UNIVERSITAT DE BARCELONA

Role of the histone demethylase PHF2 during early neurogenesis

Stella Pappa

DOCTORAL PROGRAM IN GENETICS
DEPARTMENT OF GENETICS, MICROBIOLOGY AND STATISTICS
FACULTY OF BIOLOGY
UNIVERSITY OF BARCELONA

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

Role of the histone demethylase PHF2 during early neurogenesis

---

A thesis submitted by Stella Pappa to obtain the doctoral degree by University of Barcelona

Author:
Stella Pappa

Directed by:
Dr. Marian Martinez Balbás

Tutored by:
Dr. Pedro Martinez Serra

Barcelona 2019

Acknowledgements

Looking back to the beginning of this long journey, I would like to mention and thank all the people that stand for me during these years and help me make possible to write this thesis. To start with, I would like to say to Marian how grateful I am that she gave me the opportunity to be in her lab these last 4 years. Next to you I didn't only widen my horizons into the fascinating world of science and research but I learned to believe in myself and to take decisions on my own even if they were the wrong ones. Every time I would enter your office, sometimes bit disappointed when things were not working out as expected, you gave me such motivation. By the end of our discussion, I wanted to try a million new things. You were always transmitting me your excitement of all the projects running in the lab. For all these and many more reasons I want to thank you. You are a great mentor. These years, you have created an amazing environment in the lab.

My love in Biology had started when I was very young. My mom, a biologist herself helped me growing up in an environment where one of my favourite presents was a microscope. She transmitted me her love for ecology and the importance of protecting the nature. But most of all I want to thank her because all my life I was trying to be like her. She is the person I admire most, the person that made me be like I am today. This journey would look nothing like that if it wasn't for you. Θέλω να σου πώ ακόμα μια φορά πως σε αγαπώ πάρα πάρα πολύ. To my dad, I want to say that you are the most supportive person I know. All my life you were next to me, caring more than anyone else, always encourage me to do what I want, convincing me to take great steps and great decisions. I know that having you on my life I am not scared of anything. Σε αγαπώ μέχρι το φεγγάρι, είσαι ο πιο υποστηρικτικός μπαμπάς του κόσμου. Τίποτα απο ότι έχω καταφέρει μέχρι σήμερα θα ήταν δυνατό χωρίς εσένα δίπλα μου. To my brother, I want to say that one of the hardest parts moving to Barcelona was leaving him these 5 years. Growing up together was one of the most amazing things that happened in my life. I miss you terribly. Finally, the last persons I want to acknowledge from my family are my grandparents. To the happiest person I know, who spreads happiness and joy when she is around, my grandmother or Omi as I will always call her. This smile on her leaps always makes me feel nice. Omi, Ich liebe dich! Θα θυμάμαι πάντα τις βόλτες μας για παγωτό πριν το φαγητό, που πάντα φώναζε ο παππούς. To

my grandfather, the smartest person in my family who has been teaching me all my life. Since I was little he was helping me study for school. Thank you for all the knowledge you have transmitted me. I may didn't study to be an archeologist as you were always trying to convince me but because of you I learned to love the greek culture and even enjoy the endless tours in Acropolis and the museums. Θα ήθελα πως και τι να ξανά γυρίσουμε στην Ακρόπολη και στο τέλος να με κεράσεις μία γρανίτα φράουλα. To my grandmother, Dimitra, the best cook of my family who she was always teaching me all the popular poems since I was 6 years old. I really admired the way you remembered all these by heart.

Of course this journey wouldn't look like this at all if it wasn't for the girls in the lab. Thank you all for being such good friends and sharing the excitement of your projects and achievements. Thank you for teaching me all the techniques I know, for sharing your passion for science. Thank you Mari and Alejandra for being the first that helped me when I arrived to the lab. Raquel, I have told you many times how much you have inspired me and I thank you for that. Thank you for helping me with my experiments and always care and discuss them with me. I have learned so much from you. Thank you Simona for being always so awesome lab-mate, for letting me participate to your work and discuss it with you. Thank you Marta, you are an amazing person, your way of being always so positive and happy is really inspiring. Claudia, your devotion to your work is overwhelming. I am so sure that you will have a great future ahead. Thank you Elia, for giving me the chance to teach you everything I learned in the lab. Your motivation made me feel very proud for being your teacher. Finally, I want to thank all the people from the institute that are always so friendly and open to help each other. It was a great experience to do my PhD in this institute and get to know all of you guys.

I would like to thank many of my friends from Athens that we grew up together. We may be now spread across the whole world but you have been always there for me when I needed you. Thank you Maria for the endless coffee conversations in sweet habit, thank you Stavroula for sharing my passion in science and our endless discussions that no one else understood. Thank you Foteini for the amazing time we are spending in London, every time I visit you and of course thank you Nick for being an amazing friend, even if you left Greece with 15 years old. Our friendship stayed the same either you were living in London, Boston or San Francisco. Guys I love you

all so much. We may be living all so far from each other but we always find time to spend amazing summers in Greece. I couldn't forget to thank my friends who turned out to be my new family in Barcelona. To Elena, Ruth, Miriam, Arnau and Paula that shared the same journey with me during these 5 years. Os quiero un montón, me habéis hecho sentrir como en casa. Of course I want to thank the whole team of Sam that let me write this thesis in their office. Getting to know all of you guys was amazing, you really made me feel welcome and helped me write this thesis in a really nice environment.

Last, but definitely not least I want to thank the most important person in my life, my soulmate, my life partner, my fiancé, my everything. Agapi, this thesis is dedicated to you with all my heart. One paragraph is not enough to tell you for how many reasons I thank you. Thank you for making me discover this amazing country, your culture, your language. I now feel that I have a part of Spanish in me. But most of all I want to thank you for converting me to a better person every day, for teaching me to dream big and to follow these dreams. Thank you for calming me down when I get stressed for always being there for me, for reminding me of three little birds. Adri, you are the best person I know in this world. I couldn't be more proud of you. Your help during all these years was unbelievably huge. I couldn't make it happen without you.

To Adrián, my partner in life

Contents

List of Figures

List of Tables

List of Abbreviations

Arg	**Arg**inine
ASD	**A**utism **S**pectrum **D**isorder
ATP	**A**denosine **T**hiotri**P**hosphate
BIR	**B**reak **I**nduced **R**eplication
BMP	**B**one **M**orphogenetic **P**roteins
bps	**b**ase **p**airs
BSA	**B**ovine **S**erum **A**lbumin
Cdks	**C**yclin **d**ependent **k**inases
ChIP-seq	**Ch**romatin **I**mmuno **P**recipitation **Seq**uencing
ChIP	**Ch**romatin **I**mmuno **P**recipitation
Chr	**Chr**omosome
CKIs	**C**DK **I**nhibitors
CNS	**C**entral **N**ervous **S**ystem
CO-IP	**CO** **I**mmuno**P**recipitation
DAPI	4′,6 **D**iamidino 2 **P**henylindole
DMSO	**D**i**M**ethyl **S**ulf**o**xide
DNA	**D**eoxyribonucleic **A**cid
DNTP	**D**eoxynucleotide
DOX	**DOX**ycyline
DSBs	**D**ouble **S**trand **B**reaks
dsDNA	**d**ouble **s**tranded DNA
EGF	**E**pidermal **G**rowth **F**actor
EP	**E**lectro**P**oration
ESCs	**E**mbryonic **S**tem **C**ells
FAD	**F**lavin **A**denine **D**inucleotide
FBS	**F**etal **B**ovine **S**erum
FGF	**F**ibroblast **G**rowth **F**actor
GEO	**G**ene **E**xpression **O**mnibus
GF	**G**rowth **F**actors

GO	**G**ene **O**ntology
HAT	**H**istone **A**cetyl**T**ransferases
HDACs	**H**istone **D**eacetylaces
HDM	**H**istone **D**e**M**ethylase
HH	**H**amburger **H**amilton
HKMTs	**H**istone Lysine **M**ethyl **T**ransferases
HP1	**H**eterochromatin **P**rotein 1
HR	**H**omologous **R**ecombination
IDT	**I**ntegrated **D**NA **T**echnologies
JmJC	**J**umonJi-C
LINES	**L**ong **I**nterspersed **E**lements
LTR	**L**ong **T**erminal **R**epeat
Lys	**Lys**ine
MLL	**M**ixed **L**ineage **L**eukemia
MS	**M**ass **S**pectrometry
MZ	**M**antle **Z**one
NGS	**N**ext **G**eneration **S**equencing
NHEJ	**N**on **H**omologous **E**nd **J**oining
NP	**N**eural **P**late
NPCs	**N**eural **P**rogenitor **C**ells
NSCs	**N**eural **S**tem **C**ells
NT	**N**eural **T**ube
PBS	**P**hosphate **B**uffered **S**aline
PcG	**P**olycomb **G**roup proteins
PE	**P**ost **E**lectroporation
PEV	**P**osition **E**ffect **V**ariegation
PHC	**P**ericentromeric **H**eterochromatin
PNK	**P**oly**N**ucleotide **K**inase
PTMs	**P**ost **T**ranslational **M**odifications
qPCR	**q**uantitative **P**olymerase **C**hain **R**eaction
rcf	**r**elative **c**entrifugal **f**orce
RG	**R**adial **G**lial
RIPA	**R**adioimmunoprecipitation **A**ssay buffer
RNA-seq	**RNA Seq**uencing
RNAPII	**R**NA **P**olymerase II
RT	**R**everse **T**ranscription
SD	**S**tandard **D**eviation

SDS	**S**odium **D**odecyl **S**ulfate
SEM	**S**tandard **E**rror of the **M**ean
Shh	**Sonic H**edge **H**og
SINES	**S**hort **I**nterspersed **E**lements
ssDNA	**S**ingle **S**tranded DNA
SVZ	**S**ub**V**entricular **Z**one
TBE	**T**ris **B**orate **E**DTA
TFs	**T**ranscription **F**actors
TSS	**T**ranscription **S**tart **S**ite
UV	**U**ltra **V**iolet
UCSC	**U**niversity of **C**alifornia **S**anta **C**ruz
VZ	**V**entricular **Z**one
WT	**W**ild **T**ype
XLMR	**X** Linked **M**ental **R**etardation

Chapter 1

Introduction

1.1 Chromatin dynamics

Chromatin structure is intimately connected with gene expression and cell identity. The complexity of multicellular organisms is due to their capacity to produce and maintain a multitude of different cell types that share the same genomic DNA. Such complexity requires tight regulation of gene expression to unambiguously specify and constrain the developmental paths taken by cells in the embryo. Signaling molecules and networks of expressed transcription factors (TFs) cooperate to control cellular fate (Perino and Veenstra, 2016). Multiple studies by different groups led to the widely accepted model of nucleosomal organization of chromatin (Kornberg, 1974), and visualized in 1997 by the landmark X-ray crystal structure of the histone octamer–DNA particle (Luger et al., 1997). Chromatin is the state in which DNA is packaged within the cell. The nucleosome is the fundamental unit of chromatin and it is composed of an octamer of the four core histones (H3, H4, H2A, H2B) around which 146 base pairs of DNA are wrapped. This unit is repeated throughout the genome and the chromatin is further condensed through association of additional proteins, for example the linker histone H1 (Kouzarides, 2007). Pioneering work carried out between the late 80's and the early 90's defined nucleic acids, chromatin and histone proteins, which led to the cytological distinction between euchromatin and heterochromatin. In 1928, Emil Heitz, developed chromatin staining methods and discovered that chromosomes are composed of two different regions. Regions that are not stained after telophase (euchromatin) and regions that are stained throughout the cell cycle (heterochromatin). He described that staining patterns are chromosome-specific and

later suggested that genes are found in euchromatin, whereas heterochromatin is genetically inert (Allshire and Madhani, 2018). Observations with optical and electronic microscopy led to the classical bimodal description of eukaryotic chromatin: the gene-rich, permissive to transcription euchromatin, is positioned at the center of the nucleus, while heterochromatin, that corresponds to a gene-poor and compacted form of chromatin is generally positioned at the nuclear periphery (Ea et al., 2015). One can further distinguish two types of heterochromatin: constitutive heterochromatin is conserved from one cell type to another while facultative heterochromatin is not conserved. The characteristic heterochromatic foci (dark dots) are visualized by DAPI staining of AT-rich repeat sequences is seen in the Figure 1.1.

In the next section of the introduction I will focus on the main components of the euchromatin and heterochromatin.

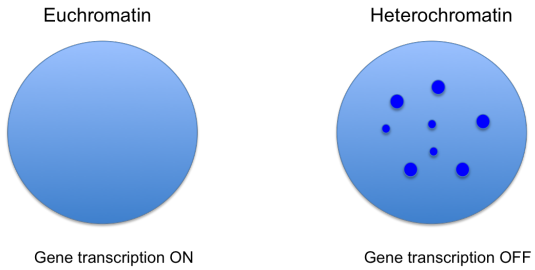


FIGURE 1.1: Schematic representation of two interphase nuclei from female mouse somatic cells. Adapted from (Allis and Jenuwein, 2016)

1.1.1 Epigenetic modifications

The British developmental biologist, Conrad Waddington was the first to introduce the term "epigenetics" to refer to the mechanisms of acquisition of stable cell fates during development (Noble and Waddington, 2015). He represented the developmental process as a series of decisions that could be represented as "valleys" and "forks" in a developmental landscape. Later on, this term has been repeatedly modified in order to be more specific. Robin Holliday defined epigenetics as the inheritance of changes in gene expression patterns and, more generally, the inheritance of any change in

gene function that does not involve a change in DNA sequence (Holliday, 1987). Chromatin is decorated by modifications and binding of different factors that are specific for the function of the sequence and its transcriptional state. Chromatin modifications act at various scales of size, from topological organization of chromatin, chromatin accessibility, and histone modifications to DNA methylation (Kornberg, 1974), as seen in Figure 1.2

In this doctoral thesis, I will focus on histone modifications and the enzymes that catalyze these modifications.

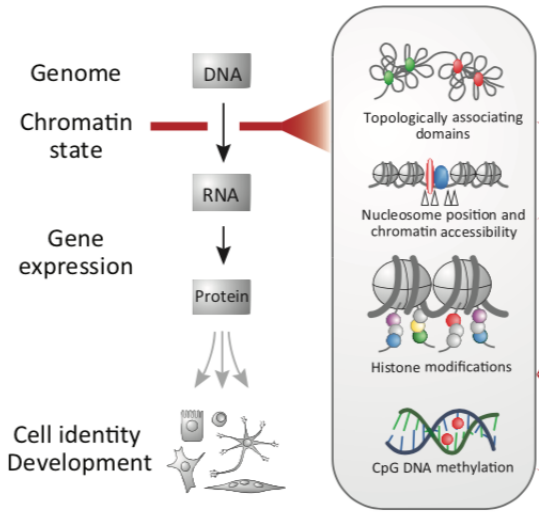


FIGURE 1.2: Overview of epigenetic modifications and their role in development. Adapted from (Perino and Veenstra, 2016)

1.1.2 Histone modifications

The abundance of lysine residues within the histone tails and their ability to be posttranslationally modified provides a massive combinatorial repertoire, denoted a "histone code", and has the potential to regulate many chromatin-templated functions (Jenuwein and Allis, 2001). Regulatory sequences together with TFs and coactivator complexes, have the ability to influence the state of chromatin by recruiting chromatin-modifying enzymes. These enzymes deposit post-translational modifications (PTMs)

on the histones. The most relevant for transcriptional regulation but also well-studied are PTMs occurring on lysines of the histone N-terminal tails, particularly of histone H3 (Ausio, Dong, and Holde, 1989).

The core histones are predominantly globular except for their N-terminal "tails", which are unstructured. A striking feature of histones, and most specifically of their tails, is the large number and type of modified residues they possess. Histones can be modified at many sites. There are over 60 different residues on histones where modifications have been detected either by specific antibodies or by mass spectrometry (Kouzarides, 2007). The most well-characterized modifications include acetylation, methylation, phosphorylation, ubiquitylation, and SUMOylation. Each of these, influence the chromatin accessibility and depending on the site, the degree and the type of modification, they have different functional outcomes, as seen in the Table 1.1. Acetylation is historically the most widely studied modification on histone tails, and enzymes that are responsible for both the addition and removal of acetyl groups have been identified and extensively studied (Kuo and Allis, 1998; Khan and Khan, 2010). Later on and more specifically during the past decade, the focus has shifted more towards the study of histone methylation. This was due to the discovery of enzymes that can catalyze the addition and removal of methyl groups to lysine (Lys) and arginine (Arg) residues on histone tails (Kooistra and Helin, 2012). Before the discovery of the first histone demethylase Lys-specific demethylase 1 (LSD1 also known as KDM1A, BHC110 and AOF2) (Shi et al., 2004), it was thought that methylation of histone residues was relatively permanent and could only be removed by histone exchange or by dilution during DNA replication.

In the next subsections, I will provide more details regarding histone methylation/demethylation because they have been studied during this doctoral thesis.

1.1.2.1 Histone methylation

Histone methylation consists in the addition of one, two or three methyl groups to Lys or Arg, and is performed by histone Lys methyl transferases (HKMTs) or Arg methyltransferases, respectively. This reaction does not change the electrical charge of the amino acid, but it appeared to have functional consequences (Rea et al., 2000). All histones have the capability to be methylated on one or more residues, with some residues to be

TABLE 1.1: Overview of different classes of chromatin modifications identified on histones and their function. Adapted from (Kouzarides, 2007)

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

targeted more frequently than others. Commonly methylated sites are Lys 4, -9, -27, -36 and -79 of histone H3 and Lys 20 of histone H4. There is variation not only in the location of methylation sites but also in the degree of methylation at these sites. In mammalian cells, up to 40–80% of all histones are dimethylated at positions H3K9, H3K27, H3K36 or H4K20. Less abundant is the mono- and trimethylation of these residues, such methylation marks occur on up to 20% of histones (Schotta et al., 2008; Jung et al., 2010). These patterns have become apparent from chromatin immunoprecipitation (ChIP) experiments using antibodies to specific histone modifications (Lee and Mahadevan, 2009). Regarding the gene regulation, methylation of H3K4, H3K36 and H3K79 is associated with actively transcribed genes, whereas H3K27me2/3, H3K9me2/3 and H4K20me2/3 are associated with inactive genes. On repressed genes, the H3K9 and H4K20 methylation marks are relatively homogeneously distributed, whereas H3K27me3 is enriched at promoters (Kooistra and Helin, 2012). Within actively transcribed genes, some methylation marks, such as H3K4me3, are mainly found at the promoter area or around the transcription start site (TSS), whereas others, such as H3K36me3, are found in the gene bodies. Interestingly, H3K27me3

and H3K4me3 coexist on some promoters called bivalent promoters. These are found on developmental genes in stem cell states and endow the cell with the capacity of a rapid activation upon developmental cues (Bernstein et al., 2006). H4K20me3 enrichment peaked specifically at pericentromeric chromatin, but not gene deserts or subtelomeric loci (Rosenfeld et al., 2009) and H3K27me3 is enriched at silent genes and subtelomeric chromatin (Rosenfeld et al., 2009). Both modifications occur less frequently at other non-genic regions (Rosenfeld et al., 2009; Pauler et al., 2009). Regarding H3K9me2/3 and H3K27, these modifications can be found in expansive domains (Pauler et al., 2009).

Taken together, these examples may be few but demonstrate that both genic and non-genic regions have unique combinations of histone modifications (Figure 1.3). These combinations allow diverse structural features to be created and modified to distinguish subtypes of heterochromatin.

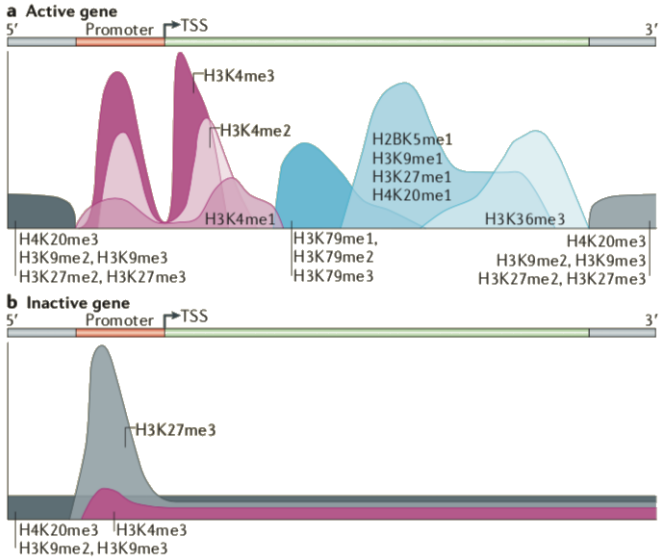


FIGURE 1.3: Schematic view of chromatin and histone methylation. Adapted from (Kooistra and Helin, 2012)

1.1.2.2 Methylation on H3K9

One histone modification that I would like to highlight in the introduction of this thesis is methylation of H3K9. As mentioned before, it can be monomethylation (me), dimethylation (me2) or trimethylation (me3). It has been described that during development H3K9me2 is implicated in the silencing of genes that play important roles in chromatin homeostasis and development (Epsztejn-Litman et al., 2008). During the differentiation of human embryonic stem cells (ESCs) into neural progenitor cells (NPCs), pluripotency-related genes and non-neural lineage genes gain the H3K9me2 mark concomitantly with the loss of H3K4me3 (Golebiewska et al., 2009). Subsequently, these genes gain the H3K9me3 mark, leading to their long-term repression in terminally differentiated neurons. In ECSs, it has been proposed that H3K9me2 increases across the genome as cells differentiate and acquire lineage specificity. They showed that mammalian cells acquire large regions of H3K9me2 during differentiation, which affect at least 30% of the genome. These regions are evolutionarily conserved and linked to changes in gene expression in a tissue-specific manner, and they are substantially lost in human cancer cell lines (Wen et al., 2009). Although this has been quite controversial (Filion and Steensel, 2010). Specifically, H3K9me2 is enriched at lineage non-specific genes, suggesting that acquisition of H3K9me2 is critical for gene silencing during differentiation (Shinkai and Tachibana, 2011). H3K9me2 demarcates heterochromatin, particularly non-genic regions. Importantly, H3K9me2 together with H3K9me3 are essential components of the constitutive heterochromatin that is found close to centromeres, telomeres and interspersed repetitive elements (Allshire et al., 1994; Saksouk, Simboeck, and Dejardin, 2015; Lehnertz et al., 2003; Nakayama et al., 2001) while it has been described that H3K9me1 is enriched at active genes (Rosenfeld et al., 2009). H3K9me2 is also prevalent in gene deserts, with little being observed at individual active or silent genes (Rosenfeld et al., 2009). H3K9me2 domains are also strongly correlated with binding of Lamin B1. Mapping of Lamin B1 interacting domains demonstrated that inactive genes are preferentially located at the nuclear periphery in association with the nuclear lamina (Guelen et al., 2008). Regarding the role of H3K9me2/3 in the pericentromeric regions I will detail it in the next section of this introduction.

1.1.2.3 Methyltransferases

There are many enzymes that catalyze methylation (Lan and Shi, 2009). The discovery of the first HKMT combined insights from the position effect variegation (PEV) modifier factors that contained an evolutionarily conserved SET domain in *Drosophila melanogaster* (Tschiersch et al., 1994) with the characterization of the mammalian orthologues (Aagaard et al., 1999). The SET domain is present in Su(var)3–9, E(z) and Trithorax proteins, all of which had been implicated in epigenetic regulation but they did not show any evidence of enzymatic activity. Later on, its catalytic activity was predicted (Jenuwein et al., 1998). Robust catalytic activity of the SET domain of recombinant SUV39H1 was shown to methylate histone H3 *in vitro*. Follow-up studies described that SUV39H1 methylates H3K9me3, selectively. (Rea et al., 2000). The SET domain of SUV39H1 provided a signature catalytic domain, and numerous SET-domain-containing proteins were then tested as potential KMTs. As discussed above, histone lysine methylation can either be activating, such as H3K4 methylation or repressive, such as SUV39H1-mediated H3K9me3. Many groups identified silenced or active chromatin states that are controlled during development (such as, H3K9me2 by G9a (Tachibana et al., 2001), H3K27me3 by Polycomb and Enhancer of zeste homologue 2 (EZH2) (Kuzmichev et al., 2002), and H3K4me3 by Trithorax and mixed-lineage leukaemia (MLL) (Yokoyama et al., 2004). Biochemically, both SUV39H1 and G9a can catalyze mono-, di-, and trimethylation reactions on H3K9 (Collins et al., 2005; Kubicek et al., 2007). Investigations of G9a-deficient and Suv39h1/h2-double-deficient cells demonstrated that SUV39H1/H2 are crucial HKMTs for H3K9me3 on pericentromeric heterochromatin, and that G9a is a major H3K9me1 and H3K9me2 HKMT of euchromatin (Tachibana et al., 2002). G9a is also involved in H3K9me3 modification *in vivo* (Yokochi et al., 2009).

1.1.2.4 Demethylases

As mentioned in the previous subsection, before the discovery of LSD1, it was thought that methylation of histone residues was relatively permanent. The identification of the enzyme family that LSD1 belongs to, and of an entirely separate and much larger family of Jumonji C (JMJC) histone demethylating enzymes, has resulted in a completely different view of histone methylation in which this modification is now considered to be

highly dynamic (Kooistra and Helin, 2012). There are two evolutionarily conserved families of histone demethylases, which utilize different reaction mechanisms to establish demethylation. These two families are: LSD demethylases and JmJC demethylases (Pedersen and Helin, 2010; Klose, Kallin, and Zhang, 2006).

1.1.2.5 LSD family

LSD1, the first reported histone demethylase, is able to demethylate H3K4me1 and H3K4me2 and can also catalyze demethylation towards H3K9me1 and H3K9me2, as well as non-histone targets. The other LSD family member, LSD2 has so far only been shown to demethylate H3K4me1 and H3K4me2 (Karytinos et al., 2009). Both, use a flavin adenine dinucleotide (FAD)- dependent amine oxidation reaction to catalyze the demethylation of their substrate. As seen in the Figure 1.4, due to the requirement of a free electron pair at the methylated residue (Lys), LSD1 is only able to demethylate mono- and dimethylated, but not trimethylated, Lys residues (Shi et al., 2004).

1.1.2.6 JmJC family

The JMJC family utilize Fe(II) and 2-oxoglutarate in the demethylation reaction in which the quaternary complex formed by the substrate bound to the catalytic domain, together with Fe2+ and α-ketoglutarate reacts with oxygen, as seen in the Figure 1.4. The oxidative decarboxylation of α-ketoglutarate coordinated with the hydroxylation of the target methyl group creates a hydroxymethyl ammonium intermediate that gives rise to formaldehyde and the demethylation product (Tsukada et al., 2006). The first JMJC histone demethylase (HDM) was identified in 2006 (Tsukada et al., 2006). Since then, this family has grown greatly in number, and it includes enzymes that target me1, me2, and me3 histones (Kooistra and Helin, 2012; Fueyo, Garcia, and Martinez-Balbas, 2015). Histone demethylases can be classified into seven subfamilies according to their structural domains and specificity for histones (Figure 1.5).

In this thesis, I will particularly introduce a subfamily of the Lysine demethylases (KDMs), the KDM7, due to its relevance in the experimental section.

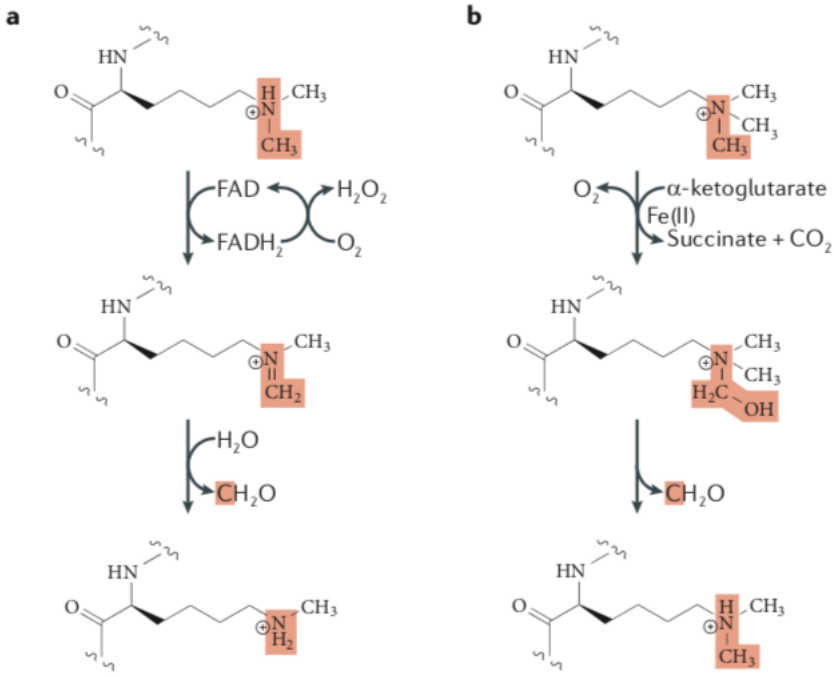


FIGURE 1.4: The catalytic mechanisms of histone demethylation. (a) The flavin adenine dinucleotide (FAD)-dependent amine oxidation reaction is used by LSD1 and LSD2. (b) The JMJC family uses a dioxygenase reaction that is dependent on Fe(II) and α-ketoglutarate. Adapted from (Kooistra and Helin, 2012)

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

FIGURE 1.5: The two histone demethylase families and their histone substrates. Adapted from (Fueyo, Garcia, and Martinez-Balbas, 2015)

1.1.2.7 KDM7 family of demethylases

A class of JmjC-domain chromatin-modifying enzymes is characterized by a single N-terminal plant homeodomain (PHD) zinc finger, a domain that was shown to be associated with methylated lysine residues (Li et al., 2006). In humans this group consists of three members: the plant homeodomain fingers 2 and 8 (PHF2 and PHF8) and KDM7A (KIAA1718) (Fortschegger and Shiekhattar, 2011) as shown in Figure 1.6. Structures

of their PHDs and JmjC-domains have now been solved by X-ray crystallography (Horton et al., 2010; Wen et al., 2010). Besides conserved zinc-chelating residues, these PHDs comprise a patch of phenylalanine and tyrosine residues called aromatic cage, which can occur in several domains and can interact with methylated lysine. Biochemical experiments demonstrated that the three proteins interact specifically with histone H3 methylated at lysine 4 via their PHDs (Feng et al., 2010; Fortschegger et al., 2010; Wen et al., 2010). Active JmjC-domains comprise several conserved amino acids that are needed for correct binding of the crucial cofactors Fe2+ and α-ketoglutarate. Besides those, the KDM7-family displays four short additional α-helices in the carboxy-terminal region of the JmjC which are essential for activity (Fortschegger and Shiekhattar, 2011). PHF8 preferentially acts on H3K9me2/1 and H4K20me1, KDM7A mainly demethylates H3K9me1/2 and H3K27me1/2 and PHF2 demethylates H3K9me1/2 (Horton et al., 2010; Liu et al., 2010; Tsukada, Ishitani, and Nakayama, 2010; Huang et al., 2010a; Fortschegger and Shiekhattar, 2011). Interestingly, little is known about the properties of the C-terminal halves of KDM7 proteins. They do not contain any known protein domains, and the homology between the three human proteins is low in this region compared to that of PHD and JmjC but all three human proteins contain a putative coiled-coil region. PHF2 and PHF8 also exhibit several phosphorylation sites which appear to be important for the regulation of their activity (Liu et al., 2010). Finally, KDM7 family has been described as a new class of transcriptional co-activators. By virtue of association with H3K4me3 and removal of H3K9me2/1, H3K27me2/1 or H4K20me1 modifications, KDM7 proteins create a more permissive chromatin environment at promoters (Fortschegger and Shiekhattar, 2011). PHF8 is the best-known member of this subfamily partly because mutations in the PHF8 gene can cause Siderius-Hamel syndrome, an X-linked mental retardation (XLMR), which is often accompanied by cleft lip and/or cleft palate (Siderius et al., 1999). The mutations mostly lead to truncations before or in the JmjC domain. The equilibrium between histone methylation and demethylation plays an essential role in development. Thus, it is not surprising that specific alterations in methyltransferase and demethylase activities correlate with neurodevelopmental disorders (Fueyo, Garcia, and Martinez-Balbas, 2015). PHF8 HDM activity was essential to promote cytoskeleton dynamics, and it was required for proper neurite outgrowth in mouse primary

cortical neurons (Asensio-Juan, Gallego, and Martinez-Balbas, 2012). Although PHF8 is the most studied member of the KDM7 subfamily, another protein of this subfamily has been associated with neural development. When KDM7A was knocked down, neural differentiation was blocked in mouse ESCs. This proneural effect was due to the direct transcriptional activation of FGF4, a signaling component implicated in neural differentiation (Huang et al., 2010a) as well as, promoted neural induction in early chick embryos where its overexpression led to an expansion of the neural plate (Huang et al., 2010b). Despite of PHF8 and KDM7A, the role of PHF2 in neural development and neurogenesis is not yet known.

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

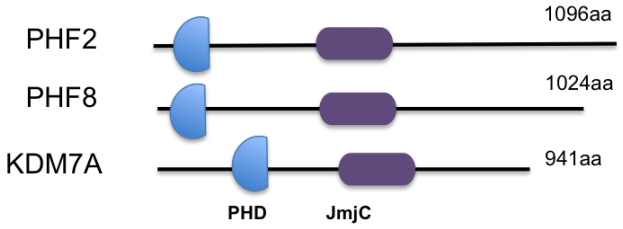


FIGURE 1.6: Schematic representation of human KDM7 family of proteins. Adapted from (Wen et al., 2010)

1.1.2.8 PHF2 demethylase

Despite the importance of H3K9 methylation during cell identity and development, little is known about the role of the enzymes responsible for this mark during neurogenesis and neural development. Biochemical studies demonstrated that PHF2 demethylates H3K9me2 upon interaction with H3K4me2/3 through its PHD domain (Wen et al., 2010). During the past years very few groups had shown interest in studying PHF2. This could be due to the fact that this protein was thought to be enzymatically inactive, because instead of the second conserved Fe2+ binding histidine, contains a tyrosine residue at this position (Horton et al., 2011). It was shown to become an active H3K9me2 demethylase through PKA-mediated phosphorylation (Baba et al., 2011). PHF2 was first characterized as a novel

PHD finger gene which maps to human Chromosome (Chr) 9q22 close to D9S196. The mouse homolog was also characterized and mapped to the region on mouse chromosome 13. The predicted human and mouse proteins are 98% identical (Hasenpusch-Theil et al., 1999). PHF2 was first identified as a candidate gene for hereditary sensory neuropathy type I (Nicholson et al., 1996) because it is expressed at high levels in the neural tube and dorsal root ganglia (Hasenpusch-Theil et al., 1999). PHF2 mutations have been found in patients with autism spectrum disorder (ASD) (Cotney et al., 2015; C Yuen et al., 2017; Iossifov et al., 2012). Regarding the physiological role of PHF2 *in vivo* in obesity and metabolic syndrome, PHF2 has been shown to be a co-activator of multiple transcription factors. Systemic Phf2 null mice showed partial neonatal death and growth retardation and exhibited less adipose tissue and reduced adipocyte numbers compared with control littermates (Okuno et al., 2013). Furthermore, working with transcription factors, PHF2 regulates different differentiation processes, among them the osteoblast differentiation (Kim et al., 2014). PHF2 plays an essential role in bone formation as shown in PHF2 transgenic mice in comparison to wild-type ones. The calvarial and limb bones of PHF2 transgenic mice developed earlier. Moreover, PHF2 is shown to regulate the adipocyte (Lee et al., 2014) as well as the erythroid differentiation processes (Yang et al., 2018). PHF2 acts as an epigenetic co-regulator of the adipogenic transcription factors C/EBPα and C/EBPδ as seen in Figure 1.7. Depletion of PHF2 dramatically repressed adipogenesis by regulating the expression of major genes related to fat cell differentiation, including CEBPA and PPARG (Lee et al., 2014). In the case of erythroid differentiation, PHF2 acts as a negative epigenetic regulator of erythroid differentiation via regulation of p53 expression by demethylation of H3K9me2 in the promoter region of p53 gene (Yang et al., 2018). In addition, this histone demethylase is implicated in cancer. Alterations in PHF2 have been identified in several types of cancer (Ghosh et al., 2013; Lee et al., 2017; Pattabiraman et al., 2016; Lee et al., 2015). Interestingly, it acts as a tumor suppressor in association with p53 in cancer development as converted the chromatin that is favorable for transcription of p53 by demethylating H3K9me2 (Lee et al., 2015) (Figure 1.8). Moreover, PHF2 represses rDNA transcription by competing with PHF8 for binding of rDNA promoter and by recruiting H3K9me2/3 methyltransferase SUV39H1 (Shi et al., 2014). Finally, it has been described

that PHF2 can form a complex with ARID5B protein. PHF2 can demethylate ARID5B at Lys 336 as a previous step for DNA binding of the complex and subsequent gene activation (Baba et al., 2011). It has also been described the role of ARID5B in PHF2 recruitment to the chondrocytic genes a study that associates PHF2 with chondrogenesis (Hata et al., 2013).

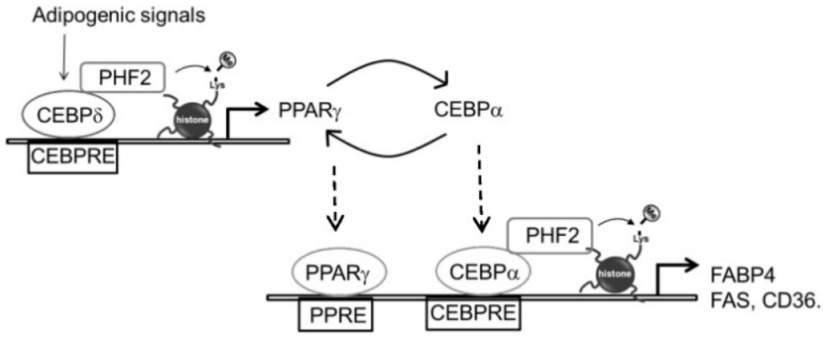


FIGURE 1.7: Schematic illustration of the role of PHF2 in C/EBPα and C/EBPδ transcriptional activation. Adapted from (Lee et al., 2014)

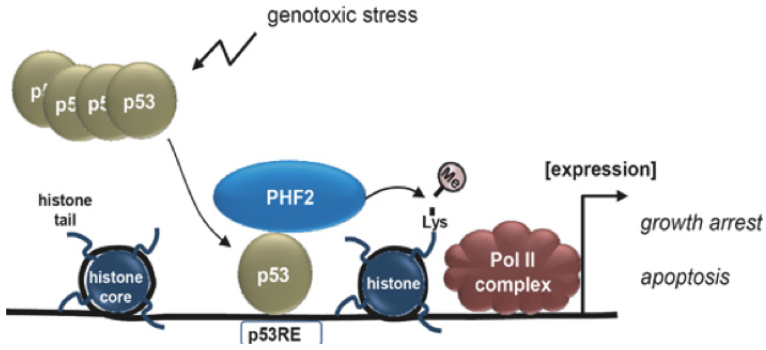


FIGURE 1.8: Schematic representation of the role of PHF2 as tumor supressor, demethylating p53 targets. Adapted from (Lee et al., 2015)

1.1.3 Chromatin changes during cell cycle

At this point of the doctoral thesis I would like to emphasize on the role of histone modifications and the enzymes that catalyze these modifications for cell cycle regulation.

1.1.3.1 Enzymes involved in cell cycle regulation

The cell cycle is a succession of coordinated events that ends up with cell division. Includes four phases (G1, S, G2 and mitosis) and it is highly regulated to ensure the correct inheritance of information to daughter cells. Cyclin-dependent kinases (Cdks) are modulated by interactions with cyclins and Cdk inhibitors (CKIs). Cooperation between these three is necessary for ensuring orderly progression through the cell cycle (Lim and Kaldis, 2013).

