' components can be controlled by active processes that will determine condensates formation or dissolution.

Indeed, this is the case for the addition or removal of PTMs to DNA, RNA, and proteins, which will regulate their valency, and so will promote or inhibit phase separation in a context-dependent manner. Molecular chaperones and RNA helicases also actively regulate condensates, seemingly towards a dissolution induced by phase separation inhibition, as observed in stress granules, ribonucleoprotein (RNP) granules and FUS condensates, although the mechanisms underlying these processes are not well understood (187).

Not only active, but also passive mechanisms dictate macromolecules' intrinsic solubility. ATP has been suggested to play such a role, acting as a hydrotrope that destabilize the aggregation of macromolecules by actively keeping a liquid-like state, even out of equilibrium (188,189). Other molecule that seems to play such a role is RNA. Although it has been previously mentioned for favoring condensate formation, it has also been reported to buffer the solubility of RNA-binding proteins (190). Lastly, the dissolution of nuclear components after nuclear envelope breakdown has been suggested as another putative mechanism for condensates disassembly (191).

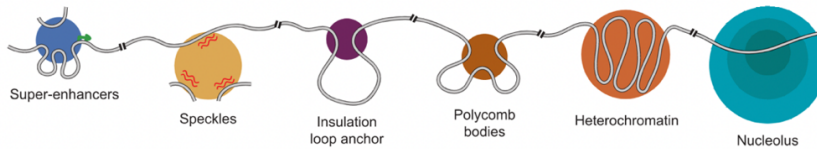
All considered, it is worth noting that several control platforms may interplay in the cell to create a complex network of regulatory mechanisms that will shape the appearance, dissolution, and physical properties of condensates in accordance with specific contexts.

## **4.5. Physiological roles of biomolecular condensates**

### **4.5.1. Nuclear biomolecular condensates**

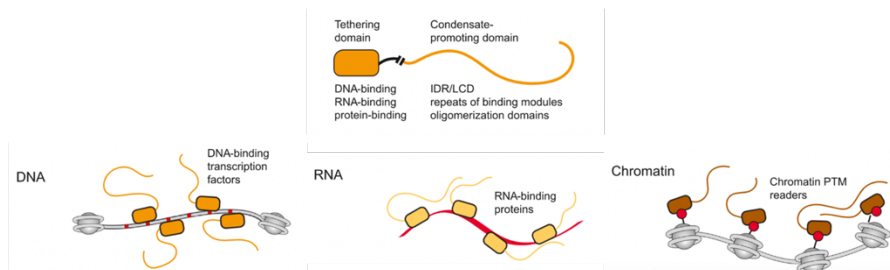
Biomolecular condensates play major roles in the regulation of nuclear processes, including DNA replication, DNA transcription, DNA repair, chromosome structure and maintenance, or RNA processing, among others. Some of the condensates that can be found in the nucleus include

the nucleolus, polycomb bodies, constitutive heterochromatin, Cajal bodies, nuclear speckles and paraspeckles, CTCF clusters, DNA damage foci, DNA replication origins, SEs, or the chromatin itself (Figure I21) (192).



**Figure I21.** Depiction of some representative nuclear condensates implicated in the regulation of nuclear processes (192).

Most nuclear condensates are formed at specific genomic loci. There are many nuclear components that instruct condensates to appear at these locations, such as bifunctional proteins composed of stable structured domains – with which they bind a specific DNA, RNA or protein with high affinity – and condensate-promoting domains, typically IDRs and LCRs. One example of this type of proteins are TFs, which are recruited to binding sites located at enhancers and promoters, until they reach the threshold to concentrate highly disordered coactivators and form transcriptional condensates (Figure I22) (192). Histone proteins also contain a structured domain that binds DNA to form the nucleosome, and an intrinsically disordered tail that establishes weak multivalent interactions, inducing different chromatin states in accordance with the regulation exerted by the PTMs that decorate these tails (Figure I22) (193). The nascent RNA processing machinery is another example of bifunctional proteins, binding both the RNA and the RNAPII phosphorylated CTD through the structured domains, and self-associating through their arginine and serine-rich IDRs (Figure I22) (194).



**Figure I22.** Representation of the bifunctional proteins that drive condensates nucleation, including TFs, RNA-binding proteins and chromatin reader proteins. Adapted from (192).

Apart from bifunctional proteins, the RNA itself can influence the appearance of condensates in specific loci, as it crowds RNA-binding proteins, rich in condensate-promoting domains, that establish RNA-RNA and RNA-proteins interactions.

As previously said, nuclear condensates are involved in the control of many nuclear processes. In the following subsections the regulation processes exerted by nuclear condensates that are most compelling for the scope of this project will be reviewed, including their role on chromatin structure, transcriptional and post-transcriptional regulation.

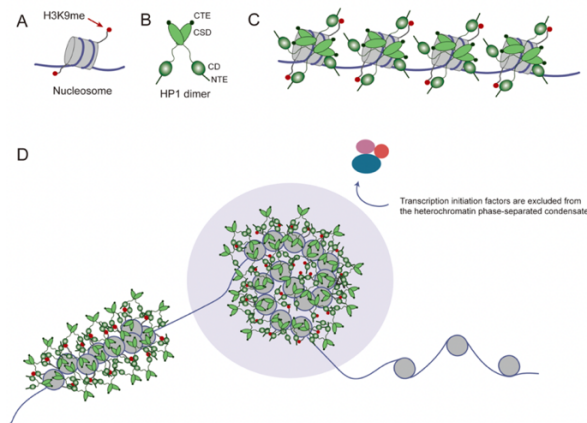
#### 4.5.1.1. Biomolecular condensates and chromatin structure

Apart from PPPS, chromatin has also been showed to undergo LLPS driven by its constituent histones tails, which promote the formation of nuclear bodies. This process is favored through the concentration of nucleosomes into chromatin nuclear bodies by the linker histone H1, consistent with the condensing role of this protein. Furthermore, the prevailing nucleosome spacing pattern –  $10n + 5$  base pairs of DNA – also promotes phase separation. On the other hand, histone acetylation by p300 disassembles chromatin condensates and decreases its phase separation in the nucleus, but multi-bromodomain proteins, such as the transcriptional regulator



BRD4, can reinduce the phase separation potential of its modified chromatin in a way in which a new phase separated state arises (193). Thus, it seems that chromatin condensates can partition the genome so that it self-organizes into subcompartments (167). Phase separation can also regulate the formation or dissolution of loops within chromatin bodies, as observed for KLF4 and UTX-condensates (195,196).

Furthermore, heterochromatin formation has also been linked to phase separation. HP1 $\alpha$ , which is known to assemble heterochromatin, establishes weak interactions with other heterochromatin components and forms condensates that become larger upon recruitment of more HP1 $\alpha$  molecules (197). This, in turn, contributes to the formation of phase separated heterochromatin bodies that exclude the transcriptional machinery (Figure I23) (198).

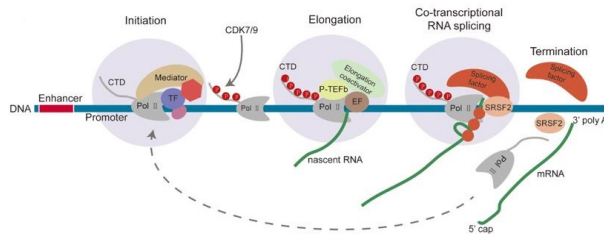


**Figure I23.** Graphic representation of the heterochromatin bodies formation and the factors they exclude (199).

Not only LLPS influences chromatin structure, but also the other way around. Chromatin organization influences the areas where droplets are formed, appearing mainly at loose and low-density regions while the surrounding chromatin shows mechanical resistance (167,193).

#### 4.5.1.2. Biomolecular condensates and transcriptional regulation

As previously commented, the organization of chromatin into different compartments translates into a reorganization of the genome in response to stimuli, which will condition gene transcription. Transcription factories are considered self-assembling organizers of the genome that comprise RNAPII, TFs, DNA cis-regulatory elements, and RNA molecules (200). Some of these components have been shown to undergo phase separation processes driven by their IDRs, giving rise to the so-called transcriptional condensates. This is the case for RNAPII, whose CTD was demonstrated to form a hydrogel nucleated by RNA in a phosphorylation-dependent manner. RNAPII can also be incorporated into the condensates formed by its elongating factor P-TEFb through direct interaction of the CTD with P-TEFb IDR. This way, the phosphorylation status of the polymerase is enhanced, and so is its elongational activity (201,202). Moreover, a model concerning phase separation has been recently proposed to explain transcription. This model points to the CTD phosphorylation as the factor that controls RNAPII shuttle between an initiation condensate, which also contains TFs and coactivators, and an elongation condensate, that is formed by nascent RNA, and elongation and RNA processing factors. Once elongation is completed, hypophosphorylated RNAPII is released from the complex and returned to the initiation condensate (201,203) (Figure I24).



**Figure I24.** Depiction of the phase separation model of the transcriptional process, according to which initiation and elongation distinct condensates exist (199).

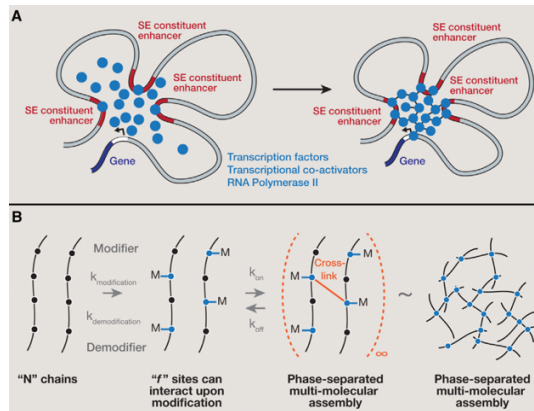
With respect to TFs, they form condensates that regulate gene expression at particular genomic loci through a trans-activator domain-dependent interaction with the DNA, Mediator and other co-activators. This is the case for OCT4, GCN4, MYC, p53, NANOG, SOX2, RAR $\alpha$ , GATA2 and Era. Taking OCT4 as an example, in ESCs it interacts with the MED1 subunit of the Mediator complex to be incorporated into its phase separated condensates on super-enhancers (204). Given the significant role it has for this thesis, the interplay between transcriptional condensates and cis-regulatory regions, particularly enhancers and super-enhancers, is discussed in the following subsection.

#### **4.5.1.2.1. Transcriptional condensates at typical enhancers and super-enhancers**

Enhancers, and especially SEs, are cis-regulatory regions that accumulate high densities of transcription factors and co-activator proteins to promote transcription. As previously mentioned, these molecules show phase separation propensity, so it is conceivable that assemblies likely arise during the formation of SEs and typical enhancers, although less frequently for the latter (Figure I25) (199).

Taking this into account, a model to depict a general cooperative mechanism through which transcriptional machinery is assembled at enhancers, specially at SEs, has been proposed. According to it, the protein and DNA elements present at enhancers and SEs are depicted as chain-like molecules that contain IDRs. These regions are enriched in residues amenable to be chemically modified or demodified, so that they can stablish both specific structure cross-linking interactions that will localize TFs to a particular locus (TF-DNA interactions), and weak multivalent cross-linking interactions (TF-coactivator interactions). This way, the formation of transcriptional condensates at specific genomic loci is promoted when the number of TF-coactivator interactions exceeds the critical concentration threshold (Figure I25). This model, that is compatible with the formation of

transcription factories, explains the assembly at a specific genomic locus by the presence of dense clusters of particular TFs binding sites. Moreover, it sustains the phase separation potential of SEs versus that of typical enhancers, as the firsts are more cooperativity prone than typical enhancers, and this favors phase separation (92,205).



**Figure 125.** Model to explain the emergence of phase-separated transcriptional condensates at SEs based on its intrinsic cooperativity (205).

This cooperativity-based model explains many of the characteristic features of SEs-driven transcription, such as the synchronous bursting activities of more than one promoter activated by the same enhancer, as the condensate would include the enhancer and both promoters; the high frequency of bursting, that correlates with a relatively constant high level of enhancer-driven transcriptional activity; SEs vulnerability to perturbation of their transcriptional cofactors, which would vary the valency of the cluster and may disassemble the separated phase; and, finally, the fidelity of SEs-transcriptional control, as their cooperativity buffers the modifier effect of environmental signals (92,205).

It is worth mentioning that, as commented above, the relative positioning of cis-regulatory regions could explain the dynamism, interconnection and environmental tunability of the transcriptional state of different spatial

regions within the nucleus; the other way around, transcriptional condensate formation has potential implications regarding genome organization, such as the promotion of the compartmentalization of actively transcribed regions (205,206).

#### **4.5.1.3. Biomolecular condensates and post-transcriptional regulation**

The formation of condensates is also associated with RNA splicing. As it is known, RNA splicing can occur co-transcriptionally for genes that are transcribed at especially high rates (207). This way, co-transcriptional RNA splicing condensates have been shown to appear when the RNAPII CTD is hyperphosphorylated, as its association to Mediator and the transcription complex is changed for the splicing machinery (201).

#### **4.5.2. Other biomolecular condensates (cytoplasm, mitochondria, and chloroplasts)**

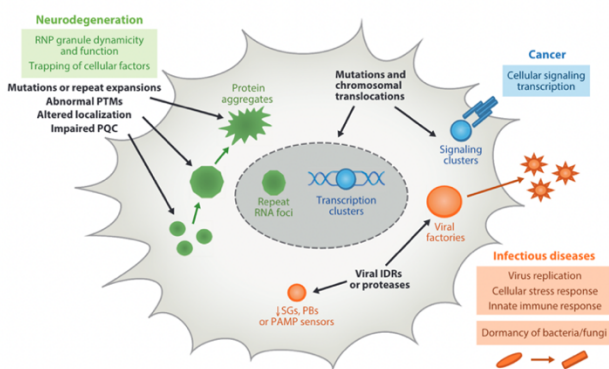
Cytosolic components are also partitioned into condensates, generally named as cytoplasmic messenger ribonucleoprotein (mRNP) granules, which are fundamental for mRNA metabolism and cellular homeostasis. P-bodies, SGs, germ granules, centrosomes and neuronal RNA granules are the main representative mRNP granules (208,209). Furthermore, RNA granules have also been found within some organelles, such as mitochondria and chloroplasts (209).

#### **4.6. Dysregulation of biomolecular condensates**

Dysregulation of phase separation leads to aberrant protein aggregates formation, which have been linked to many diseases such as neurodegeneration, cancer, and infections (Figure I26). The molecular mechanisms behind these “pathological assemblies” are still unclear, but

they involve the self-assembly of misfolded protein species into insoluble aggregates that can be amorphous, filamentous, or amyloid like (210).

Neurodegenerative diseases constitute the most prominent example of dysregulated phase separation processes. Indeed, aberrant protein aggregates have been observed in Alzheimer's, amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), Parkinson's and Huntington disease (Tau, TDP-43, FUS,  $\alpha$ -synuclein and polyQ aggregates, respectively) (211,212) (Figure I26). Apart from neurodegeneration, cancer has also been linked to pathological condensates, as demonstrated for many mutant forms of TP53, UTX or SPOP, and for the fusion proteins PML-RAR $\alpha$ , EWS-FLI1 and FUS-CHOP (213). Regarding infectious diseases, both viral factories and bacterial and fungal dormant cells constitute phase separated entities (212,214) (Figure I26).



**Figure I26.** Depiction of the involvement of phase separated molecular condensates in disease, including neurodegeneration, cancer, and infections (212).

## Objectives

The thesis objectives resulted from previous work from Dr Martínez-Balbás lab that characterized the cooperation between TGF $\beta$  and JMJD3 to regulate transcription at enhancers, promoters, and gene bodies (66,67,152). Taking into consideration these works, described in the introduction, as well as the current bibliography, the following goals were proposed for this PhD thesis to deepen into the role of histone demethylases as transcriptional regulators:

1. To elucidate the mechanism underlying JMJD3 and TGF $\beta$ -driven enhancer activation.
  - To analyze the changes in the three-dimensional structure of the chromatin triggered by the TGF $\beta$  pathway activation, as well as the contribution of JMJD3 to these changes.
  - To examine the mechanism by which TGF $\beta$ -JMJD3 axis is promoting structural changes at enhancers.
2. To determine the existence of a common mechanism used by JmjC-containing lysine demethylases (JmjC-KDMs) to promote transcriptional regulation.





# Materials and methods

In this section of the manuscript, I will provide insight into the methodology followed for the experimental work, bioinformatic and statistical analysis performed during the thesis.

## 1. Materials

### 1.1. Plasmids

All plasmids used during this doctoral thesis, as well as their origin and the experiments in which they were involved, are listed in the below table (Table M1).

Plasmid	Origin/Provider	Experiment
pCMV-VSVG	Dr Timothy Thomson	Virus production vector
pCMV-GAL-POL	Dr Timothy Thomson	Virus production vector
pLKO.1-shJMJD3 (CCTCTGTTCTTGAGGGACAAA)	Sigma	Virus production vector
pLKO.1-shSMAD3 (CCTTACCACTATCAGAGAGTA)	Sigma	Virus production vector
pLKO.1-random (CAACAAGATGAAGAGCACC)	Sigma	Virus production vector
pGL3 basic	Promega #E1761	Luciferase vector
pGL3 basic- <i>Chst8</i> Promoter (pGL3- <i>Chst8 P</i> )	Dr Raquel Fueyo	Luciferase vector
pGL3 basic-Viewpoint (pGL3-VP)	Dr Raquel Fueyo	Luciferase vector
pGL3-Viewpoint- <i>Chst8</i> Promoter (pGL3-VP- <i>Chst8 P</i> )	Cloned in this study	Luciferase vector
pGL3-E1- <i>Chst8</i> Promoter (pGL3-E1- <i>Chst8 P</i> )	Cloned in this study	Luciferase vector
pGL3-E2- <i>Chst8</i> Promoter (pGL3-E2- <i>Chst8 P</i> )	Cloned in this study	Luciferase vector
pGL3-E3- <i>Chst8</i> Promoter (pGL3-E3- <i>Chst8 P</i> )	Cloned in this study	Luciferase vector
pGL3-E3-Viewpoint - <i>Chst8</i> Promoter (pGL3-E3-VP- <i>Chst8 P</i> )	Cloned in this study	Luciferase vector
pGL3-E3-E1- <i>Chst8</i> Promoter (pGL3-E3-E1- <i>Chst8 P</i> )	Cloned in this study	Luciferase vector
pGL3-E2-E3- <i>Chst8</i> Promoter (pGL3-E2-E3- <i>Chst8 P</i> )	Cloned in this study	Luciferase vector
pR-TK Renilla	Promega	Luciferase vector
pDONR-JMJD3	Dr Sussanne Mueller	Gateway® Recombination Cloning Technology vector

pDONR-JMJD3 ΔIDR	Dr Raquel Fueyo	Gateway® Recombination Cloning Technology vector
pInducer20	Dr Naiara Akizu	Gateway® Recombination Cloning Technology vector
pInducer -JMJD3 WT	Dr Raquel Fueyo	Gateway® Recombination Cloning Technology vector
pInducer-JMJD3 HE>AA (H1390A/E1392A)	Dr Raquel Fueyo	Gateway® Recombination Cloning Technology vector
pInducer -JMJD3 ΔIDR	Dr Raquel Fueyo	Gateway® Recombination Cloning Technology vector
pCMV-mEGFP-C1	Dr Anna Aragay	Overexpression vector
pCMV-HA-JMJD3	Dr K. Helin	Overexpression vector
pCMV-HA-mEGFP-JMJD3	Cloned in this study	Overexpression vector
pCMV-HA-mEGFP-JMJD3 ΔIDR	Cloned in this study	Overexpression vector
pCIG-JMJD3 HE>AA (H1390A/E1392A)	Dr Conchi Estarás	Overexpression vector
pCMV-HA-mEGFP-JMJD3 HE>AA (H1390A/E1392A)	Cloned in this study	Overexpression vector
pCMV-mCherry-C1	Dr Anna Aragay	Overexpression vector
pcDNA-mCherry-MED15	Dr Rick Young	Overexpression vector
p3xFLAG-PHF2	Dr Jiemin Wong	Overexpression vector
p3xFLAG-PHF2 ΔCharged.	Cloned in this study	Overexpression vector
pCMV-mCherry-PHF2	Cloned in this study	Overexpression vector
pCMV-mCherry-PHF2 ΔCharged	Cloned in this study	Overexpression vector
pCDNA-FLAG-EGFP-KDM2A	Dr Till Bartke	Overexpression vector
pCDNA-KDM4B-EGFP	Dr Thomas Jenuwein	Overexpression vector

**Table M1.** List of plasmids, their origin and the type of experiment for which the vector was used. The sequence in brackets of the pIKO.1 plasmid correspond to its target.

## 1.2. Antibodies

The following table (Table M2) gather both primary and secondary antibodies used during this study, providing information about their targets, references, and dilution for the different applications in which they were used.

	Antibody target	Provider and reference	Dilution for the indicated application
Primary antibody	JMJD3 (serum)	Raised in our lab	Immunofluorescence 1:200; Western Blot 1:1000
	JMJD3	Abcam, ab38113	Immunofluorescence 1:200
	PHF2	Cell Signaling, D45A2	Immunofluorescence 1:250; Western Blot 1:1000
	HA	Abcam, ab20084	Western Blot 1:5000
	GFP	Roche, 11814460001	Immunofluorescence 1:500; Western Blot 1:500
	mCherry	Thermo Fisher, MA5-32977	Immunofluorescence 1:500; Western Blot, 1:1000
	H3K9me3	Abcam, ab8898	Immunofluorescence 1:250
	H3K27me3	Millipore #07449	ChIP 1:500
	H3K36me3	Abcam, ab9050	Immunofluorescence 1:250
	$\beta$ -TUBULIN	Millipore #MAB3408	Western Blot 1:5000
	Unspecific IgGs	Diagenode, C15410206	ChIP, same dilution than the specific IgG
Secondary antibody	Anti-Rabbit Fluor 555	Invitrogen A32727	Immunofluorescence 1:1000
	Anti-Mouse Fluor 488	Invitrogen A32731	Immunofluorescence 1:1000
	Anti-Rabbit IgG IRDye	LI-COR #926-32221	Western Blot 1:5000
	Anti-Mouse IgG IRDye	LI-COR #926-32210	Western Blot 1:5000

**Table M2.** List of primary and secondary antibodies used in this study. Providers, references and dilutions of usage are indicated.

### 1.3. Primers

Along this thesis, we have used primers for experimental procedures that require DNA amplification, such as RT-qPCRs (retrotranscription followed by quantitative polymerase chain reaction) or conventional PCR (polymerase chain reaction) for the cloning of luciferase reporters and overexpression vectors (Table M3), and for the 4C-seq experiments (Table M4).

	Region	Forward primer (FW)	Reverse primer (RV)	Restriction site
cDNA	<i>mJMJD3</i> mRNA	CTCGTTCTGAGTCTGAGGT	CCCGCTCTACAAGGGTC	-
	<i>Nrip3</i> mRNA	CTAGCTTCCAGTCTAACCAAA	TCTATTCCTCTACCACACTTCAC	-
	<i>Chst8</i> mRNA	CTTTCCCAGGTTCAAGGAC	GGTCCGAGAAAGGTTTGGAG	-
	<i>Neurog2</i> mRNA	CACAACCTAAACGCCGC	TCTTCGTGAGCTTGGCAT	-
	<i>Pepd</i> mRNA	CGAGGGCATTAGCAAGTTCA	AGCGCAAACCTCTAGTTCC	-

	<i>Fabp4</i> mRNA	TGAGTCCCCCACTTGCTTTA	CACCCTGTAAGGCTGGTGAT	-
	<i>Nrip3</i> eRNA	GGGAACACCAGGAAGC	AGCCACACTTGATACGG	-
	<i>Chst8</i> VP eRNA	TAGGAGTGGACCAGTTAGC	CTTCTCTCCCACCGTCAA	-
	<i>Chst8</i> E1 eRNA	GAGACAGTCGCCTTCTGTGC	TGGGATCTTGATGGCTACCT	-
	<i>Chst8</i> E2 eRNA	GGTTGTGGTCAGGGTG	ACTGGGCAAAGAATCTATGG	-
	<i>Chst8</i> E3 eRNA	GAGCAGTTCACCTAGATATT	TGTAAGTGTAGGATACATGG	-
	<i>Chst8</i> E4 eRNA	TGACTCACTGGCTTGGTCAG	TGTCTGGCTGCAGGTCATAG	-
	<i>Fabp4</i> eRNA	TGAGTCCCCCACTTGCTTTA	CACCCTGTAAGGCTGGTGAT	-
	<i>E2f3</i> mRNA	GGCCATTGAGGTTTACTTG	ACCGAGCAGTCACTATGTC	-
	<i>Pcna</i> mRNA	GGGTGAAGTTTTCTGCAAG	GCAAACGTTAGGTGAACAGG	-
	<i>Gapdh</i> mRNA	ATGTTCTGCATGGGTGTG	CCTTCCACGATACCAAAGTTG	-
	<i>Rps23</i> mRNA	CGTCAGGGTGCAGCTCATTA	GGCACGAACGCTGTGATCTT	-
ChIP	<i>Chst8</i> promoter	ACTACAGTACGCGTGAATT	GAGTCCCGTTTCTACACTTG	-
	<i>Hoxd8</i> promoter	CAGTCTCTGGCAGTTCTTT	CCTGTCTGTGCTTAACG	-
	<i>Ctgf</i> enhancer	TCACGCTGCTCCCTT	CTCTTGCTTCTCTGCTATG	-
	<i>Chst8</i> VP	TAGGAGTGGACCAGTTAGC	CTTCTCTCCCACCGTCAA	-
	<i>Chst8</i> E1	GAGACAGTCGCCTTCTGTGC	TGGGATCTTGATGGCTACCT	-
	<i>Chst8</i> E2	GGTTGTGGTCAGGGTG	ACTGGGCAAAGAATCTATGG	-
	<i>Chst8</i> E3	GAGCAGTTCACCTAGATATT	TGTAAGTGTAGGATACATGG	-
	<i>Chst8</i> E4	TGACTCACTGGCTTGGTCAG	TGTCTGGCTGCAGGTCATAG	-
Cloning	<i>Chst8</i> E1	ATAACGCGTAGTGGCGTCCTGTCTGTAGC	GCTACGCGTCCCAGAGCCTGTCTCTAGCT T	MluI
	<i>Chst8</i> E2	ATAGCTAGCCGGGTCTTAGGGTCTGAAAC	GCTGCTAGCCCCATATGTTCCCACTCCAA	NheI
	<i>Chst8</i> E3	ATAGGTACCGGCTGCATAACCTAGGCAAA	GCTGGTACCGCACACTACCTGTGCCTTCA	Acc65I
	mEGFP	AGCTGAGCAAAGACCCCAACG	TGGACTGGGTGCTCAGGTAGT	Acc65I
	<i>Phf2</i> ΔCharged	GAGAAGGAAGAACCTGACTCGTACTGAAGA T	TAACGAGTCAGGTTCTTCTCTCCCAGGTC TC	BamHI

**Table M3.** Primers used during the study and their application.

Region	Primer
FW <i>Chst8</i> shC 0h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTACAGTGGGTTCCCCGTCATTCATG
FW <i>Chst8</i> shC 3h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGTGAAGGTTCCCCGTCATTCATG
FW <i>Chst8</i> shJMD3 3h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCTTGGGTTCCCCGTCATTCATG
FW <i>Chst8</i> shC 0h Replicate 2	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGCCGGTTCCCCGTCATTCATG
FW <i>Chst8</i> shC 3h Replicate 2	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCGTACTGGTTCACCGTCATTCATG
FW <i>Chst8</i> shJMD3 3h Replicate 2	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTTAAGGGGTTCCCCGTCATTCATG
FW <i>Chst8</i> shJMD3 + JMD3 3h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTAGGGGGTTCACCGTCATTCATG
FW <i>Chst8</i> shJMD3 + JMD3 3h Replicate 2	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTTGAACGGTTCACCGTCATTCATG
RV <i>Chst8</i>	CAAGCAGAAGACGGCATACGACCAGCTGGTTAACACAGGA
FW <i>Ldlrad4</i> shC 0h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTACAGTGAGAACCACAGAGTAGGCATG
FW <i>Ldlrad4</i> shC 3h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGTGAAGAACCACAGAGTAGGCATG
FW <i>Ldlrad4</i> shJMD3 3h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTTTGAGAACCACAGAGTAGGCATG
FW <i>Ldlrad4</i> shC 0h Replicate 2	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTGCCAGAACCACAGAGTAGGCATG
FW <i>Ldlrad4</i> shC 3h Replicate 2	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTCGTACTAGAACCACAGAGTAGGCATG
FW <i>Ldlrad4</i> shJMD3 3h Replicate 2	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTTAAGGAGAACCACAGAGTAGGCATG
RV <i>Ldlrad4</i>	CAAGCAGAAGACGGCATACGACTTGCTCAAGTCGCTCGC
FW <i>Aopep</i> shC 0h Replicate 1	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTCCGATCTACAGTGTACAGTATGGGAATCATG

FW <i>Aopep</i> shC 3h Replicate 1	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT <b>GTGAATCACAGTATGGGAATCATG</b>
FW <i>Aopep</i> shJMD3 3h Replicate 1	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT <b>CTTGTCACAGTATGGGAATCATG</b>
FW <i>Aopep</i> shC 0h Replicate 2	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT <b>GCCTCACAGTATGGGAATCATG</b>
FW <i>Aopep</i> shC 3h Replicate 2	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT <b>CGTACTTCACAGTATGGGAATCATG</b>
FW <i>Aopep</i> shJMD3 3h Replicate 2	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT <b>TAAGGTCACAGTATGGGAATCATG</b>
RV <i>Aopep</i>	CAAGCAGAAGACGGCATA <b>CGACATCACTACGGGAAAGTCA</b>

**Table M4.** Primers used in the 4C-seq assay.

## 2. Experimental methods

### 2.1. Cell culture

Three different cell lines have been used in the study: mouse neural stem cells (NSCs), HEK293T and NIH3T3 cells. I will describe methods for working, manipulating, and treating the three of them. All manipulations were done under sterile conditions.

#### 2.1.1. Cell culture growth and maintenance

##### 2.1.1.1. Mouse neural stem cells (NSCs)

Mouse NSCs were dissected from cerebral cortices of C57BL/6J mouse fetal brains (E12.5) and cultured in poly-D-lysine (Millipore, A-003-E) (5 ug/ml, 2 hours 37°C) and laminin (Sigma, L2020) (5 ug/ml 37°C, 4 hours 37°C) precoated dishes following the previously assessed protocols (215). These cells were cultured in a medium prepared by mixing equal parts of DMEM F12 (without Phenol Red, Gibco, 31331093) and Neural Basal Media (Gibco, 12348-017), to which 1% Penicillin/Streptomycin (Gibco,

15140-122) 1% Glutamax (Gibco 35050061), N2 and B27 supplements (Gibco, 17502-048 and 17504-044 respectively), 1 mM sodium pyruvate (Gibco, 11360039), 0.1 mM non-essential amino acids (Gibco #11140035), 2 mg/l Heparin (Sigma #H-4784), 5 mM Hepes (Gibco, 15630056), 25 mg/l bovine serum albumin (Sigma #A7906) and 0.01 mM  $\beta$ -mercaptoethanol (Gibco #31350-010) were added (216).

NSCs are really sensitive so, care should be taken when manipulating them, both at the mechanic and the chemical level. Prolonged pipetting and temperature changes are not recommended in order to maintain a stable epigenome. NSCs were usually expanded in a ratio 1:5 every two days to avoid them getting confluent. To detach these cells the Accutase enzyme (Gibco, A11105-01) was added after removing the medium and rinsing with PBS (Phosphate Buffered Saline). This enzyme is a milder detacher than trypsin, so it is more adequate for the extremely sensitive nature of NSCs. After 2 minutes at 37°C, the Accutase was diluted with PBS and cells were centrifuged at 172 g for 5 minutes. Cell pellet was gently flicked and then diluted in previously warmed expansion medium supplemented with fresh growth factors EGF (Vitro, 236-EG-200ug) and FGF (Invitrogen, PHG0021), which help NSCs to maintain the ability to self-renew and originate differentiated neural cell types (217). Then, medium containing the cells was split in precoated plates.

#### **2.1.1.2. HEK293T and NIH3T3**

HEK293T cells were derived from human embryonic kidney and transformed with the large T antigen of the SV40 virus (218). NIH3T3 cells were derived from desegregated NIH Swiss mouse embryo fibroblasts and spontaneously become immortalized. 3T3 means “3-day transfer, inoculum  $3 \times 10^5$  cells, which corresponds to the protocol with which the immortal cell line was stabilized in cell culture (219).

Contrary to NSCs, both of these cells are very resistant and should be repeatedly pipetted to avoid cellular clumps. Both HEK293T and NIH3T3 cells were cultured in DMEM (Gibco, 41965-062) supplemented with 10% of fetal bovine serum (FBS) (Gibco, 10270106) and 1% of Penicillin/Streptomycin (Gibco, 15140-122) (220). They were usually expanded in a ratio 1:8 every two days, as there is no issue in getting them confluent. After removing medium and rinsing with PBS, trypsin enzyme (Sigma, T9935) diluted with EDTA was added to the plate for 1 minute at 37°C to dissociate cells. Then its activity was stopped by adding two volumes of medium and cells were diluted according to the desired split ratio.

### **2.1.2. Cells freezing and thawing**

Cells can be stored in cryotubes for weeks at -80°C and for years in liquid N<sub>2</sub>. The freezing protocol was the same as the one followed to expand cells but diluting them in medium containing 10% of sterile DMSO instead of normal medium. Then a progressive descent in temperature was achieved by keeping cells at -80°C in a box that contains isopropanol called Mr. Frosty (Nalgene, 5100-0001). The alcohol makes the temperature to lower at a 1°C/minute rate, allowing an optimal freezing process in which cells don't suffer so much.

The thawing process must be rapidly performed to avoid cell death, so the cryotube with the desired cells was introduced in a 37°C water bath until it reached a liquid state. Then the content of the vial was diluted in PBS and centrifuged at 172 g for 5 minutes to eliminate the DMSO previously used for freezing. Finally, cell pellet was diluted in the expansion medium and plated into pre-coated plates.



### **2.1.3. Genetic manipulation of growing cells**

Cells can be manipulated through the introduction of exogenous DNA that will impede the translation of cellular mRNAs or that code for certain proteins. Now I will describe the procedures performed to deliver DNA into cells.

#### **2.1.3.1. Calcium phosphate transfection**

In this protocol exogenous DNA is delivered to cells through an endocytic process of calcium phosphate precipitates that contain the target vector stuck on the surface. The way in which these precipitates were generated was by mixing HEBS (250 mM NaCl, 9 mM KCl, 1.5 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM glucose and 50 mM Hepes pH 7.12), 0.25 M CaCl<sub>2</sub> and the desired DNA vector under vortexing conditions. After 10 minutes at room temperature precipitates formed and thus the mixture was added to the growing medium drop by drop. This medium was replaced for fresh one after 6 hours, so that cells did not stress due to the acidic pH change.

#### **2.1.3.2. Lentiviral transduction**

This is a really highly efficient system in NSCs – up to 90% of infected cells achievement- so it is the one that has been used for knocking down proteins in our cells through shRNAs.

In this case the delivery is performed by lentiviral particles, which were produced in HEK293T cells by calcium phosphate cotransfecting the DNA encoding the target shRNA (Table M1 list the ones used in this study), a control random shRNA, pCMV-VSVG and pCMV-GAG-POL plasmids which encode the viral capsid and transcriptional machinery respectively. 24 hours later, supernatants containing viral particles were collected and concentrated in a sucrose bed by ultra-centrifugation at 57.000 g for 2 hours. Once concentrated, supernatant was removed and viral particles

were resuspended into NSCs medium, which was later added to growing NSCs. About  $1.5 \times 10^6$  cells were infected by production. 24 hours after infection, cells were selected with an appropriate antibiotic -in our case 2ug/mL puromycin (Sigma, P8833) for pLKO.1 plasmid-, for 48 hours.

#### **2.1.4. Cell treatments**

##### **2.1.4.1. Doxycycline cell induction**

As mentioned above cells can be manipulated by introducing exogenous DNA that code for a particular protein of interest. It is a possibility that the expression of this protein depends on the addition of a certain drug, as is the case for some of the overexpression plasmids used in this thesis (see Table M1), which rely on doxycycline adding.

NSCs stable lines containing the DNAs of interest were established by lentiviral transduction. In the case of HEK293T cells we just transfected these plasmids. In both cases, once cells were in growing conditions, doxycycline hyclate (Millipore, 324385-1GM) was added at a concentration of 1 ug/ml. After one day of induction cells were fixed or collected to perform sequential experiments (immunofluorescence, RT-qPCR or Western blot).

##### **2.1.4.2. Transforming Growth Factor $\beta$ (TGF $\beta$ ) treatment**

To activate the TGF $\beta$  pathway, this cytokine was added to the culture medium of NSCs at a concentration of 5 ng/ml for different periods of time (mainly 3 and 6 hours, enough time to observe effects at both structural and transcriptional level). TGF $\beta$ 1 was purchased from Millipore (GF111).

##### **2.1.4.3. GSK-J4 HCl inhibitor**

GSK-J4 HCl (Selleckchem, S7070) is a cell permeable drug that selectively inhibits the catalytic activity of JMJD3 and UTX H3K27me3 histone

demethylases. Specifically, we used it to analyze the dependency of our findings on the catalytic activity of JMJD3, at a concentration of 1  $\mu$ M for 6 hours following the manufacturers' recommendations.

#### **2.1.4.4. 1,6-Hexanediol treatment for live imaging cells**

1,6-Hexanediol is a chemical compound that can be used to disrupt phase separated condensates. In this study, we used it to assess the phase separated nature of the observed nuclear condensates through live imaging. HEK293T cells were transfected with HA-mEGFP-JMJD3 vector and grown in 1.5ml of DMEM media. They were imaged in several selected fields before treatment to establish a baseline. After the fifth acquisition, 1,6-Hexanediol (1,6-HD, Sigma, 240117) was added to cells at a final concentration of 6% in growing media. Cells were imaged for as much as 20 minutes of treatment before they start dying.

## **2.2. Molecular biology procedures**

### **2.2.1. Nucleic acids-related**

#### **2.2.1.1. Genomic DNA extraction**

Along this thesis it was necessary to extract genomic DNA from cells for several subsequent applications. This way, 200  $\mu$ l of buffer (10 mM Tris-HCl, 10 mM EDTA, 10 mM NaCl and 0.5% SDS) were used to lyse about  $6 \times 10^6$  cells. Lysates were incubated with 0.5 mg/ml of proteinase K (Sigma, P2308) during 1 hour at 50°C and with 1 mg/ml of RNase A (Fermentas, EN0531) during 2 hours at 50°C. To finally purify DNA from lysates, a phenol-chloroform extraction was performed, which will be explained in the following subsection.

### **2.2.1.2. Phenol chloroform extraction and ethanol precipitation**

Getting advantage of the different affinity of DNA and proteins for phenol, this technique can be used to purify DNA from complex protein-DNA mixtures.

The aqueous and the organic phase of the mixture were separated by adding 1 volume of phenol and centrifuging the mixture at 16.392 g for 3 minutes. The aqueous phase, which contains the DNA, was transferred to a clean tube. This step was repeated but using chloroform instead of phenol. Then, the DNA in the aqueous phase was ethanol precipitated so that it was concentrated and desalted. 0.1 volumes of NaAc 3 M and 1 volume of cold ethanol were added to the sample, that was incubated at -80°C for at least 20 minutes to favour precipitation. Then it was centrifuged at 16.392 g during 20 minutes at 4°C and washed with ethanol 70%. The pellet was dried and resuspended in a suitable buffer.

### **2.2.1.3. RNA extraction and DNase treatment**

Not only DNA, but also cellular RNA was necessary to perform experiments during this work. To this end, 1 ml of Trizol reagent (Invitrogen, 15596018) was used to extract the RNA from  $3 \times 10^6$  cells. Then, 200  $\mu$ l of chloroform were added and after centrifugation at 16.392 g for 5 minutes the upper phase of the tube was collected and precipitated by adding 800  $\mu$ l of isopropanol. It was centrifuged again at 16.392 g for 10 minutes. The supernatant was discarded, and pellet washed with 70% ethanol. Pellet was finally resuspended in H<sub>2</sub>O after a last centrifugation of 5 minutes at 16.392 g.

A treatment with DNase was performed after RNA extraction to avoid contamination with genomic DNA. We followed the protocol of the DNA-free Kit (Ambion, AM1906). After adding 0.1 volumes of 10X buffer to the RNA sample, it was incubated with 1  $\mu$ l of DNase during 30 minutes at 37°C. 0.1

volumes of inhibition agent were then added to stop reaction, and after 2 minutes of flicking pure RNA could be purified by centrifuging samples at 10.000 g for 1.5 minutes and transferring the supernatant to a new tube.

RNA was finally quantified using a Nanodrop device and quality was evaluated with 260/280 and 260/230 ratios. Also, its integrity was checked by agarose gel.

#### **2.2.1.4. Retrotranscription of RNA into cDNA**

In order to check transcription in a quantitative manner we have performed RT-qPCR assays. Prior to this assay purified RNA was retrotranscribed into cDNA.

Up to 1 µg of RNA (both mRNA and eRNA) were reversed transcribed to obtain cDNA using the High Capacity cDNA reverse transcription kit (Invitrogen, 4368814). RNA was incubated with random hexamers or deoxynucleotides (dNTPs) and a retrotranscriptase in a thermocycler following cycles of 25°C 10 minutes, 37°C 120 minutes and 85°C 5 minutes.

#### **2.2.1.5. qPCR**

qPCR has been used to quantify the retrotranscribed cDNA. Reactions were performed in a LightCycler 480 (Roche, 4887352001) with the following conditions: 95°C 5 minutes, 40 cycles of 95°C 5 minutes, 60°C 10 seconds, 72°C 20 seconds, melting curve of 95°C 5 seconds and 65°C one minute. They were set-up in a 96-well plate to a final volume of 10 µL containing SYBR Green (Roche, 4887352001) and specific primer pairs (see Table M3) both for mRNA – in which case they were designed spanning exon-exon junctions – and eRNA – designed against the specific coordinates of the enhancer. The experiment was carried out in a QuantStudio 5 Real-Time PCR system (ThermoFisher Scientific).

Non-template controls and standard curves with every new primer pair were run, and only primers with an efficiency of at least 95% were kept. The  $2^{-\Delta\Delta CT}$  method was the one used to analyse qPCR data, and normalization was done using the genes *Gapdh* or *Rps23* as a reference.

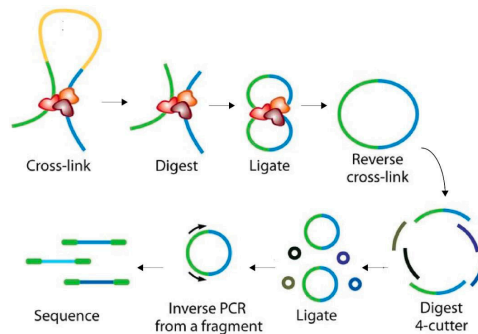
#### 2.2.1.6. PCR

Conventional PCR has been used in this thesis to amplify genomic regions for cloning purposes.

Genomic DNA was incubated in a thermoblock with dNTPs, buffer, specific primers (Table M3) and the enzyme Taq DNA polymerase (Fermentas EP0401). The temperatures and incubation times used were specific to the target region.

#### 2.2.1.7. 4C-seq assay

This chromatin conformation capture assay is meant to provide information about the regions of the genome that contact a particular region of the researcher's interest, named viewpoint (Figure M1).



**Figure M1.** Picture depicting the 4C-seq protocol. Adapted from (221).

The 4C-seq experimental protocol was based on (222,223) (Figure M1).  $12 \times 10^6$  mouse NSCs were fixed for 30 minutes using 1% of formaldehyde at room temperature. The fixing reaction was quenched with glycine 0.125 M

for 10 minutes. After 2 washes with PBS, cell pellets were resuspended in 5 ml of cytoplasmic lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100 and protease inhibitors) during 10 minutes on ice. Lysates were centrifuged for 5 minutes at 650 g and 4°C. Nuclei were resuspended in 0.5ml of NlaIII buffer with 0.3% SDS and they were incubated at 37°C and 900 rpm for one hour. After that, Triton X-100 was added to a final concentration of 2% followed by 1 hour of incubation at 37°C and 650 g. Next, DNA was digested overnight at 37°C and 650 g with 400U of NlaIII, which was afterward inactivated by adding SDS to a final concentration of 1.6% and incubating for 20 minutes at 65°C and 650 g. The digested chromatin was transferred to 50 ml tubes and 6.125 ml of 1.15X ligation buffer (50 mM Tris-HCl pH 7.6, 10 mM MgCl<sub>2</sub>, 1 mM ATP, 1mM DTT) and 1% of Triton X-100 were added and incubated during 1 hour at 37°C and 650 rpm. Digested chromatin was ligated with 100U of T4 DNA ligase for 8 hours at 16°C and then, treated with RNase A 1mg/ml for 45 minutes at 37°C. Decrosslinking step was performed by adding 1mg/ml of proteinase K and incubating at 65°C overnight. DNA was purified by standard phenol-chloroform extraction followed by ethanol precipitation and resuspended in 100 ul of H<sub>2</sub>O. Then, proper digestion and ligation were evaluated by visualizing the DNA in an agarose gel. If they were fine, a second digestion with 50U of DpnII was performed at 37°C overnight, after which enzyme and buffer were removed from the sample by phenol-chloroform extraction followed by ethanol precipitation. Once purified, DNA samples were resuspended in 500 ul of H<sub>2</sub>O. Then a second ligation was carried out by adding 200U of T4 DNA ligase in a final volume of 14ml of 1X ligation buffer. The mixture was incubated overnight at 16°C and after the last round of phenol-chloroform extraction and ethanol precipitation, the DNA was resuspended in 100 µl of H<sub>2</sub>O and purified with a Qiagen PCR purification column. Through DNA electrophoresis, the efficiencies of the second digestion and ligation were tested. Whenever the sample was properly digested and ligated, it constituted the 4C-library sample.

For *Chst8*, *Ldlrad4* and *Aopep* indicated viewpoints, inverse PCR reactions were performed with the Expand Long template PCR system (Roche, 11681834001) with the following cycling conditions: 94°C 2 minutes, 29 cycles of 94°C 10 seconds – 55°C 1 minute – 68°C 3 minutes and 68°C 5 minutes (for *Chst8* and *Ldlrad4*) and 94°C 2 minutes, 31 cycles of 94°C 10 seconds – 55°C 1 minute – 68°C 3 minutes and 68°C 5 minutes (for *Aopep*). Primer sequences are described in Table M4. The products of these reactions were sequenced in the Erasmus Center for Biomix in Rotterdam, in an Illumina HiSeq2500 sequencer with a read depth of 100 bp in the case of the *Chst8* VP and 74 bp in the other samples. 4C-seq data have been deposited in the GEO database under the accession GSE197013 (GSE197010 for *Chst8*; GSE197011 for *Aopep*; GSE197012 for *Ldlrad4*). Computational analysis will be described in the subsection 3.1.

#### **2.2.1.8. Cloning of plasmidic DNA**

##### **2.2.1.8.1. Luciferase constructs**

The aim of these cloning was to introduce some of *Chst8* enhancer regions - Viewpoint, E1, E2 and E3 - in the pGL3 basic-*Chst8* Promoter (pGL3-*Chst8 P*) vector (previously constructed by Dr Fueyo in the lab). To this aim, these regions were amplified by PCR from genomic mouse DNA using specific primers with some restriction enzymes cut sites added (please refer to Table M3 to see details). After agarose gel purifying the amplified fragments, various strategies were followed in order to clone different combinations of promoter-enhancer(s) into the pGL3 promoter vector. They will now be detailed construct by construct.

##### pGL3-Viewpoint-*Chst8* Promoter (pGL3-VP-*Chst8 P*)

Promoter region was extracted from the vector pGL3-*Chst8 P* by digestion with NcoI and BglII restriction enzymes followed by agarose gel purification. In parallel, the destination vector pGL3-VP (previously cloned by Dr Fueyo



in the lab) was opened with the same enzymes. Both the promoter and the linearized destination vector were ligated using T4 DNA ligase.

#### pGL3-E1-*Chst8* Promoter (pGL3-E1-*Chst8* P)

The DNA fragment E1 was PCR amplified and digested with Mlul enzyme. After its gel purification, it was ligated with previously Mlul digested vector pGL3-*Chst8* P by means of the T4 ligase.

#### pGL3-E2-*Chst8* Promoter (pGL3-E2-*Chst8* P)

E2 fragment was PCR amplified, digested with NheI restriction enzyme and gel purified. Then, it was ligated using the T4 DNA ligase with the vector pGL3-*Chst8* P, which was opened with NheI and agarose gel purified.

#### pGL3-E3-*Chst8* Promoter (pGL3-E3-*Chst8* P)

pGL3-*Chst8* P vector was opened with Acc65I restriction enzyme. Later, it was ligated with E3 fragment using T4 DNA ligase, which was previously PCR amplified and digested with the same enzyme.

#### pGL3-E3-VP-*Chst8* Promoter (pGL3-E3-VP-*Chst8* P)

pGL3-VP-*Chst8* P vector was opened with Acc65I restriction enzyme so that it could be ligated using T4 DNA ligase with E3 fragment, which was PCR amplified, digested using this same enzyme and agarose gel purified.

#### pGL3-E3-E1-*Chst8* Promoter (pGL3-E3-E1-*Chst8* P)

pGL3-E3-*Chst8* P was digested with Mlul enzyme and purified by agarose gel extraction. Then, it was ligated with the E1 DNA fragment, which was PCR amplified and Mlul digested, using the T4 ligase.

#### pGL3-E2-E3-*Chst8* Promoter (pGL3-E2-E3-*Chst8* P)

The vector pGL3-E3-*Chst8* P was opened with NheI restriction enzyme and agarose gel extracted, and it was ligated by means of the T4 ligase with the PCR amplified and NheI digested E2 enhancer.

### 2.2.1.8.2. Overexpression constructs

These vectors were intended to overexpress different forms of wild type (WT) and mutated JMJD3 and PHF2 proteins (Table M1). As previously reported, they were mainly used for imaging assays due to the mEGFP and mCherry proteins.

#### pCMV-HA-mEGFP-JMJD3

mEGFP was extracted from the vector pCMV-mEGFP-C1 (kindly donated by Dr Aragay) by PCR amplification with the primers detailed in Table M3, Acc65I restriction enzyme digestion and agarose gel purification. Then the vector pCMV-HA-JMJD3 was opened with Acc65I and ligated with the extracted mEGFP using T4 DNA ligase.

#### pCMV-HA-mEGFP-JMJD3 $\Delta$ IDR (140-820)

The previously obtained vector pCMV-HA-mEGFP-JMJD3 was partially digested by HindIII enzyme to remove the IDR from JMJD3. After agarose purification, the vector was self-ligated using T4 DNA ligase.

#### pCMV-HA-mEGFP-JMJD3 HE>AA (H1390A/E1392A)

JMJD3 HE>AA was obtained by digesting the pCIG-JMJD3-H1390A/E1392A vector (previously cloned by Dr Estarás) with XhoI and XagI enzymes. The desired fragment was gel purified and used to replace JMJD3 from pCMV-HA-mEGFP-JMJD3, which was previously digested with the same enzymes and the resulting open plasmid without JMJD3 was also purified by agarose gel. Both fragments were ligated with T4 DNA ligase to obtain pCMV-HA-mEGFP-JMJD3 HE>AA vector.

#### pCMV-mCherry-PHF2

PHF2 was extracted from the p3xFLAG-PHF2 vector (kindly provided by Dr Jiemin Wong) through digestion with BamHI restriction enzyme. After agarose gel extraction, PHF2 fragment was inserted into the BamHI

digested pCMV-mCherry-C1 plasmid, and they were ligated by means of the T4 DNA ligase.

#### p3xFLAG-mCherry-PHF2 ΔCharged (487-606)

p3xFLAG-PHF2 ΔCharged vector was obtained from p3xFLAG-PHF2 through a PCR amplification with specific primer pairs (Table M3) so that the deleted region was excluded. After amplification, the vector was purified and self-ligated with T4 ligase.

#### pCMV-mCherry-PHF2 ΔCharged (487-606)

mCherry-PHF2 ΔCharged plasmid was obtained by opening pCMV-mCherry-PHF2 vector with BamHI enzyme and replacing its PHF2 with the ΔCharged mutant version, which was extracted from the p3xFLAG-PHF2 ΔCharged by digestion with BamHI. After digestions, the desired fragments were agarose gel purified and ligated with T4 ligase.

### **2.2.1.9. Amplification of plasmidic DNA: mini and maxi preparations of DNA**

To amplify plasmidic DNA, the first step consisted in bacteria transformation and inoculation of an isolated colony into LB medium, so that it grew overnight. Depending on the concentration of DNA we wanted to get, we used either 5 mL or 500 mL of LB, thus, that we obtained “minipreps” – around 15 ug of DNA- or “maxipreps”- around 80 ug of DNA. This way, bacterial DNA was purified using the QIAprep Spin Miniprep Kit (for minipreps, QIAGEN 27106) and of the QIAGEN Plasmid Maxi Kit (for maxipreps, QIAGEN 12165). Both protocols consisted in the addition of buffers P1 (100 ug/mL RNase A, 50 mM Tris-HCl, 10 mM EDTA pH 8.0), P2 (200 mM NaOH, 1% SDS) and P3 (3 M KAc, pH 5.5) for resuspension, lysis and neutralization, respectively. Then DNA was specifically retained in a column from which it was eluted and precipitated with isopropanol after several washes. Once isolated, DNA was washed with 70% ethanol and after drying it was resuspended in a suitable buffer.

#### **2.2.1.10. Electrophoresis in agarose gels**

This technique consists in running DNA or RNA in an agarose gel to visualize and/or purify them. Gel was made of the desired percentage of agarose dissolved into TBE (Tris-Borate-EDTA) buffer (45 mM Tris, 45 mM boric acid and 1 mM EDTA) and Redsafe reagent (Intron, 21141), which allows the visualization of nucleic acids due to the fluorescence that it emits when binding them. Once the gel was solidified, samples containing orange-glycerol were loaded and it was run at 80 V for typically 1-4 hours – depending on the separation of bands needed. Nucleic acids were visualized using an UV-transilluminator.

#### **2.2.1.11. Luciferase reporter assay**

This experiment is used to study gene expression at a transcriptional level. It is based on the detection of light emitted by the oxidative reaction of luciferin to oxyluciferin, which is catalysed by the firefly luciferase enzyme. The gene encoding this genetic reporter is cloned into expression vectors that also contain regulatory regions, such as enhancers, so that when performing the assay, you can measure the increase of light emission whenever a regulatory element is expressed. This way, it is possible to infer that upon the presence of this target regulatory element the expression of the reporter is enhanced, and due to light measurements is even possible to quantify this increase. On the other hand, *Renilla* luciferase is simultaneously used as an internal control of the experiment, as it also acts as a genetic reporter by catalysing another luminescence reaction – for instance, from coelenterazine to coelenteramide.

$1.2 \times 10^6$  NIH3T3 cells were cotransfected both with plasmids that contain our target regions– in this case, enhancer regions – and the firefly luciferase reporter gene and with a vector containing the *Renilla* luciferase gene (see Table M1). Dual-Luciferase® Reporter Assay System kit (Promega, E1910)

was used to perform the assay, and *Renilla* luciferase served as an internal control. When confluent, cells were lysed using 500 ul Passive Lysis Buffer (PLB) and light emission measurements were taken in a luminometer by sequentially adding 100 uL of Luciferase Assay Reagent II (LAR II) to 20 ul of cell lysate. This way luciferase activity can be read. Once all samples were tested, the firefly reaction was quenched by adding 100 ul of Stop & Glo® reagent, which in parallel initiates the *Renilla* one. Finally, firefly records were normalized by the control *Renilla* ones.

### **2.2.2. Protein-related**

#### **2.2.2.1. Total protein extraction and quantification**

This protocol was done in order to determine levels of specific proteins in certain contexts. The buffer used was the highly astringent RIPA (Radioimmunoprecipitation assay buffer), able to break both cytoplasmic and nuclear membranes (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS and 50 mM Tris, pH 8.0 and protease inhibitors). Cells were incubated on ice during 20 minutes in suspension within a volume of RIPA which depended on the desired final protein concentration, but which was about 500 ul for  $6 \times 10^6$  cells. Then it was submitted to centrifugation at 4°C and maximum speed for 10 minutes in an Eppendorf 5415, finally obtaining protein extract in the supernatant of the suspension.

To quantify the amount of protein extracted we used the Bradford method, which is based on the reaction by which Coomassie Brilliant Blue G-250 (Bio-Rad, 5000001) changes its color depending on protein concentration (224). 1 ul of protein extract was mixed with 1 ml of reagent, and after 5 minutes of reaction the absorbance value of the sample was measured using a spectrometer. Protein concentration was finally obtained using a bovin serum albumin (BSA) calibration line.

### **2.2.2.2. SDS-Page electrophoresis**

This protocol is suitable to separate proteins in a gel according to their actual size, which is possible due to the net negative charge that proteins acquire when they are in solution with sodium dodecyl sulphate (SDS).

Protein extracts were mixed with Laemmli buffer (375 mM Tris-HCl, 9% SDS, 50% Glycerol, 0.03% Bromophenol blue) and 5% of  $\beta$ -mercaptoethanol. Before loading in a polyacrylamide gel, they were heated during 5 minutes at 95°C. This gel (TEMED, 10% ammonium persulfate, 10% SDS, 1.5 M Tris pH 6.8/8.8, 40% polyacrylamide of variable %, H<sub>2</sub>O) is composed of two different parts, the stacking gel – with a 5% polyacrilamide concentration and pH 6.8- and the resolving gel – with a polyacrilamide concentration that will depend on the sizes of proteins to resolve and a pH 8.8. Gel was covered with SDS-Page Running buffer (25 mM Tris-HCl, 192 mM glycine, 0.1% SDS) and ran at 25 mA until proteins were at a distance enough to distinguish them.

### **2.2.2.3. Western Blot**

Assay performed to detect a specific protein in a sample previously resolved by SDS-Page electrophoresis. Proteins were transferred from the gel to a nitrocellulose membrane using an 80 V current that makes them flow from the negative to the positive side of a multi-layered cassette. To set up the electrical reaction, a sponge, Whatman paper, gel, nitrocellulose membrane, Whatman paper and another sponge were placed into the cassette, following this order from negative to positive side. Then the cassette was covered with Transfer buffer (we used two different buffers- for conventional proteins, 25 mM Tris-HCl, 192 mM glycine, 0.05% SDS and 10% methanol; for high molecular weight proteins, such as JMJD3, the transfer buffer contained 0.1% of SDS instead of 0.05%). The transference was done at 80 V for 90 minutes for conventional proteins and 140 minutes for the high molecular weight ones.

After protein transference the nitrocellulose membrane was blocked with 5% milk during 1 hour at room temperature. After three washes with PBS-Tween 0.1%, it was incubated with a specific primary antibody (Table M2) at 4°C overnight. The following day membrane was washed again 3 times with PBS-Tween 0.1% and incubated with a secondary antibody (Table M2) at room temperature for 1 hour. These antibodies are bound to a fluorophore, so to visualize proteins we just used a LI-COR Odyssey scanner.

#### **2.2.2.4. Chromatin immunoprecipitation (ChIP)**

This protocol precipitates proteins bound to chromatin and allows the identification of the DNA bound. About  $6 \times 10^6$  cells were fixed with 1% of formaldehyde for 10 minutes, reaction was stopped by adding 0.125 M of glycine for 5 minutes. After cell lysis in 1% SDS lysis buffer (1% SDS, 10 mM EDTA pH8.0, 50 mM Tris-HCl pH8.1) chromatin was sonicated in a Bioruptor sonicator to obtain fragments of around 300 bps of DNA. Chromatin was then purified by centrifugation at maximum speed for 10 minutes and the supernatant was used to test whether sonication produced chromatin fragments of the correct size. Immunoprecipitation was done by overnight incubation at 4°C with the specific antibody of interest to a dilution of the chromatin tenfold with immunoprecipitated buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris-HCl pH8.0). A control parallel reaction with IgGs was also run. Next day antibodies and proteins bound were attached to Magna ChIP magnetic beads (Millipore, 16-661) through 44 hours at 4°C incubation, and after sequential washes with buffers TSEI (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH8.0 and 150 mM NaCl), TSEII (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH8.0 and 500 mM NaCl), TSEIII (0.25 M LiCl, 1% NP40, 1% sodium deoxycholate, 1mM EDTA and 10mM Tris-HCl pH8.0) and TE (10 mM Tris-HCl pH 8.0 and 1 mM EDTA), elution of antibody-protein-DNA complexes was accomplished by adding elution buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>) for

15 minutes. Samples were then decrosslinked by overnight incubation at 65°C, so that DNA could be recovered for posterior qPCR analysis. After incubation with 1 mg/ml RNase A for 30 minutes at 37°C and with proteinase K at 55°C for 2 hours, DNA was extracted with phenol-chloroform and ethanol precipitation. Finally, it was resuspended in 50 µl H<sub>2</sub>O.

To conclude the analysis, DNA was subjected to qPCR as previously explained (see subsection 2.2.1.5.) using specific primers (Table M3). To quantify immunoprecipitated DNA with respect to the starting chromatin we used the input material.

#### **2.2.2.5. Droplet assays in nuclear extracts**

5 µg of vectors encoding the desired gene fused to mEGFP were transfected into 20 × 10<sup>6</sup> HEK293T cells as previously described (225). After 24 hours, cells were resuspended in 10 ml HMSD50 buffer (20 mM HEPES, 5 mM MgCl<sub>2</sub>, 250 mM sucrose, 1 mM DTT, 50 mM NaCl supplemented with 0.2 mM PMSF and 5 mM sodium butyrate) and incubated for 30 min at 4°C with gentle agitation. The solution was spun down at 2.600 g at 4°C for 10 min. The supernatant was discarded and the pellet containing nuclei was washed in Mnase buffer (20 mM HEPES, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, protease and phosphatase inhibitors). The washed nuclei were resuspended in one pellet volume of Mnase buffer and treated with 1 U Mnase (Sigma, N3755) at 37°C for 10 minutes. One pellet volume of stop buffer (20 mM HEPES, 500 mM NaCl, 5 mM MgCl<sub>2</sub>, 30% glycerol, 15 mM EGTA, protease and phosphatase inhibitors) was added to stop the reaction. The solution was briefly sonicated and spun down at 2.600 g at 4°C for 10 min. The supernatant was spun down again at 2.600 g at 4°C for 5 min to clear the nuclear extract. The nuclear extract concentration was measured using a spectrophotometer (about 2 mg/ml).



Then nuclear extracts were used for droplet formation assays by 1:1 diluting them with Buffer B (10% glycerol, 20 mM HEPES). The final droplet buffer conditions were 20 mM HEPES, 150 mM NaCl, 15% glycerol, 3.75 mM EGTA, 2.5 mM MgCl<sub>2</sub>, 1.25 mM CaCl<sub>2</sub>. The reactions were incubated for 30 minutes in 8-well PCR strips and loaded onto glass bottom 384 well plate (Cellvis P384-1.5H-N) 5 minutes prior to imaging on an Automated Inverted Microscope Leica Thunder 3D Live Cell using a 63x water immersion objective (NA=1.2).

#### **2.2.2.6. Indirect immunofluorescence**

Microscopy technique which aims to visualize stained proteins. To get these images, cells grown on coverslip coated with poly-D-lysine were fixed for 20 minutes at room temperature in 4% of paraformaldehyde. Then, they were permeabilized with PBS-Triton X-100 0.5% for 10 minutes at room temperature and blocked in 5% BSA (Bovine serum albumina) for 30 minutes at room temperature. After this step the coverslip was incubated overnight at 4°C with primary antibodies (Table M2). The following day cells were incubated for at least 1 hour at room temperature with Alexa-conjugated secondary IgG antibodies (Table M2) and 0.1 ng/μl DAPI (ThermoFisher, D1306). As these antibodies emit fluorescence light, they are necessary to visualize the proteins at the microscope. For cells overexpressing mEGFP or mCherry-fusion proteins, the intrinsic fluorescence of the molecules was captured without using either primary or secondary antibodies. Images were captured by a Leica SP5 confocal microscope using LAS-AF software. Quantification of the level of fluorescence of randomly selected cells was achieved by using Fiji (226) to calculate the corrected cell fluorescence as the Integrated Density - (Area of selected cell x Mean fluorescence of background readings).

### 2.2.2.7. Immunofluorescence with DNA FISH

Fluorescence *in situ* hybridization is a cytogenetic technique that uses specific fluorescent probes to target a locus of interest. Its combination with an immunofluorescence experiment against a certain protein allowed us to observe the presence of our protein of interest in the dyed locus.

For this, the study of the endogenous protein was performed. This way, NSCs were grown on glass dishes precoated with 5 ug/ml of poly-D-lysine and 5 ug/ml of laminin and treated with TGF $\beta$  for 3 hours. After fixation in 4% paraformaldehyde containing 0.1% Triton X-100 solution for 10 minutes, cells were incubated in 10 mM glycine for 30 minutes and washed with PBS three times. Then, cells were dehydrated by performing sequential washes with 70%, 85% and 100% ethanol for 2 minutes at RT and then air dried. Probe hybridization mixture was made by mixing 8 ul of FISH Hybridization Buffer (Empire Genomics) and 2 ul of the FISH probe (see below). Then, 10 ul of mixture were added on a slide. Genomic DNA and probes were denatured at 75°C for 7 minutes and slides were incubated at 37°C in the dark overnight. The coverslip was removed from the slide and washed twice with Wash solution 1 (0.3% NP-40/0.4x SCC) at 73°C for 5 minutes and twice with Wash solution 2 (0.3% NP-40/3x SCC) at room temperature for 2 minutes.

Then immunofluorescence was performed in a slightly different way as described above. Cells were fixed again in 4% paraformaldehyde for 10 minutes and washed with PBS three times. Permeabilization was done in 0.2% Triton X-100 in PBS for 10 minutes. After washing with PBS, cells were blocked at room temperature with PBG (2% Fish gelatin, 5% BSA, 1x PBS to volume) for 1 hour. Cells were incubated with the primary antibody (see table M2) at 4°C overnight. After three washes with PBG, cells were incubated with the secondary antibody (table M2) for 45 minutes at room temperature. The coverslip was washed twice with PBG and twice with PBS, and nuclei were stained with DAPI for 10 minutes at room

temperature. Coverslips were mounted with Mowiol and images were acquired on an Andor Dragonfly Spinning Disk Confocal Microscope with an oil immersion 100x (NA=1.49) objective and a pixel size of 51 nm objective using Fusion acquisition software. Images were post-processed using Fiji (226). FISH foci were manually identified in individual z stacks through intensity thresholds set up.

The DNA FISH probe was synthesized by Empire Genomics and targets the *Chst8* enhancer cluster locus (mm10 coordinates chr7:34795935-34985109).

#### **2.2.2.8. Live imaging**

The specifications detailed in this section are common to all live imaging experiments performed. HEK293T cells transfected with HA-mEGFP-JMJD3 vector were grown in 1.5ml DMEM media on 35mm glass plates (ibidi,  $\mu$ -Dish 35 mm, high Glass Bottom 81158) coated with 5 ug/ml of poly-D-lysine.

Images were obtained on a 37°C heated stage of a Zeiss LSM780 Confocal using Zen software. Images were acquired with the Spectral (GaAsP) detector and a 40x water immersion (NA=1.2) objective. Raw images were processed using Fiji software (226) for posterior analysis and quantifications.

#### **2.2.2.9. Fluorescence recovery after photobleaching (FRAP) in live cells**

FRAP assays were done to study the dynamic and liquid-like nature of the observed nuclear condensates. After removing the fluorescence, or bleaching, of a certain region of the cell (nuclear condensates in this case), the kinetics of the fluorescence recovery were measured, reflecting the diffusion capacity of the molecules within the condensate.

A Zeiss LSM780 confocal microscope with a 40x water immersion objective (NA=1.2) and a GaAsP photomultiplier detector was used to perform FRAP experiments. Acquisition settings were optimized for fast imaging and low photobleaching, using 488 nm laser excitation power of 0.15% (AOTF), a detector gain of 780, a pixel dwell of 1.27 usec and a pixel size of 140 nm. Bleaching was performed after the acquisition of 5 images, using 488 nm laser excitation power of 100% (AOTF), a pixel dwell of 2.55 and 10 iterations, over a 6x6 pixels region of interest (ROI) focused on the interest puncta. Acquisition was set to intervals of 1 second for both the pre-bleach imaging and the post-bleach recovery time.

The quantification of the intensity recovery was calculated with Fiji (226). A macro was programmed that i) registered the whole time-lapse to avoid live cell fluctuations; ii) allowed the user to draw a region of interest (ROI) around the bleached spot; iii) fine-tuned the selected ROI by applying a threshold on the time projection of the signal spot; iv) allowed the user to select a background ROI; v) automatically segmented the target nucleus and created a ROI where to calculate the bleaching gap and the bleach depth; vi) automatically measured the intensity of the three ROIs over the time-lapse and vii) delivered the data in a \*.txt format. The code and further details can be downloaded from <https://github.com/MolecularImagingPlatformIBMB>. To minimize the error of image analysis, three rounds of quantification were collected. Statistics and fitting were performed using the easy-FRAP web (227). The experiment settings, performance and analysis were done in collaboration with Dr Rebollo (Molecular Imaging Platform, IBMB).

#### **2.2.2.10. Quantification of droplets liquid-like features**

Droplets “circularity”, “roundness”, “aspect ratio” and “convexity” shape descriptors were calculated to determine their liquid-like nature. For this, Fiji software (226) was used: the “Analyze particles” plugin quantified

“circularity”, “roundness” and “aspect ratio”, while “convexity” was calculated running the Fiji macro “Calculate Convexity and Solidarity”. After cropping images, a threshold was set to see each droplet as an individual object.

#### **2.2.2.11. Focus calling (immunofluorescence, 1,6-Hexanediol live-imaging)**

Foci were called using the “Object Counter 3D” plugin in Fiji (226). For each image, the “threshold” parameter was set so that each focus could be seen as an individual object. The showed parameters (number of foci/cell, intensity and volume) are the mean of the results obtained for each image with the “Statistics” function of the plugin (the number of cells used for the quantifications is specified in each figure).

### **3. Bioinformatic methods**

#### **3.1. 4C-seq computational analysis**

Computational analysis of the 4C-seq results were performed by Dr Fueyo in the Stanford School of Medicine, in California. I will briefly comment the details here. The resulting 4C-seq reads were processed with the pipeline pipe4C (228) using default parameters except for the trimLength that was set up to 36 bp, and the genome version for mapping that was *Mus musculus* mm10. R3Cseq (229) was employed for statistical analysis. This Bioconductor package allows the identification of interacting genomic regions and the comparison between multiple replicates and experimental conditions.

#### **3.2. Gene expression omnibus (GEO) accessions**

In this study we have taken advantage of publicly available genome-wide datasets to compare with our results and test hypothesis. The GEO (gene

expression omnibus) accession numbers of these data are registered in Table M5.

Data	Accession number
SMAD3 0,5h TGF $\beta$ ChIP-seq	GEO GSM898371
JMJD3 3h TGF $\beta$ ChIP-seq	GEO GSM937827
H3K27ac ChIP-seq	GEO GSE66961
H3K4me1 ChIP-seq	GEO GSE66961
H3K4me3 ChIP-seq	GEO GSE66961
H3K27me3 ChIP-seq	GEO GSE38269
SMC1 ChIP-seq	GSM883646

**Table M5.** GEO accession numbers for next generation sequencing data used in the thesis.

### 3.3. Capture obtaining

Several captions of both ChIP-seq and 4C-seq signals are included in this doctoral thesis. They result from the visualization of bigwig and bedgraph files of different experiments in the UCSC genome browser (230).

### 3.4. Protein sequence analysis and predictions

Protein disorder estimations were generated using three prediction algorithms, PONDR-VL3 (231), IUPred (232) and PONDR-VSL2 (233). The predictors give a value between 0 and 1 for each amino acid, where above 0.5 is predicted to lie within a disordered region of more than 50 amino acids long. To predict the phase separation property of each protein, the PSPredictor (234) and catGRANULE (235) predictors were used online.

Low-complexity domains presence was assessed using the SEG algorithm together with the MobiDB (236) database. For amino acid composition analysis, the web application Prot Pi Protein Tool (<https://www.protpi.ch/Calculator/ProteinTool>) was used. Disordered proteins were defined by the presence of a 50 residues fragment whose

IUPRED median score was at least of 0.55 and that was not found in Pfam, a protein domain database. The hydrophobicity was calculated with ExPASy website (237) using Hopp and Woods (238) scale and a sliding window of 21.

### **3.5. Amino acid composition of conserved IDRs within KDM families**

The IDR conservation between the members of each JMJC-KDM family was assessed using the Basic Local Alignment Search Tool (BLAST®) (239). This is, we first aligned IDRs of JMJC-KDMs of a certain family, determined the IDRs that were conserved within the family, and determine the proportion of each amino acid within these conserved tracks with respect to the total amount of amino acids in these tracks. The amino acid proportion within a KDM family was calculated for each amino acid in the following way: e.g., proportion of K in KDM4 conserved tracks = number of K in KDM4 conserved tracks / number of total amino acids in KDM4 conserved tracks. This was repeated for each KDM family, and once done, the mean value for each amino acid among all KDMs families was calculated (e.g., total K content in all the KDMs' IDR conserved tracks = mean of the proportion of K in KDM2 conserved tracks, KDM3 conserved tracks, KDM4 conserved tracks, KDM5 conserved tracks, KDM6 conserved tracks and KDM7 conserved tracks).

## **4. Statistical analysis**

### **4.1. Sample size**

Experiments presented in this thesis have been performed in triplicate, except for the 4C-seq assay that shows the result of two replicates.

#### **4.2. Graphics and measures of dispersion (mean and standard error of the mean)**

The software used to make calculations and construct graphics were Microsoft Excel and Prism-GraphPad. Graphics of experiments that fit a linear model have been represented as the mean, and their error bars correspond to the standard error of the mean (SEM). Regarding boxplots, boxes comprise values from Q1 to Q3 of the dataset, line corresponds to median value, and whiskers show the data range (from min. to max. values within dataset).

#### **4.3. Student's t-test**

This parametric test was used to determine the significance of experiments that follow a linear model, establishing that they are actually significant when a result represents a true hypothesis within a 95% of confidence. P-values were obtained with Microsoft Excel software, and they were represented as asterisks as follows: \* p-value < 0.05; \*\* p-value < 0.01; \*\*\* p-value < 0.001.

#### **4.4. Wilcoxon-Mann-Whitney test**

This test is frequently described as the nonparametric version of the t-test for two independent groups, and it is used to determine whether two independent samples come from equally distributed populations. This is, the test contrasts whether it is equally probable that a random value from the first population will be greater than or lesser than a random value from the second population (240). We have performed this test to analyze the significance of the differences in the 4C-seq signals between the different conditions.



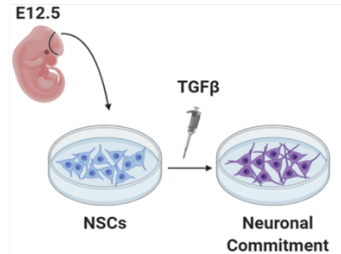
# Results

## Chapter 1: Characterization of JMJD3-mediated transcriptional regulation at enhancers during early neurogenesis

As commented in the introduction, both the 3D structure of the chromatin and the action of epigenetic factors and signaling pathways are fundamental to accomplish a proper development (241,242). Previously in the lab, we investigated the regulatory effects that the cooperation of JMJD3 with the TGF $\beta$  pathway exerted at enhancers during neurodevelopment. These studies identified some of the linear molecular components involved in neuronal enhancer activation (67). However, the role that chromatin structure might play in this regulation still needs to be unveiled. Thus, one of the goals of my doctoral thesis has been to examine the putative structural genome reorganization triggered by the TGF $\beta$ -JMJD3 crosstalk in a neurodevelopmental context. Most of the results presented in this chapter have been recently published (243).

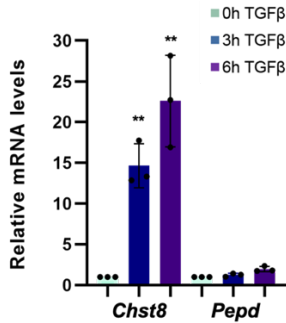
### 1.1. TGF $\beta$ promotes chromatin reorganization around the *Chst8* locus

We started this work wondering whether enhancer-promoter loops were underlying the TGF $\beta$ -dependent gene activation. To study this possibility, we decided to perform a 4C-seq experiment taking advantage of our previously mentioned model, the E12.5 mice NSCs. Although the TGF $\beta$  pathway is moderately active under basal conditions in these cells, further stimulation through the addition of the cytokine promotes activation of the neurodevelopmental program (66,67,140,159) (Figure R1).



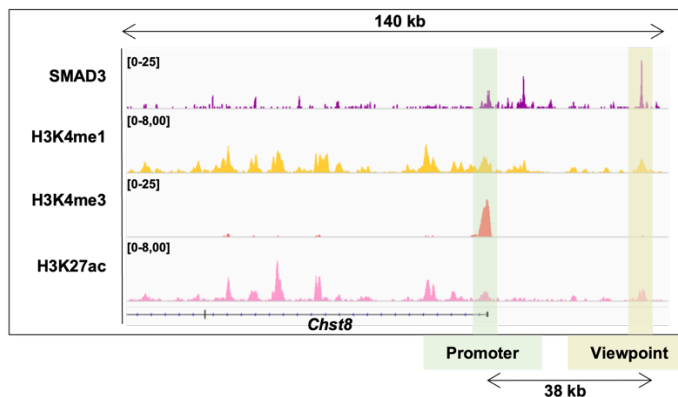
**Figure R1.** Picture depicting the model used along the study: NSCs obtained from the cortex of mice in the embryonic stage 12.5. These cells become committed to the neuronal lineage upon addition of the TGF $\beta$  cytokine.

For this chromatin conformation capture experiment, a cis-regulatory region named viewpoint (VP from now on) needs to be selected to investigate the contacts that are established between this specific locus and the rest of the genome. This way, enhancer-promoter loops can be elucidated. The VP region, that can be either a promoter or an enhancer, had to meet some criteria in order to be eligible: it had to be robustly upregulated upon TGF $\beta$  activation; it had to lie at least 30 kb apart from its cognate enhancer or promoter, as it is known that the contacts appearing close to the bait are difficult to interpret due to the preferential ligation of the viewpoint with its linear adjacent regions; and it had to regulate a relatively long gene, so that we could analyze both the contacts established with other regulatory regions and with the gene body. Taking all these requirements into account, we selected as a VP a putative enhancer of the TGF $\beta$ -regulated *Chst8* gene, whose mRNA levels increase up to ~25-fold upon TGF $\beta$ -stimulation (Figure R2).



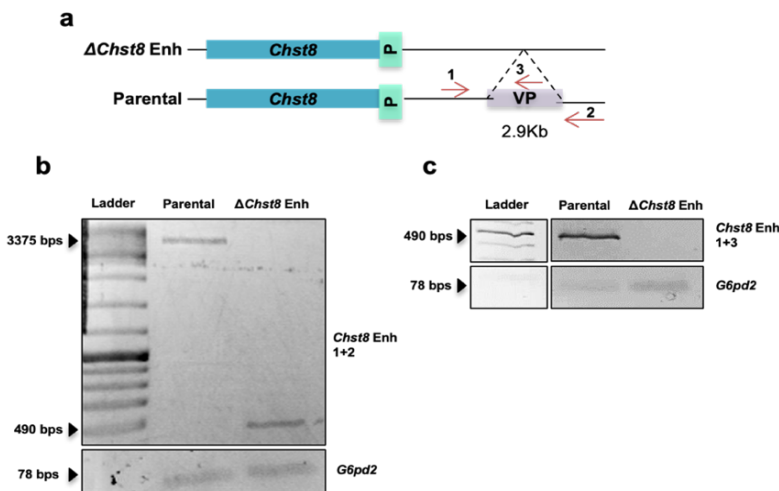
**Figure R2.** Relative transcription levels of the *Chst8* gene in NSCs stimulated with TGFβ for 3 or 6 hours. Both time points demonstrated to be sufficient to get a striking accumulation of mRNA. Its downstream gene *Pcpd* was used as a negative control.

Indeed, both the gene promoter and the VP enhancer targeted SMAD3. Furthermore, the VP lied 38 kb apart from the *Chst8* gene, allowing a reliable resolution in the experiment, and *Chst8* gene was moderately long (138 kb) (Figure R3). So, the VP enhancer met all the above-mentioned criteria to be selected.



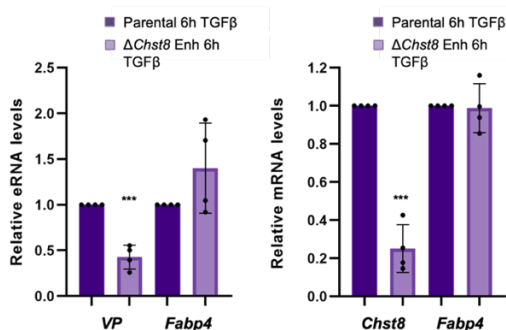
**Figure R3.** UCSC capture depicting the *Chst8* gene and its putative enhancer (VP enhancer). The SMAD3 ChIP-seq binding in NSCs treated with TGFβ is also displayed, as well as the chromatin landscape in untreated NSCs (H3K4me1, H3K27ac, and H3K4me3 marks, used to define enhancers and promoters (67).

Before proceeding with the 4C-seq experiment, we confirmed that the VP enhancer indeed activated the *Chst8* gene in response to the TGF $\beta$  signaling. To do this, we took advantage of CRISPR/Cas9-edited NSCs that were previously generated in the lab, in which the VP enhancer was removed. They were named *Chst8* enhancer-deleted ( $\Delta$ *Chst8* Enh) cells (Figure R4).



**Figure R4.** (a) Schematic representation of the editing strategy followed to generate the  $\Delta$ *Chst8* Enh NSC lines. Red arrows depict the pairs of primers used to test the deletion. (b-c) Conventional PCR results demonstrating the proper deletion of the enhancer, through parental or  $\Delta$ *Chst8* Enh DNA amplification using specific primers combinations (left, amplification with primers 1 and 2; right, amplification with primers 1 and 3).

Then, activation of both enhancer and *Chst8* gene were analyzed in parental and  $\Delta$ *Chst8* Enh lines through eRNA and mRNA levels measurements. As expected, a remarkable decrease of the VP eRNA levels, but also of *Chst8* mRNA levels, was observed in  $\Delta$ *Chst8* Enh lines when compared to the parental ones. This demonstrated that the VP indeed was an essential enhancer of the *Chst8* gene (Figure R5).



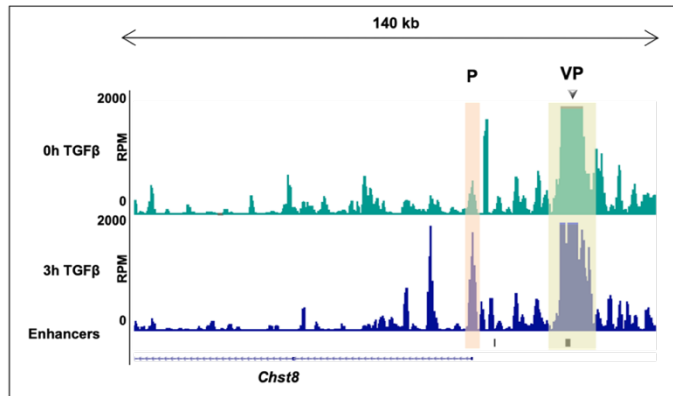
**Figure R5.** Relative eRNA (left) and mRNA (right) levels of the VP enhancer and *Chst8* gene in 6 hours TGFβ-stimulated parental and  $\Delta Chst8$  Enh NSCs. *Fabp4* enhancer and gene were used as unrelated negative controls.

Once assessed that the VP was an enhancer that regulated the *Chst8* gene, the 4C-seq experiment was performed. We included two independent biological replicates of untreated and three hours TGFβ treated NSCs. Prior to analyzing the results, the quality of the experiment was checked following the criteria described in the literature (228).

Sample	Total number of reads	% of reads containing VP sequence	% of fragments that map in cis	% of reads that map in unique sites within 1 Mb around VP
NSC 0h R1	10923463	96.58	69.2	55.66
NSC 0h R2	9632921	92.66	51.78	74.98
NSC 3h R1	7054576	97.15	68.82	62.46
NSC 3h R2	6979938	91.73	60.58	77.94

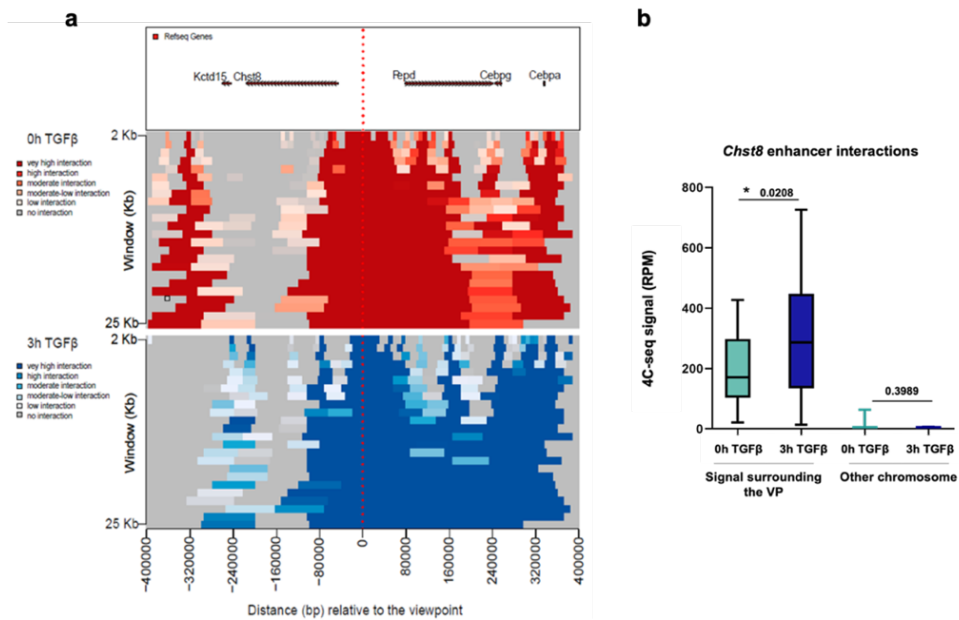
**Table R1.** Parameters that assess the quality of the 4C-seq experiment for each sample as previously described. Briefly, to consider the experiment as highly successful, the % of reads containing VP sequence should be above 90, the % of fragments that map in cis above 50, and the % of reads that map in unique sites within 1 Mb around VP above 60.

