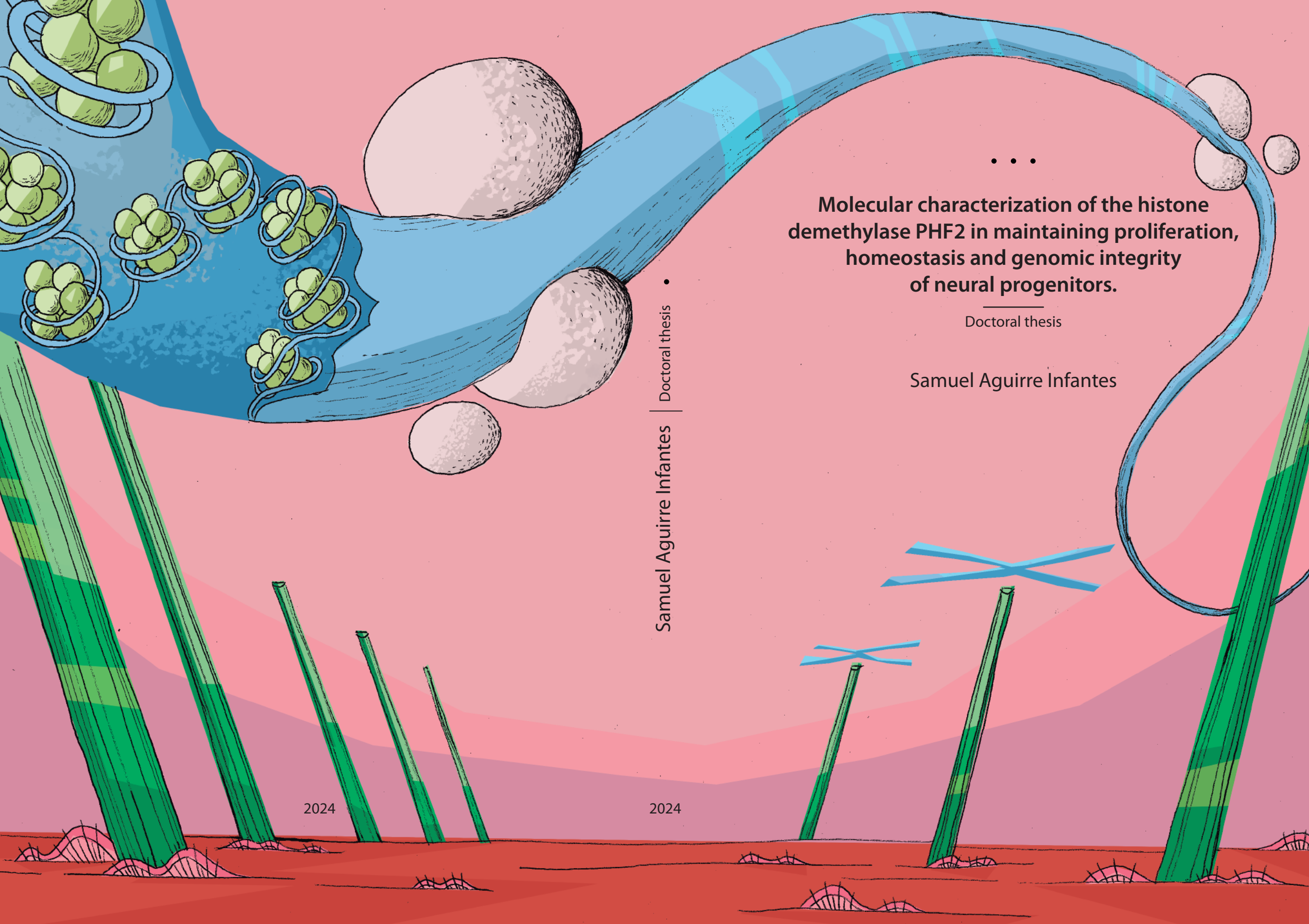
UNIVERSITAT DE BARCELONA

Molecular characterization of the histone demethylase PHF2 in maintaining proliferation, homeostasis and genomic integrity of neural progenitors

Samuel Aguirre Infantes



• • •

Molecular characterization of the histone demethylase PHF2 in maintaining proliferation, homeostasis and genomic integrity of neural progenitors.

Doctoral thesis

Samuel Aguirre Infantes

Samuel Aguirre Infantes | Doctoral thesis •

2024

2024

Doctoral program in Genetics
Faculty of Biology, Universitat de Barcelona

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

Molecular characterization of the histone demethylase PHF2 in maintaining proliferation, homeostasis and genomic integrity of neural progenitors

A thesis submitted by Samuel Aguirre Infantes to obtain the doctoral degree by the Universitat de Barcelona

Author:
Samuel Aguirre Infantes

Director:
Dr. María A. Martínez Balbás

Tutor:
Dr. Marta Pascual Berniola

Barcelona, 2024

ACKNOWLEDGEMENTS

Desde que estudié a Mendel en cuarto de la ESO con Ana Claudia como profesora supe que la genética era una de mis pasiones, con todos los cruces entre alelos dominantes y recesivos, y cómo el genotipo se veía reflejado en el fenotipo. Nadie me advirtió de lo mucho que se iba a complicar la genética hasta llegar a donde estoy ahora, escribiendo una tesis doctoral... Pero llegados a este punto, y echando la vista atrás, sí que ha habido mucha gente que se ha cruzado por mi camino por diversas razones y a los cuales me gustaría dedicar unas palabras de agradecimiento.

En primer lugar, quiero agradecer a Marian. No digo en vano que eres la mejor directora de tesis que una persona podría desear. Siempre disponible, ya sea para consultar dudas científicas, discutir experimentos, tomar un café a las once en punto (o algo más tarde algunos días…) o ponernos a "arreglar el mundo" como tanto nos gusta. Haciendo literalmente magia con los recursos que tenemos para tener una ciencia de calidad y publicar en revistas de primer nivel. Tu cercanía y espontaneidad, haciéndonos partícipes de tu vida, ha hecho que para mí no seas solamente una directora de tesis, sino una compañera y amiga más del laboratorio; porque, al fin y al cabo, el laboratorio es y será siempre parte de nuestra vida.

También me gustaría agradecer a mis mentoras en esta etapa doctoral: Marta y Simona. Aún recuerdo el día que fui a entrevistarme con Marian, siendo la comidilla del laboratorio (y de medio instituto) porque "Marian estaba entrevistando a un chico". "El chico de Marian", decían. Ese día bajamos al "pati" y tomamos café bajo el sol (cosa que

tanto os gustaba a vosotras, y que yo no aguanto… pero no era plan de ponerme a deciros dónde tomar café siendo el nuevo). Desde ahí comenzó una amistad que hoy estoy orgulloso de mantener. Marta, eres un ser de luz. Hemos estado juntos en los mejores momentos (como las bodas que se vienen, un besito a Javi), y también en los peores. Y aún en estos últimos, tú siempre has tenido una visión positiva de todo. Eres y serás la voz de la experiencia para mí. Te echo de menos. Simona, aunque estuve menos tiempo contigo, una de las cosas que me enseñaste fue a tomarme las cosas menos en serio, sobre todo cuando las cosas no salen como uno quiere (que en el laboratorio es más a menudo de lo normal), y no ha podido resultarme más útil. Sé que uno de tus sueños es vivir en Australia, y si tú y Michele lo conseguís, cuenta con una visita por mi parte. Y gracias a Stella, aunque ya te habías ido cuando llegué, fuiste la iniciadora de PHF2 y siento que me allanaste un poco el camino con tu tesis.

Por supuesto debo agradecer a mis compañeras actuales, al igual que a los que han pasado por el lab: Marta Artés, que me soporta cada día, siempre corriendo de un lado a otro. A Mónica, mi primera aprendiz y ahora una loca más sumándose a hacer un doctorado. Me siento muy orgulloso de haber sido quien te inició en esto, más aún de cómo estás evolucionando, y no puedo olvidar los ratos de cotilleo que tuvimos juntos, ya sabes tú que te tenía calada antes que todos los demás. A Ángela, Arnau y ahora Raquel, estudiantes con un futuro brillante por delante, gracias por enseñarme a mí.

También agradecer a mis compañeros del IBMB del lab de Albert en especial a Mónica y Núria, sin cuya ayuda y soporte este viaje de cuatro años no habría sido posible. A Chari, Guille, Carles, Dani, Alicia e Inma, que siempre estuvieron ahí para darme

feedback en todos nuestros *lab meetings*. A Pilar, por ser mi terapeuta personal y mi vegana-lesbiana-futura dueña de una tienda de libros-cafetería favorita. Tu personalidad es de esas que contagian felicidad, no cambies nunca, y sigue deconstruyéndote y deconstruyéndonos. A Mònica, Marcos, Alex, y MariaJo, por ser el comedic relief que uno necesita en un día gris, sin buscarlo ni esperarlo, siempre estáis ahí en el momento oportuno. A Blanca, Kim, Yara, Elisabeth, Ian, Paula, MaJe, las Crises, Miriam, Elena, Tania, y en general, a todos los miembros del IBMB que en su día a día han aportado valor, favores y brillo a estos años de doctorado.

Echando la vista un poco más atrás, tengo que agradecer a mis amigos "barceloneses" (a pesar de que ninguno es de Barcelona). A Nestor. Qué te voy a decir, amigo. Desde el día en el que te vi llegar con el pelo engominado hacia atrás y tu maletín en mano a la preinscripción de Bioquímica supe que lo nuestro iba para largo. Pero en ninguno de mis más locos sueños podría haber imaginado todo lo que hemos vivido juntos. Siempre me acuerdo de tus hojas de apuntes con los márgenes llenos de dibujos, ¿quién sino iba a ser el encargado de hacer la portada de mi tesis? A Dani y Marina. Porque las amistades buenas, son las que perduran en el tiempo sin importar cuando o dónde. Desde Sabadell hasta Estocolmo, sois dos personas increíbles que siempre están ahí cuando la ocasión lo requiere, ya sea para apoyar o para cotillear, todo se nos da bien. Gracias por tantos años de locuras juntos, de historias, de risas y de algún llanto.

Cómo no acordarme de los que me han acompañado, apoyado y animado desde que tengo uso de razón: mis amigos de Mora, mi pueblo natal, en un mar de olivos. Javi, Elena, Diego, Javi, Jaime, Cris, Victoria, Fer, Manu y Ro. Cuántas tardes hemos pasado juntos

en “El Piloto”. Cuántas veces nos hemos esperado en las escaleras del ayuntamiento (siempre llegando yo el primero). Con vosotros han sido muchas de nuestras primeras veces. Sois el soplo de aire fresco cuando Barcelona se hace demasiado grande, aunque sea de Fiesta del Olivo en Fiesta del Olivo. Somos un grupo pintoresco cuanto menos, pero dicen que en la variedad está la riqueza, ¿no? Mención especial al grupo de amigos de Almansa: José, Cuenca, Javichu, Carmen, Belén, Rubén. Quién nos iba a decir que dos semanas en Daimiel iban a ser suficientes para tener amigos de por vida. Aunque a veces no os lo diga, os pienso muy a menudo, y os echo de menos. Los amigos son la familia que uno elige, y yo tengo el carné de familia numerosa.

Y hablando de familia, está claro que hoy soy la persona que soy gracias a la mía, a los que están y a los que ya no. Gracias, mamá y papá, por haberme inculcado los valores del trabajo y del esfuerzo, la curiosidad por el saber, el amor por la cultura y el respeto por todo y todos desde el momento en que nací. Gracias por vuestro esfuerzo y por vuestros sacrificios para darme todo lo que estaba en vuestra mano y formarme como persona. Gracias por las clases de música (que no eran pocas), el karate, el tenis, el baloncesto, la natación, el teatro, las jotas, la carrera, el máster y posiblemente alguna otra actividad que se me pase ahora mismo. Gracias por tener un horario con mis actividades delante de la tabla de planchar, para saber dónde me teníais que recoger, y a dónde me teníais que llevar cada día. Gracias por demostrarme vuestro amor, por alegraros por mis éxitos, y enseñarme a hacerlo mejor cuando era necesario. Han pasado ya diez años desde el día en que mirasteis a la habitación de un niño de dieciocho que se quedaba vacía para irse a estudiar a Barcelona y pensasteis: “qué pocas veces vamos a

ver esta habitación llena de nuevo". Esta tesis y todo lo que representa es para vosotros. Gracias a mi niña, mi ratita, mi rubia, Ángela. Eres todo lo que un hermano mayor podría desear. Me siento muy orgulloso de ti, de todo lo que trabajas, de todo el amor que nos demuestras a los demás, y de la mujer en la que te estás convirtiendo. Tenerte cerca es tener un trocito de casa aquí, y escuchar tu risa es una infusión de felicidad. Prometo que nuestra relación no va a cambiar nunca. No ha pasado ni un solo día en que no haya pensado en vosotros, os quiero con toda mi alma.

Para finalizar, pero no menos importante, quiero agradecer a la persona que me aguanta desde que me levanto hasta que me acuesto. José, tú más que nadie sabes lo que este doctorado significa. Gracias por ser mi refugio en los días en los que los resultados no tenían sentido. Gracias por entenderme cuando tenía la cabeza llena de incógnitas, experimentos, y problemas, y no me apetecía hablar tanto. Gracias por todas esas cenas que pedimos cuando no nos apetece cocinar, en especial cuando son de comida asiática y la pides porque es mi preferida. Gracias por interesarte por esta tesis, porque, aunque te cueste entenderlo, siempre me preguntas qué tal han ido los experimentos, y qué significan los resultados. Gracias por los viajes que hemos hecho, y los que están por venir. Esta etapa no habría sido lo mismo si no hubieras estado a mi lado. Por último, gracias por acceder a pasar una vida conmigo. Esta tesis también es para ti.

ABSTRACT

The histone demethylase PHF2, previously studied in the laboratory, has been shown to be crucial for the expansion of neural progenitors by maintaining low levels of H3K9me2/3 at the promoters of cell cycle and developmental genes. It was also discovered that the absence of PHF2 induces the expression of repetitive elements such as Major Satellite, and that PHF2 interacts with components of heterochromatin.

The thesis focuses on the interaction of PHF2 with constitutive pericentromeric heterochromatin and its role during mitosis. Experiments were conducted on neural stem cells (NSCs) both in control conditions and depleted of PHF2 through shRNA. The effects were analyzed across three areas of chromatin: euchromatin, pericentromeric heterochromatin (PcH), and the borders of PcH.

In euchromatin, PHF2 was observed to localize at the promoters of cell cycle and developmental genes. Depletion of PHF2 resulted in increased levels of H3K9me3 and decreased accessibility at these promoters.

In heterochromatin, PHF2 was enriched in satellite repeat sequences, such as GSAT_MM. Depletion of PHF2 increased the accessibility of heterochromatin and decreased H3K9me3 levels, leading to increased transcription of repetitive sequences. Live imaging experiments confirmed these changes in methylation.

Upon observing the disorganization of heterochromatin and considering its nucleation and extension mechanisms described in the literature, the study examined the boundary region between heterochromatin and euchromatin to elucidate what was causing this disruption of the heterochromatin and further genome instability. A significant enrichment of PHF2 was observed at the borders of pericentromeric heterochromatin (from megabase 3 to 3.5 on each chromosome). Depletion of PHF2 in these regions led to an increase in H3K9me3 and a decrease in accessibility, suggesting that PHF2 balances H3K9me3 across the genome.

Eliminating PHF2 causes an imbalance between euchromatin and heterochromatin, silencing euchromatic regions and de-repressing previously repressed genes that were not targeted by PHF2. The integrity of heterochromatin can be affected either by the lack of transcriptional activity in the boundary regions or by the dilution of constitutive heterochromatin components, which are limited. This increases the expression of repetitive sequences, causing genomic instability and DNA damage.

PHF2 requires both its PHD and JmjC domains to maintain the balance of H3K9me3 and genomic integrity. This was confirmed through overexpression and rescue experiments with mutants for each of these domains in shPHF2 neural stem cells. The positively charged amino acid-rich domain, important for phase separation and gene expression, is not essential for the integrity of heterochromatin.

In collaboration with Amanda Fisher's laboratory at the University of Oxford, the role of PHF2 during mitosis was studied. It was observed that PHF2 binds to mitotic chromosomes, affecting chromosomal size: depletion of PHF2 reduces chromosome size and alters methylation marks such as H3K9me3 and H3K27me3. An increase in H3K27me3 in the body of the chromosome and a reduction in H3K9me3 levels in the chromosome centromere were observed. Temporal transcriptomics revealed that post-mitotic transcription reactivation was delayed when PHF2 was depleted, especially in histone genes.

This doctoral thesis has significantly advanced the understanding of PHF2's role in balancing H3K9me3 and maintaining genomic stability. The research has demonstrated the importance of PHF2 in chromatin regulation and the proliferation of neural progenitor cells, highlighting its critical function in both heterochromatin and mitosis.

LIST OF CONTENTS

LIST OF ABBREVIATIONS

ASD: autism spectrum disorder

ATAC: assay for transposase accessible chromatin

ATAC-seq: assay for transposase accessible chromatin sequencing

ATP: adenosine triphosphate

BH: Benjamini-Hochberg

BIR: break induced replication

BMP: bone morphogenetic protein

Bp: base pair

BSA: bovine serum albumin

Cdk: cyclin dependent kinase

ChIP: chromatin immunoprecipitation

ChIP-seq: chromatin immunoprecipitation sequencing

Chr: chromosome

CNS: central nervous system

CoIP: co-immunoprecipitation

cPAS: combinatorial probe-anchor synthesis

CPM: counts per million

DAPI: 4',6-Diamino-2-Phenylindole

DMSO: dimethyl sulfoxide

CpG: cytosine-phosphate-guanine

CpH: cytosine-phosphate-histidine

CT: control

DMEM: Dulbecco's modified eagle medium

DNA: deoxyribonucleic acid

DNMT: DNA methyltransferase

dNTP: deoxynucleotide triphosphate

DOX: doxycylcline

DSB: double strand breaks

dsDNA: double-stranded DNA

EDTA: ethylenediamine tetraacetic acid

EGTA: ethylene glycol tetraacetic acid

EGF: epidermal growth factor

EGFP: enhanced green fluorescent protein

ESC: embryonic stem cell

EZH2: enhancer of zeste homolog 2

FACS: fluorescence-activated cell sorting

FAD: flavin adenine dinucleotide

FBS: fetal bovine serum

FC: fold change

FDR: false discovery rate

FGF: fibroblast growth factor

GEO: gene expression omnibus

GF: growth factor

GO: gene ontology

GRO-seq: global run-on sequencing

H: histone

HAT: histone acetyl transferase

HDAC: histone deacetylase

HJ: Holiday junction

HKDM: histone lysine demethylase

HKMT: histone lysine methyltransferase

HP1: heterochromatin protein 1

HP1BP3: heterochromatin protein 1 binding protein 3

HR: homologous recombination

HRMT: histone arginine methyltransferase

IDR: intrinsically disordered region

IF: immunofluorescence

IGV: integrated genome viewer

JmjC: Jumonji-C domain containing

Kb: kilobase

KD: knock down

KDM: lysine demethylase

KO: knock out

LAD: lamina-associated domain

LINE: long interspersed element

LSD: lysine-specific histone demethylase

Mb: megabase

MBD: methyl CpG binding domain

mEGFP: monomeric enhanced green fluorescent protein

mRNA: messenger RNA

MS: mass spectrometry

MSC: mesenchymal stem cell

MZ: marginal zone

ncRNA: non coding RNA

NHEJ: non-homologous end joining

NPC: neural progenitor cell

NSC: neural stem cell

PBS: phosphate buffered saline

PcG: polycomb group of proteins

PcH: pericentromeric heterochromatin

PCR: polymerase chain reaction

PDL: poly-D-lysine

PHD: plant homeodomain

PI: propidium iodide

PIC: protein inhibitor cocktail

PLB: passive lysis buffer

PolyA: polyadenylic

PSF: point spread function

PTM: post-translational modification

qPCR: quantitative polymerase chain reaction

RF: replication fork

RG: radial glial cells

RIPA: radioimmunoprecipiation assay buffer

RNA: ribonucleic acid

RNAPII: RNA polymerase II

ROI: region of interest

RT: retrotranscription

Sat: satellite

SD: standard deviation

SDS: sodium dodecyl sulphate

Shh: sonic hedgehog

shRNA: short hairpin RNA

SINE: short interspersed element

SSRF: super-resolution radial fluctuations

ssDNA: single stranded DNA

SVZ: subventricular zone

TAD: topologically associated domain

TBE: tris-borate-EDTA

TET: ten eleven translocations

TF: transcription factor

TLS: translesion synthesis

tRNA: transfer RNA

TSS: transcription start site

VZ: ventricular zone

WT: wild type

ZF: zinc finger

INTRODUCTION

The term "epigenetics" was introduced in 1942 by embryologist Conrad Waddington. He related it to the 17th-century concept of "epigenesis" and defined it as the complex of developmental processes that occur between the genotype and the phenotype (1). Embryonic development relies on the intricate interplay of epigenetic regulators, transcription factors (TFs), and signaling pathways. These elements are meticulously coordinated in a spatial-temporal manner to ensure the precise gene expression programs required for the differentiation and function of each cell type (2). Signaling pathways provide the external and internal cues that guide cellular behavior, transcription factors interpret these signals and activate or repress target genes, and epigenetic regulators modify the chromatin landscape to facilitate or hinder access to these genes (3). This harmonious coordination is essential for the proper progression of developmental processes, ultimately shaping the complex architecture and functionality of the organism (4).

During the present work, I explored the function of the histone demethylase PHF2, that was previously described to be controlling neural progenitor proliferation (5), broadening the knowledge in heterochromatin stability, trimethyl K on histone H3 (H3K9me3) balance and through mitosis during early neural development.

1. CHROMATIN

Chromatin is a repetitive structure that packages the DNA within the cell nucleus. It is formed by millions of repetitions of its fundamental unit, the nucleosome, composed of 147 base pairs (bps) of DNA wrapped around an octamer of four core histones. This octamer has two copies each of histones H3, H4, H2A, and H2B. In between each nucleosome we found linker DNA, bound by histone H1, that connects these nucleosomes and further compacts the chromatin structure. Histone H1 plays a crucial role in stabilizing the nucleosome core and facilitating additional chromatin condensation (4,6,7) (Figure I1).

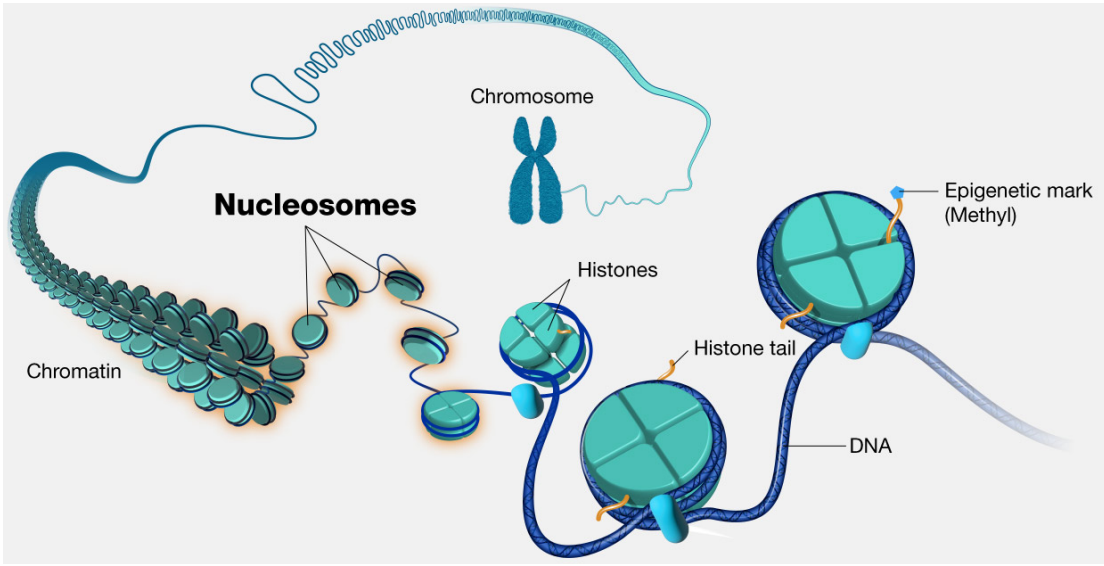


Figure I1. Representative image depicting chromatin in all its different compaction states from chromosome to nucleosome fiber. Detailed histone octamers with histone H1 and histone tail modification also are showed. Adapted from Darryl Leja, National Human Genome Research Institute.

1.1 Epigenetic modifications

Epigenetic modifications are heritable changes in gene expression that do not involve changes in the DNA sequence, such as DNA methylation and histone modifications, and play a critical role in various diseases and cellular functions (4). These epigenetic modifications and their regulation impact chromatin compaction, topological organization, and accessibility (8); and their disruption can lead to developmental conditions like autism spectrum disorder (ASD) or intellectual disability (9,10).

The modifications by which chromatin is regulated will be explained in the following subsections of this doctoral thesis.

1.1.1 DNA Methylation

DNA methylation is an epigenetic modification traditionally associated with gene repression. It involves the addition of methyl groups to GC-rich regions known as CpG islands, typically located at the 5' ends of many genes (11). When these CpG islands become methylated during development, their associated promoters are silenced. This transcriptional repression occurs both through the direct obstruction caused by the methyl group, which prevents other proteins from binding to the DNA, and through the recruitment of repressive factors such as methyl-CpG binding domain proteins (MBD) 1, MBD2, and MBD3 and methyl-CpG binding protein 2 (MeCP2) (12,13).

Various enzymes are involved in adding methyl groups to DNA, functioning to either maintain or increase global methylation levels. DNA methyltransferase 1 (DNMT1) maintains the methylation pattern across cell generations by adding methyl groups to CpG islands that are methylated in the parental strand (14). In contrast, DNMT3A and DNMT3B catalyze de novo methylation, primarily in embryonic cells (15).

DNA can also be demethylated through active or passive mechanisms. Active DNA demethylation involves the removal of methyl groups by ten-eleven translocation (TET) enzymes (16). Passive demethylation occurs through successive rounds of DNA replication without the re-addition of methyl groups (17). Several factors promote DNA demethylation, including DNA cytosine deaminases, DNA glycosylases, DNA repair factors, transcription factors, and even DNA methyltransferases (18,19).

1.1.2 Histone post-translational modifications

Histones, like DNA, are subjected to epigenetic modifications that might be addition or removal of different chemical groups (Figure I2). The different combination of these post-translational modifications has been traditionally defined as the “histone code”, and this give them the ability to regulate several chromatin-related functions (20). Histones, which are highly conserved among organisms, have two main domains: the globular domain, mediating the formation of H2A–H2B and H3–H4 dimers, and the unstructured, flexible N-terminal tail that protrudes from the nucleosome core and interacts with DNA (21). While residues in both domains can undergo PTMs (22), the N-terminal tails are particularly relevant for transcriptional regulation, replication, recombination, and DNA

repair. PTMs on histone tails, including acetylation, methylation, phosphorylation, ubiquitylation, and SUMOylation (23), are mediated by "writer" and "eraser" enzymes that add or remove these modifications respectively, and “reader”, that recognize specific binding sites, affecting histone-DNA interactions and chromatin compaction (4,24–27).

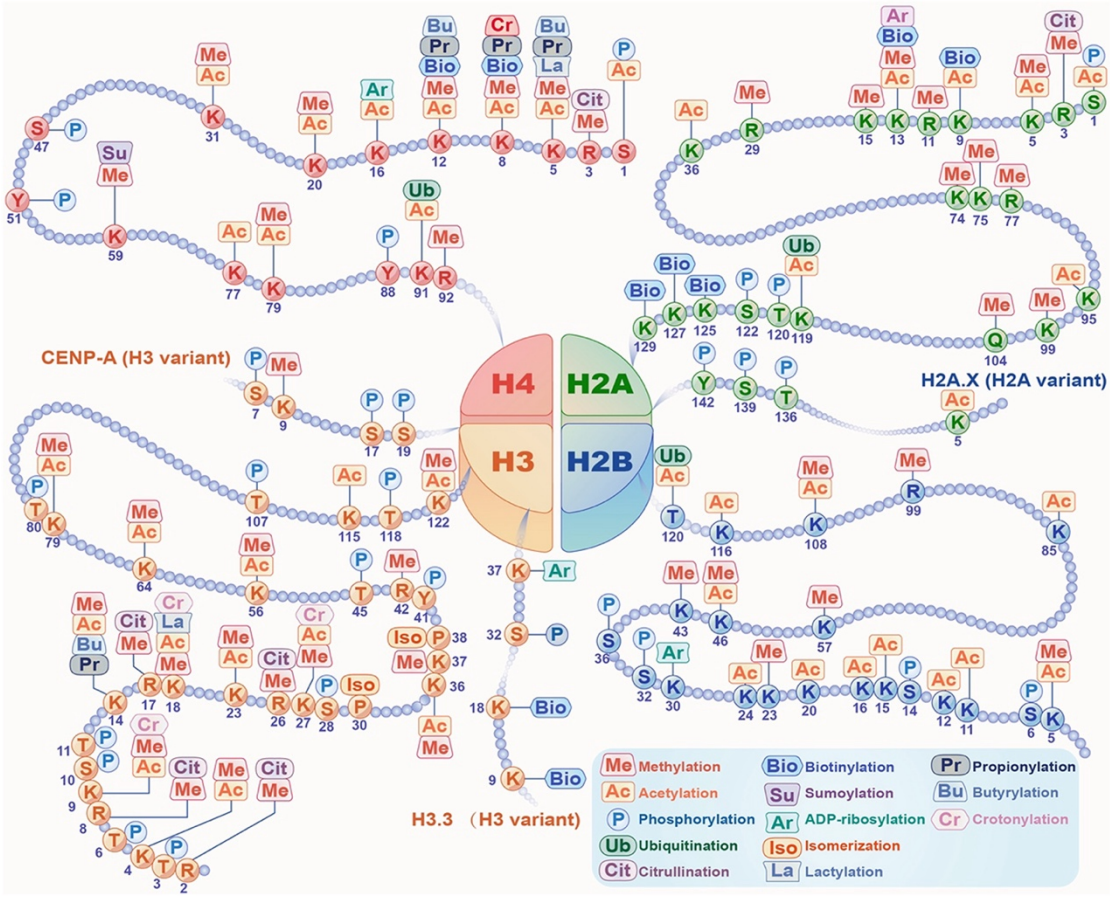


Figure I2. Summary of all histone modifications on the tails of the four core histones, H2A, H2B, H3 H4, in addition to histone variants H2A.X, CENP-A or H3.3. Adapted from (23).

Acetylation is the most extensively studied histone tail modification (28), with histone acetyl transferases (HATs) and histone deacetylases (HDACs) playing key roles. HATs use acetyl-CoA to add acetyl groups to lysine residues, such as K9, K14, K18, K23, and K27 on histone H3, weakening histone-DNA interactions and correlating with active transcription at transcription start sites (TSS) and enhancers. HDACs remove these acetyl groups, often functioning as corepressors (29,30).

More recently, focus has shifted towards histone methylation, which is significant for transcriptional regulation and involves enzymes that add (methyltransferases) and remove (demethylases) methyl groups on lysine and arginine residues (31–33). This modification was once thought permanent until the discovery of histone demethylases like lysine-specific demethylase 1 (LSD1) (KDM1A), which highlighted the dynamic nature of methylation (34).

These modifications and their regulatory enzymes are essential for proper embryonic development and cellular function, ensuring that chromatin structure and gene expression are precisely controlled. The following subsections will explain in detail methylation and demethylation, since this last one is crucial for the correct comprehension of this doctoral thesis.

1.1.2.1 Histone methylation and methyltransferases

Histone methylation involves the addition of one, two, or three methyl groups to lysines (Lys) or one or two to arginines (Arg), catalyzed by histone lysine methyltransferases (HKMTs) or arginine methyltransferases (HRMTs), respectively (32). Although this modification does not alter the electrical charge of the amino acid, it has significant functional consequences. Methylation can occur on various residues of histones, with lysines 4, 9, 27, 36, and 79 on histone H3 and lysine 20 on histone H4 being particularly prominent targets (Figure I3). Histones can be mono-, di-, or tri-methylated, and in mammalian cells, up to 40-80% of histones are dimethylated at H3K9, H3K27, H3K36, or H4K20, while mono- and tri-methylation are less common (35–37).

The association between histone methylation and gene regulation depends on the specific residue and the degree of methylation. H3K4me3 and H3K36me3 are usually linked to active transcription, whereas H3K27me2/3, H3K9me2/3, and H4K20me2/3 are associated with gene repression (38,39). For instance, H3K4me1/2 typically marks enhancers, H3K4me3 is found at promoters or around the transcription start site (TSS), and H3K36me3 marks gene bodies. Conversely, H3K27me3 is enriched at promoters of silent genes, while H3K9 and H4K20 methylation marks are more uniformly distributed across inactive regions. Notably, "bivalent" promoters, which are marked by both H3K27me3 and H3K4me3, maintain genes in a "poised" state for rapid activation or repression during development (40).

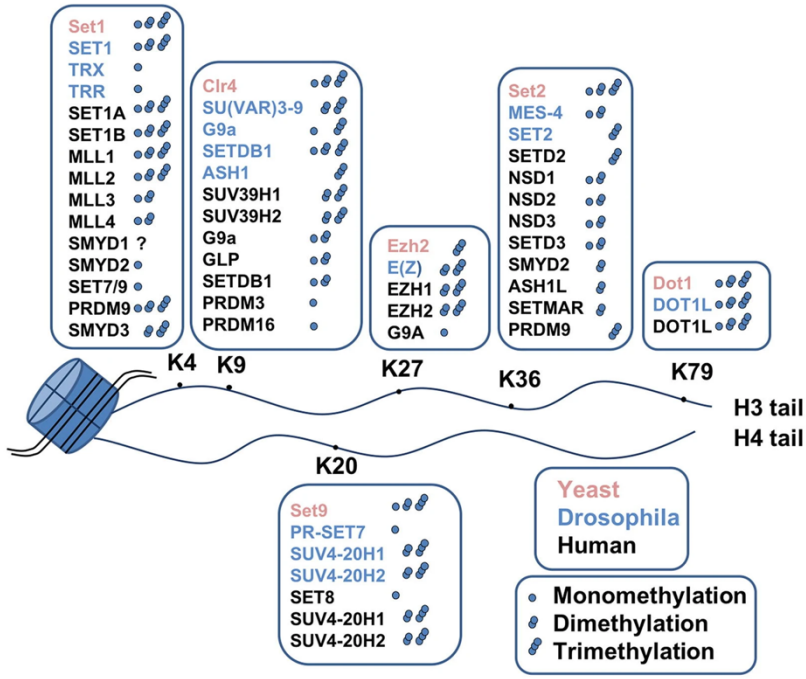


Figure I3. Representative image depicting histone H3 and H4 tails and their different residues that can be methylated, along the histone methyltransferase that deposit it. Adapted from (37).

Histone lysine methyltransferases (HKMTs) like EZH2/1 and MLL are particularly relevant, forming part of the Polycomb repressive complex 2 (PRC2) and the MLL complex, respectively. PRC2 mediates chromatin compaction and transcriptional repression by di- and tri-methylating H3K27, thus repressing cell identity and developmental genes (41). In contrast, MLL complexes promote gene activation through H3K4 methylation (42). The antagonistic functions of these complexes are crucial for the proper regulation of developmental genes and cellular identity determination (43). Also,

G9a and SUV39 are the HKMTs that deposit H3K9me2/3 in euchromatin and heterochromatin respectively (44,45).

Additionally, H4K20me3 and H3K9me3 are enriched at pericentromeric chromatin (46), while H3K27me3 is prevalent at silent genes and subtelomeric regions, although less frequent in other non-genic areas (47). H3K9me2/3 and H3K27me3 can also be found in expansive chromatin domains, contributing to the establishment of distinct epigenetic patterns that regulate chromatin structure and gene expression (48).

1.1.2.2 Histone demethylation and demethylases

For many years, histone demethylation was thought to be a passive process resulting from histone exchange or dilution during replication. However, numerous histone lysine demethylases (KDMs), or "eraser" proteins, have now been identified, targeting most methylated residues such as H3K4, H3K9, H3K27, H3K36, H4R3 and H4K20 (Figure I4) (49,50). These discoveries have revealed that histone methylation is a highly dynamic process. There are two evolutionarily conserved families of KDMs, which utilize different catalytic mechanisms to remove methyl groups: the LSD family and the Jumonji C (JmJC) domain-containing proteins family.

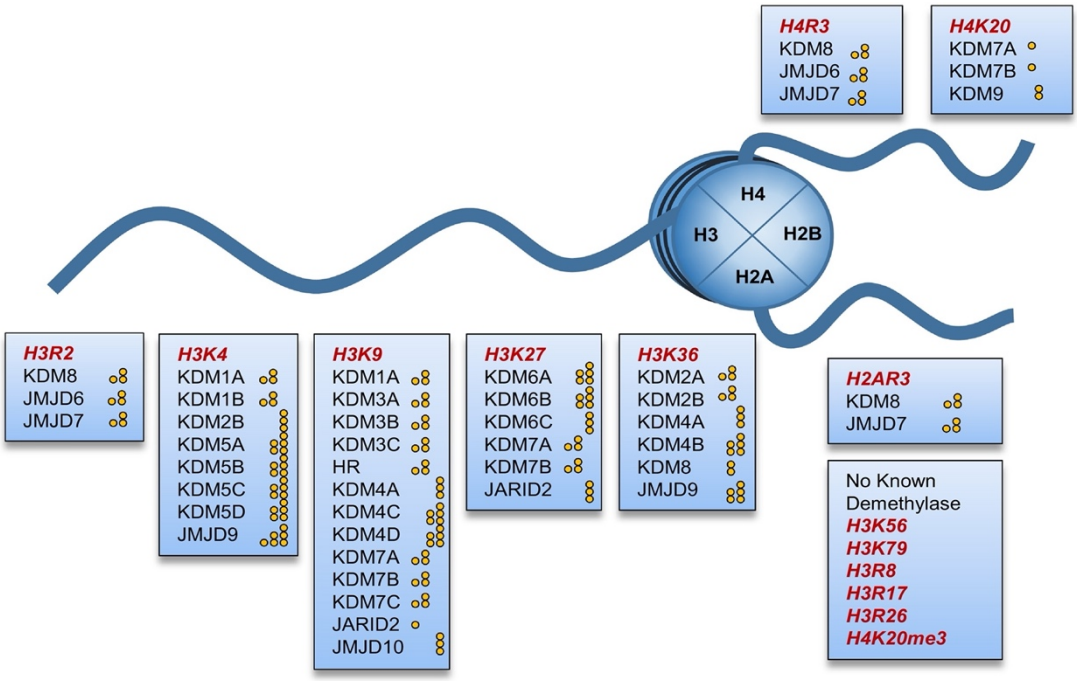


Figure I4. Representative image depicting histone H2A, H3 and H4 tails and their different residues that can be demethylated, along the histone demethylase that removes the methyl group. Adapted from (50).

The LSD family employs a flavin adenine dinucleotide (FAD)-dependent amine oxidation reaction to demethylate histones. This family includes only two enzymes, LSD1 and LSD2. LSD1, the first identified histone demethylase, can demethylate H3K4me1/2 and H3K9me1/2, as well as some non-histone targets (51). LSD2 is also capable of demethylating H3K4me1/2 but has not been shown to act on H3K9. Because of their reaction mechanism requiring of a free electron pair at the methylated residue, LSD enzymes can only demethylate mono- and dimethylated, but not trimethylated, lysine residues (34).

In contrast, the JmJC family, to which our protein of interest belongs, uses an Fe (II)-dependent mechanism and alpha-ketoglutarate as a cofactor to catalyze demethylation. This reaction converts the methyl group to a hydroxymethyl group, which is then released as formaldehyde. The JmJC family targets a broader range of methylated residues compared to the LSD family (52,53). (Figure I5)

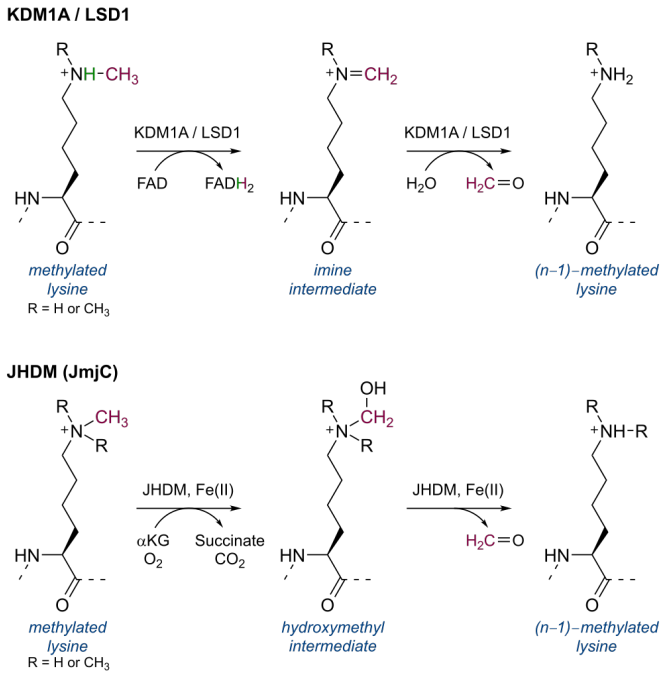


Figure I5. Schematic representation of the mechanisms of lysine demethylation from LSD (top) and JmjC (bottom) families. Adapted from (52)

The identification and study of these demethylases have significantly changed our understanding of histone methylation, revealing it to be a reversible and regulated process that plays a crucial role in gene expression and chromatin dynamics.

1.1.2.3 KDM7 family of histone demethylases

The JmjC-containing KDM family encompasses 24 JmjC-domain-containing KDMs, which are evolutionarily conserved from yeast to humans (Revisit Table I1). These enzymes play significant roles in various physiological and pathological processes, such as gene expression, embryonic stem cell renewal, cellular differentiation, X-linked mental retardation, and cancer (54–58). Their function is determined not only by the JmjC catalytic domain but also by the combination of other conserved structural domains such as PHD, Tudor, CXXC, FBOX, ARID, LRR, and JmjN. Based on structural similarities and histone specificity, these KDMs can be classified into seven subfamilies (Table I1) (33,52).

FAMILY	HDM	OTHER NAMES	HISTONE SUBSTRATE
LSD	KDM1A	KDM1, KIAA0601, LSD1, AOF2	H3K4me2/me1, H3K9me2/me1
	KDM1B	LSD2, AOF1, C6orf193	H3K4me2/me1, H3K9me2/me1
JMJC	KDM2A	JHDM1A, KIAA1004, CXXC8, FBL7, FBXL11	H3K36me2/me1
	KDM2B	JHDM1B, CXXC2, FBL10, FBXL10, PCCX2	H3K4me3, H3K36me2
	KDM3A	JHDM2A, JMJD1, JMJD1A, KIAA0742, TSGA	H3K9me2/me1
	KDM3B	JHDM2B, JMJD1B, KIAA1082, C5orf7	H3K9me2/me1
	JMJD1C	JHDM2C, KIAA1380, TRIP8	H3K9me2/me1
	KDM4A	JHDM3A, JMJD2A, KIAA0677	H3K9me3/me2, H3K36me3/me2, H1.4K26me3
	KDM4B	JHDM3B, JMJD2B, KIAA0876	H3K9me3/me2, H3K36me3/me2, H1.4K26me3
	KDM4C	JHDM3C, JMJD2C, KIAA0780, GASC1	H3K9me3/me2, H3K36me3/me2, H1.4K26me3
	KDM4D	JHDM3D, JMJD2D, FLJ10251	H3K9me3/me2, H1.4K26me3/me2
	KDM5A	JARID1A, RBP2	H3K4me3/me2
	KDM5B	JARID1B, PLU1	H3K4me3/me2
	KDM5C	JARID1C, SMCX	H3K4me3/me2
	KDM5D	JARID1D, SMCY	H3K4me3/me2
	KDM6A	UTX	H3K27me3/me2
	KDM6B	JMJD3, KIAA0346	H3K27me3/me2
	UTY		
	KDM7A	JHDM1D, KIAA1718, KDM7	H3K9me2/me1, H3K27me2/me1
	PHF8	JHDM1F, KIAA1111, ZNF422	H3K9me2/me1, H4K20me1
	PHF2	JHDM1E, KIAA0662	H3K9me2/me1, H4K20me3

Table I1. Summary of LSD and JmjC histone demethylase families, containing their members and main substrate for each one. Adapted from (49).

A notable class within this family is characterized by a single N-terminal plant homeodomain (PHD) zinc finger, a domain associated with methylated lysine residues (59). In humans, this group consists of three members: plant homeodomain fingers 2 and 8 (PHF2 and PHF8) and KDM7A (KIAA1718) (60,61) (FigureI6). Structures of their PHDs and JmjC-domains have been elucidated by X-ray crystallography (62). Besides conserved zinc-chelating residues, these PHDs contain an aromatic cage of phenylalanine and tyrosine residues, which interact with methylated lysine. Biochemical experiments have shown that these proteins specifically interact with histone H3 methylated at lysine 4 via their PHDs (63,64). The active JmjC-domains contain conserved amino acids essential for binding Fe^{2+} and α-ketoglutarate, crucial cofactors for their demethylase activity. Additionally, the KDM7-family has four short α-helices in the C-terminal region of the JmjC, necessary for activity (61). PHF8 preferentially demethylates H3K9me2/1 and H4K20me1, KDM7A mainly demethylates H3K9me1/2 and H3K27me1/2, and PHF2 demethylates H3K9me1/2 (60).

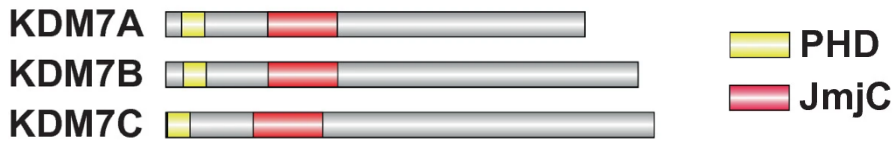


Figure I6. Structure of KDM7 subfamily of histone demethylase, where a PHD domain is located at the very N-terminal, and JmjC catalytic domain is located after the first. Adapted from (65).

Interestingly, the properties of the C-terminal halves of KDM7 proteins are less understood. They do not contain known protein domains, and the homology between the three human proteins is low in this region compared to that of the PHD and JmjC domains. However, all three human proteins have a putative coiled-coil region, and PHF2 and PHF8 exhibit several phosphorylation sites important for regulating their activity (66). The KDM7 family has been described as a new class of transcriptional co-activators, creating a more permissive chromatin environment at promoters by associating with H3K4me3 and removing H3K9me2/1, H3K27me2/1, or H4K20me1 modifications (60,61).

PHF8 is the most well-studied member of this subfamily, partly because mutations in the PHF8 gene can cause Siderius-Hamel syndrome, an X-linked mental retardation often accompanied by cleft lip and/or cleft palate. These mutations usually lead to truncations before or in the JmjC domain (67). The balance between histone methylation and demethylation is crucial for development, and specific alterations in these activities are linked to neurodevelopmental disorders (49).

In the next subsection, I will focus on PHF2 and the previous studies from our lab regarding this KDM7 member due to its relevance in this doctoral thesis.

1.1.2.4 PHF2 demethylase

Histone demethylase PHF2 contains a PHD, JmjC domain and a large instrinsically disordered region (IDR) (Figure I7). Initially, PHF2 was thought to be enzymatically inactive due to a tyrosine residue replacing the second conserved Fe2+ binding histidine.