Cdks contain a serine/threonine-specific catalytic core and they partner with regulatory subunits known as cyclins, which control kinase activity and substrate specificity. Cdk/cyclin complexes were first implicated in cell cycle control based on previous work in yeast, in which a single Cdk was found to promote transitions between different cell cycle phases through its interactions with various cyclins (Beach, Durkacz, and Nurse, 1982). Accordingly, Cdks are the ones driving cell cycle progression whereas cyclins are changed to aid the transition between cycle phases. The kinase activity of Cdk/cyclin complexes is tightly regulated by plenty of Cdk inhibitors (CKIs), which serve as brakes to halt cell cycle progression under unfavorable conditions (Morgan, 2007).

The involvement of cell cycle regulators in transcription has been a long-standing affair and one of the best-characterized examples is the retinoblastoma (Rb)/E2F pathway (Weinberg, 1995). They discovered that the tumor suppressor protein pRB interacted with E2F and blocked its transcriptional activity before the G1/S transition. This discovery established an early model for cell cycle control (Goodrich et al., 1991; Nevins et al., 1991). Apart from that, its connection with pRB, and therefore cancer, greatly increased efforts towards the understanding on how E2F controls cell proliferation. When the first E2F family members were identified, these have led many people to search for the transcriptional targets of E2F. These targets could play a role as key regulators of cell cycle progression. In search of E2F-binding sites in the promoters of genes that were already known to

be induced at the G1/S transition, many E2F targets were identified. As a consequence, E2F was found to regulate several additional genes with functions in cell cycle control (for example CCNE1 and CDC25A) and DNA replication (MCM2-7 and CDC6) (Nevins, 2001; Muller and Helin, 2000).

One example of how Cdk/cyclin complexes regulate cell cycle is highlighted in Figure 1.9. During the G1 phase of the cell cycle, the complexes of Cdk4/cyclin D (cycD) and Cdk2/cyclin E (cycE) sequentially phosphorylate Rb, a phosphorylation that leads to the activation of E2F proteins and consequently the expression of E2F-responsive genes. This cluster of genes encodes cell cycle regulators important for G1/S transition (cyclin E, cyclin A (cycA) and Cdk1), proteins involved in nucleotide biosynthesis (thymidine kinase) and components of the DNA replication machinery such as Cdc6 and the origin recognition complex subunit 1 (Orc1). During the G2 phase of the cell cycle, the complexes of Cdk2/cyclin A and Cdk1/cyclin B (cycB) phosphorylate FoxM1, leading to the relief of its self-inhibition and finally the recruitment of the histone deacetylase p300/CREB binding protein (CBP) that activates the expression of FoxM1 target genes. These genes encode cyclin B, a cell cycle regulator which is required for the execution of mitosis and interactors of the kinetochore complex required for proper chromosome segregation such as the centromere protein F (Cenpf). The effects of Cdk phosphorylation on FoxM1 can be counteracted by the phosphatase PP2A/B55α (Lim and Kaldis, 2013).

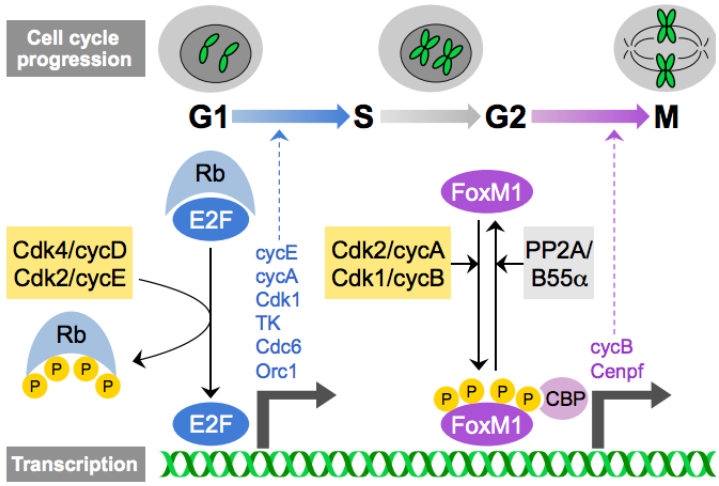


FIGURE 1.9: Schematic representation of Cdk/cyclin complexes regulate Rb/E2F- and FoxM1-mediated transcription. Adapted from (Lim and Kaldis, 2013)

1.1.3.2 Epigenetic dynamics during cell cycle

One of the major mechanism driving cell cycle regulation is the orchestrated transcriptional program of gene expression. Although cyclins and cdks are the main regulators of cell cycle, epigenetic modulators regulate gene expression throughout the cell cycle, which in turn ensures correct inheritance of the epigenetic marks to the new daughter cells. During G1, the so-called growth phase of cell cycle, the large transcriptional activity has been linked to a decrease in the levels of DNMT1 and DNMT3b (Robertson et al., 2000), followed by a decrease in global DNA methylation levels (Brown et al., 2007). The global decrease in DNA methylation is accompanied by global acetylation of nucleosomes, which facilitates the activation of different sets of genes, including the E2F family of transcription factors (Zhang and Dean, 2001; Macaluso, Montanari, and Giordano, 2006). Moreover, it has been demonstrated that the expression of several E2F regulated genes, including the G1/S expressed cyclins, was significantly decreased in cells lacking both subunits of the HDM PRC1/2 complexes EZH2 and EED

(Bracken et al., 2003). Several studies have shown that E2F4 forms a complex with the Rb family member p130 and histone deacetylases (HDACs) in quiescent and early-G1 cells, which binds to promoters containing E2F sites. This mediates local de-acetylation of H3 leading to gene repression (Takahashi, Rayman, and Dynlacht, 2000). Moreover, global methylation levels of the histone H3 and histone H4 tails have been shown to increase in late G1 and S-phase (Takahashi, Rayman, and Dynlacht, 2000). H3K27 is rapidly monomethylated on newly synthesized histones, which is probably due to the localization of EZH2 to sites of replication (Scharf, Barth, and Imhof, 2009). Furthermore, other histone modifications have been shown to be relevant to proper cell cycle progression, such as H4K20me3. It was detected at its lowest levels in S-phase, in agreement with the predicted negative correlation with H4K16Ac (Strahl and Allis, 2000). In addition, KMTs and KDMs can impact cell cycle by directly modulating expression of important cell cycle genes. One example that demonstrates that demethylases can affect these genes, is PHF8 demethylase that is regulated over cell cycle with peak levels in G1/S, which corresponds to the role of PHF8 coactivating E2F1 targets and G1/S transition (Liu et al., 2010). Finally, PTMs are well documented during mitosis in several cell types, mainly because of their importance for chromatin condensation and chromosomal segregation. Histone H3 phosphorylation is low until late interphase and spreads throughout the genome until prophase, where it constitutes a mitotic hallmark (Hendzel et al., 1997).

1.2 Chromatin states

After introducing the histone modifications and their role in chromatin dynamics, in this section I would like to focus on the two states of the chromatin and their epigenetic features. The DNA in the nucleus exists in two forms that reflect the level of activity of the cell (Figure 1.10).

1.2.1 Euchromatin

As mentioned in the previous section, by the term "euchromatin" is described the light-staining, decondensed and transcriptionally accessible regions of the genome. Euchromatin is marked by a variety of chromatin

modifications that antagonize the assembly of heterochromatin. These include the histone variant H2A.Z, which is deposited in response to the formation of nucleosome-free regions at the first nucleosome downstream of transcription start sites. Euchromatin contains the majority of single-copy genes, is the site of active gene expression, although not all genes are capable of being activated. Histone modifications associated with active transcription include lysine acetylation and methylation. Typically, histone acetylation occurs at multiple lysine residues, most commonly on histones H3 and H4 (Saksouk, Simboeck, and Dejardin, 2015). Regarding methylation, methylation at H3K4, H3K36 and H3K79 is typically associated with active transcription as mentioned in the subsection 1.1.2.1

1.2.2 Heterochromatin

Heterochromatin is the fundamental architectural feature of eukaryotic chromosomes that endows particular genomic regions with specific functional properties. The term "heterochromatin", as mentioned in the previous section corresponds to the dark-staining, condensed and gene-poor regions of the genome but now generally refers to molecular subtypes of transcriptionally repressed chromatin domains that extend beyond a single gene or regulatory element (Allshire and Madhani, 2018). The main roles of heterochromatin include asserting cell-type-specific transcription, centromere function and preventing genetic instability. The most prominent histone feature in heterochromatin is global hypoacetylation, which leads to chromatin fiber compaction. As described in the subsection 1.1.2.1 the best-studied types of heterochromatin are marked by the addition of two (me2) or three (me3) methyl groups to histone H3 lysine 9 (H3K9me) or lysine 27 (H3K27me). Another hallmark of heterochromatin is DNA methylation where a direct link between the readers of this modification and the histone lysine methylation system has been demonstrated. Methyl-CpG binding proteins also bind to a multitude of different chromatin modifying enzymes, including histone deacetylases and histone lysine methyltransferases. One example is the methyl-CpG-binding protein, MeCP2 that acts as a bridge between DNA methylation and histone methylation (Fuks et al., 2003). In this doctoral thesis I will focus on H3K9 methylation-dependent

heterochromatin, as well as centromeric and pericentromeric heterochromatin. In the next subsections I will detail the two types of heterochromatin; the constitutive and the facultative. A typical mark enriched on constitutive heterochromatin is the H3K9me3, while H3K27me3 is usually present on facultative heterochromatin (Saksouk, Simboeck, and Dejardin, 2015).

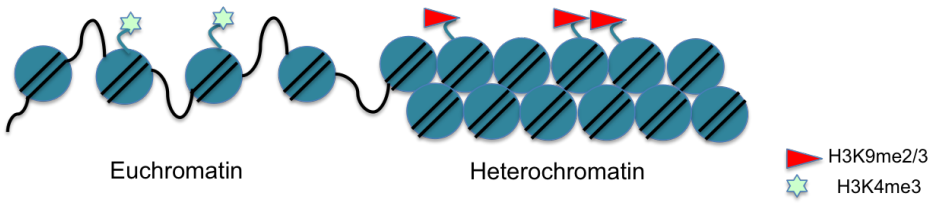


FIGURE 1.10: Chromatin is found in two states: euchromatin (transcriptionally competent) and heterochromatin (transcriptionally silent).

1.2.2.1 Facultative heterochromatin

During development, as cells become increasingly specialized, there is a trend of euchromatin converting to heterochromatin. The formation of "facultative" heterochromatin, as is called, is thought to mediate developmental control of gene activity, which becomes increasingly canalized as differentiation progresses. One well-studied example of developmental change is the inactivation of one of the two mammalian X chromosomes in female embryos near the time of gastrulation (LYON, 1962). In other words, facultative heterochromatin refers to a type of chromatin that may form at various chromosomal regions, which usually contain genes that must be kept silent upon developmental cues. It refers to locus-specific and cell-type-specific heterochromatin (Saksouk, Simboeck, and Dejardin, 2015).

1.2.2.2 Constitutive heterochromatin

Constitutive heterochromatin is believed to occur at the same genomic regions in every cell type and these regions usually do not contain genes.

Hence, is often viewed as a more static structure than facultative heterochromatin (Saksouk, Simboeck, and Dejardin, 2015). In most eukaryotes, is consistently formed throughout the cell cycle. In most organisms, the bulk of constitutive heterochromatin forms at pericentromeric regions, at telomeres and gene-poor areas that are usually made of tandem repetitions, that vary in DNA size from 5 bp to a few hundred bp and are called satellite repeats (Eymery, Callanan, and Vourc'h, 2009; Saksouk, Simboeck, and Dejardin, 2015). Constitutive heterochromatin is found also at ribosomal regions, telomeres that are constituted by a repeated short conserved DNA motif (TTAGGG) and harbor enrichment in H3K9me3. Telomeres are bound by conserved protein machineries to protect chromosomal ends from being recognized as double-strand breaks. Finally, constitutive heterochromatin is found at different loci along the chromosome such as repetitive DNA sequences that include simple repeats, DNA transposons, LTR-endogenous retroviral elements, non-LTR autonomous retrotransposons including long interspersed elements (LINES) and short interspersed elements (SINES) (Saksouk, Simboeck, and Dejardin, 2015) as seen in Figure 1.11.

In the next subsection I will focus on the centromeric and pericentromeric regions of the chromatin due to its relevance to the results section.

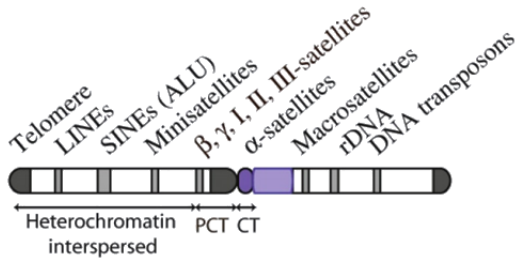


FIGURE 1.11: Schematic representation of the organization of constitutive heterochromatin in a human chromosome. Adapted from (Saksouk, Simboeck, and Dejardin, 2015)

1.2.2.2.1 Centromeric and pericentromeric heterochromatin

Chromosomes undergo major changes in structure and organization during the cell cycle. They condense during mitosis and become visible as

chromosomal constrictions. The specialized nature and environment of centromeric chromatin enables the assembly of the kinetochore, which is a large, multi-protein complex that attaches to microtubules during cell division (Fukagawa and Earnshaw, 2014). Kinetochores are maintained by a balance of histone PTMs and transcription. They ensure equal parts of genetic material between daughter cells. After cell division, chromatin decondenses, the structure and biochemical composition of centromeres change, following by the kinetochore disassembly. During mitosis, this decompaction is visualized by weak DNA staining on individual chromosomes. In interphase, densely stained chromocenters become visible in mouse cells. They correspond to pericentromeric heterochromatin bringing several chromosomes together (Muller and Almouzni, 2017). Most eukaryotic centromeres are marked by the histone H3 variant centromere protein A (CENP-A) (Muller and Almouzni, 2014). In 1985 CENP-A was discovered as one of the proteins detected by autoantibodies from patients with CREST (calcinosis, Reynaud syndrome, oesophageal dysmotility, sclerodactyly, telangiectasia), together with the centromeric proteins CENP-B and CENP-C (Earnshaw and Rothfield, 1985). Later on, it was described CENP-A as a histone H3 variant (Palmer et al., 1991). Studies on stretched chromatin fibers showed that CENP-A domains show chromatin marks typically associated with transcriptionally active regions, such as H3K4me2 or H3K36me2 (Sullivan and Karpen, 2004). Consistent with the presence of active marks at centromeres, other studies have revealed that centromeres undergo low levels of RNAP II-mediated transcription during mitosis (Chan et al., 2012). Human centromeres are flanked by pericentromeric heterochromatin that is marked by the repressive histone marks H3K9me2, H3K9me3 and H4K20me3.

The bulk of constitutive heterochromatin forms at pericentromeric regions. In contrast with telomeres, the repeat sequences making pericentromeres and their organization can greatly vary between organisms, or even between chromosomes of the same species. The human centromere has long stretches of α-satellite repeats and several CenH3 CENP-A, H3.1 and H3.3 nucleosomes. In this doctoral thesis, I will focus on the mouse pericentromeric region due to our model of study that I will introduce

later on. The mouse centromere has a centromeric region flanked by pericentromeric heterochromatin at one side. The centromeric region is transcribed into minor satellite RNA and the pericentromeric region is transcribed into major satellite RNA. Both RNAs have structural roles in their respective regions. The major satellite RNA is required to recruit heterochromatin protein 1 (HP1). Pericentromeric heterochromatin (PHC) is characterized by HP1 and SUV39H1, as seen in Figure 1.12, which as mentioned in the previous subsection is a histone-lysine N-methyltransferase that is responsible for establishing H3K9me3 and is a main component of the PHC (Muller and Almouzni, 2017). In mammals, 5meC DNA methyltransferase can be recruited with H3K9 methyl transferases so that the two reciprocally bolster each other to ensure that the DNA is rendered inaccessible (Lehnertz et al., 2003).

The balance between an open euchromatin signature and a flanking heterochromatin domain is necessary for proper chromosome segregation. The nature and dynamics of the boundary between these two mutually exclusive chromatin domains at centromeres is currently an open question (Molina et al., 2017). Aberrant expression levels of some centromeric factors involves centromeres in disease. Centromeric higher-order organization integrates with the dynamics of cell cycle-regulated events, such as duplication of centrioles, which are microtubule-organizing centres involved in the formation of the mitotic spindle (Muller and Almouzni, 2017). Finally, at the onset of mitosis, centrosome amplification that is the presence of more than two centrosomes has long been associated with multipolar spindle formation, and with the generation of genetic instability and tumorigenesis (Marthiens, Piel, and Basto, 2012).

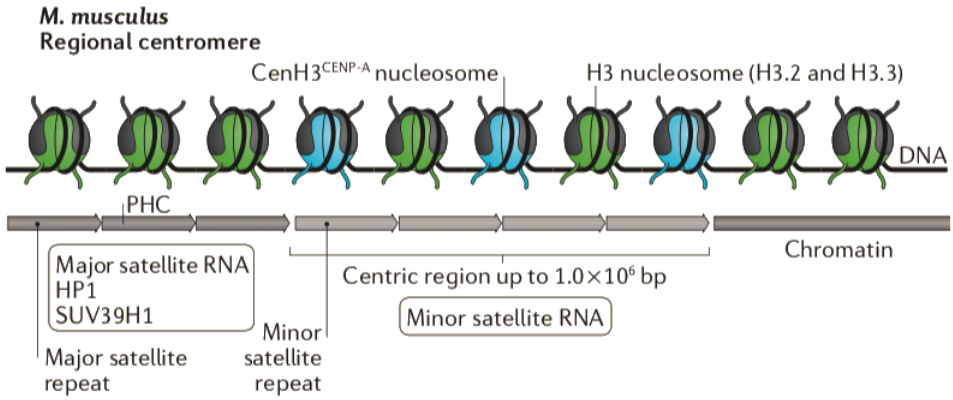


FIGURE 1.12: Schematic illustration of the mouse regional centromere. The centromeric and pericentromeric regions are present. Adapted from (Muller and Almouzni, 2017)

1.2.2.3 Core heterochromatin components and mechanisms

Pericentromeric heterochromatin is not just structurally but also functionally linked to the centromeric region. Some of the heterochromatin components have already been described. HP1 is implicated in heterochromatin formation and maintenance, as well as transcriptional regulation (Probst and Almouzni, 2011). It was the first heterochromatic component identified but also turned out to be a central one. It is also involved in regulating the binding of cohesin to centromeres, which is crucial for chromosome segregation during mitosis (Hahn et al., 2013), but also in replication. HP1 proteins contain a chromodomain reader module and also a C terminal chromoshadow domain (CSD). Dimerization of the CSD forms a binding platform for other effector proteins (Brasher et al., 2000; Cowieson et al., 2000). Thus, the reading of H3K9me by HP1 proteins provides another route to reader–modifier coupling. HP1 proteins that interact with SUV39H can bind to H3K9me3 modifications and to each other. These interactions are thought to spread the structure of pericentromeric heterochromatin and silence adjacent gene expression, as elucidated by PEV in fission yeast (Allshire et al., 1994). H3K9-metyltransferases are conserved in regional centromeres from yeast to human. Some examples include Clr4

in Schizosaccharomyces pombe, SU(VAR)3–9 in *Drosophila melanogaster*, Suv39h1, and Suv39h2 in mouse and human (Figure 1.13). Suv39h1 and Suv39h2 double knockout mice lose H3K9 trimethylation from the pericentromeric regions of the chromosomes (Peters et al., 2001). This loss is accompanied by an increase in pericentromeric transcription (Peters et al., 2003). Another biochemical marker of heterochromatin is the presence of DNA methylation. The addition of methyl groups to cytosines within the CpG dinucleotide by DNA methyltransferases contributes to the stability of pericentromeric heterochromatin (Okano et al., 1999; Bachman, Rountree, and Baylin, 2001). In mammalian cells, DNA methylation is catalyzed by two important classes of DNA methyltransferases (Bestor, 2000). DNA methyltransferase 1 (Dnmt1) methylates CpG dinucleotides in the newly synthesized daughter strand. The function of Dnmt1 is thought to be essential for maintaining DNA methylation patterns in proliferating cells (Li, Bestor, and Jaenisch, 1992). There are two members of the second class of methyltransferases, Dnmt3a and Dnmt3b, which are required for the initiation of de novo methylation during embryonic development (Okano et al., 1999).

Apart from the basic heterochromatic components, studies have shown that a critical role in heterochromatin formation and silencing may play the contribution of transcription factors. Their implication maybe be less studied, but the absence of promoter elements in these regions has led to the hypothesis of the binding of these factors outside a genic context (Bulut-Karslioglu et al., 2012; Millanes-Romero et al., 2013).

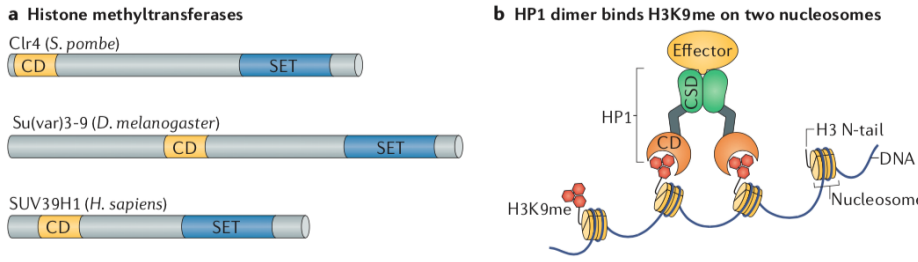


FIGURE 1.13: Schematic representation of the core heterochromatin components, the histone methyltransferases in different species and HP1 protein. Adapted from (Allshire and Madhani, 2018)

1.2.2.4 Heterochromatin spreading

Once nucleated at a particular location, the biochemical properties of heterochromatin components enable the expansion of the domain in a manner that is largely independent of DNA sequence. Heterochromatin spreading requires reader–writer coupling. As "readers" are characterized the proteins that recognize and bind chromatin through histone modification recognition domains. Proteins that contain "reader" domains can be broadly classified into four groups: chromatin architectural proteins, chromatin remodeling enzymes, chromatin modifiers, and adaptor proteins. As "writers" can be described the enzymes that add histone modifications to chromatin such as histone methyltransferases. Nucleosomes bearing H3K9me are bound by the chromodomains of H3K9me writers such as Suv39 and Clr4, and mutations in the Clr4 chromodomain impede spreading in S. pombe (Zhang et al., 2008a). However, spreading also requires the HP1-dependent recruitment of HDACs (Fischer et al., 2009). Thus, interconnections among reader, writer and eraser modules form positive feedback loops that extend heterochromatin domains.

Due to heterochromatin spreading, mechanisms that restrict its expansion are necessary to avoid erroneous and deleterious gene silencing. Mechanisms to create such barriers and interrupt lateral heterochromatin spreading include the following: generation of nucleosome depleted regions by binding of proteins such as transcription factors; processes that promote

nucleosome turnover; recruitment of anti silencing factors by ongoing transcription and associated regulatory elements; recruitment of readers with anti-silencing activity; and restricting silencing factors to their sites of prior action (Allshire and Madhani, 2018).

Heterochromatin can itself recruit the inhibitors that limit its own spreading through reader–eraser coupling as seen in Figure 1.14. One example is Epe1 (enhancement of position effect 1) a known JmjC domain chromatin-associated protein in S. pombe. Epe1 is a putative H3K9 demethylase that is recruited by the reader *Swi6* (HP1 ortholog in fission yeast) (Ayoub et al., 2003; Trewick et al., 2007). They demonstrated that this protein regulates the stability of heterochromatin domains. Other studies proposed that in Epe1 mutant cells, silent chromatin domains oscillate, expanding into the surrounding euchromatin or contracting to cause alleviation of silencing (Trewick et al., 2007). Epe1 acts in parallel with boundary elements. In addition, was found to be responsible for the rapid removal of the ectopic H3K9 methylation. They discovered that cells lacking both Epe1 and the globally acting histone acetyltransferase Mst2 display widespread ectopic heterochromatin assembly and slow growth (Wang, Reddy, and Jia, 2015). Ectopic heterochromatin formation in such double mutants suggests that the processes that trigger heterochromatin assembly at the primary genomic locations can act globally.

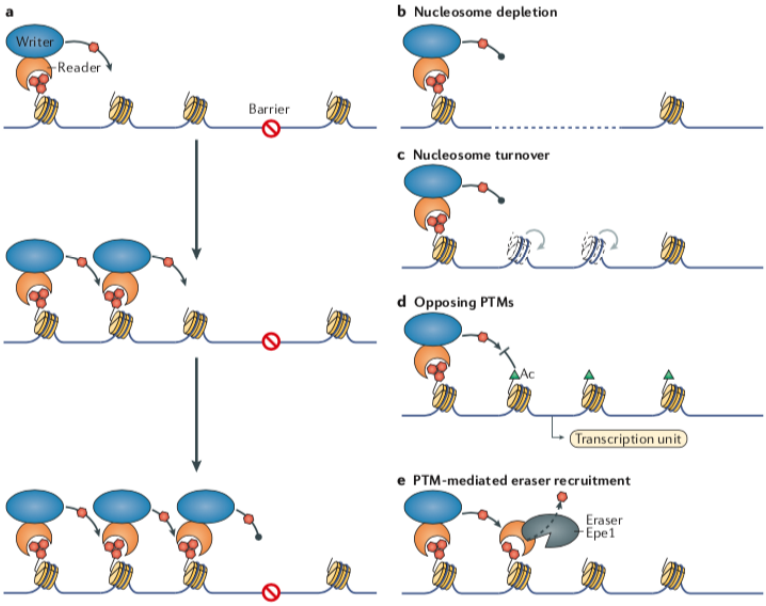


FIGURE 1.14: The regulation of heterochromatin spreading. (a) A model for the expansion of a heterochromatin domain (b) Sequences that are bound by factors that disfavour nucleosome assembly create extensive gaps (dashed line) that prevent heterochromatin from spreading. (c) Factors that promote nucleosome turnover through disassembly and reassembly and/or through cycles of histone exchange (light nucleosomes and arrows) effectively block heterochromatin domain expansion. (d) Adjacently expressed transcription units mediate the addition of active PTMs (green triangles) to histones, which prevent the intrusion of repressive PTMs and heterochromatin. (e) Erasers such as the S. pombe demethylase Epe1 are recruited by readers of repressive PTMs at the edge of heterochromatin and prevent heterochromatin expansion. Adapted from (Allshire and Madhani, 2018)

1.2.2.5 Pericentromeric transcription

In 1969 scientists had suggested that pericentromeric DNA was transcriptionally silent in differentiated mouse tissues (Flamm, Walker, and McCallum, 1969). However, in the same period were presented the first indications of a possible satellite DNA transcription in mouse tumor cells (Harel et al., 1968). Nowadays, with the increased sensitivity of molecular techniques, transcription of pericentromeric satellite repeats has been confirmed in a multitude of organisms and in various contexts, including proliferation, development differentiation, senescence, stress response, and transformation. This could either be the consequence of leaky heterochromatin, or it might reflect specific, and perhaps, conserved, functions of the transcription process or of the resulting transcripts (Saksouk, Simboeck, and Dejardin, 2015). Pericentromeric transcription includes transcription of satellite RNAs that are usually not very abundant in somatic cells, suggesting that transcription is a relatively rare event, or that RNAs are highly unstable. They are produced by RNAPII and can exist in sense (in mice: T-rich) and antisense (in mice: A-rich) orientations (Rudert et al., 1995; Valgardsdottir et al., 2005). Interestingly, sense and antisense transcripts are not necessarily present in equal quantities. This fact suggests that transcription might underlie a regulated process that is controlled by specific regulatory DNA elements rather than an unspecific side-product of decondensed pericentromeres.

1.2.2.6 Repetitive elements and their transcription

In the subsection 1.2.2.2 I introduced the repetitive elements as part of constitutive heterochromatin. Here, I would like to focus a little bit more on the consequences of their transcription.

Repetitive elements can be subdivided into tandem repeats and transposable elements, based on their sequence characteristics. Tandem repeats are head-to-tail repetitions of 2–200 bp long sequence elements, present either as microsatellites and minisatellites of 5–150 bp in length, dispersed around the genome, or as megabase long stretches of major satellite sequence around centromeres (Zeller and Gasser, 2017). Based on the intermediates formed during transposition, the transposable elements can be further subdivided into RNA and DNA transposons. RNA transposons make up the largest fraction of repetitive elements in the human genome

and are the most active transposon class. They either do or do not contain long terminal repeats (LTRs; retrovirus-like transposons). Of special importance is the non-LTR-transposon LINE-1, as it makes up 17% of our genomes and is one of the most transpositionally active transposons in the human genome (Lander et al., 2001; Boissinot and Furano, 2005). A variety of heterochromatin-generating pathways mediate transcriptional or posttranscriptional silencing of these sequences. The unscheduled transcription of repetitive elements can compromise genome integrity (Zeller et al., 2016). In the Figure 1.15, there is a summary of the contribution of the three major repeat classes to the human genome.

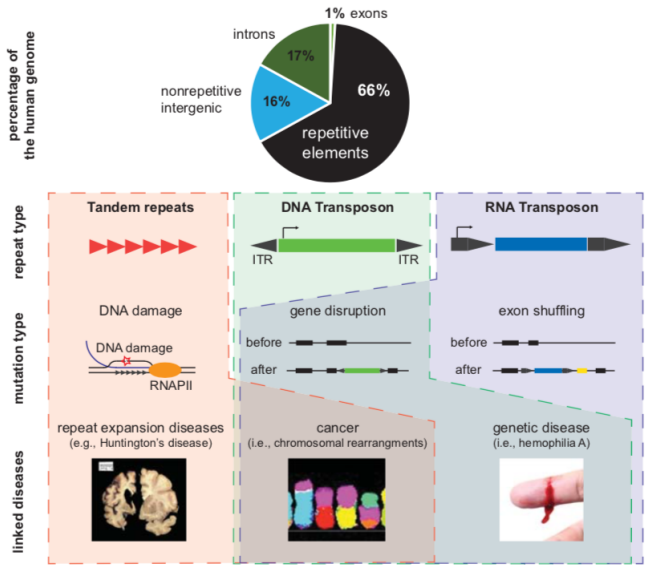


FIGURE 1.15: Summary of repeat classes and related human diseases. Adapted from (Zeller and Gasser, 2017)

Furthermore, it has been demonstrated that the unprogrammed transcription of the repeats correlate with increased rates of repeat-specific insertions and deletions, copy number variation, R-loops and enhanced sensitivity to replication stress (Zeller et al., 2016). This unprogrammed transcription might lead to unscheduled collisions of the replication and

transcription machineries, inducing DNA breaks (Bayona-Feliu et al., 2017; Hoffman et al., 2015)

1.2.2.7 Pericentromeric transcription in cancer

In this subsection I would like to focus on the importance of heterochromatin regulation and the implication of the misregulation in pathological conditions. Pathological incidences, due to misregulation of pericentromeric satellites have been reported, together with decondensation and demethylation of pericentromeric DNA. For instance, aberrant overexpression of pericentromeric satellite repeats has been described in several epithelial cancers (Eymery et al., 2009; Zhu et al., 2011; Ting et al., 2011). The first evidence of the existence of satellite transcripts in cancer originates from findings in Wilms neuroblastoma tumors and epidermal carcinoma cells (Enukashvily et al., 2007; Alexiadis et al., 2007). Numerous diseases including various types of cancers result from deregulated epigenetic mechanisms, which could also directly affect the compaction status of pericentromeres and their expression (Saksouk, Simboeck, and Dejardin, 2015). One study links the tumor suppressor gene BRCA1 to the repression of pericentromeric transcription (Zhu et al., 2011). BRCA1 knockout mice revealed a strong increase in major satellite transcripts. Another study involves the lysine-specific demethylase 2A (KDM2A) as a tumor suppressor gene, which is downregulated in prostate cancer (Frescas et al., 2008). KDM2A is a specific H3K36 demethylase and was also shown to target pericentromeres, thereby ensuring a compact and condensed chromatin structure. Finally, as DNA methylation is one of the main epigenetic marks of constitutive heterochromatin at pericentromeres, is a crucial player in transcriptional repression in mammals. Along this line, satellite transcription in mouse embryonic fibroblasts was observed upon 5-aza-dC treatment, a potent inhibitor of DNA methyltransferases (Sugimura et al., 2010). Importantly, DNA methylation is impaired in neoplasia, which is characterized by a global DNA demethylation as well as localized hypomethylation of oncogenes and hypermethylation of tumor suppressor genes (Wilson, Power, and Molloy, 2007; Saksouk, Simboeck, and Dejardin, 2015).

To conclude, the overexpression of satellite transcripts in cancer may reflect global alterations in heterochromatin silencing, leading to genome

instability and it could potentially be useful as a biomarker for cancer detection. In the following section, I will focus on genome instability and its causes, due to is relevance to the results.

1.3 Genome instability

Genetic instability in the form of mutations and chromosome rearrangements is usually associated with pathological disorders, but it is also crucial for evolution and the generation of genetic diversity. Even before the discovery of the double helix in 1953, it was known that exogenous agents, such as X-rays, ultraviolet (UV) light, and various chemicals, can cause genetic changes that can promote cancer (Friedberg, 2008). In humans, genomic instability is often associated with premature ageing, predisposition to various types of cancer and with inherited diseases (Aguilera and Gomez-Gonzalez, 2008). Cells use a number of mechanisms to preserve the genome from the mutagenic action of genotoxic agents and to guarantee faithful chromosome duplication and transmission to the offspring. In eukaryotes, cell-cycle checkpoints guarantee coordination of DNA synthesis and DNA repair with cell division. Genome instability is mainly due to sporadic replication or repair errors but can also take place in response to developmental or environmental signals. Genome instability can be enhanced by exposure to external genotoxic agents or as the result of cellular pathologies. Some of these agents include reactive oxygen species, UV radiation and environmental mutagens. Instability events such as single-stranded DNA (ssDNA) gaps or double-strand breaks (DSBs) are generated as a consequence of replication stress and cover events mediated by homologous recombination (HR), including break-induced replication (BIR), and non homologous end-joining (NHEJ) mechanisms, represent the most extended form of instability and can also be associated with mutations and chromosome loss (Aguilera and Garcia-Muse, 2013).

In continuation, I will briefly mention the causes of genome instability.

1.3.1 Causes of genome instability

Genome instability may result from failures at different steps of the DNA cycle, from replication to segregation. Among the main causes of genome instability are: replication dysfunction, failures in postreplicative repair

and homologous recombination, site-specific hotspots of genome instability, transcription replication collisions as well as anaphase bridges and chromosome breakage at M-phase (Aguilera and Garcia-Muse, 2013).

In this doctoral thesis, I will focus on some of these causes mentioned above.

1.3.1.1 Replication dysfunction

DNA replication is the most vulnerable cellular process during cell-cycle progression and it is tightly controlled at different stages from initiation to termination. Once per cell division, genome duplication is regulated at the steps of loading and activation of the replicative DNA helicase at replication origins (Boos, Frigola, and Diffley, 2012). Once replication has initiated, forks may undergo a transient pausing or a longer delay referred to as fork stalling. The replisome usually remains associated with stalled forks so that replication could restart without major consequences once the obstacle has been removed (De Piccoli et al., 2012). However, the persistent fork arrest might lead to a DSB in one of the nascent sister double-stranded DNAs (dsDNAs) or to fork regression generating a Holliday junction (Sogo, Lopes, and Foiani, 2002; Aguilera and Garcia-Muse, 2013).

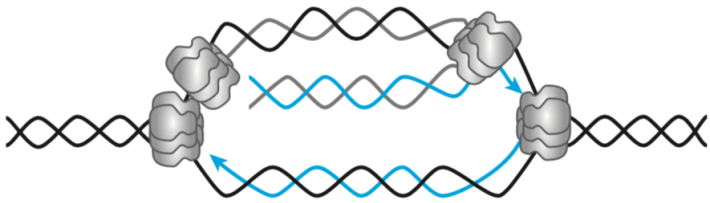


FIGURE 1.16: Replication fork collisions and breakage caused by untimely re-replication. Adapted from (Aguilera and Garcia-Muse, 2013)

1.3.1.2 Transcription replication collisions

The DNA replication and transcription machineries share a common DNA template and thus can collide with each other (French, 1992; Liu and Alberts, 1995). Replication-transcription collisions can cause replication fork

arrest, premature transcription termination, DNA breaks, and recombination intermediates threatening genome integrity. Replication forks may occasionally collapse after colliding with a RNA polymerase. Although elongating or paused/stalled RNA polymerases can constitute physical barriers that block the replication fork progression, increased positive superhelical density generated during head-on collisions between the converging replication fork and RNA polymerase II may also contribute to fork stalling and collapse (Aguilera and Garcia-Muse, 2013). Recent studies in bacteria indicate that replication-transcription collisions can cause apart of replication fork stalling also DNA breaks leading to two types of mutation signatures: duplications and deletions within the transcriptional unit and promoter-localized base substitutions (Sankar et al., 2016).

1.3.1.3 Cotranscriptional R-Loops

In this subsection I would like to focus on the importance of R-loops and their implication to DNA damage. RNA:DNA hybrid structures known as R-loops were thought to be rare byproducts of transcription with little impact in the cell (Aguilera and Garcia-Muse, 2012). The R-loop structure was first characterized in 1976 (Thomas, White, and Davis, 1976), and once formed was considered with great stability as RNA/DNA associations are thermodynamically more stable than DNA/DNA interactions (Roberts and Crothers, 1992). This may be due to the structure of the RNA/DNA . Later on, the first demonstration that R-loops exist *in vivo* came in 1995 (Drolet et al., 1995). They showed that R-loop formation occurs in a bacterial cell and is a consequence of the transcription process (Drolet et al., 1995). Since then, in the field these structures have become an increasingly expanded area of research, placing them as a potential regulator of gene expression. R-loop are three-stranded nucleic acid structures that comprise nascent RNA 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 generate 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 (Westover, Bushnell, and Kornberg, 2004). One of the major causes responsible for R-loops accumulation is the replication fork collapse or stall (Gan et al., 2011). R-loops form naturally as key intermediates in specific cellular processes, such as E. coli

plasmid replication, mitochondrial DNA replication, or immunoglobulin class switching. Beyond these specific roles, the last years, many studies suggest an alternative scenario in which R-loops are formed at a higher frequency than previously thought, thus creating a major threat in gene expression and genome integrity (Aguilera and Garcia-Muse, 2012). However, over the last decade a new era has emerged, identifying these structures as powerful regulators of gene expression since various evidence has shed light on a number of biological processes controlled by the programmed formation of R-loops (Skourti-Stathaki and Proudfoot, 2014).

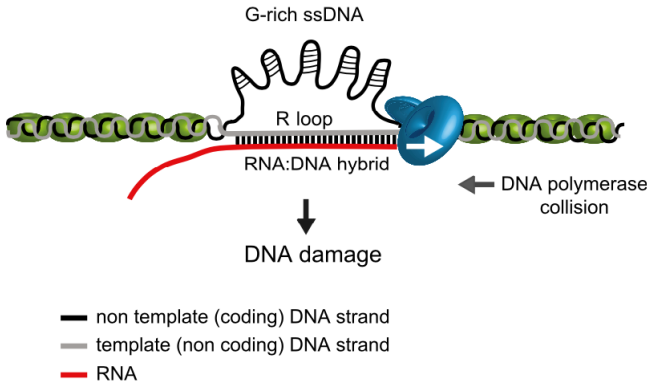


FIGURE 1.17: R-loops formation is a source of DNA damage. Adapted from (Skourti-Stathaki and Proudfoot, 2014)

R-loops can possibly lead to DNA damage by the exposure of ssDNA formed as a result of the RNA/DNA hybridization. Being more unstable, the ssDNA would then be susceptible to lesions and transcription-associated mutagenesis or transcription-associated recombination. A possible scenario, on how R-loops lead to DNA damage, is the unpaired DNA strand resulting from R-loop formation that is more susceptible to DNA damage such as spontaneous deamination of dC to dU, leading to DSBs and recombination. Other possible scenarios include a protein recognizing R-loop structure that could be involved in initiating the generation of mutagenesis. An additional mechanism suggests that transcriptional R-loops induce genomic instability by interfering with DNA replication (Aguilera and Garcia-Muse, 2012).