The results of the experiment can be seen in the captures depicted in Figure R6, which show that the proportion of cells displaying VP-*Chst8* Promoter (*Chst8* P) contacts increased upon stimulation with the cytokine. Unexpectedly, TGF $\beta$  seemed to trigger a general reorganization of the chromatin in the locus, as the appearance and disappearance of contacts between the VP and the *Chst8* gene body was also observed.



**Figure R6.** USCS Genome Browser 4C-seq profiles in NSCs untreated or treated for 3 hours with TGF $\beta$  in the *Chst8* locus. Previously defined enhancers signal is also depicted (black boxes).

Indeed, the changes triggered by the signaling pathway can be visualized in domainograms representing the frequency of 3D-contacts established between the VP and the genomic regions located 400 kb upstream or downstream the VP (Figure R7a). Furthermore, these changes can be quantified by measuring the averaged interactions occurring 500 kb around the VP (excluding the nearest  $\pm 20$  kb) in the two replicates. As expected, in the target region there was a striking increase in contacts driven by the pathway, as compared to the signal obtained in other unrelated chromosome used as a negative control, (p-values 0.0208 versus 0.3989) (Figure R7b).

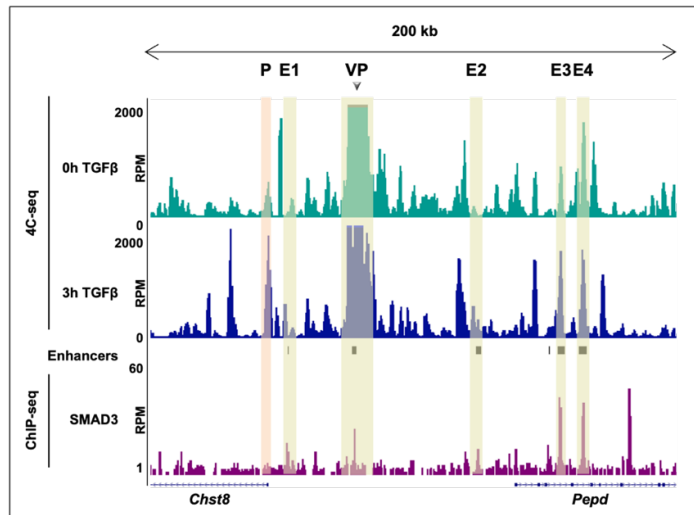


**Figure R7. (a)** r3Cseq domainogram that shows the interactions between the VP and its surrounding genomic regions, located 400 kb either upstream or downstream of the enhancer. Different shades of red (NSC 0h TGFβ) or blue (NSCs 3h TGFβ) illustrate the frequency of the interactions. The y-axis represents interactions identified using different window sizes. The plot is the result of one of the biological replicates. **(b)** Boxplot depicting the averaged 4C-signals obtained from the combination of the two biological replicates and represented as reads per million (RPMs). An independent region located in another chromosome (chr4:33076383-35216108) was used as a negative control. p-values are the result of a Wilcoxon-Mann-Whitney test.

## 1.2. TGFβ induces the formation of the *Chst8* enhancer cluster

In addition to the VP-*Chst8* P loop, other contacts were observed between the VP and inter- and intragenic enhancers. In fact, the mentioned interactions were more abundant and intense than the one between the VP and the promoter, as can be appreciated in Figure R8. Some of these

regions are intragenic enhancers located within the *Peptd* gene, that was not regulated by TGF $\beta$  (Figure R2).

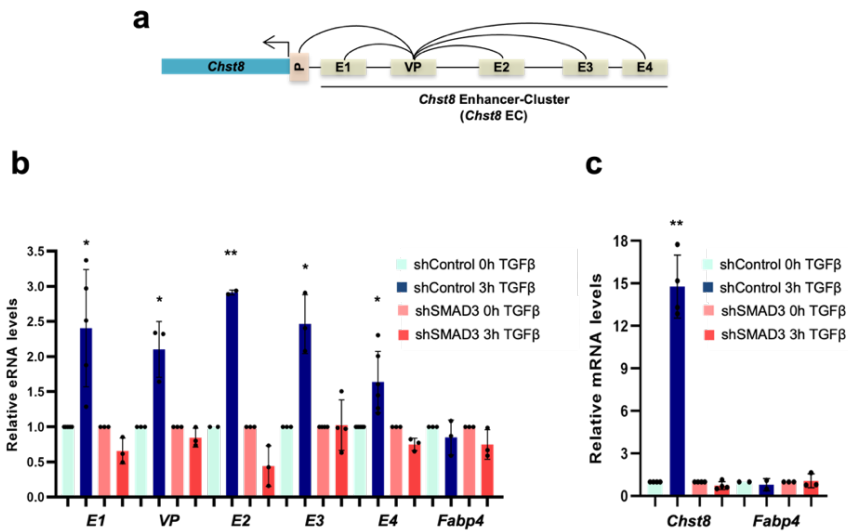


**Figure R8.** UCSC Genome Browser caption showing the contacts established between the VP and inter- and intragenic enhancer regions, mainly found within *Peptd* gene. SMAD3 ChIP-seq signals are also depicted, as well as previously defined enhancers (black boxes). The light orange box indicates enhancer-promoter contacts, while yellow boxes show enhancer-enhancer contacts.

These results pointed to a cooperation between enhancers to potentially activate *Chst8* in response to TGF $\beta$  signal. In other words, TGF $\beta$  pathway seemed to be promoting the assembly of an enhancer cluster to activate gene expression in the *Chst8* locus. This assembly was named *Chst8* enhancer cluster (EC) (Figure R9a). The enhancers that conform *Chst8* EC, which were named enhancer (E) 1, E2, E3, E4 and VP, were activated by the cytokine themselves, as they transcribed eRNA upon its addition (Figure R9b). The dependency of the EC activity for the signaling pathway was further reinforced by the fact that these enhancers didn't transcribe eRNA in cells depleted for the effector protein of the pathway SMAD3 (NSCs shSMAD3, characterized in (67)). Moreover, the pathway was not

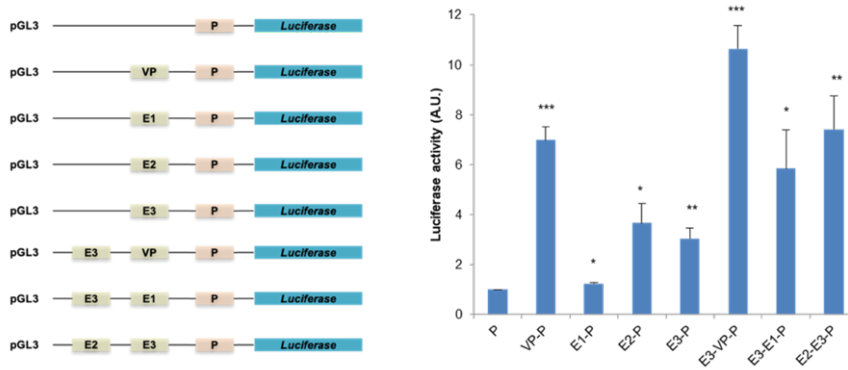


only necessary to induce *Chst8* EC activation, but also for the expression of its regulated gene *Chst8* (Figure R9c).



**Figure R9.** (a) Scheme that represents the different enhancer regions that form *Chst8* EC, together with *Chst8* promoter. (b-c) qPCR results that show the activity of *Chst8* EC constituent enhancers (b) and *Chst8* gene (c) in control and SMAD3-depleted NSCs that were stimulated with TGFβ for 3 hours. *Fabp4* enhancer (b) and gene (c) were used as negative controls.

Enhancers can cooperate in many different manners to promote gene expression (additive, synergistic, hierarchical, or competitive manner). To further characterize the cooperation mode of the described *Chst8* EC, we performed a set of gene reporter (luciferase) experiments. We constructed plasmids containing different combinations of the enhancers from the *Chst8* EC, together with *Chst8* P and the luciferase reporter gene. Figure R10 shows the plasmids used for the experiment, as well as the expression levels achieved in NIH3T3 expressing different enhancers-promoter combinations. Presumably, the regulatory elements of the *Chst8* EC acted in an additive manner.

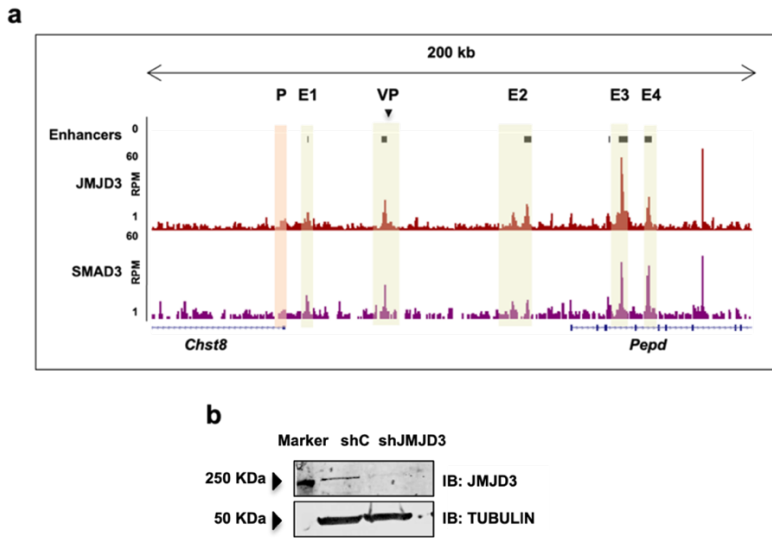


**Figure R10.** Schematic representation of the vectors used for the luciferase experiment (left), as well as the expression levels achieved in NIH3T3 cells that expressed these vectors containing different combinations of cis-regulatory elements (right).

Altogether, these results demonstrated a TGF $\beta$ -promoted reorganization of the 3D-chromatin structure at the *Chst8* locus. Furthermore, this restructuring involved the formation of the *Chst8* EC, whose enhancers acted in an additive manner to promote *Chst8* gene expression in response to TGF $\beta$  signal.

### 1.3. TGF $\beta$ -promoted *Chst8* enhancer cluster depends on JMJD3

Since JMJD3 cooperates with the TGF $\beta$  signal at enhancers (67), we decided to analyze whether JMJD3 contributed to *Chst8* EC formation. First, we reanalyzed our previously published JMJD3 ChIP-seq in NSCs treated with TGF $\beta$  for three hours (66) to confirm that JMJD3 binds *Chst8* EC enhancers (Figure R11a). Next, we depleted JMJD3 from NSCs using lentiviral vectors that expressed either control (shC) of JMJD3 specific shRNAs (shJMJD3), as previously done in the lab (67) (Figure R11b) and repeated the 4C-seq assay in these cells upon TGF $\beta$  treatment.



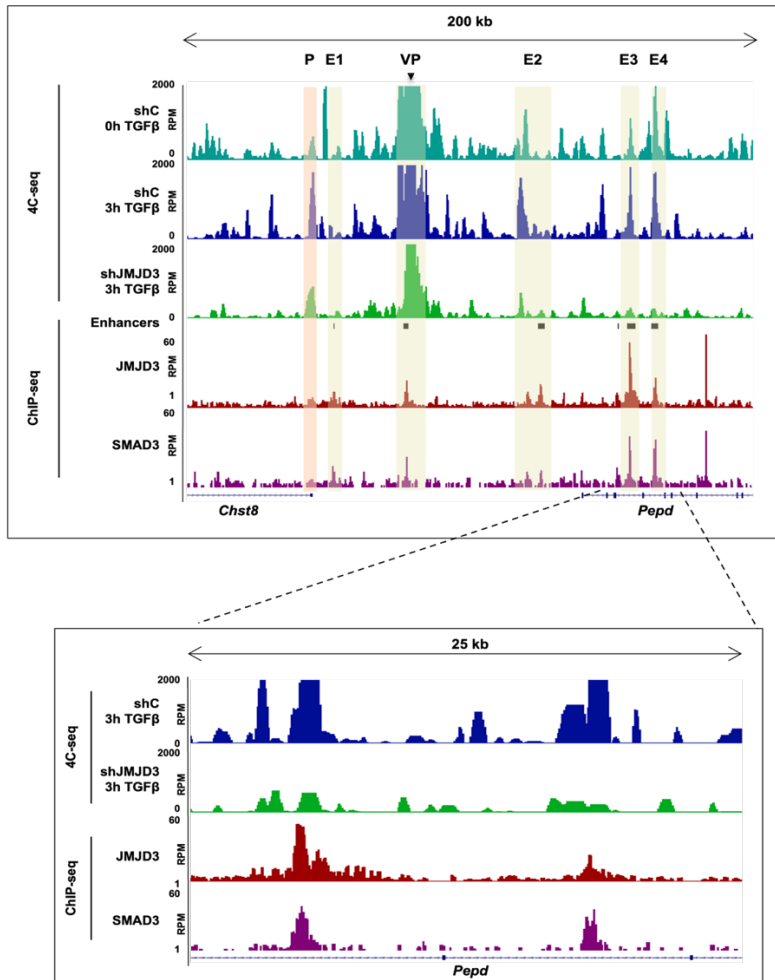
**Figure R11. (a)** UCSC Genome Browser caption showing previously defined enhancers, JMJD3 and SMAD3 ChIP-seq signals in the *Chst8* EC locus in NSCs treated for 3 hours with TGF $\beta$ . The light orange box highlights *Chst8* promoter, while yellow boxes depict the enhancers from *Chst8* EC **(b)** Western Blot showing JMJD3 levels in NSCs after infection with lentiviral vectors that express control (shC) or JMJD3 specific shRNAs (shJMJD3).

The quality of the 4C-seq experiments was once more checked as done before. Results in Table R2 demonstrated that the experiment met the defined criteria.

Sample	Total number of reads	% of reads containing VP sequence	% of fragments that map in cis	% of reads that map in unique sites within 1Mb around VP
shJMJD3 3h R1	9374517	97.15	75.09	66.91
shJMJD3 3h R2	8704966	80.01	52.06	74.83

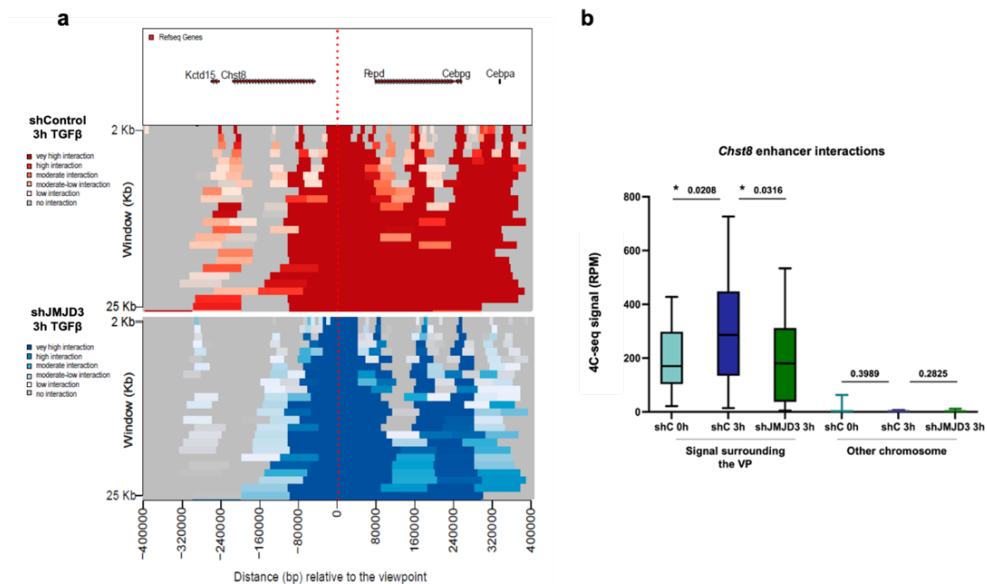
**Table R2.** Quality of the 4C-seq experiments evaluated according to the described parameters.

As seen in Figure R12, the lack of the demethylase led to a strong decrease of the contacts established between the *Chst8* EC elements, suggesting that JMJD3 was required for the EC assembly.



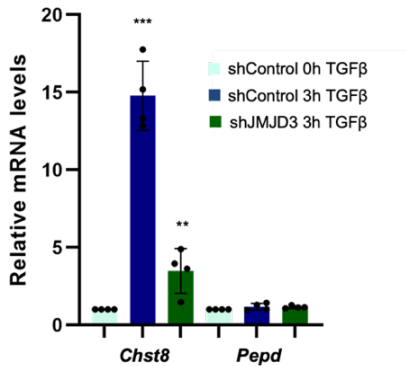
**Figure R12.** UCSC Genome Browser caption showing the 4C-seq signal in the *Chst8* EC locus in control and JMJD3-depleted NSCs treated for 3 hours with TGFβ. These signals can be better appreciated in the zoom-in caption (bottom). Enhancers, JMJD3 and SMAD3 ChIP-seq signals are also represented. The light orange box indicates enhancer-promoter contacts, while yellow boxes show enhancer-enhancer contacts.

Furthermore, a general decrease in contacts upon JMJD3 depletion was seen in domainograms representing the frequency of 3D-contacts established between the VP and the genomic regions located 400 kb upstream or downstream the VP (Figure R13a). This decrease became clear when quantifying the interactions occurring 500 kb around the VP (excluding the nearest  $\pm 20$  kb) in the two replicates. The RPMs of NSCs shJMJD3 dropped to the levels observed previous addition of the cytokine (p-values 0.0316), while in the independent negative control region there were no changes (p-values 0.2825) (Figure R13b).



**Figure R13. (a)** r3Cseq domainogram showing the 3D interactions between the VP *Chst8* enhancer and the genomic regions located 400 kb either upstream or downstream. Different shades of red (shControl 3 h TGFβ) or blue (shJMJD3 3 h TGFβ) illustrate the frequency of the interactions. The y-axis represents interactions identified using different window sizes. **(b)** Boxplot depicting the averaged RPMs from two biological independent replicates. As negative control, a region located in another chromosome (chr4:33076383-35216108) was chosen. p-values are the result of a Wilcoxon-Mann-Whitney test.

In accordance with the structural changes observed in the *Chst8* locus, gene expression was also diminished upon depletion of JMJD3 (Figure R14).



**Figure R14.** Relative mRNA levels of *Chst8* gene in NSCs control and shJMJD3 upon TGFβ treatment for 3 hours. *Pepd* gene was used as a negative control.

Thus, the presented results demonstrated that JMJD3 was necessary for the establishment of contacts, which led to the formation of *Chst8* EC upon activation of the TGFβ pathway. Moreover, *Chst8* expression was also dependent on JMJD3.

#### 1.4. TGFβ and JMJD3 drive chromatin reorganization

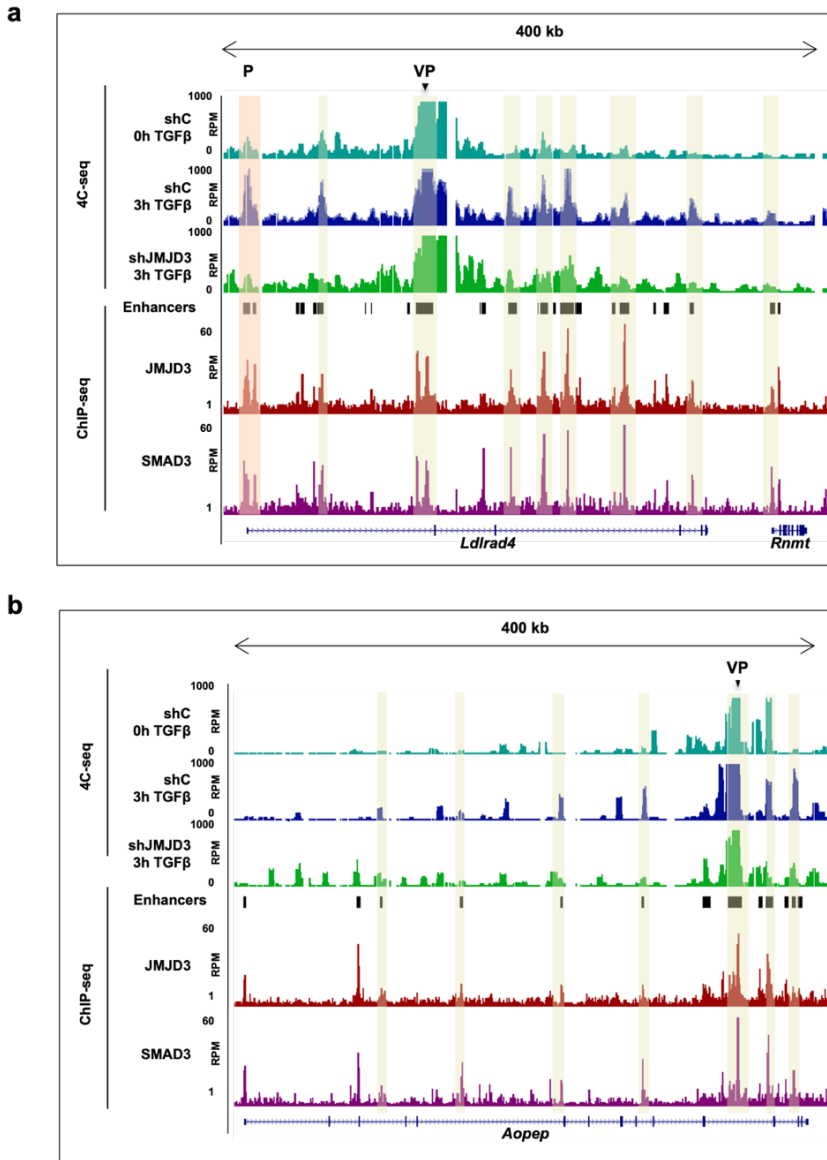
The results showed until now referred to the *Chst8* locus. In an attempt to broad our conclusions, we selected other two loci, *Ldlrad4* and *Aoepc* genes, that were regulated in a TGFβ and JMJD3 dependent manner according to previously generated data in the lab (66). By 4C-seq experiments, we analyzed the structural changes occurring upon TGFβ treatment in control and shJMJD3 NSCs. To do that, we selected two intragenic candidate enhancers targeted by SMAD3 and JMJD3, which were also surrounded by other enhancers with binding sites for both proteins.

Before analyzing 4C-seq results, as previously done, we assessed the quality of the experiment. Results in Table R3 demonstrate that the experiment mostly met the defined criteria.

	Sample	Total number of reads	% of reads containing VP sequence	% of fragments that map in cis	% of reads that map in unique sites within 1Mb around VP
<b><i>Ldlrad4</i></b>	shC 0h R1	8041622	98.48	51.51	70.69
	shC 0h R2	10701253	97.21	40.09	67.4
	shC 3h R1	8398290	97.84	65.65	76.05
	shC 3h R2	8559204	96.73	41.12	69.68
	shJMJD3 3h R1	10251577	97.82	69.76	81.33
	shJMJD3 3h R2	9326840	97.23	36.18	68.82
<b><i>Aopep</i></b>	shC 0h R1	9091971	98.24	42.54	65.5
	shC 0h R2	11147089	97.13	48.42	63.84
	shC 3h R1	9051926	96.89	62.53	77.97
	shC 3h R2	9491667	97.12	44.58	69.88
	shJMJD3 3h R1	9845743	98.09	75.13	85.75
	shJMJD3 3h R2	9354806	97.85	46.67	67.01

**Table R3.** Parameters that demonstrate the quality of *Ldlrad4* and *Aopep* 4C-seq experiment for each sample according to what was previously described.

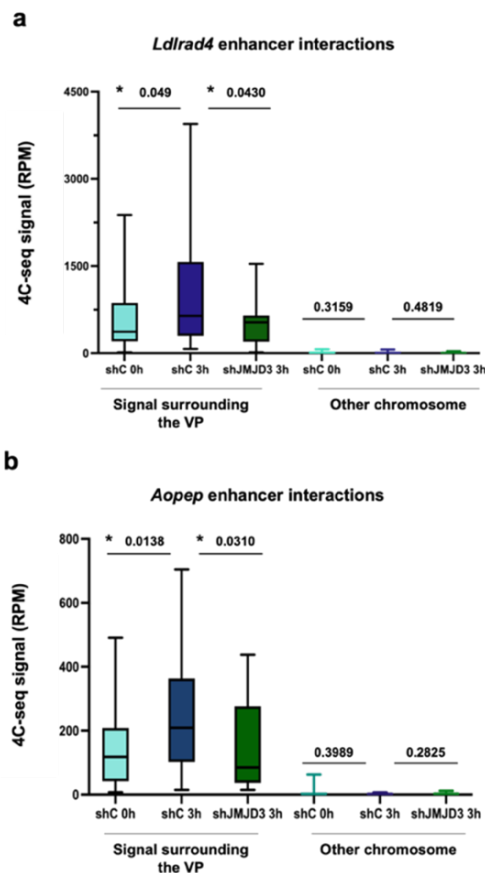
Then, we analyzed 4C-seq results. Figure R15 shows the interactions profile for *Ldlrad4* (a) and *Aopep* (b) locus. Indeed, it was clear that both VP enhancers established contacts with their surrounding cis-regulatory regions in a TGF $\beta$  and JMJD3 dependent manner.



**Figure R15.** UCSC Genome Browser caption showing interaction profile in the *Ldlrad4* (**a**) and *Aoep* (**b**) locus in control and JMJD3-depleted NSCs treated for three hours with TGF $\beta$ . Enhancers, JMJD3 and SMAD3 ChIP-seq signals are also represented. The light orange box indicates enhancer-promoter contacts, while yellow boxes show enhancer-enhancer contacts.



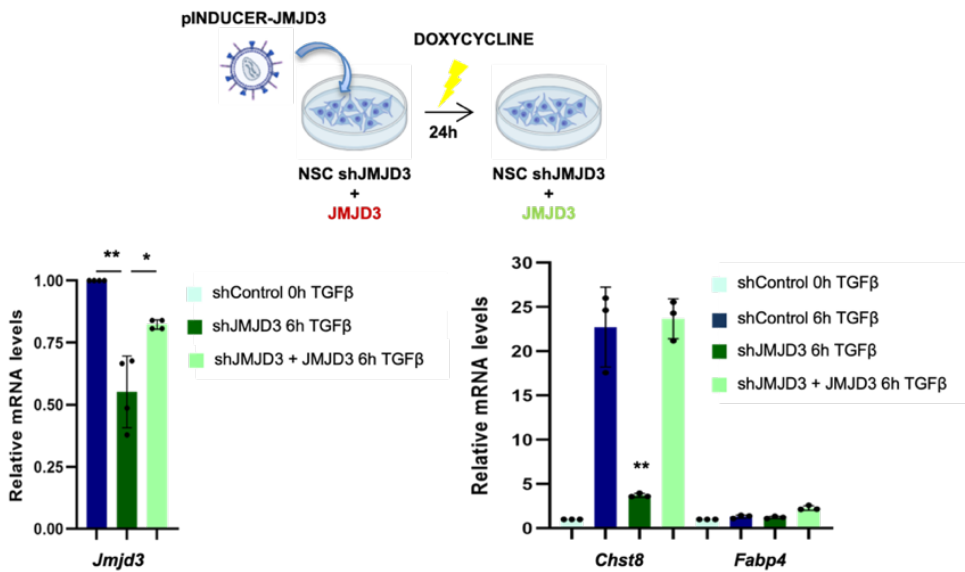
The quantifications of contacts occurring 500 kb upstream and downstream the VP (excluding the nearest  $\pm 20$  kb) in the two replicates clearly reflected this tendency. The RPMs of treated control NSCs increased with respect to the untreated cells (p-value 0.049 for *Ldlrad4* and 0.0138 for *Aoep*), while the ones observed for shJMJD3 decreased until reaching the levels seen previous addition of the cytokine (p-values 0.0430 for *Ldlrad4* and 0.0310 for *Aoep*). There were no changes in the negative control region (p-values 0.3159 and 0.4819 for *Ldlrad4*, and 0.3989 and 0.2825 for *Aoep*) (Figure R16).



**Figure R16.** Boxplot depicting the averaged RPMs from two combined biological independent replicates in the *Ldlrad4* locus (chr18:67,600,000-38,400,000) **(a)** and *Aoep* locus (chr13:62,680,000-63,600,000) **(b)**. As negative control, a region located in another chromosome (chr4:33,100,000-35,180,000) was chosen. p-values are the result of a Wilcoxon-Mann-Whitney test.

To further confirm that JMJD3 was necessary for the activation of the *Chst8* gene and for the establishment of the *Chst8* EC, we stably integrated inducible constructs containing JMJD3 into shJMJD3 NSCs (Figure R17).

This way, upon treatment with doxycycline for 24 hours, NSCs overexpressed JMJD3 up to levels that were similar to the endogenous ones (Figure R17). With this set-up, we first measured *Chst8* transcription in response to TGF $\beta$  to analyze whether gene expression was rescued upon JMJD3 re-expression. Figure R17 demonstrates that, indeed, this was what occurred.



**Figure R17.** (Top panel) Schematic representation of the experiment performed to rescue JMJD3 expression in NSCs depleted of the protein. Basically, doxycycline-inducible constructs containing JMJD3 were stably integrated into JMJD3-depleted NSCs. (Bottom panel) After 24 hours of doxycycline induction, relative expression levels of *Jmjd3* (left) and *Chst8* (right) gene in NSCs treated with TGF $\beta$  for 6 h were measured. The gene *Fabp4* was used as a negative control. Data are presented as mean values +/- SEM.

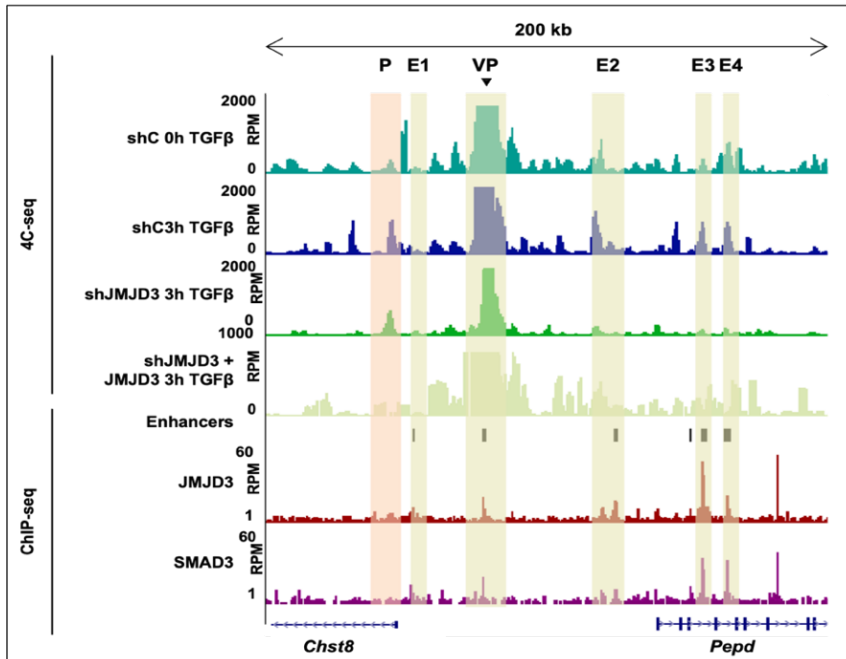
Next, we took advantage of our just generated system to assess whether the re-expression of JMJD3 also rescued the 3D-contacts that were lost in the *Chst8* locus upon JMJD3 depletion, which according to our hypothesis

should be the driving force of *Chst8* expression rescue. Thus, we repeated the 4C-seq assay in the shJMJD3 NSCs in which JMJD3 was re-expressed, in duplicate. As previously done, we first checked the quality of the experiment (Table R4).

Sample	Total number of reads	% of reads containing VP sequence	% of fragments that map in cis	% of reads that map in unique sites within 1Mb around VP
shJMJD3 + JMJD3 3h R1	9360610	95.07	68.97	76.41
shJMJD3 + JMJD3 3h R2	11266160	96.23	67.9	72.35

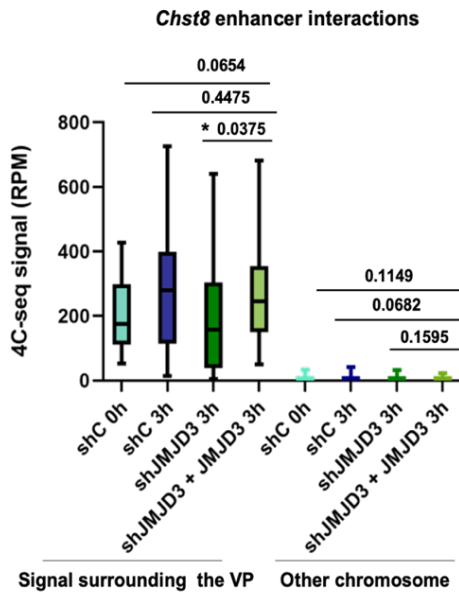
**Table R4.** Parameters used to assess the quality of the *Chst8* 4C-seq experiment for each sample.

Once confirmed the experiment was of enough quality, we analyzed the results. Figure R18 shows captions of the interaction profile obtained, which demonstrated that, as expected, the reintroduction of JMJD3 also rescued the establishment of contacts between *Chst8* EC elements.



**Figure R18.** UCSC Genome Browser captions showing the interaction profile in the *Chst8* EC conforming regions in NSCs control, JMJD3-depleted, and JMJD3 re-expressed (shJMJD3 + JMJD3) cells treated for 3 hours with TGFβ. Enhancers, JMJD3 and SMAD3 ChIP-seq signals are also represented. The light orange box indicates enhancer-promoter contacts, while yellow boxes show enhancer-enhancer contacts.

The recovery became patent when quantifying the contacts occurring 500 kb around the VP (excluding the nearest  $\pm 20$  kb, chr7:33841896-35860773) in the two replicates. The 4C-seq signal of NSCs shJMJD3 re-expressing JMJD3 increased up to the levels of control treated NSCs (p-values 0.0475) in comparison with the ones observed for the shJMJD3 NSCs (p-values 0.0375), while in the independent negative control region there were no changes (Figure R19).



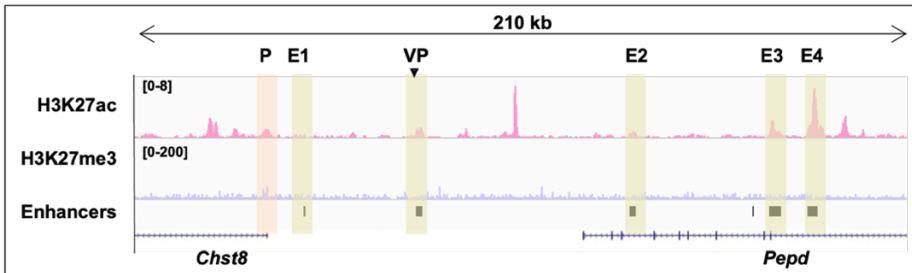
**Figure R19.** Boxplot depicting the averaged RPMs from two biological independent replicates. As negative control, a region located in another chromosome (chr4:33076383-35216108) was chose. p-values are the result of a Wilcoxon-Mann-Whitney test.

Altogether, these results suggested that the TGF $\beta$  signal and the epigenetic factor JMJD3 regulated gene expression through 3D-chromatin structure modulation.

### 1.5. JMJD3's role mediating contacts seems to be independent of its demethylase activity

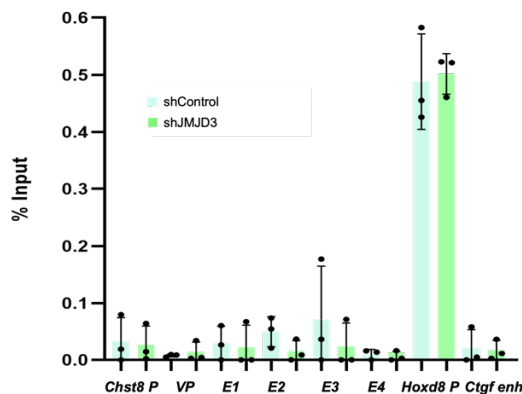
As JMJD3 seemed to be fundamental for the establishment and/or maintenance of contacts in the 3D-genome space, we deepened into the JMJD3 intrinsic features that made it amenable for such a function. Regarding the protein structure, it just had one known domain, the demethylase JmjC catalytic domain, which was previously demonstrated by our lab to be unnecessary for the activation of some neural enhancers in response to TGF $\beta$  (67). Nevertheless, we decided to analyze whether the demethylation capacity of the protein was playing a role in the *Chst8* EC JMJD3-mediated activation. To this aim, we used H3K27me3 ChIP-seq data from NSCs (66) and analyzed the methylation levels of the *Chst8* EC.

Figure R20 demonstrates that *Chst8* locus was not methylated before the activation of the TGF $\beta$  pathway.



**Figure R20.** USCS Genome Browser capture showing the H3K27me3 and H3K27ac levels in the *Chst8* locus in untreated NSCs. Previously defined enhancers are also depicted.

Furthermore, we performed ChIP-qPCR experiments in NSCs to study whether H3K27me3 increased in the locus when JMJD3 was depleted. As seen in Figure R21, this was not the case for any of the enhancers from the *Chst8* EC nor for *Chst8* promoter. The *Hoxd8* promoter and *Ctgf* enhancer were used as H3K27me3 positive and negative controls, respectively.



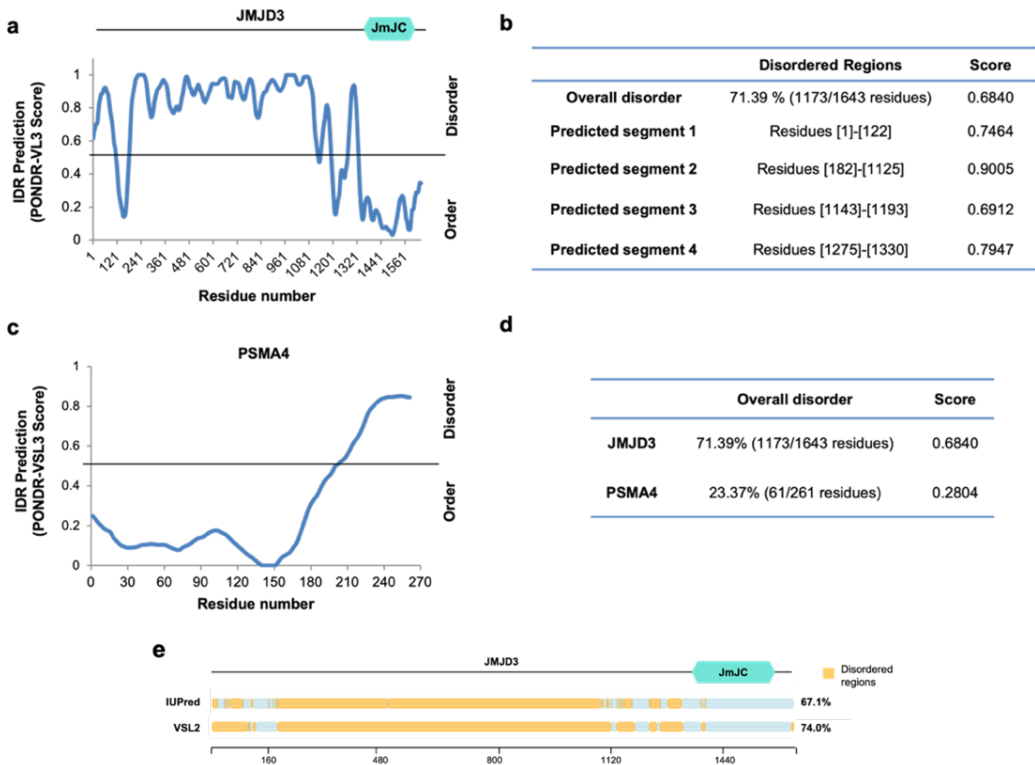
**Figure R21.** Levels of H3K27me3 in the *Chst8* promoter and EC in control and JMJD3-depleted NSCs. The promoter of *Hoxd8*, highly trimethylated, was used as a positive control; the enhancer of *Ctgf*, non-trimethylated, was used as a negative control.

All considered, JMJD3 demethylase activity did not seem to be the underlying motor driving *Chst8* EC formation nor activation.

### **1.6. JMJD3 is a highly disordered protein with potential to phase separate**

The catalytic domain of JMJD3 was discarded to be impacting the 3D-structure of chromatin at the *Chst8* locus. Thus, we decided to examine the rest of the protein, which doesn't contain any other known domain. In fact, it consists of a huge unstructured, disordered domain. Lastly, these so-called intrinsically disordered regions (IDRs) have been related to transcriptional regulation through the formation of biomolecular condensates by phase separation processes ((92,199,244–246) , see “Introduction” section 4).

Bearing this in mind, we analyzed JMJD3 amino acid sequence features, starting by the level of disorder of the protein. Taking advantage of the validated algorithm PONDR-VL3 (231), we confirmed that JMJD3 presented a highly disordered structure. Indeed, JMJD3 disordered score (0.6840) was considerably higher than the one obtained for PMS4A (0.2804), a well-known structured protein that was used as a negative control for the analysis. Furthermore, the percentage of amino acids that exist in disordered domains was also much higher in JMJD3 (71.39% for JMJD3 as compared to the 23.37% of PSMA4) (Figure R22a-d). This was further reinforced by other two predictors (IUPred (232) and VSL2 (247)), which also returned a significant proportion of disordered regions for JMJD3 (Figure R22e). This disorder prediction analysis allowed us to identify a strikingly long region that spanned 943 amino acids (residues from 182 to 1125) and presented a high disorder score (0.9). We named this region as JMJD3 IDR.

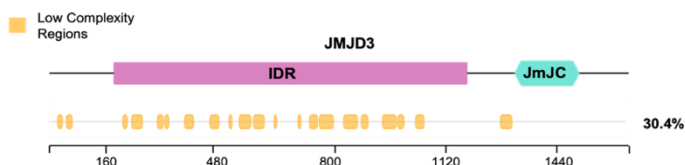


**Figure R22.** (a) Graphic that shows the disorder prediction of human JMJD3 using PONDR-VL3 algorithm. (b) Table where the disorder score and the lengths of the predicted disordered regions are indicated. The length of each disordered segment is above 50 amino acids. (c) Graphic depicting the disorder prediction of human PSMA4 using PONDR-VL3 algorithm. The protein was used as a negative control for our analysis. (d) Table comparing overall percentage of disorder and disorder scores for JMJD3 (disordered) and PSMA4 (ordered) proteins. (e) Schematic representation of JMJD3 disordered regions (yellow patches) predicted by IUPred and VSL2 algorithms, as well as the percentage they represent within the protein sequence.

Not only the disorder, but also the nature of the amino acids that compose IDRs has been shown to be fundamental for the IDR-mediated transcription



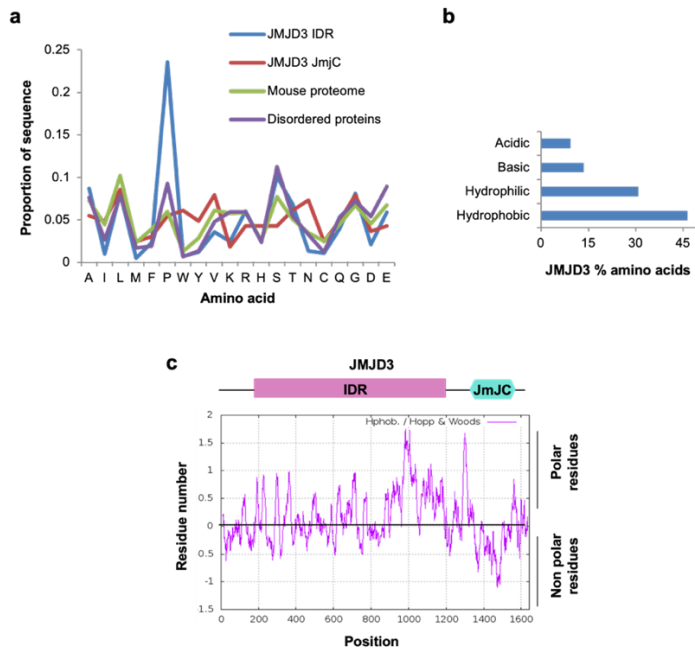
regulation. In fact, disordered regions frequently coincide with domains that are biased for certain amino acids (245,248,249). These are the so-called low-complexity domains. We used SEG algorithm (236) to check whether this was the case for JMJD3 and discovered that 30.4% of JMJD3 sequence was predicted to contain low-complexity segments, most of them within JMJD3 IDR (Figure R23).



**Figure R23.** Schematic depiction of JMJD3 low-complexity domains (yellow boxes) predicted by SEG algorithm. The percentage they represent within the protein sequence (30.4%) is included.

Moreover, when examining JMJD3 amino acid sequence composition (see “Materials and Methods”) we found a striking abundance of prolines (24% of total amino acids). In fact, these residues conformed tracks that spanned as long as 20 amino acids (Figure R24a) that were widespread conserved among vertebrates (Figure R24b).

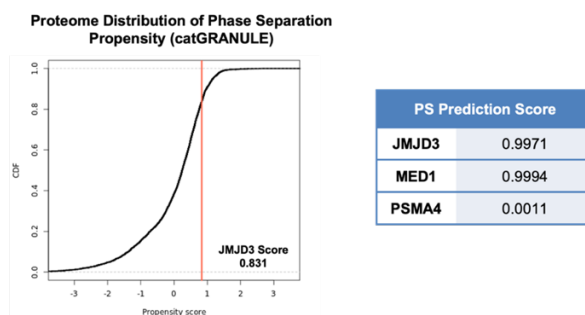




**Figure R25. (a)** Graph representing the amino acid composition of JMJD3 IDR, JMJD3 catalytic domain, mouse proteome and disordered proteins. The latter were defined by the presence of 50 residues fragment whose IUPRED median score was at least 0.55 and that was not found in Pfam (so that functional domains were avoided). **(b)** Bar plot that shows the percentages of acid, basic, hydrophobic, and hydrophilic amino acids within JMJD3 sequence. **(c)** JMJD3 hydrophobicity profile, obtained by means of the ExPASy (252) website with the Hopp and Woods scale and a sliding window of 21. Non polar residues are displayed below an score of 0, while polar residues are scored above 0.

Nowadays, it is well-known that both hydrophobic and electrostatic interactions are major players in biomolecular condensate formation (see section 4 of the “Introduction”). Indeed, prolines were described as hydrophobic residues that concatenate to generate sticky domains with the capacity to bind other proteins in a rapid and reversible manner (253). Thus, the sequence features of JMJD3 seemed to favor its potential to be involved in phase separated condensates. Given this, we used catGRANULE (235)

and PSPredictor (254) to predict the capacity of JMJD3 to phase separate. Figure R26 shows the scores that both algorithms gave to the demethylase (0.83 and 0.99), which were similar to known phase-separated proteins, such as the Mediator complex subunit MED1 (0.99), and higher than our negative control proteasome protein PSMA4 (0.001).



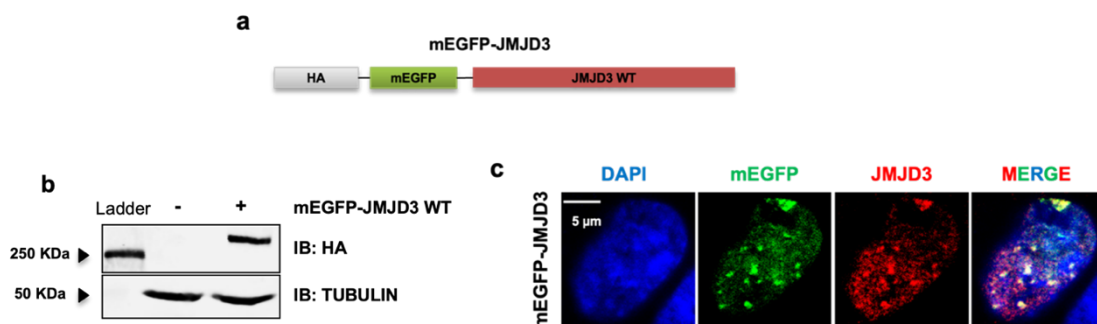
**Figure R26.** Phase separation predictions obtained with catGRANULE (left) and PSPredictor (right) algorithms. The left-handed graph shows JMJD3 phase separation propensity score with respect to the propensity distributions of the whole proteome. The right-sided table includes a comparison of the scores obtained for JMJD3, a positive control (MED1) and a negative control protein (PSMA4).

This analysis pointed to JMJD3 as a highly disordered protein able to be involved in phase separation processes.

### 1.7. JMJD3 undergoes liquid-liquid phase separation *in vitro* and *in vivo*

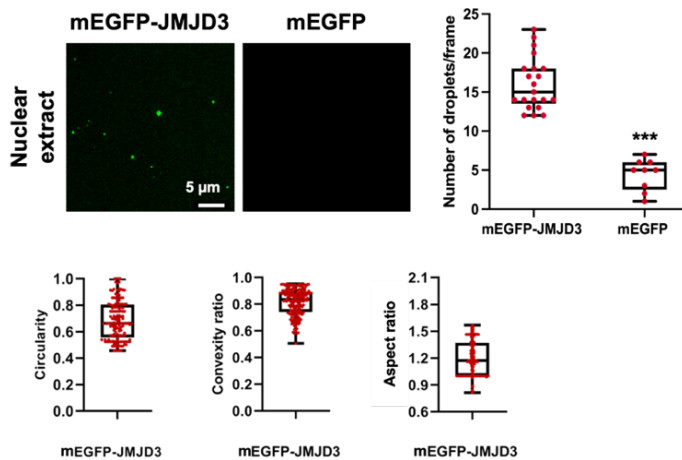
Given the above-mentioned results, we hypothesized that JMJD3 could be involved in the formation of, or be incorporated into already constituted, phase-separated condensates. To assess our hypothesis, we constructed a vector that expressed JMJD3 fused to the monomeric EGFP (mEGFP) fluorescence protein and the HA tag. Then we transfected the vector in

HEK293T cells and prepared total protein extracts, so that we could test the expression of the resulting protein (Figure R27).



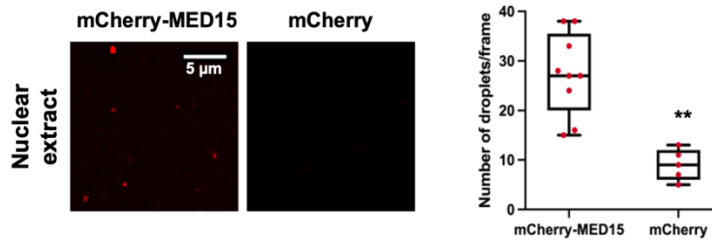
**Figure R27. (a)** Scheme representing the vector of the fluorescent JMJD3 fusion protein. **(b)** mEGFP-JMJD3 vector was transfected into HEK293T cells, and total protein extracts were prepared to determine JMJD3 and Tubulin levels by immunoblot. HA-tag and Tubulin antibodies were used, respectively. **(c)** Immunofluorescence of HEK293T cells expressing mEGFP-JMJD3 vector. Cells were stained with JMJD3 antibody, DNA was visualized with DAPI, and merged image overlaps JMJD3 staining and mEGFP intrinsic fluorescent signal.

Then, we performed an *in vitro* droplets formation assay using nuclear extracts of HEK293T cells overexpressing mEGFP-JMJD3. Figure R28 demonstrates that mEGFP-JMJD3 formed droplets that didn't appear when overexpressing mEGFP alone, so that they could be attributed to JMJD3. Moreover, these droplets showed previously defined liquid-like features (183,255,256), such as a circularity, convexity and aspect ratios close to 1, as expected for a perfect circle (this is, a liquid-like droplet) (Figure R28).



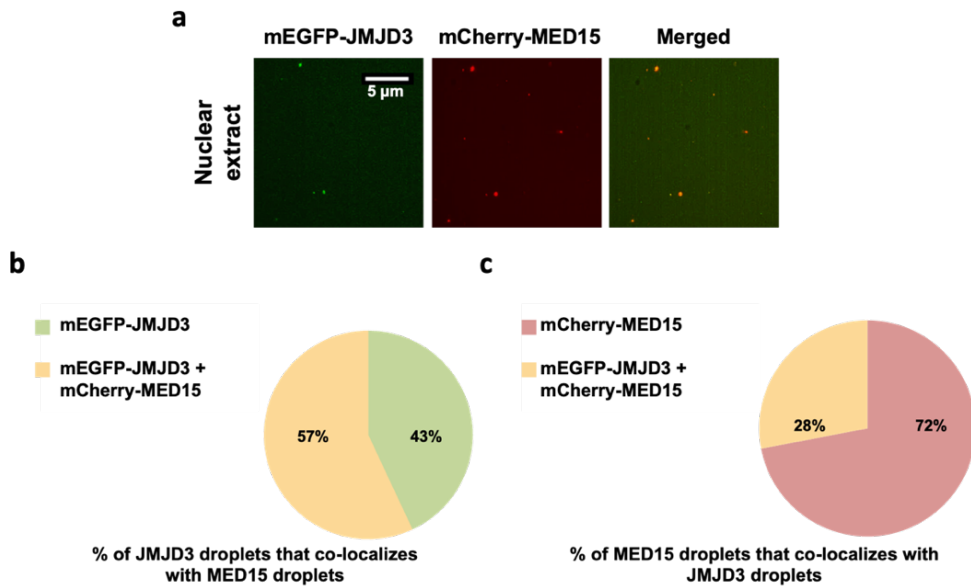
**Figure R28.** Droplet formation assay using nuclear extracts of HEK293T cells overexpressing either mEGFP-JMJD3 or mEGFP, performed at room temperature in the presence of 150 mM NaCl (top, left panels). The number of droplets per frame observed for each condition are shown in the right-handed graph. Data are the mean  $\pm$  SEM. Droplets circularity, convexity and aspect ratio were also calculated; they are depicted in the bottom boxplots. Droplets in 5 fields in each group from three biologically independent experiments were quantified.

As we were speculating that JMJD3 participates in phase separation processes at enhancers, and Mediator proteins are known to phase separate (225,246,248), we decided to check whether some of the JMJD3 droplets were shared with the subunit MED15 of the Mediator complex. First, we studied MED15 capacity to form droplets through droplet formation assays with nuclear extracts from HEK293T cells that overexpressed mCherry-MED15. Figure R29 shows the droplets formed by MED15.



**Figure R29.** Droplet formation assay using nuclear extracts of HEK293T cells overexpressing either mCherry-MED15 or mCherry proteins, done at room temperature in the presence of 150 mM NaCl (left panels). Quantifications of the number of droplets per frame are depicted in the right-handed graph. Data are the mean  $\pm$  SEM. Droplets in 5 fields in each group from three biologically independent experiments were quantified.

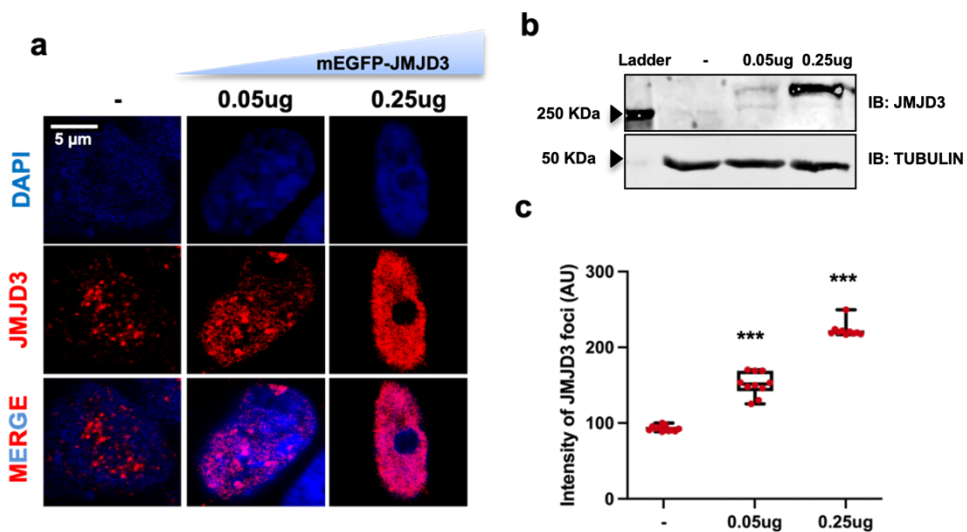
After this checking, we co-transfected HEK293T cells with both mEGFP-JMJD3 and mCherry-MED15 and performed a droplet formation assay. It can be seen in Figure R30 that, indeed, both proteins were found to co-localize at many of the observed droplets. Specifically, about 57% of JMJD3 droplets were shared with MED15, while 28% of MED15 droplets co-localized with the demethylase (Figure R30, bottom pie charts).



**Figure R30.** (a) Droplet formation assay using nuclear extracts of HEK293T cells overexpressing both mEGFP-JMJD3 and mCherry-MED15 proteins, done at room temperature in the presence of 150 mM NaCl. (b-c) Pie charts that show the percentage of JMJD3 droplets that colocalize with MED15 (b) and vice-versa (c) are displayed.

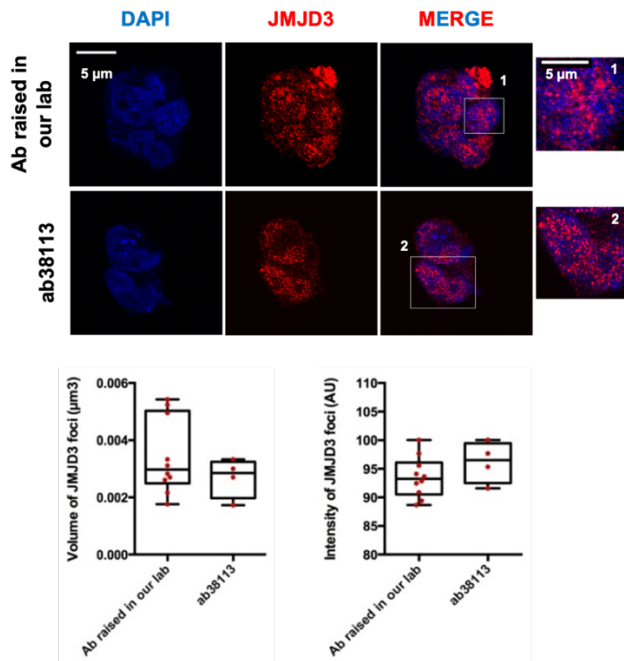
Then, we decided to check the formation of JMJD3-containing condensates in cells through immunofluorescence experiments. Figure R31 shows the formation of nuclear puncta by JMJD3, both endogenously and when overexpressed in HEK293T cells. Interestingly, we also saw that the size and intensity of nuclear puncta increased along with the nuclear JMJD3, even aggregating when high concentrations of the plasmid were expressed. This is characteristic of phase separated condensates, which, as previously commented, are highly dependent on the concentration (see section 4 of the “Introduction”).





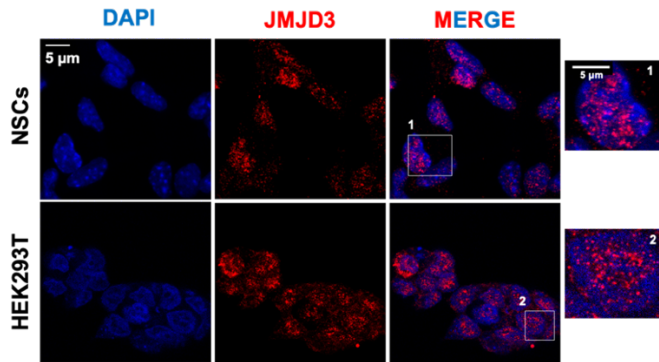
**Figure R31.** HEK293T cells were transfected with increasing amounts of mEGFP-JMJD3 plasmid. **(a)** Immunofluorescence showing the nuclear puncta formed by endogenous and overexpressed mEGFP-JMJD3, as detected by staining with JMJD3 antibody. Images also show aggregation when 0.25 ug of vector are transfected. **(b)** Western blot displaying the levels of endogenous or overexpressed JMJD3 for each condition. **(c)** The intensity of endogenous or overexpressed JMJD3 nuclear puncta was quantified. Data show the mean  $\pm$  SEM.

To discard that the observed nuclear puncta was an artifact resulting from JMJD3 overexpression or from an antibody cross-reaction, we checked again the endogenous formation of puncta through immunofluorescence with two different JMJD3 antibodies (the one raised in our lab (67), and the abcam ab38113 one; see table M3 in “Materials and Methods”). As seen in Figures R31 and R32, endogenous JMJD3 did form puncta in all the tested conditions.



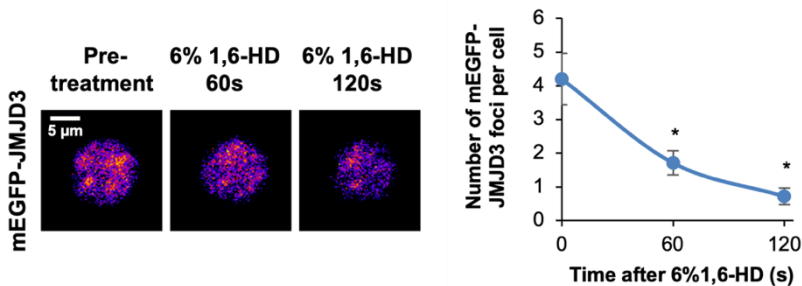
**Figure R32.** Endogenous JMJD3 nuclear puncta in HEK293T cells, as detected with two different antibodies (the one risen in our lab and a commercial one) that gave similar results in term of condensates volume and intensity (quantified at the bottom of the figure).

Furthermore, we checked whether the formation of nuclear condensates depended on the cell type. We did immunofluorescence of the endogenous JMJD3 protein in HEK293T and NSCs, just using the antibody raised in our lab. Results shown in Figure R33 demonstrate that the formation of endogenous JMJD3 nuclear puncta occurred independently of the cell type.



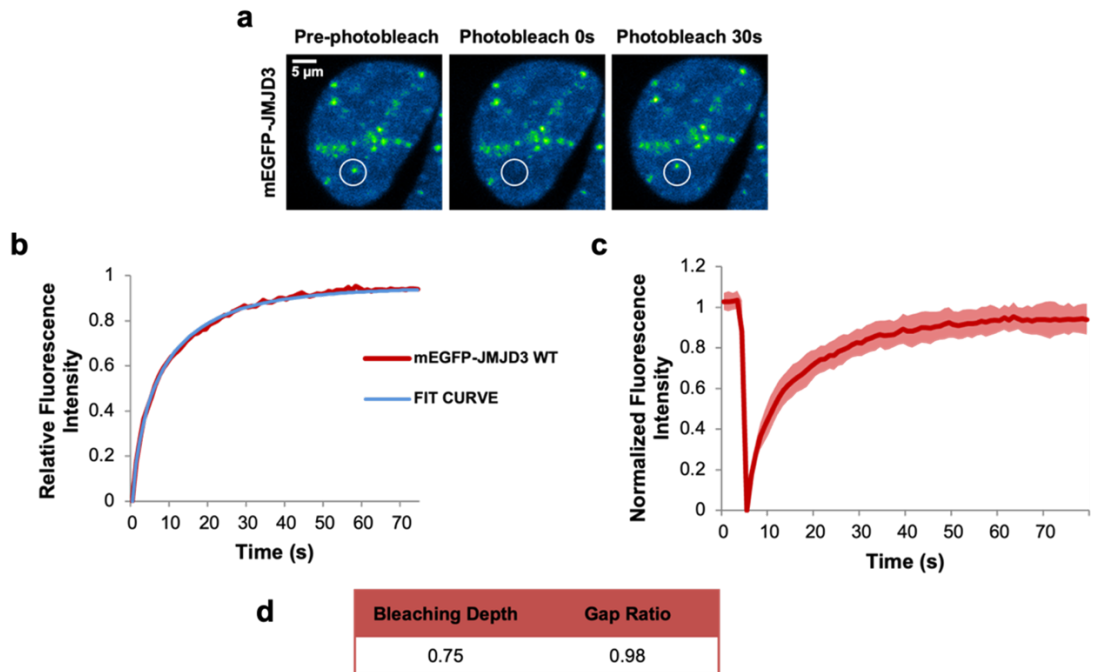
**Figure R33.** Immunostaining of endogenous JMJD3 nuclear puncta in NSCs and HEK293T cells, detected by means of the JMJD3 antibody risen in our lab.

Then, given the previously shown high hydrophobicity of JMJD3, we decided to check the sensibility of its condensates to the aliphatic alcohol 1,6-Hexanediol (1,6-HD). This compound has been demonstrated to disrupt the hydrophobic interactions underlying phase-separated droplets (257), so we expected a reduction in the number and/or size of condensates. Indeed, Figure R34 shows that these reductions were seen after treating HEK293T cells overexpressing mEGFP-JMJD3 with 6% of 1,6-HD for 60 or 120 seconds.



**Figure R34.** Left panels correspond to HEK293T cells overexpressing mEGFP-JMJD3 that were treated with 6% 1,6-HD for 5 minutes and imaged at 60 and 120 s. Nuclei were visualized with DAPI. The right-sided graph depicts the quantification of JMJD3 nuclear foci per cell along the treatment. Data are the mean  $\pm$  SEM.

JMJD3 droplets were characterized by possessing typical features of a liquid-like nature (Figure R28). To discern whether JMJD3 nuclear condensates also displayed liquid-like properties, we decided to examine their dynamic nature by measuring the FRAP rate (see “Methods”) in HEK293T cells, as the rapid diffusion of the constituent molecules of condensates has been described as a liquid-like feature (92,179,258). We observed that after photobleaching, the fluorescence recovery of mEGFP-JMJD3 condensates was almost complete on a time scale of seconds (Figure R35a-c). This was similar to the behavior observed for other proteins that form liquid-like condensates (259). The quality of the experiment was assessed by analyzing some key parameters, such as the loss of fluorescence in the bleached region (bleaching depth, which in this case was 75%) and the total fluorescence remaining in the cell after bleaching the desired region (gap ratio, 98% in our hands) (Figure R35d). We also calculated the molecular pool that underwent exchange within the bleached zone (mobile fraction), which corresponded to 0.95. This is, 95% of the molecules contained within JMJD3 condensates were mobile. Furthermore, the aggregates that appeared when high levels of JMJD3 were expressed displayed reduced mobility in these assays.



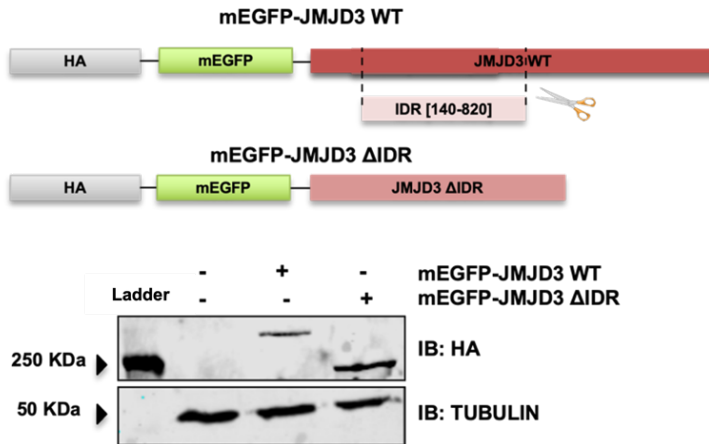
**Figure R35. (a)** FRAP experiment in HEK293T cells overexpressing mEGFP-JMJD3. Images represent pre-bleaching, bleaching ( $t=0$  s) and post-bleaching ( $t=30$  s) points. **(b)** Fit curve to a double-exponential smoothing ( $R^2=1$ ), where bleaching occurs at  $t=0$  s. Data are plotted as background-subtracted and normalized mean. **(c)** The graph represents the background-subtracted and normalized fluorescence intensities relative to a pre-bleach time point ( $t=0$  s). Data are plotted as mean normalized  $\pm$  SD. **(d)** Table that depicts the bleaching depth and gap ratio parameters.

Taken together, these data demonstrated that JMJD3 formed both *in vitro* droplets and *in vivo* nuclear condensates that showed liquid-like properties.

### 1.8. JMJD3 IDR is necessary for condensate formation

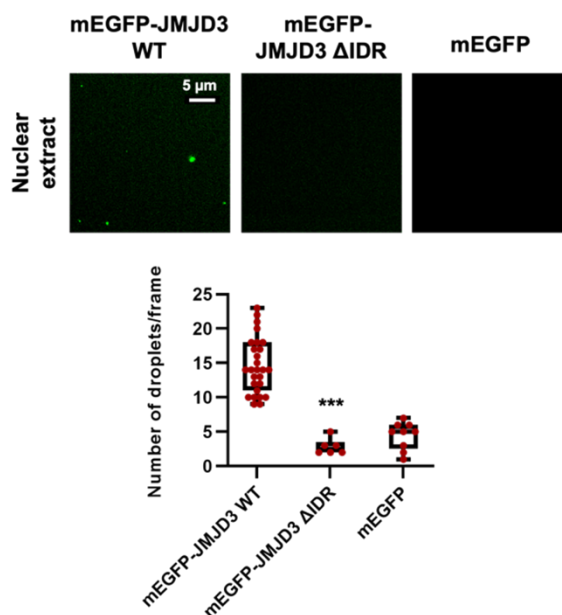
Once JMJD3 potential to form biomolecular condensates was assessed, we continued investigating whether its IDR was necessary for the protein to be involved in the process. To do this, we constructed a mutant version

of JMJD3 that lacked the IDR domain (residues 140-820) and named the new construct mEGFP-JMJD3  $\Delta$ IDR. First, we checked the expression of the resulting fluorescent fusion protein by Western Blot (Figure R36).



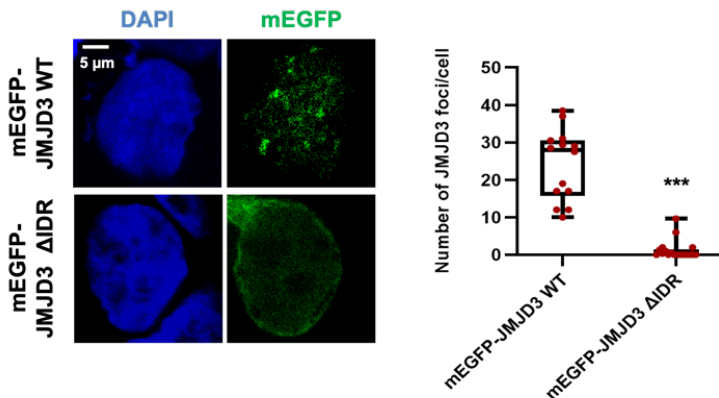
**Figure R36.** Scheme depicting the new vector constructed from the original mEGFP-JMJD3 to overexpress JMJD3  $\Delta$ IDR (top). mEGFP-JMJD3 WT and mEGFP-JMJD3  $\Delta$ IDR were transfected into HEK293T cells, total cell extracts were prepared, and the fusion proteins were detected by Western Blot using antibodies against HA (bottom).

Secondly, we performed a droplet formation assay with nuclear extracts from HEK293T cells overexpressing either mEGFP-JMJD3 WT or mEGFP-JMJD3  $\Delta$ IDR. Figure R37 shows the incompetence of JMJD3  $\Delta$ IDR to form droplets.



**Figure R37.** Droplet formation assay using nuclear extracts of HEK293T cells overexpressing either mEGFP-JMJD3 WT, mEGFP-JMJD3 ΔIDR or mEGFP proteins, performed at room temperature in the presence of 150 mM NaCl (left panels). Quantifications of the number of droplets per frame are depicted on the right-handed graph. Data are the mean ± SEM. Droplets in 5 fields in each group from three biologically independent experiments were quantified.

Finally, we proceeded to investigate JMJD3 ΔIDR capacity to form nuclear puncta in HEK293T cells. Although both, mutant and WT proteins, were distributed between nucleus and cytoplasm, the ΔIDR mutant showed some bias toward the cytoplasm. However, JMJD3 ΔIDR couldn't form any kind of condensate independently of its location (Figure R38).



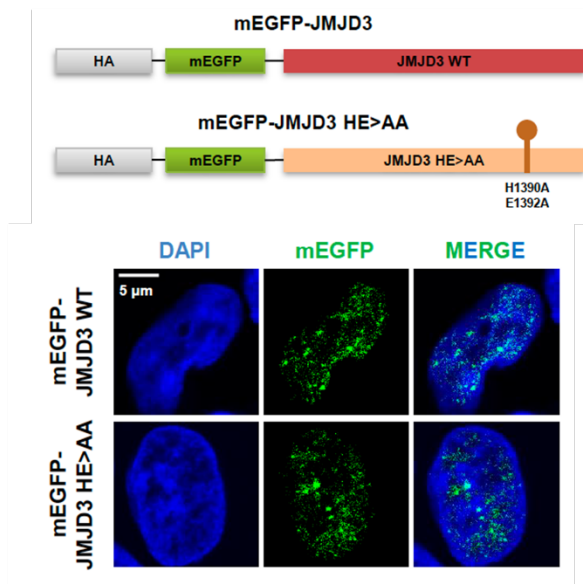
**Figure R38.** Immunofluorescence experiment of HEK293T transfected with mEGFP-JMJD3 WT or mEGFP-JMJD3  $\Delta$ IDR (left-sided panels). The intrinsic fluorescence of the molecules was followed to detect proteins. The number of JMJD3 foci per cell was quantified for each condition (right-handed boxplot). Data show the mean  $\pm$  SEM.

These data favored the hypothesis that the IDR domain was necessary for JMJD3 to form or to be part of biomolecular condensates.

### 1.9. JMJD3 demethylase activity is not necessary for condensate formation

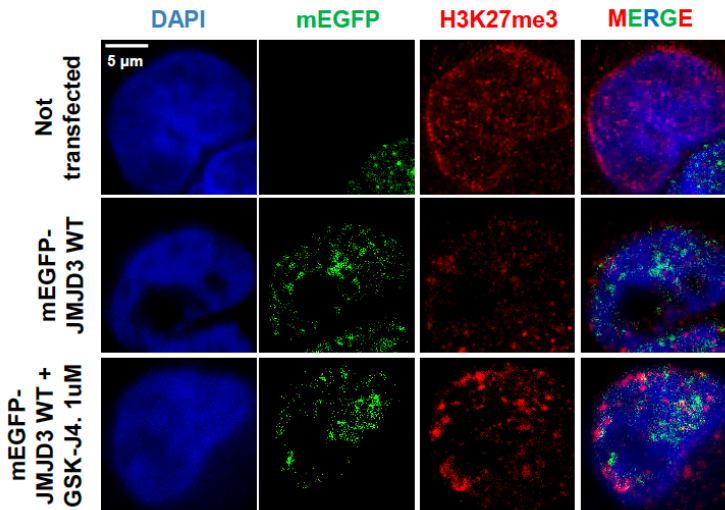
As we assessed the major role of the IDR domain for condensates formation, we decided to evaluate the relevance of JMJD3 catalytic domain for this process. Thus, we constructed a plasmid containing a fluorescent version of JMJD3 mutated at the catalytic domain (JMJD3 HE>AA), which lacked the demethylating capability and functioned as a dominant negative form of JMJD3 (66) (Figure R39). Immunofluorescence experiments done in HEK293T cells overexpressing either JMJD3 WT or JMJD3 HE>AA demonstrated that both proteins were able to form an equal number of nuclear condensates, which were also very similar in intensity and volume (Figure R39 and R41).





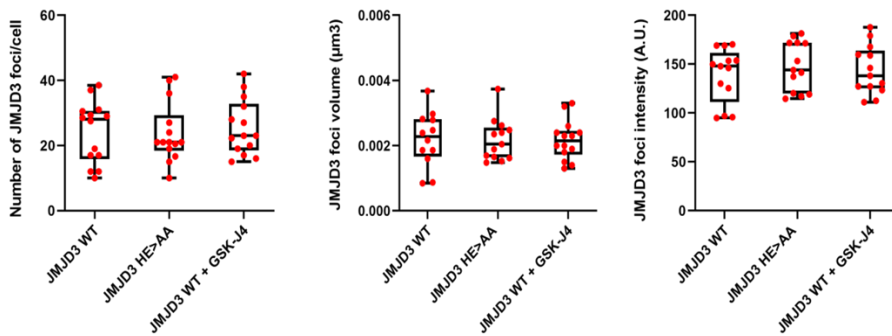
**Figure R39.** On the upper panel, scheme representing the vector constructed to overexpress the catalytic mutant form of JMJD3, JMJD3 HE>AA. Below, immunofluorescence of HEK293T cells that overexpressed JMJD3 WT or JMJD3 HE>AA. Proteins were detected by following their intrinsic fluorescence.

This result suggested that the catalytic activity of JMJD3 was not necessary for the formation of condensates. To further confirm the finding, we took advantage of a specific inhibitor of JMJD3 catalytic activity named GSK-J4 (260). After treating HEK293T cells for 6 hours with the inhibitor, we saw the same number of nuclear condensates than in untreated cells (Figure R40 and R41). This time of treatment was enough to effectively inhibit JMJD3 demethylase capacity, as reflected by the levels of H3K27me3 observed in treated cells (Figure R40).



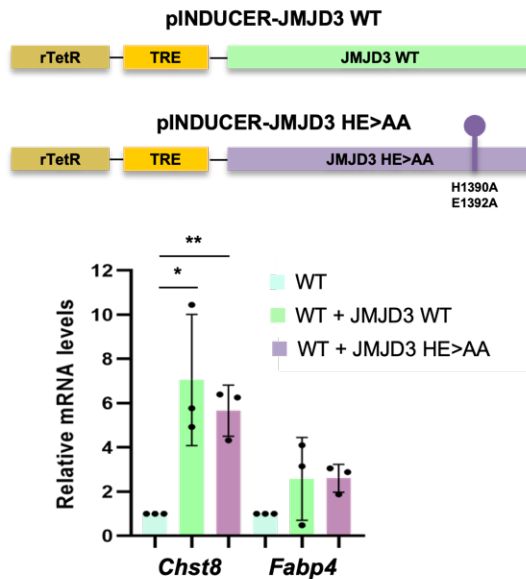
**Figure R40.** Immunofluorescence experiment of HEK293T cells overexpressing JMJD3 WT, either untreated or treated for 6 hours with 1uM of the JMJD3 catalytic activity inhibitor GSK-J4. JMJD3 condensates were detected by following the intrinsic fluorescence of the overexpression protein, H3K27me3 antibody was used to detect the histone mark levels, and DAPI served to visualize nuclei. Not transfected HEK293T cells were used as a control to detect endogenous levels of H3K27me3.

Quantifications of the number, intensity and volume of the observed condensates demonstrated that all of them (the ones formed by JMJD3 WT, either in untreated or GSK-J4 treated cells, and JMJD3 HE>AA protein) were indeed very similar (Figure R41).



**Figure R41.** Graphs showing the number, volume and intensity of the condensates formed by cells overexpressing either JMJD3 WT or JMJD3 HE>AA, and also those from cells expressing JMJD3 WT and treated with GSK-J4 inhibitor.

All these results supported that JMJD3 condensate formation capacity was independent of its catalytic activity. We further wanted to investigate whether this independency could be extended to the transcriptional activation promoted by JMJD3 at the *Chst8* locus, so we constructed inducible overexpression vectors that contained JMJD3 WT or JMJD3 HE>AA and transfected them in HEK293T cells, where the TGF $\beta$  pathway is already active. After inducing the expression of these proteins for 24 hours with doxycycline, we analyzed the transcription levels of the *Chst8* gene. As presumed, Figure R42 shows that the catalytic mutant produced similar levels of transcription as the WT protein.



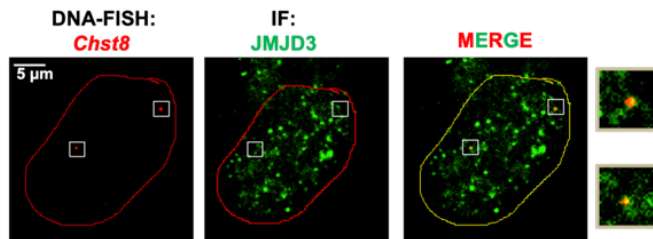
**Figure R42.** In the top part of the figure, scheme representing the inducible vectors used to overexpress JMJD3 WT or JMJD3 HE>AA upon induction with doxycycline for 24 hours. Below, relative transcription levels of *Chst8* gene were measured in HEK993T cells overexpressing JMJD3 WT or JMJD3 HE>AA. The gene *Fabp4* was used as a negative control.

Altogether, these results demonstrated that the catalytic activity of JMJD3 was not involved neither in condensates formation nor in the transcriptional activation of our studied locus.

### 1.10. JMJD3 promotes *Chst8* enhancer cluster assembly and gene transcription

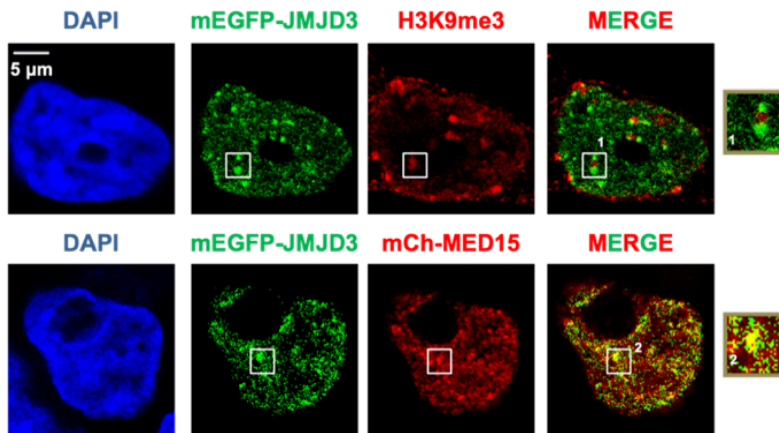
Wrapping up all our data, along this work we demonstrated that JMJD3 regulated the transcriptional activation of the *Chst8* locus by means of its 3D-structure reorganization, and that it was involved in phase separation processes. Thus, we hypothesized that JMJD3 nuclear condensates could be contributing to the JMJD3-mediated activation of the *Chst8* locus.

To address our hypothesis, we performed an immuno-FISH experiment in 3 hours TGF $\beta$ -treated NSCs to co-image endogenous JMJD3 condensates and the *Chst8* locus. The details of the DNA probe to target the locus, as well as the methodology employed, can be seen in the “Materials and Methods” section of the manuscript. As shown in Figure R43, a clear colocalization of the *Chst8* locus and JMJD3 condensates was observed.



**Figure R43.** Immuno-FISH experiment for endogenous JMJD3 (stained with JMJD3 antibody, in green) and the *Chst8* locus (DNA probe, in red) in NSCs treated for 3 hours with TGF $\beta$ . The right-sided boxes show the co-localization of JMJD3 nuclear puncta and the FISH signal (zoom-in).

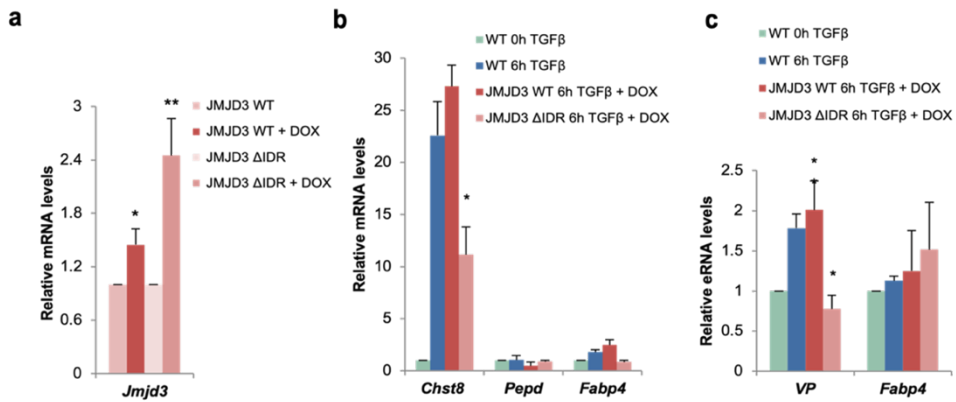
So, this result demonstrated that the JMJD3 protein that targeted *Chst8* locus was contained within condensates. Furthermore, to reinforce the putative involvement of JMJD3 condensates in transcriptional activation, we analyzed their exclusion or/and co-localization with the repressive mark H3K9me3 and the protein MED15, which was characteristic of active transcriptional sites and was previously shown to co-localize with JMJD3 droplets (Figure R30). The immunofluorescence shown in Figure R44 confirmed that, as expected, JMJD3 condensates excluded repressed sites, while co-localized with active loci.



**Figure R44.** Immunofluorescence experiment in HEK293T cells that expressed mEGFP-JMJD3 alone (upper panel) or together with mCherry-MED15 (bottom panel). Cells from the upper panel were stained with H3K9me3 antibody, while the fluorescence of mCherry and mEGFP were used to follow MED15 (for the bottom panel cells) and JMJD3, respectively. Their exclusion or co-localization with JMJD3 was analyzed, as represented by the zoomed-in boxes on the right-side of the figure. DAPI was used to visualize nuclei.

Basically, we found that JMJD3 condensates were related to transcription activation, and that they were specifically found in the *Chst8* locus, correlating the involvement of phase separation with the activation of the locus. To address whether the IDR domain, and hence JMJD3 condensates, were necessary for the expression of the VP and *Chst8*, we analyzed the transcriptional levels of these regions upon stably integrating inducible constructs to overexpress JMJD3 WT or JMJD3  $\Delta$ IDR in NSCs. Figure R45 shows that, after doxycycline and TGF $\beta$  stimulation, JMJD3  $\Delta$ IDR was not able to induce the expression of the gene *Chst8*, nor the VP enhancer, to the same levels as JMJD3 WT did, despite the mutant expression was higher than the WT one (Figure R45a). In fact, the mutant protein acted as a dominant negative, decreasing the expression levels of both *Chst8* gene (Figure 45b) and VP enhancer (Figure 45c) with respect

to the levels achieved in TGF $\beta$  treated cells that do not overexpress any form of JMJD3 (NSCs WT).



**Figure R45.** NSCs that stably overexpress the inducible constructs JMJD3 WT or JMJD3  $\Delta$ IDR were treated with doxycycline for 24 hours. After induction, the relative expression levels of *Jmjd3* gene (a), *Chst8* gene (b) and VP enhancer (c) in untreated or 6 hours TGF $\beta$ -treated cells were measured. *Pepd* and *Fabp4* genes, and *Fabp4* enhancer, were used as negative controls. Data are presented as mean values +/- SEM.

Thus, we concluded that the expression of the IDR domain, and presumably the formation of phase separated JMJD3-containing condensates, were necessary to achieve a proper TGF $\beta$ -driven transcriptional response at the *Chst8* locus.





