Characterization of DNMT3C-mediated inactivation of active Transposons in Mouse

Dissertation
Zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Srinivasa Abishek Prakash
geboren am 16 Juli 1993 in Chennai

Mainz, 2024

Dekan: **Prof. Dr. Eckhard Thines**

1. Betreuer: **Dr. Joan Barau**

2. Betreuer: **Prof. Dr. René F. Ketting**

Tag der mündlichen Prüfung: **06.11.2024**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Transposable elements (TEs) are parasitic sequences within the genome. Argonaute PIWI proteins and their associated piwi-interacting RNAs (piRNAs) are required for DNA methylation-dependent inactivation of transposable elements (TEs) in mouse embryonic germ cells. However, how DNA methylation is targeted to the context of an active promoter and which DNA methylation machinery is involved remains unclear. Mice evolved a germline-specific *de novo* DNA methyltransferase, DNMT3C, whose activity is essential for TE promoter methylation. To explore the mechanisms of DNMT3C targeting, we generated a transgenic mouse line with an endogenously tagged DNMT3C allele. We used this transgenic line to uncover DNMT3C's protein interactors in male embryonic germ cells in mice through immunoprecipitation—mass spectrometry. Among the uncovered proteins enriched in the DNMT3C interactome, we found members of the core nuclear piRNA pathway and related proteins, including MIWI2, SPOCD1, TEX15, and SPIN1, significantly enriched. We validated this connection to the nuclear piRNA pathway using immunoprecipitation followed by western blotting and nuclear colocalization in embryonic germ cells. The genome-wide targets of DNMT3C, MIWI2, and SPOCD1 in male germ cells are nearly identical. Our data shows that DNMT3C is targeted at TE promoters by the nuclear piRNA pathway.

During mouse germline development, young and active TE promoters become transcriptionally active and gain H3K4me3. H3K4me3 is a modification present at active promoters and is known to repel *de novo* DNA methylation by preventing DNMT3 proteins from binding to chromatin. Intriguingly, we discovered that DNMT3C is essential for removing this antagonistic mark from its targets, indicating that the upstream nuclear piRNA pathway's activity is insufficient to remove H3K4me3 from TE promoters. The *Lsd* demethylases LSD1 and LSD2 were enriched in the interactome data. We validated DNMT3C's interaction with LSD1. Germline conditional catalytic knock-in of *Lsd1* showed that DNMT3C DMRs are normally methylated, as observed in $Dnmt3c^{WT/WT}$. However, they are also marked by H3K4me3, as observed in $Dnmt3c^{KO/KO}$. This indicates that DNMT3C could have a non-canonical mechanism of DNA methylation where it methylates DNA marked by H3K4me3. We propose a model where DNMT3C acts downstream of the MIWI2 and SPOCD1 in the nuclear

piRNA pathway. LSD1, through its interaction with DNMT3C and SPOCD1, and in conjunction with JARID proteins, removes the H3K4me at the piRNA targets. In DNMT3C's absence, the active transposon promoters are not converted to the constitutive heterochromatin and are marked by H3K4me3, leading to a meiotic catastrophe. Our work provides mechanistic insights into DNMT3C-mediated TE inactivation and nuclear piRNA pathway organization.

ZUSAMMENFASSUNG

Transponierbare Elemente (TEs) sind parasitäre Sequenzen im Genom. Argonaute PIWI-Proteine und die mit ihnen verbundenen Piwi-interagierenden RNAs (piRNAs) sind für die DNA-Methylierung-abhängige Inaktivierung von transponierbaren Elementen (TEs) in embryonalen Keimzellen der Maus erforderlich. Unklar ist jedoch, wie die DNA-Methylierung in den Kontext eines aktiven Promotors gebracht wird und welche DNA-Methylierungsmaschinerie daran beteiligt ist. Bei Mäusen hat sich eine keimlinienspezifische de novo-DNA-Methyltransferase, DNMT3C, entwickelt, deren Aktivität für die TE-Promotor-Methylierung wesentlich ist. Um die Mechanismen des DNMT3C-Targetings zu untersuchen, haben wir eine transgene Mauslinie mit einem endogen markierten DNMT3C-Allel erzeugt. Mit Hilfe dieser transgenen Linie konnten wir die Proteininteraktoren von DNMT3C in männlichen embryonalen Keimzellen der Maus durch Immunpräzipitation und Massenspektrometrie aufdecken. Unter den aufgedeckten Proteinen, die im DNMT3C-Interaktom angereichert sind, fanden wir Mitglieder des zentralen nuklearen piRNA-Wegs und verwandte Proteine, einschließlich MIWI2, SPOCD1, TEX15 und SPIN1, signifikant angereichert. Wir validierten diese Verbindung zum nukleären piRNA-Signalweg durch Immunpräzipitation, gefolgt von Western Blotting und nukleärer Kolokalisierung in embryonalen Keimzellen. Die genomweiten Ziele von DNMT3C, MIWI2 und SPOCD1 in männlichen Keimzellen sind nahezu identisch. Unsere Daten zeigen, dass DNMT3C durch den piRNA-Weg auf TE-Promotoren ausgerichtet ist.

Während der Keimbahnentwicklung der Maus werden junge und aktive TE-Promotoren transkriptionell aktiv und gewinnen H3K4me3. H3K4me3 ist eine Modifikation, die an aktiven Promotoren vorkommt und dafür bekannt ist, de novo-DNA-Methylierung abzuwehren, indem sie DNMT3-Proteine an der Bindung an Chromatin hindert. Interessanterweise entdeckten wir, dass DNMT3C entscheidend für die Entfernung dieser antagonistischen Markierung von seinen Zielen ist, was darauf hindeutet, dass die Aktivität des vorgelagerten nuklearen piRNA-Wegs nicht ausreicht, um H3K4me3 von TE-Promotoren zu entfernen. Die Lsd-Demethylasen LSD1 und LSD2 waren in den Interaktomdaten angereichert. Wir validierten die Interaktion von DNMT3C mit LSD1. Keimbahnbedingtes katalytisches Knock-in von Lsd1 zeigte, dass

DNMT3C DMRs normal methyliert sind, wie bei *Dnmt3c*$^{WT/WT}$ beobachtet. Sie sind jedoch auch durch H3K4me3 markiert, wie dies bei *Dnmt3c*$^{KO/KO}$ beobachtet wurde. Dies deutet darauf hin, dass DNMT3C über einen nicht-kanonischen Mechanismus der DNA-Methylierung verfügen könnte, bei dem es durch H3K4me3 markierte DNA methyliert. Wir schlagen ein Modell vor, bei dem DNMT3C stromabwärts von MIWI2 und SPOCD1 im nuklearen piRNA-Weg wirkt. LSD1 entfernt durch seine Interaktion mit DNMT3C und SPOCD1 und in Verbindung mit JARID-Proteinen das H3K4me an den piRNA-Zielen. In Abwesenheit von DNMT3C werden die aktiven Transposon-Promotoren nicht in das konstitutive Heterochromatin umgewandelt und sind durch H3K4me3 markiert, was zu einer meiotischen Katastrophe führt. Unsere Arbeit bietet mechanistische Einblicke in die DNMT3C-vermittelte TE-Inaktivierung und die Organisation des piRNA-Signalwegs im Nukleus.

INTRODUCTION

3.1 TRANSPOSONS

3.1.1 Transposable Elements – A Small History

The journey of transposon research began in the mid-20th century when Barbara McClintock, working with Maize (*Zea mays*), first proposed the concept of "controlling elements" that could change position within the genome, influencing gene expression and phenotype (MCCLINTOCK, 1950, 1956; Ravindran, 2012). This ground-breaking discovery earned McClintock the Nobel Prize in 1983.

In the subsequent decades, molecular biology and genetics advancements facilitated deeper investigations into transposons across diverse organisms. Zinder and Lederberg's work built upon earlier research in microbial genetics and bacterial conjugation, hinting at the possibility of genetic elements that could transfer between bacterial cells. Their key experiment, published in 1952, demonstrated genetic exchange in *Salmonella typhimurium*. They observed that certain genetic markers could transfer between bacterial strains during mating experiments, suggesting the existence of mobile genetic elements that facilitated this transfer (ZINDER and LEDERBERG, 1952). Their pioneering study laid the foundation for understanding the molecular mechanisms of transposition and their implications for bacterial genetics and evolution.

One of the earliest and most influential studies was conducted by David Hogness and colleagues in 1974, who discovered mobile genetic elements in the genome of *Drosophila*. They observed that these elements could transpose to different locations within the genome, thereby altering gene expression and potentially contributing to phenotypic variability (Wensink et al., 1974). Allan Spradling and Gerald Rubin made a significant breakthrough when they identified a new transposable element in the genome of *Drosophila melanogaster*. This element was later named the P element due to its preference for inserting near the chromosomal telomeres (Rubin, Kidwell and Bingham, 1982; Rubin and Spradling, 1982; Spradling and Rubin, 1982). The P element, initially identified in laboratory stocks of *Drosophila*, was found to transpose autonomously and could insert itself into various genomic

locations. Studies revealed that the P element carries its own transposase gene, catalyzing its movement within the genome (O'Hare and Rubin, 1983). The element also exhibits preferences for insertion sites and has specific structural features that distinguish it from other transposons found in *Drosophila*. The introduction of the P element into *Drosophila* research revolutionized the field of insect genetics. It provided a robust tool for creating transgenic flies, mapping genes, and studying gene function in vivo. Additionally, the P element's insertion patterns and evolutionary dynamics have contributed to our understanding of genome evolution and the role of transposable elements in generating genetic diversity.

3.1.2 Classes of Transposons

Transposons are broadly divided into two classes based on the transposition life cycle: Class I transposons, or Retrotransposons, and Class II transposons, or DNA transposons (Bourque et al., 2018).

Class I transposons move via a "copy and paste" mechanism, in which the transposon sequence is first transcribed into an mRNA; this mRNA is reverse-transcribed back into DNA by the enzyme Reverse Transcriptase. This DNA copy is then integrated at a novel position in the genome. This category of TEs consists of long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and long terminal repeat (LTR) retrotransposons; they all share this property and differ in other features of their organization and replication strategies (Kazazian, 2004). LINEs and LTR retrotransposons are autonomous TEs; that is, they use encoded proteins, including reverse transcriptase and endonuclease, in transposition. SINEs, on the other hand, are non-autonomous elements; they do not code for the proteins necessary for their transposition and rely upon the machinery provided by LINEs. The best-studied human family of SINEs is the Alu elements derived from the 7SL RNA gene. Alu elements are approximately 300 base pairs (bp) long and exist in over 1 million copies of the human genome, contributing to about 10% of it (Batzer and Deininger, 2002; Deininger et al., 2003). In mice, SINEs include the B1 and B2 families, which also use the L1 machinery for transposition (Deininger et al., 2003).

Class II transposons move primarily through a "cut and paste" mechanism, by which the transposon is excised from one location in the genome and inserted into another. This process is mediated by the enzyme transposase, encoded by the transposon. DNA transposons with terminal inverted repeats (TIRs) at their ends are the transposase binding sites. Examples of Class II transposons include P elements in *Drosophila*, which is well studied, and Ac/Ds elements in maize. Unlike Class I transposons, which can amplify copy numbers in the genome, Class II transposons usually do not multiply in the copy number unless they replicate before the excision event (Feschotte and Pritham, 2007).

Helitrons are another unique family of transposable elements. They are distinctive because of their particular “peel-and-paste” rolling-circle type of replication mechanism—something notably different from the other transposons described above (Bourque et al., 2018; Thomas and Pritham, 2015). Unlike most transposable elements, Helitrons do not harbor TIRs and are characterized by conserved 5' T/C and 3' CTRR (most frequently CTAG), sequences recognized by the replication machinery. They are thought to transpose through a single-stranded DNA intermediate—certainly a replicative transposase related to those found in bacterial plasmids and phages. Helitrons in many eukaryotic genomes—from plants to fungi and animals—highlight broad evolutionary importance and adaptability (Kapitonov and Jurka, 2001; Pritham and Feschotte, 2007). These elements play an essential role in genome evolution because they allow the trapping and shuffling of fragments from host genes, which could give rise to new genes and regulatory networks (Morgante et al., 2005; Yang and Bennetzen, 2009)

3.1.2.1 Long Interspersed Nuclear Element 1 (LINE-1)

LINEs belong to a class of non-LTR retrotransposons. Among all the non-LTR retrotransposons, the LINE-1 family has been studied extensively due to its abundance and contribution to genome structure and function. LINE-1 elements comprise around 20% of the mice genome and 18% of the human genome. The mouse genome contains about 1.5 million copies of LINE-1 elements. However, the number of potentially active LINE-1 elements—the

order of 3000 copies—is thought to be at least capable of retrotransposition (DeBerardinis et al., 1998).

A LINE-1 intact component is about 5000-6000 bp long and contains two ORFs (ORF1 and ORF2) flanked by 5' and 3' UTRs. The 5' UTR includes an internal promoter that drives LINE-1 mRNA transcription. It consists of tandemly repeated monomers and numerous transcription factor binding sites. The number of monomers and their sequence are different among various LINE-1 subfamilies. ORF1 encodes an RNA-binding chaperone protein, while ORF2 encodes a protein with endonuclease and reverse transcriptase activities that are imperative for the element's retrotransposition (Kazazian, 2004). The 3' UTR usually has a polyadenylation signal, which has a part in the stability and processing of the LINE-1 mRNA. Another interesting component of the LINE1's is the antisense promoter in the 5'UTR region (Li et al., 2014; Speek, 2001). It can drive transcription of the antisense transcripts, which might influence the transcriptional activity of the flanking genes and L1 in a negative feedback loop manner (Speek, 2001). This promoter activity may generate antisense RNAs that can interfere with the expression of LINE1 elements by mechanisms such as RNA interference or the formation of double-stranded RNA, leading to the degradation of LINE1 mRNA (Speek, 2001). This antisense promoter may also influence the expression of neighbour genes by providing an alternative promoter or enhancer, therefore incorporating LINE1 activity into the regulatory network of the host genome (Faulkner et al., 2009). The activity of such an antisense promoter could also help explain the enrichment of sense and antisense piRNAs (described later in this chapter) at the 5' UTR region.

The life cycle of LINE-1 starts with the transcription of the LINE-1 DNA into RNA by the host's RNA polymerase II. This mRNA is exported into the cytoplasm, where it is translated into the two proteins encoded by LINE1 – ORF1p and ORF2p. The ORF1p is a nucleic acid-binding protein with RNA chaperone activity. It assembles into trimers and binds LINE1 mRNA in cis (Martin, 1991; Martin and Bushman, 2001). Structurally, ORF1p contains a coiled-coil domain, a C-terminal RNA recognition motif, and an N-terminal domain, which participates in

its nucleic acid chaperone function essential for annealing LINE1 mRNA to target DNA during reverse transcription (Khazina and Weichenrieder, 2009). The mRNA-bound ORF1p interacts with the protein encoded by ORF2p and composes a ribonucleoprotein particle. ORF2p is a multifunctional protein carrying an endonuclease and a reverse transcriptase domain. This ribonucleoprotein particle is then imported into the nucleus. The endonuclease activity of ORF2p nicks the genomic DNA (Mita et al., 2018; Taylor et al., 2018), hence permitting LINE1 mRNA to prime its reverse transcription at the site of insertion (Cost et al., 2002; Feng et al., 1996; Thawani et al., 2024). This reverse transcriptase domain synthesizes the cDNA from the RNA template and then integrates this new copy of LINE1 into the genome. This process has come to be called target-primed reverse transcription (TPRT) and is a hallmark of non-LTR retrotransposons (Luan et al., 1993).

Biological implications due to LINE-1 elements are immense. LINE-1 insertions can disrupt normal gene function if they insert in exonic or regulatory regions, which potentially causes phenotypic alterations and diseases to develop (Kazazian and Moran, 2017; Tubio et al., 2014). This has been known to trigger some genetic disorders exhibited through LINE-1-induced mutation (Suarez, Macia and Muotri, 2018). In addition, LINE-1 activity acts as another source of genomic instability through structural variation of deletions, duplications, and rearrangements. For instance, active LINE-1 elements cause spontaneous mutations in mouse germ cells and, therefore, can result in heritable genetic changes (Kazazian, 2004). Indeed, how LINE-1 elements may be involved in genetic instability was also proved different from other related pathogenic factors described above. Besides, despite these possible deleterious effects, the host genome has also co-opted the LINE-1 elements for its benefit. These elements have been identified with the formation of new exons they created during gene expression regulation, the process known as exonization (Zemojtel et al., 2007). This could give rise to new regulatory elements or splicing variants, thereby increasing the intricacy and complexity of gene regulation.

LINE-1 elements have additionally been implicated in contributing to shaping the epigenetic landscape such that the environment and position of genes change due to the alterations caused by the elements as well as other retrotransposons like them in the chromatin structure itself (Slotkin and Martienssen, 2007). Their evolution seems to be closely intertwined with the host genome's evolution (Sookdeo et al., 2013). LINE-1 elements, then, have been a feature of mammalian genomes for at least 100 million years, indicating that they were evolving quite successfully in the past (Smit et al., 1995). Phylogenetic analyses showed several waves of LINE-1 element expansions followed by extinction throughout mammalian evolution. New subfamilies will have unique sequence properties and active LINE-1s that mark these waves of activity. For example, L1MdA is the youngest and the most active subfamily of LINE-1s found in mice, responsible for most of the most recent LINE-1 insertions reported to have occurred in other species in 2013 by Sookdeo *et al.*, older subfamilies, such as L1MdT, have accumulated mutations and are, therefore, largely inactive (Naas et al., 1998; Platt, Vandewege and Ray, 2018). Studying a myriad of organisms beyond mammals would provide us with a much-improved vision of TE evolutionary dynamics across the Animal Kingdom (like in this study - (Ivancevic et al., 2016)).

3.1.2.2 Long Terminal Repeat (LTR) Retrotransposons

Endogenous retroviruses, or ERVs, are transposable elements inserted into the host genome because of ancient retroviral infections. Critical roles have been attributed to ERVs defining genomic diversity and regulatory complexity in mice. These elements make up about 8-10% of the mouse genome. The same is found for the human genome (Jern and Coffin, 2008; Lander et al., 2001). It is estimated that in the mouse genome, there are approximately 2,000 unique ERV loci resulting from ancient infections that occurred millions of years ago (Gifford et al., 2018). Most of these ERVs have been inactivated by mutational events and epigenetic control; however, a small number is still capable of being active and potentially affecting the host genome in many different ways.

The structure of an ERV is similar to that of exogenous retroviruses. There are long terminal repeats (LTRs) at both ends of the ERV that have promoters and regulatory sequences required to transcribe it. The coding region between the LTRs contains genes encoding for viral proteins, including gag (group-specific antigen), pol (polymerase), and env (envelope). The gag gene encodes the viral particle's structural proteins, the pol gene encodes enzymes such as reverse transcriptase, integrase, and protease, and the env gene encodes the viral envelope protein. ERVs, in general, can be quite variable in size but will usually run to between 6,000 and 11,000 base pairs (bp) overall in length (Mager and Stoye, 2015).

The life cycle of an ERV begins with its transcription from the host genome, driven by the promoter sequences in its 5' LTR. This RNA transcript can serve dual functions: mRNA for translation into viral proteins and genomic RNA for packaging into viral particles. Once translated, the viral proteins assemble into a viral particle in the cytoplasm. While most ERVs in the mouse genome are defective, so they cannot give rise to infectious particles, this classic life cycle is mimicked. Reverse transcriptase, coded by the pol gene, synthesizes the DNA version of the RNA genome, and integrase integrates it back into the DNA of the host. This has linked ERVs with retroviruses and, in turn, shown that they are active and can propagate within the genome (Mager and Stoye, 2015). Most have been shown to influence both gene expression and genomic stability in mice and possibly immune response. The regulatory sequences in the LTRs could serve as alternative promoters or enhancers for genes near their landing sites, thereby modulating gene expression in a tissue-specific or developmental stage-specific manner. Indeed, the mouse mammary tumour virus (MMTV), an ERV, has been shown to express nearby oncogenes through its LTR in mice, leading to tumorigenesis in these animals (Ross, 2010). In some cancers and autoimmune diseases, this leads to overexpression of ERVs that contribute to the progress of the diseases (Stoye, 2012). The expression of ERVs in mouse models has been associated with triggering autoimmunity due to the expression of proteins recognized as foreign by the immune system, activating an immune response against the host (Krieg, Gourley and Perl, 1992).

Evolutionarily, ERVs are old elements that drove genome evolution. ERVs are retroviral infections integrated into the germline DNA of an ancestral species and, thus, can be transmitted across generations. Therefore, like in all other mammals, multiple ERV insertions over millions of years have occurred in mice, leading to a diverse array of ERV families within the genome. Analyses of phylogenetic trees based on sequence similarity and the time of integration reveal that ERVs may be divided into several families. Some of the better-characterized ERV families in mice include intracisternal A-particle elements (Ribet et al., 2008) and musD (Ribet et al., 2007). Host genomes have co-opted sequences derived from ERVs to serve novel functions, a process that is often referred to as exaptation. This led to regulatory elements being contributed from the ERV part of the genome to the host gene network in mice, acting on host gene expression patterns. Besides, the ERV sequences were also involved in another evolutionary scenario about mammals; the capture of a founding syncytin-like gene could have been instrumental in the transition from egg-laying to placental mammals (Dupressoir, Lavialle and Heidmann, 2012).

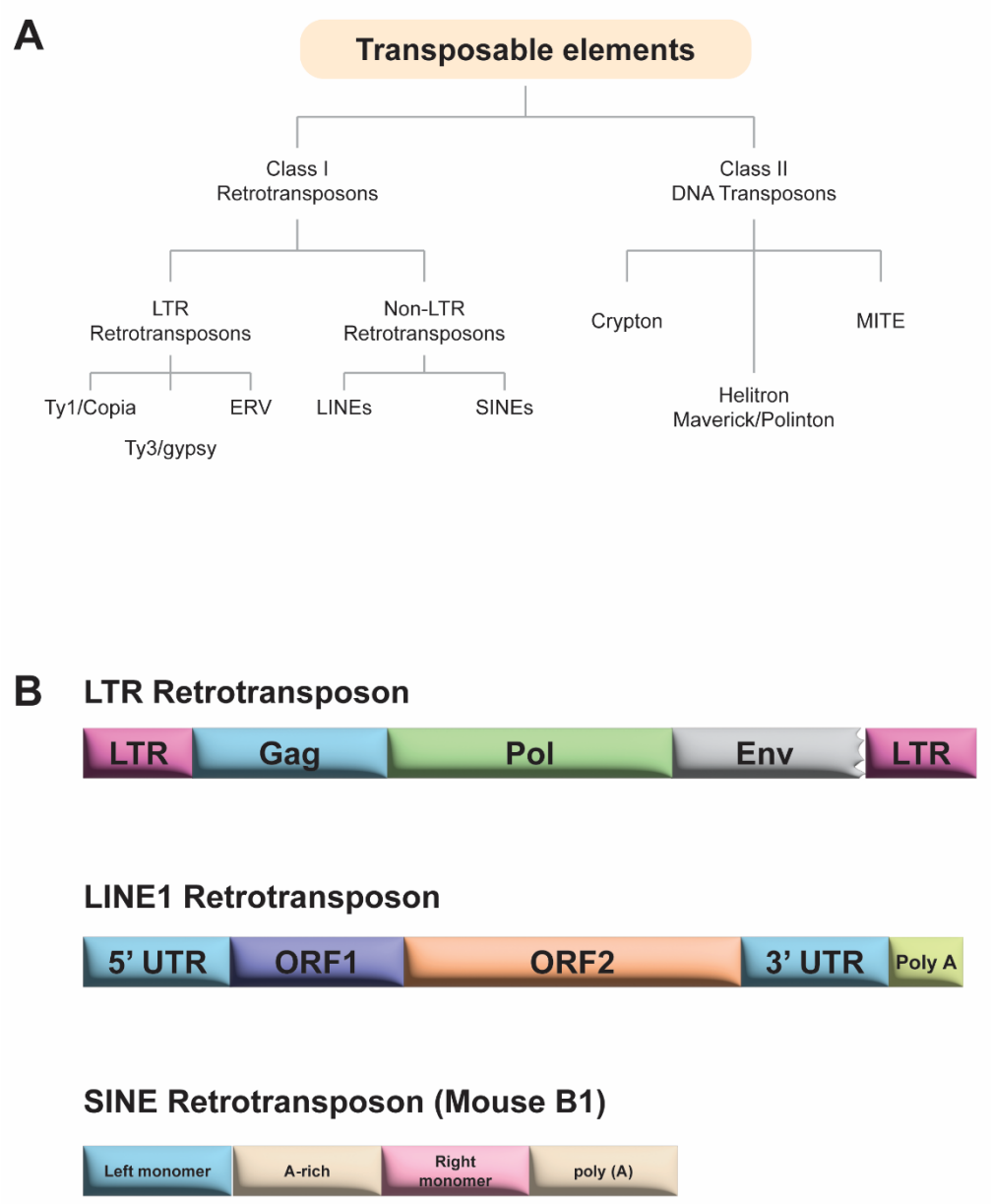


Figure 3.1 - **Transposons - Classification and Structure**. (A) Different classes of Transposons according to their mechanism of mobility. (B) Characteristic structure of Transposon DNA

3.1.3 Role of transposons in Mammalian development and evolution

In mice, transposons have varied functions from pre-implantation and embryonic development to pivotal evolutionary transitions, such as in the transition from oviparity to viviparity (murine endogenous retrovirus-L). In early mouse development, especially in pre-implantation, transposons like MERVL (murine endogenous retrovirus-L) and LINE1 play vital roles during zygotic genome activation (ZGA). ZGA refers to a critical step where the genome of a fertilized ovum becomes transcriptionally active, thereby passing from maternally deposited mRNAs into mRNAs generated from the newly formed zygotic genome. During the two-cell stage of embryogenesis, there is a transient expression of MERVL elements, which mediate expression patterns for ZGA genes essential for early embryonic development (Kigami et al., 2003). Similarly, LINE1 elements are also known to have a critical role in ZGA, given their abundance during the early stages of the embryonic development process. The LINE1 RNA transcripts are molecular scaffolds that aid in reorganizing chromatin, improving the accessibility of transcription machinery. The said scaffolding helps to bring about the expression of many genes associated with progress during early development (Percharde et al., 2018).

Transposable elements play an essential role in the mammalian immune system, hence contributing much to the genetic diversity and adaptability of host immune responses. A good example is their significance in developing antibody diversity through V(D)J recombination. This process generates a large number of different antibodies able to detect many diverse antigens specifically. Integration of TE increases the genetic diversity of recombination events associated with immunopathogenic effector cell antibodies (Kapitonov and Jurka, 2005).

TEs have been pivotal in major evolutionary transitions from egg-laying to placental mammals. TE-derived sequences are essential in modulating genes required for implantation and placental development within the endometrial cells comprising placental mammals. These have become regulatory elements that have been incorporated into the promoters and enhancers of critical genes, thus facilitating the correct gene expression regulation that is

required for a successful pregnancy (Bourque et al., 2008). The activity of evolutionarily young TEs may reflect an ongoing co-evolution that continues to impact mammalian development (Senft and Macfarlan, 2021).

3.1.4 Repression of Transposons

Despite the roles of potential positive roles of transposons discussed above, repression of these repeat elements is critical for maintaining genomic stability across various organisms. These mobile elements are silenced by different mechanisms in mice, *Drosophila* and *Caenorhabditis elegans*.

3.1.4.1 Transposon Repression in Mice

In mice, KRAB-ZFP proteins, histone modifications, DNA methylation and the piRNA pathway work together to repress transposons. Members of the KRAB-ZFP protein family, which consists of numerous zinc finger proteins, play a crucial role in recognizing and binding to transposable elements, followed by recruitment of the KAP1 co-repressor complex (Rowe et al., 2010). This leads to the formation of heterochromatin, which blocks transcriptionally active states, as explained in the following text. KAP1 also interacts with various chromatin-modifying enzymes, including SETDB1, which adds tri-methyl groups on H3K9me3 marks typical for silent chromatin.

Another crucial mechanism for transposon repression, especially in the germline, is the piRNA pathway (Aravin et al., 2006; Manakov et al., 2015). These small non-coding RNAs guide Piwi proteins to transposon transcripts, leading to their degradation and consequent silencing. In the mouse system, the piRNA pathway is essential for maintaining the integrity of the germline genome and preventing transposon mobilization (Siomi et al., 2011). This mechanism involves processing single-stranded long precursor RNAs into mature piRNAs associated with Piwi proteins, targeting transcripts of complementary transposons for degradation or transcriptional silencing by recruiting repressive chromatin marks.

Histone modifications are also important in repressing transposon activity. Apart from H3K9me3, other repressive marks, such as H3K27me3 mediated by Polycomb repressive

complex 2 (PRC2), are critical for silencing transposable elements. Such modifications compact chromatin, blocking access by transcriptional machinery, thereby inhibiting transposon activation (Bulut-Karslioglu et al., 2014).

In mice, DNA methylation is one of the most robust mechanisms to prevent transposon activity. Methylation of transposon promoters inhibits the binding of transcription factors while attracting proteins promoting heterochromatin formation. During early development, most DNA methylation patterns are reprogrammed; however, selective sequences like those of numerous transposons become methylated and remain silent (Bourc'his and Bestor, 2004; Walsh, Chaillet and Bestor, 1998).

3.1.4.2 Transposon Repression in *Drosophila*

In *Drosophila*, transposon repression primarily occurs through the piRNA pathway and heterochromatin formation. The *Drosophila* piRNA pathway has been well characterized and involves processing piRNAs from genomic piRNA clusters. These regions produce primary piRNAs that direct Piwi family proteins against transposon RNAs, leading either to degradation or transcriptional silencing (Brennecke et al., 2007; Sato and Siomi, 2020). This is essential for erasing transposable elements in germ cells and ensuring genome stability from one generation to another.

Heterochromatin formation also plays a significant role in repressing transposons. Su(var)3-9, a histone methyltransferase, catalyzes H3K9me2 and H3K9me3 modifications that are important for transcriptionally silent heterochromatin. This leads to chromatin compaction at these sites while at the same time inhibiting transcription, thereby silencing transposable elements (Elgin and Reuter, 2013).

3.1.4.3 Transposon Repression in *Caenorhabditis elegans*

Transposon repression in *Caenorhabditis elegans* involves the piRNA pathway, RNAi, and chromatin modifications. The piRNAs intersect with the RNAi pathway in *C. elegans* to regulate transposon and gene expression (Batista et al., 2008; Montgomery et al., 2021). RNAi serves as a potent mechanism for silencing transposons (Sijen and Plasterk, 2003). The

double-stranded RNA (dsRNA) processing by Dicer generates siRNAs that guide RISC in degrading transposon mRNAs.

Chromatin modifications also play a significant role in transposon repression in *C. elegans*. Histone methylation, especially H3K9me3 and deacetylation of histones, provides a repressive chromatin environment preventing transposon activation. Proteins like MET-2 or SET-25, which are histone methyltransferases responsible for catalyzing H3K9me3, are crucial for transposon silencing (Zeller et al., 2016).

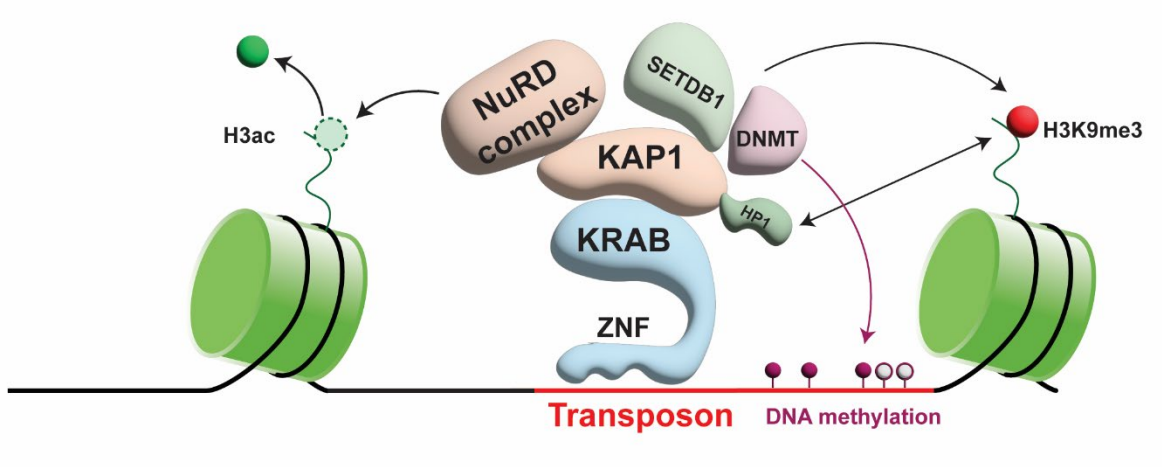


Figure 3.2: **Model of KRAB-ZFP binding to DNA and induction of heterochromatin formation**. KRAB-ZNFs recognize and bind to transposon DNA. KAP1 is recruited through the KRAB domain and interacts with the NuRD complex (which catalyzes the removal of H3ac) and histone methyltransferases like SETDB1 (which catalyzes the addition of H3K9me3). HP1 interacts with both KAP1 and H3K9me3. DNA methyltransferases (DNMTs) methylate genomic CpG sites, leading to inheritable silencing.

3.2 DNA METHYLATION

DNA methylation is a critical epigenetic mark in mammalian systems (Greenberg and Bourc'his, 2019). It is the most common modification found in DNA. DNA methylation is an epigenetic modification where a methyl group (CH_3) is added to the DNA molecule, typically at the 5' carbon of the cytosine base in a cytosine-guanine (CpG) dinucleotide sequence. In general, DNA methylation, when present at enhancers and promoters of genes, is typically associated with transcriptional repression; when present in the gene body, it is associated with transcriptional activity (Baubec and Schübeler, 2014; Smith and Meissner, 2013). It is a critical component in essential molecular processes like X-inactivation (Duncan et al., 2018; Sharp et al., 2011), germline gene inactivation in somatic cells (Borgel et al., 2010), Genomic imprinting (Li, Beard and Jaenisch, 1993) and transposon repression (Walsh, Chaillet and Bestor, 1998; Yoder, Walsh and Bestor, 1997). In this section, I explain how DNA methylation was discovered, the enzymes responsible for it, and how this epigenetic mark functions.

3.2.1 History of Discovery of DNA Methylation

The early 20th century was marked by the journey to discover DNA methylation. In 1925, Johnson and Coghill announced that a methylated nucleotide, 5-methylcytosine, existed within the DNA of *Mycobacterium tuberculosis* (Johnson and Coghill, 1925). This very early finding implied that deoxyribonucleic acid molecules could undergo chemical modifications. Further research during the 1940s verified the existence of 5-methylcytosine in higher organisms' DNA. Notably, Rollin Hotchkiss used paper chromatography to separate nucleotides from each other and discovered 5-methylcytosine content in calf thymus DNA samples (Hotchkiss, 1948). This fact enlightened people on the possibility that methylation was widespread among eukaryotes' DNA. Originally thought to exist only in mammals, later studies revealed that DNA methylation is also present in plants, invertebrates, and vertebrates. For example, research by Gruenbaum and colleagues showed that plant DNA has 5-methyl cytosine, meaning evolution has conserved it throughout different kingdoms (Gruenbaum et al., 1981). In bacteria, DNA methylation primarily occurs at adenine residues within specific sequences and serves as a

defense mechanism against foreign DNA through the restriction-modification system (Adhikari and Curtis, 2016).

Significant theoretical advances were made in the 1970s to establish a biological role for DNA methylation. Two models that laid down the basic concepts of methylation on DNA were independently formulated by Robin Holliday and John Pugh (1975) (Holliday and Pugh, 1975), and by Arthur Riggs (1975) (Riggs, 1975). Holliday and Pugh's model claimed that methylation at CpG dinucleotides could reversibly regulate gene expression. They also proposed that methylation patterns could be copied during DNA replication, ensuring maintenance and epigenetic inheritance. Besides, Riggs argued that there was potential for methylation to serve as a mechanism for regulating X-chromosome inactivation and genomic imprinting. Sager and Kitchin proposed that DNA methylation in eukaryotes could function as a Restriction-Modification system where methylation of a stretch of DNA could predispose it to be restricted by degradation or heterochromatinization. This would then lead to gene inactivation. Recombination events resulting from these double-strand breaks could provide the sequence mutations to regulate gene expression (Sager and Kitchin, 1975).

High-performance liquid chromatography (HPLC) in the 1980s enabled more precise quantitation of DNA methylation by researchers (Yuki et al., 1979). In vertebrates, DNA methylation was found to occur predominantly in CpG dinucleotides (Bird, 1986). In addition, these researchers highlighted an association between unmethylated sequences with abundant CpG content within commonly active genes and gene expression, implying that methylation might regulate gene expressions (Bird, 2002). At the end of the 1980s, the discovery of DNA methyltransferases (DNMTs, DNA MeTase) was a breakthrough that greatly enhanced the understanding of DNA methylation at a molecular level. The identification of DNMT1 as a maintenance methyltransferase that copies methylation patterns in hemimethylated DNA during replication highlighted the importance of this enzyme in maintaining epigenetic stability (Bestor et al., 1988; Bestor and Ingram, 1983). The discovery of DNMT3A and DNMT3B as de novo DNA methyltransferases in the late 1990s extended our knowledge on DNA methylation.

These enzymes have been documented to establish novel methylation patterns during differentiation processes under various environmental influences (Okano et al., 1999). Therefore, DNA methylation is dynamic and plays a vital role during the embryonic developmental period, particularly in cell differentiation.

Advancements were made in sequencing technologies from the late 20th century through the early 21st century, such as bisulfite and next-generation sequencing, which allowed for mapping genome-wide DNA methylation patterns. Consequently, novel insights were obtained regarding the complexity and variability within DNA methylation in different cell types and stages of development (Lister et al., 2009).

Today, DNA methylation is recognized as a critical player in epigenetic regulation, influencing gene expression, development, and genomic stability. Abnormal DNA methylation patterns are associated with genomic instability and various diseases, including cancer, highlighting the importance of understanding this modification's mechanisms and functions (Bouras et al., 2019; Esteller, 2008; Kulis and Esteller, 2010; Zhou and Robertson, 2016).

For further reading on the historical perspective of DNA methylation, please read this excellent review article by Alexandra L. Mattei, Nina Bailly, and Alexander Meissner - (Mattei, Bailly and Meissner, 2022).

3.2.2 DNA Methyltransferases - DNMTs

In mammals, DNA methyltransferase (DNMT) enzymes establish and maintain DNA methylation. They play essential roles in establishing and maintaining DNA methylation patterns, which are crucial for normal development, genomic stability, and cellular differentiation. Mice have four de novo DNMTs — DNMT3A, DNMT3B, DNMT3C, and DNMT3L, and a maintenance DNMT — DNMT1.

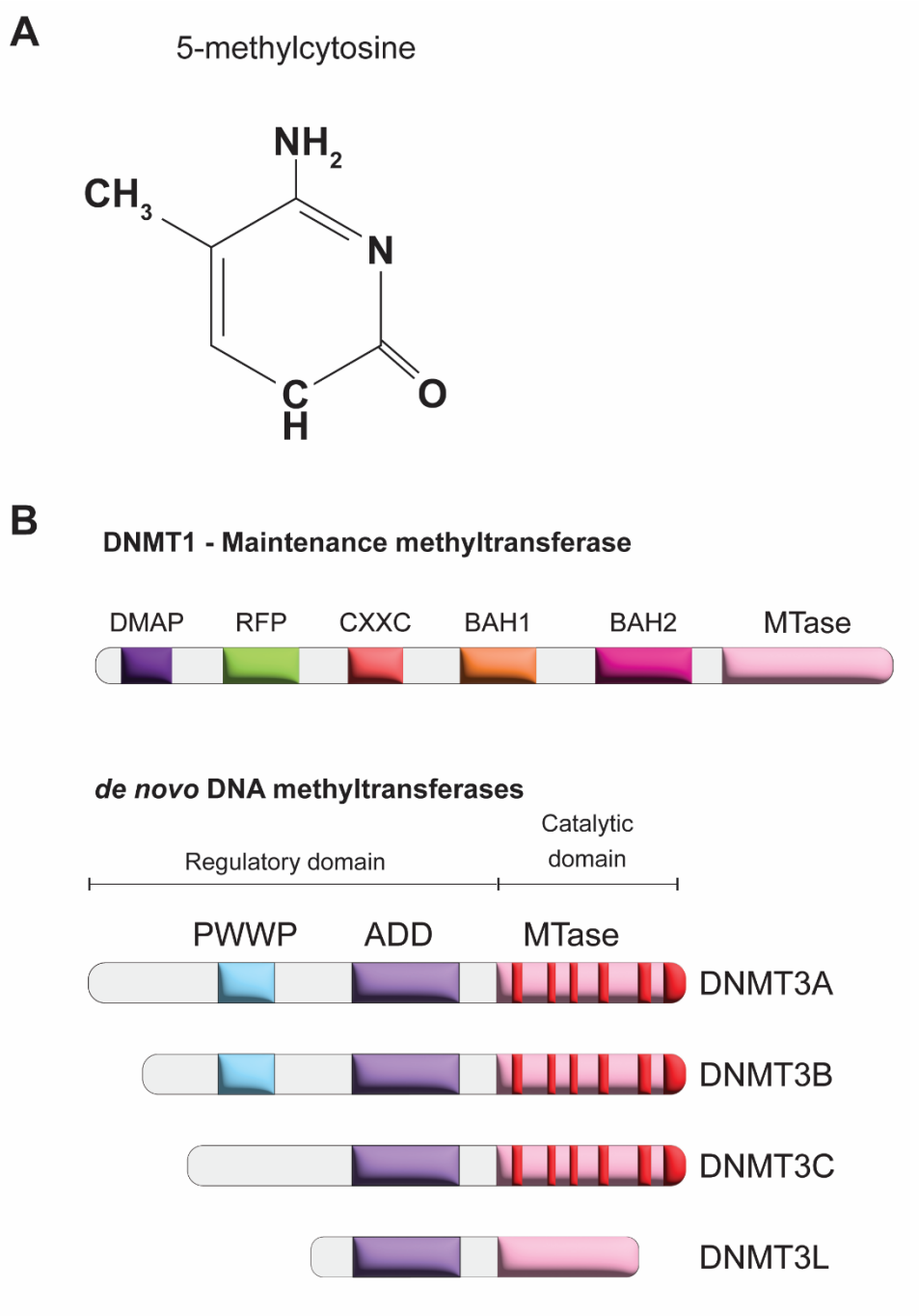


Figure 3.3: **DNA methylation and the enzymes that catalyze it**. (A) Structure of 5-methylcytosine. (B) Domain architecture of DNA methyltransferases – maintenance and *de novo*.

3.2.2.1 DNMT1

DNMT1, the first eukaryotic DNA methyltransferase discovered. The enzyme was initially characterized by its ability to methylate hemimethylated DNA strands, a process critical for maintaining methylation patterns during DNA replication (Bestor et al., 1988; Bestor, 1992; Bestor and Ingram, 1983).

DNMT1 is a large protein consisting of several functional domains. It contains a regulatory N-terminal domain, a replication foci targeting sequence (RFTS) domain, a CXXC domain, two bromo-adjacent homology (BAH) domains, and a C-terminal catalytic domain (Cheng and Blumenthal, 2008)The RFTS domain targets DNMT1 to replication foci during the S-phase, while the CXXC domain binds to unmethylated CpG dinucleotides, preventing de novo methylation. The BAH domains are involved in protein-protein interactions, and the catalytic domain carries out the methyltransferase activity by transferring a methyl group from the donor S-adenosylmethionine (SAM) to the cytosine ring of DNA (Jeltsch, 2006). Mechanistically, the RFTS domain of DNMT1 binds RFTS hemimethylated DNA while tethering the enzyme to it for selective catalysis (Pradhan et al., 1999). The protein undergoes specific structural changes such that the cytosine residues could undergo SAM-facilitated methylation.

Various isoforms of DNMT1 are formed resulting from alternative splicing. These isoforms vary in their regulatory and catalytic activities, contributing to the enzyme's versatility in different cellular contexts (Saini et al., 2017). For instance, somatic DNMT1s appear in dividing cells mostly, while oocyte-specific DNMT1o is necessary for maintaining methylation patterns during oogenesis up to early embryogenesis stages (Ratnam et al., 2002).

The primary role of DNMT1 is to preserve methylation patterns during DNA replication, which is vital in gene regulation, X chromosome inactivation, genomic imprinting and silencing transposons (Bird, 2002). Mitotic inheritance of DNA methylation is made possible by DNMT1-mediated methylation of hemimethylated DNA. A recent report showed that DNA methylation levels are not completely restored by mitosis. Their study revealed that methylation maintenance is outpaced by cell division in mouse embryonic stem cells. This (slower)

maintenance methylation could help explain the loss of DNA methylation over time throughout many cell divisions (like seen in cancer and aging cells) (Stewart-Morgan et al., 2023).

In addition to its classic function, DNMT1 involves itself in DNA repair processes and the organization of higher-order chromatin structures (Mortusewicz et al., 2005). Its functions diversify through interaction with multiple proteins and post-translational modifications that regulate its localization and activity (Estève et al., 2006; Estève et al., 2009). DNMT1 interacts with many different proteins that regulate and mediate its functions. DNMT1 forms a complex with UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), which essentially performs the function of locating DNMT1 at hemimethylated DNAs during DNA replication (Bostick et al., 2007). DNMT1 also interacts with histone-modifying enzymes like HDAC1 (histone deacetylase 1) and G9a, which marks chromatin modifications associated with repressive and inactive loci on chromosomes (Estève et al., 2006; Fuks et al., 2000).

DNMT1 activity can be regulated at different levels, including transcriptionally, post-transcriptionally or post-translationally. Transcriptional regulation is done through various transcription factors and signalling pathways that alter the expression of DNMT1 by different cellular signals (Spada et al., 2006). Examples of post-transcriptional effectors are alternative splicing and mRNA stability. However, post-translational modifiers like phosphorylation, among many others, influence DNMT1's stability, localization, and interaction with other proteins (Goll and Bestor, 2005).