1.4 The development of the nervous system

In this final section of the introduction of this doctoral thesis, I will cover the essential aspects of the nervous system development, in order to properly introduce my models of study, the *in vitro* model of NSCs and the *in vivo* model of the chicken neural tube.

The mammalian nervous system is the most complex system of any living organism. In vertebrates, the development of the nervous system is triggered by signals from a powerful "organizing" region of the early embryo. After egg fertilization, the blastocyst divides into the three primary layers: ectoderm, mesoderm and endoderm. The ectoderm is the germ layer that will give rise to the nervous system (Patthey and Gunhaga, 2014). After gastrulation, the dorsal region of the ectoderm divides symmetrically and generates the neural plate. Next, the neural plate invaginates to form the neural tube, whose anterior and posterior region will generate the brain and the spinal cord respectively (Ozair, Kintner, and Brivanlou, 2013). The width of the neural tube is occupied by a polarised neuroepithelium. Neuroepithelial cells possess a basal side attached to the pial surface and an apical side that contacts the lumen of the neural tube. As development proceeds, they multiply and form the central nervous system (CNS) germinal zones, namely the ventricle zone (VZ) and subventricular zones (SVZ) (Barry, Pakan, and McDermott, 2014).

1.4.1 The cortex development

The size and surface area of the mammalian brain are thought to be critical determinants of intellectual ability. Brain size varies between mammalian species, which is mostly the result of a disproportionate difference in size of the cerebral cortex (Finlay and Darlington, 1995). One of the most prominent features of the human brain is the fabulous size of the cerebral cortex and its folding, visible as bulges and grooves on its external surface. The cerebral cortex is a laminar tissue where neurons lie on the upper part, and the lower or inner part contains most of the wire connecting neurons between brain areas (Fernandez, Llinares-Benadero, and Borrell, 2016). In humans, it is believed to provide the uniqueness of our species, even though the forebrain structure is conserved in all vertebrates (Wilson and Houart, 2004).

At mid-gestation of the mouse between embryonic day 9 (E9) and 10 (E10), the first neurons of the CNS are born, heralding an important developmental transition in the development of the neural progenitor cells of the brain. Coincident with the acquisition of neurogenic potential, the progenitors acquire the identity of radial glial (Martynoga, Drechsel, and Guillemot, 2012). Radial glial cells have been extensively studied in order to understand CNS formation. This cell type, with its long radial process, was originally discovered on 1885 by Camillo Golgi (Rakic, 2003). Moreover, Ramon Cajal originally suggested a glial identity for these cells by demonstrating their morphological similarities with astrocytes. Although, their glial phenotype was ultimately confirmed many years later when immunohistochemistry and electron microscopy showed that these cells contain glycogen granules and GFAP, which are intracellular characteristics found only in glia (Levitt and Rakic, 1980). Neuroepithelial cells, that are termed radial neuroepithelial cells, serve as neural stem cells, expanding the CNS by self-renewing and populating it by generating neurons and glia. After completing the early phases of neurogenesis, they begin their transformation into radial glial cells. Radial glia are considered the main subtype of cortical progenitors, comprising the majority of mitotically active cells in brain ventricular zones (Lui, Hansen, and Kriegstein, 2011). These cells will differentiate in a diverse type of neurons in order to generate the cortical structure (Tiberi, Vanderhaeghen, and Ameele, 2012). Radial glia in the cortex contribute to neurogenesis directly or via immediate neuronal precursor cells (nIPC). Neurons are generated in the deeper part of the developing brain, the VZ. Next, they migrate to the pial surface so that they generate the six-layered cortical architecture. The last step in neurogenesis is the neuron terminal translocation to the marginal zone where it will maturate developing axons and dendrites (Ohtaka-Maruyama and Okado, 2015). Cortical and spinal cord radial glia contribute to gliogenesis by producing astrocytes and possibly oligodendrocytes. Finally, alterations in proliferation and/or survival of neural progenitors lead to abnormal brain size, either excessive (megalencephaly), defective (microcephaly), or imbalanced (dysplasia).

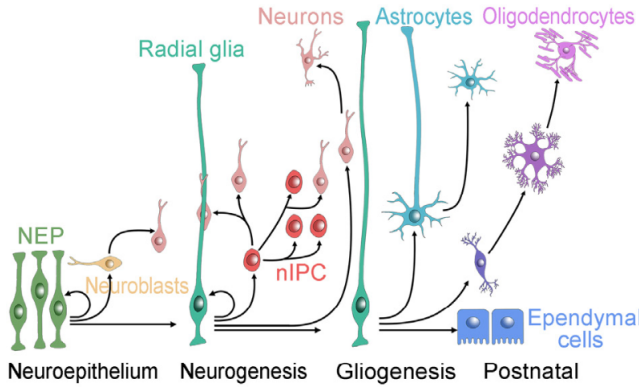


FIGURE 1.18: Neuroepithelial cells proliferate and generate neuroblasts and immature neurons. Adapted from (Barry, Pakan, and McDermott, 2014)

1.4.1.1 Molecular pathways regulating neurogenesis in the cerebral cortex

The processes of cell proliferation and differentiation occur in response to a variety of developmental cues, and their completion requires close coordination between transcription factors, extrinsic signals, and chromatin-modifying enzymes. The developmental signals and transcription factor families that control neurogenesis have been studied for many years (Guillemot, 2007). However, the last years, great attention has been given to factors that interact with chromatin as regulators of neurogenesis and neural development.

Cell-intrinsic genetic controls and cell-extrinsic signaling mechanisms ultimately influence cell fate via the precise modulation of gene expression. Molecular pathways regulate the onset, progression, and termination of neurogenesis. Moreover, multiple signaling pathways synergize to elicit the neurogenic-to-gliogenic switch. Along neurogenesis, different signaling pathways governed by growth factors (GF) act upon cells to orchestrate the balance between proliferation and differentiation. Sonic Hedgehog (Shh) pathway controls cell cycle kinetics of radial glial cells, maintaining the proliferation, survival and differentiation of neurons in the neocortex (Komada, 2012). Fibroblast growth factor (FGF) signaling influences

the extent and level of proliferation of NSCs *in vitro* and *in vivo* (Ornitz and Itoh, 2015). The canonical Wnt pathway stimulates the proliferation of the VZ neural precursors and promotes neuron fate (Chenn and Walsh, 2002). In addition, the bone morphogenetic proteins (BMP) and TGFβ participate in forebrain development. BMPs sequentially induce neurogenesis and then gliogenesis (Li, Cogswell, and LoTurco, 1998). On the other hand, the TGF pathway, promotes cell cycle exit of VZ progenitors by modulating the cell cycle proteins (Siegenthaler and Miller, 2005).

Intrinsic factors balancing proliferation versus differentiation operate mainly through a cell-autonomous manner. Some of the essential intrinsic factors are the bHLH proteins. The class II of bHLH proteins contains the proneural TFs, key determinants of neural cell fate. Proneural transcription factors are necessary and sufficient to activate neuronal differentiation programs, and in the CNS their expression is confined into specific cell populations (Wilkinson, Dennis, and Schuurmans, 2013). Other intrinsic factor is the Notch pathway, that plays an important role for the maintenance of the self-renewal in NSCs (Shimojo, Ohtsuka, and Kageyama, 2008).

1.4.1.2 NSCs as a model of our study

NSCs are multipotent cells in the nervous system that can self-renew, proliferate in an unlimited manner and differentiate into neurons, astrocytes and oligodendrocytes. Notably, our in vitro system of NSCs displays morphology and markers similar to those of the radial glial in the embryo (BLBP, RC2, GLAST, PAX6 and NESTIN among others) (Pollard et al., 2006). *In vivo*, NSCs occupy a microenvironment called niche that supplies the factors required for stem cell self-renewal and differentiation. On the contrary, in the adherent *in vitro* system mitogens such as FGF and EGF (epidermal growth factor) force the symmetric divisions of the NSCs retaining their multipotency, thus reducing the differentiating counterpart (Currle and Monuki, 2007). In this thesis, we have used adherent cultures of NSCs, due to the advantage that give us avoiding the high heterogeneity of the *in vivo* systems.

1.4.2 The neural tube development

The spinal cord is the anatomically simplest and most conserved region of the vertebrate CNS. Vertebrate neurulation is a complex morphogenetic

process that requires the coordination of many cellular and molecular events, and is regulated by more than 300 genes in mammals (Wilde, Petersen, and Niswander, 2014). Primary neurulation is the process by which the neural tube (NT) closes from an open neural plate (NP) and is achieved sequentially in distinct steps. The NT is forming when the flat sheet of neuroepithelial cells, bordered by cells that become the epidermis, rolls up and seals together to form the closed neural tube covered by a layer of surface ectoderm, as seen in Figure 1.19. Indeed, several hundred genes have been identified whose functions are required for NT closure in the mouse embryo (Harris and Juriloff, 2007). It has been described that failure of NT closure is associated with gene mutations. Some examples include Neurofibromin, Pax3, Phactr4, Jumonji and Notch pathway (Harris and Juriloff, 2007; Copp and Greene, 2010; Wallingford et al., 2013). In this process, a solid cord of NT progenitor cells in the developing tail bud becomes "canalized" to form a neuroepithelium surrounding a lumen, without formation or closure of neural folds (Copp et al., 2015). These neurulation mechanisms involve not only the neuroepithelium itself but also the surrounding tissues. Indeed, the non-neural ectoderm, mesoderm and notochord have all been implicated in regulating NT closure.

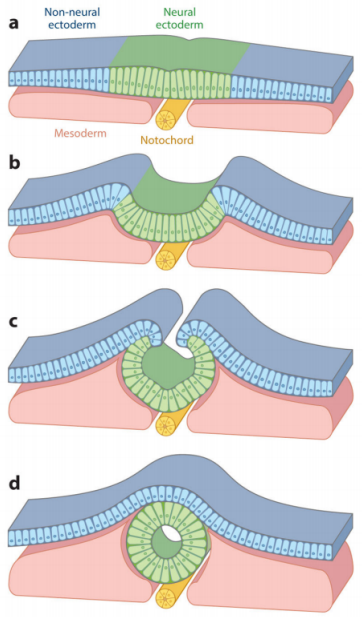


FIGURE 1.19: Schematic transverse sections that illustrate the closure of the neural tube. Adapted from (Wilde, Petersen, and Niswander, 2014)

During the NT tube closure occurs patterning of the neural tissue. In the last years, considerable progress has been made in determining the signaling activities and the genetic networks that control the patterning of the neural tissue along its two major axes: anterior–posterior and dorsal–ventral as well as the mechanisms that control the progressive acquisition of neural cell identity along the dorsal–ventral axis. Disruption of key patterning pathways such as Shh, BMP signaling, and retinoid signaling have been associated with failure of NT closure (Harris and Juriloff, 2007; Copp and Greene, 2010) The "Shh" protein is secreted from ventral sources (notochord and floor plate) and promotes ventral identity by repressing dorsal character, whereas members of both the Wnt and BMP families emanating from dorsal locations are favoring dorsal identities (Le Dreau and Marti, 2012).

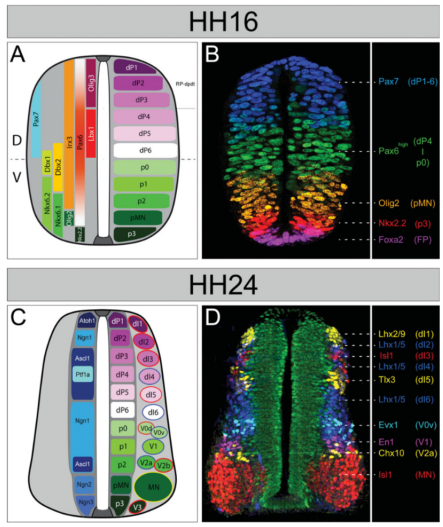


FIGURE 1.20: Dorsal–ventral patterning of the vertebrate developing spinal cord at earlier and later stage (HH). Adapted from (Le Dreau and Marti, 2012)

Determinants of dorsal–ventral identities in the chick NT include members of the homeodomain and the basic-helix–loop–helix families of TFs, which are expressed in restricted DV domains. The developing spinal cord is thus divided into 11 discrete domains of neural progenitors, with five ventral domains (called p3, pMN, p2-0 from ventral to dorsal) and six dorsal domains (called dP1–6 from dorsal to ventral) which are identified by a particular TF code, and this code determines the neuronal subtype progeny they will produce (Le Dreau and Marti, 2012), as seen in Figure 1.20.

1.4.2.1 The chick neural tube as a model of our study

Tight control of the balance between self-expanding symmetric and self-renewing asymmetric neural progenitor divisions is crucial to regulate the number of cells in the developing NT. These populations are maintained in two ways: through asymmetric divisions that generate one stem cell daughter with a developmental potential indistinguishable from that of the parental cell and a cell with a more restricted potential or through symmetric proliferative divisions that generate two stem cells and that serve to

expand the stem cell pool. Alternatively, stem cells can undergo symmetric divisions that generate two cells that enter the differentiation pathway and exhaust the stem cell pool.

The chick neural tube at early stages of the development is occupied by proliferating neural progenitors in the VZ, that undergo symmetric divisions. In later stages, these progenitors will undergo asymmetric divisions and will give rise to post-mitotic neurons that will occupy the mantle zone (MZ) (Figure 1.21). Finally, the dorsal-ventral patterning establishes defined compartments of neural progenitor cells that lead to distinct classes of neurons (Jessell, 2000).

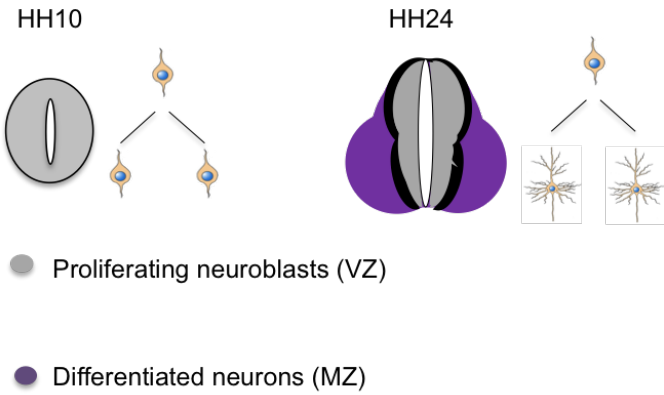


FIGURE 1.21: Schematic transverse section of the chicken NT where the ventricular zone (VZ) and the mantle zone (MZ) are depicted.

1.4.3 Epigenetic mechanisms in neural development

During the last decade, great efforts have been made to elucidate the role of epigenetic regulators that govern the transcriptional programs involved in neural development. In the subsection 1.1.2.7 of the introduction of this

thesis, I mentioned some examples of enzymes of the KDM7 family of demethylases and their implication in neural development and neurogenesis. In order to understand the role of such enzymes that catalyze specific epigenetic modifications, a lot of progress has been achieved by depleting them or by mutations in their catalytic activity. Here, I would like to depict some of these enzymes and their previously described roles.

The Polycomb group proteins (PcG) catalyzes the H3K27me3, a modification, as previously mentioned, mainly associated with transcriptional repression. When Polycomb complex members Ring1b or EZH2 were removed during cortical neurogenesis, the rate of neurogenesis was increased, and the length of the neurogenic period was extended with a corresponding delay in the onset of astrogliogenesis (Hirabayashi et al., 2004). PcG complexes are also required for differentiation of NPCs. High levels of EZH2 are required for NPC differentiation to oligodendrocytes. Conversely, EZH2 expression decreases to allow neuronal and astrocytic differentiation (Sher et al., 2008). Some other studies have provided evidence that PcG activity must be antagonized to sustain cortical neurogenesis. They have discovered that the retinoic acid pathway, which is an important determinant of the onset of neurogenesis directly induces JMJD3, an H3K27me3 demethylase, which when overexpressed is able to activate genes associated with neurogenesis in cortical progenitors (Jepsen et al., 2007). In a similar vein, it has been described that the activity of the Trithorax complex member, MLL1, in antagonizing H3K27me3 deposition on the promoter of the neurogenic transcription factor Dlx2, can sustain neurogenesis (Lim et al., 2009). In addition, the demethylase JMJD3 is has been shown highly regulated at the transcriptional level in response to different developmental, differentiation, and stress-related signals. In a developmental context, JMJD3 was up-regulated during differentiation of ESCs to a neural lineage, and it was required for neuronal commitment (Burgold et al., 2008). Moreover, in NSCs, upon signaling activation, JMJD3 was recruited by SMAD3 to the promoter of TGFβ-responsive genes to facilitate transcription elongation, which allowed the TGFβ developmental program to proceed (Estaras et al., 2012; Estaras et al., 2013) as well as *in vivo* experiments in the chick neural tube have demonstrated that TGFβ-induced neuronal differentiation was dependent on JMJD3 demethylase activity (Estaras et al., 2012). Another study, has demonstrated that JMJD3

activates neural enhancers acting together with SMAD3, the proneural factor ASCL1 and CHD8 remodeling factor upon TGFβ signal (Fueyo et al., 2018).

Finally, in addition to histone methylation, DNA methylation also has a role in sustaining embryonic neurogenesis. When Dnmt1 is mutated in the developing CNS, neurogenesis is terminated prematurely (Fan et al., 2005).

Chapter 2

Objectives

Taking into consideration the current bibliography mentioned in the introduction chapter, we have set the following objectives for this doctoral thesis.

1. To elucidate the molecular function of PHF2 demethylase in NSCs. In order to do that, we sought to determine:
   (a) The chromatin bound regions of this protein along the genome as well as the histone marks that colocalize with PHF2.
   (b) The contribution of PHF2 demethylase activity towards H3K9.
   (c) The genome-wide transcriptional profile of NSCs upon PHF2 depletion.
   (d) The effect of PHF2 in cell cycle regulation, cell proliferation and genome integrity.
2. To investigate the role of PHF2 *in vivo* in the chick embryo neural tube.
   (a) To determine the impact of PHF2 depletion in proliferation and differentiation of neural progenitors in the neural tube.
   (b) To analyze the effect of PHF2 depletion in the neural subpopulation specification.

Taking into account the above objectives, this doctoral thesis continues detailing the materials and methods.

Chapter 3

Materials and Methods

3.1 Materials

In this section I will include all the reagents I have used in this thesis, as well as tables of plasmids, antibodies and primers.

3.1.1 Reagents

All the reagents used in this thesis will be mentioned in the specific procedures, as well as the concentration used for every reagent.

3.1.2 Plasmids

In the Table 3.1, I have listed all the plasmids used in this thesis, including their provider and the experiment that were used.

3.1.3 Antibodies

In the Table 3.2 are listed all the antibodies used in this work, including their provider as well as the dilution and concentration for every application.

3.1.4 Primers

In the Figure 3.1 can be found the sequences and experimental purposes of every set of primers used in this thesis.

TABLE 3.1: List of plasmids used in this thesis.

Plasmid	Provider
pCMV-VSVG	Dr. T. Thomson
pCMV-GAL-POL	Dr. T. Thomson
p3xFLAG-PHF2	Dr. J. Wong
pLKO-Control (CAACAAGATGAAGAGCACC)	Sigma
pLKO-PHF2 (CGTGGCTATTAAAGTGTTCTA)	Sigma
pLKO-PHF2_2 (CCTTATCCACTCCCACTTGACC)	C. Navarro
pSUPER-PHF2 (GGAGCTTCGAAGTCGCACT)	S. Pappa
pSHIN-PHF2_2 (CTATGTCGGACCAGAGAGA)	S. Pappa
pcDNA3.1-HP1BP3-HA	Dr. B. Garfinkel
CMV-FLAG SUV39H1	Dr. T. Jenuwein
pINDUCER PHF2 WT	C. Navarro
pINDUCER PHF2 HID>AIA	C. Navarro

Primer sequences	Primer Forward	Primer Reverse
Primers of repetitive elements and heterochromatin		
Major sat	TGGAATATGGCGAGAAAACTG	AGGTCCTTCAGTGGGCATTT
Major sat (chr9)	TTCACGTTTTTCAGTGATTTCG	CCACTGTAGGACAAGGAATATGG
Minor sat	AATGATAAAAACCACACTGTAGAACAT	ATGTTTCTCATTGTAACTCATTGATATAC
L1	TGGCTTGTGCTGTAAGATCG	TCTGTTGGTGGTCTTTTTGTC
SINE 1	GAGCACACCCATGCACATAC	AAAGGCATGCACCTCTACCACC
IAP1	CGCTCCGGTAGAATACTTAC	TGCCATGCCGGCGAGCCTGT
B1	AGGACACTACAGGTGGGGATT	ACTGACGGCCCAACTGCT
B2	CAATTAAGGACGGGGCAGT	GTCTCCCCATTTCACAATCC
B3	GACTGTGCAGCATTGCATTC	CCTGTGATGAAAACCGAACA
Primers used for gene expression		
Phf2	GGGTGGAAGAGGAAGAGG	GCCATTCTGCACTGGTTTC
Gapdh	ATGTTCGTCATGGGTGTG	CCTTCCACGATACCAAAGTTG
Phf8	GCATACTGGAGAACCGAGAG	CGAGATTTCAAAGCAGGGTC
Sox3	CAAGATGCACAACTCCGAG	TACTTGTAGTCCGGGTACTC
E2f3	GGCCCATTGAGGTTTACTTG	ACCGAGCAGTCACTATGTC
Mcm6	AGACTGTTTCCAAGCCCT	GTACCAGTTGACAAGCTCG
Pcna	GGGTGAAGTTTTCTGCAAG	GCAAACGTTAGGTGAACAGG
Brca2	TCAGATCTCCTTGGTTCAGG	CCACTCCAAGACTTTGCTG
Ppia	ATGGCAAGACCAGCAAGAAG	TTACAGGACATTGCGAGCAG
Primers used for ChIP -qPCR		
Col2a1 TSS	GGTCTCTACCGCTCCCTCAT	GCGACCGGGAGCATATAACT
Gda intr1	ACGACGGCACCAAAGAATAC	TTGCCAGACAAGCAATCAAC
Brca2 TSS	CGGGAAACAGACACACACAC	GCAGCGGTAGCTGACTGAC
Mcm6 TSS	ACAGCTTCTGGCATTTCTCG	CCCTGTTATTGGCTGAGGTG
Aasdhppt TSS	TAAGGGTCTCCACAGGCAAC	TACTTGAGGCCTGCGACTTT
Cdkn1a TSS	CTTCGTGGCAGGGAACAATA	CACACACACACACACTTCTG
Primers used in Chicken		
PHF2	AGTATAACACCAACAGTCAGCC	ATGGAGTTATTCTTGGCCTGG
GAPDH	CGATCTGAACTACATGGTTTAC	ATCACAAGTTTCCCGTTCTC
EZH1	TTGCCAACCACTCCGTGAA	TCACCATCACAACTTTGGCATAG

FIGURE 3.1: List of primers used in this thesis.

TABLE 3.2: List of primary and secondary antibodies used in this thesis.

Antibody target	Provider and reference	Dilution used for every application
PHF2	Cell signaling, D45A2	Western Blot 1:1000, CoIP 1:50, ChIP 1:50, Immunochemistry 1:500
H3K9me2	Abcam, ab1220	ChIP 2μg/mL, Immunochemistry 1:500
H3K9me3	Abcam, ab8898	ChIP 2μg/mL, Immunochemistry 1:500
H4K20me3	Abcam, ab9053	Immunochemistry 1:500
H3K4me3	Abcam, ab8580	Immunochemistry 1:500
Anti-phospho-H2A.X	Millipore, 07-164	Immunochemistry 1:500
S9.6	Described in (Boguslawski et al., 1986)	Immunochemistry 1:500
HP1 α	Active motif, 39295	Immunochemistry 1:500
β -TUBULIN	Millipore, MAB3408	Western Blot 1:10000
H3Sp10	Millipore, 06-570	Immunochemistry 1:500
SOX2	Millipore, AB5603	Immunochemistry 1:500
β-TUBULIN III (TUJ1)	Covance, MMS-435P	Immunochemistry 1:500
MNR2	DSHB, 81.5C10	Immunochemistry 1:500
PAX6	DSHB	Immunochemistry 1:500
HA- tag	Abcam, ab20084	Western Blot 1:5000
FLAG- tag	Sigma, F3165	Western Blot 1:5000
Unspecific IgGs	Diagenode, C15410206	ChIP, same dilution than the specific IgG
Anti-Rabbit IgG IRDye	LI-COR 926-32221	Western Blot 1:5000
Anti-Mouse IgG IRDye	Amersham NA9310	Western Blot 1:5000
Anti-Rabbit Fluor 555	Alexa A31572	Immunochemistry 1:1000
Anti-Mouse Fluor 488	Alexa A21202	Immunochemistry 1:1000

TABLE 3.3: List of GEO accession numbers used in this thesis.

ChIP-seq data	Accession number from GEO
H3K4me2	GSE82943
H3K4me3	GSE82761
H3K9me3	GSE82835
E2F1	GSM935484
E2F4	GSM935365

3.1.5 Gene expression omnibus accessions

In this doctoral thesis we have used previously published NGS datasets to test our hypothesis. These types of genome-wide data are usually deposited in the GEO database (https://www.ncbi.nlm.nih.gov/geo). Accession numbers can be found at Table 3.3

3.2 Models of this study

During this work, most of the experimental work is performed in the *in vitro* model of mouse neural stem cells (NSCs), however, other *in vitro* models were used in particular experiments, such as 293T, HeLa and DF1 cells. The purpose of using these cell lines will be explained later on. Additionally, the chick neural tube was used as an *in vivo* model.

3.2.1 *In vitro* models

3.2.1.1 Mouse neural stem cells

Mouse NSCs were dissected from cerebral cortices of C57BL/6J fetal brains (E12.5) and cultured in previously coated dishes with 5 μg/mL poly-D-lysine (Millipore A-003-E) for 2 hours at 37ºC and 5 μg/mL laminine (Sigma L2020) for 4 hours at 37ºC, as indicated in previous published procedures (Currle and Monuki, 2007). NSCs in order to maintain the ability to self-renew and to generate a wide range of differentiated neural cell types were grown in a medium prepared mixing equal parts of DMEM F12 (without Phenol Red, Gibco 31331093) and Neurobasal Media (Gibco 12348-017)

containing 1% of Penicillin/Streptomycin (Gibco 15140-122), 1% Glutamax (Gibco 35050061), N2 and B27 supplements (Gibco 17502-048 and 17504-044 respectively), 0.1 mM non-essential amino acids (Gibco 11140035), 1 mM sodium pyruvate (Gibco 11360039), 5 mM Hepes (Gibco 15630056), 2 mg/L of heparin (Sigma H-4784), 25 mg/L of bovine serum albumin (Sigma A7906) and 0.01 mM of β-mercaptoethanol (Gibco 31350-010) as previously described (Currle and Monuki, 2007, Estaras et al., 2012). Following the medium preparation, addition of 20 ng/mL of fresh recombinant human EGF (R&D systems 236-EG) and 10 ng/mL of FGF (Invitrogen PHG0021) was performed.

3.2.1.2 HEK 293T cells

HEK 293T cells derive from human embryonic kidney transformed with the large T antigen of the SV40 virus (Graham et al., 1977). For lentivirus productions and protein overexpression, in this thesis, I have used these cells due to their high rate of transfection efficiency.

3.2.1.3 HeLa cells

HeLa cells is the oldest used human cell line. Derives from human cervical cancer cells (SCHERER, SYVERTON, and GEY, 1953). In this study, I have used HeLa cells for their ability of cell cycle synchronization.

3.2.1.4 UMNSAH/DF-1 cells

UMNSAH/DF-1 is a spontaneously immortalized chicken cell line derived from 10 day old East Lansing Line (ELL-0) eggs. Primary chicken embryonic fibroblasts were dissociated and grown in culture (Foster and Foster, 1997). These cells have been used for particular experiments, in order to test the efficiency of chick shRNAs that were used for loss of function experiments.

3.2.1.5 Cellular expansion

Firstly, NSCs were split in a ratio of 1:5 every two days. In order to avoid confluent conditions while growing these cells, we usually split them when reaching 80% confluency. To do so, we remove the medium and rinse with

PBS (Phosphate Buffered Saline), then the cells are detached by the addition of the Accutase enzyme (Gibco A11105-01). This enzyme is recommended for stem cells and is used as a substitute to the classical trypsine due to its milder effect. After 1 minute at 37ºC, the Accutase is diluted with PBS and cells are centrifuged at 260 rcf (relative centrifugal force) for 5 minutes. After the centrifuge, the cell pellet is gently flicked and then diluted in previously warmed expansion medium supplemented with growth factors. Then, this medium that contained the cells is split and plate in precoated plates with the poly-D-lysine and laminine.

The cell lines HEK 293T, HeLa and UMNSAH/DF-1 were cultured in DMEM (Gibco 41965-062) supplemented with 10% of fetal bovine serum (FBS) (Gibco 10270106) and 1% of Penicillin/Streptomycin. They can be passed in a ratio of 1:8 every two days. To do so, we remove the medium in order to rinse with PBS, then cells are detached using the classical trypsine. Trypsine is a chemical dissociating agent (Sigma T9935) diluted with EDTA. Addition of the trypsine follows 1 minute at 37ºC and then its activity is stopped by adding two volumes of medium. Finally the desired split ratio is applied and the cells are plated.

3.2.1.6 Thawing and freezing of cell lines

NSCs can be stored at -80ºC for weeks and in liquid N2 for years. Freezing is performed following the same protocol as in splitting but instead of diluting the cells just in regular medium they are diluted in medium that contains 10% of DMSO. For decreasing the temperature slowly a box named Mr. Frosty is used (Nalgene, 5100-0001). This box contains isopropanol, an alcohol that decreases the temperature 1ºC/minute, thus allowing an optimal freezing process. Cells thawing is rapidly performed by placing the cryotube that contains the cells at 37ºC in the water bath and as soon as it reaches the liquid state the medium with the cells is diluted in PBS and centrifuged at 260 rcf during 5 minutes, in order to eliminate the DMSO that had been previously used for freezing. Cells are then diluted and plated in the NSCs expansion medium.

Like NSCs, the rest of the cell lines we have used in this thesis can be stored at -80ºC for weeks and in liquid N2 for years. To freeze the cells, trypsinization is stopped with PBS and cells are centrifuged at 260 rcf for 1 minute. Then, the cell pellet is diluted in FBS containing 10% of DMSO and the cryotubes are introduced in the Mr. Frosty box at -80ºC, just as for

NSCs. HEK 293T, HeLa and UMNSAH/DF-1 cells thawing is identical to that of NSCs.

3.2.2 *In vivo* models

3.2.2.1 Chicken neural tube

The chicken neural tube is an *in vivo* model that has been used in this doctoral thesis in order to perform *in ovo* electroporation for loss of function experiments that will be further explained in the upcoming subsection. Eggs from White-Leghorn chickens were incubated at 38.5ºC and 70% humidity. Embryos were staged following Hamburger and Hamilton (HH) (Hamburger et al., 1992).

3.2.2.1.1 *In ovo* electroporation

Chick embryos were electroporated (EP) with purified plasmidic DNA at 3 μg/μL in H_2O using 50 ng/mL of Fast Green (Sigma F7258). Plasmidic DNA was injected into the lumen of HH11–HH12 neural tubes by electric pulse. The electrodes were placed at both sides of the neural tube and embryos were EP by an Intracel Dual Pulse (TSS-100) electroporator delivering five 50 milliseconds square pulses of 20–25 V (Volts). This electric field also causes the negatively charged nucleic acids to migrate towards the positively charged electrode, thus effectively driving the DNA of interest into the cell, leaving the part of the neural tube, where the negatively charged electrode is present, to be used as a control. (Figure 3.2).

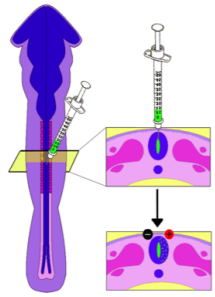


FIGURE 3.2: Schematic view of chick neural tube electroporation.

3.2.2.1.2 Immunostaining

Electroporated embryos were developed for 48 hours and then they were dissected out and processed for the indicated protocol. The brachial regions of the embryos were collected and fixed for 2 hours at 4ºC in 4% paraformaldehyde. They were embedded in 5% agarose-10% sucrose for sectioning in a Leica vibratome. Sections were blocked at room temperature for 1 hour in 1% bovine serum albumin (in PBS with 0.1% tritonX). Then, they were incubated with primary antibodies overnight and 2 hours at room temperature with the secondary IgG antibodies and 0.1 ng/μL DAPI (Thermofisher, D1306). Images were captured by a Leica SP5 confocal microscope using LAS-AF software.

3.3 Methods

In this section, I will detail all the methodology used in this doctoral thesis.

3.3.1 Genetic manipulation of growing cells

Manipulation of living systems 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.

3.3.1.1 Calcium phosphate transfection

This method allows the plasmid's delivery into the cells through endocytosis of the precipitates of calcium phosphate that contain DNA molecules that are stuck on the surface. To generate the calcium phosphate precipitates, the DNA of interest is added into a mixture of HEBS (250 mM NaCl, 9 mM KCl, 1.5 mM Na_2HPO_4, 10 mM glucose and 50 mM Hepes pH 7.12) with the addition of 0.25 M $CaCl_2$ using vortex continuously. After 15 minutes at room temperature, the mixture is ready to be added to the growing medium. After 6 hours have passed, fresh medium is supplied in order to avoid cell stress due to acidity.

3.3.1.2 Lentiviral transduction

In this doctoral thesis, the method of lentiviral transduction has been applied to transduce shRNAs that will permit the knocking down of proteins in our system, as well as for the overexpression of PHF2 (wild type) WT and PHF2 mutant for the catalytic activity (249HI251D>AIA). This delivery system is highly efficient in NSCs, achieving up to a 90% of genetically modified cells. This procedure consists in three steps: lentiviral production, lentiviral transduction and selection. Lentiviral particles are produced in HEK 293T cells by cotransfecting in four 10 cm plates the DNA encoding the shRNA (pLKO.1-shControl, pLKO.1-shPHF2, pLKO.1-shPHF2_2, pINDUCER-PHF2WT, pINDUCER-PHF2 HID>AIA Table 3.1) together with pCMV-VSVG and pCMV-GAG-POL plasmids that 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 surface, previously added in the tube, at 57000 g during 2 hours. For an increased efficiency, the centrifuge is decreasing the high speed of 57000 g very slowly. Then, supernatant is removed and the viral particles are resuspended in NSCs medium. These particles can be stored at -80ºC or be directly used for infection. Transduction of the previously produced cells consists in the addition of the medium containing viral particles to the receptor cells. In this thesis we have used NSCs as receptor cells. Approximately one production will infect 1.5×10^6 cells. 24 hours after infection, cells are selected with the correspondent antibiotic, that for pLKO.1 plasmids is puromycin (Sigma P8833) at a concentration of 2 μg/mL and for pINDUCER plasmids is neomycin (Sigma 108321) at a concentration of 600 μg/mL. Doxycycline (DOX) (Millipore, 324385) was used at 1 μg/mL for induction. After 48 hours cells can be considered selected for pLKO.1 plasmids and the knocking down of the protein can be assessed by RT-qPCR or Western Blot. In the case of pINDUCER plasmids, after 72 hours of neomycin selection, DOX can be added for 48 hours. After these 48 hours, the protein overexpression can be assessed by RT-qPCR or Western Blot.

3.3.1.3 Mutagenesis

The catalytic mutant (249HI251D>AIA) was generated from PHF2 WT by direct mutagenesis using QuikChange II Site-Directed Mutagenesis Kit (Agilent 200523). First, oligonucleotide primers were designed individually according to the desired mutation. After setting up a series of sample reactions using various amounts of dsDNA template, 5 μL of 10 $\times$ reaction buffer was used, 10 ng of dsDNA template, 125 ng of oligonucleotide primer 1, 125 ng of oligonucleotide primer 2, 1 μL of dNTP mix and H_2O to a final volume of 50 μL. Finally, 1 μL of PfuUltra High-fidelity DNA polymerase (2.5 U/μL) was added to the mix. The cycling parameters that were used are the following: 1 cycle at 95°C for 30 seconds, 12 cycles (95°C 30 seconds, 55°C 1 minute, 68°C 1 minute/kb of plasmid length). Amplification was checked by electrophoresis of 10 μL of the product on a 1% agarose gel. Dpn I digestion of the amplification products was performed using 1 μL of the Dpn I restriction enzyme (10 U/μL) directly to each amplification reaction and transformation of XL1-Blue super-competent cells was performed by transferring 1 μL of the Dpn I-treated DNA into 50 μL of the super-competent cells, following by heat pulse for 45 seconds at 42°C.

3.3.2 Molecular biology procedures

In this section I would like to explain all the molecular biology procedures and the protocols I have performed in this doctoral thesis.

3.3.2.1 Genomic DNA extraction

Approximately 6×10^6 cells were lysed in 200 μL of buffer (10 mM Tris-HCl, 10 mM EDTA, 10 mM NaCl and 0.5% SDS). Then, these lysates were incubated together with 0.5 mg/mL of proteinase K (Sigma P2308) during 1 hour at 50ºC and with 1 mg/mL of RNase A (Fermentas EN0531) during 2 hours at 50ºC. Finally, these DNA and protein mixtures were subjected to phenol-chloroform extraction for DNA purification.

3.3.2.2 Phenol-chloroform extraction and ethanol precipitation

This technique serves to purify DNA from complex protein-DNA mixtures. It is based on the different affinity of DNA and proteins for the organic solvent phenol. First, 1 volume of phenol is added to the mixture, after vortex.

Centrifugation at maximum speed during 3 minutes is performed and the aqueous with the organic phases are separated, resulting to an aqueous phase that contain the DNA which is transferred into a new tube. This step is repeated with chloroform and the aqueous phase obtained is subjected to ethanol precipitation. Ethanol precipitation is used to concentrate and desalt the DNA obtained by phenol-chloroform extraction. For this reason, 0.1 volumes of NaAc 3M and 1 volume of cold ethanol are added to the sample. Then, tubes are incubated at -80ºC to achieve precipitation. Next, centrifugation at maximum speed during 20 minutes at 4ºC is applied, following an additional wash with ethanol 70%. Finally, the pellet is dried using a speed vacuum and the DNA that is in the pellet is resuspended in a suitable buffer.

3.3.2.3 RNA extraction

RNA extraction is achieved by using Trizol reagent (Invitrogen 15596018). This RNA will be used later on for RT-qPCR analysis. Approximately 3×10^6 cells are lysed in 1 mL of Trizol. This step is followed by the addition of 200 μL of chloroform that after centrifugation at maximum speed during 5 minutes the upper phase of the tube is collected and precipitated by adding 800 μL of isopropanol. Then, another centrifugation at maximum speed during 10 minutes is performed and the supernatant is discarded. The observed pellet is washed using 70% ethanol. After a final centrifugation of 5 minutes at maximum speed, the pellet is resuspended in H_2O. To avoid contamination of genomic DNA in the resuspended RNA, the samples are treated with DNase. For this treatment, the DNA-free Kit (Ambion AM1906) has been used that is special for genomic DNA removal from RNA samples due to the inert beads that serve as inhibition agent. The protocol consists in the addition of 0.1 volumes of 10X buffer to the sample, that is incubated with 1μL of DNase during 30 minutes at 37ºC. Then, 0.1 volumes of inhibition agent are added and pure RNA can be obtained by collecting the supernatant by performing centrifugation at 10000 rcf during 1.5 minutes. After this, RNA is quantified using a Nanodrop device and its quality can be evaluated using the ratio of 260/280 and 260/230.

3.3.2.4 RNA extraction followed by high-throughput sequencing

For RNA extraction followed by RNA-sequencing, RNA was extracted using High pure RNA isolation kit (Roche 11828665001) followed by DNase I treatment, in order to achieve as pure quality of RNA as possible. For the RNA sequencing the libraries were prepared using the TruSeq Stranded Total RNA Sample Preparation kit with Ribo-Zero Human/Mouse/Rat Kit (Illumina, RS-122-2201/2) according to the manufacturer's protocol. Briefly, 500 ng of total RNA was used for ribosomal RNA depletion. Then, ribosomal RNA depleted RNA was fragmented for 4.5 minutes. The remaining steps of the library preparation were followed according to the manufacturers instructions. Final libraries were analyzed using Agilent DNA 1000 chip to estimate the quantity and check size distribution, and were then quantified by qPCR using the KAPA Library Quantification Kit (Roche, 07960204001) prior to amplification with Illumina's cBot. The libraries were sequenced on Illumina High HiSeq 2500 with paired-end 50 base pair long reads.