## **Chapter 2: Analysis of the contribution of JmjC-lysine demethylases common domains to gene expression**

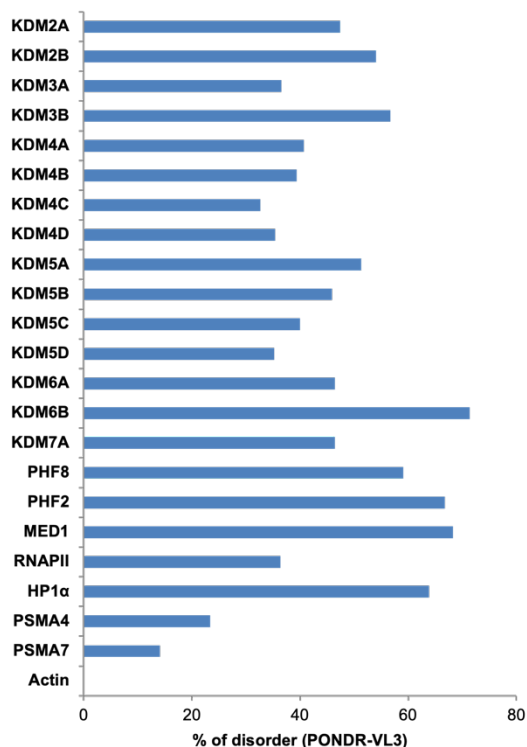
Along the first chapter of this manuscript, we described how the well-known histone demethylase JMJD3 is able to control transcriptional activation in a catalytic independent manner. The obtained results led us to hypothesize that phase separation might be a general mechanism utilized by JmjC demethylases to control their biological function. Thus, the second aim of my thesis was to examine whether these proteins possess any common structural feature, apart from their catalytic domain, that influences their fundamental roles for the proper cellular functioning, focusing specially on their transcriptional regulatory capacity.

### **2.1. JMJC-KDMs are highly disordered proteins**

This analysis started with the examination of JMJC-KDMs structure. Although most of these proteins contain multiple domains (261), a large part of them do not present a defined structure. Taking this into account, we proposed that JMJC-KDMs could belong to the group of intrinsically disordered proteins.

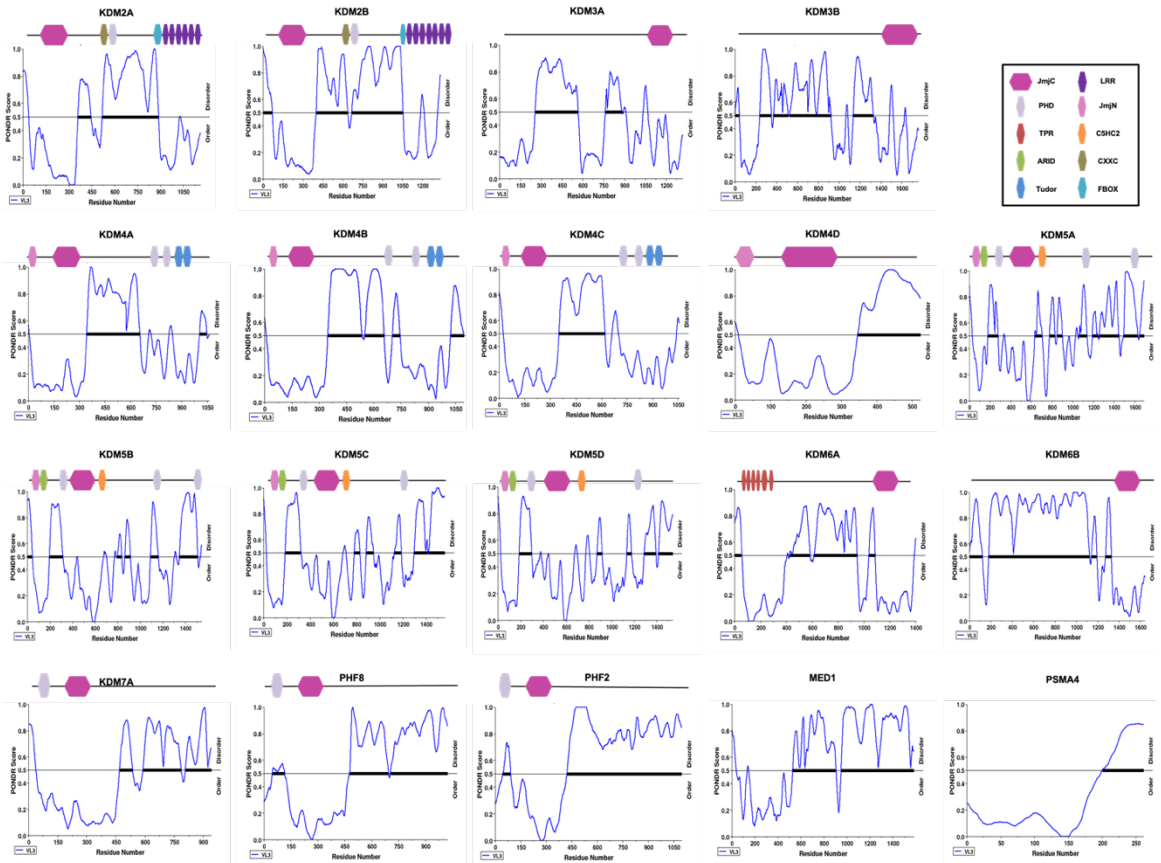
To assess our hypothesis, we searched for disordered regions in our proteins' sequences using three predictors: PONDR-VL3 (231), IUPred (232) and VSL2 (247). The three tools returned significant disorder scores, which according to PONDR-VL3 ranged from 0.68 for JMJD3 to 0.40 for KDM4C (Figure R46 and R47). Moreover, a striking proportion of amino acids belonged to disordered domains: according to PONDR-VL3, from a 71.39% for JMJD3 to a 32.67% for KDM4C (Figure R46).

IDR PREDICTION SCORE			
	VL3	VSL2	IUPred3
KDM2A	0.5024	0.5068	0.32
KDM2B	0.5662	0.5658	0.39
KDM3A	0.4116	0.5031	0.34
KDM3B	0.5174	0.6090	0.4
KDM4A	0.4449	0.4901	0.31
KDM4B	0.4450	0.4992	0.36
KDM4C	0.4023	0.4399	0.3
KDM4D	0.4147	0.4611	0.37
KDM5A	0.5084	0.5107	0.3
KDM5B	0.4661	0.4886	0.29
KDM5C	0.4692	0.4904	0.33
KDM5D	0.4274	0.4464	0.26
KDM6A	0.4124	0.5243	0.39
KDM6B	0.6840	0.7350	0.63
KDM7A	0.4725	0.5568	0.38
PHF8	0.5456	0.6031	0.44
PHF2	0.6165	0.6743	0.51
MED1	0.6422	0.7255	0.61
RNAPII	0.4169	0.5233	0.37
HP1α	0.5393	0.6191	0.42
PSMA4	0.2804	0.4045	0.22
PSMA7	0.2545	0.3753	0.33
Actin	0.2148	0.3360	0.34



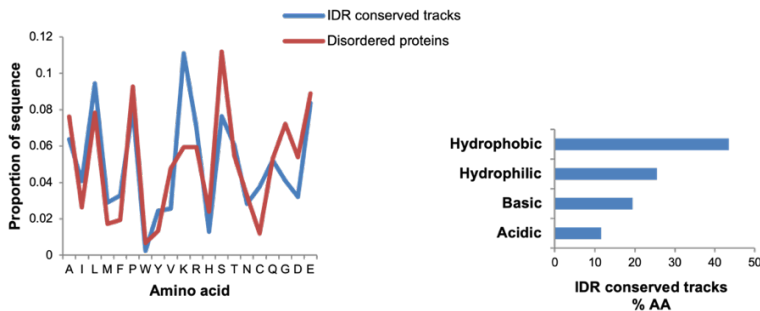
**Figure R46.** IDR prediction scores (left) and percentage of disorder (right) found in JMJC-KDMs sequences, obtained with PONDR-VL3, VSL2 and IUPred algorithms. The results for the positive control proteins MED1, RNAPII and HP1α, and for the negative control proteins PSMA4, PSMA7 and Actin, are included.

These values, both the scores and the percentage of amino acids in disordered regions, were considerably higher than the ones obtained for PSMA4, PSMA7 and actin, used as negative controls, and similar to the ones obtained for MED1, RNAPII and HP1α, used as positive controls (Figure R46 and R47).



**Figure R47.** PONDR-VL3 IDR predicted profile for JMJC-KDMs protein, as well as for the positive control MED1 and for the negative control PSMA4.

Thus, the predictions indicated that large part of JMJC-KDMs sequences consisted of IDRs. Moreover, about 99% of JMJC-KDMs identified IDRs were conserved within the family the protein belongs to, as assessed with BLAST® tool (239). These conserved IDRs showed a striking enrichment in lysine, arginine and glutamate (charged residues), as well as in the hydrophobic amino acids leucine and proline, which, as already mentioned, are associated with a phase separation capacity (see “Material and Methods”) (Figure R48).



**Figure R48.** The left-sided graph depicts the mean amino acid composition of the conserved IDRs between JMJC-KDMs families as compared to disordered proteins. The right-handed bar plot shows the percentages of hydrophobic, hydrophilic, and charged residues in these conserved IDRs.

Thus, these findings demonstrated that JMJC-KDMs were intrinsically disordered proteins, and pointed to a putative phase separating capacity for them.

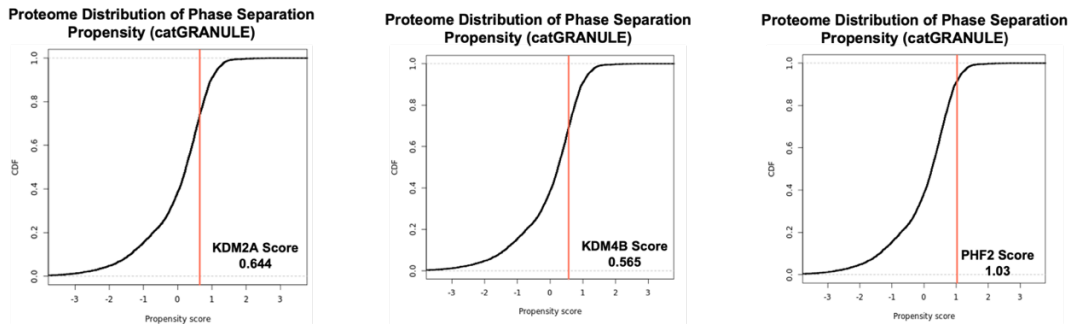
## 2.2. JMJC-KDMs are predicted to phase separate

To check whether JMJC-KDMs could undergo phase separation we took advantage of the predictors PSPredictor (254) and catGRANUE (235). As seen in Figure R49, both tools returned high scores for many of the demethylases. Taking as an example the PSPredictor results, scores ranged from 0.99 for PHF2 to 0.01 for KDM4A, results that were much higher than the one obtained for PSMA4 (0.0011).

PS Prediction Score		
	PSPredictor	catGRANULE
KDM2A	0.0613	0.644
KDM2B	0.0324	0.784
KDM3A	0.1827	0.834
KDM3B	0.836	1.031
KDM4A	0.0133	0.561
KDM4B	0.2596	0.565
KDM4C	0.0755	0.677
KDM4D	0.2173	0.626
KDM5A	0.1734	0.598
KDM5B	0.081	0.294
KDM5C	0.028	0.534
KDM5D	0.2155	0.408
KDM6A	0.976	0.883
KDM6B	0.9971	0.831
KDM7A	0.9942	0.770
PHF8	0.9725	1.045
PHF2	0.9995	1.027
MED1	0.9994	1.593
RNAPII	0.6411	1.029
HP1 $\alpha$	0.9635	0.542
PSMA4	0.0011	0.245
PSMA7	0.0792	0.304
Actin	0.0145	0.097

**Figure R49.** Phase separation predicted score obtained with two different algorithms (PSPredictor and catGRANULE). The positive and negative controls MED1, RNAPII and HP1 $\alpha$ , and PSMA4, PSMA7 and actin, are also shown.

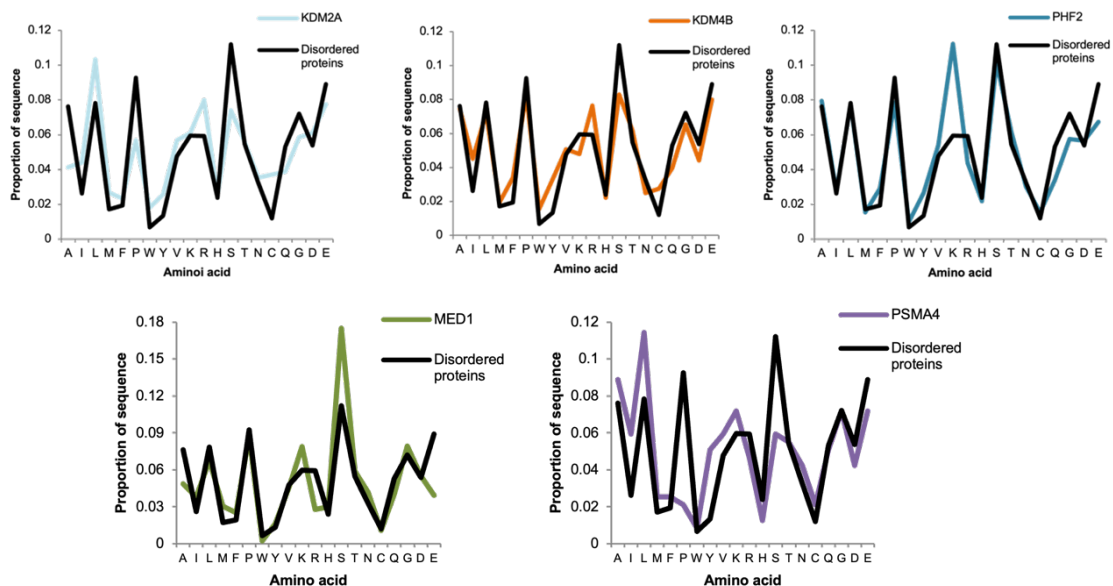
This analysis pointed to a phase separation capacity for our studied demethylases. To deepen into this possibility, we selected three representative KDMs according to their scores: PHF2, KDM2A and KDM4B, with the highest, medium and lowest scores according to both algorithms, respectively (Figure R49 and R50). The rest of our study was conducted by means of these three representative proteins.



**Figure R50.** Phase separation propensity prediction, with respect to the propensity distributions of the whole proteome, obtained with catGRANULE for the three representative JMJC-KDMs we selected for further analysis (KDM2A, KDM4B and PHF2).

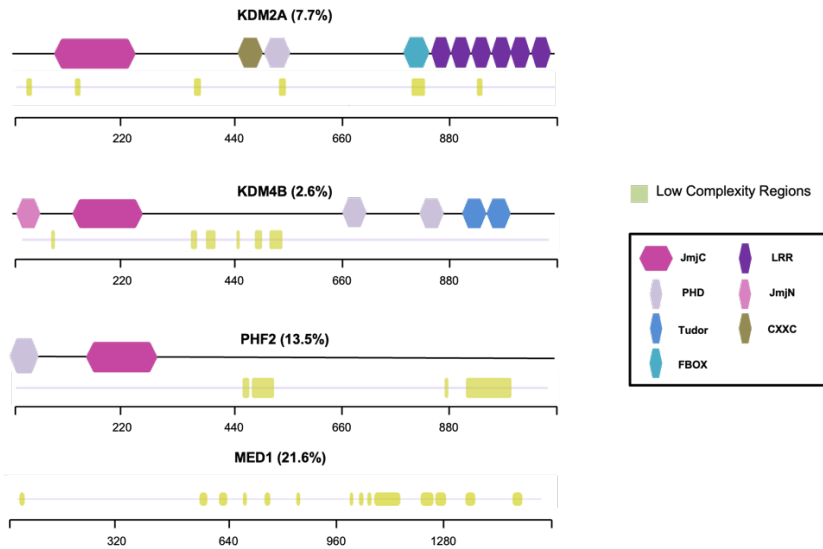
### 2.3. KDM2A, KDM4B and PHF2 sequence features prime them to phase separate

Previously, the importance of the amino acid composition for the phase separation process was highlighted. Thus, we deepened into the analysis of our selected JMJC-KDMs composition. As can be seen in Figure R51, these proteins showed an enrichment of leucine, arginine and glutamate for KDM2A, serine, lysine and glutamate for KDM4B, and serine and lysine in the case of PHF2. All these residues have been associated with a phase separation capacity (245).



**Figure R51.** Amino acid composition of KDM2A, KDM4B and PHF2, and the positive control MED1 and negative control PSMA4, as compared to the average residues in mouse disordered proteins.

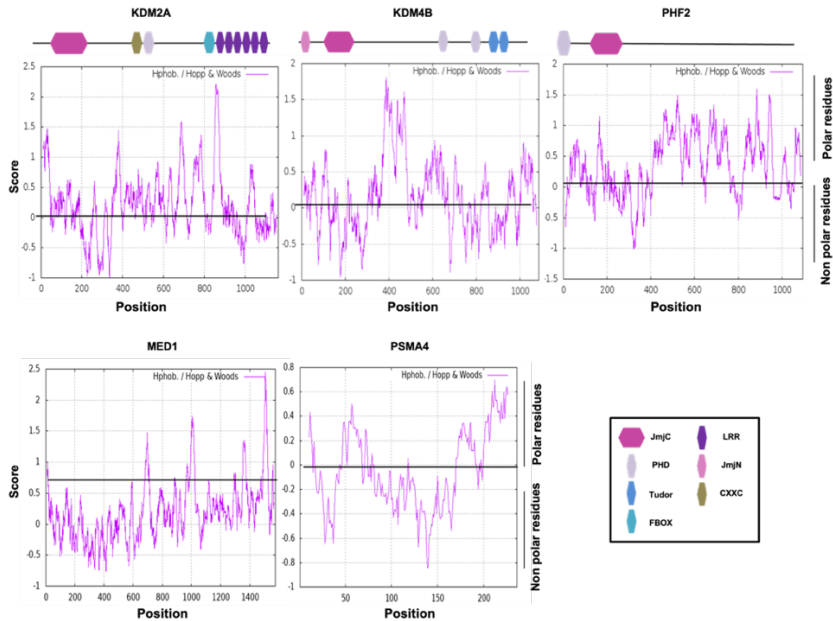
It is also well-known that disordered regions frequently coincide with low-complexity domains, which are biased toward certain types of amino acids (245,248,249). We checked then the enrichment of our selected KDMs in low-complexity domains using SEG algorithm (236). Figure R52 demonstrates that these proteins did contain low-complexity domains, with percentages ranging from a 2.6% of their sequences in the case of KDM4B to a 13.5% for PHF2. These percentages were in accordance with the ranges of disorder found within the proteins (Figure R46). As expected, low-complexity domains mostly coincided with unstructured protein regions.



**Figure R52.** Schemes representing the proportions of low-complexity domains found within the sequences of KDM2A, KDM4B, PHF2, and the negative control MED1 (green boxes, predicted by SEG algorithm).

Lastly, it is also known that, together with electrostatic, hydrophobic interactions are fundamental for phase separation to occur (92,262). Thus, we examined the hydrophobicity profiles of our selected KDMs (Figure R53). Although they predominantly contain polar residues, the three of them also show hydrophobic tracks, especially KDM4B.



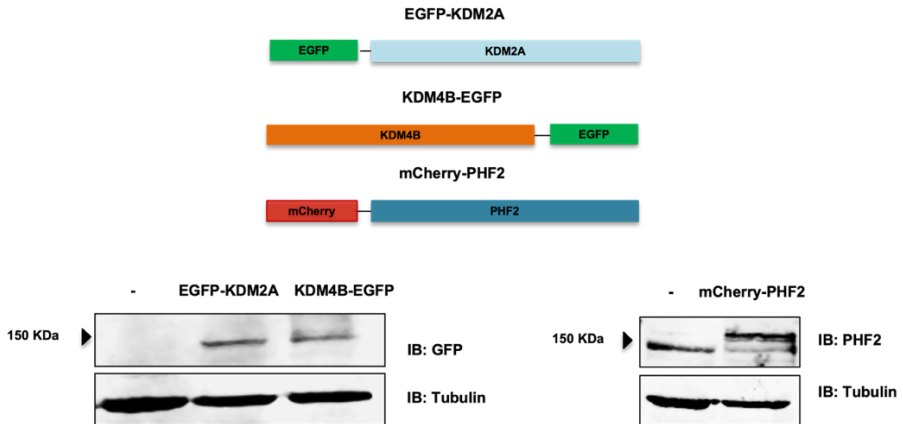


**Figure R53.** Hydrophobicity profiles of KDM2A, KDM4B, PHF2, the positive control MED1 and the negative control PSMA4, obtained through the ExPASy website (252) using the Hopp and Woods scale and a sliding window of 21.

All considered, these data highlighted that the amino acid features of JMJC-KDMs set them up to participate in phase separation processes.

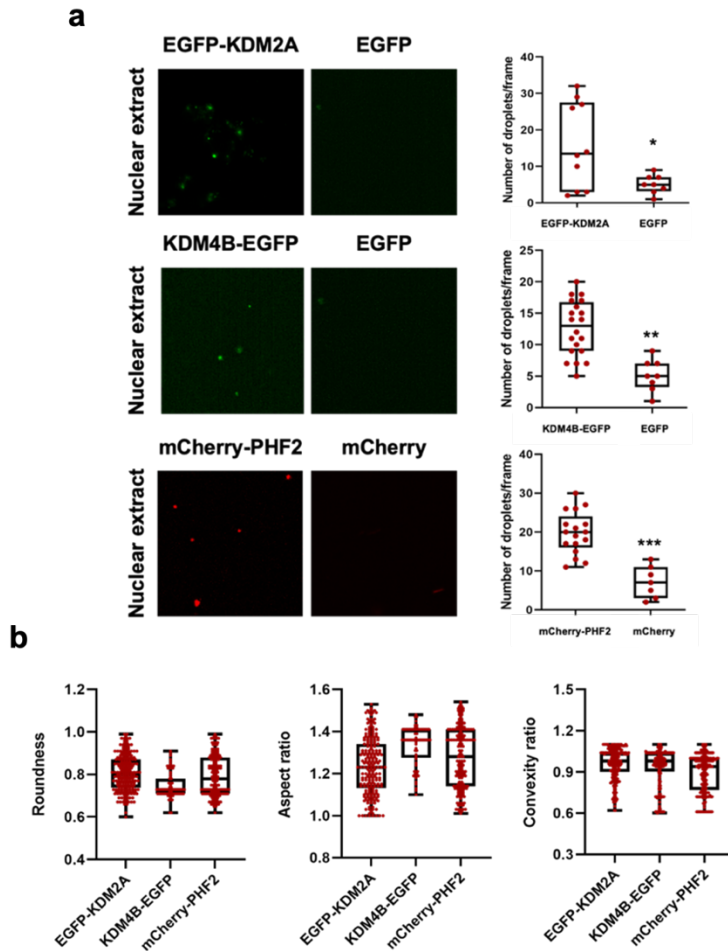
#### 2.4. KDM2A, KDM4B and PHF2 undergo liquid-liquid phase separation *in vitro* and *in vivo*

To assess whether KDMs were able to phase separate, we analyzed their capacity to form biomolecular condensates both *in vitro* and *in vivo*. To do this, we took advantage of fluorescent fusion proteins in which KDM2A and KDM4B were fused to EGFP, and PHF2 to mCherry (Figure R54, top panel). Their expression in HEK293T cells was checked by Western Blot assays (Figure R54, bottom panels).

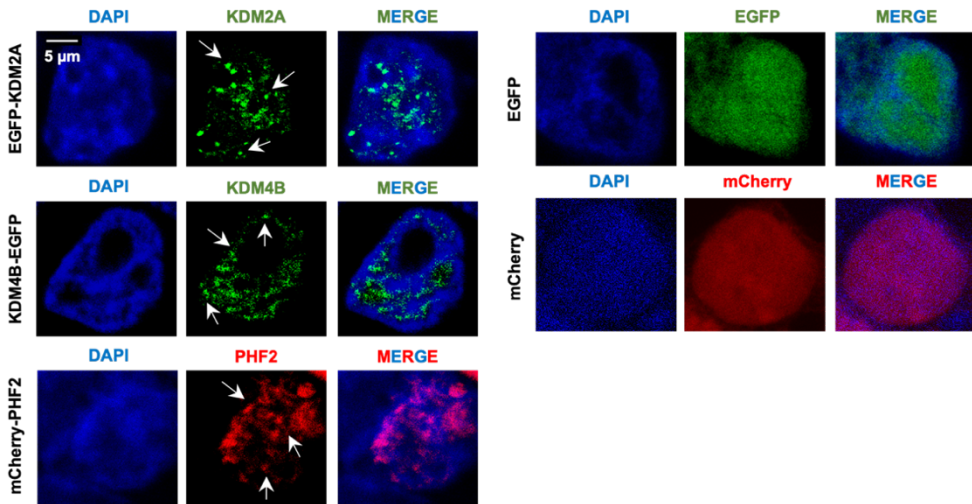


**Figure R54.** Top panel shows schemes representing the vectors used to overexpress the selected KDMs. Below, Western Blot assays performed with total extracts of HEK293T cells overexpressing either EGFP-KDM2A, KDM4B-EGFP (left) or mCherry-PHF2 (right). The fusion proteins were detected with GFP and PHF2 antibodies, respectively.

Then, we performed droplets formation assays in nuclear extracts of HEK293T cells that overexpressed the fluorescent proteins. Figure R55a shows that, indeed, the three demethylases were able to form *in vitro* droplets that did not appear with the fluorescent tag alone. Moreover, these droplets showed features characteristic of a liquid-like nature, as a roundness, aspect ratio and convexity ratio close to 1 (Figure R55b).

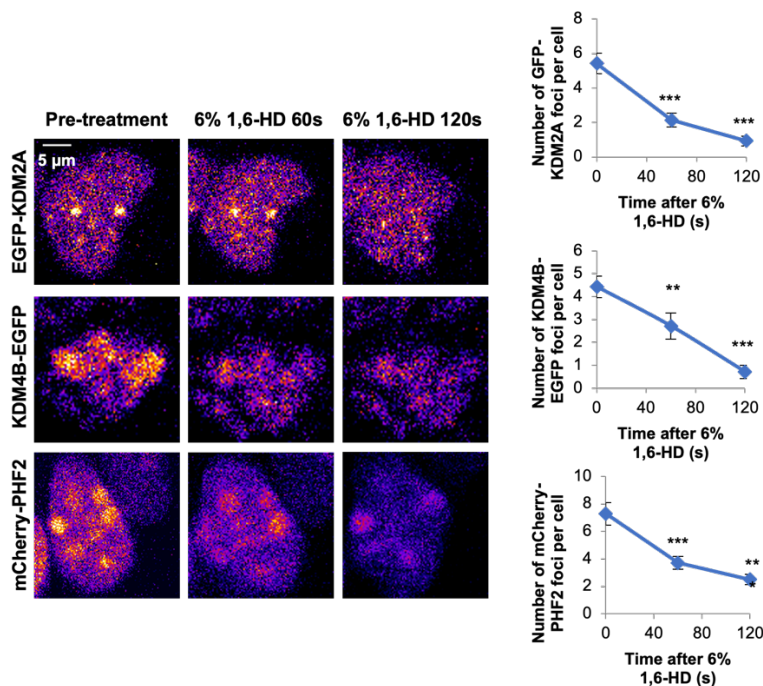


We also studied the capacity of these proteins to form nuclear puncta in HEK293T cells. Figure R56 shows that, indeed, the overexpressed KDMs formed condensates in the cell nucleus, as compared to EGFP and mCherry molecules.



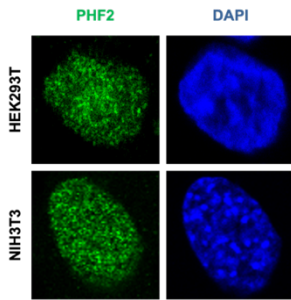
**Figure R56.** Immunofluorescence assay of HEK293T cells overexpressing EGFP-KDM2A, KDM4B-EGFP or mCherry-PHF2 plasmids (on the left). White arrows signalize puncta. EGFP and mCherry signal were included as negative controls of puncta formation (on the right). The intrinsic fluorescence of the molecules was followed to detect proteins.

It is known that hydrophobic interactions are one of the main driving forces of nuclear condensates assembly. Although these proteins were more enriched in polar regions, the three of them contained hydrophobic tracks (Figure R53). Thus, we tested the sensibility of KDM2A, KDM4B and PHF2 condensates to the aliphatic alcohol 1,6-Hexanediol, which disrupts these interactions. As can be appreciated in Figure R57, the treatment led to a reduction in the number and size of KDMs nuclear condensates.



**Figure R57.** Left panels correspond to HEK293T cells overexpressing EGFP-KDM2A, KDM4B-EGFP and mCherry-PHF2 that were treated with 6% 1,6-HD for 5 minutes and imaged at 60 and 120 s. Nuclei were visualized with DAPI. The right-sided graphs depict the quantification of KDMs nuclear foci per cell along the treatment. Data are the mean  $\pm$  SEM.

Furthermore, to rule out the possibility that the observed puncta were due to an overexpression artifact, we performed immunofluorescence experiments of the endogenous PHF2 protein in both HEK293T and NIH3T3 cells. Indeed, these experiments demonstrated that the endogenous levels of the demethylase were enough to form nuclear condensates in a cell type independent manner (Figure R58).



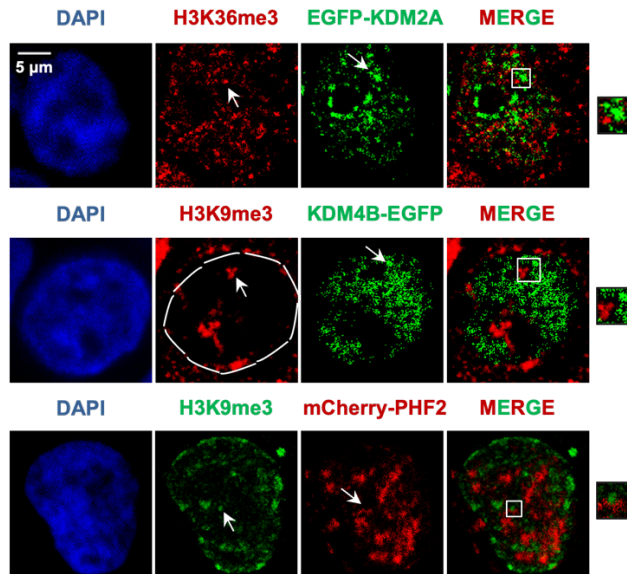
**Figure R58.** Immunostaining of endogenous PHF2 nuclear puncta in HEK293T and NIH3T3 cells, detected by staining with PHF2 antibody.

These results demonstrated that JMJC-KDMs were able to form liquid-like biomolecular condensates *in vitro* and *in vivo*, which constituted a separated phase within the cell nucleus.

## 2.5. KDM2A, KDM4B and PHF2 condensates are related to transcription

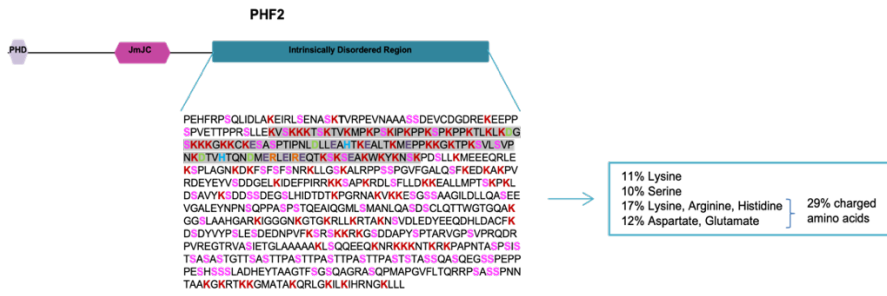
As mentioned before, the role of JMJC-KDMs is fundamental for gene expression regulation, and phase separation has also been implicated in this process. Thus, we wondered whether the observed KDMs condensates were related to the transcriptional outcomes of these proteins.

We started by performing immunofluorescence assays to examine the localization of JMJC-KDMs condensates with respect to the repressive H3K9me3 or the active H3K36me3 histone marks. Figure R59 demonstrates that KDM2A condensates were excluded from H3K36me3 foci, consistent with its role as a transcriptional repressor. On the other hand, KDM4B and PHF2 condensates excluded the repressive mark, as expected due to their transcriptional activating role.

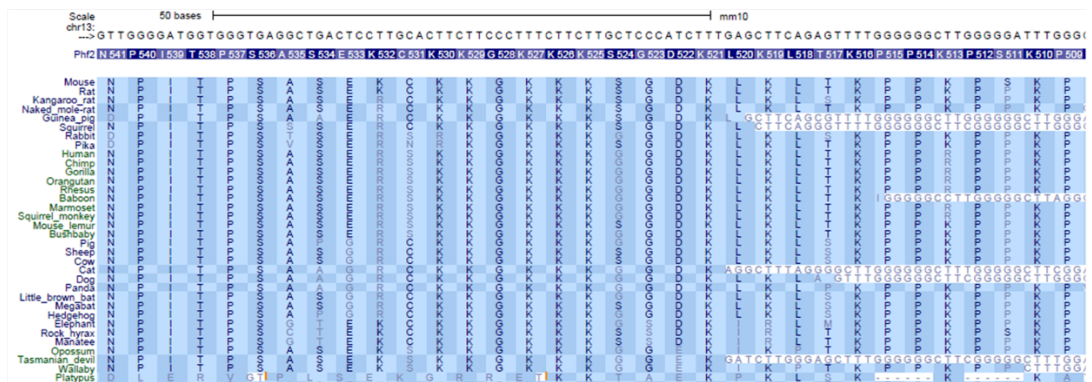


**Figure R59.** Immunofluorescence experiment in HEK293T cells that expressed EGFP-KDM2A, KDM4B-EGFP or mCherry-PHF2. Cells from the upper panel were stained with H3K36me3 antibody, while cells from the middle and bottom panel were stained with H3K9me3 antibody, and the exclusion of these histone marks with KDMs condensates was analyzed (right-sided boxes, zoom in). DAPI was used to visualize nuclei. Both marks and condensates are pointed with white arrows.

These results suggested that there was a correlation between KDMs condensates and these proteins underlying transcriptional activity. To further analyze this possibility, we selected PHF2 protein, which has been previously studied by our lab, and constructed a mutant protein that lacked part of its IDR. This mutant was depleted of a lysine-enriched region that spanned the amino acids 487-806, and that was largely conserved among mammals, as seen in Figure R60.



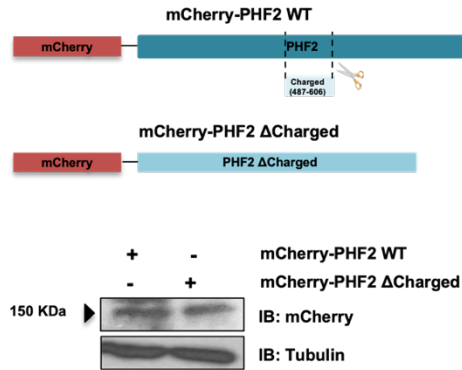
mm10: chr13:48,817,572-48,817,668



**Figure R60. (a)** Cartoon representing PHF2 IDR residues, highlighting in grey the lysine-rich track that was deleted in the mutant. The percentages of the most abundant residues found within the region are also depicted. **(b)** UCSC tracks displaying the conservation of the lysine-rich region of PHF2 IDR among vertebrate species. Multiz alignments (250) was used to analyze the conservation.

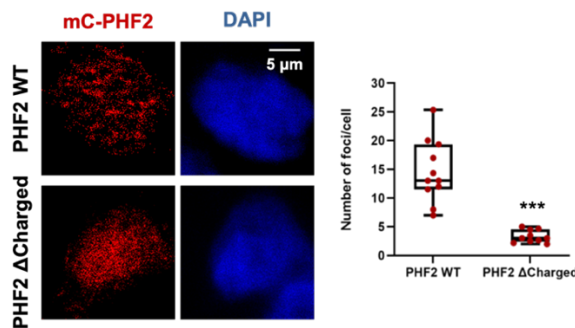
We assessed the expression of the mutant protein, which was named PHF2 ΔCharged, in HEK293T cells by Western Blot assay (Figure R61).





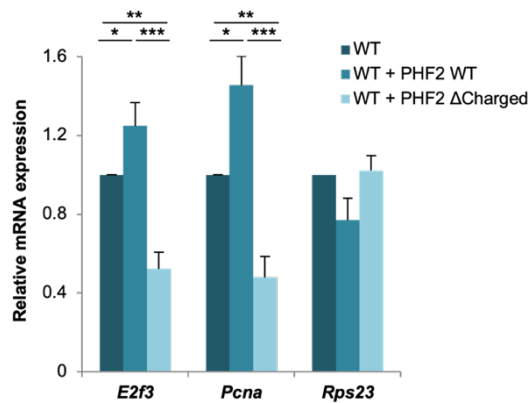
**Figure R61.** On the top, scheme depicting the mutant vector constructed from the original mCherry-PHF2 to overexpress PHF2  $\Delta$ Charged. Below, Western Blot assays performed with total extracts of HEK293T cells overexpressing either PHF2 WT or PHF2  $\Delta$ Charged mutant. Proteins were detected with mCherry antibody.

Then, we analyzed the capacity of PHF2  $\Delta$ Charged to form nuclear condensates in HEK293T cells. We performed immunofluorescence experiments in fixed cells; as expected, PHF2  $\Delta$ Charged was not able to form nuclear condensates *in vivo* (Figure R62).



**Figure R62.** Immunofluorescence experiment of HEK293T cells overexpressing either PHF2 WT or PHF2  $\Delta$ Charged (left). The intrinsic fluorescence of the molecules was followed to detect proteins. The number of PHF2 foci per cell was quantified for each condition (right boxplot). Data show the mean  $\pm$  SEM.

Considering these results, we concluded that the phase separating capacity of PHF2 was dependent on its lysine-rich IDR. Given so, we studied if this dependency was translated into any functional outcome. To do this, we analyzed the expression levels of some PHF2 target genes (*E2f3* and *Pcna*, (263)) in NIH3T3 cells that overexpressed either PHF2 WT or PHF2  $\Delta$ Charged. Figure R63 shows that the mutant protein acted as a dominant negative, interfering with gene expression, as compared to PHF2 WT protein. This was specific for PHF2 targets, as the gene *Rps23*, used as a negative control, was not affected by the mutant.



**Figure R63.** Expression levels of PHF2 target genes *E2f3* and *Pcna* were analyzed by qPCR in NIH3T3 cells that overexpressed either PHF2 WT or  $\Delta$ Charged protein. The gene *Rps23* was used as a negative control. Data are presented as mean values  $\pm$  SEM.

Altogether, we concluded that PHF2 transcriptional regulatory capacity seemed to be dependent on the expression of its IDR, specifically on its lysine-enriched domain. Hence, PHF2 transcriptional role probably relied on its potential to undergo phase separation.

## Discussion

In this section of the manuscript, I will comment on the obtained results considering the current bibliography. The section will be divided into three major parts according to the discussed topic.

### 1. Regarding chromatin three-dimensional structure and transcription regulation

#### 1.1. About the TGF $\beta$ -driven assembly of enhancer clusters

The first part of this work attempted to deepen into the putative interplay between chromatin 3D structure and transcription activation at enhancers in a neurodevelopmental context. Indeed, we demonstrated that the JMJD3-TGF $\beta$  crosstalk promotes a chromatin reorganization that correlates with gene expression in the *Chst8* locus, specifically by means of the formation of an enhancer cluster (*Chst8* EC) (Figure R9).

The TGF $\beta$  pathway was already shown to mediate transcriptional responses upon the establishment of enhancer-promoter loops, specifically through its cooperation with the cofactors BRG1, p300 and the Mediator complex (264). On the other hand, enhancer-enhancer contacts have been widely described in the literature (265). These cis-regulatory regions cooperate to achieve high levels of transcription, which could be crucial for the proper function of some genes. In our particular case, the importance of the accurate expression of the gene *Chst8* is demonstrated by its implication in a peeling skin syndrome (OMIM #616265) (266). The most-well studied type of enhancer clusters are SEs, which are characterized by especially high levels of epigenetic marks and by recruiting really large amounts of transcriptional factors. This way, they act in a synergistic manner to produce striking levels of transcription, typically of cell identity and cell fate determining genes (85). Therefore, although we considered

whether *Chst8* EC could be a SE, we discarded the possibility, as the levels of both epigenetic marks and gene expression in the locus were not extremely high as compared to other SEs (90) (Figure R3 and R9). Furthermore, *Chst8* enhancers cooperated in an additive rather than a synergistic manner (Figure R10) to promote transcription of the gene *Chst8*, which is neither a cell identity nor a cell fate regulator. Thus, we preferred to be conservative and referred to the observed cooperation of enhancers as an enhancer cluster.

We acknowledge that it would be of great interest to further characterize *Chst8* EC, particularly focusing on the putative hierarchy established between its constituent enhancers. In this regard, the current view is that the mechanistic relationship between enhancers within a cluster is highly context and locus specific. Indeed, some studies are consistent with a hierarchical structure of SEs that includes both essential and dispensable constituent enhancers. This has been observed for the  $\alpha$ -globin SE and the *SLC25A37* SE in erythroid cells, the mammary STAT5-driven *Wap* SE, and several murine ESC SEs (267–270). Nonetheless, the technical challenges of characterizing enhancer clusters on a larger scale hinder the generalization of a hierarchical organization in the genome, although efforts are being made to develop appropriate approaches (271). For the study of our particular locus, the combination of genome editing tools and chromosome conformation capture techniques would be useful to address whether any of the constituent enhancers is a master driver of the cooperative behavior or, on the contrary, any of them is dispensable. In fact, in the lab we are already working on this matter. Taking advantage of the previously described  $\Delta$ *Chst8* Enh line (Figure R4), as well as deleting other enhancers from the *Chst8* EC, we are planning to perform both gene expression and genome conformation experiments to compare the functional impact of removing these regions in TGF $\beta$ -stimulated NSCs. We believe that with these approaches we would contribute to the current scarce knowledge of the field, particularly addressing the contextual

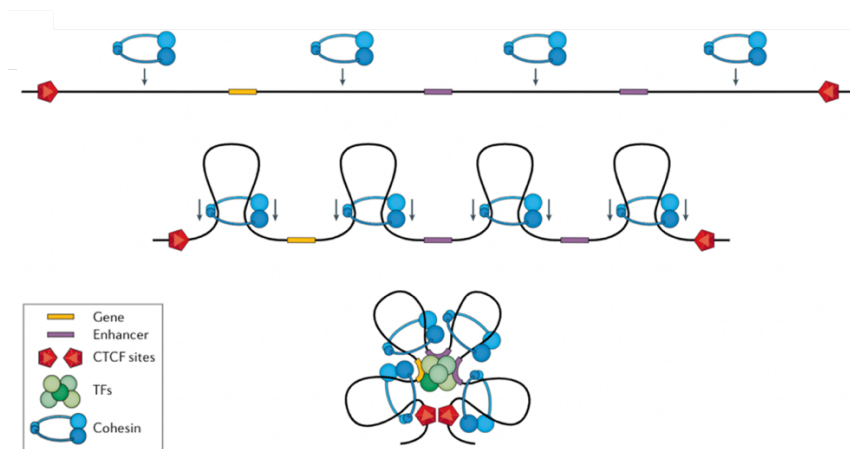
dependency of enhancers' mechanistic relationships through the TGF $\beta$  signaling activation.

### **1.2.About the contribution of JMJD3 to the enhancer cluster establishment**

This work has also shed light on an intriguing question raised upon previous work from the lab that showed the recruitment of JMJD3 to enhancers that were not marked with H3K27me3 (67). The fact that a histone demethylase binds loci depleted of its specific target mark is puzzling. However, we have proved that in a neurodevelopmental context this protein has a role at enhancers that is independent of its catalytic activity (Figures R20 and R21). Instead, it is related to its capacity to reorganize chromatin structure to promote enhancer-enhancer and enhancer-promoter contacts (Figures R12 and R18). This was not only found for the *Chst8* locus, but also for other two loci (*Ldlrad4* and *Aopep*, Figure R15), suggesting that JMJD3-structural role could have genome-wide implications. In fact, this finding is further supported by previous observations that demonstrated JMJD3-mediated enhancer-promoter looping during endoderm differentiation (272). So, although there are examples in the literature where H3K27me3 demethylation was necessary to activate enhancers (65,272), the present work and others support a major role for JMJD3 at enhancers that is not related to its catalytic activity (67,273).

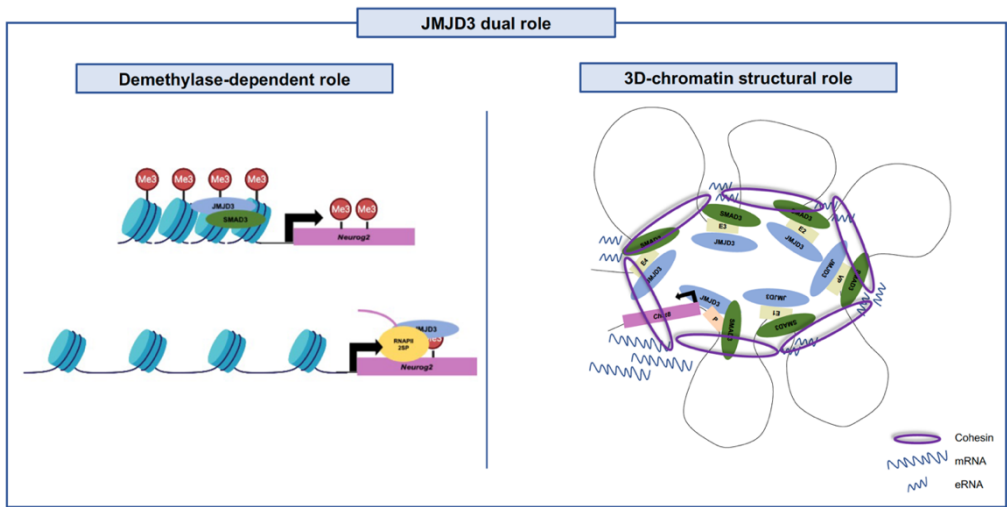
Moreover, previous work from Dr Martínez-Balbás lab demonstrated that the structural role of JMJD3 might also be independent of the LMC proteins. Although it was shown that JMJD3 co-localized with the Cohesin subunit SMC1 in many regions, including the *Chst8* locus, it did not regulate the expression of SMC1 nor of other LMC proteins. Thus, the observed structural effects were attributed to the direct action of JMJD3. Indeed, it is known that not only the loop extrusion mechanism, but also affinity interactions between the factors bound at regulatory elements are major intermediaries of their contacts. This has been shown for transcription

factors such as GATA1, KLF1 or LDB1, and cofactors such as YY1 or Mediator (274). We do now see it for the coactivator JMJD3 too. Furthermore, Cohesin has been demonstrated to co-localize with many of these proteins, such as Mediator or other transcription factors. This led to the suggestion that the loop extrusion process, apart from promoting the formation of loops itself, strengthen and provides specificity to the protein affinity-driven interactions (275,276) (Figure D1). The JMJD3-Cohesin co-localization raises the possibility of such a crosstalk between the coactivator-driven 3D-structure reorganization and the loop-extrusion mechanism.



**Figure D1.** Model that depicts the dual effect of Cohesin-mediated loop formation, both as a mechanism to extrude chromatin itself and as a stabilizer of protein affinity-driven interactions. Adapted from (274).

Thus, considering previous results together with the ones presented in this thesis, JMJD3 could be presented as a transcriptional coactivator with a dual role, a demethylase-dependent one, (66,152), and a 3D-chromatin structural one, through which the demethylase may cooperate with the Cohesin-driven loop-extrusion (Figure D2).



**Figure D2.** Model depicting the two putative roles of the activator protein JMJD3: its catalytic-related and its structural role.

### 1.3. About the interplay between three-dimensional genome structure and function

The present work has demonstrated that *Chst8* locus activation is accompanied by chromatin reorganization, but whether its expression is a cause or a consequence of the 3D-genome structure is not clear. Following the principle that genome structure exerts a major instructive role on the regulation of gene expression, we assumed that the reorganization of the locus promoted *Chst8* transcription. However, we do not really know whether the enhancers that form *Chst8* EC were activated upon the establishment of contacts, or conversely, they were previously active, and this induced the structural changes.

Nowadays the link between genome conformation and function is controversial; in fact, the literature is full of contradictory studies that aimed to clarify the functional relevance of loops and TADs as gene expression regulatory units. Although loops have been shown to facilitate a proper

gene regulation by bringing together cis-regulatory elements and genes (277), recent findings supported that the physical enhancer-promoter contact is not necessary to trigger transcription in some paradigmatic cases, such as the *Shh* or *Sox2* loci (278–280). Regarding TADs, many reports assessed that their global disruption through CTCF or Cohesin depletion had just moderate effects on transcription (281–283). However, others showed that specific disruptions by structural variants such as duplications, inversions or deletions either disassemble endogenous or promote ectopic enhancer-gene interactions, hence driving aberrant gene silencing or activation (280,284,285). On the other hand, transcription has also been shown to influence chromatin structure. Many TAD borders are set by active genes in mammalian, while in *D.melanogaster* the role of transcription for TAD organization seems to be more important than the one of CTCF itself (286). Furthermore, it has been observed that gene reactivation occurs previously to the reestablishment of topological domains in some loci upon mitotic exit (287).

Altogether, the instructive role of genome conformation over functional activation is being challenged. Alternatively, the existence of a dynamic crosstalk in which both structure and transcription modulate the activity of the other is emerging as an actual possibility (288). Indeed, this might be the case for *Chst8* EC, where the JMJD3-TGF $\beta$  cooperation could be promoting such a crosstalk.

## **2. Regarding phase separation and transcription regulation**

### **2.1. About JMJC-KDM IDRs characterization**

The discovery of a JMJD3 catalytic-independent role led us to study the protein structure, so that the mechanism underlying its remodeling capacity was envisioned. This way, we came to discover that JMJD3 is a highly disordered protein that, apart from the JmjC catalytic domain, just contains a consecution of intrinsically disordered regions. We focused on a



particularly large, proline-enriched region - JMJD3 IDR - whose structural features resembled those described for other phase separating proteins, such as MED1 (248). Indeed, JMJD3 potential to form part of nuclear condensates was determined by this domain (Figures R37 and R38). However, due to its length and complexity, we were precluded to demonstrate which specific regions of the IDR were fundamental for phase separation to occur. We will now hypothesize on this regard.

When referring to the JMJD3 IDR we usually highlight prolines, due to their abundance and the magnitude of the tracks found within the protein sequence. However, it also presents remarkably vast low-complexity, repetitive tracks of serine, threonine and charged residues that could contribute to the IDR-driven phase separation (Figure D3). Indeed, all serine, hydrophobic and charged amino acids have been demonstrated to be involved in the nucleation of phase separated condensates (92,165,251,262). In the particular case of JMJD3 IDR, four specific regions that are highly conserved among vertebrates drew our attention: a charged region (spanning 973-1016 AA), which is enriched in lysine, glutamic acid, arginine and histidine, and three hydrophobic regions (spanning 889-908, 947-967 and 1033-1073 AA) which are enriched in proline and leucine, proline, leucine, alanine and glycine, and proline and alanine, respectively. Also, the three of them contain high levels of serine within and nearby the tracks (Figure D3). It is well-known that PTMs, such as phosphorylation, can modulate the phase separation capacity of a protein. Given this, phosphorylation sites link that particular region of the protein sequence to phase separation (289). Indeed, one of JMJD3 hydrophobic tracks (889-908 AA) contains a phosphorylation site (T894), which points to a major role of this track in JMJD3 phase separation regulation. All considered, we believe that any of these regions could be fundamental for the phase separation capacity of JMJD3, although the possibility also exists that it is not one, but the combination of many low-complexity, disordered tracks, what allows the protein to form condensates.



**Figure D3.** Scheme representing JMJD3 IDR. The charged (grey) and hydrophobic (pink, blue and yellow) tracks that we believe could be important for phase separation are highlighted.

Contrary to what happened for JMJD3, we were able to demarcate a specific region of PHF2 IDR with features typically linked to phase separation. So, we addressed a track particularly enriched in charged residues (487-606 AA) that, indeed, was mediating PHF2 formation of condensates (Figures R60 and R61). Certainly, this track contains 5 serine and 1 threonine residues amenable to be phosphorylated (S511, S534, S536, T538, S567 and S570), and it is also conserved among vertebrates (Figure R60). These features could explain the fundamental role that the region plays on PHF2 phase separation. However, there are other regions in the protein that could play such a role. For instance, we have also identified a conserved track specially enriched in hydrophobic residues (958-1021 residues) that may be worth studying, as it is possible that more than one region is involved in the phase separation capacity of the protein, as previously commented.

Considering the general high degree of disorder that JMJC-KDMs show, we acknowledge that it would be of great interest for the phase separation field to extend the study to all these proteins' IDRs, so that a detailed characterization of the pattern of amino acids that drive KDMs' putative capacity to form condensates was established. This would allow researchers to modulate their putative phase separation, constituting an interesting experimental tool to deepen both into the functional consequences of KDMs phase separation and into the mechanistic insights of the process. Furthermore, phase separated condensates are being widely studied as potential unconventional therapeutic venues (290). This characterization could provide evidence of the interactions that stabilize biomolecular condensates, so that specific drugs amenable to target key amino acids could be designed for KDMs condensates and, also, for others driven by similar residues.

## **2.2. About JMJC-KDM phase separation functional outcomes regarding transcription**

The results from this work state the phase separation potential of virtually all JmjC-containing lysine demethylases. As we could not include all proteins in our study, in addition to JMJD3 we chose representative examples of JMJC-KDMs belonging to different families, in accordance with their disorder content and their phase separation capacity predictions. Thus, we came to demonstrate the formation of nuclear condensates by JMJD3, KDM2A, KDM4B and PHF2. This is supported by previous studies that showed KDM2A nuclear bodies in a heterochromatic context (291). Furthermore, UTX has also been recently demonstrated to share this capacity (196). Altogether, these data point to phase separation as a dynamic common mechanism used by JMJC-KDMs to regulate their chromatin-related functions.

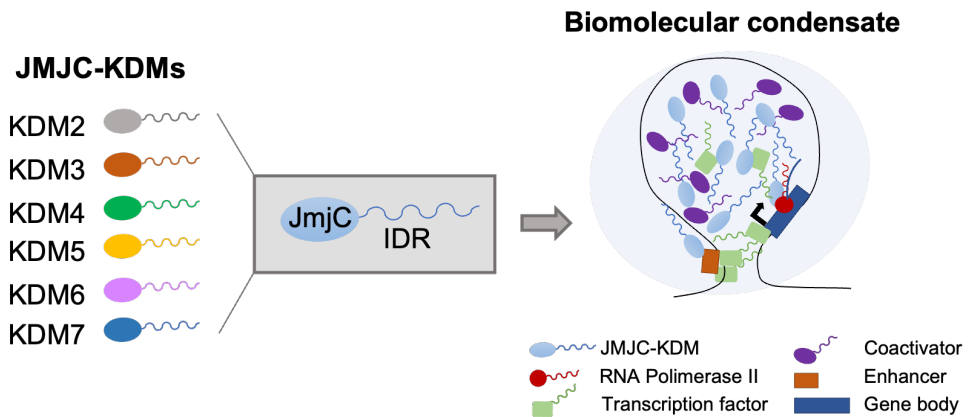
As JMJC-KDMs are widely involved in transcriptional regulation, we specifically focused our attention on the functional impact of their

condensates on this process. Phase separation has been widely related to transcription (82,192,201,204,292); in fact, the term “transcriptional condensate” has already been used to designate those biomolecular condensates that contain transcriptional machinery and induce gene expression (92,192,205). A model involving condensate formation has been proposed to explain transcriptional initiation and elongation (see Figure I24 and (201,203,293,294). According to it, two condensates would be formed during transcription: an initiation condensate, containing the non-phosphorylated RNAPII, TFs, Mediator and coactivators, and an elongation condensate, within which the phosphorylated RNAPII, RNA processing and elongation factors, and RNA itself would be found. In this thesis, we showed that, indeed, JMJC-KDM condensates correlate with the transcriptional features associated to these proteins (Figures R44 and R59). These results are further supported by the discovery that UTX assemblies are also related to its transcriptional regulatory capacity (196).

The condensate-related transcriptional model could be especially relevant for the case of JMJD3, which may play a role in both initiation and elongation condensates. Regarding its putative involvement in the initiation condensate, previous studies from the lab demonstrated its function at promoters and enhancers (66,67). In both cases, JMJD3 is recruited to the regulatory regions through the interaction with SMAD3, a transcription factor that has also been shown to form phase separated condensates in an IDR-dependent manner (204). Indeed, SMAD3 interacts with JMJD3 through its linker region, which is the one that contains the IDR (66). Furthermore, at enhancers JMJD3 and SMAD3 form an activator complex, together with the pioneer transcription factor ASCL1 and the chromatin remodeler protein CHD8, to promote gene expression (67). We have now demonstrated that JMJD3 condensates are found at the *Chst8* EC locus, whose expression also depends on the TGF $\beta$ -JMJD3 crosstalk (Figures R17 and R43), and many of these condensates also contain the essential transcriptional cofactor MED15 (Figures R30 and R44). Altogether, these data suggest that SMAD3 and JMJD3 could be engaging in multivalent

interactions, together with other transcriptional machinery -such as CHD8 and MED15-, to participate in the above-mentioned transcription initiation condensate. Moreover, JMJD3 has been demonstrated to be essential to transcription elongation due to its interaction with the elongating form of the RNAPII and with elongation factors (152,295). Given so, JMJD3 could also contribute to the elongating condensates.

Thus, the presented results demonstrate that JMJC-KDM condensates facilitate transcription, but they could do it in many ways. The most prominent hypothesis is that they constitute transcriptional hubs that concentrate cofactors, transcription factors and transcription machinery at specific loci to increase the kinetic efficiency of transcription, as proposed for other factors (201,246,296) (Figure D4). Nonetheless, other possibilities should be considered. Indeed, KDMs condensates could favor transcription by physically insulating the transcriptional machinery from its inactivators, as could occur to protect the active phospho-SMAD3 from phosphatases, for example. JMJC-KDMs condensates could also favor nucleosome unwrapping through their IDR-facilitated access to compacted heterochromatin, as seen for KLF4 during reprogramming (195). Finally, condensates could favor the establishment of contacts between distal regulatory regions, as observed in this thesis for JMJD3 (Figure R18) and in other works for UTX and YY1 (196,297). The interplay between phase separation and the genome structure will be further commented in section 3.



**Figure D4.** Cartoon model representing the common IDR of JMJC-KDMs, which allows them to participate in liquid-liquid phase separation processes to assemble condensates that regulate gene transcription.

Despite our interpretation of the results regarding JMJC-KDMs condensates functional impact, it is worth noting that the currently accepted relationship between phase separation and the promotion of transcription is starting to be challenged. Particularly, a recently published study argues against the proposed enhancement of transcription activation by TFs condensates. This work demonstrates that phase separation does not enhance transcription despite increasing the local concentration of TFs. On the contrary, the activation capacity of TFs seems to be promoted by the establishment of multivalent interactions through their activation domains independently of phase separation, but by increasing the residence time in chromatin and the recruitment of coactivators (298).

### 2.3. About JMJC-KDMs phase separation and disease

Phase separation has been linked to many pathologies, including neurodegenerative diseases, cancer, and infections (185,211,212,299). Thus, phase-separated condensates nucleation and/or disturbance may

underly the etiology of many JMJC-KDMs-related disorders. Indeed, UTX phase separation has been linked to its tumor suppressor capacity (196). Furthermore, many JMJC-KDMs pathology-leading mutations occur in their IDRs, such as many of the ones that lead to intellectual disability in KDM4B (300), the cancer-related nonsense mutations in PHF2 (301), or the point mutations that produce neurodevelopmental delays and dysmorphic features in JMJD3 (77). In the case of JMJD3, its phase separation capacity could also impact TGF $\beta$ -regulated physiological processes, such as cell proliferation and differentiation, and disorders, such as cancer and neurodegeneration (299). In this regard, aberrant phase separation processes could lead to the dysregulation of the locus studied along this work, and as previously mentioned, an improper expression of *Chst8* is associated to a peeling skin syndrome (266).

Understanding the involvement of KDMs phase separated condensates in disease is fundamental, as it could provide unconventional targets to develop new therapeutic venues. Even more, biomolecular condensates are also being studied as therapeutic tools to deliver drugs. It has been demonstrated that small particles selectively partition into condensates through physicochemical properties, so that the modulation of these properties could affect drugs concentration and pharmacodynamics (302). Thus, KDMs condensates could also be studied as putative reservoirs that deliver drugs to their target loci, such as chromatin.

### **3. Regarding the interplay between phase separation and genome structure to promote transcription**

Phase separated condensates have emerged as another layer of genome structural organization, although the relationship between the two concepts is reciprocal. On the one hand, phase separation processes are affected by chromatin structure itself. Indeed, condensates are prone to nucleate in loose and low-density chromatin regions (303), while the surrounding chromatin shows certain mechanical resistance to their assembly

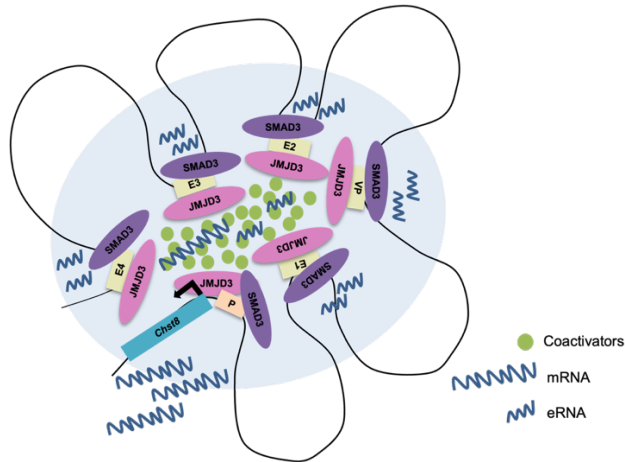
(182,304). On the other hand, membranelles organelles are also able to directly influence chromatin structure. This has been observed for BRD4, whose condensates showed a repelling effect on chromatin that led to its remodeling into different domains (193); for the polymer-polymer interactions established between histone tails and other chromatin proteins, which facilitate the formation of TADs (68,69); for UTX, whose condensates regulate dynamic changes of chromatin contacts by either promoting or suppressing loops (40), and for YY1, that mediates the formation of an enhancer cluster through phase-separated transcriptional condensates (54).

The interplay between phase separation and chromatin reorganization is also related to their potential to efficiently promote transcription. It has been shown that loci with high local concentrations of cis-regulatory regions, and hence of transcriptional machinery, nucleate membranelles organelles that promote transcription in a highly efficient manner, even buffering cell fluctuations in transcription factors and coactivators (29,46). Indeed, the establishment of contacts between those regulatory regions could be the trigger for the accumulation of cofactors and the subsequent phase separation. Conversely, the establishment of loops could be facilitated within these phase separated hubs that bring into close proximity widely distributed regions, both through affinity interactions between the abundant resident proteins and through loop extrusion mechanisms (15,40,54). Furthermore, the high amounts of factors contained within the transcriptional hub may allow cis-regulatory regions to function even when not directly touching (70).

Independently of the triggering factor, it becomes clear that phase separated condensates could be considered structural regulatory units, which integrate both conformational and transcriptional requirements to achieve gene expression in an optimized manner for the cell. Considering this, we propose the following model for our studied locus (Figure D5), where the crosstalk between JMJD3 and the TGF $\beta$  signaling pathway



induce a chromatin reorganization that implies the assembly of an enhancer cluster (*Chst8* EC). This, in turn, drives *Chst8* gene expression, presumably in an IDR-dependent manner through the formation of phase separated condensates. Indeed, our model is supported by the recently published study that demonstrates the capacity of YY1 transcriptional condensates to drive an enhancer cluster assembly and gene activation at the FOXM1 locus in a histidine rich-IDR-dependent manner (54).



**Figure D5.** Model depicting JMJD3 contribution to the transcriptional activation of the *Chst8* locus. Through its IDR-driven formation of phase-separated condensates, JMJD3 regulates the 3D-conformation of the chromatin at this locus to allow gene expression.



## Conclusions

1. TGF $\beta$  induces a general reorganization of the three-dimensional structure of the chromatin at the *Chst8* locus, triggering enhancer-enhancer and enhancer-promoter contacts.
2. TGF $\beta$  drives the formation of an enhancer cluster at the *Chst8* locus that correlates with *Chst8* gene expression.
3. *Chst8* enhancer cluster constituent enhancers cooperate in an additive manner.
4. JMJD3 is essential for the TGF $\beta$ -driven *Chst8* enhancer cluster establishment and *Chst8* gene activation.
5. TGF $\beta$  and JMJD3 cooperate to drive three-dimensional chromatin structure reorganization at *Ldlrad4* and *Aopep* loci.
6. JMJD3 plays a structural role at enhancers that is independent on its demethylase catalytic activity.
7. JMJD3 is a highly disordered protein that contains a long intrinsically disordered region (IDR) enriched in hydrophobic and charged residues.
8. JMJD3 is able to form condensates through liquid-liquid phase separation in an IDR-dependent and a catalytic-independent manner.
9. JMJD3 condensates are found at the *Chst8* locus and correlate with transcriptional features.
10. JmjC-containing KDMs (JMJC-KDMs) possess IDRs enriched in charged and hydrophobic residues, and are predicted to phase separate.
11. KDM2A, KDM4B and PHF2 form phase-separated biomolecular condensates that correlate to their role as transcriptional regulators.
12. PHF2 phase separation capacity depends on a charged region of its IDR. This dependency was also observed for the proper expression of some PHF2 regulated genes.



## Bibliography

1. Buecker C, Wysocka J. Enhancers as information integration hubs in development: Lessons from genomics. *Trends in Genetics*. 2012.
2. Kouzarides T. Chromatin Modifications and Their Function. *Cell*. 2007 Feb 23;128(4):693–705.
3. Rosa S, Shaw P. Insights into Chromatin Structure and Dynamics in Plants. *Biology* 2013, Vol 2, Pages 1378-1410. 2013 Nov 28;2(4):1378–410.
4. Ea V, Baudement MO, Lesne A, Forné T. Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. *Genes (Basel)*. 2015 Jul 27;6(3):734–50.
5. Kornberg RD. Chromatin structure: A repeating unit of histones and DNA. *Science* (1979). 1974 May 24;184(4139):868–71.
6. Bogdanović O, Lister R. DNA methylation and the preservation of cell identity. *Current Opinion in Genetics & Development*. 2017 Oct 1;46:9–14.
7. Bird AP, Wolffe AP. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell*. 1999 Nov 24;99(5):451–4.
8. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* 2012 13:7. 2012 May 29;13(7):484–92.
9. Bestor TH. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *The EMBO Journal*. 1992;11(7):2611.
10. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics* 1998 19:3. 1998;19(3):219–20.
11. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev*. 2016 Apr 1;30(7):733–50.
12. Nabel CS, Manning SA, Kohli RM. The Curious Chemical Biology of Cytosine: Deamination, Methylation and Oxidation as Modulators of Genomic Potential. *ACS Chemical Biology*. 2012 Jan 20;7(1):20.
13. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001 Aug 10;293(5532):1074–80.
14. Iwasaki W, Miya Y, Horikoshi N, Osakabe A, Taguchi H, Tachiwana H, et al. Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio*. 2013 Jan 1;3(1):363–9.
15. Lawrence M, Daujat S, Schneider R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet*. 2016 Jan 1;32(1):42–56.
16. Lee DY, Hayes JJ, Pruss D, Wolffe AP. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*. 1993 Jan 15;72(1):73–84.
17. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Research* 2011 21:3. 2011 Feb 15;21(3):381–95.
18. Khan SN, Khan AU. Role of histone acetylation in cell physiology and diseases: An update. *Clinica Chimica Acta*. 2010 Oct 9;411(19–20):1401–11.
19. Hodawadekar SC, Marmorstein R. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* 2007 26:37. 2007 Aug 13; 26(37):5528–40.

20. Yang XJ, Seto E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*. 2007 Aug 13;26(37):5310–8.
21. Das C, Lucia MS, Hansen KC, Tyler JK. CBP / p300-mediated acetylation of histone H3 on lysine 56. *Nature*. 2009 May 7;459(7243):113.
22. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 2010 Dec 14;107(50):21931–6.
23. Sawicka A, Seiser C. Sensing core histone phosphorylation - a matter of perfect timing. *Biochim Biophys Acta*. 2014;1839(8):711–8.
24. Meas R, Mao P. Histone ubiquitylation and its roles in transcription and DNA damage response. *DNA Repair (Amst)*. 2015;36:36–42.
25. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000 406:6796. 2000 Aug 10;406(6796):593–9.
26. Schotta G, Sengupta R, Kubicek S, Malin S, Kauer M, Callén E, et al. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev*. 2008 Aug 1;22(15):2048–61.
27. Jung HR, Pasini D, Helin K, Jensen ON. Quantitative Mass Spectrometry of Histones H3.2 and H3.3 in Suz12-deficient Mouse Embryonic Stem Cells Reveals Distinct, Dynamic Post-translational Modifications at Lys-27 and Lys-36. *Molecular & Cellular Proteomics : MCP*. 2010 May;9(5):838.
28. Bochyńska A, Lüscher-Firzlaff J, Lüscher B. Modes of Interaction of KMT2 Histone H3 Lysine 4 Methyltransferase/COMPASS Complexes with Chromatin. *Cells*. 2018 Mar 2;7(3).
29. Kooistra SM, Helin K. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol*. 2012 May;13(5):297–311.
30. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006 Apr 21;125(2):315–26.
31. Lan F, Shi Y. Epigenetic regulation: methylation of histone and non-histone proteins. *Science in China Series C: Life Sciences* 2009 52:4. 2009 Apr 21;52(4):311–22.
32. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature* 2011 469:7330. 2011 Jan 19;469(7330):343–9.
33. Schuettengruber B, Cavalli G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development*. 2009 Nov 1;136(21):3531–42.
34. Husmann D, Gozani O. Histone lysine methyltransferases in biology and disease. *Nat Struct Mol Biol*. 2019 Oct 1;26(10):880–9.
35. Fueyo R, García MA, Martínez-Balbás MA. Jumonji family histone demethylases in neural development. *Cell Tissue Res*. 2015 Jan 1 ;359(1):87–98.
36. Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet*. 2006 Sep 8;7(9):715–27.
37. Pedersen MT, Helin K. Histone demethylases in development and disease. *Trends Cell Biol*. 2010 Nov;20(11):662–71.

38. Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, et al. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 2009 461:7262. 2009 Sep 2; 461(7262):415–8.
39. Metzger E, Wissmann M, Yin N, Müller JM, Schneider R, Peters AHFM, et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature*. 2005 Sep 15;437(7057):436–9.
40. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*. 2004 Dec 29;119(7):941–53.
41. Karytinis A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, et al. A Novel Mammalian Flavin-dependent Histone Demethylase. *The Journal of Biological Chemistry*. 2009 Jun 26;284(26):17775.
42. Tsukada YI, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, et al. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 2005 439:7078. 2005 Dec 18;439(7078):811–6.
43. Cao X, Chen Y, Wu B, Wang X, Xue H, Yu L, et al. Histone H4K20 Demethylation by Two hHR23 Proteins. *Cell Reports*. 2020 Mar 24;30(12):4152–4164.e6.
44. Bergink S, Toussaint W, Luijsterburg MS, Dinant C, Alekseev S, Hoeijmakers JHJ, et al. Recognition of DNA damage by XPC coincides with disruption of the XPC-RAD23 complex. *J Cell Biol*. 2012 Mar 19;196(6):681–8.
45. Tateishi K, Okada Y, Kallin EM, Zhang Y. Role of Jhdm2a in regulating metabolic gene expression and obesity resistance. *Nature*. 2009 Apr 9;458(7239):757–61.
46. Loh YH, Zhang W, Chen X, George J, Ng HH. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev*. 2007 Oct 15;21(20):2545–57.
47. Jepsen K, Solum D, Zhou T, McEvelly RJ, Kim HJ, Glass CK, et al. SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. *Nature*. 2007 Nov 15;450(7168):415–9.
48. Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, et al. The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell*. 2007 Mar 23;128(6):1077–88.
49. Hsia DA, Tepper CG, Pochampalli MR, Hsia EYC, Izumiya C, Huerta SB, et al. KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proc Natl Acad Sci U S A*. 2010 May 25; 107(21):9671–6.
50. Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK, Chen S, et al. A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature*. 2007 Oct 11;449(7163):689–94.
51. de Santa F, Totaro MG, Prosperini E, Notarbartolo S, Testa G, Natoli G. The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell*. 2007 Sep 21;130(6):1083–94.
52. Min GL, Villa R, Trojer P, Norman J, Yan KP, Reinberg D, et al. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science*. 2007 Oct 19;318(5849):447–50.
53. Vandamme J, Lettier G, Sidoli S, di Schiavi E, Nørregaard Jensen O, Salcini AE. The *C. elegans* H3K27 Demethylase UTX-1 Is Essential for Normal Development, Independent of Its Enzymatic Activity. *PLoS Genetics*. 2012 May;8(5):e1002647.

54. Burchfield JS, Li Q, Wang HY, Wang RF. JMJD3 as an epigenetic regulator in development and disease. *Int J Biochem Cell Biol*. 2015 Oct 1;67:148–57.
55. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015 Jan 23;347(6220).
56. Burgold T, Spreafico F, de Santa F, Totaro MG, Prosperini E, Natoli G, et al. The Histone H3 Lysine 27-Specific Demethylase Jmjd3 Is Required for Neural Commitment. *PLOS ONE*. 2008 Aug 21;3(8):e3034.
57. Agger K, Cloos PAC, Rudkjær L, Williams K, Andersen G, Christensen J, et al. The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence. *Genes Dev* 2009 May 15 23(10):1171–6.
58. Ding Y, Yao Y, Gong X, Zhuo Q, Chen J, Tian M, et al. JMJD3: a critical epigenetic regulator in stem cell fate. *Cell Communication and Signaling*. 2021 Dec 1;19(1):1–9.
59. Burchfield JS, Li Q, Wang HY, Wang RF. JMJD3 as an epigenetic regulator in development and disease. Vol. 67, *International Journal of Biochemistry and Cell Biology*. Elsevier Ltd; 2015. p. 148–57.
60. Zhang X, Liu L, Yuan X, Wei Y, Wei X. JMJD3 in the regulation of human diseases. *Protein and Cell* 2019;10(12):864–82.
61. Li Q, Zou J, Wang M, Ding X, Chepelev I, Zhou X, et al. Critical role of histone demethylase Jmjd3 in the regulation of CD4+ T-cell differentiation. *Nat Commun*. 2014;5.
62. Satoh T, Takeuchi O, Vandenbon A, Yasuda K, Tanaka Y, Kumagai Y, et al. The Jmjd3-Irf4 axis regulates M2 macrophage polarization and host responses against helminth infection. *Nat Immunol*. 2010 Oct;11(10):936–44.
63. Burgold T, Voituron N, Caganova M, Tripathi PP, Menuet C, Tusi BK, et al. The H3K27 Demethylase JMJD3 Is Required for Maintenance of the Embryonic Respiratory Neuronal Network, Neonatal Breathing, and Survival. *Cell Reports*. 2012 Nov 29;2(5):1244–58.
64. Ohtani K, Zhao C, Dobрева G, Manavski Y, Kluge B, Braun T, et al. Jmjd3 controls mesodermal and cardiovascular differentiation of embryonic stem cells. *Circ Res*. 2013 Sep 13;113(7):856–62.
65. Dai JP, Lu JY, Zhang Y, Shen YF. Jmjd3 activates Mash1 gene in RA-induced neuronal differentiation of P19 cells. *J Cell Biochem*. 2010 Aug 15;110(6):1457–63.
66. Estarás C, Akizu N, García A, Beltrán S, de la Cruz X, Martínez-Balbás MA. Genome-wide analysis reveals that Smad3 and JMJD3 HDM co-activate the neural developmental program. *Development*. 2012 Aug 1;139(15):2681–91.
67. Fueyo R, Iacobucci S, Pappa S, Estarás C, Lois S, Vicioso-Mantis M, et al. Lineage specific transcription factors and epigenetic regulators mediate TGFβ-dependent enhancer activation. *Nucleic Acids Res*. 2018;46(7):3351–65.
68. Akizu N, Estarás C, Guerrero L, Martí E, Martínez-Balbás MA. H3K27me3 regulates BMP activity in developing spinal cord. *Development*. 2010 Sep 1;137(17):2915–25.
69. Alexaki VI, Fodelianaki G, Neuwirth A, Mund C, Kourgiantaki A, Ieronimaki E, et al. DHEA inhibits acute microglia-mediated inflammation through activation of the TrkA-Akt1/2-CREB-Jmjd3 pathway. *Mol Psychiatry*. 2018 Jun 1;23(6):1410–20.



70. Tang Y, Li T, Li J, Yang J, Liu H, Zhang XJ, et al. Jmjd3 is essential for the epigenetic modulation of microglia phenotypes in the immune pathogenesis of Parkinson's disease. *Cell Death and Differentiation*. 2014 Mar;21(3):369.
71. He XB, Kim M, Kim SY, Yi SH, Rhee YH, Kim T, et al. Vitamin c facilitates dopamine neuron differentiation in fetal midbrain through tet1- and jmjd3-dependent epigenetic control manner. *Stem Cells*. 2015 Apr 1;33(4):1320.
72. Fonseca MB, Nunes AF, Rodrigues CMP. c-Jun Regulates the Stability of Anti-Apoptotic  $\Delta$ Np63 in Amyloid- $\beta$ -Induced Apoptosis. *Journal of Alzheimer's Disease*. 2012 Jan 1;28(3):685–94.
73. Fonseca MB, Nunes AF, Morgado AL, Solá S, Rodrigues CMP. TAp63y Demethylation Regulates Protein Stability and Cellular Distribution during Neural Stem Cell Differentiation. *PLOS ONE*. 2012 Dec 14;7(12):e52417.
74. de Santa F, Narang V, Yap ZH, Tusi BK, Burgold T, Austenaa L, et al. Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. *EMBO J*. 2009 Nov;28(21):3341–52.
75. Hanisch UK. Linking STAT and TLR signaling in microglia: A new role for the histone demethylase Jmjd3. *Journal of Molecular Medicine*. 2014 Mar 1;92(3):197–200.
76. Das ND, Jung KH, Choi MR, Yoon HS, Kim SH, Chai YG. Gene networking and inflammatory pathway analysis in a JMJD3 knockdown human monocytic cell line. *Cell Biochem Funct*. 2012 Apr;30(3):224–32.
77. Stolerman ES, Francisco E, Stallworth JL, Jones JR, Monaghan KG, Keller-Ramey J, et al. Genetic variants in the KDM6B gene are associated with neurodevelopmental delays and dysmorphic features. *Am J Med Genet A*. 2019 Jul 1;179(7):1276–86.
78. Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordân R, Rohs R. Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci*. 2014;39(9):381–99.
79. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet*. 2012 Apr;13(4):233–45.
80. Luse DS. The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled? *Transcription*. 2014;5(1).
81. Liu X, Kraus WL, Bai X. Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem Sci*. 2015 Sep 1;40(9):516–25.
82. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A phase separation model predicts key features of transcriptional control. *Cell*. 2017 Mar 23;169(1):13.
83. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 2010 470:7333. 2010 Dec 15;470(7333):279–83.
84. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res*. 2011 Oct 21(10):1757–67.
85. Peng Y, Zhang Y. Enhancer and super-enhancer: Positive regulators in gene transcription. *Animal Models and Experimental Medicine*. 2018 Sep;1(3):169.

86. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010 465:7295. 2010 Apr 14;465(7295):182–7.
87. de Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 2010 May;8(5).
88. Peng Y, Zhang Y. Enhancer and super-enhancer: Positive regulators in gene transcription. *Animal Models and Experimental Medicine.* 2018 Sep;1(3):169–79.
89. Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol.* 2011 Aug;18(8):956–63.
90. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell.* 2013 Apr 11;153(2):307–19.
91. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013 Nov 7;155(4):934.
92. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. *Cell.* 2017;169:13–23.
93. Hay D, Hughes JR, Babbs C, Davies JOJ, Graham BJ, Hanssen LLP, et al. Genetic dissection of the  $\alpha$ -globin super-enhancer in vivo. *Nature Genetics* 2016 48:8. 2016 Jul 4;48(8):895–903.
94. Maston GA, Evans SK, Green MR. Transcriptional Regulatory Elements in the Human Genome. 2006 Sep 1;7:29–59.
95. Kolovos P, Knoch TA, Grosveld FG, Cook PR, Papantonis A. Enhancers and silencers: An integrated and simple model for their function. *Epigenetics and Chromatin.* 2012 Jan 9;5(1):1–8.
96. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research.* 2019 Jan 8;47(D1):D330–8.
97. Ausió J. The shades of gray of the chromatin fiber: recent literature provides new insights into the structure of chromatin. *Bioessays.* 2015 Jan;37(1):46–51.
98. Prieto EI, Maeshima K. Dynamic chromatin organization in the cell. *Essays Biochem.* 2019; 63(1):133–45.
99. Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology* 2012 13:7. 2012 Jun 22;13(7):436–47.
100. Bascom G, Schlick T. Linking Chromatin Fibers to Gene Folding by Hierarchical Looping. *Biophys J.* 2017 Feb 7;112(3):434–45.
101. Jansen A, Verstrepen KJ. Nucleosome Positioning in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews: MMBR.* 2011 Jun;75(2):301.
102. Szalaj P, Plewczynski D. Three-dimensional organization and dynamics of the genome. *Cell Biol Toxicol.* 2018 Oct 1;34(5):381–404.
103. McCord RP, Kaplan N, Giorgetti L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol Cell.* 2020 Feb;77(4):688–708.
104. Kempfer R, Pombo A. Methods for mapping 3D chromosome architecture. *Nature Reviews Genetics.* 2020 Aug 11;21(4):207–26.

105. Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet.* 2007 Feb;8(2):104–15.
106. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014 Dec 18;159(7):1665–80.
107. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009 Oct 9;326(5950):289–93.
108. Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* 2010 Jun;20(6):761–70.
109. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* 2015 518:7539. 2015 Feb 18;518(7539):331–6.
110. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 2017 551:7678. 2017 Sep 27;551(7678):51–6.
111. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012 485:7398. 2012 Apr 11;485(7398):376–80.
112. Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature.* 2017 Jul 5;547(7661):61–7.
113. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature.* 2013; 503(7475):290–4.
114. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. *Nature Reviews Genetics* 2018 19:7. 2018 Apr 24;19(7):453–67.
115. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 2012 May 17;485(7398):381–5.
116. Matharu N, Ahituv N. Minor Loops in Major Folds: Enhancer–Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLOS Genetics.* 2015;11(12):e1005640.
117. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 2016 May 31;15(9):2038–49.
118. Kadauke S, Blobel GA. Chromatin loops in gene regulation. *Biochim Biophys Acta.* 2009 Jan; 1789(1):17–25.
119. Mobley AS. Induced Pluripotent Stem Cells. In: *Neural Stem Cells and Adult Neurogenesis.* Elsevier; 2019. p. 67–94.
120. Adcock IM, Caramori G. Transcription Factors. *Asthma and COPD.* 2009 Jan 1;373–80.
121. Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordân R, Rohs R. Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci.* 2014;39(9):381–99.
122. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. *Genes Dev.* 2014 Dec 15;28(24):2679–92.

123. Raposo AASF, Vasconcelos FF, Drechsel D, Marie C, Johnston C, Dolle D, et al. Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis. *Cell Rep.* 2015 Mar 10;10(9):1544–56.
124. Spiegelman BM, Heinrich R. Biological control through regulated transcriptional coactivators. *Cell.* 2004 Oct 15;119(2):157–67.
125. Katayama Y, Nishiyama M, Shoji H, Ohkawa Y, Kawamura A, Sato T, et al. CHD8 haploinsufficiency results in autistic-like phenotypes in mice. *Nature.* 2016;537(7622):675–9.
126. Micucci JA, Sperry ED, Martin DM. Chromodomain Helicase DNA-Binding Proteins in Stem Cells and Human Developmental Diseases. *Stem Cells and Development.* 2015 Apr 15;24(8):917.
127. Sims RJ, Mandal SS, Reinberg D. Recent highlights of RNA-polymerase-II-mediated transcription. *Curr Opin Cell Biol.* 2004 Jun;16(3):263–71.
128. Harlen KM, Churchman LS. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol.* 2017 Apr 1;18(4):263–73.
129. Wilson SW, Houart C. Early steps in the development of the forebrain. *Dev Cell.* 2004 Feb;6(2):167–81.
130. Molyneaux BJ, Arlotta P, Menezes JRL, Macklis JD. Neuronal subtype specification in the cerebral cortex. *Nature Reviews Neuroscience* 2007 8:6. 2007 Jun;8(6):427–37.
131. Barry DS, Pakan JMP, McDermott KW. Radial glial cells: Key organisers in CNS development. *The International Journal of Biochemistry & Cell Biology.* 2014 Jan 1;46(1):76–9.
132. Tiberi L, Vanderhaeghen P, van den Aemele J. Cortical neurogenesis and morphogens: diversity of cues, sources and functions. *Curr Opin Cell Biol.* 2012 Apr;24(2):269–76.
133. Ohtaka-Maruyama C, Okado H. Molecular pathways underlying projection neuron production and migration during cerebral cortical development. *Frontiers in Neuroscience.* 2015;9(DEC):447.
134. Fernández V, Llinares-Benadero C, Borrell V. Cerebral cortex expansion and folding: what have we learned? *EMBO J.* 2016 May 17;35(10):1021–44.
135. Ornitz DM, Itoh N. The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip Rev Dev Biol.* 2015 May 1;4(3):215–66.
136. Shimojo H, Ohtsuka T, Kageyama R. Oscillations in notch signaling regulate maintenance of neural progenitors. *Neuron.* 2008 Apr 10;58(1):52–64.
137. Komada M. Sonic hedgehog signaling coordinates the proliferation and differentiation of neural stem/progenitor cells by regulating cell cycle kinetics during development of the neocortex. *Congenit Anom (Kyoto).* 2012 Jun;52(2):72–7.
138. Chenn A, Walsh CA. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science.* 2002 Jul 19;297(5580):365–9.
139. Siegenthaler JA, Miller MW. Transforming growth factor beta 1 promotes cell cycle exit through the cyclin-dependent kinase inhibitor p21 in the developing cerebral cortex. *J Neurosci.* 2005 Sep 21;25(38):8627–36.
140. Vogel T, Ahrens S, Büttner N, Kriegstein K. Transforming growth factor beta promotes neuronal cell fate of mouse cortical and hippocampal progenitors in vitro and in vivo: identification of Nedd9 as an essential signaling component. *Cereb Cortex.* 2010 Mar;20(3):661–71.