The mouse models with *Dnmt1* mutations provide useful information about its function. DNMT1-null mice exhibit embryonic lethality, highlighting the enzyme's essential role in development (Li, Bestor and Jaenisch, 1992). Hypomorphic mutants, which have reduced DNMT1 activity, display a range of phenotypes, including developmental defects, immunodeficiency, and tumorigenesis (Gaudet et al., 2003). Mutations in the human *DNMT1* gene are also associated with several human diseases. One prominent example is autosomal dominant cerebellar ataxia, deafness, and narcolepsy (ADCA-DN), a neurodegenerative disorder caused by mutations in DNMT1 that disrupt its methylation activity (Klein et al., 2011).

These mutations have also led to specific changes in gene expression (including genes associated with clinical phenotypes of ADCA-DN) and DNA methylation patterns in different patient-derived cell types (Davis et al., 2023). As aberrant DNA methylation is closely associated with Parkinson's disease, *DNMT1* has been shown to play a role in modulating its susceptibility (Wang, J.-Y. et al., 2023; Zhang, H.-Q. et al., 2021).

3.2.2.2 DNMT3A

DNMT3A was discovered through homology-based cloning techniques aimed at identifying novel methyltransferases. It was identified as a novel *de novo* methyltransferase, responsible for establishing new methylation patterns during development (Okano, Xie and Li, 1998).

DNMT3A is a multi-domain protein characterized by several functional regions. It contains a PWWP domain, a Pro-Trp-Trp-Pro motif that plays a role in chromatin targeting and binding to histone H3 when methylated at lysine 36 (H3K36me3) (Dhayalan et al., 2010; Ge et al., 2004; Sendzikaite, 2021). Mutations in the PWWP domains have been shown to cause aberrant recruitment of DNMT3A (Weinberg et al., 2021). Another domain is the ADD domain (ATRX-DNMT3-DNMT3L), which binds to unmodified histone H3 at lysine 4 (H3K4) and thus prevents DNMT3A from methylating actively transcribed genes (Ooi et al., 2007; Otani et al., 2009). Mutations to the ADD domain of DNMT3A decrease global CG methylation levels in mice sperm and oocytes (Kubo et al., 2024; Uehara et al., 2023). The catalytic domain of DNMT3A is located at the C-terminus. It is responsible for transferring a methyl group to the 5th carbon of cytosine residues in CpG dinucleotides, a reaction facilitated by the donor S-adenosylmethionine. This catalytic domain shares structural similarity with bacterial DNA methyltransferases, indicating a conserved mechanism across species (Jurkowska, Jurkowski and Jeltsch, 2011).

DNMT3A exists in multiple isoforms, the most prominent being DNMT3A1 and DNMT3A2. DNMT3A1 is the full-length isoform that includes all functional domains, whereas DNMT3A2 is shorter and lacks the N-terminal 219 amino acids, including the PWWP domain.

The differential expression of these isoforms is developmentally regulated; DNMT3A2 is predominantly expressed in embryonic stem cells, while DNMT3A1 is more prevalent in somatic tissues (Chen et al., 2002).

DNMT3A is critical in shaping DNA methylation profiles during embryogenesis. Recent studies have also highlighted its role in the neural system, influencing neuronal differentiation and function (Lister et al., 2013). It has also been shown to be critical for Spermatogonial Stem Cells differentiation to further Spermatogenesis in mice (Dura et al., 2022).

DNMT3A is known to collaborate with histone methyltransferases, such as G9a and Suv39h1, and deacetylases to establish repressive chromatin environments (Fuks et al., 2001; Fuks et al., 2003). These interactions help coordinate the deposition of DNA methylation with histone modifications, such as H3K9 methylation, to ensure stable gene silencing. DNMT3A has also been shown to interact with LSD1, which enables it to methylate DNA by removing the antagonistic mark H3K4me (Petell et al., 2016).

Mutations in DNMT3A have profound effects in both mice and humans. In mice, knockout studies have shown that loss of DNMT3A leads to severe developmental defects, including impaired growth and abnormal differentiation of multiple tissues, ultimately resulting in embryonic lethality (Okano et al., 1999). Conditional knockouts in specific tissues have further illustrated its role in maintaining tissue-specific methylation patterns and normal cellular functions (Kaneda et al., 2004).

In humans, mutations in DNMT3A are linked to various diseases, including developmental disorders and cancers. Germline mutations in DNMT3A cause Tatton-Brown-Rahman syndrome, characterized by overgrowth, intellectual disability, and facial dysmorphisms (Smith et al., 2021; Tatton-Brown et al., 2014). Somatic mutations in DNMT3A are frequently observed in hematological malignancies, particularly Acute Myeloid Leukemia (AML), where they are associated with poor prognosis (Ley et al., 2010).

3.2.2.3 DNMT3B

DNMT3B [DNA (cytosine-5)-methyltransferase 3 beta] was discovered alongside DNMT3A (Okano et al., 1999). The discovery of DNMT3B was a milestone in understanding how methylation patterns are determined during embryonic development.

The domain architecture of DNMT3B is similar to that of DNMT3A. The N-terminal regulatory part contains a PWWP domain which recognizes histone modifications, especially H3K36me3 (Baubec et al., 2015; Dhayalan et al., 2010). The ADD (ATRX-DNMT3-DNMT3L) domain recognizes unmethylated H3K4, ensuring that DNMT3B activity is directed to appropriate chromatin states (Morselli et al., 2015; Ooi et al., 2007). The C-terminal hosts the catalytic domain with conserved motifs needed for methyltransferase activity.

Several DNMT3B isoforms exist owing to alternative splicing. Among the most studied isoforms are DNMT3B1 and DNMT3B2, which differ in their ability to interact with various cofactors and chromatin components, affecting their functional roles (Gujar, Weisenberger and Liang, 2019). A catalytically inactive isoform of DNMT3B, DNMT3B3, acts as an accessory protein for DNA methylation (Duymich et al., 2016; Weisenberger et al., 2004). For further reading about the different isoforms of DNMT3 proteins, please refer to this review article written by Hemant Gujar and colleagues - (Gujar, Weisenberger and Liang, 2019)

DNMT3B is recruited to the gene-body of actively transcribing genes through H3K36me3-PWWP interaction. DNMT3B-dependent methylation of these intragenic regions has been shown to prevent spurious transcription initiation from within the gene-body region (Neri et al., 2017). DNMT3B-mediated methylation is critical for differentiating Epiblast-like cells (EpiLCs) into meso-endodermal (ME) progenitors. *Dnmt3b*$^{KO/KO}$ redirect EpiLCs towards the neuro-ectodermal lineages (Lauria et al., 2023). This mutation leads to embryonic lethality in mice. DNMT3B is also essential for a major portion of the *de novo* methylation in early post-implantation mouse embryos (Borgel et al., 2010). Loss of *Dnmt3b* results in de-repression of germline genes in trophoblast lineages. A recent study performed conditional knockout of *Dnmt3b* in the growing embryo but not in the extra-embryonic lineages. This rescued the

embryonic lethality phenotype of *Dnmt3b* complete knockout, suggesting that DNMT3B activity during embryogenesis is primarily required to regulate placental development (Andrews et al., 2023). Interestingly, the catalytic activity of DNMT3B is dispensable for embryogenesis (Nowialis et al., 2019). These results suggest a catalytic-independent accessory function for the predominantly expressed isoform of DNMT3B, as shown for some other catalytically inactive splice variants (Gordon, Hartono and Chédin, 2013). This accessory function of DNMT3B could then enable other DNA methyltransferases to methylate.

Like DNMT3A, DNMT3B has also been shown to interact with G9a and HDAC, linking it with transcriptional repression (Epsztejn-Litman et al., 2008; Geiman et al., 2004). Mutations in the DNMT3B gene are linked to immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome in humans. Patients with ICF syndrome exhibit hypomethylation at pericentromeric regions, leading to genomic instability and various developmental abnormalities (Ehrlich et al., 2001; Xu et al., 1999). ICF1 is characterized by widespread hypomethylation of satellite repeats and other heterochromatic loci as well (Heyn et al., 2012; Jeanpierre et al., 1993). A recent report suggests that DNMT3B is recruited to constitutive H3K9me3-marked heterochromatin in a PWWP-independent manner. Mutations of the PWWP domain that disrupt its interaction with H3K36me3 increase the recruitment of DNMT3B to H3K9me3-marked heterochromatin. They also indicate that DNMT3B's N-terminal region facilitates this recruitment by interacting with HP1α (Taglini et al., 2024). This study also tested the ICF1-causing mutation DNMT3B-S270P and showed that S270P destabilizes DNMT3B, which explains why it causes hypomethylation of heterochromatin, unlike the other PWWP mutations they analyzed.

3.2.2.4 DNMT3C

DNMT3C is the newest member in the DNMT3 *de novo* DNA methyltransferases saga. DNMT3C was discovered independently by two groups in 2016/17 through an ENU mutagenesis screen in mice (Barau et al., 2016; Jain et al., 2017).

Dnmt3c is present in all of Muroidea but is absent in primates. *Dnmt3c* resulted from a gene duplication event from *Dnmt3b* ~60 million years ago in the last common ancestor of muroid rodents (Molaro, Malik and Bourc'his, 2020). Therefore, it shares a high level of pairwise identity with DNMT3B at the protein level (70.6%). DNMT3C possesses the ADD and catalytic domains. Both these domains show near-perfect identity with the same domains from DNMT3B. The central domains and C-terminus of DNMT3B and DNMT3C have undergone extensive gene conversion, leading to the homogenization of their sequences. A significant difference in DNMT3C is that it lacks a recognizable PWWP domain. The PWWP domain, as mentioned before, is involved in recruiting DNMT3 proteins to H3K36me2/3 domains. Lack of this domain would mean DNMT3C does not methylate the gene bodies of actively transcribing genes and might allow it to specialize in functions distinct from DNMT3B. Additionally, the N-terminal of DNMT3C, but not the N-termini of DNMT3A or DNMT3B, is subject to strong diversifying selection. This signifies that DNMT3C participates in an evolutionary arms race (Molaro and Malik, 2016). The signature of diversifying selection and loss of the PWWP domains make *Dnmt3C* unique among all *Dnmt3* genes. *Dnmt3C* seems to be the most divergent of all *Dnmt3* genes in muroid rodents based on the branch lengths of the DNMT3 phylogeny, followed by *Dnmt3L*, *Dnmt3B*, and finally, *Dnmt3A*, which is the most highly conserved (Molaro, Malik and Bourc'his, 2020).

Dnmt3c has two isoforms – the long and short isoforms. The catalytically active long isoform is expressed primarily in germ cells of the developing testes at embryonic day E15.5-P1 (1 day after birth) (Sangrithi et al., 2017). The short non-coding isoform is expressed in adult testes, and its functions are unknown (Barau et al., 2016). *Dnmt3c* is necessary for the meiotic progression of male germ cells. *Dnmt3c*$^{KO/KO}$ male mice are infertile due to hypogonadism caused by germ cell loss. This is because the germ cells fail to progress through the pachytene stage of meiosis. This interruption is caused by a sudden derepression of transposons, which causes a meiotic catastrophe. This phenotype is similar to the knockout of another member of the *Dnmt3* family *Dnmt3l* (Bourc'his and Bestor, 2004). Evolutionary young transposons belonging to the L1, ERV1, and ERVK families are derepressed in the

Dnmt3c$^{KO/KO}$ male meiotic cells. Whole genome bisulfite sequencing revealed that a failure in methylation of the promoter DNA of these transposons caused this derepression. Interestingly, there is little evidence that this derepression of transposons leads to a burst of retrotransposition. The main impact of TE derepression is chromatin-mediated. Unmethylated transposon promoters are marked by H3K4me3 (Nagamori et al., 2018). This enables the recruitment of early meiotic recombination machinery components like the SPO11 transesterase. SPO11-mediated double-strand breaks (DSB) occur at these TE sequences, making them meiotic recombination hotspots. The bonafide recombination hotspots in the mice genome are depleted in the meiotic recombination machinery, indicating that these components are redirected to the unmethylated transposon sequences. This causes an improper chromosome synapsis, which leads to germ cell loss (Zamudio et al., 2015). DNMT3C was ascribed to the function of methylating the promoter DNA of these transposons. Mutant mice of the piRNA pathway like *Piwil4*$^{KO/KO}$ (Carmell et al., 2007), *Tdrd9*$^{KO/KO}$ (Shoji et al., 2009), *Mael*$^{KO/KO}$ (Soper et al., 2008), *Spocd1*$^{KO/KO}$ (Zoch et al., 2020), *Tex15*$^{KO/KO}$ (Schöpp et al., 2020), to name a few, also show the same phenotype as *Dnmt3c*$^{KO/KO}$ mice. The piRNA pathway mutants also share the molecular signature of unmethylated transposon promoters, as seen in *Dnmt3c*$^{KO/KO}$. This suggests that DNMT3C could be part of the piRNA pathway. This potential association is one of the areas I studied in this project, and the outcome is explained in the Results section (6.2.9 – 6.2.12).

All mammalian genomes face a similar pressure of transposons, but only rodents have evolved a novel DNMT3 to tackle this. In humans, this phenomenon could be explained by a yet undiscovered DNMT or the assumption of DNMT3C's function by an existing DNMT3 protein. DNMT3C, as described earlier, has a signature of strong diversifying selection at its N-terminus. In primates, DNMT3A is subjected to similarly strong diversifying selection. This could posit that it is DNMT3A that methylates transposons. Because of DNMT3A's presence in all mammals, it could be the original DNMT3 protein that methylates transposons. *Dnmt3c* could have evolved to take up this function in Muroidea, thus helping explain the absence of positive selection of *Dnmt3a* in Muroidea (Molaro, Malik and Bourc'his, 2020). Another

explanation could be that an ancestral DNMT3B could have been the protein methylating transposons and other genomic loci. However, after the gene duplication event, the “Escape from adaptive conflict” model of sub-functionalization could have placed *Dnmt3b* and *Dnmt3c* towards distinct evolutionary paths (Des Marais and Rausher, 2008; Hittinger and Carroll, 2007). This could mean ancestral *Dnmt3b* was the original *Dnmt3* responsible for methylating transposons, at least in Muroidea. The sheer number of active transposon copies in rodents could have warranted the evolution of a dedicated DNMT3 protein. Examining the transposon number and the signature of diversifying/purifying selection of *Dnmt3* genes in other mammals could provide a better understanding of this topic.

3.2.2.5 DNMT3L

Dnmt3l was identified as a gene related to the DNA methyltransferase family due to its homology with other known DNA methyltransferases, such as *Dnmt3a* or *Dnmt3b*. The discovery of DNMT3L was significant because it highlighted a non-catalytic role in regulating DNA methylation processes, particularly in the germline. DNMT3L was initially discovered by studying gene expression patterns in early embryonic development and germ cells, where it was found to be highly expressed (Bourc’his and Bestor, 2004).

DNMT3L, in contrast to active DNA methyltransferases, does not have a PWWP or a functional catalytic domain. It is almost half the size of other DNMT3 proteins. It is only equipped with an ADD domain. Unlike its catalytic counterparts, DNMT3L has only one isoform. DNMT3L mainly stimulates the activity of methyltransferases DNMT3A/B – it acts as their cofactor. This enhancement is crucial during germ cell development and early embryogenesis, where de novo methylation patterns are established. DNMT3L achieves this by attaching itself to the ADD domains situated on DNMT3A and DNMT3B, thus ensuring that they remain on chromatin and carry out their catalytic functions correctly.

Genomic imprinting is an epigenetic phenomenon where specific genes are expressed in a parent-of-origin-specific manner. This process involves the differential methylation of the allele inherited from the mother versus the one from the father, leading to the expression of

only one allele while the other is silenced. DNMT3L plays a critical role in establishing these genomic imprints during gametogenesis in mice. In male mice, DNMT3L is essential for the methylation of imprinted genes in spermatogenesis. It acts in conjunction with DNMT3A to establish paternal imprints during sperm development (Bourc'his et al., 2001). DNMT3L is also essential for establishing maternal imprints during oogenesis. The absence of DNMT3L leads to the failure of de novo methylation of imprinted genes in oocytes, which can result in severe developmental defects in the offspring (Kaneda et al., 2004). Mutations in *Dnmt3l* in mice lead to severe developmental defects, particularly in the germline. *Dnmt3l*$^{KO/KO}$ male and female mice are sterile. The *Dnmt3l*$^{KO/KO}$ males are sterile due to the derepression of transposons during meiosis, leading to germ cell loss (as seen in *Dnmt3c*$^{KO/KO}$ males) (Bourc'his and Bestor, 2004). Female mice also exhibit fertility issues. The heterozygous (*Dnmt3l*$^{KO/WT}$) progeny of homozygous females (*Dnmt3l*$^{KO/KO}$) died before midgestation (Bourc'his et al., 2001).

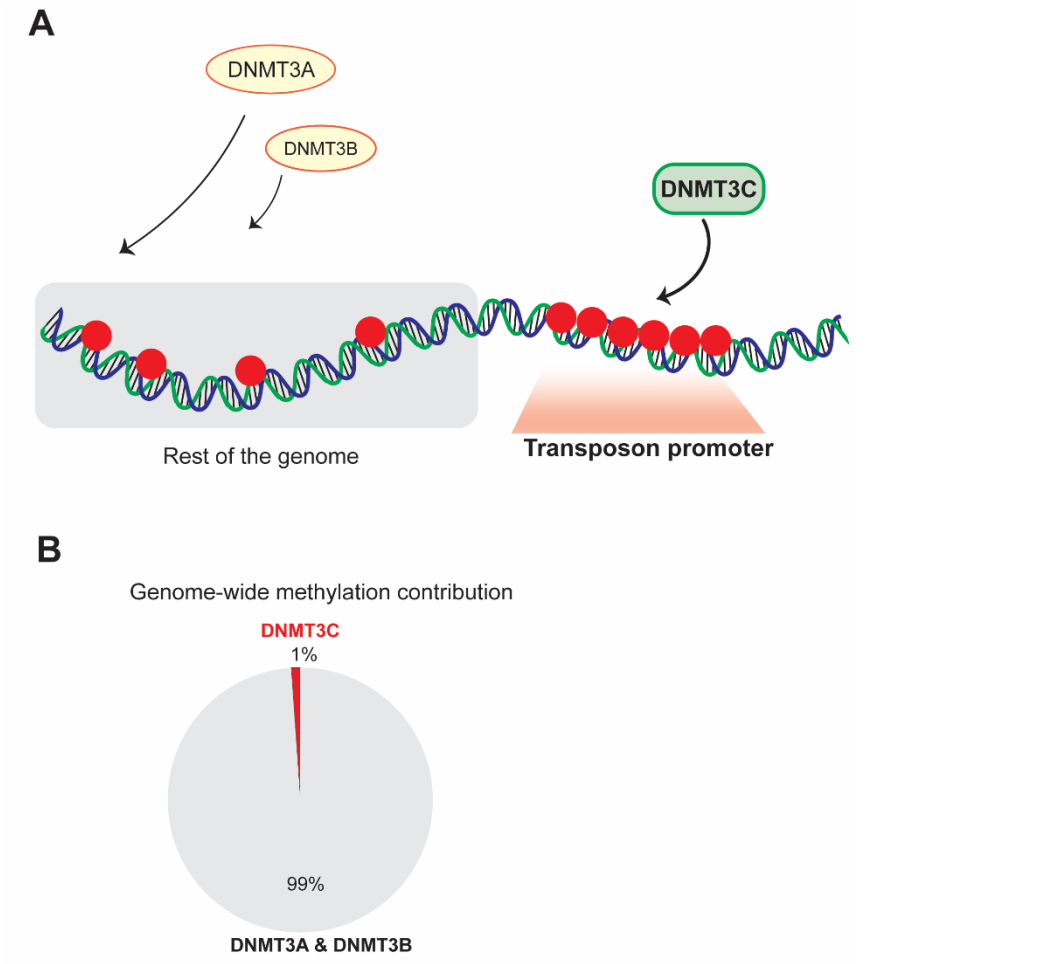


Figure 3.4: **Genome-wide contribution of DNA methylation by DNMT3 proteins.**

3.3 MOUSE MALE GERMLINE DEVELOPMENT

3.3.1 Germline Specification and Migration

Germline cell lineage is a unique cell lineage that produces gametes—sperm in males and oocytes in females. The specification of the germline begins during early embryogenesis. Primordial germ cells (PGCs), the precursors of germ cells, are first identified as a cluster of approximately 40 cells in the proximal epiblast adjacent to the extraembryonic ectoderm at embryonic day E6.5 (Saitou, Barton and Surani, 2002). PGC specification requires a unique gene expression program, which includes the upregulation of key transcription factors such as *Prdm1* (*Blimp1*), *Prdm14* and *Ap2γ* (Magnúsdóttir et al., 2013; Ohinata et al., 2005). These factors repress the somatic program and initiate the germline transcription program. Transcription factors like *Oct4*, *Sox2* and *Nanog* are also essential to maintain the pluripotency and germline identity of PGCs (Kehler et al., 2004; Yamaguchi et al., 2009).

Extra- and intracellular signals drive the specification of PGCs. Bone morphogenetic proteins (BMPs) from the extraembryonic ectoderm play a crucial role in generation of these PGCs (Ying et al., 2000; Ying and Zhao, 2001). Loss of *Prdm1* or *Bmp4* results in a significant loss of PGC number (Jan et al., 2012; Lawson et al., 1999). After specification, PGCs migrate from the base of the allantois to the developing gonads. Initially, they move passively through the hindgut endoderm and then actively migrate through the dorsal mesentery to reach the genital ridge by stage E10.5 (Anderson et al., 2000). In this journey, PGCs interact with various extracellular matrix (ECM) components (for example, fibronectin). These provide structural support and guidance cues to the migratory PGCs. A signalling molecule, SDF1 (a chemoattractant), is expressed in the gonads and attracts the PGCs by interacting with the cell surface receptor CXCR4 (Ara et al., 2003; Molyneaux et al., 2003). During migration, PGCs proliferate extensively, increasing from approximately 100 cells at E7.5 to around 25,000 by E13.5 (Tam and Snow, 1981). Studies have shown that Steel factor (also known as KIT ligand) and its receptor c-KIT play a pivotal role in supporting PGC survival and proliferation during migration (Runyan et al., 2006).

3.3.2 Prospermatogonia

Sex determination happens at E11.0-E11.25 through the expression of the *Sry* gene (Hiramatsu et al., 2009; Kashimada and Koopman, 2010). After reaching the genital ridges, PGCs continue to proliferate. Eventually, they transition into a quiescent state known as Prospermatogonia. Prospermatogonia, also known as Gonocytes, are the immediate precursors to Spermatogonia. They are characterized by their large, round nuclei, prominent nucleoli, and distinct chromatin organization. During this stage, germ cells undergo mitotic arrest in the G0/G1 phase (between E12.5 and E14.5), temporarily stopping their cell division. Significant changes in gene expression and cellular behavior mark the transition from proliferating germ cells to Prospermatogonia. They remain quiescent in the testes until shortly after birth (P2 – Day 2 postpartum), when they resume proliferation and differentiate into Spermatogonia (Kluin and Rooij, 1981; Vergouwen et al., 1991). Several factors have been implicated in the molecular mechanisms governing the transition to mitotic arrest. Between E11.5 and E12.5, the PGCs in the male genital ridge respond to signals from the masculinizing environment and commit to Spermatogenesis (Adams and McLaren, 2002). This signal causes the upregulation of $p27^{Kip1}$ and $p15^{INK4b}$ genes. These two genes are known to mediate cell cycle arrest. These proteins then inhibit CyclinE-cdk2 and CyclinD-cdk4/6, respectively, preventing hyperphosphorylation of pRB (Retinoblastoma protein) in arresting germ cells. By E14.5, hypo phosphorylated pRB activates the G1-S checkpoint across the germ cell population and mediate cell cycle arrest (Cobrinik, 2005; Western et al., 2008). Activin, TGF Beta 1 and TGF Beta 2 have also been reported to inhibit the proliferation of PGCs (Godin and Wylie, 1991; Moreno et al., 2010; Richards, Enders and Resnick, 1999). TGF Beta 1 has been shown to upregulate $p27^{Kip1}$ in mouse B cells (Kamesaki et al., 1998). In human epithelial cells, TGF-β1 stabilizes the p15INK4B-cdk4 complexes and subsequently increases cyclin E-cdk2-associated p27KIP1 (Sandhu et al., 1997). These mechanisms help explain the cell cycle arrest seen in PGCs.

During the transition to Prospermatogonia, germ cells undergo extensive epigenetic reprogramming, which involves changes in DNA methylation, histone modifications, and

chromatin remodeling. This reprogramming resets the epigenetic landscape of germ cells, establishing a unique transcriptional program essential for their differentiation and development (Seisenberger et al., 2012). This is explained more in Section 3.4.

3.3.3 Spermatogonial Stem Cells and Spermatogonia

After birth, around P3, Prospermatogonia exits its mitotic arrest phase and differentiates into Spermatogonial stem cells (SSCs). The SSCs are the foundation for Spermatogenesis. SSCs are confined to the basal compartments of seminiferous tubules and have the ability to carry out both self-renewal and differentiation. This ensures a continuous supply of germ cells necessary for spermatogenesis throughout life (Rooij and Russell, 2000). Differentiation of Spermatogonia originates at the basal lamina of the seminiferous epithelium towards the lumen of the tubule, and runs throughout the length of the tubule. SSCs are identified by the expression of specific markers including GFRα1 (GDNF family receptor alpha 1), ID4, PAX7, BCL6B, ETV5 (Fayomi and Orwig, 2018; Jan et al., 2012; Lord and Oatley, 2017; Rooij, 2017; Song and Wilkinson, 2014). In order to conserve themselves and produce progenitors, SSCs undergo self-renewal divisions. This is regulated by factors from the SSC niche, mainly Sertoli cells, which secrete GDNF. GDNF interacts with GFRα1 on SSCs, thus stimulating the downstream pathways such as PI3K/Akt and MAPK, which promote SSC survival and proliferation (Braydich-Stolle et al., 2005). Under differentiation signals, SSCs turn into undifferentiated Spermatogonia. Key transcription factors like PLZF and SOHLH1 (Spermatogenesis and oogenesis-specific basic helix–loop–helix 1) are essential for this differentiation (Ballow et al., 2006).

Spermatogonia are the differentiated progeny of SSCs. They are of three types: undifferentiated Spermatogonia (A_single, A_paired, and A_aligned), differentiating Spermatogonia (A1 to A4, Intermediate), and type B Spermatogonia. This classification is based on their morphological characteristics and their position within the seminiferous tubules (Rooij and Russell, 2000).

A_single (A_s) undifferentiated Spermatogonia are isolated cells while A_paired (A_{pr}) undifferentiated Spermatogonia are connected by cytoplasmic bridges, and finally, A_aligned (A_{al}) undifferentiated Spermatogonia are series of chains of interconnected cells. During this stage, Spermatogonia can either self-renew to replenish the SSC pool or differentiate into type A1 Spermatogonia under the influence of retinoic acid signalling (Koubova et al., 2006; Yoshida et al., 2006). The retinoic acid is synthesized by the Sertoli cells and is generated in a pulse (Gewiss, Topping and Griswold, 2019). Type A1 spermatogonia is characterized by the expression of retinoic acid-responsive genes such as cKit and Stra8. Mutation of the *cKit* locus or mutation of its ligand SCF causes a block in the differentiation of A_{al} into A1 Spermatogonia (Jan et al., 2012; Song and Wilkinson, 2014). These genes are essential for initiating differentiation and meiosis (Ikami et al., 2015). Subsequently, multiple mitotic divisions occur, resulting in the production of A2, A3, and A4 differentiated Spermatogonia. Multiple other genes also follow a similar expression pattern to cKit, including Sohlh1, Sohlh2, Cyclin D2, Stra8, and Dmrt1, and are therefore considered to be markers of differentiated Spermatogonia (Green et al., 2018; Hermann et al., 2018; Song and Wilkinson, 2014). Type A4 Spermatogonia then differentiates into a transitional stage called Intermediate Spermatogonia. They display proliferation and differentiation markers such as Ngn3 (Neurog3) and Ccnd2 (Yoshida et al., 2006). After undergoing further rounds of mitotic divisions, the intermediate Spermatogonia differentiate into Type B Spermatogonia. Type B Spermatogonia are characterized by the expression of essential genes such as Dmrt1 (Double sex and Mab-3 related transcription factor 1) and Stra8 (Matson et al., 2010). Another round of mitosis produces the Primary Spermatocytes. Primary spermatocytes are the diploid cells that enter into meiosis. This marks the end of the mitosis phase and the onset of meiosis.

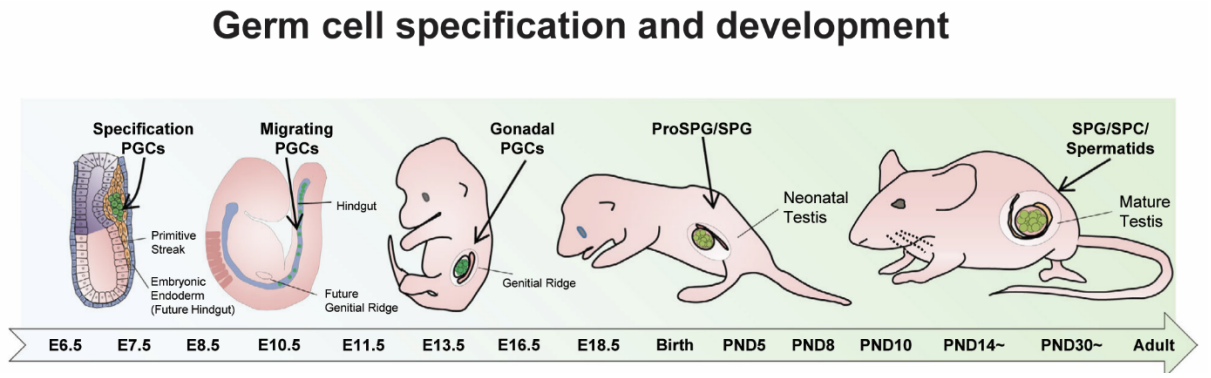


Figure 3.5 **Germ cell specification and development during mice embryonic development**. [Extracted from (Zhao et al., 2021)]. Primordial Germ cells are specified at E6.5 in the epiblast. Post specification, they migrate towards the future gonad as migratory PGCs. After migration, they populate the gonads and differentiate according to their sex. Male germ cells become Prospermatogonia and undergo epigenetic reprogramming to activate the germline transcriptional program. After birth, Prospermatogonia differentiate into Spermatogonial Stem cells (SSC). SSC forms the basis of spermatogenesis and produces mature sperm.

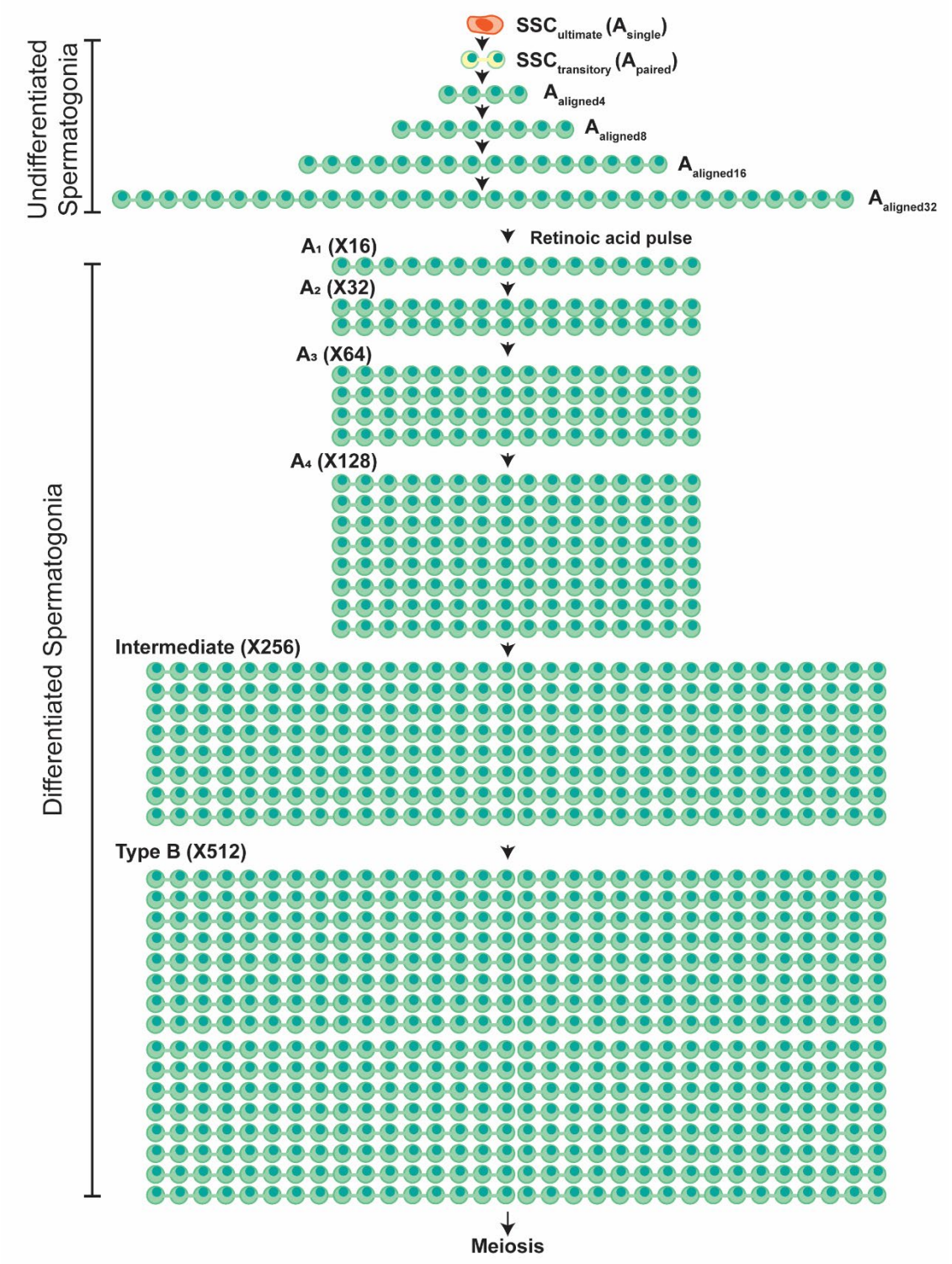


Figure 3.6: **Mitotic phase of spermatogenesis**. A_{single} (A_s) differentiates into A_{paired} (A_{pr}), which differentiates into $A_{aligned}$ (A_{al}). A_s, A_{pr} and A_{al} form the pool of undifferentiated Spermatogonia. A_{pr} and A_{al} can differentiate into A1 differentiated Spermatogonia under retinoic acid signal. A chain of 16 A_{al} undifferentiated Spermatogonia can give rise to up to 512 Type B Spermatogonia type B, which undergo meiosis to produce a total of 4,096 Spermatozoa.

3.3.4 Spermatocytes and Meiosis

Spermatocytes are the germ cells that undergo meiosis to produce haploid spermatids. Primary spermatocytes are the first germ cells to enter meiosis, followed by secondary spermatocytes after the completion of meiosis I. These secondary spermatocytes go through meiosis II at a faster pace to produce haploid spermatids (Griswold, 2016). Primary spermatocytes are characterized by large nuclei with synaptonemal complexes between homologous chromosomes during meiosis. Secondary spermatocytes, on the other hand, are smaller and have more diffuse chromatin material inside them, hence entering meiosis II for their quick completion thus, eventually resulting in gametes formation (Griswold, 2016).

Meiosis is a specialized type of cell division that produces four genetically distinct haploid cells from one diploid parent cell. It has two consecutive stages, meiosis I and meiosis II, respectively, which include prophase, metaphase, anaphase, and telophase stages.

Meiosis I

Prophase I: This is a prolonged and complex stage, subdivided into leptotene, zygotene, pachytene, diplotene, and diakinesis. Distinct chromosomal behaviors and structural formations mark each substage.

Leptotene: During the leptotene stage, chromosomes condense and become visible under a microscope. Each chromosome, initially appearing as a thin thread, starts to replicate, forming sister chromatids connected by cohesin proteins (Zickler and Kleckner, 2015). The beginning of SPO11 (meiotic topoisomerase) catalyzed DNA double-strand breaks, which are a precursor to homologous recombination, is what characterizes this phase.

Zygotene: Zygotene is a phase characterized by pairing homologous chromosomes through synapsis. SYCP1, SYCP2, SYCP3, and TEX12 proteins constitute the synaptonemal complex, which enables synapsis (Fraune et al., 2012). It is necessary for the proper positioning followed by genetic recombination of these homologous chromosomes.

Pachytene: During this phase, chromosomes become fully aligned. It is at this stage that homologous recombination or crossing over takes place, as there is an exchange of genetic material between the non-sister chromatids (Handel and Schimenti, 2010). This exchange of genetic material produces genetic diversity in the progeny. The synaptonemal complex remains intact at this stage, condensing the chromosomes and making them more distinguishable.

Diplotene: During the diplotene stage, the synaptonemal complex dissolves, and homologous chromosomes begin to separate, except at the chiasmata, where crossing over occurs. This process is called de-synapsis. The chromosomes decondense slightly, allowing for transcriptional activity that is essential for the progression of meiosis (Kleckner, 2006). This stage is also marked by increased chromosomal movement and preparation for the final stages of meiosis I.

Metaphase I: In metaphase I, the homologous chromosome pairs align along the metaphase plate. Spindle fibers from opposite poles attach to the centromeres of each homologous chromosome pair, ensuring that each daughter cell will receive one chromosome from each pair (Handel and Schimenti, 2010).

Anaphase I: During anaphase I, the homologous chromosomes are pulled apart and move to opposite poles of the cell. This separation reduces the chromosome number by half, a key feature of meiosis I. Unlike mitosis, the sister chromatids remain attached at their centromeres (Handel and Schimenti, 2010).

Telophase I and Cytokinesis: In telophase I, the chromosomes arrive at the poles, and the cell undergoes cytokinesis, resulting in two haploid daughter cells, each with half the original chromosome number but consisting of two sister chromatids. This sets the stage for meiosis II (Handel and Schimenti, 2010).

Meiosis II

Meiosis II resembles a typical mitotic division and consists of prophase II, metaphase II, anaphase II, and telophase II. In prophase II, the nuclear envelope breaks down, and a new spindle apparatus forms in each haploid cell. During metaphase II, chromosomes align along the metaphase plate. In anaphase II, the sister chromatids are finally separated and pulled to opposite poles. Telophase II and cytokinesis result in four haploid spermatids, each genetically distinct from one another and the original diploid cell (Handel and Schimenti, 2010).

3.3.5 Spermiogenesis

Spermiogenesis is the last leg of spermatogenesis when the haploid spermatids undergo morphological and physiological changes that turn them into mature spermatozoa. It is subdivided into four main phases: the Golgi phase, the cap phase, the acrosome phase, and the maturation phase.

Golgi Phase: Spermiogenesis begins with the Golgi phase. This phase is characterized by acrosome development. The Golgi apparatus produces proacrosomal granules, which combine to form a large acrosomal vesicle. This vesicle forms a cap-like structure at the anterior part of the nucleus and consists of digestive enzymes required for egg penetration during fertilization (O'Donnell, 2014). Correct acrosome positioning is critical as it dictates the future anterior-posterior axis of the sperm cell. During this phase, centrioles move to the other side of the nucleus, forming the flagellum, a tail-like structure that propels sperm forward (CLERMONT, 1972).

Cap Phase: The acrosomal vesicle spreads over the anterior half of the nucleus, forming a distinct cap. It also leads to further growth and compaction of acrosome as well pronounced alterations in the nucleus. The chromatin starts condensing, leading to highly condensed transcriptionally inactive nuclear material. Histones are replaced by protamines (arginine-rich small proteins that enable DNA to be packed more tightly) (Dadoune, 2003). This transition is essential for safeguarding genetic material through male and female reproductive

tracts. The axoneme that serves as the core structure of microtubules in the flagellum continues to elongate and grow.

Acrosome Phase: In this phase, the spermatid undergoes further elongation and reshaping. The acrosome continues to mature by covering more parts of the nucleus. Simultaneously, the nucleus becomes more condensed and elongated. Simultaneously, a transient microtubular structure, called the Manchette, is formed around this nucleus which helps to shape the head of the spermatozoan using cytoskeletal forces (O'Donnell, 2014). The mitochondria begin to accumulate at the proximal end of the flagellum resulting in the formation of the mid-piece that produces energy for spermatozoon. These organelles are helically arranged around the axoneme and hence provide ATP that is required for flagellar beating as well as motility (CLERMONT, 1972).

Maturation phase: The final maturation phase sees cytoplasmic extrusion, called the residual body. These residual bodies are engulfed by Sertoli cells. Reducing excess cytoplasm amounts becomes important since this eases the process of streamlining sperm, making it hydrodynamic and easy to progress (Russell et al., 1993). It consists of three parts: the head contains a dense nucleus that helps in fertilizing oocytes, while the mid-piece packs lots of mitochondria, providing the energy needed during the swimming of sperms through the female genital tract, whereas the tail is made up of axoneme and other structures associated with it, allows movement across these regions (O'Donnell, 2014).

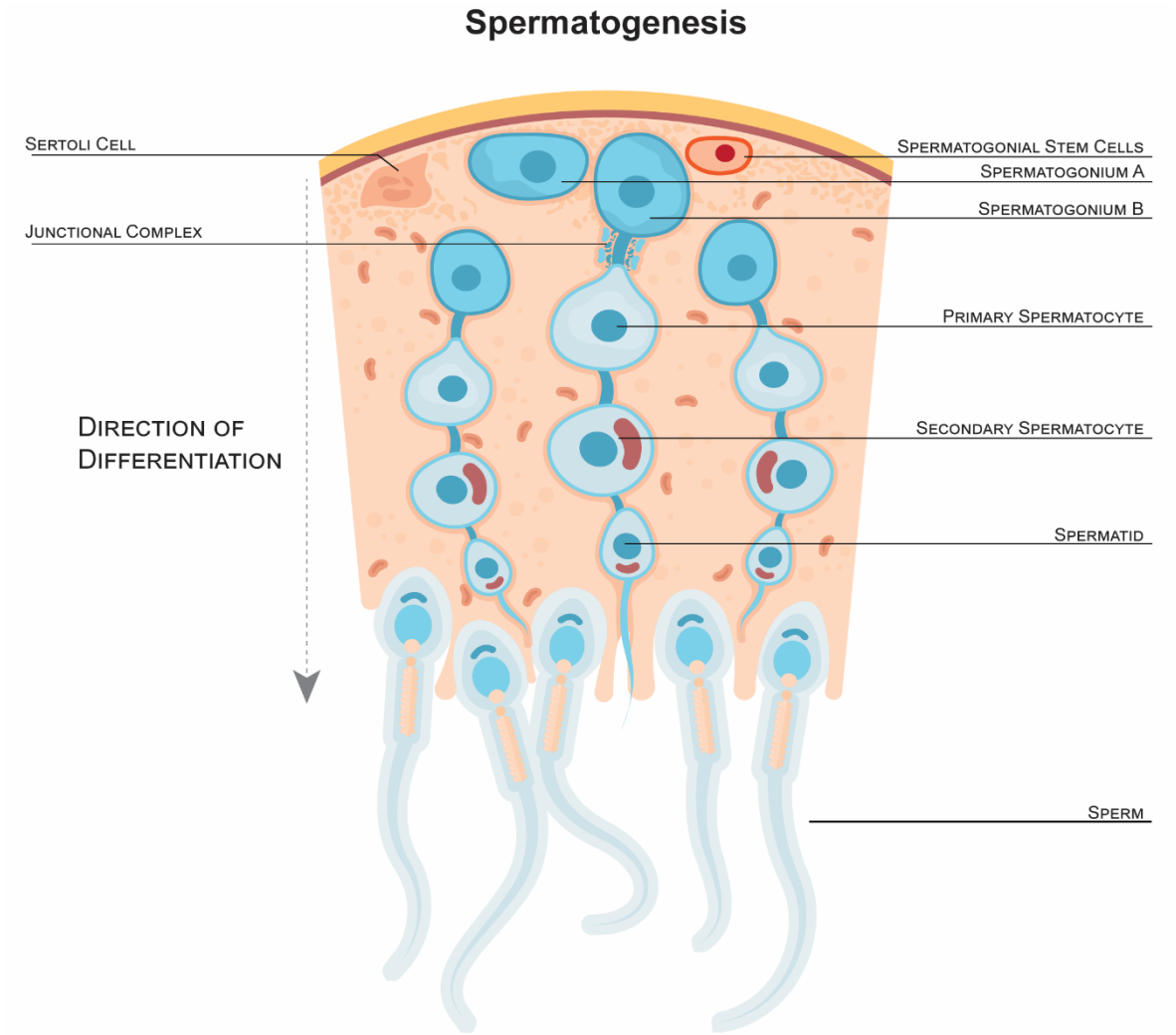


Figure 3.7 **Cartoon of a cross-section of a seminiferous tubule showing the progression of Spermatogenesis**. The differentiation of male germ cells (spermatogenesis) occurs from the basal membrane of the seminiferous tubule to the lumen

3.4 MALE GERMLINE WAVE OF EPIGENETIC REPROGRAMMING

The embryo undergoes two waves of global DNA methylation erasure and *de novo* methylation during its development. The first wave occurs in all the cells of the pre-implantation embryo right after fertilization; therefore, this is called the Embryonic wave of epigenetic reprogramming. The second wave occurs only in germ cells; therefore, it is called the Germline wave of epigenetic reprogramming (Lees-Murdock and Walsh, 2008). The Embryonic wave helps develop the embryo through totipotency from two terminally differentiated cell types (sperm and the egg), and the Germline wave helps develop the germline cells from their somatic precursors. Germline reprogramming is essential for the expression of germline genes (which are inactivated in the somatic cells), establishing a proper parent-of-origin transcriptional program (including Genomic Imprinting), and repression of transposable elements.

3.4.1 DNA demethylation of PGCs

Global DNA demethylation in PGC occurs in two steps: Passive (E6.5 – E9.5) and active demethylation (E9.5 – E13.5) (Hill, Amouroux and Hajkova, 2014; Kobayashi et al., 2013; Ohno et al., 2013; Wu and Zhang, 2017).