However, it was shown to become an active H3K9me2 demethylase through PKA-mediated phosphorylation (68,69). Biochemical studies demonstrated that PHF2 demethylates H3K9me2 upon interaction with H3K4me2/3 through its PHD domain (63). It plays a crucial role in regulating gene expression and chromatin structure by demethylating histone H3 at lysine 9 (H3K9me2/me1) (5). This modification typically represses transcription, so PHF2's activity is essential for activating specific gene sets. In addition, it has also been described as a repressor of rDNA transcription by competing with PHF8 for binding to the rDNA promoter and by recruiting H3K9me2/3 methyltransferase SUV39H1 (70).

PHF2 was first characterized as a novel PHD finger gene mapping to human chromosome 9q22 and mouse chromosome 13, with the human and mouse proteins being 98% identical. Initially identified as a candidate gene for hereditary sensory neuropathy type I due to its high expression in the neural tube and dorsal root ganglia, PHF2 mutations have since been found in patients with autism spectrum disorder (71,72). PHF2 influences a wide range of biological functions, including tumor suppression (73), metabolic regulation (74), DNA repair (75), DNA replication (76), and cellular differentiation (5,77). In metabolism, PHF2 regulates glycolysis and gluconeogenesis by interacting with key metabolic regulators and affecting their transcriptional activity (78). Additionally, PHF2 modulates the expression of genes necessary for adipocyte differentiation, such as C/EBPα and PPARγ, and regulates genes essential for neuronal differentiation and function (5).

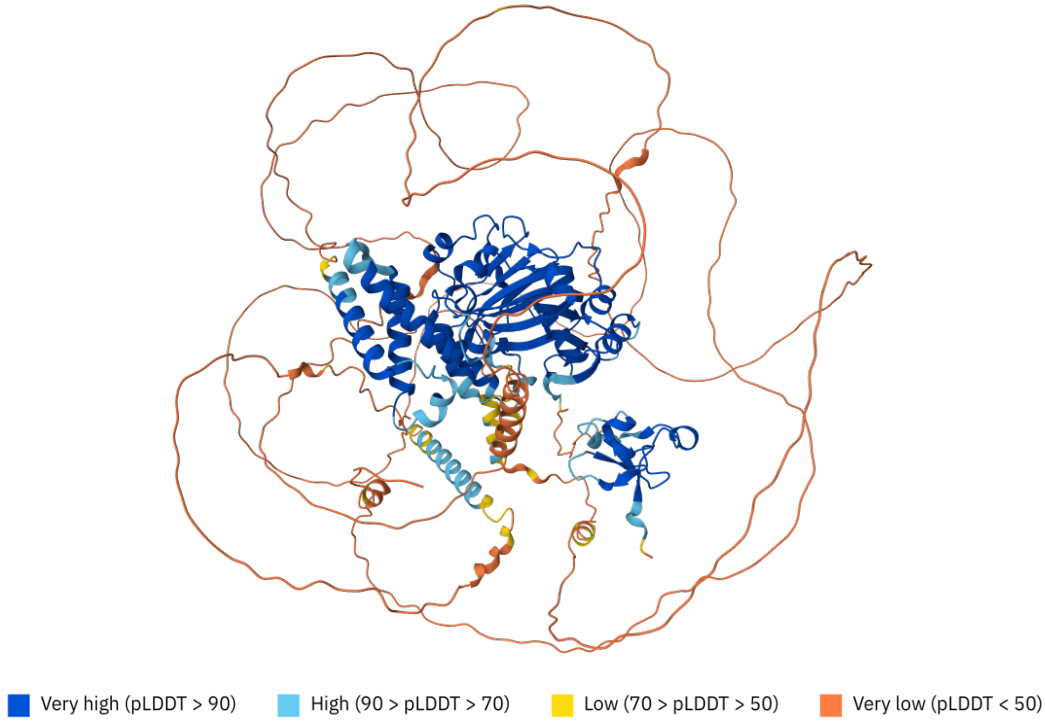


Figure I7. Predicted structure for PHF2 mouse protein. The two domains are clearly identified (JmjC on top left, and PHD bottom right), meanwhile the C terminal region appears as an intrinsically disordered region. Bottom legend shows the confident score on local Distance Difference Test. Adapted from (79)

In terms of metabolic regulation, systemic Phf2 null mice showed partial neonatal death, growth retardation, and reduced adipose tissue and adipocyte numbers (74). PHF2 also regulates osteoblast differentiation and bone formation, as evidenced by PHF2 transgenic mice developing calvaria and limb bones earlier than wild-type mice (80). Furthermore,

PHF2 plays a role in erythroid differentiation by acting as a negative epigenetic regulator through the demethylation of H3K9me2 at the promoter region of the p53 gene (81).

Pathologically, PHF2 acts as a tumor suppressor in various cancers by maintaining the expression of tumor suppressor genes. Reduced PHF2 expression or mutations are associated with poor prognosis in some cancers, suggesting its potential as a prognostic marker (73,82–84) (Figure I8). In metabolic disorders, alterations in PHF2 function contribute to conditions like obesity and type 2 diabetes, as PHF2-mediated epigenetic changes impact insulin sensitivity and glucose metabolism (74). Neurologically, dysregulation of PHF2 is linked to cognitive impairments and conditions affecting synaptic plasticity and neuronal health (5,72). Additionally, PHF2 can form a complex with ARID5B protein, demethylating ARID5B at Lys 336 to facilitate DNA binding and subsequent gene activation. This interaction is also associated with chondrogenesis (85).

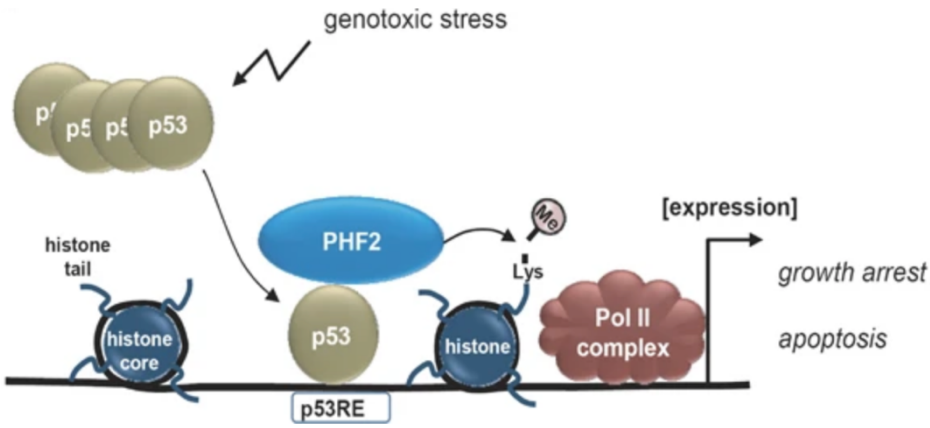


Figure I8. PHF2's role in tumor suppression in association with p53. Adapted from (73)

1.1.2.4.1 PHF2, H3K9 methylation and gene expression

As described in previous works from the lab, PHF2 is crucial for both neural progenitor proliferation and early neurogenesis in the chicken spinal cord. Through genome-wide analyses and biochemical assays, we demonstrated that PHF2 regulated the expression of essential cell cycle progression genes, particularly those involved in DNA replication, by maintaining low levels of H3K9me3 at promoters. Consequently, the depletion of PHF2 led to the accumulation of R-loops, resulting in significant DNA damage and cell cycle arrest. These findings highlighted PHF2's role as a protector of genome stability, which was crucial for the proper expansion of neural progenitors during development (5) (Figure I9).

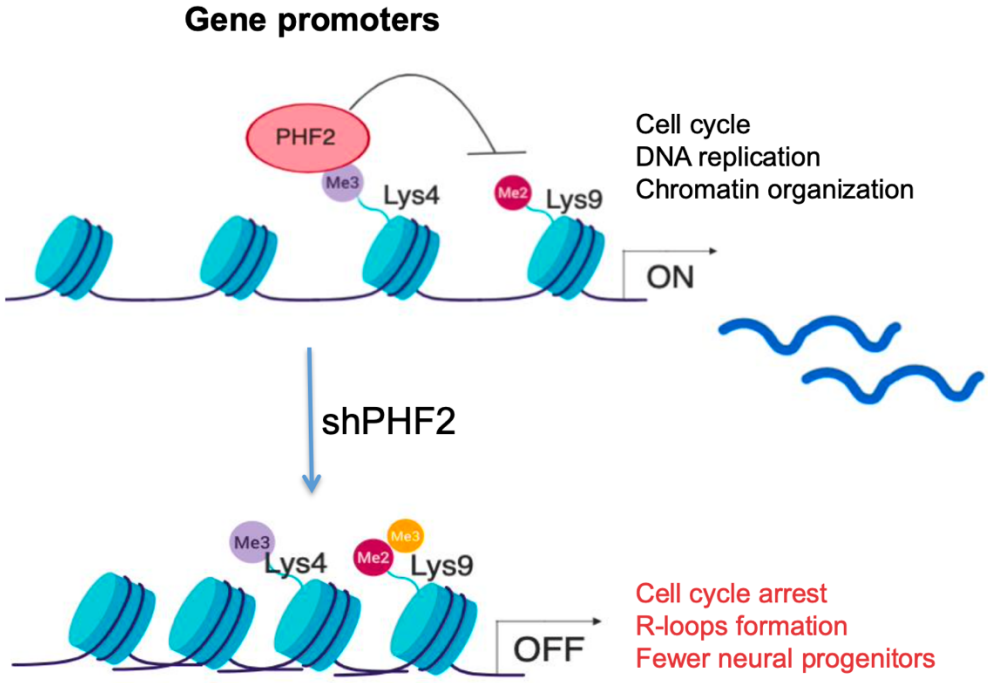


Figure I9. PHF2's role in promoters, maintaining cell cycle progression genes and DNA replication genes by keeping low levels of H3K9me.

In this thesis, I will elaborate on genome stability focusing on heterochromatin, as well as in the cell cycle studying the role of PHF2 in mitosis.

1.2 Chromatin States

Depending on the structure, the position inside the nucleus, the replication timing and the transcriptional activity, chromatin can be found in two main states: euchromatin and heterochromatin. These two states will be detailed in the following subsections.

1.2.1 Euchromatin

Euchromatin is a form of chromatin found in the nucleus of eukaryotic cells, characterized by its less condensed structure compared to heterochromatin. This relaxed configuration makes euchromatin more accessible to the cellular machinery responsible for transcribing DNA into RNA, making it typically associated with regions of the genome that are actively being transcribed. Under a microscope, euchromatin stains lightly with DNA-binding dyes, distinguishing it from the more densely packed and darker-staining heterochromatin. The structure of euchromatin is dynamic, allowing it to change in response to various signals, which is crucial for regulating gene expression during processes like cell differentiation and development. Additionally, euchromatin is usually located towards the inner parts of the nucleus, whereas heterochromatin tends to

be found at the nuclear periphery. Finally, euchromatin is decorated with specific histone modifications, mainly H3 and H4 acetylation as well as methylation of H3K4 (86,87).

1.2.2 Heterochromatin

In 1928, E. Heitz was the first to observe intensely stained and condensed chromatin areas in interphase cells and termed these structures “heterochromatin” (88). Contrary to euchromatin, heterochromatin was initially thought to be totally inactive and full of “useless” tandem repetitive elements. Recently, evidence showed that heterochromatin not only is transcriptionally active, but also is crucial for genome architecture, chromosome function and even has key roles in the pathogenesis of human diseases. Different varieties of heterochromatin are distinguished by their unique combinations of histone post-translational modifications (PTMs), which influence the recruitment of proteins and the folding of chromatin. Sequences embedded in heterochromatin often contain repetitive elements, such as satellite repeats and transposable elements. A crucial function of heterochromatin, which is generally more compact than euchromatin, is to prevent these selfish sequences from causing genetic instability (89) (Figure I10).

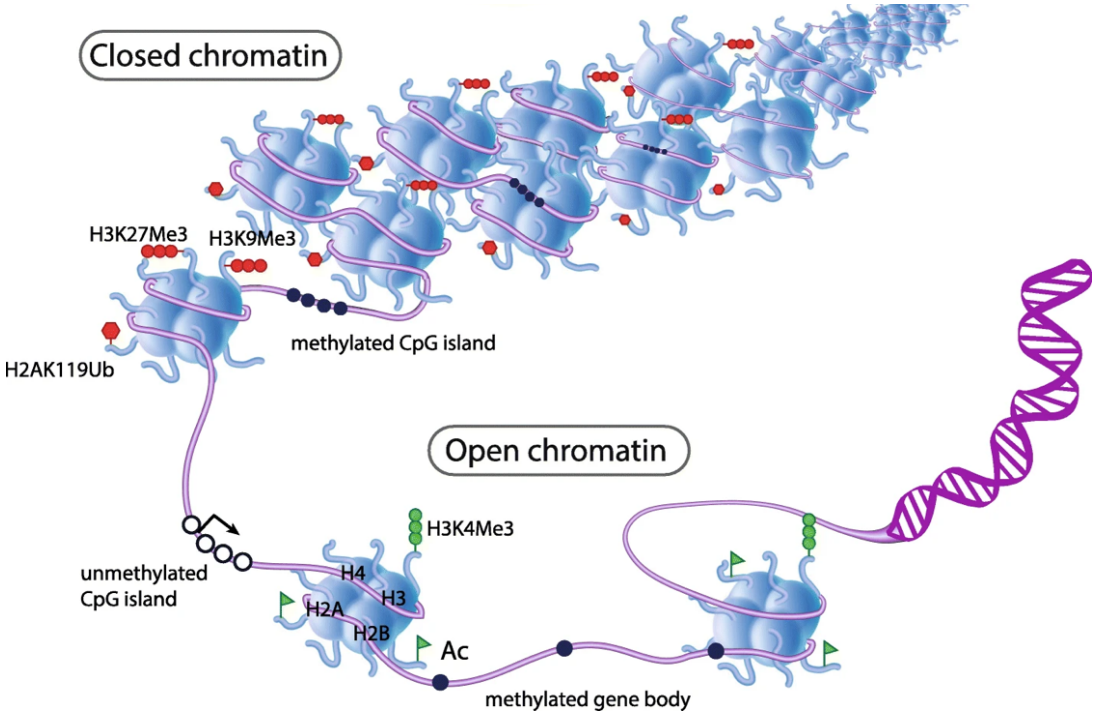


Figure I10. Illustrative representation of chromatin states and associated DNA and histone modifications. Heterochromatin (closed) is showed in the upper part, and euchromatin (open) in the bottom. Adapted from (90)

1.2.2.1 Facultative heterochromatin

Facultative heterochromatin consists of euchromatic regions that are compacted into a heterochromatin-like form in a developmentally regulated manner. Unlike constitutive heterochromatin, facultative heterochromatin is not characterized by repetitive sequences, making it distinct at the DNA sequence level. The main histone modification that characterizes it is the di- or trimethylation of histone H3 at the lysine 27 (H3K27me2/3) and binding of polycomb repressive complex members PRC1 and PRC2

(Figure I11). However, facultative heterochromatin can share many molecular signatures with constitutive heterochromatin, including histone hypoacetylation and H3-K9 methylation. (91,92).

1.2.2.2 Constitutive heterochromatin

Approximately half of the mammalian genome consists of repeat sequences such as satellite DNA, telomeric DNA, and transposable elements. Genetic and biochemical studies have revealed that histone methylation at H3K9 and its recognition by heterochromatin protein 1 (HP1) are fundamental mechanisms for the formation of heterochromatin (89). Constitutive heterochromatin is believed to occur at the same genomic regions in every cell type, typically lacking genes and often viewed as a more static structure compared to facultative heterochromatin. In most eukaryotes, constitutive heterochromatin consistently forms throughout the cell cycle, primarily at pericentromeric regions, telomeres, and gene-poor areas composed of tandem repeats, known as satellite repeats, with sizes ranging from 5 bp to a few hundred bp (93). It is also found at ribosomal regions and telomeres, which consist of the repeated short, conserved DNA motif (TTAGGG) and are enriched in H3K9me3. Telomeres are bound by conserved protein machineries that protect chromosomal ends from being recognized as double-strand breaks. Additionally, constitutive heterochromatin is found at various loci along the chromosome, encompassing repetitive DNA sequences, including simple repeats, DNA transposons, LTR-endogenous retroviral elements, and non-LTR autonomous retrotransposons such as long interspersed elements (LINES) and short interspersed elements (SINES) (94).

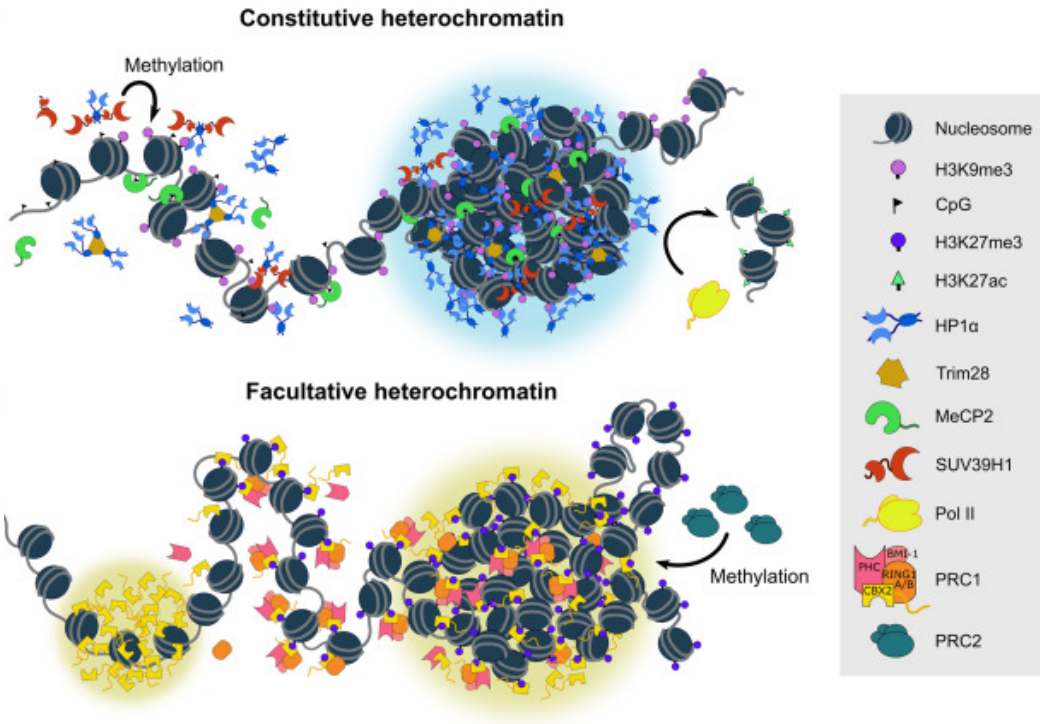


Figure I11. Schematic representation of constitutive and facultative heterochromatin as well as their associated histone modifications, and chromatin remodeling factors. Adapted from (95)

1.2.2.3 Heterochromatin components and mechanisms

Some key components of heterochromatin have been identified, including HP1, which is crucial for heterochromatin formation, maintenance, and transcriptional regulation (96). HP1, the first identified heterochromatic component, also regulates cohesin binding to centromeres, essential for chromosome segregation during mitosis and plays a role in replication (97). HP1 proteins contain a chromodomain reader module and a C-terminal

chromoshadow domain (CSD). The dimerization of the CSD forms a platform for other effector proteins (98,99). This allows HP1 proteins to read H3K9me marks and interact with SUV39H, facilitating the spread of pericentromeric heterochromatin and silencing adjacent gene expression, as shown in fission yeast (100).

H3K9-methyltransferases, which as explained before, add methyl groups to histone H3 at lysine 9, are conserved across species from yeast to humans. Examples include Clr4 in fission yeast, SU(VAR)3-9 in Drosophila melanogaster, and SUV39H1 and SUV39H2 in mice and humans. Suv39h1 and Suv39h2 double knockout mice exhibit a loss of H3K9 trimethylation in pericentromeric regions, leading to increased transcription in these regions (101,102).

Another marker of heterochromatin is DNA methylation, where DNA methyltransferases add methyl groups to cytosines within CpG dinucleotides, stabilizing pericentromeric heterochromatin. In mammals, DNA methylation is carried out by DNA methyltransferase 1 (DNMT1), which maintains methylation patterns during cell proliferation (103), and by DNMT3a and DNMT3b, which initiate de novo methylation during embryonic development (15).

Apart from these core components, transcription factors also play a critical role in heterochromatin formation and silencing, despite being less studied. The absence of promoter elements in these regions suggests that transcription factors may bind outside of a genic context, contributing to heterochromatin structure and function (104,105).

1.2.2.4 Heterochromatin nucleation, spreading and integrity

The establishment of heterochromatin domains is driven by histone modifications, initiated by the recruitment of histone-modifying enzymes to nucleation sites. This nucleation process leads to the deposition of histone H3 lysine-9 methylation (H3K9me) by SUV39, which serves as a foundation for forming densely packed territories of heterochromatin proteins.

Once initiated at a specific site, heterochromatin components possess biochemical properties that facilitate domain expansion, largely independent of DNA sequence. This spreading process relies on reader-writer coupling mechanism. Chromodomains of H3K9me readers like HP1 in mammals bind nucleosomes bearing H3K9me3 and clump them together. At the same time, HP1 is recognized by writer SUV39 and K9me3 by its chromodomain that deposits K9me3 onto the next nucleosome. This mechanism continues facilitating spreading (106,107).

However, spreading also necessitates HP1-dependent recruitment of HDACs that deacetylase nucleosomes ahead for them to be able to compact and be methylated (108). Interconnections among reader, writer, and eraser modules form positive feedback loops extending heterochromatin domains (Figure I12).

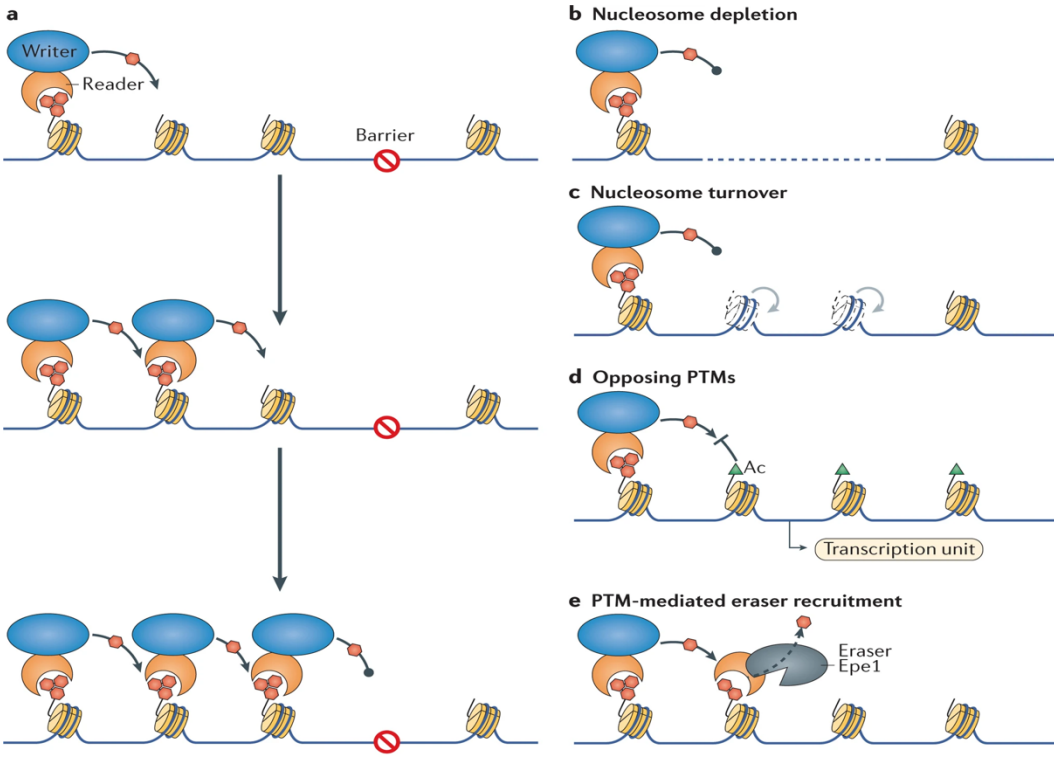


Figure I12. Illustrative representation of a model of heterochromatin expansion (a) and possible boundary elements (b-e). Adapted from (89).

To counteract excessive spreading and prevent erroneous gene silencing, mechanisms create barriers and interrupt lateral heterochromatin spreading. These include generating nucleosome-depleted regions, promoting nucleosome turnover, ncRNA and tRNA binding, and recruiting anti-silencing factors (89,109–111) (see again Figure I12). Additionally, heterochromatin can recruit inhibitors through reader-eraser coupling. For instance, Epe1, a JmjC domain chromatin-associated protein in S. pombe, acts as a putative H3K9 demethylase recruited by the reader Swi6. Epe1 regulates

heterochromatin domain stability, and its absence leads to oscillation of silent chromatin domains, causing expansion into euchromatin or alleviation of silencing. Epe1 functions alongside boundary elements and rapidly removes ectopic H3K9 methylation. Cells lacking both Epe1 and the histone acetyltransferase Mst2 display widespread ectopic heterochromatin assembly, suggesting global regulation of heterochromatin assembly processes (112). Recent studies indicate that a critical density of H3K9me3 and its associated factors is essential for the propagation of heterochromatin domains across multiple generations (89,113,114).

1.2.2.5 Repetitive elements

In both mice and humans, more than half of the genome consists of repetitive elements, primarily including tandem repeats (like satellites) and interspersed transposable elements (TEs). Tandem repeats are sequences of 2–200 bp length, arranged head-to-tail, appearing as microsatellites and minisatellites dispersed across the genome or as large stretches of major satellite sequence around centromeres (115). Transposable elements, on the other hand, can be categorized into RNA and DNA transposons based on their transposition mechanisms (Figure I13). RNA transposons, the predominant fraction in the human genome, include retrotransposons like LINEs and SINEs (including Alu elements), which move via RNA intermediates. DNA transposons move autonomously as DNA segments like Maverik elements (116). Among retrotransposons, LINE-1 is particularly noteworthy, constituting 17% of our genomes and being highly active (117,118). Heterochromatin-generating pathways regulate the transcriptional or posttranscriptional silencing of these elements, crucial for maintaining genome integrity.

Other retrotransposon classes like SVAs and LTR elements, such as ERVs, also impact host gene-regulatory networks by serving as cis-regulatory elements (119). Unlike humans, mouse repetitive elements are not fully mapped to the mouse reference genome yet (120).

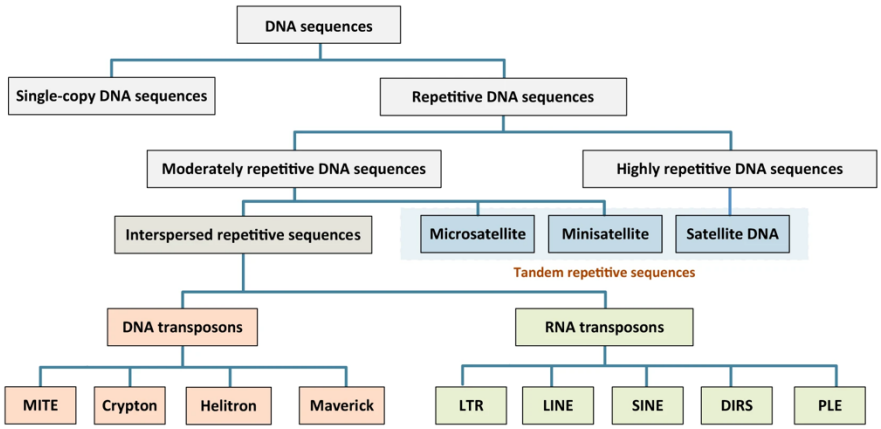


Figure I13. Repetitive elements classification. Adapted from (116)

1.2.2.5.1 Satellite repeats

Since the discovery of highly repetitive tandem DNA in the 1960s, extensive literature has explored the structure, organization, function, and evolution of these sequences. Today, satellite DNA is known to consist of abundant tandem repeats that play important roles in cellular processes, including chromosome segregation, genome organization, and chromosome end protection (121). Most satellite DNA repeat units are either of nucleosomal length or 5–10 bp long, occupying centromeric, pericentromeric, or

telomeric regions. Due to their high repetitiveness, satellite DNA sequences have largely been excluded from genome assemblies. Although few conserved satellite-specific sequence motifs have been identified, features like DNA curvature, dyad symmetries, and inverted repeats are characteristic of various satellite DNAs in several organisms (122).

Mouse chromosomes are telocentric, with centromeric satellite arrays located at or near the start of the chromosome. Despite sharing approximately 70% sequence identity in their genes, mice and humans carry completely different satellite repeats. Mus musculus centromeres consist of tandem arrays of 120-bp minor satellite (MiSat) sequences, which are more homogeneous within a Mus species compared to the α-satellites of primate species (123,124). Minor satellite (MiSat) arrays are flanked by pericentromeric major satellites (MaSat) (115), which consist of 234 bp monomers (125) (Figure I14). MiSats are associated with centromere proteins such as CENP-A, CENP-B, and CENP-C, while MaSats are associated with heterochromatin protein 1α (HP1). Both MaSat and, to a lesser extent, MiSat, have been shown to contain H3K9me3 (126–128).

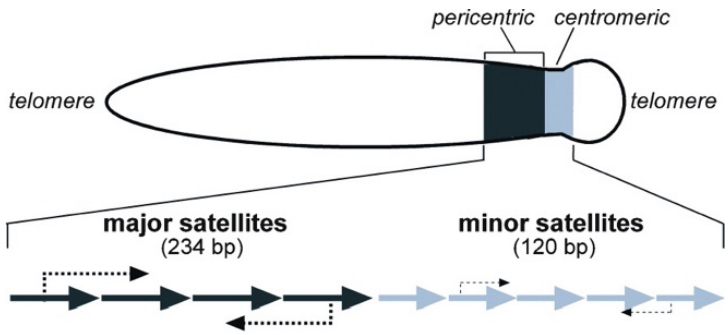


Figure I14. Major and minor satellites represented in the mouse chromosome. Adapted from (125).

1.2.2.6 Pericentromeric heterochromatin and its transcription

As previously mentioned, various heterochromatin-generating pathways mediate the transcriptional or posttranscriptional silencing of these sequences (129). While it has been demonstrated that repetitive elements transcription is crucial for a correct genome homeostasis and cell cycle progression, unscheduled transcription of repetitive elements can compromise genome integrity (130). Studies have shown that unprogrammed transcription of repeats correlates with increased rates of repeat-specific insertions and deletions, copy number variation, R-loops, and heightened sensitivity to replication stress (131). This unprogrammed transcription can lead to collisions between the replication and transcription machineries, causing DNA breaks (132,133). These issues will be further explained in the following subsections.

2. CELL CYCLE AND EPIGENETIC DYNAMICS

Since the earliest observations of cells in mitosis, the close relationship between the cell cycle and nuclear chromatin architecture has been evident. During the cell cycle, the nuclear envelope and chromatin undergo significant assembly and disassembly (Figure I14). The regulation of histone biogenesis and chromatin modification is tightly controlled in a cell cycle-dependent manner, both transcriptionally and post-transcriptionally. Chromatin binding proteins and chromatin modifications, in turn, influence the expression of key cell cycle regulators, the accessibility of DNA replication origins, DNA repair processes, and cell fate determination (134). Previous work from the laboratory showed that PHF2 depletion led to cell cycle arrest (5). In this doctoral thesis, further research will be conducted on this matter, particularly in PHF2's role in mitosis. The following subsections will provide a general cell cycle overview and how chromatin and the epigenetic modifications behave during this process.

2.1 Interphase

The interphase is a phase in the cell cycle consisting of G1, S, and G2 stages, where the cell grows, repairs DNA, and replicates its DNA in preparation for division. However, mature, differentiated cells and tissues do not remain cycling, but instead they enter G0 phase and stop dividing (135).

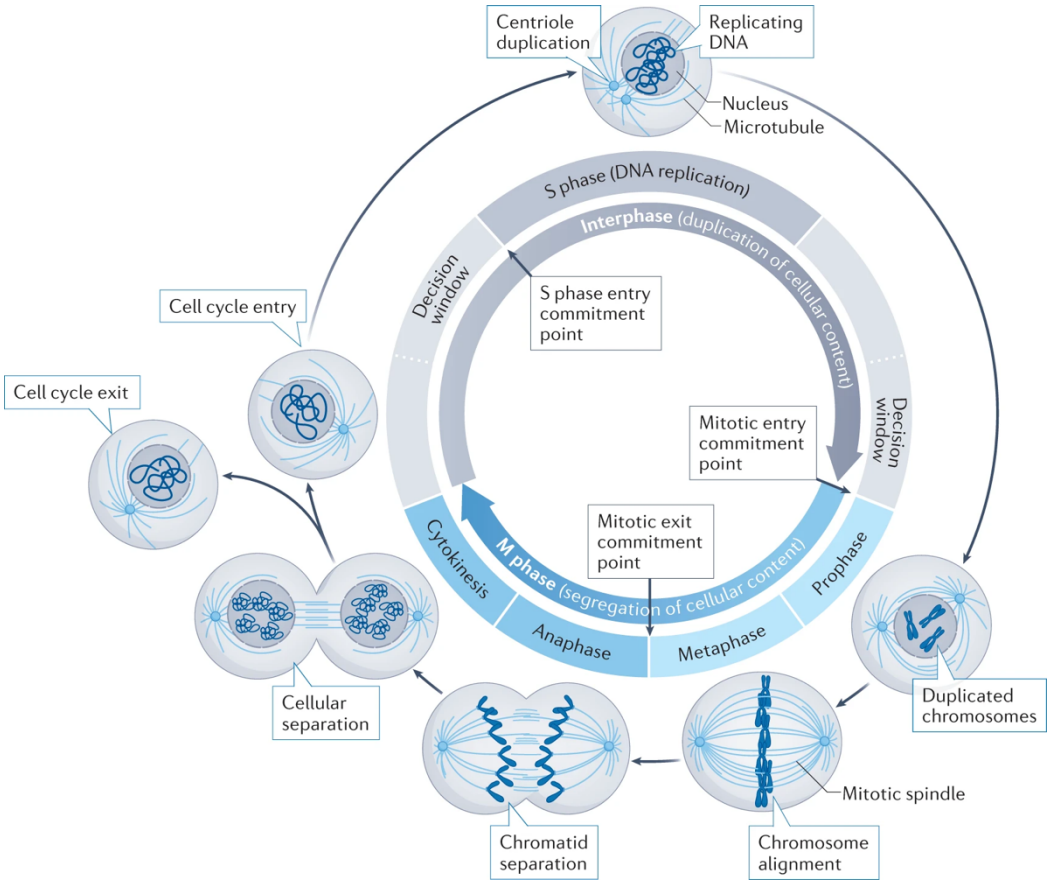


Figure I14. Schematic overview of cell cycle with the mitotic cell division explained detail. Adapted from (136).

2.1.1 Epigenetic modifications in cell cycle

Pluripotent cells exhibit a distinct open chromatin configuration, which becomes progressively restricted during development. This loss of chromatin accessibility is accompanied by increased DNA methylation and the deposition (or redistribution) of various histone marks, which influence chromatin interactions. However, most histone

modifications remain relatively stable during cell cycle progression, except some like H4K20me. Topologically associated domains (TADs) and Lamina associated domains (LADs) are largely stable during stem cell cycle, but subdomain interactions might change. The monomethylation (H3K4me1) or dimethylation (H3K4me2) marks on H3K4 have broader patterns compared with H3K4me3 and are mainly detected at intergenic sites, while trimethylation is present at promoters. H3K27 acetylation is deposited at active regulatory elements (predominantly promoters and enhancers). H3K27me3 is associated with gene silencing and is mainly deposited at CpG-rich promoters but can also be detected at intergenic regions. H3K9 dimethylation is associated with repression of gene expression and is widely distributed on chromatin in large domains that can extend to 4 Mb in size (137).

2.1.2 S phase – DNA replication

We are just beginning to comprehend how epigenetic regulators coordinate with cell cycle transitions. Current studies support a model in which cis-acting epigenetic signals present in G1 must be duplicated and incorporated into the new chromosome during the S phase. As the replication fork progresses and DNA is replicated, the resulting DNA molecule is hemi-methylated: composed of a parental strand in the original methylation state and a newly synthesized unmethylated strand. The preservation of methylation in both DNA molecules relies on the activity of DNMT1, which has a high affinity for hemi-methylated DNA and restores CpG methylation on the unmethylated strand. During DNA replication, the disassembly of nucleosomes leads to the release of stable H3-H4 tetramers that do not mix with newly synthesized histones, unlike the more dynamic

H2A-H2B dimers that can mix with new histones and partner with old and new H3-H4 tetramers. To preserve transcriptional memory, the modified parental histones that are inherited must remain bound to the same genomic positions after the replication fork passes. Parental histones with different types of post-translational modifications (PTMs), including H3K9me3 and H3K27me3, can be incorporated into newly replicated DNA and reproduce the same genome-wide patterns observed before DNA replication-coupled dissociation (138,139).

2.2 Mitosis

When the moment of cell division arrives, chromosomes undergo a 10.000-fold condensation, and the nuclear envelope disassembles (see again Figure I1) (140). This process disrupts higher-order chromatin organization, leads to the dissociation of many transcription factors, inactivates RNA polymerase II through phosphorylation, and results in a general decrease in gene transcription (141).

2.2.1 Chromosome compaction

Mitotic chromosomes exhibit widespread alterations in chromatin modifications, including reduced acetylation of histone tails, which facilitates chromatin compaction, and the accumulation of mitosis-regulatory chromatin marks, such as phosphorylation of H3 on serine-10 and serine-28. Despite these changes, it is generally assumed that once epigenetic marks are duplicated during S phase, their genome-wide distribution in the G2 phase remains unaltered on mitotic chromosomes and is inherited by cells in the G1 phase

without interfering with the regulatory processes required for mitosis completion (138,139).

2.2.2 Chromosome segregation

The centromere is crucial for chromosome segregation during mitosis as it serves as the assembly site for the mitotic kinetochore. Epigenetically defined by the presence of the histone H3 variant CENP-A and H3K9me3 enrichment, the centromere is a large chromatin-containing protein complex. This complex forms the foundation for the mitotic kinetochore, a megadalton protein assembly that binds spindle microtubules to facilitate chromatid segregation during anaphase. Beyond microtubule attachment, kinetochores also activate the mitotic checkpoint, which prevents the onset of anaphase in the presence of unattached kinetochores. Without the centromere, kinetochores would not form, and cells would be unable to segregate their chromosomes, underscoring the centromere's critical role in chromosome segregation and mitotic regulation (142).

PHF2 is also called Centromeric protein 35 (CENP35). In highly purified chromosomes it has been identified as one of the components of the centromere by multiclassifier combinatorial proteomics (143) (Figure I15).

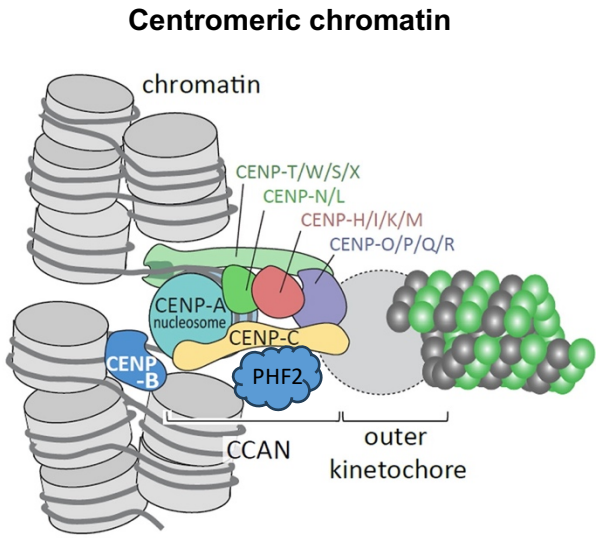


Figure I15. Schematic representation of the centromere and the kinetochore structures. Adapted from (144)

2.2.3 Transcription reactivation after mitosis

Proliferating cells experience a global reduction in transcription during mitosis yet maintain their cell identity and propagate regulatory information from mother to daughter cells (141). The restoration of the programmed gene expression pattern after mitosis may involve marking genes through chromatin remodeling or modifying activities (Figure I16). This process, known as molecular or mitotic bookmarking, involves the binding of transcription factors or retention of histone modifications during mitosis to mark specific chromatin regions and aid their reactivation in the subsequent cell cycle (145,146). A significant fraction of transcription factors remains bound to chromatin during mitosis, exhibiting faster interactions with mitotic chromatin compared to interphase, as observed

by single molecule tracking, which indicates that chromosomes are not as compacted as previously thought. Indeed, chromatin accessibility and the nucleosome landscape remain unchanged in bookmarked regions during mitosis. Proteins such as C/EBP, HSF1, TBP, GATA1, FOXA, SOX, and ESRRβ are confirmed bookmarking factors, with some identified as PHF2 targets or interactors (147).

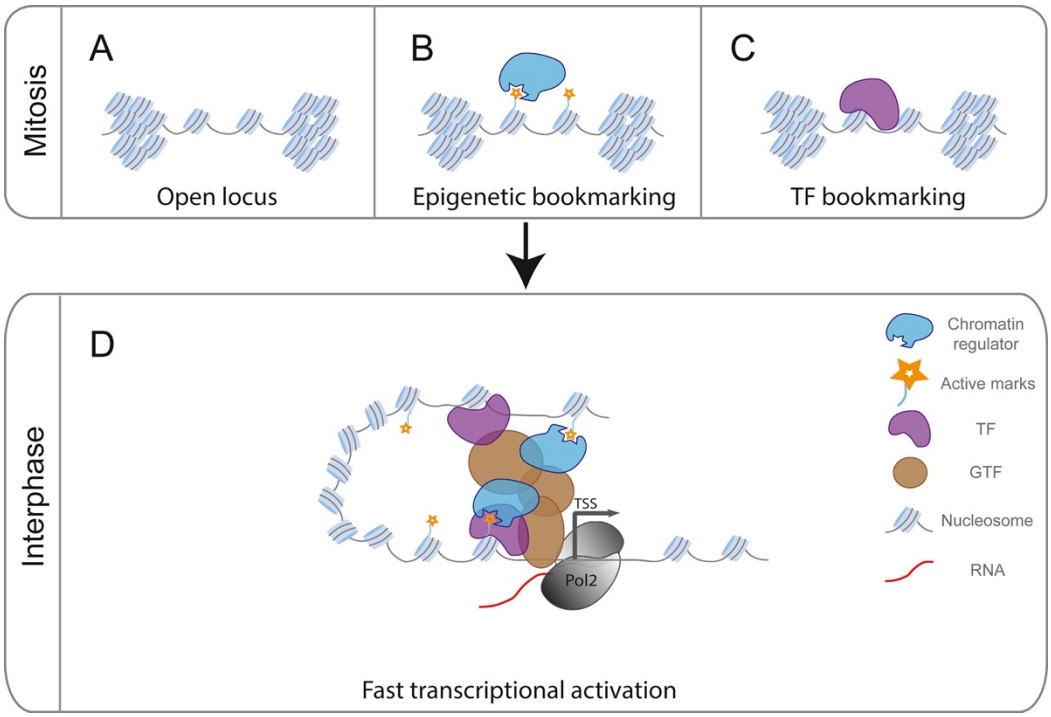


Figure I16. Schematic summary of the potential mechanisms supporting memory of active genes through mitosis (A-C) and how they would interact in transcription reactivation (D). Adapted from (148).

2.3 Genome instability

Genetic instability, manifesting as mutations and chromosome rearrangements, is typically associated with pathological disorders but is also crucial for evolution and generating genetic diversity. Even before the discovery of the double helix in 1953, it was known that exogenous agents like X-rays, ultraviolet (UV) light, and various chemicals could induce genetic changes that promote cancer. In humans, genomic instability is often linked to premature aging, predisposition to various cancers, and inherited diseases (149,150). Cells employ multiple mechanisms to protect the genome from the mutagenic effects of genotoxic agents and ensure accurate chromosome duplication and transmission to offspring. In eukaryotes, cell-cycle checkpoints coordinate DNA synthesis and repair with cell division. Genome instability primarily arises from sporadic replication or repair errors but can also occur in response to developmental or environmental signals, exposure to external genotoxic agents, single stranded DNA (ssDNA), double strand breaks (DSBs) or rates of repetitive elements transposition (151–153).

2.3.1 Causes of genome instability

As mentioned before, genome instability can result from failures at various stages of the DNA cycle, from replication to segregation. Among the main causes of genome instability are replication dysfunction, failures in post-replicative repair, and unscheduled transcription of repetitive elements, which can compromise genome integrity. Additionally, defects in homologous recombination, site-specific hotspots of genome

instability, and conflicts between the transcription and replication machineries can lead to DNA damage. Errors during chromosome segregation in mitosis, such as anaphase bridges and chromosome breakage at M-phase, further contribute to genomic instability. These factors collectively shape the complex landscape of genomic instability, affecting cellular function and organismal health (153) .

2.3.1.1 Replication dysfunction

DNA replication, out of most cellular processes, is the most vulnerable one during cell-cycle progression and is tightly controlled at various stages, from initiation to termination. Genome duplication is regulated once per cell division during the loading and activation of the replicative DNA helicase at replication origins, as well as the preservation of the replication fork by TIMELESS in the fork protection complex (154,155). After replication has initiated, replication forks may experience transient pausing or longer delays, referred to as fork stalling. When a single-stranded DNA (ssDNA) nick is encountered in the leading strand template, it leads to the immediate collapse of the replication fork (RF), resulting in a one-ended DSB (153,156) . This DSB can then facilitate the restart of replication via break-induced replication (BIR). DNA adducts or tightly bound proteins, such as components of the transcription machinery (represented by a red rectangle), can obstruct RF progression or inhibit leading-strand synthesis. If leading-strand synthesis is hindered, the synthesis of the leading and lagging strands becomes uncoupled. This can result in fork reversal, forming a Holliday junction (HJ) or a 'chicken-foot' structure (157). Restart of replication can occur through HJ-cleavage followed by BIR, or the RF may directly restart through HJ reversal after lesion

bypass via template switching or lesion repair. Lesions that block synthesis of only one DNA strand without impeding fork progression create different scenarios based on which strand is affected. A lesion blocking lagging-strand synthesis does not cause RF stalling but creates a ssDNA gap between two flanking Okazaki fragments. Conversely, a lesion blocking leading-strand synthesis can be bypassed by the RF, which restarts downstream of the lesion, leaving a ssDNA gap behind. In both situations, these ssDNA gaps can be repaired either by error-prone translesion synthesis (TLS) or by error-free homologous recombination (HR) using the sister chromatid as a template (template switching). It remains unclear whether template switching necessitates the conversion of the ssDNA gap into a DSB before repair (150) (Figure I17).