141. Ross SE, Greenberg ME, Stiles CD. Basic Helix-Loop-Helix Factors in Cortical Development. *Neuron*. 2003 Jul 3;39(1):13–25.
142. Wilkinson G, Dennis D, Schuurmans C. Proneural genes in neocortical development. *Neuroscience*. 2013 Dec 3;253:256–73.
143. Taverna E, Götz M, Huttner WB. The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu Rev Cell Dev Biol*. 2014;30:465–502.
144. Pollard SM, Conti L, Sun Y, Goffredo D, Smith A. Adherent neural stem (NS) cells from fetal and adult forebrain. *Cereb Cortex*. 2006 Jul;16
145. Spencer Currie D, Monuki ES. Flash Freezing and Cryosectioning E12.5 Mouse Brain. *Journal of Visualized Experiments : JoVE*. 2007 May;(4).
146. Conti L, Cattaneo E. Neural stem cell systems: physiological players or in vitro entities? *Nat Rev Neurosci*. 2010 Mar;11(3):176–87.
147. Gordon KJ, Blobel GC. Role of transforming growth factor-beta superfamily signaling pathways in human disease. *Biochim Biophys Acta*. 2008 Apr;1782(4):197–228.
148. Shi Y, Massagué J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*. 2003 Jun 13 [;113(6):685–700.
149. Akhurst RJ, Hata A. Targeting the TGFβ signalling pathway in disease. *Nat Rev Drug Discov*. 2012 Oct;11(10):790–811.
150. Massagué J. TGF-beta signal transduction. *Annu Rev Biochem*. 1998;67:753–91.
151. Massagué J, Seoane J, Wotton D. Smad transcription factors. *Genes & Development*. 2005 Dec 1;19(23):2783–810.
152. Estarás C, Fueyo R, Akizu N, Beltrán S, Martínez-Balbás MA. RNA polymerase II progression through H3K27me3-enriched gene bodies requires JMJD3 histone demethylase. *Mol Biol Cell*. 2013 Jan 1;24(3):351–60.
153. Ramadoss S, Chen X, Wang CY. Histone demethylase KDM6B promotes epithelial-mesenchymal transition. *J Biol Chem*. 2012 Dec 28;287(53):44508–17.
154. Kartikasari AER, Zhou JX, Kanji MS, Chan DN, Sinha A, Grapin-Botton A, et al. The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation. *EMBO J*. 2013 May 15;32(10):1393–408.
155. Massagué J. TGFβ signalling in context. *Nature Reviews Molecular Cell Biology* 2012 13:10. 2012 Sep 20;13(10):616–30.
156. David CJ, Massagué J. Contextual determinants of TGFβ action in development, immunity and cancer. *Nature Reviews Molecular Cell Biology* 2018 19:7. 2018 Apr 11;19(7):419–35.
157. Pohlers D, Brenmoehl J, Löffler I, Müller CK, Leipner C, Schultze-Mosgau S, et al. TGF-beta and fibrosis in different organs - molecular pathway imprints. *Biochim Biophys Acta*. 2009 Aug;1792(8):746–56.
158. Burgold T, Spreafico F, de Santa F, Totaro MG, Prosperini E, Natoli G, et al. The Histone H3 Lysine 27-Specific Demethylase Jmjd3 Is Required for Neural Commitment. *PLoS ONE*. 2008 Aug 21 3(8):3034.
159. García-Campmany L, Martí E. The TGFbeta intracellular effector Smad3 regulates neuronal differentiation and cell fate specification in the developing spinal cord. *Development*. 2007 Jan;134(1):65–75.

160. Fueyo Arévalo R, Martínez Balbás M, Suau León P, Universitat Autònoma de Barcelona. Departament de Bioquímica i Biologia Molecular. Study of the role of the histone demethylase JMJD3 as a transcriptional regulator.
161. Sabari BR, Dall'Agnese A, Young RA. Biomolecular Condensates in the Nucleus. *Trends in Biochemical Sciences*. 2020 Nov 1;45(11):961–77.
162. Pederson T. The Nucleolus. *Cold Spring Harbor Perspectives in Biology*. 2011 Mar 1;3(3):a000638.
163. Decker CJ, Parker R. P-bodies and stress granules: Possible roles in the control of translation and mRNA degradation. *Cold Spring Harbor Perspectives in Biology*. 2012 Sep;4(9).
164. Wu H. Higher-order assemblies in a new paradigm of signal transduction. Vol. 153, *Cell*. Elsevier B.V.; 2013. p. 287–92.
165. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: Organizers of cellular biochemistry.
166. Shin Y, Berry J, Pannucci N, Haataja MP, Toettcher JE, Brangwynne CP. Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell*. 2017;168(1–2):159–171.e14.
167. Frank L, Rippe K. Repetitive RNAs as Regulators of Chromatin-Associated Subcompartment Formation by Phase Separation. Vol. 432, *Journal of Molecular Biology*. Academic Press; 2020. p. 4270–86.
168. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*. 2009 Jun 26;324(5935):1729–32.
169. Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proc Natl Acad Sci U S A*. 2011 Mar 15;108(11):4334–9.
170. Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*. 2015 Aug 27;162(5):1066–77.
171. Flory PJ. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics*. 2004 Dec 22; 10(1):51.
172. Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* 2012 483:7389. 2012 Mar 7;483(7389):336–40.
173. Sanders DW, Kedersha N, Lee DSW, Strom AR, Drake V, Riback JA, et al. Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell*. 2020 Apr 16;181(2):306–324.e28. /
174. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. 2012 May 11;149(4):753–67.
175. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* 2005 6:3. 2005 Mar;6(3):197–208.
176. Diella F, Haslam N, Chica C, Budd A, Michael S, Brown NP, et al. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci*. 2008 May;13(17):6580–603.
177. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. Vol. 114, *Chemical Reviews*. American Chemical Society; 2014. p. 6589–631.

178. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Vol. 16, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group; 2015. p. 18–29.
179. Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends in Cell Biology*. 2018;28:420–435.
180. Zirkel A, Nikolic M, Sofiadis K, Mallm JP, Brackley CA, Gothe H, et al. HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types. *Mol Cell*. 2018 May 17;70(4):730-744.e6.
181. Larson AG, Narlikar GJ. The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. Vol. 57, *Biochemistry*. American Chemical Society; 2018. p. 2540–8.
182. Strom AR, Emelyanov A v., Mir M, Fyodorov D v., Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. *Nature*. 2017 Jul 13;547(7662):241–5.
183. Hyman AA, Weber CA, Jülicher F. Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology*. 2014;30:39–58.
184. Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A, Parker R. ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell*. 2016 Jan 28;164(3):487–98.
185. Taniue K, Akimitsu N. Aberrant phase separation and cancer. Vol. 289, *FEBS Journal*. John Wiley and Sons Inc; 2022. p. 17–39.
186. Wright RHG, le Dily F, Beato M. ATP, Mg<sup>2+</sup>, Nuclear Phase Separation, and Genome Accessibility. *Trends in Biochemical Sciences*. 2019 Jul 1;44(7):565–74.
187. Snead WT, Gladfelter AS. The Control Centers of Biomolecular Phase Separation: How Membrane Surfaces, PTMs, and Active Processes Regulate Condensation. *Molecular Cell*. 2019;76(2):295–305.
188. Brangwynne CP, Tompa P, Pappu R v. Polymer physics of intracellular phase transitions. *Nature Physics* 2015 11:11. 2015 Nov 3;11(11):899–904.
189. Patel A, Malinowska L, Saha S, Wang J, Alberti S, Krishnan Y, et al. ATP as a biological hydrotrope. *Science*. 2017 May 19;356(6339):753–6.
190. Maharana S, Wang J, Papadopoulos DK, Richter D, Pozniakovskiy A, Poser I, et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*. 2018 May 25;360(6391):918–21.
191. Weber SC, Brangwynne CP. Inverse size scaling of the nucleolus by a concentration-dependent phase transition. *Curr Biol*. 2015 Mar 2;25(5):641–6.
192. Sabari BR, Dall’Agnese A, Young RA. Biomolecular Condensates in the Nucleus. *Trends in Biochemical Sciences*. 2020;45(11):961–77.
193. Gibson BA, Doolittle LK, Schneider MWG, Jensen LE, Gamarra N, Henry L, et al. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell*. 2019 Oct 3;179(2):470-484.e21.
194. Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu R v., et al. Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol Cell*. 2020 Mar 19;77(6):1237-1250.e4.
195. Sharma R, Choi KJ, Quan MD, Sharma S, Sankaran B, Park H, et al. Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization. *Nature Communications* 2021 12:1. 2021 Sep 22;12(1):1–17.

196. Shi B, Li W, Song Y, Wang Z, Ju R, Ulman A, et al. UTX condensation underlies its tumour-suppressive activity. *Nature*. 2021;597(7878):726–31.
197. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, et al. Liquid droplet formation by HP1 $\alpha$  suggests a role for phase separation in heterochromatin. *Nature*. 2017 Jul 13;547(7662):236–40.
198. Sanulli S, Trnka MJ, Dharmarajan V, Tibble RW, Pascal BD, Burlingame AL, et al. HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature*. 2019 Nov 14;575(7782):390–4.
199. Peng L, Li EM, Xu LY. From start to end: Phase separation and transcriptional regulation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. 2020;1863(12):194641.
200. Cook PR, Marenduzzo D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res*. 2018 Nov 2;46(19):9895–906.
201. Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*. 2019 Aug 22;572(7770):543–8.
202. Lu H, Yu D, Hansen AS, Ganguly S, Liu R, Heckert A, et al. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature*. 2018 Jun 14; 558(7709):318–23.
203. Cramer P. Organization and regulation of gene transcription. Vol. 573, *Nature*. Nature Publishing Group; 2019. p. 45–54.
204. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio A v., et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*. 2018;175(7):1842-1855.e16.
205. Shrinivas K, Sabari BR, Coffey EL, Klein IA, Boija A, Zamudio A v., et al. Enhancer Features that Drive Formation of Transcriptional Condensates. *Molecular Cell*. 2019 Aug 8;75(3):549-561.e7.
206. Palikyras S, Papantonis A. Modes of phase separation affecting chromatin regulation. *Open Biol*. 2019;9(10):190167.
207. Hsin JP, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev*. 2012 Oct 1;26(19):2119–37.
208. Mitrea DM, Kriwacki RW. Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling* 2016 14:1. 2016 Jan 5;14(1):1–20.
209. Uversky VN. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr Opin Struct Biol*. 2017 Jun 1; 44:18–30.
210. Dobson CM. The Amyloid Phenomenon and Its Links with Human Disease. *Cold Spring Harb Perspect Biol*. 2017 Jun 1;9(6).
211. Zbinden A, Pérez-Berlanga M, de Rossi P, Polymenidou M. Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. Vol. 55, *Developmental Cell*. Cell Press; 2020. p. 45–68.
212. Alberti S, Dormann D. Liquid-Liquid Phase Separation in Disease. *Annual Review of Genetics* 2019; 53:171-194
213. di Masi A, Cilli D, Berardinelli F, Talarico A, Pallavicini I, Pennisi R, et al. PML nuclear body disruption impairs DNA double-strand break sensing and repair in APL. *Cell Death Dis*. 2016;7(7).



214. Cascarina SM, Ross ED. Phase separation by the SARS-CoV-2 nucleocapsid protein: Consensus and open questions. *Journal of Biological Chemistry*. 2022 Mar 1;298(3):101677.
215. Spencer Currie D, Hu JS, Kolski-Andreaco A, Monuki ES. Culture of mouse neural stem cell precursors. *Journal of Visualized Experiments*. 2006;152:152.
216. Estarás C, Akizu N, García A, Beltrán S, de la Cruz X, Martínez-Balbás MA. Genome-wide analysis reveals that Smad3 and JMJD3 HDM co-activate the neural developmental program. *Development*. 2012;139:2681–91.
217. Pollard SM, Conti L, Sun Y, Goffredo D, Smith A. Adherent neural stem (NS) cells from fetal and adult forebrain. *Cereb Cortex*. 2006 Jul;16
218. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*. 1977;36(1):59–72.
219. Todaro gj, green h. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *The Journal of Cell Biology*. 1963;17(2):299.
220. Blanco-García N, Asensio-Juan E, de La Cruz X, Martínez-Balbás MA. Autoacetylation Regulates P/CAF Nuclear Localization. *Journal of Biological Chemistry*. 2009 Jan 16;284(3):1343–52.
221. Fraser J, Williamson I, Bickmore WA, Dostie J. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiol Mol Biol Rev*. 2015 Sep;79(3):347–72.
222. Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc*. 2013 Feb; 8(3):509–24.
223. Splinter E, de Wit E, van de Werken HJG, Klous P, de Laat W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods*. 2012 Nov;58(3):221–30.
224. MM B. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. 1976 May 7;72(1–2):248–54.
225. Zamudio A v., Dall’Agnese A, Henninger JE, Manteiga JC, Afeyan LK, Hannett NM, et al. Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol Cell*. 2019 Dec 5;76(5):753-766.e6.
226. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 2012 9:7. 2012 Jun 28;9(7):676–82.
227. Koulouras G, Panagopoulos A, Rapsomaniki MA, Giakoumakis NN, Taraviras S, Lygerou Z. EasyFRAP-web: a web-based tool for the analysis of fluorescence recovery after photobleaching data. *Nucleic Acids Res*. 2018 Jul 2;46(W1):W467–72.
228. Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, de Laat W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods*. 2020 Jan 1;170:17–32.
229. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res*. 2013 Jul;41(13).

230. Karolchik D, Hinricks AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004 Jan 1;32.
231. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol.* 2005 Feb;3(1):35–60.
232. Dosztányi Z, Csizmók V, Tompa P, Simon I. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *Journal of Molecular Biology.* 2005 Apr 8;347(4):827–39.
233. Peng QY, Zhang QF. Precise positions of Phoebe determined with CCD image-overlapping calibration. *Monthly Notices of the Royal Astronomical Society.* 2006 Apr 17 [;366(1):208–12.
234. Chu X, Sun T, Li Q, Xu Y, Zhang Z, Lai L, et al. Prediction of liquid–liquid phase separating proteins using machine learning. *BMC Bioinformatics.* 2022 Dec 1;23(1):1–13.
235. Bolognesi B, Gotor NL, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, et al. A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Rep.* 2016;16(1):222–31.
236. Piovesan D, Tabaro F, Paladin L, Necci M, Mičetić I, Camilloni C, et al. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.* 2018 Jan 1;46(D1):D471–6.
237. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003 Jul 1;31(13):3784–8.
238. Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A.* 1981;78(6):3824–8.
239. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997 Sep 1;25(17):3389–402.
240. Scientifi A. The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation1. 2014 Jan 1;3(1):11.IT.3.1.
241. Furlong EEM, Levine M. Developmental enhancers and chromosome topology. *Science.* 2018 Sep 28;361(6409):1341–5.
242. Bleckwehl T, Rada-Iglesias A. Transcriptional and epigenetic control of germline competence and specification. *Curr Opin Cell Biol.* 2019 Dec 1;61:1–8.
243. Vicioso-Mantis M, Fueyo R, Navarro C, Cruz-Molina S, van Ijcken WFJ, Rebollo E, et al. JMJD3 intrinsically disordered region links the 3D-genome structure to TGFβ-dependent transcription activation. *Nature Communications* 2022 13:1. 2022 Jun 7;13(1):1–15.
244. Fuxreiter M. Fuzziness in Protein Interactions-A Historical Perspective. *J Mol Biol.* 2018 Aug 3;430(16):2278–87.
245. Pak CW, Kosno M, Holehouse AS, Padrick SB, Mittal A, Ali R, et al. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell.* 2016 Jul 7;63(1):72–85.
246. Cho WK, Spille JH, Hecht M, Lee C, Li C, Grube V, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (1979).* 2018 Jul 27;361(6400):412–5.
247. K P, P R, S V, AK D, Z O. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics.* 2006;7(1):208–12.

248. Sabari BR, Dall'Agnese A, Bojja A, Klein IA, Coffey EL, Shrinivas K, et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* (1979). 2018;361(6400):eaar3958.
249. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, et al. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*. 2015;57(5):936–47.
250. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smith AFA, Roskin KM, et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004 Apr;14(4):708–15.
251. Murthy AC, Dignon GL, Kan Y, Zerze GH, Parekh SH, Mittal J, et al. Molecular interactions underlying liquid–liquid phase separation of the FUS low-complexity domain. *Nature Structural & Molecular Biology* 2019 26:7. 2019 Jul 1;26(7):637–48.
252. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003 Jul 1;31(13):3784–8.
253. Williamson MP. The structure and function of proline-rich regions in proteins. *Biochemical Journal*. 1994;297:249–60.
254. Sun T, Li Q, Xu Y, Zhang Z, Lai L, Pei J. Prediction of Liquid-Liquid Phase Separation Proteins Using Machine Learning. *SSRN Electronic Journal*. 2020;
255. Weber SC, Brangwynne CP. Getting RNA and protein in phase. *Cell*. 2012 Jun 8;149(6):1188–91.
256. Gopal PP, Nirschl JJ, Klinman E, Holzbaub ELF. Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. *Proc Natl Acad Sci U S A*. 2017;114(12):E2466–75.
257. Kroschwald S, Maharana S, Simon A. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters (Zur)*. 2017;(May).
258. Henis YI, Rotblat B, Kloog Y. FRAP beam-size analysis to measure palmitoylation-dependent membrane association dynamics and microdomain partitioning of Ras proteins. *Methods*. 2006 Oct;40(2):183–90.
259. Phair RD, Misteli T. High mobility of proteins in the mammalian cell nucleus. *Nature* 2000 404:6778. 2000 Apr 6;404(6778):604–9.
260. Kruidenier L, Chung CW, Cheng Z, Liddle J, Che K, Joberty G, et al. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature*. 2012;488(7411):404–8.
261. Dimitrova E, Turberfield AH, Klose RJ. Histone demethylases in chromatin biology and beyond. *EMBO Rep*. 2015 Dec 1 [;16(12):1620–39.
262. Mir M, Bickmore W, Furlong EEM, Narlikar G. Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development*. 2019 Aug 6;146(19).
263. Pappa S, Padilla N, Iacobucci S, Vicioso M, de La Campa EA, Navarro C, et al. PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitors. *Proc Natl Acad Sci U S A*. 2019 Sep 24;116(39):19464–73.
264. Gaarenstroom T, Hill CS. TGF- $\beta$  signaling to chromatin: How Smads regulate transcription during self-renewal and differentiation. *Seminars in Cell & Developmental Biology*. 2014 Aug 1;32:107–18.
265. Snetkova V, Skok JA. Enhancer talk. *Epigenomics*. 2018;10(4):483.

266. Cabral RM, Kurban M, Wajid M, Shimomura Y, Petukhova L, Christiano AM. Whole-exome sequencing in a single proband reveals a mutation in the CHST8 gene in autosomal recessive peeling skin syndrome. *Genomics*. 2012 Apr;99(4):202–8.
267. Hay D, Hughes JR, Babbs C, Davies JOJ, Graham BJ, Hanssen LLP, et al. Genetic dissection of the  $\alpha$ -globin super-enhancer in vivo. *Nature Genetics* 2016 48:8. 2016 Jul 4;48(8):895–903.
268. Huang J, Liu X, Li D, Shao Z, Cao H, Zhang Y, et al. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Developmental Cell*. 2016 Jan 11;36(1):9–23.
269. Shin HY, Willi M, Yoo KH, Zeng X, Wang C, Metser G, et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet*. 2016 Aug 1;48(8):904–11.
270. Hnisz D, Schuijers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, et al. Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Molecular Cell*. 2015 Apr 16; 58(2):362–70.
271. Huang J, Li K, Cai W, Liu X, Zhang Y, Orkin SH, et al. Dissecting super-enhancer hierarchy based on chromatin interactions. *Nature Communications* 2018 9:1. 2018 Mar 5;9(1):1–12.
272. Kartikasari AER, Zhou JX, Kanji MS, Chan DN, Sinha A, Grapin-Botton A, et al. The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation. *EMBO J*. 2013 May 15;32(10):1393–408.
273. Williams K, Christensen J, Rappsilber J, Nielsen AL, Johansen JV, Helin K, et al. The Histone Lysine Demethylase JMJD3/KDM6B Is Recruited to p53 Bound Promoters and Enhancer Elements in a p53 Dependent Manner.
274. Oudelaar AM, Higgs DR. The relationship between genome structure and function. *Nature Reviews Genetics*. 2021;22(3):154–68.
275. Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, et al. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res*. 2010 May;20(5):578–88.
276. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 2010 467:7314. 2010 Aug 18;467(7314):430–5.
277. Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics*. 2019;20(8):437–55.
278. Benabdallah NS, Williamson I, Illingworth RS, Kane L, Boyle S, Sengupta D, et al. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell*. 2019 Nov 7;76(3):473–484.e7.
279. Alexander JM, Guan J, Li B, Maliskova L, Song M, Shen Y, et al. Live-cell imaging reveals enhancer-dependent sox2 transcription in the absence of enhancer proximity. *Elife*. 2019 May 1;8.
280. Smol T, Sigé J, Thuillier C, Frénois F, Brunelle P, Rama M, et al. Lessons from the analysis of TAD boundary deletions in normal population. *bioRxiv*. 2020 Apr 3;2020.04.01.021188.
281. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 2017 551:7678. 2017 Sep 27;551(7678):51–6.

282. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin loss eliminates all loop domains. *Cell*. 2017 Oct 5; 171(2):305.
283. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*. 2017 May 18;169(5):930.
284. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015 May 30;161(5):1012–25.
285. Laugsch M, Bartusel M, Rehim R, Alirzayeva H, Karaolidou A, Crispatzu G, et al. Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell*. 2019 May 2;24(5):736-752.e12.
286. Cavalheiro GR, Pollex T, Furlong EE. To loop or not to loop: what is the role of TADs in enhancer function and gene regulation? *Current Opinion in Genetics and Development*. 2021;67:119–29.
287. Zhang H, Emerson DJ, Gilgenast TG, Titus KR, Lan Y, Huang P, et al. Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* 2019 576:7785. 2019 Nov 27;576(7785):158–62.
288. van Steensel B, Furlong EEM. The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology* 2019 20:6. 2019 Mar 18;20(6):327–37.
289. Owen I, Shewmaker F. The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins. *International Journal of Molecular Sciences*. 2019 Nov 1; 20(21).
290. Biesaga M, Frigolé-Vivas M, Salvatella X. Intrinsically disordered proteins and biomolecular condensates as drug targets. *Current Opinion in Chemical Biology*. 2021 Jun 1;62:90–100.
291. Lađinović D, Novotná J, Jakšová S, Raška I, Vacík T, Lađinovi D, et al. A demethylation deficient isoform of the lysine demethylase KDM2A interacts with pericentromeric heterochromatin in an HP1a-dependent manner. 2017
292. Cramer P. Organization and regulation of gene transcription. *Nature*. 2019;573(7772):45–54.
293. Schneider N, Wieland FG, Kong D, Fischer AAM, Hörner M, Timmer J, et al. Liquid-liquid phase separation of light-inducible transcription factors increases transcription activation in mammalian cells and mice. *Science Advances*. 2021 Jan 1;7(1).
294. Wagh K, Garcia DA, Upadhyaya A. Phase separation in transcription factor dynamics and chromatin organization. *Current Opinion in Structural Biology*. 2021 Dec 1;71:148–55.
295. Chen S, Ma J, Wu F, Xiong LJ, Ma H, Xu W, et al. The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. *Genes Dev*. 2012 Jun;26(12):1364–75.
296. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription.
297. Wang W, Qiao S, Li G, Cheng J, Yang C, Zhong C, et al. A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated enhancer clusters. *Nucleic Acids Research*. 2022 May 20;50(9):4917–37.

298. Trojanowski J, Frank L, Rademacher A, Mücke N, Grigaitis P, Rippe K. Transcription activation is enhanced by multivalent interactions independent of phase separation. *Molecular Cell*. 2022 May;82(10):1878-1893.e10.
299. Battle E, Massagué J. Transforming Growth Factor- $\beta$  Signaling in Immunity and Cancer. *Immunity*. 2019 Apr;50(4):924–40.
300. Duncan AR, Vitobello A, Collins SC, Vancollie VE, Lelliott CJ, Rodan L, et al. Heterozygous Variants in KDM4B Lead to Global Developmental Delay and Neuroanatomical Defects. *Am J Hum Genet*. 2020 Dec 3;107(6):1170–7.
301. Lee JH, Yoo NJ, Kim MS, Lee SH. Histone Demethylase Gene PHF2 Is Mutated in Gastric and Colorectal Cancers. *Pathol Oncol Res*. 2017 Jul 1;23(3):471–6.
302. Klein IA, Boija A, Afeyan LK, Hawken SW, Fan M, Dall’Agnese A, et al. Partitioning of cancer therapeutics in nuclear condensates. *Science (1979)*. 2020 Jun 19;368(6497):1386–92.
303. Wang N, Liu C. Implications of liquid-liquid phase separation in plant chromatin organization and transcriptional control. *Curr Opin Genet Dev*. 2019 Apr 1;55:59–65.
304. Watson M, Stott K. Disordered domains in chromatin-binding proteins. *Essays in Biochemistry*. 2019;63(1):147–56.
305. Misteli T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*. 2020 Oct 1;183(1):28–45.
306. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A*. 2018 Jul 17;115(29):E6697–706.
307. Heist T, Fukaya T, Levine M. Large distances separate coregulated genes in living *Drosophila* embryos. *Proc Natl Acad Sci U S A*. 2019;116(30):15062–7.

## Appendix





I would like to include the publications in which I participated during my doctoral thesis.

- Vicioso-Mantis, M, Aguirre, S and Martínez-Balbás, MA. JmjC family of histone demethylases form condensates *in vivo*. Under revision.
- Vicioso-Mantis, M., Fueyo, R., Navarro, C. et al. JMJD3 intrinsically disordered region links the 3D-genome structure to TGF $\beta$ -dependent transcription activation. *Nat Commun* 13, 3263 (2022).
- Cabello-Lobato, M. J. et al. Physical interactions between MCM and Rad51 facilitate replication fork lesion bypass and ssDNA gap filling by non-recombinogenic functions. *Cell Rep.* 36, 109440 (2021).
- Iacobucci S, Padilla N, Gabrielli M, Navarro C, Lombardi M, Verderio C, de la Cruz X, Martínez-Balbás MA. The histone demethylase PHF8 regulates astrocyte differentiation and function. *Development* 15 June 2021; 148 (12): dev194951.
- Pappa S, Padilla N, Iacobucci S, Vicioso-Mantis M, de la Campa E, Navarro C, Marcos E, de la Cruz X, Martínez-Balbás MA. PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitor cells. *PNAS*, (2019) 116 (39), 19464–19473





# JMJD3 intrinsically disordered region links the 3D-genome structure to TGF $\beta$ -dependent transcription activation

Marta Vicioso-Mantis<sup>1,7</sup>, Raquel Fueyo<sup>1,5,7</sup>, Claudia Navarro<sup>1</sup>, Sara Cruz-Molina <sup>2</sup>, Wilfred F. J. van Ijcken <sup>3</sup>, Elena Rebollo<sup>4</sup>, Álvaro Rada-Iglesias <sup>2,6</sup> & Marian A. Martínez-Balbás <sup>1</sup>✉

Enhancers are key regulatory elements that govern gene expression programs in response to developmental signals. However, how multiple enhancers arrange in the 3D-space to control the activation of a specific promoter remains unclear. To address this question, we exploited our previously characterized TGF $\beta$ -response model, the neural stem cells, focusing on a ~374 kb locus where enhancers abound. Our 4C-seq experiments reveal that the TGF $\beta$  pathway drives the assembly of an enhancer-cluster and precise gene activation. We discover that the TGF $\beta$  pathway coactivator JMJD3 is essential to maintain these structures. Using live-cell imaging techniques, we demonstrate that an intrinsically disordered region contained in JMJD3 is involved in the formation of phase-separated biomolecular condensates, which are found in the enhancer-cluster. Overall, in this work we uncover novel functions for the coactivator JMJD3, and we shed light on the relationships between the 3D-conformation of the chromatin and the TGF $\beta$ -driven response during mammalian neurogenesis.

<sup>1</sup>Department of Molecular Genomics, Instituto de Biología Molecular de Barcelona (IBMB), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona 08028, Spain. <sup>2</sup>Center for Molecular Medicine Cologne (CMMC), University of Cologne, Robert-Koch-Strasse 21, 50931 Cologne, Germany. <sup>3</sup>Center for Biomics, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands. <sup>4</sup>Molecular Imaging Platform, Instituto de Biología Molecular de Barcelona (IBMB), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona 08028, Spain. <sup>5</sup>Present address: Department of Chemical and Systems Biology, Stanford School of Medicine, Stanford University, Stanford, CA 94305, USA. <sup>6</sup>Present address: Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), CSIC/University of Cantabria, Santander, Spain. <sup>7</sup>These authors contributed equally: Marta Vicioso-Mantis, Raquel Fueyo. ✉email: [mmbbmc@ibmb.csic.es](mailto:mmbbmc@ibmb.csic.es)

During mammalian neurogenesis, neural stem cell (NSC) progenitors differentiate into neurons in response to different signaling pathways<sup>1</sup>. Upon developmental pathway activation, transcription factors are recruited to the chromatin, and together with epigenetic regulators, they activate cis-regulatory elements that will establish cell-specific gene expression patterns<sup>2–4</sup>. In mammals, promoters are usually regulated by more than one enhancer, and in fact, the number of enhancers in the mouse genome is one order of magnitude larger than the number of promoters<sup>5,6</sup>. This complex and sometimes-redundant configuration is crucial to ensure precise spatial-temporal control of the gene expression. However, how these multiple enhancers are orchestrated to regulate their target genes is still an open debate. Clusters of enhancers, also named super-enhancers by others<sup>7</sup>, are regions of euchromatin that are characterized by a high density of binding motifs, where transcription factors and cofactors such as Mediator, RNA-polymerase II (RNAPII) or chromatin remodelers colocalize (for review<sup>8</sup>). In recent times, it has been proposed that these clusters of enhancers facilitate transcriptional activation by promoting liquid-liquid phase separation (LLPS), a process by which molecules are condensed and concentrated in membrane-less compartments<sup>9–11</sup>. These condensates have been proposed to be formed by dynamic and weak multivalent interactions, that are characteristic of proteins that contain intrinsically disordered regions (IDR)<sup>12–18</sup>. Thus, intrinsically disordered proteins or regions have been suggested to drive the formation or to be incorporated into these biomolecular condensates. Despite the importance of enhancer clusters in cell identity establishment, we are still far from totally understanding the mechanisms by which they control gene transcription. Many research articles have revealed the importance of the 1D and 3D structure of the chromatin in development (reviewed in<sup>19–22</sup>). Nonetheless, the field lacks a specific assessment of the impact of individual developmental pathways on chromatin re-organization and function of specific loci.

To fill this gap, we have analyzed the chromatin reorganization that underlies the transforming growth factor beta (TGF $\beta$ ) pathway activation during neuronal commitment. We and others have demonstrated that in response to TGF $\beta$ , neural progenitors lose multipotency and commit to the neuronal lineage both in vivo and in vitro<sup>23–26</sup>. To do that, SMAD2/3, the major effectors of the pathway, cooperate with specific cofactors to regulate transcription. Particularly, SMAD3 interacts with the lysine demethylase (KDM) JMJD3<sup>23,27,28</sup>, a Jumonji C (JmjC) domain-containing enzyme that catalyzes the histone 3 lysine 27 trimethylation (H3K27me3) removal<sup>29,30</sup> and has been linked to numerous developmental processes (reviewed in<sup>31,32</sup>). In cortical progenitor cells, we have previously shown that JMJD3 cooperates with the TGF $\beta$  pathway to induce neuronal differentiation<sup>23,33</sup>. In this context, JMJD3 and SMAD3 together bind and trigger the activation of neural cis-regulatory elements, presumably guided by the pioneer lineage-specific transcription factor ASCL1, and cooperating with the chromatin remodeler CHD8<sup>33</sup>. Although some of the linear molecular components involved in the TGF $\beta$ -mediated enhancer activation have been identified, the relevance of their interactions at the 3D-level is still to be uncovered.

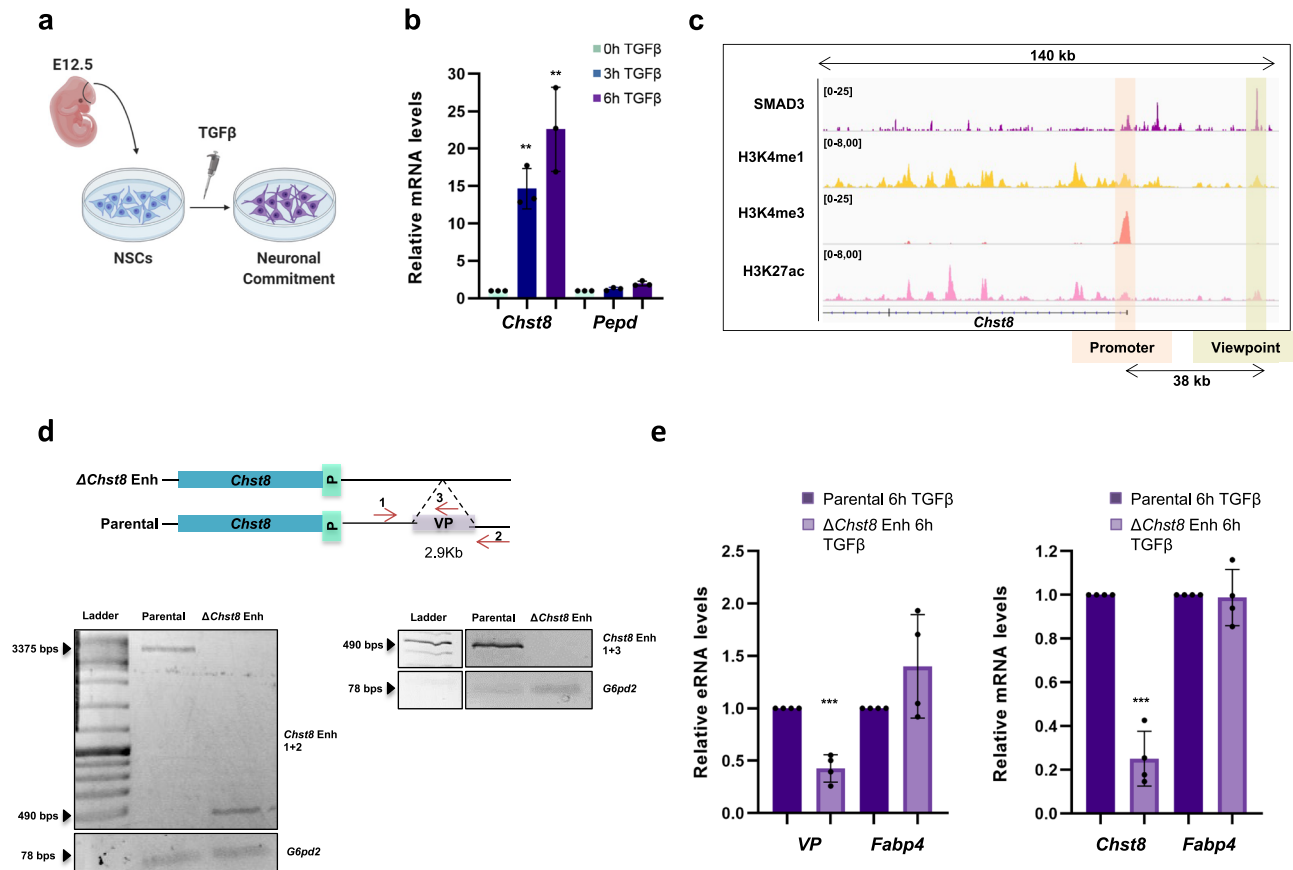
Here, we perform 4C-seq experiments, and we illustrate that TGF $\beta$  drives enhancer-enhancer contacts that facilitate an enhancer cluster assembly, and ultimately gene activation. Upon TGF $\beta$  stimulation, we observe that the establishment of multi-enhancer interactions requires the coactivator JMJD3. Using live-cell imaging and molecular biology techniques, we demonstrate that a proline-rich IDR contained in JMJD3 is essential to induce LLPS, and we report a correlation between the JMJD3-containing molecular condensates and the enhancer driven gene activation.

With our work, we reveal that JMJD3, containing a disordered domain, lies at the edge of chromatin structure and function upon TGF $\beta$  stimulation of NSCs.

## Results

**TGF $\beta$  drives enhancer cluster assembly.** The three-dimensional proximity between cis-regulatory regions has been systematically described as an intrinsic feature of the genome organization<sup>6</sup>. Nonetheless, the impact that the genome structure exerts over its function is still an open debate<sup>22</sup>. Within this framework, we hypothesize that the 3D-structure of the chromatin could be playing a role in the signal-dependent regulation of the TGF $\beta$ -responsive enhancers. To test our hypothesis, we drew upon our well-characterized TGF $\beta$  model of study, the E12.5 mouse NSCs. In these cells, the TGF $\beta$  signaling pathway is moderately active under basal conditions to allow neural progenitor proliferation<sup>23–26</sup>; however, further TGF $\beta$  stimulation leads to the activation of hundreds of enhancers and genes that induce neuronal commitment in vitro and in vivo<sup>23–26,33</sup> (Fig. 1a). Using this system, we asked whether the activation of the enhancers that occurs upon TGF $\beta$  stimulation entails 3D-chromatin changes. For this purpose, we performed 4C-seq experiments using as a viewpoint (VP) a TGF $\beta$ -regulated enhancer that lies 38 kb downstream of the carbohydrate sulfotransferase 8 (*Chst8*) gene. The rationale to select this gene was the following: first, *Chst8* is robustly upregulated upon TGF $\beta$  treatment; as indicated in Fig. 1b, the mRNA levels of the *Chst8* gene increase up to ~25-fold upon TGF $\beta$ -stimulation. Second, the distance between the VP and the *Chst8* gene allows for a reliable resolution in 4C-seq experiments (38 kb) (Fig. 1c), as it is a known-fact that one of the 4C-seq caveats is the preferential ligation of the VP with its 1D closest regions, making the contacts that appear adjacent to the bait problematic to interpret. Third, *Chst8* is a moderately long gene (138 kb), thus permitting the analysis not only of the contacts at the promoter level, but also potential interactions between the VP enhancer and the *Chst8* gene body. Before analyzing which regions contact the selected VP, we confirmed that the VP is a TGF $\beta$ -responsive *Chst8* enhancer. For that purpose, we used CRISPR/Cas9 technology to delete the VP enhancer (Fig. 1d) and measured the enhancer activity and the *Chst8* transcriptional response to TGF $\beta$ -stimulation. To this end, we evaluated the transcription of enhancer RNA (eRNA), which serves as a read-out of enhancer activation<sup>34</sup>. Results in Fig. 1e show a remarkable decrease of both enhancer activity and *Chst8* induction upon TGF $\beta$ -stimulation in the *Chst8* enhancer-deleted ( $\Delta$ *Chst8* Enh) cells, demonstrating that the VP enhancer is an essential cis-regulatory element of the *Chst8* gene.

After testing that our selected VP is a bona fide enhancer of the *Chst8* gene, we performed two independent biological replicates of a 4C-seq experiment, where we analyzed the 3D-interactions between the VP and the genome before and after 3 h of TGF $\beta$  addition. The quality of the experiments was assessed following the criteria described in<sup>35</sup> (Supplementary Data 1). The UCSC browser capture in Fig. 2a shows the obtained profiles for untreated and TGF $\beta$  treated NSCs. As expected, the proportion of cells displaying contacts between the VP enhancer and the *Chst8* promoter increased upon TGF $\beta$  treatment (see region under light orange). However, we also observed novel contacts between the VP and the *Chst8* gene body, these contacts were not particularly characterized by any type of regulatory element, but they suggest that TGF $\beta$  triggers a re-organization of the chromatin at this region (Fig. 2a and Supplementary Fig. 1a). Indeed, in the two biological replicates, the height of peaks located 500 kb upstream or downstream of the VP displayed a significant increase when NSCs cells were treated with TGF $\beta$  (p-value 0.0208), pointing to



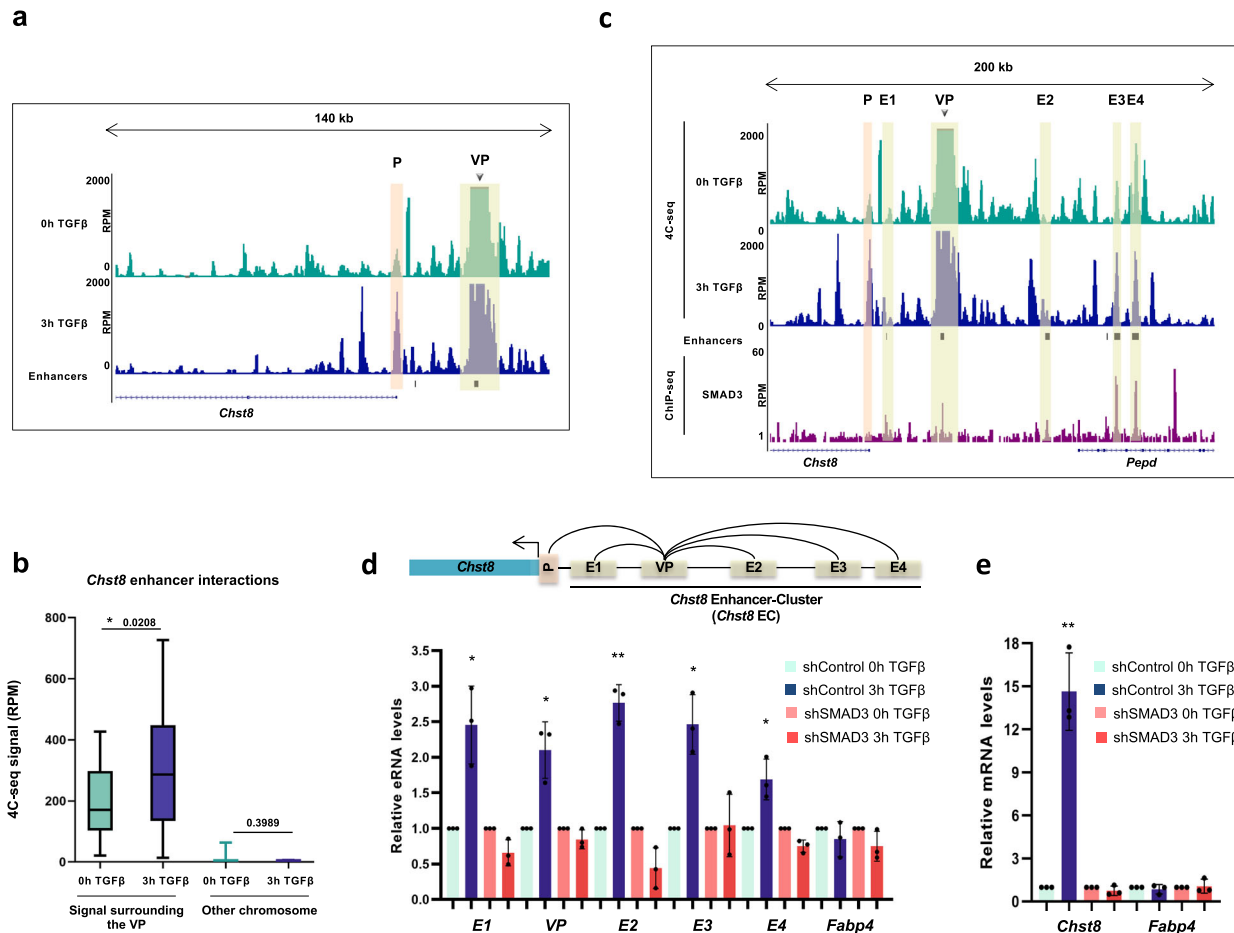
**Fig. 1 The VP is an essential *Chst8* enhancer.** **a** Schematic view of the model used in this study. NSCs were dissected from cerebral cortices of mouse fetal brains (E12.5) and cultured ex vivo (see methods). TGFβ addition leads to neuronal commitment. **b** NSCs were treated with TGFβ. Total RNA was prepared and the levels of the mRNA of the indicated genes were determined by qPCR. Values were normalized to the housekeeping gene *Gapdh* and the figure shows values relative to time 0 h. Results are the mean of three biologically independent experiments. Data are presented as mean values  $\pm$  SEM.  $^{**}p < 0.01$  (*P* values were calculated using two-tailed Student's *t* test,  $p = 0.001131677$  and  $p = 0.006143072$ ). Source data are provided as a Source Data file. **c** UCSC captures showing the chromatin landscape and SMAD3 binding around the *Chst8* gene promoter and the *Chst8* putative enhancer (VP) in NSCs. Tracks display ChIP-seq in NSCs treated with TGFβ (SMAD3) or untreated NSCs (H3K4me1, H3K27ac, and H3K4me3). Promoter and VP enhancer are shaded in light orange and yellow respectively. **d** Schematic representation of the CRISPR/Cas9 experimental approach used to delete the *Chst8* putative enhancer in NSCs. Two gRNAs flanking the *Chst8* enhancer region were used to create the deletion (2.9 kb). Red arrows represent primers to test the deletion. PCRs using *Chst8* deletion and *G6pd2* pairs of primers are shown at the bottom of the figure in parental and  $\Delta$ *Chst8* enh NSC lines. Results are representative of three independent experiments. Source data are provided as a Source Data file. **e** Parental and  $\Delta$ *Chst8* enh cell lines were treated with TGFβ for 6 h. Total RNA was prepared and the levels of eRNA of the VP enhancer (left) or *Chst8* mRNA (right) were determined by qPCR. mRNA and eRNA levels of *Fabp4* were used as a control. Values were normalized to the *Gapdh* gene, and figure shows values relative to parental line. Data are presented as mean values  $\pm$  SEM. Results are representative of three biological independent experiments.  $^{***}p < 0.001$  (*P* values were calculated using two-tailed Student's *t* test,  $p = 2.2027E-05$  and  $p = 1.71676E-08$ ). Source data are provided as a Source Data file.

TGFβ as a driver of cis-regulatory region contacts (Fig. 2b and Supplementary Fig. 1b).

Interestingly, in addition to the contacts observed between the VP and the *Chst8* gene, we also identified several contacts occurring between the VP and inter- and intragenic enhancers located within the *Pepd* gene, a gene that lies 63 kb far from the VP, at its telomeric part (see regions under yellow in Fig. 2c). Surprisingly, the number and intensity of contacts between the VP and the enhancers located at its downstream region were higher than between the VP and the *Chst8* gene promoter. The gene *Pepd* is not regulated by TGFβ (Fig. 1b), but its intragenic VP-contacting enhancers are bound by the TGFβ transcription factor SMAD3 upon TGFβ treatment (Fig. 2c), pointing to a structural role of *Pepd* in the convergence of TGFβ-regulated enhancers that could potentially be cooperating to activate TGFβ-responsive gene promoters. These results indicate that TGFβ drives enhancer-enhancer contacts that lead to the assembly of an

enhancer cluster; we named this assembly *Chst8* enhancer cluster (EC).

To confirm that the identified contacting regions within the *Chst8* EC were TGFβ-responsive enhancers, we analyzed whether they became active upon TGFβ pathway induction. To this end, we evaluated the transcription of eRNAs by qPCR upon TGFβ addition. We named the different enhancers of the EC enhancer (E)1, E2, E3, E4 and VP (Fig. 2c, d). Results in Fig. 2d show that the tested regions transcribed eRNAs in response to TGFβ. To prove the TGFβ-dependency of the EC activation we tested the eRNAs transcription in cells lacking the TGFβ pathway effector SMAD3, that were previously characterized by our lab<sup>33</sup>. To this end, we measured the eRNA molecules transcribed from these enhancers upon TGFβ treatment in the control cells and in the SMAD3 depleted cells (shSMAD3). In concordance with the previous results, eRNAs were hardly induced in the shSMAD3 cells compared to the control cell line (Fig. 2d). Similarly, the



**Fig. 2 TGFβ drives enhancer-cluster assembly.** **a** UCSC Genome Browser 4C-seq profiles generated in NSCs before and upon TGFβ addition are shown at the *Chst8* promoter and gene body. The light orange box indicates enhancer-promoter contact. The yellow box indicates the VP enhancer (dark arrow). **b** Boxplot displaying the averaged values obtained from two biological independent replicates of RPM signals of the peaks located 500 kb around the VP - excluding the nearest  $\pm 20$  kb - (mm10 chr7:33841896-35860773) in NSCs untreated or treated for 3 h with TGFβ. An independent region located in another chromosome (mm10 chr4:33076383-35216108) was tested as a negative control. Boxes comprise values from Q1 to Q3 of the dataset; line corresponds to median value; whiskers show the data range (from min. to max. values within dataset). Depicted quantifications were performed for  $n = 2$  biologically independent samples.  $p$ -values are the result of a Wilcoxon-Mann-Whitney test. **c** UCSC Genome Browser captures showing 4C-seq profiles in NSCs untreated or treated (3 h) with TGFβ spanning a 200 kb distance around the VP enhancer (dark arrow). ChIP-seq signals of SMAD3 are shown. The positions of enhancers (defined in<sup>33</sup>) are also displayed. The light orange box indicates enhancer-promoter contacts; yellow boxes show enhancer-enhancer contacts. **d, e** The top panel shows a scheme summarizing the enhancer-enhancer and enhancer-promoter contacts identified in the 4C-seq experiment. The bottom panel shows the treatment of control NSCs or shSMAD3 NSCs for 3 h with TGFβ. **d** shows eRNA levels from the indicated enhancers and **e** shows mRNA from the indicated genes quantified by RT-qPCR. Transcription values were normalized to the housekeeping gene *Gapdh* and the figure shows values relative to the untreated samples. Progesterone-responsive *Fabp4* eRNA was used as a negative control. Results are the mean of three biologically independent experiments. Data are presented as mean values  $\pm$  SEM. \* $p < 0.05$ ; \*\* $p < 0.01$  ( $P$  values were calculated using two-tailed Student's  $t$  test,  $p = 0.03229859$  ( $E1$ ),  $p = 0.04082807$  ( $VP$ ),  $p = 0.00839842$  ( $E2$ ),  $p = 0.02618772$  ( $E3$ ),  $p = 0.01592669$  ( $E4$ ) and  $p = 0.00113168$  ( $Chst8$ )). Source data are provided as a Source Data file.

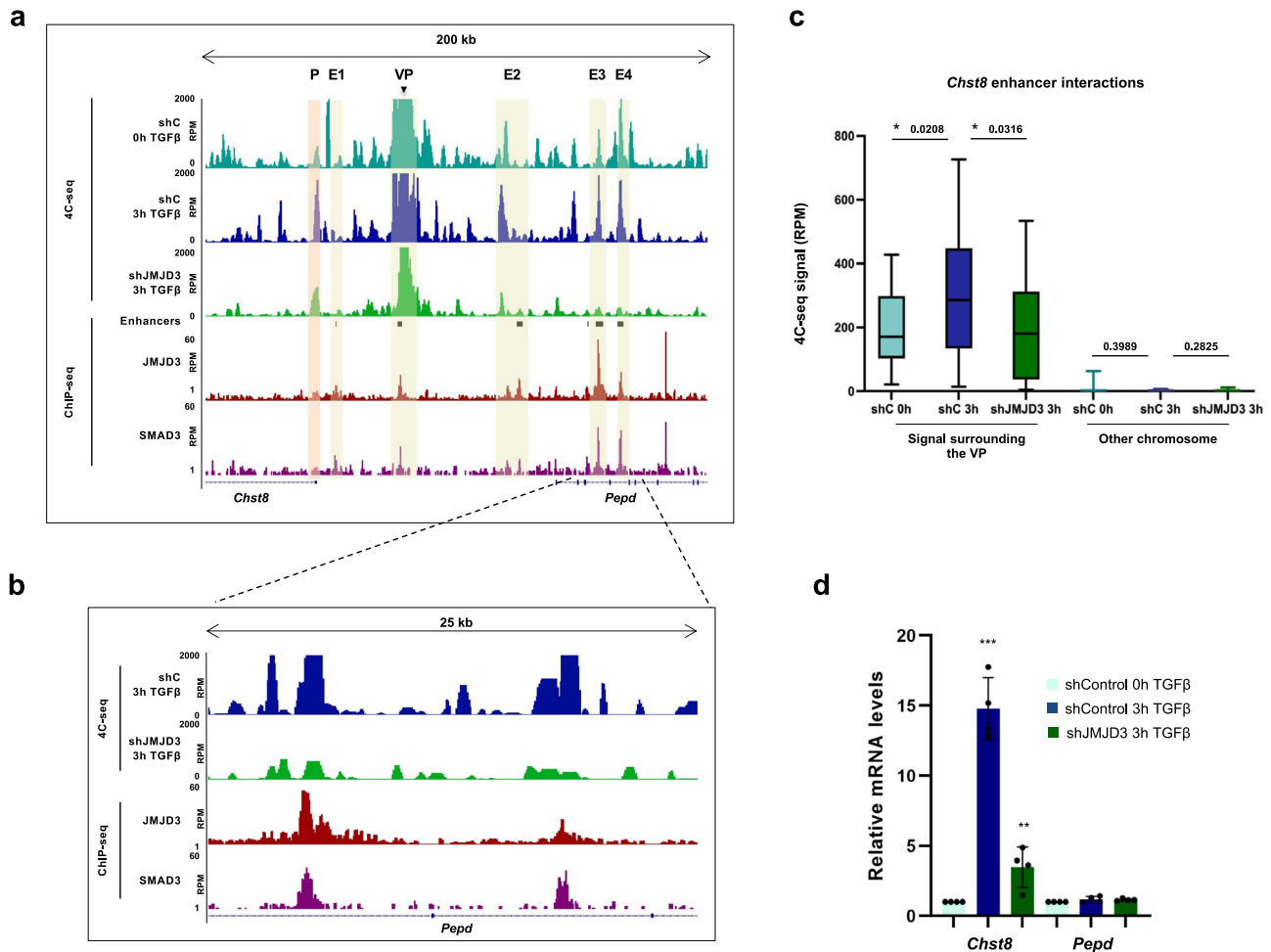
*Chst8* gene was not expressed upon TGFβ addition in the shSMAD3 NSCs (Fig. 2e).

Altogether, these results demonstrate that TGFβ-driven gene activation entails a reorganization of the chromatin structure. Moreover, this reorganization results in the formation of the *Chst8* EC that is potentially involved in the regulation of genes upon TGFβ.

#### TGFβ-mediated enhancer-cluster assembly depends on JMJD3.

Previous work from our lab has demonstrated that the histone KDM JMJD3 functions as a cofactor for SMAD3 in the TGFβ-driven activation of neuronal enhancers in NSCs<sup>33</sup>. For this reason, we decided to test whether JMJD3 was also occupying the enhancers involved in the *Chst8* EC by analyzing our previously

published JMJD3 ChIP-seq performed upon TGFβ stimulation<sup>23</sup>. As shown in Fig. 3a, all the contacting regions belonging to the *Chst8* EC (VP, E1, E2, E3 and E4) are occupied by JMJD3, consistent with the presence of enhancers at these regions. Next, motivated by the fact that the demethylase catalytic activity of JMJD3 is not involved in enhancer activation in our model<sup>33</sup>, we decided to address whether JMJD3 could be playing a structural role at enhancers, contributing to the *Chst8* EC assembly. For this purpose, we efficiently depleted JMJD3 from NSCs (shJMJD3 NSCs) using lentivirus containing JMJD3-specific shRNAs (Supplementary Fig. 2a and refs. 23,33), and then, we performed a 4C-seq assay upon TGFβ treatment using the *Chst8* VP enhancer. Figures 3a–c, and Supplementary Fig. 2b, c show the striking effect that the depletion of JMJD3 causes in the 3D-structure of



**Fig. 3 TGFβ-mediated enhancer-cluster assembly depends on JMJD3.** **a** UCSC Genome Browser captures show 4C-seq profiles spanning 200 kb around the VP enhancer (black arrow) in NSCs untreated or treated (3 h) with TGFβ. ChIP-seq signals of SMAD3 and JMJD3 upon TGFβ stimulation (0.5 and 3 h, respectively) are shown. The location of the members of the EC is also indicated with yellow boxes. **b** Capture showing a zoom into a region where TGFβ-induced contacts are lost in JMJD3-depleted (shJMJD3) NSCs. **c** Boxplot displays the averaged values obtained from two biological independent replicates of RPM signals of the peaks located 500 kb around the VP - excluding the nearest ±20 kb - (mm10 chr7:33841896-35860773) in control or shJMJD3 NSCs untreated or treated with TGFβ during 3 h. An independent region located in another chromosome (mm10 chr4:33076383-35216108) was used as a negative control.  $n = 2$  biologically independent replicates were quantified.  $p$ -values are the result of a Wilcoxon-Mann-Whitney test. **d** Control NSCs or shJMJD3 NSCs were treated for 3 h with TGFβ. Then, total RNA was prepared, and the levels of the mRNA of the indicated genes were determined by qPCR. Values were normalized to the housekeeping gene *Gapdh*. The figure shows values relative to time 0 h. Results are the mean of three biologically independent experiments. Data are presented as mean values  $\pm$  SEM.  $***p < 0.001$ ;  $**p < 0.01$  ( $P$  values were calculated using two-tailed Student's  $t$  test,  $p = 0.001131677$  and  $p = 0.003848794$ ). Source data are provided as a Source Data file.

the chromatin. Upon JMJD3 removal, we observed that the genomic contacts between the VP and the surrounding regions were severely reduced (Fig. 3c and Supplementary Fig. 2c). In particular, the contacts between enhancers belonging to the *Chst8* EC region were abolished upon JMJD3 depletion (Fig. 3a, b); indicating that JMJD3 is directly or indirectly required for *Chst8* EC assembly. Accordingly, *Chst8*, gene expression was markedly reduced (Fig. 3d).

To broaden our conclusions and to confirm that TGFβ-driven gene activation encompasses a chromatin structure reorganization that depends on JMJD3, we analyzed the 3D-chromatin status of other two candidate enhancers that potentially regulate the TGFβ-responsive and JMJD3-dependent genes, *Ldlrad4* and *Aopep*<sup>23</sup>. These enhancers are located within the *Ldlrad4* and *Aopep* genes, display SMAD3 and JMJD3 binding, and are surrounded by other SMAD3/JMJD3-bound enhancers, suggesting that they could potentially engage in enhancer clusters. Indeed, 4C-seq assays using these cis-regulatory regions as VPs

(see quality of the experiments assessed as described in<sup>35</sup>, Supplementary Data 2 and 3) show contacts between the *Ldlrad4* VP and the *Ldlrad4* promoter (Supplementary Fig. 3a, shaded in orange), and between the VPs and the surrounding enhancers (Supplementary Fig. 3a, c, shaded in yellow). In agreement with our hypothesis, upon TGFβ treatment, the frequency of the contacts between cis-regulatory elements —enhancers or promoters— increased [Supplementary Fig. 3a, b (*Ldlrad4*), c, d (*Aopep*)]. Furthermore, these contacts remarkably diminished when JMJD3 was depleted (Supplementary Fig. 3a–d). Altogether these results corroborate that the TGFβ pathway and JMJD3 are involved in 3D-chromatin structure regulation.

As JMJD3 is a coactivator<sup>23,33,36</sup>, we decided to rule out the possibility of an indirect transcriptional effect triggered by the lack of JMJD3 in NSCs, that could potentially be affecting the expression of the proteins involved in loop formation<sup>37,38</sup>. To this end, we analyzed gene expression data from our previously published microarray experiments<sup>23</sup>, and we show in the

Supplementary Fig. 4a that neither TGF $\beta$  nor JMJD3 regulate the expression of some of the most characteristic proteins involved in loop formation (CTCF, SMC1/3, RAD21, PDS5A/B, and WAPL). Among these proteins, the Cohesin complex has been described to play a role in the establishment of dynamic contacts during gene transcription by being the main motor of the loop extrusion process<sup>39,40</sup>, thus, we analyzed the presence of the SMC1 subunit of the Cohesin complex in the *Chst8* EC using previously published SMC1 ChIP-seq data in NSCs<sup>41</sup>. Interestingly, we noticed that some of the regions occupied by JMJD3 were also bound by SMC1 (Supplementary Fig. 4b). This observation prompted us to check whether JMJD3 co-occupies genomic regions with SMC1 in a genome-wide manner. The Venn diagram in Supplementary Fig. 4c shows that these proteins do not colocalize widely across the genome (only 19% of JMJD3 peaks overlap SMC1); this can be explained by the involvement of JMJD3 in transcription elongation, that leads to many JMJD3-bound regions falling outside the cis-regulatory elements that engage in 3D-interactions<sup>36,42</sup>. Nonetheless, at the co-bound regions, JMJD3 and SMC1 demonstrate widespread peak overlapping (Supplementary Fig. 4d).

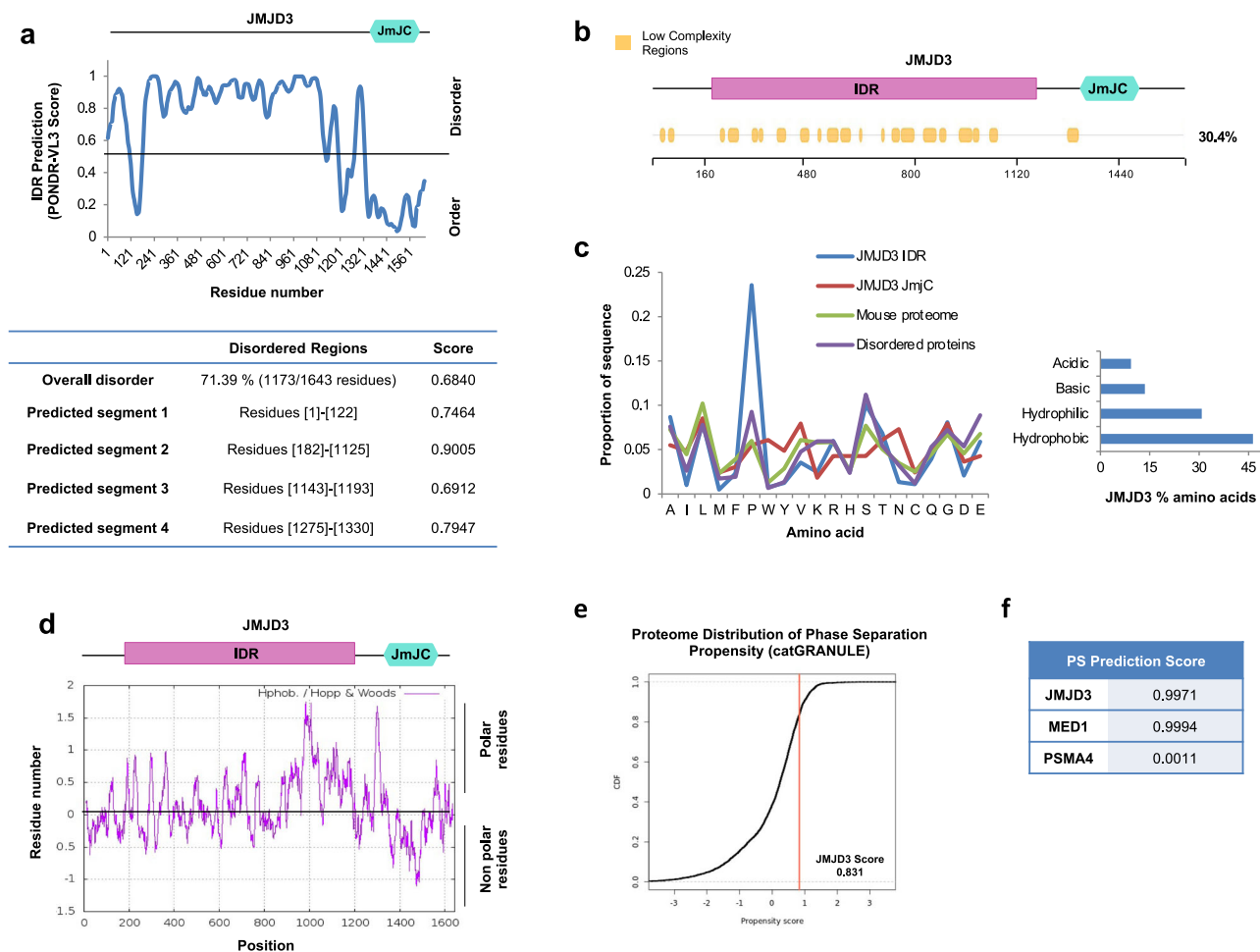
The results described above indicate that JMJD3 is necessary for the establishment and/or maintenance of contacts between cis-regulatory regions. JMJD3 lacks DNA binding capacity, and structurally it only has one known domain, the demethylase catalytic domain JmjC. Previous work from our laboratory has shown that its demethylase domain is not required for the activation of a subset of enhancers in response to TGF $\beta$ <sup>33</sup>. To deeply understand whether demethylation of the H3K27me3 is involved in the *Chst8* enhancer cluster activation, we analyzed H3K27me3 ChIP-seq data from NSCs<sup>36</sup>. The results in Supplementary Fig. 5a show that the *Chst8* locus lacks H3K27me3 prior to TGF $\beta$  stimulation. Moreover, ChIP-qPCR experiments show that JMJD3 depletion did not lead to an increase in H3K27me3 levels neither at the *Chst8* promoter nor at the scrutinized enhancers (Supplementary Fig. 5b). Of note, Supplementary Fig. 5b corroborates that the levels of H3K27me3 in these regions are negligible when compared to a classic H3K27me3-controlled promoter (*Hoxd8*), suggesting that changes on the H3K27me3 levels are unlikely a force driving the *Chst8* EC formation or activation.

**JMJD3 is a highly disordered protein.** As the investigated genomic loci are not marked by H3K27me3 prior to TGF $\beta$  activation, we hypothesized that JMJD3 could be impacting the 3D-structure of the chromatin through its unstructured domain. In the last years, numerous works have shed light on the impact on transcriptional regulation of these unstructured regions, named intrinsically disordered regions (IDR)<sup>43,44</sup>. With this in mind, we questioned whether JMJD3 with its unstructured domain could belong to the group of the intrinsically disordered proteins. To assess this, we analyzed the amino acid sequence of JMJD3 searching for disordered regions using the following previously validated algorithms: PONDR-VL3<sup>45</sup>, IUPred<sup>46</sup> and VSL2<sup>47</sup> (see methods). Overall, the three algorithms agreed on the highly significant disorder score of JMJD3 (Fig. 4a and Supplementary Fig. 6a, b). Specifically, PONDR-VL3 showed a median disorder score of 0.68 for JMJD3 (Fig. 4a), a value considerably higher than the 0.28 obtained when analyzing PSMA4, a well-known structured protein used as an ordered protein control (Supplementary Fig. 6b). In addition, more than 70% of JMJD3 amino acids (71.39% using PONDR-VL3) exist in disordered domains (Fig. 4a and Supplementary Fig. 6a). Again, this value is higher than the proteasome component PSMA4, used as a negative control (23.37%) (Supplementary Fig. 6b). Looking at

the different disordered fragments of JMJD3, we observed a remarkably long region with no defined structure that contains 943 amino acids (residues from 182 to 1125) and that was predicted to have a disorder score of 0.90, the highest observed in our data (Fig. 4a). From now on, we will refer to this region as JMJD3 IDR.

In addition to disorder, the nature of the amino acid composition has also been shown to play an important role in IDR-mediated transcription regulation. Moreover, disordered regions frequently coincide with low-complexity domains that are biased for certain amino acids<sup>10,13,14,18</sup>. To check whether this was the case for JMJD3 we used the SEG algorithm<sup>48</sup> (see methods) looking for JMJD3 complexity prediction. As shown in Fig. 4b, 30.4% of JMJD3 was predicted to contain low-complexity segments. Furthermore, by examining its amino acid composition, we found a remarkable abundance of prolines (24% of the total amino acids in the protein) (Fig. 4c and Supplementary Fig. 6c), that displays widespread conservation among mammals (Supplementary Fig. 6d). Strikingly, we found proline tracks as long as 20 residues (Supplementary Fig. 6c). Other structural features of the JMJD3 IDR are its high content of charged residues (21% of the protein) (Fig. 4c, d, and Supplementary Fig. 6c), its regions different to the proline tracks that also display high hydrophobicity (Fig. 4d), and its high serine content when compared to the average in the mouse proteome, and similar to other described IDRs (Fig. 4c)<sup>15,49</sup>. Proline residues have been described as highly hydrophobic amino acids whose concatenation generates sticky domains that bind rapidly and reversibly to other proteins<sup>50</sup>. It is known that hydrophobic interactions, as well as electrostatic ones, are relevant for biomolecular condensate formation. Thus, the amino acid composition of JMJD3 seemed to favor its potential to be involved in the so-called phase separation process<sup>9,51</sup>. Given this, we used catGRANULE<sup>52</sup> and PSPredictor<sup>53</sup> algorithms (see methods) to predict JMJD3 phase separation ability. Both tools returned high scores (0.83 and 0.99 respectively) for JMJD3 (Fig. 4e, f), similar to proteins known to be involved in phase separation (e.g. MED1, 0.99) and higher than the proteasome protein PSMA4 (0.001) (Fig. 4f). Altogether, these data point to JMJD3 as a highly disordered protein with the potential to undergo phase separation.

**JMJD3 undergoes LLPS in vitro and in vivo.** LLPS is a physicochemical process that consists on the demixing of a fluid into a diluted phase and a dense phase. It is well known that proteins mediating phase separation contain IDRs, and it is starting to be uncovered the role that these IDRs play on transcription regulation as mediators of biomolecular condensation<sup>12–15,17,54–56</sup>. On the grounds of these recent discoveries, we hypothesized that JMJD3 could be contributing to the establishment of 3D-contacts by nucleating protein and nucleic acid scaffolds to form membrane-less condensates through LLPS. To test this idea, we decided to perform in vitro droplet assays using a construct that expressed JMJD3 fused to monomeric EGFP (mEGFP) and HA (Supplementary Fig. 7a). First, we tested that the resulting fluorescence protein had the predicted molecular weight and was well recognized by JMJD3 antibody when ectopically expressed (Supplementary Fig. 7a, b). Next, we expressed mEGFP-JMJD3 in HEK293T cells and performed an in vitro droplets assay using nuclear extracts. Our data shows that the mEGFP-JMJD3 protein forms droplets that do not appear when we overexpress mEGFP alone, reflecting that the droplets can be attributed to JMJD3, and not to the mEGFP tag (Fig. 5a). The droplets showed features [circularity, convexity and aspect ratio] that are characteristic of a liquid-like nature (Fig. 5a, bottom panels). Interestingly, some of



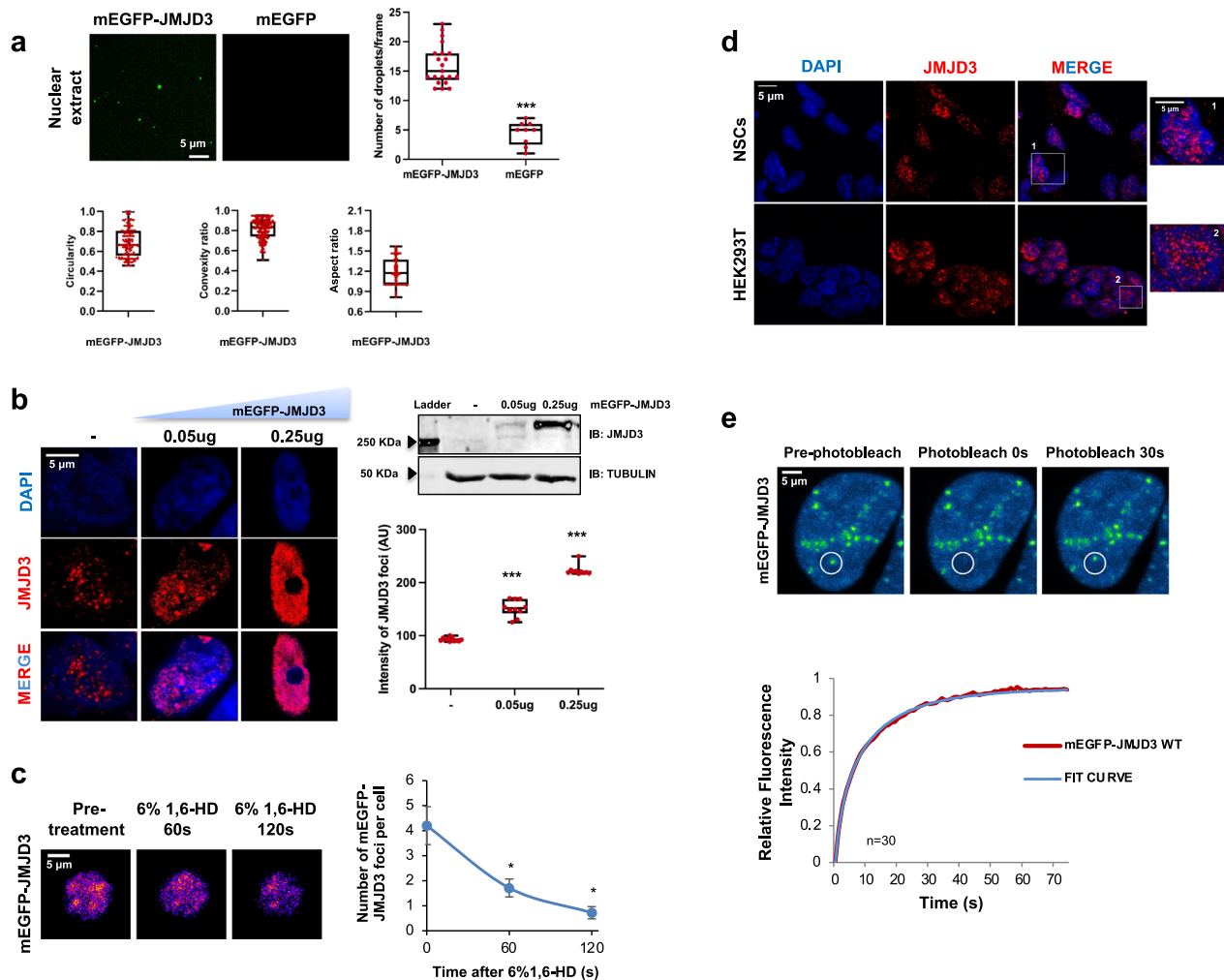
**Fig. 4 JMJD3 is a highly disordered protein.** **a** Disorder prediction of human JMJD3 using PONDRL-VL3 algorithm. In the bottom panel, the disorder score and the lengths of the predicted disordered regions are indicated (length of disordered segments >50 amino acids). A schematic representation of JMJD3 described domains is shown on top of the graphic. **b** Analysis of the presence of low-complexity domains in JMJD3 using the SEG algorithm. The percentage of low complexity regions is indicated on the right side. Low complexity regions are depicted in yellow. A schematic representation of JMJD3 described domains is shown on top of the panel. **c** Amino acid composition of JMJD3 IDR, JMJD3 catalytic domain, mouse proteome and disordered proteins defined by the presence of a 50 residues fragment whose IUPRED median score is at least 0.55 and that is not found in Pfam (so that functional domains are avoided). The percentages of acid, basic, hydrophobic, and hydrophilic amino acids of JMJD3 are shown on the right panel. **d** JMJD3 hydrophobicity profile was determined using the ExPASy website with the Hopp and Woods scale and a sliding window of 21. **e, f** The potential of JMJD3 to phase separate was determined using catGRANULE (on the left) (**e**) and PSpredictor (on the right) (**f**) algorithms.

these in vitro droplets were shared with MED15, a well-known component of enhancers that forms nuclear condensates (Supplementary Fig. 7c, d). When we overexpressed mEGFP-JMJD3, mCherry-MED15 or both we observed that 57% of JMJD3 droplets colocalized with MED15 (Supplementary Fig. 7d).

In fixed cells, overexpression of mEGFP-JMJD3 formed nuclear puncta (Fig. 5b), and the intensity of these puncta increased along with the amount of JMJD3 inside the cell (Fig. 5b); in fact, at 0.25  $\mu$ g of plasmid overexpressed, we started to observe aggregates (see below). Next, we tested the sensitivity of these condensates to the aliphatic alcohol 1,6-hexanediol, a chemical compound that has been demonstrated to disrupt the hydrophobic interactions that sustain the phase-separated droplets<sup>57</sup>. We observed that the treatment led to a reduction in the number and size of JMJD3 puncta (Fig. 5c). Importantly, endogenous JMJD3 also formed nuclear puncta as detected by immunofluorescence using two different antibodies against JMJD3 in NSCs and HEK293T cells (Fig. 5b, without transfection, d and Supplementary Fig. 8a), ruling out the possibility that the observed puncta could be an overexpression artifact. These data

suggest that JMJD3 condensates occur at endogenous levels and that these condensates represent a separated phase inside the cell.

Liquid-like condensates have been suggested to exhibit a remarkable dynamic nature, and their internal molecules have been described to diffuse rapidly<sup>12,58</sup>. Based on this, we sought to analyze whether the JMJD3 puncta exhibited liquid-like properties by analyzing the rate of fluorescence recovery after photobleaching (FRAP) of the overexpressed mEGFP-JMJD3<sup>9,59</sup>. After photobleaching mEGFP-JMJD3 puncta recovered fluorescence almost completely on a time scale of seconds (Fig. 5e, Supplementary Fig. 8b and Supplementary Movie 1), in agreement with what is observed for other proteins that form either liquid-like condensates (BRD4 and MED1) or membraneless organelles<sup>60</sup>. We also calculated the mobile fraction (the molecular pool that undergo exchange within the FRAP zone), which corresponds to 0.95 for this protein. Moreover, the aggregates that appear when high levels of proteins are overexpressed (mentioned above) showed reduced mobility in FRAP assays (Supplementary Movie 2). These data suggest that JMJD3 droplets exhibit liquid-like properties and that its



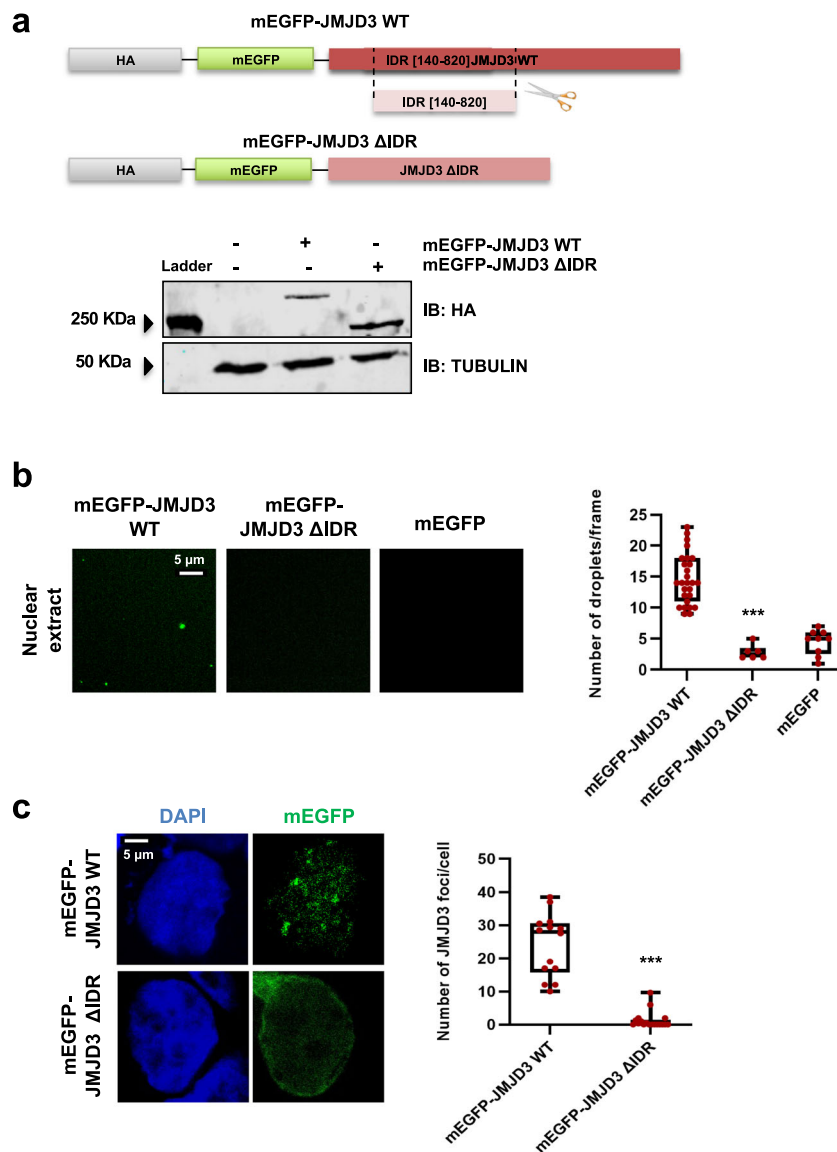
**Fig. 5 JMJD3 undergoes LLPS in vitro and in vivo.** **a** mEGFP and mEGFP-JMJD3 proteins were analyzed using droplet-formation assays in nuclear extracts at room temperature with 150 mM NaCl. Quantifications of the number of droplets per frame, circularity, convexity and aspect ratio (AR) are displayed. Data are the mean  $\pm$  SEM. Boxes comprise values from Q1 to Q3 of the dataset; line corresponds to median value; whiskers show the data range (from min. to max. values within dataset). \*\*\* $p < 0.001$  (Student's  $t$  test,  $p = 1.09982E-06$ ). Droplets in 5 fields in each group from three biologically independent experiments were quantified,  $n = 150$ . Scale bar, 5  $\mu\text{m}$ . **b** Confocal microscopy images of HEK293T cells transfected with mEGFP-JMJD3. Quantifications of the intensity of JMJD3 puncta are shown on the right. Data show the mean  $\pm$  SEM. Boxes comprise values from Q1 to Q3 of the dataset; line corresponds to median value; whiskers show the data range. \*\*\* $p < 0.001$  (Student's  $t$  test,  $p = 9.98785E-06$  and  $p = 7.15724E-17$ ).  $n = 50$  transfected cells in each group were quantified; Images are representative of 3 biologically independent experiments. Scale bar, 5  $\mu\text{m}$ . Western blot displays the levels of overexpressed JMJD3. **c** HEK293T cells were transfected with 0.05  $\mu\text{g}$  mEGFP-JMJD3, treated with 6% 1,6-HD for 5 min and imaged at 60 and 120 s. Nuclei were visualized with DAPI (blue). Quantification of the nuclear puncta per cell is shown on the right. Data are the mean  $\pm$  SEM. \* $p < 0.05$  (Student's  $t$  test,  $p = 0.0182428$  and  $p = 0.01162199$ ).  $n = 130$  transfected cells were quantified; Images are representative of three biologically independent experiments. Scale bar, 5  $\mu\text{m}$ . **d** NSCs and HEK293T cells were fixed, and endogenous JMJD3 was visualized by immunostaining assay. The images are representative of three biologically independent experiments. Scale bar, 5  $\mu\text{m}$ . **e** FRAP assay in HEK293T cells expressing 0.05  $\mu\text{g}$  of mEGFP-JMJD3. Images are representative of three biological replicates. Quantification shows the curve fit results of FRAP data for mEGFP-JMJD3 to a double-exponential smoothing ( $R^2 = 1$ ), where bleaching events occurs at  $t = 0$  s. Data are plot as background-subtracted and normalized mean ( $n = 27$  cells). Scale bar, 5  $\mu\text{m}$ . Source data are provided as a Source Data file.

conforming molecules exchange rapidly between the condensates and the surrounding.

**JMJD3 IDR is necessary for condensate formation.** As the condensation of molecules into liquid-like droplets has been related to the IDRs present in the conforming proteins<sup>61</sup>, we chose to investigate whether the proline-rich IDR of JMJD3 is necessary for being part of biological condensates. To do this, we deleted the proline-rich IDR domain (amino acids 140–820) from our previously characterized mEGFP-JMJD3 plasmid, and we named this new construct mEGFP-JMJD3  $\Delta$ IDR (Fig. 6a). We

ectopically expressed this protein (Fig. 6a) and analyzed its ability to form droplets in nuclear extracts. The results in Fig. 6b demonstrate that the mEGFP-JMJD3  $\Delta$ IDR protein did not form droplets in vitro. We next investigated the competence of mEGFP-JMJD3  $\Delta$ IDR to form puncta in fixed cells. These two versions of the protein were distributed between nucleus and cytoplasm, even though the JMJD3  $\Delta$ IDR shows some bias for the cytoplasm. Our immunofluorescence experiments revealed that the JMJD3 mutant was unable to form puncta (Fig. 6c). These data support that the IDR domain and probably the prolines are necessary for JMJD3 phase separation. Even though we acknowledge that identifying the precise amino acids involved in





**Fig. 6 JMJD3 IDR is essential for condensate formation.** **a** mEGFP-JMJD3 and mEGFP-JMJD3  $\Delta$ IDR expression vectors were transfected into HEK293T (0.05ug). 24 h later total protein extracts were prepared and the JMJD3 (HA) and TUBULIN levels were determined by immunoblot. The image shown is representative of two independent experiments. Source data are provided as a Source Data file. **b** mEGFP, mEGFP-JMJD3 and mEGFP-JMJD3  $\Delta$ IDR proteins were analyzed using droplet-formation assays in nuclear extracts at room temperature in the presence of 150 mM NaCl. Quantifications of the droplets are displayed on the right. Data are the mean  $\pm$  SEM. Boxes comprise values from Q1 to Q3 of the dataset; line corresponds to median value; whiskers show the data range (from min. to max. values within dataset). \*\*\* $p < 0.001$  ( $P$  values were calculated using one-tailed Student's  $t$  test,  $p = 0.000113231$ ). Droplets in 5 fields in each group from three biologically independent experiments were quantified. Scale bar, 5  $\mu$ m. Source data are provided as a Source Data file. **c** Confocal microscopy images of HEK293T cells transfected with 0.05 ug mEGFP-JMJD3 or mEGFP-JMJD3  $\Delta$ IDR. The images are representative of three biologically independent experiments. Quantifications of the number of JMJD3 puncta are shown on the right. Data show the mean  $\pm$  SEM. Boxes comprise values from Q1 to Q3 of the dataset; line corresponds to median value; whiskers show the data range (from min. to max. values within dataset). \*\*\* $p < 0.001$  ( $P$  values were calculated using one-tailed Student's  $t$  test,  $p = 2.5178E-09$ ).  $n = 20$  transfected cells in each group were quantified. Scale bar, 5  $\mu$ m. Source data are provided as a Source Data file.

the condensation would benefit our work, both the length and the complexity of the IDR domain prevented this analysis. Nonetheless, our data suggest that not only one type of amino acid but also several contribute to JMJD3 phase separation.