At the time of specification, PGCs DNA methylation levels are similar to their somatic neighbors (~80%). After their specification, PGCs proliferate extensively and migrate towards the future gonad. The passive dilution of DNA methylation is the primary mechanism of PGCs losing their DNA methylation. This happens through the impairment of the maintenance of DNA methylation by excluding DNMT1-UHRF1 from the nucleus. The *de novo* methylation enzymes DNMT3A and DNMT3B are also downregulated (Kurimoto et al., 2008). The bulk of the genomic DNA is demethylated by E9.5, and the global DNA methylation level falls from ~80% to ~30% (Kagiwada et al., 2013; Seisenberger et al., 2012). However, some regions retain DNA methylation, including imprinting control regions (ICRs), meiotic gene promoters, X chromosome-linked promoters, and IAP ERVs. It has been shown that DNMT1 protects these regions from demethylation as they lose their DNA methylation upon conditional *Dnmt1*

deletion in the PGCs. Also, *Dnmt1*-cKO causes female PGCs to prematurely enter meiosis, potentially through the upregulation of *Dazl* (Hargan-Calvopina et al., 2016). Interestingly, IAPs are not derepressed in the *Dnmt1*-cKO PGCs. This could be because IAPs, unlike the meiotic gene promoters and ICRs, are highly enriched in H3K9me3 and H3K27me3 repressive histone marks. Loss of H3K9me3 on IAPs led to a loss of DNA methylation and H3K27me3, suggesting that resistance to DNA demethylation of IAPs could be due to H3K9me3 (Liu et al., 2014).

The TET enzymes actively catalyze the second phase of demethylation (Hajkova et al., 2010). TET enzymes are ten-eleven translocation (TET) methylcytosine dioxygenases. TETs catalyze the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), and then the oxidation of 5hmC to 5-formylcytosine (5fC) and 5fC to 5-carboxylcytosine (5caC) (Ito et al., 2011). TET1 is the primary enzyme responsible for the demethylation as *Tet1*$^{KO/KO}$ germ cells showed an 85% reduction in the observed 5hmC (5-hydroxy methylcytosine) levels. By E13.5, *Tet1*$^{KO/KO}$ and WT germ cells reached near-complete depletion of 5mC levels, suggesting a redundant function of other TET enzymes like TET2. The specific regions resistant to passive demethylation, including ICRs and meiotic gene promoters, were demethylated at E13.5 by TET1 (Hackett et al., 2013). TET1 also protects germline reprogramming-responsive genes (GRR) when they are already unmethylated. GRR genes are a list of 45 genes that became demethylated during germline epigenetic reprogramming and showed progressive transcriptional activation in males and females. Erroneous DNA methylation of these GRR genes led to defects in germline development. TET1 has also been shown to bind to the promoters of GRR genes and activate their expression, as their upregulation is considerably lower in *Tet1*$^{KO/KO}$ PGCs than in wild-type PGCs. Interestingly, TET1 absence does not affect the DNA demethylation of these genes, suggesting that TET1 has functions independent of its catalytic activity (Hill et al., 2018). The PGC genome reaches a minimum of 5-7% of genome-wide DNA methylation at E13.5, and the only sequences that remain significantly methylated are IAPs and some other ERVs (Guibert, Forné and Weber, 2012; Saitou, Kagiwada and Kurimoto, 2012; Saitou and Yamaji, 2012). This global erasure of DNA methylation in PGCs is also conserved in humans, indicating that germline reprogramming is a general characteristic

of mammalian development (Gkountela et al., 2015; Guo, F. et al., 2015; Tang et al., 2015; Tang et al., 2016).

DNA damage responses have also been shown to play a role in germline development (Bloom and Schimenti, 2020; Hajkova et al., 2010). PolyADP-ribosylation (PARP) inhibition impairs genome-wide and locus-specific DNA demethylation in PGCs. Impairment of PARP activity causes a significant reduction of expression of *Tet1* (Ciccarone et al., 2012). Parp1 also regulates DNA methylation patterns by regulating DNMT1 expression (Zampieri et al., 2009) and activity (Reale et al., 2005). A recent report showed that DNA Translesion synthesis (TLS) is essential for early germ cell development. Deficiencies in TLS initiation and extension caused complete failure in both activation of the germ cell transcriptional program and in DNA demethylation (Shah et al., 2024).

3.4.2 DNA re-methylation in Prospermatogonia

Genome-wide demethylation is common to both male and female PGCs. After sex determination, the male germ cells (Prospermatogonia) gain male-specific DNA methylation patterns from E14.5 to birth (Kubo et al., 2015). The methylation goes from ~7% to ~80%. This genome-wide methylation takes place in two waves. The first wave methylates most of the genome from E13.5 to E16.5, and the evolutionarily young transposable elements are methylated in a second wave (Molaro et al., 2014). The first major wave is catalyzed by DNMT3A, and the second minor wave by DNMT3C (Barau et al., 2016; Dura et al., 2022). The germline-specific cofactor DNMT3L is also essential for *de novo* DNA methylation establishment (Bourc'his and Bestor, 2004).

The first major wave of DNA methylation was found to be dependent on DNMT3A and DNMT3L (Bourc'his and Bestor, 2004; Kaneda et al., 2004). Germline conditional knockout of *Dnmt3B* did not affect male germ cell development and fertility, and thus, DNMT3B was suggested to have no function in the *de novo* methylation. It is possible for DNMT3B to have some catalytic-independent functions which are yet to be found. DNMT3A methylates most of the genomic compartments, including the paternal ICR, except *Rasgrf1* (which behaves like a

Transposable element) (Dura et al., 2022). This clearly indicated the two distinct mechanisms of DNA methylation required for the two waves. This is further strengthened by the fact that the targets of DNMT3A and DNMT3C are mutually exclusive. *Dnmt3l*$^{KO/KO}$ Prospermatogonia gain significantly less DNA methylation than WT, but still, they do possess some, indicating that DNMT3A could perform its function, albeit to a lesser extent, in the absence of DNMT3L. DNMT3A's PWWP domain has been shown to direct DNA methylation through its interaction with H3K36me2 (Weinberg et al., 2019). The lysine methyltransferase NSD1 deposits H3K36me2 in euchromatic regions. In a recent study, H3K36me2 (deposited by NSD1) has been shown to be instrumental to direct genome-wide DNA methylation in Prospermatogonia. As these regions lose DNA methylation in a *Dnmt3a*$^{KO/KO}$ Prospermatogonia, we can infer that DNMT3A methylates regions marked by NSD1-deposited H3K36me2. In oocytes, H3K36me3, deposited by the SETD2 methyltransferase, is required for DNA methylation. Thus the authors argue that this difference in recruiting mechanism could drive the sexually dimorphic pattern of DNA methylation in mature mouse gametes (Shirane et al., 2020). This mechanism seems to be much simpler than for the second wave of methylation.

The second wave of methylation is dependent on DNMT3C. It is also critically dependent on the piRNA pathway, in particular, the Fetal piRNA pathway (explained in Section 3.5). DNA demethylation in PGCs causes the derepression of evolutionarily young transposons. Their mRNA transcripts are funneled into a cytoplasmic perinuclear granule called a Nuage. In the nuage, these transcripts are digested into 24-30 nucleotide (nt) small RNAs called piRNAs by the piRNA proteins MIWI2 and MILI (Ozata et al., 2019). Antisense piRNAs loaded MIWI2 translocates to the nucleus and directs DNA methylation to the transposon loci through piRNA-mediated complementarity (Chuma and Nakano, 2013; Kuramochi-Miyagawa, 2008). Since DNMT3C methylates the promoters of the young and active transposon, there could be a link between DNMT3C and the piRNA pathway. Recent studies have shown that DNMT3 proteins have particular flanking nucleotide preferences for *de novo* methylation. DNMT3A prefers to methylate CpGs followed by C or T, while DNMT3B prefers CpGs followed by G or A. This was shown to be a consequence of structural

interactions between DNMT3 and DNA and not differential recruitment (Mallona et al., 2020). A recent report found that DNMT3C also has certain flanking nucleotide preferences. DNMT3C seems to prefer motifs where the CpG dinucleotide is preceded by a cytosine at -1 or -2 positions. They also report that such motifs are more enriched in the young retrotransposon in mice, which are DNMT3C targets (Dossmann et al., 2024). Such preferences could also contribute to DNMT3C-mediated methylation, and it would be interesting to find out more. A strong positive selection of DNMT3C's N terminal domain could also help explain such unique preferences (Molaro, Malik and Bourc'his, 2020). Although minor, this second wave of DNA methylation is more targeted and complex than the first wave. Muroidea could have developed a dedicated pathway and a novel DNMT3 protein to deal with transposons due to their high load in the genome and high mutagenic potential.

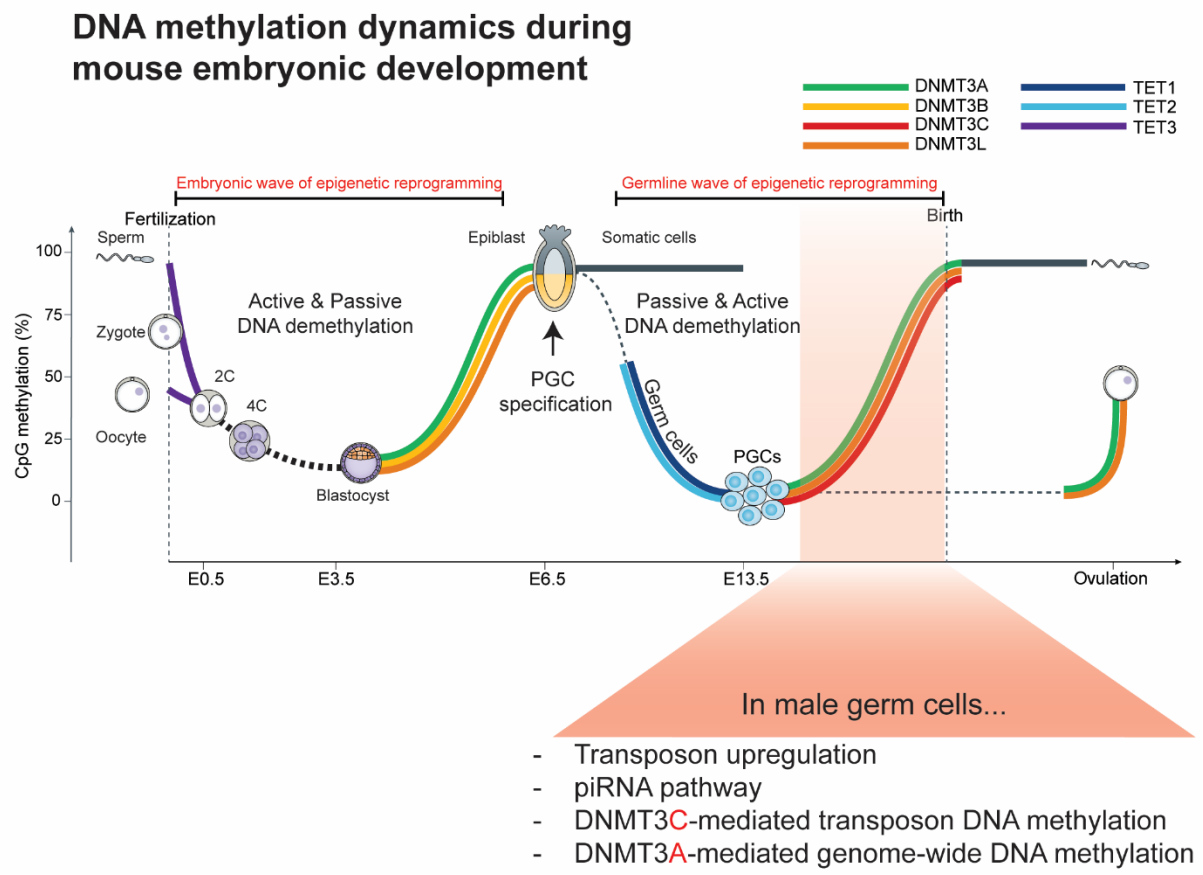
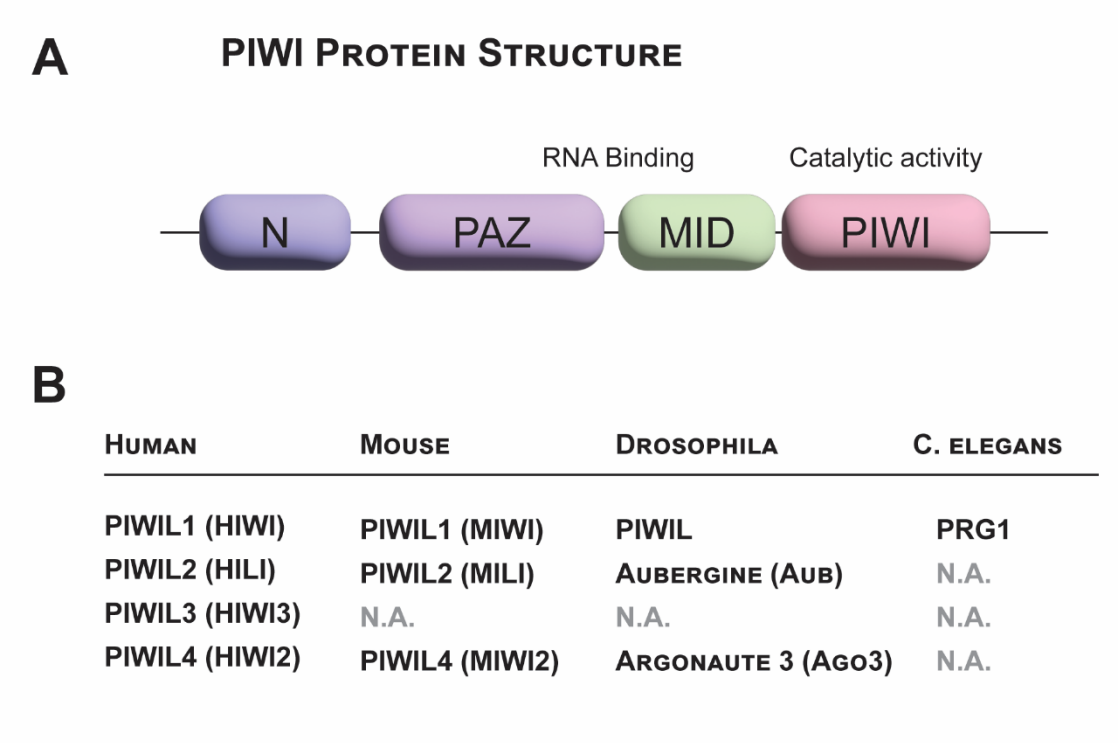


Figure 3.8: **Genome-wide DNA methylation dynamics during mouse embryonic development**. [Extracted and adapted from (Greenberg and Bourc'his, 2019)]. TET3 is active in the fertilized zygote, leading to active DNA demethylation. This is followed by passive demethylation (dashed line). DNA methylation reaches a low point at the blastocyst stage, followed by DNMT3A- and DNMT3B-mediated *de novo* DNA methylation after blastocyst implantation. This DNA de- and re-methylation wave is called the "Embryonic wave of epigenetic reprogramming." Germ cells are specified in the epiblast. They undergo two waves of DNA demethylation: passive and active (mediated by TET1 and TET2). Prospermatogonia become highly methylated before birth through the activity of DNMT3A, DNMT3C, and DNMT3L. This DNA de- and re-methylation wave is called the "Germline wave of epigenetic reprogramming." In this wave, the piRNA pathway is active (explained later), DNMT3C methylates the transposons, and DNMT3A methylates the remaining genome. The oocyte gains methylation after meiosis and before ovulation through the activity of DNMT3A and DNMT3L in mice.

3.5 THE PIWI INTERACTING RNA (PIRNA) PATHWAY

PIWI-interacting RNAs (piRNAs) are a distinct class of small non-coding RNAs, typically 24-35 nucleotides in length, that associate with PIWI (P-element-induced wimpy testis) proteins, a subclass of the Argonaute protein family. This pathway plays a vital role in silencing transposable elements, regulating gene expression, and safeguarding the integrity of the germline genome (Sun, Lee and Li, 2022). Depending on the developmental expression window, piRNAs are classified into fetal piRNAs, pre-pachytene piRNAs, and pachytene piRNAs. Fetal piRNAs are expressed in the Prospermatogonia, which is mostly made up of transposon transcripts. Pachytene piRNAs contain very few TE-derived sequences. In the following sections, I will explain the biology of fetal piRNAs, as they are responsible for transposon repression.



HUMAN	MOUSE	DROSOPHILA	C. ELEGANS
PIWIL1 (HIWI)	PIWIL1 (MIWI)	PIWIL	PRG1
PIWIL2 (HILI)	PIWIL2 (MILI)	AUBERGINE (AUB)	N.A.
PIWIL3 (HIWI3)	N.A.	N.A.	N.A.
PIWIL4 (HIWI2)	PIWIL4 (MIWI2)	ARGONAUTE 3 (AGO3)	N.A.

Figure 3.9 **PIWI proteins**. (A) Domain architecture of PIWI proteins. (B) PIWI homolog proteins across various model organisms.

3.5.1 piRNA biogenesis and post-transcriptional gene silencing

The piRNA biogenesis begins with the transcription of its precursor. When it comes to transposons, individual transposon loci have been shown to give rise to precursors that generate piRNAs (Aravin et al., 2008). This is in contrast to the *Drosophila* model, where piRNA clusters generate precursor RNAs that primarily give rise to piRNAs. The biogenesis of piRNAs can be divided into two parts: primary and secondary biogenesis. Primary biogenesis produces piRNAs from precursors, and secondary biogenesis amplifies piRNAs through a ping-pong mechanism.

In mice Prospermatogonia, the evolutionary young transposons are transcribed by RNA polymerase II. As transposons try to mimic host genes, they could utilize the same nuclear export machinery as genes to exit the nucleus. Once in the cytoplasm, these transcripts become the piRNA precursors. Since transposon proteins, specifically LINE1 ORF1 protein, are visualized by immunofluorescence in the germ cells, the exported transposon mRNA undergoes translation. The identification mechanism of which transcripts become piRNA precursors is still unknown. Ribosomes have been implicated in pachytene piRNA production (Sun et al., 2020). As they translate the transposon mRNA in Prospermatogonia, ribosomes could play a role in the identification of precursor targets and their processing. The precursors are first transported into a membrane-less granule called the intermitochondrial cement (IMC). IMCs are membrane-less phase-separated clusters that form in between mitochondria and keep them together. They are present in Prospermatogonia, Spermatogonia, and meiotic Spermatocytes until the pachytene stage. It is also known as the pi-body and is the cytoplasmic location of primary piRNA biogenesis. The proteins enriched in IMC mainly belong to the Tudor domain family and Piwi proteins. The Tudor domain specifically recognizes arginine or lysine-methylated ligands to facilitate protein-protein interaction (Maurer-Stroh et al., 2003). PIWI (P-element-induced wimpy testis) proteins are an evolutionarily conserved subclade of the Argonaute (AGO) family, which are expressed largely in mammalian germ cells and are essential for germ cell development (Czech et al., 2018; Tóth et al., 2016). Mice have three PIWI proteins: MIWI (PIWIL1), MILI (PIWIL2), and MIWI2 (PIWIL4). MIWI2 is expressed only

in Prospermatogonia (E14.5-P3). MIWI is expressed only in spermatocytes from P14 onwards until the round spermatid stage. MILI is expressed in both Prospermatogonia and spermatocytes, and its expression coincides with MIWI2 and MIWI. Interestingly, mice do not have PIWIL3, an enzyme that is essential for the piRNA pathway in hamsters and human oocytes (Ishino et al., 2021; Zhang, H. et al., 2021). This could explain the dispensability of the piRNA pathway for female fertility in mice but not in hamsters and oocytes.

The post-transcriptional processing of all primary piRNA precursors can be simplified to three steps: 5′ end formation, PIWI protein loading, and 3′ end formation. Once the piRNA precursor enters the IMB, it is bound by the RNA helicase MOV10L1. MOV10L1 feeds these precursors to the MitoPLD (Mitochondrial phospholipase) endonuclease (aka PLD6 or ZUCCHINI homolog *MmZuc*) (Huang et al., 2011; Watanabe et al., 2011). MitoPLD is located on the outer membrane of mitochondria, and it has been shown to possess endoribonuclease activity for single-stranded RNAs *in vitro* (Ipsaro et al., 2012). MitoPLD cleaves the precursor RNA, now called pre-pre-piRNA, generating the monophosphorylated pre-piRNA 5' end. The single-stranded pre-pre piRNA is cleaved into thousands of non-overlapping/phased fragments. MOV10L1 is thought to load PIWI proteins onto the pre-pre-piRNAs (Zheng et al., 2010). The cleaved pre-pre-piRNAs are loaded onto MILI, which triggers the next cleavage between the 3′ end of the pre-piRNAs and the 5′ end of the pre-pre-piRNAs. The 3’ end of the pre-piRNAs are trimmed to reach the final size of the piRNAs. This final size of the mature piRNA is determined by the footprint of the bound PIWI protein (MILI ~ 26 nt, MIWI2 ~ 28 nt, and MIWI ~ 29 nt). TDRKH, a mitochondria-anchored Tudor family protein, enables the trimming of MILI-bound piRNAs (Saxe et al., 2013). TDRKH associates with the MIWI/MIWI2 by binding to their symmetrically dimethylated arginine. Through this binding, TDRKH acts as a scaffold for interactions between PIWI–piRNA complexes and the trimming enzyme. The 3′ end trimming protein in mice is a Poly(A)-specific ribonuclease-like domain-containing protein 1 (PNLDC1) (Bronkhorst and Ketting, 2018; Ding et al., 2017; Nishimura et al., 2018; Zhang et al., 2017). This exoribonuclease trims the 3’ end of piRNA intermediates to their mature length. The exposed 3’ end is then methylated in the 2’ position (2’-O-methylation) by HEN

methyltransferase 1 (HENMT1) (Lim et al., 2015). The 2'-O-methylation prevents 3' tailing (addition of non-templated nucleotides to the 3' end of RNAs) and thus protects the 3' end, conferring stability to the mature piRNAs. Loss of 2′-O-methylation induces RNA tailing on piRNA precursors in worms and triggers pre-piRNA decay (Pastore et al., 2021). Another study found out that in yeast, tailing of Argonaute-bound small RNAs degrades the small RNAs. Failure in such a degradation results in the accumulation of 'noise' small RNAs on Argonaute and uncontrolled RNAi (Pisacane and Halic, 2017). Interestingly, a study reports that 3' uridylation or adenylation of pachytene piRNAs in mice facilitates the assembly of the MIWI/piRNA complex for efficient target regulation (Zhao et al., 2022). This could be explained by the role of pachytene piRNAs to post-transcriptionally regulate protein-coding mRNAs during mouse spermatogenesis. Corroborating this hypothesis, the terminal uridydyl transferase 4/7 (TUT4/7) mediated 3' uridylation of miRNA, piRNA, and mRNA regulates the clearance of zygotene-expressed mRNA transcripts in pachytene cells (Morgan et al., 2019). In essence, the mature, 3' trimmed 2'-O-methylated MILI-bound piRNAs are called primary piRNAs, and this marks the end of primary piRNA biogenesis. These primary piRNAs have a signature bias towards the first nucleotide in the 5' being a Uridine (1U). Since MitoPLD does not show any preference for cleaving before uridines *in vitro*, there could be a co-factor that works co-ordinately with MitoPLD to generate the 1U bias.

MILI binds to the target RNA through the (primary) piRNA-mediated complementarity. It then slices the target RNA transcript after position 10 of the piRNA sequence, creating the 5' end of the secondary piRNA (Fazio et al., 2011). Due to this, the secondary piRNAs have a bias towards the 10^{th} nucleotide being Adenine (10A). This cleavage triggers the 3' end formation, trimming, and 2'-O-methylation to produce mature secondary piRNAs. Tudor domain-containing 1 (TDRD1) and TDRD12 are MILI interactors that ensure fidelity of the interaction between MILI. TDRD1 recognizes arginine dimethylation in MILI (Chuma et al., 2006) and may regulate the entry of transcripts into piRNA biogenesis pathways. The loss of TDRD1 has no effect on the abundance of MILI-bound piRNA but rather their identity changes: ribosomal RNA- and genic-derived piRNA proportions increase, and transposon-derived

piRNAs reduce. FKBP6 was shown to interact with MILI and TDRD1. It also associates with the molecular chaperone HSP90AA1 and, through this, could recruit the chaperone machinery for secondary piRNA biogenesis (Xiol et al., 2012). FKBP6 was also shown to be essential for spermatogenesis in humans (Wyrwoll et al., 2022). TDRD1-associated protein, TDRD12, also forms a complex with the piRNA-bound MILI. It might facilitate the conformation changes required for the exchange of piRNA intermediates between MILI and MIWI2 (Pandey et al., 2013). In *Tdrd12*$^{KO/KO}$, MILI-bound piRNAs were unchanged, but the MIWI2-bound piRNAs were completely absent. DDX4 (Mouse Vasa Homolog) is a classic mouse germ cell marker (Kuramochi-Miyagawa et al., 2010). DDX4 is also required for RNP remodeling during the loading of secondary piRNA intermediates onto MIWI2 (Wenda et al., 2017). Another member essential for secondary piRNA biogenesis is GTSF1. The second germ cell membrane-less phase-separated organelle involved in piRNA biogenesis is the piP body. piP-bodies harbor MIWI2, TDRD9, and MAEL, which are required for secondary piRNA biogenesis. piP-bodies are lost after birth, but IMC structures remain until the pachytene stage. MIWI2/MAEL and MILI occupy these distinct germinal granules (Aravin et al., 2009). GTSF1 colocalizes with MIWI2 and TDRD9 in the piP bodies. In *Gtsf1*$^{KO/KO}$ mutants, the target RNA of MILI-piRNA is left unsliced at the cleavage site, and the generation of secondary piRNAs from this transcript becomes defective (Yoshimura et al., 2018).

These secondary piRNAs can target the primary piRNA precursor transcripts to generate more secondary piRNAs, resulting in a piRNA-specific “Ping-Pong” loop (Brennecke et al., 2007). MILI-derived secondary piRNAs in the pi-body are loaded onto MIWI2 in the pi-P body. Thus, the ping-pong loop oscillates between these two germinal granules. Such a ping-pong loop is only possible when both Sense and Antisense transcripts of a particular precursor sequence exist. This is highly possible for repetitive sequences like transposon sequences as they drive their transcription in Sense, and their insertion orientation at genic regions could derive Antisense transcripts driven by genic promoters. Therefore, this loop acts as an adaptive immune system, enriching its specific weapons (piRNAs) and acting on abundant target transcripts (transposons) until they are diminished. Through such a loop, transposon mRNA

will be degraded, and this is termed as post-transcriptional gene silencing (PTGS) leg of the piRNA pathway. Interestingly, MILI alone is enough to carry out the ping-pong and PTGS (Fazio et al., 2011). Cleavage by MILI triggers a phased processing of the target RNA into non-overlapping fragments. These fragments become piRNAs and can be bound by MILI and MIWI2 (Yang et al., 2016). MILI carries out the PTGS in the cytoplasm, and MIWI2 carries out the transcriptional gene silencing (TGS) (Manakov et al., 2015).

For further reading, please refer to the following review articles - (Chuma and Nakano, 2013; Li, Zhang and Liu, 2021; Sun, Lee and Li, 2022; Wang et al., 2020).

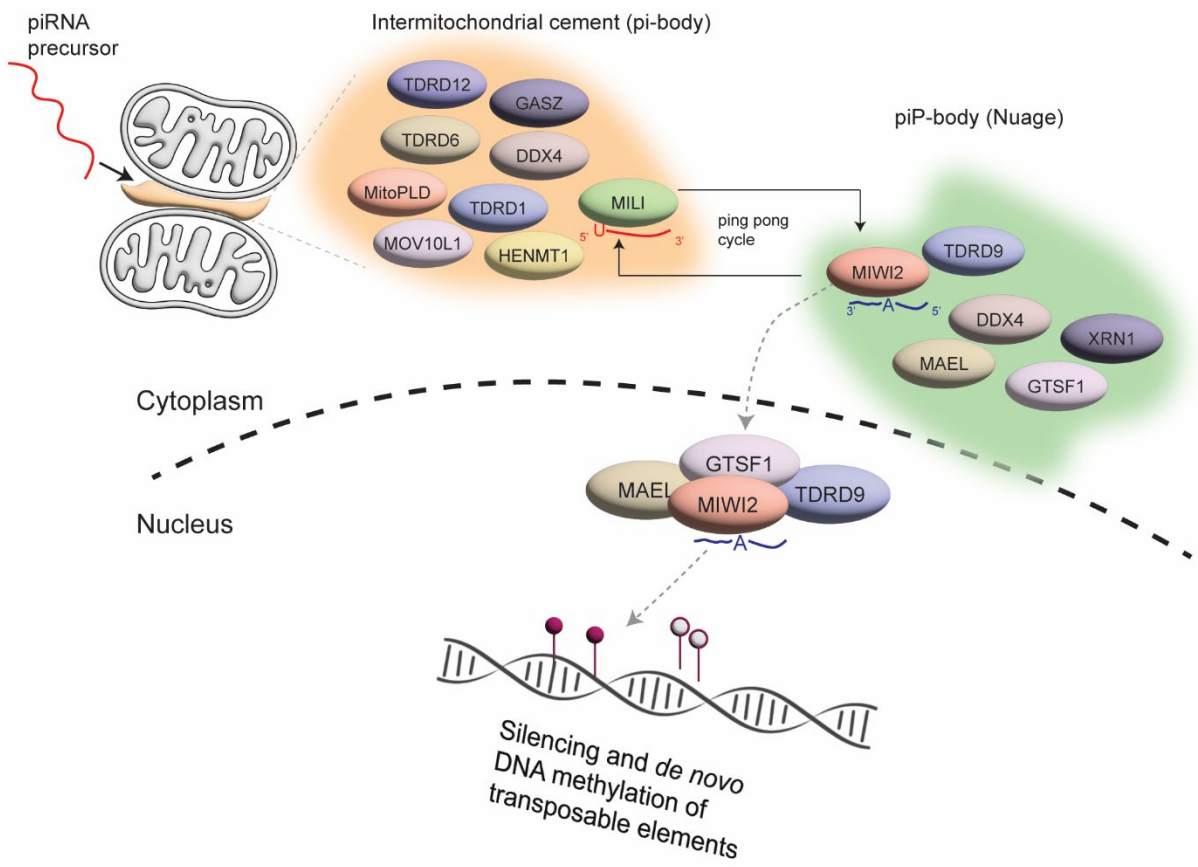


Figure 3.10: **Primary and Secondary piRNA biogenesis**. piRNA precursors are exported from the nucleus and fed into the Intermitochondrial cement (IMC). They are processed by factors in the IMC into piRNA intermediates in a process called Primary piRNA biogenesis. They are loaded onto MILI, and the 3' end is trimmed and 2'-O-methylated, generating mature primary piRNA with a 1U signature. The MILI-bound piRNA then engages in the ping-pong cycle, leading to the cleavage of complementary antisense transcripts, producing secondary piRNA with 10A signature, and then are loaded to either MILI or MIWI2. Loading of MIWI2 with secondary piRNA induces its translocation into the nucleus, where it functions in the silencing and methylation of transposable elements together with other nuclear factors.

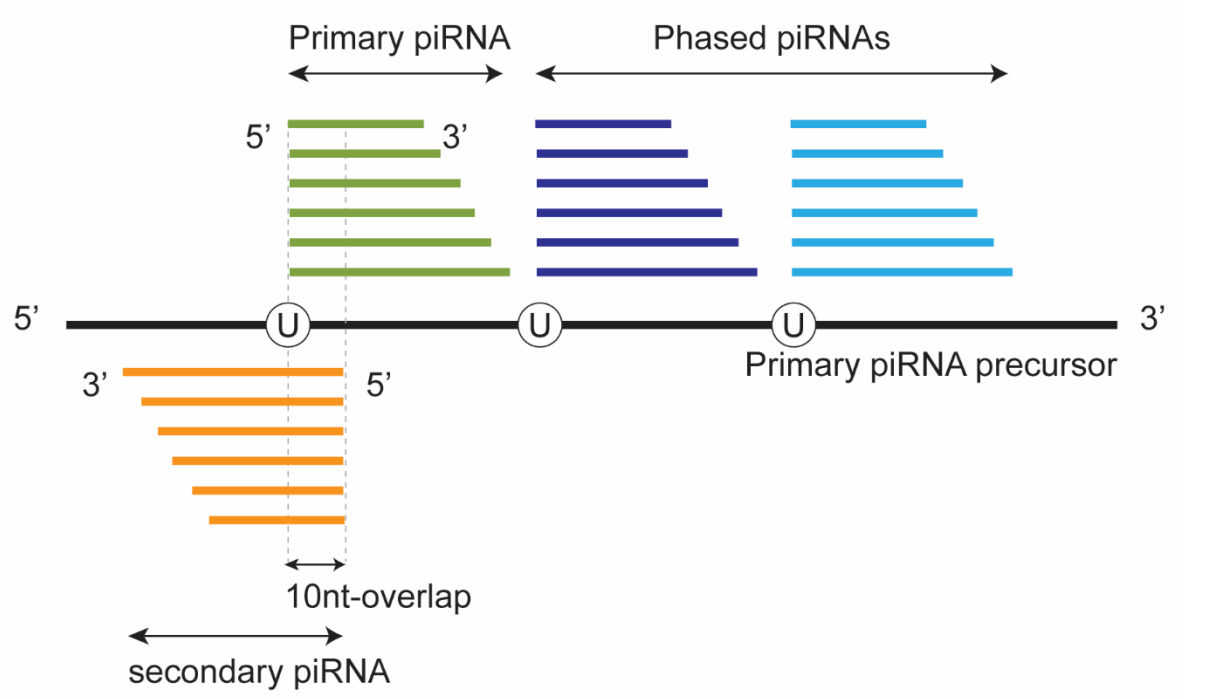


Figure 3.11: **Illustration of piRNA biogenesis**: Primary piRNAs are generated from the precursor by MILI cleavage. This leads to phased piRNA biogenesis downstream of the precursor. The primary piRNA mediates cleavage of the secondary piRNA precursor to generate secondary (antisense) piRNAs, which bear a 10-nt overlap with the primary piRNA that generated them. [Adapted from (Jensen et al., 2020)]

3.5.2 MIWI2-mediated transcriptional gene silencing

Once loaded with secondary piRNAs, MIWI2 translocates to the nucleus along with its interactors and mediates transcriptional gene silencing on the primary piRNA precursor transcribing loci. *Miwi2*$^{KO/KO}$ male germ cells showed defective synapsis due to high transposon derepression at the pachytene stage, leading to germ cell loss (Carmell et al., 2007). MIWI2 has been shown to be an effector of DNA methylation of retrotransposons. MIWI2 identifies the actively transcribing transposon loci through piRNA-mediated complementarity. To confirm that the recruitment of MIWI2 to these loci results in DNA methylation, Kojima-Kita and colleagues created a transgenic mouse expressing MIWI2-Zinc finger (MIWI2-ZF) fusion protein. The ZF recognized the promoter of a LINE-1A gene. Tethering MIWI2 to the LINE-1A promoter brought DNA methylation to these loci in a *Piwil2*$^{KO/KO}$ background (which lacked all DNA methylation at these loci). Although the DNA methylation levels were less than that seen in *Piwil2*$^{WT/WT}$ germ cells, it was specific, as other LINE-1 families did not show this increase in DNA methylation (Kojima-Kita et al., 2016). The "piRNA" part of the MIWI2-piRNA targeting

complex was also confirmed as removing a retrotransposon DNA sequence from a piRNA-dependent region abolishes DNA methylation. In the same study, it was also shown that MIWI2 binds to the RNAs transcribed from these piRNA-dependent regions (Watanabe et al., 2018). Below, I explore the roles of certain essential proteins which associate with MIWI2 and are critically involved in the nuclear piRNA pathway.

3.5.2.1 TDRD9

TDRD9, a TDRD family member, is an essential partner of MIWI2. TDRD9 complexes with MIWI2 through its Tudor domain and binds to the symmetrically dimethylated arginines in MIWI2. TDRD9 colocalizes with MIWI2 in the piP bodies and also in the nucleus. This MIWI2-TDRD9 compartmentalization is controlled by the MILI-TDRD1 complex as TDRD9 and MIWI2 were dispersed in the cytoplasm, and granular assemblies were lost in the absence of MILI or TDRD1. The study that discussed *Tdrd9*$^{KO/KO}$ mice shows that TDRD9 likely does not bind to piRNAs. *Tdrd9*$^{KO/KO}$ male germ cells show an aberrant piRNA profile, indicating that TDRD9 also impacts piRNA biogenesis. Interestingly, the ATPase activity of TDRD9 is not essential for piRNA biogenesis and for MIWI2 nuclear localization. However, it is essential for TDRD9's nuclear localization and its potential role in the transcriptional silencing of retrotransposons, identifying TDRD9 as a nuclear effector of the piRNA pathway (Shoji et al., 2009; Wenda et al., 2017). Recent MIWI2-interaction proteomics revealed TEX15 and SPOCD1 as MIWI2 interactors (Schöpp et al., 2020; Zoch et al., 2020).

3.5.2.2 TEX15

TEX15 (Testes-expressed protein 15) is a nuclear protein found in vertebrates from fish to humans. The piRNA biogenesis is intact in *Tex15*$^{KO/KO}$ mice. MIWI2 also retains its nuclear localization in *Tex15*$^{KO/KO}$ mutants, indicating that there is no defect in piRNA biogenesis. Loss of TEX15 causes demethylation in LINE1 and IAP transposon promoter regions, as seen for *Piwil2*$^{KO/KO}$ and *Piwil4*$^{KO/KO}$ male germ cells (Schöpp et al., 2020). It also shows defective chromosomal synapsis due to the impaired localization of recombination proteins RAD51 and DMC1 in the pachytene stage (Yang et al., 2008). This phenotype also resembles what we see

for the other piRNA pathway mutants. As explained before, this is caused by the relocation of recombination hotspots to transposons caused by the loss of DNA methylation (Zamudio et al., 2015). TEX15 was also shown to interact with MILI. Notably, TEX15's nuclear localization is also not dependent on MIWI2 (Yang et al., 2020). All the above evidence suggests that TEX15 acts as a nuclear effector of the piRNA pathway. A new study has found that the DUF3715 domain of TEX15 in TEX15 is involved in RNA-directed transposon silencing. TEX15 shares the DUF3715 domain with TASOR and TASORB proteins, which form the HUSH complex in humans. The HUSH complex is a protein complex that protects the genome from retrotransposons by recruiting histone methyltransferases, catalyzing H3K9me3 and additional proteins involved in chromatin compaction (Müller and Helin, 2024). Chimeric TASOR proteins with their DUF3715 domains replaced with human TEX15-DUF3715 domain or zebrafish TEX15-DUF3715 domain could restore LINE-1 silencing in a *Tasor*$^{KO/KO}$ mouse embryonic stem cell line. Although this does not explicitly convey the role of this domain in mouse TEX15's piRNA participation, their results demonstrate an essential role for this domain in transposon silencing (Schöpp et al., 2023). TEX15 is also essential for normal spermatogenesis in human males (Okutman et al., 2015; Zhang et al., 2022). Interestingly, TEX15 is absent from the SPOCD1 IP-MS dataset, suggesting that TEX15 is likely to function upstream of the latter or in an arm of MIWI2 pathways parallel to the SPOCD1 complex. Various other piRNA pathway mutants have also been shown to cause non-obstructive azoospermia in humans (Ozturk, 2023; Stallmeyer et al., 2024). Understanding their role in mice could shed light on their functions in human male germ cell development.

3.5.2.3 SPOCD1

SPOCD1, another MIWI2 interactor, also enables MIWI2 activity in the nucleus. It is also expressed in the male germ cells with the same expression pattern as MIWI2. SPOCD1 does not impact the piRNA biogenesis and MIWI2's nuclear localization. The *Spocd1*$^{KO/KO}$ mutants show defects in DNA methylation of retrotransposons in male germ cells and, thus, defects in chromosomal synapsis at the pachytene stage (Zoch et al., 2020). Like MIWI2, SPOCD1 was also demonstrated to be involved in *de novo* methylation of evolutionary young

IAPEy and MMERVK10C retrotransposons as well as young LINE1 families, L1MdA, L1MdT, and L1MdGf. A recent study showed that SPOCD1 is also required for human male fertility and transposon silencing (Zoch et al., 2024).

SPOCD1 protein has two protein domains – SPOC (SPEN ortholog and paralog C-terminal) domain and TFIIS-central domain. Another protein that contains a SPOC domain is SPEN (SHARP, MINT), which is involved in X-chromosome inactivation (XCI) (Kaufmann and Wutz, 2023). XCI and transposon repression are quite similar – both are RNA-mediated transcriptional repression (*Xist* lncRNA for XCI and piRNAs for transposons) and inactivation mechanisms (mediated by DNA methylation). A recent study showed that SPEN's SPOC domain is a major effector of the gene-silencing function of SPEN (Dossin et al., 2020). SPEN's SPOC domain interacts with the NCoR/SMRT complex proteins (NCoR1, NCoR2, HDAC3), NuRD complex (MTA1), RNA polymerase II, and the m6A methylation machinery (METTL3, WTAP). They also noted that SPEN binding peaks on X-chromosome overlap with RNA polymerase II phosphorylated on serine 5 (pS5), which is associated with transcription initiation (Kimura and Sato, 2022). This binding is possible through the SPOC domain. The NCoR/SMRT complex promotes transcriptional repression and NuRD complex functions in chromatin remodeling. By interacting with (1) initiating RNA polymerase II, transcriptional repressive complexes through its SPOC domain, and (2) with the *Xist* lncRNA, SPEN mediates XCI. That study also showed that tethering just the SPOC domain to *Xist* lncRNA resulted in substantial gene silencing across the X chromosome in SPEN-depleted mouse embryonic stem cells. (Dossin et al., 2020). Another interesting study that was published in the same year found that SPEN binds to ancient retroviral RNA, and recruits' transcription repression machinery to the repeat loci. Loss of SPEN activates a subset of ERV elements in mouse embryonic stem cells, with gain of chromatin accessibility, active histone modifications, and ERV RNA transcription. SPEN binds to the A-repeat region in *Xist* lncRNA. *Xist* repeats are presumed to have originated from the insertion and duplication of transposable elements, with the A-repeat showing similarity to ERVs (Elisaphenko et al., 2008). This study suggests that *Xist* may have incorporated the ERV sequences to co-opt the already existing powerful TE-silencing

machinery for XCI (Carter et al., 2020). This brings XCI and transposon silencing closer than previously thought and there could be more parallels to be drawn.

The SPOC domain of a human protein, PHF3 (PHD-finger protein 3), was shown to bind to the modified C-terminal domain (CTD) of RNA polymerase II. More specifically, the PHF3's SPOC domain binds to the phosphorylated Serine-2 (pS2) mark in the heptad repeat (YSPTSPS) of the CTD. The pS2 mark is associated with elongating RNA polymerase II (Kimura and Sato, 2022). Consequently, PHF3 loss results in increased Pol II stalling and reduced elongation rate (Appel et al., 2021). The same group also did a more comprehensive study of SPOC domains of different human proteins DIDO, human SPEN, and RBM15 and established SPOC as a CTD binding domain. Human SPEN and RBM15 bound to initiating RNAP II (having pS5) and PHF3 and DIDO bound to elongating RNAP II (having pS2). Human SPEN positively regulates transcription initiation through its SPOC domain, whereas PHF3, DIDO, and RBM15 regulate transcription elongation and mRNA stability. Overall, they show that the SPOC domain serves as a major interface between the transcription machinery and regulators of transcription and co-transcriptional processes (Appel and Benedum et al., 2023; Appel and Franke et al., 2023).

Phylogenetic analysis showed that *Spocd1* originated from a duplication of *Phf3*. Given PHF3's involvement with elongating RNAP II, SPOCD1 could also interact with the elongating RNAP II at transcribing transposon loci and establish transcriptional repression. To explore its function in the mouse piRNA pathway, the group that identified SPOCD1 performed SPOCD1-affinity proteomics in testes from E16.5 *Spocd1*$^{HA/+}$ mice. Members of *de novo* methylation machinery, DNMT3A, and DNMT3L, as well as components of BAF and NURD chromatin remodeling complexes, were significantly enriched. Expressing tagged versions of each protein in a HEK293T cell line, they validated these interactions using immunoprecipitation-western blot. DNMT3C was not significantly enriched in the IP-MS dataset. However, an interaction of SPOCD1 and DNMT3C was seen from the IP-WB experiments in the HEK293T cell heterologous system. Another interesting interactor of SPOCD1 is SPIN1. SPIN1 was one

of the most highly enriched proteins in the SPOCD1 IP-MS. It is an epigenetic reader of the bivalent mark – H3K4me3/H3K9me3 (Du et al., 2021). Interestingly, following demethylation in PGCs, the retrotransposons become transcriptionally active and gain H3K4me3 marks. Simultaneously, they are also marked by H3K9me3 at their promoters, thereby showcasing bivalency (Dura, 2021; Yamanaka et al., 2019). SPIN1 could bind to this bivalent mark and help tether SPOCD1 and the other nuclear piRNA pathway components to the promoter of transposons. A recent study showed that SPIN1 and its interactor SPINDOC proved essential for proper DNA methylation maintenance during mouse embryonic development. They also show that SPIN1-SPINDOC represses TEs in a DNA methylation-independent manner (Emeline Roger, 2023).