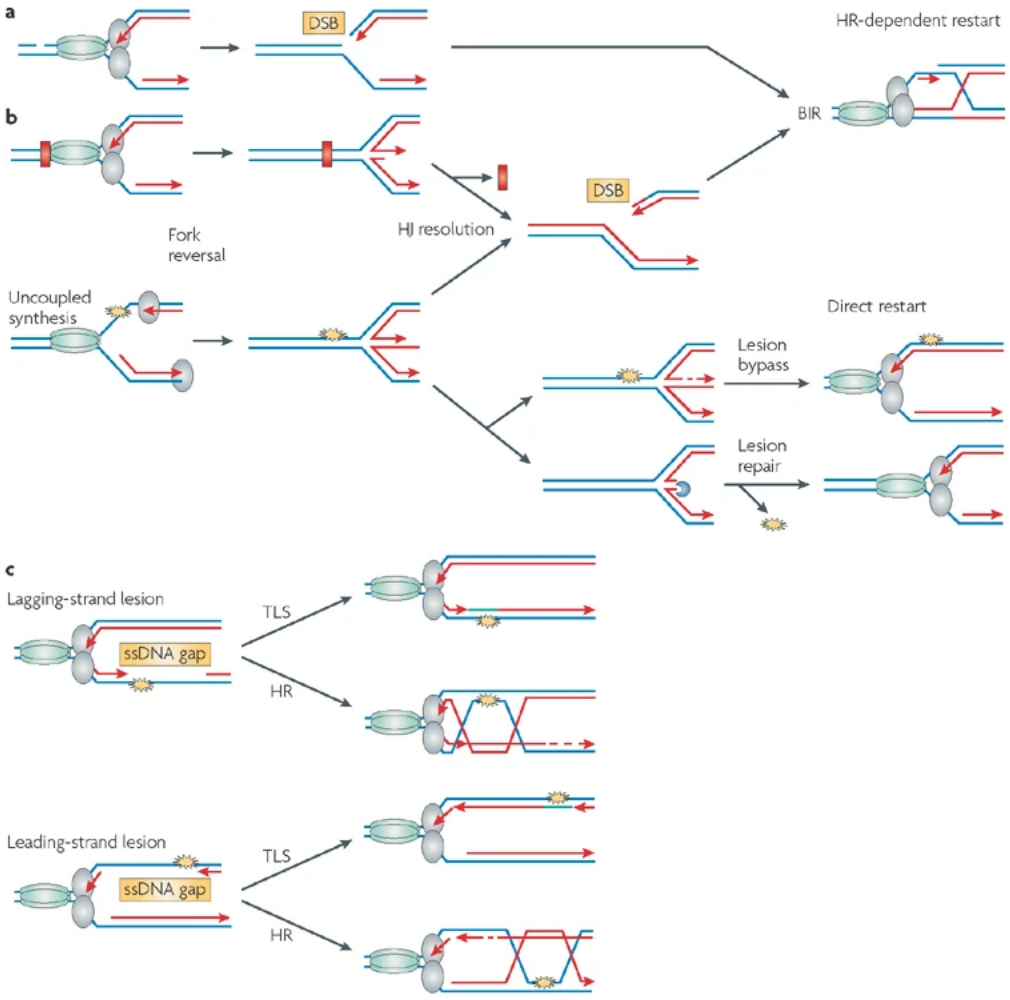


Figure I17. Illustration depicting how double stranded DNA breaks (a,b) and ssDNA gaps (c) can impact transcription and subsequent DNA repair. Adapted from (150).

2.3.1.2 Transcription-replication collisions

DNA replication and transcription use the same DNA template, which can lead to collisions between these two processes (158,159). These collisions can occur either head-

on or co-directionally, depending on the direction of movement of both the replication and transcription complexes. Head-on collisions, which require additional factors to resolve, are associated with greater genome damage. In contrast, co-directional collisions are typically managed by the replisome itself, which can use the mRNA transcript as a primer to continue replication (160). Such replication-transcription collisions can result in replication fork arrest, premature transcription termination, DNA breaks, and recombination intermediates, all of which threaten genome integrity. When replication forks collide with RNA polymerase, they may occasionally collapse. Both elongating and paused/stalled RNA polymerases can serve as physical barriers that hinder replication fork progression. Additionally, the increased positive super-helical density generated during head-on collisions between the replication fork and RNA polymerase II can also contribute to fork stalling and collapse (153). Bacterial studies show that these collisions not only cause replication fork stalling but also DNA breaks, leading to two types of mutation signatures: duplications and deletions within the transcriptional unit, and promoter-localized base substitutions (161).

2.3.1.3 Co-transcriptional R-loops

RNA:DNA hybrid structures known as R-loops were initially thought to be rare byproducts of transcription with little cellular impact. The R-loop structure was first characterized in 1976 and was considered highly stable, as RNA/DNA associations are thermodynamically more stable than DNA/DNA interactions (162,163). This stability is attributed to the structure of the RNA/DNA hybrid. Since then, R-loops have become an increasingly prominent area of research, highlighting their potential role as regulators of

gene expression. R-loops are three-stranded nucleic acid structures consisting of a nascent RNA strand hybridized with the DNA template, leaving the non-template DNA single-stranded. The RNA strand is generated by RNA polymerase II transcribing a C-rich DNA template to produce a G-rich transcript. The most accepted mechanism for R-loop formation is the "thread-back" model, where the nascent RNA transcript invades the DNA duplex as soon as it exits the RNA polymerase. One major cause of R-loop accumulation is replication fork collapse or stalling, although there are multiple sources of R-loop formation (Figure I18) (164,165).

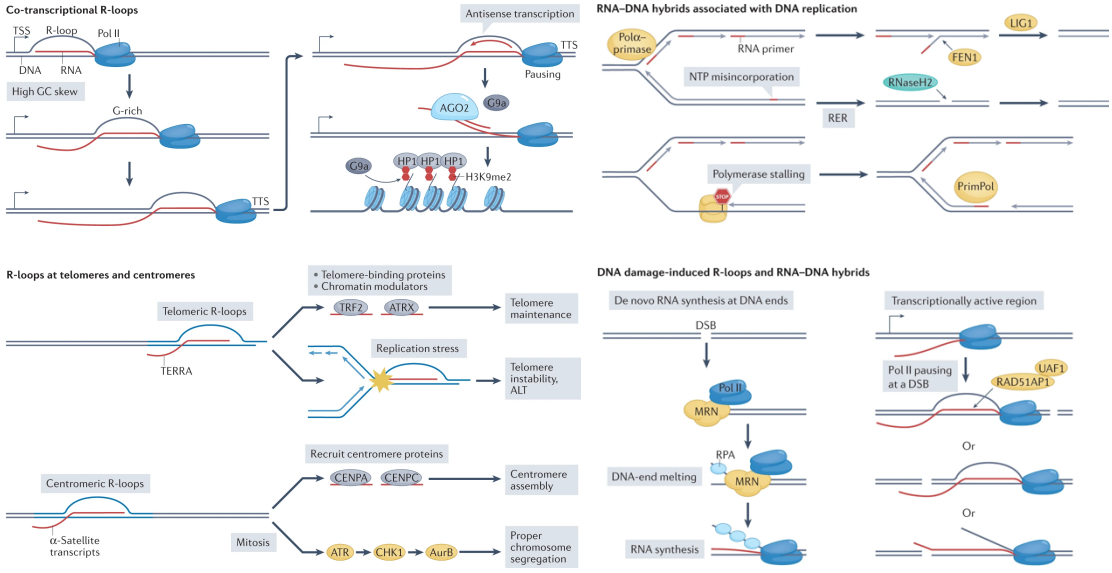


Figure I18. Schematic representation of the multiple sources of R-loop formations. Adapted from (165)

R-loops form naturally as key intermediates in specific cellular processes, such as E. coli plasmid replication, mitochondrial DNA replication, and immunoglobulin class switching. However, recent studies from our lab suggest that R-loops were accumulated when PHF2 was depleted, posing a significant threat to gene expression and genome integrity (5). Over the past decade, research has identified these structures as powerful regulators of gene expression, with evidence showing that programmed R-loop formation controls numerous biological processes (166). A recent study proposes that the primary hindrance to replication fork advancement is the elongating RNA Polymerase II entangled in an R-loop, rather than the resultant DNA-RNA hybrids. When the regulation of R-loops is disrupted, pathological R-loops accumulate, generating genome instability and contributing to diseases such as neurodegeneration and cancer (165). Additionally, the accumulation or depositing of the H3K9me2 repressive mark occurs, and heterochromatin protein 1γ (HP1γ) is recruited, further solidifying Pol II pausing prior to efficient transcriptional termination (167,168).

2.3.2 Consequences of genome instability

The previously described sources of genome instability like R-loops, have the potential to cause DNA damage by ssDNA formed as a result of RNA/DNA hybridization. This ssDNA, being inherently less stable, becomes susceptible to lesions and can undergo transcription-associated mutagenesis or recombination. One possible scenario for how R-loops induce DNA damage involves the unpaired DNA strand resulting from R-loop formation, which is more prone to DNA damage such as spontaneous deamination of dC to dU. This process can lead to DSBs and recombination events. Other potential scenarios

include the recognition of the R-loop structure by proteins, which could initiate mutagenesis. Additionally, transcriptional R-loops may induce genomic instability by interfering with DNA replication. This instability can manifest in various ways, including damage to microsatellite integrity, increased frequencies of base pair mutations, chromosome instability, altered or bypassed cell cycle checkpoints, and ultimately, cancer (169,170).

3. THE CORTEX DEVELOPMENT

In this section, I will outline the key aspects of neural cortex development to effectively introduce the in vitro model of mouse neural stem cells (NSCs) utilized in this research. The mammalian nervous system stands as the most intricate system among living organisms. The intellectual capacity of the mammalian brain is believed to be closely linked to its size, predominantly influenced by the size of the cerebral cortex. This region of the brain, responsible for cognitive function, consciousness, and sensory perception, is particularly prominent in humans due to its size and complex folding (171,172). Although the cortex is a conserved structure across all vertebrates, it is thought to be central to what makes the human species unique. Comprising six distinct layers of various neuronal subtypes and glial cells, the cortex forms a highly organized structure, with each area dedicated to specific motor, sensory, and cognitive functions. It has been described that histone demethylase PHF2 is necessary for memory consolidation via the CREB signaling pathway (173). Disruptions in the proliferation or survival of neural progenitors

can lead to abnormal brain sizes, resulting in conditions such as megacephaly, microcephaly, or dysplasia, each associated with distinct pathological outcomes (174).

3.1 Mouse neural development – Corticogenesis

Mouse corticogenesis begins at embryonic day 11 (E11). At this stage, cortical progenitors populate the proliferative zones of the dorsal telencephalon: the ventricular zone (VZ) and the subventricular zone (SVZ). Initially forming a polarized neuroepithelium, these progenitors undergo successive maturation stages to generate all cortical cell types. This process in mammals is driven by intrinsic and extrinsic signals that regulate the spatial and temporal generation of neurons (175). The most pertinent neurodevelopmental process for this work is neurogenesis, which starts when neuroepithelial progenitors (NE), acting as NSCs, begin proliferating to expand the pool of cortical progenitors. NE cells give rise to radial glial cells (RG), the primary cortical progenitor subtype, which will ultimately produce all cortical neurons. RG cells divide asymmetrically, creating another identical cell and a more differentiated intermediate progenitor cell. These progenitors eventually undergo symmetric terminal divisions to produce two differentiating daughter cells, gradually slowing and stopping progenitor expansion (176). Neurons are initially generated in the VZ, the deeper part of the developing brain, and then migrate to the surface to form the six-layered cortical structure. The final step in neurogenesis involves neuron translocation to the marginal zone (MZ), where they mature and develop axons and dendrites (Figure I19) (177).

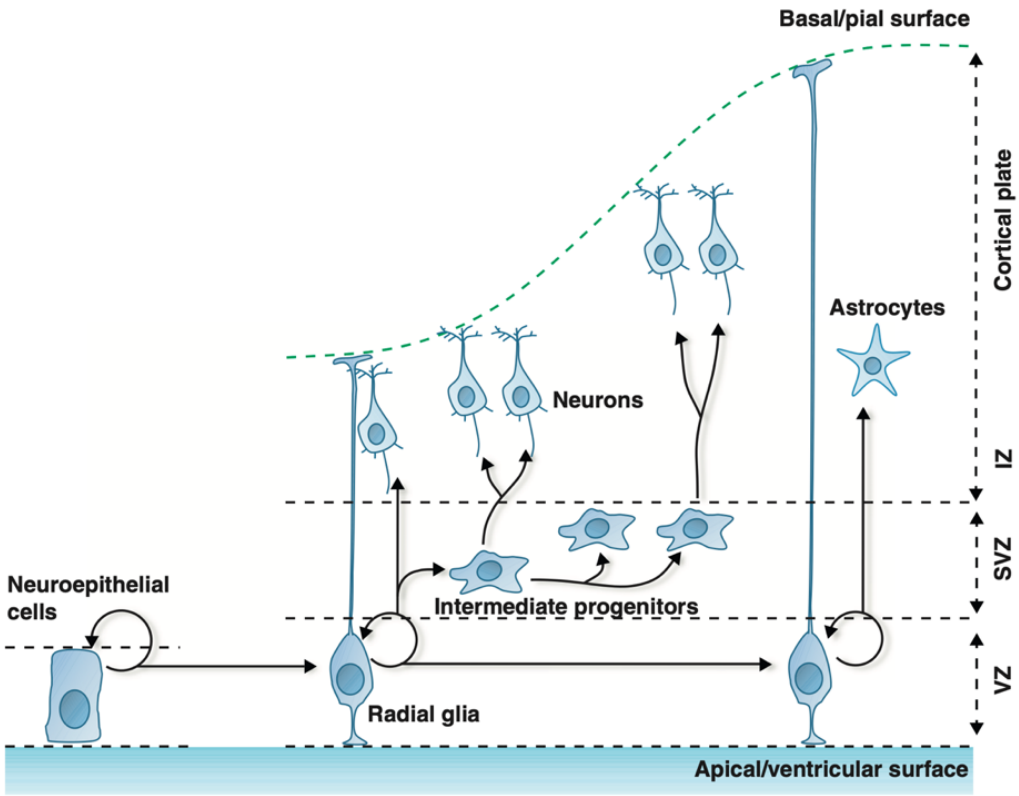


Figure I19. Illustrative explanation of mouse neural development. Adapted from (176).

The differentiation of cortical progenitors into neurons is influenced by various signaling pathways and intrinsic factors, including proneural transcription factors (TFs), epigenetic regulators, cell polarity, and mitotic molecules. Key signaling pathways include the fibroblast growth factor (FGF) pathway (178), which promotes NSC proliferation both in vitro and in vivo; the Notch pathway, essential for maintaining NSC self-renewal capacity (179); the Sonic Hedgehog (Shh) pathway, which controls the cell cycle of RG (180); and the canonical Wnt pathway, which induces the proliferation of VZ neural precursors and determines neuron fate (181). Additionally, the BMP and TGFβ pathways

play roles in forebrain development, with TGFβ signaling promoting cell cycle exit of VZ progenitors and neuronal cell fate in cortical and hippocampal progenitors (182–184). Intrinsic factors, primarily operating in a cell-autonomous manner, are also crucial. Essential intrinsic factors include proneural TFs such as ASCL1, neurogenins (NEUROG1, 2, and 3), and neurogenic and oligodendrocyte differentiation proteins (NEUROD1, 2, 4, and 6, and OLIG1, 2, and 3) (185). These TFs are necessary and sufficient to initiate neuronal differentiation programs, with their expression being restricted to specific CNS cell populations (186). Additionally, corticogenesis is influenced by cell polarity and the various modes of cell division (187–189) .

3.2 Epigenetic mechanisms in neural development

Epigenetic modifications play a crucial role in guiding NSC differentiation and lineage commitment by coordinating the spatial and temporal expression of transcriptional regulators. These modifications are primarily produced by DNA methyltransferases (DNMTs) and histone methyltransferases (KMTs). For example, DNMT1 and DNMT3 maintain DNA methylation, regulating the division of neural progenitor cells (190). It has been thoroughly demonstrated that in mice, 75% of DNA methylation occurs at CpG sites, with the remainder methylation at non-CpG (CpH) sites. CpH methylation occurs de novo during neuronal maturation in both mice and humans (191). Knockdown of DNMT3A in neurons causes loss of methylation at many CpH sites, but not at CpG sites, suggesting that neuronal CpH methylation is more dynamic and actively maintained by DNMT3A (192). SETDB1 (also known as ESET), which methylates histone H3 at lysine 9 (H3K9), is prominently expressed during early mouse brain development but its

expression diminishes over time. The deletion of SETDB1 during embryonic neurogenesis results in reduced H3K9 trimethylation and dysregulation of gene expression, including genes like Sox9 and IAP (193).

Histone modifications also play a significant role in neurogenesis. In embryonic stem cells (ESCs), housekeeping genes are marked for transcription initiation by H3K4me3, while developmental genes exhibit both active (H3K4me3) and repressive (H3K27me3) marks, a state described as "bivalent" or "poised" (166). The regulation of developmental and pluripotency genes is closely linked to the activity of Polycomb group proteins (PcG), which are responsible for H3K27me3 deposition at genes involved in cell lineage commitment in mammalian development (194). PcG proteins also control the transition from neurogenic to astrogenic phases by modulating the expression of *Ngn1*(192). Moreover, deletion of Enhancer of Zeste homolog 2 (EZH2), the enzyme responsible for H3K27 methylation, in NSCs results in de-repression of a large set of neuronal genes, and impaired neuronal differentiation (195,196)During cell differentiation, repressive histone marks (H3K9me3 and H3K27me3) increase and expand from covering around 4% of genes in ESCs to 12%-16% in fully differentiated cells (166). In contrast, H3K9 methylation mediated by Suv39h1 and Suv39h2 controls NPC differentiation in the adult hippocampus. The presence of H3K9me3 foci is significantly higher in neural progenitor cells compared to neural stem cells (NSCs) and decreases as neurons mature. Inhibition of Suv39h1/2 in adult hippocampal progenitors reduces neuronal differentiation while promoting proliferation (196). It has also been demonstrated that the histone demethylase PHF8 regulates astrocyte differentiation and function during neurogenesis (197).

3.3 Neural stem cells as an ex vivo model of this study

In this thesis, NSCs from the embryonic cortex of mice at stage E12.5 were used as a model system. NSCs are multipotent cells in the nervous system capable of self-renewal, unlimited proliferation, and differentiation into neurons, astrocytes, and oligodendrocytes. This in vitro model mirrors the morphology and markers observed in radial glia of the mouse embryo, including BLBP, RC2, GLAST, PAX6, and NESTIN (5,49). NSCs are present in the developing cortex and specific regions of the adult mammalian CNS (198). In vivo, they reside in a microenvironment known as the niche, which provides essential factors for their self-renewal and differentiation. Ex vivo, the absence of this niche necessitates the addition of mitogens such as EGF and FGF to maintain their multipotency and limit differentiation (199). However, these mitogens can alter the cells' transcriptional programs. Additionally, in vitro NSC cultures face challenges such as loss of positional identity and limited capacity to differentiate into various neuronal subtypes.

The complexity and heterogeneity of events during vertebrate neurogenesis make studying transcriptional repetitive elements regulation or dynamic processes like mitosis difficult in the developing embryo. Using adherent cultures of NSCs helps to avoid the heterogeneity of in vivo systems, despite the inherent limitations of the ex vivo model.

OBJECTIVES

Taking into consideration the current knowledge of the field described in the introduction chapter, and the previous work performed at Dr. Martínez-Balbás's lab that characterized the *in vitro* and *in vivo* role of PHF2 in promoting cell cycle progression and neural progenitor proliferation (5), the following objectives have been set for this doctoral thesis:

1. To elucidate the role of PHF2 in heterochromatin

In order to do that, we sought to study:

1.1. Interaction of PHF2 with heterochromatin components

1.2. The contribution of PHF2 to heterochromatin stability and genome integrity

2. To explore the contribution of PHF2 during mitosis

Characterizing:

2.1. PHF2 interaction with mitotic chromatin

2.2. Functional consequences in mitosis of PHF2 depletion

MATERIALS AND METHODS

1. MATERIALS

1.1 Plasmids

All plasmids used and generated during this doctoral thesis, as well as their origin and experiment in which they were used are listed on the following table (Table M1).

Plasmid	***Origin***	***Experiment***
pCMV-VSVG	Dr. Timothy Thomson	Virus production vector
pCMV-GAL-POL	Dr. Timothy Thomson	Virus production vector
pLKO-shCT (CAACAAGATGAAGAGCACC)	Sigma	Virus production vector
pLKO-shPHF2 (CGTGGCTATTAAAGTGTTCTA)	Sigma	Virus production vector
p3xFLAG-PHF2	Dr. Jiemin Wong	Overexpression vector
p3xFLAG-PHF2ΔPHD	Samuel Aguirre	Overexpression vector
p3xFLAG-PHF2ΔCharged	Samuel Aguirre	Overexpression vector
p3xFLAG-PHF2 HID>AIA	Dr Stella Pappa	Overexpression vector
pmEGFP	Dr Carme Gallego	Overexpression vector
pmEGFP-PHF2	Mónica Barrios	Overexpression vector
pLIV-PHF2	Samuel Aguirre	Virus production vector
pLIV-PHF2ΔPHD	Samuel Aguirre	Virus production vector
pLIV-PHF2ΔCharged	Samuel Aguirre	Virus production vector

pInducer-PHF2	Claudia Navarro	Virus production vector
pInducer-PHF2 HID>AIA	Claudia Navarro	Virus production vector
pCDNA-EGFP-PHF2	Samuel Aguirre	Overexpression vector
p3xFlag-MaSat-NLS-ZF18-mVenusN	Dr. Albert Jeltsch	Complementation sensor
p3xFlag-mVenusC-HP1b-chromodomain	Dr. Albert Jeltsch	Complementation sensor

Table M1. List of plasmids, origin, and experiment in which they were used during this doctoral thesis. The sequence in brackets for the pLKO vectors correspond to their respective target.

1.2 Antibodies

A comprehensive list of the primary and secondary antibodies used in this study, including their working conditions, provider and reference is provided below (Table M2).

Antibody target	***Provider and reference***	***Dilution for every application***
PHF2	Cell Signaling, D45A2	Immunofluorescence 1:500; Western Blot 1:1000; CoIP 1:50, ChIP 1:50
H3K9me2	Abcam, ab1220	Immunofluorescence 1:250
H3K9me3	Abcam, ab8898	Immunofluorescence 1:250; ChIP 1:50
H3K4me3	Abcam, ab8580	ChIP 1:50
H3K27me3	Millipore, #07449	Immunofluorescence 1:250

HP1α	Active motif, 39295	ChIP 1:50
γ-phospho-H2A.X (S139)	Merck, 05-636	Immunofluorescence 1:250
PHF8	Life tech, A301-772A	Immunofluorescence 1:250
KIAA1718	Merck, ABE498	Immunofluorescence 1:250
β-Tubulin	Millipore, MAB3408	Western Blot 1:5000
Unspecific IgGs	Diagenode, C15410206	ChIP, CoIP, same dilution than specific IgGs
Anti-Rabbit Fluor 488	Abcam, ab150073	Immunofluorescence 1:1000
Anti-Rabbit Fluor 568	Abcam, ab175470	Immunofluorescence 1:1000
Anti-Rabbit Fluor 647	Abcam, ab150075	Immunofluorescence 1:1000
Anti-Mouse Fluor 488	Invitrogen, A21202	Immunofluorescence 1:1000
Anti-Mouse Fluor 568	Invitrogen, A11004	Immunofluorescence 1:1000
Anti-Mouse Fluor 633	Invitrogen, A21052	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 doctoral thesis. Providers, references and dilution of usage are indicated.

1.3 Primers

During this thesis, I have used primers for several experimental procedures such as retrotranscription followed by quantitative polymerase chain reaction (RT-qPCRs) for relative gene expression quantification or chromatin immunoprecipitation (ChIP) quantification, and conventional polymerase chain reaction (PCR) for the cloning of several vectors. All of them are listed in Table M3.

Region	*Sense Primer*	*Antisense Primer*
Phf2	CCCTGGAGTCTTTCTCACAC	CCGTTCCGATGGATCTTCAAG
MajorSat	TGGCGAGAAAACTGAAAATCACG	TCTTGCCATATTCCACGTCCTAC
MinorSat	AATGATAAAAACCACACTGTAGAACAT	ATGTTTCTCATTGTAACTCATTGATATAC
L1	TGGCTTGTGCTGTAAGATCG	TCTGTTGGTGGTCTTTTTGTC
SINE	GAGCACACCCATGCACATAC	AAAGGCATGCACCTCTACCACC
IAP1	CGCTCCGGTAGAATACTTAC	TGCCATGCCGGCGAGCCTGT
Gda	ACGACGGCACCAAAGAATAC	TTGCCAGACAAGCAATCAAC
Gapdh	ATGTTCGTCATGGGTGTG	CCTTCCACGATACCAAAGTTG
MajorSat (ChIP)	TGGAATATGGCGAGAAAACTG	AGGTCCTTCAGTGGGCATTT
MinorSat (ChIP)	AATGATAAAAACAGTGATTTCG	ATGTTTCTCATTGTAACTCATTGATATAC
Col2a1 (ChIP)	GGTCTCACCGCTCCCTCAT	GCGACCGGGAGCATATAACT
Brca2 (ChIP)	CGGAAACAGACACACACAC	GCAGCGGTAGCTGACTGAC
Mcm6 (ChIP)	ACAGCTTCTGGCATTTCTCG	CCCTGTTATTGGCTGAGGTG
P21 (ChIP)	TAAGGACGTCCCACTTTGCC	GACCTCCTGTGCCTTTACCC
Mdm2 (ChIP)	CCCTGACGCAGGCTTTAGAA	CTGGCAATATTCGTGCTGGC
Gapdh (ChIP)	CGGGATTGTCTGCCCTA	GGAGGTTTCTGCACGG
E2F3 (ChIP)	CCTGGCAGAACACCTGGATT	GCTCTCAGAGGCCAGAAAGG
CcnB1 (ChIP)	TCGGAGGCCTTAAGTCGGTA	GAATCCCAGCTCTTGGCACT
Ngn2 (ChIP)	CACAACCTAAACGCCGC	TCTTCGTGAGCTTGGCAT
Cd47 (ChIP)	GGGATGGGAGACTGGTTTGG	CTGACCAGTGGAACCCCATC
Elfn1 (ChIP)	TAGAGAAAGAGCGAGGGTGCT	CTACGCCTACTCAAGCCTCG
Kazn (ChIP)	TCCTTGGGGAGTGGAACTCA	AACAACGGAAGTGCACTGGA
Fgf12 (ChIP)	CCTCTGACCAGGGCTGTGTA	TTGGTGCAGAAATGGGCACTAT
Stat1 (ChIP)	TACACCATACCACGTGCCTG	GCGTGTCCAACCATACCCAG
B1 (ChIP)	GTCACCGCATAGAATGAATTTGG	TAGACAAACTCTCAGCCAAACCAACTGT
B2 (ChIP)	TCCGGATCCACTCTACTCCAATCCTACCT	ACTAAGGGCACATTACTTGTTATGTTTC

B3 (ChIP)	GTTTTGGACATGGCTTTAATGAGTTGGT	TTAAATGCACGGAACGAGAGTCGGC
B4 (ChIP)	ATGGCCAAGAAGGCATCAGTGAG	AATGGCAAGGAGGCGGTGGAATCTG
Charged cloning	TAACGAGTCAGGTTCTTCCTTCTCCCGGTC TC	GAGAAGGAAGAACCTGACTCGTTACTG AAGAT

Table M3. List of primers used during this study and their application.

1.4 Deposited data

In this doctoral thesis several previously published datasets have been used to test and support our hypothesis.

1.4.1 GEO

Gene Expression Omnibus (GEO) is a public genome-wide data repository from which we have reanalyzed several ChIP-seq and RNA-seq data that are depicted in Table M4.

Dataset	***GEO Accesion number***
NSCs H3K9me2 (ChIP-seq)	GSE122263
ESCs H1 (ChIP-seq)	GSE49564
NSCs RNA Pol II (ChIP-seq)	GSE66961
NSCs RNA-seq in shCT and shPHF2	GSE122264
ESC Lamin B (ChIP-seq)	GSE17051

Table M4. List of previously deposited genome-wide data and their respective GEO accession numbers used in this thesis.

1.4.2 Other databases

In addition to the GEO ChIP-seq data, other databases have been used like the non-coding RNA (ncRNA), transfer RNA (tRNA) annotation in the mouse genome mm10; as well as Transponsable and Repetitive elements such as RepeatMasker – Tetranscripts (Jin et al 2015) annotation.

2. EXPERIMENTAL METHODS

2.1 Cell culture

During this doctoral thesis three different cell lines have been used: Mouse neural stem cells (NSCs), HEK293T and NIH3T3 cells. In the following pages, I will describe methods for working, manipulating and treating the three of them. All of the cell culture-related procedures were done in sterile conditions. In addition, cell lines were periodically checked for mycoplasma contamination.

2.1.1 Cell culture growth and maintenance

2.1.1.1 Neural stem cells

Mouse NSCs were dissected from cerebral cortices from E12.5 mouse fetal brains of C57BL/6J strain. Once disgregated, cells were cultured in poly-D-lysine (Milipore, A-003-E) (5 µg/mL, 2 hours 37ºC) and laminin (Merck, L2020) (5µg/mL, 6 hours 37ºC) precoated dishes following the previously assessed potocols (200). The culture media was prepared by mixing equal parts of Dulbecco's Modified Eagle Medium F12 (DMEM) (Gibco, 31331093) and Neurobasal medium (Gibco, 12348017), to which 1% Penicilin/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 β-mercaptoethanol (Gibco #31350-010) were added (201).

NSCs, due to their commited but not differenciated stem state. They have been extensively characterized for their ability to proliferate and differentiate into a wide range of neural cell types in culture (5,200–202), However they are really sensitive to temperature, pipetting or other chemical or mechanical stimuli. So, care should be taken when manipulating them in order to maintain a stable culture. 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). After 2 min at 37ºC, the Accutase was diluted with previously warmed PBS and cells were centrifuged at 900 rpm for 3 min. Cell pellet was gently flicked and then diluted in previously warmed expansion medium supplemented with fresh growth factors EGF 20ng/mL (Vitro, 236-EG-200µg) and FGF 10ng/mL (Invitrogen, PHG0021), which help NSCs to maintain the ability to self-renew. Then, medium containing the cells was split in precoated plates. They also preserve the ability to generate a wide range of differentiated neural cell types under appropriated conditions and originate differentiated neural cell types (203,204).

1.1.1.1 Stablished cell lines

HEK293T cells were derived from human embryonic kidney and transformed with the large T antigen of the SV40 virus (205), these cells were used mainly for lentivirus production. NIH3T3 cells were derived from desegregated NIH Swiss mouse embryo fibroblasts and spontaneously become immortalized. 3T3 means "3-day transfer, inoculum 3 x 105 cells, which corresponds to the protocol with which the immortal cell line was stabilized in cell culture (206).

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) (207). 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 min at 37ºC 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 Cell freezing and thawing

Cells can be stored in cryotubes for weeks at -80ºC and for years in liquid N2. The freezing protocol was the same as the one followed to expand cells but diluting them in medium containing 10% of sterile di-methyl-sulfoxide (DMSO) (Merck, D2650-5X5ML) instead of normal medium. Then a progressive descent in temperature was achieved by keeping cells at -80ºC in a box that contains isopropanol called Mr. Frosty (Nalgene, 5100-0001). The alcohol makes the temperature to lower at a 1ºC/min rate, allowing an optimal freezing process in which cells are frozen without stress or damage. The thawing process must be rapidly performed to avoid cell death, so the cryotube with the desired cells was introduced in a 37ºC water bath until it reached a liquid state. Then the content of the vial was diluted in PBS and centrifuged at 900 rpm for 5 min to eliminate the DMSO previously used for freezing. Finally, the cell pellet was diluted in the expansion medium and plated into plates, pre-coated if necessary.

2.1.3 Genetic manipulation of growing cells

Cell genetic manipulation can be achieved by introducing exogenous DNA that will finally code for specific proteins or impede the translation of mRNAs. In this section, I will detail the experimental procedures that have been used for DNA delivery into the cells.

2.1.3.1 Calcium phosphate and other transfection methodologies

In this protocol, the exogenous DNA is delivered into HEK293T cells through endocytosis of the calcium phosphate precipitates that contain the plasmid stuck on the surface. In order to generate this calcium phosphate precipitates, the DNA of interest is added into a mixture of Hepes buffered saline (HEBS) (250 mM NaCl, 9mM KCl, 1.5 mM Na_2HPO_4, 10 mM Glucose and 50 mM Hepes pH 7.12), 0.25 M $CaCl_2$ and the desired DNA under vortexing conditions. After 15 min at room temperature (RT), the precipitates are pipetted up and down several times, and the mixture is added drop-wisely to the plate with growing medium. After 4 to 6 hours, this media is replaced for fresh one, so that cells do not stress due to medium acidification.

NIH3T3 cells on the other hand are much more difficult to transfect by the previous method, so instead transfection agents like FUGene HD (Promega, E2311) and Lipofectamine 3000 (Invitrogen, L3000001) were used following manufacturer's instructions.

2.1.3.2 NSCs nucleofection

NSCs genetic manipulation for non-lentiviral plasmids was done through nucleofection. This technology utilizes a specific combination of cell type-specific solutions and optimized electrical parameters that enables DNA transfer into the cells' nucleus. For this purpose, we used Lonza program A-033 and Amaxa Mouse NSC Nucleofector Kit (Lonza, VPG-1004) and followed manufacturer's instructions.

2.1.3.3 Lentiviral transduction

This delivery method is highly efficient, achieving up to 90% of infected cells. That is why in this doctoral thesis, lentiviral transduction has been applied to transduce shRNAs that will permit the knocking down of proteins in our systems, as well as for the overexpression of PHF2 WT and mutants in NSCs. This procedure consists in three steps: lentivirus production, lentiviral transduction, and selection. Lentiviral particles were produced in HEK293T cells by calcium phosphate co-transfecting the DNA encoding the target shRNA or the exogenous overexpression DNA (Table M1 list the ones used in this study), pCMV-VSVG and pCMV-GAG-POL plasmids which encode the viral capsid and transcriptional machinery respectively. After 24 hours, the supernatants that contain the generated lentiviral particles are collected and centrifuged in a sucrose layer previously added to the tube, at 24000 rpm during 2.5 hours 4ºC. For an increased efficiency the centrifuge speed is decreased at a very low rate. Then, supernatant is removed, and viral particles are resuspended in the corresponding medium for usage or storage at -80ºC. Lentiviral transduction consists in the addition of the medium

containing viral particles to the target cells, could be either NIH3T3 or NSCs in this thesis. After 24 hours, cells are selected with the correspondent antibiotic, that for PLKO plasmids is puromycin (Merck, P8833) at 2 µg/mL, and for pInducer plasmids is neomycin (Merck, 108321) at 600 µg/mL. If induction is needed, doxycycline (DOX) (Millipore, 324285) was used at 1 µg/mL. Cells are considered selected usually after 48 hours of selection, and the knocking down efficiency can be assessed by RT-qPCR or Western Blot.

2.1.4 Cell treatments

2.1.4.1 Doxycycline cell induction

If induction was needed, DOX was used at 1 µg/mL, and a minimum of 48 hours of induction was performed. After these 48 hours, the protein overexpression can be assessed by RT-qPCR or Western Blot.

2.1.4.2 Demelcocine treatment

Mitotic arrest on NIH3T3 cells was done via demelcocine treatment. Demelcocine inhibits microtubule polymerization in the cell, therefore causing a metaphase arrest. This protocol allowed us to achieve metaphase mitotic chromosomes. To achieve this, demelcocine (Merck, D1925-100ML) was added to a final concentration of 0.1 µg/mL into growing NIH3T3 for a maximum time of 8 hours. Following this incubation where around 60% of the cells are arrested in metaphase, only these latter cells were collected by selective detachment tapping the plate or flask, achieving a 90% of alive metaphase arrested cells.

2.1.5 Cell synchronization

NIH3T3 cells can be synchronized and therefore a synchronization protocol was developed for this purpose. Initially cells were plated at 50% confluence for serum starvation (FBS 0.5%) during 24 hours at 37ºC. This way, most of the cells enter in a G0-like state and cell cycle is set in the starting point. Then, serum was restored at 10% and cells were collected after 6 hours followed by IP staining that showed 85% of G1 cells. If M phase wanted to be collected, 10% serum was incubated for 16 hours, and then Nocodazole (Merck, M1404) 0.1 µg/mL was added for up to 6 hours. After incubation, fresh media was added to wash out nocodazole for 1.5 hours. Cells were collected by selective detachment and IP staining (see following 2.2.1.2.) showed 75% of M phase cells. Both G1 and M phase cells were also confirmed by DNA staining with DAPI and FACs analysis (Figure M1).

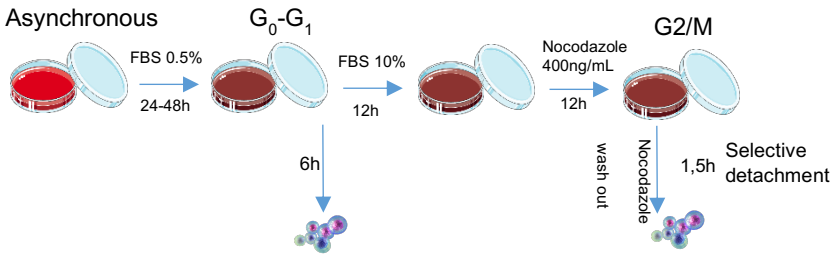


Figure M1. Illustrative representation of the cell synchronization protocol used during this doctoral thesis.

2.2 Molecular biology procedures

2.2.1 Nucleic acids related

2.2.1.1 Genomic DNA extraction

For this thesis, genomic DNA extraction from cells was a crucial step for various subsequent applications. Initially, 200 µl of buffer containing 10 mM Tris-HCl, 10 mM EDTA, 10 mM NaCl, and 0.5% SDS were utilized to lyse approximately 6x10^6 cells. The lysates underwent incubation with 0.5 mg/ml of proteinase K (Sigma, P2308) for 1 hour at 50ºC, followed by treatment with 1 mg/ml of RNase A (Fermentas, EN0531) for 2 hours at the same temperature. Finally, DNA purification from the lysates involved a phenol-chloroform extraction process, which will be elaborated on in the following pages of this chapter.

2.2.1.2 Propidium Iodide (PI) staining and FACS

In order to achieve representative G1 and M phase cell populations, propidium iodide (Merck, 537059) staining and fluorescence activated cell sorting (FACS) were performed. PI was added to cells in suspension to a final concentration of 50 µg/mL allowing fluorescence intensity (above 630 nm) to be used as an indicator of cellular DNA content. Cell populations were passed through FACS and their DNA content was split into G1 (one copy) or M (two copies).

2.2.1.3 Native mitotic chromosome preparation and staining for flow sorting

This preparation was done in collaboration with Amanda Fisher's group in the Department of Biochemistry at Oxford University and therefore this protocol was described previously in Djeghloul et al 2020 (208) in which they modified a previously established protocol (209). Briefly, mitotic arrested NIH3T3 cells were prepared as described before in 2.1.4.2 of this section and centrifuged at 1200 rpm 5 min. Cell pellet was then resuspended in a hypotonic solution (KCl 75 mM, spermidine 0.5 mM, spermine 0.2 mM, $MgSO_4$ 10mM, pH 8.0) and incubated 20 min at RT that would create holes in the cell membranes. After that, cells were centrifuged again, resuspended in ice cold polyamine isolation buffer (TRIS 15 mM, EDTA 2 mM, EGTA 0.5 mM, KCl 80 mM, DTT 3 mM, Triton 0.25X, spermine 0.2 mM, spermidine 0.5 mM, pH 7.5) and incubated on ice for 15 min. Condensed mitotic chromosomes were then released by vortexing 30 seconds and syringing the suspension with a 22.5 needle 6 times. Cellular debris were pelleted down by centrifugation at 1000 rpm 2 min 4ºC and supernatant was filtered using a 20 µm mesh filter. Finally, chromosomes were stained with Hoescht 33258 (Merck, 94403) 5 µg/mL, Chromomycin A3 (Merck, C2659) 50 µg/mL and $MgSO_4$ 10 mM overnight at 4ºC. At least 1 hour before chromosome sorting sodium citrate 10 mM and sodium sulphite 25 mM was added.

2.2.1.4 Chromosome flow sorting

Chromosomes underwent examination via flow cytometry utilizing a Becton Dickinson Influx instrument (BD FACS software version 1.2.0.142), which was equipped with spatially separated air-cooled lasers. Hoechst 33258 was excited by a Spectra Physics Vanguard 355 nm laser with a power output of 350 mW. Fluorescence from Hoechst

33258 was collected using a 400 nm long pass filter combined with a 500 nm short pass filter. Chromomycin A3 was excited using a Melles Griot 457 nm laser with a power output of 300 mW, and its fluorescence was collected using a 500 nm long pass filter combined with a 600 nm short pass filter. Forward scatter was measured using a Coherent Sapphire 488 nm laser with a power output of 200 mW, serving as the trigger signal for data collection. Chromosomes were sorted at an event rate of 15,000 per second using a 70 µm nozzle tip, with a drop drive frequency set to ~96 kHz and a sheath pressure of 65 pounds per square inch. Isolated chromosomes were collected in DNA low-binding tubes containing an excess of polyamine buffer.

2.2.1.5 Phenol chloroform extraction and ethanol precipitation

Taking advantage of the differing affinity of DNA and proteins for phenol, this technique serves to purify DNA from complex protein-DNA mixtures. The mixture underwent separation into aqueous and organic phases by the addition of 1 volume of phenol, followed by centrifugation at 16,392 g for 3 min. The aqueous phase, containing the DNA, was carefully transferred to a clean tube. This process was repeated using chloroform instead of phenol. Subsequently, DNA in the aqueous phase was ethanol-precipitated to concentrate and desalt it. To achieve this, 0.1 volumes of NaAc 3 M and 1 volume of ice-cold ethanol were added to the sample, which was then incubated at -80ºC for at least 30 min to facilitate precipitation. Afterward, the sample was centrifuged at 16,392 g for 20 min at 4ºC and washed with 70% ethanol. Finally, the pellet was dried and resuspended in an appropriate buffer.

2.2.1.6 RNA extraction and DNAse treatment

Not only DNA but also cellular RNA was essential for conducting experiments in this study. For RNA extraction from 3x10^6 cells, 1 mL of Trizol reagent (Invitrogen, 15596018) was employed. Subsequently, 200 µL of chloroform was added, and after centrifugation at 16,392 g for 5 min, the upper phase was collected and precipitated by adding 800 µL of isopropanol. Following centrifugation at 16,392 g for 10 min, the supernatant was discarded, and the pellet was washed with 70% ethanol. Finally, the pellet was resuspended in H2O after a final centrifugation of 5 min at 16,392 g. To prevent contamination with genomic DNA, a DNase treatment was conducted post RNA extraction, following 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 µL of DNase for 30 min at 37ºC. Subsequently, 0.1 volumes of inhibition agent were added to halt the reaction, and after 2 min of flicking, pure RNA was purified by centrifuging samples at 10,000 g for 1.5 min and transferring the supernatant to a new tube. RNA quantification was performed using a Nanodrop device, and its quality was assessed with 260/280 and 260/230 ratios. Additionally, its integrity was verified by agarose gel electrophoresis.

2.2.1.7 Retrotranscription of RNA into cDNA

To quantitatively assess transcription, RT-qPCR assays were conducted. Prior to this assay, purified RNA was reverse transcribed into cDNA. Up to 1 µg of RNA, comprising both mRNA and eRNA, was reverse transcribed to generate cDNA using the High-Capacity cDNA reverse transcription kit (Invitrogen, 4368814). The RNA was incubated

with random hexamers or deoxynucleotides (dNTPs) and a retrotranscriptase in a thermocycler, following cycles of 25ºC for 10 min, 37ºC for 120 min, and 85ºC for 5 min.

2.2.1.8 qPCR

During this doctoral thesis, qPCR was employed to quantify the retrotranscribed cDNA. Reactions were conducted in a QuantStudio 5 Real-Time PCR system (ThermoFisher Scientific) under the following conditions: initial denaturation at 95ºC for 5 min, followed by 40 cycles of denaturation at 95ºC for 5 seconds, annealing at 60ºC for 10 seconds, and extension at 72ºC for 20 seconds. A melting curve analysis was performed with a denaturation step at 95ºC for 5 seconds and annealing at 65ºC for one min. Reactions were set up in a 96-well plate to a final volume of 10 µL, containing SYBR Green (Roche, 4887352001) and specific primer pairs (refer to Table M3). For mRNA, primers were designed to span exon-exon junctions. Non-template controls and standard curves were run with every new primer pair, and only primers with an efficiency of at least 95% were retained. The $2^{-\Delta\Delta CT}$ method was employed to analyze qPCR data, and normalization was performed using the reference gene *Gapdh*. After chromatin immunoprecipitation, ChIPed DNA may be quantified via qPCR as well using primers specified in table M3. To quantify immunoprecipitated DNA relative to the starting chromatin, we employed the input material and IgGs were used to subtract background noise by non-specific DNA precipitation.

2.2.1.9 PCR

Conventional PCR has been used in this thesis to amplify genomic regions for cloning purposes. Genomic or plasmid DNA was incubated in a thermoblock with dNTPs, buffer, specific primers (Table M3) and the enzyme Taq DNA polymerase (Fermentas EP0401) or Phusion High-Fidelity DNA polymerase (NEB, M0530S). The temperatures and incubation times used were specific to the target region and following manufacturer's indications.

2.2.1.10 Cloning of plasmidic DNA

2.2.1.10.1 PHF2 mutants' construction

In order to gain insight into PHF2 function, several mutants were constructed. The first one was described in Pappa et al 2019 (5) and consists of a catalytic mutant where two of the main amino acids in the active center were mutated (HID>AIA). The second one was described in Vicioso-Mantis et al 2022 (210) where one of the intrinsically disordered regions (IDR) of PHF2 was identified as a rich positively charged area. This charged region was eliminated by PCR on P3xFlag-PHF2 using complementary ended primers (Table M3) and thus creating a PHF2ΔCharged mutant. Lastly, the remaining defined PHD domain was targeted. P3xFlag-PHF2 plasmid was used to obtain PHF2ΔPHD mutant by using the restriction enzymes HindIII and PasI. Each of the generated mutants were checked by agarose electrophoresis with several restriction patterns and by sequencing.

2.2.1.10.2 mEGFP-PHF2 construction

The overexpression of a fluorescent tagged protein of interest was achieved by using the restriction enzyme SalI on both the pmEGFP plasmid and the P3xFlag-PHF2 plasmid. The appropriate size bands were selected from an agarose electrophoresis, mixed, ligated and transformed into supercompetent cells. The resulting vector was checked by restriction, sequencing and under the fluorescence microscope once expressed in HEK293T cells. This cloning was done by supervised MSc student Monica Barrios.

2.2.1.10.3 pLIV-based PHF2 mutant's construction

Given the necessity of expressing the mutants in NSCs with high efficiency, lentiviral constructions of them were performed. PLIV-PHF2, and PLIV-PHF2ΔCharged plasmids were constructed using BamHI restriction enzyme from previous p3xFlag plasmids, and were introduced into pLIV plasmid opened with the same restriction enzyme. PLIV-PHF2ΔPHD plasmid was constructed following the same strategy as per p3Flag-PHF2ΔPHD.

2.2.1.11 Amplification of plasmidic DNA: mini and maxi preps

To amplify plasmidic DNA, the initial step involved bacterial transformation and inoculation of an isolated colony into LB medium, allowing it to grow overnight. Depending on the desired DNA concentration, either 5 mL or 500 mL of LB medium was utilized, resulting in "minipreps" (yielding around 15 µg of DNA) or "maxipreps" (yielding around 80 µg of DNA), respectively. Subsequently, bacterial DNA was purified using the QIAprep Spin Miniprep Kit (for minipreps, QIAGEN 27106) or the QIAGEN

Plasmid Maxi Kit (for maxipreps, QIAGEN 12165). Both protocols involved the sequential addition of buffers P1 (containing 100 µg/mL RNase A, 50 mM Tris-HCl, 10 mM EDTA pH 8.0), P2 (containing 200 mM NaOH, 1% SDS), and P3 (containing 3 M KAc, pH 5.5) for resuspension, lysis, and neutralization, respectively. Next, DNA was selectively retained in a column, from which it was eluted and precipitated with isopropanol after several washes. Following isolation, the DNA was washed with 70% ethanol, dried, and then resuspended in a suitable buffer.