Once known the relevance of the IDR domain, we analyzed if the catalytic activity of JMJD3 plays a role in the formation of condensates. To do that, we used a plasmid encoding JMJD3 mutated at the catalytic domain (JMJD3 HE > AA) (Supplementary Fig. 9a). This mutant lacks the capability of demethylating the H3K27me3 mark and functions as a dominant negative form of JMJD3<sup>23</sup>. Experiments in fixed cells revealed that

the JMJD3 HE > AA mutant formed nuclear puncta of the same intensity and volume than those of the wild type version of JMJD3, suggesting that the catalytic activity is not essential for condensate formation (Supplementary Fig. 9a, c). To further prove that the catalytic activity of JMJD3 is not required for condensate formation, we employed a specific inhibitor of the JMJD3 catalytic activity named GSK-J4<sup>62</sup>. We treated the cells for 6 h, a period of time that was enough to effectively inhibit JMJD3 (Supplementary Fig. 9b). Using the GSK-J4 inhibitor, we observed puncta formation of the same intensity and volume than those of the non-treated cells (Supplementary Fig. 9b, c). Moreover, the

overexpression of JMJD3 HE>AA led to the same *Chst8* transcriptional activation as the overexpression of JMJD3 WT in HEK293T cells, where TGF $\beta$  is active (Supplementary Fig. 9d). Altogether our data illustrate that the catalytic activity of JMJD3 is not required for either condensate formation or *Chst8* full transcriptional activation.

**JMJD3 promotes gene transcription and enhancer-cluster assembly.** As our data indicates that JMJD3 can form nuclear condensates, which have been widely related to transcription<sup>8,63–65</sup>, we hypothesized that this ability could be contributing to the JMJD3-mediated transcriptional activity on the *Chst8* locus. To test this idea, we co-imaged the *Chst8* locus and JMJD3 foci by performing an immuno-FISH experiment. A clear colocalization of JMJD3 condensates with the *Chst8* locus was detected (Fig. 7a). We also observed that JMJD3 nuclear condensates are excluded from regions marked by the repressive mark H3K9me3 (Fig. 7b), conversely, active sites of transcription marked by MED15 colocalize with JMJD3 condensates (Fig. 7b).

Finally, to conclude that JMJD3-mediated contacts indeed contribute to *Chst8* transcription, we stably integrated in the shJMJD3 NSCs a construct that, upon doxycycline treatment, overexpresses JMJD3 (Fig. 7c top and methods). After 24 h of induction, these cells displayed JMJD3 expression levels similar to the endogenous levels (Fig. 7c, left). In these conditions, we measured the expression of *Chst8* and we performed 4C-seq experiments to analyze the genomic contacts in response to TGF $\beta$ . The right panel on Fig. 7c shows a full rescue of *Chst8* transcription upon induction of JMJD3 in the shJMJD3 NSCs without affecting *Fabp4*, a negative control. As hypothesized, after reintroduction of JMJD3, the genomic contacts were recovered (Fig. 7d, e), in particular, contacts between the *Chst8* EC were efficiently restored (Fig. 7d).

Altogether, these results support a model in which JMJD3 facilitates the assembly of the described *Chst8* EC, probably through the formation of IDR-driven phase-separated condensates enriched in key factors that enable gene expression.

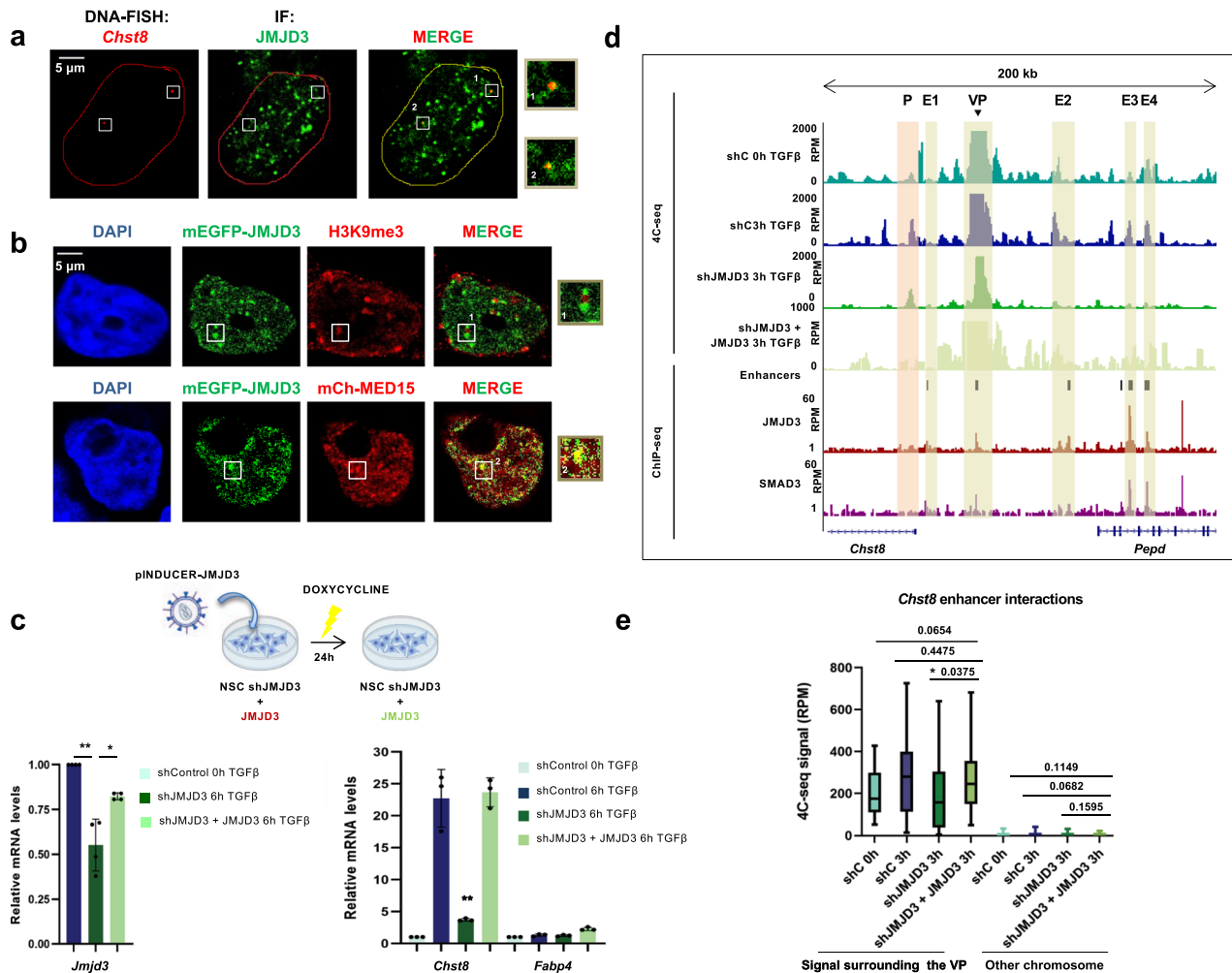
## Discussion

In this work, we provide a molecular description of an enhancer cluster formation in response to the TGF $\beta$  signaling pathway during neurogenesis. Our data uncover an unforeseen role of TGF $\beta$  reorganizing the chromatin fiber in a JMJD3 histone demethylase-dependent manner. JMJD3 promotes the establishment of enhancer–enhancer and enhancer–promoter contacts that ultimately modulate *Chst8* enhancer activity, and thus the NSCs gene expression program.

Mammalian promoters are normally surrounded and regulated by multiple enhancers. Enhancer–enhancer contacts have been described in the literature (for review<sup>66</sup>) and demonstrated by chromatin conformation capture techniques, and cell imaging<sup>67,68</sup>. Nonetheless, the contribution of individual enhancers to enhancer clusters and their impact on target gene regulation is still an open debate. Our results show that upon TGF $\beta$  treatment, the *Chst8* locus is reorganized in the 3D space (Fig. 2a, b), bringing into proximity cis-regulatory regions to facilitate an accurate *Chst8* transcriptional response (Fig. 2c, e). Interestingly, an allelic variant of a *Chst8* exon has been involved in a peeling skin syndrome (OMIM #616265), highlighting the importance of a controlled *Chst8* gene response<sup>69</sup>. Even though we have focused on the *Chst8* locus, we have also demonstrated by 4C-seq assays, that TGF $\beta$  reorganizes other loci entailing TGF $\beta$ -responsive genes in a similar manner (Supplementary Fig. 3). Altogether, these results highlight the essential contribution of the TGF $\beta$  pathway as a major force driving the 3D organization of the chromatin.

In this work, we have provided an answer to an open question from our previous work<sup>33</sup>: why is JMJD3 required at enhancers that are not marked by H3K27me3? Here, we reveal a novel function for JMJD3 mediating the TGF $\beta$ -driven enhancer–promoter and enhancer–enhancer contacts. This role agrees with previous literature that elegantly demonstrated that JMJD3 facilitates enhancer–promoter looping during endoderm differentiation<sup>70</sup>. How can JMJD3 facilitate TGF $\beta$ -driven transcriptional response through 3D-chromatin organization? Past work in our laboratory demonstrated that JMJD3 is required for TGF $\beta$ -driven activation at different levels: promoter, enhancer, and gene body (see below); additionally, some functions were demonstrated to be dependent on the demethylase catalytic activity, whereas others were not<sup>23,33</sup>. Here, we provide a further molecular explanation for this contribution, unveiling the previously unknown JMJD3 IDR as a crucial protein region that enables the formation of biomolecular condensates (Fig. 6). Interestingly, recent data demonstrated that UTX, another member of the same family of KDMs, undergoes LLPS driven by an IDR domain. In agreement with our work, Shi et al. have demonstrated that the IDR facilitates higher-order chromatin interactions and mediates tumor suppression in a catalytic independent manner<sup>71</sup>. With our study, we propose that JMJD3 phase separates in NSCs by establishing multivalent interactions through its IDR and that this condensation might be an important mechanism to enable a precise TGF $\beta$ -response. Indeed, recent studies have suggested a model involving condensate formation to explain both the initiation and the elongation transcriptional stages<sup>8,63–65</sup>. Following this model, initiation requires the condensate assembly of Mediator, transcription factors, coactivators and non-phosphorylated RNAPII. The second step consists in an elongation condensate arrangement that includes phosphorylated RNAPII, RNA processing and elongating factors, and RNA itself. Once at the end of the gene, hypophosphorylated CTD is released from the condensates so that it can be re-incorporated into the initiation condensates<sup>63,72</sup>. Having this model in mind, we speculate that JMJD3 could be participating in both the initiation and the elongation condensates. JMJD3 is found at promoters and enhancers in response to TGF $\beta$  where it interacts with SMAD3. This transcription factor also undergoes LLPS through its IDR<sup>73</sup>. Previous data from the laboratory<sup>23</sup> demonstrated that SMAD3 interacts with JMJD3 through its linker region, which contains an IDR, suggesting that SMAD3 and JMJD3 IDRs may be engaging in multivalent interactions. Interestingly, the contacting genomic regions in the *Chst8* ECs were bound by JMJD3 and SMAD3, suggesting that both factors may contribute to the formation of the initiation condensates required for transcription activation. Moreover, it has been suggested that phase separated condensates are formed at EC<sup>10,11,74,75</sup>. In agreement, our data indicates that JMJD3 contributes to condensates that take place at EC potentially together with the enhancer machinery (e.g. general factors such as Mediator, RNAPII, or TGF $\beta$  pathway-specific factors such as SMAD3 or CHD8). On the other hand, our lab and others have previously demonstrated that JMJD3 interacts with the elongating form of the RNAPII<sup>36</sup> and with elongation factors<sup>42</sup> and that it is essential for the elongation stage<sup>36,42</sup>. Thus, we speculate that JMJD3 could favor elongation by promoting or forming part of the elongating condensates.

Based on our results, we hypothesize that JMJD3 condensates could facilitate transcription in different ways. JMJD3 might concentrate SMAD3, CHD8, and MED15, and the general elongation factors in a compartment to make transcription kinetically more efficient as it has been proposed for other factors<sup>11,61,63</sup>. Alternatively, it might physically insulate TGF $\beta$ -responsive transcriptional machinery from its regulators to prevent inactivation (e.g., insulating the active phospho-SMAD3 from phosphatases). Finally,



**Fig. 7 JMJD3 promotes *Chst8* gene transcription and enhancer-cluster assembly.** **a** Immuno-FISH for JMJD3 protein (green) and *Chst8* locus (red) on HEK293T cells. *Chst8* FISH signal colocalizes with JMJD3 condensates (on the right). Results are representative of three independent experiments. **b** In HEK293T cells expressing mEGFP-JMJD3 (transfection of 0.05  $\mu$ g, green) the localization of the H3K9me3 mark and MED15 was analyzed using immunofluorescence staining with an anti-H3K9me3 antibody (red). MED15 localization was analyzed following the red signal in cells co-expressing mEGFP-JMJD3 and mCherry-MED15. Nuclei were visualized with DAPI (blue). Colocalizations are shown in yellow. Scale bar, 5  $\mu$ m. The images are representative of three independent experiments with similar results. **c** JMJD3-depleted NSCs (shJMJD3) expressing JMJD3 (shJMJD3 + JMJD3) were treated with TGF $\beta$  for 6 h. Total RNA was prepared and the mRNA expression levels of the *Chst8* gene were determined by qPCR. mRNA levels of the *Fabp4* gene were used as a negative control. Values were normalized to the *Gapdh* gene. Data are presented as mean values  $\pm$  SEM. \* $p$  < 0.05, \*\* $p$  < 0.01 ( $P$  values were calculated using two-tailed Student's  $t$  test,  $p$  = 0.00850428 and  $p$  = 0.03608055 (*Jmjd3*);  $p$  = 0.00971346 (*Chst8*)). Results are representative of four independent experiments. Source data are provided as a Source Data file. **d** UCSC Genome Browser captures show 4C-seq profiles spanning 200 kb around the VP *Chst8* enhancer (black arrow) in control, shJMJD3 and shJMJD3 + JMJD3 NSCs treated with TGF $\beta$  during 3 h. ChIP-seq signals of SMAD3 and JMJD3 upon TGF $\beta$  stimulation are shown. The location of the members of the EC is also indicated with yellow boxes. **e** Boxplot displays the averaged values obtained from two biological independent replicates of RPM signals of the peaks located 500 kb around the VP enhancer - excluding the nearest  $\pm$ 20 kb - (mm10 chr7:33841896-35860773) in control, shJMJD3 and shJMJD3 + JMJD3 NSCs treated with TGF $\beta$  during 3 h. An independent region located in another chromosome (mm10 chr4:33076383-35216108) was used as a negative control. Boxes comprise values from Q1 to Q3 of the dataset; line corresponds to median value; whiskers show the data range (from min. to max. values within dataset).  $p$ -values are the result of a Wilcoxon-Mann-Whitney test.

it is conceivable that the JMJD3-mediated chromatin 3D organization could be a critical determinant that allows biomolecular condensate formation at active genome loci. The enhancer-enhancer and enhancer-promoter contacts that contain large amount of transcription factor binding sites might work as a required nucleation step for the biomolecular condensates.

In these or other scenarios, as TGF $\beta$  is an essential regulator of cell proliferation, survival, differentiation, and plays a critical role in cancer and neurodegenerative disorders<sup>76</sup>, understanding how the phase separation of its coactivator JMJD3 impacts its regulatory processes could undoubtedly contribute to understanding

the crosstalk between diseases and development, and potentially provide new therapeutic targets.

**Methods**

**Cell culture and cell treatments.** Briefly, mouse NSCs were dissected from cerebral cortices of C57BL/6J mouse fetal brains (E12.5) and cultured in poly-D-lysine (5  $\mu$ g/ml, 2 h 37  $^{\circ}$ C) and laminin (5  $\mu$ g/ml 37  $^{\circ}$ C, 4 h 37  $^{\circ}$ C) precoated dishes<sup>77</sup> and have subsequently been maintained in culture as a stable cell line. NSCs were grown with a medium prepared by mixing equal parts of DMEM F12 (without Phenol Red, Gibco) and Neural Basal Media (Gibco), Glutamax (1%), N2 and B27 supplements (Gibco), sodium pyruvate (1 mM), non-essential amino acids (0.1 mM), Heparin (2 mg/l), Hepes (5 mM), bovine serum albumin (25 mg/l) and

β-mercaptoethanol (0.01 mM)<sup>23</sup>. Fresh recombinant human Epidermal Growth Factor (EGF) (R&D systems) and Fibroblast Growth Factor (FGF) (Invitrogen) to 20 ng/ml and 10 ng/ml final concentrations respectively were added to the media. TGFβ (Millipore) was used at a final concentration of 5 ng/ml. Human HEK293T cells were cultured in DMEM supplemented with 10% of fetal bovine serum (Gibco) and 1% of Penicillin/Streptomycin<sup>78</sup>.

**Antibodies and reagents.** Antibodies used were anti: JMJD3<sup>28</sup> (raised in the laboratory using amino acids 798–1095, dilution 1:200 for immunofluorescence (IF), 1:1000 for western blot (WB)); and Abcam, ab38113, dilution 1:250 for IF), DAPI (ThermoFisher, D1306, dilution 1:500 for IF), β-TUBULIN (Millipore, MAB3408, dilution 1:5000 for WB), HA tag (Abcam, ab20084, dilution 1:5000 for WB), H3K9me3 (Abcam, ab8898, dilution 1:250 for IF), H3K27me3 (Millipore, 07449, dilution 1:500) and Goat anti-Rabbit IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 488 (Invitrogen, A32731, dilution 1:1000 for IF), Goat anti-Mouse IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 555 (Invitrogen, A32727, dilution 1:1000 for IF). TGFβ and doxycycline hyclate were acquired from Millipore (GF111 and 324385 respectively). The doxycycline was used at a concentration of 1 μg/ml for 24 h. GSK-J4 inhibitor was acquired from Selleckchem (GSKJ4 HCl S7070) and used at a concentration of 1 μM for 6 h.

**Plasmids.** Specific lentiviral vectors were purchased from Sigma: pLKO.1-random (CAACAAGATGAAGAGACC), pLKO.1-shSMAD3 (CCTTACCACTATCA-GAGAGTA), and pLKO.1-shJMJD3 (CCTCTGTTCTTGAGGGACAAA). mEGFP was amplified by PCR from the pCMV-mEGFP-C1 vector and cloned into the pCMV-HA-JMJD3 plasmid by using an Acc65I restriction site. Primer sequences are described in Supplementary Table 1. pCMV-HA-mEGFP-JMJD3 ΔIDR (140–820) was obtained by digesting pCMV-HA-mEGFP-JMJD3 vector with HindIII enzyme and self-ligating the resulting fragment. pCMV-HA-mEGFP-JMJD3 HE > AA (H1390A/E1392A) was obtained by replacing JMJD3 from pCMV-HA-mEGFP-JMJD3 vector with JMJD3 HE > AA (H1390A/E1392A). This mutant version was cut from pCIG-JMJD3-H1390A/E1392A vector with XhoI and XagI enzymes. pInducer-JMJD3 WT and pInducer-JMJD3 HE > AA (H1390A/E1392A) vectors were obtained through an LR recombination reaction taking advantage of the Gateway Recombination Cloning Technology.

**Lentiviral transduction.** Lentiviral transduction to generate the knockdown cell lines was carried out as previously described<sup>79</sup>. Briefly, lentiviral particles are produced in HEK293T cells by cotransfecting plasmids encoding the shRNA together with pCMV-VSVG and pCMV-GAG-POL plasmids. After 24 h, supernatants containing lentiviral particles are collected and centrifuged in a sucrose bed at 57000 x g for 2 h. Then, viral particles are resuspended in NSCs medium, and they are directly used for NSCs infection. One day after infection, cells are selected with puromycin (Sigma #P8833) at a concentration of 2 μg/ml for several days.

**4C-seq assay.** The 4C-seq experimental protocol was based on<sup>80,81</sup>. Briefly,  $12 \times 10^6$  mNSCs were fixed for 30 min using 1% of formaldehyde at room temperature. The fixing reaction was quenched with glycine 0.125 M for 10 min. After 2 washes with PBS, cell pellets were resuspended in 5 ml of cytoplasmic lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100 and protease inhibitors) during 10 min on ice. Lysates were centrifuged for 5 min at 650 x g and 4 °C. Nuclei were resuspended in 0.5 ml of NlaIII buffer with 0.3% SDS and they were incubated at 37 °C and 900 rpm for one hour. After that, Triton X-100 was added to a final concentration of 2% followed by 1 h of incubation at 37 °C and 900 rpm. Next, DNA was digested overnight at 37 °C and 900 rpm with 400 U of NlaIII, which was afterward inactivated by adding SDS to a final concentration of 1.6% and incubating for 20 min at 65 °C and 900 rpm. The digested chromatin was transferred to 50 ml tubes and 6.125 ml of 1.15X ligation buffer (50 mM Tris-HCl pH 7.6, 10 mM MgCl<sub>2</sub>, 1 mM ATP, 1 mM DTT) and 1% of Triton X-100 were added and incubated during 1 h at 37 °C and 1000 rpm. Digested chromatin was ligated with 100 U of T4 DNA ligase for 8 h at 16 °C and then, treated with RNase A 1 mg/ml for 45 min at 37 °C. Decrosslinking step was performed by adding 1 mg/ml of proteinase K and incubating at 65 °C overnight. DNA was purified by standard phenol-chloroform extraction followed by ethanol precipitation and resuspended in 100 μl of H<sub>2</sub>O. At this point, proper digestion and ligation were evaluated by visualizing the DNA in an agarose gel. A second digestion with 50 U of DpnII was performed at 37 °C overnight. Enzyme and buffer were removed from the sample by phenol-chloroform extraction followed by ethanol precipitation and once purified, DNA samples were resuspended in 500 μl of H<sub>2</sub>O. A second ligation was carried out by adding 200 U of T4 DNA ligase in a final volume of 14 μl of 1X ligation buffer. The mixture is incubated overnight at 16 °C and after the last round of phenol-chloroform extraction and ethanol precipitation, the DNA was resuspended in 100 μl of H<sub>2</sub>O and purified with a Qiagen PCR purification column. The efficiencies of the second digestion and ligation are tested by DNA electrophoresis, when correct, this sample comprises the 4C-library sample. Using the indicated viewpoints, inverse PCR reactions were performed using the Expand Long template PCR system (Roche #11681834001) with the following cycling conditions: 94 °C 2 min, 29 cycles of 94 °C 10 s – 55 °C 1 min –

68 °C 3 min and 68 °C 5 min (for *Chst8* and *Ldlrad4*) and 94 °C 2 min, 31 cycles of 94 °C 10 s – 55 °C 1 min – 68 °C 3 min and 68 °C 5 min (for *Aopep*). Primer sequences are described in Supplementary Table 2. The products of these reactions were sent for sequencing to the Erasmus Center for Biomics in Rotterdam, in an Illumina HiSeq2500 sequencer with a read depth of 100 bp in the case of the *Chst8* VP and 74 bp in the other samples. 4C-seq data have been deposited in the GEO database under the accession GSE197013 (GSE197010 for *Chst8*; GSE197011 for *Aopep*; GSE197012 for *Ldlrad4*).

**Computational analysis of the 4C-seq experiment.** The obtained sequencing reads were processed using the 4C-seq pipeline named pipe4C<sup>35</sup> using default parameters except for the trimLength that was set up to 36 bp, and the genome version for mapping that was *Mus musculus* mm10. Further statistical analysis was performed with R3Cseq<sup>82</sup>, a Bioconductor package that allows the identification of interacting genomic regions and the comparison between multiple replicates and experimental conditions.

**RNA extraction and qPCR.** RNA was extracted using TRIZOL reagent (Invitrogen), following the manufacturer instructions. Reverse transcription was performed with 200–1000 ng of RNA using High Capacity cDNA reverse transcription kit (Invitrogen). qPCR was carried out with SYBR Green (Roche) in a QuantStudio 5 Real-Time PCR system (ThermoFisher Scientific) using specific primer pairs (see Supplementary Table 1).

**ChIP assays.** Chromatin immunoprecipitation (ChIP) assays were performed as previously described<sup>83</sup> with modifications:  $6 \times 10^6$  NSCs were fixed with formaldehyde 1% during 10 min. Fixation was stopped by the addition of 0.125 M of glycine. Cells were lysed in 1% SDS lysis buffer (1% SDS; 10 mM EDTA pH 8.0; 50 mM Tris-HCl pH 8.1). A sonication step to fragment the chromatin was performed in a Bioruptor sonicator (Diagenode) and shredded chromatin was used for each immunoprecipitation using the H3K27me3 antibody. Magna ChIP Protein A Magnetic Beads (Millipore) were used to capture the immunocomplex. DNA was purified by phenol-chloroform followed by ethanol precipitation and analyzed by qPCR with SYBR Green (Roche) in a QuantStudio 5 Real-Time PCR system (ThermoFisher Scientific) using specific primers (see Supplementary Table 1).

**ChIP-seq data acquisition.** ChIP-seq data were downloaded from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) (Accessions used in this paper are specified in Supplementary Table 3). ChIP-seq captions were obtained from UCSC genome browser<sup>84,85</sup>.

**Western blot.** Immunoblotting was performed using standard procedures. An ECL kit (Amersham) was used to visualize the results. Uncropped and unprocessed scans of the performed Western Blots are provided in the Source Data file.

**CRISPR-Cas9.** In order to delete the *Chst8* viewpoint (VP) enhancer, primer pairs of gRNA (Supplementary Table 1) were designed flanking the mm10 coordinates chr7:34846279–34849157 using the online tool <http://crispr.mit.edu/>. Selected primer pairs have an off-target score of 80 (left) and 90 (right) and an on-target score of 69 (left) and 65 (right). gRNAs were cloned into the pX330-U6-Chimeric\_BB-CBh-hSpCas9 vector (Addgene #42230) using BbsI sites. Plasmids were nucleofected in NSCs with an Amaxa Nucleofector (Lonza) following manufacturer instructions. After puromycin selection (0.8 μg/ml) and detection analysis with conventional PCR, heterogeneous population carrying a majority of homozygotic deletions was used for experiments.

**Droplet assays in nuclear extracts.** 5 μg of vector encoding *Jmjd3* cDNA fused to mEGFP were transfected into  $20 \times 10^6$  HEK293T cells as previously described<sup>86</sup>. Nuclear extracts were prepared at a concentration of 2 mg/ml and they were used for droplet formation assays by diluting them 1:1 with buffer (10% glycerol, 20 mM HEPES). The final droplet buffer conditions were 20 mM HEPES, 150 mM NaCl, 15% glycerol, 3.75 mM EGTA, 2.5 mM MgCl<sub>2</sub>, and 1.25 mM CaCl<sub>2</sub>. The reactions were incubated for 30 min and loaded onto a glass-bottom 384-well plate (Cellvis P384-1.5H-N) 5 min before imaging on an Automated Inverted Microscope Leica Thunder 3D Live Cell using a 63x water immersion objective (NA = 1.2).

**Quantification of droplets liquid-like features.** Droplets shape descriptors “circularity” and “aspect ratio” were quantified using the “Analyze particles” plugin in Fiji; “convexity” was calculated running the “Calculate Convexity and Solidarity” macro in Fiji<sup>12,87,88</sup>. Each image was cropped, and a threshold was set so that each droplet could be seen as an individual object. The results showed in the figure correspond to the measurement of 150 droplets.

**Fluorescence recovery after photobleaching (FRAP) in live cells.** HEK293T cells were transfected with 0.05 μg of mEGFP-JMJD3 vector and grown on glass dishes coated with 5 μg/ml of poly-D-lysine in 1.5 ml of DMEM media as previously described in this manuscript. FRAP was performed on a Zeiss LSM780

confocal microscope equipped with a 40x water immersion objective (NA = 1.2) and a GaAsP photomultiplier detector. Acquisition settings were optimized for fast imaging and low photobleaching, using 488 nm laser excitation power of 0.15% (AOTF), a detector gain of 780, a pixel dwell of 1.27  $\mu$ sec and a pixel size of 140 nm. Bleaching was performed after 5 previous images by using 488 nm laser excitation power of 100% (AOTF), a pixel dwell of 2.55 and 10 iterations, over a 6 × 6 pixels region of interest (ROI) focused on the interest spot. Acquisition was set to intervals of 1 s for both the pre-bleach imaging and the post-bleach recovery time.

Intensity recovery quantification was performed using Fiji<sup>89</sup>. A macro was programmed that (i) registered the whole time-lapse to avoid live cell fluctuations; (ii) allowed the user to draw a ROI around the bleached spot; (iii) fine-tuned the selected ROI by applying a threshold on the time projection of the signal spot; (iv) allowed the user to select a background ROI; (v) automatically segmented the target nucleus and created a ROI where to calculate the bleaching gap and the bleach depth; (vi) automatically measured the intensity of the three ROIs over the time-lapse and (vii) delivered the data in a \*.txt format. The code and further details can be downloaded from <https://github.com/MolecularImagingPlatformIBMB>. For each cell, three rounds of image quantification were collected, as to minimize the experimental error due to image analysis quantification. Statistics and fitting were performed using the easy-FRAP web<sup>90</sup>.

**1,6-Hexanediol treatment for live imaging of cells.** HEK293T cells were transfected with 0.05  $\mu$ g of HA-EGFP-JMJD3 vector and grown on glass dishes coated with 5  $\mu$ g/ml of poly-D-lysine in 1.5 ml of DMEM media as previously described in this manuscript. They were imaged before treatment on a 37 °C heated stage of a Zeiss LSM780 Confocal using Zen software to establish a baseline. The Spectral (GaAsP) detector and a 40x water immersion (NA = 1.2) objective were used. After the fifth acquisition, 1,6-Hexanediol (#240117, Sigma) was added to cells at a final concentration of 6% in normal media, and images were again taken for 5 min of continuous treatment. Raw images were processed using Fiji software for posterior analysis and quantifications. Representative and consistent images of puncta disassembly at 60 and 120 s are presented.

**Focus calling (Immunofluorescence, 1,6-Hexanediol treatment).** Foci were called using the “Object Counter 3D” plugin in Fiji. For each image, the “threshold” parameter was set so that each focus could be seen as an individual object. The parameters showed (number of foci/cell, intensity, and volume) are the mean of the results obtained for each image with the “Statistics” function of the plugin (the number of cells used for the quantifications is specified in each figure legend).

**Indirect immunofluorescence.** Immunostaining assays were carried out as previously described<sup>91</sup>. Briefly, cells were fixed with 4% paraformaldehyde for 20 min and permeabilized with PBS-Triton X-100 (0.5%). Cells were blocked for 1 h at room temperature in 0.5% BSA (in PBS with 0.1% Triton X-100) before overnight incubation at 4 °C with primary antibodies. Next, cells were incubated with Alexa-conjugated secondary IgG antibodies (Jackson ImmunoResearch) and 0.1 ng/ $\mu$ l DAPI (Sigma) for 2 h at room temperature. For cells overexpressing mEGFP or mCherry-fusion proteins, the intrinsic fluorescence of the molecules was captured without using either primary or secondary antibodies. Images were captured by Leica SP5 confocal microscope using LAS-AF software.

**Immunofluorescence with DNA FISH.** Mouse NSCs were grown on glass dishes pre-coated with 5  $\mu$ g/ml of poly-D-lysine and 5  $\mu$ g/ml of laminin and treated with TGF $\beta$  for 3 h as previously described in this manuscript. After fixation in 4% paraformaldehyde containing 0.1% Triton X-100 solution for 10 min, cells were incubated in 10 mM glycine for 30 min and washed with PBS three times. Then, cells were dehydrated by performing sequential washes with 70%, 85% and 100% ethanol for 2 min at RT and then air dried. Probe hybridization mixture was made by mixing 8  $\mu$ l of FISH Hybridization Buffer (Empire Genomics) and 2  $\mu$ l of the FISH probe (see below). The 10  $\mu$ l of mixture was added on a slide. Genomic DNA and probes were denatured at 75 °C for 7 min and slides were incubated at 37 °C in the dark overnight. The coverslip was removed from the slide and washed twice at 73 °C for 5 min with Wash solution 1 (0.3% NP-40/0.4x SCC) and twice at room temperature for 2 min with Wash solution 2 (0.3% NP-40/3x SCC). Then immunofluorescence was performed. Cells were fixed again in 4% paraformaldehyde for 10 min and washed with PBS three times. Permeabilization was done in 0.2% Triton X-100 in PBS for 10 min. After washing with PBS, cells were blocked at room temperature with PBG for 1 h. Then, cells were incubated with the primary antibody (anti-JMJD3 ab38113) at 4 °C overnight. After three washes with PBG, cells were incubated with the secondary antibody (Alexa anti-rabbit 488) for 45 min at room temperature. The coverslip was washed twice with PBG and twice with PBS, and nuclei were stained with DAPI for 10 min at room temperature. Coverslips were mounted with Mowiol and images were acquired on an Andor Dragonfly Spinning Disk Confocal Microscope with an oil immersion 100x (NA = 1.49) objective and a pixel size of 51 nm objective using Fusion acquisition software. Images were post-processed using Fiji<sup>92</sup>. FISH foci were manually identified in individual z

stacks through intensity thresholds in FIJI. The DNA FISH probe was synthesized by Empire Genomics and targets the *Chst8* enhancer cluster locus (mm10 coordinates chr7:34795935-34985109).

**Sequence analysis and predictions.** Protein disorder estimations were generated using three prediction algorithms, PONDR-VL3<sup>45</sup>, IUPred<sup>46</sup> and PONDR-VSL2<sup>47</sup>. The predictors give a value between 0 and 1 for each amino acid, where above 0.5 is predicted to lie within a disordered region of more than 50 amino acids long. To predict the phase separation property of each protein, PSPredictor<sup>53</sup> and catGRANULE<sup>52</sup> predictors were used online.

Low-complexity domains presence was assessed using SEG algorithm together with MobiDB database<sup>48</sup>. For amino acid composition analysis, the web application Prot Pi Protein Tool <https://www.protpi.ch/Calculator/ProteinTool> was used. Disordered proteins were defined by the presence of a 50 residues fragment whose IUPRED median score was at least of 0.55 and that was not found in Pfam, a protein domain database. The hydrophobicity was calculated with the ExPASy website<sup>93</sup> using the Hopp and Woods scale<sup>94</sup> and a sliding window of 21.

**Statistical analysis.** Quantitative data were expressed as mean and standard error mean (SEM) (for immunofluorescence quantifications and RNA transcription experiments). The significance of differences was assessed using the Student’s *t* test (\* $p < 0.05$ ; \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

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

## Data availability

The data that support this study are available from the corresponding author upon reasonable request. The 4C-seq data generated in this study have been deposited in the GEO database under accession code GSE197013. The ChIP-seq data used in this study are available in the GEO database under accession code GSM898371, GSM937827, GSE66961, GSE66961, GSE66961, GSM883646, GSE38269. Source data are provided with this paper.

## Code availability

The code and further details for FRAP image analysis can be downloaded from <https://github.com/MolecularImagingPlatformIBMB>.

Received: 6 November 2020; Accepted: 5 May 2022;  
Published online: 07 June 2022

## References

- Tiberi, L., Vanderhaeghen, P. & van den Aemele, J. Cortical neurogenesis and morphogens: diversity of cues, sources and functions. *Curr. Opin. Cell Biol.* **24**, 269–276 (2012).
- Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: Lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).
- Schaffner, W. Enhancers, enhancers - From their discovery to today’s universe of transcription enhancers. *Biol. Chem.* **396**, 311–327 (2015).
- Weake, V. M. & Workman, J. L. Inducible gene expression: diverse regulatory mechanisms. *Nat. Rev. Genet.* **11**, 426–437 (2010).
- Moore, J. E. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
- Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
- Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
- Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
- Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Sci. (80-.)* **361**, eaar3958 (2018).
- Cho, W. K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Sci. (80-.)* **361**, 412–415 (2018).
- Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).

13. Nott, T. J. et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
14. Pak, C. W. et al. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **63**, 72–85 (2016).
15. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
16. Fuxreiter, M. Fuzziness in protein interactions—a historical perspective. *J. Mol. Biol.* **430**, 2278–2287 (2018).
17. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* eaaf4382. <https://doi.org/10.1126/science.aaf4382> (2017).
18. Wheeler, R. J. & Hyman, A. A. Controlling compartmentalization by non-membrane-bound organelles. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170193 (2018).
19. Bleckwehl, T. & Rada-Iglesias, A. Transcriptional and epigenetic control of germline competence and specification. *Curr. Opin. Cell Biol.* **61**, 1–8 (2019).
20. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Sci. (80-.)* **361**, 1341–1345 (2018).
21. Maeso, I., Acemel, R. D. & Gómez-Skarmeta, J. L. Cis-regulatory landscapes in development and evolution. *Curr. Opin. Genet. Dev.* **43**, 17–22 (2017).
22. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
23. Estarás, C. et al. Genome-wide analysis reveals that Smad3 and JMJD3 HDM co-activate the neural developmental program. *Development* **139**, 2681–2691 (2012).
24. García-Campmany, L. & Martí, E. The TGF $\beta$  intracellular effector Smad3 regulates neuronal differentiation and cell fate specification in the developing spinal cord. *Development* **134**, 65–75 (2007).
25. Roussa, E. et al. Transforming growth factor  $\beta$  is required for differentiation of mouse mesencephalic progenitors into dopaminergic neurons in vitro and in vivo: ectopic induction in dorsal mesencephalon. *Stem Cells* **24**, 2120–2129 (2006).
26. Vogel, T., Ahrens, S., Büttner, N. & Kriegelstein, K., Transforming Growth Factor  $\beta$  Promotes Neuronal Cell Fate of Mouse Cortical and Hippocampal Progenitors In Vitro and In Vivo: Identification of Nedd9 as an Essential Signaling Component. *Cerebral Cortex [Internet]*. **20**, 661–671 (2010).
27. Dahle, Ø., Kumar, A. & Kuehn, M. R. Nodal signaling recruits the histone demethylase Jmjd3 to counteract Polycomb-mediated repression at target genes. *Sci. Signal.* **3**, ra48 (2010).
28. Kim, S. W. et al. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.* **357**, 492–504 (2011).
29. Agger, K. et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731–734 (2007).
30. De Santa, F. et al. The Histone H3 Lysine-27 Demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* **130**, 1083–1094 (2007).
31. Burchfield, J. S., Li, Q., Wang, H. Y. & Wang, R.-F. JMJD3 as an epigenetic regulator in development and disease. *Int. J. Biochem. Cell Biol.* **67**, 148–157 (2015).
32. Ding, Y. et al. JMJD3: a critical epigenetic regulator in stem cell fate. *Cell Commun. Signal.* **19**, 72 (2021).
33. Fueyo, R. et al. Lineage specific transcription factors and epigenetic regulators mediate TGF $\beta$ -dependent enhancer activation. *Nucleic Acids Res.* **46**, 3351–3365 (2018).
34. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* **17**, 207–223 (2016).
35. Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17–32 (2020).
36. Estarás, C., Fueyo, R., Akizu, N., Beltrán, S. & Martínez-Balbás, M. A. RNA polymerase II progression through H3K27me3-enriched gene bodies requires JMJD3 histone demethylase. *Mol. Biol. Cell* **24**, 351–360 (2013).
37. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev.* **30**, 881–891 (2016).
38. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2017).
39. Davidson, I. F. et al. DNA loop extrusion by human cohesin. *Sci. (80-.)* **366**, 1338–1345 (2019).
40. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
41. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
42. Chen, S. et al. The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. *Genes Dev.* **26**, 1364–1375 (2012).
43. Dunker, A. K., Bondos, S. E., Huang, F. & Oldfield, C. J. Intrinsically disordered proteins and multicellular organisms. *Semin. Cell Dev. Biol.* **37**, 44–55 (2015).
44. Wallmann, A. & Kesten, C. Common functions of disordered proteins across evolutionary distant organisms. *Int. J. Mol. Sci.* **21**, 2105 (2020).
45. Peng, K. et al. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol.* **3**, 35–60 (2005).
46. Dosztányi, Z., Csizmók, V., Tompa, P. & Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **347**, 827–839 (2005).
47. Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. & Obradovic, Z. Length-dependent prediction of protein in intrinsic disorder. *BMC Bioinforma.* **7**, 208–224 (2006).
48. Piovesan, D. et al. MobiDB3.0: More annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.* **46**, D471–D476 (2018).
49. Murthy, A. C. et al. Molecular interactions underlying liquid–liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).
50. Williamson, M. P. The structure and function of proline-rich regions in proteins. *Biochem. J.* **297**, 249–260 (1994).
51. Mir, M., Bickmore, W., Furlong, E. E. M. & Narlikar, G. Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development* **146**, (2019).
52. Bolognesi, B. et al. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.* **16**, 222–231 (2016).
53. Sun, T. et al. Prediction of liquid-liquid phase separation proteins using machine learning. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3515387> (2020).
54. Mitrea, D. M. et al. Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation. *Nat. Commun.* **9**, 842–854 (2018).
55. Lin, Y., Currie, S. L. & Rosen, M. K. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* **292**, 19110–19120 (2017).
56. Hennig, S. et al. Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J. Cell Biol.* **210**, 529–539 (2015).
57. Kroschwald, S., Maharana, S. & Simon, A. Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments. *Matters* 10.19185, 201702000010 (2017).
58. Boeynaems, S. et al. Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
59. Henis, Y. I., Rotblat, B. & Kloog, Y. FRAP beam-size analysis to measure palmitoylation-dependent membrane association dynamics and microdomain partitioning of Ras proteins. *Methods* **40**, 183–190 (2006).
60. Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604–609 (2000).
61. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Sci. (80-.)* **361**, eaar2555 (2018).
62. Kruidenier, L. et al. A selective jumoni H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404–408 (2012).
63. Guo, Y. E. et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
64. Wagh, K., Garcia, D. A. & Upadhyaya, A. Phase separation in transcription factor dynamics and chromatin organization. *Curr. Opin. Struct. Biol.* **71**, 148–155 (2021).
65. Schneider, N. et al. Liquid-liquid phase separation of light-inducible transcription factors increases transcription activation in mammalian cells and mice. *Sci. Adv.* **7**, eabd3568 (2021).
66. Snetkova, V. & Skok, J. A. Enhancer talk. *Epigenomics* **10**, 483–498 (2018).
67. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
68. Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Mol. Cell* **71**, 468–480 (2018).
69. Cabral, R. M. et al. Whole-exome sequencing in a single proband reveals a mutation in the CHST8 gene in autosomal recessive peeling skin syndrome. *Genomics* **99**, 202–208 (2012).
70. Kartikasari, A. E. R. et al. The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation. *EMBO J.* **32**, 1393–1408 (2013).
71. Shi, B. et al. UTX condensation underlies its tumour-suppressive activity. *Nature* **597**, 726–731 (2021).

72. Lu, H. et al. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318–323 (2018).
73. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855 (2018).
74. Shrinivas, K. et al. Enhancer features that drive formation of transcriptional condensates. *Mol. Cell* **75**, 549–561.e7 (2019).
75. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science (80-)*. **361**, eaar2555 (2018).
76. Batlle, E. & Massagué, J. Transforming growth factor- $\beta$  signaling in immunity and cancer. *Immunity* **50**, 924–940 (2019).
77. Spencer Currell, D., Hu, J. S., Kolski-Andreaco, A. & Monuki, E. S. Culture of mouse neural stem cell precursors. *J. Vis. Exp.* **152**, 152 (2006).
78. Blanco-García, N., Asensio-Juan, E., De La Cruz, X. & Martínez-Balbás, M. A. Autoacetylation regulates P/CAF nuclear localization. *J. Biol. Chem.* **284**, 1343–1352 (2009).
79. Asensio-Juan, E. et al. The histone demethylase PHF8 is a molecular safeguard of the IFN $\kappa$  response. *Nucleic Acids Res.* **45**, 3800–3811 (2017).
80. Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Methods* **58**, 221–230 (2012).
81. Stadhouders, R. et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.* **8**, 509–524 (2013).
82. Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E. & Lenhard, B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res.* **41**, e132–e132 (2013).
83. Valls, E. et al. Involvement of chromatin and histone deacetylation in SV40T antigen transcription regulation. *Nucleic Acids Res.* **35**, 1958–68 (2007).
84. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
85. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
86. Zamudio, A. V. et al. Mediator condensates localize signaling factors to key cell identity genes. *Mol. Cell* **76**, 1–14 (2019).
87. Gopal, P. P., Nirschl, J. J., Klinman, E. & Holzbaur, E. L. F. Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. *Proc. Natl Acad. Sci.* **114**, E2466–E2475 (2017).
88. Weber, S. C. & Brangwynne, C. P. Getting RNA and protein in phase. *Cell* **149**, 1188–1191 (2012).
89. Schindelin, J. et al. Fiji - an Open platform for biological image analysis. *Nat. Methods* **9**, 676–682 (2009).
90. Koulouras, G. et al. EasyFRAP-web: a web-based tool for the analysis of fluorescence recovery after photobleaching data. *Nucleic Acids Res.* **46**, W467–W472 (2018).
91. Akizu, N. et al. EZH2 regulates neuroepithelium structure and neuroblast proliferation by repressing p21. *Open Biol.* **6**, 150227 (2016).
92. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
93. Gasteiger, E. et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
94. Hopp, T. P. & Woods, K. R. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl Acad. Sci. USA* **78**, 3824–3828 (1981).

## Acknowledgements

We thank Dr. Benedetta Bolognesi for advice, reagents, and technical help. We also thank Drs G. Vicent, M. Beato, J. Bernues, F. Azorin, A. Jordá, A. Aragay, A. Vaquero, S.

Sánchez-Molina and J Mora for reagents. Drs F. De Lilly, C. Gallego, G. Jiménez, S. and L. Pascuali for technical help. Dr. K. Helin for kindly providing the pCMV-HA-JMJD3 vector. We thank Dr. R. Young, Dr. E. Wanker, and M. Zenkner for sharing the MED15 plasmid with us. We acknowledge Leica microsystems for support to the Molecular Imaging Platform through a Thunder 3D Live Cell collaboration agreement. We also acknowledge BioRender platform, which was used for the creation of some figures (Figs. 1a and 7c were created with BioRender.com). This study was supported by the grants BFU2015-69248-P and PGC2018-096082-B-I00 to MMB from the Spanish Ministry of Science and Technology. EQC2018-004541-P from MICIU, Co-financed by the European Regional Development Fund (FEDER) to IBMB. R.F. and M.V. were recipients of an FPU (Spanish Ministry of Education and Science) and FI (Co-funded by Secretaría de Universidades e Investigación del Departamento de Empresa y Conocimiento de la Generalidad de Cataluña and EU – “European Social Fund investing in your future”) fellowships respectively. R.F. and M.V. were also awarded a traveling fellowship from the Development journal of the Company of Biologists and The EMBO respectively.

## Author contributions

M.V. and R.F. performed experiments, contributed to experimental design, analyzed the data and revised the paper; C.N. and S.C.-M. conducted experiments and revised de manuscript; W.F.J.v.l. contributed to the genomic sequencing design and performance; E.R. and A.R.-I. designed experiments, analyzed the data and revised the paper. M.A.M.B. and R.F. conceived the project. M.A.M.B. analyzed the data and wrote the paper with editing support from all authors.

## 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/s41467-022-30614-y>.

**Correspondence** and requests for materials should be addressed to Marian A. Martínez-Balbás.

**Peer review information** *Nature Communications* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work.

**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.

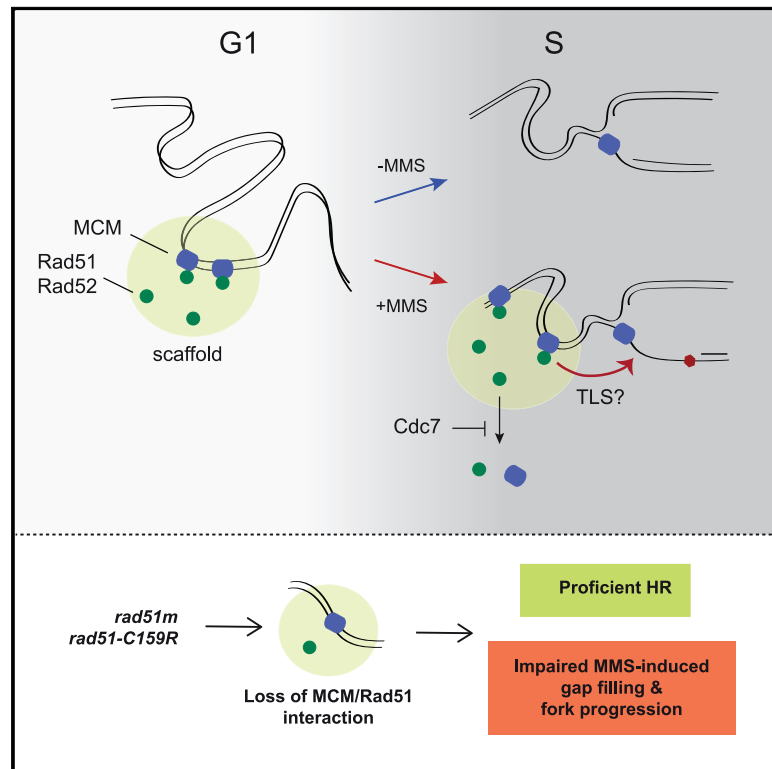


**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

# Physical interactions between MCM and Rad51 facilitate replication fork lesion bypass and ssDNA gap filling by non-recombinogenic functions

## Graphical abstract



## Authors

María J. Cabello-Lobato,  
Cristina González-Garrido,  
María I. Cano-Linares, ...,  
Román González-Prieto, Helle D. Ulrich,  
Félix Prado

## Correspondence

felix.prado@cabimer.es

## In brief

Cabello-Lobato et al. find that MCM interacts dynamically with Rad51 and Rad52 within a nuclease-insoluble nucleoprotein scaffold. These interactions are established in G1 and maintained in the S phase by Cdc7 in the presence of replication-blocking lesions to assist stressed forks through non-recombinogenic functions.

## Highlights

- Rad51 and Rad52 interact with MCM in a nuclease-insoluble nucleoprotein scaffold
- MCM/Rad51/Rad52 accumulation is regulated by cell cycle and replicative DNA damage
- Cdc7 prevents Rad51/Rad52 release from the scaffold under replicative DNA damage
- MCM/Rad51 promotes MMS-induced gap filling and fork progression by non-HR processes





## Article

# Physical interactions between MCM and Rad51 facilitate replication fork lesion bypass and ssDNA gap filling by non-recombinogenic functions

María J. Cabello-Lobato,<sup>1,3,4</sup> Cristina González-Garrido,<sup>1,3</sup> María I. Cano-Linares,<sup>1</sup> Ronald P. Wong,<sup>2</sup> Aurora Yáñez-Vílchez,<sup>1</sup> Macarena Morillo-Huesca,<sup>1</sup> Juan M. Roldán-Romero,<sup>1</sup> Marta Vicioso,<sup>1</sup> Román González-Prieto,<sup>1,5</sup> Helle D. Ulrich,<sup>2</sup> and Félix Prado<sup>1,6,\*</sup>

<sup>1</sup>Centro Andaluz de Biología Molecular y Medicina Regenerativa–CABIMER, Consejo Superior de Investigaciones Científicas; Universidad de Sevilla; Universidad Pablo de Olavide; Seville, Spain

<sup>2</sup>Institute of Molecular Biology (IMB), Mainz, Germany

<sup>3</sup>These authors contributed equally

<sup>4</sup>Present address: Manchester Cancer Research Centre, Division of Cancer Sciences, University of Manchester, Manchester M20 4GJ, UK

<sup>5</sup>Present address: Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands

<sup>6</sup>Lead contact

\*Correspondence: [felix.prado@cabimer.es](mailto:felix.prado@cabimer.es)

<https://doi.org/10.1016/j.celrep.2021.109440>

## SUMMARY

The minichromosome maintenance (MCM) helicase physically interacts with the recombination proteins Rad51 and Rad52 from yeast to human cells. We show, in *Saccharomyces cerevisiae*, that these interactions occur within a nuclease-insoluble scaffold enriched in replication/repair factors. Rad51 accumulates in a MCM- and DNA-binding-independent manner and interacts with MCM helicases located outside of the replication origins and forks. MCM, Rad51, and Rad52 accumulate in this scaffold in G1 and are released during the S phase. In the presence of replication-blocking lesions, Cdc7 prevents their release from the scaffold, thus maintaining the interactions. We identify a *rad51* mutant that is impaired in its ability to bind to MCM but not to the scaffold. This mutant is proficient in recombination but partially defective in single-stranded DNA (ssDNA) gap filling and replication fork progression through damaged DNA. Therefore, cells accumulate MCM/Rad51/Rad52 complexes at specific nuclear scaffolds in G1 to assist stressed forks through non-recombinogenic functions.

## INTRODUCTION

Replicative stress is a major source of genomic instability, which is associated with cancer and genetic diseases (Hills and Diffley, 2014). Cells are endowed with different mechanisms to deal with DNA lesions that hinder the advance of replication forks. The DNA damage tolerance (DDT) response facilitates replication fork progression through damaged DNA, postponing the repair of blocking lesions for later stages to timely complete genome duplication. The homologous recombination (HR) machinery has multiple and critical roles in this process through mechanisms that are mostly error free, in contrast to translesion synthesis (TLS) mechanisms, which can be mutagenic (Branzei and Psakhye, 2016; Prado, 2014, 2018). HR proteins escort the fork under unperturbed conditions (Alabert et al., 2014; González-Prieto et al., 2013; Hashimoto et al., 2010; López-Contreras et al., 2013) and assist replication forks in response to replicative stress. In mammalian cells, the recombinase RAD51 is required for the formation and protection of reversed forks as intermediates to promote replication fork bypass of blocking lesions (Bhat and Cortez, 2018; Zellweger et al., 2015). Reversed forks are rarely detected in

yeast cells treated with the alkylating agent methyl methanesulfonate (MMS) or UV light, which cause DNA blocking lesions (Giannattasio et al., 2014; Lopes et al., 2006). However, yeast Rad51 also prevents the accumulation of single-stranded DNA (ssDNA) fragments at the forks (Hashimoto et al., 2010; Lopes et al., 2006), and together with its mediator Rad52 are required for fork progression in the presence of alkylated bases (Alabert et al., 2009; González-Prieto et al., 2013; Vázquez et al., 2008).

It has long been established in the yeast *Saccharomyces cerevisiae* that HR proteins are required to fill in the stretches of ssDNA generated during DNA replication in the presence of DNA adducts (Gangavarapu et al., 2007; Jachymczyk et al., 1977; Prakash, 1981; Resnick et al., 1981). This process requires multiple recombination factors, including Rad51 and Rad52, and relies on the transient formation of sister chromatid junctions (SCJs) through strand-exchange reactions (Branzei et al., 2008; Liberi et al., 2005; Mankouri et al., 2007; Vanoli et al., 2010). In addition, we have recently shown that Rad51 and Rad52 facilitate TLS through non-recombinogenic functions (Cano-Linares et al., 2021). In *S. cerevisiae*, the repair of these ssDNA gaps occurs post-replicatively at DNA regions that are



spatially separated from replication forks (González-Prieto et al., 2013; Wong et al., 2020).

Previous studies in *Schizosaccharomyces pombe* and human cells showed that Rad51 and Rad52 physically interact with the minichromosome maintenance (MCM) helicase complex in the absence and presence of drugs that impair DNA replication (Baillis et al., 2008; Shukla et al., 2005). The MCM helicase is a conserved ring-shaped complex formed by six related subunits (Mcm2 to Mcm7 in *S. cerevisiae*) with essential roles in DNA replication. The loading of MCM at chromatin occurs during late mitosis and G1 and requires the initial binding of the origin recognition complex (ORC) to a replication origin, followed by the recruitment of Cdc6 and the final entry of an MCM/Cdt1 haptamer. In this pre-replicative complex (pre-RC), the MCM helicase is loaded as an inactive head-to-head dimer that encircles dsDNA. This complex is activated during S phase through a process that requires the activities of the cyclin- and Dbf4-dependent kinases (Cdc28 and Cdc7 in *S. cerevisiae*, respectively) and the binding of the replication factors Cdc45 and GINS (forming the CMG complex). During these steps, the double hexamer is split, and each active ring is opened and re-closed around the leading ssDNA template to promote DNA unwinding as the CMG complex translocates during replication elongation (Deegan and Diffley, 2016; Labib, 2010; Li and O'Donnell, 2018). Remarkably, the number of MCM complexes loaded at chromatin exceeds the number of replication origins and ORCs by a factor of 10 to 50. These excess MCM complexes spread over the surrounding chromatin where they have different roles in response to replication stress (Das and Rhind, 2016; Hyrien, 2016).

We show in the yeast *S. cerevisiae* that Rad51 and Rad52 display dynamic interactions with MCM in a nuclease-insoluble scaffold in which MCM, but not Rad51, seems to be directly bound to DNA. Binding to this scaffold is regulated through the cell cycle in a DNA-damage-dependent manner. In unperturbed cells, they accumulate in G1 and are released during S phase; however, in the presence of replication stress, Cdc7 prevents the release of Rad51 and Rad52 from the scaffold, thus, maintaining their association with MCM. Functionally, the interaction between MCM and Rad51 promotes replication fork advance and ssDNA gap repair through non-recombinogenic mechanisms.

## RESULTS

### MCM interaction with Rad51 and Rad52 is regulated by the cell cycle, DNA damage, and Cdc7

To determine whether the interactions of MCM with Rad51 and Rad52 are conserved in *S. cerevisiae*, we performed coimmunoprecipitation (CoIP) experiments using a GFP-tagged allele of Mcm4, which did not affect cell growth. Rad51 coimmunoprecipitated with Mcm4, regardless of whether the lysates were pre-treated with the nucleases MNase I (Figure 1A) or benzonase (Figure 1B), indicating that the interaction was not mediated by DNA. This CoIP also showed an association of Mcm4 with Rad52 (Figure 1B). However, the absence of Rad52 did not prevent the formation of the MCM/Rad51 complex, despite Rad52 being required for Rad51 binding to replication forks and ssDNA lesions (González-Prieto et al., 2013). In fact, the amount of precipitated Rad51 increased in *rad52Δ* cell extracts (Figure 1A,

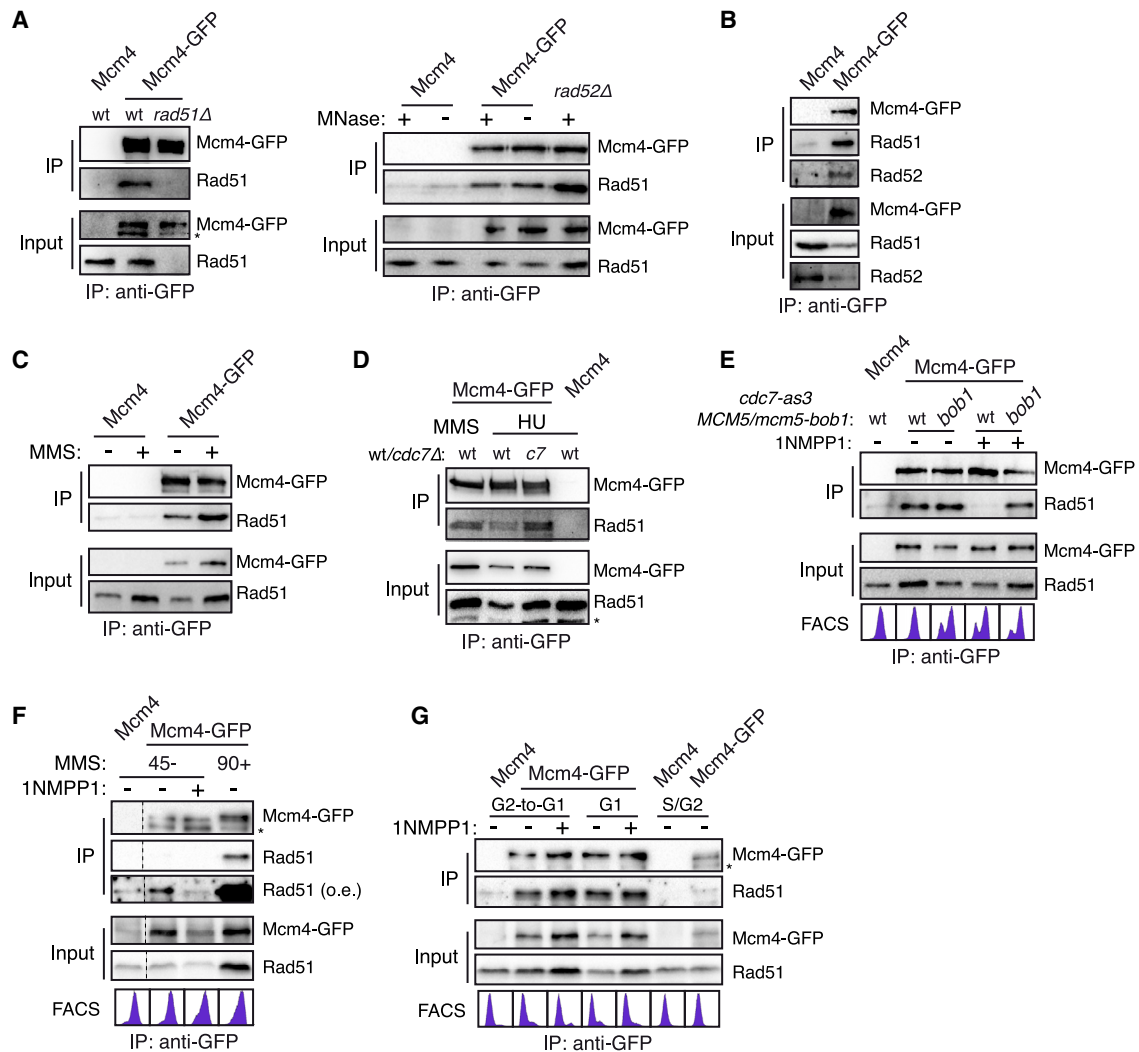
right panel), suggesting dynamic interactions between Rad51, Rad52, and MCM. Next, we performed the CoIP analysis in asynchronous cultures treated or not for 2 h with either MMS or hydroxyurea (HU), which causes deoxyribonucleotide triphosphate (dNTP) depletion. The MCM/Rad51 interaction was observed in the absence and presence of either DNA-damaging agent, although it was stronger in response to MMS (Figures 1C and 1D).

Cdc7 is a major regulator of MCM activity (Labib, 2010). To determine whether the interaction between Rad51 and Mcm4 requires the kinase activity of Cdc7, we used an allele of *CDC7* (*cdc7-as3*) that is sensitive to the ATP analog inhibitor 1NMPP1 (Wan et al., 2006). Because Cdc7 is essential for replication initiation, the inhibitor was added 30 min after G1 release, and the interaction was tested 60 min later. The experiment was performed in the presence of MMS to maintain cells with comparable cell-cycle profiles. The addition of 1NMPP1 prevented Rad51 from coimmunoprecipitating with Mcm4, indicating that the MCM/Rad51 association requires the continuous kinase activity of Cdc7 (Figure 1E). The requirement of Cdc7 for replication initiation can be bypassed by a mutation in *Mcm5* (*mcm5-bob1*) (Hardy et al., 1997). Likewise, Cdc7 activity was not required to maintain the MCM/Rad51 interaction both in *cdc7-as3 mcm5-bob1* treated with an inhibitor and in *cdc7Δ mcm5-bob1* cells in the presence of MMS or HU (Figures 1D, 1E, and S1A), suggesting that Cdc7 maintains the MCM/Rad51 interaction in response to DNA damage by acting upon the MCM complex. When these experiments were repeated in cells released in the absence of DNA damage, the MCM/Rad51 interaction was detected in unperturbed S/G2 cells only after over-exposure of the western blot (Figure 1F, lanes 2 and 3). Furthermore, the *mcm5-bob1* mutation was not sufficient to maintain the MCM/Rad51 interaction in the absence of DNA damage (Figure S1B). Rad51 also coimmunoprecipitated with Mcm4 in G1-arrested cells (Figure 1G), even though there is no Cdc7 activity in that phase (Weinreich and Stillman, 1999). Interestingly, Mcm4 is modified under conditions that promote its interaction with Rad51 (Figures 1F, 1G, and S1C), in part because of phosphorylation, as determined by phosphatase λ treatment (Figure S1D). Altogether, our CoIP analyses show an interaction between MCM and Rad51 in G1 that is lost during unperturbed S/G2; however, in the presence of replicative DNA damage, the kinase activity of Cdc7 maintains the interaction.

### The MCM/Rad51 complex localizes to a nuclease-insoluble nuclear scaffold

To determine the location of the MCM/Rad51 interaction, we first separated, by cell lysis and centrifugation, the soluble (supernatant) and insoluble (pellet containing the chromatin) fractions from cells growing asynchronously under unperturbed conditions; proper fraction was confirmed by western blots against histone H4 (insoluble fraction) and Pgk1 (soluble fraction) (Figure 2A, left panel). The amount of Rad51 interacting with Mcm4 was much greater in the insoluble, than in the soluble, fraction, even though both factors were more abundant in the soluble fraction (Figure 2A, right panel).

Nuclease digestion of the pellet solubilizes most of the chromatin; however, there is a residual nuclease-insoluble chromatin

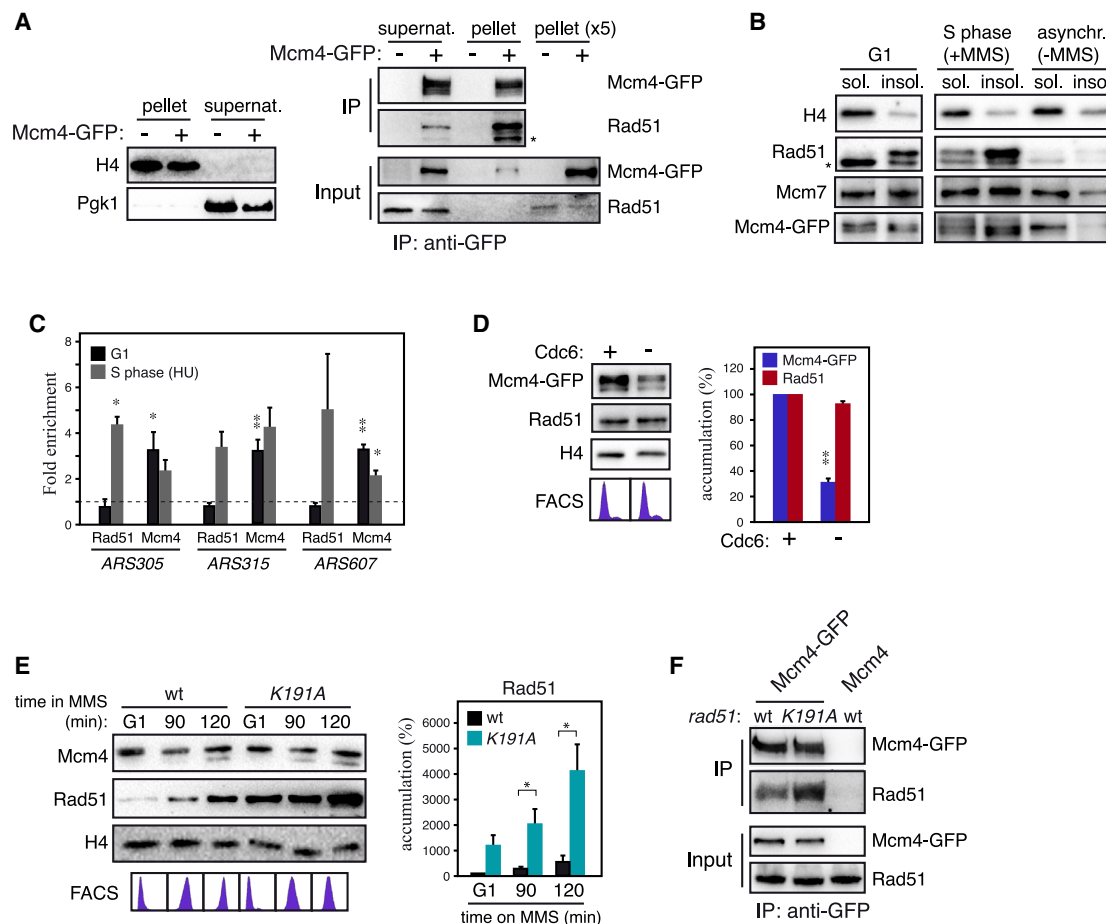


**Figure 1. MCM interacts with Rad51 and Rad52 through mechanisms regulated by cell cycle, DNA damage, and Cdc7**

(A) Rad51 interacts with Mcm4, independent of Rad52. CoIP was performed in asynchronous cultures.  
 (B) Mcm4 interacts with Rad52. Interactions were detected regardless of whether extracts had been treated or not with MNase I (A) or benzonase (B). CoIP was performed in asynchronous cultures.  
 (C) Mcm4 interacts with Rad51, both with and without 0.015% MMS for 2 h. CoIP was performed in asynchronous cultures.  
 (D) Mcm4 interacts with Rad51 in *cdc7Δ mcm5-bob1* and wild-type cells treated with 0.2 M HU for 2 h. Wild-type cells treated with 0.025% MMS for 2 h were included as controls. CoIP was performed in asynchronous cultures.  
 (E) The Mcm4/Rad51 interaction in cells released into MMS during the S phase depends on Cdc7 kinase activity, and *mcm5-bob1* bypasses that requirement. *cdc7-as3* cells were synchronized in the G1, released into the S phase in the presence of 0.015% MMS for 90 min, and treated with 15  $\mu$ M 1NMPP1 30 min after G1 release.  
 (F) Mcm4/Rad51 interaction is eliminated during an unperturbed S phase. G1-synchronized cells were released in the absence or presence of 0.015% MMS for 45 and 90 min, respectively, and were treated or not with 15  $\mu$ M 1NMPP1 30 min after G1 release (–MMS). An over-exposure (o.e.) of the Rad51 gel is shown.  
 (G) The Mcm4/Rad51 interaction occurs in G1, independent of Cdc7 activity. The analysis was performed in G1-arrested cells, either coming from G1 and released into fresh medium with  $\alpha$ -factor and 15  $\mu$ M 1NMPP1 60 min later (G2-to-G1) or maintained in G1 for 30 min with and without an inhibitor (G1). Cells synchronized in G1 and released into fresh medium for 60 min were included as controls (S/G2).  
 Physical interactions of MCM with Rad51 and Rad52 were determined by immunoprecipitation of Mcm4-GFP and western blot analyses. All experiments were repeated at least twice with similar results. The asterisks indicate a degradation product. Dashed lines indicate spliced images.

fraction that precipitates with the scaffold protein Top2, the checkpoint protein Rad53, and the replication factors ORC, Dbf4, and Sgs1 (Frei and Gasser, 2000; Pasero et al., 1999). To determine in what fraction Rad51 and MCM were present under the conditions in which they interact, the pellets obtained af-

ter fractionation of G1-arrested and MMS-released cells were digested with benzonase, and nuclease-soluble and insoluble fractions were analyzed by centrifugation, followed by western blot. A control with undamaged cells from an asynchronous culture was included for comparison. As expected, nuclease



**Figure 2. The MCM/Rad51 complex localizes to a nuclease-insoluble nuclear scaffold**

(A) Immunoprecipitation of Rad51 with Mcm4-GFP using MNase-I-treated lysates from insoluble (pellet) and soluble (supernatant) fractions from asynchronously growing cells (right panels). Similar cell equivalents of the insoluble and soluble fractions were loaded for the fractionation controls (left panels) and CoIP samples (right panels). For the input of the CoIP samples, one and five equivalents of the soluble fraction were loaded for the insoluble fraction. The asterisk indicates an unspecific band as determined in a *rad51Δ* strain (data not shown). The experiments were repeated twice with similar results.

(B) Rad51 and MCM accumulate in a nuclease-insoluble fraction both in cells arrested in G1 and in cells released in the presence of 0.025% MMS for 120 min. An unperturbed asynchronous culture was included as a control. The pellet obtained after cell fractionation was further fractionated into nuclease-soluble and -insoluble fractions by treatment with benzonase and centrifugation, and the amount of histone H4, Rad51, Mcm7, and Mcm4-GFP was determined by western blot. The asterisk indicates an unspecific band as determined in a *rad51Δ* strain (data not shown). The experiments were repeated twice with similar results.

(C) MCM/Rad51 interaction is prevented at the pre-RC. Rad51 and Mcm4-GFP accumulation at replication origins in cells synchronized in G1 and released into the S phase for 1 h in the presence of 0.2 M HU. The means  $\pm$  SEM of three independent ChIP experiments are shown. One and two asterisks represent significant differences with the untagged strain; \* $p < 0.05$  and \*\* $p < 0.01$ ; two-tailed Student's t test.

(D) *CDC6* expression is required for MCM, but not for Rad51, accumulation in the insoluble fraction in G1, as determined in *GAL::CDC6* cells synchronized in G1 under conditions of Cdc6 repression. The amount of Mcm4 and Rad51 (normalized to H4) left in the insoluble fraction in Cdc6-depleted cells relative to Cdc6-expressing cells is shown. The means  $\pm$  SEM from three independent experiments are shown. Two asterisks represent a significant difference with the Cdc6-expressing strain; \*\* $p < 0.01$ ; one-sample t test.

(E) The DNA-binding domain of Rad51 is dispensable for its accumulation in the insoluble fraction. G1-synchronized *rad51-K191A* and wild-type cells were released into the S phase in the presence of 0.025% MMS. The binding of MCM and Rad51 to the insoluble fraction was determined by cell fractionation and western blot analyses (see Figure S11 for fractionation controls). Histone H4 was used to normalize the amount of each protein. The mean  $\pm$  SEM from five independent experiments are shown. One asterisk represents a significant difference; \* $p < 0.05$ ; two-tailed Student's t test.

(F) MCM/Rad51 interaction augments in a *rad51-K191A* mutant. CoIP was performed in asynchronous cultures treated with 0.025% MMS for 2 h. Similar results were obtained in three independent experiments.

treatment released most of the chromatin, as determined by the enrichment of histone H4 in the soluble fraction (Figure 2B). Rad51 remained in the nuclease-insoluble pellet. Indeed, the helicase MCM, which is released to the solubilized fraction in un-

perturbed asynchronous cultures (Figure 2B) (Liang and Stillman, 1997; Pasero et al., 1999), was also partially retained in the nuclease-insoluble pellet in G1- and MMS-treated S-phase cells (Figure 2B).

These results suggest that MCM and Rad51 accumulate in a nuclease-insoluble fraction under conditions that promote their interaction. To determine whether MCM and Rad51 bind to DNA in this fraction, we first used chromatin immunoprecipitation (ChIP) to probe their accumulation at three different replication origins in G1-arrested cells. As a positive control, their presence was also monitored at forks arrested in the early S phase by means of an HU treatment. Whereas both Mcm4 and Rad51 accumulated at stalled forks in the S phase, only Mcm4 was detected at the origins in G1 (Figure 2C). Neither Mcm4 nor Rad51 significantly accumulated at three different positions close to *ARS305* in G1 (Figure S1E). These results indicate that the MCM/Rad51 interaction does not occur at the pre-RC in G1.

To determine whether MCM was bound to DNA in the insoluble fraction, we analyzed its accumulation in the absence of Cdc6, essential for MCM loading onto DNA (Cocker et al., 1996). For that, *GAL1::CDC6* cells were synchronized in G1 in the presence of glucose to repress the expression of Cdc6. Under these conditions, cells were unable to exit G1. This approach reduced the amount of Mcm4 in the insoluble fraction to ~30% (Figures 2D and S1F), suggesting that most MCM in this fraction is bound to DNA. Notably, the accumulation of Rad51 was unaffected under these conditions (Figure 2D). Indeed, neither Rad51 nor Rad52 was required for Mcm4 accumulation in this fraction in G1 (Figures S1G and S1H).

To test whether Rad51 was bound to DNA in the insoluble fraction, we employed a mutant (*rad51-K191A*) defective in DNA binding (Fung et al., 2006; Van Komen et al., 2000; Li et al., 2007; Morgan et al., 2002; Sung and Stratton, 1996) and followed its accumulation both in G1 and during S phase in the presence of MMS. The amount of Rad51-K191A in the insoluble fraction was strongly increased as compared with the wild-type protein (Figures 2E and S1I), suggesting that Rad51 is preferentially bound to a nuclear scaffold in a DNA-independent manner and that Rad51 is dynamically exchanged between that scaffold and DNA. Finally, we observed that the increase of Rad51 at that nuclear scaffold in the *rad51-K191A* mutant augmented its interaction with MCM (Figure 2F), further supporting the dynamism of these interactions. Altogether, these results suggest that MCM and Rad51 are independently recruited and interact dynamically in a nuclease-insoluble nucleoprotein scaffold. Although Rad51 associates to that compartment in a DNA- and MCM-independent manner, MCM seems to be bound to DNA.

### Cell cycle, DNA damage, and Cdc7 regulate the accumulation of MCM, Rad51, and Rad52 in the nuclear scaffold

To better understand the regulation of the MCM/Rad51/Rad52 interactions, we followed their binding to the insoluble fraction under various conditions (Figure 3; see Figure S2 for fractionation controls). We first analyzed their binding during the cell cycle under unperturbed conditions. As previously reported, Mcm4 accumulated in the insoluble fraction in G1 and was transiently lost during the S phase (Figure 3A) (Aparicio et al., 1997; Liang and Stillman, 1997; Tanaka et al., 1997). Rad51 and Rad52 displayed a similar pattern of binding during the cell cycle (Figures 3A and S3A). That reduction in binding to the insoluble frac-

tion during replication was not due to lower levels of total protein (Figure S3B).

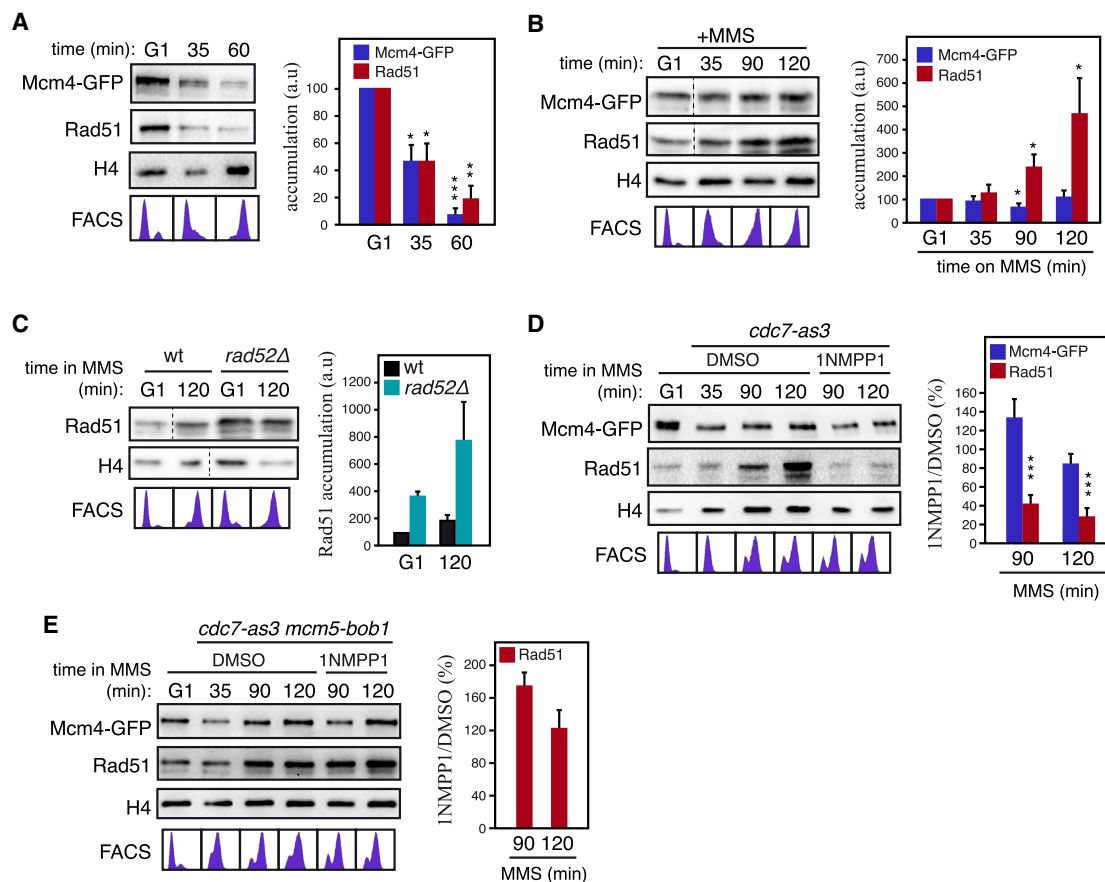
As we anticipated in Figure 2E, MCM, Rad51, and Rad52 were not released from the insoluble fraction during S/G2 in the presence of MMS; indeed, the amount of Rad51 and Rad52 increased with time (Figures 3B and S3A). It is worth noting that Mcm4 remained bound, even at very late times in MMS, during which bulk DNA was largely replicated (Figure 3B, FACS profiles), indicating that its maintenance was not due to replication inhibition.

These results suggest a functional connection between the physical interactions of MCM with Rad51 and Rad52 and their binding to the insoluble fraction. To further explore that possibility, we studied the genetic requirements of that binding. Fractionation analyses showed that Rad52 is not required for Rad51 binding to the insoluble fraction; indeed, the amount of Rad51 at that fraction increased in *rad52Δ* (Figure 3C). Next, we studied the role of Cdc7 in cells released into the S phase in the presence of MMS. Rad51 and Rad52 were lost from the insoluble fraction in the absence of Cdc7 kinase activity (Figures 3D and S3C), and that phenotype was suppressed by the *mcm5-bob1* mutation (Figure 3E). In the case of MCM, the results were not conclusive. MCM was maintained in the insoluble fraction after Cdc7 inhibition in most kinetics (Figure 3D); however, it was released to the same extent as Rad51 in some cases (5 of 16; see Figure S3D), suggesting a partial requirement of Cdc7 for a stable binding of MCM to the insoluble fraction. Overall, our results suggest that the formation of an MCM/Rad51/Rad52 complex is mechanistically linked to the binding of their components to a nuclear scaffold.

### Cdc7 is required to maintain Rad51 in the insoluble fraction, independent of the MCM/Rad51 interaction

Restricting Rad52 expression to G2/M prevents Rad51 from binding to MMS-induced ssDNA gaps and interferes with its repair (González-Prieto et al., 2013), suggesting that the recruitment of the recombination proteins to ssDNA gaps during DDT must occur during the S phase. This is a substantial mechanistic distinction from double-strand break (DSB)-induced HR; in which, Rad51 and Rad52 are recruited to ssDNA independent of DNA replication (Alabert et al., 2009; Barlow and Rothstein, 2009). To better understand this process, we performed a search for *rad51* mutants sensitive to MMS and resistant to ionizing radiation (IR). We obtained five *rad51* alleles in centromeric plasmids that were tested for their ability to interact with MCM as a putative mechanism to explain the coupling between replication and Rad51 binding to ssDNA gaps during DDT. One of them (*rad51m*) was defective in its interaction with MCM (Figure S4A). To rule out the possibility that these phenotypes were due to its plasmid location, we reproduced them in a strain in which chromosomal *RAD51* was replaced with the *rad51m* allele (Figures 4A and 4B).

The kinetics of MCM and Rad51 binding to the insoluble fraction, both in the absence and presence of MMS, were similar in *rad51m* and wild-type cells (Figures 4C, 4D, S4B, and S4C). These results confirm that the MCM/Rad51 interaction is not required for its recruitment to the insoluble fraction (Figure 2D) and suggest that Cdc7 maintains the MCM/Rad51 interaction



**Figure 3. Cell cycle, DNA damage, and Cdc7 regulate the accumulation of MCM, Rad51, and Rad52 in the insoluble fraction**

(A) Mcm4 and Rad51 accumulate in the insoluble fraction in G1 and are released during the S phase.

(B) Mcm4 and Rad51 are maintained in the insoluble fraction during the S phase in the presence of 0.025% MMS.

(C) Rad52 is not required for Rad51 binding to the insoluble fraction in G1 and MMS-released cells.

(D) Inhibition of Cdc7 kinase activity prevents the maintenance of Rad51 in the insoluble fraction in response to DNA damage. G1-synchronized cells were released into the S phase in the presence of 0.025% MMS; they were treated with 15  $\mu$ M 1NMPP1 35 min after G1 release.

(E) The *mcm5-bob1* mutation bypasses the requirement of Cdc7 for Rad51 maintenance in the insoluble fraction under replicative stress. G1-synchronized cells were released into the S phase in the presence of 0.025% MMS; they were treated with 15  $\mu$ M 1NMPP1 35 min after G1 release.

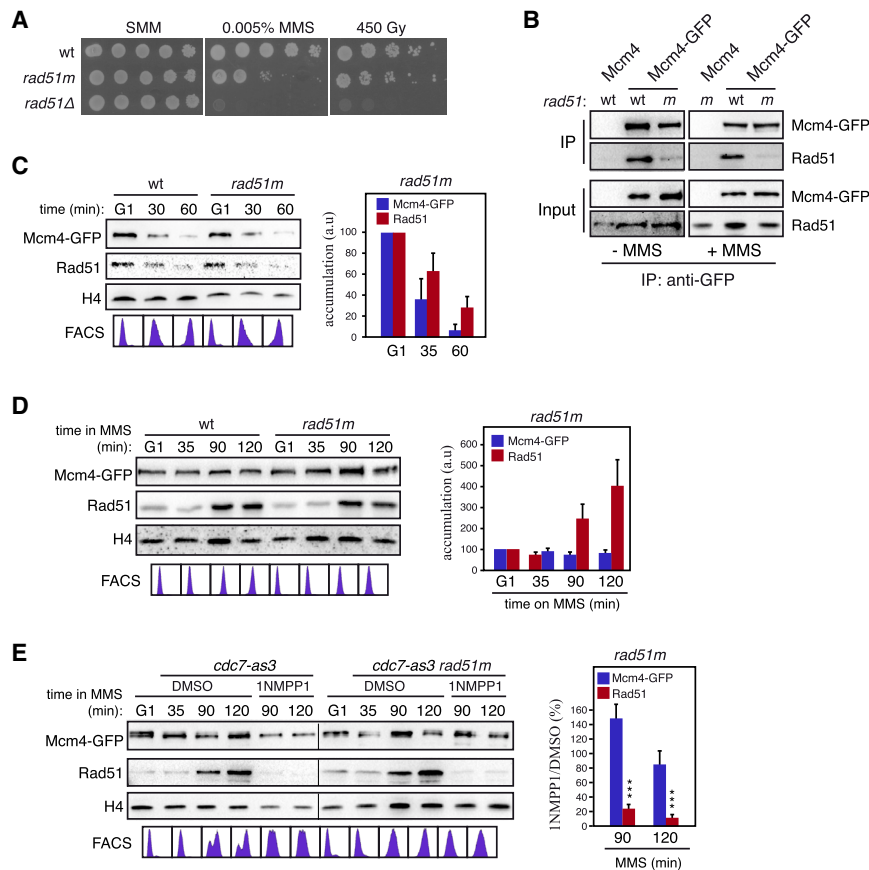
MCM and Rad51 binding to the insoluble fraction was determined by cell fractionation and western blot. See Figure S2 for fractionation controls (A–E). Dashed lines indicate spliced images. Histone H4 was used to normalize the amount of each protein. The mean of 4 (A), 11 for Mcm4-GFP and 14 for Rad51 (B), 2 (C and E) and 11 for 90-min, and 6 for 120-min (D) independent experiments are plotted. Either the SEM (A, B, and D) or range (C and E) are shown. Values are normalized to G1 (A–C) or to the absence of the inhibitor (D and E), taking as 100. Statistically significant differences relative to G1 (A and B) or to the absence of the inhibitor (D) are shown; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ; one-sample t test.

by preventing Rad51 from dissociating from the nuclear scaffold. In agreement with that, the binding of Rad51 to the insoluble fraction was reduced to a similar extent in *rad51m* and wild-type cells upon Cdc7 inhibition (Figures 4E and S4D). Therefore, Cdc7 is required to maintain Rad51 in the nuclear scaffold in the presence of replicative DNA damage but is, independent of the MCM/Rad51 interaction.