3.3.2.5 Retrotranscription

For quantification of gene expression I have performed RT-qPCR experiments. This technique allows the quantification of the RNA levels. Reverse transcription (RT) of mRNA was performed using 500 ng of RNA and the High Capacity cDNA reverse transcription kit (Invitrogen 4368814). The protocol consists in RNA incubation together with random hexamers, DNTPs and the retrotranscriptase enzyme in a thermocycler, following these parameters: 25ºC 10 minutes, 37ºC 120 minutes and 85ºC 5 minutes. The complementary DNA (cDNA) that is generated can be stored at -20ºC or -80ºC.

3.3.2.6 Quantitative PCR (qPCR)

After RNA retrotranscription, qPCR was used to quantify the cDNA that was obtained. qPCR experiments are based on the quantification of the emitted fluorescence by a fluorophore that binds DNA as PCR proceeds. A higher presence of cDNA will result in more emitted fluorescence and vice versa. qPCR reactions were set-up manually in a volume of 10 μL using the SYBR Green Kit (Roche 4887352001). Reactions were carried out in

96-well plates in a LightCycler 480 (Roche) using the following cycling conditions: 95ºC 5 minutes, 40 cycles of (95ºC 5 minutes, 60ºC 10 seconds, 72ºC 20 seconds) melting curve (95ºC 5 seconds, 65ºC 1 minute). The specific primer pairs were designed spanning exon-exon junctions of a region conserved between splice variants (mRNA). Primers were designed using Integrated DNA Technology (IDT) (https://eu.idtdna.com/pages). Primer sequences can be found in Table 3.1. To validate qPCR results, we run non-template controls and standard curves with every new primer pair, and only primers with an efficiency of 95% or higher were used for further experiments. qPCR data were analyzed using the 2-ΔΔCT method. Outliers were defined as values that differ more than 0.5 cycles from the other two wells in the triplicates and when identified, they were discarded. If the non-template controls showed amplification with a Ct=37 or lower, the reaction was repeated. In order to normalize qPCR results we usually used *Gapdh* as a reference gene because after testing different reference genes this has been the most constant and reliable between our conditions. For ChIP-qPCR, the specific primers were designed around the region of transcription start sites (TSS) of the specific gene using primer3 (http://primer3.ut.ee/). Primer sequences can be found in Table 3.1.

3.3.2.7 Cloning of plasmidic DNA

In order to generate plasmidic vectors that were not available in the laboratory, we used different cloning strategies. In this section, the obtaining of every construct will be detailed. It is important to clarify that all the constructs have been sequenced by Sanger sequencing after cloning in GATC Biotech Company (Germany).

3.3.2.7.1 pLKO.1 constructs

As mentioned in the Table 3.1, one shRNA against PHF2 was purchased in Sigma. In order to find efficient shRNAs to decrease the protein levels of mouse PHF2 (shPHF2_2), shRNAs against PHF2 were designed and then introduced into the pLKO.1 puro empty vector (Addgene 8453). Firstly, the shRNA oligos were annealed by heating at 95ºC in a thermoblock and then allowing them to cool down. Then, phosphorylation of these oligos was performed with the T4 Polynucleotide kinase (PNK) (Fermentas EK0031)

in the provided buffer and supplementing the reaction with ATP (adenosine triphosphate). In parallel, using the restriction enzymes EcoRI and AgeI, 2 μg of the destination vector was cut. Later on, purification of the digested vector was performed from an agarose gel and the annealed phosphorylated oligos and the digested vector were ligated using the enzyme T4 DNA ligase in the provided buffer (Fermentas 15224017)

3.3.2.7.2 pINDUCER20 constructs

The overexpression of proteins of our interest was achieved by lentiviral transduction of pINDUCER vectors obtained by the Gateway cloning. The Gateway Technology is a universal cloning method based on the site-specific recombination properties of bacteriophage lambda. This technology enables us to efficiently transfer DNA-fragments between plasmids using a proprietary set of recombination sequences, the "Gateway att" sites, and two proprietary enzyme mixes, called "BP Clonase" and "LR Clonase". Gateway donor vector that contains the sequence of PHF2 WT and PHF2 catalytic mutant (HID>AIA) was obtained by a "BP reaction" performed using the PHF2 and pDONR221 by Gateway cloning creating a Gateway entry clone from an attB-flanked PCR product. The pDONOR vector containing PHF2 was used later to transfer its insert to the gateway destination vector pINDUCER20. pINDUCER20 is a Tet-inducible lentiviral vector for ORF expression. An "LR reaction" was performed with pDONR221-PHF2 WT or catalytic mutant (HID>AIA) and pINDUCER20, resulting to the pINDUCER PHF2 WT and pINDUCER PHF2 (HID>AIA). For the "BP" and "LR reactions" we added the following components: 1–7μL attB-PCR product (10 ng/μL; final amount 15–150 ng), 1 μL pDONOR vector (150 ng/μL), and TE buffer, pH 8.0 to 8 μL in the case of "BP reaction" and 1–7 μL pDONOR PHF2 (50-150 ng), 1 μL pINDUCER20 vector (150 ng/μL) and TE buffer, pH 8.0 to 8 μL in the case of "LR reaction". We next added 2 μL of the BP Clonase II enzyme mix (Invitrogen 11789-020) to the reaction, in case of "BP reaction" or 2 μL of the LR Clonase II enzyme mix (Invitrogen 11791-020) in the case of "LR reaction", for about 2 minutes. We incubated the reactions at 25°C for 1 hour and we added 1 μL of the Proteinase K solution to each sample to terminate the reaction. After vortexing briefly, we incubated the samples at 37°C for 10 minutes and transformation of supercompetent cells was performed.

3.3.2.7.3 pSUPER and pSHIN constructs

The depletion of PHF2 protein by in *ovo* electroporation in the chick embryo neural tube was performed by shRNAs cloned in pSUPER and pSHIN vectors. pSHIN vector contains the pSUPER and the EGFP expression cassette. The sequence of the oligos are shown in Table 3.1. First of all, the shRNA oligos were annealed by heating at 95ºC in a thermoblock and then allowing them to cool down. Then, phosphorylation of these oligos was performed with the T4 Polynucleotide kinase (PNK) in the provided buffer and supplementing the reaction with ATP. In parallel, using the restriction enzymes HindIII and BglII, 1 μg of the destination vector (pSUPER or pSHIN) was cut. After that, purification of the digested vector was performed from an agarose gel and the annealed phosphorylated oligos and the digested vector were ligated using the enzyme T4 DNA ligase in the provided buffer.

3.3.2.8 Amplification of plasmidic DNA

The desired DNA must be amplified in order to obtain enough material to perform the specific techniques. For this amplification, we used Escherichia coli strain DH5α. Transformation is achieved by mixing 30 μL of competent bacteria with DNA of concentration ranging from 20 ng to 200 ng. This mixture is kept on ice during 30 minutes. Heat pulse at 42ºC during for 45 seconds was performed following 2 minutes on ice. This fast change of temperature allows the transient opening of pores in the bacterial cell wall thus permitting the entrance of the plasmidic DNA. Afterwards, bacteria are recovered by incubating them in 1 mL of LB (lysogeny broth) at 37ºC and agitation. Then, bacteria can be either plated in agar plates to obtain isolated colonies or grown in bigger volumes of LB to continue with mini- or maxipreps of DNA.

3.3.2.8.1 Mini and Maxi preparations of DNA

In molecular biology, solutions of DNA containing around 300 ng/μL of DNA are known as "minipreps", likewise, solutions of DNA containing approximately 1 μg/μL or more are known as "maxipreps". To generate these solutions of DNA from transformation of bacteria, 5 mL or 500 mL of LB are inoculated with either an isolated colony or 5 mL of miniculture.

After overnight growth, the bacterial mass is subjected to DNA purification with the alkaline lysis method following the QIAprep Spin Miniprep Kit for minipreps (QIAGEN 27106) and the QIAGEN Plasmid Maxi Kit for maxipreps (QIAGEN 12165). The alkaline lysis method has the following steps: resuspension, lysis and neutralization. The protocol consists in the addition of three buffers (P1, P2 and P3) corresponding to the three steps mentioned above (Buffer P1: 100 μg/mL RNase A, 50 mM Tris-HCl, 10 mM EDTA pH 8.0; buffer P2: 200 mM NaOH, 1% SDS; buffer P3: KAc 3M, pH 5.5), then the lysate is passed through a column that specifically retains DNA, and after two washes, DNA is eluted and precipitated with isopropanol. Finally, a last wash of the DNA with 70% ethanol is performed and drying is applied in a speed vacuum following resuspension in the suitable buffer.

3.3.2.9 DNA electrophoresis

This technique is used to purify and visualize DNA prior or after other applications. Firstly, an agarose gel of the desired percentage is prepared by adding agarose together with TBE (Tris-Borate-EDTA) buffer (45 mM Tris, 45 mM boric acid and 1 mM EDTA). Heating in the microwave is applied until the agarose is dissolved. In the solution is added the Redsafe reagent (Intron 21141) that permits the visualization of nucleic acids due to the green fluorescence that emits upon DNA and RNA binding. Then, this mixture is solidified in an electrophoretic chamber and after the addition of TBE buffer, the samples of DNA containing orange-glycerol are loaded. Gels typically run at 80V during 1 hour and visualization is achieved by using an UV-transilluminator.

3.3.2.10 Protein extraction

To perform protein extraction followed by Western blot, an astringent buffer called RIPA (Radioimmunoprecipitation assay buffer) was used taking into account that breaks cytoplasmic as well as nuclear membranes (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS and 50 mM Tris, pH 8.0 and protease inhibitors). Approximately, for 6×10^6 cells, 500 μL of RIPA buffer were added but this volume can vary according to concentration requirements and the surface area of the dish were the cells are plated. This suspension was incubated on ice during 20 minutes and then was

centrifuged at maximum speed during 10 minutes at 4ºC. The supernatant contained the protein extract.

3.3.2.11 Protein quantification by Bradford

Using the procedure of Bradford we are able to measure the total amount of protein present in a sample. This method is based on the reaction occurring between the proteins of the sample and a reagent called "Coomassie Brilliant Blue G-250", a reagent that change its color depending on the protein concentration (Bradford, 1976). To measure the concentration, 1 μL of the protein extract is mixed with 200 μL of Bradford solution (Bio-Rad 5000001) and 800 μL of H_2O, after 3 minutes, the colorimetric reaction is measured in a spectrometer. The concentrations of the samples are calculated using the absorbance value of the sample and a Bradford protein assay calibrate line.

3.3.2.12 SDS-Page electrophoresis

This procedure is used to separate proteins in a gel according to their sizes. The SDS (sodium dodecyl sulfate) detergent provides with net negative charge to all the proteins in solution, thus ensuring that the migration will occur exclusively according to their sizes. This technique is normally followed by Western blot that will be explained below. To prepare the samples, Laemmli buffer (375 mM Tris-HCl, 9% SDS, 50% Glycerol, 0.03% Bromophenol blue) is added to the samples with 5% of β-mercaptoethanol following heating at 95ºC during 5 minutes. Next, these samples are loaded in a polyacrylamide gel that contains two different gels: the stacking and the resolving. The stacking gel has a pH of 6.8 and a polyacrylamide concentration of 5%. The resolving gel has a pH of 8.8 and a variable polyacrylamide concentration depending on the sizes of the proteins of interest. Proteins with higher molecular weight, need a lower polyacrylamide concentration. For most of the experiments performed we used 8-10% polyacrylamide concentration. Afterwards, we set up the chamber with the gel, and then by adding the SDS-Page running buffer in the chamber (25 mM Tris-HCl, 192 mM glycine, 0.1% SDS) we start running the gel by setting the power at 25 mA.

TABLE 3.4: Polyacrylamide gel recipes for SDS-PAGE. The represented volumes are indicated in ml.

Resolving gel	8%	10%	Stacking	5%
H20	4.2	4.8	H20	7.25
40% Polyacrylamide	1.6	2.5	40% Polyacrylamide	1.25
1.5M Tris pH 8.7	2	2.5	1.5M Tris pH 6.8	1.25
10% SDS	0.08	0.1	10% SDS	0.1
10% APS	0.08	0.1	10% APS	0.1
TEMED	0.008	0.01	TEMED	0.01

3.3.2.13 Western blot

After SDS-Page electrophoresis, Western blot procedure is performed to specifically detect a protein of interest in the sample that was resolved in the gel. To begin with, a multilayered cassette is used, in which we place from the positive side to the negative side the following: one sponge, Whatman paper, a nitrocellulose membrane, the gel proceeding from the SDS-Page, another Whatman paper and another sponge. This cassette is introduced in a chamber and is fully covered by Transfer buffer (25 mM Tris-HCl, 192 mM glycine, 0.05% SDS and 10% methanol). After the chamber, the cassette and buffer are set, the power source is set at 80 V during 90 minutes. After this time, the proteins have been transferred to the nitrocellulose membrane. To specifically identify a protein in the nitrocellulose membrane we first block the membrane with Bovine serum albumin (BSA) 5% during 1 hour at room temperature. Then, overnight incubation of the membrane with a specific primary antibody (Table 3.2) at 4ºC, is performed. The next day, the membrane is washed using PBS-Tween 0.1% and incubated with a secondary antibody (Table 3.2) during 1 hour at room temperature. After this point the Western blot can be revealed. In this thesis we have used the fluorimetric method to reveal the Western blot. This method has been used to reveal abundant proteins or proteins with very sensitive antibodies. The secondary antibodies we used, were bound to a fluorophore. Revealing is performed in an LI-COR Oddissey scanner. At this point, is possible the visualization of the detected proteins.

3.3.2.14 Co-immunoprecipitation

This procedure is used for the detection of direct or indirect interactions between proteins. This method starts with an immunoprecipitation experiment, that consists in the binding of an antibody to its target in order to enrich the amount of target in a fraction. This experiment can be performed with overexpressed or with endogenous proteins. In this doctoral thesis we performed co-immunoprecipitations with overexpressed proteins using the buffer IPH (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40) for their extraction. After the protein extraction, a specific antibody is added to immunoprecipitate the target protein. At this point is important to mention that we always run a parallel reaction with IgGs of the same species in order to control the unspecific binding of the sample to the IgG per se. After overnight incubation at 4ºC, agarose beads of protein A (Millipore 16-125) or protein G (Millipore IP04) are used, depending on the primary antibody species, following the binding to the primary antibody-protein complex. After 4 hours of incubation at 4ºC, the beads are recovered and eluted in Laemmli buffer by shaking at 900 rpm and 55ºC during 15 minutes. Finally, this buffer contains the eluted proteins and is denatured as explained in SDS-Page subsection, to proceed with electrophoresis and Western blot.

3.3.2.15 Mass spectrometry

In this subsection I will detail the sample preparation, chromatographic and mass spectrometric analysis and finally the analysis of the data. At this point, it is important to mention that in this thesis I performed a single replicate of this experiment and further validation of the results should be performed.

3.3.2.15.1 Sample preparation

Mass spectrometry (MS) is a sensitive technique used to detect, identify and quantitate molecules based on their mass-to-charge (m/z) ratio. This method serves to identify direct or indirect protein-protein interactions. Total protein extract was performed as described for co-immunoprecipitation experiments in the subsection above. An anti-PHF2 specific antibody (Table 3.2) was used to immunoprecipitate the endogenous PHF2 protein.

An unspecific rabbit IgG antibody was used to identify unspecific binding. Protein A agarose beads were used as described above and after 4 hour of incubation at 4ºC were washed three times with 500 μL of 200 mM NH_4HCO_3 (ABC) and then 60 μL of 6 M Urea/ 200 mM ABC were added. Samples were reduced with dithiothreitol (30 nmol, 37 ºC, 60 min), alkylated in the dark with iodoacetamide (60 mmol, 25ºC, 30 min) and diluted to 1 M Urea with 200 mM ABC for trypsin digestion (1 μg, 37ºC, 8h, Promega, V5113). After digestion, peptide mix was acidified with formic acid and desalted with a MicroSpin C18 column (The Nest Group, Inc) prior to MS analysis.

3.3.2.15.2 Chromatographic and mass spectrometric analysis

Samples were analyzed using a LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EasyLC 1000 (Thermo Fisher Scientific (Proxeon), Odense, Denmark). Peptides were loaded onto the 2 cm Nano Trap column with an inner diameter of 100 μm packed with C18 particles of 5 μm particle size (Thermo Fisher Scientific) and were separated by reversed-phase chromatography using a 25 cm column with an inner diameter of 75 μm, packed with 1.9 μm C18 particles (Nikkyo Technos Co., Ltd. Japan). Chromatographic gradients started at 93% buffer A and 7% buffer B with a flow rate of 250 nl/min for 5 minutes and gradually increased 65% buffer A and 35% buffer B in 60 minutes. After each analysis, the column was washed for 15 minutes with 10% buffer A and 90% buffer B. Buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile. The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.1 kV and source temperature at 300°C. Ultramark 1621 was used for external calibration of the FT mass analyzer prior the analyses, and an internal calibration was performed using the background polysiloxane ion signal at m/z 445.1200. The acquisition was performed in data-dependent acquisition (DDA) mode and full MS scans with 1 micro scans at resolution of 60.000 were used over a mass range of m/z 350-2000 with detection in the Orbitrap. Auto gain control (AGC) was set to 1E6, dynamic exclusion (60 seconds) and charge state filtering disqualifying singly charged peptides was activated. In each cycle of DDA analysis, following each survey scan, the top twenty most intense ions with multiple charged ions above a threshold ion count of 5000 were

selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via collision-induced dissociation (CID) were acquired in the Ion Trap, AGC was set to 1e4, isolation window of 2.0 m/z, activation time of 10 ms and maximum injection time of 100 ms was used. All data were acquired with Xcalibur software v2.2. Digested bovine serum albumin (New England Biolabs P8108S) was analyzed between each sample to avoid sample carryover and to assure stability of the instrument and QCloud (Chiva et al., 2018) has been used to control instrument longitudinal performance during the project.

3.3.2.15.3 Data analysis

At this point I would like to clarify that the data analysis has been performed by the proteomics service were the MS experiment was performed. The acquired spectra were analyzed using the Proteome Discoverer software suite (v1.4, Thermo Fisher Scientific) and the Mascot search engine (v2.5, Matrix Science), (Perkins et al., 1999). The data was searched against a Swiss-Prot mouse database (as in October 2017, 17083 entries) plus a list of common contaminants and all the corresponding decoy entries. For peptide identification a precursor ion mass tolerance of 7 ppm was used for MS1 level, trypsin was chosen as enzyme and up to three missed cleavages were allowed. The fragment ion mass tolerance was set to 0.5 Da for MS2 spectra. Oxidation of methionine and N-terminal protein acetylation were used as variable modifications whereas carbamidomethylation on cysteines was set as a fixed modification. False discovery rate (FDR) in peptide identification was set to a maximum of 5%.

3.3.2.16 Chromatin immunoprecipitation

Regarding to this procedure, I have performed ChIP experiments to focus on particular regions of the genome using qPCR with specific pair of primers. Moreover, in order to identify the genome-wide distribution of a specific protein as well as post-translational modifications we have used, ChIP experiments followed by high-throughput sequencing. The data analysis of ChIP-seq is described in detail in the subsection 3.3.4.1.

3.3.2.16.1 Chromatin immunoprecipitation followed by qPCR

This procedure permits the immunoprecipitation of proteins bound to the chromatin in a direct or indirect way, followed by the identification of the DNA bound using qPCR. For every ChIP, 6×10^6 NSCs were fixed with 1% of formaldehyde during 10 minutes, this fixation was stopped by adding 0.125 M of glycine for 5 minutes. After the fixation, the cells were lysed in 1% SDS lysis buffer (1% SDS, 10 mM EDTA pH 8.0, 50 mM Tris-HCl pH 8.1) to proceed to the chromatin sonication in order to obtain fragments of around 300 bps of DNA to ensure resolution in the DNA detection. It is important to clarify that the sonication step was performed in a Bioruptor sonicator with variable parameters in order to obtain 200-500 bps chromatin fragments. One example of the parameters used is the following: 30 cycles of 30 seconds on and 30 seconds off at high potency. Then, the sonicated chromatin is purified by centrifugation at maximum speed during 10 minutes and the recovered supernatant is prior used for a sonication test in which the evaluation of the correct size of the chromatin fragments is performed. As mentioned above, fragments of around 200-500 bps are considered acceptable, if not obtained, further sonication of the DNA has to be applied. Once chromatin is properly shredded, the immunoprecipitation step starts by diluting the chromatin ten times using the immunoprecipitation buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris-HCl pH8.0) and adding the specific antibody for the protein of interest. As in co-immunoprecipitations, an identical parallel reaction is performed using an unspecific IgG. After overnight incubation at 4ºC the antibody-protein complexes are captured using Magna ChIP magnetic beads (Millipore 16-661) during 4 hours at 4ºC. Then, antibody-protein-DNA complexes are washed with buffers TSEI (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0 and 150 mM NaCl), TSEII (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0 and 500 mM NaCl), TSEIII (0.25 M LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCl pH 8.0) and TE respectively (10 mM Tris-HCl pH 8.0 and 1 mM EDTA) and finally eluted using elution buffer (1% SDS, 0.1 M $NaHCO_3$) during 15 minutes. At this point, DNA must be obtained and used for qPCR. To do so, samples are subjected to decrosslinking at 65ºC overnight. Then, the samples are treated with RNase A at 1 mg/mL during 30 minutes at 37ºC and with proteinase K at 55ºC during 2 hours. After this, DNA is purified

by phenol-chloroform extraction followed by ethanol precipitation. Usually DNA is resuspended in 50 μL of H_2O. Finally, ChIP DNA is analyzed by qPCR as explained in subsection 3.3.2.6, using specific primers (Table 3.1). Percentage of input material is used for the quantification of the immunoprecipitated DNA with respect to the total starting chromatin.

3.3.2.16.2 Chromatin immunoprecipitation followed by high-throughput sequencing

Chromatin immunoprecipitation and sample preparation for sequencing were done essentially as previously described in the subsection 3.3.2.16.1. Sonication step was performed in a Bioruptor sonicator and shredded chromatin was used for each immunoprecipitation with PHF2 (Cell signaling D45A2) and H3K9me2 (Abcam ab1220) antibodies. After decrosslinking, DNA was purified by ethanol precipitation. Purified DNA was used to prepare the library that was sequenced in a HiSeq 2000 sequencer (Illumina).

3.3.3 Cellular biology procedures

3.3.3.1 Indirect immunofluorescence

For this procedure, NSCs were grown in a coverslip followed by fixation in 4% of paraformaldehyde for 20 minutes at room temperature and then permeabilized with methanol at room temperature for 10 minutes. Blocking is then performed in 1% BSA and finally the coverslip is incubated overnight at 4ºC with primary antibodies (Table 3.2). In order to visualize the proteins under the microscope, the cells are incubated for 2 hours at room temperature with Alexa-conjugated secondary IgG antibodies (Table 3.2) and 0.1 ng/μL DAPI. These antibodies emit fluorescence that will be detected in the microscope. The images were captured by a Leica SP5 confocal microscope using LAS-AF software and quantification was achieved by counting cells in randomly located fields. The level of cell fluorescence in a specific region was measured using ImageJ tool (Schneider, Rasband, and Eliceiri, 2012).

3.3.3.2 Cell proliferation assay

This method was used to measure the proliferation rate of NSCs transduced with shControl and shPHF2. In detail, 1.7×10^5 NSCs were counted and seeded in precoated 6 well plates. After 24, 48 and 72 hours, cells were counted using the Neubauer chamber. Errors in the range of 20%-30% are common in this method due to pipetting and chamber volume errors. Due to these errors, multiple replicates were performed and the final number was calculated as the mean of all replicates. At this point is important to mention that after every counting, the cells were seeded in bigger sized plates so confluency will not interfere with their proliferation rate.

3.3.3.3 Cell synchronization

In this thesis we have used HeLa cells as an efficient model for cell cycle synchronization. The method we have used is the combination of thymidine treatment following the nocodazole treatment. Cells were first synchronized with a double thymidine block procedure before releasing into the nocodazole block. While it is possible to treat asynchronously growing HeLa cells with nocodazole directly, the yield and purity of the mitotic population are rather low. On the one hand, many cells remain in interphase if the nocodazole treatment is too short. On the other hand, cells may undergo mitotic slippage and apoptosis following a long nocodazole treatment (Ma and Poon, 2017). Synchronization in M phase was achieved by thymidine-nocodazole treatment including incubation in thymidine for 18 hours, washing with PBS twice, and supplementation with fresh medium with nocodazole for 12 hours. Thymidine (Sigma T9250) and nocodazole (Sigma 31430-18-9) were used at final concentrations of 2 mM and 100 ng/mL, respectively. After 12 hours of nocodazole treatment, the majority of the cells are arrested in mitosis as nocodazole is interfering with the polymerization of microtubules. Using this as a starting point on our experiment, counting 0 hours from cells release in fresh medium without nocodazole, we collected cells every two hours (Lim et al., 2013).

3.3.3.4 Cell cycle analysis by flow cytometry

Cell cycle analysis by quantitation of DNA content was one of the earliest applications of flow cytometry. This approach reveals distribution of cells

in three major phases of the cycle (G1 vs S vs G2/M) and makes it possible to detect apoptotic cells. This method is based on the use of DNA-binding dyes. In this thesis, propidium iodide staining was used as it binds to double stranded DNA by intercalating between base pairs. Cells were fixed in cold 70% ethanol and stored at -20°C until they were ready for staining. Cells were washed twice with cold PBS and pelleted by centrifugation. The cell pellet was resuspended in PBS containing 200 μL of 1 mg/mL propidium iodide and 2 mg RNase A and incubated at 37°C for 15 minutes. Then, the samples were added to the Gallios flow cytometer instrument (Beckman Coulter) and the red fluorescence (620/30 nm) was projected on a 1024 monoparametrical histogram and the aggregates were excluded. For the analysis the Wincycle software (Phoenix Flow Systems, San Diego, CA) was used.

3.3.4 Bioinformatic methods

Along my doctoral thesis I have collaborated with the team of Dr. Xavier de la Cruz for some bioinformatic analysis. Furthermore, I have performed some of the bioinformatic work using the menu-based bioinformatic platform Galaxy (Afgan et al., 2016). In the following subsections I will explain which tools were used for every analysis.

3.3.4.1 ChIP-sequencing analysis

In this subsection I will detail the tools I have used to analyze the ChIP-seq experiments we performed as well as the previously published ChIP-seq experiments.

50 base pairs sequences were mapped using Bowtie2 (Langmead and Salzberg, 2012) to the Mus musculus genome release 10 (mm10). Bam files were filtered to remove duplicates using SAMtools (Li et al., 2009) and peaks were called using MACS (Zhang et al., 2008b) with an effective genome size of 1870000000.0 and a p-value of 10^{-10} for PHF2 ChIP-seq and a p-value of 0.001 and a q-value of 0.05 for H3K9me2 ChIP-seq. The Bioconductor package ChIPseeker (Yu, Wang, and He, 2015) was used to annotate the genes of each peak. Specifically, the function annotatePeak matches peaks with genomic features extracted from mm10 (UCSC) and calculates the proportion of peaks matching each feature. It was also used to calculate the genome-wide distribution of the peaks. Lift-Over tool was

used to convert the genome coordinates (Hinrichs et al., 2006). ChIP-seq data have been deposited in the GEO database under the accession accession GSE122263.

3.3.4.1.1 Gene ontology analysis

The software Gene ontology was used to assign biological process categories of the input genes using as background the whole Mus Musculus genome (Ashburner et al., 2000; Consortium, 2017; Mi et al., 2017).

3.3.4.1.2 ChIP-sequencing data acquisition and analysis

During this doctoral thesis we have downloaded previously published ChIP-seq data for further analysis from GEO (accessions in Table 3.3). It is important to mention, that for all of the accessions, bed and wig files were already available and were used for heatmaps and Venn diagrams construction.

3.3.4.2 Venn diagrams construction

For the construction of Venn diagrams we have used the intersect intervals tool (Quinlan and Hall, 2010) in order to calculate the overlapping between two bed files, for at least 50-100 bases of overlap. Later on, the Venn diagrams were constructed using R studio. A p-value statistically evaluating the overlaps accompanies the diagrams. These p-values were obtained by generating a comparable random sample with shuffleBed and applying afterwards an equal proportions test.

3.3.4.3 Heatmaps construction

To generate heatmaps showing the colocalization of two proteins or a protein and a PTM around the transcription start site we first computed a matrix using the coordinates of one protein (as bed files) and the ChIP-seq intensities of the second protein or the PTM of our interest (as bigwig files). This matrix is generated using the computeMatrix command from Deeptools (Ramirez et al., 2016). Then, the matrix is plotted in a heatmap with plotheatmap tool. This tool allows the obtaining of the graphic (Ramirez et al., 2016).

3.3.4.4 Captures construction

In this thesis, I have included captions of the ChIP-seq signals of PHF2 protein and PTMs and captions of the RNA-seq signals for shControl and shPHF2 samples to observe specific genomic coordinates. These captions were obtained by loading wig/bigwig files of the different experiments in the IGV genome browser from the Broad Institute (Robinson et al., 2011).

3.3.4.5 Box-plots construction

Along the Results section, I have constructed Box-plots to represent the corresponding immunostaining quantification. Column entering replicate data followed by t-test was performed using GraphPad Prism version 8.00 for Mac, La Jolla California USA (https://www.graphpad.com).

3.3.4.6 RNA-sequencing analysis

For the RNA-seq analysis I collaborated with the team of Dr. Xavier de la Cruz for the bioinformatic analysis. Here I will mention the tools that they used for this analysis. Alignment of paired-end 50 base pair long reads was performed using HISAT2 (Kim, Langmead, and Salzberg, 2015), assignment of aligned reads to genes using HTSeq (Anders, Pyl, and Huber, 2015) and differential expression analysis using DESeq2 (Love, Huber, and Anders, 2014). RNA-seq data have been deposited in the GEO database under the accession GSE122264.

3.3.4.6.1 Expression heatmap construction

In the Results chapter I have included a heatmap in order to represent the expression profile of some genes of interest following the RNA-seq analysis. For the heatmap construction the Heatmapper tool (Babicki et al., 2016) was used.

3.3.5 Statistical analysis

In this last section of the Methodology I will provide details on which statistical tests have been applied to assess the reproducibility and significance of the results.

3.3.5.1 Sample size

In this doctoral thesis, ChIP-qPCR and immunostaining experiments have been performed in triplicate. In specific cases like validation of the obtained results with the second shRNA, the number of replicates have been two. Another exception is the genome-wide experiments, ChIP-seq, RNA-seq and MS. These experiments just show the results of one replicate in the case of ChIP-seq, two replicates that we included in the same run in the case of RNA-seq and one replicate in the case of MS. These experiments have been performed in one replicate due to the high cost of these techniques. Nonetheless, further validation procedures have been performed.

3.3.5.2 Standard deviation and standard error of the mean

Along the experimental work, the graphs have been represented as the mean. Error bars correspond to the standard deviation (SD) in the case of indirect immunofluorescences and RT-qPCR assays, and to the standard error of the mean (SEM) in the case of ChIPs. The numeric values have been calculated using Microsoft Excel software.

3.3.5.3 Student's t-test

To assess the significance of the results we have performed the Student's t-test. We have established that an experiment is statistically significant when within a 95% of confidence the result represents a true hypothesis. Asterisks represent the different p-values resulting from this test and calculation of the values was carried out with Microsoft Excel software. In the case of indirect immunofluorescence experiments and RT-qPCR assays the asterisks representing the p-value are the following: * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. In the case of ChIP-qPCR assays the asterisks representing the p-value are the following: * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

3.3.5.4 Equal proportion tests

Equal proportion tests were used to test the significance of the data that follow a non-parametric model. This test permits the evaluation of the differences in the proportions of a specific feature in different groups. This test

was performed to analyze the significance of Venn Diagrams. For this purpose, calculation was performed with R Studio using the prop.test function.

Chapter 4

Results

Along this doctoral thesis, I have obtained results regarding the function of PHF2 using *in vitro* (NSCs) and *in vivo* (the chicken neural tube) models. To begin with, in order to focus on the molecular mechanisms of this protein we have used NSCs.

4.1 *In vitro* characterization of PHF2 function in NSCs

4.1.1 Genomic distribution of PHF2

4.1.1.1 Genome-wide occupancy of PHF2

In order to gain insight into the biological role of PHF2 in NSCs, we first wanted to check the binding of this protein along the genome. For this reason, we performed ChIP-seq experiments in NSCs for the endogenous protein using a specific antibody that recognizes PHF2 (Table 3.2). Interestingly, from all the peaks obtained, we found that 97% of PHF2 peaks were localized on gene promoters and almost all of them around 1kb from the TSS (Figure 4.1). This result indicates that PHF2 may play a role regulating the expression of these genes.

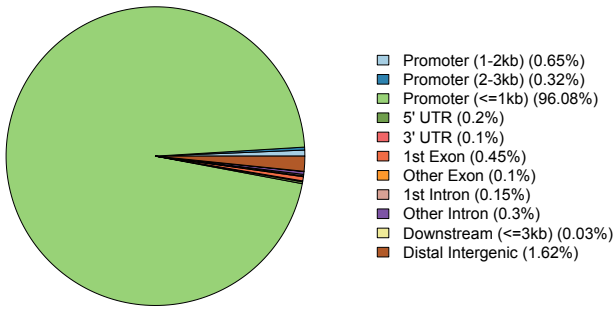


FIGURE 4.1: Genomic distribution of PHF2 ChIP-seq peaks in NSCs.

Next, we checked the gene ontology of biological processes of the genes that PHF2 was occupying and we observed that binds genes related with mRNA processing, mitotic nuclear division, regulation of cell cycle, histone modification and DNA repair. These categories were among others the most significantly enriched categories of biological processes (Figure 4.2).

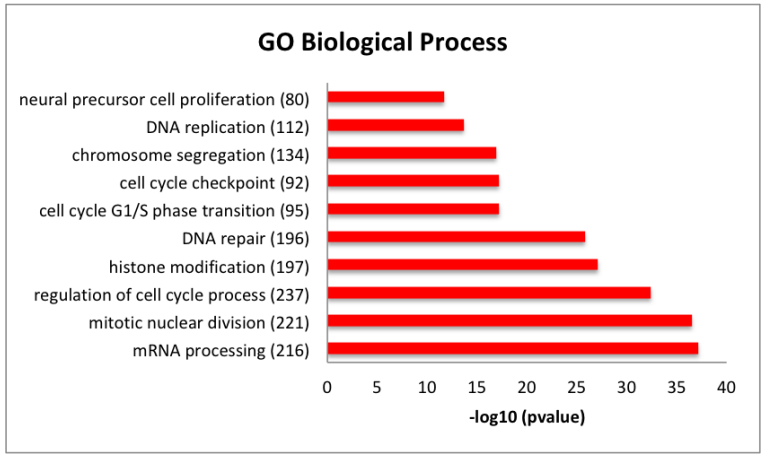


FIGURE 4.2: Gene ontology analysis showing Biological Processes of PHF2-bound genes was performed using as a background the whole Mus musculus genome. The number of genes in each category is depicted.

4.1.1.2 PHF2 colocalization with histone marks

It has been proposed that PHF2, through its PHD domain, interacts with H3K4me2/3 histone marks, thus, we analyzed the colocalization of PHF2 and H3K4me2/3 at the genome-wide level using previously published ChIP-seq data for H3K4me2/3 from NSCs (Table 3.3). Doing that, we identified 5978 (99.7%) (p-value<2.2e-16) and 5983 (99.8%) (p-value<2.2e-16) PHF2-bound regions that also contained H3K4me3 or H3K4me2 respectively. The calculated p-value is the result of an equal proportions test performed between H3K4me3 and H3K4me2 peaks and a random set. In order to check the colocalization profile of PHF2 and H3K4me2/3, we next constructed heatmaps showing the ChIP-seq intensities of the H3K4me2/3 in the PHF2 peaks (Figure 4.3). We observed a high overlapping that suggests a required H3K4me2/3 mark for an effective PHF2 binding at promoters as it has been previously described (Wen et al., 2010). To further visualize this result, a randomly picked gene (Mcm6) that contains PHF2 in its promoter is shown in the Figure 4.4. The absolute colocalization with H3K4me2/3 marks is observed.

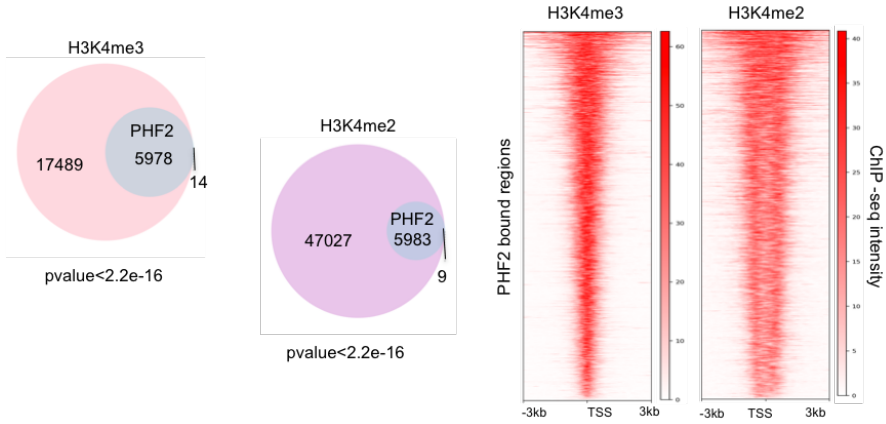


FIGURE 4.3: Venn diagrams (left) showing overlap between PHF2-bound and H3K4me3 and H3K4me2-marked regions. Heatmaps (right) are depicting PHF2 binding to H3K4me3 and H3K4me2-marked promoters in NSCs 3kb around the TSS. Scales indicate ChIP-seq intensities.

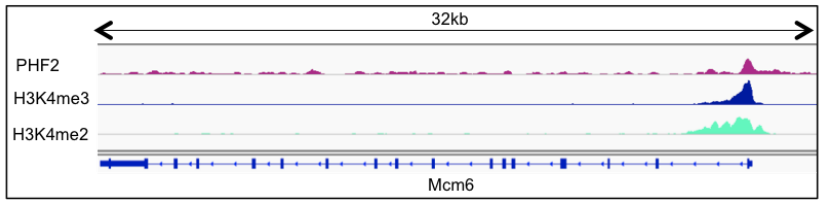


FIGURE 4.4: IGV capture showing PHF2, H3K4me3 and H3K4me2 peaks in Mcm6 gene in NSCs.

4.1.1.3 Overlapping of PHF2 with E2F transcription factors

Next, to better understand how PHF2 is targeted to the chromatin, we performed bioinformatic analysis that revealed the most statistically significant predicted PHF2 binding sites. Among these sites, we identified E2F/E2F4 and SOX2/3 DNA binding motifs, as seen in Figure 4.5. These transcription factors are essential to control the cell cycle progression and

progenitor self-renewal, suggesting a potential role of PHF2 in cell proliferation.

FIGURE 4.5: Motif enrichment analysis of PHF2 ChIP-seq peaks in NSCs using Homer known motif showing the top enriched motifs.

In addition, to reinforce these data, we analyzed the colocalization of PHF2 and E2F1 or E2F4 using previously published ChIP-seq data from HeLa S3 cells. The results showed that the 49.8% (p-value<2.2e-16) and 50.3% (p-value<2.2e-16) of E2F1 and E2F4-bound regions respectively also contained PHF2 (Figure 4.6).