2.2.1.12 Electrophoresis in agarose gels

This technique involves running DNA or RNA on an agarose gel for visualization and/or purification purposes. The gel was prepared by dissolving the desired percentage of agarose in TBE (Tris-Borate-EDTA) buffer, consisting of 45 mM Tris, 45 mM boric acid, and 1 mM EDTA. Additionally, Red-safe reagent (Intron, 21141) was included in the gel mixture, enabling the visualization of nucleic acids by emitting fluorescence upon binding to them. Once the gel was solidified, samples containing orange-glycerol were loaded, and the gel was electrophoresed at 90 V for typically 1-3 hours, depending on the required separation of bands. Subsequently, nucleic acids were visualized using a UV-transilluminator.

2.2.2 Protein related

2.2.2.1 Total protein extraction

This protocol was conducted to assess the levels of specific proteins in particular contexts. The buffer utilized was the highly stringent RIPA (Radioimmunoprecipitation

assay buffer), capable of disrupting both cytoplasmic and nuclear membranes. This buffer consisted of 150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, and 50 mM Tris (pH 8.0), supplemented with protease inhibitors. Cells were suspended in RIPA buffer on ice for 20 min, with the volume of RIPA buffer dependent on the desired final protein concentration, typically around 500 µl for 6x10^6 cells. Subsequently, to eliminate cellular debris, the suspension was centrifuged at 4ºC and maximum speed for 10 min using an Eppendorf 5415 centrifuge, yielding protein extract in the supernatant.

2.2.2.2 Protein quantification

To quantify the amount of protein extracted, we employed the Bradford method, which relies on the reaction where Coomassie Brilliant Blue G-250 (Bio-Rad, 5000001) changes its color in response to protein concentration. Specifically, 1 µL of protein extract was mixed with 1 mL of the Bradford reagent. After allowing 5 min for the reaction to proceed, the absorbance value of the sample was measured using a spectrometer. Finally, the protein concentration was determined using a calibration curve generated with bovine serum albumin (BSA) at defined concentrations.

2.2.2.3 SDS-Page electrophoresis

This protocol is suitable for separating proteins in a gel based on their actual size, facilitated by the net negative charge proteins acquire when in solution with sodium dodecyl sulfate (SDS). Protein extracts were mixed with Laemmli buffer (containing 375 mM Tris-HCl, 9% SDS, 50% glycerol, and 0.03% bromophenol blue) along with 5% β-mercaptoethanol. Before loading onto a polyacrylamide gel, they were heated at 95ºC for

5 min. The gel, composed of TEMED, 10% ammonium persulfate, 10% SDS, 1.5 M TRIS pH 6.8/8.8, variable % polyacrylamide, and water, consisted of two distinct parts: the stacking gel (with a 5% polyacrylamide concentration and pH 6.8) and the resolving gel (with a polyacrylamide concentration depending on the sizes of proteins to be resolved and pH 8.8). The gel was covered with SDS-PAGE running buffer (containing 25 mM Tris-HCl, 192 mM glycine, and 0.1% SDS) and run at 25 mA until proteins migrated sufficiently to distinguish them.

2.2.2.4 Western Blot

This assay was performed to detect a specific protein in a sample previously resolved by SDS-PAGE electrophoresis and involved transferring proteins from the gel to a nitrocellulose membrane using an 80 V current, causing them to 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, another Whatman paper, and another sponge were arranged in the cassette, in that order from the negative to the positive side. The cassette was then covered with transfer buffer. For conventional proteins the transfer buffer contained 25 mM Tris-HCl, 192 mM glycine, 0.05% SDS, and 10% methanol. The transference was carried out for 90 min. Following protein transference, the nitrocellulose membrane was blocked with 5% skim milk for 1 hour at RT. After three washes with PBS-Tween 0.1%, it was incubated with a specific primary antibody (refer to Table M2) overnight at 4ºC. The following day, the membrane was washed again three times with PBS-Tween 0.1% and incubated with a secondary antibody (refer to Table M2) at RT for 1 hour. These antibodies were conjugated to a fluorophore, allowing

visualization of proteins using a LI-COR Odyssey scanner model ODY-1871. Fiji (211) software was used for immunoblotting relative quantification following NIH guidelines.

2.2.2.5 Indirect immunofluorescence

2.2.2.5.1 Cell immunofluorescence

This microscopy technique aims to visualize stained proteins within cells and was carried as described (212). Cells grown on coverslips coated with Poly-D-lysine were fixed for 20 min at RT in 4% paraformaldehyde. Subsequently, they were permeabilized with methanol for 10 min at RT and then blocked in 5% BSA for 30 min at RT. Following blocking, the coverslips were incubated overnight at 4ºC with primary antibodies (refer to Table M2). The next day, after thorough washes, cells were incubated for at least 2 hours at RTwith Alexa-conjugated secondary IgG antibodies (refer to Table M2) and after washes, cells were incubated with DAPI 0.1 ng/µl (Sigma) for 1 hour, then mounted with ProLong Glass Antifade Mountant (Thermofisher, P36980). These secondary antibodies emit fluorescence, enabling the visualization of proteins under the microscope. Confocal images were acquired in a Dragonfly 505 multimodal spinning-disk confocal microscope (Andor Technologies, Inc) and Fusion Imaging Software. Laser excitation and image acquisition were performed sequentially for each channel. Exposure duration and laser potency were adjusted accordingly, ensuring the absence of oversaturated pixels. Confocal 3D images were captured as Z-stacks with intervals of 0.13µm and saved as 16-bit images.

2.2.2.5.2 Chromosome immunofluorescence

Flow-sorted chromosomes (chromosomes 19 and X) were cytocentrifuged onto Poly-L-lysine-coated slides (VWR) using a Cytospin3 (Shandon) at 1300 r.p.m. for 10 min at RT. The chromosome samples were then blocked with 6% normal goat serum for 1 hour at RT and incubated overnight at 4°C in a humid chamber with primary antibodies (see Table M2). After incubation, the chromosomes were washed with a buffer containing 10 mM HEPES, 2 mM MgCl2, 100 mM KCl, and 5 mM EGTA, followed by incubation with appropriate secondary antibodies (refer to Table M2) for 1 hour at RT. The immuno-stained chromosomes were mounted in Vectashield antifade mounting medium with DAPI (Vector Laboratories, H-1200-10). Wide-field epi-fluorescence microscopy was conducted on an Olympus IX70 inverted microscope using a UPlanApo 100×/1.35 oil objective lens. Again, laser excitation and image acquisition were performed sequentially for each channel and saved as 16-bit images.

2.2.2.6 Live imaging

For live cell imaging NIH3T3 cells were cultured in an IBIDI® µ-35mm dish, and 24 hours before imaging were transfected with mEGFP-PHF2 vector or co-transfected with the VenusN and VenusC reporter vectors. VenusN vector also code for a zinc finger protein that binds specifically to MajorSat; while VenusC vector encodes for the HPI chromodomain that binds specifically to H3K9me3. That way, we were able to follow specifically the K9 trimethylation at the MajorSat (213,214). Life DNA probe DRAQ5 (ThermoFisher Scientific, 62251) was used to monitor DNA in the far-red spectrum, thus

avoiding laser damage to the cells. Preparations were imaged using the previously mentioned Dragonfly 505 multimodal spinning-disk confocal microscope and a specific stage by Andor. Time frames were set up differently for every experiment according to our needs, and the intrinsic fluorescence of the molecules was captured without using either primary or secondary antibodies.

2.2.2.7 Super resolution imaging

Super-resolution images were acquired in a Dragonfly 505 multimodal spinning-disk confocal microscope (Andor Technologies, Inc). Super-Resolution Radial Fluctuations (SRRF) algorithm (7,215,216) was applied using the SRRF-Stream+ module (Andor) operated from the Fusion software. SRRF operates under the assumption that an image comprises point sources convolved with a point spread function (PSF), characterized by a higher degree of local radial symmetry than the background (radiality). Radiality, in this context, pertains to the local radial gradient's convergence towards the center of the fluorophore signal. The SRRF algorithm magnifies each image pixel into sub-pixels, assigning each subpixel a non-binary value reflecting the probability of containing a fluorophore. Additionally, temporal analysis of radiality enables noise reduction and resolution enhancement. Noise is distinguished from specific signals through temporal correlation, as noise lacks correlation over time, unlike fluorophores. Furthermore, the highest degree of temporal correlation is typically observed at the center of radiality peaks generated by fluorophore signals. The following parameters were used for all conditions: 1x ring radius, 5x radiality magnification, 100 frames for a single confocal plane of 1024x1024.

2.2.3 Genome wide techniques

2.2.3.1 Assay for Transposase-Accessible Chromatin followed by sequencing

ATAC-seq is a method designed to capture open chromatin sites utilizing a straightforward two-step protocol, suitable for analyzing samples ranging from 500 to 50,000 cells. This technique offers insights into the intricate relationships among genomic regions of open chromatin, DNA-binding proteins, individual nucleosomes, and the compaction of chromatin at a nucleotide resolution (217). Briefly, to prepare nuclei, we initially spun 50,000 cells at 500 × g for 5 min. Afterwards, we conducted a wash using 50 µL of cold 1x PBS followed by centrifugation at 500 × g for 5 min. Subsequently, the cells were lysed using cold lysis buffer containing 10 mM Tris-Cl (pH 7.4), 10 mM NaCl, 3 mM MgCl2, and 0.1% IGEPAL CA-630. Following lysis, the nuclei were spun at 500 × g for 10 min using a refrigerated centrifuge. To minimize cell loss during the nuclei preparation process, we employed a fixed-angle centrifuge and carefully pipetted away from the pellet after centrifugations. Following the nuclei preparation, the pellet was promptly resuspended in the transposase reaction mix consisting of 25 µL of 2x TD buffer, 2.5 µL of Transposase (Illumina, 20034197), and 22.5 µL of nuclease-free water. The transposition reaction was then allowed to proceed for 30 min at 37°C. Immediately after transposition, the sample underwent purification using a Qiagen MinElute kit (QIAGEN, 28004). For library construction, Nextera DNA Library Prep Kit (FC-121-1030), along with unique dual index barcodes for each sample and NEBNext High-Fidelity 2x PCR Master Mix (New England Lab, M0541) were used,

then library fragments were amplified using 1x NEBnext PCR master mix (NEB, M0541S). The PCR conditions included an initial denaturation step at 72°C for 5 min, followed by an initial denaturation at 98°C for 30 seconds. Thermocycling was performed at 98°C for 10 seconds, 63°C for 30 seconds, and 72°C for 1 min. To mitigate GC and size bias in the PCR, we monitored the reaction using qPCR to prevent saturation. Initially, the full libraries were amplified for 5 cycles. After this, an aliquot of the PCR reaction was taken, and 10 µL of the PCR cocktail with SYBR Green at a final concentration of 0.6x was added. This reaction was run for 20 cycles to determine the additional cycles needed for the remaining 45 µL reaction. The libraries were subsequently purified using a Qiagen PCR cleanup kit (QIAGEN, 28104), resulting in a final library concentration of approximately 30 nM in 20 µL. Finally, libraries were amplified for a total of 10–12 cycles. Sequencing was done in a NextSeq 2000 and 50M reads per sample were acquired.

2.2.3.2 Chromatin Immunoprecipitation (ChIP)

ChIP assays were conducted following the protocol outlined by Valls et al. 2007, (218) with some modifications: 1 x 10^6 NSCs underwent cross-linking with 0.4% Cross-link Gold (Diagenode, C01019027) in PBS for 30 min, followed by fixation with 1% methanol-free formaldehyde for 10 min and quenching with 0.125 mM glycine for 10 min. Cells were lysed successively in lysis buffers 1 (50 mM Hepes. 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton), 2 (10 mM Tris, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA), and 3 (10 mM Tris, 100mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine) all of these steps were

performed for 10 min, and each buffer was supplemented with cOmplete protease inhibitor cocktail (Roche, 04693116001). Chromatin fragmentation was achieved using a Bioruptor Pico (Diagenode, B01080010) sonicator until DNA fragments were around 180-200 bps, and immunoprecipitation was performed with the antibody of interest or general IgGs (refer to table M2) overnight at 4ºC. The same amount of ChIPed DNA was used in each immunoprecipitation reaction, and 1% of input was saved. The following day, immuno-complex was captured using magnetic beads (Magna ChIP™ Protein A Magnetic Beads (Millipore, 16-661) that were previously blocked with 0.5% BSA, and a minimum of five washes were performed with RIPA wash buffer (50 mM Hepes, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-deoxycholate) using a magnetic holder and waiting for magnetic beads to settle on the tube side, and decrosslinking was done overnight at 65ºC. Subsequently, RNAse A 0.2 mg/mL (ThermoFisher Scientific, EN0531) 30 min 37ºC and Proteinase K 0.2 mg/mL (Merck, P2308-25MG) 2 hours at 55ºC were added to digest RNA and protein respectively. Finally, DNA purification via phenol-chloroform extraction was performed and ChIPed DNA was eluted in a convenient buffer.

2.2.3.2.1 ChIP followed by sequencing (ChIP-seq)

Chromatin immunoprecipitation and sample preparation for sequencing were conducted essentially as previously described in the subsection 2.2.3. 2.. Library preparation and sequencing was performed externally at Beijing Genomic Institute (BGI, China). Briefly, after a standard quality check, qualified ChIP and Input samples underwent end-repair and then 3’ adenylation. Adaptors were ligated to the ends of these 3’ adenylated

fragments. Fragments were PCR-amplified, and PCR products were purified and selected with the Agencourt AMPure XP-Medium kit. The double-stranded PCR products were heat-denatured and circularized by the splint oligo sequence. The single-strand circular DNA (ssCir DNA) was formatted as the final library and then quality-checked. The library was amplified to produce DNA nanoballs (DNBs), each containing more than 300 copies of one molecule. The DNBs were loaded into the patterned nanoarray, and single-end 50-base reads were generated using combinatorial Probe-Anchor Synthesis (cPAS).

2.2.3.3 Global Run-On followed by sequencing (GRO-seq)

The cellular levels of a given RNA molecule are influenced by the intricate interplay of transcription, processing, and degradation. Thus, alterations at both the transcriptional and posttranscriptional levels can impact RNA abundance as detected by RNA-seq. Over the years, various RNA-seq-based methodologies have emerged to specifically capture nascent RNA transcripts synthesized by actively engaged RNA polymerases. Among these techniques, global run-on sequencing (GRO-seq) has become one of the most widely utilized (219,220). In this thesis, we performed a three-time point captured GRO-seq: M phase, early G1 (after 2.5 hours of Nocodazole washout) and G1.2 (after 5 hours of washout). To assess the nascent-RNA capture we used Click-iT™ Nascent RNA Capture Kit (ThermoFisher Scientific, C10365) following manufacturer's specifications: Ethylene uridine (EU) ribonucleotide analogs containing a reactive alkyl group are introduced into cells. Once incorporated, cells are lysed and RNA is isolated. The azide-biotin undergoes click chemistry, and streptavidin magnetic beads are then used to capture the newly synthesized RNA pool. The captured RNAs are amplified, and the

resulting cDNAs are used for library construction and sequencing externally at BGI (China). For each sample, 40M of paired-end clean reads were sequenced.

2.3 Data analysis

2.3.1 Image analysis

For confocal z-stacked images, the maximal intensity Z-projection was calculated. The fluorescence intensity per cell was measured using FIJI ImageJ Software (211). To correct the cell fluorescence, the following formula was applied: *Integrated Density - (Area of selected cell × Mean fluorescence of background readings)*. Multinucleated cells quantification was achieved by counting cells in randomly located fields marked with EGFP expression (see Table M1) and counting number of nuclei per cell staining the DNA with DAPI. Nuclear distance to lamina of PHF2, PHF8 and KIAA1718 was measured using the fluorescence intensity plot of each channel and linear measure in microns from a clear fluorescence peak between channels was quantified.

2.3.2 Nuclear foci analysis

Nuclear foci were identified using the "Object Counter 3D" plugin in FIJI. In each image, the "threshold" parameter was adjusted to ensure that each focus appeared as a separate object. The parameters reported (number of foci per cell, intensity, and volume) represent the average values obtained from the results of each image using the "Statistics" function of the plugin. The number of cells utilized for quantification is specified in each figure.

Nuclear distance to DAPI chromocenters of PHF2, PHF8 and KIAA1718, as well as distance of PHF2 to chromocenters in SRRF images was measured using the FIJI plug in Radial Profile. This plugin generates a profile plot showing normalized integrated intensities around concentric circles based on the distance from a point in the image. The point is automatically determined as the center of the rectangle enclosing the current Region of Interest (ROI), although its position can be adjusted using a dialog box. At any given distance from the point, the intensity is calculated as the sum of pixel values around a circle centered at the point with the distance as its radius. This integrated intensity is then divided by the number of pixels within the circle, ensuring normalized values that are comparable. The x-axis of the profile plot can be displayed in either pixel values or spatial calibration units based on the input image, and linear measure in microns from peaks between channels was quantified.

2.3.3 Chromosome imaging analysis

For individually sorted native mitotic chromosome analysis, Z-projections of maximum intensity were calculated using FIJI software (211). Then, specific ROIs were defined taking DAPI staining as a reference on centromeres (high intensity DAPI) and whole chromosomes. Finally, fluorescence intensity was measured as described in subsection 2.3.1 and area of the ROI was calculated to depict chromosome and centromere size.

2.3.4 ATAC-seq analysis

Paired-end reads were aligned to the mm10 genome using Bowtie2 v2.3.5.1 (221) with the parameters -X2000 and -m1. These settings permitted the alignment of fragments up

to 2 kb (-X2000) and ensured that only uniquely aligning reads were retained (-m1). SAMtools (v1.9) (222) was employed to filter out the low-quality reads with the flag 1796, remove reads mapped in the mitochondrial chromosome, and discard those with a MAPQ score below 30. For peak calling and footprinting analyses, the read start sites were adjusted to reflect the center of the transposon binding event. The resulting BAM files were sorted, and deepTools2 (223) was utilized to generate counts per million (CPM) normalized signal tracks in bedGraph and bigWig format using the command "bamCoverage-samFlagInclude 64-normalizeUsing CPM. Peak calling was done using MACS2 v2.1.2 (224) as broad peaks at q-value < 0.01. Differential peaks between experimental conditions were identified using DeSEQ2 software at FDR < 0.05 (225).

2.3.5 ChIP-seq analysis

Single-end reads underwent quality checking using FastQC (v0.11.9) and were aligned to the mouse mm10 reference genome using Bowtie2 v2.3.5.1 (221) with default options. SAMtools (v1.9) (222) utilities were employed to filter out low-quality reads with the flag 3844. Genome coverage for Input, PHF2, H3K4me3 and H3K9me3 was calculated and normalized by reads per million using BEDTools v2.28.0 (226), and regions with zero coverage were reported in the ChIP-Seq annotation (genomecov -ibam -bga -scale). MACS2 v2.1.2 (224) was utilized to subtract input coverage from PHF2, H3K4me3 and H3K9me3 to generate signal tracks (bdgcmp -m subtract). The resulting BAM files were sorted and deepTools (223) was used to generate input-subtracted and counts per million (CPM) normalized signal tracks (bamCompare --operation subtract --normalizeUsing CPM --scaleFactorsMethod None) in bedgGraph and bigWig format.

2.3.6 Repetitive elements analysis

BEDTools (v2.28.0) (226) was utilized to calculate the mean ATAC-seq and ChIP-seq signal within TEtranscripts (227) mm10 repeat annotation or genome-wide 10 Kb bins using the "map -o mean" option. The TEtranscripts annotation comprises a GTF file customly generated from the RepeatMasker annotation, containing over 3.7 million repeats organized into three levels: Class, Family, and Repeat. Repeats overlapping with regions in the mm10 ENCODE BlackList were excluded from the analysis. Heatmaps were generated using the R package pheatmap, employing the "euclidean" distance measure and the "complete" cluster method for clustering rows and columns.

2.3.7 GRO-seq analysis

GRO-seq analysis was done by supervised JAE Intro student Arnau Noguera. Reads were aligned to the mouse reference genome (mm10) Bowtie2 v2.3.5.1 (221) with default parameters and specifying strand-specific information (--rna-strandness RF). SAMtools v1.11) (222) was employed to sort BAM files and filter for properly paired-end reads (-f 2). Aligned reads were mapped to Ensembl gene annotation with TEtranscripts v2.1.4 (--sortByPos --mode multi --stranded reverse) (227). DESeq2 v1.26.0 (225) was then used to identify differentially expressed genes between shCT and shPHF2 cells. Moderated log2 fold change values were calculated using the shrinkage method to facilitate ranking and visualization without arbitrary filters on low count repeats. The Benjamini-Hochberg (BH) method was applied to correct for multiple testing and control the proportion of

false positives or FDR. Gene expression changes were deemed significant if the absolute value of the log2(FC) was greater than 1.4 and the adjusted P-value was less than 0.05.

2.3.8 Data quality and representation

2.3.8.1 Scatter plots, bar plots, boxplots and violin plots

In the Results section, I created scatter plots, bar plots, box plots and violin plots to visualize the corresponding data. Scatter plots were used for qPCR when n=2. Bar plots were used for relative mRNA expression and percentage of multinucleated cells. Boxplots were used for input subtracted ChIP-seq signal. Violin plots were used for immunostaining data. Replicate data was entered into columns using GraphPad Prism version 9.00 for Windows, La Jolla California USA (https://www.graphpad.com).

2.3.8.2 Heatmaps

ChIP-Seq profiles and at meta-repeats were constructed and visualized using deepTools (v3.5.1) (computematrix, plotHeatmap, plotProfile) (223). All regions were scaled to the same size (scale-regions mode). Heatmaps were performed by using the R package pheatmap.

2.3.8.3 Captures obtaining

Several captions of both ChIP-seq and ATAC-seq signals are included in this doctoral thesis. They result from the visualization of bigwig and bedgraph files of different experiments in the Integrated Genomics Viewer (IGV) from the Broad Institute (228).

2.3.8.4 GO analysis

The software Gene ontology was used to assign biological process categories of the input genes using as background the whole Mus Musculus genome (229).

2.4 Statistical analysis

2.4.1 Sample size

In this doctoral thesis, ChIP-qPCR, RT-qPCR and immunostaining experiments were conducted in triplicate. In specific cases, such as validating H3K9me3 ChIP-seq results with the HPIα, two replicates were performed. ChIP-seq experiments were done in quadruplicate for PHF2 and in duplicate for H3K9me3 and H3K4me3 in both shCT and shPHF2 conditions, as well as PHF2 in G1 and M conditions. RNA-seq and GRO-seq were also performed in duplicate for each condition.

2.4.2 Measures of dispersion (Mean and SD)

The software utilized for calculations and graphical representations were Microsoft Excel and GraphPad Prism 9. Experiments conforming to a linear model were depicted as the mean, with error bars representing the standard deviation (SD). Boxplots were employed to show values ranging from Q1 to Q3 of the dataset, with the median value indicated by a line within the box, and whiskers illustrating the data range from the minimum to maximum values within the dataset.

2.4.3 Student's t-test

To evaluate the significance of the results between two conditions in ChIP-qPCR, RT-qPCR and immunostaining data, we conducted the Student's t-test. We consider an experiment statistically significant when the result falls within a 95% confidence interval, indicating a true hypothesis. Asterisks denote different p-values resulting from this test, calculated using GraphPad Prism 9 software. The asterisks representing the p-value are as follows: * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; **** p-value < 0.0001.

2.4.4 Wilcoxon signed-rank test and Mann-whitney U test

When experiments did not conform to a linear model were depicted as the mean, Wilcoxon signed-rank test and Mann-Whitney U test were used. Essentially, the tests assess whether it's equally likely for a randomly selected value from the first population to be greater or lesser than a randomly selected value from the second population. We employed this test to evaluate the significance of differences in the ChIP-seq signals across various conditions. The asterisks representing the p-value are as follows: * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; **** p-value < 0.0001.

2.4.5 Fisher's exact test

Fisher's exact test was assessed for evaluating the difference between mono and multinucleated cells. The asterisks representing the p-value are as follows: * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; **** p-value < 0.0001.

1.2 Data availability

All the genome-wide data generated during this doctoral thesis have been deposited in open-access databases. PHF2, H3K9me3, H3K4me3 ChIP-seq data and ATAC-seq data have been deposited in Gene Expression Omnibus (GEO) database (230) under the accession number GSE242385. Previous mass spectrometry data have been deposited to the ProteomeXchange consortium via the PRIDE (231) partner repository under the dataset identifier PXD051336.

RESULTS

1. CHAPTER 1: PHF2's role in heterochromatin stability and PcH boundaries

As described in the introduction, the expansion of neural progenitor cells is a critical process in early neurogenesis, facilitating the generation of the diverse array of neural cells required for proper development. Recent studies have highlighted the essential role of the histone demethylase PHF2 in the expansion of NSCs, both in vitro and in vivo (5). Moreover, results from our lab also demonstrated that PHF2 interacts with chromatin components like SUV39 or HP1BP3 in NSCs through MS and CoIP, and its depletion led to satellite expression. This chapter aims to delve into the mechanistic insights underlying PHF2's influence on heterochromatin and genomic stability as well as to characterize the functional domains necessary for each function.

Most of the findings detailed in this chapter have been recently published in the scientific journal EMBO Reports under the title:

"PHF2-mediated H3K9me balance orchestrates heterochromatin stability and neural progenitor proliferation" DOI: 10.1038/s44319-024-00178-7

1.1 PHF2 is enriched in heterochromatic genomic regions

Based on the identification of interactions between PHF2 and heterochromatin components, we aimed to investigate the potential enrichment of PHF2 at heterochromatic genomic regions. To address this, we performed PHF2 chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments in quadruplicate using NSCs that were validated by quantitative polymerase chain reaction (qPCR) (see Figure R1).

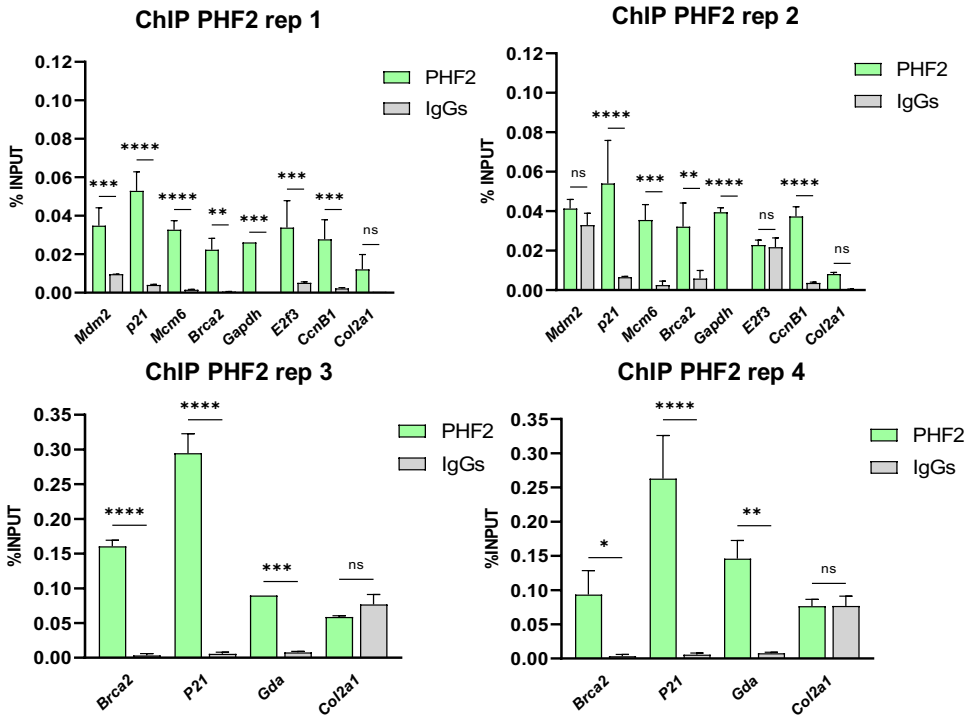


Figure R1. ChIPs of PHF2 in NSCs were analyzed by qPCR at promoter regions defined previously as targets of PHF2 before sequencing. The promoter of *Col2a1* gene was used as negative control. Results are the mean of three replicates. Error bars represent SD. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

We then analyzed PHF2 binding patterns within repetitive regions included in the TEtranscripts annotation (227), which contains over 3.7 million repeats organized into three levels: Class, Family, and Repeat group. To validate the heterochromatic nature of the classes of repeats of interest, we performed ChIP-seq for H3K9me3 and H3K4me3 histone marks in NSCs and compared them to PHF2 ChIP-seq. As expected, these sequences showed depletion for H3K4me3 and enrichment for H3K9me3 (Figure R2).

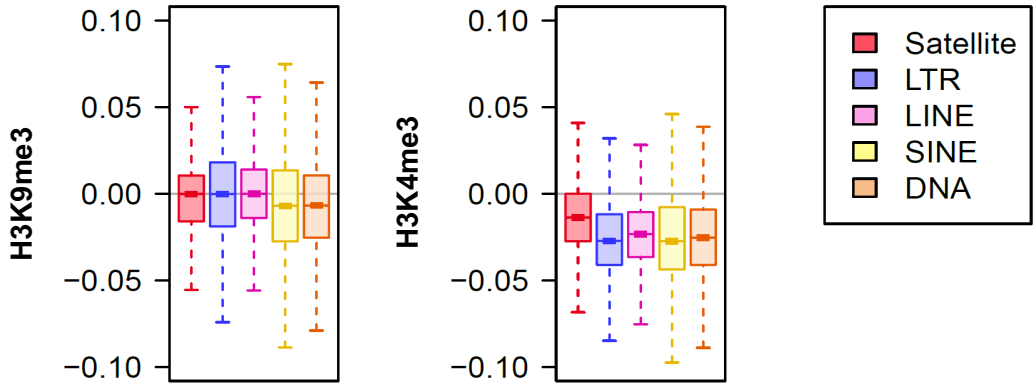


Figure R2. ChIP of H3K9me3 and H3K4me3 were done in NSCs and their enrichment and exclusion of the TEtranscripts annotation was calculated.

Our results indicate that PHF2 binding sites were enriched at specific repetitive elements such as satellites and long interspersed nuclear elements (LINEs) (Figure R3).

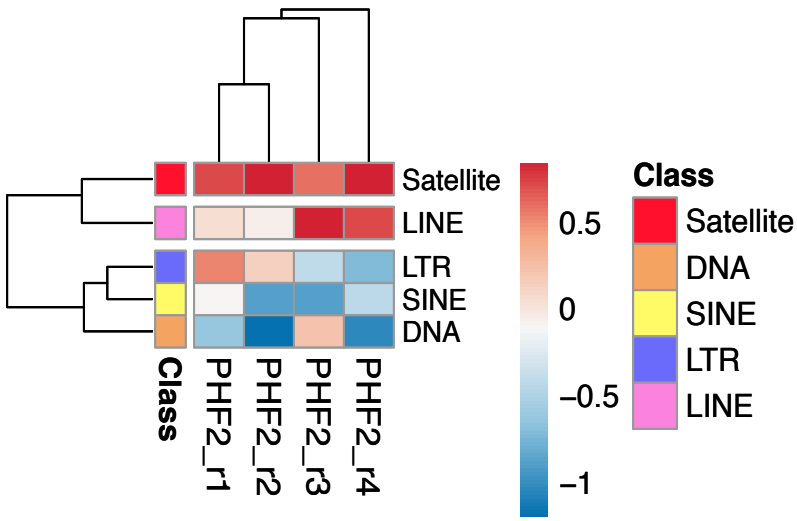


Figure R3. Heatmap showing enrichment or depletion of PHF2 input-subtracted ChIP-seq signal in repetitive elements classes of Satellite, DNA, SINE, LTR and LINE. Four biological replicates were done.

1.2 PHF2 is enriched in pericentromeric satellite repeats

Given PHF2's reported association with centromeric regions (143), we focused on the presence of PHF2 at various satellite repeats as described by the RepeatMasker tool. The results presented in the following Figure R4 show PHF2 enrichment at the major satellite (GSAT_MM), which correlated with low levels of H3K9me2 (the histone mark targeted

by PHF2), as determined by previously published H3K9me2 ChIP-seq data (5). This also correlated with high levels of H3K9me3, as expected for a heterochromatic region.

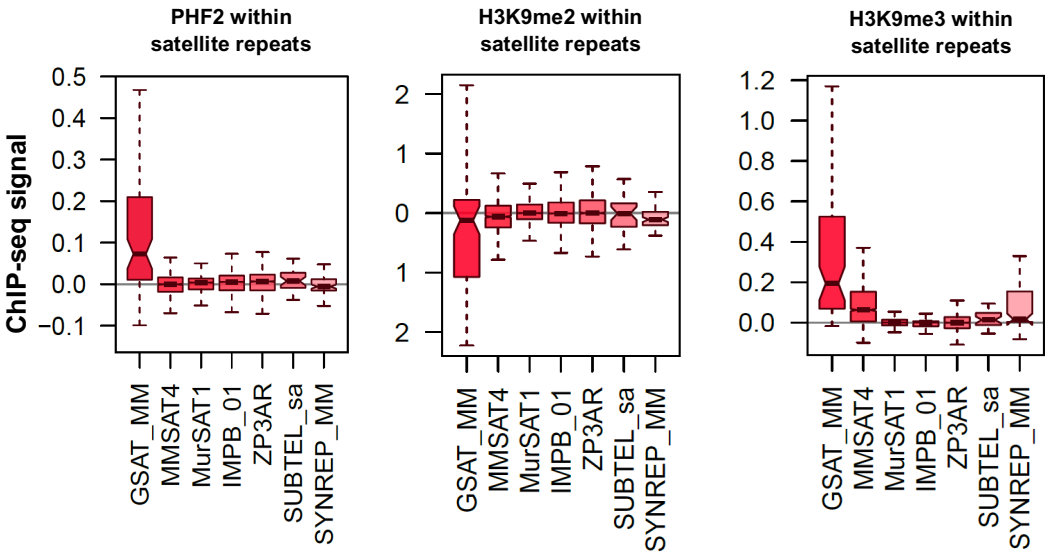


Figure R4. Boxplot depicting PHF2, H3K9me2 and H3K9me3 input-subtracted ChIP-seq signal enrichment at seven groups of Satellite repeats. GSAT_MM corresponds to MajorSat. PHF2 and H3K9me3 are enriched and H3K9me2 shows exclusion of this group. Four replicates were used for PHF2, and two replicates for H3K9me3 and H3K9me2.

To further confirm the genomic binding or proximity of PHF2 to satellites, we performed PHF2 ChIP followed by qPCR (ChIP-qPCR) targeting major and minor satellite repeats on chromosome 16, one of the best-mapped regions in the mouse genome. Our findings demonstrated significant enrichment of PHF2 at both major and minor satellites, albeit

lower than certain transcription start sites (TSS) previously identified as PHF2 targets, while no enrichment was observed at the *Col2a1* promoter region, used as a negative control (Figure R5).

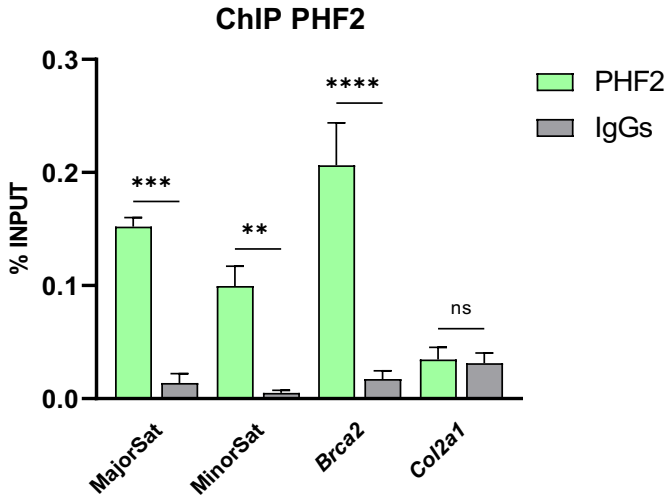


Figure R5. ChIP of PHF2 was analyzed by qPCR at the indicated genome regions. Data from qPCR were normalized and expressed as percentage of input. Results are the mean of four biologically independent experiments. Errors bars represent SD. **p<0.01; ***p<0.001; ****p<0.0001.

To broaden our conclusions, we analyzed the location enrichment of PHF2 at heterochromatic genomic regions in immortalized mouse fibroblasts (NIH3T3) using PHF2 ChIP-seq assays, obtaining similar results (Figure R6).

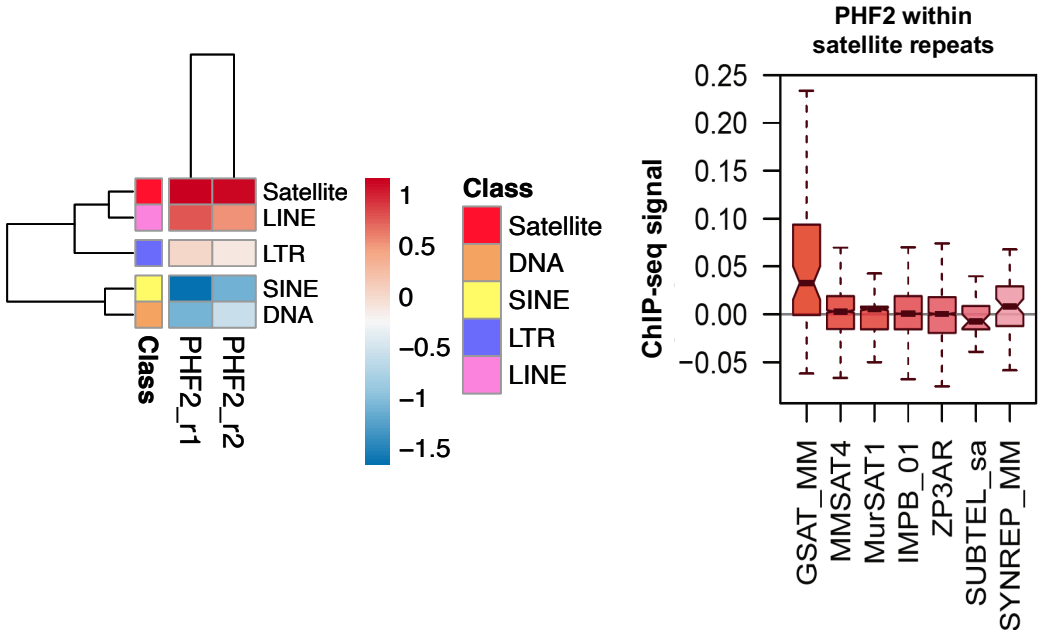


Figure R6. NIH3T3 cells were used instead of NSCs this time and the same analysis as in figures R6 and R7 were done. Results are from two biologically independent replicates.

In summary, these findings provide compelling evidence that PHF2 is enriched in centromeric satellite repeats, supporting its potential involvement in heterochromatin-associated processes.

1.3 PHF2 maintains heterochromatic silencing

To understand the potential role of PHF2 on heterochromatin, we depleted PHF2 in NSCs by employing lentivirus expressing shRNA specifically targeting PHF2, that resulted in

a partial reduction in PHF2 mRNA levels (as shown in Figure R7 left panel) as well as protein expression (Figure R7, right panel) while leaving the expression of other family members, such as PHF8, unaffected (Figure R7).

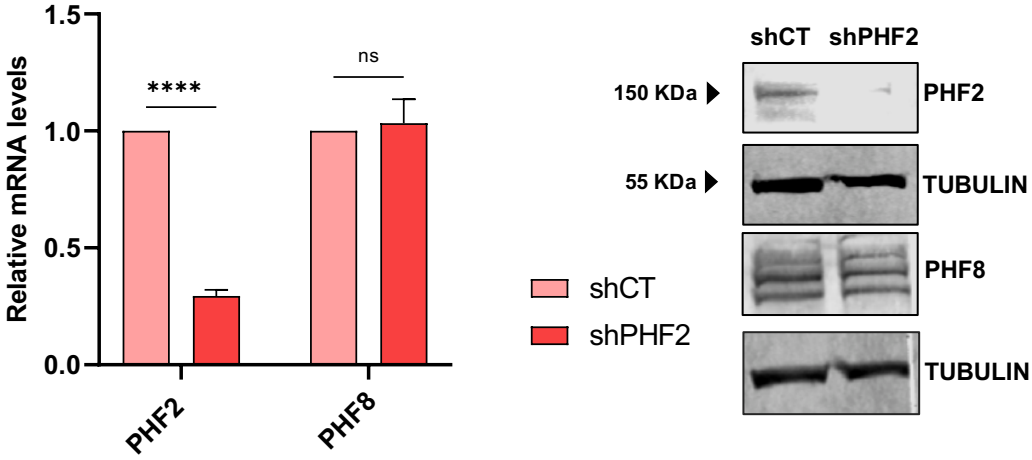


Figure R7. Left panel depicts relative transcription levels of the *Phf2* gene in NSCs after lentiviral infection and puromycin selection which shows a clear downregulation of its expression. Right panel shows PHF2's immunoblots in NSCs control (shCT) and depleted (shPHF2) conditions. Its family member PHF8 was used as negative control. TUBULIN immunoblot was used as protein load control.

Given the crucial role of heterochromatin stability in cell division and development, and based on previous studies, we speculated that PHF2 might also contribute to maintaining heterochromatin integrity. To explore that, we conducted qPCR assays to evaluate the expression levels of various heterochromatic repeats in both control and PHF2-depleted

neural stem cells (NSCs). Our results revealed a significant increase in the transcription of major and minor satellite repeats upon PHF2 depletion, while the expression of the negative control gene (*Gda*) remained unchanged. Additionally, we observed a slight increase in the expression of analyzed interspersed repetitive elements (LINES and SINES), though to a lesser extent (Figure R8).

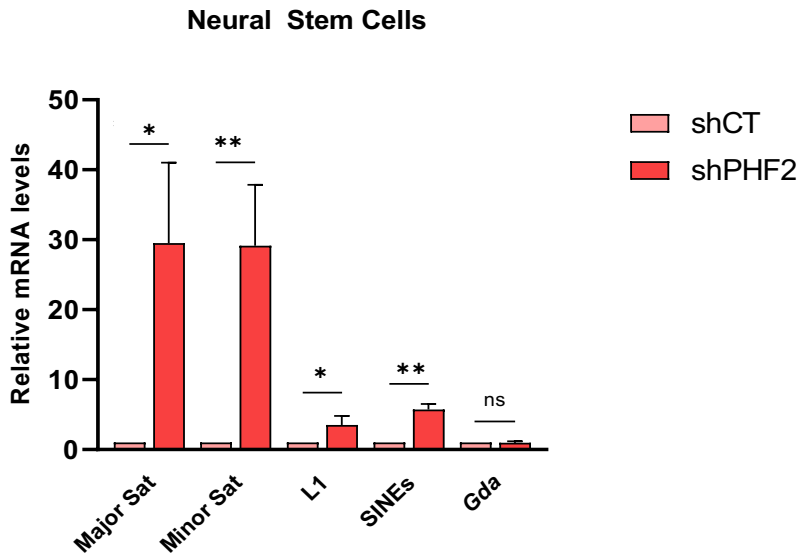


Figure R8. Bar plot depicting relative mRNA transcription levels of selected repetitive elements in NSCs control and PHF2-depleted. A striking mRNA accumulation coming from these elements is observed. The gene *Gda* was used as negative control. Results are from three biologically independent experiments. Error bars show SD. *p<0.05; **p<0.01.

To further investigate the impact of PHF2 depletion on heterochromatic repeat expression, we extended our analysis to immortalized mouse embryonic fibroblast NIH3T3 cells. Here, we noted an induction of pericentromeric repeats transcription; however, the overall expression levels were consistently lower compared to those observed in NSCs (Figure R9).

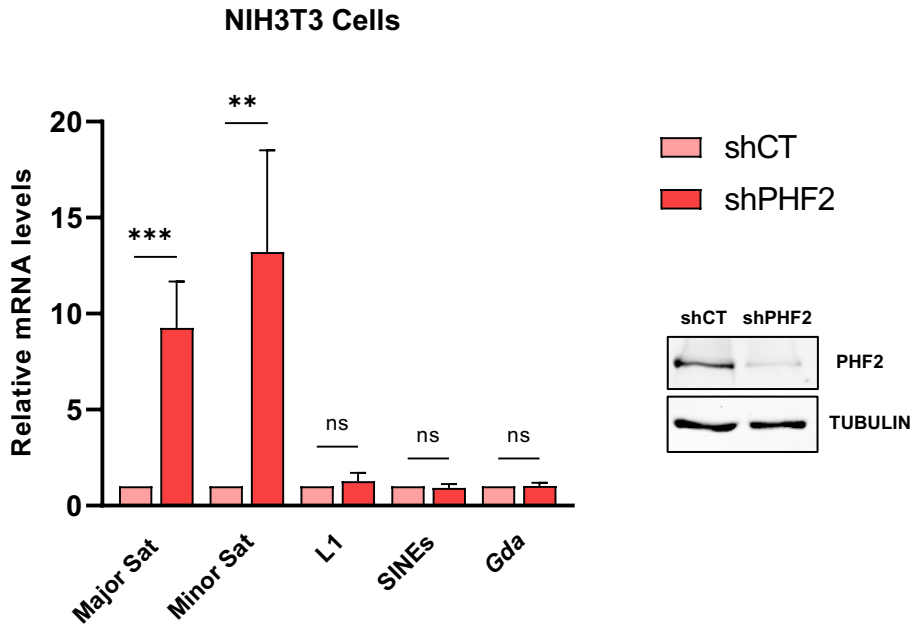


Figure R9. The same experiment than R10, but a different cell line (NIH3T3) was used (left). Right panel shows PHF2's immunoblots in NIH3T3 control (shCT) and depleted (shPHF2) conditions. TUBULIN immunoblot was used as protein load control. Results are from three biologically independent experiments. Error bars show SD. **p<0.01; ***p<0.001.

1.4 PHF2 depletion impacts PcH organization

To explore the impact of PHF2 depletion on the arrangement of pericentromeric heterochromatin (PcH), we assessed the size of PcH clusters, known as chromocenters, in mouse cells (232,233). We achieved this by quantifying the volume of DAPI-stained DNA and H3K9me3-marked PcH foci in both control and PHF2-depleted NSCs (Figure R10) and NIH3T3 cells (Figure R11). Remarkably, we noted an enlargement in the volume of chromocenters, indicating changes in PcH organization following PHF2 depletion. These observations provide additional evidence supporting the role of PHF2 influencing PcH organization.