### The interaction between MCM and Rad51 facilitates the replication of damaged DNA

DNA content analyses revealed defects in replication fork progression in the presence of MMS in the *rad51m* mutant (Figure 4D, FACS profiles). To better characterize that phenotype, we followed DNA replication at shorter time intervals. The

*rad51m* mutant did not display replication defects under unperturbed conditions (Figures S5A and S5B); however, it displayed a gradual delay in the replication of MMS-damaged DNA, which became evident at late times after G1 release (Figure 5A). The absence of Cdc7 activity also affected replication fork progression, which was severely aggravated in *rad51m cdc7-as3* cells (Figure 4E, FACS profiles). Likewise, *rad51m* and *cdc7-as3* single mutants exhibited a delay in completing DNA replication and entering into the following cell cycle after MMS treatment, whereas the *rad51m cdc7-as3* double mutant was unable to complete replication (Figure 5B). These replicative defects were confirmed by pulse field gel electrophoresis (PFGE), in which only completely replicated chromosomes can enter into the gel (Figures 5C and S5C).



**Figure 4. Cdc7 is required to maintain Rad51 in the insoluble fraction, independent of the MCM/Rad51 interaction**

(A) MMS and IR sensitivity of *rad51m*, *rad51Δ*, and wild-type cells.

(B) The Rad51m mutant protein is defective in its interaction with MCM, as determined by CoIP from asynchronous cultures treated or not with 0.025% MMS for 2 h. The experiment was repeated twice with similar results.

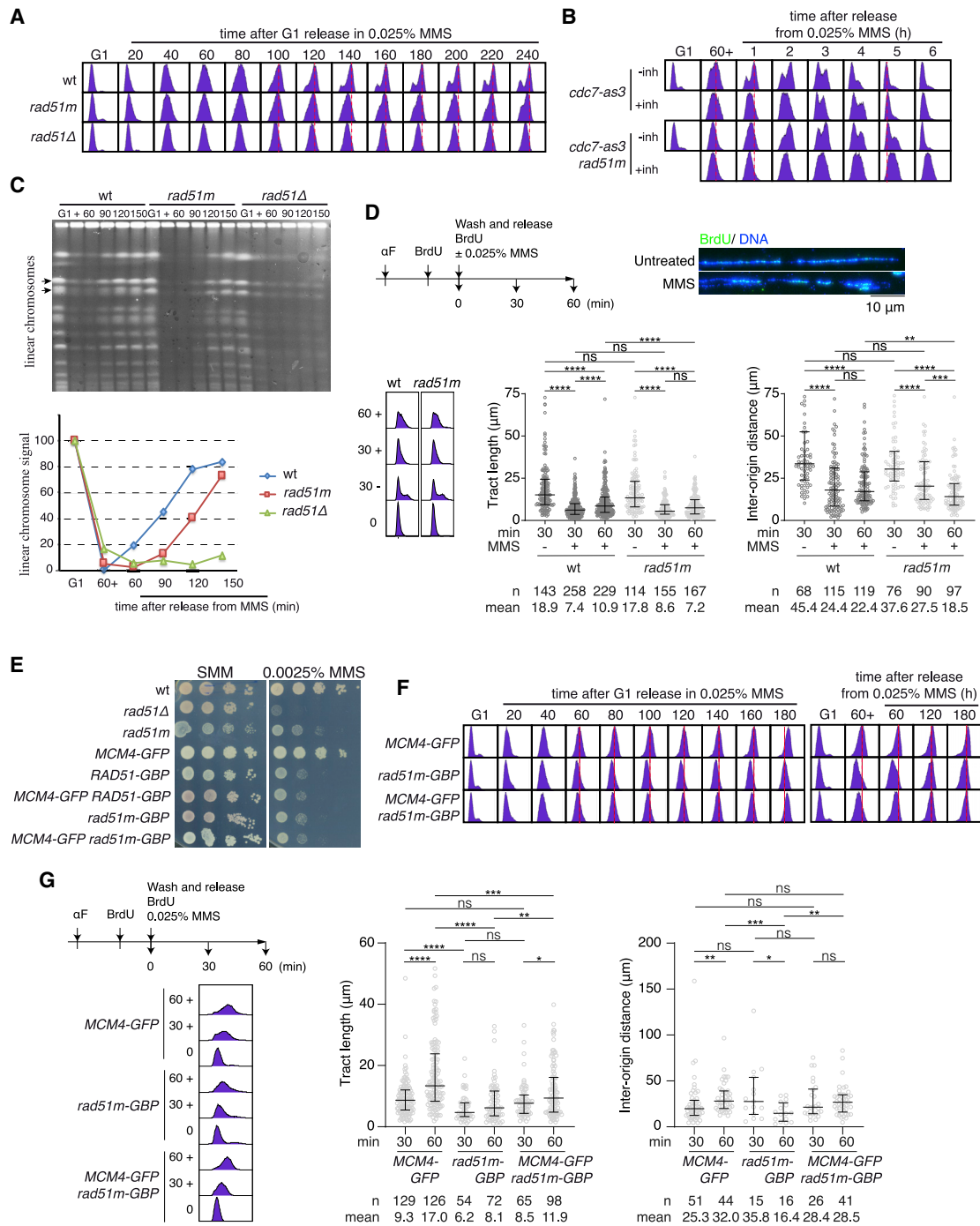
(C–E) Mcm4 and Rad51 accumulate in the insoluble fraction in G1 and are released during the S phase in *rad51m* cells (C), unless they are released in the presence of 0.025% MMS (D). (E) Inhibition of Cdc7 kinase activity prevents the maintenance of Rad51 in the insoluble fraction in *rad51m* cells released into 0.025% MMS and treated with 15 μM 1NMPPI1 35 min after G1 release. MCM and Rad51 binding to the insoluble fraction in *rad51m* cells released into 0.025% MMS and treated with 15 μM 1NMPPI1 35 min after G1 release. MCM and Rad51 binding to the insoluble fraction was determined by cell fractionation and western blot. See Figure S4B–S4D for fractionation controls. Histone H4 was used to normalize the amount of each protein. The mean of 2 (C), 6 (D), and 7 (E) independent experiments are plotted. Either the range (C) or SEM (D and E) are shown. Values are normalized to G1 (C and D) or the absence of inhibitor (E), taking as 100. Three asterisks represent a significant difference relative to the absence of inhibitor; \*\*\**p* < 0.001; one-sample *t* test.

Next, we used DNA fiber assays to analyze the replication fork speed (Figure 5D). We observed no significant difference in replication tract lengths between wild-type and *rad51m* cells 30 min after G1 release under unperturbed condition. Replication tract lengths in both strains decreased to a similar extent 30 min after release into MMS compared with the unperturbed conditions. However, although wild-type cells significantly increased replication tract lengths 60 min after G1 release into MMS, *rad51m* mutant cells were unable to accelerate DNA synthesis at the same time point (Figure 5D, left graph). Interestingly, we observed clusters of smaller replication tracts after 30 min in MMS, leading to a decrease in inter-origin distances (Figure 5D, right graph). This is probably due to firing of dormant origins upon replication stress, which has been described in mammalian cells (Ge et al., 2007). Although inter-origin distances remained unchanged in wild-type cells from 30 to 60 min after release into MMS, this parameter was further significantly reduced in the *rad51m* mutant (Figure 5D, right graph). These data suggest that the *rad51m* mutant is partially defective in supporting replication fork progression through MMS-damaged template but compensates for that defect by firing an increasing number of dormant origins. This might explain why the replication defect of *rad51m* is aggravated in the absence of Cdc7 activity (Figure 5B).

To demonstrate that the replication defects of the *rad51m* mutant were associated with its inability to interact with MCM,

we forced the MCM/Rad51 interaction in the *rad51m* mutant by co-expressing Mcm4 and Rad51m tagged with GFP and GBP (GFP-binding protein), respectively. Tagging Rad51 with GBP caused strong MMS sensitivity (Figure 5E), suggesting a loss of functionality in recombinational repair, as previously reported for Rad51-GFP (Waterman et al., 2019). Rad51m-GBP cells progressed more slowly than did Mcm4-GFP and wild-type cells in the presence of MMS (Figure 5F). Importantly, co-expression of Rad51m-GBP and Mcm4-GFP partially restored the speed of DNA replication (Figure 5F, left). Likewise, cells co-expressing Rad51m-GBP and Mcm4-GFP reached G2/M earlier than cells expressing only Rad51m-GBP did after resuming replication after MMS treatment (Figure 5F, right), suggesting that the interaction between MCM and Rad51 facilitates DNA replication in the presence of blocking DNA lesions.

DNA fiber analyses showed that the Mcm4-GFP chimera partially restores the Rad51m-GBP defect in replication through damaged DNA by accelerating the elongation rate as inferred from the increase in the replication tract length from 30 to 60 min (Figure 5G, left graph). This increase in fork progression was concomitant with a suppression in the firing of additional replication origins (Figure 5G, right graph). Of note, the Mcm4-GFP chimera partially restored the replication defects of the *rad51m-GBP* mutant but suppressed neither the MMS sensitivity nor the HR defect (Figures 5E and S5D), suggesting that the MCM/Rad51 complex does not require the recombinational function of Rad51 to promote replication fork progression.



**Figure 5. The interaction between MCM and Rad51 facilitates for DNA replication in the presence of MMS**

(A and B) Cell-cycle progression of the indicated strains synchronized in G1 and released in the presence of 0.025% MMS for different times (A) or released into 0.025% MMS for 1 h and then into fresh medium after MMS inactivation (B). In (B), cells were treated with 15  $\mu$ M 1NMP1P1 for 30 min after G1 release and every 2 h to keep Cdc7 inactive.  $\alpha$ -Factor was also added every 2 h to prevent cells from entering a new cycle.

(C) Cell-cycle progression determined by PFGE analysis of the indicated strains synchronized in G1, released into 0.025% MMS for 1 h, and then, into fresh medium after MMS inactivation. Quantified chromosomes are shown with arrows. The plot shows the mean signal relative to G1, taken as 100.

(D) Replication tract length (left) and inter-origin distance (right) in *rad51m* and wild-type cells growing under unperturbed and MMS conditions, as determined by DNA fiber analysis. G1-synchronized cells were incubated with BrdU for 15 min, washed twice, and released into the S phase in the presence of BrdU with or without 0.025% MMS. Bars represent the median and the 25th and 75th percentiles. Statistically significant differences according to Mann-Whitney tests are

(legend continued on next page)



**The *rad51m* mutant is proficient in HR but partially defective in the repair of MMS-induced ssDNA gaps at Rad52-associated DNA repair centers**

To determine whether the MCM/Rad51 interaction was relevant for MMS-induced DNA damage repair, we followed the resolution of MMS-induced Rad52 foci, which can be detected at the end of the S and G2 phases (González-Prieto et al., 2013). For that, G1-synchronized cells were released into the S phase in the presence of 0.033% MMS for 1 h and, then, into fresh medium after MMS inactivation to allow for DNA damage repair. About 50% of the *rad51m* and wild-type cells accumulated Rad52 foci 1–2 h after MMS inactivation (Figure S6A). Importantly, although the percentage of wild-type cells with Rad52 foci dropped to ~5% 4–5 h later (85% resolution), it only dropped to 25% in *rad51m* cells (45% resolution) (Figures 6A, left panel, and S6A). A similar defect was observed when we calculated the fluorescence signal at foci per cell, which integrates the percentage of cells that retain foci as well as the number and intensity of the foci (Figures 6A, right panel, and S6B). This result suggests that ssDNA lesions are not efficiently repaired in the *rad51m* mutant. To further confirm that, we followed the kinetics of replication protein A (RPA) foci formation and resolution. The RPA complex (formed by the Rfa1–3 subunits) covers ssDNA lesions, regardless of the mechanism of repair. The efficiency of ssDNA gap filling, inferred from both the percentage of cells with foci and the RPA signal per cell at the end of the time course relative to the peak during the time course, was severely compromised in the *rad51m* mutant (Figures 6B, S6C, and S6D); indeed, RPA foci resolution was affected even at a dose of 0.01% MMS (Figure S6E), which had no effect on *rad51m* viability (Figure S6F). The defect in RPA foci resolution was greater than the one observed with Rad52 foci, likely because of detection levels, as suggested by the fact that the peak of cells with RPA foci was approximately twice that of the peak of Rad52 foci (compare Figures 6A and 6C, left panels).

To determine whether the defect in DNA repair of the *rad51m* mutant was specific for MMS, we followed the formation and resolution of RPA foci in cells irradiated in G1 with 35 Gy and released into the S phase for different times (Figures 6C, S6G, and S6H). Although the repair was slower in the *rad51m* mutant than it was in the wild-type cells at the beginning of the kinetics (Figures 6C and S6H, right panels; compare the RPA signal at 2 h), the efficiency of gap repair was similar in both strains, as determined by the percentage of cells with foci and the signal intensity at the end of the kinetics (Figure 6C, left and right panels, respectively). Therefore, the *rad51m* mutant was defective in the repair of MMS-induced ssDNA gaps but not of IR-induced DSBs, even though the lethality induced with 35 Gy was greater than that induced with 0.033% MMS (Figure 6D). Notably, this defect in ssDNA gap filling was not associated with a significant loss of viability (Figure S6F).

To determine what molecular step was compromised in the *rad51m* mutant, we first tested the interaction of Rad51 with Rad52. The amount of Rad52 bound to Rad51 was strongly increased in the mutant as compared with that of the wild-type cells (Figure 6E), further supporting the dynamism of the MCM/Rad51/Rad52 interactions. Next, we analyzed Rad51 binding to MMS-damaged DNA by chromatin-endogenous cleavage (ChEC). In this approach, cells expressing a chimera of Rad51 fused to MNase I (Rad51-MN) are permeabilized with digitonin and treated with Ca<sup>2+</sup> ions for different time points to activate the nuclease; then, total DNA is analyzed by agarose gel electrophoresis. The rationale behind this approach is that the chimera will generate a detectable cut only if it is targeted to an uncut DNA fragment (González-Prieto et al., 2013, 2021). Although the binding of Rad51 to the insoluble fraction is independent of Rad52 (Figure 3C), Rad52 is essential for MMS-induced DNA cleavage by Rad51-MN (González-Prieto et al., 2013). Rad51m-MN and Rad51-MN displayed similar DNA digestion profiles (Figure 6F), suggesting that the interaction with MCM is not required for Rad51 binding to ssDNA lesions. Similar conclusions were obtained by analyzing Rad51 binding to HU-stalled replication forks in *rad51m* and wild-type cells by CHIP (Figure S6I).

The repair of MMS-induced DNA lesions by HR is associated with the formation of SCJs, which can be detected as X-shaped structures by two-dimensional (2D) electrophoresis in *sgs1Δ* cells that are defective in their dissolution (Liberi et al., 2005). The *sgs1Δ* and *sgs1Δ rad51m* cells displayed similar kinetics of X-shaped molecule accumulation (Figure 6G), indicating that the interaction with MCM is not required for the DNA strand-exchange activity of Rad51.

Finally, we measured HR using an unequal sister chromatid exchange (uSCE) system (Fasullo and Davis, 1987). For that, exponentially growing cells were treated either with MMS for 4 h or with IR, and the frequency of recombinants was determined before and after DNA damage. The *rad51m* mutant was not significantly affected in HR in response to MMS or IR (Figure 6H). Therefore, the interaction of Rad51 with MCM is not required for the recombinational repair of ssDNA lesions and DSBs. Altogether, these results indicate that the *rad51m* mutant is impaired in a non-recombinogenic mechanism of ssDNA gap filling.

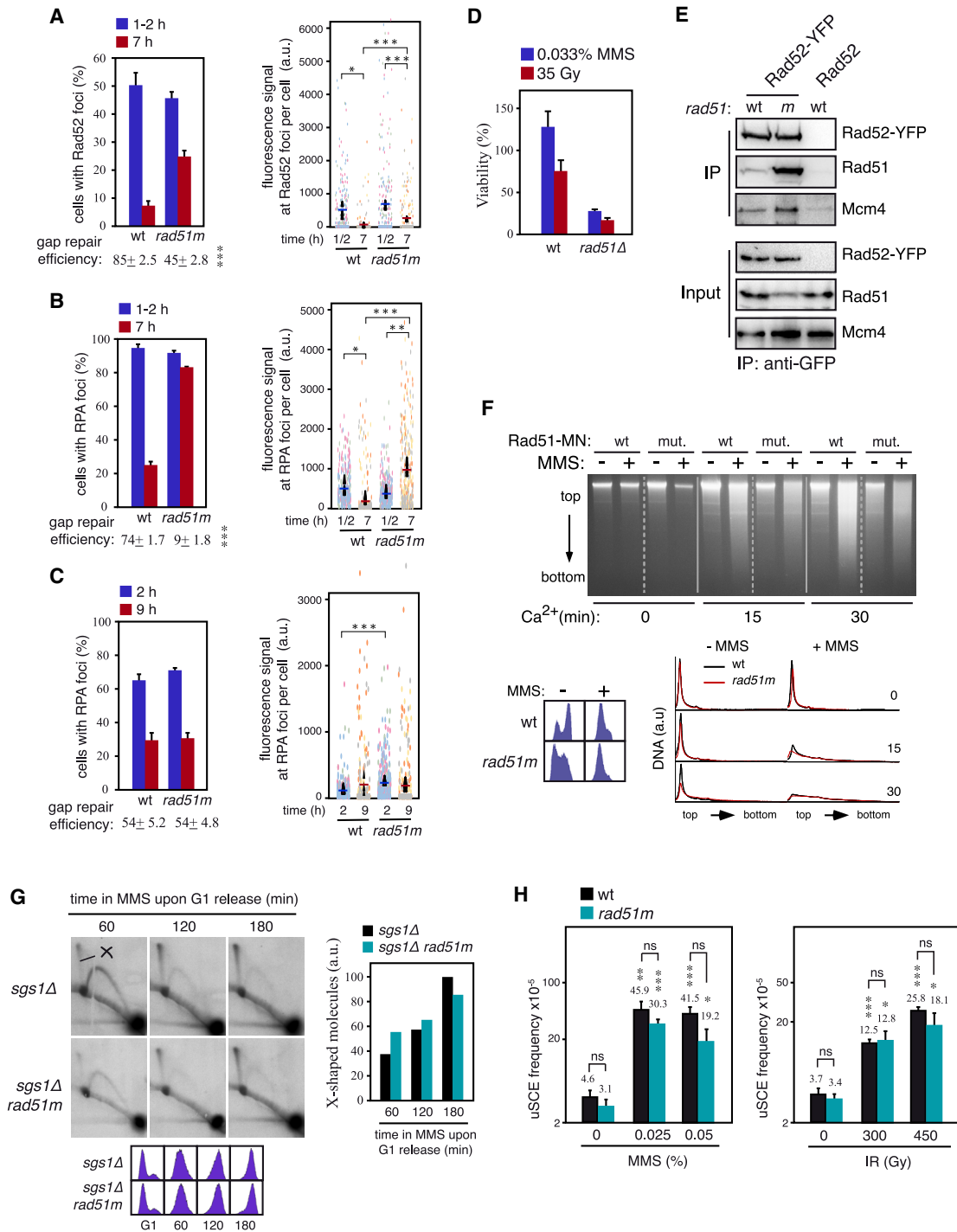
We could not use the Mcm4-GFP/Rad51m-GBP system to demonstrate that the *rad51m* defect in DNA repair is associated with the disruption of the MCM/Rad51 interaction because the Rad51m-GBP chimera is defective in HR. Because the *rad51m* allele contains five amino acid substitutions (Figure S7A), we decided to determine which one is responsible for disrupting the MCM/Rad51 interaction. The *rad51-C159R* mutant displayed a similar defect in MCM binding as the *rad51m* mutant had (Figure S7B). In comparison with *rad51m*,

shown; \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001. Cell cycle profiles and representative images of DNA fibers stained with antibodies against BrdU (green) and single-stranded DNA (blue) from wild-type cells 30 min after G1 release are shown.

(E) MMS sensitivity of the indicated strains.

(F) Cell-cycle progression in the presence (left) and after treatment for 1 h with 0.025% MMS (right) of cells expressing Mcm4-GFP, Rad51m-GBP, or both.

(G) Replication tract length (left) and inter-origin distance (right) in *MCM4-GFP*, *rad51m-GBP*, and *MCM4-GFP rad51m-GBP* cells growing in the presence of 0.025% MMS as indicated in (D). The experiments were repeated twice with similar results.



**Figure 6. The *rad51m* mutant is proficient in HR but partially defective in the repair of MMS-induced ssDNA gaps at Rad52-associated DNA repair centers**

(A–C) DNA repair efficiency of *rad51m* and wild-type cells, as determined by analyzing the formation and resolution of MMS-induced Rad52-YFP (A), MMS-induced Rfa1-YFP (B), and IR-induced Rfa1-YFP (C) foci. In (A) and (B), G1-synchronized cells were released into the S phase in the presence of 0.033% MMS for 1 h and then into fresh medium after MMS inactivation. In (C), cells were synchronized in G1, irradiated with 35 Gy, and released into the S phase. Complete kinetics are shown in Figures S6A, S6C, and S6G. The percentage of cells with foci (left panels) and the fluorescent signal at foci per cell (right panels) at 1/2 (peak) and 7/9 hours from the time course was determined with MetaMorph software. The efficiency of ssDNA gap repair was calculated as (100 – percentage of cells with foci at the end of the time course × 100)/maximal percentage of cells with foci during the time course. Color dots represent independent experiments.

(D) Cell viability of strains treated with either 0.033% MMS for 1 h or 35 Gy as determined by the number of colony-forming cells before and after DNA damage.

(legend continued on next page)

the *rad51-C159R* mutant was less sensitive to MMS (Figure S7C), which might explain the mild sensitivity of the *rad51m* mutant to acute doses of both MMS and IR (Figure S6F). However, and importantly, both mutants displayed similar defects in the repair of MMS-induced Rad52 foci (Figures S7D). Therefore, the disruption of the MCM/Rad51 interaction is genetically linked to a point mutation that causes a non-recombinogenic defect in MMS-induced DNA damage repair.

## DISCUSSION

Here, we have studied the regulation and biological meaning of the physical interactions between the helicase MCM and the recombination factors Rad51 and Rad52 in *S. cerevisiae*. We observed that Rad51 and Rad52 display dynamic interactions with MCM complexes located outside of the replication origins and forks in a nuclease-insoluble scaffold in which Rad51 accumulates in a MCM- and DNA-binding-independent manner. These interactions are detected in G1 and are lost in S/G2, unless DNA replication occurs in the presence of replicative DNA lesions; in this case, Cdc7 maintains the integrity of the MCM/Rad51/Rad52 complex by preventing the release of Rad51/Rad52 from the scaffold. Functionally, the MCM/Rad51 interaction facilitates the fork advance and ssDNA repair through non-recombinogenic activities.

### Cell cycle, DNA damage, and Cdc7 regulate the physical interactions between Rad51/Rad52 and MCM in a nuclease-insoluble scaffold

MCM, but not Rad51, was detected at replication origins in G1, whereas both factors accumulated at the forks (Figure 2C). This indicates that the MCM/Rad51 association is prevented at the pre-RC and that Rad51 binds to the fork once replication is initiated. The interaction of MCM with Rad51 and Rad52 was rarely detected in an unperturbed S phase (Figure 1F), suggesting that it is also prevented at the replication forks. Indeed, Rad51 requires Rad52 to bind to the fork (González-Prieto et al., 2013) but not to MCM (Figures 1A). This raises a question about the location and molecular nature of these interactions. Our fractionation analysis showed that Rad51 and MCM accumulate and interact with each other in a nuclease-insoluble fraction (Figures 2A and 2B). In principle, this fraction might be nuclease-insoluble chromatin. However, Rad51 requires neither its DNA binding activity nor MCM to accumulate in that fraction or to interact with MCM (Figures 2D, 2F, 4C, and 4D), indicating that Rad51 does not associate directly with DNA. In contrast to Rad51, the amount of MCM is strongly

reduced in the insoluble fraction if its loading onto DNA is impaired by repressing Cdc6 expression (Figure 2D). This suggests that MCM binds to nuclease-insoluble chromatin, where it interacts with Rad51. These data point to a model in which Rad51 and Rad52 interact with DNA-bound MCM in a nucleoprotein scaffold (Figure 7). These interactions would be highly dynamic considering their enrichment in the absence of some of the interacting partners: Rad51 with both MCM and the scaffold in *rad52Δ* (Figures 1A and 3C) and *rad51-K191R* mutants (Figures 2E and 2F), and Rad51 with Rad52 in the *rad51m* mutant (Figure 6E).

MCM helicases are assembled at every cell cycle during mitosis using newly synthesized subunits (Braun and Breeden, 2007), loaded onto replication origins at the end of mitosis and G1 (Aparicio et al., 1997; Donovan et al., 1997; Liang and Stillman, 1997; Tanaka et al., 1997), and prevented from binding to chromatin during the rest of the cell cycle (Nguyen et al., 2001). Because we did not detect Rad51 at the replication origins, the association between Rad51 and MCM may involve an excess of helicases that are loaded in G1 at replication origins and spread to the surrounding chromatin. However, genome-wide analyses have shown that MCM is concentrated at replication origins (Das et al., 2015; Wyrick et al., 2001), suggesting that this MCM/Rad51/nucleoprotein scaffold may not be associated with specific DNA positions and/or is not detected by those genomic approaches as a consequence of their insolubility. Alternatively, MCM might accumulate and interact with Rad51 and Rad52 outside of the chromatin fiber, considering that ~30% of Mcm4 remains in the insoluble fraction after repressing Cdc6. Future biochemical and genomic analyses will be required to address this point.

Remarkably, this nuclease-insoluble fraction is characterized by the presence of Top2, Rad53, Sgs1, the ORC, and the Cdc7-regulator Dbf4 (Frei and Gasser, 2000; Pasero et al., 1999), which are involved in replication fork stability under replication stress (Branzei and Foiani, 2010). In mammalian cells, replication origins and components of the pre-RC, including MCM, are associated during G1 and early S phases with the nuclear scaffold, defined as a salt-resistant or nuclease-insoluble nuclear fraction (Hesketh et al., 2015; Wilson and Coverley, 2013). According to those data, the DNA would be spooled through static replication machinery, and the newly synthesized DNA would be extruded as two loops (Wilson and Coverley, 2013). In a hypothetical similar frame, the yeast MCM/Rad51/Rad52 complexes would remain located near the forks during replication for assistance under conditions of replicative stress (Figure 7).

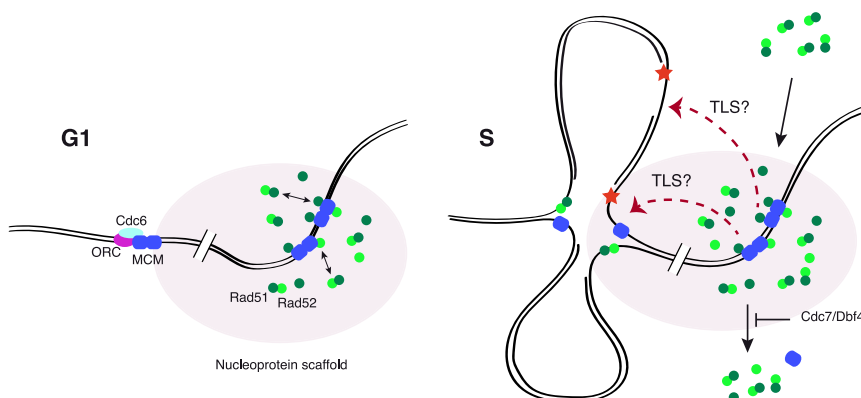
(E) The amount of Rad51 that interacts with Rad52 augments in the *rad51m* mutant. ColP was performed in asynchronous cultures treated with 0.025% MMS for 2 h. The experiments were repeated twice with similar results.

(F) The Rad51m mutant protein is proficient in binding to replicative ssDNA lesions, as determined by ChEC analysis of exponentially growing cells incubated with 0.05% MMS for 2 h. Total DNA from cells permeabilized and treated with Ca<sup>2+</sup> for different times is shown, as well as the DNA content and DNA digestion profiles. The experiments were repeated twice with similar results.

(G) The Rad51m mutant protein is proficient in the SCJ formation, as determined by 2D gel analysis of X-shaped molecules in cells synchronized in G1 and released in the presence of 0.033% MMS. The amount of X-shaped molecules (spike), relative to the total amount of molecules, with the highest value set at 100, is shown. The experiments were repeated twice with similar results.

(H) uSCE frequency in *rad51m* and wild-type cells upon exposure to increasing doses of MMS for 4 h (left) and IR (right).

The means ± SEM of three independent experiments are shown. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001; unpaired two-tailed Student's t test (A–D and H).



**Figure 7. The interaction of MCM with Rad51 facilitates replication fork advance and ssDNA gap repair in the presence of MMS**

Rad51 and Rad52 interact physically and dynamically with MCM in a nuclease-insoluble scaffold where Rad51 accumulates in a MCM- and DNA-binding-independent manner. Rad51/Rad52-interacting MCM helicases are loaded onto DNA in G1 and spread out of replication origins, although we cannot formally rule out that the interactions involve MCM molecules located outside of the DNA. MCM, Rad51, and Rad52 accumulate in G1 and are released during unperturbed replication. In the presence of replication-blocking lesions, MCM remains and Rad51/Rad52 augments through the recruitment of additional molecules and the kinase activity of Cdc7, which

prevents Rad51/Rad52 (and partially, MCM) from releasing the scaffold. These physical interactions help stressed replication forks by facilitating their advance and the filling of ssDNA gaps by non-HR mechanisms, which might include TLS (dashed arrows).

The interactions of Rad51 with both MCM and the nuclease-insoluble fraction during replicative DNA damage require Cdc7 activity (Figures 1E and 3D), suggesting that they are mechanistically linked. In this regard, the fact that the Rad51m mutant binds to the insoluble fraction (Figures 4C and 4D) demonstrates that the integrity of the MCM/Rad51 complex is not required for the binding of Rad51 to that scaffold. Rather, it suggests that the binding of Rad51 to that scaffold is a pre-requisite for its interaction with MCM and that the Cdc7 activity helps to maintain Rad51 and Rad52 in that scaffold under replicative stress. How Cdc7 performs that task is currently unknown, but the fact that Rad51 and Rad52 fall off the insoluble fraction after inactivating Cdc7 in cells that have already triggered early origins and progressed through the S phase (Figures 3D and S3C) suggests that it is not associated with its replication-initiation function. The Cdc7 requirement to maintain Rad51 in the insoluble fraction was observed even in the *rad51m* mutant (Figure 4E), indicating that the mechanism by which Cdc7 facilitates Rad51 binding to this nuclear fraction is independent of the MCM/Rad51 interaction. However, the requirement of Cdc7 activity for Rad51 binding to the insoluble fraction was bypassed by a mutation in a MCM complex subunit (*mcm5-bob1*) (Figure 3E). This suggests that Cdc7 performs this task by acting upon MCM, likely by generating a structural context that facilitates Rad51 binding; actually, this function might not be specific for Rad51 and Rad52, as suggested by the requirement of Cdc7 to maintain the replisome-associated factor Tof1 in the insoluble fraction and the bypass of this requirement by the *mcm5-bob1* mutation (Bastia et al., 2016). The *mcm5-bob1* mutation causes a conformational change in the MCM helicase, which rescues the lethality associated with the lack of Cdc7 (Hoang et al., 2007). Mcm2 phosphorylation by Cdc7 weakens its interaction with Mcm5, suggesting a potential mechanism for helicase opening and ssDNA extrusion during replication initiation that would be mimicked by the Mcm5-bob1 protein (Bruck and Kaplan, 2015). However, it is unlikely that Cdc7 facilitates Rad51 binding to the insoluble fraction by generating a local accumulation of ssDNA because the DNA binding activity of

Rad51 is dispensable for its binding to this fraction. Alternatively, MCM phosphorylation might facilitate the formation of a nucleoprotein scaffold or phase-separated liquid compartment as those reported recently at DSB repair centers (Kilic et al., 2019; Miné-Hattab et al., 2021)

#### Functional role of the MCM/Rad51 interaction in replication fork advance and ssDNA repair

A remarkable finding of this work is the cell-cycle kinetics of Rad51 and Rad52 binding to the nuclear scaffold: they accumulate in G1 and are released during the S phase, even though HR is inactive in G1 and active in the S phase (Heyer et al., 2010). This kinetics parallels that of the helicase MCM (Aparicio et al., 1997; Liang and Stillman, 1997; Tanaka et al., 1997); indeed, MCM, Rad51, and Rad52 also display similar patterns of binding to this scaffold in the presence of MMS, remaining bound to damaged DNA (Figures 3A, 3B, and S3A). As previously discussed, this kinetics of binding might facilitate physical interactions between Rad51 and Rad52 with MCM helicases located outside of pre-RCs and forks. This population of helicases vastly outnumbers the number of replication origins (Donovan et al., 1997), raising a question about their functionality that has been partially resolved by their role in activating dormant origins under replicative stress (Ge et al., 2007; Ibarra et al., 2008; Maki et al., 2011). We have ruled out a role for the MCM/Rad51 interaction in the activation of dormant origins, which actually are triggered in the *rad51m* mutant. We propose that Rad51 and Rad52 aggregate, together with excess MCM helicases (and likely additional factors), at specific nucleoprotein scaffolds in G1 for replication assistance. These aggregates would be removed during the S phase under unperturbed conditions, likely to avoid the toxicity of Rad52/Rad51/MCM-DNA interactions (Shah et al., 2010). In response to replicative DNA damage, Cdc7 would maintain those physical interactions to assist stressed replication forks by facilitating their advance and the repair of the stretches of ssDNA generated during lesion bypass (Figure 7). In this frame, the specific requirement of Cdc7 during the S phase would be a response to DNA damage to maintain the interactions and assist stressed forks.

How the MCM/Rad51 interaction promotes replicative and repair functions and whether they are mechanistically related is unknown. They seem to occur through non-recombinogenic mechanisms, which points to TLS. This might explain the mild sensitivity to MMS of both *rad51m* and, more specifically, *rad51-C159R*. One possibility is that these physical interactions provide a platform for the landing of replication and repair factors at the proximity of stressed forks. Recently, we showed that Rad51 and Rad52 have a non-recombinogenic role in TLS by facilitating the recruitment of the Rad6/Rad18 complex to chromatin (Cano-Linares et al., 2021). Proliferating cell nuclear antigen (PCNA) ubiquitylation by Rad6/Rad18 is necessary for the recruitment of TLS polymerases (Bienko et al., 2005) and facilitates replication fork advance in the presence of a damaged template (Ortiz-Bazán et al., 2014). In this frame, the MCM/Rad51 interaction might speed up DNA synthesis in the presence of blocking lesions by promoting TLS at the fork. Whether the MCM/Rad51 interaction promotes the recruitment of these and/or additional repair/replication factors or operates through different mechanisms will require further studies.

In sum, a DDT mechanism has evolved in yeast cells that relies on the accumulation of MCM/Rad51/Rad52 complexes in a nucleoprotein scaffold before replication firing to facilitate the assistance to stressed replication forks. Physical interactions of MCM with Rad51 and Rad52 are also detected in mammalian cells (Bailis et al., 2008; Shukla et al., 2005). Therefore, they seem to have a conserved role whose study may help provide an understanding about how cells deal with replicative stress.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Plasmids
  - Growth conditions
  - Search for *rad51* mutants
  - DNA damage sensitivity and cell viability
  - Genetic recombination
  - DNA repair foci analysis
  - Flow cytometry and budding analyses
  - Pulse field gel electrophoresis (PFGE)
  - DNA fiber assay
  - 2D-gel electrophoresis
  - *In vivo* ChEC analyses
  - Chromatin immunoprecipitation analysis
  - Fractionation analyses
  - Coimmunoprecipitation
  - Western blot

## ● QUANTIFICATION AND STATISTICAL ANALYSIS

- Image processing and analysis
- Statistical analyses

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.109440>.

## ACKNOWLEDGMENTS

We thank Gislene Pereira, Heinrich Leonhardt, Philippe Pasero, Nancy M. Hollingsworth, Michael T. Fasullo, Hiroshi Masumoto, and Lorraine Symington for various strains and reagents and Pedro San Segundo, Felipe Cortes Ledesma, and Fernando Monje Casas for critical reading of the manuscript. This work was supported by grants BFU2015-63698-P and PGC2018-099182-B-I00 (to F.P.) from the Spanish government, and project-ID 393547839-SFB1361 (to H.D.U.) from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). M.J.C.-L., M.I.C.-L., C.G.-G., A.Y.-V., and R.G.-P were recipients of pre-doctoral training grants from the Spanish government.

## AUTHOR CONTRIBUTIONS

Investigation, M.J.C.-L., C.G.-G., M.I.C.-L., R.P.W., A.Y.-V., M.M.-H., J.M.R.-R., M.V., and R.G.-P.; conceptualization, F.P.; writing – original draft, F.P.; writing – review & editing, F.P. and H.D.U.; funding acquisition, F.P. and H.D.U.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 24, 2020

Revised: May 28, 2021

Accepted: July 1, 2021

Published: July 27, 2021

## SUPPORTING CITATIONS

The following references appear in the Supplemental information: Gallego-Sánchez et al. (2012); Ogi et al. (2008).

## REFERENCES

- Alabert, C., Bianco, J.N., and Pasero, P. (2009). Differential regulation of homologous recombination at DNA breaks and replication forks by the Mrc1 branch of the S-phase checkpoint. *EMBO J.* 28, 1131–1141.
- Alabert, C., Bukowski-Wills, J.-C., Lee, S.-B., Kustatscher, G., Nakamura, K., de Lima Alves, F., Menard, P., Mejlvang, J., Rappsilber, J., and Groth, A. (2014). Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nat. Cell Biol.* 16, 281–293.
- Alvaro, D., Sunjevaric, I., Reid, R.J.D., Lisby, M., Stillman, D.J., and Rothstein, R. (2006). Systematic hybrid LOH: a new method to reduce false positives and negatives during screening of yeast gene deletion libraries. *Yeast* 23, 1097–1106.
- Aparicio, O.M., Weinstein, D.M., and Bell, S.P. (1997). Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* 91, 59–69.
- Bailis, J.M., Luche, D.D., Hunter, T., and Forsburg, S.L. (2008). Minichromosome maintenance proteins interact with checkpoint and recombination proteins to promote s-phase genome stability. *Mol. Cell. Biol.* 28, 1724–1738.

- Barlow, J.H., and Rothstein, R. (2009). Rad52 recruitment is DNA replication independent and regulated by Cdc28 and the Mec1 kinase. *EMBO J.* **28**, 1121–1130.
- Bastia, D., Srivastava, P., Zaman, S., Choudhury, M., Mohanty, B.K., Bacal, J., Langston, L.D., Pasero, P., and O'Donnell, M.E. (2016). Phosphorylation of CMG helicase and Top1 is required for programmed fork arrest. *Proc. Natl. Acad. Sci. USA* **113**, E3639–E3648.
- Bertazzi, D.T., Kurtulmus, B., and Pereira, G. (2011). The cortical protein Lte1 promotes mitotic exit by inhibiting the spindle position checkpoint kinase Kin4. *J. Cell Biol.* **193**, 1033–1048.
- Bhat, K.P., and Cortez, D. (2018). RPA and RAD51: fork reversal, fork protection, and genome stability. *Nat. Struct. Mol. Biol.* **25**, 446–453.
- Bienko, M., Green, C.M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kanouche, P., Wider, G., Peter, M., Lehmann, A.R., et al. (2005). Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821–1824.
- Branzei, D., and Foiani, M. (2010). Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol.* **11**, 208–219.
- Branzei, D., and Psakhye, I. (2016). DNA damage tolerance. *Curr. Opin. Cell Biol.* **40**, 137–144.
- Branzei, D., Vanoli, F., and Foiani, M. (2008). SUMOylation regulates Rad18-mediated template switch. *Nature* **456**, 915–920.
- Braun, K.A., and Breeden, L.L. (2007). Nascent transcription of MCM2-7 is important for nuclear localization of the minichromosome maintenance complex in G1. *Mol. Biol. Cell* **18**, 1447–1456.
- Bruck, I., and Kaplan, D.L. (2015). Conserved mechanism for coordinating replication fork helicase assembly with phosphorylation of the helicase. *Proc. Natl. Acad. Sci. USA* **112**, 11223–11228.
- Cano-Linares, M.I., Yáñez-Vilches, A., García-Rodríguez, N., Barrientos-Moreno, M., González-Prieto, R., San-Segundo, P., Ulrich, H.D., and Prado, F. (2021). Non-recombinogenic roles for Rad52 in translesion synthesis during DNA damage tolerance. *EMBO Rep.* **22**, e50410.
- Clemente-Ruiz, M., and Prado, F. (2009). Chromatin assembly controls replication fork stability. *EMBO Rep.* **10**, 790–796.
- Cocker, J.H., Piatti, S., Santocanale, C., Nasmyth, K., and Diffley, J.F. (1996). An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* **379**, 180–182.
- Das, S.P., and Rhind, N. (2016). How and why multiple MCMs are loaded at origins of DNA replication. *BioEssays* **38**, 613–617.
- Das, S.P., Borrmann, T., Liu, V.W.T., Yang, S.C.-H., Bechhoefer, J., and Rhind, N. (2015). Replication timing is regulated by the number of MCMs loaded at origins. *Genome Res.* **25**, 1886–1892.
- Deegan, T.D., and Diffley, J.F.X. (2016). MCM: one ring to rule them all. *Curr. Opin. Struct. Biol.* **37**, 145–151.
- Donovan, S., Harwood, J., Drury, L.S., and Diffley, J.F. (1997). Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc. Natl. Acad. Sci. USA* **94**, 5611–5616.
- Fasullo, M.T., and Davis, R.W. (1987). Recombinational substrates designed to study recombination between unique and repetitive sequences in vivo. *Proc. Natl. Acad. Sci. USA* **84**, 6215–6219.
- Feser, J., Truong, D., Das, C., Carson, J.J., Kieft, J., Harkness, T., and Tyler, J.K. (2010). Elevated histone expression promotes life span extension. *Mol. Cell* **39**, 724–735.
- Foiani, M., Marini, F., Gamba, D., Lucchini, G., and Plevani, P. (1994). The B subunit of the DNA polymerase alpha-primase complex in *Saccharomyces cerevisiae* executes an essential function at the initial stage of DNA replication. *Mol. Cell. Biol.* **14**, 923–933.
- Frei, C., and Gasser, S.M. (2000). The yeast Sgs1p helicase acts upstream of Rad53p in the DNA replication checkpoint and colocalizes with Rad53p in S-phase-specific foci. *Genes Dev.* **14**, 81–96.
- Fung, C.W., Fortin, G.S., Peterson, S.E., and Symington, L.S. (2006). The rad51-K191R ATPase-defective mutant is impaired for presynaptic filament formation. *Mol. Cell. Biol.* **26**, 9544–9554.
- Gallego-Sánchez, A., Andrés, S., Conde, F., San-Segundo, P.A., and Bueno, A. (2012). Reversal of PCNA ubiquitylation by Ubp10 in *Saccharomyces cerevisiae*. *PLoS Genet.* **8**, e1002826.
- Gangavarapu, V., Prakash, S., and Prakash, L. (2007). Requirement of RAD52 group genes for postreplication repair of UV-damaged DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 7758–7764.
- Ge, X.Q., Jackson, D.A., and Blow, J.J. (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev.* **21**, 3331–3341.
- Giannattasio, M., Zwicky, K., Follonier, C., Foiani, M., Lopes, M., and Branzei, D. (2014). Visualization of recombination-mediated damage bypass by template switching. *Nat. Struct. Mol. Biol.* **21**, 884–892.
- González-Prieto, R., Muñoz-Cabello, A.M., Cabello-Lobato, M.J., and Prado, F. (2013). Rad51 replication fork recruitment is required for DNA damage tolerance. *EMBO J.* **32**, 1307–1321.
- González-Prieto, R., Cabello-Lobato, M.J., and Prado, F. (2021). In vivo binding of recombination proteins to non-DSB DNA lesions and to replication forks. *Methods Mol. Biol.* **2153**, 447–458.
- Hardy, C.F., Dryga, O., Seematter, S., Pahl, P.M., and Sclafani, R.A. (1997). mcm5/cdc46-bob1 bypasses the requirement for the S phase activator Cdc7p. *Proc. Natl. Acad. Sci. USA* **94**, 3151–3155.
- Hashimoto, Y., Ray Chaudhuri, A., Lopes, M., and Costanzo, V. (2010). Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nat. Struct. Mol. Biol.* **17**, 1305–1311.
- Hecht, A., and Grunstein, M. (1999). Mapping DNA interaction sites of chromosomal proteins using immunoprecipitation and polymerase chain reaction. *Methods Enzymol.* **304**, 399–414.
- Hesketh, E.L., Knight, J.R.P., Wilson, R.H.C., Chong, J.P.J., and Coverley, D. (2015). Transient association of MCM complex proteins with the nuclear matrix during initiation of mammalian DNA replication. *Cell Cycle* **14**, 333–341.
- Heyer, W.-D., Ehmsen, K.T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139.
- Hills, S.A., and Diffley, J.F.X. (2014). DNA replication and oncogene-induced replicative stress. *Curr. Biol.* **24**, R435–R444.
- Hoang, M.L., Leon, R.P., Pessoa-Brandao, L., Hunt, S., Raghuraman, M.K., Fangman, W.L., Brewer, B.J., and Sclafani, R.A. (2007). Structural changes in Mcm5 protein bypass Cdc7-Dbf4 function and reduce replication origin efficiency in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 7594–7602.
- Hyrien, O. (2016). How MCM loading and spreading specify eukaryotic DNA replication initiation sites. *F1000Res.* **5**, 2063.
- Ibarra, A., Schwob, E., and Méndez, J. (2008). Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc. Natl. Acad. Sci. USA* **105**, 8956–8961.
- Jachymczyk, W.J., Chlebowicz, E., Swietlinska, Z., and Zuk, J. (1977). Alkaline sucrose sedimentation studies of MMS-induced DNA single-strand breakage and rejoining in the wild type and in UV-sensitive mutants of *Saccharomyces cerevisiae*. *Mutat. Res.* **43**, 1–10.
- Kilic, S., Lezaja, A., Gatti, M., Bianco, E., Michelenia, J., Imhof, R., and Altmeier, M. (2019). Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments. *EMBO J.* **38**, e101379.
- Labib, K. (2010). How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev.* **24**, 1208–1219.
- Li, H., and O'Donnell, M.E. (2018). The eukaryotic CMG helicase at the replication fork: emerging architecture reveals an unexpected mechanism. *BioEssays* **40**, 1700208–1700209.
- Li, X., Zhang, X.-P., Solinger, J.A., Kianitsa, K., Yu, X., Egelman, E.H., and Heyer, W.-D. (2007). Rad51 and Rad54 ATPase activities are both required

- to modulate Rad51-dsDNA filament dynamics. *Nucleic Acids Res.* 35, 4124–4140.
- Liang, C., and Stillman, B. (1997). Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in *cdc6* mutants. *Genes Dev.* 11, 3375–3386.
- Liberi, G., Maffioletti, G., Lucca, C., Chiolo, I., Baryshnikova, A., Cotta-Ramusino, C., Lopes, M., Pelliccioli, A., Haber, J.E., and Foiani, M. (2005). Rad51-dependent DNA structures accumulate at damaged replication forks in *sgs1* mutants defective in the yeast ortholog of BLM RecQ helicase. *Genes Dev.* 19, 339–350.
- Lisby, M., Rothstein, R., and Mortensen, U.H. (2001). Rad52 forms DNA repair and recombination centers during S phase. *Proc. Natl. Acad. Sci. USA* 98, 8276–8282.
- Longtine, M.S., III, McKenzie, A., III, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* 14, 953–961.
- Lopes, M., Foiani, M., and Sogo, J.M. (2006). Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol. Cell* 21, 15–27.
- López-Contreras, A.J., Ruppen, I., Nieto-Soler, M., Murga, M., Rodríguez-Acebes, S., Remeseiro, S., Rodrigo-Perez, S., Rojas, A.M., Méndez, J., Muñoz, J., and Fernandez-Capetillo, O. (2013). A proteomic characterization of factors enriched at nascent DNA molecules. *Cell Rep.* 3, 1105–1116.
- Lord, S.J., Velle, K.B., Mullins, R.D., and Fritz-Laylin, L.K. (2020). SuperPlots: communicating reproducibility and variability in cell biology. *J. Cell Biol.* 219, e202001064.
- Lundin, C., North, M., Erixon, K., Walters, K., Jenssen, D., Goldman, A.S.H., and Helleday, T. (2005). Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. *Nucleic Acids Res.* 33, 3799–3811.
- Maki, K., Inoue, T., Onaka, A., Hashizume, H., Somete, N., Kobayashi, Y., Murakami, S., Shigaki, C., Takahashi, T.S., Masukata, H., and Nakagawa, T. (2011). Abundance of prereplicative complexes (Pre-RCs) facilitates recombinational repair under replication stress in fission yeast. *J. Biol. Chem.* 286, 41701–41710.
- Mankouri, H.W., Ngo, H.-P., and Hickson, I.D. (2007). Shu proteins promote the formation of homologous recombination intermediates that are processed by Sgs1-Rmi1-Top3. *Mol. Biol. Cell* 18, 4062–4073.
- McCullum, E.O., Williams, B.A.R., Zhang, J., and Chaput, J.C. (2010). Random mutagenesis by error-prone PCR. *Methods Mol. Biol.* 634, 103–109.
- Miné-Hattab, J., Heltberg, M., Villemeur, M., Guedj, C., Mora, T., Walczak, A.M., Dahan, M., and Taddei, A. (2021). Single molecule microscopy reveals key physical features of repair foci in living cells. *eLife* 10, e60577.
- Morgan, E.A., Shah, N., and Symington, L.S. (2002). The requirement for ATP hydrolysis by *Saccharomyces cerevisiae* Rad51 is bypassed by mating-type heterozygosity or RAD54 in high copy. *Mol. Cell Biol.* 22, 6336–6343.
- Mortensen, U.H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. *Proc. Natl. Acad. Sci. USA* 93, 10729–10734.
- Naumov, G.I., Naumova, E.S., Lantto, R.A., Louis, E.J., and Korhola, M. (1992). Genetic homology between *Saccharomyces cerevisiae* and its sibling species *S. paradoxus* and *S. bayanus*: electrophoretic karyotypes. *Yeast* 8, 599–612.
- Nguyen, V.Q., Co, C., and Li, J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* 411, 1068–1073.
- Ogi, H., Wang, C.-Z., Nakai, W., Kawasaki, Y., and Masumoto, H. (2008). The role of the *Saccharomyces cerevisiae* Cdc7-Dbf4 complex in the replication checkpoint. *Gene* 414, 32–40.
- Ortiz-Bazán, M.Á., Gallo-Fernández, M., Saugar, I., Jiménez-Martín, A., Vázquez, M.V., and Tercero, J.A. (2014). Rad5 plays a major role in the cellular response to DNA damage during chromosome replication. *Cell Rep.* 9, 460–468.
- Pasero, P., Duncker, B.P., Schwob, E., and Gasser, S.M. (1999). A role for the Cdc7 kinase regulatory subunit Dbf4p in the formation of initiation-competent origins of replication. *Genes Dev.* 13, 2159–2176.
- Prado, F. (2014). Homologous recombination maintenance of genome integrity during DNA damage tolerance. *Mol. Cell Oncol.* 7, e957039.
- Prado, F. (2018). Homologous recombination: to fork and beyond. *Genes (Basel)* 9, 603.
- Prado, F., and Aguilera, A. (1995). Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the RAD1, RAD10, and RAD52 genes. *Genetics* 139, 109–123.
- Prado, F., and Aguilera, A. (2005). Partial depletion of histone H4 increases homologous recombination-mediated genetic instability. *Mol. Cell Biol.* 25, 1526–1536.
- Prakash, L. (1981). Characterization of postreplication repair in *Saccharomyces cerevisiae* and effects of *rad6*, *rad18*, *rev3* and *rad52* mutations. *Mol. Gen. Genet.* 184, 471–478.
- Resnick, M.A., Boyce, J., and Cox, B. (1981). Postreplication repair in *Saccharomyces cerevisiae*. *J. Bacteriol.* 146, 285–290.
- Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., Backmann, N., Conrath, K., Muyltermans, S., Cardoso, M.C., and Leonhardt, H. (2006). Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat. Methods* 3, 887–889.
- Rothbauer, U., Zolghadr, K., Muyltermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrapp for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell. Proteomics* 7, 282–289.
- Schmid, M., Durussel, T., and Laemmli, U.K. (2004). ChIC and ChEC: genomic mapping of chromatin proteins. *Mol. Cell* 16, 147–157.
- Shah, P.P., Zheng, X., Epshtein, A., Carey, J.N., Bishop, D.K., and Klein, H.L. (2010). Swi2/Snf2-related translocases prevent accumulation of toxic Rad51 complexes during mitotic growth. *Mol. Cell* 39, 862–872.
- Sheff, M.A., and Thorn, K.S. (2004). Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* 21, 661–670.
- Shukla, A., Navadgi, V.M., Mallikarjuna, K., and Rao, B.J. (2005). Interaction of hRad51 and hRad52 with MCM complex: a cross-talk between recombination and replication proteins. *Biochem. Biophys. Res. Commun.* 329, 1240–1245.
- Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122, 19–27.
- Sung, P., and Stratton, S.A. (1996). Yeast Rad51 recombinase mediates polar DNA strand exchange in the absence of ATP hydrolysis. *J. Biol. Chem.* 271, 27983–27986.
- Tanaka, T., Knapp, D., and Nasmyth, K. (1997). Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell* 90, 649–660.
- Van Komen, S., Petukhova, G., Sigurdsson, S., Stratton, S., and Sung, P. (2000). Superhelicity-driven homologous DNA pairing by yeast recombination factors Rad51 and Rad54. *Mol. Cell* 6, 563–572.
- Vanoli, F., Fumasoni, M., Szakal, B., Maloisel, L., and Branzei, D. (2010). Replication and recombination factors contributing to recombination-dependent bypass of DNA lesions by template switch. *PLoS Genet.* 6, e1001205.
- Vázquez, M.V., Rojas, V., and Tercero, J.A. (2008). Multiple pathways cooperate to facilitate DNA replication fork progression through alkylated DNA. *DNA Repair (Amst.)* 7, 1693–1704.
- Wan, L., Zhang, C., Shokat, K.M., and Hollingsworth, N.M. (2006). Chemical inactivation of *cdc7* kinase in budding yeast results in a reversible arrest that allows efficient cell synchronization prior to meiotic recombination. *Genetics* 174, 1767–1774.

- Waterman, D.P., Zhou, F., Li, K., Lee, C.-S., Tsabar, M., Eapen, V.V., Mazzella, A., and Haber, J.E. (2019). Live cell monitoring of double strand breaks in *S. cerevisiae*. *PLoS Genet.* *15*, e1008001–e1008022.
- Weinreich, M., and Stillman, B. (1999). Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. *EMBO J.* *18*, 5334–5346.
- Wilson, R.H.C., and Coverley, D. (2013). Relationship between DNA replication and the nuclear matrix. *Genes Cells* *18*, 17–31.
- Wong, R.P., García-Rodríguez, N., Zilio, N., Hanulová, M., and Ulrich, H.D. (2020). Processing of DNA polymerase-blocking lesions during genome replication is spatially and temporally segregated from replication forks. *Mol. Cell* *77*, 3–16.e4.
- Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* *294*, 2357–2360.
- Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J.A., Herrador, R., Vindigni, A., and Lopes, M. (2015). Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *J. Cell Biol.* *208*, 563–579.



STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Mouse monoclonal anti GFP	Clontech	Cat# 632381
Rabbit polyclonal anti-Rad51	Santa Cruz	Cat# sc-33626
Rabbit polyclonal anti-Rad51	Abcam (= <a href="https://www.abcam.com">Antibodies.com</a> )	Cat# ab63798 (= A284)
Rat monoclonal anti-HA	Roche	Cat# 11867423001 (3F10)
Mouse monoclonal anti-HA	Roche	Cat# 11666606001 (12CA5)
Goat polyclonal anti-Mcm7	Santa Cruz	Cat# sc-6688
Mouse monoclonal anti-Mcm4	Santa Cruz	Cat# sc-166036
Rabbit polyclonal anti-Rad52	( <a href="#">Mortensen et al., 1996</a> )	N/A
Mouse monoclonal anti-Pgk1	Invitrogen	Cat# 22C5D8
Rabbit polyclonal anti-H4	Abcam	Cat# ab10158
Rat monoclonal anti-BrdU	Abcam	Cat# ab6326
Mouse monoclonal anti-single stranded DNA	Sigma-Aldrich	Cat# MAB3034
Goat anti-rat Cy5	Abcam	Cat# ab6565
Goat anti-mouse Cy3	Abcam	Cat# ab6946
<b>Chemicals, peptides, and recombinant proteins</b>		
Methyl-methanesulfonate (MMS)	Merck	Cat# 129925-256
Hydroxyurea (HU)	Material Blanco de Laboratorio, S.L.	Cat# H9120
$\alpha$ -factor	AB BCN S.L.	N/A
$\alpha$ -factor	Proteogenix SAS	Cat# GM-PT301350-95
Pronase	Merck	Cat# 10165921001
Nocodazole	Merck	Cat# M1404-50MG
PP1 Analog II (1NMPP1)	Merck	Cat# 529581-1
PP1 Analog II (1NMPP1)	Santa Cruz	Cat# sc-203214
MNase I	Merck	Cat# N3755-200UN
benzonase	Merck	Cat# 70746-4
5-Bromo-2'-deoxyuridine (BrdU)	Sigma	Cat# B5002
Agarose, NuSieve GTG Agarose (low-melting temp)	Lonza	Cat# 50081
$\lambda$ phosphatase	New England Biolabs	Cat# P0753S
<b>Critical commercial assays</b>		
GFP trap magnetic beads	Chromotek	Cat# gtma-100
dynabeads protein G	Invitrogen	Cat# 10004D
QuickChange XL Site-directed Mutagenesis Kit	Agilent	Cat# 200517
<b>Experimental models: Organisms/strains</b>		
<i>S. cerevisiae</i> : Strain background: W303	ATCC	ATCC: 208353
Other yeast strains: <a href="#">Table S1</a>	This study	N/A
<b>Oligonucleotides</b>		
See <a href="#">Table S2</a>	This study	N/A
<b>Recombinant DNA</b>		
pRS551-L120A,V181A	A gift from Nancy M. Hollingsworth, Stony Brook University	N/A
pRS51s	A gift from Andrés Aguilera, Seville University	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pRSRAD51g	This study	N/A
pRS313-51.54	This study	N/A
pRS306-r51.54	This study	N/A
p313r51-C159R	This study	N/A
pWJ1213	(Alvaro et al., 2006)	N/A
pWJ1344	(Alvaro et al., 2006)	N/A
pFA6a-MN-HIS3MX6	(Schmid et al., 2004)	N/A
pKT209	(Sheff and Thorn, 2004)	N/A
pFA6A-3HA- HIS3MX6	(Longtine et al., 1998)	N/A
pFA6a-GBP-KanMX4 (pHA29)	(Bertazzi et al., 2011; Rothbauer et al., 2006, 2008)	N/A
<b>Software and algorithms</b>		
Image Lab™	Biorad	<a href="http://www.bio-rad.com/es-es/product/image-lab-software?ID=KRE6P5E8Z">http://www.bio-rad.com/es-es/product/image-lab-software?ID=KRE6P5E8Z</a>
ImageGauge	Fujifilm	<a href="https://www.bioz.com/result/sciencelab98imagegaugesoftware/product/FUJIFILM">https://www.bioz.com/result/sciencelab98imagegaugesoftware/product/FUJIFILM</a>
ImageJ	Fiji	<a href="https://imagej.net/software/fiji/">https://imagej.net/software/fiji/</a>
Prism	GraphPad	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
Scatter SuperPlots	(Lord et al., 2020)	N/A

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Félix Prado ([felix.prado@cabimer.es](mailto:felix.prado@cabimer.es))

**Materials availability**

All unique/stable reagents generated in this study are available from the Lead Contact without restriction.

**Data and code availability**

- All data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

Yeast strains used in this study are listed in Table S1. Tagged and deletion strains were constructed by a PCR-based strategy (Longtine et al., 1998). The uSCE system was backcrossed five times into the W303 background (Cano-Linares et al., 2021). The integrative plasmids pRS551-L120A,V181A (a gift from N. Hollingsworth) and pRS306-r51.54 (see Method details) were used to replace *CDC7* and *RAD51* with *cdc7-as3* (W303cdc7as3-2) and *rad51m* (w303.51.54-7), respectively. Briefly, a wild-type strain was transformed with either pRS551-L120A,V181A or pRS306-r51.54 (cut with EcoRI) and grown first in medium without uracil to select the integration event (that duplicates the gene) and then in medium with 5-fluoroorotic acid to select strains that had lost one of the two copies and the intervening sequence. *cdc7-as3* and *rad51m* strains were selected by 1NMPP1-dependent cell growth inhibition and MMS sensitivity, respectively, and confirmed by DNA sequencing.

**METHOD DETAILS**

**Plasmids**

pWJ1213 and pWJ1344 are a centromeric plasmid expressing *RAD52-YFP* (Alvaro et al., 2006). pFA6a-MN-HIS3MX6 (Schmid et al., 2004), pKT209 (Sheff and Thorn, 2004), pFA6A-3HA- HIS3MX6 (Longtine et al., 1998) and pHA29 (pFA6a-GBP-KanMX4)

(Bertazzi et al., 2011; Rothbauer et al., 2006, 2008) are plasmids for protein tagging with MNaseI, eGFP, HA and GBP, respectively. pRS551-L120A,V181A (a gift from N. Hollingsworth) is an integrative plasmid to replace *CDC7* with *cdc7-as3*. pRSRAD51g and pRS51s are centromeric plasmids expressing *RAD51* and either *URA3* or *HIS3*, respectively. pRSRAD51g was constructed by inserting a 2.4 kb PCR fragment containing the *RAD51* allele at the BamHI-HindIII site of pRS316 (Sikorski and Hieter, 1989). pRSR51s (A. Aguilera's lab) was constructed by inserting the genomic EcoRV-BamHI fragment containing the *RAD51* allele at the EcoRV-BamHI site of pRS313 (Sikorski and Hieter, 1989). pRS313-51.54 is identical to pRSR51s but it contains the *rad51m* allele (see Search for [rad51 mutants](#)). pRS306-r51.54 is an integrative plasmid containing the *rad51m* allele. It was constructed by inserting the BamHI-HindIII fragment from pRS313-r51.54 (containing *rad51m*) at the BamHI-HindIII site of pRS306 (Sikorski and Hieter, 1989). The p313r51-C159R plasmid was generated by directed mutagenesis of the *RAD51* gene in the pRSR51s plasmid with the QuickChange XL Site-Directed Mutagenesis Kit (200517; Agilent).

### Growth conditions

Yeast cells were grown at 30°C in supplemented minimal medium (SMM), except for experiments that required synchronization in metaphase, which were performed in YPAD rich medium with 15 µg/ml nocodazole for 1 h. For G1 synchronization, cells were grown to mid-log phase and  $\alpha$ -factor was added twice at 60 min intervals at either 2 (*BAR1* strains) or 0.25 µg/ml (*bar1Δ* strains). Then, cells were washed three times and released into fresh medium with 50 µg/ml pronase (to remove  $\alpha$ -factor) in the absence or presence of MMS at the indicated concentrations. To eliminate the MMS before releasing cells into fresh medium, samples were treated with 2.5% sodium thiosulfate to inactivate it and then washed three times. For Cdc6 depletion, *GAL::CDC6* cells were synchronized in G1 with  $\alpha$ -factor, released into fresh medium with 1% DMSO and 15 µg/ml nocodazole for 90 min to arrest in metaphase, incubated with 2% glucose and fresh nocodazole for 1 h, and released with 2% glucose and  $\alpha$ -factor for 2.5 h.

### Search for *rad51* mutants

An *in vivo* library enriched in *rad51* mutants was constructed by co-transforming a *rad51Δ* strain containing pRSRAD51g (centromeric plasmid expressing the *RAD51* and *URA3* genes) with 1) the plasmid pRS51s (centromeric plasmid expressing the *RAD51* and *HIS3* genes) linearized at the *RAD51* ORF with EcoRI and 2) a mutagenic PCR product of *RAD51* (McCullum et al., 2010). Cells were plated onto SMM – His to select transformants that recircularized pRS51s. Recircularization by HR with either the PCR product or pRSRAD51g was ~4 and ~20 times more efficient than recircularization by HR with just pRSRAD51g or recircularization by NHEJ (estimated in a *rad51Δ* strain containing pRS316 instead of pRSRAD51g). Next, transformants were replica-plated onto SMM medium containing 5'-fluoroorotic acid to select for cells that had lost pRSRAD51g. This strategy restricts the expression of *RAD51* to the *in vivo* cloning and allows the screening of *rad51* mutants. To search for mutants sensitive to MMS and resistance to IR, cells were either grown onto SMM medium at different MMS concentrations or plated onto SMM medium, irradiated at different doses and then grown under unperturbed conditions. Plasmids from positive clones were isolated; the phenotype was confirmed with new *rad51Δ* transformants and the mutations determined by DNA sequencing of the plasmids.

### DNA damage sensitivity and cell viability

MMS sensitivity was determined by spotting ten-fold serial dilutions of the same number of mid-log growing cells onto medium with or without the drug (chronic damage). For IR sensitivity spotted cells were irradiated and then grown under unperturbed conditions (acute damage). Cell viability was determined from exponentially growing cultures treated either with MMS or IR as the frequency of cells able to form colonies after DNA damage relative to untreated cells, taken as 100.

### Genetic recombination

HR was determined by measuring the frequency of His<sup>+</sup> recombinants generated by uSCE in a chromosomal-integrated system (Fasullo and Davis, 1987). Recombination frequencies were determined by fluctuation tests as previously reported (Prado and Aguilera, 1995) but from liquid cultures. Briefly, six independent cell cultures (started with a colony) were grown to the same mid-log phase and then treated with different concentrations of MMS for 4 hours or irradiated with IR at different doses. Cells from untreated and treated cultures were plated with the appropriate dilutions onto SMM without histidine and SMM to calculate recombinants and total viable cells (as colony-forming units), respectively. The frequency of HR was calculated using the median of recombinants and the mean of total cells. To have a more accurate value, the mean and SEM of at least 3 independent fluctuation tests are given.

### DNA repair foci analysis

The percentage of cells with Rad52 or RPA foci was determined as described previously (Lisby et al., 2001). Cells expressing Rfa1-YFP or transformed with plasmid pWJ1344 or pWJ1213 (expressing Rad52-YFP) were grown in liquid culture under the indicated conditions, fixed with 2.5% formaldehyde in 0.1M potassium phosphate pH 6.4 for 10 minutes, washed twice with 0.1M potassium phosphate pH 6.6 and resuspended in 0.1M potassium phosphate pH 7.4. Finally, cells were fixed with 80% ethanol for 10 minutes, resuspended in H<sub>2</sub>O or DAPI and visualized with a Leica CTR6000 fluorescence microscope. The percentage of cells with foci was counted directly on the processed samples under the microscope or on acquired images. In this case, six contrast and fluorescence images along the z axis (0.49 µm length each) were acquired with a CCD camera (Leica DFC350 FX) to find well-defined foci. Images

were processed and analyzed with the MetaMorph software (Molecular devices). A total number of approximately 100 cells were analyzed for each time point and experiment.

### Flow cytometry and budding analyses

DNA content analysis was performed by flow cytometry as reported previously (Prado and Aguilera, 2005). Cells were fixed with 70% ethanol, washed with phosphate-buffered saline (PBS), incubated with 1 mg of RNaseA/ ml PBS, and stained with 5  $\mu$ g/ml propidium iodide. Samples were sonicated to separate single cells and analyzed in a FACSCalibur flow cytometer. The percentage of budded cells was determined by counting 100 cells for each time point and experiment.

### Pulse field gel electrophoresis (PFGE)

Replicating and complete chromosomes were resolved by PFGE (Biorad; 120o field angle; 6 V/cm; 14oC; initial block: switch time of 70 s for 16 h; final block: switch time of 120 s for 12 h). Total DNA from cultures growing at the indicated conditions was extracted in low-melting agarose (50081; Lonza) plugs as described (Naumov et al., 1992), but the incubation with Proteinase K was performed in buffer L (10 mM TrisHCl pH 8.0, 100 mM EDTA pH 8.0, 20 mM NaCl, SDS1%) at 30°C for 24 h to avoid heat-induced breakage of methylated DNA during sample preparation (Lundin et al., 2005). Gels were stained with ethidium bromide, and the signals were acquired in a Fuji FLA5100 and quantified with the ImageGauge software (Fujifilm).

### DNA fiber assay

G1-synchronized cells were incubated with 0.4 mg/ml BrdU for 15 minutes, washed twice in water and released into fresh medium containing 0.4 mg/ml BrdU and 50  $\mu$ g/ml pronase with or without 0.025% MMS. Cells were harvested at indicated time points to isolate genomic DNA for molecular combing. Briefly, yeast cells were spheroplasted with zymolyase to over 90% completion.  $0.3 \times 10^9$  cells were embedded in low melting point agarose and incubated with 1 mg/ml Proteinase K in 125 mM EDTA pH9.5 containing 1% Sarkosyl for 2 days with one change of buffer in between. Plugs were washed thoroughly in TE buffer with 100 mM NaCl and once with 100 mM MES buffer pH6.0 and 100 mM NaCl. Plugs were melted in MES buffer with NaCl at 68 C for 15 minutes and digested overnight with Agarase (3 units per plug) at 45°C. Agarase was inactivated at 65 C for 10 minutes and the DNA solution was cooled down to room temperature. DNA fibers were combed on silanized coverslips (Combicoverslips, Genomic Vision) at a constant speed of 350  $\mu$ m/s using DNA combing apparatus from Genomic Vision. Slides were baked at 65°C for at least 2 hours, denatured in 0.5M NaOH 1M NaCl for 15 minutes at room temperature and washed thrice in PBS. Slides were dehydrated in 70%, 90% and 100% EtOH and air-dried. For immunofluorescence staining, slides were blocked in BlockAid (Thermo Fisher Scientific) for 15 minutes at 37°C and incubated sequentially with 1:20 monoclonal rat anti-BrdU antibody (clone BU1/75; ab6326; Abcam), 1:50 monoclonal mouse anti-single stranded DNA antibody (clone 16-19; MAB3034; Sigma-Aldrich), and 1:100 goat anti-rat Cy5 (ab6565; Abcam) and goat anti-mouse Cy3 (ab6946; Abcam) secondary antibodies at 37°C for 1 hour with three PBS-T (1  $\times$  PBS with 0.1% Tween-20) washes in between each incubation. Slides were dehydrated in 70%, 90% and 100% and air-dried before mounting on Prolong Diamond mounting medium (Thermo Fisher Scientific). Images were obtained with the AF7000 widefield fluorescence microscope (Leica) equipped with a 63  $\times$  (NA = 1.4) oil immersion objective, ORCA-Flash 4.0 V2 digital CMOS camera (Hamamatsu), LED light source (SOLA, lumencor) and LAS AF software (Leica). Tile-scans were acquired with Z stacks of 4 planes with a step size of 0.2  $\mu$ m. Cy3 signals were obtained with the N3 filter and Cy5 signals with the Y5 filter. Lengths of BrdU-labeled replication tracts were measured with ImageJ FIJI software. Inter-origin distances were calculated by measuring the distances between the middle point of each replication tract (replication origin) on the same fiber.

### 2D-gel electrophoresis

Replication intermediates were analyzed by 2D-gel electrophoresis from cells arrested with sodium azide (0.1% final concentration) and cooled down on ice as reported (Clemente-Ruiz and Prado, 2009). Briefly, total DNA was isolated with the G2/CTAB protocol, digested with *EcoRV* and *HindIII*, resolved by neutral/neutral two-dimensional gel electrophoresis, blotted to Hybond<sup>TM</sup>-XL membranes, and analyzed by hybridization with an *ARS305* proximal <sup>32</sup>P-labeled probe (probe A). All signals were acquired in a Fuji FLA5100 and quantified with the ImageGauge software (Fujifilm).

### In vivo ChEC analyses

Chromatin endogenous cleavage (ChEC) of Rad51-MN and Rad51m-MN cells was performed as reported (González-Prieto et al., 2013, 2021) from cultures grown in the presence or absence of MMS and arrested with sodium azide (0.1% final concentration). For cleavage induction, digitonin-permeabilized cells were incubated with 2 mM CaCl<sub>2</sub> at 30°C under gentle agitation. Total DNA was isolated and resolved into 0.8% TAE 1  $\times$  agarose gels. Gels were scanned in a Fuji FLA5100, and the signal profile quantified using the ImageGauge software (Fujifilm). The area of the DNA digestion profiles was equalized to eliminate DNA loading differences.

### Chromatin immunoprecipitation analysis

Chromatin immunoprecipitation (ChIP) assays were performed as described (Hecht and Grunstein, 1999). Each sample was processed, split and incubated with antibodies either against GFP (632381, Clontech) or Rad51 (sc-33626, Santa Cruz or

ab63798, Abcam (= A284, [Antibodies.com](https://www.abcam.com)). Protein enrichment at each specific region was calculated as the ratio between the immunoprecipitated sample and the input in a *MCM4-GFP RAD51* strain relative to the same ratio in a *MCM4 rad51Δ* strain. Oligonucleotide sequences for the real-time PCR amplifications performed on purified DNA before (input) or after (immunoprecipitated) incubation with the antibodies are shown in [Table S2](#).

### Fractionation analyses

Fractionation was performed as described for chromatin fractionation in young yeast cells ([Feser et al., 2010](#)) with some modifications. Samples (15–30 ml) from mid-log phase cultures were collected by centrifugation, washed with cold 0.1mM Tris pH 9.4, 10mM DTT, and incubated for 15 min in 1 mL of the same buffer on ice. Cells were then washed with cold spheroplasting buffer (20mM HEPES pH 7.4, 1.2mM sorbitol, Roche Complete EDTA free protease inhibitor cocktail) and incubated with 1 mL of the same buffer with 210 μg zymoliasse 20T for 1 h at 30°C. The spheroplasts were collected, washed twice with cold washing buffer (20mM Tris pH 7.4, 20mM KCl, 1M sorbitol, 0.1 μM spermine, 0.25 μM spermidine, protease inhibitors), and resuspended in 1 mL lysis buffer (20mM Tris pH 7.4, 20mM KCl, 0.4 M sorbitol, 0.1 μM spermine, 0.25 μM spermidine, 1% Triton X-100, protease inhibitors) for 5 min on ice. An aliquot (80 μl) was removed for the total sample, and the remaining sample was centrifuged for 15 min at 13000 g at 4°C to separate soluble (supernatant) and insoluble (pellet) fractions. Each pellet was washed with 0.5 mL cold lysis buffer and resuspended in 80 μL of water, and insoluble, soluble and total samples were mixed with SDS buffer for western blot analyses. Similar volumes were loaded for each time point for each kinetic, and similar cell equivalents of the insoluble and soluble fractions were loaded for the fractionation controls. To study nuclease-soluble and insoluble pellet fractions, the pellet generated after cell fractionation was resuspended in 100 μL of buffer (1mM Tris pH 7.4, 1mM MgCl<sub>2</sub>), incubated with 250 units of benzonase for 30 min at 37°C, and centrifuged 5 min at 13000 g. Similar volumes of soluble and non-soluble (resuspended in 100 μL of water) fractions were mixed with SDS buffer and loaded for western analyses.

### Coimmunoprecipitation

CoIP was performed with 100–150 mL samples from mid-log phase cultures (O.D. ~0.75) that had been lysed with a Multi Beads Shocker (Yasui Kikai) at 4°C with 1x vol glass beads in 0.5 mL NP40 lysis buffer (50mM Tris pH 7.5, 150mM NaCl, 1% NP40) with protease inhibitors (1mM PMSF, 2mM DTT and Roche Complete EDTA free). Lysates were cleared by two consecutive centrifugation steps for 5 min at 1000 g (4°C) and, in case of nuclease treatment, mixed with MNaseI (15mM Tris, 50mM NaCl, 1.4mM CaCl<sub>2</sub>, 0.2mM EDTA, 0.2mM EGTA pH 8.0) or benzonase buffer (50mM Tris pH 8.0, 1mM MgCl<sub>2</sub>) and either 2.5 u MNaseI or 25 u benzonase and incubated for 20 min at 37°C. Samples were then collected by centrifugation for 15 min at 13000 g (4°C) and the total amount of protein at the supernatant was quantified by a Bradford assay. An aliquot of each sample was removed for the input. For GFP-based CoIPs, GFP trap magnetic beads (gtma-100; Chromotek) were incubated overnight at 4°C with similar amount of proteins (~12–16 mg) in NP40 lysis buffer, washed extensively with either standard (Mcm4-GFP CoIPs) or modified NP40 lysis buffer (50mM Tris pH 7.5, 500mM NaCl, 1% NP40) (Rad51-YFP CoIP). For the Mcm4-HA immunoprecipitation, samples were pre-cleared with dynabeads protein G (10004D, Invitrogen) during 2h at 4°C, and then incubated first with 0.5 μg/ml HA antibody high affinity (3F10, Roche; 11867423001) (overnight at 4°C) and then with dynabeads protein G (2h at 4°C), washed extensively with modified NP40 lysis buffer (50mM Tris pH 7.5, 1M NaCl, 1% NP40). Finally, samples were analyzed by western blot with the corresponding antibodies. Unless otherwise indicated, CoIP was performed with MNase I-treated lysates.

For CoIP from fractionated samples, chromatin and soluble fractions were prepared as above from 200 mL mid-log phase cultures except that spheroplasts were resuspended in 1 mL lysis buffer containing 1% NP40 instead of 1% Triton X-100. After centrifugation, the chromatin fraction was resuspended in 1 mL of the same buffer, and aliquots from the soluble and chromatin fractions were removed for the fractionation controls (H4 and Pgc1). The soluble and chromatin fractions were lysed with a Multi Beads Shocker at 4°C with 1x volume glass beads and the lysates were then processed as indicated for CoIP analyses.

### Western blot

Protein samples were resolved by 8% (Mcm4-GFP, Mcm4-HA, Mcm7, Rad52-YFP, Rad51 and Rad52) or 15% (Pgc1 and H4) SDS-PAGE, and probed with antibodies against GFP (632381, Clontech), HA (Roche Refs: 11666606001 and 11867423001), Mcm4 (sc-166036, Santa Cruz), Mcm7 (sc-6688, Santa Cruz), Rad51 (sc-33626, Santa Cruz and ab63798, Abcam), Rad52 ([Mortensen et al., 1996](#)), Pgc1 (22C5D8, Invitrogen) or H4 (ab10158, Abcam). Yeast protein extracts to analyze total amount of Mcm4-GFP and Rad51 ([Figure S4A](#)) were prepared using the TCA protocol as described ([Foiani et al., 1994](#)). All western signals were acquired and quantified in a ChemiDoc MP image system and quantified with the Image Lab™ software (Biorad).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Image processing and analysis

The acquirement, processing and analysis of DNA repair foci, western blots, DNA fibers, and gels from 2D, ChEC and PFGE are specified in the corresponding [Method details](#).

**Statistical analyses**

Statistical analyses were performed using the Prism software (GraphPad). Mean, SEM, sample size and statistical tests are indicated in the Figure legends. Sample size was not predetermined using statistical methods. Given the reduced sample size, the analyses were performed assuming that they follow normal distributions. Scatter SuperPlots were done as recently reported ([Lord et al., 2020](#)).

## RESEARCH ARTICLE

# The histone demethylase PHF8 regulates astrocyte differentiation and function

Simona Iacobucci<sup>1</sup>, Natalia Padilla<sup>2,\*</sup>, Martina Gabrielli<sup>3,\*</sup>, Claudia Navarro<sup>1</sup>, Marta Lombardi<sup>3</sup>, Marta Vicioso-Mantis<sup>1</sup>, Claudia Verderio<sup>3</sup>, Xavier de la Cruz<sup>2</sup> and Marian A. Martínez-Balbás<sup>1,‡</sup>

## ABSTRACT

Epigenetic factors have been shown to play a crucial role in X-linked intellectual disability (XLID). Here, we investigate the contribution of the XLID-associated histone demethylase PHF8 to astrocyte differentiation and function. Using genome-wide analyses and biochemical assays in mouse astrocytic cultures, we reveal a regulatory crosstalk between PHF8 and the Notch signaling pathway that balances the expression of the master astrocytic gene *Nfia*. Moreover, PHF8 regulates key synaptic genes in astrocytes by maintaining low levels of H4K20me3. Accordingly, astrocytic-PHF8 depletion has a striking effect on neuronal synapse formation and maturation *in vitro*. These data reveal that PHF8 is crucial in astrocyte development to maintain chromatin homeostasis and limit heterochromatin formation at synaptogenic genes. Our studies provide insights into the involvement of epigenetics in intellectual disability.

**KEY WORDS:** PHF8, Histone demethylation, Chromatin transcription, Astrocyte differentiation, Synapse, XLID

## INTRODUCTION

X-linked intellectual disability (XLID) includes a diverse group of cognitive disorders ranging from mild intellectual deficits to severe cognitive impairments (Chelly et al., 2006; Ropers and Hamel, 2005). Intellectual disability (ID), as well as other neurodevelopmental disorders, is characterized by anomalies in the establishment and function of synaptic circuits. Historically, ID research focused mainly on neurons, although astrocytes make a crucial contribution to synapse formation (Cresto et al., 2019; McGann et al., 2012), maturation and elimination (Araque et al., 2014). Only recently have genetic studies underscored the potential role of astrocytes in neurodevelopmental disorders, such as Down, Rett and Fragile X syndromes (Cresto et al., 2019). Thus, astrocyte contribution to ID is still largely unexplored.

Large-scale genetic analyses revealed that a strikingly considerable number of genes mutated in XLID encode regulators

of chromatin activity and structure (Chelly et al., 2006; Kramer and van Bokhoven, 2009). In particular, gene mutation screening and linkage analysis of familial ID have identified the histone demethylase (HDM) plant homeodomain finger (PHD) protein 8 (PHF8) as a factor associated with XLID (Koivisto et al., 2007; Loenarz et al., 2010; Qiao et al., 2008). PHF8 belongs to the KDM7 family of HDMs that is formed by PHF2, PHF8 and KIAA1718 (KDM7A) in humans (Li et al., 2006). The members of this family contain an amino-terminal PHD finger that recognizes and binds methylated lysines (Fortschegger et al., 2010; Horton et al., 2010; Kleine-Kohlbrecher et al., 2010; Tsukada et al., 2010; Wen et al., 2010) and a Jumonji-C (JmjC) domain that catalyzes lysine demethylation (Fortschegger and Shiekhattar, 2011). Previous studies have demonstrated that PHF8 removes mainly mono- and dimethyl-lysine 9 on histone H3 (H3K9me1/2) and monomethyl-lysine 20 on histone H4 (H4K20me1) (Fortschegger et al., 2010; Horton et al., 2010; Kleine-Kohlbrecher et al., 2010; Liu et al., 2010). Deletions and point mutations in the PHF8 catalytic domain cause Siderius–Hamel syndrome, characterized by mild XLID with cleft lip and/or a cleft palate (Abidi et al., 2007; Koivisto et al., 2007; Laumonier et al., 2005; Siderius et al., 1999). Moreover, PHF8 might contribute to this phenotype by targeting genes such as *JARID1C* and *MSX1* (Kleine-Kohlbrecher et al., 2010; Qi et al., 2010), which are involved in XLID and neural development, respectively; indeed, PHF8 depletion was shown to impair the neuronal differentiation of murine P19 cells (Qiu et al., 2010). In addition, *in vivo* studies demonstrated that genetic silencing of the *Phf8* homolog in zebrafish caused apoptosis of neural cells and craniofacial anomalies (Qi et al., 2010; Tsukada et al., 2010), and led to a global increase in H3K9me2 and compromised locomotion in *Caenorhabditis elegans* (Kleine-Kohlbrecher et al., 2010).

Although it is well established that PHF8 plays a role in neural differentiation, how mutations in the PHF8 catalytic domain cause ID remains unknown. Recent studies have shown that *Phf8*-null mice display deficiencies in learning and memory as a result of alterations of the RSK-mTOR-S6K pathway (Chen et al., 2018) and are resistant to anxiety- and depression-like behaviors because of dysregulation of serotonin receptor *Htr1a* and *Htr2a* expression (Walsh et al., 2017). Interestingly, the PHF8-mediated transcriptional changes are extremely subtle in neurons; this might indicate that alterations not only in neurons, but also in other cells may account for the PHF8 lack-of-function phenotype. Thus, we investigated the contribution of PHF8 to astrocyte differentiation and function at the molecular level.

Here, we show that PHF8 directs the expression of some key regulators of astrocyte differentiation, such as *Nfia*, and that depletion of PHF8 has a striking effect on synapse formation and maturation. *Phf8*-depleted astrocytes show impaired transcription of genes crucial for synaptogenesis; this phenotype can be rescued by overexpressing wild type (WT) PHF8 but not the catalytic mutant. Interestingly, PHF8 depletion or loss of catalytic activity leads to a

<sup>1</sup>Department of Molecular Genomics, Instituto de Biología Molecular de Barcelona (IBMB), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona 08028, Spain. <sup>2</sup>Research Unit in Clinical and Translational Bioinformatics, Vall d'Hebron Institute of Research (VHIR), Passeig de la Vall d'Hebron, 119; E-08035 Barcelona, Spain. Institut Català per la Recerca i Estudis Avançats (ICREA), Barcelona 08018, Spain. <sup>3</sup>CNR Institute of Neuroscience, via Vanvitelli 32, 20129 Milan, Italy.