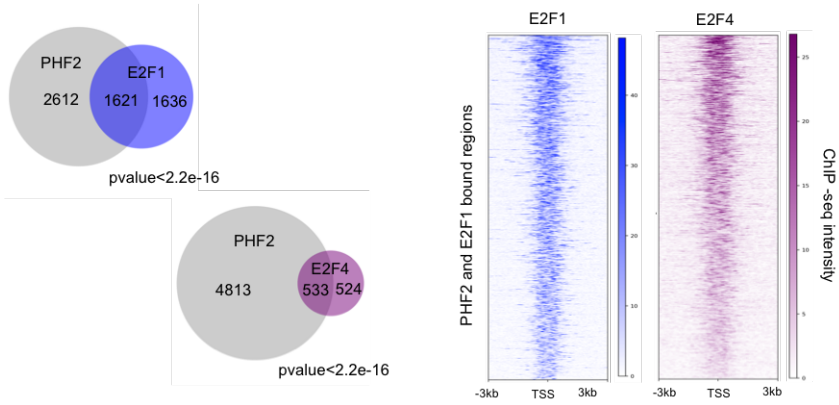


FIGURE 4.6: Venn diagrams (left) showing peak overlapping between PHF2-bound and E2F1 and E2F4 bound regions in published ChIP-seq in HeLa S3 cells. Heatmaps representation (right) of E2F1 and E2F4 binding on the PHF2 and E2F1 co-occupied regions.

4.1.2 PHF2 mediates H3K9 demethylation

4.1.2.1 Genome-wide occupancy of H3K9me2

As mentioned in the introduction of this thesis, previous studies demonstrated that PHF2 demethylates mainly H3K9me2 (Wen et al., 2010). We thus analyzed the distribution of H3K9me2 in NSCs by performing ChIP-seq experiments in order to compare the genomic distribution of this histone mark with PHF2 genomic associated regions. We observed that the majority of the peaks were localized on distal intergenic regions (58.61%) as well as on intronic regions (34.06%).

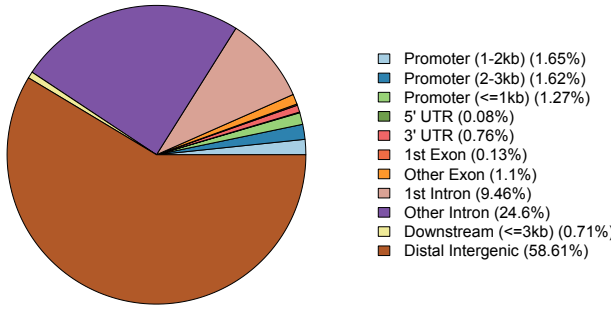


FIGURE 4.7: Genomic distribution of H3K9me2 ChIP-seq peaks in NSCs.

Furthermore, we checked for colocalization of the peaks of PHF2 with the ones of H3K9me2 and we didn't observe any peak overlapping. The

great majority (99.99%) (p-value<2.3e-6) of PHF2-bound regions were totally excluded from the H3K9me2 positive regions in NSCs as seen in Figure 4.8, indicating a potential role of PHF2 as a demethylase of H3K9me2. One example is depicted in the capture, where is highlighted that the presence of PHF2 is following the absence of H3K9me2 (Figure 4.9).

It is worth noting, that very few peaks of H3K9me2 are occupying promoter regions (4.54%), so a high overlapping between PHF2 and H3K9me2 could not be expected even if PHF2 was not demethylating H3K9me2.

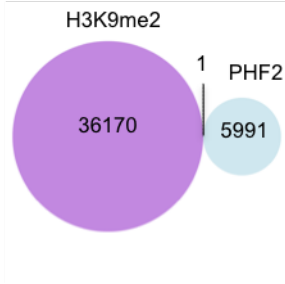


FIGURE 4.8: Venn diagram showing overlap between PHF2-bound and H3K9me2-marked regions in NSCs.

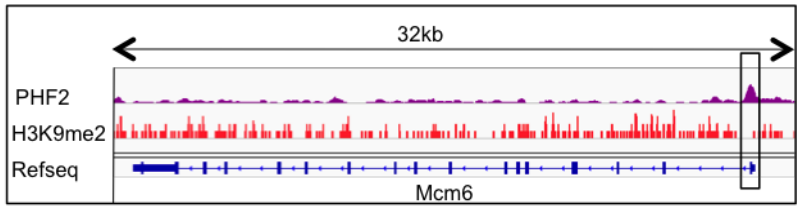


FIGURE 4.9: IGV capture showing PHF2 peaks and H3K9me2-marked regions in Mcm6 gene. The promoter region in the box shows no H3K9me2 enrichment where PHF2 binds.

Apart from H3K9me2, similar results were observed for H3K9me3-enriched regions identified by previously published ChIP-seq data from NSCs (Table 3.3). In the Figure 4.10 we can observe only 52 regions of PHF2 that also contain H3K9me3. In the introduction we mentioned that H3K9me2/3 is related with gene silencing although H3K4me2/3 is mainly

occupying promoters of active genes. These results demonstrate that PHF2 occupies genes lacking H3K9me2/3 repressive marks and suggest that PHF2 might demethylate these residues.

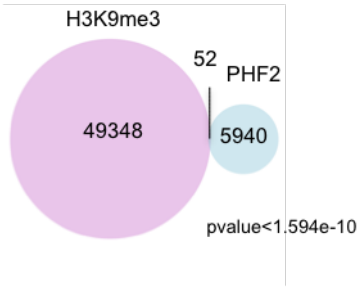


FIGURE 4.10: Venn diagram showing overlap between PHF2-bound and H3K9me3-marked regions in NSCs.

4.1.2.2 Generation of PHF2 KD NSCs

To evaluate the impact of PHF2 depletion on histone modifications *in vivo*, NSCs were transduced with lentivirus containing specific PHF2 shRNA. These shRNAs efficiently decreased the protein as well as the RNA levels of PHF2 compared to the control shRNA (shControl). For this reason we generated two different shPHF2 cell lines (shPHF2, shPHF2_2). We then checked by qPCR that the reduction of PHF2 did not affect the expression of its homologous, PHF8, as seen in Figure 4.11.

At this point, I would like to clarify that as shPHF2 was more efficient depleting PHF2 we have mainly used this shRNA, although the majority of the results have been confirmed using the second shRNA (shPHF2_2).

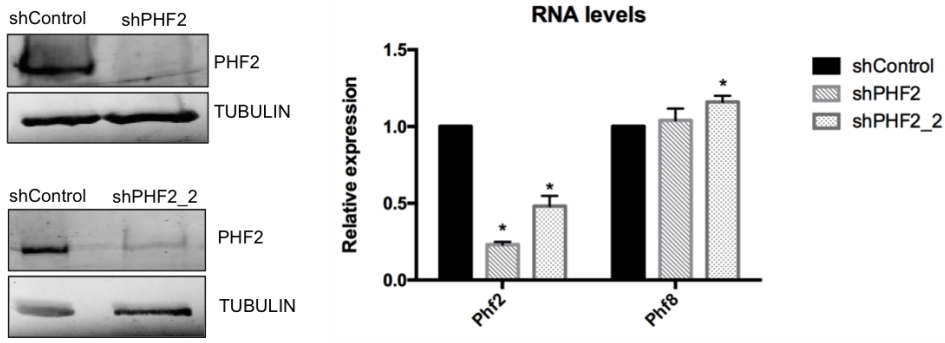


FIGURE 4.11: Western blots showing the protein (left) and column graph showing the mRNA levels (right) of PHF2 in shControl, shPHF2 and shPHF2_2 NSCs. Error bars indicate SD *P< 0.05 (Student's test)

4.1.2.3 Impact of PHF2 depletion in H3K9 methylation

To gain insight about the role of the demethylase activity of PHF2, after viral transduction with shRNAs, we analyzed the levels of H3K9me2 by immunofluorescence assays. In the Figure 4.12 we observed a slight increase on H3K9me2 levels as it has been demonstrated in other cellular contexts (Wen et al., 2010). As H3K9me2 mark serves as a substrate for the histone methyltransferase SUV39H, we tested the levels of H3K9me3 in PHF2-depleted NSCs. The results are clearly showing an increase on H3K9me3 levels (Figure 4.12). Cell fluorescence for H3K9me2 staining and for H3K9me3 foci was measured using ImageJ. Intriguingly, both the intensity and the number of H3K9me3 foci increased upon PHF2 depletion.

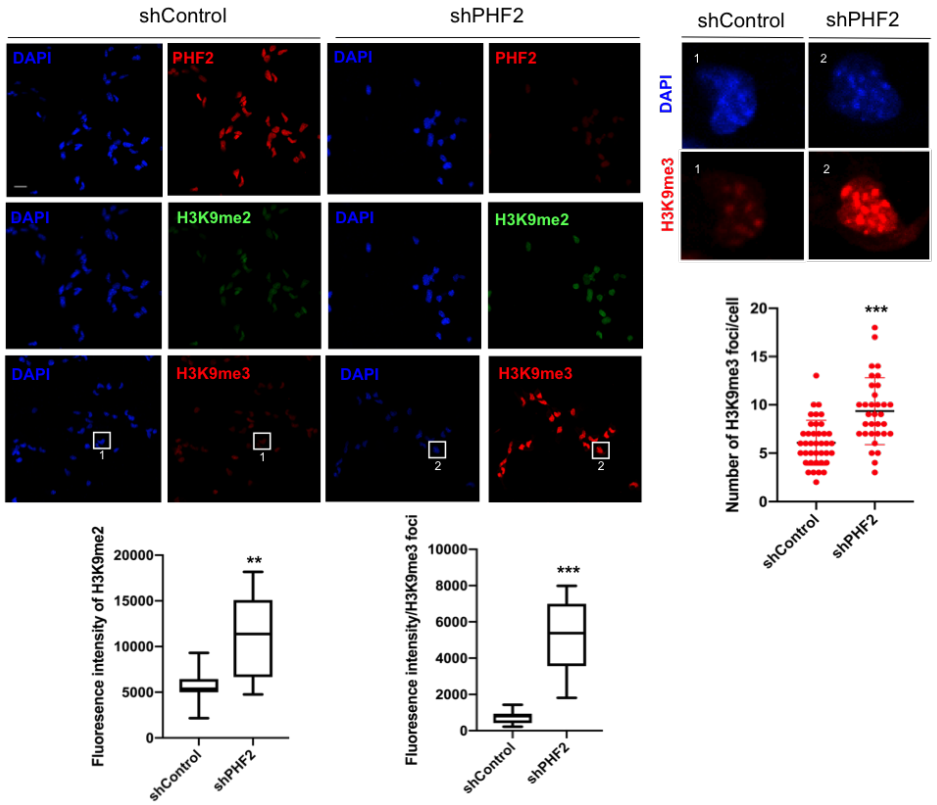


FIGURE 4.12: Immunostaining of shControl and shPHF2 cell lines. Cells were fixed and stained with PHF2, H3K9me2 and H3K9me3 antibodies and DAPI. Scale bar indicates 20 μm. Zoom in showing H3K9me3 foci in shControl and shPHF2. Boxplots represent the quantification of the fluorescence intensities as well as the number of H3K9me3 foci/cell in shControl and shPHF2 cells. **P < 0.01; ***P < 0.001 (Student's t-test)

Moreover, we sought to test whether PHF2 is also important to prevent H3K9me3 increase at local level. To do so, we chose some PHF2-target regions identified in the ChIP-seq experiment and tested the effect of PHF2 depletion on H3K9me3 levels by ChIP-qPCR assays. A clear increase on

H3K9me3 mark was observed, at the analyzed promoters, without affecting the intragenic region of Gda gene (a non PHF2-target used as a negative control) (Figure 4.13), in accordance with the global increase noticed in Figure 4.12. These data strongly suggest that PHF2 prevents H3K9me3 accumulation, limiting ectopic heterochromatin formation, particularly at promoter regions.

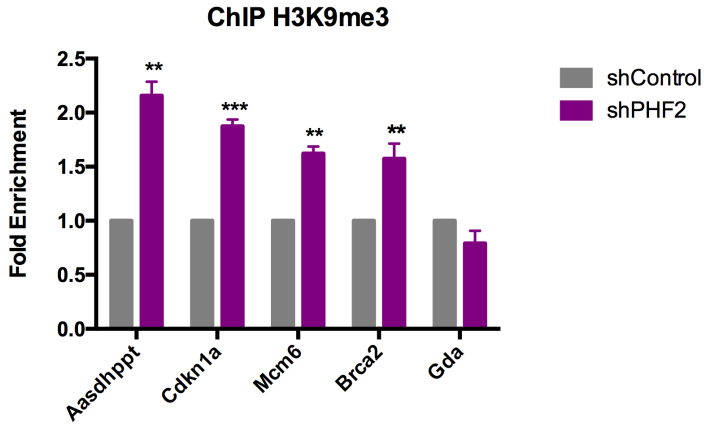


FIGURE 4.13: ChIP of H3K9me3 in shControl and shPHF2 cells was analyzed by qPCR at the indicated gene promoters that were identified by PHF2 ChIP-seq as PHF2 targets. Data from qPCR were normalized to the input and expressed as fold enrichment over the data obtained in shControl. Results are the mean of two to three biologically independent experiments. Errors bars represent SEM. **P<0.05; ***P<0.01 (Student's t-test)

4.1.2.4 Impact of PHF2 depletion in HP1α

The global increase of H3K9me3 led as to check whether the protein responsible for H3K9me3 recognition was affected at the global levels as well. We observed a clear accumulation of HP1α in PHF2-depleted cells without increase in other histone marks, such as H3K4me3 (Figure 4.14). These data may suggest that depletion of PHF2 led either to an increase of HP1α levels or to a change on its localization at global levels.

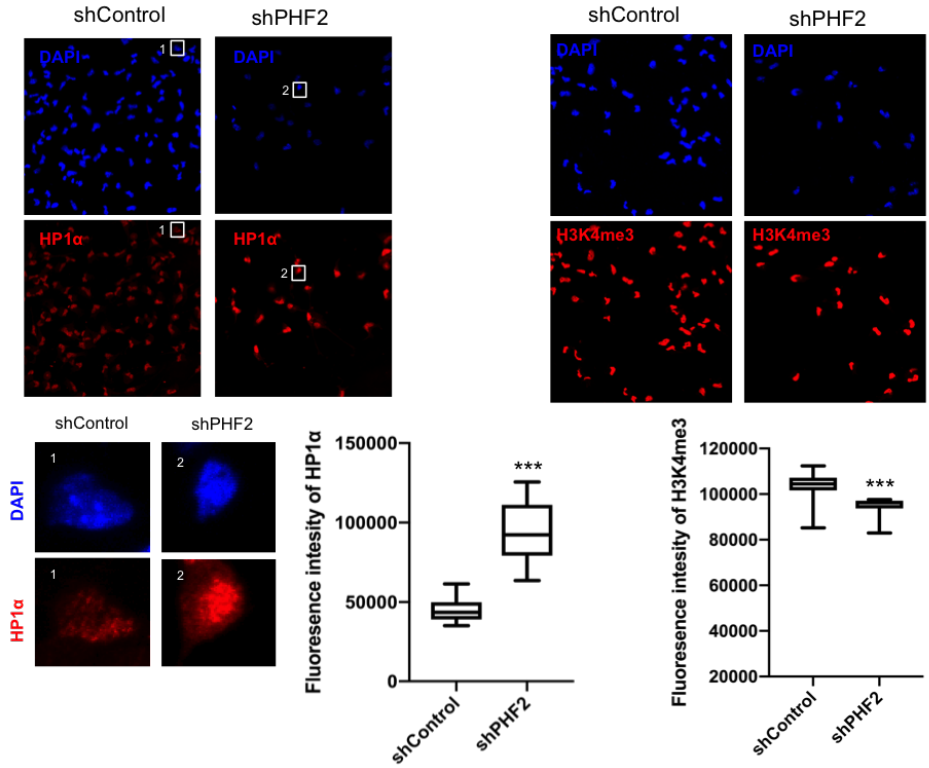


FIGURE 4.14: Immunostaining of shControl and shPHF2 cell lines. Cells were fixed and stained with HP1α and H3K4me3 antibodies and DAPI. Zoom in showing HP1α in shControl and shPHF2. Cell fluorescence was measured using ImageJ. Boxplots represent the quantification of the fluorescence intensities. ***P <0.001 (Student's t-test)

4.1.3 PHF2 regulates gene expression

4.1.3.1 Transcriptional profile of PHF2-depleted cells

In order to gain further knowledge into the function of PHF2 in NSCs, we identified the PHF2 associated transcriptional profile by RNA-seq in PHF2-depleted NSCs (shPHF2) compared with control (shControl) cells. We identified 1729 transcripts that significantly changed their expression (log2 fold change (FC)>0.8 and p-value <0.01) in shPHF2 cells compared

with shControl cells. Two biological replicates of PHF2-depleted cells were used and compared to control cells. The differential expressed genes, the upregulated and the downregulated ones, are represented in Figure 4.15.

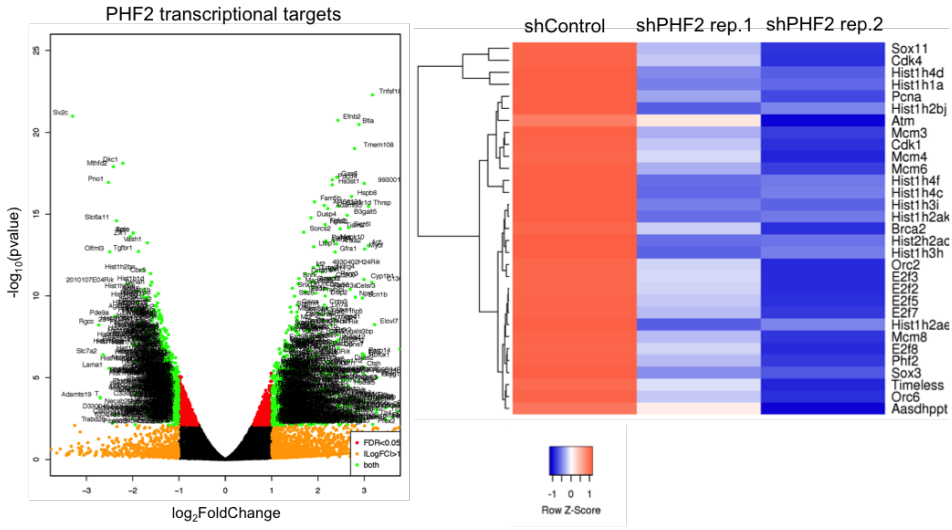


FIGURE 4.15: Volcano plot (left) represents PHF2 transcriptional targets identified by RNA-seq in shControl and shPHF2 NSCs. The green dots represent all the genes with FDR<0.05 and log2FoldChange>1. Heatmap (right) is showing some downregulated genes identified by RNA-seq in shControl and shPHF2. Two biological replicates of shPHF2 cells were used for RNA-seq. All the genes showed p-value<0.01 and log2FoldChange>0.8.

4.1.3.2 Identification of PHF2 direct targets

Next, we sought to identify which of the transcriptional regulated genes upon PHF2 depletion had as well this demethylase bound on their promoter region. For this reason we compared the previously obtained ChIP-seq data with the RNA-seq data and we identified the PHF2 direct transcriptional targets. Among the genes that showed a PHF2-dependency for transcriptional regulation (log2 fold change (FC)>0.8) in the RNA-seq experiment, 601 (34,8%) were bound by PHF2. Some randomly picked genes

are highlighted in the Figure 4.16. Interestingly, the proportion of downregulated transcripts (68%) was higher than the upregulated ones (32%). This percentage was almost maintained at log2FC>1 in accordance with the activator role proposed for PHF2 (Lee et al., 2014).

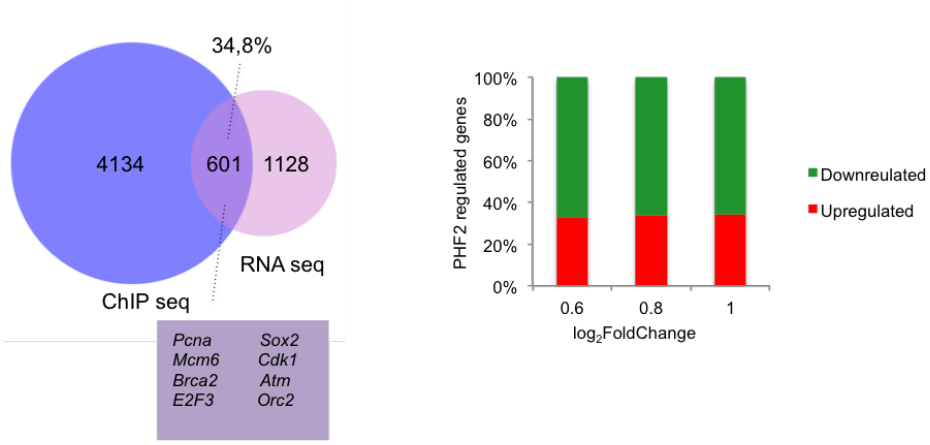


FIGURE 4.16: Venn diagram (left) showing overlapping between PHF2-bound genes and PHF2 transcriptional targets. Some of these genes are depicted. Graph (right) depicting the percentage of upregulated and downregulated genes in the shPHF2 according to the RNA-seq that contain PHF2 in their promoter classified by log2 of fold change.

4.1.3.3 Gene ontology terms of PHF2 direct targets

In addition, enrichment analysis of GO terms over these 601 PHF2-direct target genes showed that the most enriched were associated with cell cycle categories, particularly G1-S transition (E2f2/3/7/8, Cdc7, Cdc25a, Cdk4, Mcm3/4/8), DNA replication (Orc1/2/6, Pcna), mitosis (Cdk1, Smc2/3/4, Aurkb, Topo2a), as well as chromatin activity (Cenpa, Kdm1b, Hat1, Parp1, Prmt5) (Figure 4.17). Intriguingly, some of them were E2F target genes (Ccnd1, Cdc25a, Pcna, Mcm3/4/6/8, Smc4, including E2f family genes). One example of the RNA-seq profile of an E2F target gene that contained PHF2 is shown. In the capture, the signal of the RNA-seq in all the exons of Mcm6 gene is presented in shControl compared with shPHF2 cells (Figure 4.17). To further characterize the differences between control and PHF2 KD

NSCs, we performed an enrichment analysis of GO of downregulated and upregulated genes to identify those biological processes most sensitive to PHF2 depletion. The results revealed that the downregulated genes upon PHF2 depletion were strongly associated to cell cycle progression, chromatin activity and DNA repair among others. The same analysis of the upregulated genes did show functional categories related to morphogenesis, signal transduction and developmental process (Figure 4.18).

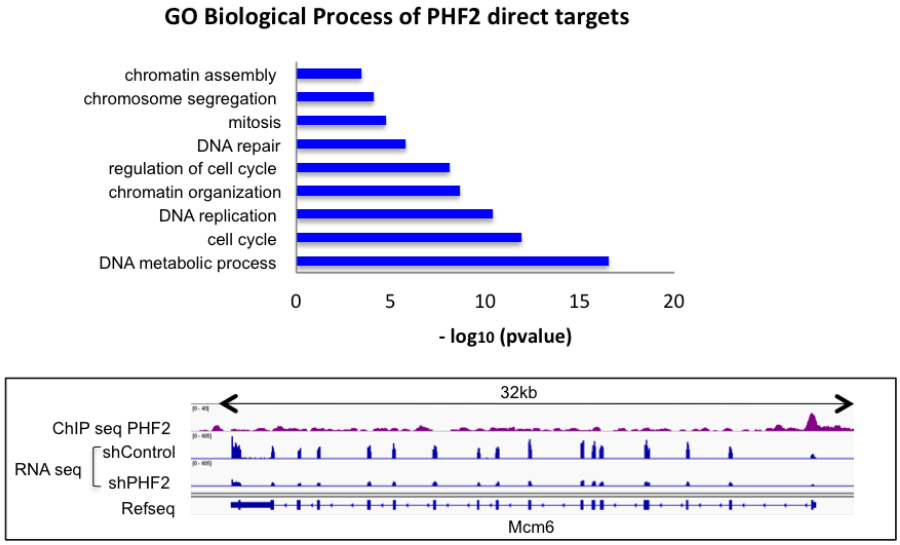


FIGURE 4.17: Gene ontology analysis showing Biological Process of the PHF2 direct targets (top), was performed using as a background the whole Mus musculus genome. IGV capture (bottom) showing PHF2 peaks and RNA levels in shControl and shPHF2 in Mcm6 gene.

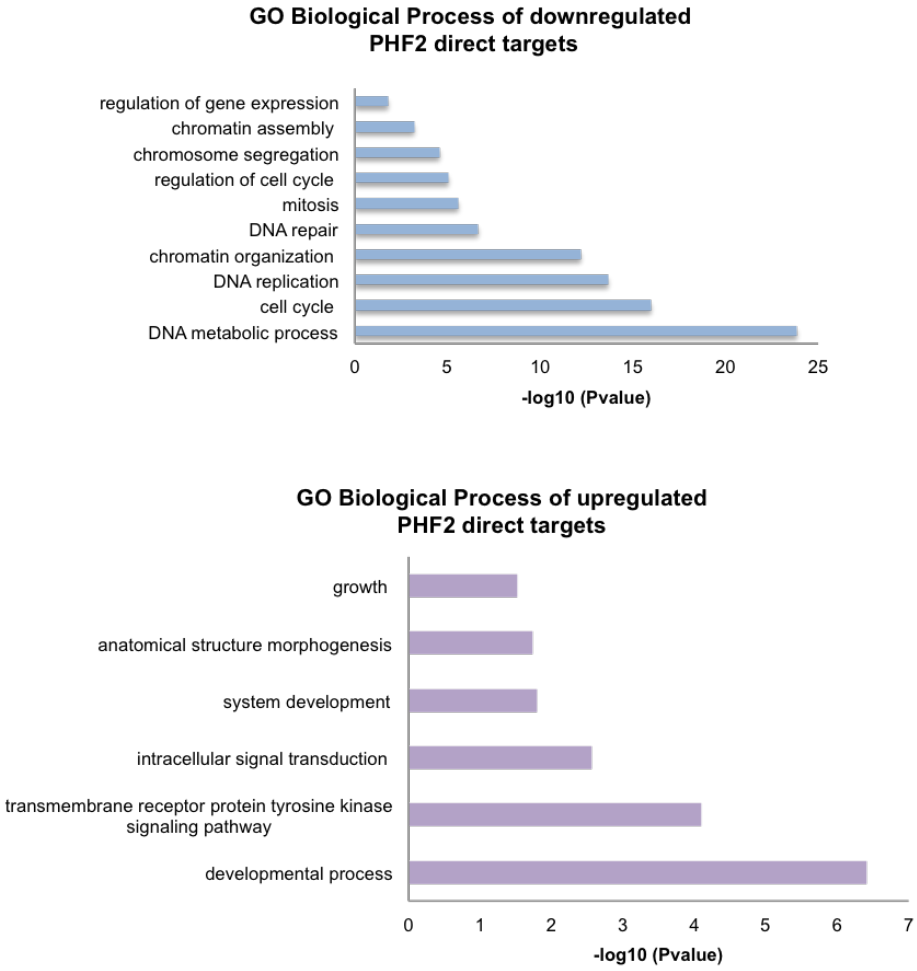


FIGURE 4.18: Gene ontology analysis showing Biological Process of PHF2 of downregulated (top) and upregulated (bottom) direct targets, was performed using as a background the whole Mus musculus genome.

Furthermore, we confirmed the previous results upon depletion of PHF2 by using another shRNA. NSCs were infected with lentivirus expressing shControl or shPHF2_2 and we tested by qPCR analysis some randomly

chosen genes (Figure 4.19).

Altogether, these data strongly suggest that PHF2 binds to the cell cycle gene promoters to fine-tune their chromatin activity and facilitate their transcription

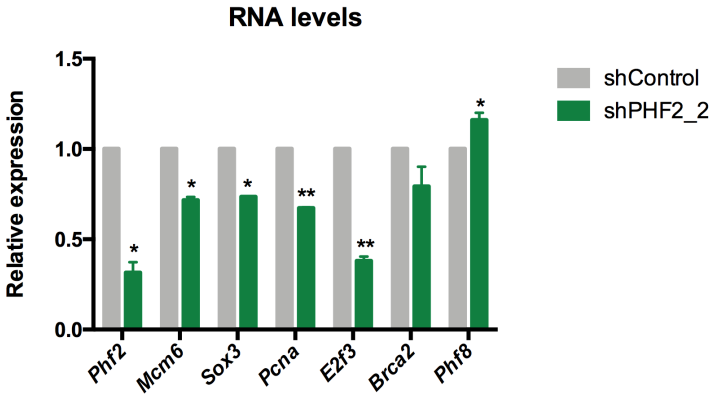


FIGURE 4.19: mRNA levels of PHF2, and randomly picked genes that appeared downregulated in the RNA-seq of shPHF2 were analyzed using shPHF2_2 by qPCR. Errors bars represent SD. *P< 0.05; **P< 0.01 (Student's t-test)

4.1.3.4 PHF2 regulates cell proliferation

The above-described findings support the idea that PHF2 is important to regulate cell cycle genes. To further study the potential role of PHF2 in cell proliferation, we analyzed the consequences of its reduction. PHF2-depleted NSCs exhibited a striking decrease on cell growth. Moreover, flow-cytometry analysis demonstrated a delay in G1/S transition (G1 shControl 42.2%, shPHF2 53.5%) and revealed a slight increase in G2-M phase population upon PHF2 depletion (Figure 4.20). Interestingly, the same findings of a delay in G1/S transition have been observed for PHF8 (Liu et al., 2010).

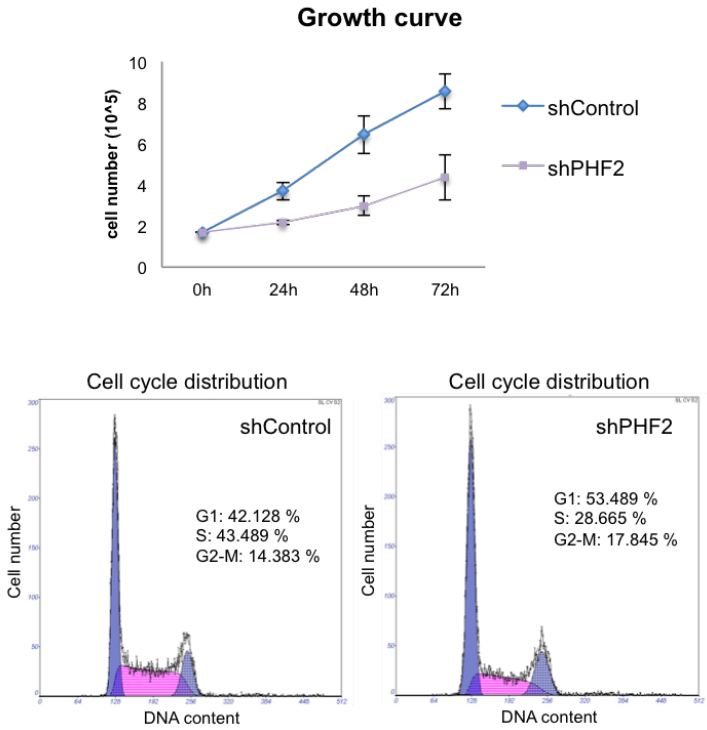


FIGURE 4.20: Growth curve showing the proliferation rate of NSCs infected with lentivirus expressing shRNA control (shControl) or shRNA for PHF2 (shPHF2) from 0 to 72 hours after plating. Flow-cytometry analysis (bottom) of NSCs Control and PHF2-depleted cells previously stained with propidium iodide.

4.1.3.5 Expression of PHF2 is regulated during cell cycle

Moreover, we sought to test if the expression of PHF2 was regulated throughout the cell cycle. For this reason we used HeLa cells because they can be easily synchronized, as mentioned in the subsection 3.3.3.3. Therefore, we observed that the expression of PHF2 reached higher levels during G1/S transition (Figure 4.21). Nonetheless, the lowest expression levels of PHF2 occur at the end of the S phase. CYCLIN E1 was used as a control for the synchronization process as it reaches the highest level of expression at G1 phase.

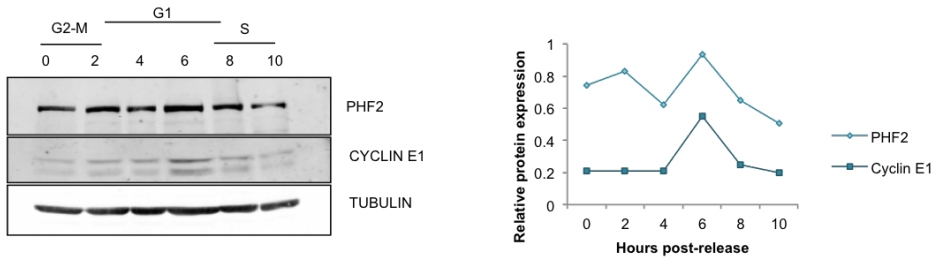


FIGURE 4.21: HeLa cells were synchronized at M phase and harvested at 2 hours intervals upon release for 10 hours. Total protein were prepared and the PHF2, G1/S marker CYCLIN E1 and TUBULIN levels were determined by immunoblot (left). ImageJ quantification of data was carried out with normalization against TUBULIN (right).

4.1.3.6 PHF2 binding at ORC1 promoter during G1/S transition

The expression of PHF2 is regulated during G1/S transition, as described above. In the subsection 4.1.3.3 we showed that PHF2 directly regulates genes implicated in G1/S transition and DNA replication. For this reason, we sought to test if PHF2 binding to the promoter region of one of these genes is regulated during the G1/S transition. To do so, ChIP-qPCR of PHF2 was performed in synchronized HeLa cells at G1 and at S phases. As seen in Figure 4.22, PHF2 binding at Gapdh promoter (a PHF2 target gene that is ubiquitously expressed during the cell cycle) is similar at 6 and 12 hours upon nocodazole release whereas the binding at Orc1 promoter increases during the S phase of the cell cycle, correlating with Orc1 induction.

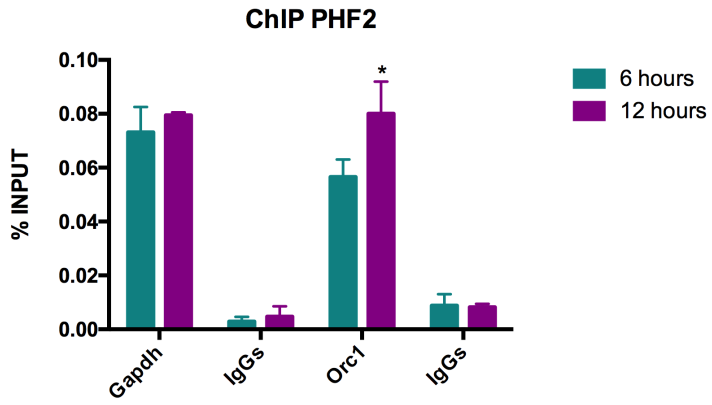


FIGURE 4.22: HeLa cells were synchronized at M phase and harvested after 6 and 12 hours upon release. ChIP of PHF2 at 6 hours (G1 phase) and 12 hours (S phase) after release was analyzed by qPCR at the indicated gene promoters. Gapdh was used as a control. Data from qPCR were normalized to the Input. Errors bars represent SEM. *P<0.1 (Student's t-test)

4.1.4 PHF2 maintains pericentromeric heterochromatin integrity

4.1.4.1 Expression of repetitive elements upon PHF2 depletion

As mentioned in the introduction, H3K9me2/3 is a fundamental component of pericentromeric heterochromatin. The mouse centromere has a centromeric region flanked by pericentromeric heterochromatin at one side. The centromeric region is transcribed into minor satellite RNA and the pericentromeric region is transcribed into major satellite RNA. In addition, PHF2 is the only member of the KDM7 family described as a protein associated to the kinetochore (Ohta et al., 2010). For this reason we depleted PHF2 in NSCs and tested by qPCR the transcription coming from major and minor satellites and other interspersed repetitive elements along the genome, including L1-LINES, SINES and IAP1. We found that PHF2 depletion strikingly increased the transcription of the major and minor satellites but not of the Ppia gene used as a negative control. A slight increase was also observed in the rest of the interspersed repetitive elements (Figure 4.23).

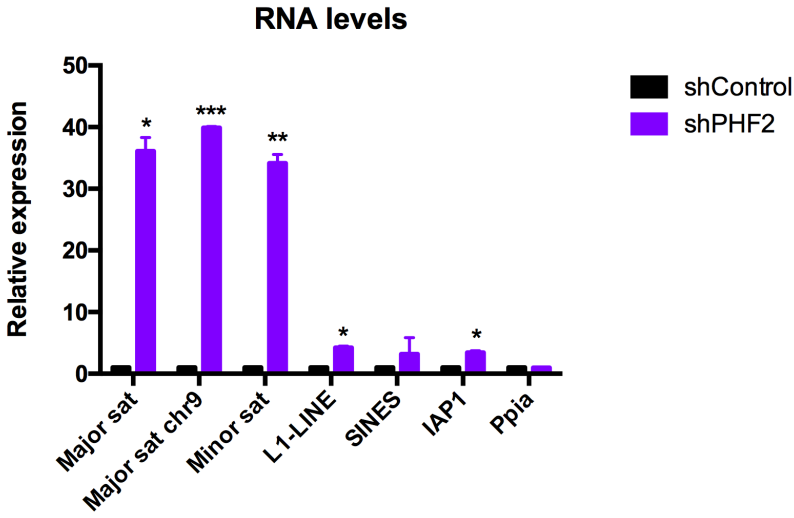
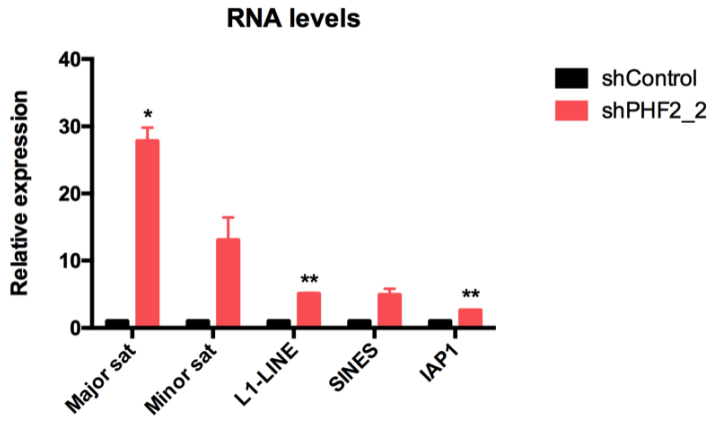


FIGURE 4.23: RNA levels of major and minor satellites as well as other repetitive sequences in shControl and shPHF2 NSCs were analyzed by qPCR. Data were normalized to Gapdh reference gene and figure shows values relative to shControl. Ppia gene was used as a negative control. Error bars indicate SD. *P< 0.05; **P< 0.01; ***P<0.001 (Student's t-test)

To confirm these data we used another shRNA. Similar effects were observed when PHF2 was depleted using shPHF2_2 (Figure 4.24). Moreover, a clear enrichment on transcripts coming from major satellite and in particular from the major satellite of chromosome 9 was detected in the RNA-seq experiment comparing the control and PHF2-depleted cells.



Number of hits	shControl	shPHF2 rep1	shPHF2 rep2
Major satellites	498	899	1795
Major satellites chr9	170	415	757

FIGURE 4.24: RNA levels of major and minor satellites as well as other repetitive sequences in shControl and shPHF2_2 NSCs were analyzed by qPCR (top). Data were normalized to Gapdh reference gene and figure shows values relative to shControl. Error bars indicate SD. *P< 0.05; **P< 0.01 (Student's t-test). Scheme representing the number of BLAST hits for major satellites and major satellites of chromosome 9 in the RNA-seq for shControl and shPHF2 (bottom)

4.1.4.2 Implication of the PHF2 catalytic activity in the expression of repetitive elements

As PHF2 acts as demethylase of H3K9me2 and this mark is a fundamental component of pericentromeric heterochromatin together with H3K9me3 we sought to test the implication of the PHF2 catalytic activity in heterochromatin integrity. For this reason, we established NSC lines that expressed either PHF2 WT or the catalytic dead mutant (HID>AIA) (Methods 3.3.1.3) that works as a dominant negative form of PHF2 upon addition of Doxycycline. As seen in Figure 4.25 both PHF2 WT and PHF2 HID>AIA were equally overexpressed upon Doxycycline induction.