3.5.2.4 C19ORF84H

C19ORF84H is an uncharacterized protein of unknown function. It was highly enriched in the SPOCD1 interactome and significantly enriched in the MIWI2 interactome (Zoch et al., 2020). C19ORF84H is also expressed at the same developmental window as SPOCD1, MIWI2, and DNMT3L. It does not possess any domains of known functions, and large swathes of the protein are disordered. Intrinsically disordered proteins (IDPs) and regions (IDRs) have been recognized as readily available sites for protein-protein interactions and posttranscriptional modification hotspots (Chakrabarti and Chakravarty, 2022; Haynes et al., 2006; Iakoucheva et al., 2002; Kim et al., 2008). C19ORF84H could play such a role for the nuclear piRNA pathway proteins, bridging the specifying mechanism (MIWI2 loaded with piRNAs) with the effector mechanism (SPOCD1, DNA methylation, and other transcriptional machinery). C19ORF84H is a testis-specific protein and shows nuclear localization. *C19orf84h*$^{KO/KO}$ shows the same phenotype as *Spocd1*$^{KO/KO}$, *Miwi2*$^{KO/KO}$, and *Dnmt3c*$^{KO/KO}$ mice – arrest of spermatocytes at pachytene stage, causing germ cell loss and therefore male-specific infertility. C19ORF84H does not impact piRNA biogenesis, indicating it is a nuclear effector of the piRNA pathway (Konieczny, 2022). Transposons are highly derepressed in the *C19orf84h*$^{KO/KO}$ spermatocytes at the pachytene stage. This is because *C19orf84h*$^{KO/KO}$ mice failed to methylate the promoters of active retrotransposons in the Prospermatogonia stage.

C19ORF84H's interaction with SPOCD1 was validated by expressing both in the HEK293T cell line. The C-terminus of SPOCD1 (amino acids 825–1,015) and a 10 amino acid patch (81-90) in C19ORF84H are involved in this interaction. It was shown by immunofluorescence that C19ORF84H, SPOCD1, and DNMT3C are diffused in the nucleoplasm at E14.5-E17.5 and form foci in the Prospermatogonia nucleus at E18.5. C19ORF84H foci colocalizes with the foci of DNMT3C and SPOCD1. Anti-HA IP-MS of DNMT3C-HA (*Dnmt3c*$^{HA/+}$) from E16.5 testes showed SPOCD1 and C19ORF84H significantly enriched. However, DNMT3C was not significantly enriched in the C19ORF84H reciprocal IP-MS. MIWI2's interactome was not changed significantly in the absence of C19ORF84H. MIWI2's interaction with the piRNA-biogenesis factors TDRD5 and RNF17 was affected in *C19orf84h*$^{KO/KO}$ mutant. Given the fact that C19ORF84H does not affect piRNA biogenesis and MIWI2's nuclear localization, the loss of such protein interactions for MIWI2 does not translate to biological significance. Even SPOCD1 lost only a handful of protein interactors in the *C19orf84h*$^{KO/KO}$ mutant. But among these were MTA2 and DNMT3L. MTA2 is a subunit of the NuRD chromatin remodeling complex whose role in the piRNA pathway is not known. DNMT3L is a core member of the DNA methylation complex and a critical member of the transposon methylation in Prospermatogonia (Bourc'his et al., 2001; Bourc'his and Bestor, 2004). Loss of SPOCD1-DNMT3L interaction signifies the role of C19ORF84H in bridging the DNA methylation machinery to the piRNA pathway (Konieczny, 2022; Zoch et al., 2024).

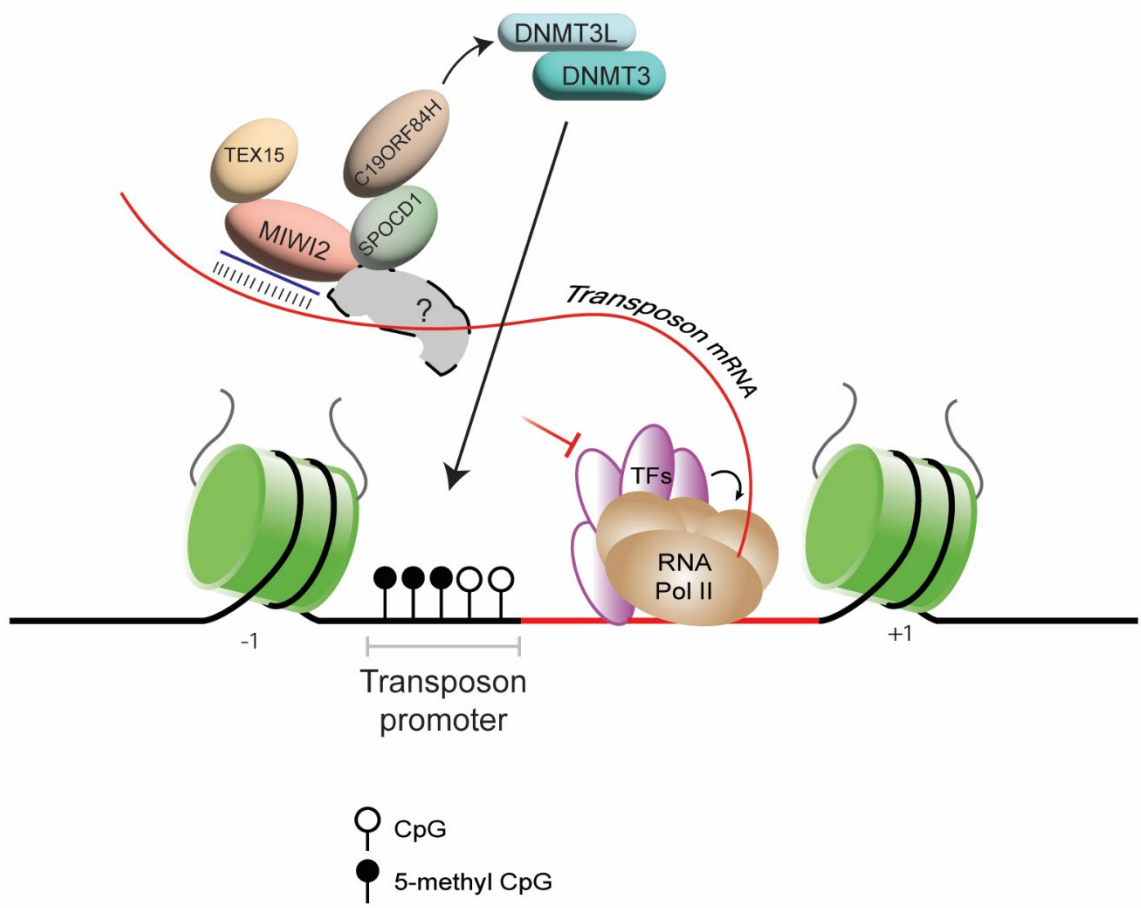


Figure 3.12: **Hypothetical model of MIWI2-mediated transcriptional silencing**. MIWI2 recognizes the nascent transposon transcript through piRNA-mediated complementarity. Through its interaction with SPOCD1, it could interact with the transcriptional machinery. C19ORF84H could help recruit the DNA methylation machinery.

AIM OF THE THESIS

Embryonic development is a symphony where multiple participants must dance in concert and tandem for things to fall into their appropriate place. Germline development is fascinating in its own right as it cycles between its germline and somatic fates every generation. How a very small man can cast a very large shadow, so does a tiny time window in the germline development cast a very large influence in the entire process. DNMT3C, through its methylation of transposons, is critical during that window. In my thesis, I wanted to know how DNMT3C fits into the puzzle of transposon methylation. *Dnmt3c* originated as a gene duplication event from *Dnmt3b*. They share a high degree of identity at the amino acid level. DNMT3C also shares a considerable level of identity with DNMT3A. Yet, only DNMT3C can methylate transposon promoters. Transposon promoters comprise ~1% of the Mouse genome, while transposons comprise ~50%. DNMT3A methylates evolutionarily old transposons and the body of young and active transposons. However, it cannot methylate the promoter DNA, which is merely a few hundred base pairs away. This was highly intriguing to me. The two questions I started with for my PhD thesis are as follows.

1. How is DNMT3C guided to the transposon promoters? In other words, what gives DNMT3C its specificity?

2. Why DNMT3A and DNMT3B cannot replace DNMT3C?

MATERIALS AND METHODS

5.1 MOUSE STRAINS AND HOUSING

Mice were housed in Greenline cages (Tecniplast GM500) with a 12h light/12h dark cycle at 24 ± 2 °C and 50–70% humidity, nest-building paper, cotton materials, and free access to food and water. All animal lines were kept on the C57Bl/6J genetic background. The mouse strains used in this study are.

Strain name	Description	Reference
Dnmt3C$^{myc/myc}$	Knock-in of 2x-myc tag at DNMT3C N-terminus	This study
Dnmt3C$^{KO/KO}$	Knockout of *Dnmt3C* gene	(Barau et al., 2016)
Dnmt3C$^{KO/KO}$ *Oct4*$^{GFP/GFP}$	Knockout of *Dnmt3C* combined with IRES-GFP allele downstream of *Oct4* (*Pou5f1*) gene stop codon. This strain was used for isolating male primordial germ cells (Stages E13.5-P2)	(Barau et al., 2016; Yoshimizu et al., 1999)
Lsd1$^{fl/fl}$	*Lsd1* allele by flanking exon 1 with loxP sites. For generating conditional knockout of *Kdm1a*.	(Duteil et al., 2017)
Lsd1$^{fKI/fKI}$	To replace the WT *Lsd1* allele with a knock-in cassette containing exons carrying triple point mutation, making LSD1 catalytically inactive.	(Duteil et al., 2016)
Prdm1$^{CreTg/WT}$	Cre driver line – *Cre* gene expression is coupled to *Prdm1* expression.	(Ohinata et al., 2005)
Vasa$^{Cre/WT}$	Cre driver line – *Cre* gene expression is coupled to *Vasa* expression.	(Gallardo et al., 2007)

Dnmt3C$^{myc/myc}$ mice generation is explained in the Results section – 6.2.1.

5.2 MOUSE EMBRYONIC STEM CELL CULTURE

E14 Mouse Embryonic stem cells (mESCs) were cultured in a medium with the following composition: Dulbecco's Modified Eagle Medium (Thermo Fisher Scientific 21969035), 15% Fetal calf serum (ES approved), 1mM Sodium Pyruvate (Thermo Fisher Scientific 11360039), 0.1mM MEM non-essential amino acids (Thermo fisher Scientific 11140050), 2mM L-Glutamine (Thermo Fisher Scientific 25030024), 100 U/mL Pen Strep (Thermo Fisher Scientific 15140122), 0.1mM 2-Mercaptoethanol, 1000U/mL Leukemia inhibitory factor (LIF, inhouse), 1 mM MEK inhibitor PD0325901 (Axon Medchem 1408), 3 mM Gsk3 inhibitor CHIR99021 (Axon Medchem 1386). All mESC cells were grown in gelatin-coated plates. The media was changed daily, and the cells were passaged every second day with TrypLE treatment.

The cell lines used in this study are E14 WT mESC, E14 TKO (*Dnmt1*$^{KO/KO}$, *Dnmt3a*$^{KO/KO}$, *Dnmt3b*$^{KO/KO}$) mESC, TKO mESCs expressing CAG-promoter driven *2xmyc-Dnmt3C* from Rosa26 locus.

5.3 IMMUNOPRECIPITATION AND WESTERN BLOTTING

We used the Myc-Trap® Magnetic Agarose (Chromotek ytma) for the Myc pulldowns and the Pierce™ Anti-HA Magnetic Beads (Thermo Fisher Scientific 88836) for the HA pulldowns. The Myc-Trap beads consist of a Myc Nanobody/VHH coupled to magnetic agarose beads, and the Anti-HA beads are coated with a mouse monoclonal anti-HA antibody.

Briefly, the beads were washed thrice with the Immunoprecipitation buffer (NS150 – 20mM HEPES pH 7.6, 150mM NaCl, 2mM $MgCl_2$, 1X Protease inhibitor cocktail, 0.2% Igepal CA-630 (Sigma Aldrich I3021), 20% Glycerol) by incubating with gentle end-over-end rotation, 5 minutes/wash. The cleared lysate from mESC cells or embryonic testes was added to the washed beads. The bead-lysate suspension was incubated at 4°C for two hours with end-over-end rotation. The flow-through was collected in a separate tube, and the beads were washed with the IP buffer thrice with the same parameters as the earlier washes.

The bound proteins were eluted for western blotting by boiling the beads at 95°C for 10 minutes in 2x NuPAGE™ LDS Sample Buffer (Thermo Fisher Scientific NP0007) containing 100mM DTT (Sigma Aldrich). Proteins in the supernatant were size separated on an SDS-PAGE gel. The proteins were then transferred onto the Nitrocellulose membrane using a wet tank blotting system (Bio-Rad, USA). The membrane was blocked with 5% (w/v) BSA (Carl Roth 8076.3) prepared in 1X TBST (20mM Tris-HCl pH 8.0, 50mM NaCl, 0.1% Tween-20) for 1 hour at room temperature. The membranes were incubated with different primary antibodies in the blocking buffer overnight at 4°C. The membranes were washed thrice with 1X TBST, 10 minutes per wash. Anti-rabbit (Thermo Fisher Scientific 31460) or Anti-mouse (Thermo Fisher Scientific 31430) Secondary antibody conjugated with HRP diluted in 1X TBST was added to the membranes and incubated for 1 hour at room temperature followed by three washes. The blots were developed using SuperSignal™ West Pico PLUS or SuperSignal™ West Femto chemiluminescent reagents (Thermo Fisher Scientific – 34580 or 34095). The Gel Doc XR+ Gel Documentation System (Bio-Rad) was used for imaging the gels and blots.

5.4 SAMPLE PREPARATION FOR MASS SPECTROMETRY

IP-enriched proteins were eluted by boiling the beads in 2x NuPAGE™ LDS Sample Buffer containing 1 mM DTT for 10 min at 70 °C. Chloroacetamide (Sigma Aldrich 22790) was added to a concentration of 5 mM, and the samples were incubated in the dark for 40 min at room temperature. Afterward, samples were separated on an SDS-PAGE gel and stained using the Colloidal Blue Staining Kit (Life Technologies LC6025). Each lane was cut into 2 pieces for separate processing, and each piece was further cut into 1 mm-sized squares. The gel pieces were destained 4 times with 1 mL destaining solution (50% Ethanol, 50 mM ammonium bicarbonate (ABC) buffer pH 8.0) for 20 minutes and dehydrated by incubating 2 times with 1 mL 100% EtOH for 10 min, each with rotation. To digest the proteins, 50µL of trypsin solution (125 ng/µL trypsin in 25 mM ABC buffer, pH 8.0) was added to the pieces and incubated at 500 rpm for 2 min. The pieces were covered with 25 mM ABC buffer and incubated overnight at 37°C. The next day, the digest was stopped by adding 50µL of peptide extraction

buffer 1 (30% ACN, 3% TFA). After 20 min incubation at 500 rpm, the supernatant was transferred to a fresh tube. Peptide extraction was further carried out by sequential incubation with extraction buffer 1, peptide extraction buffer 2 (80% ACN, 0.5% acetic acid), and ACN using a volume big enough to cover the gel pieces (~100µL). Supernatants from the same gel piece were combined, concentrated by vacuum centrifugation at 45 °C to a volume of 100µL, and subjected to StageTip purification (Rappsilber, Mann and Ishihama, 2007).

Peptide purification was carried out using self-made and extremely economical stop-and-go-extraction tips (StageTips). In short, StageTips were built by cutting out 2 disks from a C18 47 mm Empore extraction disks (CDS Analytical) with a 17-gauge Hamilton syringe and placed into a 200 µL pipette tip. StageTips were equilibrated with each 25µL of methanol, followed by 25µL stage tip buffer B (80% ACN, 0.1% formic acid) and 2 x 25µL stage tip buffer A (0.1% formic acid) by centrifuging at 500 x g. The samples were loaded on the tip and washed with 50µL buffer A in the same fashion. The peptides were eluted with 50µL elution buffer (50% ACN, 0.1% formic acid) into 96-well sample plates by applying pressure with a syringe and vacuum concentrated to a volume of 2µL. Finally, 3µL of 0.1% FA was added.

5.5 MASS SPECTROMETRY DATA ACQUISITION

Peptide fractions were analyzed on quadrupole Orbitrap mass spectrometers (Exploris 480) equipped with a UHPLC system (EASY-nLC 1200, Thermo Scientific). Peptide samples were loaded onto C18 reversed-phase columns (length 55 cm, inner diameter 75 µm; Reprosil Pur 1.9 µm, Dr Maisch) and eluted in a linear gradient from 3 to 42% acetonitrile containing 0.1% formic acid in 90 min. The mass spectrometer was operated in Data-dependent acquisition (DDA) mode, automatically switching between MS1 and MS2 acquisition. Survey full-scan MS spectra (m/z 300–1650) and fragment spectra were acquired in the Orbitrap (Bekker-Jensen et al., 2020; Michalski et al., 2011; Olsen et al., 2007). The 15 most intense ions were sequentially isolated and fragmented by higher-energy C-trap dissociation (HCD) (Olsen et al., 2007). Peptides with unassigned charge states, as well as with charge states less than +2 were excluded from fragmentation.

5.6 PROTEOMIC DATA ANALYSIS USING MAXQUANT AND PERSEUS

Raw data files were analyzed using MaxQuant (version 1.6.14) (Cox and Mann, 2008). Parent ion and MS2 spectra were searched against a database containing 63,666 mouse protein sequences obtained from the UniProtKB released in January 2020 using the Andromeda search engine (Cox et al., 2011). Spectra were searched with a mass tolerance of 6 p.p.m in MS mode, 20 p.p.m in HCD MS2 mode, strict trypsin specificity, and allowing up to two miscleavages. Cysteine carbamidomethylation was searched as fixed, while N-terminal acetylation and methionine oxidation were searched as variable modifications. Match between runs was activated. The dataset was adjusted to a 1% false discovery rate on a peptide spectrum match level based on a target-decoy approach using reversed protein sequences.

Statistical analysis and data filtering was performed using Perseus software v1.6.14.0 (Tyanova et al., 2016). Proteins “only identified by site” , “reverse hits”, or “potential contaminants” were removed. Only proteins identified with at least two peptides, including at least one unique peptide, were used for the analysis. Peptides were further filtered for at least two valid values in at least one of the treatment groups. Missing values were randomly imputed from a normal distribution as implemented in Perseus (downshift = 1.8 s.d., width = 0.3 s.d.). To identify significantly enriched proteins, an FDR-controlled t-test based on the SAM package was performed (Li and Tibshirani, 2013). The FDR cut-off was set to 5% and s0 to 0.1.

5.7 GENOMIC DNA EXTRACTION FOR PCR GENOTYPING

Genomic DNA for PCR genotyping for different mouse strains was performed using two approaches. The quicker approach was as follows.

1. To the ear clip or neonatal phalange, 50µL or 30µL of extraction solution was added, respectively.
    a. Extraction solution – 100mM NaOH, 2.5mM EDTA.
2. The tissue was dissolved at 95°C for 10 minutes.
    a. The tube was checked regularly to avoid complete evaporation

3. The contents were spun down, and the tube was placed on ice. For one PCR reaction, 1-1.5µL of this gDNA solution was used for a 25µL PCR reaction. The gDNA was stored at 4°C for a maximum of one day or frozen at -20°C for a week.

The slower approach was as follows.

1. To the ear clip or neonatal phalange, 50µL or 30µL respectively of lysis buffer was added.
    a. Lysis buffer - 100mM Tris-HCL pH 8.0, 5mM EDTA, 0.2% SDS, 200mM NaCl, 0.2mg/mL Proteinase K.
2. The tissue was digested 4h to overnight at 50-55°C.
3. Chloroform, 50µL, was added to each tube and homogenized by vortexing followed by centrifugation at top speed for 10 min at room temperate.
4. The aqueous supernatant was transferred to a new clean tube containing 150µL of 100% Ethanol (Fisher Scientific 17740239)
5. The contents were mixed by inverting the tube a few times followed by centrifugation for 10 min at 4°C at top speed.
6. The supernatant was carefully discarded. The Genomic DNA (gDNA) pellet was air-dried and resuspended in 50µL of 1X low EDTA TE buffer (10mM Tris pH 8.0, 0.5mM EDTA). For each 25µL PCR reaction, 1-1.5µL of gDNA was used. This gDNA was stored indefinitely at -20°C.

5.8 GENOTYPING PCR

PCR recipe

Component	**Volume (µL)**
5X GoTaq Green buffer	5
10mM dNTPs	0.5
Forward primer (10µM) (Table 1)	0.5

Reverse primer (10µM)	0.5
Template DNA	1
Screen Blend DNA polymerase	0.25
Nuclease free Water	17.25

PCR Program

Temperature (°C)	**Time**	Cycles
94	2 minutes	1 X
94	30 seconds	35 X
Annealing Temperature	30 seconds	
68	**Extension time**	
68	5 minutes	1X
16	Hold	-

Table 1: **Genotyping primers, Annealing, and extension parameters**

Allele	**Genotyping primers (5' → 3')**	**Annealing temp (°C)**	**Extension time (s)**	**Amplicon size (bp)**
Dnmt3C$^{myc/myc}$	F: CCATGGTGGGAAAGTTGAG	51	30	WT: 217
	R: GACTTGGGTCAAAGGAACAGT			KI: 289
Dnmt3C$^{KO/KO}$	WT F: TGGCTCCCACAGATCTTACC	58	30	WT: 216
	KO F: AACCCTCCAAATCCTTAGGC			KO: 128
	R: ACTATAGCCCATCAGTAAAGCC			
Sry	F: TTGTCTAGAGAGCATGGAGGGCCATGTCAA	60	30	WT: 273
	R: CCACTCCTCTGTGACACTTAAGCCCTCCGA			
Xist	F: CACCAGACTGCCCTGAGAAA			WT: 167
	R: TTCTGTCCAGCCAGCTCAAG			
Oct4$^{GFP/GFP}$	F: CAAGGCAAGGGAGGTAGACA			WT: 434

	WT R: TGCCAGACAATGGCTATGAG	55	45	GFP: 234
	GFP R: CCAAAAGACGGCAATATGGT			
Lsd1^{cKO}	WT F: CCTCAGTAGGCCTGGTTTGT	56	40	WT: 346
	cKO F: CCGTGGAAATTCGTGCACTC			Lox: 490
	R: TTGGTTTTGGTTGACCCTTC			cKO: 400
Lsd1^{cKI}	F: CCAGCTGCTTGTTGGTGC	60	30	WT: 217
	R: TGGAGTGAAGTGGTTACCTGC			Lox: 336
				cKI: 300

5.9 GERM CELL ISOLATION

5.9.1 Prospermatogonia (Gonocytes) isolation

Mice with the *Oct4*$^{GFP/GFP}$ allele were used for Prospermatogonia isolation. Testes from stages E14.5-P2 were dissected from embryos and kept in 1X PBS pH 7.4 (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4). Testes from embryonic stage E14.5 were isolated along with the mesonephros and testes from stages E16.5-P2 were isolated and decapsulated using fine forceps. They were transferred to a tube containing Collagenase digestion buffer (25µL/pair of testes – 25mM HEPES pH 7.3, 2µL Collagenase stock for every 20µL of HBSS (with Ca+ and Mg+), 2X MEM Non-Essential Amino Acids Solution, 2X Sodium Pyruvate in Hanks' Balanced Salt Solution) and digested at 37°C for 10 minutes with constant shaking and occasional pipetting. TrypLE™ Express Enzyme (100µL/pair of testes) was added, and the tube was incubated at 37°C for 10 minutes with constant shaking for further digestion to a single cell suspension. The TrypLE™ was neutralized by adding FBS (35µL/pair of testes). The cells were centrifuged at 700g for 3 minutes and washed with wash buffer (25mM HEPES pH 7.3, 1X PBS pH 7.4, 1.5% BSA, 10% FBS). Finally, the cells were resuspended in FACS buffer (25mM HEPES pH 7.3, 1X PBS pH 7.4, 1.5% BSA, 2µg/mL DAPI) and filtered with 40µm cell strainer (pluriSelect 43-10040-60). Sorting was performed with a 100µm nozzle using a BD FACSAria™ III Cell Sorter (BD Biosciences). DAPI-negative and GFP-positive singlet cell populations are the Prospermatogonia. They were sorted into 1.5mL

microcentrifuge tubes (Sarstedt 72.706) containing collection buffer (1X PBS pH 7.4, 0.5% BSA).

5.9.2 Spermatogonial stem cells and Spermatogonia isolation

Testes from juvenile (P10 – P15) males were dissected and washed briefly in 1X PBS pH 7.4. They were decapsulated and the seminiferous tubules were pulled apart in Collagenase digestion buffer (300µL/pair of testes) using fine forceps. This suspension was then transferred to a 1.5mL microcentrifuge tube and incubated at 37°C for 10 minutes with occasional pipetting. TrypLE™ Express Enzyme (1200µL/pair of testes) was added for further digestion to obtain a single cell suspension. The TrypLE™ was neutralized by adding FBS (420µL/pair of testes). The cells were filtered through a 40µm cell strainer to remove big clumps. The cell suspension was then centrifuged at 700g for 3 minutes, washed twice with wash buffer (25mM HEPES pH 7.3, 1X PBS pH 7.4, 1.5% BSA), and finally resuspended in 500µL of wash buffer. For FACS sorting, cells were stained with CD326 (Ep-CAM)- Alexa Fluor® 647 (BioLegend 118212) and β-2-Microglobulin Antibody – PE (Santa Cruz Biotechnology sc-32241 PE) and DAPI. Sorting was performed with a 100µm nozzle using a BD FACSAria™ III Cell Sorter (BD Biosciences) or Bigfoot Spectral Cell Sorter (Thermo fisher Scientific). The DAPI-negative, Ep-CAM-positive, and β-2-Microglobulin-negative singlet population are the Spermatogonial stem cells and Spermatogonia. They were sorted into 1.5mL microcentrifuge tubes containing collection buffer or 1X PBS pH 7.4.

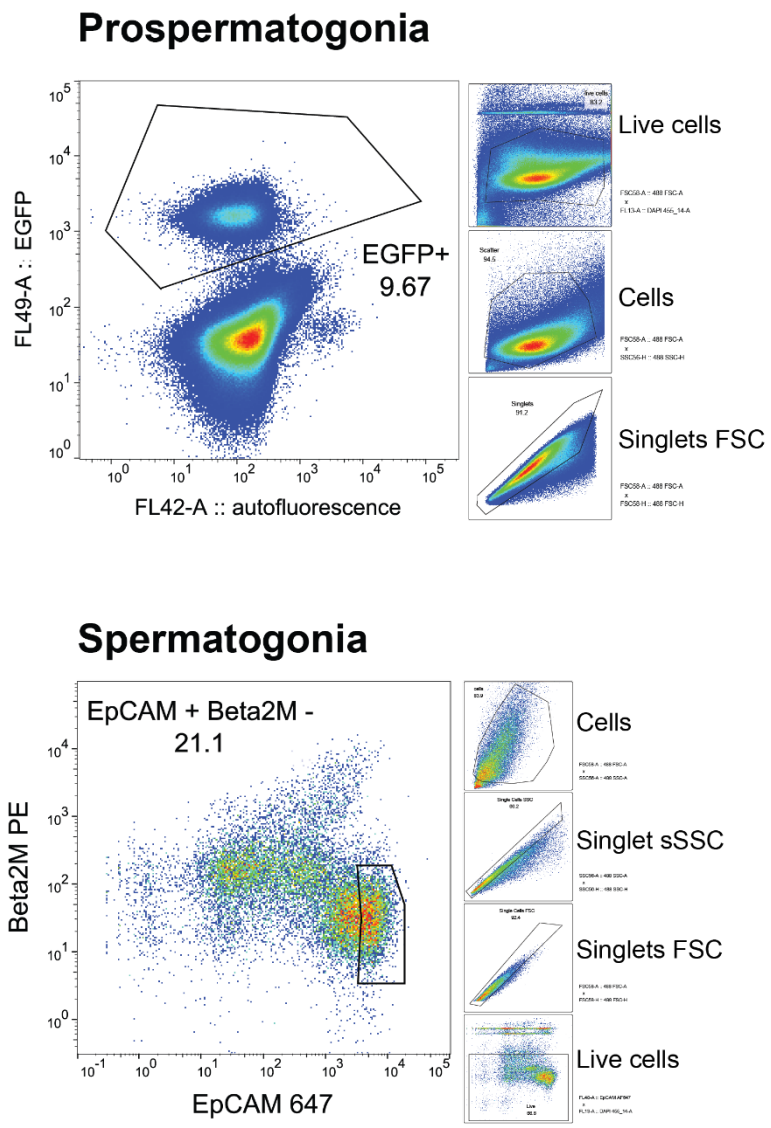


Figure 5.1: **FACS profiles for Prospermatogonia and Spermatogonia**. The sorted population is shown inside the gates.

5.10 CLEAVAGE UNDER TARGETS AND RELEASE USING NUCLEASE

Cleavage Under Targets and Release Using Nuclease (CUT&RUN) protocol was modified from Skene et al., 2017 (Skene and Henikoff, 2017).

5.10.1 CUT&RUN Protocol

1. Concanavalin A coated magnetic beads (Polysciences 86057-10), 10 - 15µL/sample, was used. The beads were activated by washing them thrice in Bead Activation

buffer. The beads were resuspended in bead activation buffer in a volume which was 10% of the volume of the cell suspension.

   a. Bead activation buffer – 20mM HEPES pH 7.3, 10mM KCl, 1mM $CaCl_2$, 1mM $MnCl_2$.

2. The activated beads were gently added to FACS-sorted germ cells (in collection buffer) or mESCs (in 1X PBS) without creating bubbles.
3. Cells were allowed to bind to beads for 20 minutes at room temperature by gently end-over-end rotation.
4. The tube was briefly spun down. The beads were separated on a magnetic rack (Thermo Fisher Scientific 12321D) and the supernatant was discarded.
5. Dig wash buffer, 25µL/sample, was added to the beads and mixed by gentle pipetting.
   a. Dig wash buffer - 20 mM HEPES pH 7.3, 150 mM NaCl, 0.5 mM Spermidine, 1X Protease inhibitors, x % Digitonin.
   b. Digitonin stock solution 5% was prepared in water and heated at 99°C for 2 minutes or until all digitonin dissolved. Final concentration used – 0.05% for all germ cells and 0.1% for mESCs.
6. The cell-bead suspension was split into 25µL aliquots in PCR strips.
7. Primary antibodies were prepared in 1:50 dilution in Dig wash buffer and 25µL was added to each tube. Final volume per tube was 50µL.
8. The cells with primary antibody were incubated overnight at 4°C.
9. Next day, the beads were washed with the Dig wash buffer. For each wash, the beads were separated on a magnetic rack (New England Biolabs S1515S). Supernatant was discarded and 75µL of Dig wash buffer was added and the tubes were kept aside for 4 minutes without any mixing. After 4 minutes, beads were separated, and another wash was performed.
10. After washes, 75µL of Protein A-MNase fusion protein (produced by the Recombination Protein Platform of the Institut Curie, 0.785 mg/mL) in Dig wash buffer

was added to the samples and incubated at room temperature for 1 hour on a nutator (Corning CGW6720).

   a. During this incubation period, digestion solution (Dig wash buffer with 9mM $CaCl_2$) was prepared and kept on ice.

11. Two washes were performed with Dig wash buffer and the beads were resuspended in 75µL of Dig wash buffer. The tubes were cooled down on ice for 5 minutes.
12. Ice cold digestion solution, 25µL, was added to each sample and mixed immediately by flicking the tubes. Final $CaCl_2$ concentration is 2.25mM. The tubes were kept back on ice and incubated for 30 minutes.
    a. As MNase only needs Ca^{2+} ions for activity, lowering the temperature reduces background cleavage caused by diffusion of the fusion protein.
13. After digestion, 100µL of 2X Stop solution was added to each sample and mixed by pipetting.
    a. 2X Stop solution – 20mM HEPES pH 7.3, 340 mM NaCl, 20 mM EDTA, 6 mM EGTA, 50ug/mL RNase A, 0.05% or 0.1% Digitonin.
14. Samples were incubated in a 37°C water bath for 20 minutes to release the DNA fragments into the supernatant.
15. Cells in beads were separated on the magnetic rack and the supernatant was collected in a fresh 1.5mL microcentrifuge tube.
16. Proteins were digested by adding 2µL of 10% SDS and 1.5µL of 20mg/mL Proteinase and incubating at 55°C for 1 hour.
17. DNA was extracted using the DNA Clean & Concentrator-5 kit (Zymo D4014) according to the manufacturer's instructions. The DNA was eluted in 15µL of elution buffer and stored at -20°C.

5.10.2 CUT&RUN Library Preparation

Libraries from CUT&RUN DNA were prepared using the "xGen™ ssDNA & Low-Input DNA Library Preparation Kit" from Integrated DNA Technologies (10009817) according to the

manufacturer's instructions. Indexing primers from the "xGen UDI Primer Plate 2, 8nt, 96rxn" kit (Integrated DNA Technologies 10009816) were used for unique dual indexing.

5.11 CLEAVAGE UNDER TARGETS AND TAGMENTATION

Cleavage Under Targets and TAGmentation (CUT&TAG) protocol was modified from Kaya-Okur et al., 2019 (Kaya-Okur et al., 2019). The protocol is as follows.

5.11.1 Transposome Preparation

1. Mosaic end-adapter A (ME-A), Mosaic end-adapter B (ME-B), and Mosaic end–reverse oligonucleotides (ME-Reverse) were diluted to 200 µM in annealing buffer (10mM Tris pH8, 50mM NaCl, 1mM EDTA).

Oligonucleotide	**Sequence (5' → 3')**
Mosaic end-reverse	[PHO]CTGTCTCTTATACACATCT
Mosaic end-Adapter A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Mosaic end-Adapter B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

2. Each pair of oligos, ME-A+ME-Reverse and ME-B+ME-Reverse, were mixed separately in a PCR tube, resulting in a 100 µM annealed product.
3. The tubes were placed in a PCR machine with the following program.

Temperature (°C)	**Time**	**Cycles**	**∆T(°C)**
95	5 minutes	1X	
95↓	1 minute	71X	-1°C/cycle
25	5 minutes	1X	
4	Hold	-	

4. Annealed ME-A+ME-Reverse, 8µL, and annealed ME-B+ME-Reverse, 8µL, were mixed with 100µL of 5.5µM Protein A-Transposase Tn5 fusion protein (produced by the Protein production team at Institute of Molecular Biology.
5. The contents were mixed by gentle pipetting and the tube was mixed by gentle end-over-end rotation for 1 hour at room temperature and then stored at -20°C.

5.11.2 CUT&TAG Protocol

1. Concanavalin A coated magnetic Beads and the cells were prepared the same way as described for CUT&RUN.
2. Primary antibody incubation was also performed the same way as for CUT&RUN.
3. Next day, the beads were washed with the Dig wash buffer. For each wash, the beads were separated on a magnetic rack. Supernatant was discarded and 75µL of Dig wash buffer was added and the tubes were kept aside for 4 minutes without any mixing. After 4 minutes, beads were separated, and another wash was performed.
4. After washes, 50µL of Guinea pig anti-Rabbit IgG secondary antibody (antibodies-online.com ABIN101961) 1:100 dilution in Dig wash buffer was added and the samples were incubated at room temperature for 1 hour on the nutator.
5. Two washes with 75µL of Dig wash buffer was performed.
6. After washes, 75µL of the adapter loaded Protein A-Tn5 diluted 1:100 in Dig 300 buffer was added to the samples and incubated for 1 hour at room temperature on the nutator.
    a. Dig 300 buffer - 20 mM HEPES pH 7.3, **300 mM NaCl**, 0.5 mM Spermidine, 1X Protease inhibitors, x % Digitonin.
    b. Digitonin – 0.05% for all germ cells and 0.1% for mESCs.
7. Two washes with 100µL of Dig 300 buffer was performed.
8. For Tagmentation, 100µL of tagmentation buffer (Dig 300 buffer with 10mM $MgCl_2$) was added to each sample and incubated in a 37°C water bath for 1 hour.

9. After tagmentation, 5µL of 0.5M EDTA, 1.5µL of 10% SDS, and 1.25µL of 20mg/mL Proteinase K were added to each sample. The tubes were mixed by vortexing and incubated in a 55°C water bath for 1 hour.
10. After incubation, the tubes were vortexed heavily again. Beads were separated on the magnetic rack, and the supernatant was used for DNA purification.
11. DNA was extracted using the DNA Clean & Concentrator-5 kit (Zymo D4014) according to the manufacturer's instructions. The DNA was eluted in 25µL of elution buffer, stored at -20°C, or used directly for library preparation.

5.11.3 CUT&TAG Library Preparation

1. As the adapters were already inserted into the DNA by the Transposome, the library preparation is simply a PCR step.

   PCR Recipe

Component	**Volume (µL)**
5X Q5 reaction buffer	10
5X Q5 GC enhancer	10
10mM dNTP	1
10µM indexing Primer mix (Forward & Reverse)	1.25
Template DNA	25
Q5 DNA polymerase	0.5
Nuclease free Water	2.25

   PCR Program

Temperature (°C)	**Time**	**Cycles**
72	5 minutes	1X
98	30 seconds	1X
98	10 seconds	16 X for germ cells 14X for mESC
63	30 seconds	

72	1 minute	1X
4	Hold	-

5.11.4 Library Purification.

1. In-house prepared SPRI beads (Cytiva GE65152105050250) were used to clean up the PCR. The beads were equilibrated to room temperature for 30 minutes before use.
2. The PCR tubes were removed from the cycler and briefly spun down. For the clean-up, 1.3X volume (65µL) of the SPRI bead slurry was added and mixed by vortexing for 20 seconds.
3. The tubes were kept at room temperature for 20 minutes.
4. The tubes were kept on the magnetic rack for 2 minutes or until the solution became clear.
5. Supernatant was discarded, and the beads were washed twice while still on the magnetic rack with 200µL of 80% freshly prepared Ethanol.
6. The tubes were briefly spun and placed on the magnetic rack, and any remaining Ethanol was aspirated using a 20µL pipette.
7. Tubes were removed from the magnetic rack, and the beads were resuspended in 15-20µL of low EDTA TE buffer (10mM Tris pH 8.0, 0.5mM EDTA)
8. Elution was performed at room temperature for 10 minutes.
9. The tubes were placed on the magnetic rack, and the supernatant containing the library was transferred to clean 1.5 mL microcentrifuge tubes.

5.12 CUT&TAG – BS

CUT&TAG with Tagmentation-based Bisulfite Sequencing (CUT&TAG-BS) protocol was adapted and modified from Li et al., *Cel Rep Methods* 2021 (Li, Grimm and Wade, 2021).

5.12.1 Transposome Preparation with Methylated Adapters

1. The oligonucleotides were resuspended in annealing buffer (10mM Tris-HCl pH 8. , 50mM NaCl, 1mM EDTA) to a final concentration of 200mM.

Oligonucleotide name	Sequence (5' → 3')	Modifications
Tn5mC-Apt1	TcGTcGGcAGcGTcAGATGTGTATAAGAGAcAG	c: 5C-methylated, p: phosphate, ddC: dideoxycytidylate, invT: inverted deoxythymidylate
Tn5mC1.1-A1block	pCTGTCTCTTATACAddC	
Tn5mC-ReplO1	pcTGTcTcTTATAcAcATcTccGAGccCAcGAGAci nvT	

2. Equal volumes of the oligos Tn5mC-Apt1 and Tn5mC1.1-A1block were mixed in a PCR tube and annealed in a PCR using the annealing program.

   Annealing Program

Temperature (°C)	Time	Cycle	ΔT(°C)
95	2 minutes	1X	
90 ↓	5 minutes	14X	-5°C/cycle
25	5 minutes	1X	
4	Hold	-	

3. Annealed oligos, 16µL, were mixed with 100µL of 5.5µM Protein A-Transposase Tn5 fusion protein (produced by the Protein production team at the Institute of Molecular Biology.
4. The contents were mixed by gentle pipetting, and the tube was mixed by gentle end-over-end rotation for 1 hour at room temperature and then stored at -20°C.

5.12.2 CUT&TAG-BS Protocol

The CUT&TAG-BS protocol is the same as the CUT&TAG protocol with two modifications – (1) Protein A-Tn5 loaded with the methylated adapters is used instead of the Protein A-Tn5 loaded with unmethylated adapters. (2) The elution volume for DNA isolation in the last step is 12µL.

5.12.3 Oligonucleotide Replacement and Gap Repair

1. As the Tn5mC-A1Block oligonucleotide is unmethylated and not covalently bound to the genomic DNA, it was replaced with a methylated adapter oligonucleotide Tn5mC-Repl01. The 9-nucleotide gap is repaired by DNA polymerase and DNA ligase.

Component	**Volume (µL)**
CUT&TAG-BS Eluted DNA	12
10X Ampligase buffer	2
Tn5mC-Repl01	2
10mM dNTP mix	2

Step	**Temperature**	**Time**	**Ramp (°C/s)**	**Cycles**
Denature	50	1	6	1
Anneal	45	10	6	
Anneal	45 ↓		-0.1	1
	37	Hold		-

2. With the tubes in the PCR block, 1µL of T4 DNA polymerase (New England Biolabs M0203S) and 2.5µL of Ampligase (Lucigen) were added. The contents were mixed well with a 20µL pipette.
3. The reaction was incubated at 37°C for 30 minutes in the PCR machine for gap repair.

5.12.4 Bisulfite Treatment and DNA Purification

DNA from the previous step was bisulfite converted using the EZ DNA Methylation-Gold Kit from Zymo (D5006) using the manufacturer's instructions. The DNA was eluted in 13µL of elution buffer.

5.12.5 Library Preparation

1. Library preparation for CUT&TAG-BS is also a PCR as the adapters are already on the DNA molecules. As these DNA molecules consist of Uracil due to Bisulfite conversion, we used a DNA polymerase that tolerates Uracil residues.

   PCR Recipe

Component	**Volume (µL)**
KAPA HiFi HotStart Uracil+ReadyMix (2X)	15
10mM dNTP	1
10µM indexing Primer mix (Forward & Reverse)	1.25
Adapter ligated Template DNA	13

   PCR Program

Temperature (°C)	**Time**	**Cycle**
95	3 minutes	1X
98	20 seconds	15X
62	15 seconds	
72	1 minute	
72	1 minute	1X
4	Hold	-

5.12.6 Library Purification

Library purification for CUT&TAG-BS was the same as for CUT&TAG.

5.13 DNA METHYLATION ANALYSIS BY PYROSEQUENCING

Targeted analysis of DNA methylation at retrotransposon promoters was performed by Bisulfite conversion and Pyrosequencing.

1. FACS sorted Spermatogonia and Spermatogonial Stem cells were pelleted and resuspended in 12µL of 1X PBS.
2. The DNA within the cells was converted directly to bisulfite using the EZ DNA Methylation-Direct Kit (Zymo D5020) according to the manufacturer's instructions. The DNA was eluted in 10µL of elution buffer and stored at -20°C.
3. PCR was performed to amplify the targets using the following parameters.

   PCR recipe

Component	**Volume (µL)**
5X Green GoTaq® Flexi Buffer	10
25mM MgCl2 solution	3
10mM dNTP	1
10µM Forward primer (Table 2)	1
10µM Reverse primer	1
GoTaq® DNA Polymerase (5u/µl)	0.25
Template DNA	1
Nuclease free water	32.75

PCR Program

Temperature (°C)	**Time**	**Cycle**
94	45 seconds	1X
94	15 seconds	30X
Annealing temperature	25 seconds	
72	15 seconds	
72	5 minutes	1X
4	Hold	-

Table 2: **Primers and Amplicon parameters for Pyrosequencing**

Target	**Primers (5' → 3')**	**Annealing Temperature (°C)**	**Amplicon size (bp)**
LINE1 A	F : AGATTGAGGTATATAGGGAAGTAGGTT	55	172
	R: [Btn]ATCCACTCACCAAAAATCTTAAAAT		
	Seq: GGTATATAGGGAAGTAGGTTA		
LINE1T	F: GGTTGGGGAGGAGGTTTAAGTTATA	55	130
	R: [Btn]CTACCTATTCCAAAAACTATCAAATTCTCT		
	Seq: GGGAGGAGGTTTAAGTTATAGTA		
IAP	F: GAGGGTGGTTTTTTATTTTATGTGT	58	82
	R: [Btn]ATCACTCCCTAATTAACTACAACC		
	Seq: TTTTTATTTTATGTGTTTTGTTTTT		

4. Pyrosequencing and analysis were performed using the Pyromark Q48 Autoprep instrument (Qiagen 9002471) and the PyroMark Q48 Autoprep Software version 2.4.2. The program was set using the following parameters, and the assay was performed according to the manufacturer's instructions.

Table 3: **Sequence information for Pyrosequencing**.

Target	LINE1A	LINE1T	IAP
Sequence before Bisulfite conversion (5' → 3')	CCCGGGCCTGATCTGGGGC ACAAGTCCCTTCCGCTCGA CTCGAGACTCGAGCTCCGG GCTACCTTGCCAGCAGAGT CTTGCCCAACACCCGCAAG GGCCCACACGGGACTCCC CA	GCAGCGGTCGCCATCTTGG TCCGGGACCCGCCGAACTT GGGAAATTAGTCTGAACAG GTGAGAGGGTGCGCCAGA GAACCTGACAGCTTCTGGA	CCCGTGACGACAACTCGGC CGATGGGCTGCAGCCAATC AAGGAGT
Sequence to analyze (5' → 3')	TTYGGGTTTGATTTGGGGTA TAAGTTTTTTTYGTTYGATTY GAGATTYGAGTTTYGGGTTA TTTTGTTAGTAGAGTTTTGTT TAATATTYGTAAGGGTTTATA YGGGATTTTTTA	GTAGYGGTYGTTATTTTGGT TYGGGATTYGTYGAATTTGG GAAATTAGTTTGAATAGGTG AGAGGGTGYGTTAGAGAAT TTGATAGTTTTTGGA	TTYGTGAYGATAATTYGGTY GATGGGTTGTAGTTAATTAA GGAGT
Dispensation order (generated by the Pyromark Q48 Autoprep software)	ATCTGTGATGGCTATATGTT CTAGTCTGTATCTGAGTATC TGATGTTCTGTATTGTAGTA GAGTTGTATGATCTGTAGTA TGATCGATT	AGCTATGTCAGTCGTATTAG TCTGTATCTAGTCTGATGATA GTGATAGTGAGAGTAGTCG TA	GTCTGTGTATCGACTGATCT AGTCTGAT
Number of CpGs	8	6	4

5.14 QUANTITATIVE PCR

Quantitative PCR (qPCR) was performed for CUT&RUN and CUT&TAG samples from mESC. qPCR) was performed using Power SYBR™ Green PCR Master Mix (Thermo Fisher Scientific 4367659) on a ViiA7 Real-Time PCR System (Applied Biosystems). The assay was performed in technical triplicates, and the samples were normalized to their respective IgG controls using the ΔΔ Ct approach. The qPCR primers for LSD1-CUT&RUN/TAG were designed by reanalyzing published LSD1 ChIP in mESCs (Cao et al., 2018).