\*These authors contributed equally to this work

‡Author for correspondence (mmbmc@ibmb.csic.es)

© M.G., 0000-0003-4958-541X; M.V., 0000-0002-4069-3951; M.A.M., 0000-0003-0173-0964

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Handling Editor: James Briscoe

Received 17 July 2020; Accepted 15 April 2021

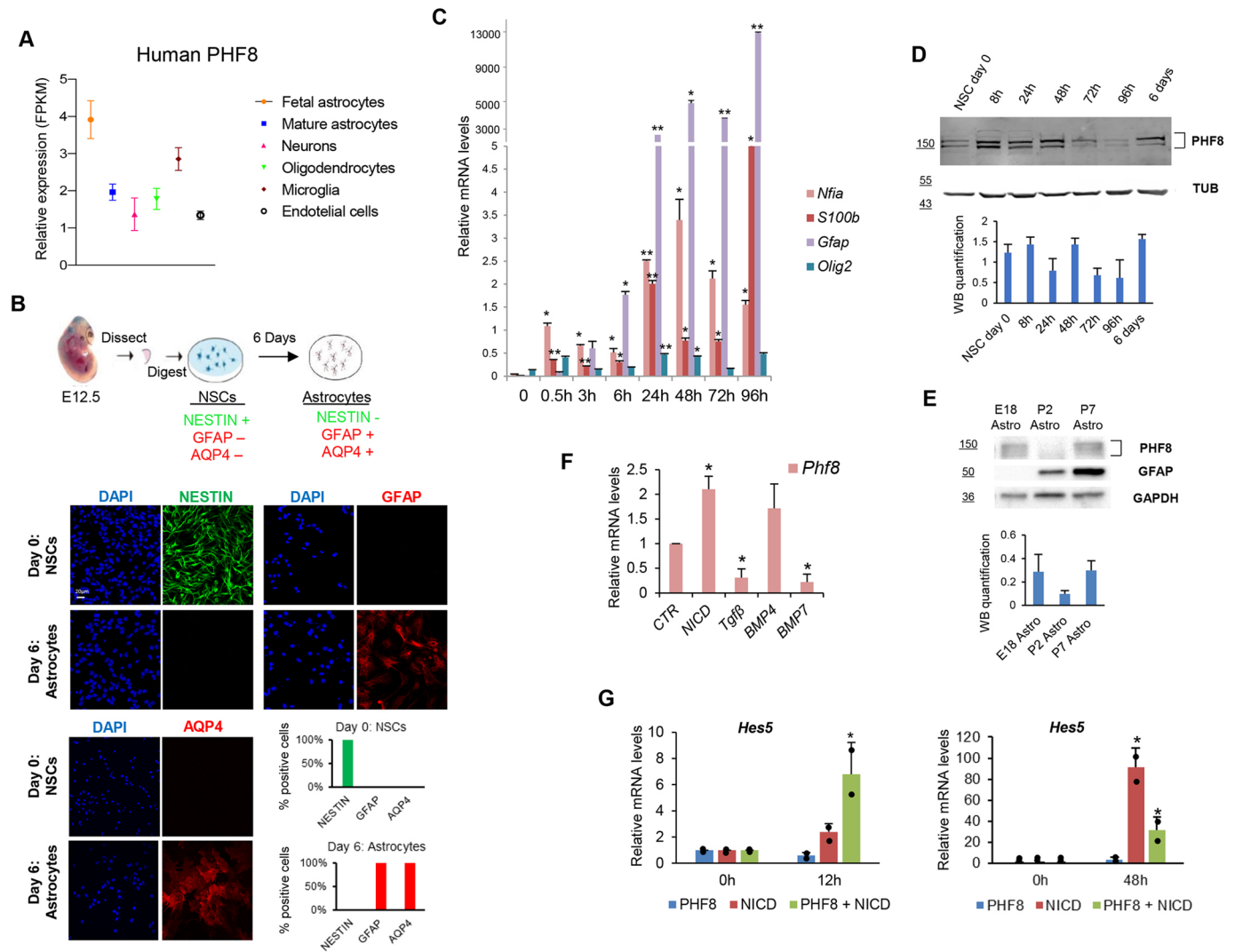
global increase in heterochromatin-associated histone marks in astrocytes. Our data suggest that the loss of PHF8 in astrocytes impairs synaptogenesis because of the formation of ectopic heterochromatin.

## RESULTS

### PHF8 expression during astrocyte differentiation

Analysis of publicly available data of human (Zhang et al., 2016) and mouse (Zhang et al., 2014) neural cell populations showed that

astrocytes express high levels of PHF8 (Fig. 1A; Fig. S1A). Interestingly, this expression was particularly high in fetal astrocytes (Fig. 1A), suggesting a potential role of PHF8 in early events of astrocyte development. To evaluate the functional relevance of PHF8 during astrocyte differentiation, we isolated neural stem cells (NSCs) from cortices of mouse embryos at embryonic day (E) 12.5 (Estarras et al., 2012; Sun et al., 2001) and differentiated them to astrocytes following the protocol described in the Materials and Methods (Fig. 1B). After 6 days in astrocytic



**Fig. 1. PHF8 expression fluctuates during astrocyte differentiation.** (A) PHF8 expression in different human neural cells. Expression level is shown by fragments per kilobase of transcript per million reads mapped (FPKM). Publicly available data from human RNA-seq experiments (<http://www.brainrnaseq.org>; Zhang et al., 2016). Data are mean  $\pm$  s.d. (B) Top: Schematic view of the model used to study the function of PHF8 during astrocyte differentiation. Bottom: Immunostaining assay of NSCs (day 0) and astrocytes (day 6) using nestin, AQP4 and GFAP antibodies and DAPI. Data shown are representative of three biological independent experiments; more than 50 cells were quantified per replicate. The graphs represent the percentage of cells expressing GFAP, AQP4 or nestin at day 0 (NSCs, green) and day 6 (astrocytes, red). (C) NSCs were maintained in astrocytic differentiation medium for different lengths of time. Total RNA was prepared at the indicated times and the levels of *Nfia*, *S100b*, *Gfap* and *Olig2* were determined by qPCR. Expression values were normalized to the housekeeping gene *Gapdh*; values shown are relative to time 0 h. *Olig2* mRNA was used as a negative control. Data are mean  $\pm$  s.d.;  $n=3$  biological replicates. \* $P<0.05$ ; \*\* $P<0.01$ . (D) NSCs were maintained in astrocytic differentiation medium for the indicated times. Total protein extracts were prepared and the PHF8 levels were determined by immunoblot. Alpha-tubulin antibody (TUB) was used as the loading control. Data are mean  $\pm$  s.d.; blot is representative of two biological independent experiments. (E) PHF8 western blot assays of MACS-isolated astrocytes from E18, P2 and P7 mouse brain. GAPDH antibody was used as a loading control. Data are mean  $\pm$  s.d.; blot is representative of two biological independent experiments. (F) HEK 293T cells were transfected with NICD, Smad3S/D (the constitutive active form of Smad3), BMP7, BMP4 or empty vector (CTR) as indicated. Total mRNA was purified and the *PHF8* levels were determined by qPCR. Expression values were normalized to the housekeeping gene *GAPDH*. Data are mean  $\pm$  s.d.;  $n=3$  independent biological replicates. \* $P<0.05$ . (G) HEK 293T cells were transfected with a vector expressing *Phf8* together or not with NICD. Total mRNA was purified and the *HES5* levels at 12 h (left) and 48 h (right) were established by qPCR. Values shown are relative to time 0 h. Expression values were normalized to the housekeeping gene *GAPDH*. Data are mean  $\pm$  s.d.;  $n=2$  independent biological replicates. Values from each are indicated by dots. \* $P<0.05$ . Scale bar: 20  $\mu$ m.



medium, almost 100% of NSCs had lost expression of the progenitor marker nestin and acquired high levels of the astrocyte-associated markers glial fibrillary acidic protein (GFAP) and aquaporin-4 (AQP4) (Fig. 1B). No signal for  $\beta$ -Tubulin 3 (TUB $\beta$ 3), a marker of neuronal differentiation, was noticed at that time (Fig. S1B), although some TUB $\beta$ 3-expressing cells were detected during early differentiation timepoints (Fig. S1B). Accordingly, a progressive increase in *Nfia* and *S100b* expression was detected throughout the neuronal differentiation process (Fig. 1C). Immunostaining of H4K20me1/3, H3K27me3 and H3K9me2 histone marks associated with heterochromatin formation also increased during this differentiation (Fig. S1C).

To gain insight into the role of PHF8 in astrocyte differentiation, we analyzed *Phf8* expression during this process. The results showed an increase in PHF8 protein levels during early differentiation and in differentiated astrocytes (6 days of differentiation) (Fig. 1D). PHF8 upregulation *in vivo* was demonstrated by western blot assays of magnetic-activated cell sorting (MACS)-isolated astrocytes from early development (E18) and intact postnatal (P) mouse brain (P2 and P7) (Fig. 1E). These data suggest a potential contribution of PHF8 to both astrocyte differentiation and function.

Next, we sought to identify the signaling pathway that could be responsible for *Phf8* upregulation during astrocyte differentiation. Given that Notch, TGF $\beta$  and BMP signals are involved in astrocyte development, we activated these pathways and determined the *Phf8* mRNA levels by qPCR. The results showed a clear *Phf8* induction after transfection of the Notch intracellular domain (NICD) (Fig. 1F). The Notch pathway is crucial for astrocyte differentiation induction (Ge et al., 2002; Martini et al., 2013); thus, we investigated whether PHF8 modulates Notch activity. To do so, we overexpressed *Phf8* upon Notch activation (by NICD expression) and analyzed the expression of the well-known Notch target *Hes5* by qPCR. PHF8 facilitated Notch target activation over short time frames, but inhibited it over longer time frames (Fig. 1G). This repression was also observed in overexpression and depletion experiments using a luciferase reporter vector fused to the *Hes5* promoter (Fig. S1D,E).

Altogether, these data demonstrate that PHF8 expression is modulated during astrocyte differentiation. Moreover, they suggest the existence of a regulatory feedback loop between PHF8 and Notch signaling.

### PHF8 regulates transcription during astrocyte differentiation

To gain further understanding of the function of PHF8 in astrocytes, we analyzed the PHF8-dependent transcriptional profile by RNA-sequencing (RNA-seq). To this end, NSCs were transduced with a lentivirus containing either a specific PHF8 short hairpin (sh)RNA that efficiently decreased the PHF8 protein levels (Fig. S2A) or a control shRNA (see Materials and Methods). Next, control (CTR) and PHF8-depleted (PHF8-KD) NSCs were differentiated into astrocytes (6 days of differentiation) (Fig. 2A). Astrocytes derived from PHF8-KD NSCs (Astro PHF8-KD) exhibited decreased PHF8 transcription compared with those from control NSCs (Astro-CTR), as indicated by qPCR (Fig. 2A, bottom panel). The transcriptional profiles of two Astro-CTR and two Astro PHF8-KD samples (Fig. S2B,C) showed that 4987 transcripts had significantly altered expression [ $\log_2$  fold change (FC) >0.5 and <-0.5, respectively;  $P < 0.08$ ] in two biologically independent experiments (Fig. 2B,C). Among these transcripts, 2899 (58%) were downregulated and 2087 (42%) were upregulated upon PHF8 depletion. When the  $\log_2$  FC was increased to 2, the percentage of downregulated genes (694)

increased to 76% (Fig. 2D; Fig. S2D), in agreement with the activator role proposed for PHF8. Changes in the expression of some selected genes were validated by qPCR (Fig. S2E).

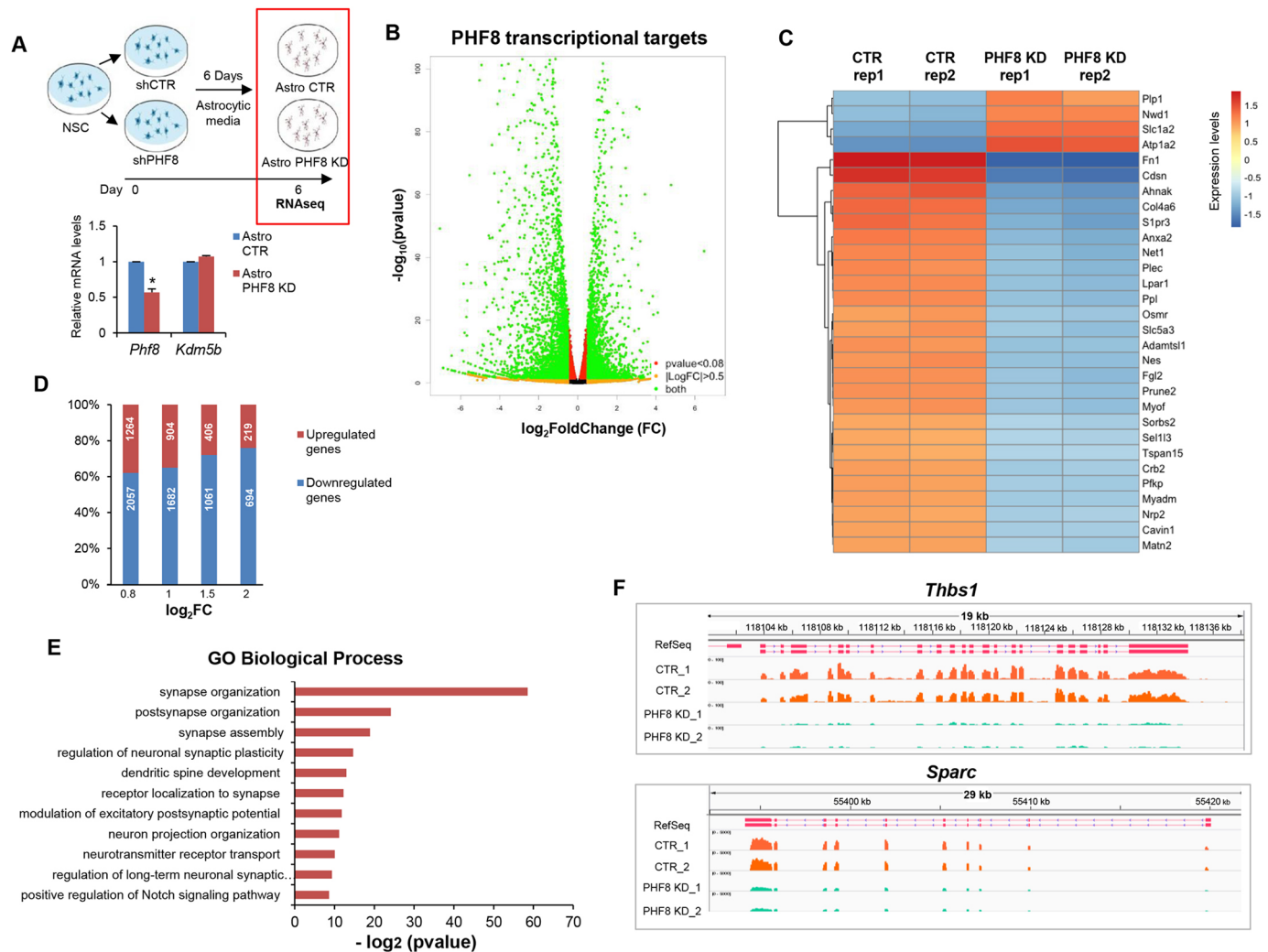
To further characterize the difference between CTR and PHF8-KD astrocytes, we performed a gene ontology (GO) enrichment analysis of regulated genes to identify those biological processes most sensitive to PHF8 depletion. The analysis revealed changes in genes related to biological processes involved in synapse formation and maturation (Fig. 2E), including the astrocytic genes *Gpc4*, *Sparc*, *Thbs1*, *Nrxn1*, *Pcdh8* and *Sdc4* (Cheng et al., 2016; Farhy-Tselnicker et al., 2017; Kucukdereli et al., 2011; Singh et al., 2016) (Fig. 2F). Interestingly, synaptic genes were both downregulated and upregulated in PHF8-depleted astrocytes (Fig. S2F), although downregulated genes showed higher FC values. Importantly, 57% of the astrocytic genes involved in neuron-astrocyte interplay at synapses (Hillen et al., 2018) were affected in PHF8-KD astrocytes (Fig. S2F,G), suggesting that the synapse might be altered. In addition, some Notch targets (*Notch3*, *Hes5*, *Dll3*, *Dll1* and *Cd44*) were misregulated in PHF8-KD astrocytes (Fig. S2H,I). Interestingly, the Notch signaling pathway is crucial for the induction of astrocyte differentiation (Ge et al., 2002; Martini et al., 2013). Finally, an essential gene for astrogenesis, *Nfia* (Deneen et al., 2006), was downregulated in PHF8-depleted astrocytes (Fig. S2E,J), pointing to a significant role of PHF8 in astrocytic differentiation.

Altogether, these data demonstrate that PHF8 facilitates the transcription of astrogenic and synaptogenic genes.

### PHF8 binds to astrogenic and synaptogenic genes

To gain insight into the contribution of PHF8 to gene regulation, we next determined its biological substrates by performing chromatin immunoprecipitation coupled with sequencing (ChIP-seq) in astrocytes after 6 days of differentiation (Fig. 3A). Upon normalization to the input, 8401 peaks ( $P = 0.001$ ) were detected in ChIP data for PHF8 (Fig. 3B). The analysis of the genomic distribution of PHF8 peaks revealed that 46.8% of the peaks localized on distal intergenic regions (see example in Fig. S3A). The remaining 53% were located along the genome and were particularly enriched at introns (Fig. 3B,C). Interestingly, the PHF8 distribution in postmitotic astrocyte cells was noticeably different to that previously described in embryonic stem cells (ESCs), in which the majority of the PHF8-binding sites resided at promoters (Fig. S3B). Comparing PHF8 genomic location in astrocytes and ESCs, we observed that the peaks at promoters in ESCs corresponded to those at introns and intergenic regions in astrocytes (Fig. S3B). GO analysis of PHF8-bound regions indicated that PHF8 was associated with genes involved in neural development, particularly in astrogenesis (*Nfia*) and synaptogenesis (*Sparc* and *Gpc4*), among others (Fig. 3D).

To better understand how PHF8 is targeted to chromatin, we performed bioinformatics analysis and identified that one of the most statistically significant predicted PHF8-binding sites was the RBPJ1 DNA-binding motif (Fig. 3E). RBPJ1 is an effector of the Notch signaling pathway, which is essential in astrocyte differentiation (Ge et al., 2002; Martini et al., 2013). Interestingly, RBPJ1 motif and PHF8 binding were identified at the *Phf8* gene (Fig. 3F), supporting the idea that PHF8, in addition to regulating Notch targets (Fig. S2G,H), is itself a target. In fact, the RBPJ1-PHF8 interaction has been previously identified in another cellular context (Yatim et al., 2012). To reinforce these data, we performed GO analysis of the PHF8 and RBPJ1-bound regions; the results showed that they bind to genes involved in synapse assembly and



**Fig. 2. PHF8 regulates transcription in astrocytes.** (A) Schematic of the experiment to study the transcriptional profile of PHF8 in astrocytes (top panel). NSCs were infected with lentivirus expressing either control shRNA (shcontrol) or shRNA specific for PHF8 (shPHF8). After 6 days in astrocytic medium, total RNA was purified and the *Phf8* levels were determined by qPCR. Expression values were normalized to the housekeeping gene *Gapdh*. Data are mean  $\pm$  s.d.;  $n=3$  independent biological replicates. \* $P<0.05$ . (B) Volcano plot representing PHF8 transcriptional targets identified by RNA-seq in Astro-CTR and Astro PHF8-KD. The green dots represent all genes with  $P<0.08$  and a  $\log_2$  FC  $>0.5$  and  $<-0.5$ . (C) Heat map showing the top-30 regulated genes identified by RNA-seq in shcontrol and shPHF8 NSCs. Two biological replicates of shPHF8 cells were used for RNA-seq. All genes showed  $P<0.08$  and a  $\log_2$  FC  $>0.5$  and  $<-0.5$ . (D) Percentage and number of upregulated and downregulated genes in Astro PHF8-KD compared with Astro-CTR with  $P<0.08$  and classified by increasing  $\log_2$  FC. (E) GO analysis showing the biological processes associated with the PHF8-regulated genes ( $P<0.08$  and  $\log_2$  FC  $>0.5$  and  $<-0.5$ ) was performed using as a background the whole *Mus musculus* genome. (F) Integrated Genomics Viewer (IGV) capture showing RNA levels in Astro-CTR and Astro PHF8-KD in *Thbs1* and *Sparc*.

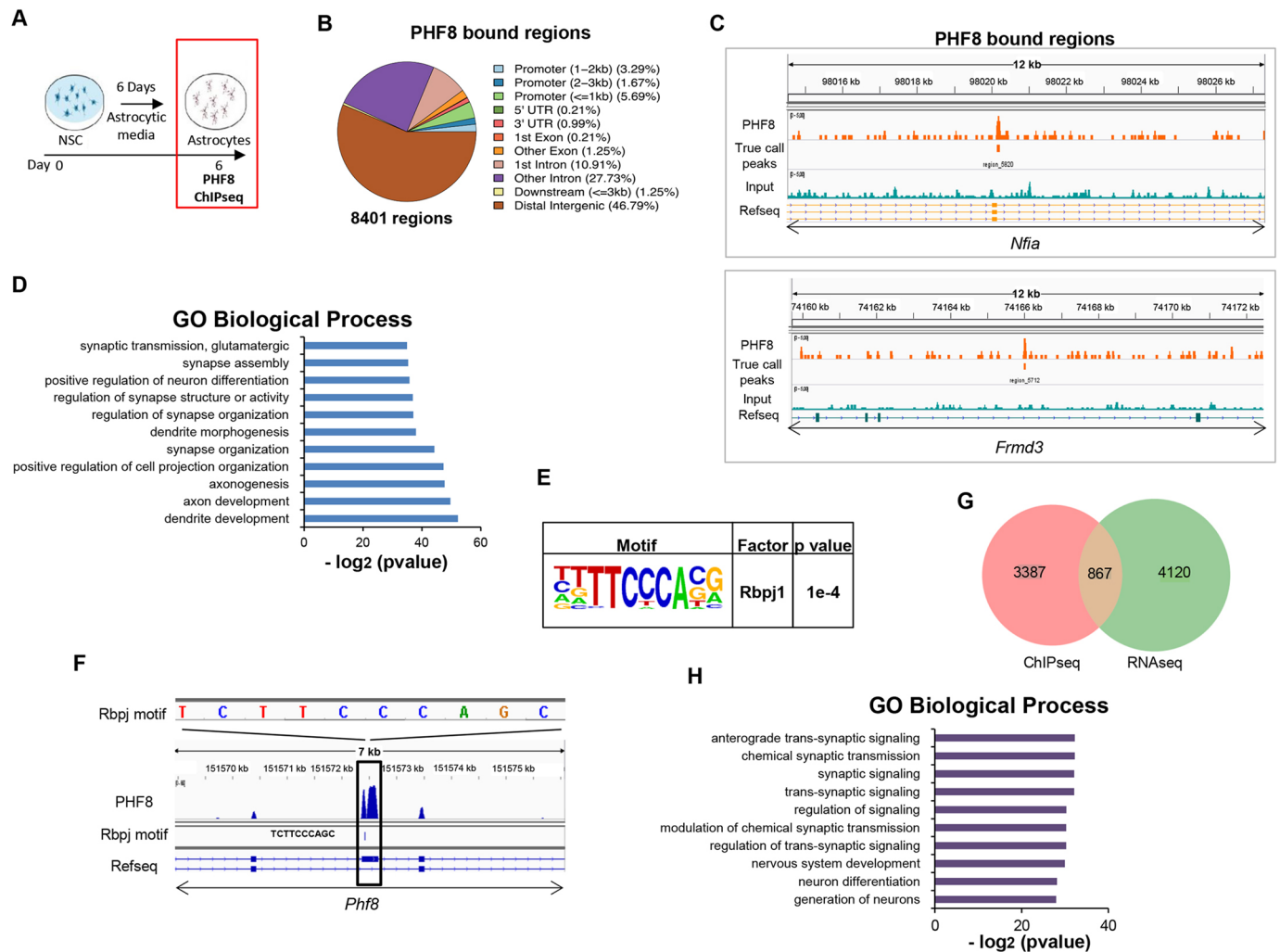
function (Fig. S3C). PHF8 ChIP-seq was validated by qPCR analysis of seven randomly chosen genes (Fig. S3D). Moreover, by ChIP qPCR, we confirmed the binding of PHF8 to genes essential for astrocyte differentiation and function (*Nfia* and *Sparc*) during early stages of differentiation (day 1) (Fig. S3E).

Next, we identified the direct transcriptional targets of PHF8 by comparing the genes bound by PHF8 in the ChIP-seq experiment (4254) with the transcriptional profile ( $\log_2$  FC  $>0.5$  and  $<-0.5$ , respectively;  $P<0.08$ , 4986 transcripts). Among the genes bound by PHF8, 867 (20.3%) showed a PHF8 dependency for transcriptional regulation in the RNA-seq experiment (Fig. 3G). GO analysis of the PHF8-direct target genes showed that the most enriched terms were again related to astrocytic differentiation and synapse assembly and function (Fig. 3H).

Collectively, these data show that PHF8 regulates genes related to astrogenesis and synaptogenesis in astrocytes.

### PHF8 depletion alters the astrocyte transcriptional profile

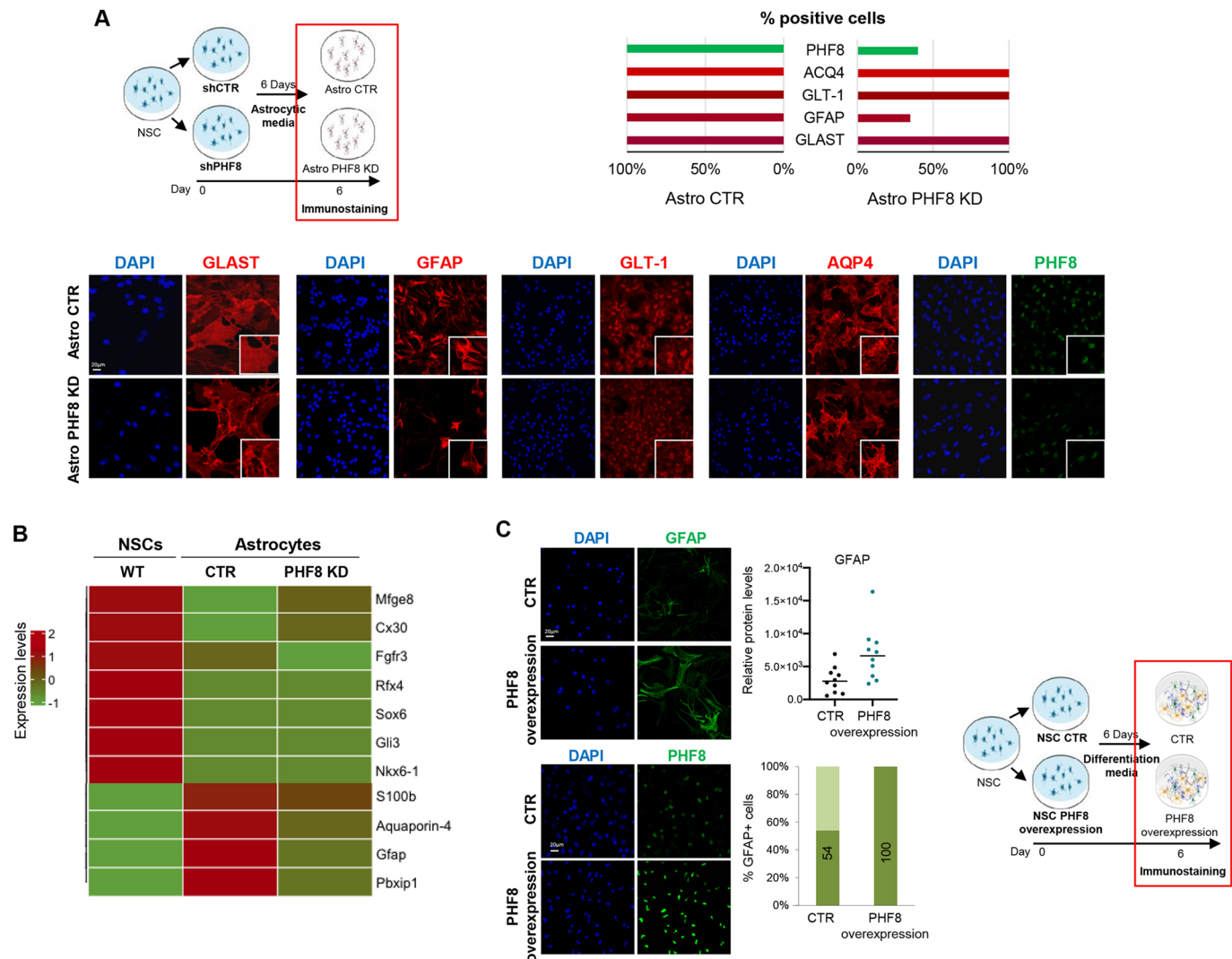
Given that PHF8 regulates the expression of key genes involved in astrocyte differentiation, we explored whether PHF8 depletion alters astrocyte differentiation. Immunostaining for the well-known astrocytic markers GLAST, GFAP, GLT-1 and AQP4 revealed that PHF8 depletion caused a decrease in GFAP protein expression, with no alteration in the other astrocytic markers (Fig. 4A). We also tested for immunoreactivity of other neural cell markers: oligodendrocytes (OLIG2, NG2 and GPR17), neurons (TUB $\beta$ 3), microglia (IBA1) and NSCs (nestin) (Fig. S4A). A percentage of Astro PHF8-KD was positive for either TUB $\beta$ 3 (13%) or OLIG2 (33%) (Fig. S4A). Notably, under our differentiation conditions, 12% of Astro-CTR expressed OLIG2, whereas this was higher in Astro PHF8-KD (33%). The lack of signal for other oligodendrocyte markers (GPR17 and NG2) (Fig. S4A) and the presence of astrocytic markers (GLAST, GLT-1 and AQP4) (Fig. 4A) indicated



**Fig. 3. PHF8 binds to astrogenic and synaptogenic genes.** (A) Schematic of the ChIP-seq experiment to identify the binding sites of PHF8 in astrocytes. PHF8 antibody used for immunoprecipitation was previously used in ChIP assays (see Materials and Methods). (B) Genomic distribution of PHF8 ChIP-seq peaks in astrocytes. In total, 8401 regions were identified. (C) Integrated Genomics Viewer (IGV) captures showing PHF8 peaks in *Nfia* and *Frmd3* in astrocytes. (D) GO analysis showing the biological processes associated with the PHF8-bound genes (4254) using as a background the whole *Mus musculus* genome. (E) Motif enrichment analysis of PHF8 ChIP-seq peaks in astrocytes using 'Homer known motif' showing an enriched motif. (F) IGV captures showing PHF8 peaks and the RBPJ1-binding motif in *Phf8* in astrocytes. (G) Venn diagram showing overlap between PHF8-bound genes (4254) and PHF8 transcriptional targets with  $P < 0.08$  and  $\log_2 FC > 0.5$  and  $< -0.5$  (4987 transcripts). Of the PHF8-bound genes identified by ChIP-seq, 20.3% were differentially expressed by RNA-seq. (H) GO analysis showing the biological processes associated with the PHF8-direct target genes using as a background the whole *Mus musculus* genome.

that these cells were not oligodendrocytes. Interestingly, the TUB $\beta$ 3 marker was not detected in PHF8-KD astrocytes after a longer differentiation time (25 days), suggesting a delay in the differentiation process (Fig. S6C). Moreover, a small percentage of Astro PHF8-KD expressed multiple lineage markers [TUB $\beta$ 3 and GLAST (9.6%); OLIG2 and GLAST (33.0%)] (Fig. S4B). To further understand the nature of astrocytes developed from PHF8-depleted NSCs, we compared the transcriptional profile of NSCs (GSE88173; ENCODE Project Consortium, 2012) with those of Astro-CTR and Astro PHF8-KD. Cells resulting from differentiation upon PHF8 depletion had an astrocyte signature, although they misexpressed some genes compared with control astrocytes (Fig. 4B). Moreover, the transcriptional profiles of both Astro-CTR and Astro PHF8-KD were clearly different from those of oligodendrocytes (Fig. S4C) or neurons (Fig. S4D). Altogether, these data suggest that PHF8-depleted NSCs differentiated *in vitro* into astrocytes that showed a distinct transcriptional profile compared with that of control astrocytes.

A possibility that could explain the differences observed in transcription is that the levels of PHF8 contribute to astrocyte heterogeneity. To explore this hypothesis, we analyzed PHF8 expression in the five astrocyte subtypes defined by Batiuk and collaborators (Batiuk et al., 2020). Given that these data were obtained from cortex and hippocampus, they are the most appropriate *in vivo* model to compare with our cortical NSC-derived astrocytes. The results indicated that PHF8 was expressed at similar levels in astrocyte subtypes (Fig. S5A). In addition, we sought to determine whether PHF8 depletion primes any enrichment of the subtypes defined by Batiuk and collaborators. Considering that the identified astrocyte subtypes are characterized by the unique expression of transcripts, we analyzed whether the genes upregulated in PHF8-depleted astrocytes were significantly enriched in any of the five astrocyte groups. No significant enrichment on any population was observed (Fig. S5B), suggesting that, upon PHF8 depletion, a reshaped transcriptional profile was obtained that did not correspond to any defined astrocyte subtypes *in vivo*. Altogether, these data



**Fig. 4. PHF8-depleted NSCs differentiate into distinct astrocytes.** (A) Control and PHF8-depleted NSCs were differentiated to astrocytes over 6 days to generate Astro-CTR and Astro PHF8-KD, respectively. Cells were fixed and stained with GLAST, GFAP, GLT-1, AQP4 and PHF8 antibodies and DAPI. The percentage of cells expressing these markers in each population is shown. (B) Heat map showing astrocytic gene expression identified by RNA-seq in Astro-CTR and Astro PHF8-KD compared with NSCs (GSE88173). All the genes showed  $P < 0.08$  and  $\log_2 FC > 0.5$  and  $< -0.5$ . (C) Control and PHF8-overexpressing NSCs were maintained in medium without growing factors for 6 days (see schematic). Cells were fixed and stained with GFAP antibody and DAPI. The relative levels of GFAP (determined by ImageJ) per cell and percentage of GFAP-expressing cells are shown. Images are representative of at least three biological independent experiments. Scale bars: 20  $\mu\text{m}$ .

suggest that PHF8 did not contribute to generate astrocyte heterogeneity *in vitro*; however, more definitive studies are required to understand this complex issue fully.

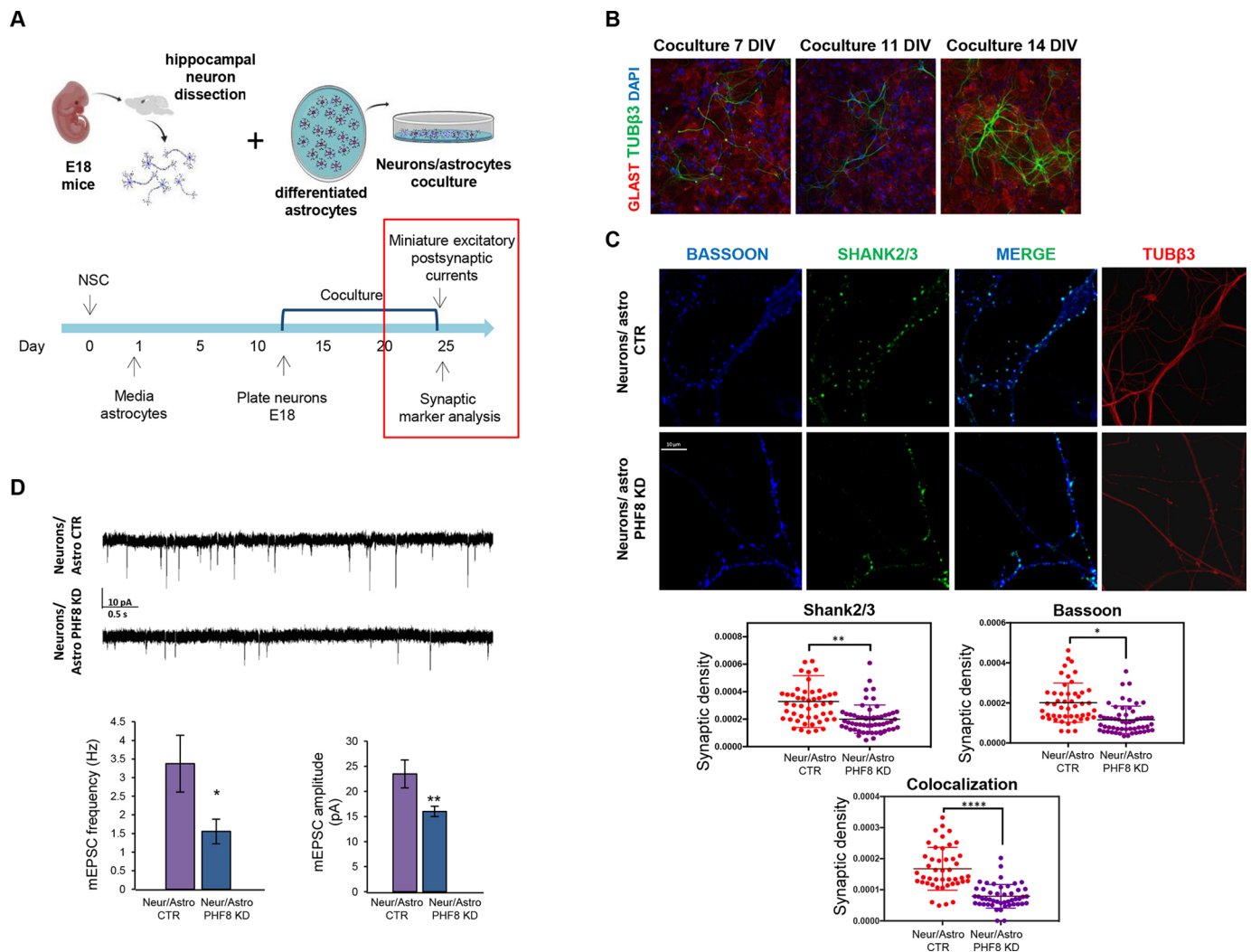
We next investigated the capacity of Astro PHF8-KD to resume proliferation and maintain the stem cell state upon differentiation. We first compared the transcriptional profile of NSCs (GSE88173; ENCODE Project Consortium, 2012) with that of Astro-CTR or Astro PHF8-KD; Astro PHF8-KD did not express either proliferation- or stemness-related genes (Fig. S5C). Moreover, immunostaining assays did not reveal a signal for the progenitor marker nestin in Astro PHF8-KD (Fig. S4A). Finally, the ability of Astro PHF8-KD to proliferate was analyzed; the data revealed that these cells lost the ability to enter the cell cycle (as Astro CTR) under the differentiation conditions used in the study (Fig. S5D). Altogether, these data suggest that PHF8-depleted NSCs differentiate into distinct astrocytes.

Next, we investigated the contribution of *Phf8* to astrocytic fate, by establishing a NSC line that overexpressed PHF8 in an inducible

manner and culturing the cells in a medium without growth factor. Under basal *Phf8* expression and without growth factor, NSCs differentiated into neurons, astrocytes and oligodendrocytes, as expected (Jori et al., 2007). However, when *Phf8* overexpression was induced, a clear increase in GFAP-, but not OLIG2-expressing cells was observed (Fig. 4C; Fig. S5E), indicating that PHF8 promotes NSC differentiation towards astrocytes.

### PHF8 depletion impairs neuronal synapse

Given that PHF8 depletion led to profound defects in synaptogenic gene expression, we next investigated the function of astrocytic PHF8 in synapse formation. We cultured primary hippocampal neurons on PHF8-depleted or control astrocytes and quantified the density and function of excitatory synapses. Control or PHF8-depleted NSCs were first differentiated towards astrocytes for 10 days. Then, primary neurons were plated on the differentiated cells and maintained in co-culture for 14 days, to analyze the density of excitatory synapses and measure basal synaptic transmission (Fig. 5A). Immunofluorescence



**Fig. 5. PHF8 depletion impairs neuronal synapse formation.** (A) Schematic of the neuron/astrocyte co-culture experiment. (B) Immunostaining assay showing GLAST, TUBβ3 and DAPI levels in co-cultures after 7, 11 and 14 DIV. (C) Immunostaining showing bassoon and SHANK2/3 staining in neurons cultured on Astro-CTR or Astro PHF8-KD (top). The synaptic density was determined by the colocalization of both markers (bottom). Data shown are representative of three biological independent experiments; more than 50 cells were quantified per replicate. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\*\* $P < 0.0001$ . (D) Representative traces of mEPSCs from neurons cultured on astrocytes CTR or PHF8 KD and histogram showing the mean frequency ( $P = 0.042$  Mann–Whitney Rank Sum Test) and amplitudes of mEPSCs ( $P = 0.009$  Mann–Whitney Rank Sum Test). Data are mean  $\pm$  s.d.; \* $P < 0.05$ , \*\* $P < 0.01$ . Scale bars: 20  $\mu$ m in B; 10  $\mu$ m in C.

analysis of the co-cultures for TUBβ3 revealed progressive enlargement of the neuron cell bodies and maturation of the dendritic tree over time, at 7, 11 and 14 days *in vitro* (DIV) (Fig. 5B). No differences in neuron maturation or survival were detected between neurons co-cultured with PHF8-KD versus Astro-CTR (Fig. S6A,B). Finally, the maintenance of astrocyte identity after 25 days in culture was confirmed by immunostaining assays using AQP4, GLAST, TUBβ3 and OLIG2 markers (Fig. S6C).

To prove the involvement of astrocytic PHF8 in synapse formation, we first analyzed the density of excitatory synapses. Immunostaining for the presynaptic active zone marker bassoon and the postsynaptic density marker SHANK2 showed a significant decrease in the density of both pre- and postsynaptic puncta as well as of juxtaposed pre- and postsynaptic terminals (normalized per dendritic length, see Materials and Methods), in neurons co-cultured with PHF8-KD astrocytes compared with neurons co-cultured with Astro-CTR (Fig. 5C). As a further control, we examined synaptic density in neurons cultured in the absence of

astrocytes. The analysis showed a similar decrease in the density of bassoon-positive presynaptic terminals in purified neurons and neurons co-cultured with PHF8-depleted astrocytes compared with neurons co-cultured with Astro-CTR (Fig. S6D). The same held true for SHANK2-positive postsynaptic terminals and bassoon/SHANK2 colocalizing puncta (Fig. S6D).

To evaluate the impact of astrocytic PHF8 depletion on synaptic transmission, we measured miniature excitatory postsynaptic currents (mEPSCs), through whole-cell patch-clamp electrophysiological recordings, on 14-day-old neurons co-cultured with either control or PHF8-depleted astrocytes. Importantly, only neurons with similar resting membrane potentials were analyzed. mEPSC analysis revealed a significant decrease in both the frequency and amplitude of the miniature excitatory events upon astrocytic PHF8 depletion (Fig. 5D). A variation in mEPSC frequency is usually related to presynaptic changes leading to altered probability of neurotransmitter release, or to a change in the number of synapses, in line with immunofluorescence analysis for pre- and postsynaptic markers

reported herein. By contrast, variation in mEPSC amplitude is probably the result of postsynaptic modifications, such as an altered number of postsynaptic receptors.

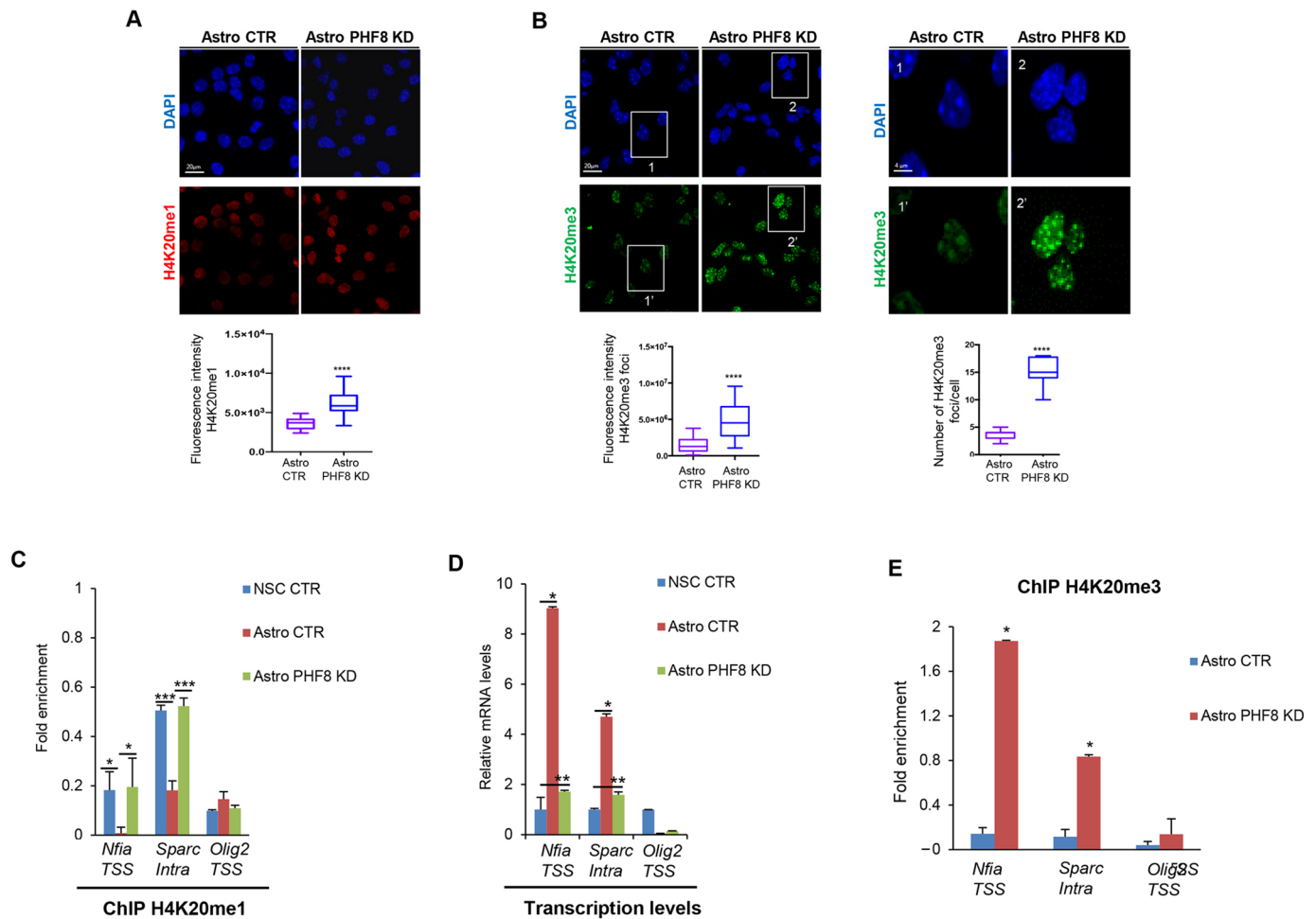
Altogether, these data demonstrate that PHF8 deficiency in astrocytes induces profound alterations in the formation and function of excitatory synapses *in vitro*.

### PHF8 maintains low levels of H4K20me1/3 at astrogenic and synaptogenic genes

Previous studies have shown that H4K20me1 is the main substrate of PHF8 demethylating activity (Liu et al., 2010). Thus, using immunofluorescence analysis, we evaluated the impact of PHF8 depletion on H4K20me1 global levels in astrocytes, finding a slight increase in H4K20me1 levels (Fig. 6A; Fig. S7A), as previously demonstrated in other cellular contexts (Liu et al., 2010). Given

that the H4K20me1 mark serves as substrate for the histone methyltransferase SUV420H1 (KMT5B), we also tested the levels of H4K20me3 and found a clear increase in the H4K20me3 heterochromatic mark. Intriguingly, both the intensity and number of H4K20me3 foci increased upon PHF8 depletion (Fig. 6B), whereas no changes in other histone marks associated with transcriptional activation, such as H3K4me3, were detected (Fig. S7B).

Given that depletion of PHF8 in astrocytes led to an increase in global H4K20me3–heterochromatin-related marks, we investigated whether PHF8 is important in preventing the accumulation of H4K20me at PHF8-regulated genes during astrocyte differentiation. We chose two PHF8-target genes identified by ChIP-seq and essential for astrogenesis (*Nfia*) and synaptogenesis (*Sparc*) and tested the effect of PHF8 depletion on H4K20me1 levels in NSCs,



**Fig. 6. PHF8 maintains low levels of H4K20me1/3 at key astrogenic and synaptogenic genes.** (A,B) Astro-CTR and Astro PHF8-KD were fixed and immunostained using H4K20me1 (A) and H4K20me3 (B) antibodies and DAPI. Magnifications show H4K20me3 foci in Astro-CTR and Astro PHF8-KD. Cell fluorescence for H4K20me1 and H4K20me3 staining and H4K20me3 foci was measured using ImageJ. Images are representative of at least three biological independent experiments. More than 30 cells per population were quantified. Box plots represent the quantification of the fluorescence intensities and the number of H4K20me3 foci/cell in Astro-CTR and Astro PHF8-KD. Boxes show magnified regions on the right. \*\*\*\* $P < 0.0001$ . (C) Levels of H4K20me1 histone marks in NSCs, Astro-CTR and Astro PHF8-KD were determined by ChIP-qPCR at the indicated genes. The *Olig2* TSS region devoid of H4K20me1 was used as a negative control. Data from qPCR were normalized to the input, the IgG values were subtracted and the final data were expressed as fold enrichment over the values obtained in the shCTR. 'Intra' refers to the intragenic region identified in the PHF8 ChIP-seq assays. Data are mean $\pm$ s.d.; \* $P < 0.05$ ; \*\*\* $P < 0.001$ . (D) Expression levels of the indicated genes in NSCs, Astro-CTR and Astro PHF8-KD were determined by qPCR. Values were normalized to the housekeeping gene *Gapdh* and are shown relative to time 0 h. *Olig2* mRNA was used as a negative control. Data are mean $\pm$ s.d.;  $n = 3$  independent biological replicates. \* $P < 0.05$ ; \*\* $P < 0.01$ . (E) The levels of H4K20me3 in Astro-CTR and Astro PHF8-KD were determined by ChIP-qPCR. The *Olig2* TSS region was used as a negative control. Data from qPCR were normalized to the input, the IgG values were subtracted and the final data were expressed as fold enrichment over the values obtained in shCTR. 'Intra' refers to intragenic region identified in the PHF8 ChIP-seq assay. Errors bars represent s.d. \* $P < 0.05$ . Scale bars: 4  $\mu$ m in B (magnifications); 20  $\mu$ m in B (main images). TSS, transcription start site.

control astrocytes and PHF8-depleted astrocytes by ChIP-qPCR. A clear decrease in H4K20me1 mark was noticed upon astrocyte differentiation (Fig. 6C) (comparing NSCs and Astro-CTR) that correlated with gene activation (Fig. 6D); no changes at the *Olig2* promoter (a non-PHF8 target used as a negative control) were observed (Fig. 6C). Interestingly, there was no decrease in H4K20me1 in PHF8-depleted astrocytes (Fig. 6C), indicating that PHF8 catalytic activity might be involved in their regulation. We also analyzed the consequences of PHF8 depletion on H4K20me3 levels. A clear increase in H4K20me3 was observed in PHF8 KD cells (Fig. 6E), in agreement with the increased global levels (Fig. 6B). Given that PHF8 also targets the H3K9me2 histone mark (Horton et al., 2010; Kleine-Kohlbrecher et al., 2010), we analyzed whether PHF8 acts by demethylating H3K9me2 on the regions identified by ChIP-seq. PHF8 was not found to be responsible for maintaining low levels of H3K9me2 at the analyzed genes in astrocytes, because no increase in H3K9me2 was observed upon PHF8 depletion (Fig. S7C).

### PHF8 HDM activity is important for astrocyte differentiation

Given that changes in H4K20me1/3 levels correlated with transcriptional changes in genes involved in astrocytic differentiation and synapses (Fig. 6C-E), we assessed the role of PHF8 HDM activity in the observed phenotypes. We established PHF8-KD NSC cell lines that overexpressed either WT PHF8 or a PHF8 mutant lacking HDM activity (mutant H247A) (see Materials and Methods) (Fig. S7D). First, we rescued the defects on astrocyte differentiation by analyzing GFAP protein in differentiated cells. GFAP expression was recovered upon PHF8 WT overexpression, but not upon overexpression of the catalytic mutant (Fig. 7A). Next, we tested the role of the catalytic activity in the control of synaptic and astrocytic gene expression. qPCR demonstrated that PHF8 WT, but not the PHF8 mutant, was able to rescue the expression levels of the tested astrogenic and synaptogenic genes regulated by PHF8 during differentiation without affecting the expression of *Kdm5b* (used as a negative control) (Fig. 7B). Finally, the importance of PHF8 HDM activity was demonstrated by the rescue of H4K20me3 levels after overexpression of PHF8 WT, but not of the catalytic mutant (Fig. 7C). Notably, in the case of overexpression of the PHF8 catalytic mutant, an apparent increase in H4K20me3 intensity was observed, further highlighting the importance of the enzymatic activity of PHF8 in preventing heterochromatin mark accumulation during astrocyte differentiation. These data strongly suggest that PHF8 demethylates H4K20me1 at genes crucial for astrocytic differentiation and synapsis formation, such as *Nfia* and *Sparc*, respectively, to facilitate their transcriptional activation. Taken together, our data indicate that the major role of astrocytic PHF8 is to demethylate H4K20me1, preventing ectopic heterochromatin formation during differentiation.

### DISCUSSION

Our study reveals an unexpected role of the XLID gene *Phf8* in astrocytes, the most abundant glial cells in the mammalian brain. Our data demonstrated that PHF8 directly regulates the expression of the astrogenic master gene *Nfia* and of genes that are essential for synaptic formation and function (Fig. 7D). Indeed, depletion of astrocytic PHF8 resulted in decreased density and strength of excitatory synapses in neuron-astrocyte co-cultures *in vitro*.

### PHF8 targets *Nfia* during astrocyte differentiation

We identified the transcription factor NFIA as a PHF8 transcriptional target (Fig. 7D). The essential role of NFIA in

activating the expression of astrocyte-specific genes and facilitating astrocytic differentiation has been extensively reported (Deneen et al., 2006; Piper et al., 2010; Shu et al., 2003). In particular, *in vivo* studies have demonstrated that *Nfia*<sup>-/-</sup> mice have normal expression of astrocyte markers, but decreased levels of GFAP in the cortex and the hippocampus (das Neves et al., 1999), similar to our observations in PHF8-KD astrocytes *in vitro*. More recent studies have demonstrated that NFIA occupies and regulates the *Gfap* promoter prior to the induction of astrocyte differentiation (Cebolla and Vallejo, 2006; Namihira et al., 2009; Piper et al., 2010). Recently, a study demonstrated that *Nfia* loss in astrocytes leads to diminution of the synaptic function (Huang et al., 2020). Thus, the downregulation of *Nfia* observed in PHF8-depleted astrocytes could be responsible, at least in part, for the phenotype and impaired synaptic transmission of PHF8-KD astrocytes. Furthermore, it has been previously reported that *Nfia* is a Notch target gene (Namihira et al., 2009; Piper et al., 2010). Thus, it is likely that PHF8 regulates *Nfia* expression through the activation of this pathway. We and others (Yatim et al., 2012) demonstrated that PHF8 is a modulator of the Notch signaling pathway. Moreover, we have revealed that *Phf8* is itself a Notch target, suggesting the existence of a regulatory feedback mechanism responsible for *Nfia* transcriptional control, which deserves further studies.

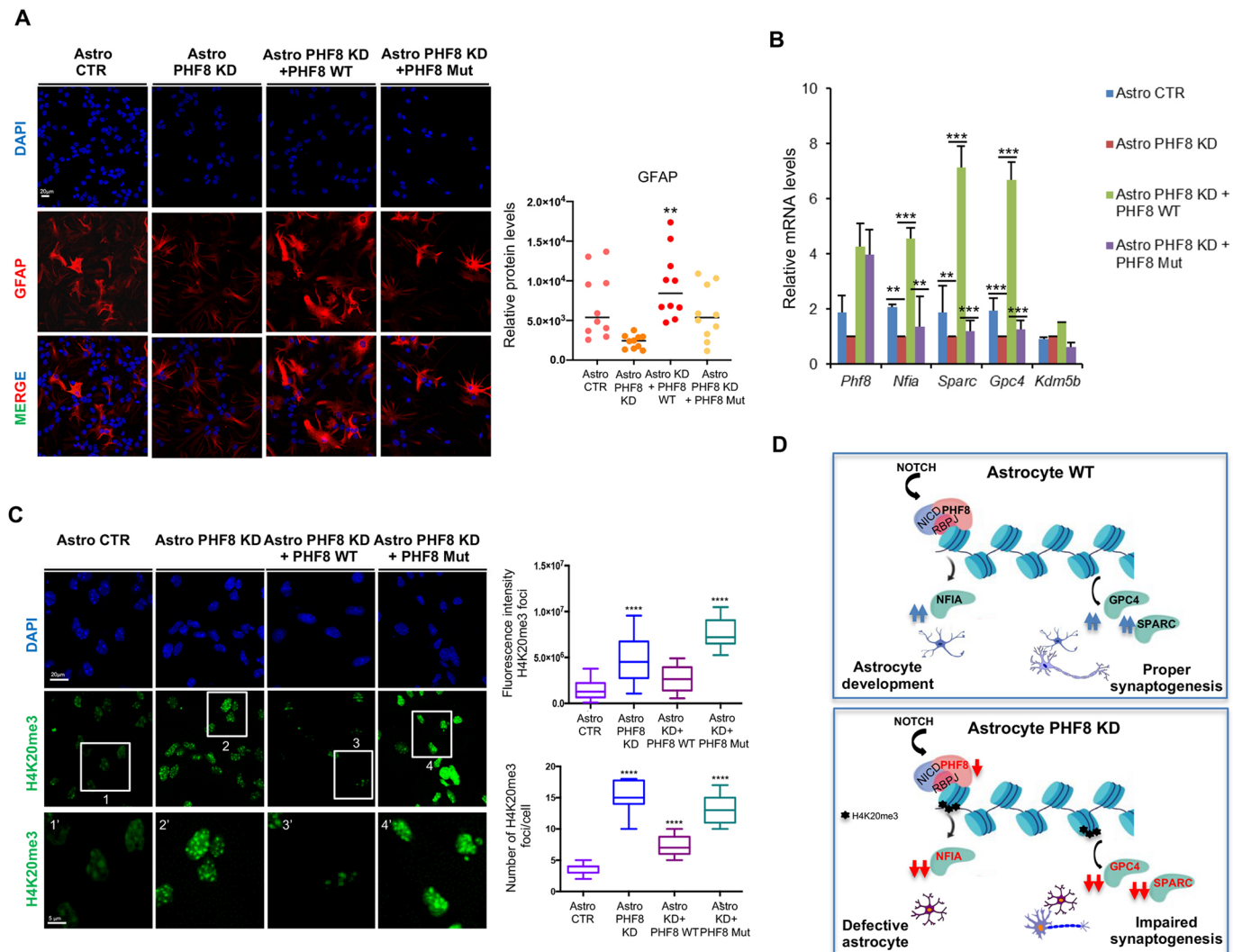
### A subpopulation of PHF8 KD astrocytes expresses OLIG2

Upon PHF8 depletion, a subpopulation of astrocytes expressed high levels of OLIG2 markers. Previous studies showed that some astrocytes express OLIG2 in the gray matter of the mouse spinal cord, thalamus and forebrain (Barnabe-Heider et al., 2010; Griemsmann et al., 2015; Guo et al., 2011; Ohayon et al., 2019; Tatsumi et al., 2018). Thus, PHF8 levels might contribute to the redirection of astrocyte subtypes. However, our transcriptomic analysis revealed that PHF8-depleted astrocytes were not enriched in the astrocyte subtypes (A1-A3) expressing OLIG2 (Batiuk et al., 2020), indicating that PHF8 depletion was not involved in the specification of those astrocyte subtypes.

In addition to *Olig2*, we observed the increased expression of another typical oligodendrocyte gene in PHF8-depleted astrocytes: *Plp1*. Given that *Plp1* is not a PHF8 direct target, its transcriptional alteration is probably an indirect consequence of the primary PHF8-induced transcriptional changes. Interestingly, NFIA has also been involved in oligodendrocyte differentiation; thus, its downregulation, resulting from PHF8 loss, might indirectly lead to the upregulation of some oligodendrocyte genes, as we observed from the RNA-seq results.

### PHF8 regulates the expression of synaptic genes

In addition to *Nfia*, genes involved in synapse formation were found to be regulated directly by astrocytic PHF8 in our study (Fig. 7D). It is well known that the structural and functional interactions of astrocytes with neurons at synapses are necessary for proper brain function (Araque et al., 2014). In particular, astrocytes participate in synaptic plasticity, i.e. the ability of synapses to strengthen or weaken over time, a key process underlying cognitive performance (Bliss and Collingridge, 1993), by promoting synapse formation or pruning aberrant synapses (Clarke and Barres, 2013; Liddel and Barres, 2015). PHF8 defects were previously linked to XLID (Koivisto et al., 2007; Loenarz et al., 2010; Qiao et al., 2008), although the underlying mechanism was poorly understood. Previous studies have shown that *Phf8*-KO mice are deficient in learning and memory (Chen et al., 2018) and are resistant to anxiety- and depression-like behaviors (Walsh et al., 2017).



**Fig. 7. PHF8 HDM activity is important for astrocyte differentiation.** (A) Control and PHF8-depleted NSCs were differentiated to astrocytes over 6 days. Expression of either WT PHF8 or the catalytic mutant (H247A) (Mut) in Astro PHF8-KD was then induced by doxycycline addition and the cells were allowed to differentiate for a further 6 days. The cells were immunostained using GFAP antibody and DAPI (left). The relative protein levels of GFAP per cell were determined using ImageJ (right).  $**P < 0.01$ . (B) WT PHF8 and the catalytic domain were expressed in Astro PHF8-KD and the expression levels of the indicated genes were determined by qPCR. Expression values were normalized to the housekeeping gene *Gapdh*. Values are shown relative to Astro PHF8-KD. *Kdm5b* mRNA was used as a negative control. Data are mean  $\pm$  s.d.;  $n = 3$  independent biological replicates.  $**P < 0.01$ ;  $***P < 0.001$ . (C) WT PHF8 and the catalytic mutant were expressed in Astro PHF8-KD; the levels of H4K20me3 were then established by immunostaining assays. Magnified images show H4K20me3 foci. Images are representative of three biological independent experiments. Fluorescence intensity of H4K20me3 and H4K20me3 foci number were determined using ImageJ. Box plots represent the quantification of the fluorescence intensities and the number of H4K20me3 foci/cell in each population. Boxes show magnified regions below.  $****P < 0.0001$ . (D) Model depicting the contribution of PHF8 to astrocyte differentiation and function. PHF8 directly regulates the expression of the master regulator of astrocyte differentiation *Nfia* as well as genes involved in synapse differentiation and function. Depletion or alteration of PHF8 catalytic activity leads to distinct astrocytes that have deficient synaptic function. Scale bars: 5  $\mu$ m in C (magnifications); 20  $\mu$ m in C (main images).

However, only subtle PHF8-mediated transcriptional changes were observed in neurons and the effects on glial cells were not evaluated. Here, we showed that PHF8 induces profound transcriptional changes in astrocytes, which, in turn, cause important alterations of synaptic transmission *in vitro*. Interestingly, our results demonstrated that synaptic genes are both up- (*Slc1a2* and *Nrxn1*) and downregulated (*Sparc* and *Gpc4*) in PHF8-depleted astrocytes (although downregulated genes showed higher FC values; Fig. S2F) suggesting that PHF8 might be involved in balancing the expression of these genes in astrocytes to regulate synaptic function. Thus, our results suggest that astrocytes might be the main cellular target of PHF8.

### PHF8 prevents ectopic heterochromatin formation

Importantly, our study also unveiled the molecular mechanism behind PHF8-mediated transcriptional changes. Our results demonstrated that PHF8 maintains the expression of astrocytic and synaptic genes by maintaining low levels of the H4K20me1 histone mark. PHF8 depletion led to ectopic heterochromatin formation at both global and local levels. Thus, PHF8 HDM activity is crucial for maintenance of the H4K20me1/H4K20me3 equilibrium. Interestingly, the heterochromatic mark H4K20me3 increased, in terms of both its global level and number of foci, upon PHF8 depletion. These data suggest that H4K20me1 is used as a substrate by the histone methyltransferases SUV420H1/2 to



generate H4K20me3. Consistent with this finding, PHF2, another member of the KDM7 family, limits the accumulation of another heterochromatic mark, H3K9me3, at promoters of cell cycle-related genes in NSCs (Pappa et al., 2019). By contrast, our results demonstrated that depletion of PHF8 led to elevated H4K20me1 levels at the synaptic genes, which correlated with low transcription levels. These results are in agreement with previous studies demonstrating that depletion of PHF8 in neurons resulted in the downregulation of cytoskeleton genes by increasing H4K20me1 levels (Asensio-Juan et al., 2012).

Our work significantly advances current knowledge of the physiological role of PHF8 in astrocyte differentiation and synaptic formation *in vitro*, suggesting that PHF8 may be a key regulator of astroglialogenesis and synaptogenesis. These findings pave the way for the development of pharmacological interventions aimed at improving cognitive function in XLID. Moreover, they prompt us to investigate the contribution of H4K20 methylation not only to neurodevelopmental disorders, but also to other pathological conditions, such as cancer, in which PHF8 is involved. This will increase our understanding of the crosstalk between epigenetics, development and disease.

## MATERIALS AND METHODS

### Cell culture and differentiation

Mouse NSCs were dissected from cerebral cortices of C57BL/6J mouse fetal brains (E12.5) and cultured in poly-D-lysine (5 µg/ml, 2 h at 37°C)- and laminin (5 µg/ml, 4 h at 37°C)-precoated dishes following previously published procedures (Currel et al., 2007). NSCs were grown in medium containing equal parts of Dulbecco's Modified Eagle Medium/Nutrient Mixture (DMEM F12; without Phenol Red; Thermo Fisher Scientific) and Neurobasal medium (Thermo Fisher Scientific) with penicillin/streptomycin (5%), GlutaMAX (1%), N2 and B27 supplements (Thermo Fisher Scientific), sodium pyruvate (1 mM), nonessential amino acids (0.1 mM), heparin (2 mg/l), HEPES (5 mM), bovine serum albumin (BSA; 25 mg/l) and β-mercaptoethanol (0.01 mM), as previously described (Estarras et al., 2012). Fresh recombinant human epidermal growth factor (EGF) (R&D Systems) and fibroblast growth factor (FGF) (Invitrogen) to final concentrations of 20 ng/ml and 10 ng/ml, respectively, were added to the media. Under these conditions, NSCs maintain the ability to self-renew and to differentiate into a wide range of neural cell types (Currel et al., 2007; Pollard et al., 2006).

For NSC differentiation into astrocytes, the medium was replaced with a differentiation medium (astrocytic medium) containing DMEM/F-12, 5% N2, GlutaMAX and 10% fetal bovine serum (FBS); fresh astrocytic medium was supplied every 2 days.

Human HEK 293T cells were maintained in culture under standard conditions (Blanco-Garcia et al., 2009), cultured in DMEM supplemented with 10% FBS.

### Astrocyte-neuron co-cultures

Primary neuronal cultures were obtained from the hippocampus of 18-day-old fetal C57BL/6 wild-type mice (Charles River). Briefly, dissociated cells were plated onto previously differentiated astrocytes (10 days) in 12-well multiwell plates at a seeding density of  $0.3 \times 10^6$  and maintained in Neurobasal medium supplemented with 2% B27 (Life Technologies), penicillin/streptomycin (1%), L-glutamine (0.5 mM) and glutamate (12 µM). Cultures were maintained in standard conditions at 37°C and 5% CO<sub>2</sub>. After 3 days *in vitro*, the medium was partially replaced by fresh medium.

### Antibodies and reagents

The antibodies and reagents used were: anti-PHF8 (Abcam, ab36068; western blot 1:1000; ChIP 1:500; immunocytochemistry 1:500), anti-H4K20me1 (Abcam, ab9051 ChIP 2 µg/ml; immunocytochemistry 1:500), anti-H4K20me3 (Abcam, ab9053; ChIP 2 µg/ml; immunocytochemistry 1:500),

anti-H3K9me2 (Abcam, ab1220; ChIP 2 µg/ml; immunocytochemistry 1:500), anti-H3K27me3 (Millipore, 07449; 1:500), H3K4me3 (Abcam, ab8580; 1:500), anti-GAPDH (Synaptic Systems, 247002; 1:500), anti-β-tubulin 3 (TUJ1; BioLegend, MMS-435P; 1:500), anti-GFAP (Agilent Dako, z0334, and Synaptic Systems, 173 004; 1:500), anti-nestin (Abcam, ab5968; 1:500), anti-OLIG2 (Merck, AB9610; 1:500), anti-alpha-tubulin (Abcam, ab4074; 1:10,000), anti-GLAST (Abcam, ab416; 1:500), anti-aquaporin 4 (Abcam, ab125049; 1:500), anti-EAAT2 (Abcam, ab41621; 1:500), anti-NG2 chondroitin sulfate proteoglycan (Merck, AB5320; 1:500), anti-GPR17 (Cayman Chemical, 10136; 1:500), anti-IBA-1 (Thermo Fisher Scientific, PA5-27436; 1:500), anti-bassoon (Synaptic Systems, 141004; 1:500), anti-SHANK2 (Synaptic Systems, 162202; 1:500), anti-SHANK3 (Synaptic Systems, 162304; 1:500) and DAPI (Thermo Fisher Scientific, D1306).

### Plasmids and recombinant proteins

Previously published specific lentiviral vectors were either purchased from Sigma-Aldrich or cloned in the pLKO.1-puro vector (Sigma-Aldrich) using the AgeI and EcoRI sites (brackets indicate the target sequence): pLKO-random (CAACAAGATGAAGAGCACC) and pLKO-mPHF8\_1 (GCAGGTAATGGGAGAGGTT). PHF8 human cDNA from pEF6-HA-PHF8 (kindly provided by Dr C. Leonarz (Institute of Pharmaceutical Sciences, University of Freiburg, Germany) was cloned into the pINDUCER vector (#44012, Addgene) between the attB1 and attB2 sites. Both PHF8 WT and catalytic mutant H247A were induced upon doxycycline addition (1 µg/ml). pCIG-FLAG-SMAD3-S/D, BMP4 and BMP7 were kindly provided by Dr Elisa Martí (IBMB, CSIC, Barcelona, Spain). Primer sequences are described in Table S1.

### Lentiviral transduction

Lentiviral transduction was carried out as previously described (Asensio-Juan et al., 2017). Briefly, HEK 293T cells were transfected with a mix of packaging, envelop and shRNA transfer vector DNAs (6, 5 and 7 µg, respectively). The medium was collected 24-30 h later and the virus concentrated by ultracentrifugation [26,000 rpm (121.139 g), 2 h at 4°C]. Viral particles were then used to transduce NSCs. After 24 h, cells were selected with either puromycin (2 µg/ml; Merck, P8833) (pLKO.1 vectors) or geneticin (600 µg/ml; Merck, 345810) (pINDUCER vectors). After selection, 99-100% of the cells expressed the shRNA.

### ChIP assays

ChIP assays were performed as previously described (Valls et al., 2007) with modifications:  $10^6$  NSCs were fixed with formaldehyde (1%) for 10 min. Fixation was stopped by adding 0.125 mM glycine. Cells were lysed in lysis buffer (1% SDS; 10 mM EDTA, pH 8.0; 50 mM Tris-HCl, pH 8.1). Chromatin fragmentation was performed in a Bioruptor sonicator (Diagenode) before immunoprecipitation. The immunocomplex was captured using Magna ChIP Protein A Magnetic Beads (Millipore). After decrosslinking, DNA was purified by ethanol precipitation. ChIP DNA was quantified by qPCR with SYBR Green (Roche) in a LightCycler 480 PCR system (Roche) using the primers indicated in Table S1. The input percentage was used for quantification of the immunoprecipitated material with respect to the starting chromatin. Integrative Genomics Viewer (IGV) (<https://software.broadinstitute.org/software/igv/>) was used to visualize a particular region.

### Indirect immunofluorescence and quantification

Immunofluorescence assays were performed as previously described (Sanchez-Molina et al., 2014). Cells were fixed for 20 min in 4% paraformaldehyde and permeabilized with PBS-Triton X-100 (0.1%) before blocking at room temperature for 1 h in 1% BSA (in PBS with 0.1% Triton X-100). Primary antibodies were used overnight at 4°C. Finally, cells were incubated for 2 h at room temperature with Alexa-conjugated secondary antibodies and DAPI (0.1 ng/µl) (Sigma-Aldrich). Images were captured using a Leica SP5 confocal microscope with LAS-AF software. Two different methods of quantification were performed depending on the experiment. In the case of histone marks and PHF8 immunostaining, the

fluorescence intensity per cell was measured using ImageJ and the corrected cell fluorescence was calculated using the formula: integrated density—(area of selected cell×mean fluorescence of background readings).

For H4K20me3, the number of foci per cell was also calculated. Finally, the percentage of positive cells per field was determined for the cell-lineage markers nestin, GLAST, GFAP, TUBβ3 and OLIG2. The maximum and minimum cell body diameters of neurons were measured on immunofluorescence confocal images of co-cultures with PHF8-KD or CTR astrocytes, fixed and stained for TUBβ3 and DAPI, using the straight-line tool of ImageJ software followed by the Analyze>Measure command. Neuronal cell density was evaluated in the same images by counting the number of neuronal cell bodies in the acquired area. The quantification for the co-culture experiments was performed as indicated above.

### RNA extraction and qPCR

TRIzol reagent (Invitrogen) was used to extract RNA following the manufacturer's instructions. Reverse transcription was performed with 2 μg of RNA using High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) and qPCR was performed with SYBR Green (Roche) in a LightCycler 480 (Roche). Specific primer pairs are detailed in Table S1.

### Western blot

Immunoblotting was performed using standard procedures (Towbin et al., 1979) and visualized by means of an ECL kit (Merck, GERPN2106)

### MACS isolation of astrocyte

Purified astrocytes were isolated from E18 embryos or P2 or P7 mouse whole brains by MACS (Miltenyi Biotec) with anti-GLAST (ACSA 2) MicroBeads according to the manufacturer's instructions. Cells were lysed with a buffer containing 1% SDS, 2 mM EDTA pH 7.4, 10 mM Tris-HCl pH 7.4 and protease inhibitor (1:100; Merck, P8340).

### ChIP-seq

ChIP was carried out as previously described (Fueyo et al., 2018) with minor modifications, described as follows. The sonication step was performed in a Bioruptor sonicator. The input sample, corresponding to the 10% of the total material, was reserved at this point. The PHF8 antibody (Abcam, ab36068) used for immunoprecipitation was the same as that used in the ChIP assays described above. After decrosslinking, DNA was purified by ethanol precipitation. The libraries were prepared and sequenced in a HiSeq 2000 Sequencing System (Illumina) (Table S2). In total, 50 base pairs sequences were mapped using Bowtie2 (Langmead and Salzberg, 2012) to the *Mus musculus* genome release 10 (mm10); files were filtered to remove duplicates and peaks were called using MACS (Zhang et al., 2008) and an effective genome size of 1.87 Gb and a *P*-value of 0.001 for PHF8 ChIP-seq. The Bioconductor package ChIPseeker (Yu et al., 2015) was used to annotate the genes of each peak. Specifically, the function annotatePeak matches peaks with genomic features extracted from mm10 (UCSC) and calculates the proportion of peaks matching each feature. Homer tool (<http://homer.ucsd.edu/homer/motif/>) was used to identify DNA-binding motif enrichment. For further details, see Table S2. ChIP-seq data have been deposited in the GEO database under the accession GSE141969.

### RNA-seq

RNA was extracted using High Pure RNA Isolation Kit (Roche) followed by DNaseI treatment from two biological independent samples. Libraries were prepared using the TruSeq Stranded Total RNA Sample Preparation Kit with Ribo-Zero Human/Mouse/Rat Kit (Illumina) according to the manufacturer's protocol. Briefly, 500 ng of total RNA was used for ribosomal RNA (rRNA) depletion. Then, rRNA-depleted RNA was fragmented for 4.5 min. The remaining steps of the library preparation were followed according to the manufacturer's instructions. Final libraries were analyzed using the Agilent DNA 1000 Kit to estimate the quantity and check the size distribution, and were then quantified by qPCR using the KAPA Library Quantification Kit (Roche) prior to amplification with cBot (Illumina). The libraries were sequenced on Illumina High HiSeq 2500

Sequencing System with paired-end 50 base pair-long reads (Table S2). Alignment was performed using STAR (Dobin et al., 2013); aligned reads were assigned to genes using HTSeq (Anders et al., 2015) and differential expression analysis was performed using DESeq2 (Love et al., 2014). RNA-seq data have been deposited in the GEO database under the accession GSE141970.

### Analysis of synapses

The percentage of synapses with pre- and postsynaptic terminals was determined in a single confocal plane using ImageJ software. Regions of interest (ROIs) were drawn using the freehand line tool on TUBβ3 images and the length of the segments was measured using the analyze function. Bassoon, SHANK2 and double-positive puncta were counted by generating merge images of bassoon/SHANK2/TUBβ3. Synapse density was calculated dividing the number of puncta by the ROI length, which corresponds to the dendrite length.

### Electrophysiology

Whole-cell voltage clamp recordings were performed using a MultiClamp 700A amplifier (Molecular Devices), a 1320A Digidata (Molecular Devices), coupled to pCLAMP 10 Software (Molecular Devices) and an inverted Axiovert 200 microscope (Zeiss). mEPSCs were recorded from DIV 12-14 neurons in Krebs-Ringer's HEPES solution (KRH) (125 mM NaCl, 5 mM KCl, 1.2 mM MgSO<sub>4</sub>, 1.2 mM KH<sub>2</sub>PO<sub>4</sub>, 2 mM CaCl<sub>2</sub>, 6 mM D-glucose and 25 mM HEPES/NaOH, pH 7.4), supplemented with 1 μM tetrodotoxin (Abcam, ab120055) and 20 μM bicuculline (Abcam, ab120110). Experiments were performed at room temperature (20-25°C), setting the holding potential at -70 mV and using the following internal solution: (130 mM potassium gluconate, 10 mM KCl, 1 mM EGTA, 10 mM HEPES, 2 mM MgCl<sub>2</sub>, 4 mM MgATP, 0.3 mM Tris-GTP; pH 7.4, adjusted with KOH). Recording pipettes were pulled from patch-clamp borosilicate capillary glass (World Precision Instruments) to a tip resistance of 3-5 MΩ using a two-stage vertical puller (Narishige). Traces were sampled at 10 kHz and filtered at 2 kHz. Series resistance was monitored during recording. mEPSCs were detected offline using Clampfit software (Molecular Devices) setting a threshold of 7 pA.

Resting membrane potential was evaluated immediately after breaking the cell-attached patch, by setting the amplifier to zero current mode (*I*=0).

### Statistical analysis

Quantitative data were expressed as mean±standard deviation (s.d.). The Mann-Whitney Rank Sum Test was used to quantify the frequency and amplitude of the excitatory postsynaptic currents. The Jaccard index was used to determine the similarity between the samples; it measures the similarity between two nominal attributes by taking the intersection of both and dividing it by their union (Vorontsov et al., 2013). The significance of differences between two groups was assessed using the Student's *t*-test; two-way ANOVAs were used for three or more groups and the relative viabilities were compared using Tukey's multiple comparisons test (\**P*<0.05, \*\**P*<0.01, \*\*\**P*<0.001, \*\*\*\**P*<0.0001).

### Acknowledgements

We would like to thank Dr Christoph Loenarz for the HA-PHF8 vector; Drs A. Jordá, F. Azorín, J. Bernues, G. Vicent, A. Vaquero and E. Martí for reagents; M. Arbones and M. J. Barallobre for advice; and team members for helpful comments and suggestions.

### Competing interests

The authors declare no competing or financial interests.

### Author contributions

Conceptualization: S.I., C.V., M.A.M.-B.; Methodology: S.I., N.P., M.G., C.N., M.L., C.V., X.C., M.A.M.-B.; Software: N.P., X.C.; Validation: S.I., M.V.-M.; Formal analysis: N.P., X.C.; Investigation: S.I., M.G., C.N., M.L.; Resources: C.V., M.A.M.-B.; Data curation: M.G., X.C., M.A.M.-B.; Writing - original draft: M.A.M.-B.; Writing - review & editing: S.I., N.P., C.V., X.C.; Supervision: M.A.M.-B.; Project administration: M.A.M.-B.; Funding acquisition: C.V., M.A.M.-B., X.C.

## Funding

This study was supported by grants from the Spanish Ministry of Economy (Ministerio de Economía, Industria y Competitividad, Gobierno de España; BFU2015-69248-P and PGC2018-096082-B-I00 to M.A.M.-B.) and Ministerio de Ciencia e Innovación (PID2019-111217RB-I00 to X.C) and a Travelling Fellowship from Boehringer Ingelheim to S.I. Open access funding provided by Consejo Superior de Investigaciones Científicas. Deposited in PMC for immediate release.

## Data availability

ChIP-seq data have been deposited in the Gene Expression database under accession number GSE141969. RNA-seq data have been deposited in the Gene Expression Omnibus database under accession number GSE141970.