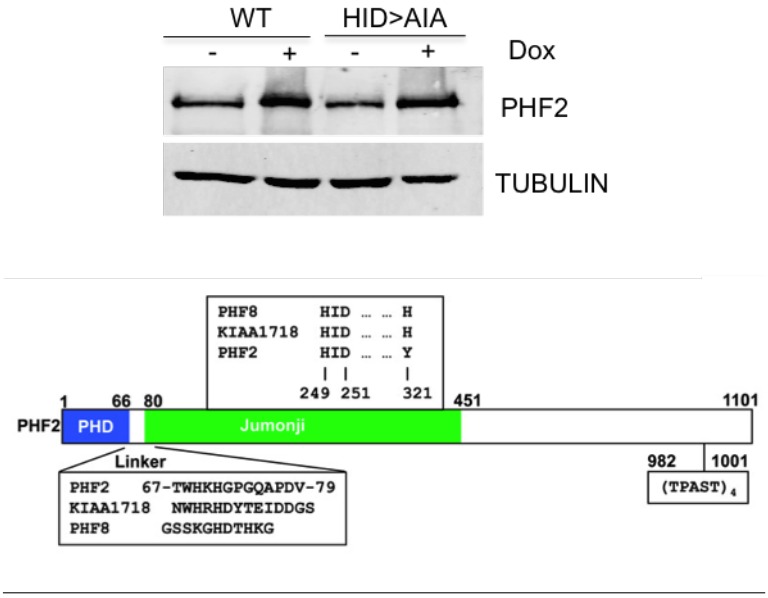


FIGURE 4.25: NSCs were infected with lentivirus expressing pINDUCER vectors for overexpression of PHF2 WT and PHF2 catalytic mutant (HID>AIA) after doxycycline induction. PHF2 and TUBULIN levels were determined by immunoblot before (-) and after (+) doxycycline addition (top). The linker sequences and the iron binding residues of the KDM7 family are indicated (bottom). Adapted from (Horton et al., 2011)

After obtaining these NSC lines, we wanted to test whether the overexpression of the PHF2 WT or PHF2 catalytic mutant, that works as a dominant negative, was affecting the transcription of centromeric and pericentromeric repeats. Induction of PHF2 WT did not alter the transcription from the satellite repeats, however overexpression of PHF2 catalytic mutant (HID>AIA) clearly induced transcription from both major and minor satellites (Figure 4.26). These data suggest that the catalytic activity of this demethylase is required to prevent aberrant transcription from pericentromeric repeats.

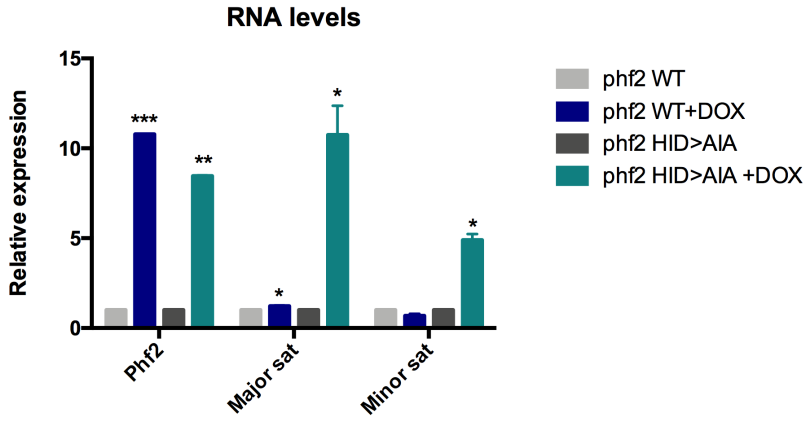


FIGURE 4.26: RNA levels of major and minor satellites in NSCs expressing pINDUCER PHF2 WT and PHF2 catalytic mutant (HID>AIA) before (-) and after (+) Doxycycline induction were analyzed by qPCR. Data were normalized to Gapdh reference gene and figure shows values for PHF2 WT +DOX relative to PHF2 WT and for PHF2 HID>AIA +DOX relative to PHF2 HID>AIA. Error bars indicate SD. *P < 0.05; **P < 0.01; ***P<0.001 (Student's t-test)

4.1.4.3 PHF2 is located in pericentromeric and centromeric region

These results above prompted us to test whether PHF2 directly regulated transcription from satellite repeats by physically contacting them. To do so, we analyzed the binding of PHF2 by ChIP-qPCR assays. As seen in Figure 4.27, PHF2 was enriched in both, major and minor satellites, but not in Col2a1 promoter used as a negative control.

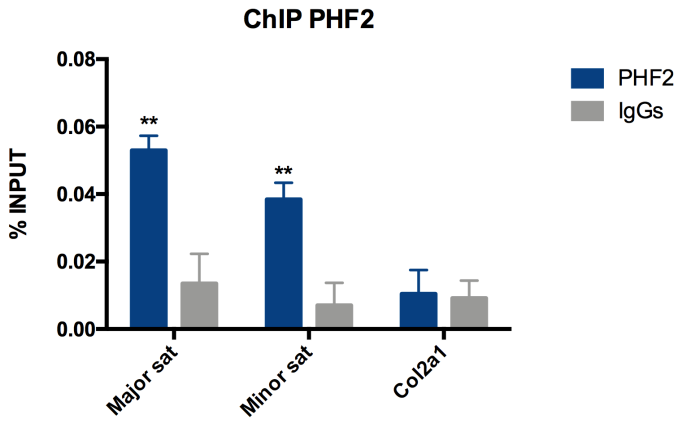


FIGURE 4.27: ChIP of PHF2 in NSCs analyzed by qPCR at major and minor satellites. The promoter of Col2a1 gene was used as negative control. Results are the mean of two biologically independent experiments. Errors bars represent SEM. **P<0.05 (Student's t-test)

4.1.4.4 Interaction between PHF2 and the Heterochromatin Protein 1-Binding Protein 3

In order to identify proteins that can possibly interact with PHF2 we performed Mass spectrometry (MS) analysis. To do that, we immunoprecipitated the endogenous protein from NSC extracts and the associated factors were identified by MS analysis. Among these proteins, we identified the heterochromatin protein 1-binding protein 3 (HP1BP3). In order to test this interaction, we overexpressed HP1BP3-HA and PHF2 in HEK 293T cells. We have used this model taking advantage of its high transfection rate. The interaction between these two proteins was detected by co-immunoprecipitation (Co-IP) experiments (Figure 4.28).

Protein identified by MS	Score	Coverage
Heterochromatin protein 1-binding protein 3 (HP1BP3)	30.63	2.35

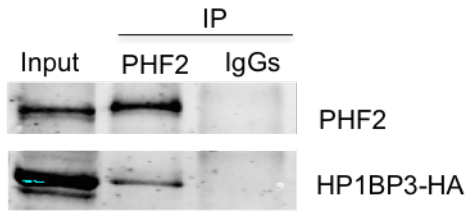


FIGURE 4.28: Endogenous PHF2 protein was extracted from NSC cells in order to perform MS. HP1BP3 was identified, among others, as a possible PHF2 interactor. The score and the coverage are shown in the table (top). HEK 293T cells were transfected using overexpression vectors for PHF2 and HP1BP3-HA. PHF2 was immunoprecipitated using PHF2 antibody and the presence of HP1BP3 in the immunopellet was determined by immunoblot using HA antibody (bottom). IgGs were used as negative control.

4.1.4.5 Interaction between PHF2 and the Methyltransferase SUV39H1

Following these findings, we tested whether PHF2 interacts with main components of pericentromeric heterochromatin. Indeed, it has been previously described that PHF2 binds to a component of heterochromatin, SUV39H1 (Shi et al., 2014). For this reason, we overexpressed SUV39H1-FLAG and PHF2 in HEK 293T cells. The interaction between these two proteins was detected by Co-IP experiments (Figure 4.29).

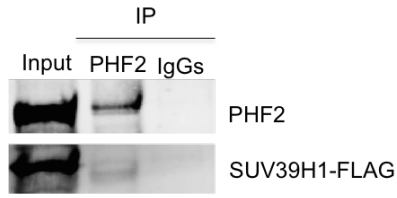


FIGURE 4.29: HEK 293T cells were transfected using overexpression vectors for PHF2 and SUV39H1-FLAG. PHF2 was immunoprecipitated using PHF2 antibody and the presence of SUV39H1 in the immunopellet was determined by immunoblot using FLAG antibody. IgGs were used as negative control.

4.1.4.6 Effect of PHF2 depletion on pericentromeric H3K9me2/3

As mentioned in the subsection 4.1.4.2, the catalytic activity of PHF2 is required to prevent the transcription from pericentromeric repeats. Thus, we sought to determine whether changes on H3K9me2/3 occur concomitant with the unprogrammed transcription. To do so, we chose the major satellite of chromosome 9 to analyze the surrounding non-repetitive genomic region (to 35kb from the pericentromeric repeats) (Figure 4.30) and the major satellite.

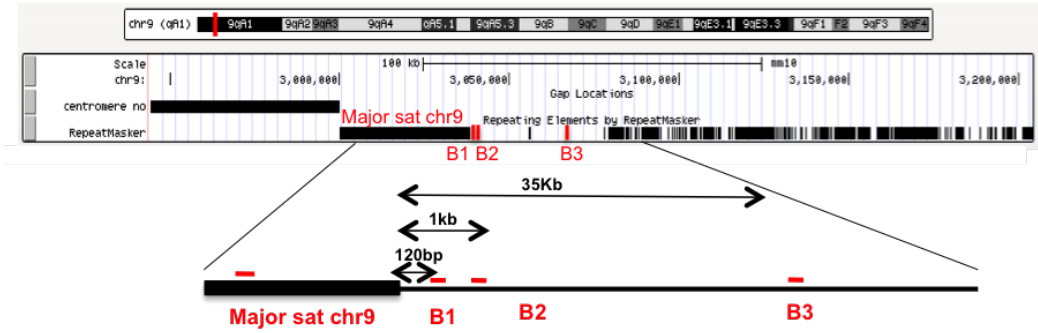


FIGURE 4.30: Scheme from UCSC genome browser is showing the genomic location of major satellite in chromosome 9 as well as the location of the primers used (B1-B3) for the nearby region, pointing out the distance to the major satellite.

The levels of H3K9me2 and H3K9me3 at these regions in control and PHF2-depleted NSCs were analyzed by ChIP-qPCR. The graphs in Figure 4.31 and 4.32 showed profound alterations on the distribution and levels of these marks. H3K9me2 decreased along the analyzed regions with the exception of Gda intragenic region that is marked by H3K9me2/3 but devoid of PHF2 and was used as negative control while H3K9me3 increased at the surrounded 1kb locus. These results demonstrated that PHF2 is important to preserve pericentromeric heterochromatin stability by maintaining the H3K9me2 and H3K9me3 equilibrium. Moreover, our data in Figure 4.32 suggest that PHF2 might prevent heterochromatin from spreading beyond the pericentromeric repeats, and probably limits ectopic heterochromatin formation, just as we observed at promoters (Figure 4.13). Intriguingly, a similar role has been proposed for another JMJC family member in yeast called Epe1 (Ayoub et al., 2003; Isaac et al., 2007; Trewick et al., 2007). In the discussion chapter I will further focus on the changes of H3K9me2/3 and the similarities between PHF2 and the yeast protein Epe1.

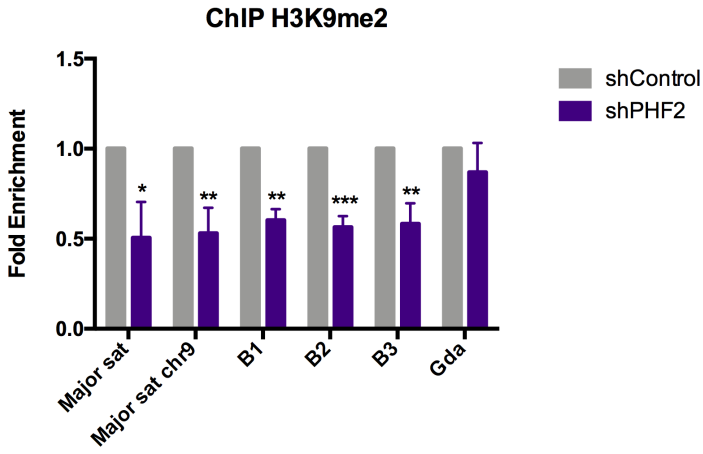


FIGURE 4.31: ChIP of H3K9me2 in shControl and shPHF2 cells was analyzed by qPCR at the major satellites, the specific major satellite of chromosome 9 and the nearby region (B1-B3). Data from qPCR were normalized to the input and expressed as fold enrichment over the data obtained in shControl. Results are the mean of three biologically independent experiments. Errors bars represent SEM. *P<0.1; **P<0.05; ***P<0.01 (Student's t-test)

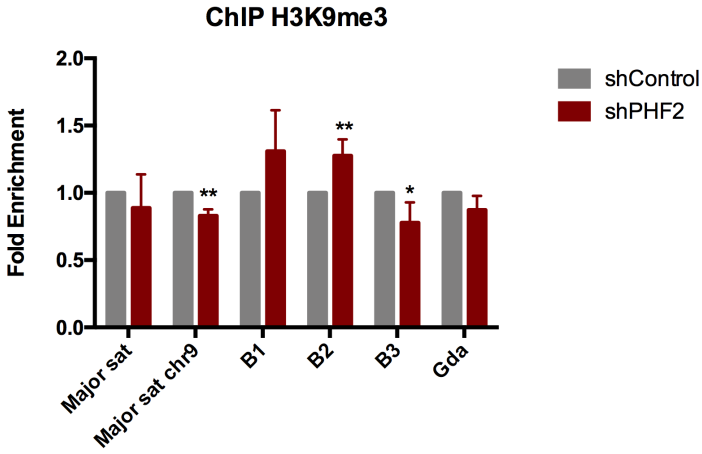


FIGURE 4.32: ChIP of H3K9me3 in shControl and shPHF2 cells was analyzed by qPCR at the major satellites, the specific major satellite of chromosome 9 and the nearby region (B1-B3). Data from qPCR were normalized to the input and expressed as fold enrichment over the data obtained in shControl. Errors bars represent SEM. *P<0.1; **P<0.05 (Student's t-test)

4.1.5 PHF2 depletion leads to DNA damage and genome instability

4.1.5.1 Loss of PHF2 leads to γH2Ax accumulation

As PHF2 depletion led to aberrant pericentromeric heterochromatin transcription, we tested whether this fact induced DNA damage that we measured by the γH2Ax content. Upon PHF2 KD, a clear accumulation of γH2Ax was detected (Figure 4.33, panel II). This increase was not observed in cells transduced with the shControl RNA (Figure 4.33, panel I). Interestingly, the observed DNA damage was rescued by overexpression of the PHF2 WT (Figure 4.33, panel III) but not by overexpression of the catalytic mutant that acts as a dominant negative (HID>AIA) (Figure 4.33, panel IV); indeed, in the latter case, an apparent increase in γH2Ax reactivity was detected, suggesting an important role of the enzymatic activity of PHF2 preventing damage.

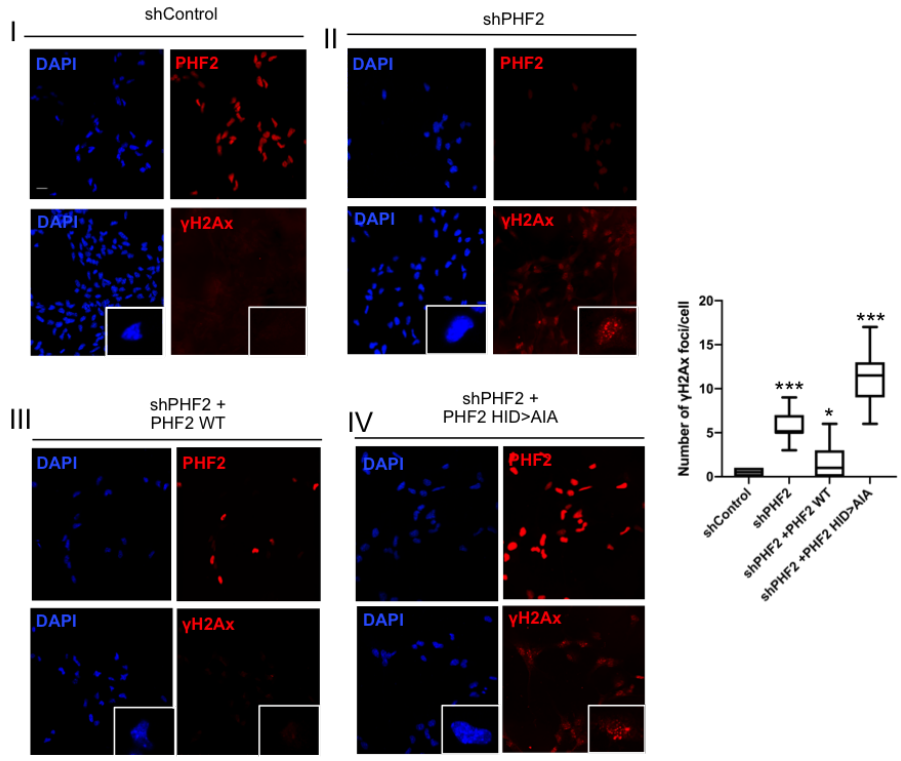


FIGURE 4.33: Immunostaining of NSCs expressing shControl (I), shPHF2 (II), shPHF2 together with PHF2 WT (III) and shPHF2 together with PHF2 (HID>AIA) (IV). Cells were fixed and stained with PHF2 and γH2Ax antibodies and DAPI. Enlarged images are showing individual cells. More than 30 cells were quantified. Data shown are representative of 2-3 biologically independent experiments. Scale bar indicates 20μm. Boxplots represent the number of γH2Ax foci/cell. *P < 0.05; ***P < 0.001 (Student's t-test)

4.1.5.2 PHF2 depletion together with overexpression of PHF2 catalytic mutant causes R-loops accumulation

As R-loops accumulation, particularly in heterochromatin regions, is usually associated with DNA damage we wanted to investigate whether γH2Ax

reactivity observed in PHF2-depleted cells was related to R-loops accumulation. For this purpose, we performed immunofluorescence experiments in control, PHF2-depleted, and PHF2-depleted that overexpressed the PHF2 (HID>AIA) mutant NSCs. In the last, the γH2Ax signal was higher as seen in Figure 4.33 (panel IV). For this purpose we have used the S9.6 antibody that specifically recognizes RNA-DNA hybrids. The data in Figure 4.34 showed a clear increase in S9.6 reactivity that colocalized with the γH2Ax positive regions. Interestingly, γH2Ax foci partially overlapped or were very close to dense DAPI staining regions (Figure 4.34), arguing that DNA damage provoked by the lack of PHF2's activity preferentially accumulates in heterochromatin.

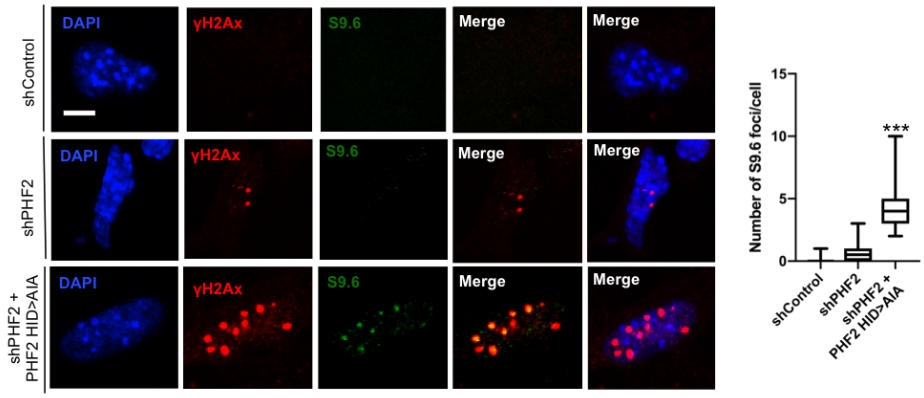


FIGURE 4.34: Immunostaining of a representative enlarged cell expressing shControl, shPHF2 and shPHF2 together with PHF2 (HID>AIA). Cells were fixed and stained with γH2Ax and S9.6 antibodies and DAPI. More than 30 cells were quantified. Data shown are representative of 2-3 biologically independent experiments. Scale bar indicates 5μm. Boxplots represent the number of S9.6 foci/cell. ***P < 0.001 (Student's t-test)

4.1.5.3 Loss of PHF2 induces the formation of multi-nucleated cells

DNA damage is often linked with genome instability; accordingly to that, we noticed an increased frequency of chromosome segregation defects, in particular, anaphase chromatin bridges and multinuclear cells in PHF2-depleted compared with control NSCs (Figure 4.35). Altogether these data

demonstrated that deficiency of PHF2 induced DNA damage and R-loops accumulation that led to chromosome segregation defects and ultimately genome instability.

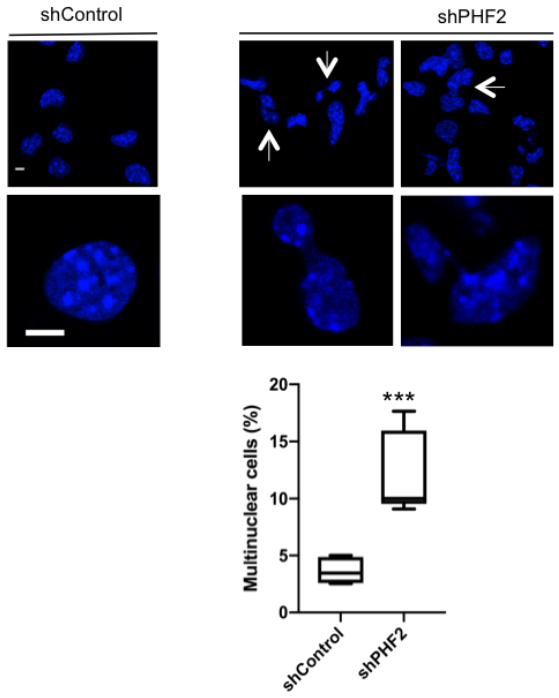


FIGURE 4.35: Formation of multi-nucleated cells and segregation defects are shown for shControl and shPHF2 cells. 50-100 cells were quantified. Data shown are representative of 4 independent experiments. Scale bar indicates 5μm. Boxplots represent the percentage of multinuclear cells. ***P < 0.001 (Student's t-test).

4.2 *In vivo* characterization of PHF2 function

The results described in the section above support the idea that PHF2 activates genes essential for neural progenitor proliferation in NSCs. In this section we will test the notion that PHF2 is essential for progenitor proliferation *in vivo* using as a model the neural tube of the chick embryo.

4.2.1 Impact of PHF2 depletion in the chick embryo neural tube

4.2.1.1 PHF2 KD in DF1 chicken cells

First of all, in order to analyze the function of PHF2 in early neurogenesis, we cloned two shRNAs for chick PHF2 (cPHF2) in a bicistronic vector, and tested their ability to deplete cPHF2. To do that, we transiently transfected DF1 chicken cells and analyzed by qPCR the levels of cPHF2. In the Figure 4.36, we observed a reduction of PHF2 upon shPHF2 expression without affecting the mRNA levels of EZH1, used as a control.

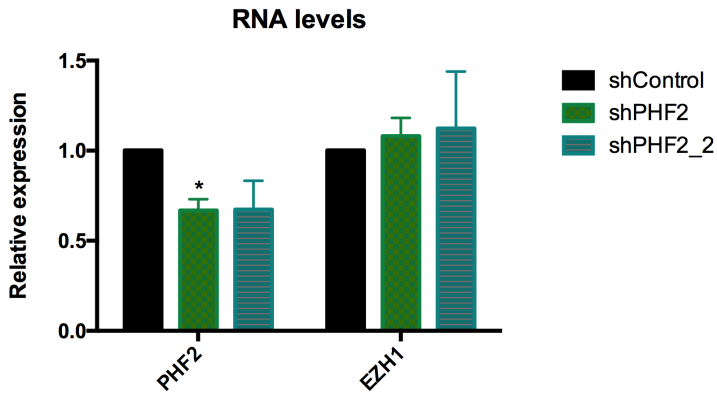


FIGURE 4.36: DF1 chicken cells were transfected with shControl and two different shRNAs for the chick PHF2 (shcPHF2, shcPHF2_2). 60-70% of DF1 cells were transfected. Total RNA was purified and the mRNA levels of PHF2 were analyzed by qPCR. EZH1 gene was used as a negative control. Errors bars represent SD. *P<0.05 (Student's t-test).

4.2.1.2 PHF2 depletion in the chick embryo neural tube

After confirming the efficiency of both shRNAs, we knocked down the expression of PHF2 in Hamburger and Hamilton (HH) stage 11-12 in chick embryo neural tubes by *in ovo* electroporation (Figure 3.2) of the shRNAs for cPHF2 or a control shRNA (shControl). Remarkably, cPHF2 KD resulted in a neural tube reduced in size (Figure 4.37). The observed phenotype was stronger using the shcPHF2, compared with the phenotype

obtaining by shcPHF2_2. Nonetheless, it is important to mention that despite this difference, both shRNAs gave the same phenotype. For the rest of the experiments performed, we chose to use the shcPHF2.

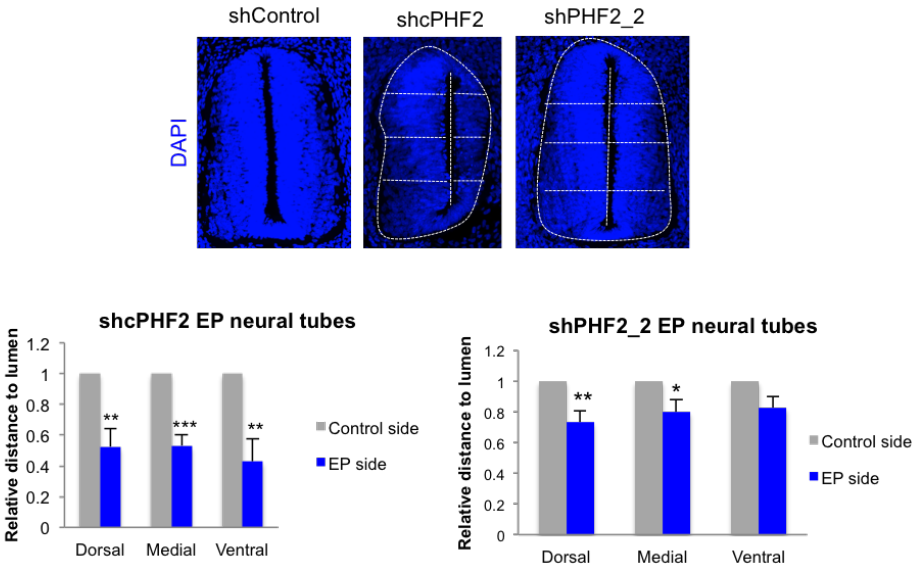


FIGURE 4.37: HH11–12 embryos were electroporated with shRNA Control (shControl) or shRNA for cPHF2 (shcPHF2) cloned into pSUPER vector and GFP-expressing vector. Transversal sections of electroporated neural tubes stained with DAPI 48 hours post-electroporation (PE). Graphs show the quantification of the size of the control side and shcPHF2/shcPHF2_2 electroporated side. To do so, we measured the dorsal, medial and ventral distances to the lumen on each side, relative to the length of the central line of the lumen. Data represent the mean of four to five embryos (from two to four biologically independent experiments). Error bars indicate SD. *P<0.05; **P<0.01; ***P<0.001 (Student's t-test)

4.2.1.3 Impact of PHF2 loss in neural progenitor proliferation

In the previous section, as our results indicated that PHF2 is essential for neural progenitor proliferation *in vitro*, we hypothesized that cPHF2 plays an active role in maintaining neural progenitors proliferating. To explore

this idea, HH11-12 neural tubes were electroporated with shcPHF2 or shControl and the effect on neural progenitor proliferation was analyzed after 48 hours. We evaluated neural progenitors entry into mitosis by analyzing the presence of H3S10p. Neural tubes electroporated with shcPHF2 showed a reduction in H3S10p (PH3)-positive mitotic cells (ratio of PH3+ cells EP side/control side: shControl 106,5±36,7, shcPHF2 65,3±21; P<0.05) (Figure 4.38). Accordingly, reduction of the neural progenitor marker SOX2 (ratio of SOX2+cells EP site/control site: shControl 101,6±10,3, shcPHF2 67,3±14; P<0.05) (Figure 4.38) indicates that cPHF2 is required for neural progenitor self-renewal in the neural tube.

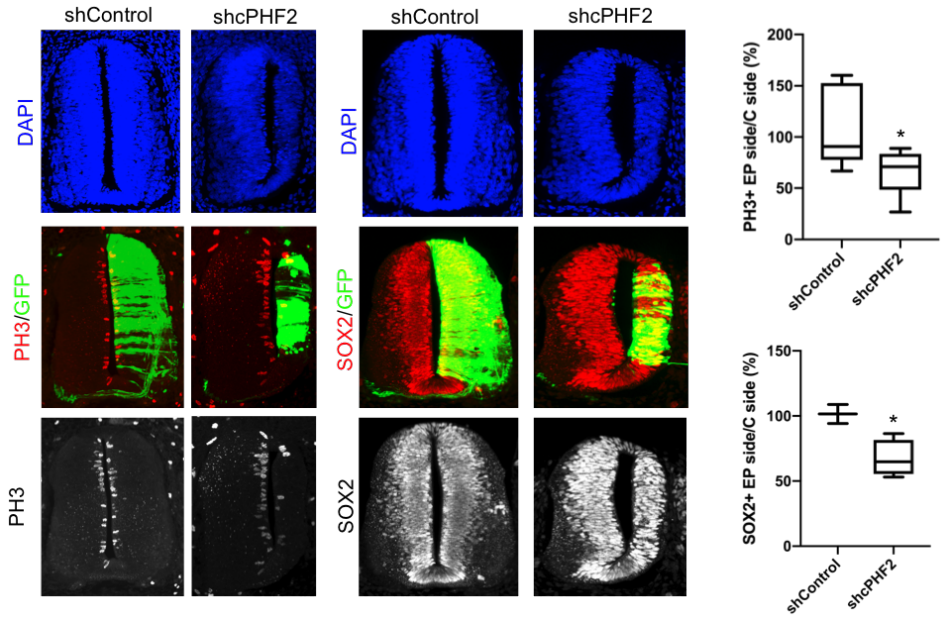


FIGURE 4.38: Transversal sections of neural tubes from HH11–12 embryos electroporated *in ovo* with shControl or shcPHF2 and stained for H3S10P (PH3) and Sox2 48 hours PE. Boxplots are showing the quantification of the corresponding immunostaining. Data represent the mean of four to twelve embryos (from three to four biologically independent experiments). *P<0.05 (Student's t-test)

4.2.1.4 Effect of PHF2 depletion in neural progenitor differentiation

We next explored the possibility that the inhibition of proliferation observed upon cPHF2 reduction would also correspond with a premature differentiation of neural progenitors. Neural tubes electroporated *in ovo* with shcPHF2 and stained for TUJ1, a neuronal differentiation marker, neither showed premature differentiation nor ectopic localization (Figure 4.39). Only a clear reduction of TUJ1 positive cells was observed, probably reflecting the decrease on the progenitor population.

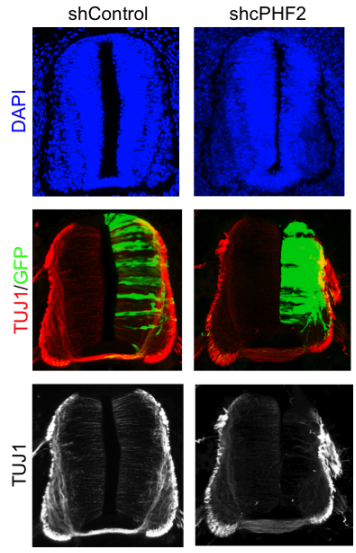


FIGURE 4.39: Transversal sections of neural tubes from HH11–12 embryos electroporated *in ovo* with shControl or shcPHF2 and stained for TUJ1, neuronal marker 48 hours PE. Figure shows a representative image of four to twelve embryos (from three to four biologically independent experiments)

4.2.1.5 Impact of PHF2 loss in neural subpopulations

Moreover, we investigated whether the neural progenitor proliferation impairment affected in a similar way the neural subpopulations along the dorsal-ventral axis. To do so, we depleted cPHF2 and analyzed different populations by using different neural markers: MNR2 for ventral neurons

(motoneurons) and Pax6 for dorsal progenitors. The results in Figure 4.40 showed that PHF2 depletion impairs the generation of all analyzed cell types. Overall, these results point to an essential role of cPHF2 promoting early neurogenesis by controlling progenitors proliferation.

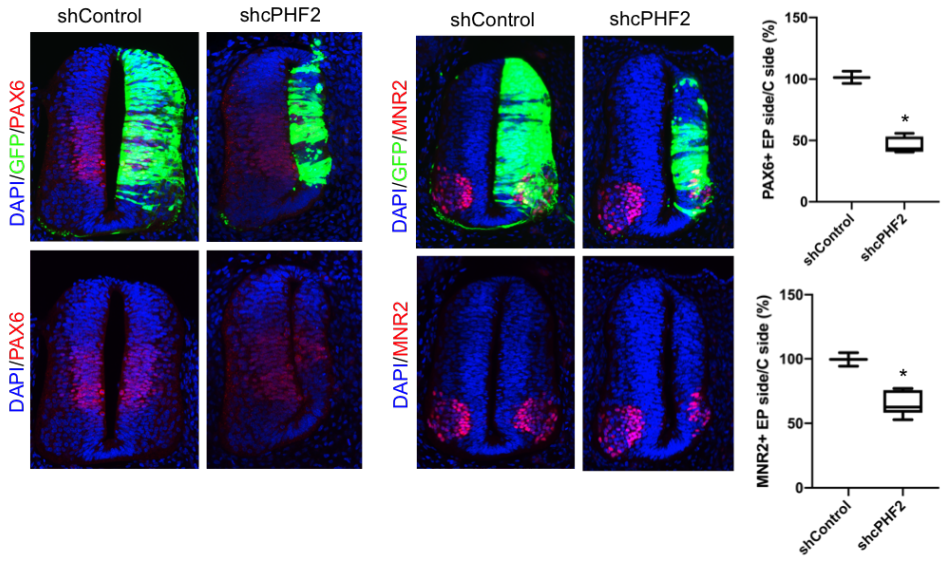


FIGURE 4.40: Transversal sections of neural tubes from HH11–12 embryos electroporated *in ovo* with shControl or shcPHF2 and stained for PAX6 and MNR2 in 48 hours PE. Boxplots are showing the quantification of the corresponding immunostaining. *P<0.05 (Student's t-test).

4.2.1.6 Implication of PHF2 loss in cell cycle distribution

In addition, we wanted to test if similar to PHF2-depleted NSCs, the electroporated cells in the chick embryo neural tube showed an alteration in the phases of the cell cycle. For this reason we performed flow-cytometry analysis of the GFP-positive cells from neural tubes electroporated with shControl and shcPHF2. The results in the Figure 4.41 showed a similar distribution of the cell cycle in Control and PHF2-depleted cells. In conclusion, we did not observe a delay in G1/S transition in disagreement with the PHF2-depleted NSCs.

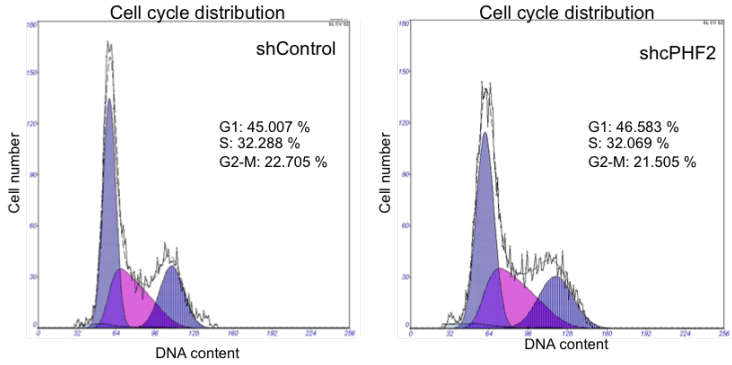


FIGURE 4.41: Flow-cytometry analysis of GFP-positive cells of neural tubes electroporated with shControl and shcPHF2. The percentage of cells in each phase of the cell cycle is depicted.

4.2.1.7 Effect of PHF2 depletion in H3K9me2 levels

Furthermore, we sought to test the global levels of H3K9me2 upon depletion of PHF2 in the chick embryo neural tube. As seen in the Figure 4.42 we did not observe any difference in the H3K9me2 levels in the electroporated cells compared with the non-electroporated ones.

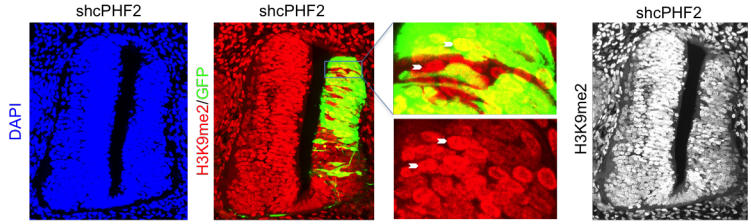


FIGURE 4.42: Transversal sections of neural tubes from HH11–12 embryos electroporated *in ovo* with shcPHF2 and stained for H3K9me2 in 48 hours PE. Amplification of a region is depicted where one electroporated and one non-electroporated cell is indicated with white arrows.

Chapter 5

Discussion

In this chapter of my doctoral thesis, I would like to discuss the obtained results, taking into account the current bibliography.

5.1 Regarding the *in vitro* characterization of PHF2 function in NSCs

5.1.1 PHF2 binds promoters decorated with H3K4me2/3 marks

Beginning this thesis, we wanted to describe the genome-wide occupancy of the histone demethylase PHF2. Performing ChIP-seq experiments, revealed us that almost all the peaks were localized on gene promoters. This result is in agreement with another study of PHF2 in primary cultured hepatocytes. They described by ChIP-seq as well, that 79% of PHF2 peaks were predominantly localized and covered the transcription start site on its target gene promoters (Bricambert et al., 2018) and 13% of the peaks were occupying introns. The percentage of PHF2 peaks in TSS, in this study, is a bit lower than the one we obtained in NSCs considering that it was performed in different cell lines and under different conditions. At this point, I would like to point out that the number of PHF2 peaks obtained by our ChIP-seq in NSCs were 5992, whereas the number obtained in primary hepatocytes were 11822 (Bricambert et al., 2018). This difference that may reflect the way of ChIP-seq analysis, could also explain the slight difference between the percentage of the peaks occupying promoters and introns between these two experiments.

Despite PHF2 localization, we found a great overlapping with H3K4me2/3 histone marks. Within actively transcribed genes, the methylation mark

H3K4me3, is mainly found at the promoter area or around the transcription start site (Bernstein et al., 2006), so a great overlapping between H3K4me3 histone mark and PHF2 was highly expected. In addition, biochemical studies had demonstrated the interaction of this histone demethylase with H3K4me2/3 thought its PHD domain (Wen et al., 2010) and finally in primary cultured hepatocytes was shown that 90% of PHF2-bound promoters were marked with H3K4me3 (Bricambert et al., 2018). All these data are in agreement with our observation that 99.7% and 99.8% of PHF2-bound regions contained H3K4me3 or H3K4me2 respectively.

5.1.2 PHF2 targets cell cycle genes and overlaps with E2F transcription factors

The GO terms of Biological Processes of PHF2 peaks showed that among the most significantly enriched categories, PHF2 binds many cell cycle genes and genes involved in DNA metabolic process. These data are in accordance with previous studies in primary cultured hepatocytes (Bricambert et al., 2018), were they identified by GO analysis that PHF2 binds promoters of genes implicated in metabolic process (2.12e–88), transcription (5.47e–65) and cell cycle (2.28e–58) among the top three most enriched categories.