Quantitative PCR program

Step	Temperature (°C)	Time	Cycle
Hold	50	2 minutes	1X
	95	10 minutes	1X
PCR	95	15 seconds	40X
	60	1 minute	
Melt curve	95	15 seconds	1X
	60	1 minute	1X
	95	15 seconds	1X

Primers used for Quantitative PCR for LSD1 targets

Target name	Primer sequences (5' → 3')	Amplicon size (bp)
Pou5f1	F: CAGATGCATAACAAAGGTGC	113
	R: CACAAAGCTTCCTCAATAGC	
Ckb	F: GGTGACCTCAAACAAAATGC	106
	R: TTGAGATGTATATGAGCTGCC	
Lrrc34	F: AGGCTACTAACTTCTTGGGG	107
	R: TCTGAAGAAAACAGCGTGG	
	F: GTCAGCAAAATCTTTCTGGC	

Negative control 1	R: CTGAGACAAAAGGAAGGACC	108
Negative control 2	F: AAAGAGAGAAGGGAGAGAGG	102
	R: CTCTCTGTCTTCTCCTCTCC	

5.15 HISTOLOGY

Embryonic testes were fixed in 1% Formaldehyde in 1X PBS pH 7.4 for 30 minutes at 4°C. Adult testes were fixed in 4% Formaldehyde in 1X PBS pH 7.4 overnight at 4°C. Post-fixation, the testes were placed in 15% sucrose solution in 1X PBS pH 7.4 overnight. Followed by an overnight incubation in 30% sucrose solution. Then the tissues were equilibrated in Tissue-Tek® O.C.T. Compound (Sakura 4583) for 1 hour before embedding in it. The cryoblocks were frozen on dry ice and stored at -80°C until sectioning.

The testes were sectioned into 9-10µm sections using a Cryostat (Leica CM3050 S). The sections were placed on a Superfrost Plus glass slide (Fisher Scientific 12312148) and stored at -20°C for the short term or -80°C for the long term.

5.16 IMMUNOFLUORESCENCE STAINING

The immunofluorescence protocol for testes went as follows. The immunofluorescence protocol for mESCs starts from step 4 in the below protocol.

1. Slides were taken out of the freezer and allowed to thaw at room temperature for 5-10 minutes.
2. A hydrophobic barrier pen (Biozol VEC-H-4000) was used to draw borders around the section.
3. The tissue was rehydrated by incubating it in 1X PBS pH 7.4 in a Coplin jar for 10 minutes at room temperature.
4. The sections were then kept in a humidified chamber and 40µL of Blocking and Permeabilization buffer was added for each section. This incubation was carried out for 40 minutes – 1 hour. All incubations were conducted in the humified chamber.

   a. Blocking and Permeabilization buffer (BP) – 1X PBS pH 7.4, 10% Horse serum, 3% BSA, 1X Protease inhibitor cocktail, 0.2% Triton X-100.

5. Primary antibody solution was prepared in BP buffer. Sections were incubated with the primary antibody at 4°C overnight.
6. Next day, the slides were washed with washing buffer in a Coplin jar for 15 minutes with gentle shaking. The washing buffer was changed twice.
   a. Washing buffer – 1X PBS pH 7.4, 0.05% Tween 20.
7. The sections were then incubated with a secondary antibody conjugated with a fluorophore for 1 hour at room temperature in the dark.
8. Tissue was washed as in step 6.
9. DAPI, 1 µg/ml, was added to the tissue and incubated for 5 minutes.
10. The tissue was washed in 1X PBS for 15 minutes.
11. After final washes, the sections were mounted with Immu-Mount™ (Fisher Scientific 10662815). The slides were stored at 4°C until imaging.

5.17 MICROSCOPY

Images were acquired using a Leica SP5 confocal microscope. All images corresponding to the same experiment were acquired and processed afterward, applying the same settings. Image editing was performed using the ImageJ software (Schneider, Rasband and Eliceiri, 2012).

5.18 BIOINFORMATIC ANALYSIS

5.18.1 RNA seq data analysis

Raw data files were downloaded from the European Nucleotide Archive (www.ebi.ac.uk/ena/) (Sangrithi et al., 2017). Reads were mapped using STAR (v2.7.3a) (Dobin et al., 2013) allowing up to 4% of the mapped bases to be mismatched (--outFilterMismatchNoverLmax 0.04 --outFilterMismatchNmax 999) and using 49 nt round splice junctions to construct the splice junctions database (--sjdbOverhang 49). The genome

assembly and annotation of GENCODE (Frankish et al., 2019) release M26 were used during mapping. Secondary hits were removed using Samtools (v1.10) (Danecek et al., 2021). Fragments per gene were counted using featureCounts of the Subread tool suite (v2.0.0) (Liao, Smyth and Shi, 2014). Parameter -p was set to count fragments instead of reads, parameter -s0 was used to indicate that the data is non-stranded and the gene annotation file gencode.vM26.primary_assembly.annotation.gtf of GENCODE release M26 was utilized when running featureCounts. Counts per gene were normalized by the total number of mapped fragments per sample (in millions) and the gene length (in thousands). Normalized counts are referred to as FPKMs (Fragments Per Kilobase of gene per Million fragments mapped).

5.18.2 Whole Genome Bisulfite Sequencing Analysis

Published WGBS datasets were taken from Barau et al., Science 2016, and Zoch et al., *Nature* 2020. (Barau et al., 2016; Zoch et al., 2020). All files downloaded from Zoch et al, *Nature* 2020 were concatenated based on the Array Express. Reference genome GRCm38 was downloaded from Gencode and prepared with Bismark with --bowtie2 flag. All files were trimmed with TrimGalore 0.6.6 with flags --length 25 --trim-n --clip_R2 5 –paired. Trimmed fastq files were mapped to the reference genome with Bismark (Krueger and Andrews, 2011) with option --score_min L,0,-0.4. Next, files were sorted with samtools, deduplicated with Picard MarkDupplicates REMOVE_DUPLICATES=true, and filtered with samtools view -q 30. Methylation calls were extracted with the Bismark methylation_extraction tool with bedgraph files as an output. Differentially methylated regions (DMR) were called separately for *Piwil4*$^{KO/KO}$, *Spocd1* $^{KO/KO,}$ and *Dnmt3C*$^{KO/KO}$ mutant compared with their respective WT controls with DSS tool with flags delta=0.25, minCG=5, minlen=200, dis.merge=500, p.threshold=0.05 and subsequently annotated with TxDb.Mmusculus.UCSC.mm10.knownGene. Transposable elements information was downloaded from repeatmasker database (mm10 - Dec 2011 - RepeatMasker open-4.0.6 - Dfam 2.0). L1 elements were filtered for length > 5 kb and LTR elements > 500 bp. CpG methylation data metaplots were processed using deepTools v3.5.4; only individual element annotations with contiguous sizes greater than 5 kb were analyzed for LINE1s and 500bp for MERVL and IAPEz (to remove solo LTR elements).

5.18.3 CUT&RUN / CUT&TAG data analysis

The CUT&RUN / CUT&TAG libraries were sequenced in the Illumina NextSeq platform with 150bp paired-end read sequencing.

Raw sequencing reads were filtered to remove any read pairs with a mean base quality score of less than 20 at either read end. Adapters were removed using TrimGalore 0.6.6. Trimmed reads were mapped to the Mus musculus reference genome mm10 using Bowtie2 v 2.5.3 (Langmead and Salzberg, 2012), reporting randomly one position. CUT&TAG fragments are generated according to their DNA accessibility to Tn5 digestion. Due to this, they can share exact starting and ending positions. Therefore, duplicates observed in CUT&TAG are not necessarily a consequence of overamplification by PCR (Henikoff et al., 2020; Zheng, Ahmad and Henikoff, 2020) and thus, for CUT&TAG duplicates were not removed. Samples were normalized to their respective IgG control. Visualization of the data using heatmaps and metaplots was performed using deepTools v3.5.4 (Ramírez et al., 2016) using the Galaxy platform ('The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update', 2024).

5.18.4 CUT&TAG – BS data analysis

CUT&TAG – BS data analysis was performed using an existing protocol (Li, Grimm and Wade, 2021).

Raw sequencing reads were filtered to remove any read pairs with a mean base quality score of less than 20 at either read end. Adapters were removed via Cutadapt v1.12 (Martin, 2011) with parameters “-a CTGTCTCTTATACAC -A CTGTCTCTTATACAC -O 5 -q 0 -m 20 -p”. For any read pairs in which the adapter was identified and removed, an additional nine bases were trimmed from the 3’ end of read1, as they correspond to the 9-base gap region filled in with unmethylated nucleotides. Genomic alignment was performed by Bismark v0.23.0 (Krueger and Andrews, 2011) with parameters “-X 1000 --non_bs_mm” (all other parameters as default) against the mm10 reference assembly (GRCm38) with Bowtie2 v2.3.0 (Langmead and Salzberg, 2012) as the underlying mapper. Positional methylation bias information was

reported by the Bismark v0.23.0 bismark_methylation_extractor tool with parameters “-p --include_overlap --mbias_only”. Per-residue methylation data was collected by the Bismark v0.23.0 bismark_methylation_extractor tool with parameters “-p --ignore_r2 9 --comprehensive --mbias_off --bedGraph --cytosine_report”. The bisulfite conversion rate per library was calculated by assessing methylated and unmethylated cytosine counts at the first 9 bases of read2.

RESULTS

6.1 DNMT3C Interactome in Mouse Embryonic Stem Cells

6.1.1 Optimization of DNMT3C isolation and co-immunoprecipitation.

To study the interactome of DNMT3C in mESCs, I used the cell lines *Dnmt*-tKO (*Dnmt1*$^{KO/KO}$, *Dnmt3a*$^{KO/KO}$, *Dnmt3b*$^{KO/KO}$) and *Dnmt*-tKO: *Dnmt3C* (stable-integration of CAG promoter-driven *Dnmt3c*$^{myc/myc}$ gene at the Rosa26 safe-harbor locus in addition to tKO). I used the *Dnmt*-tKO mESC system and not the WT mESC system for two reasons. (1) The closest cell culture model for male Prospermatogonia was mouse embryonic stem cells. (2) Within mESCs, *Dnmt*-tKO mESCs have very similar methylation patterns to Prospermatogonia after the wave of DNA demethylation (Reik, Dean and Walter, 2001). This has been shown to activate certain germline reprogramming responsive (GRR) genes in Prospermatogonia and mESCs (Hill et al., 2018). In combination with the fact that any other possible factors needed for DNA methylation are not diluted away from DNMT3C by DNMT3A or DNMT3B, makes *Dnmt*-tKO: *Dnmt3C* the closest *in-vitro* system to study DNMT3C's interactome.

I produced cell lysates using RIPA buffer (50mM Tris pH 8, 150mM NaCl, 1% Igepal CA-630, 0.5% Sodium deoxycholate, 0.1% SDS). I then performed immunoprecipitation using this lysate and Myc-Trap® beads (Proteintech) and confirmed the expression of the myc-tagged DNMT3C protein (Figure 6.1). The 2xmyc-DNMT3C protein is enzymatically functional as the genome-wide methylation percentages increase from 0% in tKO to 30.4% in *Dnmt*-tKO: *Dnmt3C* as measured by Luminometric methylation assay (Dura, 2021). Therefore, this *in vitro* system was suitable to study DNMT3C's protein interactome.

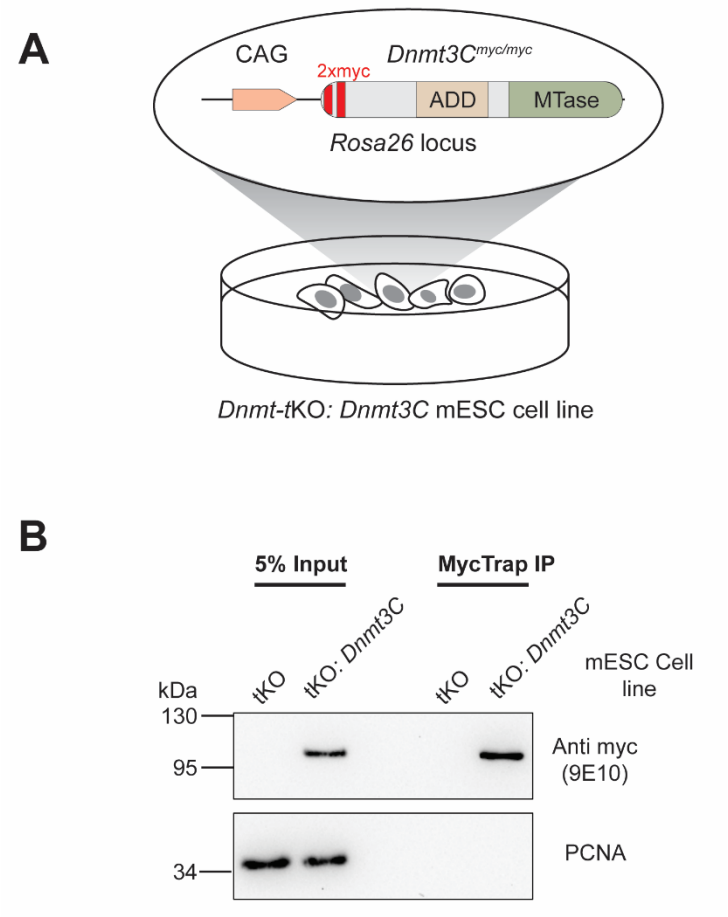


Figure 6.1: ***Dnmt-t*KO: *Dnmt3c* mESC cell line generation**. (A) Schematic of Dnmt3c transgene inserted into the Rosa26 locus. (B) Confirmation of 2xmyc-DNMT3C expression through immunoprecipitation

To achieve this, I fractionated *Dnmt*-tKO: *Dnmt3C* mESC lysate into cytoplasmic, nuclear soluble, and insoluble chromatin fractions. As shown in Figure 6.2A, most of the DNMT3C is associated with the insoluble chromatin. This could be because of DNMT3C's overexpression or affinity for heterochromatin. I also performed immunoprecipitation (MycTrap and control Rabbit IgG magnetic beads) on the RIPA lysate, NS420 nuclear soluble fractions, to check if there was any DNMT3C solubilized in the NS420 buffer. In addition, I also added the Sm nuclease to digest the DNA in order to check if that could help solubilize DNMT3C. As shown in Figure 6.2B, we can clearly see that no DNMT3C is solubilized with either NS420 or NS420 + Sm nuclease conditions. To solubilize DNMT3C, I treated the chromatin pellet with nuclear solubilization buffers with high salt concentration (NS buffers with 800mM/1200mM NaCl). This solubilized a significant portion of DNMT3C (Figure 6.2C). However, as these salt

concentrations are too high for an immunoprecipitation experiment, I tested if DNMT3C solubilization can occur at lower NaCl concentrations. A gradient of salt concentrations from 400-800mM NaCl was used to test this. As shown in Figure 6.2D, solubilization of DNMT3C needs a minimum of 600mM NaCl concentration. Such stringent isolation conditions of DNMT3C, similar to histone protein isolation, provide insight into its affinity for heterochromatin. This agrees with previous observations of overexpressed FLAG-tagged -DNMT3B and -DNMT3C localizing to high DNA-density regions (heterochromatin) in mESCs (Barau et al., 2016). Also, hypermethylated regions in *Dnmt*-tKO: *Dnmt3C* mESC are marked with H3K9me3, a characteristic heterochromatin marker (Dura, 2021), revealing DNMT3C's strong affinity for heterochromatin.

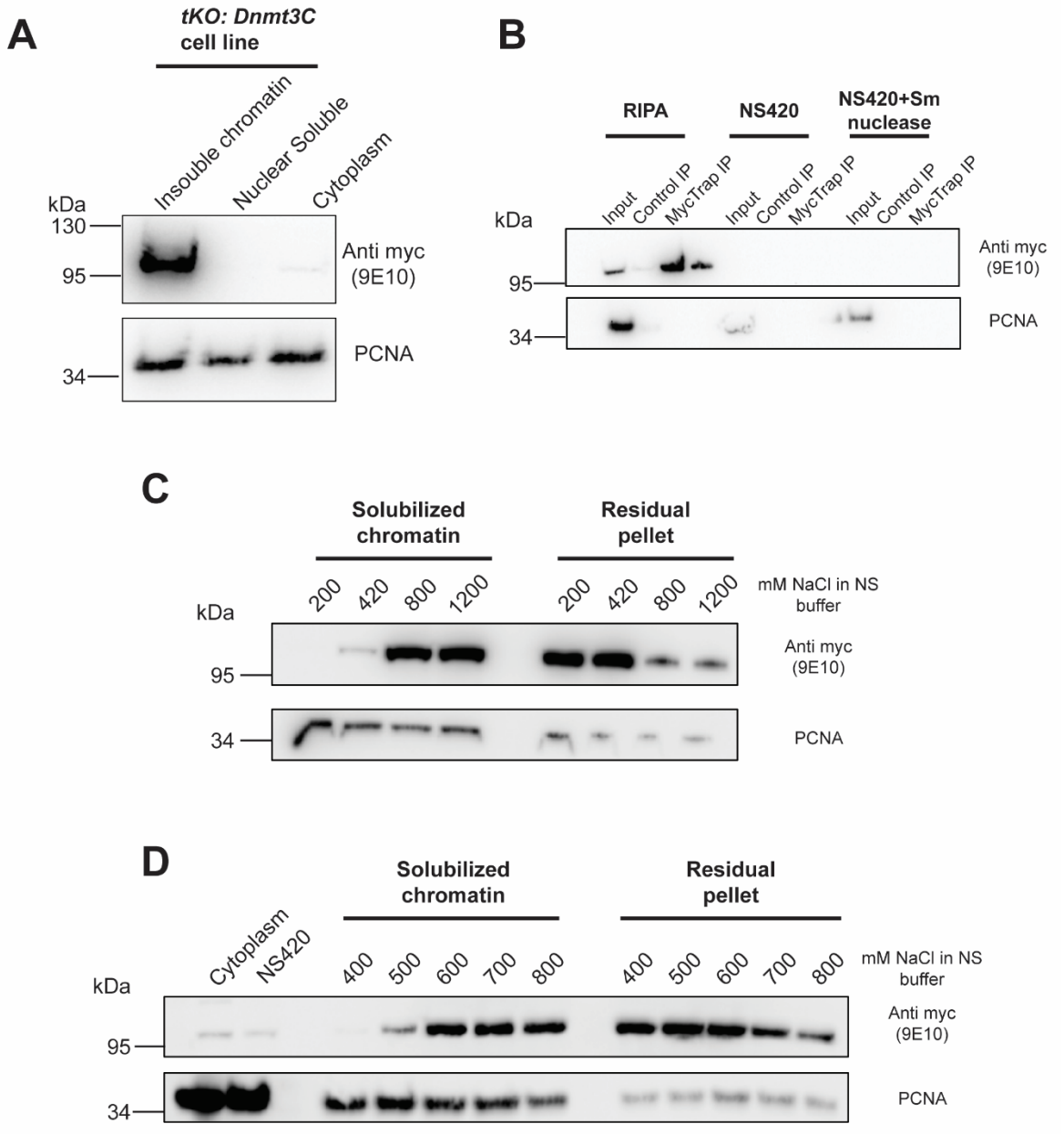


Figure 6.2: **Insolubility of DNMT3C** (A) DNMT3C is associated with the insoluble chromatin. (B) IP of nuclear soluble fraction shows a negligible amount of DNMT3C presence. (C) DNMT3C is solubilized from the chromatin at higher NaCl concentrations. (D) DNMT3C requires a minimum of 600mM NaCl concentration for solubilization

DNMT3C, being nuclear, could have interactors in both the nuclear soluble and insoluble chromatin fractions. I pooled the nuclear soluble and solubilized nuclear insoluble fractions to capture as many interactions as possible. I adjusted the salt concentration to 150mM NSON buffer (Nuclear solubilization buffer without NaCl). MycTrap agarose beads were added to this mixture and incubated overnight. The change in salt concentration led to a

precipitate formation, which, unfortunately, contaminated the IP by binding to the beads (Figure 6.3A). Therefore, I modified the nuclear isolation protocol from a 2-step isolation (NS420 treatment of nuclei followed by NS600 treatment of chromatin pellet) into a 1-step isolation (NS600 treatment of nuclei). The NaCl concentration was adjusted to 150mM with NSON buffer, and the mixture was incubated on a rotating wheel overnight at 4°C. The precipitated proteins were removed by centrifugation (14000g, 20 minutes, 4°C). The supernatant was used for immunoprecipitation with MycTrap magnetic agarose beads. This time, the IP was much cleaner (Figure 6.3B).

I performed the IP from both *Dnmt*-tKO and *Dnmt*-tKO: *Dnmt3C* using the magnetic MycTrap beads (Figure 6.3C). To validate the experiment, I probed the potential interaction of DNMT3C with DNMT3L. DNMT3L is an inactive cofactor for DNMT3A and DNMT3B proteins (Bourc'his et al., 2001; Chedin, Lieber and Hsieh, 2002; Suetake et al., 2004). Although not essential, it is a stimulatory factor for DNA methylation by DNMT3A and DNMT3B, and it is essential for the methylation at regions that DNMT3C targets. Supporting this, *Dnmt3l*$^{KO/KO}$ male mice show the same phenotype as *Dnmt3c*$^{KO/KO}$ male mice (Bourc'his et al., 2001) . The amino acid residues in DNMT3A and DNMT3B that promote their interaction with DNMT3L are conserved in DNMT3C as well (Barau et al., 2016). With a MycTrap immunoprecipitation followed by western blot, I showed biochemically that DNMT3C interacts strongly with DNMT3L (Figure 6.3D). This validated my approach, and I finalized the isolation protocol for DNMT3C. (1) Cytoplasmic fractionation using a hypotonic buffer, (2) nuclear isolation using NS600 buffer, (3) adjusting the salt concentration and overnight incubation to reconstitute protein-protein interactions, (4) removal of precipitate, (5) MycTrap immunoprecipitation using supernatant.

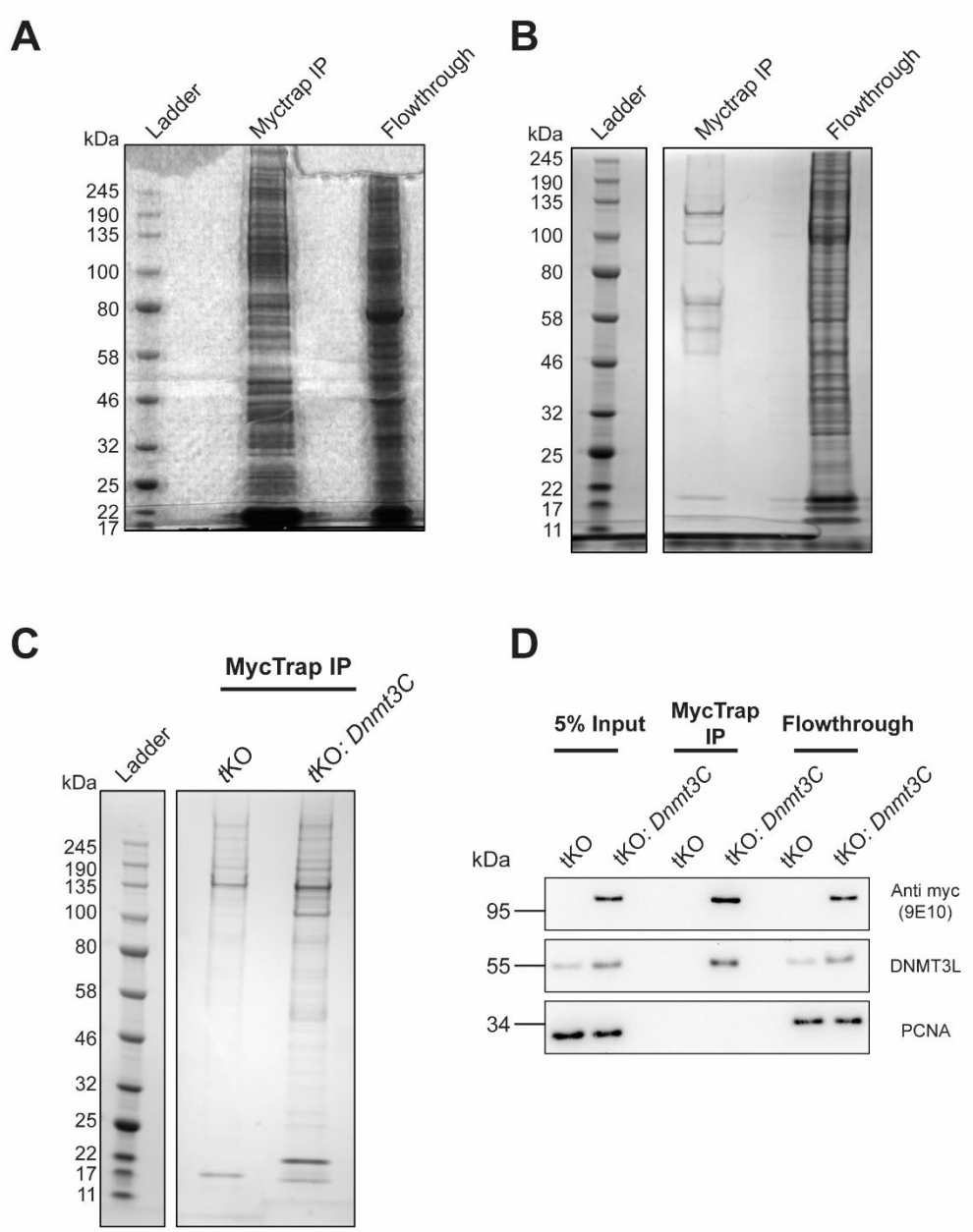


Figure 6.3: **Optimizing immunoprecipitation of DNMT3C**. (A) Combining nuclear soluble and solubilized chromatin and performing immunoprecipitation caused precipitate formation on the beads. (B) Clearing the precipitate led to a much cleaner IP. (C) MycTrap IP *t*KO: *Dnmt3c* shows more proteins co-immunoprecipitated than *t*KO. The 100kDa band should be DNMT3C. (D) DNMT3L is co-immunoprecipitated along with DNMT3C, confirming that DNMT3C and DNMT3L interact.

6.1.2 Immunoprecipitation-Mass spectrometry of DNMT3C

Using the modified protocol, the IP-MS was conducted with 4 technical replicates (Figure 6.4A). The bound proteins were eluted in 2X LDS Buffer containing 1mM DTT. Our collaborator Matthias Ostermaier from Dr. Petra Beli's lab performed the mass spectrometry. The volcano plot is given in Figure 6.4B.

There were 134 genes significantly enriched (p-value ≤ 0.05, Fold change ≥ 2) in our IP-MS dataset, with DNMT3C and DNMT3L being the most highly enriched proteins. DNMT3B was also significantly enriched. This is likely a misassignment of DNMT3C peptides as DNMT3B. The cell lines used in the experiment are *Dnmt3b*$^{KO/KO}$. DNMT3C and DNMT3B have very similar sequences, which may lead to difficulties in distinguishing them based on short peptides. The detection of DNMT3L with such a high p-value and high enrichment and Co-IP validation confirms that the IP-MS experiment worked, as the interaction between DNMT3C and DNMT3L was already confirmed in the previous IP-WB experiment (Figure 6.4B).

6.1.3 Analysis of DNMT3C IP-MS data from mESC

Many enriched proteins (107/134) are also expressed in the Prospermatogonia at stage E18.5 (Figure 6.4C). I performed gene ontology enrichment analysis (https://geneontology.org/) to determine the enriched protein families and functions (Figure 6.4D). As expected, GO analysis revealed “DNA methyltransferase activity” as the highest enriched GO class. Since DNMT3C and DNMT3L are essential for the methylation of Transposon promoters in Prospermatogonia, and they are highly enriched in our IP-MS, the GO class of “retrotransposon silencing by DNA methylation-dependent heterochromatin formation (GO:0141095)” is also understandably one of the highest enriched in our dataset.

Among the other GO classes enriched, a similar pattern occurred in all three GO modules: Biological Process, Molecular Function, and Cellular component. They can be combined into these three major topics: “maturation of ribosomal transcripts,” “ribosomal biogenesis,” and “ribosomal assembly and maturation.” DNMT3C has been shown to localize to high DNA-density regions when overexpressed in mESCs (Barau et al., 2016). The

nucleolus is the largest substructure in the cell, and it is the site of ribosome biogenesis (Cooper, 2000). Although the nucleolus is the most active site of cellular transcription, it is also an attractive compartment for nuclear heterochromatic regions (Correll, Bartek and Dundr, 2019). Thus, identifying several nucleolar proteins (ribosomal proteins – RPLs) as strong interactors of DNMT3C supports this previously recorded observation of DNMT3C's nucleolar association.

A number of DNMT3C interacting candidates were found to be WD40-repeat (WDR) proteins (WDR75, WDR74, WDR36, WDR43, WDR46, DCAF13, BOP1, PWP2, TBL3, etc.) (Figure 6.4E). WD40 repeat (WDR) domains act as protein-protein interaction scaffolds in multiprotein complexes (Schapira et al., 2017). They are a large protein family known to have a critical role in many essential biological functions, including signal transduction, transcription regulation, and apoptosis. Proteins like WDR74 (Maserati et al., 2011), WDR36 (Gallenberger et al., 2011), and Dcaf13 (Liu et al., 2019) have been shown to have critical functions in embryonic development. Several WDR proteins have also been implicated in nucleolar functions, required for ribosomal RNA synthesis and Ribosomal maturation (Hirai et al., 2013; Iwanami et al., 2008). DNMT3C interaction with WDR proteins could be driven by its nucleolar localization. It could be hypothesized that these WDR proteins help form a protein complex involving DNMT3C and aid in its function. This interaction could also be an artifact due to the overexpression of DNMT3C. This could explain the number of different WDR proteins enriched, as DNMT3C is not expected to be part of many protein complexes.

Another potential interesting candidate of DNMT3C is CSE1L (Figure 6.4B), an exportin-2 nuclear transport factor essential for mouse embryonic development (Bera et al., 2001). A si-RNA + Reporter screen in a human cell line for factors involved in silencing a methylated promoter revealed CSE1L as one of the top candidates. It was also shown to be essential for silencing many endogenously methylated genes. CSE1L depletion does not cause demethylation of the promoters. Rather, the nuclear import function of CSE1L is implicated in its transcriptional silencing function. The genes derepressed by CSE1L depletion

largely overlapped with methylated genes that were also reactivated by treatment with histone deacetylase inhibitors (HDACi). NOVA1, HDAC1, HDAC2, and HDAC8, genes known as silencing factors, became delocalized into cytosol upon CSE1L depletion (Dong et al., 2018). DNMT3C interaction with CSE1L in mESCs might help maintain the repressed state of the methylated genes. A sub-section of DNMT3C methylated regions in *Dnmt*-tKO: *Dnmt3C* mESC are ERVs, particularly ERV1 and ERVK (Dura, 2021). These are some of the most active transposon subfamilies in the mouse (Maksakova et al., 2006). Without DNA methylation, these transposon families are repressed through repressive histone modifications H3K9me3 and H3K27me3 (Walter et al., 2016). DNMT3C overexpression might methylate the ERVK elements through H3K9me3 recruitment and keep them repressed through its interaction with CSE1L. This also provides another potential crosstalk opportunity between de novo DNA methylation and the chromatin environment.

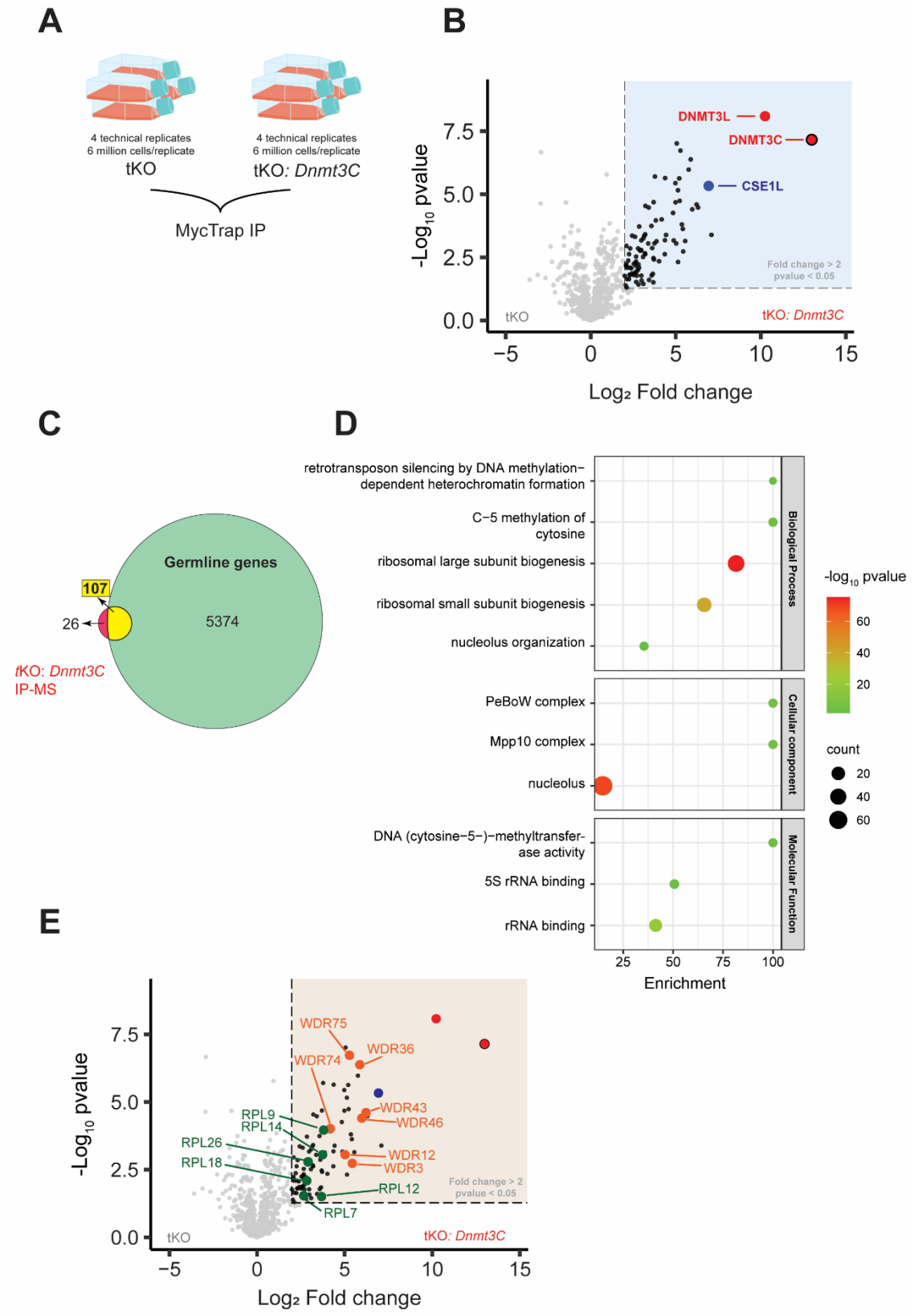


Figure 6.4: **MycTrap Immunoprecipitation of DNMT3C and analysis** (A) Strategy of the IP. (B) Volcano plot result from the mass spectrometry showing the bait protein, bonafide interactor DNMT3L, and an interesting candidate CSE1L. (C) Intersection of the IP-MS dataset with germline genes expressed in the male germline at stage E18.5. (D) Gene ontology analysis of the significantly enriched proteins in DNMT3C IP-MS dataset. (E) Volcano plot showing WDR and RPL proteins enriched in the IP-MS.

6.2 DNMT3C INTERACTOME IN PROSPERMATOGONIA

6.2.1 Generation of *Dnmt3C*$^{myc/myc}$ transgenic mice

The *Dnmt3c* gene occupies Chromosome 2 (chr2:153,696,652-153,729,907; 19 exons) in the mouse genome. Due to the high degree of similarity of DNMT3C with the DNMT3B protein sequence, generating a DNMT3C-specific antibody would be difficult. Therefore, the endogenous *Dnmt3c* gene was tagged at its 5' end with a 2xmyc epitope tag using CRISPR/Cas9 genome editing technology. The transgenic mice were generated at the Institute Curie, Paris, France.

The oligonucleotide template strand was 200bp with left and right homology arms of 64bp each and the recombinant **insert** (Myc tag – linker – Myc tag – linker) with 72bp. (5' ctctcatgtccctgtctccttctttccttcctatccattctggccttctcccacagacaacaat**GGAGCAAAAACTCATATCA GAAGAAGACCTTGGTGGTGAGCAGAAGCTGATCTCTGAAGAGGACCTTGGAGGT**aggg gaggtagcagacacctcagtaatgaggaggatgtcagtggatgtgaggactgtattatc 3') (Figure 6.5A). A schematic representation of the knock-in strategy is given in Figure 6.5B. The mixture of Cas9 mRNA, the guide RNA (5' CTCCCACAGACAACAATGAG 3'), and the template was microinjected into hybrid zygotes (C57BL/6 x DBA/2). Injected embryos were implanted into pseudopregnant females. The pups were genotyped by phalange clipping (Figure 6.5C), and the mice homozygous for the Myc allele were kept as the transgenic founder animals. The WT and KI PCR bands were sequenced separately by Sanger sequencing to ensure no mutations were found in the recombinant insert sequence (Figure 6.5D) and the WT allele (data not shown). These founder animals were backcrossed with C65B6/J mice. The heterozygous F1 generation offspring from this breeding backcross were mated to get the homozygous F2 generation. Two mice lines were derived and maintained from this transgenesis – *Dnmt3c*$^{myc/myc}$ and *Dnmt3c*$^{WT/WT}$ (the WT animal from the heterozygous F1 generation cross). The *Dnmt3c*$^{WT/WT}$ was used as a control for all the *Dnmt3c*$^{myc/myc}$ mice experiments. A schematic of the 2xmyc-DNMT3C protein is given in Figure 6.5E.

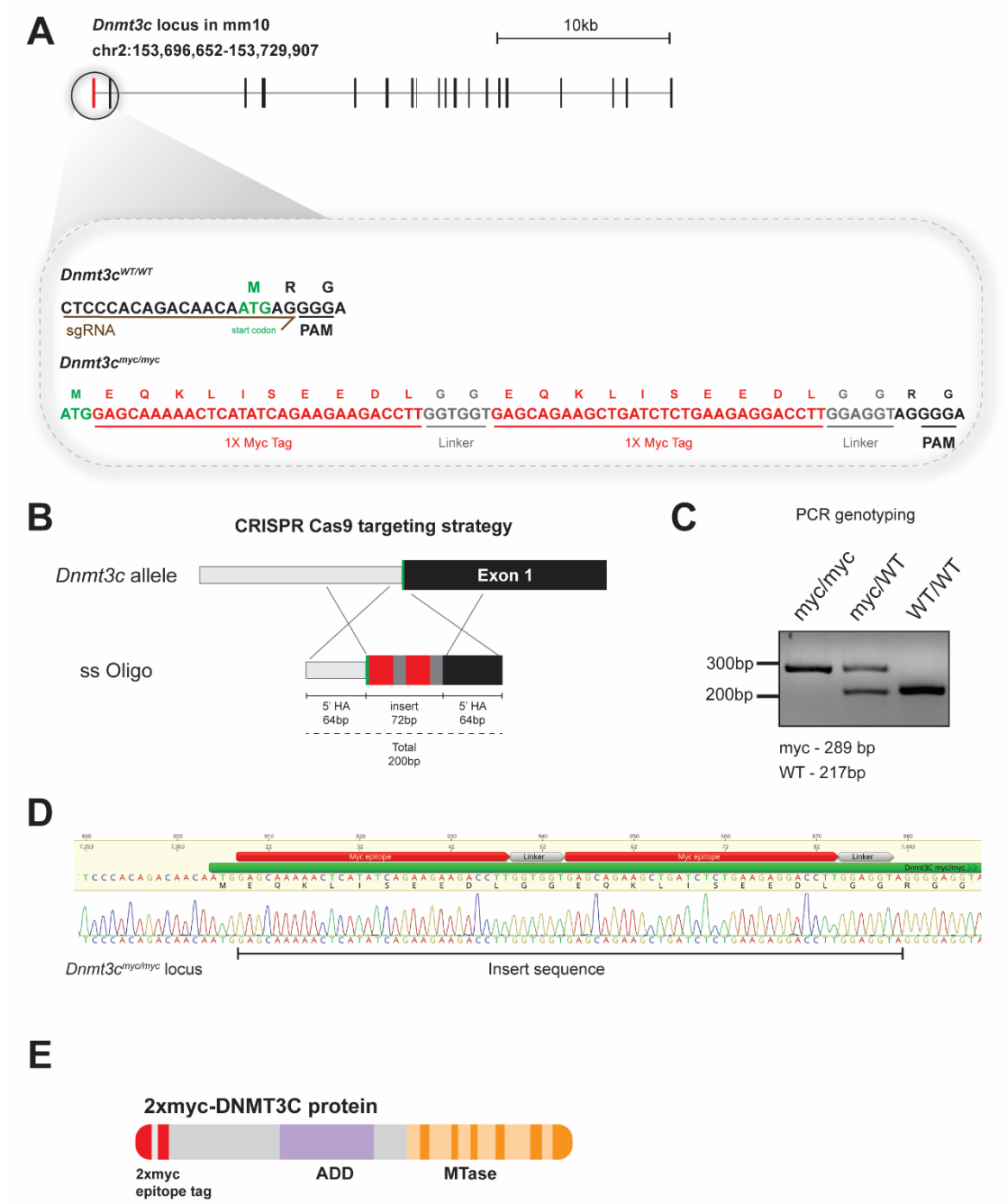


Figure 6.5: ***Dnmt3c$^{myc/myc}$* transgenic mice generation**. (A) Schematic representation of *2xmyc-Dnmt3c* generation. The sgRNA (brown), insert sequence (in red and grey), and the start codon (green) are highlighted. (B) CRISPR Cas9 targeting strategy to generate the *Dnmt3c^{myc}* allele with a short single-stranded DNA oligo of 200 nucleotides containing 5′ and 3′ homology arms (5′HA and 3′HA) of 64 nucleotides and insert of 72 nucleotides. (C) Representative image of PCR genotyping results for *Dnmt3c$^{myc/myc}$*, *Dnmt3c$^{myc/WT,}$* and *Dnmt3c$^{WT/WT}$* mice (D) Sanger sequencing chromatogram result of the *Dnmt3c^{myc}* allele showing the protein sequence as well (E) Schematic representation of the 2xmyc-DNMT3C protein.

6.2.2 Validation and characterization of 2xmyc-DNMT3C expression

Dnmt3c is expressed in the embryonic male germ cells in the developmental window E12.5 – P11 (Figure 6.6A; data analyzed by Anke Busch, Bioinformatics, Institute of Molecular Biology) (Sangrithi et al., 2017). Therefore, I wanted to confirm the expression of 2xmyc-DNMT3C at stage E18.5 by RIPA lysis of the testes, followed by western blotting. When probing for the myc tag with the 9E10 anti-myc antibody (Evan et al., 1985), we can see a strong band around 130kDa present in both the *Dnmt3c*$^{WT/WT}$ and *Dnmt3c*$^{myc/myc}$ testes samples and a shadow band around 95kDa present only in *Dnmt3c*$^{myc/myc}$ samples (Figure 6.6B). The molecular weight of 2xmyc-DNMT3C is around 85kDa. Therefore, the band around 95kDa in this blot was considered to be 2xmyc-DNMT3C. To be sure, I performed immunoprecipitation with MycTrap beads on the RIPA testes lysate supernatant, followed by western blotting. I also ran a MycTrap IP from *Dnmt*-tKO: *Dnmt3C* mESCs alongside as a positive control. As shown in Figure 6.6C, I confirmed that the 95kDa band represents the 2xmyc-DNMT3C protein as it is present in only the IP of the *Dnmt3c*$^{myc/myc}$ samples and not in the *Dnmt3c*$^{WT/WT}$ samples.

I extracted proteins from testes from E14.5 to P2 stages of *Dnmt3c*$^{myc/myc}$ mice using RIPA and performed MycTrap immunoprecipitation followed by western blotting to understand the dynamics of 2xmyc-DNMT3C expression. As shown in Figure 6.6D, we can see that 2xmyc-DNMT3C is most highly expressed between E16.5 and E18.5. This expression dynamic understandably coincides with the genome-wide methylation wave of epigenetic reprogramming (Tseng et al., 2015) and correlates well with the mRNA expression of *Dnmt3c.*

I also confirmed the germ cell-specific expression of 2xmyc-DNMT3C using immunofluorescence microscopy, using testes cross-sections from stage E18.5 (Figure 6.7).