## Peer review history

The peer review history is available online at <https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.194951>

## References

- Abidi, F. E., Miano, M. G., Murray, J. C. and Schwartz, C. E. (2007). A novel mutation in the PHF8 gene is associated with X-linked mental retardation with cleft lip/cleft palate. *Clin. Genet.* **72**, 19-22. doi:10.1111/j.1399-0004.2007.00817.x
- Anders, S., Pyl, P. T. and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169. doi:10.1093/bioinformatics/btu638
- Araque, A., Carmignoto, G., Haydon, P. G., Oliet, S. H. R., Robitaille, R. and Volterra, A. (2014). Gliotransmitters travel in time and space. *Neuron* **81**, 728-739. doi:10.1016/j.neuron.2014.02.007
- Asensio-Juan, E., Gallego, C. and Martínez-Balbás, M. A. (2012). The histone demethylase PHF8 is essential for cytoskeleton dynamics. *Nucleic Acids Res.* **40**, 9429-9440. doi:10.1093/nar/gks716
- Asensio-Juan, E., Fueyo, R., Pappa, S., Iacobucci, S., Badosa, C., Lois, S., Balada, M., Bosch-Presegué, L., Vaquero, A., Gutiérrez, S. et al. (2017). The histone demethylase PHF8 is a molecular safeguard of the IFN $\gamma$  response. *Nucleic Acids Res.* **45**, 3800-3811. doi:10.1093/nar/gkw1346
- Barnabé-Heider, F., Göritz, C., Sabelström, H., Takebayashi, H., Pfrieger, F. W., Meletis, K. and Frisén, J. (2010). Origin of new glial cells in intact and injured adult spinal cord. *Cell Stem Cell* **7**, 470-482. doi:10.1016/j.stem.2010.07.014
- Batiuk, M. Y., Martirosyan, A., Wahis, J., de Vin, F., Marneffe, C., Kusserow, C., Koeppen, J., Viana, J. F., Oliveira, J. F., Voet, T. et al. (2020). Identification of region-specific astrocyte subtypes at single cell resolution. *Nat. Commun.* **11**, 1220. doi:10.1038/s41467-019-14198-8
- Blanco-García, N., Asensio-Juan, E., de la Cruz, X. and Martínez-Balbás, M. A. (2009). Autoacetylation regulates P/CAF nuclear localization. *J. Biol. Chem.* **284**, 1343-1352. doi:10.1074/jbc.M806075200
- Bliss, T. V. P. and Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31-39. doi:10.1038/361031a0
- Cebolla, B. and Vallejo, M. (2006). Nuclear factor- $\kappa$ B regulates glial fibrillary acidic protein gene expression in astrocytes differentiated from cortical precursor cells. *J. Neurochem.* **97**, 1057-1070. doi:10.1111/j.1471-4159.2006.03804.x
- Chelly, J., Khelifaoui, M., Francis, F., Chérif, B. and Bienvenu, T. (2006). Genetics and pathophysiology of mental retardation. *Eur. J. Hum. Genet.* **14**, 701-713. doi:10.1038/sj.ejhg.5201595
- Chen, X., Wang, S., Zhou, Y., Han, Y., Li, S., Xu, Q., Xu, L., Zhu, Z., Deng, Y., Yu, L. et al. (2018). Phf8 histone demethylase deficiency causes cognitive impairments through the mTOR pathway. *Nat. Commun.* **9**, 114. doi:10.1038/s41467-017-02531-y
- Cheng, C., Lau, S. K. M. and Doering, L. C. (2016). Astrocyte-secreted thrombospondin-1 modulates synapse and spine defects in the fragile X mouse model. *Mol. Brain* **9**, 74. doi:10.1186/s13041-016-0256-9
- Clarke, L. E. and Barres, B. A. (2013). Emerging roles of astrocytes in neural circuit development. *Nat. Rev. Neurosci.* **14**, 311-321. doi:10.1038/nrn3484
- Cresto, N., Pillet, L. E., Billuart, P. and Rouach, N. (2019). Do astrocytes play a role in intellectual disabilities? *Trends Neurosci.* **42**, 518-527. doi:10.1016/j.tins.2019.05.011
- Currie, D. S., Hu, J. S., Kolski-Andreaco, A. and Monuki, E. S. (2007). Culture of mouse neural stem cell precursors. *J. Vis. Exp.* **2**, e152. doi:10.3791/152
- das Neves, L., Duchala, C. S., Godinho, F., Haxhiu, M. A., Colmenares, C., Macklin, W. B., Campbell, C. E., Butz, K. G. and Gronostajski, R. M. (1999). Disruption of the murine nuclear factor I-A gene (Nfia) results in perinatal lethality, hydrocephalus, and agenesis of the corpus callosum. *Proc. Natl. Acad. Sci. USA* **96**, 11946-11951. doi:10.1073/pnas.96.21.11946
- Deneen, B., Ho, R., Lukaszewicz, A., Hochstim, C. J., Gronostajski, R. M. and Anderson, D. J. (2006). The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron* **52**, 953-968. doi:10.1016/j.neuron.2006.11.019
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21. doi:10.1093/bioinformatics/bts635
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74. doi:10.1038/nature11247
- Estaras, C., Akizu, N., Garcia, A., Beltran, S., de la Cruz, X. and Martínez-Balbás, M. A. (2012). Genome-wide analysis reveals that Smad3 and JMJD3 HDM co-activate the neural developmental program. *Development* **139**, 2681-2691. doi:10.1242/dev.078345
- Farhy-Tselnicker, I., van Casteren, A. C. M., Lee, A., Chang, V. T., Aricescu, A. R. and Allen, N. J. (2017). Astrocyte-Secreted Glypican 4 regulates release of neuronal Pentraxin 1 from Axons to induce functional synapse formation. *Neuron* **96**, 428-445.e13. doi:10.1016/j.neuron.2017.09.053
- Fortschegger, K. and Shiekhattar, R. (2011). Plant homeodomain fingers form a helping hand for transcription. *Epigenetics* **6**, 4-8. doi:10.4161/epi.6.1.13297
- Fortschegger, K., de Graaf, P., Outchkourov, N. S., van Schaik, F. M. A., Timmers, H. T. M. and Shiekhattar, R. (2010). PHF8 targets histone methylation and RNA polymerase II to activate transcription. *Mol. Cell. Biol.* **30**, 3286-3298. doi:10.1128/MCB.01520-09
- Fueyo, R., Iacobucci, S., Pappa, S., Estarás, C., Lois, S., Vicioso-Mantis, M., Navarro, C., Cruz-Molina, S., Reyes, J. C., Rada-Iglesias, Á. et al. (2018). Lineage specific transcription factors and epigenetic regulators mediate TGF $\beta$ -dependent enhancer activation. *Nucleic Acids Res.* **46**, 3351-3365. doi:10.1093/nar/gky093
- Ge, W., Martinowich, K., Wu, X., He, F., Miyamoto, A., Fan, G., Weinmaster, G. and Sun, Y. E. (2002). Notch signaling promotes astroglialogenesis via direct CSL-mediated glial gene activation. *J. Neurosci. Res.* **69**, 848-860. doi:10.1002/jnr.10364
- Griemsmann, S., Höft, S. P., Bedner, P., Zhang, J., von Staden, E., Beinhauer, A., Degen, J., Dublin, P., Cope, D. W., Richter, N. et al. (2015). Characterization of Panglial Gap junction networks in the thalamus, neocortex, and hippocampus reveals a unique population of glial cells. *Cereb. Cortex* **25**, 3420-3433. doi:10.1093/cercor/bhu157
- Guo, F., Maeda, Y., Ma, J., Delgado, M., Sohn, J., Miers, L., Ko, E. M., Bannerman, P., Xu, J., Wang, Y. et al. (2011). Macrogliar plasticity and the origins of reactive astroglia in experimental autoimmune encephalomyelitis. *J. Neurosci.* **31**, 11914-11928. doi:10.1523/JNEUROSCI.1759-11.2011
- Hillen, A. E. J., Burbach, J. P. H. and Hol, E. M. (2018). Cell adhesion and matricellular support by astrocytes of the tripartite synapse. *Prog. Neurobiol.* **165-167**, 66-86. doi:10.1016/j.pneurobio.2018.02.002
- Horton, J. R., Upadhyay, A. K., Qi, H. H., Zhang, X., Shi, Y. and Cheng, X. (2010). Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* **17**, 38-43. doi:10.1038/nsmb.1753
- Huang, A. Y.-S., Woo, J., Sardar, D., Lozzi, B., Bosquez Huerta, N. A., Lin, C.-C. J., Felice, D., Jain, A., Paulucci-Holthausen, A. and Deneen, B. (2020). Region-specific transcriptional control of astrocyte function oversees local circuit activities. *Neuron* **106**, 992-1008.e9. doi:10.1016/j.neuron.2020.03.025
- Jori, F. P., Galderisi, U., Napolitano, M. A., Cipollaro, M., Cascino, A., Giordano, A. and Melone, M. A. B. (2007). RB and RB2/P130 genes cooperate with extrinsic signals to promote differentiation of rat neural stem cells. *Mol. Cell. Neurosci.* **34**, 299-309. doi:10.1016/j.mcn.2006.11.009
- Kleine-Kohlbrecher, D., Christensen, J., Vandamme, J., Abarrategui, I., Bak, M., Tommerup, N., Shi, X., Gozani, O., Rappsilber, J., Salcini, A. E. et al. (2010). A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation. *Mol. Cell* **38**, 165-178. doi:10.1016/j.molcel.2010.03.002
- Koivisto, A. M., Ala-Mello, S., Lemmelä, S., Komu, H. A., Rautio, J. and Järvelä, I. (2007). Screening of mutations in the PHF8 gene and identification of a novel mutation in a Finnish family with XLMR and cleft lip/cleft palate. *Clin. Genet.* **72**, 145-149. doi:10.1111/j.1399-0004.2007.00836.x
- Kramer, J. M. and van Bokhoven, H. (2009). Genetic and epigenetic defects in mental retardation. *Int. J. Biochem. Cell Biol.* **41**, 96-107. doi:10.1016/j.biocel.2008.08.009
- Kucukdereli, H., Allen, N. J., Lee, A. T., Feng, A., Ozlu, M. I., Conatser, L. M., Chakraborty, C., Workman, G., Weaver, M., Sage, E. H. et al. (2011). Control of excitatory CNS synaptogenesis by astrocyte-secreted proteins Hevin and SPARC. *Proc. Natl. Acad. Sci. USA* **108**, E440-E449. doi:10.1073/pnas.1104977108
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923
- Laumonier, F., Holbert, S., Ronce, N., Faravelli, F., Lenzner, S., Schwartz, C. E., Lespinasse, J., Van Esch, H., Lacombe, D., Goizet, C. et al. (2005). Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. *J. Med. Genet.* **42**, 780-786. doi:10.1136/jmg.2004.029439
- Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D. and Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91-95. doi:10.1038/nature04802
- Liddelow, S. and Barres, B. (2015). SnapShot: astrocytes in health and disease. *Cell* **162**, 1170-1170.e1. doi:10.1016/j.cell.2015.08.029
- Liu, W., Tanasa, B., Tyurina, O. V., Zhou, T. Y., Gassmann, R., Liu, W. T., Ohgi, K. A., Benner, C., Garcia-Bassets, I., Aggarwal, A. K. et al. (2010). PHF8

- mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508-512. doi:10.1038/nature09272
- Loenarz, C., Ge, W., Coleman, M. L., Rose, N. R., Cooper, C. D. O., Klose, R. J., Ratcliffe, P. J. and Schofield, C. J.** (2010). PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an N<sup>ε</sup>-dimethyl lysine demethylase. *Hum. Mol. Genet.* **19**, 217-222. doi:10.1093/hmg/ddp480
- 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**, 550. doi:10.1186/s13059-014-0550-8
- Martini, S., Bernoth, K., Main, H., Ortega, G. D., Lendahl, U., Just, U. and Schwanbeck, R.** (2013). A critical role for Sox9 in notch-induced astrogliogenesis and stem cell maintenance. *Stem Cells* **31**, 741-751. doi:10.1002/stem.1320
- McGann, J. C., Liyo, D. T. and Mandel, G.** (2012). Astrocytes conspire with neurons during progression of neurological disease. *Curr. Opin. Neurobiol.* **22**, 850-858. doi:10.1016/j.conb.2012.03.009
- Namihira, M., Kohyama, J., Semi, K., Sanosaka, T., Deneen, B., Taga, T. and Nakashima, K.** (2009). Committed neuronal precursors confer astrocytic potential on residual neural precursor cells. *Dev. Cell* **16**, 245-255. doi:10.1016/j.devcel.2008.12.014
- Ohayon, D., Escalas, N., Cochard, P., Glise, B., Danesin, C. and Soula, C.** (2019). Sulfatase 2 promotes generation of a spinal cord astrocyte subtype that stands out through the expression of Olig2. *Glia* **67**, 1478-1495. doi:10.1002/glia.23621
- Pappa, S., Padilla, N., Iacobucci, S., Vicioso, M., Álvarez de la Campa, E., Navarro, C., Marcos, E., de la Cruz, X. and Martínez-Balbás, M. A.** (2019). PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitors. *Proc. Natl. Acad. Sci. USA* **116**, 19464-19473. doi:10.1073/pnas.1903188116
- Piper, M., Barry, G., Hawkins, J., Mason, S., Lindwall, C., Little, E., Sarkar, A., Smith, A. G., Moldrich, R. X., Boyle, G. M. et al.** (2010). NFIA controls telencephalic progenitor cell differentiation through repression of the Notch effector Hes1. *J. Neurosci.* **30**, 9127-9139. doi:10.1523/JNEUROSCI.6167-09.2010
- Pollard, S. M., Conti, L., Sun, Y., Goffredo, D. and Smith, A.** (2006). Adherent neural stem (NS) cells from fetal and adult forebrain. *Cereb. Cortex* **16** Suppl. 1, i112-i120. doi:10.1093/cercor/bhj167
- Qi, H. H., Sarkissian, M., Hu, G.-Q., Wang, Z., Bhattacharjee, A., Gordon, D. B., Gonzales, M., Lan, F., Ongusaha, P. P., Huarte, M. et al.** (2010). Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature* **466**, 503-507. doi:10.1038/nature09261
- Qiao, Y., Liu, X., Harvard, C., Hildebrand, M. J., Rajcan-Separovic, E., Holden, J. J. A. and Lewis, M. E. S.** (2008). Autism-associated familial microdeletion of Xp11.22. *Clin. Genet.* **74**, 134-144. doi:10.1111/j.1399-0004.2008.01028.x
- Qiu, J., Shi, G., Jia, Y., Li, J., Wu, M., Dong, S. and Wong, J.** (2010). The X-linked mental retardation gene PHF8 is a histone demethylase involved in neuronal differentiation. *Cell Res.* **20**, 908-918. doi:10.1038/cr.2010.81
- Ropers, H.-H. and Hamel, B. C. J.** (2005). X-linked mental retardation. *Nat. Rev. Genet.* **6**, 46-57. doi:10.1038/nrg1501
- Sánchez-Molina, S., Estarás, C., Oliva, J. L., Akizu, N., Asensio-Juan, E., Rojas, J. M. and Martínez-Balbás, M. A.** (2014). Regulation of CBP and Tip60 coordinates histone acetylation at local and global levels during Ras-induced transformation. *Carcinogenesis* **35**, 2194-2202. doi:10.1093/carcin/bgu111
- Shu, T., Butz, K. G., Plachez, C., Gronostajski, R. M. and Richards, L. J.** (2003). Abnormal development of forebrain midline glia and commissural projections in *Nfia* knock-out mice. *J. Neurosci.* **23**, 203-212. doi:10.1523/JNEUROSCI.23-01-00203.2003
- Siderius, L. E., Hamel, B. C. J., van Bokhoven, H., de Jager, F., van den Helm, B., Kremer, H., Heineman-de Boer, J. A., Ropers, H.-H. and Mariman, E. C. M.** (1999). X-linked mental retardation associated with cleft lip/palate maps to Xp11.3-q21.3. *Am. J. Med. Genet.* **85**, 216-220. doi:10.1002/(SICI)1096-8628(19990730)85:3<216::AID-AJMG6>3.0.CO;2-X
- Singh, S. K., Stogsdill, J. A., Pulimood, N. S., Dingsdale, H., Kim, Y. H., Pilaz, L.-J., Kim, I. H., Manhaes, A. C., Rodrigues, W. S., Jr., Pamukcu, A. et al.** (2016). Astrocytes assemble Thalamocortical Synapses by Bridging NRX1 $\alpha$  and NL1 via Hevin. *Cell* **164**, 183-196. doi:10.1016/j.cell.2015.11.034
- Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X., Fan, G. and Greenberg, M. E.** (2001). Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell* **104**, 365-376. doi:10.1016/S0092-8674(01)00224-0
- Tatsumi, K., Isonishi, A., Yamasaki, M., Kawabe, Y., Morita-Takemura, S., Nakahara, K., Terada, Y., Shinjo, T., Okuda, H., Tanaka, T. et al.** (2018). Olig2-Lineage astrocytes: a distinct subtype of astrocytes that differs from GFAP astrocytes. *Front. Neuroanat.* **12**, 8. doi:10.3389/fnana.2018.00008
- Towbin, H., Staehelin, T. and Gordon, J.** (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.
- Tsukada, Y.-i., Ishitani, T. and Nakayama, K. I.** (2010). KDM7 is a dual demethylase for histone H3 Lys 9 and Lys 27 and functions in brain development. *Genes Dev.* **24**, 432-437. doi:10.1101/gad.1864410
- Valls, E., Blanco-Garcia, N., Aquizu, N., Piedra, D., Estarás, C., de la Cruz, X. and Martínez-Balbás, M. A.** (2007). Involvement of chromatin and histone deacetylation in SV40 T antigen transcription regulation. *Nucleic Acids Res.* **35**, 1958-1968. doi:10.1093/nar/gkl1113
- Vorontsov, I. E., Kulakovskiy, I. V. and Makeev, V. J.** (2013). Jaccard index based similarity measure to compare transcription factor binding site models. *Algorithms Mol. Biol.* **8**, 23. doi:10.1186/1748-7188-8-23
- Walsh, R. M., Shen, E. Y., Bagot, R. C., Anselmo, A., Jiang, Y., Javidfar, B., Wojtkiewicz, G. J., Cloutier, J., Chen, J. W., Sadreyev, R. et al.** (2017). Phf8 loss confers resistance to depression-like and anxiety-like behaviors in mice. *Nat. Commun.* **8**, 15142. doi:10.1038/ncomms15142
- Wen, H., Li, J., Song, T., Lu, M., Kan, P.-Y., Lee, M. G., Sha, B. and Shi, X.** (2010). Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J. Biol. Chem.* **285**, 9322-9326. doi:10.1074/jbc.C109.097667
- Yatim, A., Benne, C., Sobhian, B., Laurent-Chabalier, S., Deas, O., Judde, J.-G., Lelievre, J.-D., Levy, Y. and Benkirane, M.** (2012). NOTCH1 nuclear interactome reveals key regulators of its transcriptional activity and oncogenic function. *Mol. Cell* **48**, 445-458. doi:10.1016/j.molcel.2012.08.022
- Yu, G., Wang, L.-G. and He, Q.-Y.** (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383. doi:10.1093/bioinformatics/btv145
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al.** (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. doi:10.1186/gb-2008-9-9-r137
- Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keefe, S., Phatnani, H. P., Guarnieri, P., Caneda, C., Ruderisch, N. et al.** (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929-11947. doi:10.1523/JNEUROSCI.1860-14.2014
- Zhang, Y., Sloan, S. A., Clarke, L. E., Caneda, C., Plaza, C. A., Blumenthal, P. D., Vogel, H., Steinberg, G. K., Edwards, M. S. B., Li, G. et al.** (2016). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37-53. doi:10.1016/j.neuron.2015.11.013



# PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitors

Stella Pappa<sup>a</sup>, Natalia Padilla<sup>b,c,1</sup>, Simona Iacobucci<sup>a,1</sup>, Marta Vicioso<sup>a</sup>, Elena Álvarez de la Campa<sup>b,c</sup>, Claudia Navarro<sup>a</sup>, Elia Marcos<sup>a</sup>, Xavier de la Cruz<sup>b,c,2</sup>, and Marian A. Martínez-Balbás<sup>a,2</sup>

<sup>a</sup>Department of Molecular Genomics, Instituto de Biología Molecular de Barcelona, Consejo Superior de Investigaciones Científicas, 08028 Barcelona, Spain; <sup>b</sup>Clinical and Translational Bioinformatics, Vall d'Hebron Institute of Research, E-08035 Barcelona, Spain; and <sup>c</sup>Institut Català per la Recerca i Estudis Avançats, 08018 Barcelona, Spain

Edited by Robert E. Kingston, Massachusetts General Hospital/Harvard Medical School, Boston, MA, and approved August 6, 2019 (received for review February 25, 2019)

**Histone H3 lysine 9 methylation (H3K9me) is essential for cellular homeostasis; however, its contribution to development is not well established. Here, we demonstrate that the H3K9me2 demethylase PHF2 is essential for neural progenitor proliferation in vitro and for early neurogenesis in the chicken spinal cord. Using genome-wide analyses and biochemical assays we show that PHF2 controls the expression of critical cell cycle progression genes, particularly those related to DNA replication, by keeping low levels of H3K9me3 at promoters. Accordingly, PHF2 depletion induces R-loop accumulation that leads to extensive DNA damage and cell cycle arrest. These data reveal a role of PHF2 as a guarantor of genome stability that allows proper expansion of neural progenitors during development.**

histone demethylation | chromatin transcription | neuronal progenitor proliferation | PHF2

**D**uring neural development, multipotent progenitor cells self-renew and ultimately originate specialized neurons and glial cells (1, 2). The chromatin acting factors are essential players in both proliferation and cell differentiation events during embryo development. This epigenetic regulation is achieved by stabilizing chromatin structure that allows establishing heritable gene expression patterns. The epigenetic control is mainly mediated by covalent modifications of histones and DNA (3). Recently, histone methylation/demethylation has received special attention as an essential regulator of gene expression and genome stability during development (4). One critical histone modification during development is dimethylation of histone H3 at lysine 9 (H3K9me2), implicated in the silencing of genes playing important roles in chromatin homeostasis and development (5). In embryonic stem (ES) cells, it has been proposed that H3K9me2 increases across the genome as cells differentiate and acquire lineage specificity (6), although this is contentious (7). In addition, H3K9me2 together with H3K9me3 are essential components of the constitutive heterochromatin (8–10).

Despite the importance of this modification during development, little is known about the role of the enzymes responsible for this mark in neurogenesis. PHF2 is a member of the KDM7 histone demethylase (HDM) family (11). It contains a plant homeodomain (PHD) in the N-terminal and the Jumonji-C (JMJC) domain, which has demethylase enzymatic activity (12). Biochemical studies demonstrated that PHF2 demethylates H3K9me2 upon interaction with H3K4me2/3 through its PHD domain (13). PHF2 was first identified as a candidate gene for hereditary sensory neuropathy type I (14) because it is expressed at high levels in the neural tube and dorsal root ganglia (15). The physiological role of PHF2 in vivo is not yet clear, but it is a coactivator of multiple transcription factors (11). Working with them, PHF2 regulates various differentiation processes (16–18). Alterations in PHF2 have been identified in several cancer types (19–21). In-

terestingly, PHF2 mutations have been found in patients with autism spectrum disorder (ASD) (22–24). Despite PHF2 high expression in the neural tube and its implication on mental diseases, PHF2 involvement in neural development has not yet been explored.

In this study, we analyzed the role of PHF2 in neural stem cell (NSC) biology. We found that PHF2 is essential for progenitor proliferation in vitro and in vivo, in the chicken spinal cord. PHF2 binds and regulates cell cycle gene promoters, particularly those involved in DNA replication and cell cycle progression. Moreover, PHF2 depletion induces R-loop accumulation, DNA damage, and cell cycle arrest. These data reveal a role for PHF2 as a safeguard of genome stability during development.

## Results

**PHF2 Binds Promoters and Mediates H3K9me2 Demethylation.** To gain insight into the biological substrate of PHF2 in NSCs, we performed chromatin immunoprecipitation coupled with sequencing

## Significance

**During early neurogenesis, progenitors maintain the balance between self-renewal and differentiation. This equilibrium is preserved by the coordinated action of developmental cues and epigenetic mechanisms. We analyzed the role of the H3K9me2 histone demethylase PHF2 in neural stem cells biology in vitro and in early neurogenesis in the chicken neural tube. We found that PHF2 is essential for proliferation of progenitors. Additionally, PHF2 maintains the genome integrity, preventing R-loop accumulation and DNA damage. Our work contributes to the clarification of both the physiological and molecular roles of PHF2 during early neurogenesis. It opens avenues to understand the role of H3K9 methylation not only in development but also in diseases such as autism, cancer, and obesity where PHF2 has been involved.**

Author contributions: S.P., X.d.l.C., and M.A.M.-B. designed research; S.P., S.I., M.V., C.N., and E.M. performed research; X.d.l.C. and M.A.M.-B. contributed new reagents/analytic tools; S.P., N.P., S.I., M.V., E.Á.d.l.C., C.N., E.M., X.d.l.C., and M.A.M.-B. analyzed data; and M.A.M.-B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: RNA-seq and ChIP-seq data have been deposited in the GEO database under accessions [GSE122264](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122264) and [GSE122263](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122263), respectively.

<sup>1</sup>N.P. and S.I. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [xavhir@gmail.com](mailto:xavhir@gmail.com) or [mmbbmc@ibmb.csic.es](mailto:mmbbmc@ibmb.csic.es).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1903188116/-DCSupplemental](https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1903188116/-DCSupplemental).

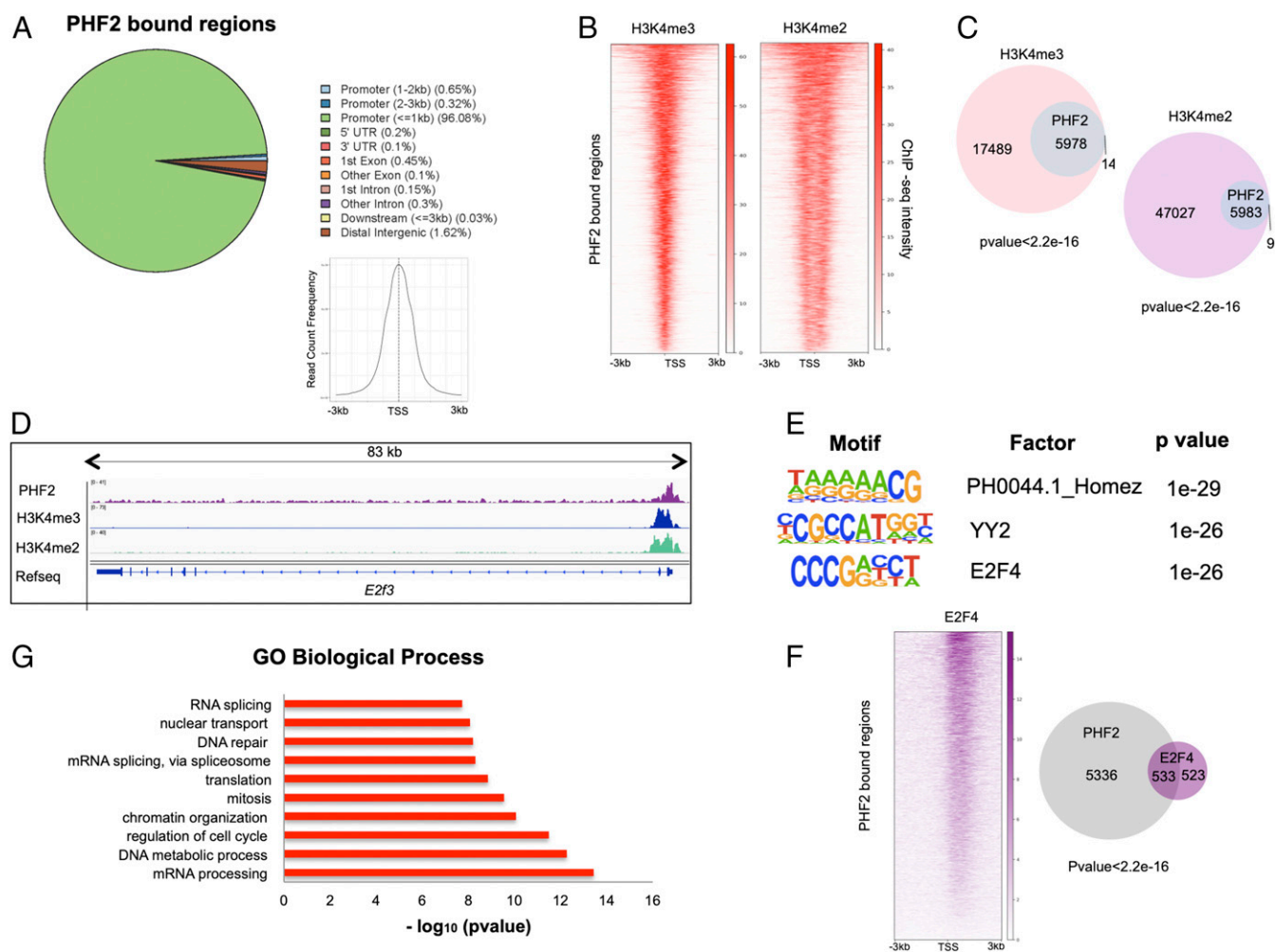
First published September 5, 2019.

(ChIP-seq) experiment. We detected 5,992 peaks normalized to the input ( $P$  value of  $1e-10$ ) in ChIP data for PHF2. *SI Appendix, Fig. S1A* shows validation of ChIP-seq results by qPCR for a random set of PHF2 targets. Then, we examined the genomic distribution of the PHF2 peaks. Our results showed that 97% of PHF2 peaks localized on gene promoters, around the transcription start site (TSS) (Fig. 1A), as it has been reported in a different cellular context (25).

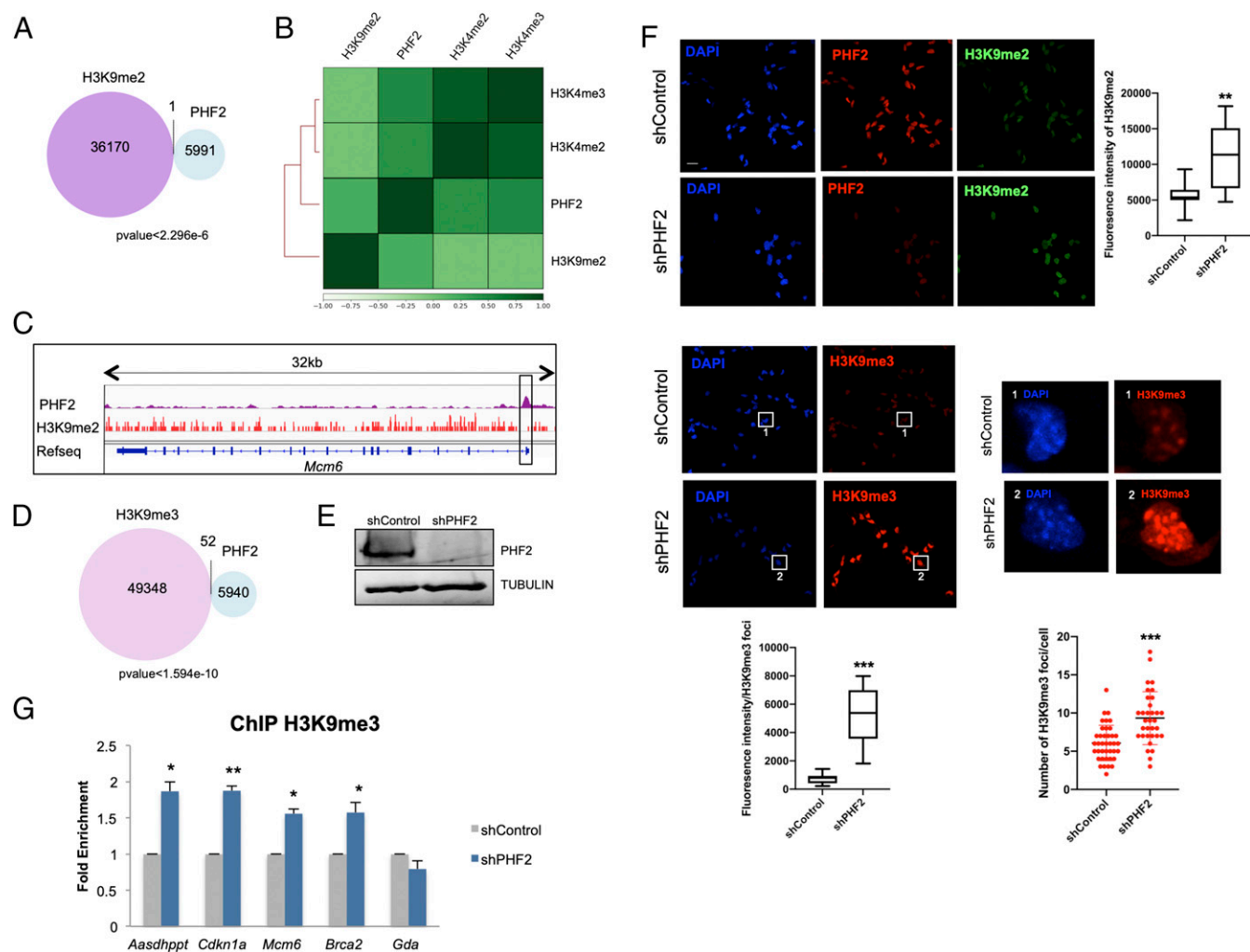
It has been proposed that PHF2 through its PHD domain interacts with H3K4me2/3 histone marks; thus, we analyzed the colocalization of PHF2 and H3K4me2/3 at the genome-wide level using previously published H3K4me2/3 ChIP-seq data from NSCs. Doing that, we identified 5,978 (99.7%) ( $P$  value  $<2.2e-16$ ) and 5,983 (99.8%) ( $P$  value  $<2.2e-16$ ) PHF2-bound regions that also contained H3K4me3 or H3K4me2, respectively (Figs. 1B–D and 2B and *SI Appendix, Fig. S1B*). The high overlapping suggests a required H3K4me2/3 mark for effective PHF2 binding at promoters as it is described (13, 25). To better understand how PHF2 is targeted to the chromatin, we performed bioinformatics analysis that revealed the most statistically significant predicted PHF2 binding sites using the Homer de novo motif

research tool. We identified as the top 3: Homez, YY2, and E2F4 DNA binding motifs (Fig. 1E). These transcription factors, particularly E2F4 and YY2, are essential to control progenitor self-renewal and cell cycle progression (26–28), suggesting a potential role of PHF2 in cell proliferation. To reinforce these data, we analyzed the colocalization of PHF2 and E2F4 using previously published ChIP-seq data in HeLa S3 cells. The results showed that the 50.3% ( $P$  value  $<2.2e-16$ ) of E2F4-bound regions, respectively, also contained PHF2 (Fig. 1F). These results reveal a potential role of PHF2 in regulating E2F-mediated transcription as it has been described for another KDM7 family member, PHF8 (29). Accordingly, gene ontology (GO) analysis of PHF2-bound regions indicated that PHF2 was associated with the promoter region of genes involved in cell cycle regulation, DNA repair, RNA processing, and chromatin organization, among others (Fig. 1G).

Previous studies demonstrated that PHF2 demethylates mainly H3K9me2 (13, 25). We thus analyzed the distribution of H3K9me2 in NSCs by performing ChIP-seq experiments and compared it with PHF2 genomic-associated regions. As expected for a H3K9me2 demethylase, the vast majority (99.99%) ( $P$  value  $<2.3e-6$ ) of



**Fig. 1.** PHF2 binds promoters in NSCs. (A) Genomic distribution of PHF2 ChIP-seq peaks in NSCs showing that PHF2 mainly binds promoter regions around the TSS. (B) Heatmaps depicting PHF2 binding to H3K4me3- and H3K4me2-marked promoters in NSCs 3 kb around the TSS. Scales indicate ChIP-seq intensities. (C) Venn diagrams showing overlap between PHF2-bound and H3K4me3 and H3K4me2-marked regions.  $P$  value is the result of an equal proportions test performed between H3K4me3 and H3K4me2 peaks and a random set. (D) IGV capture showing PHF2, H3K4me3, and H3K4me2 peaks in NSCs. (E) Motif enrichment analysis of PHF2 ChIP-seq peaks in NSCs using the Homer de novo motif research tool showing the 3 top enriched motifs. (F) Heatmap representation of PHF2 and E2F4 co-occupied regions (Left) and Venn diagrams (Right) showing peak overlapping between PHF2-bound and E2F4-bound regions in published ChIP-seq in HeLa S3 cells. (G) Gene ontology analysis showing the biological process of the PHF2-bound genes was performed using as a background the whole *Mus musculus* genome.



**Fig. 2.** PHF2 mediates H3K9me2 demethylation. (A) Venn diagram showing overlap between PHF2-bound and H3K9me2-marked regions in NSCs. (B) Clustered heatmap showing Pearson correlation of H3K9me2, H3K4me2, H3K4me3, and PHF2 ChIP-seq samples based on read coverage within genomic regions. (C) IGV capture showing PHF2 peaks and H3K9me2-marked regions in the *Mcm6* gene. The promoter region in the box shows no H3K9me2 enrichment where PHF2 binds. (D) Venn diagram depicting overlap between PHF2-bound and H3K9me3-marked regions in NSCs. (E) NSCs were infected with lentivirus expressing shRNA control (shControl) or shRNA specific for PHF2 (shPHF2). After 48 h, total protein extracts were prepared and the PHF2 and TUBULIN levels were determined by immunoblot. (F) Immunostaining of shControl and shPHF2 cell lines. Cells were fixed and stained with PHF2, H3K9me2, and H3K9me3 antibodies and DAPI (Scale bar: 20  $\mu$ m). Zoom in showing H3K9me3 foci in shControl and shPHF2. Cell fluorescence for H3K9me2 staining and H3K9me3 foci was measured using ImageJ. Boxplots represent the quantification of the fluorescence intensities as well as the number of H3K9me3 foci/cell in shControl and shPHF2 cells.  $^{**}P < 0.01$ ;  $^{***}P < 0.001$  (Student's *t* test). (G) ChIP of H3K9me3 in shControl and shPHF2 cells was analyzed by qPCR at the indicated gene promoters that were identified by PHF2 ChIP-seq as PHF2 targets. An intragenic region of the *Gda* gene marked by H3K9me2/3 but devoid of PHF2 was used as negative control. Data from qPCR were normalized to the input and expressed as fold enrichment over the data obtained in shControl. Results are the mean of 2 to 3 biological independent experiments. Error bars represent SEM.  $^{*}P < 0.05$ ;  $^{**}P < 0.01$  (Student's *t* test).

PHF2-bound regions were totally excluded from the H3K9me2 positive regions in NSCs (Fig. 2A–C and *SI Appendix*, Fig. S2A). Accordingly, clustered heatmaps showed a clear exclusion of the H3K9me2 mark and H3K4me2/3-associated regions (Fig. 2B). Moreover, a strong overlapping of PHF2 peaks with H3K4me2/3-associated regions was observed, in agreement with the data presented in Fig. 1B and C. Altogether these results suggest that PHF2 binds to genomic regions marked with histone modifications related to transcriptional activation. Similarly to H3K9me2, H3K9me3-enriched regions identified by previously published ChIP-seq were excluded from the PHF2 positive regions in NSCs (Fig. 2D). To evaluate the impact of PHF2 depletion on histone modifications in vivo, NSCs were transduced with lentivirus containing specific PHF2 shRNA that efficiently decreased the PHF2 levels (Fig. 2E and *SI Appendix*, Fig. S2B) or with a control shRNA (shControl) (see *Materials and Methods*). The reduction of PHF2 did not affect

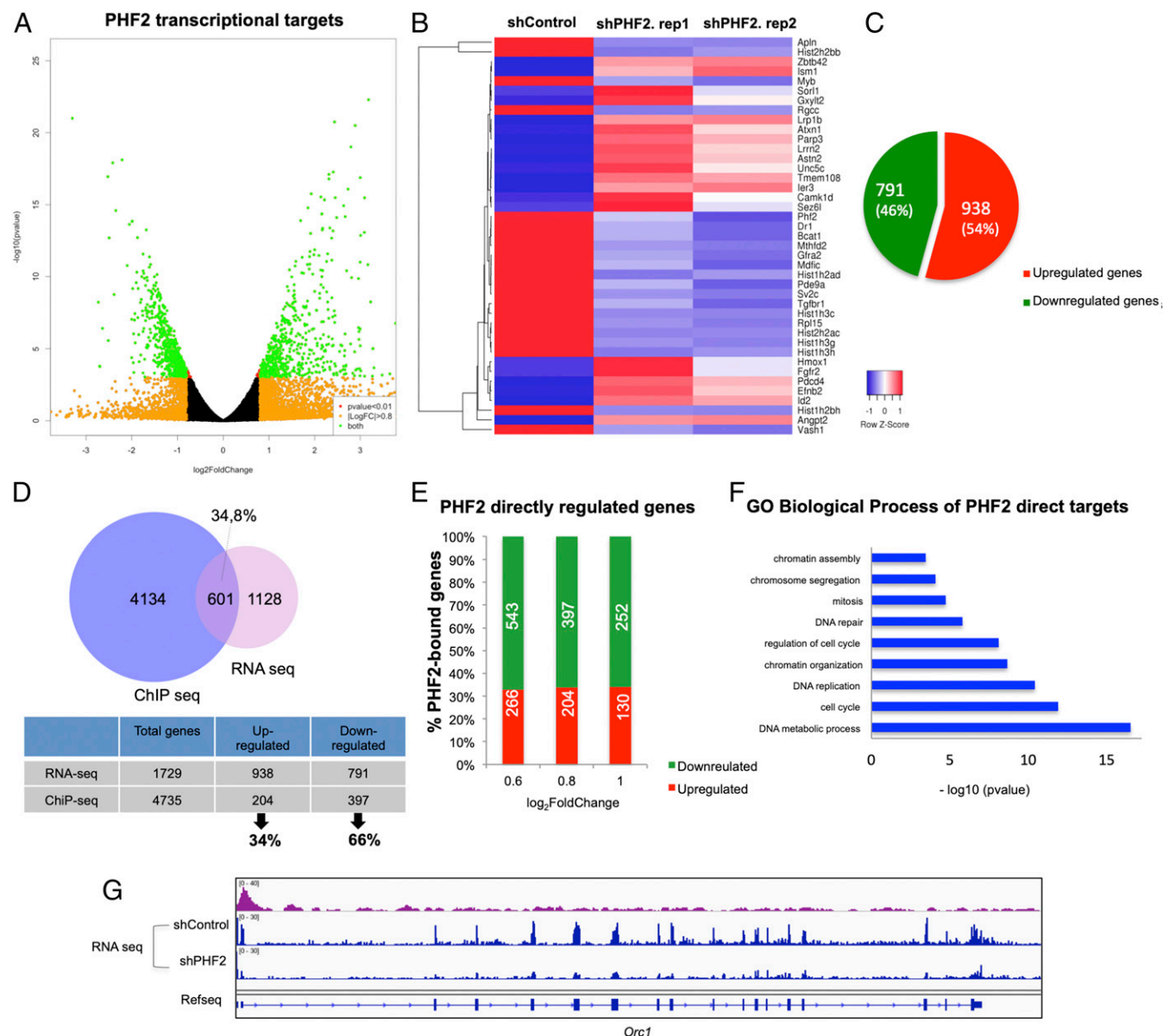
the expression of its homologous PHF8 (*SI Appendix*, Fig. S2B). After viral transduction, the levels of H3K9me2 were analyzed by immunofluorescence assays. In Fig. 2F we observed a slight increase in H3K9me2 levels as it has been demonstrated in other cellular contexts (13). As the H3K9me2 mark serves as a substrate for the histone methyltransferase SUV39H, we tested the levels of H3K9me3 in PHF2-depleted NSCs. Results in Fig. 2F clearly showed an increase in H3K9me3. Intriguingly, both the intensity and the number of H3K9me3 foci increased upon PHF2 depletion. Accordingly, a clear accumulation of HP1 $\alpha$  was also detected in PHF2-depleted cells (*SI Appendix*, Fig. S2C) without increase in other histone marks such as H3K4me3 (*SI Appendix*, Fig. S2D).

These data strongly suggest that depletion of PHF2 led to an increase in heterochromatin-related marks at global levels. Thus, we sought to test whether PHF2 is also important to prevent H3K9me3 increase at a local level. To do so, we chose some PHF2-target regions

identified in the ChIP-seq experiment and tested the effect of PHF2 depletion on H3K9me3 levels by ChIP-qPCR assays. A clear increase in the H3K9me3 mark was observed (Fig. 2G), at the analyzed promoters, without affecting the intragenic region of the *Gda* gene (a non-PHF2 target used as a negative control), in accordance with the global increase noticed in Fig. 2F. These data strongly suggest that PHF2 prevents H3K9me3 accumulation, limiting ectopic heterochromatin formation, particularly at promoter regions.

**PHF2 Regulates Cell Cycle Gene Transcription.** To gain further knowledge into PHF2's function in NSCs, we identified the

PHF2-associated transcriptional profile by RNA-seq. We found 1,729 transcripts that significantly changed their expression ( $\log_2$  fold change [FC]  $> 0.8$  and FC  $< -0.8$  and  $P$  value  $< 0.01$ ), in the 2 biological independent experiments in PHF2-depleted NSCs (shPHF2) compared with control (shControl) cells (Fig. 3A and B). Among these, 791 (45.8%) were down-regulated and 938 (54.2%) up-regulated upon PHF2 depletion (Fig. 3C). These results were confirmed by depletion of PHF2 by using another shRNA against a distinct region of PHF2 (see *Materials and Methods*) as we tested by qPCR analysis of some randomly chosen genes (*SI Appendix, Fig. S3A*). Next, we identified the



**Fig. 3.** PHF2 regulates cell cycle gene transcription. (A) Volcano plot represents PHF2 transcriptional targets identified by RNA-seq in shControl and shPHF2 NSCs. The green dots represent all of the genes with false discovery rate (FDR)  $< 0.01$  and  $\log_2$  FC  $> 0.8$  and  $\log_2$  FC  $< -0.8$ . (B) Heatmap showing the top 20 regulated genes identified by RNA-seq in shControl and shPHF2. Two biological replicates of shPHF2 cells were used for RNA-seq. All of the genes showed  $P$  value  $< 0.01$  and  $\log_2$  FC  $> 0.8$  and  $\log_2$  FC  $< -0.8$ . (C) Diagram showing the number and percentage of up- and down-regulated genes in the RNA-seq experiment comparing shControl and shPHF2 NSCs. (D) Venn diagram showing overlapping between PHF2-bound genes and PHF2 transcriptional targets. A total of 34.8% of differentially expressed genes identified by RNA-seq with  $\log_2$  FC  $> 0.8$  and  $\log_2$  FC  $< -0.8$  were also PHF2 direct targets identified by ChIP-seq. From these, 34% were up-regulated and 66% down-regulated. (E) Graph depicting the percentage and number of up-regulated and down-regulated genes in the shPHF2 according to the RNA-seq that contain PHF2 in their promoter classified by  $\log_2$  fold change. (F) Gene ontology analysis showing the "biological process" of the PHF2 direct targets was performed using as a background the whole *M. musculus* genome. (G) IGV capture showing PHF2 peaks and RNA levels in shControl and shPHF2 in the *Orc1* gene.

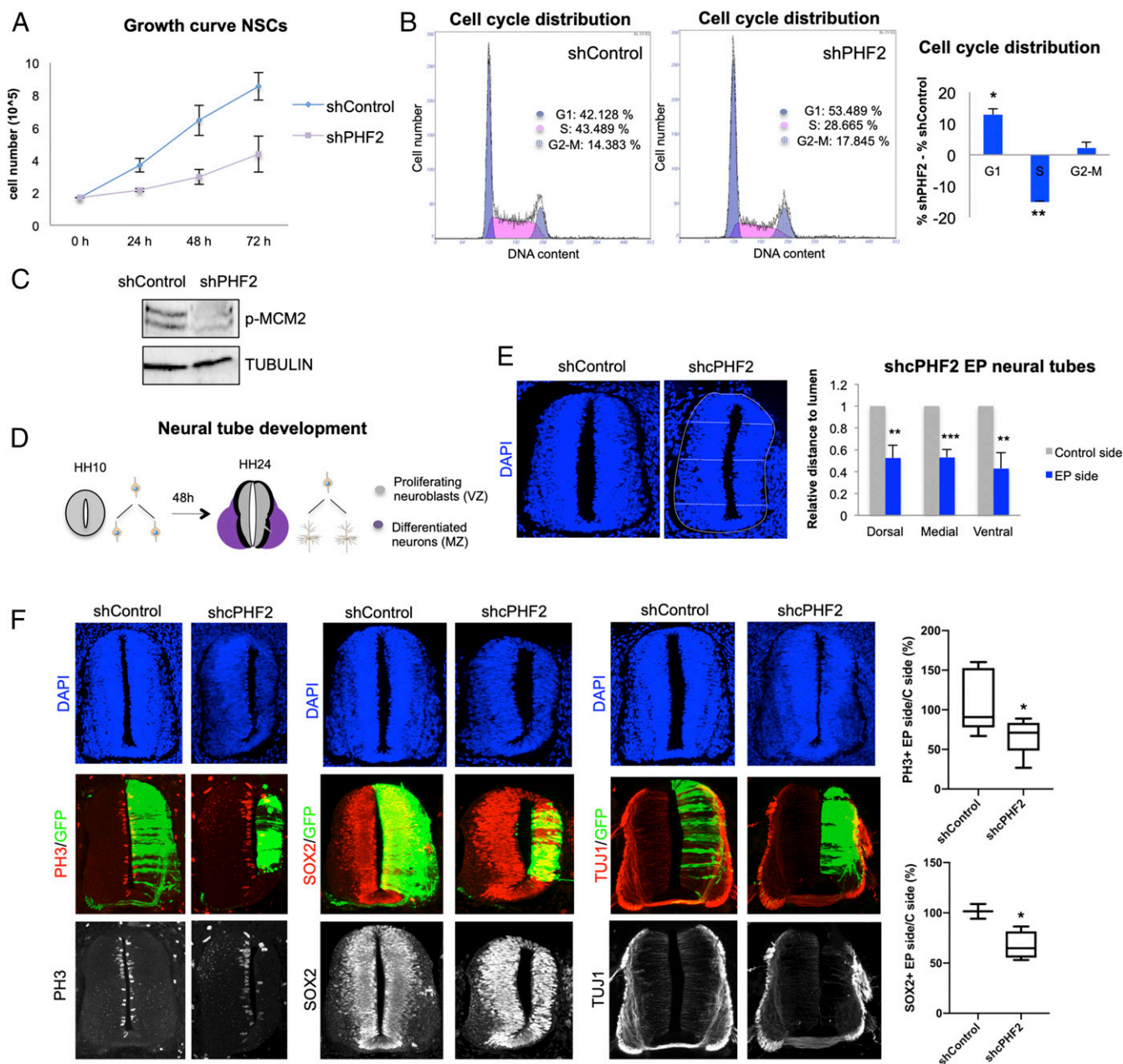


PHF2 direct transcriptional targets by comparing the PHF2-associated transcriptional profile with the ChIP-seq data. Among the genes that showed a PHF2 dependency for transcriptional regulation ( $\log_2$  FC > 0.8 and  $\log_2$  FC < -0.8) in the RNA-seq experiment, 601 (34.8%) were bound by PHF2 (Fig. 3D). Interestingly, the proportion of direct down-regulated transcripts (397, 66%) was higher than the direct up-regulated ones (204, 34%) (Fig. 3D). This percentage was almost maintained at different FCs (Fig. 3E). Moreover, half of the down-regulated genes in the RNA-seq experiment (50.2%) contain PHF2 bound at the promoter region as we determined in the ChIP-seq experiment (*SI Appendix*, Fig. S3B), while only 21.7% of the up-regulated genes were direct targets of PHF2 (*SI Appendix*, Fig. S3C). These data are in accordance with the activator role proposed for PHF2 (17). Nevertheless, the fact that 34% of the direct targets became up-regulated upon PHF2 depletion, suggest a potential role of PHF2 contributing to transcriptional repression or preventing activation as it has been demonstrated for another member of the KDM7 family, PHF8, in a different cellular context (30). Enrichment analysis of GO terms over the 601 PHF2 direct target genes showed that the most enriched were associated with cell cycle categories, particularly G1/S transition (*E2f2/3/7/8*, *Cdc7*, *Cdc25a*, *Cdk4*, and *Mcm3/4/8*), DNA replication (*Orc1/2/6* and *Pcna*), mitosis (*Cdk1*, *Smc2/3/4*, *Aurkb*, and *Topo2a*), as well as chromatin activity (*Cenpa*, *Kdm1b*, *Hat1*, *Parp1*, and *Prmt5*) (Fig. 3 F and G and *SI Appendix*, Fig. S4A). Intriguingly, some of them were E2F target genes (*Ccnd1*, *Cdc25a*, *Pcna*, *Mcm3/4/6/8*, and *Smc4*, including *E2f* family genes) (Fig. 3 F and G and *SI Appendix*, Fig. S4B). To further characterize the differences between control and PHF2 knockdown (KD) NSCs, we performed a GO enrichment analysis of down-regulated, up-regulated, and unregulated direct target genes to identify those biological processes most sensitive to PHF2 depletion. The results revealed that PHF2 down-regulated genes were strongly associated with cell cycle progression, chromatin activity, and DNA repair among others (*SI Appendix*, Figs. S3D and S4B). The same analysis of the up-regulated genes did show functional categories related to morphogenesis, signal transduction, and developmental process (*SI Appendix*, Fig. S3E) or to RNA processing and chromatin organization among others in the case of unregulated ones (*SI Appendix*, Fig. S3F). It is important to notice that the number of genes and the statistically significant values of up-regulated and unregulated gene-associated categories were much lower ( $P$  values ranging from  $-\log_{10}$  2–6) compared with the down-regulated genes ( $P$  values  $-\log_{10}$  4–23). This fact, led us to focus on the potential role of PHF2 as a transcriptional activator of cell cycle and DNA repair genes in this particular neural context. Finally, we confirmed by immunoblot the decrease of some proteins whose genes were down-regulated in PHF2-depleted cells (*SI Appendix*, Fig. S4C). Interestingly, we did not detect any reduction in the histone protein levels, although their mRNAs were down-regulated upon PHF2 depletion (*SI Appendix*, Fig. S4D), probably due to the tight control of the histone levels inside the cells. Altogether, these data strongly suggest that PHF2 binds to the cell cycle gene promoters to fine tune their chromatin activity and facilitate their transcription.

We next sought to analyze the transcriptional consequences of PHF2 overexpression in NSCs. To do that, we established a NSC line that expressed PHF2 wild type (WT) upon addition of doxycycline (see *Materials and Methods*). We induced the PHF2 expression and the levels of some direct PHF2 targets, identified in the RNA-seq and ChIP-seq experiments, were analyzed by qPCR. No changes or slight alterations of the expression (up and down) were observed upon PHF2 overexpression (*SI Appendix*, Fig. S4E).

**PHF2 Regulates Cell Proliferation.** To further analyze the potential role of PHF2 in cell proliferation, we examined the consequences of its reduction. PHF2-depleted NSCs exhibited a striking decrease in cell growth (Fig. 4A). Moreover, flow-cytometry analysis demonstrated a delay in G1/S transition (%G1 shControl 42.2%, shPHF2 53.5%) upon PHF2 depletion (Fig. 4B). Moreover, the levels of the Mcm2 factor (the putative helicase essential for DNA replication initiation and elongation in eukaryotic cells) phosphorylated at S40 required for the initiation of DNA replication were lower in PHF2-depleted cells compared with control progenitors (Fig. 4C). Interestingly, the expression of PHF2 was regulated throughout the cell cycle with higher levels at G1/S transition (*SI Appendix*, Fig. S5A) and its recruitment to the S-phase-activated gene promoter ORC1 also increased at G1/S transition (*SI Appendix*, Fig. S5B). These data reinforce the idea that PHF2 contributes to cell cycle progression by facilitating the expression of cell cycle progression genes, particularly those involved in DNA replication.

**PHF2 Depletion Blocks Neurogenesis in the Spinal Cord.** The findings described above support the idea that PHF2 activates genes essential for neural progenitor proliferation (Fig. 3F and *SI Appendix*, Fig. S4 A and B). Thus, we sought to test this notion in vivo using the chicken embryo neural tube as a model. Structurally, 2 main zones can be distinguished in a transversal section of the neural tube: the ventricular zone (VZ), where proliferating progenitors reside, and the mantle zone (MZ), where final differentiated neurons accumulate (Fig. 4D). To analyze the function of PHF2 in early neurogenesis, we first cloned 2 shRNAs for chick PHF2 (cPHF2) in a bicistronic vector containing GFP sequence, which reduced PHF2 levels (*SI Appendix*, Fig. S5C). Then, we knocked down PHF2's expression in Hamburger and Hamilton stage 10–12 (HH10–12) chicken embryo spinal cords by in ovo electroporation of the shRNAs for cPHF2 or a control shRNA (shControl). Remarkably, cPHF2 KD resulted in a neural tube reduced in size (Fig. 4E and *SI Appendix*, Fig. S5D). As the 2 cPHF2 shRNAs gave the same phenotype, we chose the shcPHF2, that provided better cPHF2 depletion (*SI Appendix*, Fig. S5C), to perform the rest of the experiments. The observed reduced size of the neural tube could mainly be due to cell death or to proliferation defects. As we did not observe apoptotic cells in the electroporated neural tubes and our previous results indicated that PHF2 is essential for neural progenitor proliferation in vitro, we hypothesized that cPHF2 plays an active role in maintaining neuroblast proliferating. To explore this idea, HH10–12 neural tubes were electroporated with shcPHF2 or shControl and the effect on neural progenitor proliferation was analyzed after 48 h. We evaluated neural progenitor entry into mitosis by analyzing the presence of H3S10p. Neural tubes electroporated with shcPHF2 showed a reduction in H3S10p (PH3)-positive mitotic cells (ratio of PH3+ cells electroporated [EP] side/control side: shControl  $106.5 \pm 36.7$ , shcPHF2  $65.3 \pm 21$ ;  $P < 0.05$ ) (Fig. 4F). Accordingly, reduction of the neural progenitor marker SOX2 (ratio of SOX2+ cells EP site/control site: shControl  $101.6 \pm 10.3$ , shcPHF2  $67.3 \pm 14$ ;  $P < 0.05$ ) (Fig. 4F) indicates that cPHF2 is required for neural progenitor self-renewal in the neural tube. We next explored the possibility that the inhibition of proliferation observed upon cPHF2 reduction would also correspond with a premature differentiation of neuroblasts. Neural tubes electroporated in ovo with shcPHF2 and stained for TUJ1, a neuronal differentiation marker, showed neither premature differentiation nor ectopic localization (Fig. 4F). Only a clear reduction of TUJ1-positive cells was observed, probably reflecting the decrease on the progenitor population. Finally, we investigated whether the neuroblast proliferation impairment affected similarly to all neural subpopulations along the dorsal–ventral axis. To do that, we depleted cPHF2 and analyzed different populations by using different neural markers: MNR2 for ventral neurons (motoneurons)



**Fig. 4.** PHF2 depletion blocks neurogenesis in the spinal cord. (A) Growth curve showing the proliferation rate of NSCs infected with lentivirus expressing shRNA control (shControl) or shRNA for PHF2 (shPHF2) from 0 to 72 h. (B) Flow cytometry analysis of NSCs control and PHF2-depleted cells previously stained with propidium iodide. (C) Total protein extracts were prepared from NSCs infected with lentivirus expressing shRNA control (shControl) or shRNA specific for PHF2 (shPHF2) and the phospho-MCM2 (p-MCM2) and TUBULIN levels were determined by immunoblot. (D) Schematic representation indicating the regions occupied by proliferating neural progenitors (VZ) and postmitotic neurons (MZ) in HH10 and HH24 chicken embryo spinal cord. (E) HH10–12 embryos were electroporated with shControl or shRNA for cPHF2 (shcPHF2) cloned into pSUPER vector and GFP-expressing vector. Transverse sections of electroporated neural tubes are indicated above stained with DAPI 48 h postelectroporation (PE). Graphs show the quantification of the size of the control side and shcPHF2-electroporated side. To do so, we measured the dorsal, medial, and ventral distances to the lumen on each side, relative to the length of the central line of the lumen. Data represent the mean of 4 to 5 embryos (from 2 to 4 biological independent experiments). Error bars indicate SD.  $^{**}P < 0.01$ ;  $^{***}P < 0.001$  (Student's *t* test). (F) Transverse sections of neural tubes from HH10–12 embryos electroporated in ovo with shControl or shcPHF2 and stained for H3S10P (PH3), Sox2, or TUJ1 48 PE. Boxplots are showing the quantification of the corresponding immunostaining. Data represent the mean of 4 to 12 embryos (from 3 to 4 biological independent experiments).  $^{*}P < 0.05$  (Student's *t* test).

and Pax6 for dorsal progenitors. The results in *SI Appendix, Fig. S5E* show that PHF2 depletion impairs the generation of all analyzed cell types. Overall, these results point to an essential role of cPHF2 promoting early neurogenesis by controlling progenitor proliferation.

Although we did not observe any effect on neuroblast differentiation upon PHF2 depletion, it has been demonstrated that PHF2 plays an important role in differentiation in other models

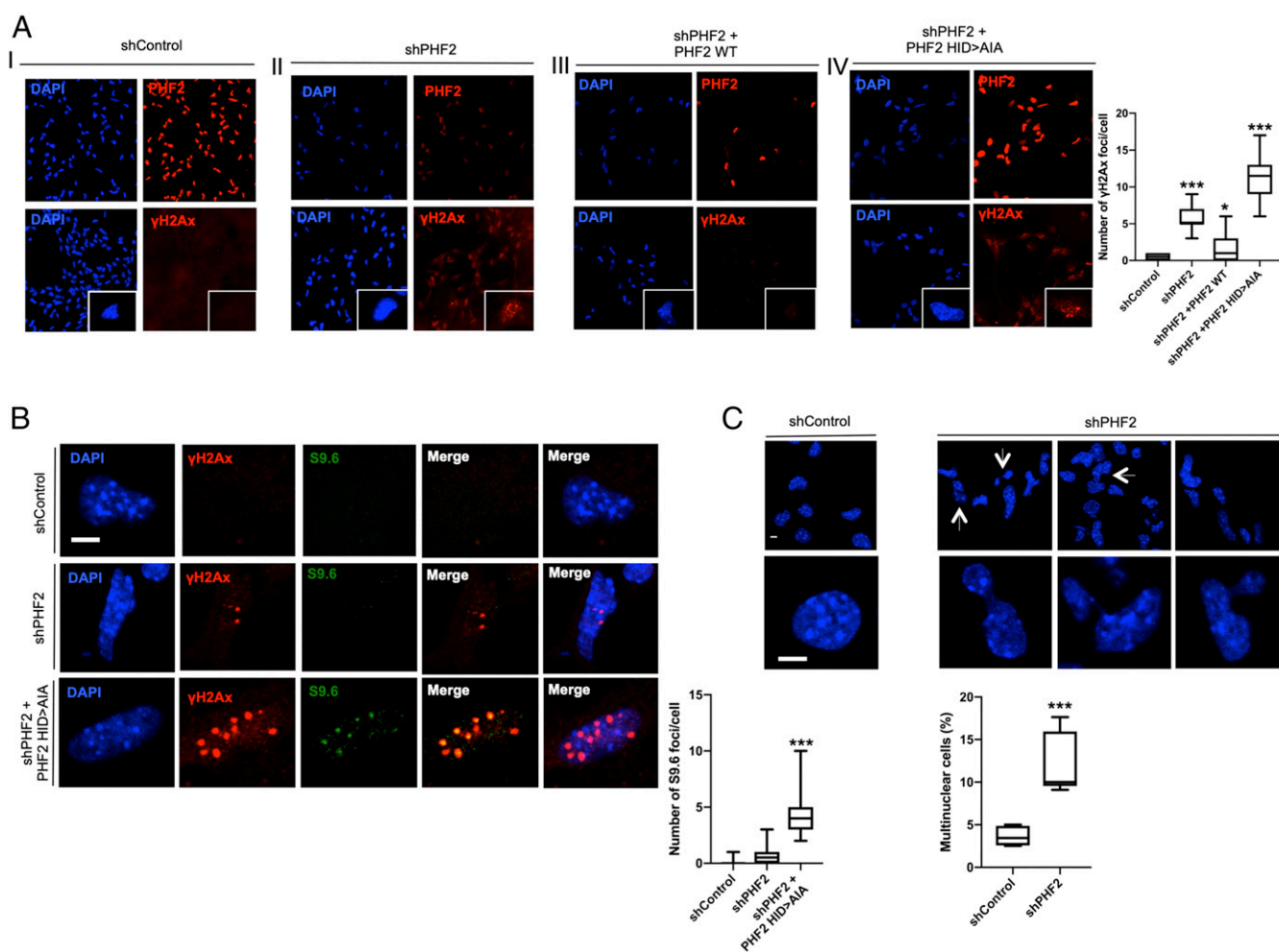
(16–18). To deeply understand the potential contribution of PHF2 to neuronal commitment, we first analyzed the PHF2 expression during neuronal differentiation. Our results in *SI Appendix, Fig. S6A* indicated that PHF2 levels slightly increased during differentiation (*SI Appendix, Fig. S6A*). Next, we analyzed the consequences of overexpressing PHF2 during neuronal differentiation that we measured by the induction of the neuronal marker TUJ1 and the repression of the pluripotency gene NESTIN. Results

in *SI Appendix, Fig. S6 B and C* indicate that PHF2 overexpression did not affect neuronal differentiation in vitro.

As PHF2 depletion led to cell cycle arrest, we finally tested whether PHF2 overexpression overcame G1/S cellular checkpoint imposed by growth factors removal or neuronal differentiation induction. Cell cycle reentry was analyzed by measuring the levels of some cell cycle-related genes by qPCR and cell cycle analysis by flow cytometry. Results in *SI Appendix, Fig. S7 A–C* suggest that, although we observed the induction of some genes, PHF2 overexpression did not overcome the cell cycle arrest imposed by either growth factor removal or neuronal differentiation induction. Altogether these data strongly indicate that the major role of PHF2 at early neurogenesis is to facilitate neural progenitor proliferation.

**PHF2 Depletion Leads to DNA Damage and Genome Instability.** As PHF2 depletion caused profound defects on replication machinery expression (*E2f*, *Cyclin E*, *Orc1/2*, and *Mcm2/7*) as well as repair components (*Brac1/2*, *Rad51b*, and *ATR/M*) that finally led to cell cycle arrest, we tested whether this fact induced DNA damage that we measured by the  $\gamma$ H2Ax content. The histone

variant H2AX is phosphorylated at the Ser-139, forming  $\gamma$ H2Ax as an early cellular response to the DNA double-strand breaks. Then, we quantified the  $\gamma$ H2Ax content as a measure of DNA damage in control and PHF2-depleted cells. Upon PHF2 KD, a clear accumulation of  $\gamma$ H2Ax was detected (Fig. 5A, panel II). This increase was not observed in cells transduced with the shControl RNA (Fig. 5A, panel I). Interestingly, the observed DNA damage was rescued by overexpression of the PHF2 WT (Fig. 5A, panel III) but not by overexpression of the catalytic dead mutant (249H > A) (Fig. 5A, panel IV, and Fig. 5B); indeed, in the latter case, an apparent increase in  $\gamma$ H2Ax reactivity, accompanied by a higher decrease of replication-related gene expression (*SI Appendix, Fig. S8*), was detected, suggesting an important role of PHF2's enzymatic activity in preventing damage. R-loop (RNA–DNA hybrids formed by reannealing of nascent transcripts to their DNA template, leaving the nontemplated strand as single-stranded DNA) accumulation is associated with DNA replication mistiming that increases collisions with the transcriptional machinery (31), leading to DNA damage (31, 32). Therefore, we investigated whether  $\gamma$ H2Ax reactivity observed in PHF2-depleted



**Fig. 5.** PHF2 depletion leads to DNA damage and genome instability. (A) Immunostaining of NSCs expressing shControl (I), shPHF2 (II), shPHF2 together with PHF2 WT (III), and shPHF2 together with PHF2 (249H > A) (IV). Cells were fixed and stained with PHF2 and  $\gamma$ H2Ax antibodies and DAPI. Enlarged images are showing individual cells. More than 30 cells were quantified. Data shown are representative of 2 to 3 biological independent experiments (Scale bar: 20  $\mu$ m). Boxplots represent the number of  $\gamma$ H2Ax foci/cell. \* $P$  < 0.05; \*\*\* $P$  < 0.001 (Student's  $t$  test). (B) Immunostaining of a representative enlarged cell expressing shControl, shPHF2, and shPHF2 together with PHF2 (249H > A). Cells were fixed and stained with  $\gamma$ H2Ax and S9.6 antibodies and DAPI. More than 30 cells were quantified. Data shown are representative of 2 to 3 biological independent experiments (Scale bar: 5  $\mu$ m). Boxplots represent the number of S9.6 foci/cell. \*\*\* $P$  < 0.001 (Student's  $t$  test). (C) Formation of multinucleated cells and segregation defects are shown for shControl and shPHF2 cells. A total of 50 to 100 cells were quantified. Data shown are representative of 4 independent experiments (Scale bar: 5  $\mu$ m). Boxplots represent the percentage of multinuclear cells. \*\*\* $P$  < 0.001 (Student's  $t$  test).

cells was related to R-loop accumulation. For this purpose, we performed immunofluorescence experiments in control, PHF2 depleted, and PHF2-depleted that overexpressed the PHF2 (249H > A) mutant NSCs (where the  $\gamma$ H2Ax signal was higher, see Fig. 5A, panel IV) using the S9.6 antibody that specifically recognizes RNA–DNA hybrids without cross-reacting with single- or double-stranded DNA. The data in Fig. 5B show a clear increase in S9.6 reactivity that colocalized with the  $\gamma$ H2Ax-positive regions.

DNA damage is often linked with genome instability; accordingly, we noticed an increased frequency of chromosome segregation defects, in particular, anaphase chromatin bridges and multinuclear cells in PHF2-depleted compared with control NSCs (Fig. 5C). Altogether these data demonstrate that PHF2's deficiency induced DNA damage and R-loop accumulation that led to chromosome segregation defects and ultimately genome instability.

## Discussion

Our data reveal an unforeseen role of PHF2 during development. We demonstrate that this HDM binds to the cell cycle gene promoters facilitating their transcription and preventing genome instability. In that way, PHF2 allows neural stem cell proliferation during progenitor expansion. To do that, PHF2 required its catalytic activity. PHF2 prevents H3K9me3 accumulation, limiting ectopic heterochromatin formation particularly at promoter regions.

Our data indicate that the histone demethylase PHF2 plays a pivotal role in the neural development by promoting the proliferation of neural progenitors. Interestingly, a synergy for regulation of the key proliferation factor E2F was observed. PHF2 directly promotes the expression of several members of the E2F family and at the same time might cooperate with them to facilitate their transcriptional activity. These data are in agreement with previous studies reporting that E2F target genes were repressed by H3K9me2/3 marks and HP1 factor (33, 34). Thus, PHF2 might prevent H3K9me2/3 accumulation at the cell cycle gene promoters before S entrance, and at the same time it facilitates their demethylation to promote cell cycle progression. Interestingly, previous studies have shown that the absence of other member of the KDM7 family, PHF8, also impaired G1/S transition in conjunction with E2F family factors (29).

Our results indicate that PHF2 depletion was associated with the accumulation of RNA:DNA hybrids, DNA damage, and genome instability. The link between R-loops and DNA damage is well established (32). One of the major causes responsible for R-loop accumulation is the replication fork collapse or stall (31, 35–37). Thus, the replication temporal changes due to replication machinery alteration generated by PHF2 depletion might lead to unscheduled collisions of the replication and transcription machineries, inducing DNA breaks (37–40). On the other hand, the observed decrease of expression of factors involved in R-loop resolution and DNA damage repair (ATM/ATR, Rad51, and BRCA1) in PHF2-depleted cells that could contribute to accumulation of double-strand breaks and genome instability. Finally, it has been proposed the demethylation of H3K9 may be an important step in the repair of double-strand breaks. Thus, the global increase in H3K9me3 observed in PHF2-depleted progenitors (Fig. 2F) could delay or impair DNA repair, as it has been described for KDM4B demethylase (41–43). Interestingly, PHF2 associated with p53 and facilitates its activity (44). p53 is an essential tumor suppressor that maintains genomic stability. p53, in addition to the cell cycle checkpoint control, maintains genomic integrity and replication fidelity by preventing DNA topological conflicts between transcription and replication (45). Thus, PHF2 depletion might decrease p53 activity, leading to accumulation of transcription–replication conflicts and genomic instability.

PHF2 has been related to ASD (22–24), but the function of PHF2 in neural development is still poorly understood. *Phf2* KD

mice showed partial neonatal death, growth retardation, and reduced body weight. Interestingly, the brain weights of *Phf2* KD mice were larger than wild-type littermates (46). Macrocephaly is a common phenotype associated with ASD. Intriguingly, we observed that *Phf2* KD in cortical progenitors and in the neural tube results in defective neural progenitor proliferation. A similar paradox has been described for another chromatin factor strongly associated with ASD, the remodeling factor CHD8 (47–49). Although at this moment we do not have an explanation for this phenomenon, it would be interesting to elucidate the role of PHF2 in the development of other cell types such as astrocytes. Furthermore, our results raise the possibility that some of the ASD-associated phenotypes in patients carrying PHF2 mutations may be caused by defects during early neural development.

Our work paves the way for investigating the contribution of PHF2 to genomic stability and transcriptional regulation in other cellular contexts. In particular, PHF2 has been widely involved in cancer. Several studies use H3K9me modifier enzymes as targets in cancer treatment (50, 51). Moreover, alteration of the demethylase activity of PHF2 has been suggested as a new target to treat disorders linked to diet-induced obesity, due to its essential role in regulating adipogenesis (46). Our results from PHF2-depleted cells, as well as published data, removing H3K9me2/3 histone methyltransferases (HMTs) (52), indicate that alteration or removal of H3K9 methylation might not be a suitable therapeutic strategy. In both cases, genomic instability might be a drawback in these treatments. Therefore, our work helps to improve our understanding of the multiple cross-talks between epigenetics, development, and diseases.

## Materials and Methods

**Cell Culture and Differentiation.** Mouse NSCs were prepared from cerebral cortices of C57BL/6J mouse fetal brains (embryonic day [E]12.5). They were cultured in poly-D-lysine (5  $\mu$ g/mL, 2 h, 37 °C) and laminin (5  $\mu$ g/mL, 4 h, 37 °C) precoated dishes (53, 54). Cells were maintained in culture as previously described (55) with fibroblast growth factor (FGF) and epidermal growth factor (EGF) to 10 and 20 ng/mL, respectively. Human 293T cells were maintained in culture under standard conditions (56). Chicken UMNSAH/DF1 and HeLa cells were cultured in DMEM supplemented with 10% FBS. NSC differentiation protocol is described in *SI Appendix, Materials and Methods*.

**Antibodies and Reagents.** Antibodies used were anti: PHF2 (Cell Signaling, D45A2), H3K9me3 (Abcam, ab8898), H3K9me2 (Abcam, ab1220), H3K4me3 (Abcam, ab8580), H3Sp10 (Millipore, 06-570), anti-phospho-H2A.X (Millipore, 07-164),  $\beta$ -tubulin III (TUJ1, Covance, MMS-435P), SOX2 (Millipore, AB5603),  $\beta$ -tubulin (Millipore, MAB3408), MNR2 (DSHB, 81.5C10), HP1 $\alpha$  (Euromedex), NESTIN (Abcam, ab5968), p-MCM2 (Abcam, ab133243), ATR (Santa Cruz, sc-21848), Rad51 (Santa Cruz, sc8349), histones H4 (Abcam, ab10158), H3 (Abcam, ab1791), H2B (Abcam, ab1790) and PAX6 (DSHB). Mouse monoclonal S9.6 is described in ref. 57. DAPI was obtained from Thermo Fisher (1306). Doxycycline (Millipore, 324385) was used at 1  $\mu$ g/mL.

**Plasmids.** PHF2 cDNA from p3xFLAG-PHF2 (kindly provided by Jiemin Wong, East China Normal University, Shanghai, China) was cloned into pInducer between attB1 and attB2 sites. Both, PHF2 WT and mutant were induced upon doxycycline addition (1  $\mu$ g/mL). pLKO.1 lentiviral vectors expressing short hairpin RNA against mPHF2 (CGTGGCTATTAAGTGTCTA), shPHF2 or control (CAACAAGATGAAGAGCACC) were purchased from Sigma and (CCTTATCCACTCCACTTGACC) shPHF2.2 was cloned in pLKO.1 lentiviral vector. DNA sequences coding cPHF2 short hairpin RNAs (GGAGCTTCGAAGTCGCACT) shcPHF2 and (CTATGTCGGACCAGAGAGA) shcPHF2.2 were cloned into pSUPER or pSHIN vectors (58), as indicated. pSHIN vector contains the pSUPER and the EGFP expression cassette.

**RNA Extraction and qPCR Assays.** RNA was extracted using TRIZOL reagent (Invitrogen). High-capacity cDNA reverse transcription kit (Invitrogen) and 50 ng of RNA were utilized for reverse transcription. qPCR assays were performed with SYBR Green (Roche) in a LightCycler 480 (Roche) machine using specific primer pairs (*SI Appendix, Table S1*).

**Chick in Ovo Electroporation and Immunostaining.** Eggs from White-Leghorn chickens were used in the in ovo electroporation experiments. They were incubated at 38.5 °C and 70% humidity. Embryo developmental stage was determined following HH (59). Embryos were electroporated with the indicated DNAs at 3 μg/μL with 50 ng/mL of Fast Green as previously described (54, 60). Expanded protocol is described in *SI Appendix, Materials and Methods*.

**Lentiviral Transduction.** Lentiviral transduction was carried out as previously described (30, 61). Extended protocol is provided in *SI Appendix, Materials and Methods*.

**Indirect Immunofluorescence.** Immunofluorescence assays were performed basically as previously described (61, 62). Cells were fixed in 4% paraformaldehyde for 20 min and permeabilized using PBS-Triton X-100 (0.1%). Then, they were blocked at room temperature for 1 h in 1% BSA (in PBS with 0.1% Triton X-100). Primary antibodies were incubated overnight at 4 °C. Finally, Alexa-conjugated secondary IgG antibodies and DAPI (0.1 ng/μL) (Sigma) were used for 2 h at room temperature. Images were obtained using a Leica SP5 confocal microscope by LAS-AF software.

**Western Blot.** Immunoblotting was performed using standard procedures and visualized using the ECL kit (Amersham).

**ChIP Assays.** ChIP assays were performed as previously described (63). See extended protocol in *SI Appendix, Materials and Methods*.

**Cell Cycle Analysis by Flow Cytometry.** Cells were fixed in cold 70% ethanol and stored at –20 °C until they were ready for staining. Cells were washed twice with cold PBS and pelleted by centrifugation. The cell pellet was resuspended in PBS containing 200 μL of 1 mg/mL propidium iodide and 2 mg RNase A and incubated at 37 °C for 15 min.

**ChIP-Seq Procedure.** Chromatin immunoprecipitation as well as sample preparation for sequencing from one replicate were done as previously described (54). For details see expanded protocol in *SI Appendix, Materials and Methods and Table S2*. ChIP-seq data have been deposited in the GEO database under the accession GSE122263 (access token ovizqewwblullyx).

**RNA-Seq Procedure.** RNA was extracted using the High Pure RNA isolation kit from Roche followed by DnaseI treatment from 2 biological independent

samples. Libraries were prepared using the TruSeq Stranded Total RNA Sample Preparation kit with Ribo-Zero Human/Mouse/Rat Kit (Illumina, RS-122–2201/2) according to the manufacturer's protocol. Briefly, 500 ng of total RNA was used for ribosomal RNA depletion. Then, after removing ribosomal RNA, the remaining RNA was fragmented for 4.5 min. The remaining steps of the library preparation were followed according to the manufacturer's instructions. Final libraries were analyzed using an Agilent DNA 1000 chip to estimate the quantity and check size distribution, and then quantified by qPCR using the KAPA Library Quantification kit (Roche, 07960204001) prior to amplification with Illumina's cBot. The libraries were sequenced on Illumina High HiSeq 2500 with paired-end 50 base pair long reads (*SI Appendix, Table S2*). Alignment was performed using HISAT2 (64), assignment of aligned reads to genes was performed using HTSeq (65), and differential expression analysis was performed using DESeq2 (66). RNA-seq data have been deposited in the GEO database under the accession GSE122264 (access token anmrkicwntqjisp).

**ChIP-Seq Data Acquisition and Analysis.** ChIP-seq data were downloaded from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) (Accessions used in this paper are specified in *SI Appendix, Table S3*). See further explanations in *SI Appendix, Materials and Methods*.

**Statistical Analysis.** Quantitative data were expressed as mean and SD (for immunofluorescence countings and RNA transcription experiments) and as mean and SEM (for ChIPs). The significance of differences between groups was assessed using the Student's *t* test.

**Data and Materials Availability.** RNA-seq and ChIP-seq data have been deposited in the GEO database under accessions GSE122264 and GSE122263, respectively. All other materials are available upon request. All the experiments have been approved by the Bioethics Committee of the Consejo Superior de Investigaciones Científicas.

**ACKNOWLEDGMENTS.** We thank Drs. Jiemin Wong and Jiwen Li for the PHF2 plasmid; Dr. Borisy for the pSHIN vector; and Drs. A. Jordá, F. Azorin, J. Bernues, L. L. Espinás, S. de la Luna, A. Vaquero, N. Agell, J. Roig, G. Vicent, M. Beato, M. Arbonés, and E. Martí for reagents; and all members of the M.A.M.-B. laboratory for helpful comments and suggestions. This study was supported by grants BFU2012-34261 and BFU2015-69248-P (to M.A.M.-B.) from the Spanish Ministry of Economy and Fondation Jérôme Lejeune (to M.A.M.-B.).

- L. Tiberi, P. Vanderhaeghen, J. van den Aemele, Cortical neurogenesis and morphogens: Diversity of cues, sources and functions. *Curr. Opin. Cell Biol.* **24**, 269–276 (2012).
- S. Temple, The development of neural stem cells. *Nature* **414**, 112–117 (2001).
- T. Kouzarides, Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
- R. Fueyo, M. A. García, M. A. Martínez-Balbás, Jumonji family histone demethylases in neural development. *Cell Tissue Res.* **359**, 87–98 (2015).
- S. Epsztejn-Litman *et al.*, De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat. Struct. Mol. Biol.* **15**, 1176–1183 (2008).
- B. Wen, H. Wu, Y. Shinkai, R. A. Irizarry, A. P. Feinberg, Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* **41**, 246–250 (2009).
- G. J. Filion, B. van Steensel, Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. *Nat. Genet.* **42**, 4; author reply 5–6 (2010).
- J. Nakayama, J. C. Rice, B. D. Strahl, C. D. Allis, S. I. Grewal, Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
- R. C. Allshire, J. P. Javerzat, N. J. Redhead, G. Cranston, Position effect variegation at fission yeast centromeres. *Cell* **76**, 157–169 (1994).
- N. Saksouk, E. Simboeck, J. Déjardin, Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3 (2015).
- Y. Okuno, K. Inoue, Y. Imai, Novel insights into histone modifiers in adipogenesis. *Adipocyte* **2**, 285–288 (2013).
- K. Fortschegger, R. Shiekhattar, Plant homeodomain fingers form a helping hand for transcription. *Epigenetics* **6**, 4–8 (2011).
- H. Wen *et al.*, Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J. Biol. Chem.* **285**, 9322–9326 (2010).
- G. A. Nicholson *et al.*, The gene for hereditary sensory neuropathy type I (HSN-I) maps to chromosome 9q22.1–q22.3. *Nat. Genet.* **13**, 101–104 (1996).
- K. Hasenpusch-Theil *et al.*, PHF2, a novel PHD finger gene located on human chromosome 9q22. *Mamm. Genome* **10**, 294–298 (1999).
- H. J. Kim *et al.*, Plant homeodomain finger protein 2 promotes bone formation by demethylating and activating Runx2 for osteoblast differentiation. *Cell Res.* **24**, 1231–1249 (2014).
- K. H. Lee, U. I. Ju, J. Y. Song, Y. S. Chun, The histone demethylase PHF2 promotes fat cell differentiation as an epigenetic activator of both C/EBP $\alpha$  and C/EBP $\beta$ . *Mol. Cells* **37**, 734–741 (2014).
- J. Yang *et al.*, Epigenetic regulation of megakaryocytic and erythroid differentiation by PHF2 histone demethylase. *J. Cell. Physiol.* **233**, 6841–6852 (2018).
- D. R. Pattabiraman *et al.*, Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. *Science* **351**, aad3680 (2016).
- A. Ghosh *et al.*, Association of FANCC and PTCH1 with the development of early dysplastic lesions of the head and neck. *Ann. Surg. Oncol.* **19** (suppl. 3), S528–S538 (2012).
- C. Lee, B. Kim, B. Song, K. C. Moon, Implication of PHF2 expression in clear cell renal cell carcinoma. *J. Pathol. Transl. Med.* **51**, 359–364 (2017).
- R. K. C. Yuen *et al.*, Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
- I. Iossifov *et al.*, De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
- J. Cotney *et al.*, The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat. Commun.* **6**, 6404 (2015).
- J. Bricambert *et al.*, The histone demethylase Phf2 acts as a molecular checkpoint to prevent NAFLD progression during obesity. *Nat. Commun.* **9**, 2092 (2018).
- X. N. Wu *et al.*, Methylation of transcription factor YY2 regulates its transcriptional activity and cell proliferation. *Cell Discov.* **3**, 17035 (2017).
- T. Lammens, J. Li, G. Leone, L. De Veylder, Atypical E2Fs: New players in the E2F transcription factor family. *Trends Cell Biol.* **19**, 111–118 (2009).
- S. Tahmasebi *et al.*, Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12360–12367 (2016).
- W. Liu *et al.*, PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508–512 (2010).
- E. Asensio-Juan *et al.*, The histone demethylase PHF8 is a molecular safeguard of the IFN $\gamma$  response. *Nucleic Acids Res.* **45**, 3800–3811 (2017).
- W. Gan *et al.*, R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
- J. Sollier, K. A. Cimprich, Breaking bad: R-Loops and genome integrity. *Trends Cell Biol.* **25**, 514–522 (2015).
- S. J. Nielsen *et al.*, Rb targets histone H3 methylation and HP1 to promoters. *Nature* **412**, 561–565 (2011).
- I. Panteleeva *et al.*, HP1 $\alpha$  guides neuronal fate by timing E2F-targeted genes silencing during terminal differentiation. *EMBO J.* **26**, 3616–3628 (2007).

35. S. Tuduri *et al.*, Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324 (2009).
36. A. Helmrich, M. Ballarino, L. Tora, Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* **44**, 966–977 (2011).
37. A. M. Deshpande, C. S. Newlon, DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030–1033 (1996).
38. E. A. Hoffman, A. McCulley, B. Haarer, R. Arnak, W. Feng, Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome Res.* **25**, 402–412 (2015).
39. H. Merrikkh, C. Machón, W. H. Grainger, A. D. Grossman, P. Soultanas, Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470**, 554–557 (2011).
40. A. Bayona-Feliu, A. Casas-Lamesa, O. Reina, J. Bernués, F. Azorin, Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. *Nat. Commun.* **8**, 283 (2017).
41. S. U. Colmenares *et al.*, Drosophila histone demethylase KDM4A has enzymatic and non-enzymatic roles in controlling heterochromatin integrity. *Dev. Cell* **42**, 156–169.e5 (2017).
42. L. C. Young, D. W. McDonald, M. J. Hendzel, Kdm4b histone demethylase is a DNA damage response protein and confers a survival advantage following  $\gamma$ -irradiation. *J. Biol. Chem.* **288**, 21376–21388 (2013).
43. H. Zheng, L. Chen, W. J. Pledger, J. Fang, J. Chen, p53 promotes repair of heterochromatin DNA by regulating JMJD2b and SUV39H1 expression. *Oncogene* **33**, 734–744 (2014).
44. K. H. Lee *et al.*, PHF2 histone demethylase acts as a tumor suppressor in association with p53 in cancer. *Oncogene* **34**, 2897–2909 (2015).
45. C. Q. X. Yeo *et al.*, p53 maintains genomic stability by preventing interference between transcription and replication. *Cell Rep.* **15**, 132–146 (2016).
46. Y. Okuno *et al.*, Epigenetic regulation of adipogenesis by PHF2 histone demethylase. *Diabetes* **62**, 1426–1434 (2013).
47. B. J. O’Roak *et al.*, Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
48. S. De Rubeis *et al.*; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium, Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
49. R. Bernier *et al.*, Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* **158**, 263–276 (2014).
50. Y. Kondo *et al.*, Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells. *PLoS One* **3**, e2037 (2008).
51. T. Wagner, M. Jung, New lysine methyltransferase drug targets in cancer. *Nat. Biotechnol.* **30**, 622–623 (2012).
52. P. Zeller *et al.*, Histone H3K9 methylation is dispensable for *Caenorhabditis elegans* development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat. Genet.* **48**, 1385–1395 (2016).
53. D. S. Currle, J. S. Hu, A. Kolski-Andreaco, E. S. Monuki, Culture of mouse neural stem cell precursors. *J. Vis. Exp.*, e152 (2007).
54. R. Fuego *et al.*, Lineage specific transcription factors and epigenetic regulators mediate TGF $\beta$ -dependent enhancer activation. *Nucleic Acids Res.* **46**, 3351–3365 (2018).
55. C. Estarás *et al.*, Genome-wide analysis reveals that Smad3 and JMJD3 HDM co-activate the neural developmental program. *Development* **139**, 2681–2691 (2012).
56. N. Blanco-García, E. Asensio-Juan, X. de la Cruz, M. A. Martínez-Balbás, Autoacetylation regulates P/CAF nuclear localization. *J. Biol. Chem.* **284**, 1343–1352 (2009).
57. S. J. Boguslawski *et al.*, Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89**, 123–130 (1986).
58. S. Kojima, D. Vignjevic, G. G. Boris, Improved silencing vector co-expressing GFP and small hairpin RNA. *Biotechniques* **36**, 74–79 (2004).
59. V. Hamburger, H. L. Hamilton, A series of normal stages in the development of the chick embryo. 1951. *Dev. Dyn.* **195**, 231–272 (1992).
60. N. Akizu *et al.*, EZH2 regulates neuroepithelium structure and neuroblast proliferation by repressing p21. *Open Biol.* **6**, 150227 (2016).
61. E. Asensio-Juan, C. Gallego, M. A. Martínez-Balbás, The histone demethylase PHF8 is essential for cytoskeleton dynamics. *Nucleic Acids Res.* **40**, 9429–9440 (2012).
62. S. Sánchez-Molina *et al.*, Regulation of CBP and Tip60 coordinates histone acetylation at local and global levels during Ras-induced transformation. *Carcinogenesis* **35**, 2194–2202 (2014).
63. E. Valls *et al.*, Involvement of chromatin and histone deacetylation in SV40 T antigen transcription regulation. *Nucleic Acids Res.* **35**, 1958–1968 (2007).
64. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
65. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).