E2F transcription factors are essential to control cell cycle progression. Our data demonstrated that PHF2 is occupying promoters of genes that are targets of E2F1 and E2F4 transcription factors. Interestingly, previous studies have shown that another member of the KDM7 family, PHF8, was also showed to occupy E2F1 target genes (Liu et al., 2010). Furthermore, E2F1 ChIP-seq analysis in HeLa cells indicated that more than 79% of E2F1-bound promoters corresponded to those binding by PHF8, suggesting a role of PHF8 in regulating E2F1 target genes.

At this point I would like to clarify that upon the absence of ChIP-seq data for E2F transcription factors in NSCs, in order to perform overlapping analysis, we have used previously published data from HeLa S3. However, we need to take into account that the composition of E2F complexes varies between cell types thus, it is likely that E2F activity also varies between different types of cells. Indeed, they have previously demonstrated that E2F sites in synthetic promoters can act as either positive or negative elements in a cell type-dependent fashion (Weintraub et al., 1995). For this reason,

comparison of ChIP-seq peaks between PHF2 and E2F transcription factors in NSCs may show a slight difference from the analysis performed in this thesis.

5.1.3 PHF2 regulates cell cycle gene expression and G1/S transition

PHF2 binds to cell cycle gene promoters and facilitates their transcription, as discovered by comparing the ChIP-seq with the RNA-seq data, upon depletion of PHF2. Interestingly, we observed that the proportion of downregulated transcripts (68%) was higher than the upregulated ones (32%) independently of the applying fold change in the RNA-seq. These data are in accordance with the activator role already proposed for PHF2 (Lee et al., 2014). In addition, among the differentially expressed genes we found genes that were E2F target genes (Ccnd1, Cdc25a, Pcna, Mcm3/4/6/8, Smc4, including E2f family). Moreover, by flow-cytometry, we have shown that PHF2 regulates G1/S transition. More specifically, upon PHF2 depletion the percentage of cells that remain in G1 phase of the cell cycle is higher than the control cells, in agreement with previous studies that have shown that the absence of PHF8 also impaired G1–S transition in conjunction with E2F1 (Liu et al., 2010), as we observed for PHF2. Altogether these data strongly support the idea that PHF2 is a new regulator of G1/S transition.

By performing experiments using HeLa cells as a model for cell cycle synchronization, we observed that the expression of PHF2 reached higher levels during G1/S transition. This result led us to think that PHF2 may play a role regulating cell cycle genes that show a peak of their expression in G1/S transition. We investigated this possibility by checking the binding of PHF2 at the ORC1 gene promoter in cells that were in G1 phase and in S phase. We observed that the binding of PHF2 at the ORC1 promoter region increases in cells that are in the S phase of cell cycle. However, cells in G1 phase showed PHF2 binding at the ORC1 promoter. This result could be due to the fact of bad synchronization. Having as a background the cells that are not in the same phase of the cell cycle as the rest of synchronized cells, could explain this result. Another possibility could be that PHF2 binds ORC1 promoter at G1 phase keeping it ready for posterior activation in S phase. Taking into consideration the fact that we have only tested

one gene promoter, further experiments are needed to demonstrate that actually PHF2 binds on these promoters during the S phase of cell cycle, thus regulating their expression.

5.1.4 PHF2 demethylates H3K9me2

At this part of the discussion, I would like to focus on the demethylation substrate of PHF2. As detailed in the introduction of this thesis, PHF2 has been shown in many studies to possess demethylase activity towards H3K9me2 (Wen et al., 2010; Lee et al., 2015; Baba et al., 2011). However, our data showed that apart from H3K9me2, a clear exclusion between PHF2 and H3K9me3 peaks was also observed. We identified only 1 and 52 regions of PHF2 that also contained H3K9me2 and H3K9me3, respectively. We concluded that the obtaining results could indicate a possible role of PHF2 occupying genes lacking repressive histone marks. These data are in agreement with the data published by another group claimed that genome-wide occupancy of H3K9me2, H3K9me3, and PHF2 shows an inverse correlation between the presence of the histone marks and the demethylase (Pattabiraman et al., 2016). They suggested that PHF2 plays a role not only as a H3K9me2 demethylase but as a H3K9me3 demethylase, as well.

The statement of their conclusion and more specifically that PHF2 can possess demethylase activity towards H3K9me3, could be in agreement with our immunofluorescence results upon PHF2 depletion, where we observed a slight increase in the global levels of H3K9me2 in accordance with previous studies (Wen et al., 2010) but surprisingly a greater increase in the global levels of H3K9me3. It is important to mention that H3K9me2 mark serves as a substrate for the histone methyltransferase SUV39H. For this reason if PHF2 demethylates H3K9me2, upon its depletion we would expect a clear increase in the global levels of H3K9me2, however, if the methyltransferase SUV39H uses the excess of H3K9me2 as a substrate to convert it to H3K9me3, the final effect we would observe is the great increase of the H3K9me3 mark. Another interesting point is that since the active JmjC centers of the KDM7 family of demethylases cannot accommodate trimethylated lysines, these do not constitute viable substrates for these enzymes (Fortschegger and Shiekhattar, 2011). Intriguingly, JmjCs that exhibit aspartate as Fe2+ chelating residue seem to be specific for me2/1, while glutamate at this position confers specificity for me3/2.

Taking all these data into consideration, the global increase in the H3K9me3 mark could be either explained by the role of PHF2 as a H3K9me3 demethylase, were upon its depletion a clear accumulation of this PTM is observed or by its role as a H3K9me2 demethylase where SUV39H could use the produced H3K9me2 in order to generate more H3K9me3. For biochemical reasons, as seen in Figure 5.1, the scenario that PHF2 could act as a H3K9me3 demethylase seems quite unexpected but further work is needed to figure out exactly how the depletion of PHF2 leads to such a great increase of H3K9me3 mark.

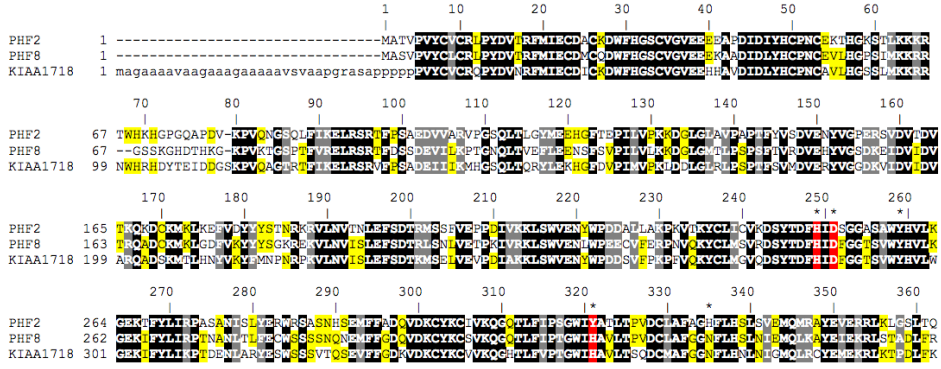


FIGURE 5.1: Schematic representation of the sequence alignment of the KDM7 family members, including PHF2 is indicated. Three Fe(II) binding residues (HxD/E ... Y) are highlighted in red background. Adapted from (Horton et al., 2011)

5.1.5 PHF2 maintains pericentromeric heterochromatin integrity

Based on the data presented in the results chapter, we proposed that the crucial role of H3K9 methylation is to suppress transcription on a global level but also to stabilize repetitive sequences, particularly at pericentromeric region. Accordingly, PHF2-depleted cells showed high H3K9me3 levels associated to transcriptional repression. Intriguingly, at heterochromatin we observed an increased transcription that correlated with altered H3K9me2/3 status and spreading into the surrounding region. These results suggest that fine-tuning H3K9 methylation is essential to control transcription at euchromatin but also to stabilize repetitive sequences in eukaryotic genomes.

Interestingly, the homologous JmjC domain protein Epe1 prevents heterochromatin instability in yeast (Trewick et al., 2007; Zofall and Grewal, 2006). Similar to our results, Epe1 is required for heterochromatin integrity since its loss impaired silencing at centromeres. Apparently contrary to this, Epe1 restricts heterochromatin domain (Trewick et al., 2007; Zofall and Grewal, 2006), as we also observed. How does PHF2 depletion cause these intuitively opposite effects? We have demonstrated that PHF2 binds satellite heterochromatin and prevents repeat transcription in a catalytic dependent manner. After PHF2 depletion we observed an erratic behaviour of heterochromatin with decreased H3K9me2 levels, maintenance of H3K9me3 in the pericentromeric repeats and spreading of H3K9me3 beyond the heterochromatin domains and overall, unprogrammed transcription of pericentromeric repeats. These might be due to a deficit in heterochromatin regulation. These results demonstrate that in the absence of PHF2, pericentromeric heterochromatin crumples, leading to aberrant repeat transcription (Figure 4.23), and at the same time heterochromatin is able to expand beyond the major satellite repeats (Figure 4.32). These data suggest that PHF2 provides stability by preventing the normal and dynamic assemble and disassemble of heterochromatin. Similar mechanism has been proposed for Epe1 protein (Trewick et al., 2007). Interestingly, unprogrammed transcription from the heterochromatic repeats has been reported in Caenorhabditis elegans embryos which lack all H3K9 methylation (Zeller and Gasser, 2017). Moreover, alteration of H3K4me3 levels at the heterochromatic region by inactivation of Jarid1C gene has been shown to lead to the aberrant expression of heterochromatin generating genomic instability (Rondinelli et al., 2015).

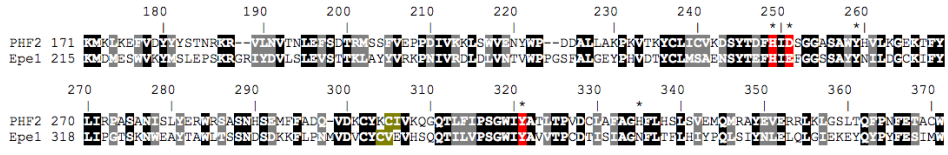


FIGURE 5.2: Schematic representation of the sequence alignment of PHF2 and Epe1 protein in yeast. Three Fe(II) binding residues (HxD/E ... Y) are highlighted in red background. Adapted from (Horton et al., 2011)

5.1.6 PHF2 binding in pericentromeric heterochromatin

Another very intriguing question would be how PHF2 is targeted to the centromeric and pericentromeric region. As detailed in the introduction chapter, the centromeric region is transcribed into minor satellite RNA whereas the pericentromeric is transcribed into major satellite RNA. The major satellite RNA is required to recruit heterochromatin protein 1 (HP1). PHC is characterized by HP1 and SUV39H1, the histone N-methyltransferase that is responsible for establishing H3K9me3 and is a main component of the PHC (Muller and Almouzni, 2017). As seen by ChIP-qPCR experiments, PHF2 showed enrichment mainly in major satellite repeats but in minor satellite repeats as well (Figure 4.27). How PHF2 is targeted in the centromeric and pericentromeric region? Some studies revealed that CENP-A domains show chromatin marks typically associated with transcriptionally active regions, such as H3K4me2 or H3K36me2 (Sullivan and Karpen, 2004). As the PHD domain of PHF2 has been demonstrated to interact with H3K4me2/3 (Wen et al., 2010), this interaction could be possible on the centromeric region, as well as on the promoters of the genes. Further investigation of this hypothesis should be performed.

Apart from the centromeric region that shows similar features with the open chromatin, we tried to answer the question of how PHF2 is targeted in the major satellites in the PHC. For this reason we performed MS in order to identify possible interactors, proteins that are mainly localized in the heterochromatin. Among other proteins, we have highlighted the appearance of HP1BP3 in the MS, where we tested the interaction between PHF2 and HP1BP3 using overexpressed proteins. HP1BP3 has been demonstrated to resemble similar domains with H1. It contains functional motifs not found in H1 histones, including an acidic stretch and a consensus HP1-binding motif (Garfinkel et al., 2015). In conclusion, PHF2 through the interaction with HP1BP3 and HP1 could be targeted to the PHC. However, another interaction between PHF2 and SUV39H1 was also described in this thesis, using again overexpressed proteins. Additionally, this interaction has been previously described (Shi et al., 2014). Taking all these data into account, future work is needed to clarify the proteins that play an important role on PHF2 binding and stabilization in the PHC.

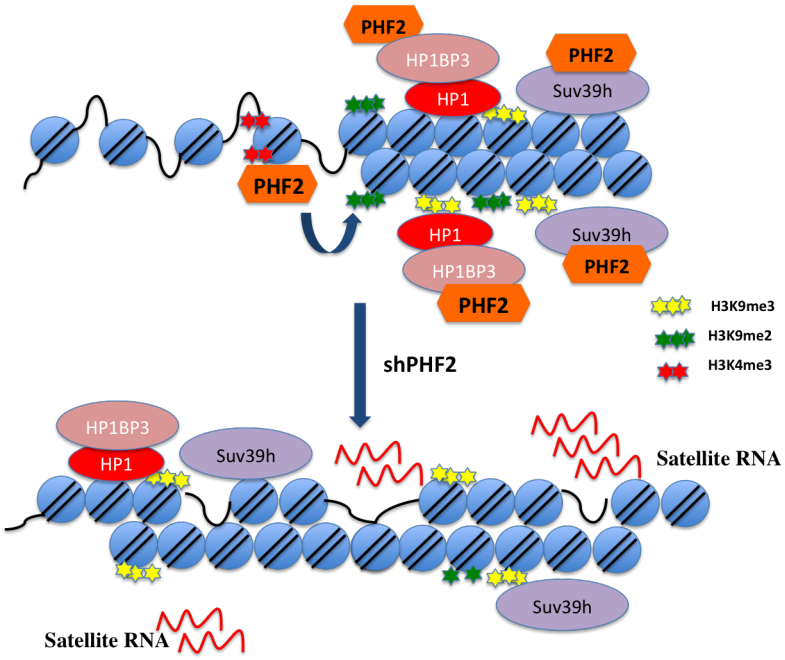


FIGURE 5.3: Schematic representation of pericentromeric region in Control and PHF2-depleted cells, where PHF2 is depicted together with the main components of the PHC and the methylation marks.

5.1.7 PHF2 depletion leads to DNA damage and genome instability

Furthermore, the results of this thesis indicate that expression of satellites repeats due to PHF2 depletion was associated with the accumulation of RNA:DNA hybrids, DNA damage and genome instability. The link between R-loops and DNA damage has been extensively studied and well established during the last years (Sollier and Cimprich, 2015). One of the major causes responsible for R-loops accumulation is the replication fork collapse or stall (Deshpande and Newlon, 1996; Gan et al., 2011; Helmrich,

Ballarino, and Tora, 2011). Thus, the unprogrammed transcription at repeats might lead to unscheduled collisions of the replication and transcription machineries, inducing DNA breaks (Bayona-Feliu et al., 2017; Deshpande and Newlon, 1996). Another possibility to explain the observed accumulation of RNA:DNA hybrids is the increased amount of aberrant transcripts due to PHF2 depletion that cannot be efficiently cotranscriptionally processed, leading to R-loops formations, as it has been previously proposed (Dominguez-Sanchez et al., 2011). On the other hand, it has been proposed that the demethylation of H3K9 may be an important step in the repair of double-strand breaks. Thus, the global increase in H3K9me3 observed in PHF2-depleted progenitors (Figure 4.12) could contribute to delay or impair DNA repair, particularly at heterochromatin, as it has been described for KDM4B demethylase (Colmenares et al., 2017; Young, McDonald, and Hendzel, 2013; Zheng et al., 2014).

Finally, it might be also possible that expression of factors involved in R-loops resolution and DNA damage repair were impaired in PHF2-depleted cells. In fact, we have observed a decrease of ATM/ATR and BRCA1 factors in the RNA-seq experiment that could contribute to the accumulation of double-strand breaks and genome instability in PHF2-depleted cells. Then, the accumulation of the observed DNA damage in PHF2-depleted neural progenitors might be the result of both, in one hand the increase on DNA damage, mainly due to repeats transcription and in another hand the deficiency of the DNA damage to be repaired due to high levels of H3K9me3 and low activity of enzymes involved in repair. Taking all the above into account, more experimental work is needed to demonstrate these hypotheses.

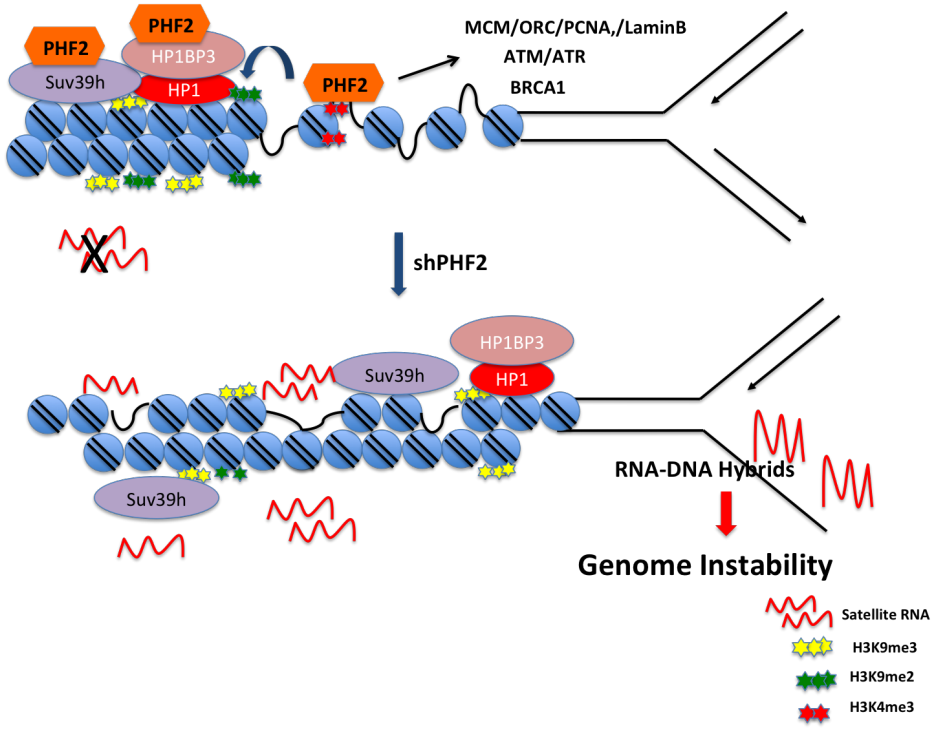


FIGURE 5.4: Illustration of the of pericentromeric region in Control and PHF2-depleted cells, where the effect of PHF2 depletion in genome instability is presented.

5.2 Regarding the *in vivo* characterization of PHF2 function

In the above section of this thesis, I discussed about the binding of PHF2 to cell cycle gene promoters and how this facilitates their transcription. Apart from that, we have showed that PHF2 is essential to maintain the heterochromatin integrity. Together, these functions ensure the generation of progenitors that guarantees proper neural development.

5.2.1 PHF2 depletion affects neural progenitor proliferation

Using as *in vivo* model of early neurogenesis the chick embryo neural tube, we demonstrated that PHF2 is required for neural progenitor self renewal. Depletion of PHF2 led to a reduction of PH3-positive mitotic cells without enhancing the differentiation process in the neural tube. There are many remaining open questions in this context. First of all, by DAPI staining, we did not observe cell death in the neural tubes lacking PHF2. Furthermore, these cells did not show any alteration in the phases of the cell cycle, neither showed a peak of apoptosis in the cell cycle distribution profile (Figure 4.41). As mentioned in the introduction, tight control of the balance between self-expanding symmetric and self-renewing asymmetric neural progenitor divisions is crucial to regulate the number of cells in the developing neural tube. An interesting question would be if PHF2 controls the balance between these types of divisions. Further experiments will be needed to test this hypothesis.

5.3 Regarding the contribution of PHF2 in diseases

In the introduction chapter was mentioned that PHF2 has been related to ASD (Cotney et al., 2015; Iossifov et al., 2012). However, the function of this protein in neural development is still poorly understood. Phf2 KD mice showed partial neonatal death, growth retardation, and reduced body weight. Interestingly, the brain weights of Phf2 KD mice were larger than wild-type littermates (Okuno et al., 2013). Macrocephaly is a common phenotype associated to ASD. Intriguingly, we observed that Phf2 KD in cortical progenitors and in the neural tube results in defective neural progenitor proliferation and differentiation. Similar paradox has been described for another chromatin factor strongly associated to ASD, the remodeling factor CHD8 (Bernier et al., 2014; De Rubeis et al., 2014; O'Roak et al., 2012). CHD8 is essential in cortical development by promoting neural progenitors proliferation and neurogenesis (Breuss and Gleeson, 2016; De Rubeis et al., 2014; Katayama et al., 2016). Although, at this moment the explanation to this phenomenon is not possible, but it would be interesting to elucidate the role of PHF2 in development of other cell types such as astrocytes. Alteration of PHF2 activity could affect the size or properties of these cells, leading to the macrocephaly observed in patients with ASD.

These observations suggest that PHF2 maintains neural progenitor proliferation through regulation of cell cycle genes and pericentromeric heterochromatin integrity. Furthermore, these results raise the possibility that some of the ASD associated phenotypes in patients carrying PHF2 mutations may be caused by defects during early neural development.

Finally our work paves the way for investigating the contribution of H3K9 methylation to genomic stability and transcriptional regulation in other cellular contexts. In particular, PHF2 has been widely involved in cancer. Several studies use H3K9me modifiers enzymes as targets in cancer treatment (Kondo et al., 2008; Wagner and Jung, 2012). Moreover, alteration of the demethylase activity of PHF2 has been suggested as a new target to treat disorders linked to diet-induced obesity, due to its essential role regulating adipogenesis (Okuno et al., 2013). Our results from PHF2-depleted cells, as well as published data, removing H3K9me2/3 histone methyltransferases (HMTs) (Zeller et al., 2016), indicate that alteration or removal of H3K9 methylation might not be a suitable therapeutic strategy. In both cases genomic instability might be a drawback in these treatments. In fact, PHF2 act as a tumor suppressor in association with p53 (Lee et al., 2015) and its depletion has been recurrently found in different cancer types (Ghosh et al., 2013; Lee et al., 2017; Pattabiraman et al., 2016; Lee et al., 2015).

5.4 General view of the results

During neural development, the transcriptional programs orchestrated by the signaling pathways, TFs and epigenetic regulators lead to phenotypical changes that provide NSCs with neuronal features (Temple, 2001). The chromatin acting factors are essential players in both, proliferation and cell differentiation events during embryo development. In this context, this doctoral thesis has contributed to reveal an unforeseen role of PHF2 during development. We demonstrate that this histone demethylase binds to cell cycle gene promoters facilitating their transcription and suppresses unprogrammed transcription from pericentromeric repeats. To do so, PHF2 requires its catalytic activity. PHF2 allows neural stem cells proliferation and preserves heterochromatin integrity during progenitor expansion. PHF2

fine-tunes H3K9 methylation levels, ensuring genome stability and chromatin homeostasis. Finally, this work helps to move forward our understanding of the multiple crosstalks between epigenetics and developmental programs.

Chapter 6

Conclusions

1. PHF2 binds gene promoters related to mRNA processing, regulation of cell cycle, histone modification and DNA repair in NSCs.
2. PHF2 partially overlaps with E2F1 and E2F4-bound regions in NSCs.
3. PHF2 colocalizes with H3K4me2/3 and it is excluded from regions marked with H3K9me2/3 in NSCs.
4. PHF2 mediates H3K9 demethylation at global and promoter levels in NSCs.
5. PHF2 directly regulates the expression of 601 genes in NSCs. These are associated with cell cycle categories such as G1-S transition, DNA replication, mitosis and chromatin activity.
6. PHF2-depleted cells show impaired cell proliferation and delay in G1/S transition in NSCs.
7. PHF2 reduction leads to an increased transcription of major and minor satellite repeats in NSCs.
8. PHF2 binds major and minor satellite repeats.
9. HP1BP3 has been identified as a new interactor of PHF2.
10. PHF2 preserves pericentromeric heterochromatin stability by maintaining the H3K9me2/3 equilibrium in NSCs.
11. PHF2 lack of function leads to γH2Ax and R-loops accumulation in NSCs.

12. PHF2 is vital to maintain genome stability.
13. PHF2 is essential for neural progenitor self-renewal in the chick embryo neural tube.
14. PHF2 is important for neural subpopulation specification in the chick embryo neural tube.

Bibliography

Aagaard, L., G. Laible, P. Selenko, M. Schmid, R. Dorn, G. Schotta, S. Kuhfittig, A. Wolf, A. Lebersorger, P. B. Singh, G. Reuter, and T. Jenuwein (1999). "Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31". In: *EMBO J.* 18.7, pp. 1923–1938.

Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. ?ech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Gruning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, and J. Goecks (July 2016). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update". In: *Nucleic Acids Res.* 44.W1, W3–W10.

Aguilera, A. and T. Garcia-Muse (2012). "R loops: from transcription byproducts to threats to genome stability". In: *Mol. Cell* 46.2, pp. 115–124.

— (2013). "Causes of genome instability". In: *Annu. Rev. Genet.* 47, pp. 1–32.

Aguilera, A. and B. Gomez-Gonzalez (2008). "Genome instability: a mechanistic view of its causes and consequences". In: *Nat. Rev. Genet.* 9.3, pp. 204–217.

Alexiadis, V., M. E. Ballestas, C. Sanchez, S. Winokur, V. Vedanarayanan, M. Warren, and M. Ehrlich (2007). "RNAPol-ChIP analysis of transcription from FSHD-linked tandem repeats and satellite DNA". In: *Biochim. Biophys. Acta* 1769.1, pp. 29–40.

Allis, C. D. and T. Jenuwein (Aug. 2016). "The molecular hallmarks of epigenetic control". In: *Nat. Rev. Genet.* 17.8, pp. 487–500.

Allshire, R. C. and H. D. Madhani (2018). "Ten principles of heterochromatin formation and function". In: *Nat. Rev. Mol. Cell Biol.* 19.4, pp. 229–244.

Allshire, R. C., J. P. Javerzat, N. J. Redhead, and G. Cranston (1994). "Position effect variegation at fission yeast centromeres". In: *Cell* 76.1, pp. 157–169.

Anders, S., P. T. Pyl, and W. Huber (2015). "HTSeq–a Python framework to work with high-throughput sequencing data". In: *Bioinformatics* 31.2, pp. 166–169.

Asensio-Juan, E., C. Gallego, and M. A. Martinez-Balbas (2012). "The histone demethylase PHF8 is essential for cytoskeleton dynamics". In: *Nucleic Acids Res.* 40.19, pp. 9429–9440.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium". In: *Nat. Genet.* 25.1, pp. 25–29.

Ausio, J., F. Dong, and K. E. van Holde (1989). "Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome". In: *J. Mol. Biol.* 206.3, pp. 451–463.

Ayoub, N., K. Noma, S. Isaac, T. Kahan, S. I. Grewal, and A. Cohen (2003). "A novel jmjC domain protein modulates heterochromatization in fission yeast". In: *Mol. Cell. Biol.* 23.12, pp. 4356–4370.

Baba, A., F. Ohtake, Y. Okuno, K. Yokota, M. Okada, Y. Imai, M. Ni, C. A. Meyer, K. Igarashi, J. Kanno, M. Brown, and S. Kato (2011). "PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B". In: *Nat. Cell Biol.* 13.6, pp. 668–675.

Babicki, S., D. Arndt, A. Marcu, Y. Liang, J. R. Grant, A. Maciejewski, and D. S. Wishart (July 2016). "Heatmapper: web-enabled heat mapping for all". In: *Nucleic Acids Res.* 44.W1, W147–153.

Bachman, K. E., M. R. Rountree, and S. B. Baylin (2001). "Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin". In: *J. Biol. Chem.* 276.34, pp. 32282–32287.

Barry, D. S., J. M. Pakan, and K. W. McDermott (2014). "Radial glial cells: key organisers in CNS development". In: *Int. J. Biochem. Cell Biol.* 46, pp. 76–79.

Bayona-Feliu, A., A. Casas-Lamesa, O. Reina, J. Bernues, and F. Azorin (Aug. 2017). "Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin". In: *Nat Commun* 8.1, p. 283.

Beach, D., B. Durkacz, and P. Nurse (1982). "Functionally homologous cell cycle control genes in budding and fission yeast". In: *Nature* 300.5894, pp. 706–709.

Bernier, R., C. Golzio, B. Xiong, H. A. Stessman, B. P. Coe, O. Penn, K. Witherspoon, J. Gerdts, C. Baker, A. T. Vulto-van Silfhout, J. H. Schuurs-Hoeijmakers, M. Fichera, P. Bosco, S. Buono, A. Alberti, P. Failla, H. Peeters, J. Steyaert, L. E. L. M. Vissers, L. Francescatto, H. C. Mefford, J. A. Rosenfeld, T. Bakken, B. J. O'Roak, M. Pawlus, R. Moon, J. Shendure, D. G. Amaral, E. Lein, J. Rankin, C. Romano, B. B. A. de Vries, N. Katsanis, and E. E. Eichler (2014). "Disruptive CHD8 mutations define a subtype of autism early in development". In: *Cell* 158.2, pp. 263–276.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells". In: *Cell* 125.2, pp. 315–326.

Bestor, T. H. (2000). "The DNA methyltransferases of mammals". In: *Hum. Mol. Genet.* 9.16, pp. 2395–2402.

Boguslawski, S. J., D. E. Smith, M. A. Michalak, K. E. Mickelson, C. O. Yehle, W. L. Patterson, and R. J. Carrico (1986). "Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids". In: *J. Immunol. Methods* 89.1, pp. 123–130.

Boissinot, S. and A. V. Furano (2005). "The recent evolution of human L1 retrotransposons". In: *Cytogenet. Genome Res.* 110.1-4, pp. 402–406.

Boos, D., J. Frigola, and J. F. Diffley (2012). "Activation of the replicative DNA helicase: breaking up is hard to do". In: *Curr. Opin. Cell Biol.* 24.3, pp. 423–430.

Bracken, A. P., D. Pasini, M. Capra, E. Prosperini, E. Colli, and K. Helin (2003). "EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer". In: *EMBO J.* 22.20, pp. 5323–5335.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding". In: *Anal. Biochem.* 72, pp. 248–254.

Brasher, S. V., B. O. Smith, R. H. Fogh, D. Nietlispach, A. Thiru, P. R. Nielsen, R. W. Broadhurst, L. J. Ball, N. V. Murzina, and E. D. Laue (2000). "The

structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer". In: *EMBO J.* 19.7, pp. 1587–1597.

Breuss, M. W. and J. G. Gleeson (Oct. 2016). "When size matters: CHD8 in autism". In: *Nat. Neurosci.* 19.11, pp. 1430–1432.

Bricambert, J., M. C. Alves-Guerra, P. Esteves, C. Prip-Buus, J. Bertrand-Michel, H. Guillou, C. J. Chang, M. N. Vander Wal, F. Canonne-Hergaux, P. Mathurin, V. Raverdy, F. Pattou, J. Girard, C. Postic, and R. Dentin (May 2018). "The histone demethylase Phf2 acts as a molecular checkpoint to prevent NAFLD progression during obesity". In: *Nat Commun* 9.1, p. 2092.

Brown, S. E., M. F. Fraga, I. C. Weaver, M. Berdasco, and M. Szyf (2007). "Variations in DNA methylation patterns during the cell cycle of HeLa cells". In: *Epigenetics* 2.1, pp. 54–65.

Bulut-Karslioglu, A., V. Perrera, M. Scaranaro, I. A. de la Rosa-Velazquez, S. van de Nobelen, N. Shukeir, J. Popow, B. Gerle, S. Opravil, M. Pagani, S. Meidhof, T. Brabletz, T. Manke, M. Lachner, and T. Jenuwein (2012). "A transcription factor-based mechanism for mouse heterochromatin formation". In: *Nat. Struct. Mol. Biol.* 19.10, pp. 1023–1030.

Burgold, T., F. Spreafico, F. De Santa, M. G. Totaro, E. Prosperini, G. Natoli, and G. Testa (2008). "The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment". In: *PLoS ONE* 3.8, e3034.

C Yuen, R. K., D. Merico, M. Bookman, J. L Howe, B. Thiruvahindrapuram, R. V. Patel, J. Whitney, N. Deflaux, J. Bingham, Z. Wang, G. Pellecchia, J. A. Buchanan, S. Walker, C. R. Marshall, M. Uddin, M. Zarrei, E. Deneault, L. D'Abate, A. J. Chan, S. Koyanagi, T. Paton, S. L. Pereira, N. Hoang, W. Engchuan, E. J. Higginbotham, K. Ho, S. Lamoureux, W. Li, J. R. MacDonald, T. Nalpathamkalam, W. W. Sung, F. J. Tsoi, J. Wei, L. Xu, A. M. Tasse, E. Kirby, W. Van Etten, S. Twigger, W. Roberts, I. Drmic, S. Jilderda, B. M. Modi, B. Kellam, M. Szego, C. Cytrynbaum, R. Weksberg, L. Zwaigenbaum, M. Woodbury-Smith, J. Brian, L. Senman, A. Iaboni, K. Doyle-Thomas, A. Thompson, C. Chrysler, J. Leef, T. Savion-Lemieux, I. M. Smith, X. Liu, R. Nicolson, V. Seifer, A. Fedele, E. H. Cook, S. Dager, A. Estes, L. Gallagher, B. A. Malow, J. R. Parr, S. J. Spence, J. Vorstman, B. J. Frey, J. T. Robinson, L. J. Strug, B. A. Fernandez, M. Elsabbagh, M. T. Carter, J. Hallmayer, B. M. Knoppers, E. Anagnostou, P. Szatmari, R. H. Ring, D. Glazer, M. T. Pletcher, and

S. W. Scherer (2017). "Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder". In: *Nat. Neurosci.* 20.4, pp. 602–611.

Chan, F. L., O. J. Marshall, R. Saffery, B. W. Kim, E. Earle, K. H. Choo, and L. H. Wong (2012). "Active transcription and essential role of RNA polymerase II at the centromere during mitosis". In: *Proc. Natl. Acad. Sci. U.S.A.* 109.6, pp. 1979–1984.

Chenn, A. and C. A. Walsh (2002). "Regulation of cerebral cortical size by control of cell cycle exit in neural precursors". In: *Science* 297.5580, pp. 365–369.

Chiva, C., R. Olivella, E. Borras, G. Espadas, O. Pastor, A. Sole, and E. Sabido (2018). "QCloud: A cloud-based quality control system for mass spectrometry-based proteomics laboratories". In: *PLoS ONE* 13.1, e0189209.

Collins, R. E., M. Tachibana, H. Tamaru, K. M. Smith, D. Jia, X. Zhang, E. U. Selker, Y. Shinkai, and X. Cheng (2005). "In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases". In: *J. Biol. Chem.* 280.7, pp. 5563–5570.

Colmenares, S. U., J. M. Swenson, S. A. Langley, C. Kennedy, S. V. Costes, and G. H. Karpen (July 2017). "Drosophila Histone Demethylase KDM4A Has Enzymatic and Non-enzymatic Roles in Controlling Heterochromatin Integrity". In: *Dev. Cell* 42.2, pp. 156–169.

Consortium, The Gene Ontology (Jan. 2017). "Expansion of the Gene Ontology knowledgebase and resources". In: *Nucleic Acids Res.* 45.D1, pp. D331–D338.

Copp, A. J. and N. D. Greene (2010). "Genetics and development of neural tube defects". In: *J. Pathol.* 220.2, pp. 217–230.

Copp, A. J., N. S. Adzick, L. S. Chitty, J. M. Fletcher, G. N. Holmbeck, and G. M. Shaw (Apr. 2015). "Spina bifida". In: *Nat Rev Dis Primers* 1, p. 15007.

Cotney, J., R. A. Muhle, S. J. Sanders, L. Liu, A. J. Willsey, W. Niu, W. Liu, L. Klei, J. Lei, J. Yin, S. K. Reilly, A. T. Tebbenkamp, C. Bichsel, M. Pletikos, N. Sestan, K. Roeder, M. W. State, B. Devlin, and J. P. Noonan (2015). "The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment". In: *Nat Commun* 6, p. 6404.

Cowieson, N. P., J. F. Partridge, R. C. Allshire, and P. J. McLaughlin (2000). "Dimerisation of a chromo shadow domain and distinctions from the

chromodomain as revealed by structural analysis". In: *Curr. Biol.* 10.9, pp. 517–525.

Currle, D. S. and E. S. Monuki (2007). "Flash freezing and cryosectioning E12.5 mouse brain". In: *J Vis Exp* 4, p. 198.

De Piccoli, G., Y. Katou, T. Itoh, R. Nakato, K. Shirahige, and K. Labib (2012). "Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases". In: *Mol. Cell* 45.5, pp. 696–704.

De Rubeis, S. et al. (2014). "Synaptic, transcriptional and chromatin genes disrupted in autism". In: *Nature* 515.7526, pp. 209–215.

Deshpande, A. M. and C. S. Newlon (1996). "DNA replication fork pause sites dependent on transcription". In: *Science* 272.5264, pp. 1030–1033.

Dominguez-Sanchez, M. S., S. Barroso, B. Gomez-Gonzalez, R. Luna, and A. Aguilera (2011). "Genome instability and transcription elongation impairment in human cells depleted of THO/TREX". In: *PLoS Genet.* 7.12, e1002386.

Drolet, M., P. Phoenix, R. Menzel, E. Masse, L. F. Liu, and R. J. Crouch (1995). "Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I". In: *Proc. Natl. Acad. Sci. U.S.A.* 92.8, pp. 3526–3530.

Ea, V., M. O. Baudement, A. Lesne, and T. Forne (2015). "Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization". In: *Genes (Basel)* 6.3, pp. 734–750.

Earnshaw, W. C. and N. Rothfield (1985). "Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma". In: *Chromosoma* 91.3-4, pp. 313–321.

Enukashvily, N. I., R. Donev, I. S. Waisertreiger, and O. I. Podgornaya (2007). "Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells". In: *Cytogenet. Genome Res.* 118.1, pp. 42–54.

Epsztejn-Litman, S., N. Feldman, M. Abu-Remaileh, Y. Shufaro, A. Gerson, J. Ueda, R. Deplus, F. Fuks, Y. Shinkai, H. Cedar, and Y. Bergman (2008). "De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes". In: *Nat. Struct. Mol. Biol.* 15.11, pp. 1176–1183.

Estaras, C., N. Akizu, A. Garcia, S. Beltran, X. de la Cruz, and M. A. Martinez-Balbas (2012). "Genome-wide analysis reveals that Smad3 and JMJD3 HDM co-activate the neural developmental program". In: *Development* 139.15, pp. 2681–2691.

Estaras, C., R. Fueyo, N. Akizu, S. Beltran, and M. A. Martinez-Balbas (2013). "RNA polymerase II progression through H3K27me3-enriched gene bodies requires JMJD3 histone demethylase". In: *Mol. Biol. Cell* 24.3, pp. 351–360.

Eymery, A., M. Callanan, and C. Vourc'h (2009). "The secret message of heterochromatin: new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription". In: *Int. J. Dev. Biol.* 53.2-3, pp. 259–268.

Eymery, A., B. Horard, M. El Atifi-Borel, G. Fourel, F. Berger, A. L. Vitte, A. Van den Broeck, E. Brambilla, A. Fournier, M. Callanan, S. Gazzeri, S. Khochbin, S. Rousseaux, E. Gilson, and C. Vourc'h (2009). "A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells". In: *Nucleic Acids Res.* 37.19, pp. 6340–6354.

Fan, G., K. Martinowich, M. H. Chin, F. He, S. D. Fouse, L. Hutnick, D. Hattori, W. Ge, Y. Shen, H. Wu, J. ten Hoeve, K. Shuai, and Y. E. Sun (2005). "DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling". In: *Development* 132.15, pp. 3345–3356.

Feng, W., M. Yonezawa, J. Ye, T. Jenuwein, and I. Grummt (2010). "PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation". In: *Nat. Struct. Mol. Biol.* 17.4, pp. 445–450.

Fernandez, V., C. Llinares-Benadero, and V. Borrell (May 2016). "Cerebral cortex expansion and folding: what have we learned?" In: *EMBO J.* 35.10, pp. 1021–1044.