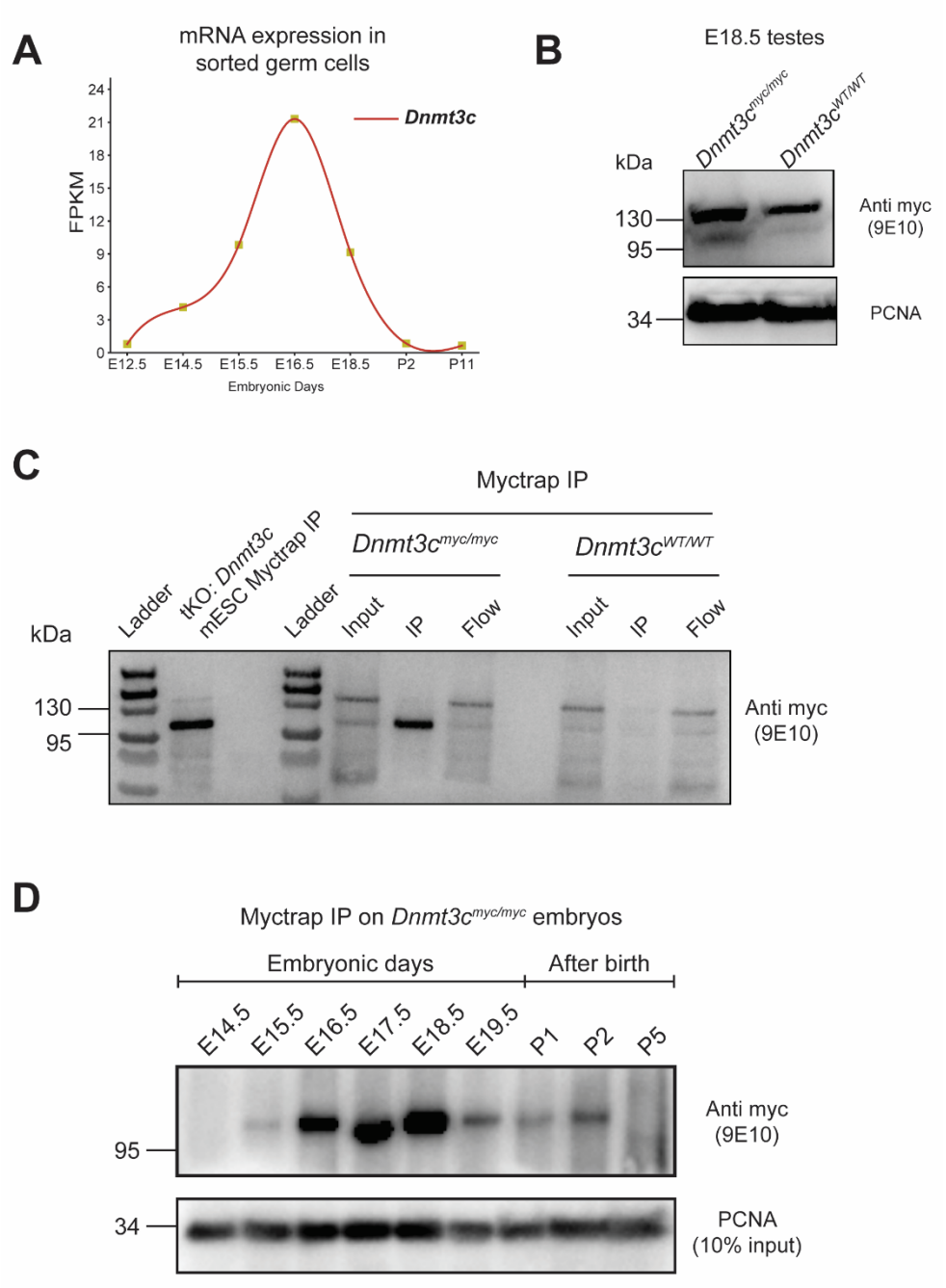


Figure 6.6: **Validation of 2xmyc-DNMT3C expression through immunoprecipitation-western blot**. (A) mRNA expression of *Dnmt3c* in the germline. It reaches a peak around E16.5/17.5 and reduces after. (B) Western blot of RIPA lysate from testes of *Dnmt3c*$^{myc/myc}$ and *Dnmt3c*$^{WT/WT}$ mice. A faint band is seen specifically in the *Dnmt3c*$^{myc/myc}$ samples around 95kDa, considered 2xmyc DNMT3C. (C) MycTrap IP of testes lysate confirming the band around 95kDa is 2xmyc-DNMT3C. The IP from *Dnmt-t*KO: *Dnmt3c* mESC is shown as a positive control (D) Expression dynamic of 2xmyc-DNMT3C during the germline wave of epigenetic reprogramming and after birth.

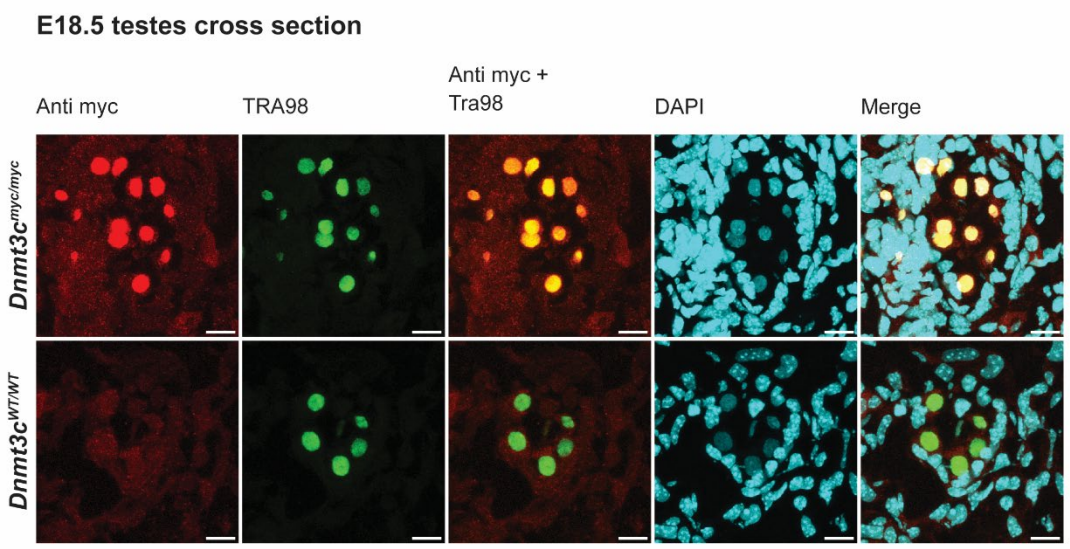


Figure 6.7: **Validation of 2xmyc-DNMT3C expression in germ cells through immunofluorescence**. TRA98 is used as a germ cell marker. The expression of 2xmyc-DNMT3C (seen through staining with Anti-myc antibody) is specific to *Dnmt3c*$^{myc/myc}$ testes and is seen only in the germ cells, not the somatic gonadal cells. The scale bar is 10μm.

6.2.3 *Dnmt3c*$^{myc/myc}$ mice have normal fertility

I examined the DNA methylation at promoter regions of LINE1-A, LINE1-T, and IAPEz elements (targets of DNMT3C) in male germ cells to check if tagging DNMT3C affects its enzymatic activity. I sorted Spermatogonial stem cells and Spermatogonia from P15 testes of *Dnmt3c*$^{myc/myc}$, *Dnmt3c*$^{WT/WT}$, and *Dnmt3c*$^{KO/KO}$ mice. Bisulfite conversion and Pyrosequencing showed no statistical difference between DNA methylation levels of these transposon promoters in *Dnmt3c*$^{myc/myc}$ and *Dnmt3c*$^{WT/WT}$ mice (Figure 6.8A). This confirms that the tagging does not affect DNMT3C functionality. Loss of methylation in these promoters in *Dnmt3c*$^{KO/KO}$ confirms that these are DNMT3C targets. The breeding performance of the *Dnmt3c*$^{myc/myc}$ genotype was similar to that of the *Dnmt3c*$^{WT/WT}$, with a similar number of litters per female and litter sizes (Figure 6.8B). Therefore, it can be concluded that the tagging of DNMT3C does not create any fertility issues.

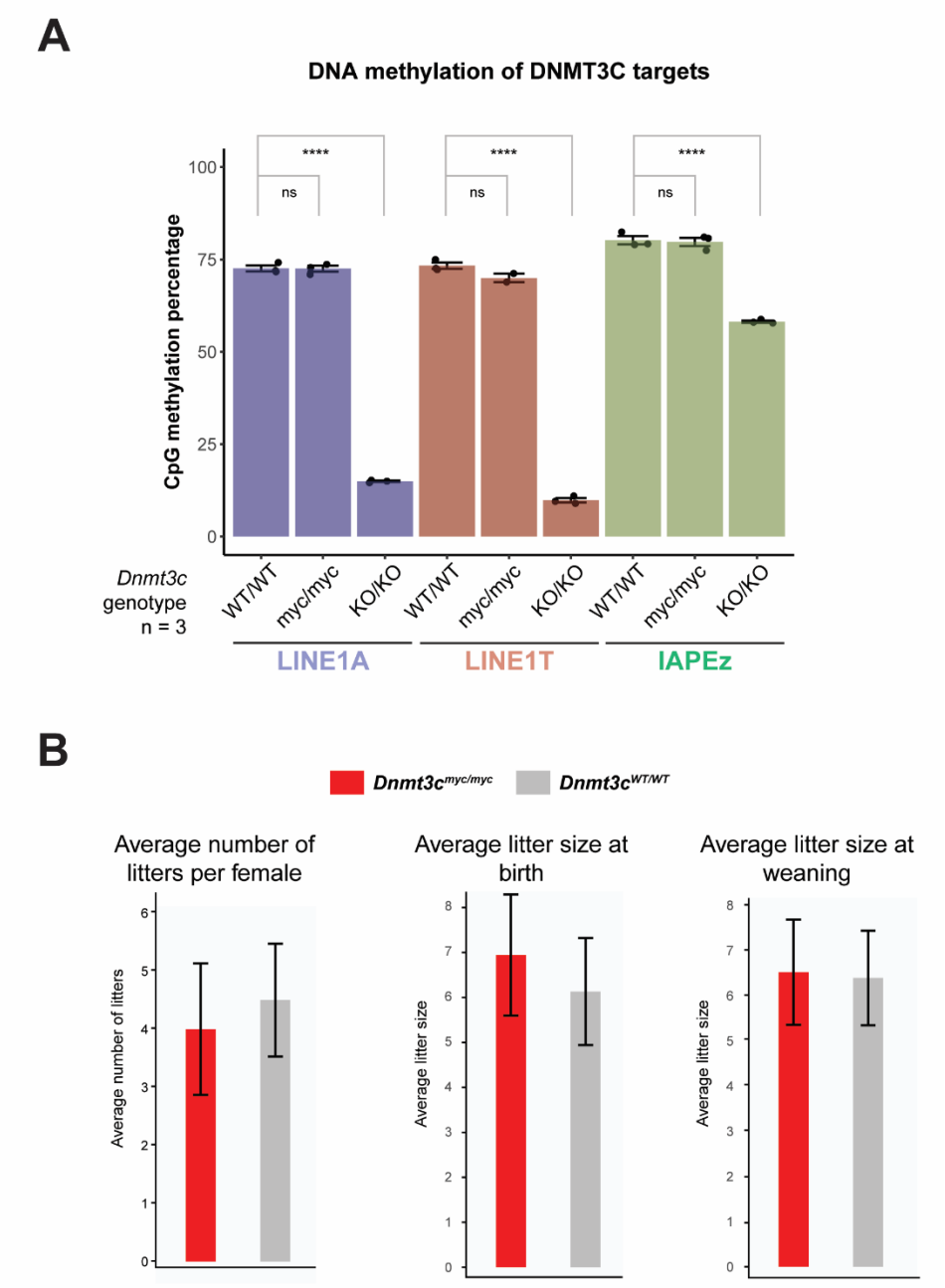


Figure 6.8: **Tagging *Dnmt3c* does not affect its function and fertility of mice**. (A) Graph showing DNA methylation at promoter regions of L1-A, L1-T, and IAPEz elements in sorted testicular germ cells at P15 in different genotypes. As expected, the methylation is very low in *Dnmt3c*$^{KO/KO}$ mice, as they are DNMT3C targets. There was no significant change in methylation in *Dnmt3c*$^{WT/WT}$ and *Dnmt3c*$^{myc/myc}$ juvenile germ cells, showing tagging of *Dnmt3c* does not affect its function (B) The breeding performance of *Dnmt3c*$^{myc/myc}$ and *Dnmt3c*$^{WT/WT}$ are very similar, implying that tagging of *Dnmt3c* has not affected fertility either.

6.2.4 Optimization of DNMT3C isolation for IP-MS

Combining the fact that most of the DNA methylation mediated by DNMT3C happens between E16.5 and E18.5 (Dura, 2021), and the highest expression of 2xmyc-DNMT3C is at E17.5 and E18.5 (Figure 6.6D); I chose to extract testes from E17.5 and E18.5 stages to probe for 2xmyc-DNMT3C's interactome.

Before I joined his lab, Dr Joan Barau had already started collecting *Dnmt3c*$^{myc/myc}$ E17.5 testes for the IP-MS. He fractionated the testes into cytoplasmic fractions and nuclei. The nuclei were treated with NS420 buffer and further fractionated into nuclear soluble and insoluble chromatin. As we discovered in the previous section, most 2xmyc-DNMT3C is associated with the insoluble chromatin fraction and needs a minimum of 600mM NaCl concentration for solubilization. *Dnmt3c* is lowly expressed in germ cells, comprising only a small proportion of the testes. Therefore, the total amount of DNMT3C protein is very low. To preserve the protein structure and the protein-protein interactions, I decided not to subject the insoluble chromatin from testes samples to a high salt concentration. Instead, I decided to test other chemical and mechanical protein extraction methods. The least number of embryos to visualize 2xmyc-DNMT3C expression in the IP-western blot was two (data not shown). So, I used testes from 2 embryos at stages E17.5/E18.5 for optimizing 2xmyc-DNMT3C's solubilization.

Firstly, I used buffers with different NaCl and detergent concentrations. I used buffers NS200, NS300, NS200 with 0.5% Igepal CA630, NS200 with 0.5% NP-40, and NS200 with 0.5% TritonX-100, using RIPA as the reference for maximum protein isolation. The insoluble chromatin pellet from E18.5 testes was treated with these buffers, the tubes were vigorously shaken for 30 minutes in an Intelli mixer with setting UU-70, and the supernatant was used for immunoprecipitation. As shown in Figure 6.9A and Figure 6.9B, only marginal amounts (15% of what was isolated in RIPA extraction) of DNMT3C protein were solubilized in any of the tested non-denaturing conditions; the NS300 buffer performed best. Other studies that performed immunoprecipitation of mouse embryonic testes proteins treated the testes with

RNase A (Schöpp et al., 2020) or Benzonase (Zoch et al., 2020), and this treatment increased the number of proteins enriched in their IP-MS assay. I treated our chromatin fraction with NS300 buffer containing either RNase A (New England Biolabs T3018L) or Sm nuclease (produced by the Protein Production Core facility at the Institute of Molecular Biology) or higher Igepal CA-630. Unfortunately, neither Sm nuclease nor RNase increased DNMT3C solubility. NS300 with 0.5% Igepal CA-630 and NS400 + 0.5% Igepal CA-630 had the equivalent highest DNMT3C solubilization among the buffers tested. I chose the NS300 + 0.5% Igepal CA-630 due to the lower salt concentration (Figure 6.9C and D).

The rationale behind adding nucleases is to digest the DNA, helping solubilize DNA/Chromatin-associated proteins. Because the nucleases did not aid in DNMT3C solubilization, I tried sonication as a physical method to solubilize the chromatin (Kustatscher et al., 2014). As I wanted to test many sonication conditions, I used *Dnmt-t*KO: *Dnmt3c* mESC as a model. The insoluble chromatin from the mESC was resuspended in NS300 buffer and then sonicated for 0-9 cycles (10 different conditions), with the parameters for one cycle being—20 seconds ON, 40 seconds OFF, and high amplitude at 4°C on a Bioruptor (Diagenode). More time was chosen for the OFF setting to prevent overheating of the samples and protein denaturation. We can see from Figure 6.9E that sonication helps solubilize DNMT3C, and a maximum amount of DNMT3C is already solubilized with five cycles of sonication.

The final protocol for the isolation of DNMT3C from testes went as follows: Testes from E17.5 or E18.5 stages were decapsulated and then fractionated into cytoplasmic, nuclear soluble and insoluble chromatin fractions. The insoluble chromatin fraction was treated with NS300 buffer containing 0.5% Igepal CA-630 and sonicated for 5 cycles (1 cycle – 20 seconds ON, 40 seconds OFF, high amplitude, 4°C). Post-sonication, they were vigorously shaken in the Intelli mixer (setting UU-70) for 30 minutes at 4°C and then centrifuged to give the solubilized chromatin fraction.

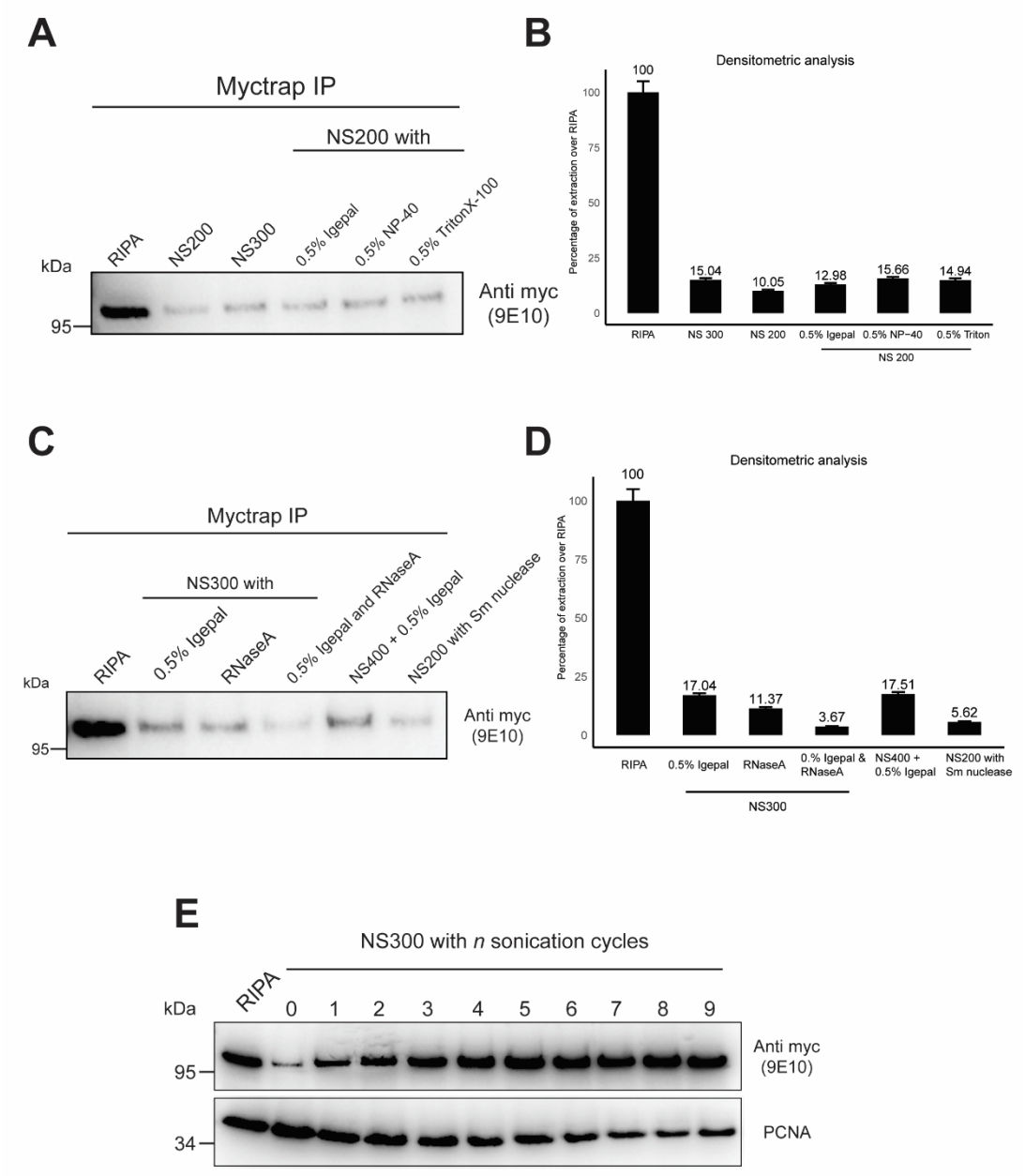


Figure 6.9: **Optimization of DNMT3C solubilization from testes extracts**. (A) MycTrap IP of solubilized chromatin from E18.5 testes from 2 embryos for each test. The chromatin was solubilized under different buffer conditions, salt concentrations, and detergents. (B) Densitometric analysis from the western blot in (A) using Image Lab software. This shows that the NS300 buffer has the highest solubilization in non-denaturing conditions. (C) MycTrap IP of solubilized chromatin from E18.5 testes from 2 embryos for each test. NS300 buffer with different detergents and other conditions were also tested. (D) Densitometric analysis from the western blot in (C) shows that NS300+0.5% Igepal CA-630 and NS400+0.5% Igepal CA-630 had the highest DNMT3C solubilization. (E) Sonication of the *Dnmt-t*KO: *Dnmt3c* mESC insoluble chromatin in NS300 buffer increases solubilization with increasing cycles until five cycles. After five cycles, no more DNMT3C is solubilized.

6.2.5 Immunoprecipitation and Mass spectrometry from testes

We collected testes from 409 E17.5/E18.5 embryos in total from *Dnmt3c*$^{myc/myc}$ and *Dnmt3c*$^{WT/WT}$ mice. Because this material was extremely precious, considering the time and effort it took for collection, I combined the cytoplasmic, nuclear soluble, and solubilized chromatin fractions for immunoprecipitation. This was done to ensure that we could identify as many potential DNMT3C interactors as possible.

As the number of *Dnmt3c*$^{myc/myc}$ testes was more than the *Dnmt3c*$^{WT/WT}$ testes, I combined all the testes and immunoprecipitation with MycTrap beads with 205 embryos, and as the control, immunoprecipitation with magnetic agarose beads with 204 embryos. The total soluble fraction (cytoplasmic + nuclear soluble + solubilizer chromatin fractions) was incubated overnight with gentle end-over-end rotations to reconstitute the protein-protein interactions. Centrifugation was performed to remove the precipitated proteins and the supernatant was used for a MycTrap IP or a control IP. The supernatant was incubated with the beads for 2 hours at 4°C with gentle end-over-end rotation. Both the control and MycTrap IPs were performed using technical triplicates. The beads were washed with NS300 buffer for 10 minutes. Bound proteins were eluted in 2X-LDS buffer containing 1mM DTT by heating at 70°C for 10 minutes (Figure 6.10).

As the material was highly precious, I wanted to extract as much information as possible. I incubated the flowthrough of the previous IPs with recombinantly produced fusion protein GFP-DNMT3C. A GFP IP of *Dnmt-t*KO mESC RIPA extract with or without recombinant GFP-DNMT3C clearly showed that GFP-DNMT3C can be used to pull down DNMT3L (Figure 6.11). After overnight incubation of GFP-DNMT3C with the testes IP flowthrough to reconstitute the protein interactions, I performed a GFP-IP using magnetic GFP binder beads (produced by the Protein Production Core facility at the Institute of Molecular Biology). These IPs were also performed in technical triplicates.

Our collaborators, Matthias Ostermaier and Katharina Mayr from Petra Beli's lab (at the Institute of Molecular Biology), prepared the samples for Mass spectrometry and analyzed

the results. The resulting volcano plot of enriched proteins in the Myc IP-MS and the GFP IP-MS are given in Figure 6.10 and Figure 6.11. Only 11 proteins were common between the Anti-Myc IP-MS and the Anti-GFP IP-MS. Therefore, I chose to proceed with the Myc IP-MS for further analysis.

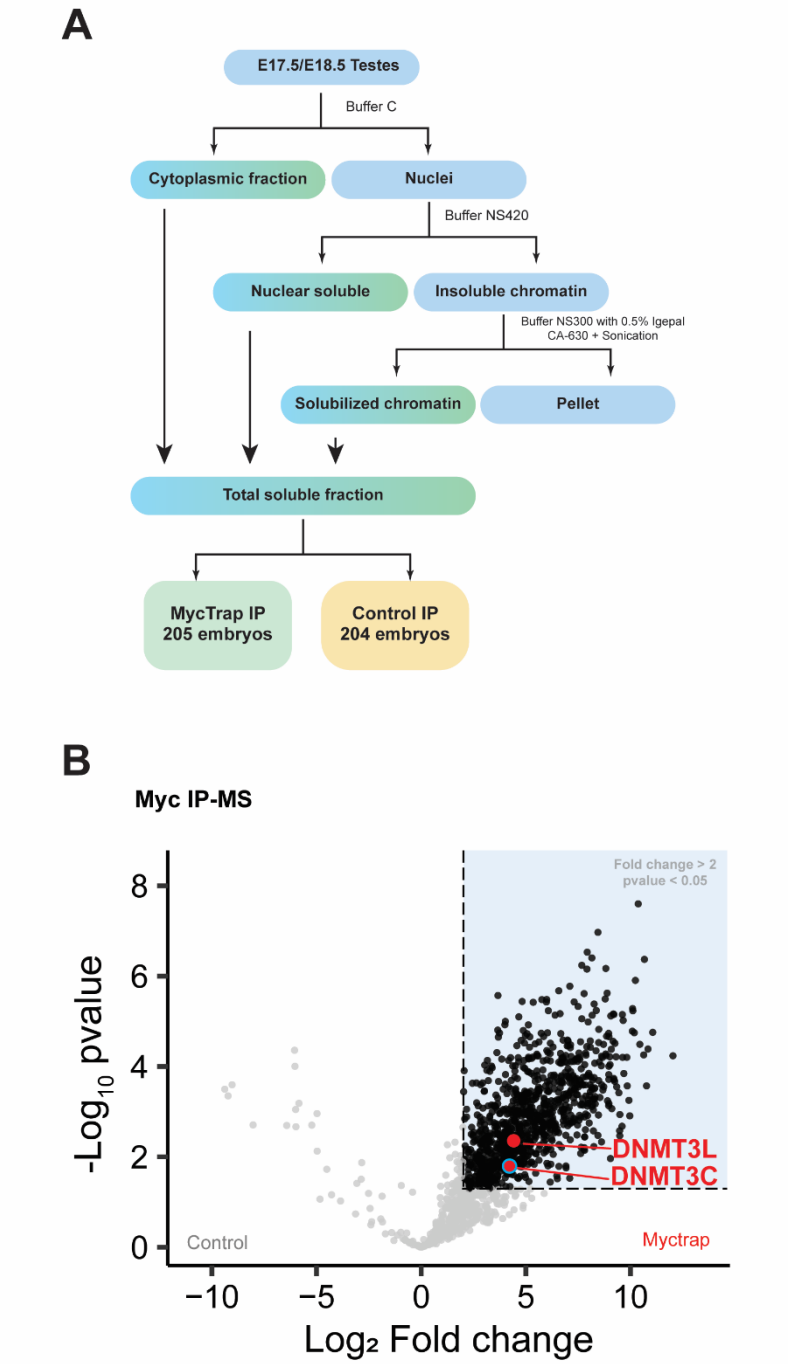


Figure 6.10: **Myc IP-MS from E17.5/E18.5 testes**. (A) Optimized protocol workflow for 2xmyc-DNMT3C protein isolation and immunoprecipitation. (B) Volcano plot result of the Myc IP-MS showing the bait protein DNMT3C and the bonafide interactor DNMT3L.

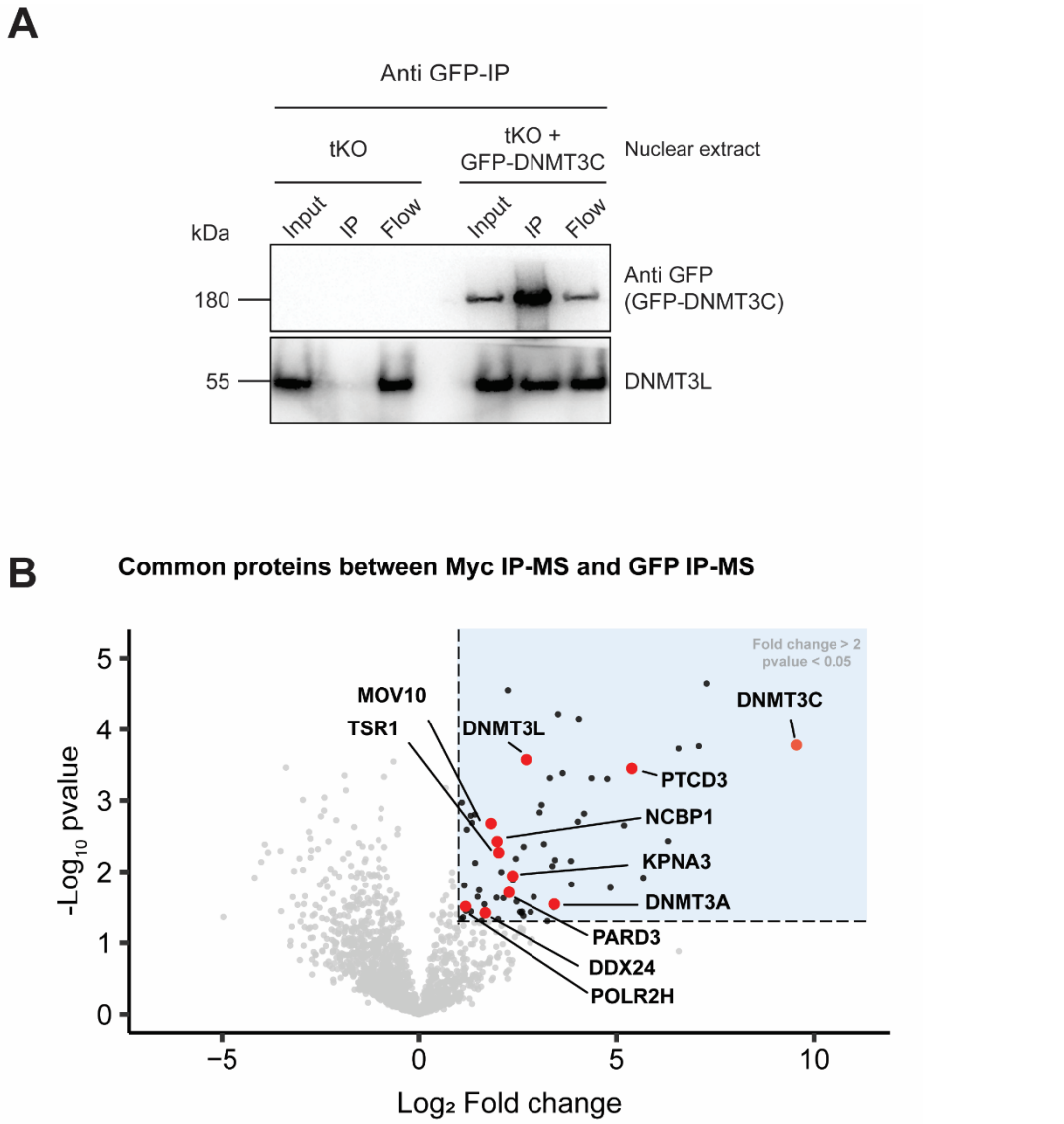


Figure 6.11: **GFP IP-MS of DNMT3C**. (A) Western blot results show that GFP-DNMT3C, when incubated with nuclear extract from *Dnmt: t*KO mESC, can pull down the bonafide DNMT3C interactor DNMT3L. (B) Volcano plot resulting from the GFP IP-MS indicating the proteins common with Myc IP-MS, including the bait and DNMT3L.

6.2.6 Analysis of 2xmyc-DNMT3C Prospermatogonia IP-MS data

The bait protein DNMT3C is significantly enriched, and the bonafide DNMT3C-interactor DNMT3L is also significantly enriched (Fold change ≥ 2, p-value ≤ 0.05), validating that the IP experiment worked. Immunoprecipitation followed by western blotting showed that DNMT3L co-immunoprecipitates with DNMT3C (Figure 6.12A). With this result, I confirmed that DNMT3C and DNMT3L also interact in Mouse Prospermatogonia.

Surprisingly, DNMT3C, being the bait, was not the highest enriched protein. There could be a few possible reasons for this. (1) DNMT3C and DNMT3B share 70.6% identity in their protein sequence. Some DNMT3C peptides could have been misidentified as DNMT3B. (2) DNMT3C is very lowly expressed. Therefore, there could be only a few quantifiable peptides available. Combining with the identity with DNMT3B, the number of unique peptides could have been reduced further. (3) The control magnetic agarose beads were bound to very few proteins; thus, the control IP was extremely clean. This would create a problem with the downstream analysis, based on the enrichment of proteins in the samples over the control. Since the control had very few proteins, this might have led to inaccurate imputation during analysis, leading to the erroneous calculation of the fold enrichment values. The last point could also be the reason for enriching so many proteins in the dataset.

In total, 1031 proteins were significantly enriched in the IP-MS dataset. The less stringent conditions during protein extraction and the fact that proteins were isolated from whole testes and not only germ cells could have led to the enrichment of many proteins. To gain insight into the DNMT3C interactome, I filtered the data using two parameters: cellular localization and gene expression. DNMT3C is a germline nuclear protein, so its interactors should also be expressed in the germline concurrently with *Dnmt3c* and present in the nucleus. Proteins in the IP-MS dataset whose localization was “nucleus” and “unknown” (data obtained from Uniprot) were filtered. This new dataset was further intersected with the genes expressed in Prospermatogonia at stage E18.5 (genes with FPKM expression greater than or equal to *Dnmt3c*) (Sangrithi et al., 2017). This strategy reduced the IP-MS dataset to 325 proteins, with

221 significantly enriched proteins (Fold change ≥ 2, p-value ≤0.05). The volcano plot after filtering the data is given in Figure 6.12. The total list of enriched proteins in the DNMT3C IP-MS dataset is given in Figure 6.13.

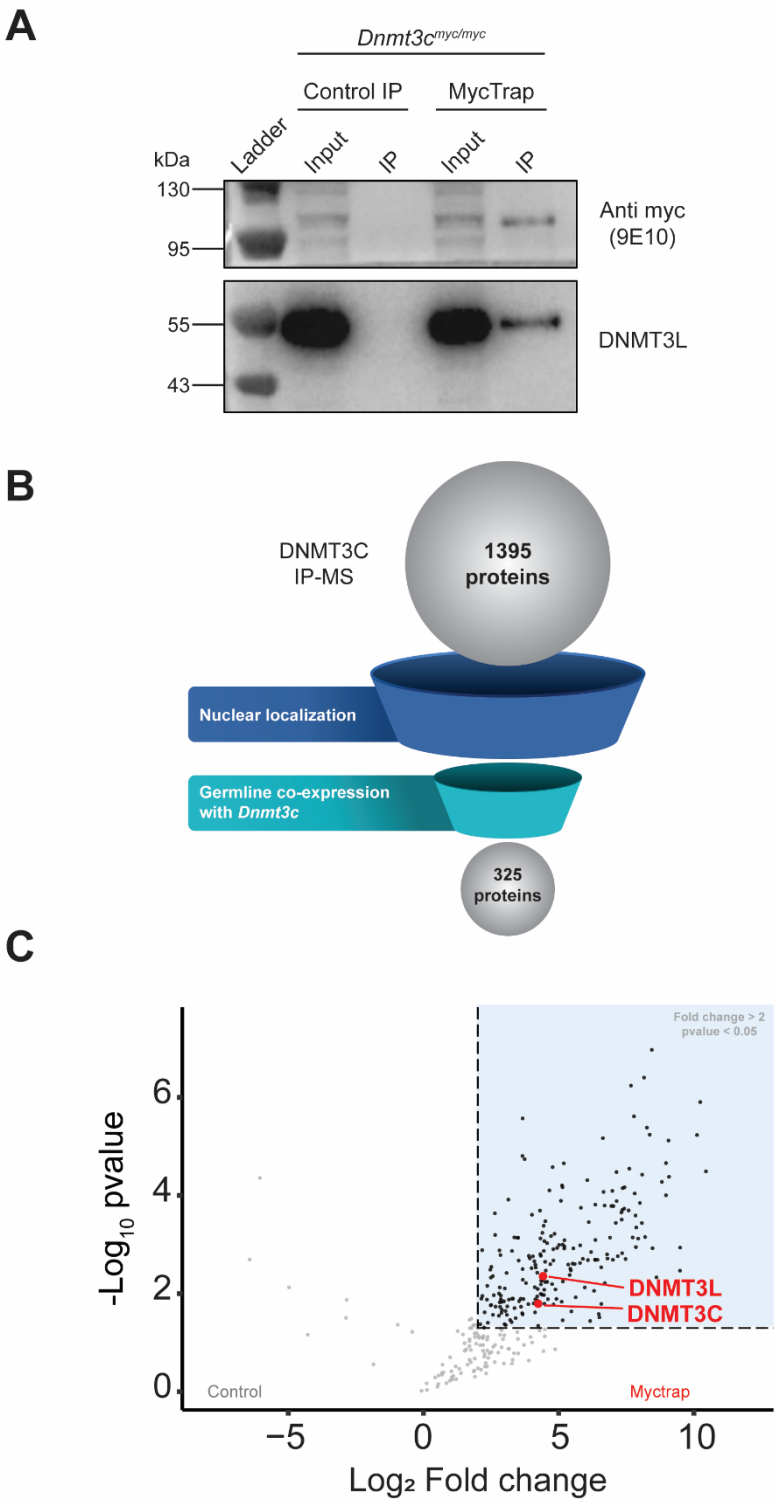


Figure 6.12: **Validation and filtration of the Myc IP-MS dataset**. (A) Western blotting showing DNMT3L co-immunoprecipitates with DNMT3C from testes material. (B) Strategy to filter the IP-MS dataset. (C) Volcano plot of Myc IP-MS after filtration.

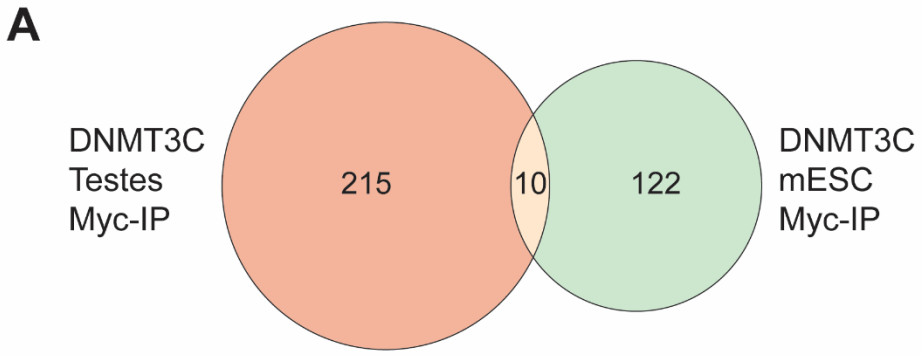
LIST OF PROTEINS ENRICHED IN DNMT3C IP-MS
(Fold change > 2, p value < 0.05)

CPSF1	WRNIP1	SYMPK	MORC3	FLII
SF3B1	MLH1	FBXO38	DNMT1	DDX1
SF3B3	PSMC4	LUC7L3	LSM2	LUC7L2
UFL1	LUZP1	EP400	SREBF2	ILF2
ZMYM2	SRRM2	ZSCAN18	PPAN	SDAD1
CPSF2	USP15	POLR3D	ASH2L	AKAP8L
NUP107	SNRNP200	SF3B6	LSD2	RBM39
SPOCD1	CDK5RAP3	SCAF8	SF3A1	NOL9
SRRT	POLR1C	SF3B4	MIER2	AEBP1
TCF25	CPSF3	MPHOSPH8	PCBP1	AKAP8
PNN	STAT2	TDRD9	NFYC	PDCD11
PRPF40A	YEATS4	PRPF4B	MYO1C	MCM2
LSD1	SNRNP27	FLCN	NEK9	CACTIN
DHX15	TEX15	MCM3	CCNT1	SRPK1
RBM25	HSPA9	POLR2H	RAD21	CRNKL1
SLTM	MATR3	MAPKAP1	UBAP2L	SEC13
PATL1	TCF20	SNRPD2	PRPF3	MCM7
TIMELESS	RUVBL1	NUP93	AQR	KANSL2
AHCTF1	EFTUD2	BAG3	CDC5L	ZHX1
PALLD	TAF6	TAF2	PPP1CB	FMR1
U2SURP	POLR2A	PSMC6	MYBBP1A	BAG6
PSMC5	EHMT2	SMC3	HSPA8	GNL2
NUP133	ACTL6A	PELP1	SFSWAP	UBE2O
POLR3B	CSTF3	DCAF7	CTNNB1	SNRNP70
ANAPC4	NUMA1	IPO5	TSR1	SSRP1
ATE1	TFIP11	HDAC1	RNF17	CNOT3
CIC	PSMC3	CHD4	ACD	ZC3H13
VIM	RALBP1	DNMT3L	HNRNPH2	KIF20B
ATXN2L	KANSL3	TAF1	SMARCA4	CSTF1
DDX42	DNMT3A	LSM4	MED15	CCNL2
WDR33	MIWI2	KPNA3	SART3	STAG3
ZXDC	UBR4	RBM10	MAGED1	PAN2
THRAP3	HNRNPF	POLR3C	NEDD4	MCM6
IK	SIN3A	PPP1CA	PRICKLE1	NKAP
DMAP1	PER1	POLR3F	TGM2	NCBP1
HDAC2	RBM17	SMC1A	PRPF6	
DDB1	POLR2E	PSMC1	SNRPE	
PRPF8	HELLS	MPHOSPH10	KAT2A	
EIF4ENIF1	NAT10	MAU2	CCAR2	
MED14	MCM5	ASH1L	MGAILF3	
RNPS1	RBBP4	KPNA2	CLP1	
SETX	CSNK1D	DNMT3C	U2AF1	
POLR2B	OGT	MED23	XRN2	
CPSF4	WDR5	LSM3	SON	
SMU1	RUVBL2	MED17	FOXO1	
SPIN1	POLDIP3	ZBTB17	RBM5	

Figure 6.13: **List of proteins enriched in the DNMT3C IP-MS after filtration using nuclear localization and germline expression**. The list is given in the order of fold enrichment from top to bottom and then left to right.

6.2.7 Comparing the mESC and Testes DNMT3C IP-MS

Next, I compared the IP-MS datasets from mESC and Testes to judge whether mESC was a good model for studying DNMT3C interactors. Figure 6.14 shows that only ten proteins were commonly enriched in both IP-MS, excluding the bait DNMT3C. The bonafide interactor DNMT3L was commonly enriched. Other than that, the most interesting proteins were the RNA polymerase II subunit POLR2A and the chromatin remodeling NuRD complex member SMARCA4. The lack of many commonly enriched proteins tells us we are not capturing meaningful interactions of DNMT3C in mESC. With this, we can say that mESC is not a good model for studying DNMT3C's interactions in the context of its functions in Prospermatogonia.



B **COMMON ENRICHED PROTEINS BETWEEN TESTES AND mESC IP**

DNMT3L
POLR2A
NAT10
MPHOSPH10
PSMC5
PELP1
THRAP3
TAF2
PSMC4
SMARCA4

Figure 6.14: **Comparison between the Testes and mESC DNMT3C IP-MS**. (A) Venn diagram showing the number of shared proteins between the two IP-MS datasets. (B) Commonly enriched between the two IPs.

6.2.8 Gene Ontology Enrichment Analysis of the filtered Testes IP-MS data

I performed Gene ontology enrichment analysis to analyze the data. One of the GO classes that was highest enriched was, justifiably, “C−5 methylation of cytosine”. GO classes belonging to “RNA polymerase II transcription” were also highly enriched. The transposons that DNMT3C targets are expressed during the germline wave of epigenetic reprogramming (Zamudio et al., 2015). As DNMT3C methylates the promoters of these transposons, it could interact with the RNA polymerase II-associated transcriptional machinery gathering around the promoter simultaneously. The association of DNMT3C with the RNA polymerase II complex could also explain the enrichment of proteins belonging to the “RNA splicing complex” and “mRNA cleavage and polyadenylation specificity factor complex” (Figure 6.15). It is not clear from the data whether DNMT3C directly interacts with the transcriptional machinery. Another GO class that was enriched was “DNA methylation−dependent heterochromatin formation”. This explains the participation of DNMT3C in converting the actively transcribing transposon loci to heterochromatin. A previous study has shown that DNMT3C targets are marked by the histone marks H3K4me3 and H3K9me3 during the epigenetic reprogramming (Dura, 2021). This is concordant with the transcriptional activity of these regions. Post-DNA methylation, these regions are marked with only H3K9me3, representing its conversion into transcriptionally inactive heterochromatin. This explains the enrichment of such a GO class in the IP-MS dataset.

One of the most intriguing GO classes enriched was “piRNA−mediated heterochromatin formation.” As explained before, the piRNA pathway is a germline small RNA pathway that is a crucial mechanism through which the host silences and inactivates transposons. This was very interesting to us as *Dnmt3c*$^{KO/KO}$ and *Piwil4*$^{KO/KO}$ mice showed the same phenotype, male-specific sterility caused by germ cell loss due to early meiotic arrest (Barau et al., 2016; Carmell et al., 2007). The piRNA pathway has two modes: cytoplasmic and nuclear. In the *Dnmt3c*$^{KO/KO}$ male germline, only the DNA methylation of the transposons was affected, not the biogenesis of piRNAs. Many known members of the nuclear piRNA pathway were enriched in the DNMT3C IP-MS filtered dataset, namely MIWI2, TDRD9, SPOCD1, and

TEX15 (Figure 6.15 & Figure 6.16). The expression pattern of all these nuclear piRNA proteins and DNMT3C is also the same (Figure 6.16), a strong indication that DNMT3C is part of the nuclear piRNA pathway, like in the case of SPOCD1 (Zoch et al., 2020) and C19ORF84H (Zoch et al., 2024). The following section discusses the interaction between DNMT3C and the piRNA pathway proteins and the molecular phenotype of their absence.

Two more intriguing GO classes enriched are “nuclear pore complex assembly” and “nuclear import signal receptor activity”. The nuclear pore complex is known to bind to heterochromatin and lamin-associated domains (Shevelyov, 2023). Nucleoporins (Nup) make up the nuclear pore complex. Various Nups and nuclear export factors have been shown to participate in transposon silencing in *Drosophila* (Czech et al., 2013). NUP54 and NUP58 are essential for transposon silencing by enabling transport and subsequent piRNA biogenesis from the *flamenco* piRNA locus in *Drosophila melanogaster* (Munafò et al., 2021). Another study showed that Nup358 was essential for PIWI’s nuclear entry and piRNA pathway function in *Drosophila melanogaster*, as germline knockdown disrupted piRNA biogenesis and activated transposons (Parikh, Lin and Gangaraju, 2018). The nuclear localization of piRNA-loaded MIWI2 is also critical in the mouse piRNA pathway. MIWI2 could interact with NUPs for its nuclear translocation. Through its interactions with MIWI2, DNMT3C could indirectly interact with the NUPs. Interestingly, there is a nucleoporin NUP93 that is enriched in the IP-MS data of DNMT3C (this study), MIWI2 and SPOCD1 (Zoch et al., 2020). This could provide insight into the organization and initiation of the nuclear piRNA pathway. NUPs, through their interaction with piRNA pathway proteins and DNMT3C, could aid in initiating the nuclear piRNA pathway and heterochromatin formation at the transposon loci.