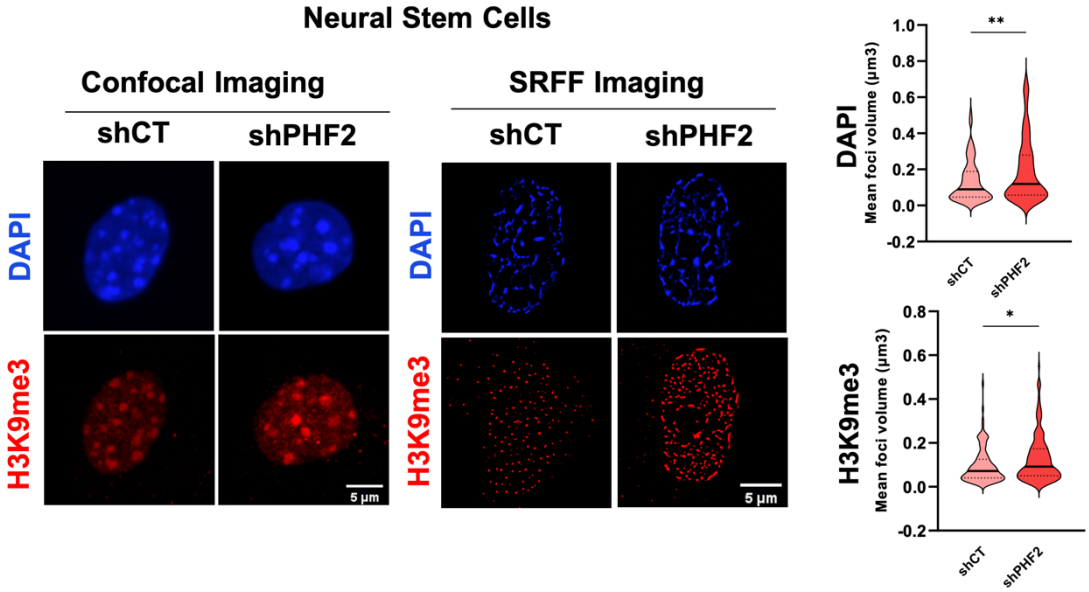


Figure R10. Immunofluorescence experiment in shCT and shPHF2 NSCs. Cells were stained with H3K9me3 antibody and DAPI, then confocal (left panel) and super-resolution (middle panel) images were acquired. Foci volume was quantified for both DAPI and H3K9me3 channels from n=50 cells and represented in violin plots (right panel) *p<0.05; **p<0.01.

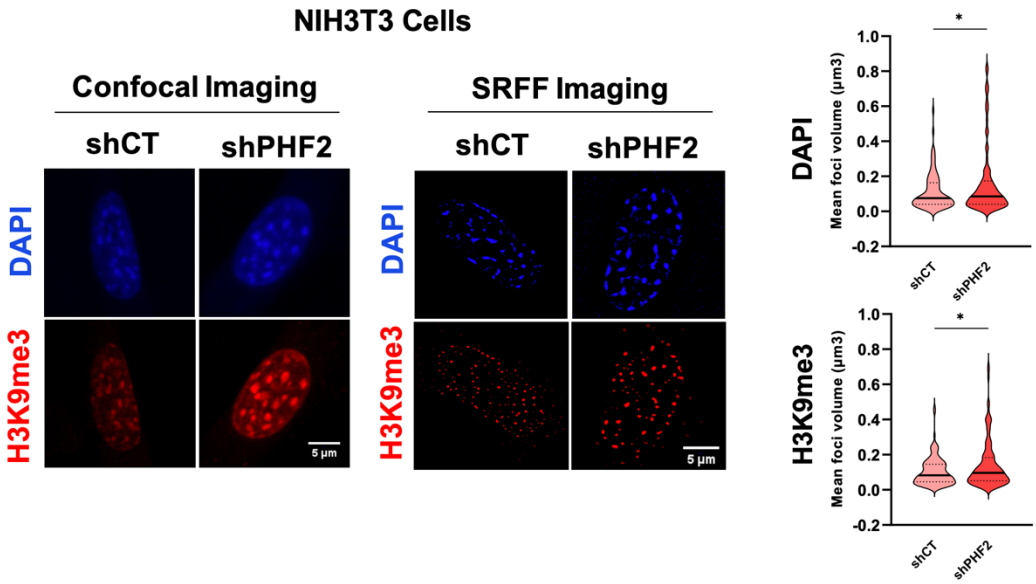


Figure R11. The same than in figure R12, but NIH3T3 cells were used. Identical outcome was observed in another cell line. *p<0.05.

1.5 PHF2 balances H3K9me3 levels

To delve into how PHF2 preserves heterochromatin stability, we explored the effect of PHF2 depletion on H3K9me3 levels. We conducted duplicate H3K9me3 ChIP-seq analyses in shCT and shPHF2 NSCs.

1.5.1 H3K9me3 changes at promoters upon PHF2 depletion

Initially, we assessed global changes in chromatin methylation by dividing the genome into 15 kb bins. This analysis unveiled that PHF2 depletion resulted in stronger H3K9me3 signals in 168 bins and weaker signals in 22 bins compared to control cells

(Figure R12, left panel). Additionally, we noted increased H3K9me3 levels around PHF2 binding sites (Figure R12, right panel).

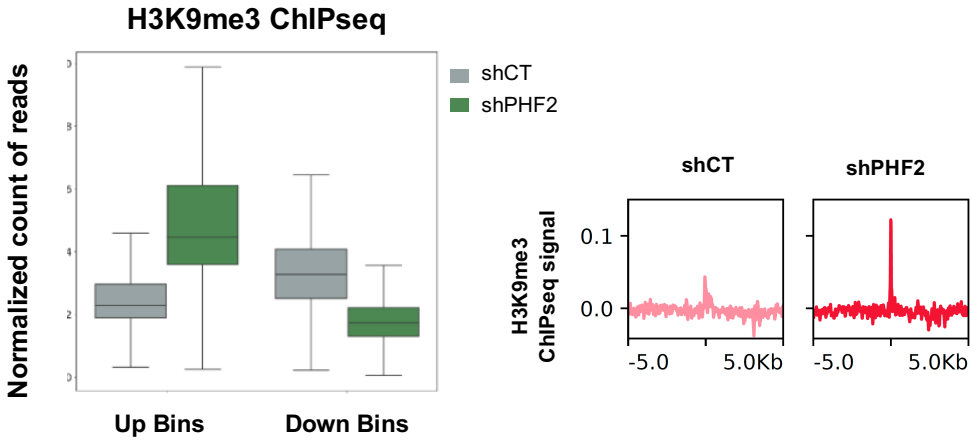


Figure R12. H3K9me3 reads distributed in bins that gained or lose signal. Upon PHF2 depletion we see an overall increase in H3K9me3 (left panel) as well as an increase in PHF2 specific binding sites (right panel). Results from two independent replicates.

These observations corroborate with H3K9me3 immunostaining assays, indicating an overall elevation in bulk H3K9me3 methylation upon PHF2 depletion as showed in Figures R10 and R11, as well as in previous work from the lab (5). Intriguingly, the bins exhibiting increased H3K9me3 levels upon PHF2 depletion were notably enriched at TSS. (Figure R13).

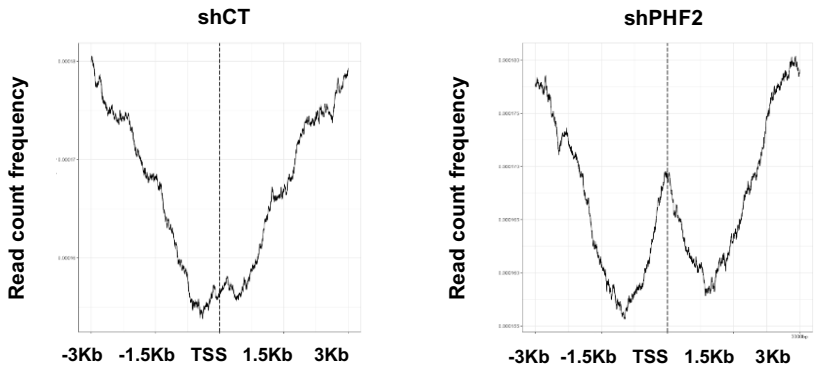


Figure R13. H3K9me3 ChIP-seq read count frequency around genomic TSS locations in shCT and shPHF2 NSCs. A striking increase in H3K9me3 is observed upon PHF2 depletion. Results from two independent replicates.

Moreover, gene ontology (GO) analysis of these regions revealed their involvement in cell cycle, DNA transcription, and other essential processes associated with cell proliferation and gene expression (Figure R14).

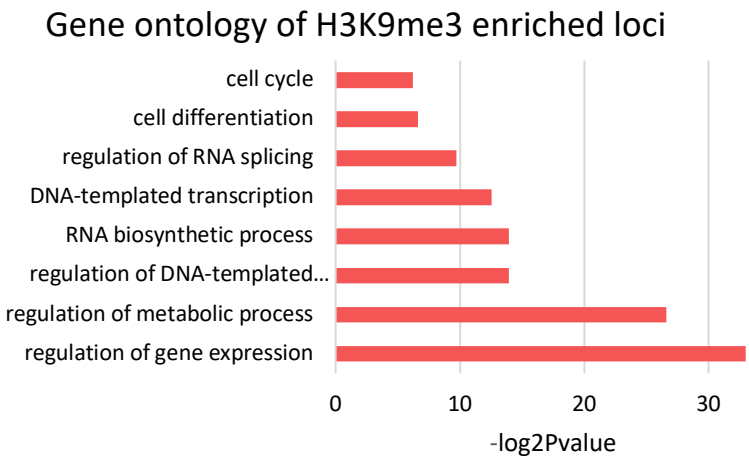


Figure R14. Gene ontology analysis showing biological process of the H3K9me3 enriched loci upon PHF2 depletion in NSCs using the Mus musculus genome (mm10) as a reference.

These findings underscore the role of PHF2 in maintaining low levels of H3K9me3 at regions critical for cell proliferation, as previously indicated (5).

1.5.2 H3K9me3 changes in satellites upon PHF2 depletion

However, a distinct scenario emerged when we investigated the satellite repeats, notably the major satellite. Here, we observed a reduction in H3K9me3 ChIP-seq levels in the satellite groups analyzed in figure R7, specifically GSAT_MM (Figure R15). These changes were confirmed by qPCR assays in the two individual ChIP-seq experiments (Figure R16).

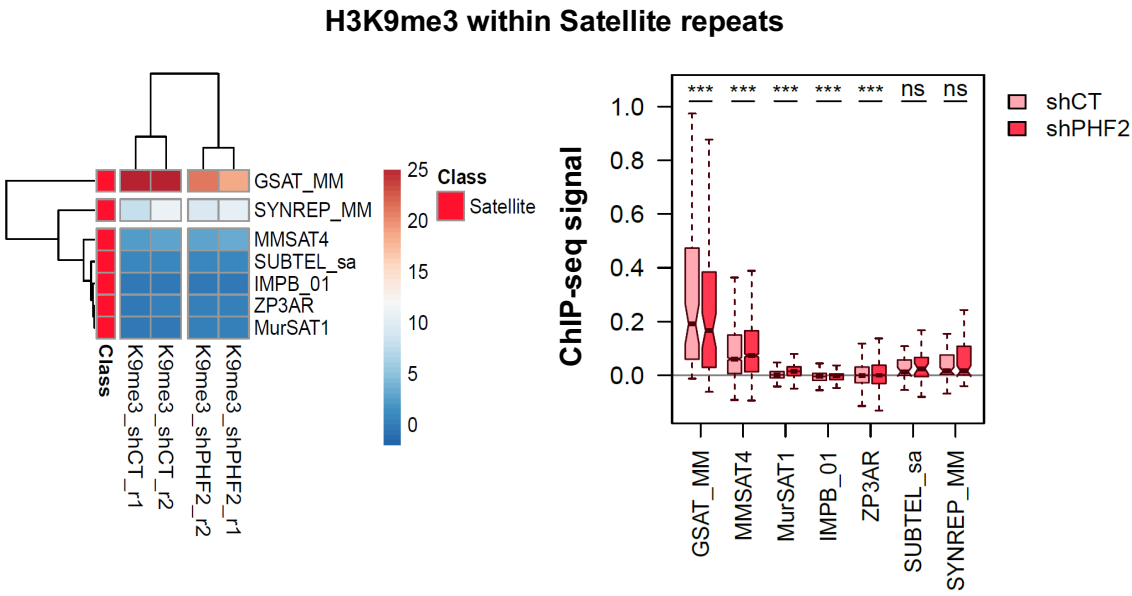


Figure R15. Heatmap showing enrichment or depletion of PHF2 input-subtracted ChIP-seq replicate-separated signal in repetitive elements groups of satellites (left panel). Two replicates were merged, mapped in the same satellite groups, quantified and box plotted (right panel). A decrease in GSAT_MM (MajorSat) was observed. ***p<0.001.

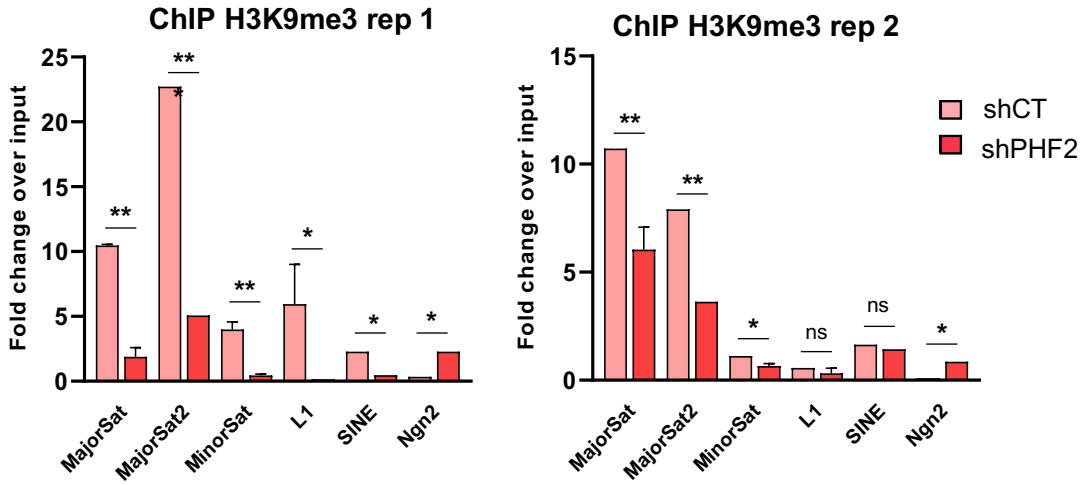


Figure R16. ChIP of H3K9me3 in control (shCT) and PHF2 depleted (shPHF2) NSCs was analyzed by qPCR at the major and minor satellites, albeit other repetitive elements. Data from qPCR were normalized to the input, IgGs subtracted, and expressed as fold enrichment. *Ngn2* was used as a control. Error bars represent SD. *p<0.05; **p<0.01.

Given the association between H3K9me3 and HP1α, we ought to determine if our data was accompanied by a corresponding decrease in this protein's binding (Figure R17), offering a molecular basis for the observed activation of PcH transcription, and we confirmed these results.

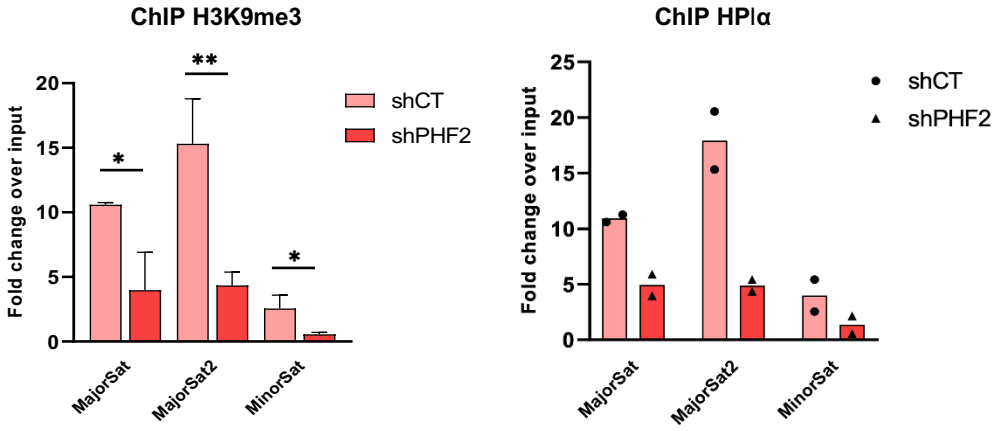


Figure R17. ChIPs of H3K9me3 (left) and HPIα (right) in control (shCT) and PHF2 depleted (shPHF2) NSCs were analyzed by qPCR at the major and minor satellites. Data from qPCR were normalized to the input, IgGs subtracted, and expressed as fold enrichment. Three biologically independent experiments were used for H3K9me3 and two for HPIα. Error bars represent SD. *p<0.05; **p<0.01.

1.5.3 H3K9me3 specifically decreases in MajorSat repeats

To confirm the decrease in H3K9me3 at heterochromatin regions containing major satellites upon PHF2 depletion, we employed modular fluorescence complementation sensors. These sensors enabled us to specifically detect H3K9me3 signals at the major satellite genomic site in live cells, as depicted in Figure R18 (Also see Methods 2.2.2.6) (213,214).

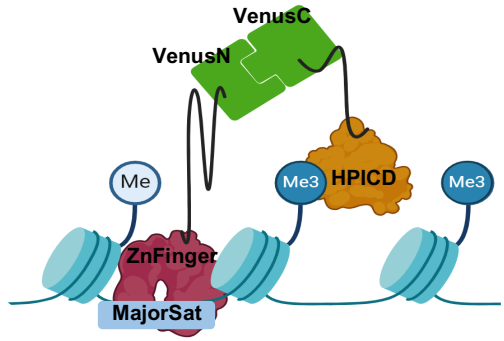


Figure R18. Schematic model of modular fluorescence complementation sensors.

Live-cell imaging of both control and PHF2-depleted cells unveiled a reduction in H3K9me3 levels at major satellite regions compared to the rest of the cellular levels (Figure R19). Moreover, we noted that methylated foci exhibited a larger volume in PHF2-depleted cells compared to control cells, consistent with the observations in Figures R10 and R11.

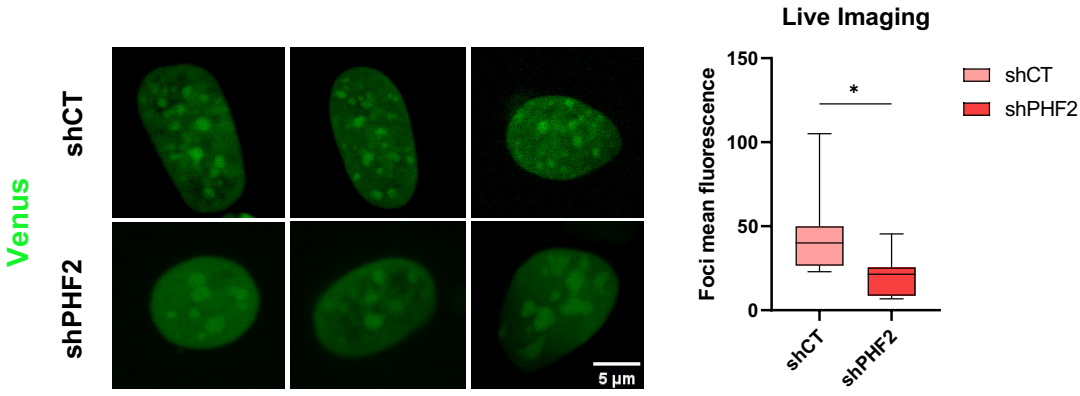


Figure R19. Confocal live cell images of NIH3T3 shCT and shPHF2 cells transfected with modular fluorescence complementation vectors. Foci mean fluorescence was quantified for Venus fluorophore from n=50 cells. Data acquired from three biologically independent experiments. *p<0.05

1.5.4 H3K9me3 changes at PcH boundaries upon PHF2 depletion

The genomic boundaries are important regions that play a critical role in maintaining heterochromatin stability. Thus, upon observing a loss in methylation in the repeats, we decided to analyze the H3K9me3 levels near these boundaries. For this assessment, we divided the mouse genome into 10 Kb bins and examined the changes in H3K9me3 levels across the 5-megabase (Mb) region from the centromeric end of each mouse chromosome in control and PHF2-depleted samples. Interestingly, contrary to the observations at the satellite repeats (described in Figures R19-R22), we detected a significant increase in

H3K9me3 levels, particularly within the first 500 Kb next to the chromosome centromeres (3-3.5 Mb) upon PHF2 depletion (Figure R20).

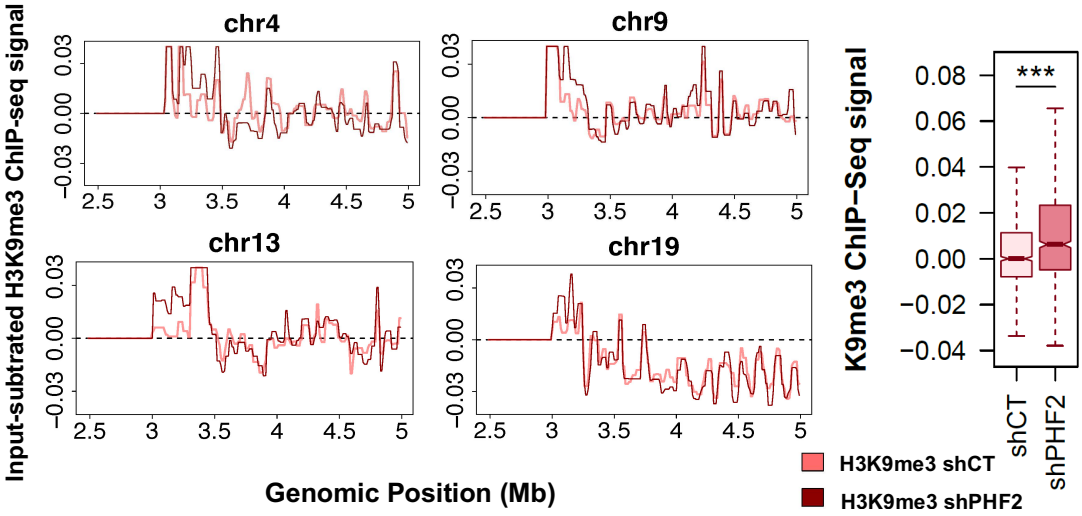


Figure R20. Line plot depicting input-subtracted ChIP-seq signal of H3K9me3 in shCT and shPHF2 NSCs at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 4, 9, 13 and 19 are shown as a representative sample, but all of them were considered for the boxplot statistical calculation. ***p<0.001

Our findings suggest that the depletion of PHF2 triggers an increase in H3K9me3 within PcH boundaries (Figure R21) and other genomic regions, especially TSS (Pappa et al., and Figure R13), potentially leading to the destabilization of heterochromatin, particularly if the availability of heterochromatin components is limited, thereby facilitating increased transcription of repetitive elements.

1.6 HPIα confirms H3K9me3 disbalance upon PHF2 depletion

Since H3K9me3 is recognized by HP1α, which facilitates silencing through a read-write mechanism (234,235), we decided to investigate the binding of HP1α in these boundary regions using ChIP-qPCR. The results indicated that, in line with the increase in H3K9me3 levels, the binding of HP1α was also elevated (Figure R21).

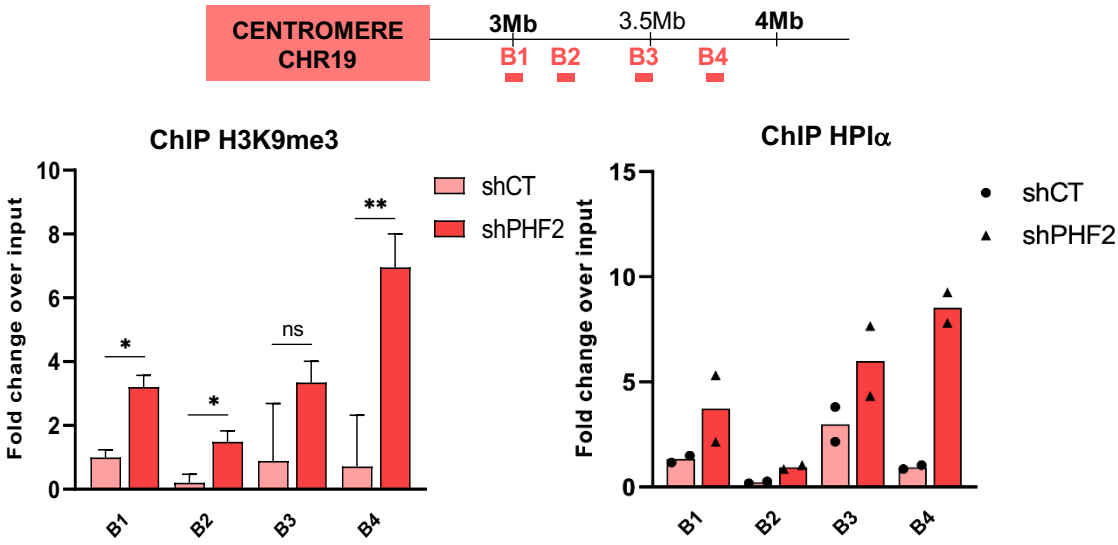


Figure R21. ChIPs of H3K9me3 (left) and HPIα (right) in control (shCT) and PHF2 depleted (shPHF2) NSCs were analyzed by qPCR at the selected locations of Chr19 boundaries (upper panel). Data from qPCR were normalized to the input, IgGs subtracted, and expressed as fold enrichment. Three biologically independent experiments were used for H3K9me3 and two for HPIα. Error bars represent SD. *p<0.05; **p<0.01.

Additionally, no significant changes in H3K4me3 levels were observed upon PHF2 depletion either in the vicinity of PcH or at the global level (Figure R22).

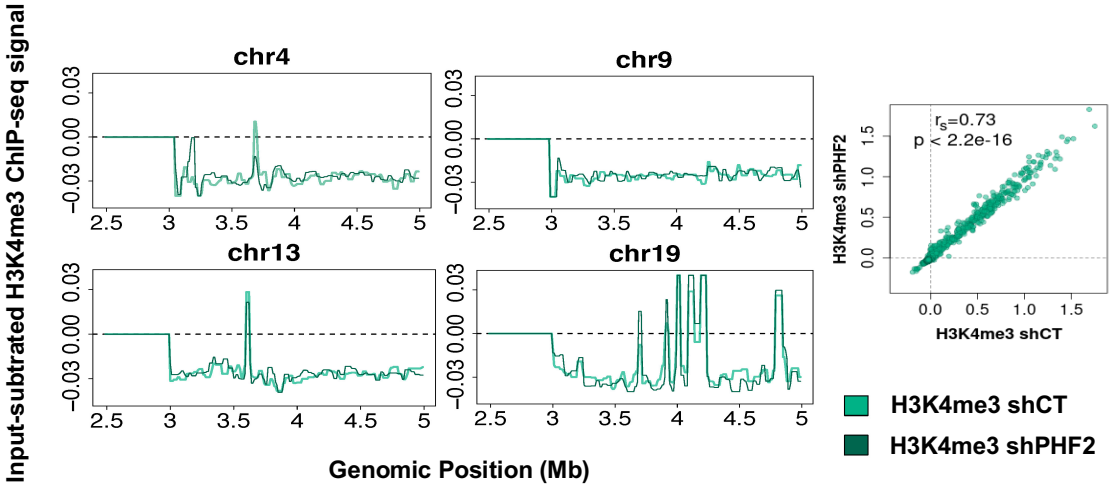


Figure R22. Line plot depicting input-subtracted ChIP-seq signal of H3K4me3 in shCT and shPHF2 NSCs at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 4, 9, 13 and 19 are shown as a representative sample, but all of them were considered for the scatter plot statistical calculation. ***p<0.001

These repetitive elements transcription may be caused by the lack of availability of heterochromatin components as previously mentioned. To test that, we focused on specific genes marked by H3K9me3 in control NSCs, which were subsequently found to be upregulated in the RNA-seq upon PHF2 depletion. To ascertain whether this increased

transcription following PHF2 depletion is mediated indirectly through the redistribution of heterochromatin components, we turned our attention to HP1α. ChIP-qPCR analysis of HP1α revealed a decrease in its recruitment to analyzed heterochromatic loci upon PHF2 depletion, while an increase was observed over the boundary regions (Figure R23).

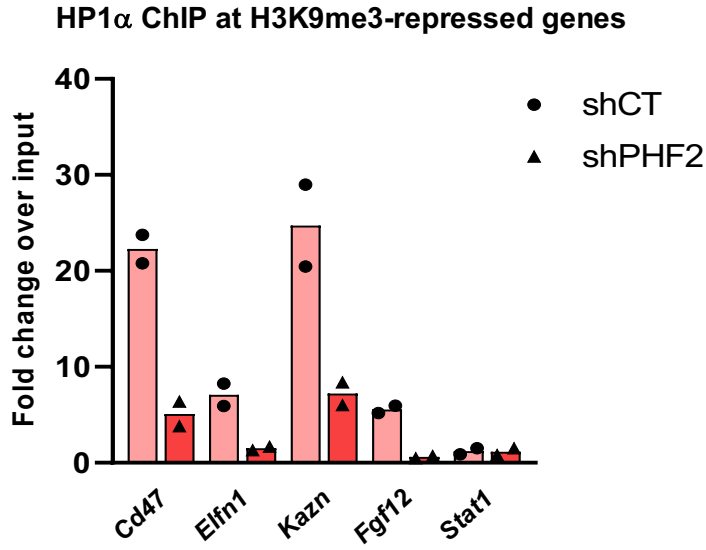


Figure R23. ChIP of HP1α in control (shCT) and PHF2 depleted (shPHF2) NSCs was analyzed by qPCR at the promoter regions of H3K9me3 repressed genes that were activated upon PHF2 depletion. The promoter region of *Stat1* was used as a negative control. Data from qPCR were normalized to the input, IgGs subtracted, and expressed as fold enrichment. Two biologically independent experiments were used.

This observation suggests a potential reorganization of HP1α, and likely other heterochromatin components, upon PHF2 depletion.

1.7 PHF2 is enriched in PcH boundaries

After detecting alterations in the H3K9me3 pattern around the PcH, we opted to reanalyze the PHF2 ChIP-seq data to investigate the presence of PHF2 at these regions. Intriguingly, we observed a notable enrichment of PHF2 in a 5 Mb region from the centromeric end of each mouse chromosome, particularly prominent within the first 500 Kb adjacent to the centromere (3-3.5 Mb). Specifically, the ChIP-seq signal within 1,000 bins corresponding to these initial pericentromeric regions (3-3.5 Mb) of mouse chromosomes was significantly higher than within 1,000 random genomic windows (Figure R24).

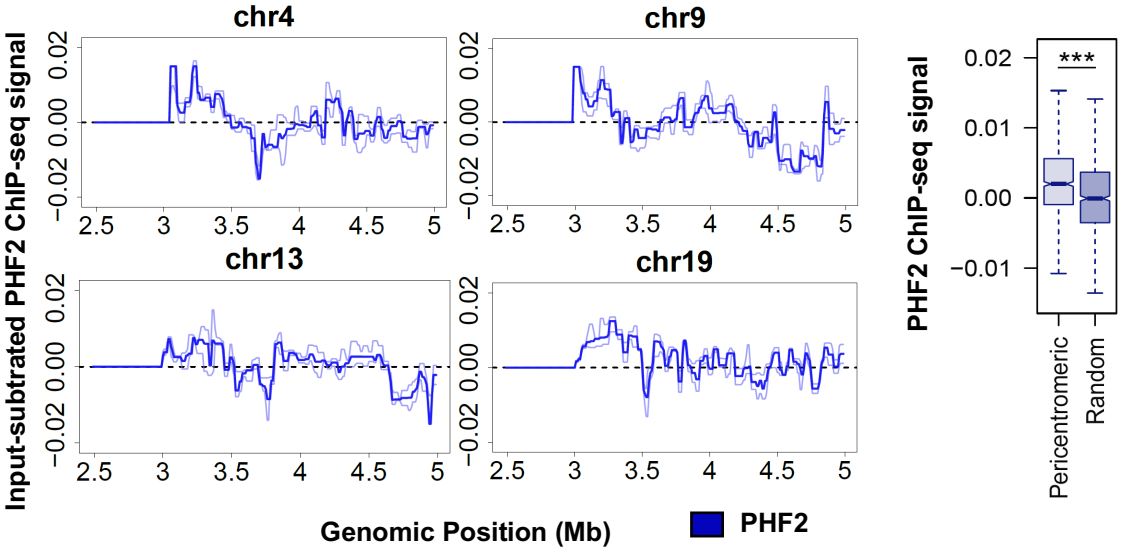


Figure R24. Line plot depicting input-subtracted ChIP-seq signal of PHF2 in NSCs at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 4, 9, 13 and 19 are shown as a representative sample, but all of

them were considered for the boxplot statistical calculation vs 1000 random regions of the genome. ***p<0.001

To further confirm this result, the enrichment of PHF2 in the boundaries region was assessed by ChIP-qPCR with the primers used in Figure R24 for the chromosome 19 (Figure R25).

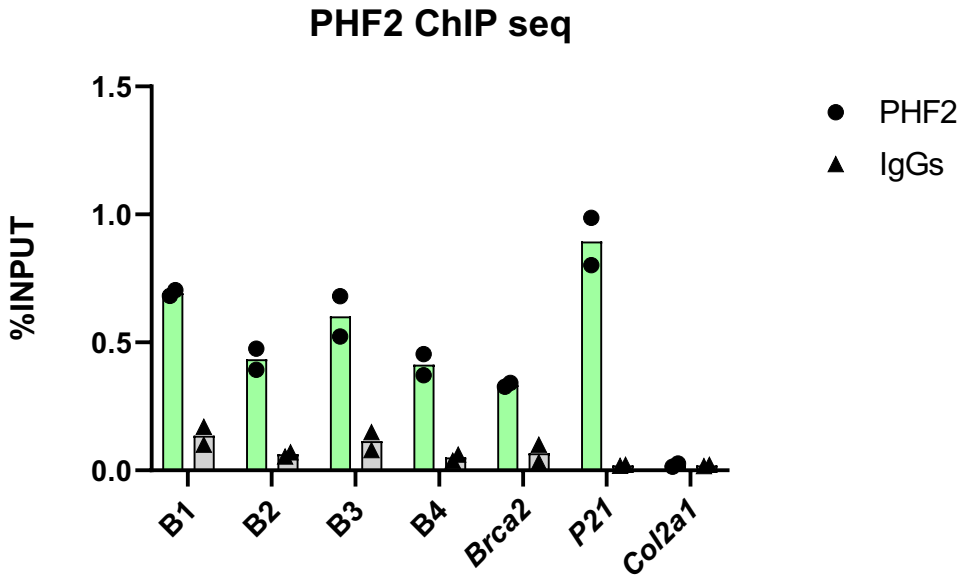


Figure R25. ChIP PHF2 in NSCs was analyzed by qPCR at the selected locations of Chr19 boundaries. Data from qPCR were normalized and expressed as percentage of input. The promoter region of *Brca2* and *P21* genes were used as positive control, and the promoter region of *Col2a1* was used as negative control. Two biologically independent experiments were used.

1.8 PHF2 is close to the chromocenter and nuclear lamina

PHF2 immunostaining assays and super-resolution microscopy helped us discern the exact location it occupies inside the cell nucleus. Results showed that it was found to be in close proximity to, but not colocalized with, chromocenters (Figure R26).

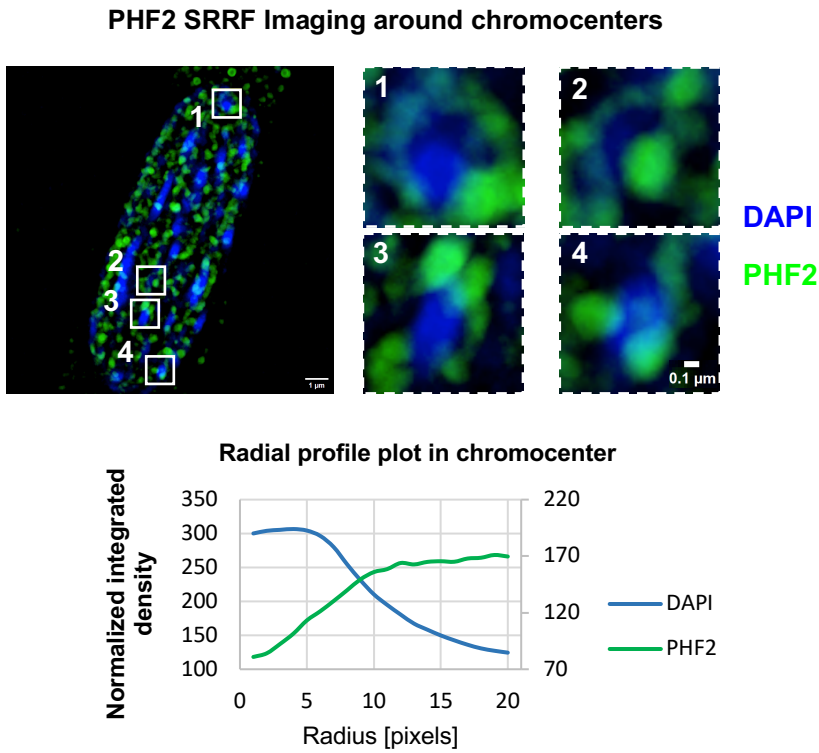


Figure R26. Super-resolution imaging where cells were stained with PHF2 antibody (green) and DAPI (blue). Several nuclear foci were selected for a higher magnification panel (right 1-4) where PHF2 appears surrounding chromocenters. Radial profile intensity around foci ROI was quantified and plotted for each channel (bottom panel).

Indeed, PHF2 was found to be closer to the chromocenter and nuclear lamina compared to other members of the HDM7 family, including PHF8 and KIAA1718, used as a control (Figure R27).

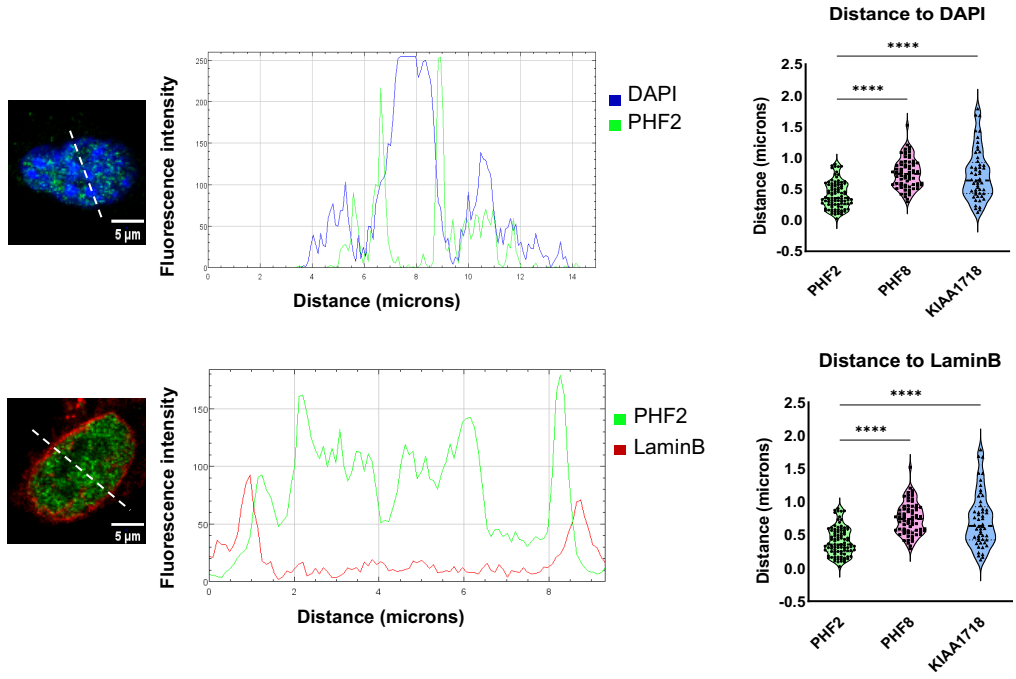


Figure R27. Immunofluorescence experiment where cells were stained with PHF2 antibody (green) and either Lamin antibody (Red) or DAPI (blue). Then, confocal images were acquired. Intensity across a nuclear section was quantified for both channels, represented in line plots (middle panel) and distance between a clear peak measured. The same was done with PHF8 and KIAA1718, and distance were represented in violin plots from n=50 cells (right panels). ****p<0.0001.

1.9 PcH boundary's transcription is affected upon PHF2 depletion

Given the established association between chromatin dynamics at heterochromatin boundaries and heterochromatin stability across various organisms (129,236–239) we sought to investigate the effect of PHF2 depletion on transcription within these regions. To achieve this, we reanalyzed previously published RNA-seq data from NSCs (5). Our analysis unveiled a reduction in transcription levels in the pericentromeric regions following PHF2 depletion (Figure R28), suggesting a potential direct or indirect involvement of PHF2 in maintaining transcriptional activity in these regions.

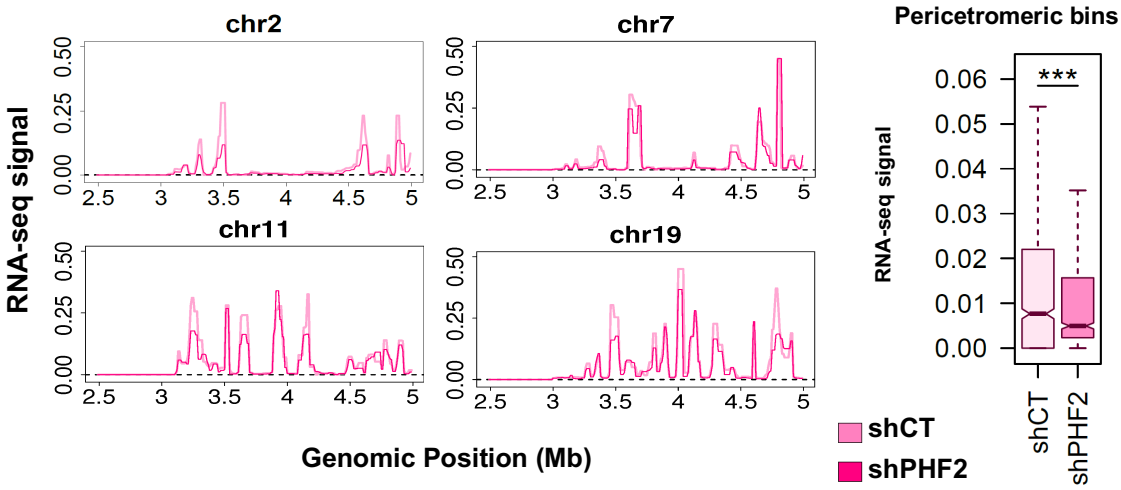


Figure R28. Line plot depicting RNA-seq signal of shCT and shPHF2 NSCs at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 2, 7, 11 and 19 are shown as a representative sample, but all of them were considered for the boxplot statistical calculation of shCT vs shPHF2. ***p<0.001

Notably, these regions exhibited a higher level of RNA Pol II binding compared to randomly selected regions (Figure R29), accompanied by a diminished level of histone H1 binding (Figure R30), indicative of heightened transcriptional activity relative to the genome-wide average.

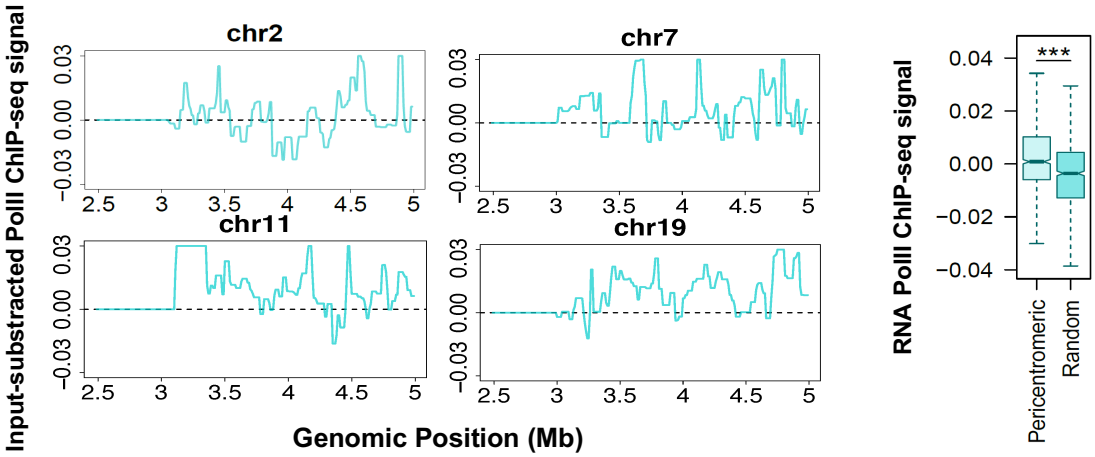


Figure R29. Line plot depicting input-subtracted ChIP-seq signal of RNA Pol II in NSCs at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 2, 7, 11 and 19 are shown as a representative sample, but all of them were considered for the boxplot statistical calculation vs 1000 random regions of the genome. ***p<0.001

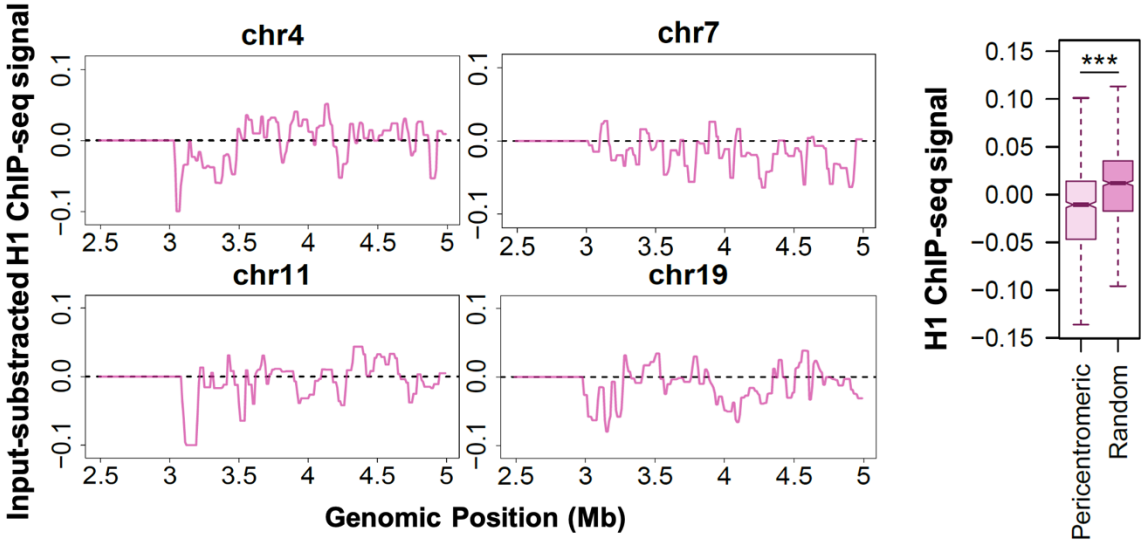


Figure R30. Line plot depicting input-subtracted ChIP-seq signal of H1 in NSCs at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 4, 7, 11 and 19 are shown as a representative sample, but all of them were considered for the boxplot statistical calculation vs 1000 random regions of the genome. ***p<0.001

Consequently, we observed a transcriptional decline upon PHF2 depletion (Figure R28), alongside the noted redistribution of heterochromatin components (HP1α) due to unbalanced H3K9me3 (Figure R23).

1.10 PcH boundaries comprehensive characterization

In order to understand the characteristics of the boundary regions, we thoroughly analyzed the genomic area 3-5 Mb downstream of the repeats. Initial assessments indicated the absence of significant G/C features in these regions (Figure R31).

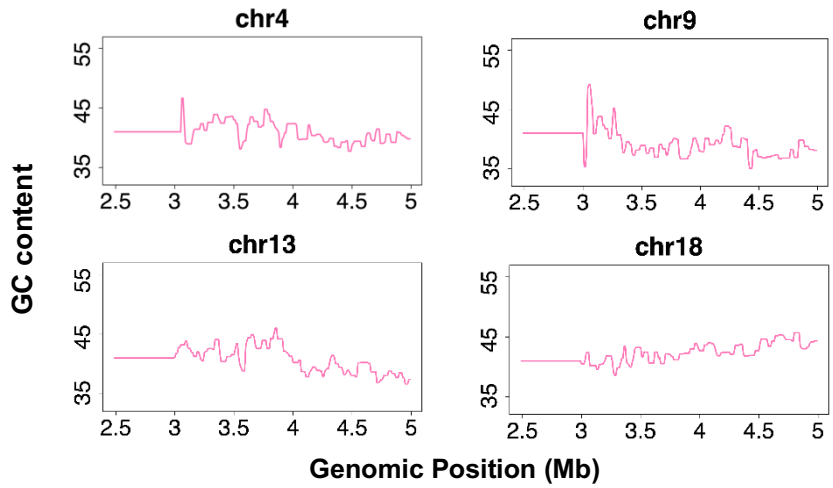


Figure R31. Line plot depicting GC content at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 4, 9, 13 and 18 are shown as a representative sample, but all of them were considered for the statistical calculation vs 1000 random regions of the genome, where no difference was observed.