Filion, G. J. and B. van Steensel (2010). "Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells". In: *Nat. Genet.* 42.1, pp. 5–6.

Finlay, B. L. and R. B. Darlington (1995). "Linked regularities in the development and evolution of mammalian brains". In: *Science* 268.5217, pp. 1578–1584.

Fischer, T., B. Cui, J. Dhakshnamoorthy, M. Zhou, C. Rubin, M. Zofall, T. D. Veenstra, and S. I. Grewal (2009). "Diverse roles of HP1 proteins in

heterochromatin assembly and functions in fission yeast". In: *Proc. Natl. Acad. Sci. U.S.A.* 106.22, pp. 8998–9003.

Flamm, W. G., P. M. Walker, and M. McCallum (1969). "Some properties of the single strands isolated from the DNA of the nuclear satellite of the mouse (Mus musculus)". In: *J. Mol. Biol.* 40.3, pp. 423–443.

Fortschegger, K. and R. Shiekhattar (2011). "Plant homeodomain fingers form a helping hand for transcription". In: *Epigenetics* 6.1, pp. 4–8.

Fortschegger, K., P. de Graaf, N. S. Outchkourov, F. M. van Schaik, H. T. Timmers, and R. Shiekhattar (2010). "PHF8 targets histone methylation and RNA polymerase II to activate transcription". In: *Mol. Cell. Biol.* 30.13, pp. 3286–3298.

Foster, Douglas N and Linda K Foster (1997). *Immortalized cell lines for virus growth*. US Patent 5,672,485.

French, S. (1992). "Consequences of replication fork movement through transcription units in vivo". In: *Science* 258.5086, pp. 1362–1365.

Frescas, D., D. Guardavaccaro, S. M. Kuchay, H. Kato, A. Poleshko, V. Basrur, K. S. Elenitoba-Johnson, R. A. Katz, and M. Pagano (2008). "KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state". In: *Cell Cycle* 7.22, pp. 3539–3547.

Friedberg, E. C. (2008). "A brief history of the DNA repair field". In: *Cell Res.* 18.1, pp. 3–7.

Fueyo, R., M. A. Garcia, and M. A. Martinez-Balbas (2015). "Jumonji family histone demethylases in neural development". In: *Cell Tissue Res.* 359.1, pp. 87–98.

Fueyo, R., S. Iacobucci, S. Pappa, C. Estaras, S. Lois, M. Vicioso-Mantis, C. Navarro, S. Cruz-Molina, J. C. Reyes, A. Rada-Iglesias, X. de la Cruz, and M. A. Martinez-Balbas (2018). "Lineage specific transcription factors and epigenetic regulators mediate TGFÎ²-dependent enhancer activation". In: *Nucleic Acids Res.* 46.7, pp. 3351–3365.

Fukagawa, T. and W. C. Earnshaw (2014). "The centromere: chromatin foundation for the kinetochore machinery". In: *Dev. Cell* 30.5, pp. 496–508.

Fuks, F., P. J. Hurd, D. Wolf, X. Nan, A. P. Bird, and T. Kouzarides (2003). "The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation". In: *J. Biol. Chem.* 278.6, pp. 4035–4040.

Gan, W., Z. Guan, J. Liu, T. Gui, K. Shen, J. L. Manley, and X. Li (2011). "R-loop-mediated genomic instability is caused by impairment of replication fork progression". In: *Genes Dev.* 25.19, pp. 2041–2056.

Garfinkel, B. P., N. Melamed-Book, E. Anuka, M. Bustin, and J. Orly (2015). "HP1BP3 is a novel histone H1 related protein with essential roles in viability and growth". In: *Nucleic Acids Res.* 43.4, pp. 2074–2090.

Ghosh, A., G. P. Maiti, M. N. Bandopadhyay, J. Chakraborty, J. Biswas, S. Roychoudhury, and C. K. Panda (2013). "Inactivation of 9q22.3 tumor suppressor genes predict outcome for patients with head and neck squamous cell carcinoma". In: *Anticancer Res.* 33.3, pp. 1215–1220.

Golebiewska, A., S. P. Atkinson, M. Lako, and L. Armstrong (2009). "Epigenetic landscaping during hESC differentiation to neural cells". In: *Stem Cells* 27.6, pp. 1298–1308.

Goodrich, D. W., N. P. Wang, Y. W. Qian, E. Y. Lee, and W. H. Lee (1991). "The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle". In: *Cell* 67.2, pp. 293–302.

Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn (1977). "Characteristics of a human cell line transformed by DNA from human adenovirus type 5". In: *J. Gen. Virol.* 36.1, pp. 59–74.

Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel (2008). "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions". In: *Nature* 453.7197, pp. 948–951.

Guillemot, F. (2007). "Spatial and temporal specification of neural fates by transcription factor codes". In: *Development* 134.21, pp. 3771–3780.

Hahn, M., S. Dambacher, S. Dulev, A. Y. Kuznetsova, S. Eck, S. Worz, D. Sadic, M. Schulte, J. P. Mallm, A. Maiser, P. Debs, H. von Melchner, H. Leonhardt, L. Schermelleh, K. Rohr, K. Rippe, Z. Storchova, and G. Schotta (2013). "Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin". In: *Genes Dev.* 27.8, pp. 859–872.

Hamburger, V., H. L. Hamilton, V. Hamburger, and H. L. Hamilton (1992). "A series of normal stages in the development of the chick embryo. 1951". In: *Dev. Dyn.* 195.4, pp. 231–272.

Harel, J., N. Hanania, H. Tapiero, and L. Harel (1968). "RNA replication by nuclear satellite DNA in different mouse cells". In: *Biochem. Biophys. Res. Commun.* 33.4, pp. 696–701.

Harris, M. J. and D. M. Juriloff (2007). "Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects". In: *Birth Defects Res. Part A Clin. Mol. Teratol.* 79.3, pp. 187–210.

Hasenpusch-Theil, K., B. P. Chadwick, T. Theil, S. K. Heath, D. G. Wilkinson, and A. M. Frischauf (1999). "PHF2, a novel PHD finger gene located on human chromosome 9q22". In: *Mamm. Genome* 10.3, pp. 294–298.

Hata, K., R. Takashima, K. Amano, K. Ono, M. Nakanishi, M. Yoshida, M. Wakabayashi, A. Matsuda, Y. Maeda, Y. Suzuki, S. Sugano, R. H. Whitson, R. Nishimura, and T. Yoneda (2013). "Arid5b facilitates chondrogenesis by recruiting the histone demethylase Phf2 to Sox9-regulated genes". In: *Nat Commun* 4, p. 2850.

Helmrich, A., M. Ballarino, and L. Tora (2011). "Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes". In: *Mol. Cell* 44.6, pp. 966–977.

Hendzel, M. J., Y. Wei, M. A. Mancini, A. Van Hooser, T. Ranalli, B. R. Brinkley, D. P. Bazett-Jones, and C. D. Allis (1997). "Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation". In: *Chromosoma* 106.6, pp. 348–360.

Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, and W. J. Kent (2006). "The UCSC Genome Browser Database: update 2006". In: *Nucleic Acids Res.* 34.Database issue, pp. D590–598.

Hirabayashi, Y., Y. Itoh, H. Tabata, K. Nakajima, T. Akiyama, N. Masuyama, and Y. Gotoh (2004). "The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells". In: *Development* 131.12, pp. 2791–2801.

Hoffman, E. A., A. McCulley, B. Haarer, R. Arnak, and W. Feng (2015). "Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription". In: *Genome Res.* 25.3, pp. 402–412.

Holliday, R. (1987). "The inheritance of epigenetic defects". In: *Science* 238.4824, pp. 163–170.

Horton, J. R., A. K. Upadhyay, H. H. Qi, X. Zhang, Y. Shi, and X. Cheng (2010). "Enzymatic and structural insights for substrate specificity of

a family of jumonji histone lysine demethylases". In: *Nat. Struct. Mol. Biol.* 17.1, pp. 38–43.

Horton, J. R., A. K. Upadhyay, H. Hashimoto, X. Zhang, and X. Cheng (2011). "Structural basis for human PHF2 Jumonji domain interaction with metal ions". In: *J. Mol. Biol.* 406.1, pp. 1–8.

Huang, C., Y. Xiang, Y. Wang, X. Li, L. Xu, Z. Zhu, T. Zhang, Q. Zhu, K. Zhang, N. Jing, and C. D. Chen (2010a). "Dual-specificity histone demethylase KIAA1718 (KDM7A) regulates neural differentiation through FGF4". In: *Cell Res.* 20.2, pp. 154–165.

Huang, C., J. Chen, T. Zhang, Q. Zhu, Y. Xiang, C. D. Chen, and N. Jing (2010b). "The dual histone demethylase KDM7A promotes neural induction in early chick embryos". In: *Dev. Dyn.* 239.12, pp. 3350–3357.

Iossifov, I., M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y. H. Lee, G. Narzisi, A. Leotta, J. Kendall, E. Grabowska, B. Ma, S. Marks, L. Rodgers, A. Stepansky, J. Troge, P. Andrews, M. Bekritsky, K. Pradhan, E. Ghiban, M. Kramer, J. Parla, R. Demeter, L. L. Fulton, R. S. Fulton, V. J. Magrini, K. Ye, J. C. Darnell, R. B. Darnell, E. R. Mardis, R. K. Wilson, M. C. Schatz, W. R. McCombie, and M. Wigler (2012). "De novo gene disruptions in children on the autistic spectrum". In: *Neuron* 74.2, pp. 285–299.

Isaac, S., J. Walfridsson, T. Zohar, D. Lazar, T. Kahan, K. Ekwall, and A. Cohen (2007). "Interaction of Epe1 with the heterochromatin assembly pathway in Schizosaccharomyces pombe". In: *Genetics* 175.4, pp. 1549–1560.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code". In: *Science* 293.5532, pp. 1074–1080.

Jenuwein, T., G. Laible, R. Dorn, and G. Reuter (1998). "SET domain proteins modulate chromatin domains in eu- and heterochromatin". In: *Cell. Mol. Life Sci.* 54.1, pp. 80–93.

Jepsen, K., D. Solum, T. Zhou, R. J. McEvilly, H. J. Kim, C. K. Glass, O. Hermanson, and M. G. Rosenfeld (2007). "SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron". In: *Nature* 450.7168, pp. 415–419.

Jessell, T. M. (Oct. 2000). "Neuronal specification in the spinal cord: inductive signals and transcriptional codes". In: *Nat. Rev. Genet.* 1.1, pp. 20–29.

Jung, H. R., D. Pasini, K. Helin, and O. N. Jensen (2010). "Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36". In: *Mol. Cell Proteomics* 9.5, pp. 838–850.

Karytinos, A., F. Forneris, A. Profumo, G. Ciossani, E. Battaglioli, C. Binda, and A. Mattevi (2009). "A novel mammalian flavin-dependent histone demethylase". In: *J. Biol. Chem.* 284.26, pp. 17775–17782.

Katayama, Y., M. Nishiyama, H. Shoji, Y. Ohkawa, A. Kawamura, T. Sato, M. Suyama, T. Takumi, T. Miyakawa, and K. I. Nakayama (Sept. 2016). "CHD8 haploinsufficiency results in autistic-like phenotypes in mice". In: *Nature* 537.7622, pp. 675–679.

Khan, S. N. and A. U. Khan (2010). "Role of histone acetylation in cell physiology and diseases: An update". In: *Clin. Chim. Acta* 411.19-20, pp. 1401–1411.

Kim, D., B. Langmead, and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements". In: *Nat. Methods* 12.4, pp. 357–360.

Kim, H. J., J. W. Park, K. H. Lee, H. Yoon, D. H. Shin, U. I. Ju, S. H. Seok, S. H. Lim, Z. H. Lee, H. H. Kim, and Y. S. Chun (2014). "Plant homeodomain finger protein 2 promotes bone formation by demethylating and activating Runx2 for osteoblast differentiation". In: *Cell Res.* 24.10, pp. 1231–1249.

Klose, R. J., E. M. Kallin, and Y. Zhang (Sept. 2006). "JmjC-domain-containing proteins and histone demethylation". In: *Nat. Rev. Genet.* 7.9, pp. 715–727.

Komada, M. (2012). "Sonic hedgehog signaling coordinates the proliferation and differentiation of neural stem/progenitor cells by regulating cell cycle kinetics during development of the neocortex". In: *Congenit Anom (Kyoto)* 52.2, pp. 72–77.

Kondo, Y., L. Shen, S. Ahmed, Y. Boumber, Y. Sekido, B. R. Haddad, and J. P. Issa (2008). "Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells". In: *PLoS ONE* 3.4, e2037.

Kooistra, S. M. and K. Helin (2012). "Molecular mechanisms and potential functions of histone demethylases". In: *Nat. Rev. Mol. Cell Biol.* 13.5, pp. 297–311.

Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA". In: *Science* 184.4139, pp. 868–871.

Kouzarides, T. (2007). "Chromatin modifications and their function". In: *Cell* 128.4, pp. 693–705.

Kubicek, S., R. J. O'Sullivan, E. M. August, E. R. Hickey, Q. Zhang, M. L. Teodoro, S. Rea, K. Mechtler, J. A. Kowalski, C. A. Homon, T. A. Kelly, and T. Jenuwein (2007). "Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase". In: *Mol. Cell* 25.3, pp. 473–481.

Kuo, M. H. and C. D. Allis (1998). "Roles of histone acetyltransferases and deacetylases in gene regulation". In: *Bioessays* 20.8, pp. 615–626.

Kuzmichev, A., K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein". In: *Genes Dev.* 16.22, pp. 2893–2905.

Lan, F. and Y. Shi (2009). "Epigenetic regulation: methylation of histone and non-histone proteins". In: *Sci. China, C, Life Sci.* 52.4, pp. 311–322.

Lander, E. S. et al. (2001). "Initial sequencing and analysis of the human genome". In: *Nature* 409.6822, pp. 860–921.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2". In: *Nat. Methods* 9.4, pp. 357–359.

Le Dreau, G. and E. Marti (2012). "Dorsal-ventral patterning of the neural tube: a tale of three signals". In: *Dev Neurobiol* 72.12, pp. 1471–1481.

Lee, B. M. and L. C. Mahadevan (2009). "Stability of histone modifications across mammalian genomes: implications for 'epigenetic' marking". In: *J. Cell. Biochem.* 108.1, pp. 22–34.

Lee, J. H., N. J. Yoo, M. S. Kim, and S. H. Lee (2017). "Histone Demethylase Gene PHF2 Is Mutated in Gastric and Colorectal Cancers". In: *Pathol. Oncol. Res.* 23.3, pp. 471–476.

Lee, K. H., U. I. Ju, J. Y. Song, and Y. S. Chun (2014). "The histone demethylase PHF2 promotes fat cell differentiation as an epigenetic activator of both C/EBPÎ± and C/EBPÎ´". In: *Mol. Cells* 37.10, pp. 734–741.

Lee, K. H., J. W. Park, H. S. Sung, Y. J. Choi, W. H. Kim, H. S. Lee, H. J. Chung, H. W. Shin, C. H. Cho, T. Y. Kim, S. H. Li, H. D. Youn, S. J. Kim, and Y. S. Chun (2015). "PHF2 histone demethylase acts as a tumor suppressor in association with p53 in cancer". In: *Oncogene* 34.22, pp. 2897–2909.

Lehnertz, B., Y. Ueda, A. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A. H. Peters (2003). "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin". In: *Curr. Biol.* 13.14, pp. 1192–1200.

Levitt, P. and P. Rakic (1980). "Immunoperoxidase localization of glial fibrillary acidic protein in radial glial cells and astrocytes of the developing rhesus monkey brain". In: *J. Comp. Neurol.* 193.3, pp. 815–840.

Li, E., T. H. Bestor, and R. Jaenisch (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality". In: *Cell* 69.6, pp. 915–926.

Li, H., S. Ilin, W. Wang, E. M. Duncan, J. Wysocka, C. D. Allis, and D. J. Patel (2006). "Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF". In: *Nature* 442.7098, pp. 91–95.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079.

Li, W., C. A. Cogswell, and J. J. LoTurco (1998). "Neuronal differentiation of precursors in the neocortical ventricular zone is triggered by BMP". In: *J. Neurosci.* 18.21, pp. 8853–8862.

Lim, D. A., Y. C. Huang, T. Swigut, A. L. Mirick, J. M. Garcia-Verdugo, J. Wysocka, P. Ernst, and A. Alvarez-Buylla (2009). "Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells". In: *Nature* 458.7237, pp. 529–533.

Lim, H. J., N. V. Dimova, M. K. Tan, F. D. Sigoillot, R. W. King, and Y. Shi (2013). "The G2/M regulator histone demethylase PHF8 is targeted for degradation by the anaphase-promoting complex containing CDC20". In: *Mol. Cell. Biol.* 33.21, pp. 4166–4180.

Lim, S. and P. Kaldis (2013). "Cdks, cyclins and CKIs: roles beyond cell cycle regulation". In: *Development* 140.15, pp. 3079–3093.

Liu, B. and B. M. Alberts (1995). "Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex". In: *Science* 267.5201, pp. 1131–1137.

Liu, W., B. Tanasa, O. V. Tyurina, T. Y. Zhou, R. Gassmann, W. T. Liu, K. A. Ohgi, C. Benner, I. Garcia-Bassets, A. K. Aggarwal, A. Desai, P. C. Dorrestein, C. K. Glass, and M. G. Rosenfeld (2010). "PHF8 mediates

histone H4 lysine 20 demethylation events involved in cell cycle progression". In: *Nature* 466.7305, pp. 508–512.

Love, M. I., W. Huber, and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biol.* 15.12, p. 550.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution". In: *Nature* 389.6648, pp. 251–260.

Lui, J. H., D. V. Hansen, and A. R. Kriegstein (2011). "Development and evolution of the human neocortex". In: *Cell* 146.1, pp. 18–36.

LYON, M. F. (1962). "Sex chromatin and gene action in the mammalian X-chromosome". In: *Am. J. Hum. Genet.* 14, pp. 135–148.

Ma, H. T. and R. Y. Poon (2017). "Synchronization of HeLa Cells". In: *Methods Mol. Biol.* 1524, pp. 189–201.

Macaluso, M., M. Montanari, and A. Giordano (2006). "Rb family proteins as modulators of gene expression and new aspects regarding the interaction with chromatin remodeling enzymes". In: *Oncogene* 25.38, pp. 5263–5267.

Marthiens, V., M. Piel, and R. Basto (2012). "Never tear us apart–the importance of centrosome clustering". In: *J. Cell. Sci.* 125.Pt 14, pp. 3281–3292.

Martynoga, B., D. Drechsel, and F. Guillemot (2012). "Molecular control of neurogenesis: a view from the mammalian cerebral cortex". In: *Cold Spring Harb Perspect Biol* 4.10.

Mi, H., X. Huang, A. Muruganujan, H. Tang, C. Mills, D. Kang, and P. D. Thomas (Jan. 2017). "PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements". In: *Nucleic Acids Res.* 45.D1, pp. D183–D189.

Millanes-Romero, A., N. Herranz, V. Perrera, A. Iturbide, J. Loubat-Casanovas, J. Gil, T. Jenuwein, A. Garcia de Herreros, and S. Peiro (2013). "Regulation of heterochromatin transcription by Snail1/LOXL2 during epithelial-to-mesenchymal transition". In: *Mol. Cell* 52.5, pp. 746–757.

Molina, O., N. Kouprina, H. Masumoto, V. Larionov, and W. C. Earnshaw (2017). "Using human artificial chromosomes to study centromere assembly and function". In: *Chromosoma* 126.5, pp. 559–575.

Muller, H. and K. Helin (2000). "The E2F transcription factors: key regulators of cell proliferation". In: *Biochim. Biophys. Acta* 1470.1, pp. 1–12.

Muller, S. and G. Almouzni (2014). "A network of players in H3 histone variant deposition and maintenance at centromeres". In: *Biochim. Biophys. Acta* 1839.3, pp. 241–250.

— (Mar. 2017). "Chromatin dynamics during the cell cycle at centromeres". In: *Nat. Rev. Genet.* 18.3, pp. 192–208.

Nakayama, J., J. C. Rice, B. D. Strahl, C. D. Allis, and S. I. Grewal (2001). "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly". In: *Science* 292.5514, pp. 110–113.

Nevins, J. R. (2001). "The Rb/E2F pathway and cancer". In: *Hum. Mol. Genet.* 10.7, pp. 699–703.

Nevins, J. R., S. P. Chellappan, M. Mudryj, S. Hiebert, S. Devoto, J. Horowitz, T. Hunter, and J. Pines (1991). "E2F transcription factor is a target for the RB protein and the cyclin A protein". In: *Cold Spring Harb. Symp. Quant. Biol.* 56, pp. 157–162.

Nicholson, G. A., J. L. Dawkins, I. P. Blair, M. L. Kennerson, M. J. Gordon, A. K. Cherryson, J. Nash, and T. Bananis (1996). "The gene for hereditary sensory neuropathy type I (HSN-I) maps to chromosome 9q22.1-q22.3". In: *Nat. Genet.* 13.1, pp. 101–104.

Noble, D. and C. Waddington (2015). "Conrad Waddington and the origin of epigenetics". In: *J. Exp. Biol.* 218.Pt 6, pp. 816–818.

Ohta, S., J. C. Bukowski-Wills, L. Sanchez-Pulido, F. d. e. L. Alves, L. Wood, Z. A. Chen, M. Platani, L. Fischer, D. F. Hudson, C. P. Ponting, T. Fukagawa, W. C. Earnshaw, and J. Rappsilber (2010). "The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics". In: *Cell* 142.5, pp. 810–821.

Ohtaka-Maruyama, C. and H. Okado (2015). "Molecular Pathways Underlying Projection Neuron Production and Migration during Cerebral Cortical Development". In: *Front Neurosci* 9, p. 447.

Okano, M., D. W. Bell, D. A. Haber, and E. Li (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development". In: *Cell* 99.3, pp. 247–257.

Okuno, Y., F. Ohtake, K. Igarashi, J. Kanno, T. Matsumoto, I. Takada, S. Kato, and Y. Imai (2013). "Epigenetic regulation of adipogenesis by PHF2 histone demethylase". In: *Diabetes* 62.5, pp. 1426–1434.

Ornitz, D. M. and N. Itoh (2015). "The Fibroblast Growth Factor signaling pathway". In: *Wiley Interdiscip Rev Dev Biol* 4.3, pp. 215–266.

O'Roak, B. J., L. Vives, W. Fu, J. D. Egertson, I. B. Stanaway, I. G. Phelps, G. Carvill, A. Kumar, C. Lee, K. Ankenman, J. Munson, J. B. Hiatt, E. H. Turner, R. Levy, D. R. O'Day, N. Krumm, B. P. Coe, B. K. Martin, E. Borenstein, D. A. Nickerson, H. C. Mefford, D. Doherty, J. M. Akey, R. Bernier, E. E. Eichler, and J. Shendure (2012). "Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders". In: *Science* 338.6114, pp. 1619–1622.

Ozair, M. Z., C. Kintner, and A. H. Brivanlou (2013). "Neural induction and early patterning in vertebrates". In: *Wiley Interdiscip Rev Dev Biol* 2.4, pp. 479–498.

Palmer, D. K., K. O'Day, H. L. Trong, H. Charbonneau, and R. L. Margolis (1991). "Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone". In: *Proc. Natl. Acad. Sci. U.S.A.* 88.9, pp. 3734–3738.

Pattabiraman, D. R., B. Bierie, K. I. Kober, P. Thiru, J. A. Krall, C. Zill, F. Reinhardt, W. L. Tam, and R. A. Weinberg (2016). "Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability". In: *Science* 351.6277, aad3680.

Patthey, C. and L. Gunhaga (2014). "Signaling pathways regulating ectodermal cell fate choices". In: *Exp. Cell Res.* 321.1, pp. 11–16.

Pauler, F. M., M. A. Sloane, R. Huang, K. Regha, M. V. Koerner, I. Tamir, A. Sommer, A. Aszodi, T. Jenuwein, and D. P. Barlow (2009). "H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome". In: *Genome Res.* 19.2, pp. 221–233.

Pedersen, M. T. and K. Helin (2010). "Histone demethylases in development and disease". In: *Trends Cell Biol.* 20.11, pp. 662–671.

Perino, M. and G. J. Veenstra (Sept. 2016). "Chromatin Control of Developmental Dynamics and Plasticity". In: *Dev. Cell* 38.6, pp. 610–620.

Perkins, D. N., D. J. Pappin, D. M. Creasy, and J. S. Cottrell (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data". In: *Electrophoresis* 20.18, pp. 3551–3567.

Peters, A. H., D. O'Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schofer, K. Weipoltshammer, M. Pagani, M. Lachner, A. Kohlmaier, S. Opravil, M. Doyle, M. Sibilia, and T. Jenuwein (2001). "Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability". In: *Cell* 107.3, pp. 323–337.

Peters, A. H., S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, J. H. Martens, and T. Jenuwein (2003). "Partitioning and plasticity of repressive histone methylation states in mammalian chromatin". In: *Mol. Cell* 12.6, pp. 1577–1589.

Pollard, S. M., L. Conti, Y. Sun, D. Goffredo, and A. Smith (2006). "Adherent neural stem (NS) cells from fetal and adult forebrain". In: *Cereb. Cortex* 16 Suppl 1, pp. i112–120.

Probst, A. V. and G. Almouzni (2011). "Heterochromatin establishment in the context of genome-wide epigenetic reprogramming". In: *Trends Genet.* 27.5, pp. 177–185.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features". In: *Bioinformatics* 26.6, pp. 841–842.

Rakic, P. (2003). "Elusive radial glial cells: historical and evolutionary perspective". In: *Glia* 43.1, pp. 19–32.

Ramirez, F., D. P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dundar, and T. Manke (July 2016). "deepTools2: a next generation web server for deep-sequencing data analysis". In: *Nucleic Acids Res.* 44.W1, W160–165.

Rea, S., F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis, and T. Jenuwein (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases". In: *Nature* 406.6796, pp. 593–599.

Roberts, R. W. and D. M. Crothers (1992). "Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition". In: *Science* 258.5087, pp. 1463–1466.

Robertson, K. D., K. Keyomarsi, F. A. Gonzales, M. Velicescu, and P. A. Jones (2000). "Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells". In: *Nucleic Acids Res.* 28.10, pp. 2108–2113.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov (2011). "Integrative genomics viewer". In: *Nat. Biotechnol.* 29.1, pp. 24–26.

Rondinelli, B., D. Rosano, E. Antonini, M. Frenquelli, L. Montanini, D. Huang, S. Segalla, K. Yoshihara, S. B. Amin, D. Lazarevic, B. T. The, R. G. Verhaak, P. A. Futreal, L. Di Croce, L. Chin, D. Cittaro, and G. Tonon

(2015). "Histone demethylase JARID1C inactivation triggers genomic instability in sporadic renal cancer". In: *J. Clin. Invest.* 125.12, pp. 4625–4637.

Rosenfeld, J. A., Z. Wang, D. E. Schones, K. Zhao, R. DeSalle, and M. Q. Zhang (2009). "Determination of enriched histone modifications in non-genic portions of the human genome". In: *BMC Genomics* 10, p. 143.

Rudert, F., S. Bronner, J. M. Garnier, and P. Dolle (1995). "Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development". In: *Mamm. Genome* 6.2, pp. 76–83.

Saksouk, N., E. Simboeck, and J. Dejardin (2015). "Constitutive heterochromatin formation and transcription in mammals". In: *Epigenetics Chromatin* 8, p. 3.

Sankar, T. S., B. D. Wastuwidyaningtyas, Y. Dong, S. A. Lewis, and J. D. Wang (July 2016). "The nature of mutations induced by replicationâ€"transcription collisions". In: *Nature* 535.7610, pp. 178–181.

Scharf, A. N., T. K. Barth, and A. Imhof (2009). "Establishment of histone modifications after chromatin assembly". In: *Nucleic Acids Res.* 37.15, pp. 5032–5040.

SCHERER, W. F., J. T. SYVERTON, and G. O. GEY (1953). "Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix". In: *J. Exp. Med.* 97.5, pp. 695–710.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis". In: *Nat. Methods* 9.7, pp. 671–675.

Schotta, G., R. Sengupta, S. Kubicek, S. Malin, M. Kauer, E. Callen, A. Celeste, M. Pagani, S. Opravil, I. A. De La Rosa-Velazquez, A. Espejo, M. T. Bedford, A. Nussenzweig, M. Busslinger, and T. Jenuwein (2008). "A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse". In: *Genes Dev.* 22.15, pp. 2048–2061.

Sher, F., R. Rossler, N. Brouwer, V. Balasubramaniyan, E. Boddeke, and S. Copray (2008). "Differentiation of neural stem cells into oligodendrocytes: involvement of the polycomb group protein Ezh2". In: *Stem Cells* 26.11, pp. 2875–2883.

Shi, G., M. Wu, L. Fang, F. Yu, S. Cheng, J. Li, J. X. Du, and J. Wong (2014). "PHD finger protein 2 (PHF2) represses ribosomal RNA gene transcription by antagonizing PHF finger protein 8 (PHF8) and recruiting methyltransferase SUV39H1". In: *J. Biol. Chem.* 289.43, pp. 29691–29700.

Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, and Y. Shi (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1". In: *Cell* 119.7, pp. 941–953.

Shimojo, H., T. Ohtsuka, and R. Kageyama (2008). "Oscillations in notch signaling regulate maintenance of neural progenitors". In: *Neuron* 58.1, pp. 52–64.

Shinkai, Y. and M. Tachibana (2011). "H3K9 methyltransferase G9a and the related molecule GLP". In: *Genes Dev.* 25.8, pp. 781–788.

Siderius, L. E., B. C. Hamel, H. van Bokhoven, F. de Jager, B. van den Helm, H. Kremer, J. A. Heineman-de Boer, H. H. Ropers, and E. C. Mariman (1999). "X-linked mental retardation associated with cleft lip/palate maps to Xp11.3-q21.3". In: *Am. J. Med. Genet.* 85.3, pp. 216–220.

Siegenthaler, J. A. and M. W. Miller (2005). "Transforming growth factor beta 1 promotes cell cycle exit through the cyclin-dependent kinase inhibitor p21 in the developing cerebral cortex". In: *J. Neurosci.* 25.38, pp. 8627–8636.

Skourti-Stathaki, K. and N. J. Proudfoot (2014). "A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression". In: *Genes Dev.* 28.13, pp. 1384–1396.

Sogo, J. M., M. Lopes, and M. Foiani (2002). "Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects". In: *Science* 297.5581, pp. 599–602.

Sollier, J. and K. A. Cimprich (2015). "Breaking bad: R-loops and genome integrity". In: *Trends Cell Biol.* 25.9, pp. 514–522.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications". In: *Nature* 403.6765, pp. 41–45.

Sugimura, K., Y. Fukushima, M. Ishida, S. Ito, M. Nakamura, Y. Mori, and K. Okumura (2010). "Cell cycle-dependent accumulation of histone H3.3 and euchromatic histone modifications in pericentromeric heterochromatin in response to a decrease in DNA methylation levels". In: *Exp. Cell Res.* 316.17, pp. 2731–2746.

Sullivan, B. A. and G. H. Karpen (2004). "Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin". In: *Nat. Struct. Mol. Biol.* 11.11, pp. 1076–1083.

Tachibana, M., K. Sugimoto, T. Fukushima, and Y. Shinkai (2001). "Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3". In: *J. Biol. Chem.* 276.27, pp. 25309–25317.

Tachibana, M., K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, M. Ohki, M. Fukuda, N. Takeda, H. Niida, H. Kato, and Y. Shinkai (2002). "G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis". In: *Genes Dev.* 16.14, pp. 1779–1791.

Takahashi, Y., J. B. Rayman, and B. D. Dynlacht (2000). "Analysis of promoter binding by the E2F and pRB families in vivo: distinct E2F proteins mediate activation and repression". In: *Genes Dev.* 14.7, pp. 804–816.

Temple, S. (2001). "The development of neural stem cells". In: *Nature* 414.6859, pp. 112–117.

Thomas, M., R. L. White, and R. W. Davis (1976). "Hybridization of RNA to double-stranded DNA: formation of R-loops". In: *Proc. Natl. Acad. Sci. U.S.A.* 73.7, pp. 2294–2298.

Tiberi, L., P. Vanderhaeghen, and J. van den Ameele (2012). "Cortical neurogenesis and morphogens: diversity of cues, sources and functions". In: *Curr. Opin. Cell Biol.* 24.2, pp. 269–276.

Ting, D. T., D. Lipson, S. Paul, B. W. Brannigan, S. Akhavanfard, E. J. Coffman, G. Contino, V. Deshpande, A. J. Iafrate, S. Letovsky, M. N. Rivera, N. Bardeesy, S. Maheswaran, and D. A. Haber (2011). "Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers". In: *Science* 331.6017, pp. 593–596.

Trewick, S. C., E. Minc, R. Antonelli, T. Urano, and R. C. Allshire (2007). "The JmjC domain protein Epe1 prevents unregulated assembly and disassembly of heterochromatin". In: *EMBO J.* 26.22, pp. 4670–4682.

Tschiersch, B., A. Hofmann, V. Krauss, R. Dorn, G. Korge, and G. Reuter (1994). "The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic

regulators of homeotic gene complexes". In: *EMBO J.* 13.16, pp. 3822–3831.

Tsukada, Y., T. Ishitani, and K. I. Nakayama (2010). "KDM7 is a dual demethylase for histone H3 Lys 9 and Lys 27 and functions in brain development". In: *Genes Dev.* 24.5, pp. 432–437.

Tsukada, Y., J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers, P. Tempst, and Y. Zhang (2006). "Histone demethylation by a family of JmjC domain-containing proteins". In: *Nature* 439.7078, pp. 811–816.

Valgardsdottir, R., I. Chiodi, M. Giordano, F. Cobianchi, S. Riva, and G. Biamonti (2005). "Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells". In: *Mol. Biol. Cell* 16.6, pp. 2597–2604.

Wagner, T. and M. Jung (2012). "New lysine methyltransferase drug targets in cancer". In: *Nat. Biotechnol.* 30.7, pp. 622–623.

Wallingford, J. B., L. A. Niswander, G. M. Shaw, and R. H. Finnell (2013). "The continuing challenge of understanding, preventing, and treating neural tube defects". In: *Science* 339.6123, p. 1222002.

Wang, J., B. D. Reddy, and S. Jia (2015). "Rapid epigenetic adaptation to uncontrolled heterochromatin spreading". In: *Elife* 4.

Weinberg, R. A. (1995). "The retinoblastoma protein and cell cycle control". In: *Cell* 81.3, pp. 323–330.

Weintraub, S. J., K. N. Chow, R. X. Luo, S. H. Zhang, S. He, and D. C. Dean (1995). "Mechanism of active transcriptional repression by the retinoblastoma protein". In: *Nature* 375.6534, pp. 812–815.

Wen, B., H. Wu, Y. Shinkai, R. A. Irizarry, and A. P. Feinberg (2009). "Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells". In: *Nat. Genet.* 41.2, pp. 246–250.

Wen, H., J. Li, T. Song, M. Lu, P. Y. Kan, M. G. Lee, B. Sha, and X. Shi (2010). "Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation". In: *J. Biol. Chem.* 285.13, pp. 9322–9326.

Westover, K. D., D. A. Bushnell, and R. D. Kornberg (2004). "Structural basis of transcription: separation of RNA from DNA by RNA polymerase II". In: *Science* 303.5660, pp. 1014–1016.

Wilde, J. J., J. R. Petersen, and L. Niswander (2014). "Genetic, epigenetic, and environmental contributions to neural tube closure". In: *Annu. Rev. Genet.* 48, pp. 583–611.

Wilkinson, G., D. Dennis, and C. Schuurmans (2013). "Proneural genes in neocortical development". In: *Neuroscience* 253, pp. 256–273.

Wilson, A. S., B. E. Power, and P. L. Molloy (2007). "DNA hypomethylation and human diseases". In: *Biochim. Biophys. Acta* 1775.1, pp. 138–162.

Wilson, S. W. and C. Houart (2004). "Early steps in the development of the forebrain". In: *Dev. Cell* 6.2, pp. 167–181.

Yang, J., J. Ma, Y. Xiong, Y. Wang, K. Jin, W. Xia, Q. Chen, J. Huang, J. Zhang, N. Jiang, S. Jiang, and D. Ma (2018). "Epigenetic regulation of megakaryocytic and erythroid differentiation by PHF2 histone demethylase". In: *J. Cell. Physiol.* 233.9, pp. 6841–6852.

Yokochi, T., K. Poduch, T. Ryba, J. Lu, I. Hiratani, M. Tachibana, Y. Shinkai, and D. M. Gilbert (2009). "G9a selectively represses a class of late-replicating genes at the nuclear periphery". In: *Proc. Natl. Acad. Sci. U.S.A.* 106.46, pp. 19363–19368.

Yokoyama, A., Z. Wang, J. Wysocka, M. Sanyal, D. J. Aufiero, I. Kitabayashi, W. Herr, and M. L. Cleary (2004). "Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression". In: *Mol. Cell. Biol.* 24.13, pp. 5639–5649.

Young, L. C., D. W. McDonald, and M. J. Hendzel (2013). "Kdm4b histone demethylase is a DNA damage response protein and confers a survival advantage following Î³-irradiation". In: *J. Biol. Chem.* 288.29, pp. 21376–21388.

Yu, G., L. G. Wang, and Q. Y. He (2015). "ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization". In: *Bioinformatics* 31.14, pp. 2382–2383.

Zeller, P. and S. M. Gasser (2017). "The Importance of Satellite Sequence Repression for Genome Stability". In: *Cold Spring Harb. Symp. Quant. Biol.* 82, pp. 15–24.

Zeller, P., J. Padeken, R. van Schendel, V. Kalck, M. Tijsterman, and S. M. Gasser (Nov. 2016). "Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability". In: *Nat. Genet.* 48.11, pp. 1385–1395.

Zhang, H. S. and D. C. Dean (2001). "Rb-mediated chromatin structure regulation and transcriptional repression". In: *Oncogene* 20.24, pp. 3134–3138.

Zhang, K., K. Mosch, W. Fischle, and S. I. Grewal (2008a). "Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin". In: *Nat. Struct. Mol. Biol.* 15.4, pp. 381–388.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu (2008b). "Model-based analysis of ChIP-Seq (MACS)". In: *Genome Biol.* 9.9, R137.

Zheng, H., L. Chen, W. J. Pledger, J. Fang, and J. Chen (2014). "p53 promotes repair of heterochromatin DNA by regulating JMJD2b and SUV39H1 expression". In: *Oncogene* 33.6, pp. 734–744.

Zhu, Q., G. M. Pao, A. M. Huynh, H. Suh, N. Tonnu, P. M. Nederlof, F. H. Gage, and I. M. Verma (2011). "BRCA1 tumour suppression occurs via heterochromatin-mediated silencing". In: *Nature* 477.7363, pp. 179–184.

Zofall, M. and S. I. Grewal (2006). "Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats". In: *Mol. Cell* 22.5, pp. 681–692.

Appendix A

Appendix

A.1 Publications

In this appendix I would like to include the publications I have participated during my doctoral thesis.

1. Fueyo, R., S. Iacobucci, S. Pappa, C. Estaras, S. Lois, M. Vicioso-Mantis, C. Navarro, S. Cruz-Molina, J. C. Reyes, A. Rada-Iglesias, X. de la Cruz, and M. A. Martinez-Balbas (2018). "Lineage specific transcription factos and epigenetic regulators mediate TGFbeta-dependent enhancer activation." Nucleic Acids Res 46(7): 3351-3365

2. Asensio-Juan, E*., R. Fueyo*, S. Pappa, S. Iacobucci, C. Badosa, S. Lois, M. Balada, L. Bosch-Presegue, A. Vaquero, S. Gutierrez, C. Caelles, C. Gallego, X. de la Cruz, and M. A. Martinez-Balbas (2017). "The histone demethylase PHF8 is a molecular safeguard of the IFNgamma response." Nucleic Acids Res 45(7): 3800-3811