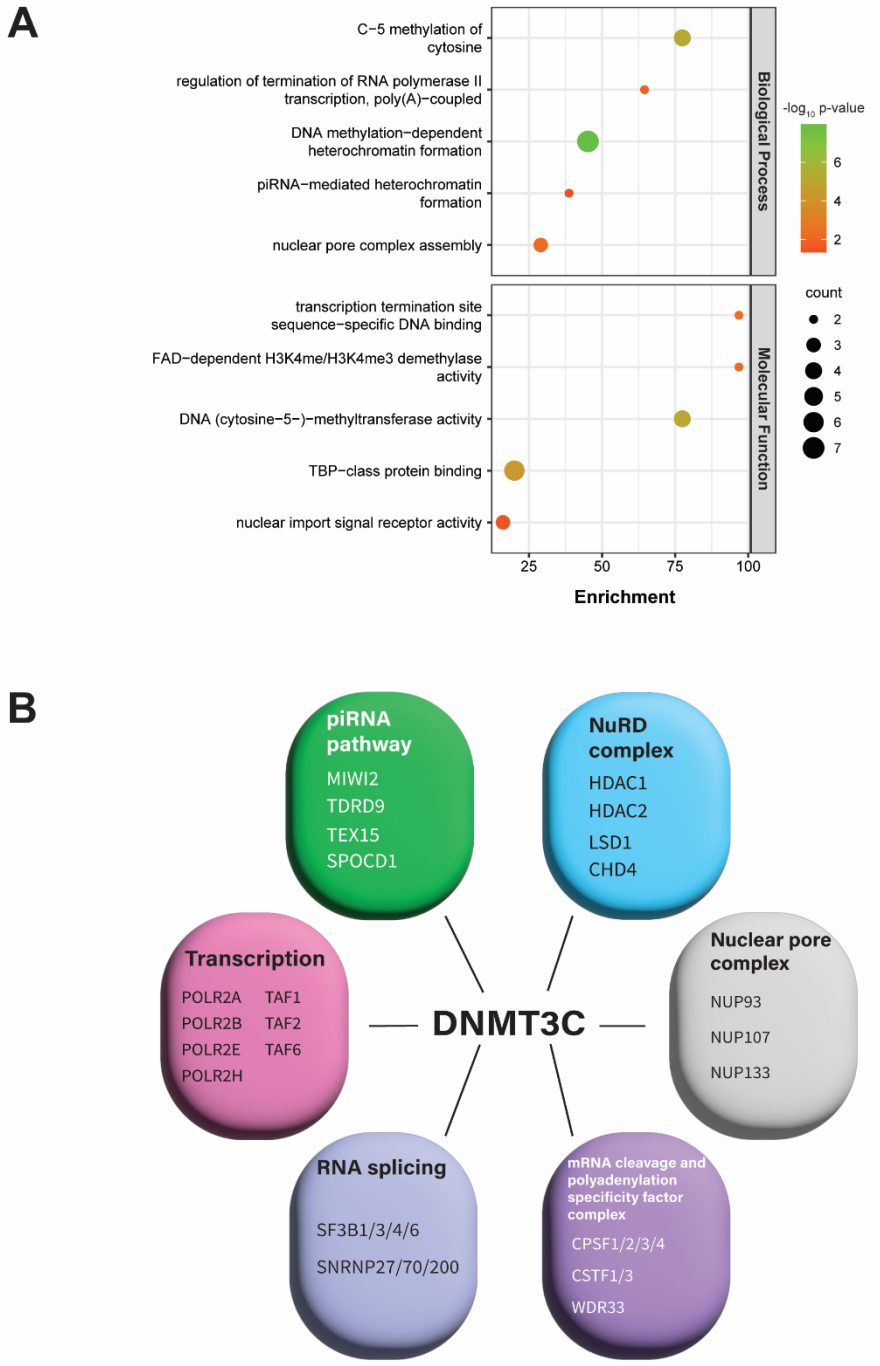


Figure 6.15: **Gene ontology enrichment analysis of Myc IP-MS**. (A) The result of the GO analysis (divided by Biological Process and Molecular Function) shows some of the most enriched GO classes. (B) Proteins enriched in the Myc IP-MS and belong to interesting GO classes.

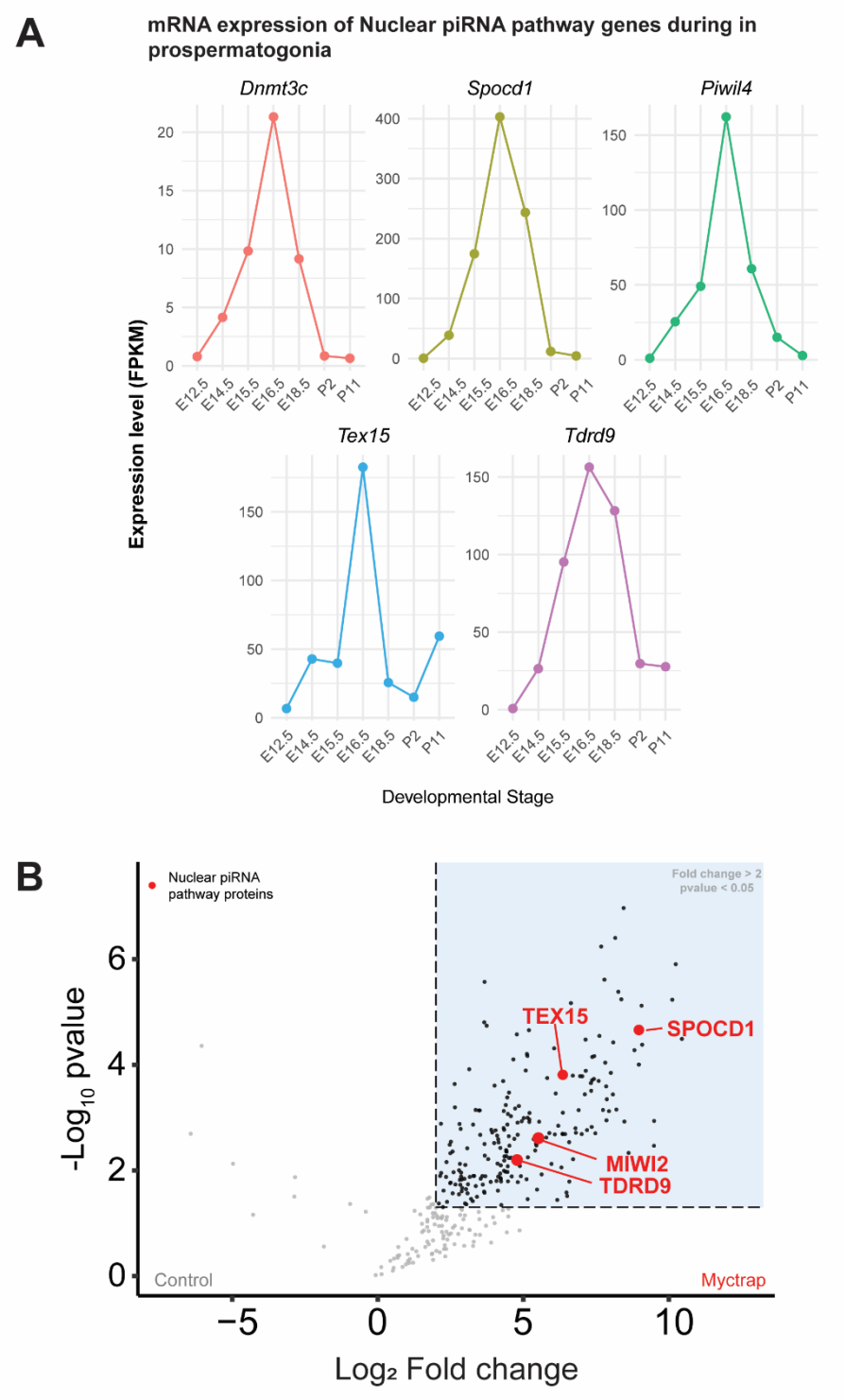


Figure 6.16: **Nuclear piRNA pathway and DNMT3C**. (A) Line graphs showing the mRNA expression pattern of DNMT3C is the same as various nuclear piRNA pathway members. (B) Volcano plot of Myc IP-MS highlighting the nuclear piRNA pathway members.

6.2.9 DNMT3C HA-IP also has nuclear piRNA pathway members enriched

Another study performed IP-MS of DNMT3C from the Dónal O'Carroll laboratory (Zoch et al., 2024). They tagged DNMT3C at its N terminus with an HA tag and performed an HA-IP from E16.5 fetal testes using the same protocol used for the SPOCD1 IP (Zoch et al., 2020). In brief, they lysed the testes in hypotonic lysis buffer, treated it with 50 U/ml benzonase, cleared it for 5 min at 21,000g, and added it to cross-linked anti-HA magnetic beads. The authors chose this IP protocol to remain consistent and comparable with previous IP datasets their lab generated. They found 83 proteins as highly enriched (fold change > 2 and p-value <0.05). Comparing their dataset with my DNMT3C Myc-IP dataset yielded only five proteins in common, including the bait DNMT3C (Figure 6.17A). Among those were the bonafide interactor DNMT3L and, more interestingly, the nuclear piRNA pathway members SPOCD1 and TDRD9 (Figure 6.17C). Notably, C19ORF84H was enriched in the HA-IP but not in my Myc-IP. The same study also performed C19ORF84H IP. Comparing the IP-MS's of DNMT3C and C19ORF84H showed only five proteins in common, among which were SPOCD1 and SPIN1 (Figure 6.17 B and D). This indicates that SPOCD1 could play a central role in the nuclear piRNA pathway, bridging interactions between effectors. This evidence from two independent studies further strengthens the connection of DNMT3C with the nuclear piRNA pathway. The gene ontology enrichment analysis of HA-IP-enriched proteins showed "piRNA-mediated retrotransposon silencing by heterochromatin formation," "transcription elongation by RNA polymerase II," and "transcription factor TFIIH complex" as some of the most enriched GO classes. Similar GO classes were also enriched in the Myc-IP.

A few reasons could explain the lack of significant overlap between the two DNMT3C IP datasets. (1) DNMT3C solubilization needs more stringent steps. Therefore, extracting proteins using a hypotonic buffer solution might not have solubilized much DNMT3C protein, thus not enriching many potential interactors. (2) The HA-IP was performed at E16.5, an earlier time point than the Myc-IP (E17.5/18.5). As seen from Figure 6.6D, DNMT3C's expression is maximum at E17.5 and E18.5. As DNMT3C-mediated DNA methylation starts from E16.5 (Dura et al., 2022), much of DNMT3C at E16.5 might not have been committed to DNA

methylation and thus might not yet interact with the relevant proteins. (3) As discussed earlier in this chapter, using magnetic agarose beads as the control in Myc-IP might have enriched many proteins due to incorrect imputations. This could also help exacerbate the difference between Myc-IP and HA-IP.

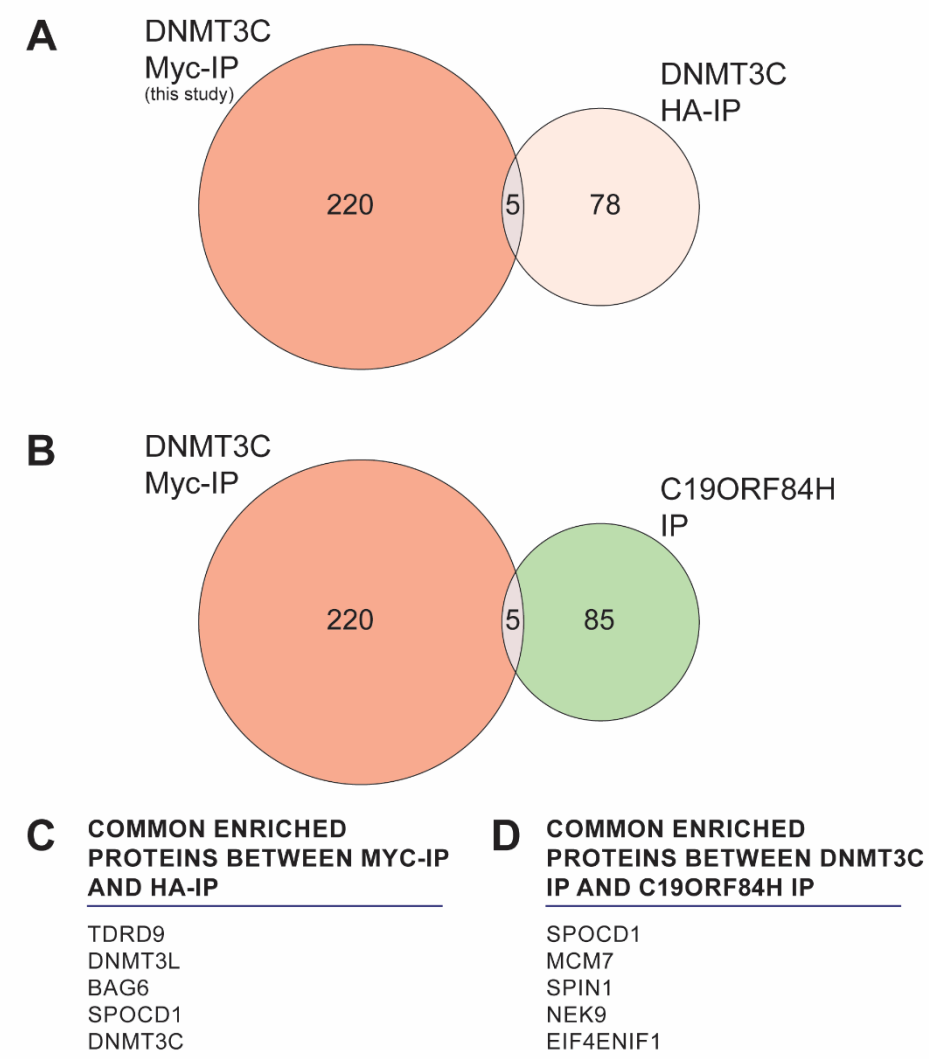
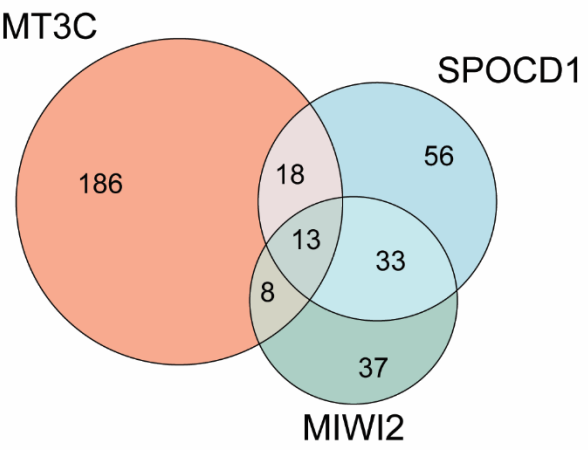


Figure 6.17: **Comparison of Myc IP with HA IP and C19ORF84H IP**. (A) Venn diagram comparing DNMT3C Myc-IP (this study) with DNMT3C HA-IP. (B) Venn diagram comparing DNMT3C Myc-IP with C19ORF84H IP. (C) Common proteins between DNMT3C Myc-IP and DNMT3C HA-IP. (D) Common proteins between DNMT3C IP and C19ORF84H IP.

6.2.10 Common proteins between DNMT3C, SPOCD1 and MIWI2 IP-MS

Overall, many interesting proteins were shared between the IP-MS datasets of DNMT3C (this study), SPOCD1, and MIWI2. DNMT3C IP-MS had 31 enriched proteins common with the SPOCD1 IP-MS dataset and 21 enriched proteins with the MIWI2 IP-MS dataset (Figure 6.18). Thirteen proteins were common in all three datasets. One among them is SPIN1. SPIN1 is a bivalent chromatin mark reader of H3K4me3 and H3K9me3 (Du et al., 2021). H3K4me3 and H3K9me3 mark DNMT3C targets (transposon promoters) during germline reprogramming (Dura, 2021). SPIN1 could bind to these bivalent marks and tether the effector complex to the transposon promoters. YEATS4 and IK also feature among common proteins. YEATS4 is a chromatin reader component of the NuA4 histone acetyltransferase (HAT) complex, and IK is a component of the spliceosome involved in pre-mRNA splicing. These proteins could be enriched in the datasets because the MIWI2-directed effector complex acts on transcriptionally active loci, and thus, interactions with the core transcription complex and associated proteins are highly probable. Notably, POLR2A and POLR2B, core subunits of the DNA-directed RNA polymerase II complex, are shared between DNMT3C and SPOCD1 IP-MS datasets but not found in the MIWI2 IP-MS. Chromatin remodelers like CHD4, SMARCA4, SIN3A, and HDAC1 are also common interactors. This corroborates that the active and open chromatin is converted to a closed heterochromatin state (Uneme et al., 2024; Walter et al., 2016; Yamanaka et al., 2019). As discussed earlier, NUP93 is also enriched in all three datasets. Additionally, IPO5 (Importin 5), a nuclear transport receptor, is also commonly enriched. This could participate in the nuclear translocation of piRNA-bound MIWI2 (Figure 6.18).

A



B

COMMON PROTEN INTERACTORS IN THE IP-MS DATASETS

DNMT3C & SPOCD1 (#18)	DNMT3C & MIWI2 (#8)	SPOCD1 & MIWI2 (#33)	DNMT3C & SPOCD1 & MIWI2 (#13)
CPSF4	VIM	SUMO2	SPIN1
ZMYM2	SETX	ACTN4	SIN3A
DNMT3A	TEX15	SMARCA5	SMARCA4
PPP1CA	TAF6	SMARCD2	IPO5
NOL9	UBR4	MBD3	YEATS4
KDM1A	CRNKL1	SOD1	NUP93
RBBP4	HSPA8	MTA2	DNMT3L
HDAC2	RNF17	RBBP5	CHD4
ACTL6A		STIP1	HDAC1
PSMC5		CHAMP1	IK
POLR2A		RRP8	BAG6
KDM1B		SGTA	MIWI2
NUMA1		NR2F6	SPOCD1
PPP1CB		SMARCB1	
ZXDC		NME1	
TAF1		MEAF6	
TFIP11		BCAS3	
POLR2B		PSMB1	
		INTS1	
		KDM4B	
		BABAM1	
		SARNP	
		SUMO3	
		SNX6	
		DPPA4	
		SFPQ	
		VRK3	
		CSE1L	
		TPP2	
		RALY	
		PPIA	
		SNRNP40	
		HSPA1A	

Figure 6.18: **Common proteins between DNMT3C, SPOCD1, and MIWI2 IP-MS datasets**. (A) Venn diagram showing the number of proteins enriched in each IP-MS dataset and the number of common proteins. (B) List of common proteins in different combinations.

6.2.11 DNMT3C and nuclear piRNA pathway proteins interact in male germ cells

DNA methylation of the transposons is assumed to be the last step in the nuclear piRNA pathway. piRNA-loaded MIWI2 identifies the actively transcribing transposon loci and triggers the downstream pathway of transcriptional silencing and inactivation of the promoter. DNMT3C could be guided to its targets by interacting with MIWI2 and other nuclear piRNA pathway proteins. For validation of these interactions, I performed co-immunoprecipitation- western blotting. In the IP, along with DNMT3C, MIWI2 and SPOCD1 were also co-immunoprecipitated (Figure 6.19A). I also performed immunofluorescence staining in testes at stage E17.5 from *Dnmt3c*$^{myc/myc}$ testes. As the DNA methylation proceeds, DNMT3C forms foci in the nuclei. As shown in Figure 6.19B, DNMT3C and MIWI2 are spread out and in foci. They colocalize in many spots in the nuclei of germ cells. This further confirms that DNMT3C interacts with MIWI2 in germ cells. The study that performed DNMT3C HA-IP has shown that DNMT3C colocalizes with SPOCD1 and C19ORF84H foci in germ cells (Zoch et al., 2024). These results conclusively show that DNMT3C interacts with the nuclear piRNA pathway proteins in germ cells during the germline wave of epigenetic reprogramming.

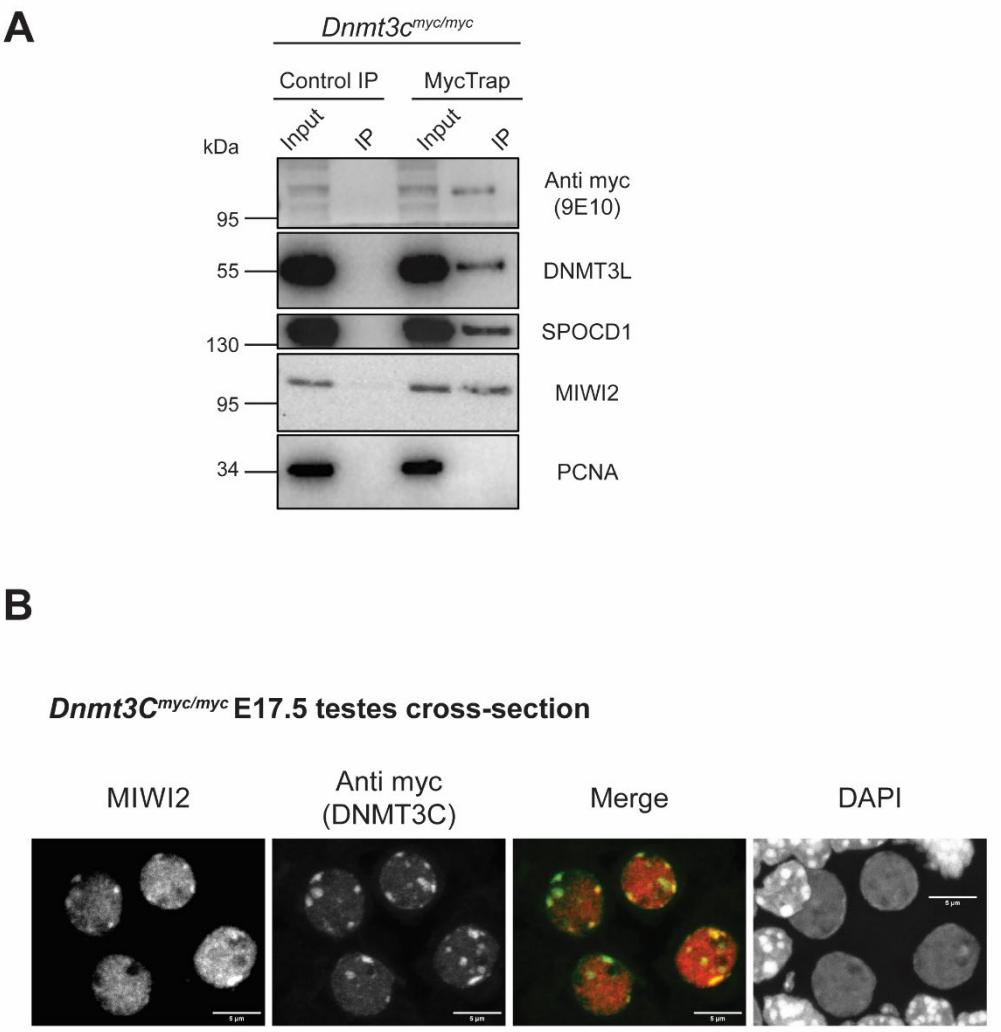


Figure 6.19: **DNMT3C interacts with nuclear piRNA pathway proteins**. (A) Western blots show that pulling down DNMT3C co-immunoprecipitates MIWI2 and SPOCD1. SPOCD1 antibody was obtained from the Dónal O'Carroll laboratory, and the MIWI2 antibody was obtained from the Haruhiko Siomi lab (Hirano, Hasuwa and Siomi, 2017). (B) Immunofluorescence shows that DNMT3C and MIWI2 colocalize in many foci in germ cell nuclei at E17.5. The scale bar is 5μm.

6.2.12 DNMT3C, MIWI2, and SPOCD1 have the same genomic targets

Dnmt3c$^{KO/KO}$, *Piwil4*$^{KO/KO}$, and *Spocd1*$^{KO/KO}$ share the same organism-level phenotype – male-specific sterility. I checked the DNA methylation defects in the three mutants to see if they have the same molecular phenotype. To do this, my colleagues Ania and Ishita analyzed the published whole-genome DNA methylation data for all three knockouts [*Dnmt3c*$^{KO/KO}$ - (Barau et al., 2016); *Piwil4*$^{KO/KO}$, and *Spocd1*$^{KO/KO}$ - (Zoch et al., 2020)].

The two studies generated their methylation data using different techniques (WGBS and EM-Seq). They had different average minimum read coverages over each cytosine in a CpG dinucleotide context (6 for *Dnmt3c*$^{KO/KO}$, 3 for *Piwil4*$^{KO/KO}$, and *Spocd1*$^{KO/KO}$). Therefore, for comparative purposes, the DMR (differentially methylated region) calling was performed with a minimum coverage of 3 reads for every cytosine in a CpG dinucleotide context. The WT control for respective studies was used for the DMR calling of the mutants. Due to the repetitive nature of the transposon promoters, the DMR calling algorithm might identify different regions of the promoter as a DMR for different mutants. To account for this inconsistency, we extended the DMRs by 1000bp (500bp upstream and 500bp downstream), as the average size for transposon promoters ranges around 700bp.

The DMRs of all three mutants majorly comprise repeats, in particular, the autonomous transposons LINEs and LTR/ERVs (Figure 6.20). Only a tiny proportion of genic elements (Promoters, 5' and 3' UTRs, CDS, downstream) and non-coding RNA were affected. When the DMRs of the three mutants were compared, a significant proportion of them were found to be common. These overlapping DMRs are found at the promoters of evolutionarily young transposons (Figure 6.21). Interestingly, some DMRs of *Spocd1*$^{KO/KO}$ and *Piwil4*$^{KO/KO}$ mutants do not overlap with *Dnmt3c*$^{KO/KO}$ DMRs. These DMRs are called "SPOCD1 specific DMRs" and "MIWI2 specific DMRs" respectively. This suggested that other DNMTs might be methylating these regions. DNMT3A is a *de novo* DNA methyltransferase that methylates most of the germline genome. Although DNMT3B is simultaneously expressed in the germline, its effect on DNA methylation seems negligible, as most genome-wide DNA methylation in the germline is lost in *Dnmt3a*$^{cKO/cKO}$ conditional knockout mutants (Dura et al., 2022). To find out which DNMT3 enzyme methylates these *Piwil4*$^{KO/KO}$ and *Spocd1*$^{KO/KO}$ specific DMRs, I examined the levels of DNA methylation at these *specific* DMRs in WT, *Dnmt3a*$^{KO/KO}$, and *Dnmt3c*$^{KO/KO}$ Prospermatogonia at stage E18.5 (Dura et al., 2022). The metaplots in Figure 6.22 clearly show that these DMRs lack methylation in the *Dnmt3c*$^{KO/KO}$ mutants. This indicates that DNMT3C, and not DNMT3A, methylates these regions. These regions might show up as *Piwil4*$^{KO/KO}$ and *Spocd1*$^{KO/KO}$ specific DMRs in our comparison because of the differences in

experimental techniques, batch effects in sequencing, etc. In conclusion, the DNA methylation data indicates that all three proteins target the same genomic regions and that DNMT3C is the *de novo* DNA methyltransferase responsible for piRNA-directed DNA methylation.

The best way to compare the three mutants would be to perform the experiment at the same stage, using the same technique and sequencing simultaneously to avoid batch effects.

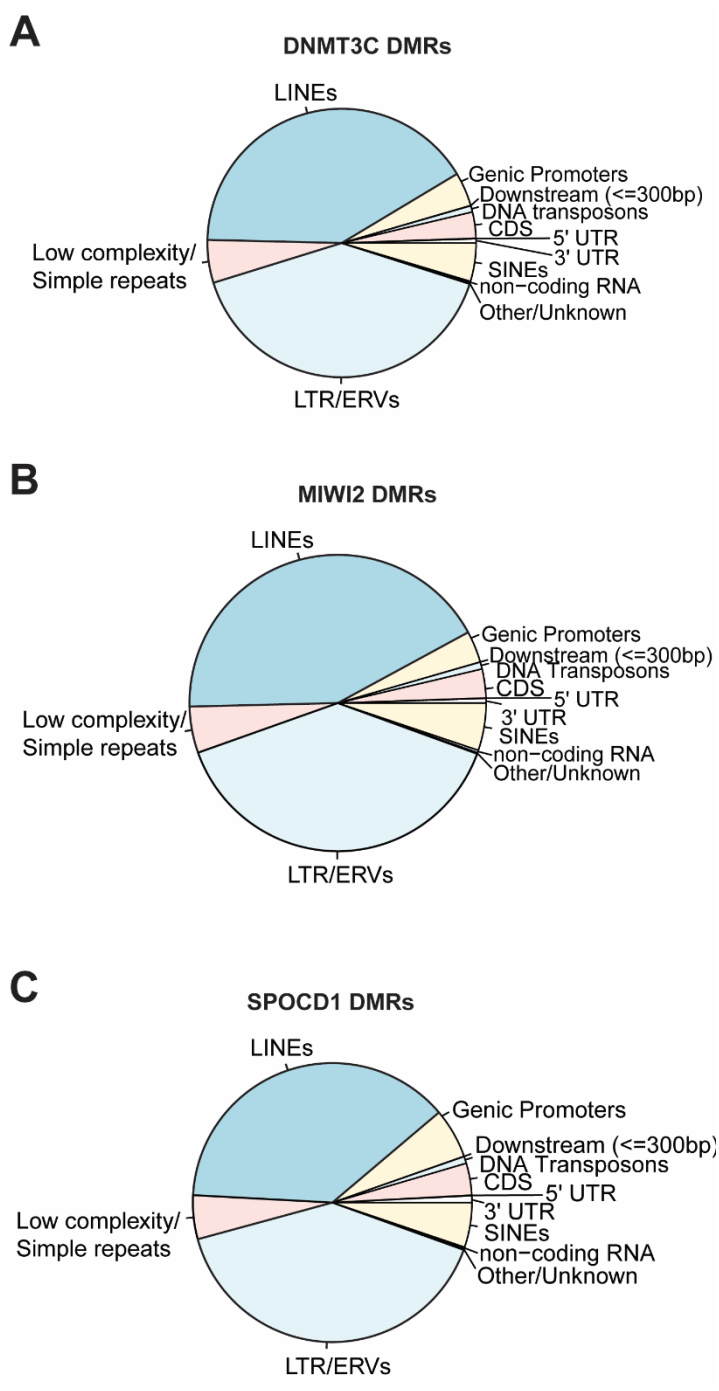


Figure 6.20: **DMR Annotation**. The DMRs of DNMT3C (A), MIWI2 (B), and SPOCD1 (C) are majorly made up of active transposon families LINEs and LTR/ERVs.

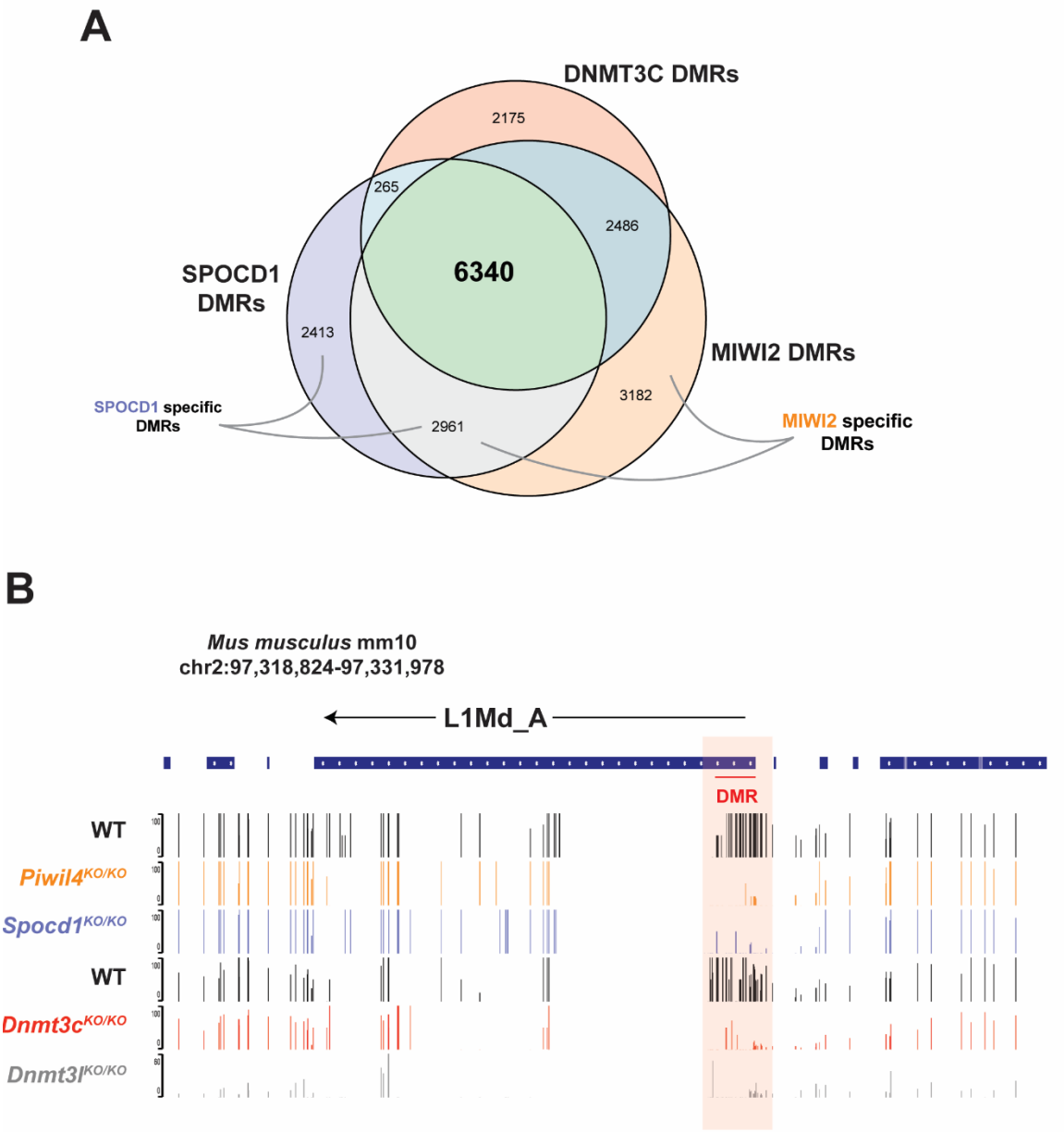


Figure 6.21: **DMR comparison between *Dnmt3c*$^{KO/KO}$, *Spocd1*$^{KO/KO}$ and *Piwil4*$^{KO/KO}$ mutants**. (A) Venn diagram showing the overlap of the DMRs. (B) Representative LINE1_MdA loci showing DNA methylation loss at its promoters in different mutants. WT for each study is represented right above the mutants.

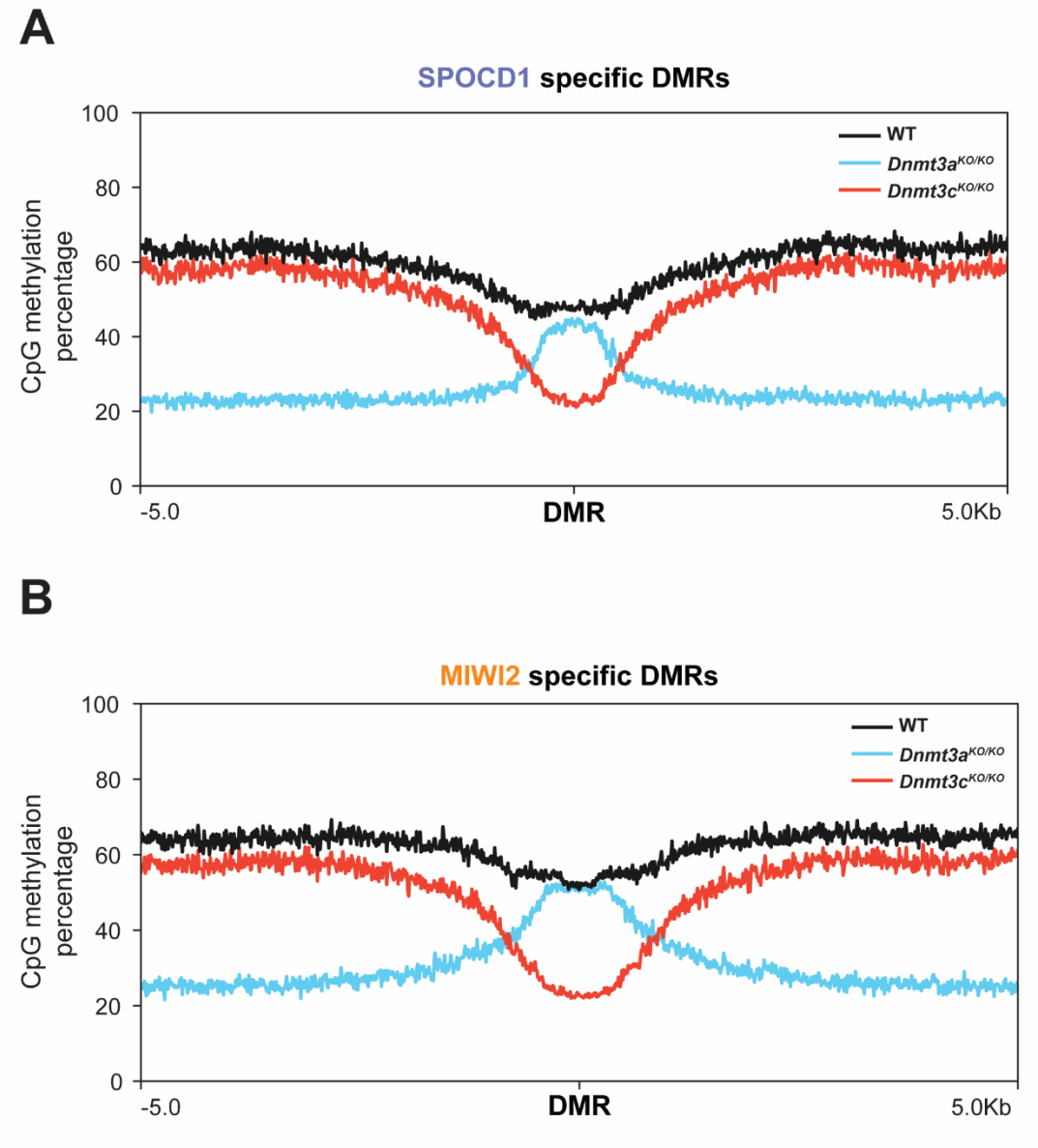


Figure 6.22: **CpG DNA methylation in Prospermatogonia at stage E18.5 from WT, *Dnmt3a*$^{KO/KO}$, and *Dnmt3c*$^{KO/KO}$ in SPOCD1 and MIWI2 -specific DMRs regions**. This data shows that DNA methylation is defective in all DMRs when DNMT3C is absent. [Data from (Dura et al., 2022)]

6.2.13 Hypothetical model of DNMT3C mode of action

DNMT3C does not affect the piRNA biogenesis (Barau et al., 2016). Additionally, the nuclear localization of MIWI2 is not affected in *Dnmt3c*$^{KO/KO}$ mutants (Figure 6.23). This shows that the cytoplasmic piRNA pathway proceeds normally without DNMT3C.

Combining the protein-protein interaction, nuclear co-localization, and DMR comparison data, I can answer my first question regarding how DNMT3C's targets are specified – the nuclear piRNA pathway targets DNMT3C. In this model, piRNA-loaded MIWI2 translocates into the nuclei and marks the locations of actively transcribing transposons by piRNA-mediated complementarity. DNMT3C interacts with the members of the nuclear piRNA pathway and is specified to the transposon loci, where it methylates the promoter DNA (Figure 6.23).

The second question I wanted to address was why DNMT3A or DNMT3B cannot replace DNMT3C. The interaction of DNMT3C with piRNA pathway proteins could explain this. We can assume that the piRNA-mediated transposon inactivation is a step-wise process – step 1 being the silencing of transcription from the transposon promoter and step 2 being promoter inactivation by DNA methylation. If this holds, DNMT3A or DNMT3B can methylate the silenced transposon promoter due to the removal of active histone modifications, especially the DNA methylation antagonistic mark H3K4me. The rationale is that *Dnmt3a* and *Dnmt3b* are expressed at orders of magnitude higher than *Dnmt3c*. Additionally, DNMT3A methylates most of the genome, including around the regions targeted by DNMT3C. Since we do not see DNMT3A/DNMT3B's participation, I wanted to check if DNMT3C has other functions, making it irreplaceable. MIWI2 is essential for remodeling the active histone mark H3K4me2 at its targets (Nagamori et al., 2018). Such H3K4me3 enrichment over IAPs and LINE1's was also seen in the *Dnmt3l*$^{KO/KO}$ mutant (Zamudio et al., 2015). This prompted me to study the relationship of DNMT3C with this histone modification at its targets.

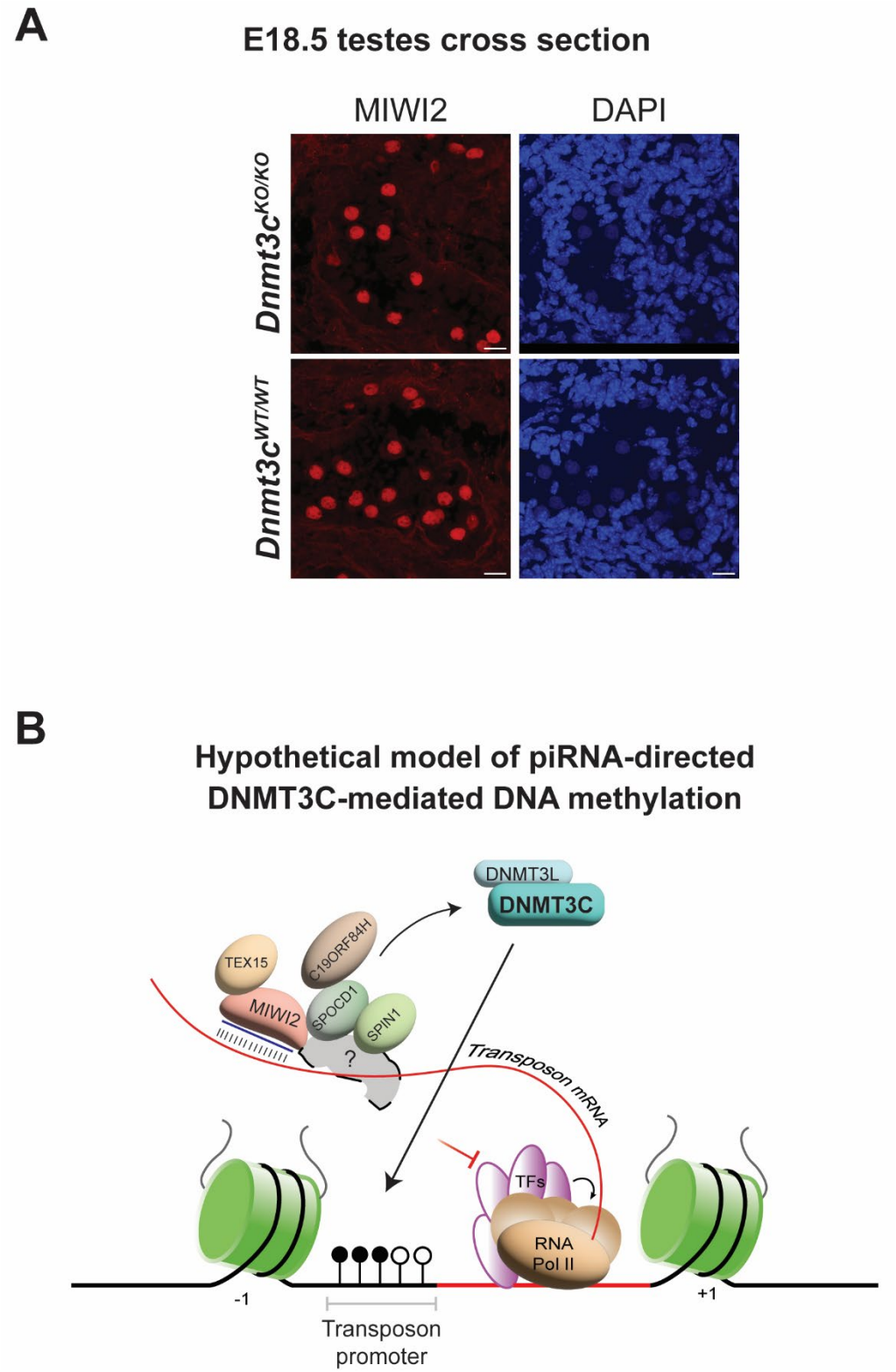


Figure 6.23: **Hypothetical model of DNMT3C action**. (A) An immunofluorescence image of an E18.5 testes cross section showing MIWI2 nuclear localization is unaffected in *Dnmt3c*$^{KO/KO}$. The scale bar is 10μm. (B) Illustration of the hypothetical model of DNMT3C action. MIWI2 recognizes the nascent transposon transcript. SPOCD1, SPIN1, C19ORF84H, TEX15, and other nuclear piRNA pathway factors assemble and recruit the DNMT3C-DNMT3L complex to methylate the transposon promoter.