However, it is well-documented that noncoding transfer RNA (tRNA) genes serve as boundary elements (129,238,240). Consequently, we examined the presence of tRNAs and noncoding RNAs (ncRNAs), revealing an enrichment of these genomic elements across the boundary regions (Figure R32).

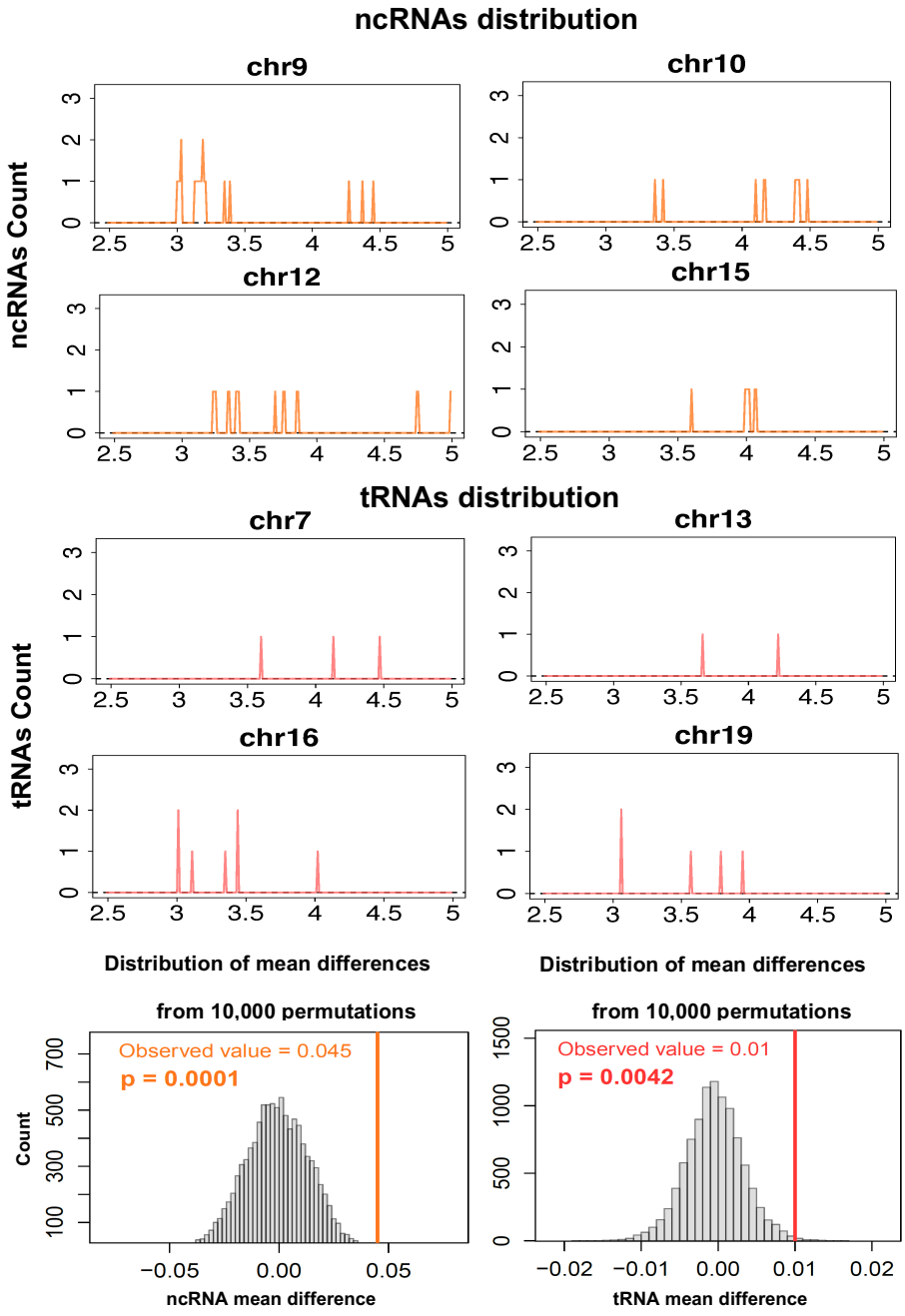


Figure R32. Line plot depicting ncRNAs (top) and tRNAs (middle) count at the boundary regions (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb

region. Chromosomes in the figure were selected as a representative sample, but all of them were considered for the statistical calculation from 10000 permutations of the genome (bottom).

Unfortunately, our attempts to evaluate the potential role of PHF2 in their expression were hindered by the lack of detection in the RNA-seq data.

Overall, these findings suggest that PHF2 plays a pivotal role in maintaining balanced levels of H3K9me3. This is crucial for stabilizing PcH, potentially through the regulation of transcription at the boundaries, and/or by modulating the distribution of heterochromatin components.

1.11 PHF2 regulates chromatin accessibility

To delve into how the depletion of PHF2 impacts heterochromatin stability, we conducted ATAC-seq assays on control and two sets of PHF2-depleted NSCs. These assays offer insights into chromatin accessibility. Our results indicated a slightly higher decrease in accessibility in regions affected by PHF2 depletion compared to those with increased accessibility (Figure R33), consistent with PHF2's function as an H3K9me2 demethylase.

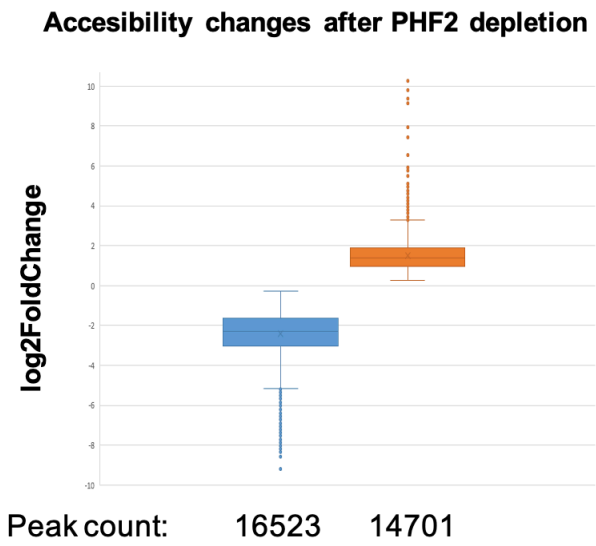


Figure R33. ATAC-seq was performed in shCT and shPHF2 NSCs. Then, box plot depicting number of peaks that gained or lose accessibility was represented. A higher number and higher negative Fold Change was observed in the "accessibility loss" group, as expected.

1.11.1 ATAC-seq changes at promoters upon PHF2 depletion

Further analysis revealed that the regions losing accessibility were notably enriched in promoters, especially at transcription start sites (TSS) (Figure R34) and associated with genes crucial for cell cycle progression and DNA transcription (Figure R35), aligning with PHF2's known role in NSCs (5).

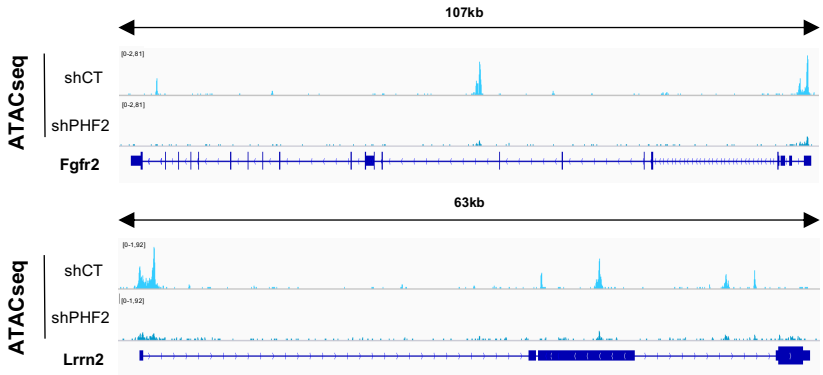


Figure R34. IGV captures showing ATAC-seq signal at specific genes. A clear decrease in accessibility around the promoter area of the genes is observed when PHF2 is depleted.

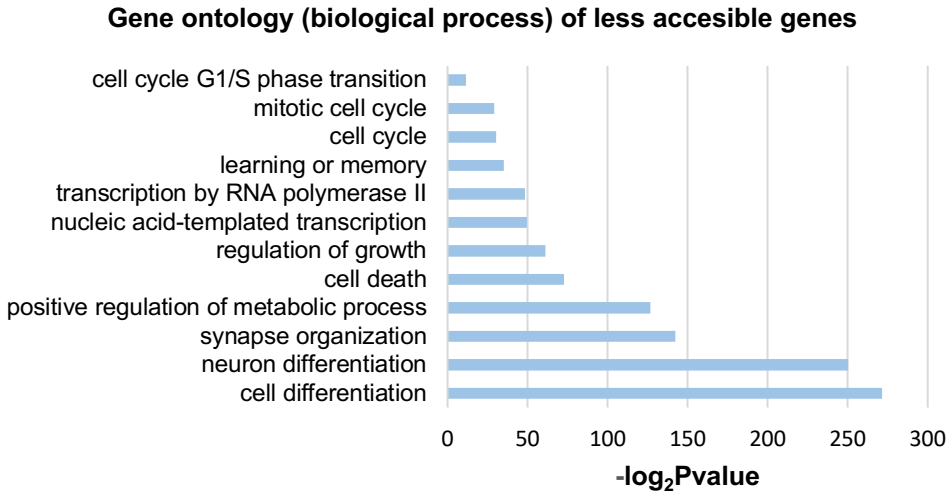


Figure R35. Gene ontology (biological process) analysis was performed with the genes that were significantly less accessible in shPHF2 condition than in shCT, aligning with PHF2's role in NSCs.

Conversely, regions gaining accessibility were predominantly found in intergenic regions (Figure R36), and regions where PHF2 binds showed reduced accessibility upon its depletion (Figure R37), indicating its role in maintaining accessibility at these sites, consistent with previous findings (5).

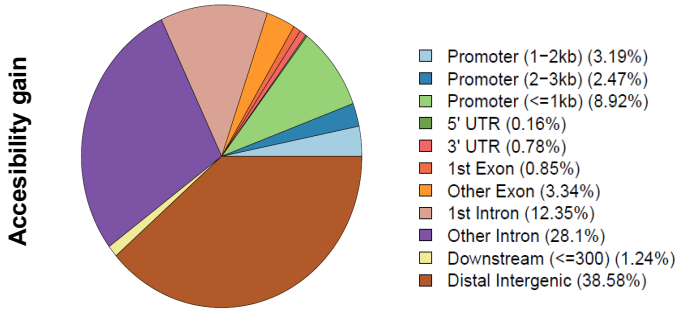


Figure R36. Genomic distribution of ATAC-seq peaks that showed increased accessibility. Most of them (38.58%) are in intergenic regions.

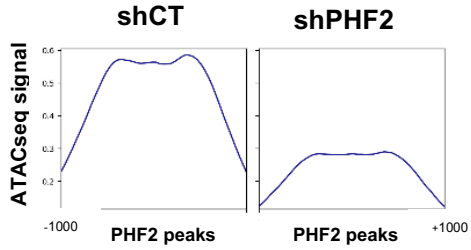


Figure R37. ATAC-seq signal around PHF2 ChIP-seq peaks was plotted with 1Kb up and downstream of each peak. A decrease in accessibility around PHF2's binding sites is observed, supporting PHF2's role as a H3K9me2 demethylase.

1.11.2 ATAC-seq changes at PcH and satellites upon PHF2 depletion

Considering PHF2's potential involvement in heterochromatin destabilization, we focused on PcH, specifically examining global changes in accessibility at satellite repeats. Intriguingly, our analysis reveals a significant increase in accessibility at these repeats (Figure R38).

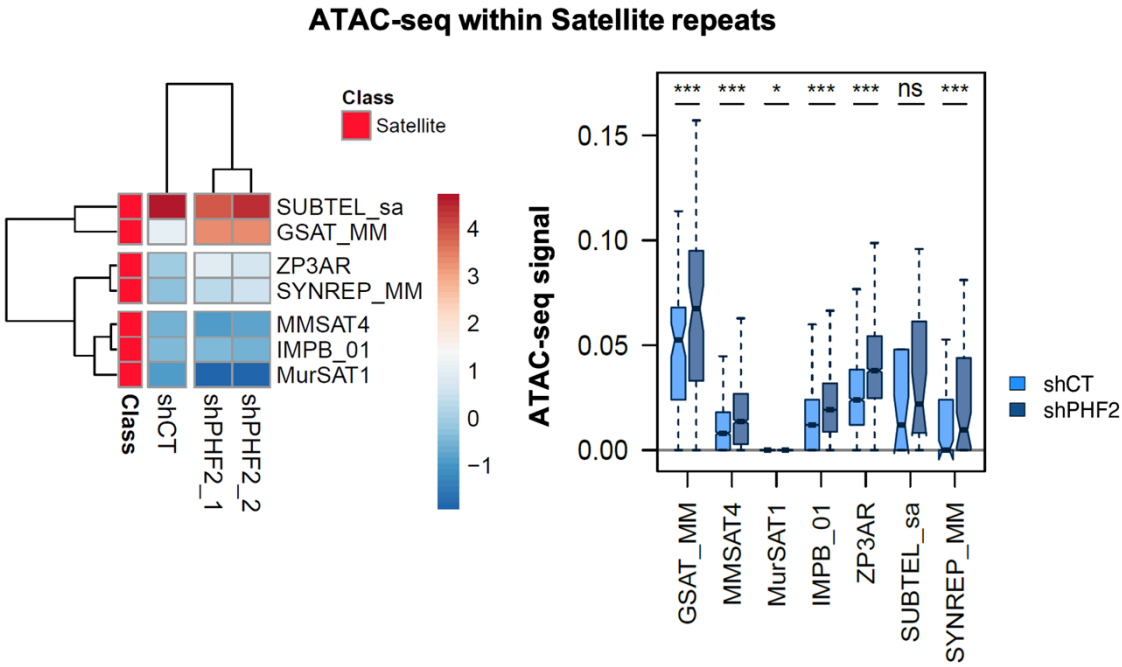


Figure R38. Heatmap showing ATAC-seq replicate-separated signal in repetitive elements groups of satellites (left panel). Two replicates were merged, mapped in the same satellite groups, quantified and box plotted (right panel). An increase in GSAT_MM (MajorSat), albeit other satellites was observed. *p<0.05; ***p<0.001.

1.11.3 ATAC-seq changes at boundaries upon PHF2 depletion

We also investigated accessibility changes across boundary regions of interest in control and PHF2-depleted samples. Interestingly, unlike the satellite repeats (Figure R38), we observed a slight but significant decrease in ATAC signal near chromosome centromeres upon PHF2 depletion (Figure R39), consistent with elevated H3K9me3 levels in this region (Figure R20).

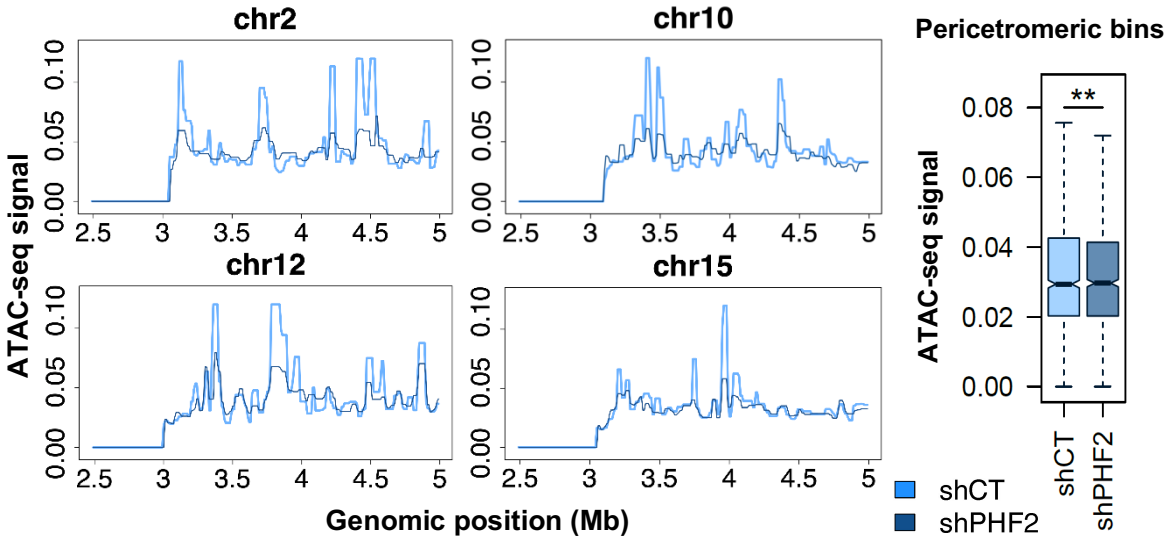


Figure R39. Line plot depicting ATAC-seq signal of shCT and shPHF2 NSCs at the boundary region (3-5Mb from the start of chromosome) and specifically at 3-3.5Mb region. Chromosomes 2, 10, 12 and 15 are shown as a representative sample, but all of them were considered for the boxplot statistical calculation of shCT vs shPHF2. **p<0.01

These findings suggest that PHF2 maintains accessibility at promoters it occupies but leads to increased accessibility at satellite repeats upon its depletion.

1.12 PcH stability relies on the PHD and JmjC domains within PHF2

To delve into the mechanism by which PHF2 regulates heterochromatin silencing, our focus was on pinpointing the specific domains within PHF2 responsible for the observed unscheduled transcription from PcH. PHF2 contains two distinct protein domains: the PHD domain, recognized for its H3K4me3-binding capability, and the JmjC domain, which exhibits demethylase activity (49). Additionally, a recently identified charged region within the IDR has been associated with nuclear condensate formation, a critical factor in transcriptional control (210). To evaluate the involvement of these domains in heterochromatic repeat transcription, we first confirmed their comparable expression levels in NIH3T3 cells (Figure R40).

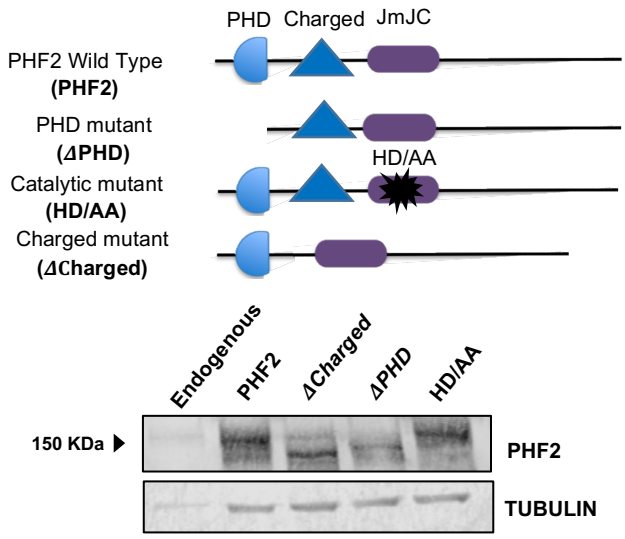


Figure R40. Schematic view of PHF2 wild type protein structure, and the generated mutants during this doctoral thesis (top). Each of them was transfected in NIH3T3 cells,

total protein extracted and immunoblotted using anti PHF2 antibody. Tubulin was used as load control.

1.2.1 Satellite transcription induction

To further characterize these mutants, and test out hypothesis about the heterochromatic repeats transcription, we overexpressed of the PHD (ΔPHD) and catalytic (HD/AA) PHF2 mutants which led to increased satellite transcription, contrasting with the PHF2 wild type (WT) or ΔCharged mutant (Figure R41).

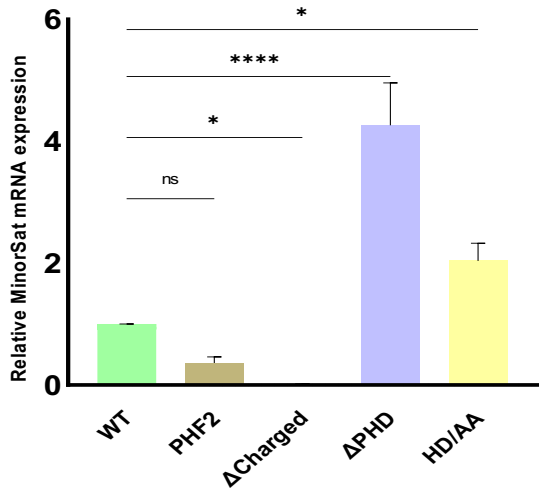


Figure R41. Bar plot depicting relative mRNA transcription levels of MinorSat in NIH3T3 cells control and transfected with either PHF2 WT, ΔCharged, ΔPHD or HD/AA mutant. A striking mRNA accumulation coming from these elements is observed when the ΔPHD or the HD/AA mutant are transfected. Results are from three biologically independent experiments. Error bars show SD. *p<0.05; ****p<0.0001.

1.2.2 Satellite transcription rescue

Following this, we conducted rescue experiments by introducing either PHF2 WT or specific PHF2 mutants into PHF2-depleted NSCs or NIH3T3 cells, and then evaluated the expression of satellite sequences. As illustrated, both PHF2 WT and the charged mutant within the IDR (ΔCharged) effectively restored normal satellite expression levels. Conversely, the mutants for catalytic and PHD domains failed to rescue transcription and instead exacerbated the aberrant PcH transcription (Figure R42).

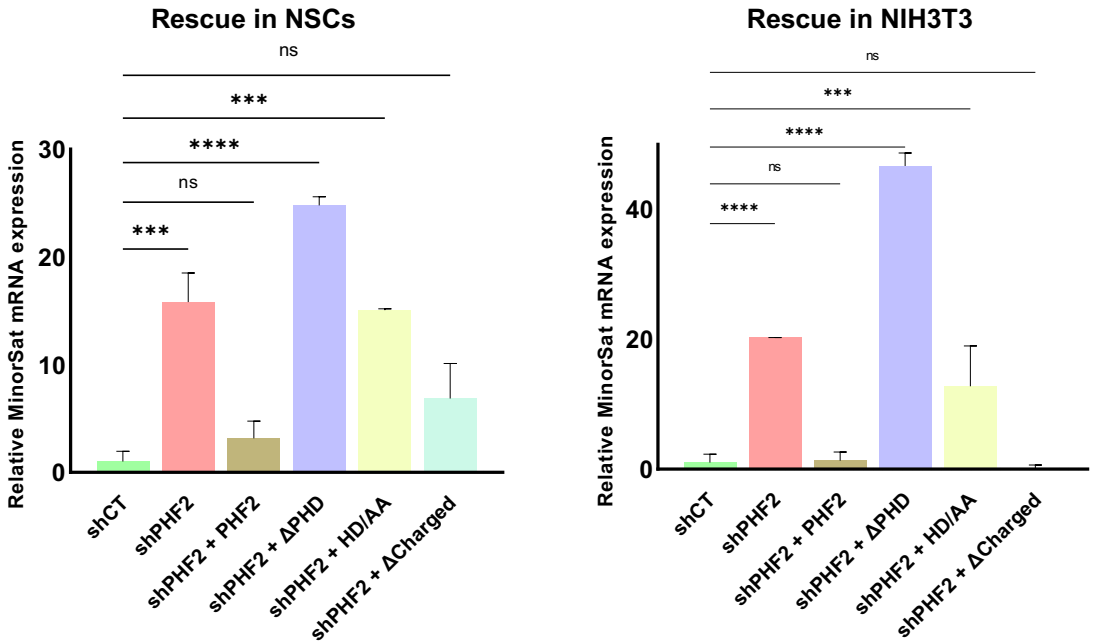


Figure R42. Bar plot depicting relative mRNA transcription levels of MinorSat in NSCs (left) and NIH3T3 (right) cells. Both cell types were PHF2 depleted and then nucleofected or transfected respectively with either PHF2 WT, ΔCharged, ΔPHD or HD/AA mutant. Transfection of WT and ΔCharged mutant rescued the phenotype, but not the ΔPHD or the HD/AA mutants. Results are from three biologically independent experiments. Error bars show SD. ***p<0.001; ****p<0.0001.

1.2.3 Increased nuclear foci volume rescue

Additionally, in line with the results presented in Figures R10 and R11, both PHF2 WT and the ΔCharged mutant effectively reversed the increased volume of H3K9me3 foci observed in PHF2-depleted cells. Conversely, the catalytic and PHD domains failed to produce this effect (Figure R43).

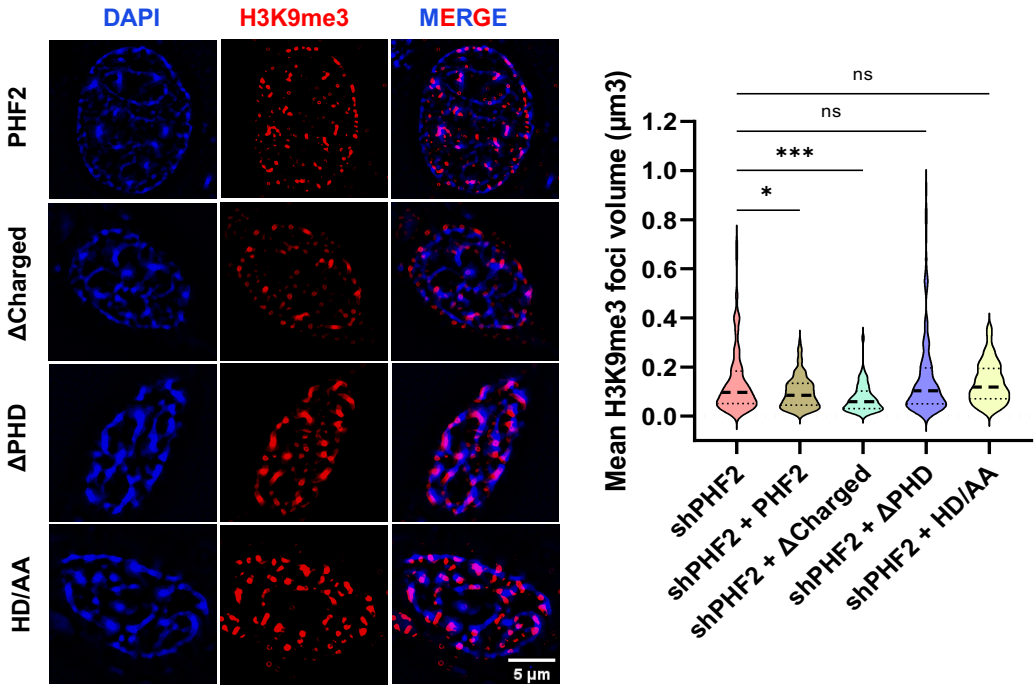


Figure R43. Immunofluorescence experiment on shPHF2 NIH3T3 cells where either PHF2 WT, ΔCharged, ΔPHD or HD/AA mutants were transfected. Cells were stained with H3K9me3 (red) antibody and DAPI (blue), then super-resolution images were acquired (left panel). Foci volume was quantified for H3K9me3 channel from n=50 cells and represented in violin plots (right panel). *p<0.05; ***p<0.001.

Taken together, these findings offer evidence supporting the involvement of PHF2 catalytic activity, likely in maintaining the balance of H3K9me3, and the PHD domain, possibly in facilitating the recognition of euchromatin histone marks, in regulating the transcriptional status of pericentromeric satellites.

1.13 PHF2-induced satellite transcription results in DNA damage

PHF2 depletion has been associated with the accumulation of double-strand breaks (DSBs) and R-loops (5). The DNA damage stemming from unscheduled transcription of PcH is a significant contributor to this phenomenon. Hence, we opted to investigate the correlation between repeat transcription, DSBs, and genome instability. To achieve this, we utilized the aforementioned PHF2 mutants, some of which were capable of rescuing unscheduled repeat transcription while others were not. We assessed their potential to rescue DNA damage induced by PHF2 depletion, as detected through immunostaining with γH2Ax. The results depicted in Figure R44 revealed that the ΔCharged mutant, recognized for its significance in gene transcriptional activity (210), completely restored DNA breaks, akin to PHF2 WT. However, the PHD and catalytic mutants, which induce repeat expression, failed to do so.

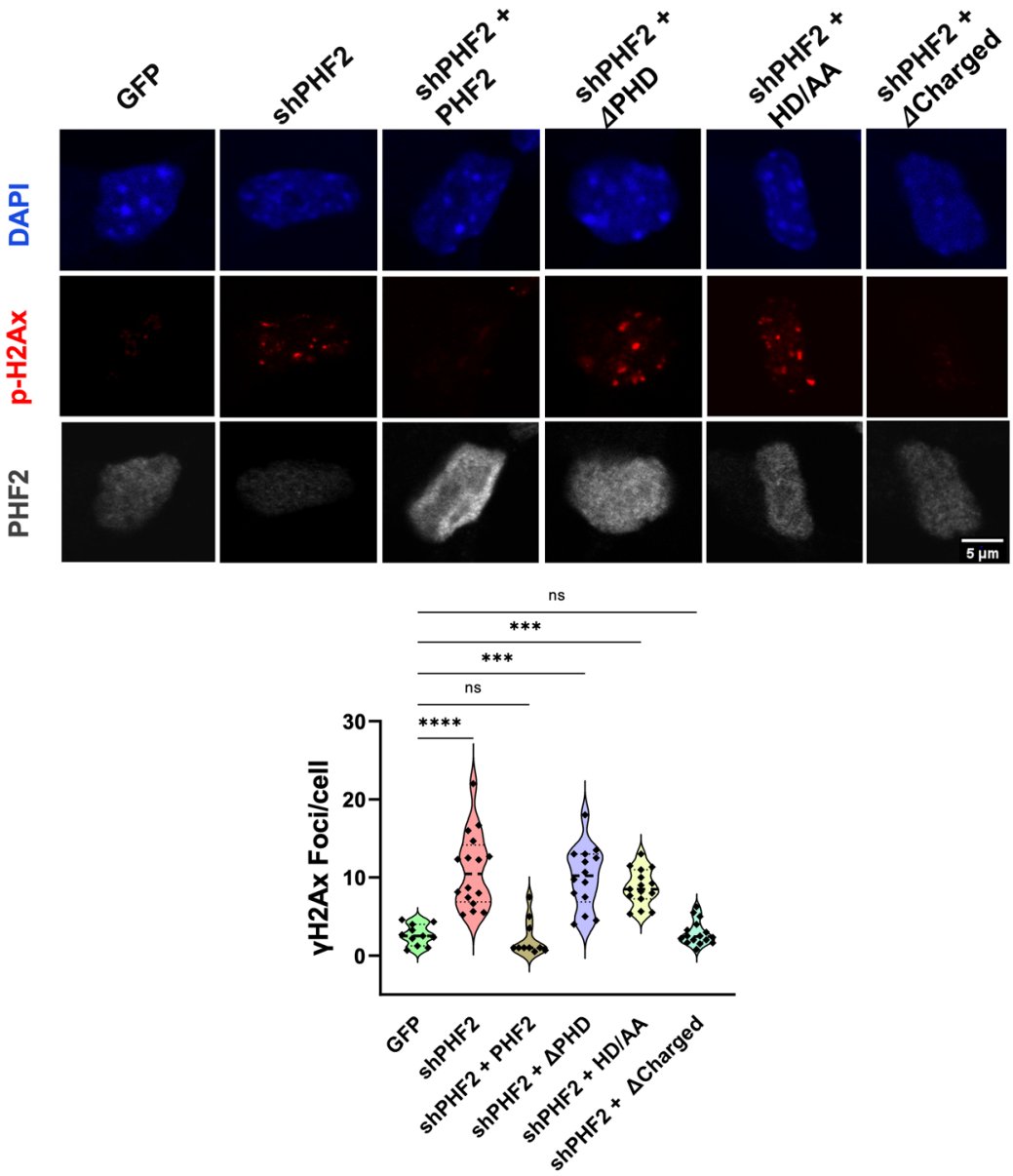


Figure R44. Immunofluorescence experiment on shCT and shPHF2 NSCs where either PHF2 WT, ΔCharged, ΔPHD or HD/AA mutants were nucleofected. Cells were stained with γH2Ax (red), PHF2 (grey) antibody and DAPI (blue), then confocal images were acquired (top panel). Number of γH2Ax foci per cell was quantified from n=50 cells and represented in violin plots (bottom panel). ***p<0.001; ****p<0.0001.

This observation was further reinforced by Figure R45, demonstrating that only the overexpression of the PHD and catalytic PHF2 mutants, but not the WT, resulted in DNA damage.

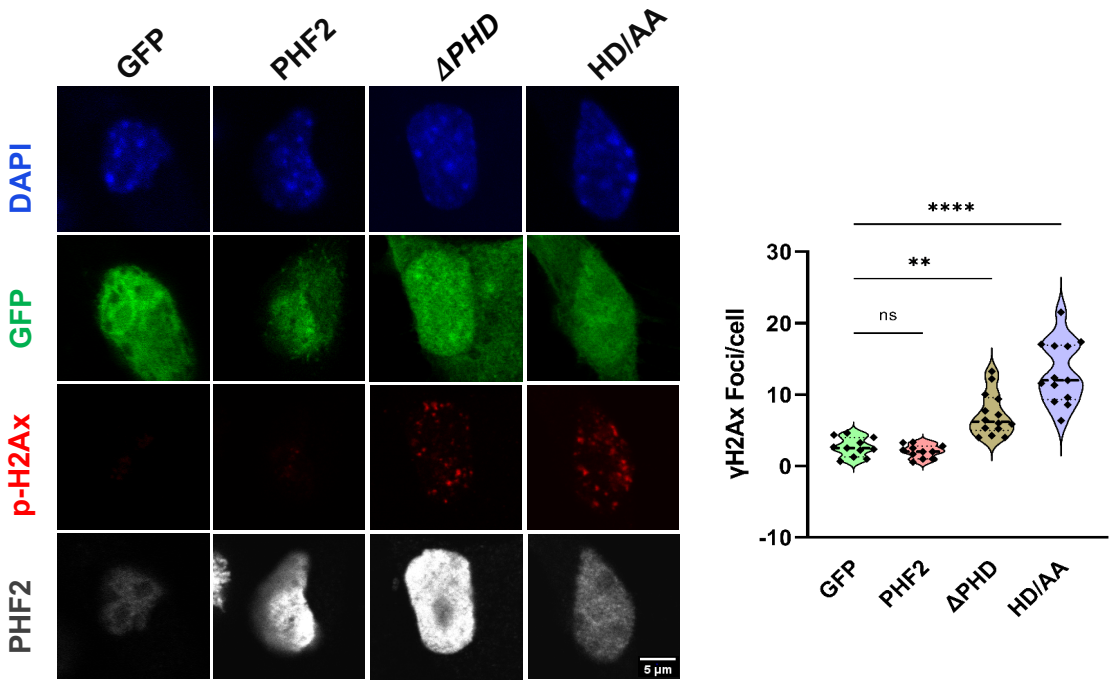


Figure R45. Immunofluorescence experiment on NSCs where either PHF2 WT, ΔCharged, ΔPHD or HD/AA mutants were nucleofected. GFP is showed as a positive transfection control. Cells were stained with γH2Ax (red), PHF2 (grey) antibody and DAPI (blue), then confocal images were acquired (left panel). Number of γH2Ax foci per cell was quantified from n=50 cells and represented in violin plots (right panel). **p<0.01; ****p<0.0001.

1.14 PHF2-induced satellite transcription results in genomic instability

It has been previously described that PHF2 depletion induced genomic instability (5). To test the functional role of PHF2 in this phenotype, we nucleofected NSCs that had been previously depleted of PHF2 with the described mutants. The genomic instability was effectively mitigated by PHF2 WT and the ΔCharged mutant, while the PHF2 PHD or catalytic domain mutants failed to provide rescue (Figure R46).

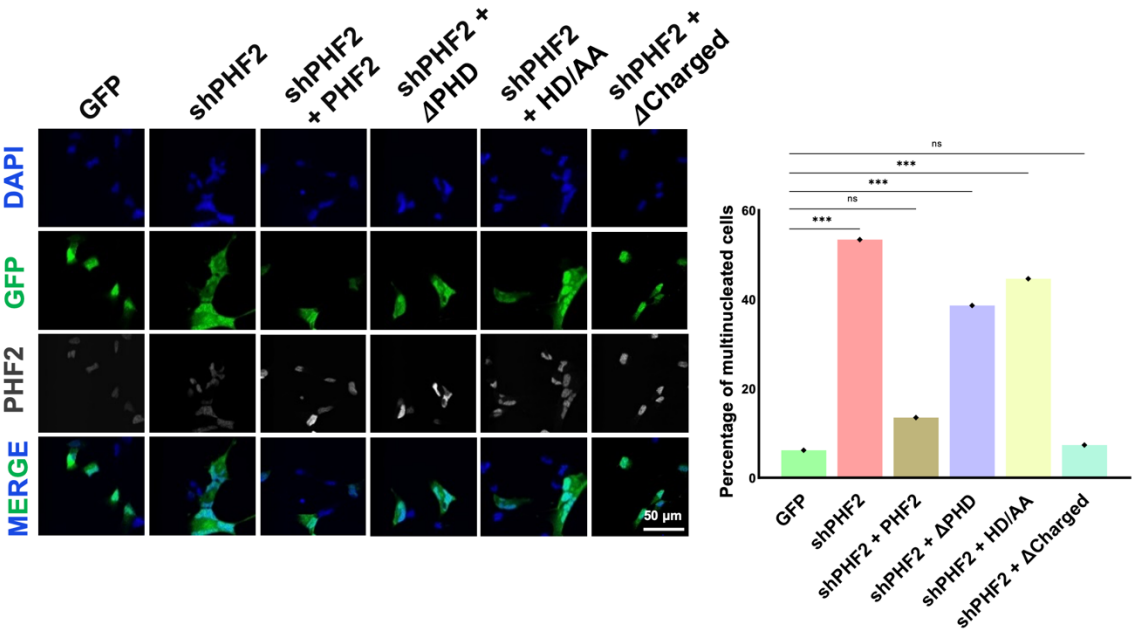


Figure R46. Immunofluorescence experiment on shCT and shPHF2 NSCs where either PHF2 WT, ΔCharged, ΔPHD or HD/AA mutants were nucleofected. GFP is showed as a positive transfection control and used to discern cell shape. Cells were stained with PHF2 (grey) antibody and DAPI (blue), then confocal images were acquired (left panel). Percentage of multinucleated cells was quantified from n=50 cells and represented in violin plots (right panel). ***p<0.001.

These findings imply that in addition to its established role at gene promoters, PHF2 also holds a pivotal function in preserving PcH stability through its PHD and catalytic domains. This newly identified role of PHF2 serves to safeguard against genome instability, highlighting the multifaceted nature of its functions.

2. CHAPTER 2: PHF2's role in mitosis

In this second chapter, I will elaborate on PHF2's role in the cellular process of mitosis. As it has been described before in our laboratory, PHF2 is necessary for a correct neural progenitor expansion and cell cycle progression, and its depletion leads to a delay in neural progenitor proliferation; on the other hand, PHF2 has been described as a kinetochore associated protein (143). Moreover, Fisher's group at Oxford (UK), has identified proteins that were significantly enriched in chromosome-sorted fractions compared to mitotic extracts, and PHF2 was identified, among other proteins like Sox2, Hist1h1a, Smc1, Dnmt1, PRC1 and Ezh2 (Table R1) (208).

Protein ID	Gene name	Mol weight	Score	Intensity
Q9WTU0	Phf2	120,81	58,425	2468900000
Q3TMX7	Qsox2	77,774	6,0591	320290000
P43275	Hist1h1a	21,785	90,8	2423800000
Q9CU62	Smc1a	143,23	323,31	2649200000
P13864	Dnmt1	183,19	122,39	4987500000
Q99K43	Prc1	70,289	43,44	2359800000
Q61188	Ezh2	85,291	41,844	760650000

Table R1. Table showing chromosome-sorted fraction proteomics results. Chromosomes from fibroblasts were used for this experiment and peptides were identified by mass spectrometry.

Based on these data we decided to analyze the potential role of PHF2 during the mitotic phase of the cell cycle using NIH3T3 cells. In order to study the mitosis in depth, and to increase our n of mitotic cells, we performed cell synchronization as described in the materials and methods section. Following that, propidium iodide staining and FACS analysis confirmed that our populations were representative of G1(84,176%) and M phase (74,861%) (Figure R47)

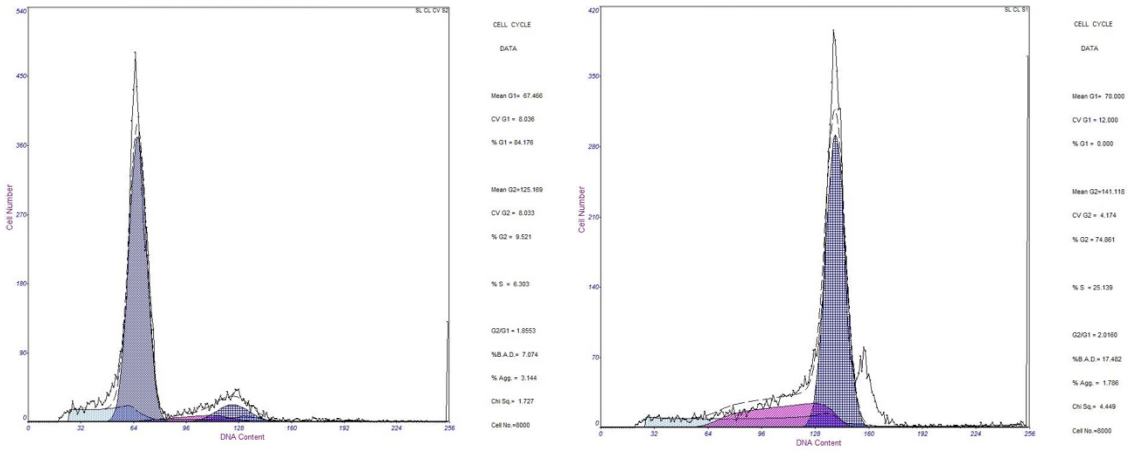


Figure R47. NIH3T3 G1 and M (manually shaken off) cell populations were stained with propidium iodide and FACS sorted to analyze their DNA content.

2.1 PHF2 remains bound to mitotic chromosomes

During mitosis, chromatin condenses, transcription is shut off and most transcription factors and cofactors have been reported to be excluded from chromosomes (141). Surprisingly, immunoblots and immunostaining experiments showed that PHF2, was not

degraded or excluded from the chromatin, but instead retained to chromatin during mitosis (Figure R48 and R49).

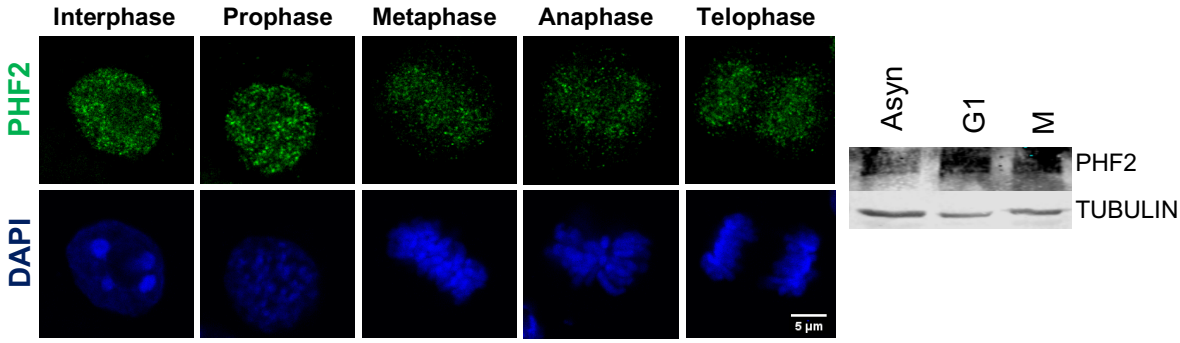


Figure R48. NIH3T3 cells were synchronized to both G1 and M phase, and protein extracts were immunoblotted against PHF2. Tubulin was used as load control (right panel). Immunofluorescence experiment on NIH3T3 cells that were stained with PHF2 (green) antibody and DAPI (blue), then confocal images were acquired in every step of the mitosis. A clear enrichment following DNA shape is observed in PHF2's signal (left panel).

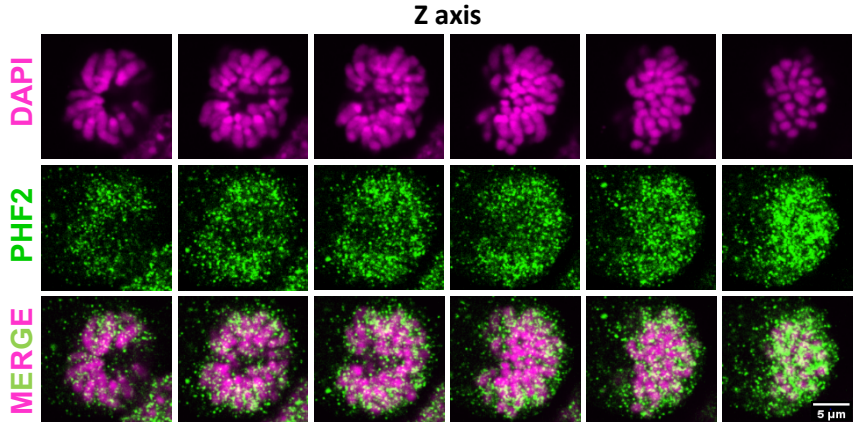


Figure R49. Detailed metaphase immunofluorescence experiment on NIH3T3 cells that were stained with PHF2 (green) antibody and DAPI (pink), then confocal images were acquired, and several confocal planes are showed to clarify PHF2's distribution.

To confirm these data and to avoid the used of crosslinking agents in collaboration with Prof. A Fisher's group, we sorted individual unfixed native mitotic chromosomes for imaging analysis. We then separately purify 3 representative mitotic chromosome/s from NIH3T3 cells following their stablished protocol. Chromosomes 19, 3+X and 1+2 were easily isolable by FACS (Figure R50), although the resolution was not as high as when using ESCs (241).

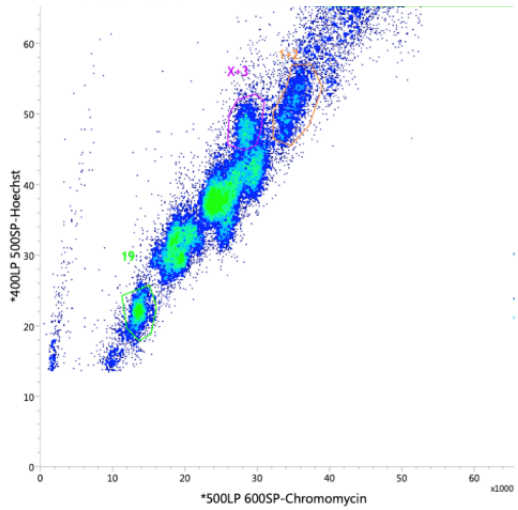


Figure R50. Flow karyotype of mitotic chromosomes isolated from NIH3T3 cells. Gates used to isolate chromosomes 19, X+3 or 1+2 are indicated. Image is representative of three independent experiments.

To accurately check if PHF2 is bound to native mitotic chromosomes, individual chromosomes were cyto-spined and incubated with PHF2 antibody and DAPI stain for imaging (Figure R51). The results show that PHF2 is bound to native mitotic chromosomes.

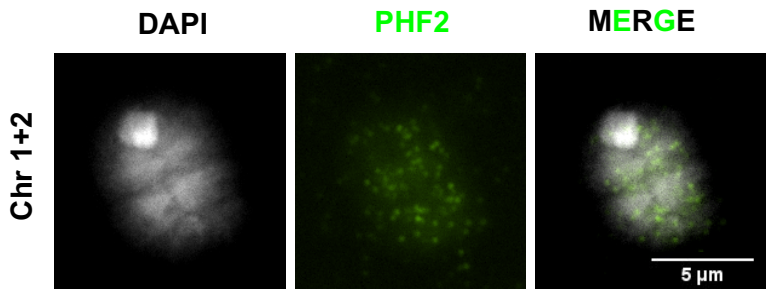


Figure R51. Representative images of chromosomes 1+2 from different NIH3T3 cells are shown, where DAPI stain (grey) and PHF2 antibody (green) were used. Images are representative of two independent experiments.

2.2 mEGFP-PHF2 follows chromosome dynamics

To analyze PHF2 dynamics during mitosis, we constructed a recombinant protein in which we fused PHF2 to a monomeric enhanced green fluorescent protein (PHF2-mEGFP). We then checked by immunoblot and immunofluorescence (Figure R52, left panel) the correct expression of mEGFP-PHF2. The fusion protein remains bound to the mitotic chromatin at different stages of mitosis (Figure R52, right panel).

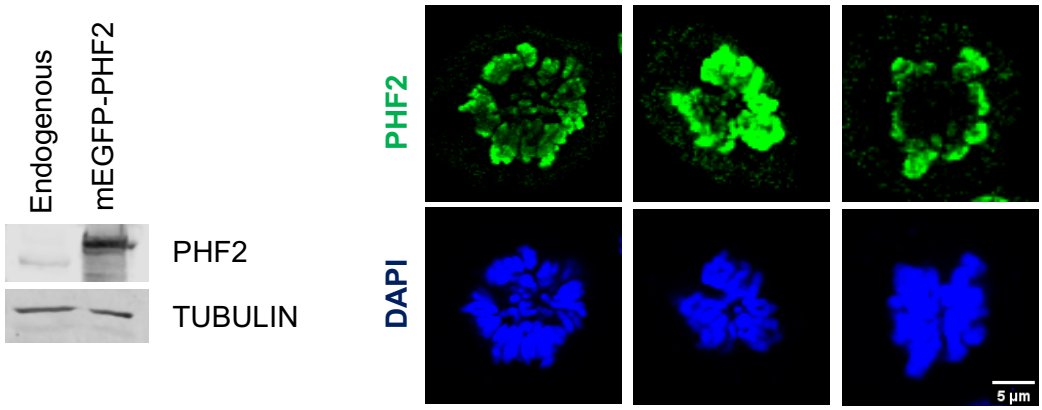


Figure R52. HEK293T cells were transfected with a random vector or mEGFP-PHF2 and protein extracts were immunoblotted against PHF2. Tubulin was used as load control (left panel). NIH3T3 cells were transfected with mEGFP-PHF2, stained with DAPI and confocal images were acquired (right panel).

We then recorded mEGFP-PHF2 expression in live NIH3T3 cells and observed that it follows chromosome dynamics in mitosis. (Figure R53).

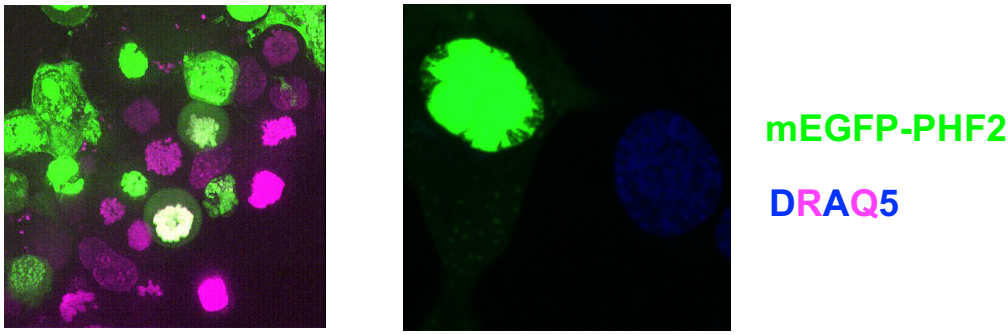


Figure R53. NIH3T3 cells were synchronized and transfected with mEGFP-PHF2. Then cells were live imaged and confocal time series were taken. Green channel corresponds to mEGFP-PHF2, and blue/pink channel corresponds to DNA stain DRAQ5. PHF2 follows chromosome dynamics during mitosis.

2.3 PHF2 relocates during mitosis

Following the retention of PHF2 to mitotic chromatin, we aimed to investigate the specific genomic regions bound by PHF2 in mitosis. To address this, we performed PHF2 ChIP-seq experiments in duplicate using NIH3T3 that had been synchronized to either G1 phase or M phase.

Our findings demonstrated binding of PHF2 at both G1 and M phases, although a differential location was observed. While G1 binding was very similar to what it had been previously described for asynchronous cells in Pappa et al 2019 (5), mitotic binding presented a shift towards intergenic regions (Figure R55).

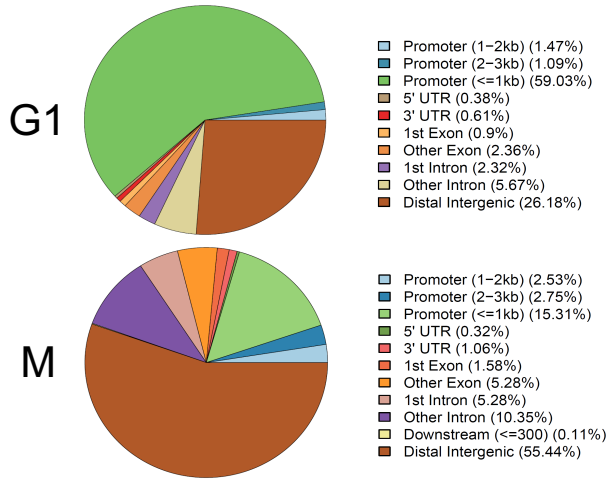


Figure R55. Spatial positioning (ChIPSeeker) analysis was performed with the genomic binding locations of PHF2 replicates merged in G1 or M phase showing a differential binding between these two phases.

Concordantly with differential spatial positioning, only a small percentage of each condition overlapped when Venn's Diagram was plotted (Figure R56), confirming that PHF2 relocates during mitosis.

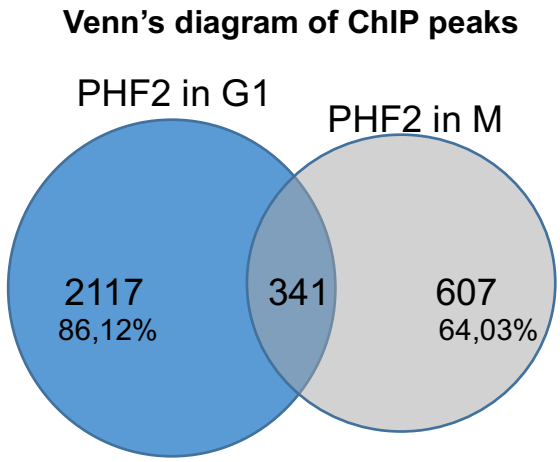


Figure R56. Venn's diagram depicting number of peaks in G1 (blue) and M (grey) conditions as well as common peaks between them.

Unfortunately, our GO analysis did not throw any significant enrichment in gene categories, given the small number of genes in which there was PHF2 binding.

2.4 PHF2 mitotic exclusive binding sites show higher colocalization with Lamin B

To elucidate if PHF2 was playing a structural role in mitosis, therefore joining intronic regions and possibly binding close or in contact with the Lamin associated domains (LADs), we separated our PHF2 ChIP information into groups: Asynchronous, G1 exclusive, G1 and M common peaks, and M exclusive peaks. These groups were then merged with previously deposited Lamin B ChIP-seq. M exclusive peaks showed higher percentage of colocalization with Lamin B (Table R2), suggesting a possible structural adjuvant effect of PHF2 in mitosis.

Sample	% in LADs
Asynchronous	7,9105474
G1 exclusive	10,6258442
G1+M common	12,5326371
M exclusive	19,4991055

Table R2. Table showing percentage of colocalization of different PHF2 peaks populations with LADs. M exclusive peaks show the highest percentage.

2.5 Mitotic chromosomes are smaller after PHF2 depletion

Once we were sure that PHF2 remains bound to mitotic chromatin we sought to determine the functional and structural consequences of its depletion. To that we stablished a PHF2 KD NIH3T3 cell line (revisit Figure R9), and then we separately purify 3 representative mitotic chromosome/s as described in previous Figure R50 (Figure R57)..

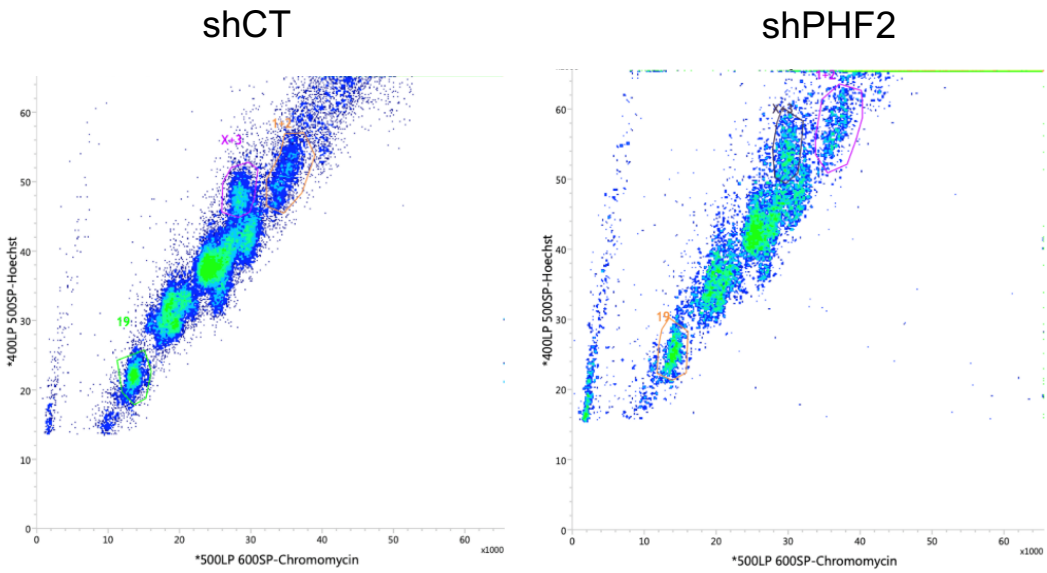


Figure R57. Flow karyotype of mitotic chromosomes isolated from shCT and shPHF2 NIH3T3 cells. Gates used to isolate chromosomes 19, X+3 or 1+2 are indicated. Images are representative of three independent experiments.

To check if PHF2 was depleted, individual chromosomes were cyto-spined and incubated with PHF2 antibody and DAPI stain for imaging (Figure R58).

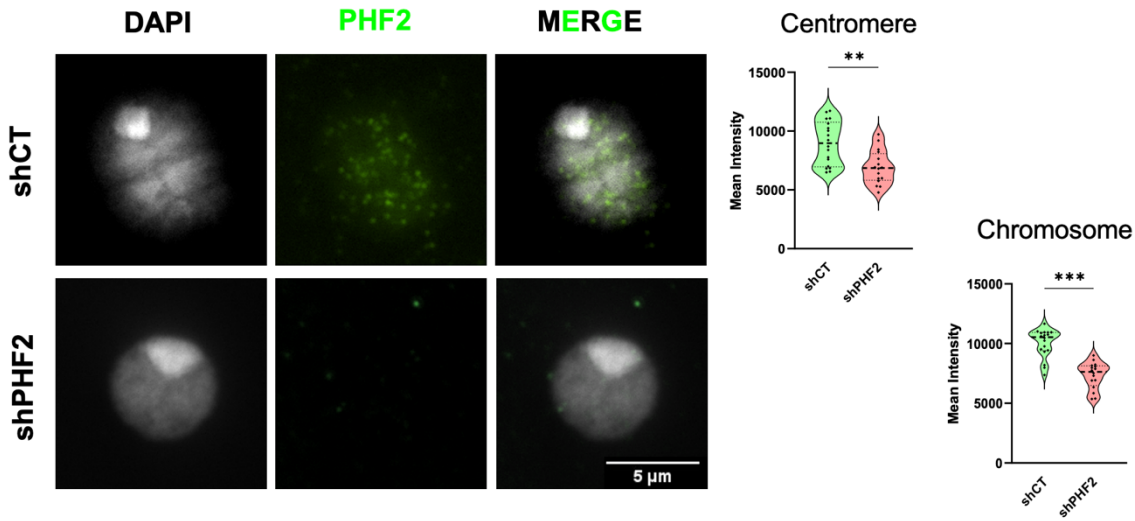


Figure R58. Representative images of chromosomes from different NIH3T3 cells are shown, where DAPI stain (grey) and PHF2 antibody (green) were used. DAPI stain was used to discern the chromosome body and centromere (high intensity). PHF2 intensity was measured in centromere and chromosome separately from at least 100 chromosomes per replicate, and violin plots are calculated with data from three independent experiments. **p<0.01; ***p<0.001

Next, individual chromosomes were measured by microscopy using ImageJ software to determine their total area and estimate the size of DAPI-bright pericentric domains. These analyses showed that mitotic chromosomes 19 and X+3 lacking PHF2 were significantly smaller than the corresponding chromosomes from shCT cells (Figure R59).

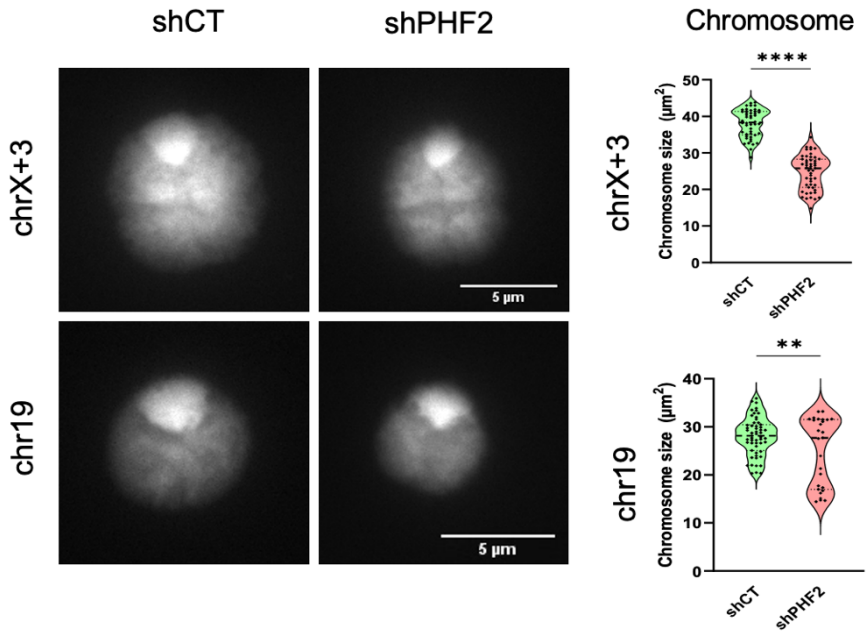


Figure R59. Representative images of chromosomes 19 and X+3 from shCT and shPHF2 NIH3T3 cells are shown, where DAPI stain (grey) indicate the centromere (high intensity) and chromosome body. Chromosome size was calculated by measuring at least 100 individual chromosomes for each condition and chromosome number, over three independent experiments. Violin plots represent a decrease in chromosome size upon PHF2 depletion. **p<0.01; ****p<0.0001

2.6 Centromeric H3K9me3 is reduced upon PHF2 depletion in chromosomes

After observing a decrease in chromosome size when PHF2 was depleted, we ought to study the epigenetic changes most related to chromosome size and compaction (H3K9me3 and H3K27me3), in mitotic cells upon PHF2 depletion. When we performed chromosome sorting and examined K9 methylation levels in the centromere area, we obtained a decrease in H3K9me3 levels, concordantly to what described in Figures R15 and R16 (Figure R60).

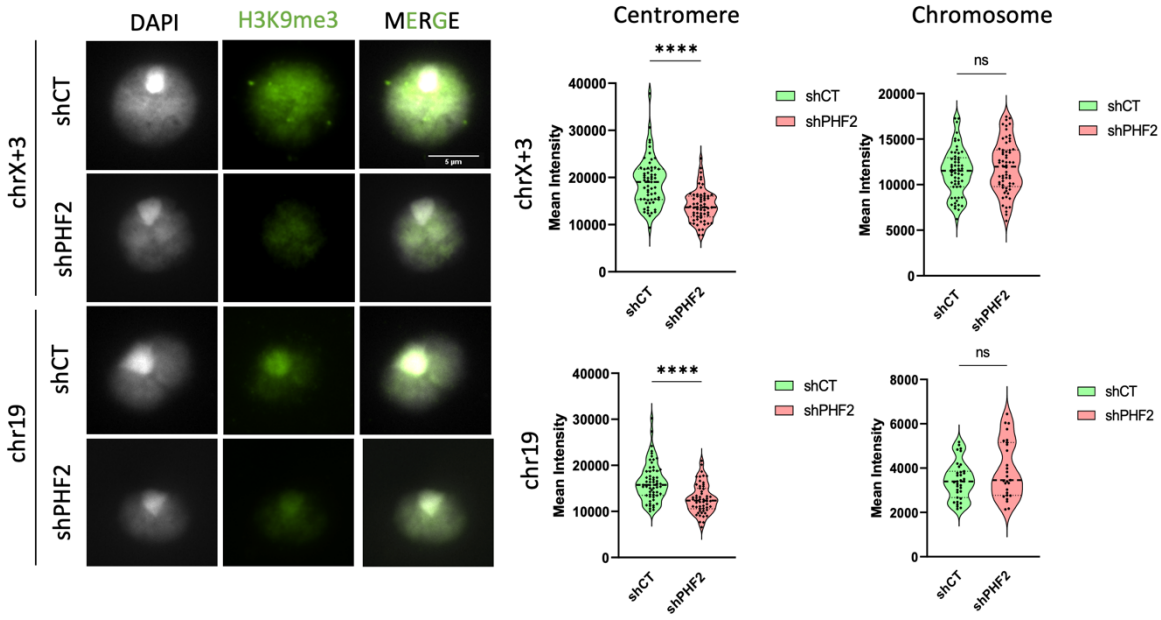


Figure R60. Representative images of chromosomes 19 and X+3 from shCT and shPHF2 NIH3T3 cells are shown, where DAPI stain (grey) indicate the centromere (high intensity) and chromosome body, and H3K9me3 antibody (green) was incubated. Chromosome and centromere areas were defined as mentioned and H3K9me3 intensity levels were measured from at least 100 individual chromosomes and centromeres for

each condition and chromosome number, over three independent experiments. Violin plots represent a decrease in centromere H3K9me3 intensity upon PHF2 depletion. ****p<0.0001

2.7 PHF2 depletion increases H3K27me3 in mitotic chromosomes

Interestingly, H3K27me3 immunostaining on sorted chromosomes showed an increase in this mark in the chromosome body (Figure R61), which correlates with decrease in chromosome size.

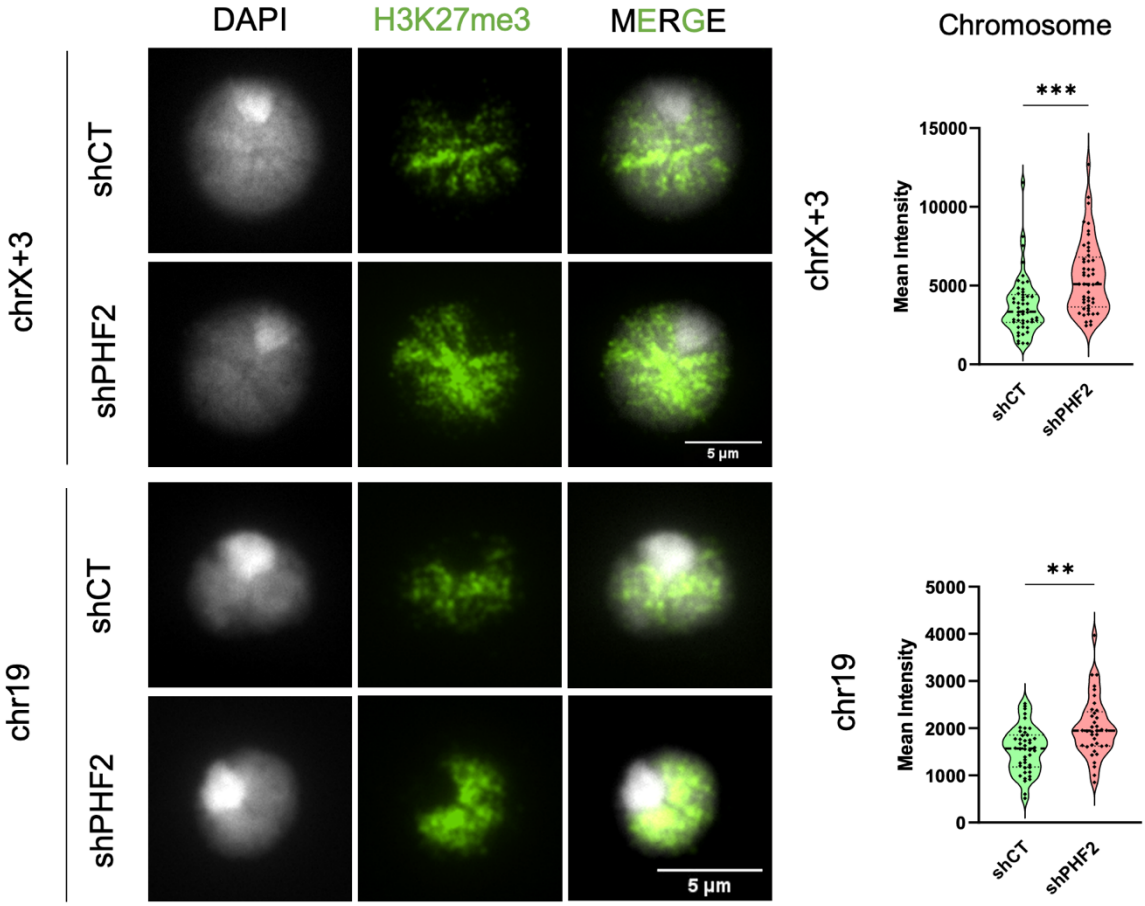


Figure R61. Representative images of chromosomes 19 and X+3 from shCT and shPHF2 NIH3T3 cells are shown, where DAPI stain (grey) indicate the centromere (high intensity) and chromosome body, and H3K27me3 antibody (green) was incubated. H3K27me3 intensity levels were measured from at least 100 individual chromosomes for each condition and chromosome number, over three independent experiments. Violin plots represent an increase in H3K27me3 intensity upon PHF2 depletion. **p<0.01; ***p<0.001

2.8 H3K9me3 and H3K27me3 increase in mitotic fixed cells upon PHF2 depletion

We also wanted to check these results on a cellular level. Concordantly with previous results in this thesis and other works from the lab (5), we detected a slight increase in general mitotic H3K9me3 levels as well as H3K27me3 levels (Figure R62 and R63).

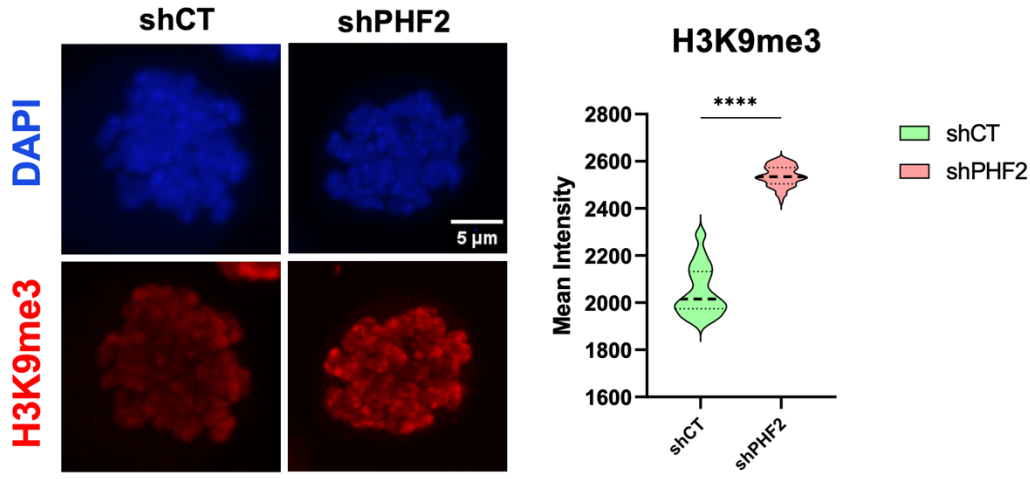


Figure R62. Immunostaining experiment from shCT and shPHF2 mitotic NIH3T3 cells is shown, where DAPI stain (blue) and anti H3K9me3 (red) antibody were incubated. Mean intensity was quantified for H3K9me3 channel from n=50 cells and represented in violin plots (right panel). ****p<0.0001

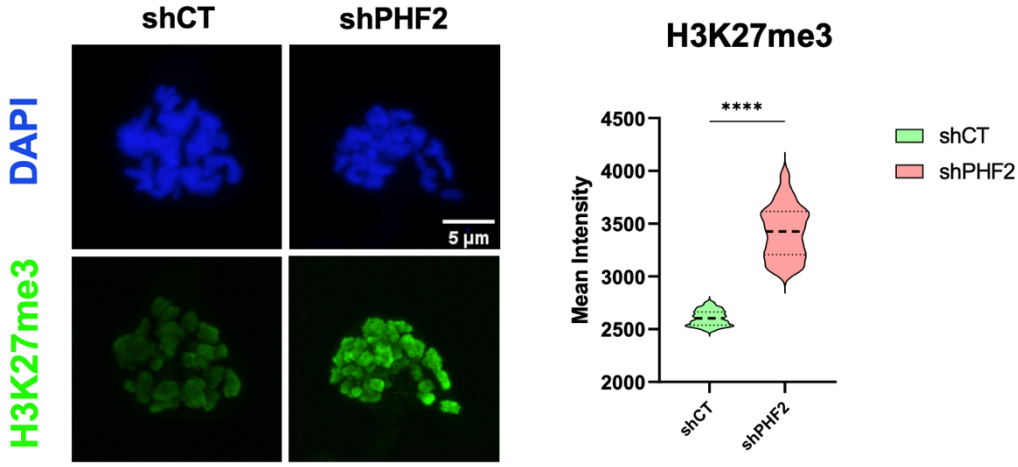


Figure R63. Immunostaining experiment from shCT and shPHF2 mitotic NIH3T3 cells is shown, where DAPI stain (blue) and anti H3K27me3 (green) antibody were incubated.. Mean intensity was quantified for H3K27me3 channel from n=50 cells and represented in violin plots (right panel). ****p<0.0001

These data suggest that PHF2 is important to maintain the size of mitotic chromosomes, probably maintaining the balance of H3K27me3 and H3K9me3 levels.

2.9 Gene expression is delayed in early G1 after a PHF2-depleted mitosis

Given the described role of PHF2 in regulating transcription (5), and the complexity of transcription reactivation after mitosis (146,242), we sought to elucidate if PHF2 was contributing to the transcription reactivation at G1 phase. To do that, we performed Global Run On followed by sequencing (GRO-seq) in shCT and shPHF2 NIH3T3 cells at early G1 (G1.1) and later G1 (G1.2). Effectively, we observed a delay in the transcription reactivation in early G1 compared to later G1 described by a bigger accumulation of downregulated genes in this phase (Figure R64).

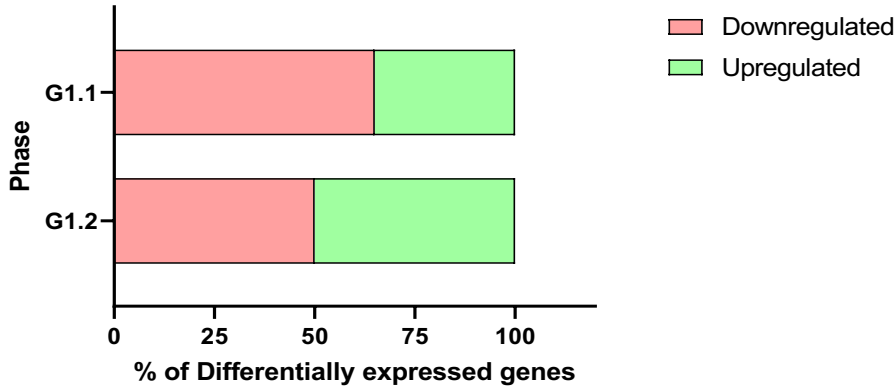


Figure R64. GRO-seq of shCT and shPHF2 NIH3T3 in G1.1 and G1.2 time point was done. Percentage of differentially expressed genes in each time point were plotted in bar graph as part of a whole. Early G1 showed accumulation of downregulated genes vs late G1. Results are the mean of two biologically independent experiments.

Moreover, gene ontology (GO) analysis of the downregulated genes in G1.2 revealed their involvement in essential processes associated with chromatin organization (Figure R65).

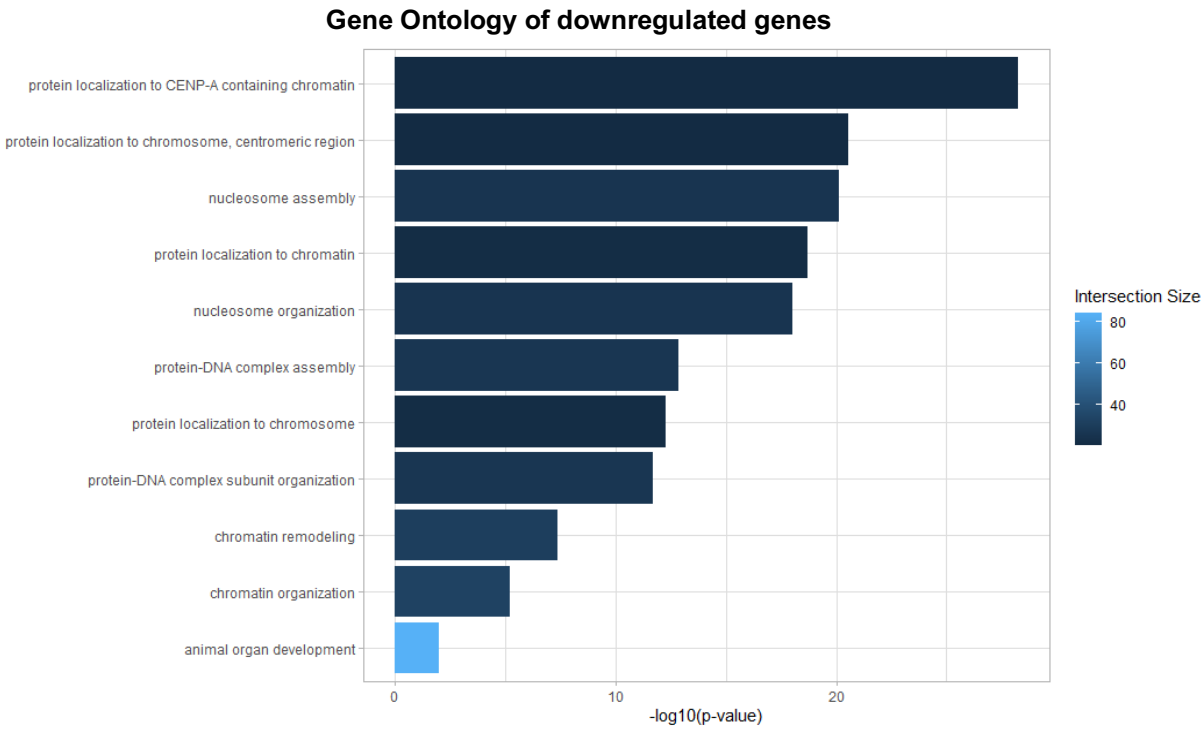


Figure R65. Gene ontology analysis showing biological process of the downregulated genes upon PHF2 depletion in G1 NIH3T3 cells using the Mus musculus genome (mm10) as a reference.

In particular, the expression gene clusters codifying for histones were affected both in early G1 and later G1 (Table R3 and Figure R66) suggesting that PHF2 might be necessary for a correct amount of histone transcription and further load into the chromatin.

SYMBOL	ENSEMBL	SYMBOL	ENSEMBL	SYMBOL	ENSEMBL
Hist2h3c1	ENSMUSG00000093769	Hist1h4k	ENSMUSG00000064288	Prdm16	ENSMUSG00000039410
Hist1h1e	ENSMUSG00000051627	Hist1h4a	ENSMUSG00000060093	Hist1h2ai	ENSMUSG00000071516
Hist1h4d	ENSMUSG00000061482	Hist1h1b	ENSMUSG00000058773	Hist1h2ak	ENSMUSG00000063021
Hist1h2bn	ENSMUSG00000095217	Hist2h3c2	ENSMUSG00000081058	Hist1h4j	ENSMUSG00000067455
Hist2h2aa2	ENSMUSG00000063954	Hist1h2af	ENSMUSG00000061991	Hist1h2bk	ENSMUSG00000062727
Zdbf2	ENSMUSG00000027520	Hist1h2ae	ENSMUSG00000069272	Hist1h2bj	ENSMUSG00000069300
Kdm5d	ENSMUSG00000056673	Hist1h2br	ENSMUSG00000069303	Hist1h4i	ENSMUSG00000060639
Hdac9	ENSMUSG00000004698	Hist1h2ag	ENSMUSG00000069301	Med10	ENSMUSG00000021598
Hist1h4h	ENSMUSG00000060981	Hist1h1c	ENSMUSG00000036181	Hist1h4n	ENSMUSG00000069305
Hist1h2ao	ENSMUSG00000094248	Hist1h2bq	ENSMUSG00000069307	Hist1h2ad	ENSMUSG00000071478
Hist1h2ap	ENSMUSG00000094777	Eya4	ENSMUSG00000010461	Hist1h4b	ENSMUSG00000069266
Hist1h4c	ENSMUSG00000060678	Hist1h3a	ENSMUSG00000069265	Hist1h2ah	ENSMUSG00000069302
Ercc6l2	ENSMUSG00000021470	Phf19	ENSMUSG00000026873	Hist2h2bb	ENSMUSG00000105827
Hist1h4f	ENSMUSG00000069274	Hist1h2an	ENSMUSG00000069309	Bcl11a	ENSMUSG00000000861
Hist1h1d	ENSMUSG00000052565	Hist1h2bc	ENSMUSG00000018102	Hist1h1a	ENSMUSG00000049539
Nup210	ENSMUSG00000030091	Fancg	ENSMUSG00000028453	Hist2h3b	ENSMUSG00000074403

Table R3. Summary table depicting chromatin and nucleosome-related genes that are downregulated in G1.2 in the GRO-seq using shCT and shPHF2 NIH3T3 cells.

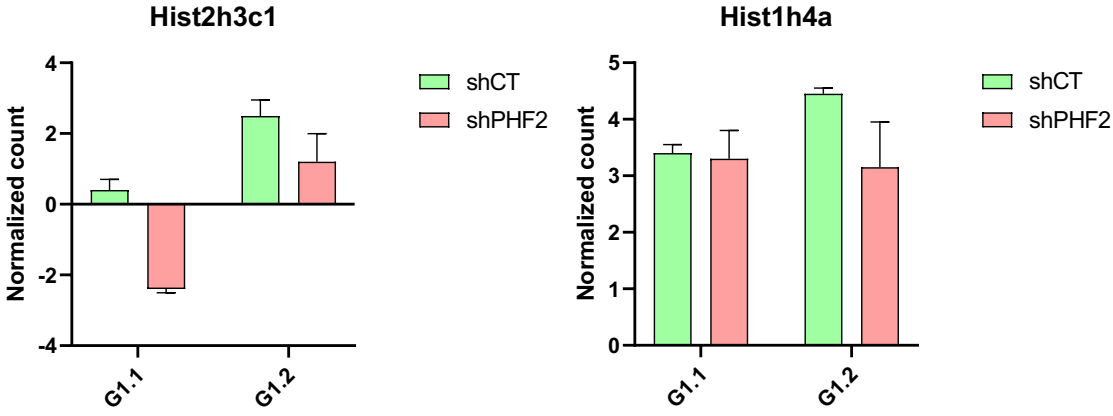


Figure R66. Bar plot depicting normalized count of histone H1 and H2 expression in shCT and shPHF2 synchronized NIH3T3 cells. Graph show lower levels of transcription in early G1 phase that weren't recovered in later G1. Three biologically independent experiments were used for each condition.

The results exposed in this second chapter imply that in addition to its established role at gene promoters, and the previously demonstrated role in heterochromatin (Chapter 1 of this thesis), PHF2 regulates mitotic chromosome size and when depleted, there is a delayed transcription reactivation in G1 phase, especially histone transcription. This could be related to its differential binding in M phase that may have implications on chromatin structure, although further research is needed to confirm that hypothesis.

DISCUSSION

1. Regarding PHF2's role in heterochromatin stability

Grasping the function of histone modifications in preserving chromatin integrity during neural development and their link to disease pathology remains a formidable challenge in the field. Among all histone epigenetic modifications, H3K9 methylation stands out as a crucial and conserved mark involved in chromatin silencing (12,114,125), particularly centromeric heterochromatin. In this thesis it has been demonstrated that PHF2-mediated H3K9 demethylation is essential for stabilizing pericentromeric heterochromatin (PcH) repetitive sequences. The previous chapter 1 of results shows that PHF2 is enriched at the boundaries of PcH, aiding in the silencing of pericentric repeats in neural stem cells (NSCs). Through this silencing, PHF2 plays a critical role in preventing DNA damage accumulation and maintaining genome stability, which are vital for the proliferation of neural progenitor cells. These insights help to understand the molecular mechanisms governing chromatin stability and their influence on cellular processes during development.

1.1 About PHF2 balancing K9me and maintaining heterochromatin stability

During this thesis, it has been demonstrated that PHF2 prevents repetitive elements transcription, and this transcription may be caused due to heterochromatin spreading in the boundary regions (Figures R20 and R21). Heterochromatin spreading is a gradual process that involves the generation of heterochromatic territories enriched in proteins such as HP1, HDAC, SUV39, and H1. These factors, whose abundance acts as a limiting factor (234,243,244) interact with a read-write mechanism that involves H3K9me3 density levels as a critical guide component of heterochromatin spreading (114,245). Therefore, the presence of a histone K9 demethylase like PHF2 may contribute to limiting this spreading since it has been previously described (245,246). Our findings indicate that maintaining a precise balance of H3K9me3 throughout the genome is essential for the stability of repetitive sequences. An increase in this mark in boundary regions and other genomic areas can potentially promote abnormal repeat transcription, likely through the dilution or redistribution of heterochromatin components, as demonstrated by our observations with HP1α (Figure R21 and R23). After PHF2 depletion, we observed an increase in the H3K9me3 mark over the PcH boundaries, leading to aberrant repeat transcription. This suggest that PHF2 provides stability to heterochromatin by preventing the dynamic assembly and disassembly of heterochromatin, which is crucial for controlling PcH spreading (106).

An alternative mechanism by which PHF2 maintains heterochromatin stability could be maintaining transcription levels at the PcH boundaries. When PHF2 is depleted, a decrease in transcription has been demonstrated in Figure R28. Numerous studies have

demonstrated an interaction between RNA-mediated mechanisms and transcription that limit the spreading of heterochromatin acting like a boundary element both in yeast and mammals (129,236–239). Based on that and giving PHF2'role as a transcriptional regulator (5), it could be modulating PcH stability by maintaining the correct amount of RNA levels in this PcH boundary. However, these changes in RNA levels in the boundary region might be a consequence of H3K9me3 spreading and not a cause per se. In addition to that, is it possible that PHF2 might cause other molecular outcomes like histone modifications, changes in genome architecture or nuclear positioning therefore regulating heterochromatin spreading, and thus allowing an exacerbate repetitive elements transcription. These described mechanisms might collaborate to ensure heterochromatin stability.

Interestingly, it has been described that PHF2 homologous JmjC domain-containing protein in *Schizosaccharomyces pombe*, Epe1, prevents heterochromatin instability (247,248). Consistently with the results exposed in this doctoral thesis, Epe1 is essential for silencing at centromeres and heterochromatin integrity, and its depletion led to impaired silencing. Furthermore, even though Epe1's catalytic activity has not been confirmed, it restricts heterochromatin domains (112,248). In mammals, KDM4B antagonizes H3K9me3 at PcH heterochromatin (249,250). This may imply a conserved role across species for JmjC domain-containing proteins in regulating heterochromatin stability and preventing the spread of heterochromatic regions.

Alternative mechanisms may also contribute to PHF2-mediated heterochromatin stability. The positioning of heterochromatin domains at the nuclear periphery facilitates

heterochromatin nucleation and spreading. Peripheral tethering is believed to create a specialized nuclear subdomain that promotes the efficient loading of factors involved in histone turnover suppression. Loss of factors involved in peripheral tethering has been shown to increase histone turnover, leading to a reduction in H3K9me3 density and defective heterochromatin propagation (251,252). Based on this, it would be interesting to test whether PHF2 modulates histone turnover, promoting the maintenance of H3K9me3 density and heterochromatin stability. Notably, the major histone mark associated with lamina association is H3K9me2, which is targeted by PHF2. Therefore, the loss of PHF2 could disrupt the anchoring of heterochromatin to the lamina, resulting in heterochromatin destabilization.

1.2 About PHF2 safeguarding DNA from damage and genomic instability

The results depicted in this doctoral thesis demonstrate that PHF2's depletion causes DNA damage and genome instability. The unprogrammed transcription of pericentromeric heterochromatin repeats due to PHF2 depletion could lead to collisions between the replication and transcription machineries, resulting in DNA breaks and R-loops accumulation that led to genomic instability and chromosomal aberrations (132,133).

Additionally, the absence of PHF2 results in an increase in H3K9me3 both globally and in the PcH boundary, which may impede DNA repair, particularly at heterochromatin regions, since it has been described that demethylation of H3K9 and decompaction is

needed for the proper repair of DSBs. This has been reported for the KDM4B demethylase (253,254). Another potential mechanism is that the depletion of PHF2 could lead to an increase in aberrant transcripts that are not efficiently processed cotranscriptionally, resulting in the formation of R-loops (255).

Furthermore, our RNA-seq experiment indicates a decrease in key factors involved in R-loop resolution and DNA damage repair, such as ATM/ATR and BRCA1, in PHF2-depleted cells, as well as transcription in the PcH boundary area, where there is an enrichment in tRNAs and ncRNAs (Figure R32). In addition to that, PHF2 has been confirmed to have a role in DNA repair (75). This lack of DNA repair associated factors could contribute to accumulation of double-strand breaks and genome instability. Therefore, the accumulation of DNA damage in PHF2-depleted neural progenitors might result from increased DNA damage due to repeat transcription and impaired DNA repair due to high levels of H3K9me3 and reduced activity of repair enzymes.

Moreover, as discussed in the previous subchapter, PHF2 might modulate histone turnover. Peripheral tethering creates a specialized nuclear subdomain that promotes the efficient loading of factors involved in histone turnover suppression. Loss of these factors increases histone turnover, reducing H3K9me3 density and disrupting heterochromatin propagation (251,252). PHF2 might modulate histone turnover maintaining H3K9me3 density and heterochromatin stability.

Finally, our research provides a foundation for further exploration into the role of H3K9 methylation in maintaining genomic stability and regulating gene expression in various

cellular and disease contexts. The involvement of PHF2 in cancer has been extensively described, and targeting enzymes that modify H3K9me has been proposed as a potential therapeutic intervention in cancer treatment (256,257). However, this thesis results and previous experiments from the lab suggest that modifying or eliminating H3K9 methylation may induce genomic instability, posing a significant drawback for these treatment approaches. This approach might be interesting if the treatment was specifically directed to maintain H3K9me3 levels in the repetitive centromeric heterochromatin. In summary, PHF2, in addition to its described role in euchromatin and promoters, its located in the PcH boundary regions safeguarding heterochromatin integrity and preventing repetitive elements transcription that would lead to heterochromatin components dilution, DNA damage and genomic instability (Figure D1).

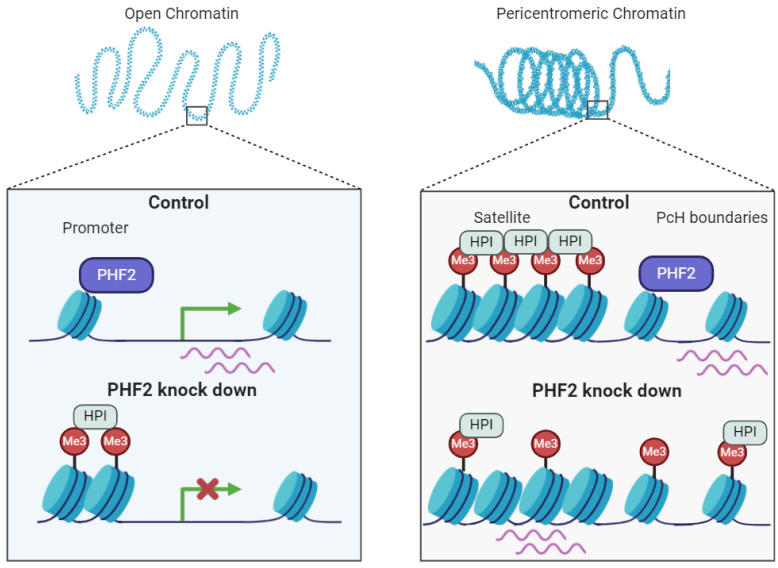


Figure D1. Model of PHF2 acting in the promoters and the pericentromeric heterochromatin boundary regions.

2. Regarding PHF2's role in mitosis

Historically, it had been described that mitosis is a process in which transcription shuts down and transcription factors and cofactors are displaced from the chromosomes (141). However, in the recent years it has been shown that a subset of factors remains bound to the chromosomes not only ensuring that proper compaction is maintained but also enabling a proper reactivation of transcription after mitosis (208,258,259). In the second chapter of results of this doctoral thesis we have confirmed that PHF2 is bound to mitotic chromosomes. This way, PHF2 was maintaining a correct level of chromosome compaction and a quick transcription reactivation of essential genes like histones (Figures R59 and R64).

2.1 About PHF2's role in chromosome compaction

Mitotic chromosome folding involves formation of increasingly compacted helically arranged nested loop arrays, with condensin II essential for helical winding and condensin I, modulating the organization within each helical turn (260). However, maintenance of this compaction through mitosis requires the interplay of multiple factors like DNA methylation, proteins and protein complexes like PRC2, and appropriate levels of histone modifications such as H3K9me3 and H3K27me3 (208,261). SUV39H1 is the methylase that deposit H3K9me3 and in Djeghloul et al. 2023, it has been showed that knock out of this protein affects also H3K9me3 levels making them almost inexistent in mitosis. This change in H3K9me3 levels was accompanied with an increase in

H3K27me3 levels and with a reduction in chromosome size as well (241). In this thesis we observe similar results with PHF2 KD. Surprisingly, although we see an increase in H3K9me3 levels in mitotic cells depleted of PHF2 (concordantly with previous results), we observe a decrease in this mark restricted to the centromere area when measured in mitotic chromosomes (Figure R60). This way, as it has been described in the first chapter of the results, there would be a decrease in H3K9me3 levels in PcH (Figures R17 and R18).

Mechanistically, PHF2 might be occupying specific genomic sites in mitosis (as depicted in figure R55) demethylating H3K9me2 or interacting with other regulatory proteins to prevent an excessive compaction. However, further experiments with the catalytic, PHD and Charged mutant would be interesting to do in order to elucidate the specific protein domains that are needed for PHF2 to regulate chromosome size.

Since H3K27me3 is enriched in mitotic chromosomes when PHF2 is depleted, it is possible that this increase is the main cause of observed reduction in size. The specific relation between H3K27me3 levels and PHF2 has not been established yet, but an indirect effect caused by low levels of H3K9me3 has been reported in Djeghloul et al. 2023, demonstrating that there is a correlation between these two histone modifications and that one can substitute or overlap functions with the other (241). To test that it would be interesting to repeat the mitotic chromosome measurements in control and shPHF2 cells but this time adding a EZH2 inhibitor, and test if that way we can rescue the phenotype of reduced chromosome size.

2.2 About PHF2's role in transcription reactivation

During mitosis, maintaining cellular identity poses a significant challenge for all dividing cells, including stem cells. These cells have a rapid cell cycle with a brief G1 phase that is crucial for initiating new transcriptional programs (262). Consequently, a critical unresolved question is how stem cells sustain their pluripotent identity after each cell cycle (263). H3K9me3 is important for mitotic chromosome structure and the efficient retention of a cohort of TFs like Esrrb, Sox2, Oct4, Tead4 or Tbx3 during mitosis (241,263). These pioneer factors and H3K27ac act as bookmarkers of the genomic loci that are needed to be expressed early on in G1 and maintain cell identity. In addition, H3K27me3 and H3K9me3, are significantly retained during mitosis. These marks are crucial for bookmarking repressive regions because the associated repressors detach from condensed chromosomes during mitosis (263). We have demonstrated that PHF2 depletion altered both H3K9me3 and H3K27me3 levels in mitotic chromosomes. This effect in such relevant mitotic marks could affect further transcription reactivation due to the inability for the RNA Pol II or other pioneer factors to bind to the chromatin. In addition to that, PHF2 might be indirectly causing this delay in transcription activation because the reduced chromosome size. A smaller, more compacted (marked by H3K27me3) chromosome would take more time to return to its natural, interphase chromatin state.

Other possible mechanism could be the regulation of LADs establishment. LADs in the mammalian genome control gene activity by establishing nuclear compartmentalization and silencing of chromatin at the nuclear lamina (264). Since LADs disappear in mitosis,

they need to be re-established in G1 to discern which chromatin loci needs to be repressed and which needs to be expressed by regulation of its nuclear localization. On top of that, PHF2 M exclusive peaks were highly colocalized with LADs (Table R2). In this case, PHF2 might be exerting a similar function to the one described in the PcH boundaries, keeping control of the accurate establishment of the LADs marking active loci, possibly in cooperation with CTCF or other LAD-related proteins.

It is necessary to mention that all the experiments performed in synchronized cells (both G1 and M phases) were in constitutive shCT and shPHF2 conditions and this could be affecting our M-specific results in terms of accumulation of effects during interphase. An interesting approach that has been tried in the laboratory is the construction of an endogenously CRISPR tagged cell line, with an inducible and reversible degron system for PHF2. Finally, the ultimate approach to study the role of PHF2 specifically in mitosis would be to tag this protein endogenously with a cyclin destruction box specific for this time frame, that way we would accomplish the protein degradation without the addition of any external drug.

3. Integrated hypothesis of PHF2 as the genome integrity safeguard

PHF2 has been described critical for neural progenitor proliferation, DNA damage repair and heterochromatin integrity (5,75,265). When impaired, these molecular processes are likely to have an impact in health. PHF2 has been related to ASD (266–268), and mice lacking this protein showed growth retardation, reduced body weight and partial neonatal

death (74). Taken together, these observations suggests that PHF2 maintains neural progenitor proliferation through regulation of cell cycle genes, keeping the homeostasis of the genome, heterochromatin and chromosome integrity, and proper transcription reactivation after mitosis.

When studying the downregulated genes in early G1, we observed that specifically histone genes were severely affected (Table R3 and Figure R66). This might be the junction between heterochromatin stability, DNA damage and cell cycle arrest. Since there is a delay in histone transcription, the histone amount and further load in DNA replication during S phase would be much slower and more demanding, probably leading to DNA Pol II stalling. This phenomenon has been described to cause R-loops and DSB (162,165,168), and therefore creating DNA damage. When PHF2 is not present, DNA repair is impaired, therefore damage is accumulated. At this point, two outcomes can occur: cell cycle checkpoints are activated and the cell stops dividing, which explain cell cycle arrest and fewer neural progenitors (giving phenotypes like ASD), or cell cycle checkpoints are overpassed and genomic instability is transmitted to the daughter cells (giving phenotypes like cancer).

Finally, this work depicts the role of PHF2 balancing H3K9me3 levels in the genome and ensuring chromatin homeostasis, accurate cell division and cell cycle progression, shedding light in how epigenetics in development reshape the future of the cell and the individual.

CONCLUSIONS

1. **PHF2 is enriched in heterochromatic genomic regions, specifically pericentromeric satellite repeats, in addition to its previously described binding at promoters.**
2. **PHF2 silences repetitive elements and maintains PcH organization.**
3. **PHF2 balances H3K9me3 levels in the genome, including promoters, heterochromatin and PcH boundaries.**
4. **PHF2 is enriched in PcH boundaries where it maintains appropriate levels of transcription.**
5. **Redistribution of limited heterochromatin components like HP1 leads to pericentromeric satellite transcription.**
6. **PHF2 regulates chromatin accessibility of cell cycle gene promoters, PcH and PcH boundary regions.**
7. **PcH stability as well as DNA damage and genomic instability of NSCs, rely on the PHD and JmjC domains within PHF2.**
8. **PHF2 remains bound to chromosomes.**
9. **PHF2 binds different genome loci in interphase compared to M phase.**
10. **PHF2 regulates chromosome size probably by maintaining proper levels of H3K9me3 and H3K27me3.**
11. **PHF2 ensures accurate transcription reactivation after mitosis, specially for chromatin organization-related genes like histones.**

BIBLIOGRAPHY

1. Deichmann U. Epigenetics: The origins and evolution of a fashionable topic. Dev Biol. 2016;416(1):249–54.

2. Delcuve GP, Rastegar M, Davie JR. Epigenetic control. Journal of Cellular Physiology. 2009; 219(2):243–50.

3. Kiefer JC. Epigenetics in development. Developmental Dynamics. 2007; 236(4):1144–56.

4. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007;128(4):693–705.

5. Pappa S, Padilla N, Iacobucci S, Vicioso-Mantis M, Álvarez de la Campa E, Navarro C, et al. PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitors. PNAS. 2019;116(39):14964-73.

6. Kornberg RD. Chromatin Structure: A Repeating Unit of Histones and DNA. Science. 1974;184(4139):868–71.

7. Salinas-Pena M, Rebollo E, Jordan A. Imaging analysis of six human histone H1 variants reveals universal enrichment of H1.2, H1.3, and H1.5 at the nuclear periphery and nucleolar H1X presence. Elife. 2023;12:RP91306.

8. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Research. 2011;21(3):381–95.

9. Zhubi A, Cook EH, Guidotti A, Grayson DR. Epigenetic mechanisms in autism spectrum disorder. International Review of Neurobiology. 2014;115:203–44.

10. Jiang C chen, Lin L shan, Long S, Ke X yan, Fukunaga K, Lu Y mei, et al. Signalling pathways in autism spectrum disorder: mechanisms and therapeutic implications. Nature Signal Transduction and Targeted Therapy. 2022;7(1):229.

11. Bogdanović O, Lister R. DNA methylation and the preservation of cell identity. Current Opinion in Genetics and Development. 2017;46:9–14.

12. Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. The Lancet. 2018;392(10149):777–86.

13. Du Q, Luu PL, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: Readers of the epigenome. Epigenomics. 2015;7(6):1051–73.

14. Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). PNAS. 2021;108(22):9055–9.

15. Okano M, Bell DW, Haber DA, Li E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell. 1999;99(3):247-57.

16. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 2016;30(7):733–50.

17. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472–9.

18. Miyajima Y, Noguchi S, Tanaka Y, Li JR, Nishimura H, Kishima M, et al. Prediction of transcription factors associated with DNA demethylation during human cellular development. Chromosome Research. 2022;30(1):109–21.

19. Bochtler M, Kolano A, Xu GL. DNA demethylation pathways: Additional players and regulators. BioEssays. 2017;39(1):1–13.

20. Jenuwein T, David Allis C. Translating the Histone Code. Science. 2001;293(5532):1074–80.

21. 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;3:363–9.

22. Lawrence M, Daujat S, Schneider R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends in Genetics. 2016;32(1):42–56.

23. Yang Y, Zhang M, Wang Y. The roles of histone modifications in tumorigenesis and associated inhibitors in cancer therapy. Journal of the National Cancer Center. 2022;2(4):277–90.

24. Gillette TG, Hill JA. Readers, writers, and erasers: Chromatin as the whiteboard of heart disease. Circulation Research. 2015;116(7):1245–53.

25. Das C, Lucia MS, Hansen KC, Tyler JK. CBP/p300-mediated acetylation of histone H3 on lysine 56. Nature. 2009;459(7243):113–7.

26. Meas R, Mao P. Histone ubiquitylation and its roles in transcription and DNA damage response. DNA Repair. 2015;36:36–42.

27. Sawicka A, Seiser C. Sensing core histone phosphorylation - A matter of perfect timing. Biochimica et Biophysica Acta. 2014; 1839(8):711–8.

28. Kuo MH, David Allis C. Roles of histone acetyltransferases and deacetylases in gene regulation. BioEssays. 1998;20(8):615-26.

29. Khan SN, Khan AU. Role of histone acetylation in cell physiology and diseases: An update. Clinica Chimica Acta. 2010;411(19-20):1401–11.

30. Yang XJ, Seto E. HATs and HDACs: From structure, function and regulation to novel strategies for therapy and prevention. Oncogene. 2007;26(37):5310–8.

31. Taylor-Papadimitriou J, Burchell JM. Histone Methylases and Demethylases Regulating Antagonistic Methyl Marks: Changes Occurring in Cancer. Cells. 2022;11(7):1113

32. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. Experimental and Molecular Medicine. 2017;49(4):324

33. Dimitrova E, Turberfield AH, Klose RJ. Histone demethylases in chromatin biology and beyond. EMBO Rep. 2015;16(12):1620–39.

34. 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;119(7):941–53.

35. Jambhekar A, Dhall A, Shi Y. Roles and regulation of histone methylation in animal development. Nature Reviews Molecular Cell Biology. 2019;20(10):625–41.

36. Black JC, Van Rechem C, Whetstine JR. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. Molecular Cell. 2012;48(4):491–507.

37. Zhang Y, Sun Z, Jia J, Du T, Zhang N, Tang J, et al. Overview of Histone Modification. Advances in Experimental Medicine and Biology. 2021;1283:1-16

38. 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;22(15):2048–61.

39. Jung JW, Lee S, Seo MS, Park SB, Kurtz A, Kang SK, et al. Histone deacetylase controls adult stem cell aging by balancing the expression of polycomb genes and jumonji domain containing 3. Cellular and Molecular Life Sciences. 2010;67(7):1165–76.

40. Kooistra SM, Helin K. Post-translational modifications: Molecular mechanisms and potential functions of histone demethylases. Nature Reviews Molecular Cell Biology. 2012;13(5):297–311.

41. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Vol. 469, Nature. 2011;469(7330):343–9.

42. Jiang H. The complex activities of the SET1/MLL complex core subunits in development and disease. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms. 2020;1863(7):194560

43. Dorafshan E, Kahn TG, Schwartz YB. Hierarchical recruitment of Polycomb complexes revisited. Nucleus. 2017;8(5):496–505.

44. Casciello F, Windloch K, Gannon F, Lee JS. Functional role of G9a histone methyltransferase in cancer. Front Immunol. 2015;6:487.

45. Stirpe A, Guidotti N, Northall SJ, Kilic S, Hainard A, Vadas O, et al. SUV39 SET domains mediate crosstalk of heterochromatic histone marks. Elife. 2021;10:e62682.

46. Herlihy CP, Hahn S, Hermance NM, Crowley EA, Manning AL. Suv420 enrichment at the centromere limits Aurora B localization and function. J Cell Sci. 2021;134(15):jcs249763.

47. Pauler FM, Sloane MA, Huang R, Regha K, Koerner M V., Tamir I, et al. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res. 2009;19(2):221–33.

48. Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications – writers that read. EMBO Rep. 2015;16(11):1467–81.

49. Fueyo R, García MA, Martínez-Balbás MA. Jumonji family histone demethylases in neural development. Cell and Tissue Research. 2015;359(1):87–98.

50. Davis K, Azarcon P, Hickenlooper S, Bia R, Horiuchi E, Szulik MW, et al. The role of demethylases in cardiac development and disease. J Mol Cell Cardiol. 2021;158:89–100.

51. Karytinos A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, et al. A novel mammalian flavin-dependent histone demethylase. Journal of Biological Chemistry. 2009;284(26):17775–82.

52. Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet. 2006;7(9):715–27.

53. 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. 2006;439(7078):811–6.

54. Tateishi K, Okada Y, Kallin EM, Zhang Y. Role of Jhdm2a in regulating metabolic gene expression and obesity resistance. Nature. 2009;458(7239):757–61.

55. 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;21(20):2545–57.

56. Jepsen K, Solum D, Zhou T, McEvilly RJ, Kim HJ, Glass CK, et al. SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. Nature. 2007;450(7168):415–9.

57. 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;128(6):1077–88.

58. 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;107(21):9671–6.

59. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature. 2006;442(7098):91–5.

60. Shao P, Liu Q, Heng Qi H. KDM7 Demethylases: Regulation, Function and Therapeutic Targeting. Advances in Experimental Medicine and Biology. 2023;1433:167–84.

61. Fortschegger K, Shiekhattar R. Plant homeodomain fingers form a helping hand for transcription. Epigenetics. 2011;6(1):4–8.

62. Horton JR, Upadhyay AK, Qi HH, Zhang X, Shi Y, Cheng X. Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat Struct Mol Biol. 2010;17(1):38–44.

63. Wen H, Li J, Song T, Lu M, Kan PY, Lee MG, et al. Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. Journal of Biological Chemistry. 2010;285(13):9322–6.

64. Feng W, Yonezawa M, Ye J, Jenuwein T, Grummt I. PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. Nat Struct Mol Biol. 2010;17(4):445–50.

65. Cao J, Yan Q. Lysine Demethylation in Pathogenesis. Adv Exp Med Biol. 2023;1433:1-14.

66. Liu W, Tanasa B, Tyurina O V., Zhou TY, Gassmann R, Liu WT, et al. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature. 2010;466(7305):508–12.

67. Siderius LE, Hamel BCJ, Van Bokhoven H, De Jager F, Van Den Helm B, Kremer H, et al. X-Linked Mental Retardation Associated With Cleft Lip/Palate Maps to Xp11.3-q21.3. J. Med. Genet. 1999;85(3):216-20.

68. Dong Y, Hu H, Zhang X, Zhang Y, Sun X, Wang H, et al. Phosphorylation of PHF2 by AMPK releases the repressive H3K9me2 and inhibits cancer metastasis. Signal Transduct Target Ther. 2023;8(1):95.

69. Baba A, Ohtake F, Okuno Y, Yokota K, Okada M, Imai Y, et al. PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. Nat Cell Biol. 2011;13(6):668–75.

70. Shi G, Wu M, Fang L, Yu F, Cheng S, Li J, et al. PHD Finger Protein 2 (PHF2) Represses Ribosomal RNA Gene Transcription by Antagonizing PHF Finger Protein 8 (PHF8) and Recruiting Methyltransferase SUV39H1. J Biol Chem. 2014;289(43):29691–700.

71. Hasenpusch-Theil K, Chadwick BP, Theil T, Heath SK, Wilkinson DG, Frischauf AM. PHF2, a novel PHD finger gene located on human Chromosome 9q22. Mammalian Genome. 1999;10(3):294–8.

72. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De Novo Gene Disruptions in Children on the Autistic Spectrum. Neuron. 2012;74(2):285–99.

73. Lee KH, Park JW, Sung HS, Choi YJ, Kim WH, Lee HS, et al. PHF2 histone demethylase acts as a tumor suppressor in association with p53 in cancer. Oncogene. 2015;34(22):2897–909.

74. Okuno Y, Ohtake F, Igarashi K, Kanno J, Matsumoto T, Takada I, et al. Epigenetic regulation of adipogenesis by PHF2 histone demethylase. Diabetes. 2013;62(5):1426–34.

75. Alonso-De Vega I, Paz-Cabrera MC, Rother MB, Wiegant WW, Checa-Rodríguez C, Hernández-Fernaud JR, et al. PHF2 regulates homology-directed DNA repair by controlling the resection of DNA double strand breaks. Nucleic Acids Res. 2020;48(9):4915–27.

76. Feng J, Chuah YH, Liang Y, Cipta NO, Zeng Y, Warrier T, et al. PHF2 regulates genome topology and DNA replication in neural stem cells via cohesin. Nucleic Acids Res. 2024;gkae457.

77. Lee KH, Ju U Il, Song JY, Chun YS. The histone demethylase PHF2 promotes fat cell differentiation as an epigenetic activator of both C/EBPα and C/EBPδ. Mol Cells. 2014;37(10):734–41.

78. Ortega-Prieto P, Postic C. Carbohydrate sensing through the transcription factor ChREBP. Front Genet. 2019;10:472.

79. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9.

80. Kim HJ, Park JW, Lee KH, Yoon H, Shin DH, Ju U Il, et al. Plant homeodomain finger protein 2 promotes bone formation by demethylating and activating Runx2 for osteoblast differentiation. Cell Res. 2014;24(10):1231–49.

81. Yang J, Ma J, Xiong Y, Wang Y, Jin K, Xia W, et al. Epigenetic regulation of megakaryocytic and erythroid differentiation by PHF2 histone demethylase. J Cell Physiol. 2018;233(9):6841–52.

82. Ghosh A, Ghosh S, Maiti GP, Mukherjee S, Mukherjee N, Chakraborty J, et al. Association of FANCC and PTCH1 with the development of early dysplastic lesions of the head and neck. Ann Surg Oncol. 2012;19(S3):S528-38.

83. Pattabiraman DR, Bierie B, Kober KI, Thiru P, Krall JA, Zill C, et al. Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. Science. 2016;351(6277):3680.

84. Lee C, Kim B, Song B, Moon KC. Implication of PHF2 expression in clear cell renal cell carcinoma. J Pathol Transl Med. 2017;51(4):359–64.

85. Hata K, Takashima R, Amano K, Ono K, Nakanishi M, Yoshida M, et al. Arid5b facilitates chondrogenesis by recruiting the histone demethylase Phf2 to Sox9-regulated genes. Nat Commun. 2013;4:2850.

86. Hilbert L, Sato Y, Kuznetsova K, Bianucci T, Kimura H, Jülicher F, et al. Transcription organizes euchromatin via microphase separation. Nat Commun. 2021;12(1):1360.

87. Tamaru H. Confining euchromatin/heterochromatin territory: Jumonji crosses the line. Genes and Development. 2010;24(14):1465–78.

88. Passarge E. Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation was Recognized Fifty Years Ago. Am J Hum Genet. 1979;31(2):106–15.

89. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. Nature Reviews Molecular Cell Biology. 2018;19(4):229–44.

90. Holliday H, Baker LA, Junankar SR, Clark SJ, Swarbrick A. Epigenomics of mammary gland development. Breast Cancer Research. 2018;20(1):100.

91. Gilbert N, Boyle S, Sutherland H, de Las Heras J, Allan J, Jenuwein T, et al. Formation of facultative heterochromatin in the absence of HP1. EMBO Journal. 2003;22(20):5540–50.

92. Wiles ET, Selker EU. H3K27 methylation: a promiscuous repressive chromatin mark. Current Opinion in Genetics and Development. 2017;43:31–7.

93. Saksouk N, Simboeck E, Déjardin J. Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin. 2015;8:3.

94. Nishibuchi G, Déjardin J. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Research. 2017;25(1):77–87.

95. Ng WS, Sielaff H, Zhao ZW. Phase Separation-Mediated Chromatin Organization and Dynamics: From Imaging-Based Quantitative Characterizations to Functional Implications. Int J Mol Sci. 2022;23(14):8039.

96. Probst A V., Almouzni G. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. Trends in Genetics. 2011;27(5):177–85.

97. Hahn M, Dambacher S, Dulev S, Kuznetsova AY, Eck S, Wörz S, et al. Suv4-20h2 mediates chromatin compaction and is important for cohesion recruitment to heterochromatin. Genes Dev. 2013;27(8):859–72.

98. Brasher S V, Smith BO, Fogh RH, Nietlispach D, Thiru A, Nielsen PR, et al. The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. The EMBO Journal. 2000;19(7):1587-97.

99. Cowieson NP, Partridge JF, Allshire RC, Mclaughlin PJ. Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. Current Biology. 2000;10(9):517–25.

100. Allshire RC, Javerzat JP, Redhead NJ, Cranston G. Position effect variegation at fission yeast centromeres. Cell. 1994;76(1):157–69.

101. HFM Peters A, Kubicek S, Mechtler K, O RJ, AHA Derijck A, Perez-Burgos L, et al. Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. Molecular Cell. 2003;12(6):1577-89.

102. Peters AH, O'Carroll D, Scherthan H, Metchler K, Sauer S, Schöfer C, et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell. 2001;107(3):323–37.

103. Li E, Bestor TH, Jaenisch R. Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality. Cell. 1992;69(6):915–26.

104. Millanes-Romero A, Herranz N, Perrera V, Iturbide A, Loubat-Casanovas J, Gil J, et al. Regulation of heterochromatin transcription by snail1/LOXL2 during epithelial-to-mesenchymal transition. Mol Cell. 2013;52(5):746–57.

105. Bulut-Karslioglu A, Perrera V, Scaranaro M, De La Rosa-Velazquez IA, Van De Nobelen S, Shukeir N, et al. A transcription factor-based mechanism for mouse heterochromatin formation. Nat Struct Mol Biol. 2012;19(10):1023–32.

106. Zhang K, Mosch K, Fischle W, Grewal SIS. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat Struct Mol Biol. 2008;15(4):381–8.

107. Padeken J, Methot SP, Gasser SM. Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. Nature Reviews Molecular Cell Biology. 2022;23(9):623–40.

108. Fischer T, Cui B, Dhakshnamoorthy J, Zhou M, Rubin C, Zofall M, et al. Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast. PNAS. 2009;106(22):8998–9003.

109. Cohen AL, Jia S. Noncoding RNAs and the borders of heterochromatin. Wiley interdisciplinary reviews RNA. 2014;5(6):835–47.

110. Wang J, Lawry ST, Cohen AL, Jia S. Chromosome boundary elements and regulation of heterochromatin spreading. Cellular and Molecular Life Sciences. 2014;71(24):4841–52.

111. Hamali B, Amine AAA, Al-Sady B. Regulation of the heterochromatin spreading reaction by trans-acting factors. Open Biology. 2023;13(11):230271.

112. Ayoub N, Noma K ichi, Isaac S, Kahan T, Grewal SIS, Cohen A. A Novel jmjC Domain Protein Modulates Heterochromatization in Fission Yeast. Mol Cell Biol. 2003;23(12):4356–70.

113. Grewal SIS. The molecular basis of heterochromatin assembly and epigenetic inheritance. Mol Cell. 2023;83(11):1767–85.

114. Cutter DiPiazza AR, Taneja N, Dhakshnamoorthy J, Wheeler D, Holla S, S Grewal SI, et al. Spreading and epigenetic inheritance of heterochromatin require a critical density of histone H3 lysine 9 tri-methylation. PNAS. 2021;118(22):e2100699118.

115. Komissarov AS, Gavrilova E V., Demin SJ, Ishov AM, Podgornaya OI. Tandemly repeated DNA families in the mouse genome. BMC Genomics. 2011;12:531.

116. Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, et al. Repetitive DNA sequence detection and its role in the human genome. Communications Biology. 2023;6(1):954.

117. Salinas-Pena M, Serna-Pujol N, Jordan A. Genomic profiling of six human somatic histone H1 variants denotes that H1X accumulates at recently incorporated transposable elements. Nucleic Acids Res. 2024;52(4):1793–813.

118. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19(1):199.

119. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014;515(7527):355–64.

120. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A V, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44–53.

121. Thakur J, Packiaraj J, Henikoff S. Sequence, chromatin and evolution of satellite DNA. International Journal of Molecular Sciences. 2021;22(9):4309.

122. Garrido-Ramos MA. Satellite DNA: An evolving topic. Genes. 2017;8(9):230.

123. Jones KW. Chromosomal and Nuclear Location of Mouse Satellite DNA in Individual Cells. Nature. 1970;225:912–5.

124. Packiaraj J, Thakur J. DNA satellite and chromatin organization at mouse centromeres and pericentromeres. Genome Biol. 2024;25(1):52.

125. Lehnertz B, Ueda Y, Derijck AAHA, Braunschweig U, Perez-Burgos L, Kubicek S, et al. Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. Current BIology. 2003;13(14):1192–200.

126. Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. Journal of Cell Biology. 2004;166(4):493–505.

127. Iwata-Otsubo A, Dawicki-McKenna JM, Akera T, Falk SJ, Chmátal L, Yang K, et al. Expanded Satellite Repeats Amplify a Discrete CENP-A Nucleosome Assembly Site on Chromosomes that Drive in Female Meiosis. Current Biology. 2017;27(15):2365-2373.

128. Almouzni G, Probst A V. Heterochromatin maintenance and establishment: Lessons from the mouse pericentromere. Nucleus. 2011;2(5):332-8.

129. Ebersole T, Kim JH, Samoshkin A, Kouprina N, Pavlicek A, White RJ, et al. tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. Cell Cycle. 2011;10(16):2779–91.

130. Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N. Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics. 2014;15:583.

131. Zeller P, Gasser SM. The Importance of Satellite Sequence Repression for Genome Stability. Cold Spring Harb Symp Quant Biol. 2017;82:15–24.

132. Hoffman EA, McCulley A, Haarer B, Arnak R, Feng W. Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. Genome Res. 2015;25(3):402–12.

133. Bayona-Feliu A, Casas-Lamesa A, Reina O, Bernués J, Azorín F. Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. Nat Commun. 2017;8(1):283.

134. Ma Y, Kanakousaki K, Buttitta L. How the cell cycle impacts chromatin architecture and influences cell fate. Frontiers in Genetics. 2015;6:19.

135. Karataş A. A Phase Overshadowed by Mitotic Division: Interphase. J Biol Educ. 2023;57(2):317–30.

136. Matthews HK, Bertoli C, de Bruin RAM. Cell cycle control in cancer. Nature Reviews Molecular Cell Biology. 2022;23(1):74–88.

137. Atlasi Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development. Nature Reviews Genetics. 2017;18(11):643–58.

138. Espinosa-Martínez M, Alcázar-Fabra M, Landeira D. The molecular basis of cell memory in mammals: The epigenetic cycle. Sci. Adv. 2024;10(9):eadl3188.

139. Probst A V., Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nature Reviews Molecular Cell Biology. 2009;10(3):192–206.

140. Güttinger S, Laurell E, Kutay U. Orchestrating nuclear envelope disassembly and reassembly during mitosis. Nat Rev Mol Cell Biol. 2009;10:178–91.

141. Martinez-Balbás MA, Dey A, Rabindran SK, Ozato K, Wu C. Displacement of Sequence-Specific Transcription Factors from Mitotic Chromatin. Cell. 1995;83(1):29-38.

142. Westhorpe FG, Straight AF. The centromere: Epigenetic control of chromosome segregation during mitosis. Cold Spring Harb Perspect Biol. 2015;7(1):a015818.

143. Ohta S, Bukowski-Wills JC, Sanchez-Pulido L, Alves F de L, Wood L, Chen ZA, et al. The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. Cell. 2010;142(5):810–21.

144. Nagpal H, Fierz B. The Elusive Structure of Centro-Chromatin: Molecular Order or Dynamic Heterogenetity? Journal of Molecular Biology. 2021;433(6):166676.

145. John S, Workman JL. Bookmarking genes for activation in condensed mitotic chromosomes. BioEssays. 1998;20(4):275–9.

146. Palozola KC, Donahue G, Liu H, Grant GR, Becker JS, Cote A, et al. Mitotic transcription and waves of gene reactivation during mitotic exit. Science. 2017;358(6359):119–22.

147. Sarnataro S, Riba A, Molina N. Regulation of transcription reactivation dynamics exiting mitosis. PLoS Comput Biol. 2021;17(10):e1009354.

148. Bellec M, Radulescu O, Lagha M. Remembering the past: Mitotic bookmarking in a developing embryo. Current Opinion in Systems Biology. 2018;11:41–9.

149. Friedberg EC. A brief history of the DNA repair field. Cell Res. 2008;18(1):3–7.

150. Aguilera A, Gómez-González B. Genome instability: A mechanistic view of its causes and consequences. Nature Reviews Genetics. 2008;9(3):204–17.

151. Pal S, Tyler JK. Epigenetics and aging. Science Advances. 2016;2(7):e1600584.

152. Feng JX, Riddle NC. Epigenetics and genome stability. Mammalian Genome. 2020;31(5–6):181–95.

153. Aguilera A, García-Muse T. Causes of genome instability. Annual Review of Genetics. 2013;47:1–32.

154. Boos D, Frigola J, Diffley JFX. Activation of the replicative DNA helicase: Breaking up is hard to do. Current Opinion in Cell Biology. 2012;24(3):423–30.

155. Patel JA, Zezelic C, Rageul J, Saldanha J, Khan A, Kim H. Replisome dysfunction upon inducible TIMELESS degradation synergizes with ATR inhibition to trigger replication catastrophe. Nucleic Acids Res. 2023;51(12):6246–63.

156. Sogo JM, Lopes M, Foiani M. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. Science. 2002;297(5581):599–602.

157. De Piccoli G, Katou Y, Itoh T, Nakato R, Shirahige K, Labib K. Replisome Stability at Defective DNA Replication Forks Is Independent of S Phase Checkpoint Kinases. Mol Cell. 2012;45(5):696–704.

158. French S. Consequences of replication fork movement through transcription units in vivo. Science. 1992;258(5086):1362–5.

159. Liu B, Alberts BM. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. Science. 1995;267(5201):1131–7.

160. Brüning JG, Marians KJ. Bypass of complex co-directional replication-transcription collisions by replisome skipping. Nucleic Acids Res. 2021 Sep 27;49(17):9870–85.

161. Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD. The nature of mutations induced by replication-transcription collisions. Nature. 2016;535(7610):178–81.

162. Thomas M, White RL, Davis RW. Hybridization of RNA to double-stranded DNA: formation of R-loops. PNAS. 1976;73(7):2294–8.

163. Roberts RW, Chrothers DM. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science. 1992;258(5087):1463–6.

164. Westover KD, Bushnell DA, Kornberg RD. Structural Basis of Transcription: Separation of RNA from DNA by RNA Polymerase II. Science. 2004;303(5660):1014–6.

165. Petermann E, Lan L, Zou L. Sources, resolution and physiological relevance of R-loops and RNA–DNA hybrids. Nature Reviews Molecular Cell Biology. 2022;23(8):521–40.

166. Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes and Development. 2014;28(13):1384–96.

167. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature. 2014;516(7531):436–9.

168. Marabitti V, Valenzisi P, Lillo G, Malacaria E, Palermo V, Pichierri P, et al. R-Loop-Associated Genomic Instability and Implication of WRN and WRNIP1. International Journal of Molecular Sciences. 2022;23(3):1547.

169. Yao Y, Dai W. Genome instability and cancer. J Carcinog Mutagen. 2014;5:73–90.

170. Golubov A. Genome instability in bacteria: Causes and consequences. Genome Stability: From Virus to Human Application. Elsevier; 2021;26:73–90.

171. Wilson SW, Houart C. Early Steps in the Development of the Forebrain. Dev Cell. 2004;6(2):167–81.

172. Molyneaux BJ, Arlotta P, Menezes JRL, Macklis JD. Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci. 2007;8(6):427–37.

173. Kim H jin, Hur SW, Park JB, Seo J, Shin JJ, Kim S young, et al. Histone demethylase PHF2 activates CREB and promotes memory consolidation. EMBO Rep. 2019;20:1–18.

174. Al-Mousawi J, Boskovic A. Transcriptional and epigenetic control of early life cell fate decisions. Curr Opin Oncol. 2022;34(2):148–54.

175. Barry DS, Pakan JMP, McDermott KW. Radial glial cells: Key organisers in cns development. International Journal of Biochemistry and Cell Biology. 2014;46(1):76–9.

176. Tiberi L, Vanderhaeghen P, van den Ameele J. Cortical neurogenesis and morphogens: Diversity of cues, sources and functions. Curr Opin Cell Biol. 2012;24(2):269–76.

177. Fernández V, Llinares-Benadero C, Borrell V. Cerebral cortex expansion and folding: what have we learned? EMBO J. 2016;35(10):1021–44.

178. Ornitz DM, Itoh N. The fibroblast growth factor signaling pathway. Wiley Interdiscip Rev Dev Biol. 2015;4(3):215–66.

179. Shimojo H, Ohtsuka T, Kageyama R. Oscillations in Notch Signaling Regulate Maintenance of Neural Progenitors. Neuron. 2008;58(1):52–64.

180. 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. Congenital Anomalies. 2012;52(2):72–7.

181. Chenn A, Walsh CA. Regulation of Cerebral Cortical Size by Control of Cell Cycle Exit in Neural Precursors. Sience. 2002;297(5580):365–89.

182. Ohtaka-Maruyama C, Okado H. Molecular pathways underlying projection neuron production and migration during cerebral cortical development. Frontiers in Neuroscience. 2015:9:447.

183. Siegenthaler JA, Miller MW. Transforming growth factor β1 promotes cell cycle exit through the cyclin-dependent kinase inhibitor p21 in the developing cerebral cortex. Journal of Neuroscience. 2005;25(38):8627–36.

184. Vogel T, Ahrens S, Büttner N, Krieglstein K. Transforming growth factor β 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. 2010;20(3):661–71.

185. Ross SE, Greenberg ME, Stiles CD. Basic Helix-Loop-Helix Factors in Cortical Development. Neuron. 2003;39(1):13-25.

186. Wilkinson G, Dennis D, Schuurmans C. Proneural genes in neocortical development. Neuroscience. 2013;253:256–73.

187. Kojima Y, Tam OH, Tam PPL. Timing of developmental events in the early mouse embryo. Seminars in Cell and Developmental Biology. 2014;34:65–75.

188. Cadwell CR, Bhaduri A, Mostajo-Radji MA, Keefe MG, Nowakowski TJ. Development and Arealization of the Cerebral Cortex. Neuron. 2019;103(6):980–1004.

189. Taverna E, Götz M, Huttner WB. The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. Annual review of cell and developmental biology. 2014;30(9):465–502.

190. Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD, et al. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat Neurosci. 2010;13(4):423–30.

191. Sergeeva A, Davydova K, Perenkov A, Vedunova M. Mechanisms of human DNA methylation, alteration of methylation patterns in physiological processes and oncology. Gene. 2023;875:147487.

192. Yao B, Christian KM, He C, Jin P, Ming GL, Song H. Epigenetic mechanisms in neurogenesis. Nature Reviews Neuroscience. 2016;17(9):537–49.

193. Park J, Lee K, Kim K, Yi SJ. The role of histone modifications: from neurodevelopment to neurodiseases. Signal Transduction and Targeted Therapy. 2022;7:217.

194. Li Y, Mo Y, Chen C, He J, Guo Z. Research advances of polycomb group proteins in regulating mammalian development. Frontiers in Cell and Developmental Biology. 2024;12:1383200.

195. Duan R, Du W, Guo W. EZH2: A novel target for cancer treatment. Journal of Hematology and Oncology. 2020;13(1):104.

196. Park J, Lee K, Kim K, Yi SJ. The role of histone modifications: from neurodevelopment to neurodiseases. Signal Transduction and Targeted Therapy. 2022;7:217.

197. Iacobucci S, Padilla N, Gabrielli M, Navarro C, Lombardi M, Vicioso-Mantis M, et al. The histone demethylase PHF8 regulates astrocyte differentiation and function. Development. 2021;148(12):dev194951.

198. Ma DK, Bonaguidi MA, Ming GL, Song H. Adult neural stem cells in the mammalian central nervous system. Cell Research. 2009;19(6):672–82.

199. Hu F, Wang X, Liang G, Lv L, Zhu Y, Sun B, et al. Effects of epidermal growth factor and basic fibroblast growth factor on the proliferation and osteogenic and neural differentiation of adipose-derived stem cells. Cell Reprogram. 2013;15(3):224–32.

200. Spencer Currle, Jia Sheng Hu, Aaron Kolski-Andreaco ESM. Culture of mouse neural stem cell precursors. Journal of Visualized Experiments. 2006;132(2):3559.

201. 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;24(3):351–60.

202. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 2006;443(7108):167–72.

203. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, et al. Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. Cell. 2012;151(1):206–20.

204. Iacobucci S, Padilla N, Gabrielli M, Navarro C, Lombardi M, Vicioso-Mantis M, et al. The histone demethylase PHF8 regulates astrocyte differentiation and function. Development. 2021;148(12):dev194951.

205. Graham FL, Smileyt J, Russell WC, Nairn R. Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. J gen Virol. 1977;36:59–66.

206. Todaro GJ, Green H. Quantitative studies of the growth of mouse embryo cells in culture and their development into stablished lines. Journal of Cell Biology. 1963;May(17):299–313.

207. Blanco-García N, Asensio-Juan E, De La Cruz X, Martínez-Balbás MA. Autoacetylation regulates P/CAF nuclear localization. Journal of Biological Chemistry. 2009;284(3):1343–52.

208. Djeghloul D, Patel B, Kramer H, Dimond A, Whilding C, Brown K, et al. Identifying proteins bound to native mitotic ESC chromosomes reveals chromatin repressors are important for compaction. Nat Commun. 2020;11(1):4118.

209. Le Scouarnec S, Gribble SM. Characterising chromosome rearrangements: Recent technical advances in molecular cytogenetics. Heredity (Edinb). 2012 Jan;108(1):75–85.

210. Vicioso-Mantis M, Aguirre S, Martinez-Balbás MA. JmjC Family of Histone Demethylases Form Nuclear Condensates. Int J Mol Sci. 2022;23(14):7664.

211. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: An open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676–82.

212. Sánchez-Molina S, Estarás C, Oliva JL, Akizu N, Asensio-Juan E, Rojas JM, et al. Regulation of CBP and Tip60 coordinates histone acetylation at local and global levels during Ras-induced transformation. Carcinogenesis. 2014;35(10):2194–202.

213. Lungu C, Pinter S, Broche J, Rathert P, Jeltsch A. Modular fluorescence complementation sensors for live cell detection of epigenetic signals at endogenous genomic sites. Nat Commun. 2017;8(1):649.

214. Panatta E, Butera A, Mammarella E, Pitolli C, Mauriello A, Leist M, et al. Metabolic regulation by p53 prevents R-loop-associated genomic instability. Cell Rep. 2022;41(5):111568.

215. Gustafsson N, Culley S, Ashdown G, Owen DM, Pereira PM, Henriques R. Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. Nat Commun. 2016;7(12471):1–9.

216. Culley S, Albrecht D, Jacobs C, Pereira PM, Leterrier C, Mercer J, et al. Quantitative mapping and minimization of super-resolution optical imaging artifacts. Nat Methods. 2018;15(4):263–6.

217. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213–8.

218. Valls E, Blanco-García N, Aquizu N, Piedra D, Estarás C, de la Cruz X, et al. Involvement of chromatin and histone deacetylation in SV40T antigen transcription regulation. Nucleic Acids Res. 2007;35(6):1958–68.

219. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008 Dec;322(5909):1845–8.

220. Lopes R, Agami R, Korkmaz G. GRO-seq, A Tool for Identification of Transcripts Regulating Gene Expression. Methods in Molecular Biology. 2017;1543:45-55.

221. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

222. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

223. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(1):160–5.

224. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

225. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

226. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

227. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics. 2015;31(22):3593–9.

228. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011;29(1):24–6.

229. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. Nature Genetics. 2000;25(1):25–9.

230. Clough E, Barrett T. The Gene Expression Omnibus database. Methods in Molecular Biology. 2016;1418:93–110.

231. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50(1):543–52.

232. Jones K. Chromosomal and Nuclear Location of Mouse Satellite DNA in Individual Cells. Nature. 1970;225(5236):912–5.

233. Pardue ML, Gall JG. Chromosomal Localization of Mouse Satellite DNA. Science. 1970;168(3937):1356–8.

234. Grewal S. The molecular basis of heterochromatin assembly and epigenetic inheritance. Mol Cell. 2023;83(11):1767–85.

235. Martens JHA, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO Journal. 2005;24(4):800–12.

236. Donze D, Kamakaka R. RNA polymerase II and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. EMBO Journal. 2001;20(3):520–31.

237. Lunyak V, Prefontaine G, Núñez E, Cramer T, Ju B, Ohgi K, et al. Developmentally Regulated Activation of a SINE B2 Repeat as Domain Boundary in Organogenesis. Science. 2007;317(5835):245–8.

238. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, et al. Human tRNA genes function as chromatin insulators. EMBO Journal. 2012;31(2):330–50.

239. Scott KC, Merrett SL, Willard HF. A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. Current Biology. 2006;16(2):119–29.

240. Ebersole T, Kim JH, Samoshkin A, Kouprina N, Pavlicek A, White RJ, et al. tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. Cell Cycle. 2011;10(16):2779–91.

241. Djeghloul D, Dimond A, Cheriyamkunnel S, Kramer H, Patel B, Brown K, et al. Loss of H3K9 trimethylation alters chromosome compaction and transcription factor retention during mitosis. Nat Struct Mol Biol. 2023;30(4):489-501.

242. Sarnataro S, Riba A, Molina N. Regulation of transcription reactivation dynamics exiting mitosis. PLoS Comput Biol. 2021;17(10):e1009354.

243. Haldar S, Saini A, Nanda JS, Saini S, Singh J. Role of Swi6/HP1 self-association-mediated recruitment of Clr4/Suv39 in establishment and maintenance of heterochromatin in fission yeast. Journal of Biological Chemistry. 2011;286(11):9308–20.

244. Zofall M, Sandhu R, Holla S, Wheeler D, Grewal SIS. Histone deacetylation primes self-propagation of heterochromatin domains to promote epigenetic inheritance. Nat Struct Mol Biol. 2022;29(9):898–909.

245. Ragunathan K, Jih G, Moazed D. Epigenetic inheritance uncoupled from sequence-specific recruitment. Science. 2015;348(6230):1258699.

246. Aygün O, Mehta S, Grewal SIS. HDAC-mediated suppression of histone turnover promotes epigenetic stability of heterochromatin. Nat Struct Mol Biol. 2013;20(5):547–54.

247. Raiymbek G, An S, Khurana N, Gopinath S, Larkin A, Biswas S, et al. An H3K9 methylation dependent protein interaction regulates the non-enzymatic functions of a putative histone demethylase. Elife. 2020;9:e53155.

248. Trewick SC, Minc E, Antonelli R, Urano T, Allshire RC. The JmjC domain protein Epe1 prevents unregulated assembly and disassembly of heterochromatin. EMBO Journal. 2007;26(22):4670–82.

249. Decombe S, Loll F, Caccianini L, Affannoukoué K, Izeddin I, Mozziconacci J, et al. Epigenetic rewriting at centromeric DNA repeats leads to increased chromatin accessibility and chromosomal instability. Epigenetics Chromatin. 2021;14(1):35.

250. Fodor BD, Kubicek S, Yonezawa M, O'Sullivan RJ, Sengupta R, Perez-Burgos L, et al. Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. Genes Dev. 2006;20(12):1557–62.

251. Murawska M, Greenstein RA, Schauer T, Olsen KCF, Ng H, Ladurner AG, et al. The histone chaperone FACT facilitates heterochromatin spreading by regulating histone turnover and H3K9 methylation states. Cell Rep. 2021;37(5):109944.

252. Noma K ichi, David Allis C, S Grewal SI. Transitions in Distinct Histone H3 Methylation Patterns at the Heterochromatin Domain Boundaries. Science. 2001;293(5532):1150–5.

253. Janssen A, Colmenares SU, Karpen GH. Heterochromatin: Guardian of the Genome. Annu Rev Cell Dev Biol. 2018;34:265–88.

254. Zheng H, Chen L, Pledger WJ, Fang J, Chen J. P53 promotes repair of heterochromatin DNA by regulating JMJD2b and SUV39H1 expression. Oncogene. 2014;33(6):734–44.

255. Domínguez-Sánchez MS, Barroso S, Gómez-González B, Luna R, Aguilera A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS Genet. 2011;7(12):e1002386.

256. Wagner KM, McReynolds CB, Schmidt WK, Hammock BD. Soluble epoxide hydrolase as a therapeutic target for pain, inflammatory and neurodegenerative diseases. Pharmacology and Therapeutics. 2017;180:62–76.

257. Kondo Y, Shen L, Ahmed S, Boumber Y, Sekido Y, Haddad BR, et al. Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells. PLoS One. 2008;3(4):e2037.

258. Swedlow JR, Hirano T. The Making of the Mitotic Chromosome: Modern Insights into Classical Questions. Molecular Cell. 2003;11(3):557-69.

259. Caravaca JM, Donahue G, Becker JS, He X, Vinson C, Zaret KS. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. Genes Dev. 2013;27(3):251–60.

260. Gibcus JH, Samejima K, Goloborodko A, Samejima I, Naumova N, Nuebler J, et al. A pathway for mitotic chromosome formation. Science. 2018;359(6376):eaao6135.

261. Zhiteneva A, Bonfiglio JJ, Makarov A, Colby T, Vagnarelli P, Schirmer EC, et al. Mitotic post-translational modifications of histones promote chromatin compaction in vitro. Open Biol. 2017;7(9):170076.

262. Dalton S. Linking the Cell Cycle to Cell Fate Decisions. Trends in Cell Biology. 2015;25(10):592–600.

263. Huang X, Wang J. Mitotic Bookmarking: Maintaining the Stem Cell Identity during Mitosis. Cell Stem Cell. 2017;20(6):741–2.

264. Zullo JM, Demarco IA, Piqué-Regi R, Gaffney DJ, Epstein CB, Spooner CJ, et al. DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. Cell. 2012;149(7):1474–87.

265. Aguirre S, Pappa S, Serna-Pujol N, Padilla N, Iacobucci S, Nacht AS, et al. PHF2-mediated H3K9me balance orchestrates heterochromatin stability and neural progenitor proliferation. EMBO Rep. 2024. Jun 18.

266. Cotney J, Muhle RA, Sanders SJ, Liu L, Willsey AJ, Niu W, et al. The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat Commun. 2015;6:6404.

267. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee YH, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A. 2015;112(41):5600–7.

268. Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental disorder risk genes with autism and developmental disability biases. Nat Genet. 2017;49(4):515–26.

APPENDIX

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

- Aguirre S, Pappa S, Serna-Pujol N, Padilla N, Iacobucci S, Nacht AS, Vicent GP, Jordan A, de la Cruz X, Martínez-Balbás MA. PHF2-mediated H3K9me3 balance orchestrates heterochromatin stability and neural progenitor proliferation. EMBO R. 2024 Jun 18. DOI: 10.1038/s44319-024-00178-7

- Vicioso-Mantis M, Aguirre S, Martínez Balbás MA. JmjC Family of Histone Demethylases Form Nuclear Condensates. International Journal of Molecular Sciences. 2022; 23(14):7664. DOI: 10.3390/ijms23147664