6.3 RELATIONSHIP BETWEEN DNMT3C AND H3K4ME3

6.3.1 DNMT3C is essential for remodelling H3K4me3 at its targets

DNMT3C DMRs gain active histone modification, such as H3K4me3, and chromatin accessibility, as they are demethylated during the germline wave of epigenetic reprogramming (Yamanaka et al., 2019) and are transcriptionally active (Zamudio et al., 2015). These mRNAs feed into the piRNA pathway, ultimately marking these loci for inactivation. As the embryo development continues, the transposons lose these active marks (Figure 6.24) and gain repressive histone modifications (Dura, 2021; Yamanaka et al., 2019). I wanted to see if DNMT3C is involved in this chromatin remodeling at its targets. My colleague, Jessica, profiled H3K4me3 using CUT&TAG in sorted Spermatogonial stem cells and Spermatogonia from *Dnmt3c*$^{KO/WT}$ and *Dnmt3c*$^{KO/KO}$ littermates. I analyzed the data and found out that in *Dnmt3c*$^{KO/KO}$ mutants, the H3K4me3 is still present, while in the *Dnmt3c*$^{KO/WT}$ littermate controls, this mark is virtually absent (Figure 6.25A). This effect is specific to DNMT3C DMRs as there is no change in H3K4me3 status at gene promoters, which are not targets of DNMT3C (Figure 6.25B). This could be because of two reasons – (1) H3K4me3 is usually remodeled during embryonic development in *Dnmt3c*$^{KO/KO}$, but DNMT3C DMRs gain this mark at later stages of development post-birth due to the absence of DNA methylation, OR (2) H3K4me3 was never removed at the DNMT3C targets in *Dnmt3c*$^{KO/KO}$. To test this, I profiled for H3K4me3 in sorted Prospermatogonia at stage P2 (2 days after birth) using CUT&TAG. I chose the time point P2 because the wave of piRNA-dependent and -independent DNA methylation would have just finished, and the Prospermatogonia are yet to differentiate into the Spermatogonial stem cells or Spermatogonia. Analyzing this CUT&TAG data, I saw the same at P2 as for P10 (Figure 6.25 C and D). In the *Dnmt3c*$^{KO/KO}$ mutant, H3K4me3 is retained at DNMT3C targets, but it is almost completely removed in the *Dnmt3c*$^{KO/WT}$ littermate controls. This tells us that DNMT3C is essential for removing this antagonistic mark, H3K4me3, from its targets during embryonic development.

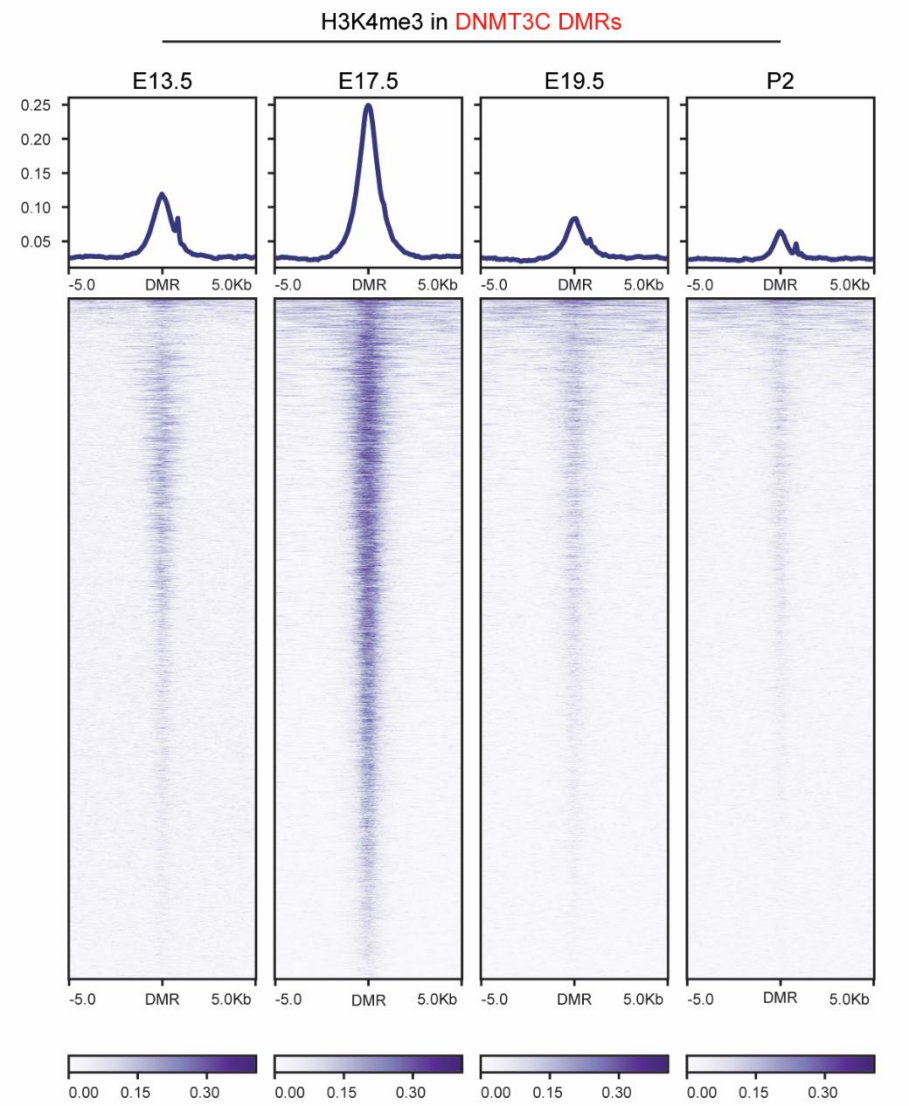


Figure 6.24: **H3K4me3 profiling at DNMT3C DMRs in *Dnmt3c*$^{WT/WT}$ Prospermatogonia**. [Data from (Yamanaka et al., 2019)]

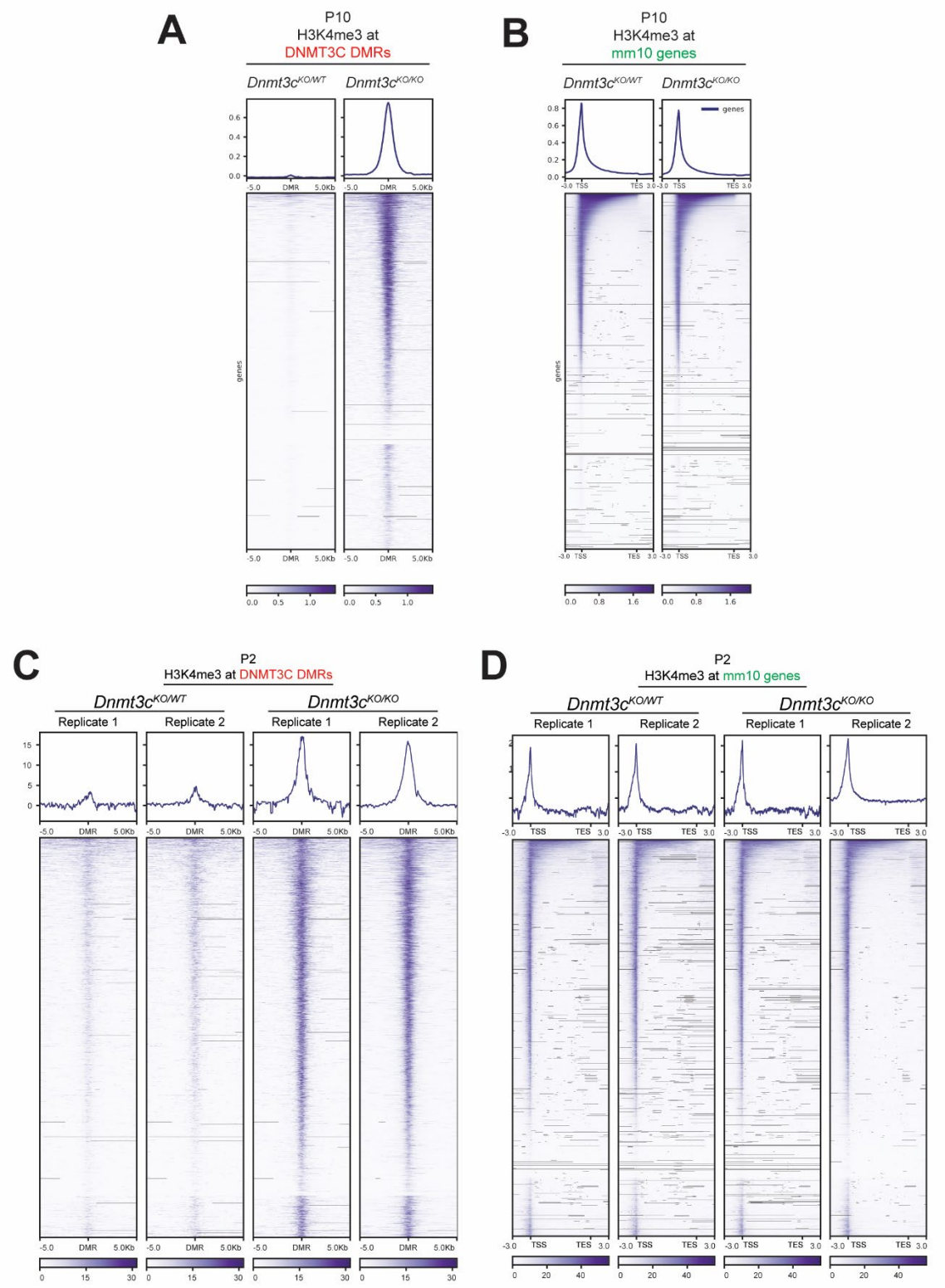


Figure 6.25: **Profiling of H3K4me3 in P10 and P2 Spermatogonia**. H3K4me3 at DNMT3C DMRs (A) and mm10 genes at stage P10 (B) (Data from Jessica Leismann). H3K4me3 at DNMT3C DMRs (C) and mm10 genes at stage P2 (D). Both data show that DNMT3C is critical for removing H3K4me3 at its targets.

This helps explain the irreplaceability of DNMT3C. Without DNMT3C, H3K4me3 is not removed from its targets, which can inhibit the activity of DNMT3A or DNMT3B by preventing their necessary interaction with the chromatin (Figure 6.26).

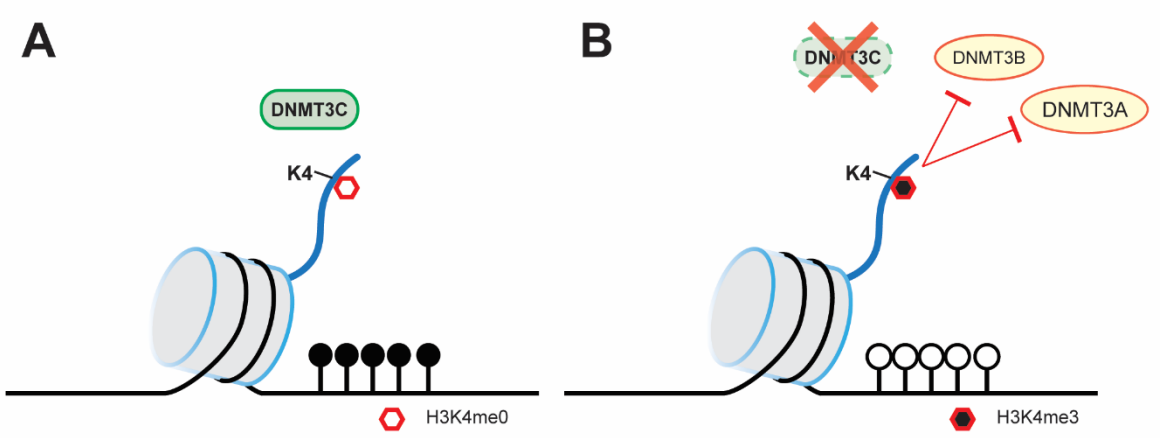


Figure 6.26: **Schematic illustration showing why DNMT3A/B cannot methylate transposon promoters**. (A) H3K4me3 is usually removed from transposon promoters in the WT, and DNMT3C methylates the DNA. (B)However, in *Dnmt3c*$^{KO/KO}$, H3K4me3 is retained, which can repel DNMT3A/3B from methylating that region.

6.3.2 H3K4me is remodeled only at the late stages of embryonic development

I profiled H3K4me2 and H3K4me3 in sorted Prospermatogonia at stages E14.5, E16.5, E18.5 using CUT&TAG. DNMT3C DMRs are unmethylated at E14.5-16.5, and subsequently, they become transcriptionally active by E16.5. In agreement with published studies (Yamanaka et al., 2019), DNMT3C DMRs gain H3K4me3 and H3K4me2 at E16.5 compared to E14.5 (Figure 6.27). There is also very little difference in the H3K4me3 & H3K4me2 marks between *Dnmt3c*$^{KO/KO}$ and *Dnmt3c*$^{KO/WT}$ mice at these stages. This is expected as DNMT3C expression starts at E15.5 (Figure 6.27), and the DNA methylation begins at E16.5 and proceeds through E19.5 (Dura, 2021).

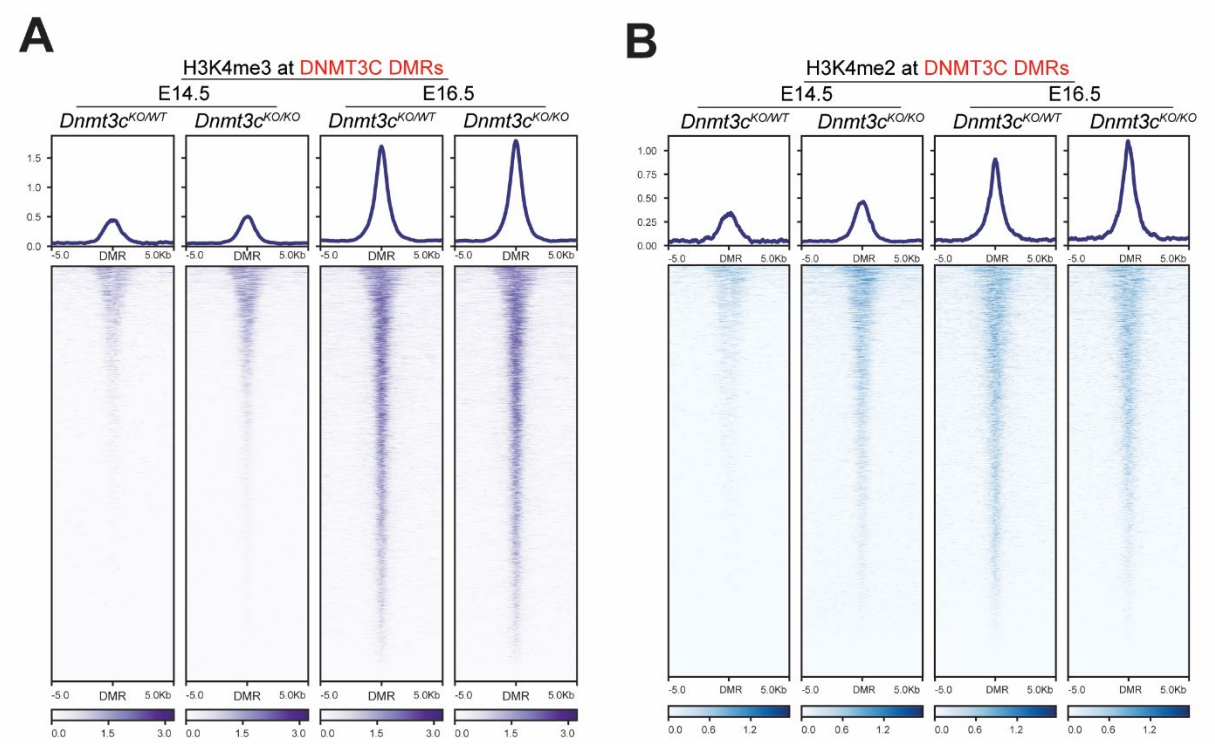


Figure 6.27: **Heatmaps showing profiles of H3K4me3 and H3K4me2 at stages E14.5 and E16.5 in *Dnmt3c*$^{KO/WT}$ and *Dnmt3c*$^{KO/KO}$.**

Intriguingly, there is very little difference in the H3K4me3 levels at E18.5 between *Dnmt3c*$^{KO/KO}$ and *Dnmt3c*$^{KO/WT}$ (Figure 6.28A). Levels of H3K4me2 are slightly higher in *Dnmt3c*$^{KO/KO}$ than in *Dnmt3c*$^{KO/WT}$ at DNMT3C DMRs (Figure 6.28C). These differences were specific to DNMT3C DMRs as H3K4me3 and H3K4me2 levels over mm10 genes did not show any differences between *Dnmt3c*$^{KO/KO}$ and *Dnmt3c*$^{KO/WT}$ littermates (Figure 6.28 B and D). This was surprising because the DNMT3C DMRs had already been methylated, albeit not entirely, by E18.5 (Dura et al., 2022) . Given the relationship between DNMT3 proteins and H3K4me, I expected H3K4 demethylation to precede DNA methylation (Guo, X. et al., 2015) and, therefore, significantly higher H3K4me3/me2 at DNMT3C DMRs in *Dnmt3c*$^{KO/KO}$. My CUT&TAG data shows that H3K4me was not remodeled by stage E18.5 but completely remodeled three days later by stage P2. Therefore, I added another time point in between and performed CUT&TAG at E19.5 for H3K4me3 (Figure 6.29 A and B). This showed that H3K4me3 is highly retained in *Dnmt3c*$^{KO/KO}$ but almost completely removed in *Dnmt3c*$^{KO/WT}$, as

we observed for P2. It is highly intriguing that DNMT3C possibly methylates DNA marked by H3K4me3 and that residual H3K4me3 is remodeled within one day of DNA methylation completion. This goes against the canonical mechanism of DNA methylation and suggests a non-canonical mechanism for DNMT3C.

This unexpected result could be possible for a few reasons. The H3K4me3 signal from *Dnmt3c*$^{KO/KO}$ at E18.5 may arise from cells that have yet to methylate the DNMT3C DMRs (as DNMT3C-mediated methylation is incomplete yet). This germ cell heterogeneity could be masked when performing bulk sequencing. Additionally, DNMT3C DMRs are repetitive regions. This could erroneously distribute the signal throughout all the DMRs, even if only a handful of DMRs contribute to the signal. After mapping the CUT&TAG data to the mm10 genome, I normalized the data using the RPKM method because of the lack of spike-in DNA. H3K4me3 CUT&TAG at E18.5 will be repeated with spike-in DNA normalization, or single-cell CUT&TAG can be performed to study this better.

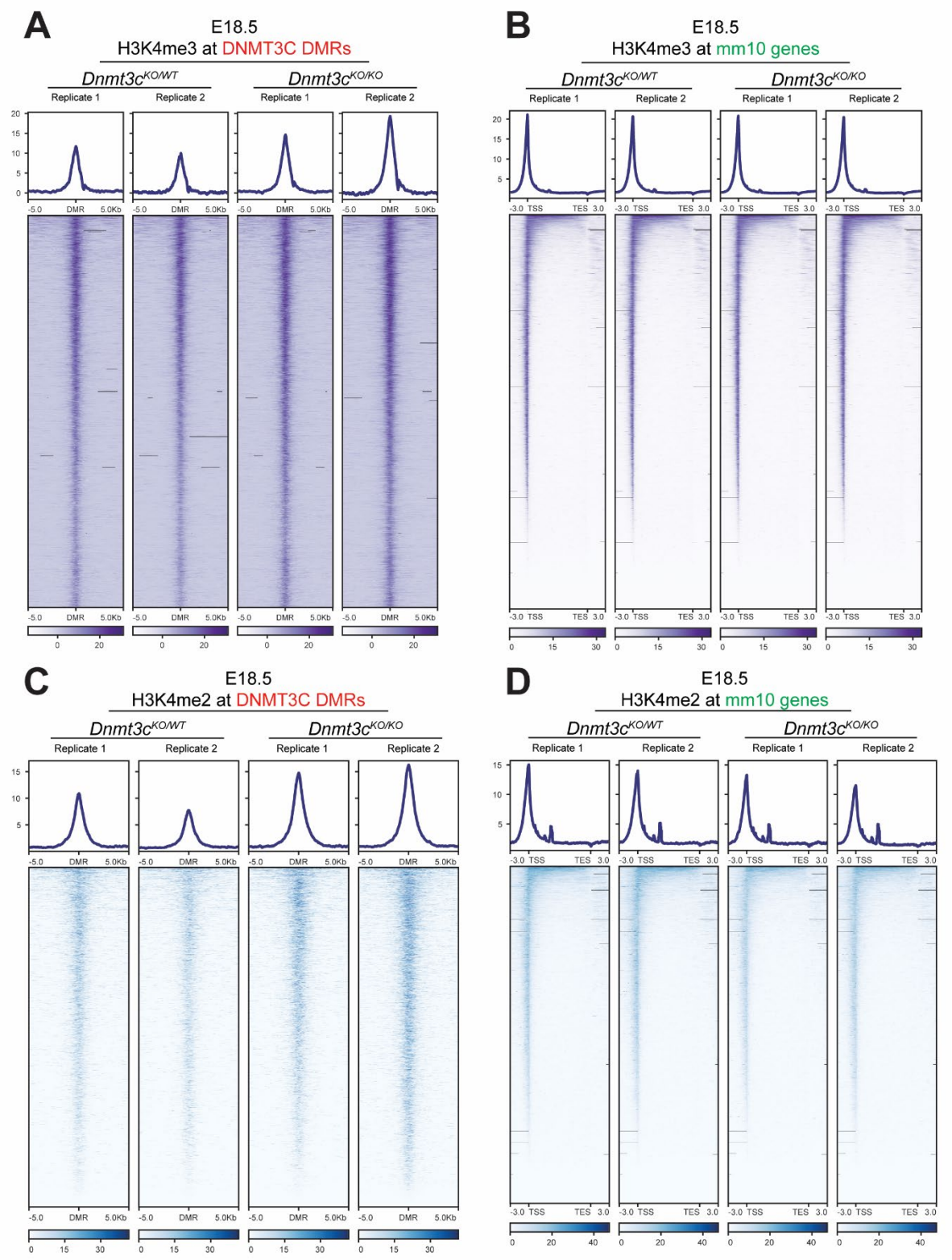


Figure 6.28: **Heatmaps showing H3K4me3 and H3K4me2 profiling at E18.5**. H3K4me3 at DNMT3C DMRs (A) show little difference between *Dnmt3c*$^{KO/WT}$ and *Dnmt3c*$^{KO/KO}$ littermates. H3K4me2 at DNMT3C DMRs (C) shows a slightly more significant difference between *Dnmt3c*$^{KO/WT}$ and *Dnmt3c*$^{KO/KO}$ littermates. These differences are specific to DNMT3C DMRs as both marks do not show any difference in mm10 genes in *Dnmt3c*$^{KO/WT}$ and *Dnmt3c*$^{KO/KO}$ littermates (B and D).

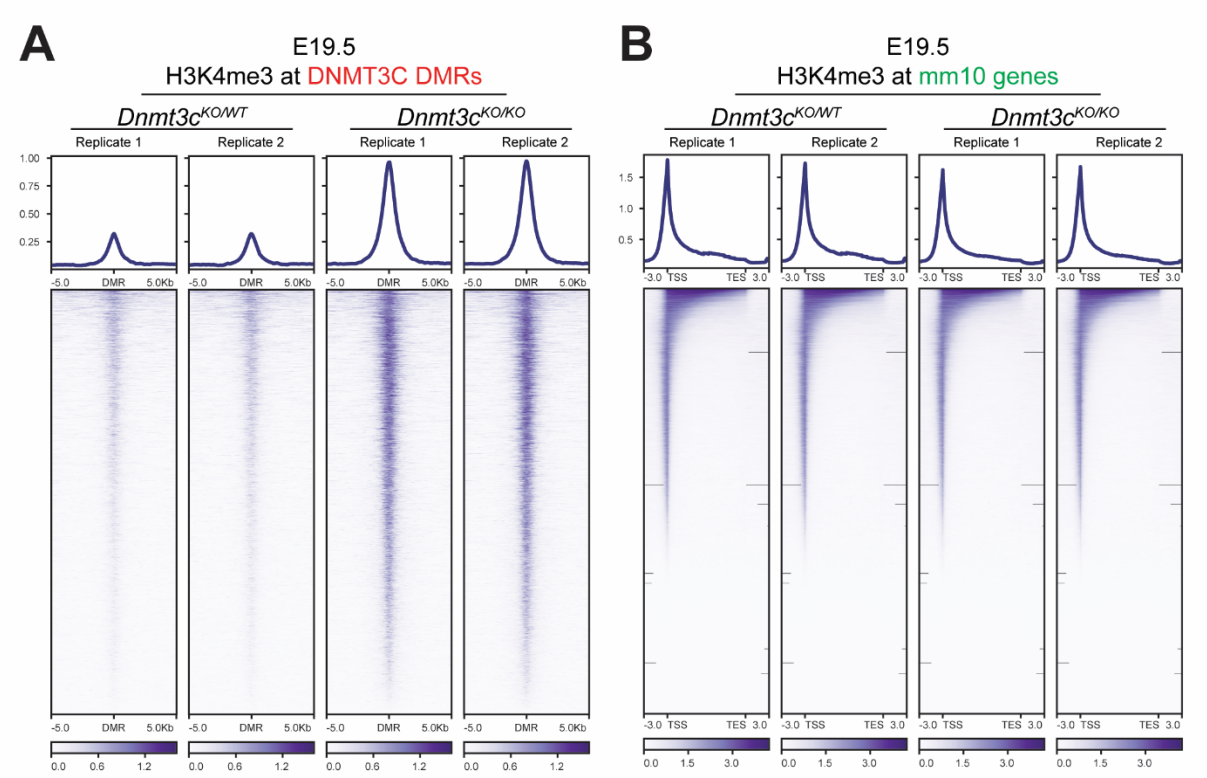


Figure 6.29: **Heatmaps showing H3K4me3 profiling at E19.5**. Heatmaps showing profiling of H3K4me3 at DNMT3C DMRs (A) and mm10 genes (B) sorted Prospermatogonia from *Dnmt3c*$^{KO/WT}$ and *Dnmt3c*$^{KO/KO}$ littermates at E19.5.

6.4 DNMT3C – CANONICAL OR NON-CANONICAL DNA METHYLATION

6.4.1 DNMT3C might have reduced autoinhibition

A recent structural study focusing on human DNMT3B protein showed that specific amino acid substitutions in the ADD domain caused steric hindrance when the ADD domain interacts with the catalytic domain of DNMT3B (Boyko, K. et al., 2022). Pulldown assay showed that DNMT3B's ADD domain weakly interacted with the catalytic domains of DNMT3B and DNMT3A, whereas DNMT3A's ADD domain interacted strongly with both the catalytic domains. This suggests that the DNMT3B's ADD domain causes the weak interaction, and therefore, the ADD-Catalytic domain auto-inhibition seen in DNMT3A and DNMT3L could be reduced in DNMT3B. Their study found the amino acid residues that extended the conformation of the autoinhibitory loop in human DNMT3B as L465, F466, and M468 (Boyko, K. et al., 2022). Multiple sequence alignments of the ADD domains from human DNMT3A/B and Mouse DNMT3A/B/C/L showed that Mouse DNMT3B and DNMT3C showed the same amino acid substitutions as human DNMT3B (Figure 6.30A). This data suggests that Mouse DNMT3C's ADD domain could also bind weakly to its catalytic domain.

Structural comparison of the experimentally determined Human DNMT3B's ADD domain (PDB: 7O45) (Boyko, K. M. et al., 2022) with the AlphaFold predicted structures of Mouse DNMT3C's ADD domain shows that the latter does not have the extended conformation of the former (Figure 6.30B) (Jumper et al., 2021). This could signify no steric hindrance during ADD-CD interaction and, therefore, no reduced autoinhibition. However, comparing the PDB structure with Human DNMT3B's AlphaFold predicted structure shows the same lack of extended conformation even though they are the same proteins. DNMT3C could also have the extended conformation that AlphaFold cannot predict. This reduced autoinhibition could enable DNMT3C to methylate DNA marked by H3K4me3, provided it is stabilized by other protein-protein interactions at its targets. Experimentally determining the structure of DNMT3C's ADD domain will shine more light on this possibility.

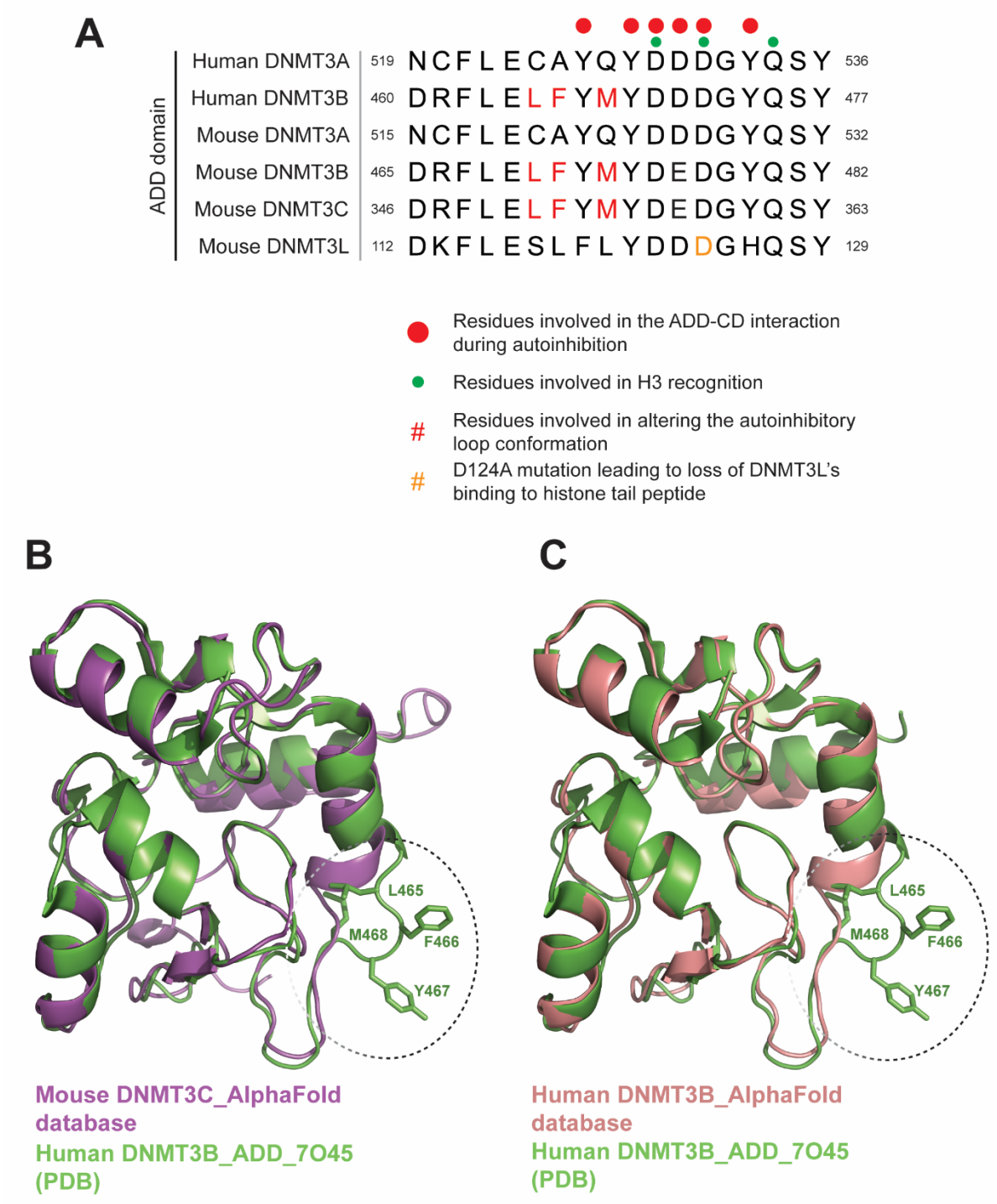


Figure 6.30: **Sequence and structural comparison of different ADD domains**. (A) Multiple sequence alignment of ADD domain sequences of different DNMT3 proteins. Mouse DNMT3B and DNMT3C possess the same amino acid substitutions that human DNMT3B has (in red), which cause steric hindrance during ADD-CD interaction. (B) & (C) Comparison of the structure of ADD domains of Human DNMT3B (PDB: 7O45) with Mouse DNMT3C (from Alpha Fold database) (B) OR with Human DNMT3B (from Alpha Fold database) (C). The experimentally determined structure of DNMT3B within the dotted circle differs from the Alpha Fold prediction of Human DNMT3B and Mouse DNMT3C in the same way, even though the amino acid sequences are the same.

6.4.2 The piRNA pathway could also help anchor DNMT3C-DNMT3L

In the DNMT3B structural study, they also report that the hydrophobic amino acid substitutions in the Human DNMT3B ADD domain lead to the formation of an additional protein-protein interaction surface in the ADD domain (Boyko, K. et al., 2022). The DNMT3C's ADD domain might not bind the histone tail when H3K4 is methylated due to the conservation of the histone binding pocket across the different DNMT3 proteins. However, this potential additional protein-protein interaction in the ADD domain could initiate interactions with nearby proteins, stabilizing DNMT3C at its targets. Interestingly, abolishing the binding of DNMT3L to the histone tail does not abolish DNA methylation at transposon promoters. The D124A amino acid substitution in DNMT3L's ADD domain (Figure 6.30A– orange residue) results in the loss of the DNMT3L-Histone tail peptide interaction when H3K4 is unmethylated (Vlachogiannis et al., 2015). Whole-genome bisulfite sequencing of Prospermatogonia at stage P1 from *Dnmt3l*^{D124A} mutant showed a significant loss of DNA methylation at DNMT3C DMRs but not complete removal as seen in *Dnmt3c*$^{KO/KO}$ (at DMRs) or *Dnmt3l*$^{KO/KO}$ (at & around DMRs) mutants (Figure 6.31A). As DNMT3L also participates in DNMT3A-mediated methylation, loss of DNA methylation in *Dnmt3l*^{D124A} mutant is more widespread across the genome than seen in *Dnmt3c*$^{KO/KO}$ (Figure 6.31B). This indicates that DNMT3C can still methylate its targets to a certain extent even when the DNMT3C-DNMT3L complex is not anchored by DNMT3L-Histone interaction. DNMT3C's ADD-Histone interaction and stabilization through protein interactions with other nuclear piRNA pathway members could make such a mechanism possible. Structural studies of the DNMT3C-DNMT3L complex alone and interacting with the H3 tail should provide more detail on this hypothesis.

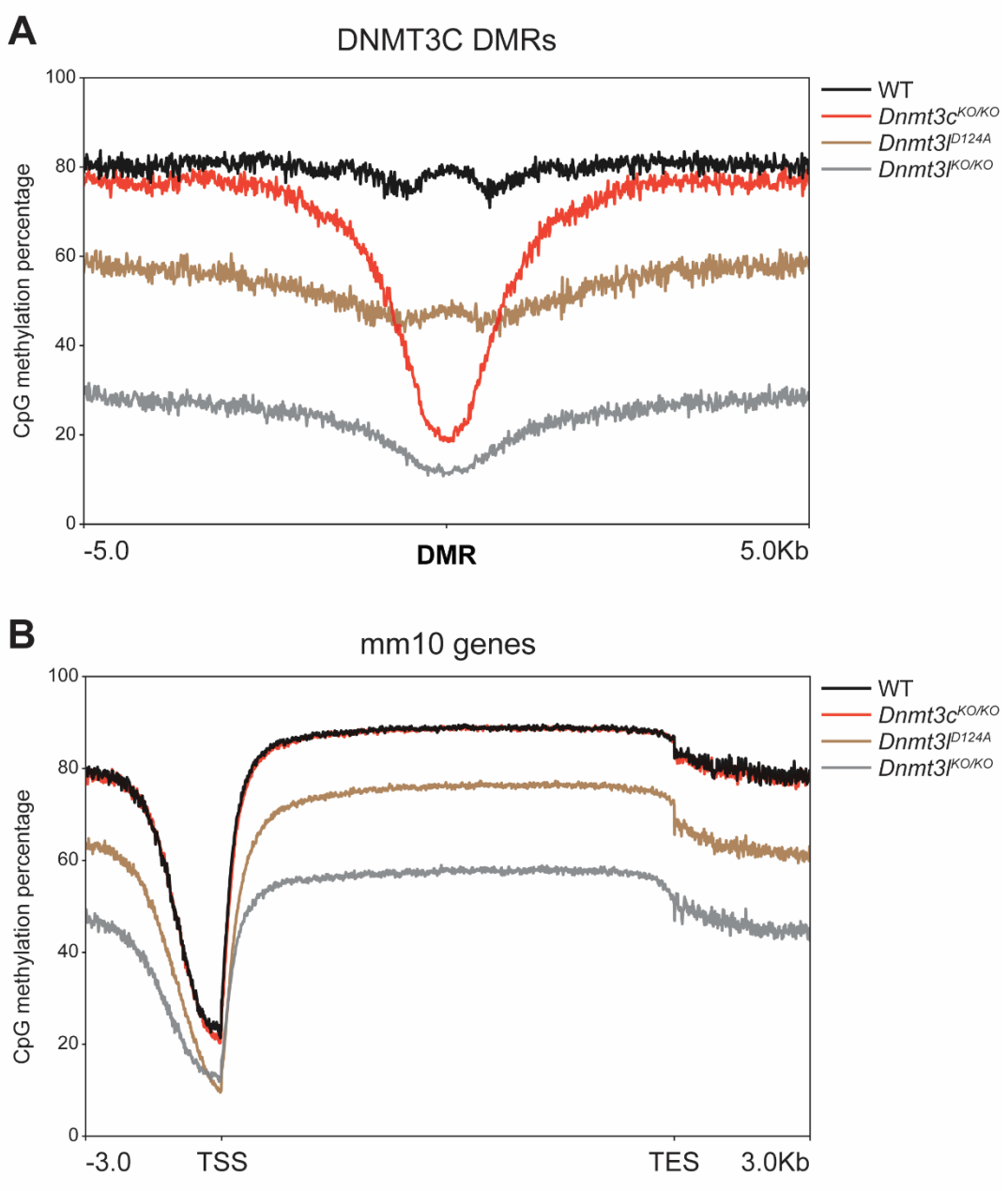


Figure 6.31: **piRNA pathway could anchor DNMT3C DNA methylation in *Dnmt3l*^{D124A} mutant**. Metaplots showing DNA methylation in different *Dnmt3c*$^{KO/KO}$, *Dnmt3l*^{D124A}, and *Dnmt3l*$^{KO/KO}$ mutants at DNMT3C DMRs (A) and at mm10 genes (B). [Data from (Dura et al., 2022; Vlachogiannis et al., 2015)].

6.4.3 CUT&TAG-BS shows canonical DNA methylation by DNMT3C

To examine if DNMT3C targets are DNA-methylated and marked by H3K4me3 at E18.5 in Prospermatogonia, I performed CUT&TAG-BS (CUT&TAG-Bisulfite sequencing). CUT&TAG-BS enables simultaneous measurement of DNA methylation and Histone modification on the same DNA molecule. CUT&TAG generates tagmented DNA at binding sites of a given histone modification, and the CUT&TAG DNA is then subjected to tagmentation-based bisulfite sequencing, resulting in a base-resolution DNA methylation map at genomic regions bound by the given histone modification (Li, Grimm and Wade, 2021). I performed CUT&TAG-BS on WT Prospermatogonia at E18.5 for H3K4me3 and H3K9me3. DNA methylation and H3K9me3 are positively correlated, so I used H3K9me3 as a positive control (Lehnertz et al., 2003; Rose and Klose, 2014). As observed before (Figure 6.28A), DNMT3C DMRs are occupied by H3K4me3 at E18.5 (Figure 6.32A). The DMRs are also marked by H3K9me3 (Figure 6.32A) (Dura, 2021; Yamanaka et al., 2019). The mm10 genes are marked by H3K4me3 at their TSS but not with H3K9me3 (Figure 4.30B), which helped me validate the CUT&TAG.

Upon analyzing the methylation data, we see that the H3K4me3-marked DNA molecules have very low average methylation (~7%) at DNMT3C DMRs (Figure 6.32C). In comparison, the H3K9me3-marked DNA molecules have a higher average DNA methylation (~45%) at DNMT3C DMRs (Figure 6.32D). This tells us that DNMT3C methylates H3K9me3-marked DNA but not H3K4me3-marked DNA. Therefore, DNMT3C, like DNMT3A/B/L, follows the canonical mechanism for DNA methylation. Another significant indication from this experiment is the difference in DNA methylation between the H3K4me3- and H3K9me3-marked DNA. This difference in DNA methylation indicates the heterogeneity and non-synchronous nature of the Prospermatogonia population. As germline reprogramming initiates, DNMT3C DMRs are marked by H3K4me3. At E18.5, some Prospermatogonia cells could be advanced in reprogramming and, thus, could have already removed H3K4me3 at DNMT3C DMRs by the nuclear piRNA pathway. This enables DNMT3C to methylate its targets in these cells, as seen in Figure 6.32D. Some cells might lag in reprogramming and, therefore, have

not yet removed H3K4me3 at DNMT3C DMRs. This mark would repel DNA methylation, so the DNA methylation levels at DNMT3C DMRs will be lower, as seen in Figure 6.32C. Such heterogeneity and the non-synchronous nature of Prospermatogonia reinstate the importance of repeating the H3K4me3 CUT&TAG at E18.5 with spike-in normalization or performing single-cell CUT&TAG (Bartosovic, Kabbe and Castelo-Branco, 2021).

Transposon promoters (LINE1A, LINE1T, and IAP) have been shown to have an average of ~65% DNA methylation by Pyrosequencing (Dura, 2021). I targeted these same regions in my Pyrosequencing assay and showed that the methylation of these promoters depends on DNMT3C (Figure 6.8A). However, in my H3K9me3 CUT&TAG-BS, I observe only ~45% DNA methylation at DNMT3C DMRs. This difference is most likely due to the differences in experimental setups.

The sequential and structural differences in DNMT3C's ADD domain discussed above could still create differences in conformation and activity. Such differences, along with the differences in its N-terminal region, could have made DNMT3C more suitable for its unique participation in the piRNA pathway. A detailed study of the enzymatic dynamics with the full-length DNMT3C protein and its histone tail binding potential will provide better insights. Since DNMT3C uniquely only participates in the piRNA pathway, studying its activity in the context of the piRNA pathway and not in isolation becomes more critical. This means reconstituting, at least, the critical elements, if not the entire nuclear piRNA pathway along with DNMT3C-DNMT3L by recombinant expression in a more malleable *in vitro* system (like mouse embryonic stem cells) would be the best way to study this system. Primordial stem cell-like cells (PGCLCs) could also be a potential system to study the piRNA pathway; however, only through recombinant expression, as PGCLCs lack piRNAs (Ramakrishna et al., 2022).

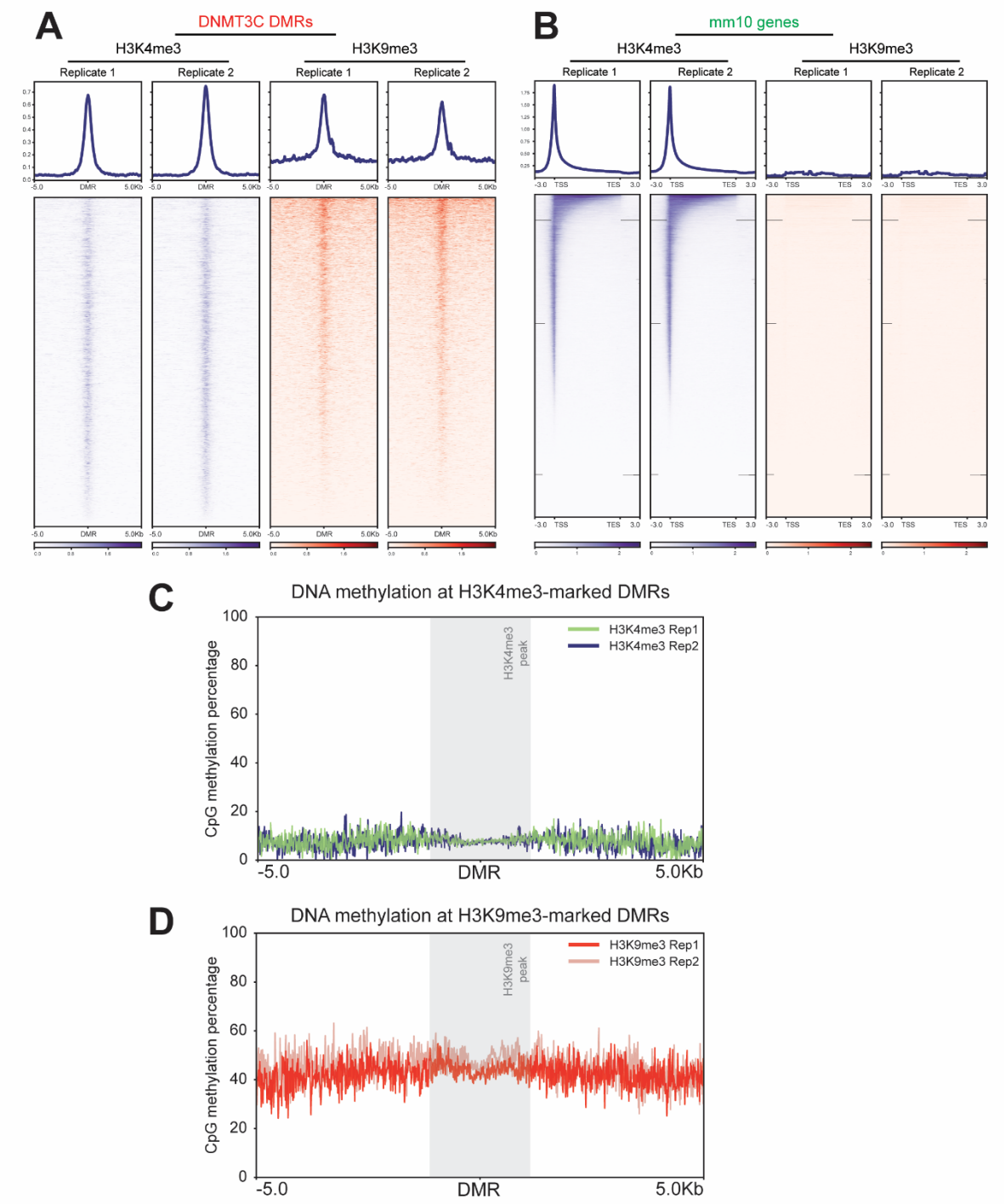


Figure 6.32: **CUT&TAG-BS shows canonical methylation by DNMT3C**. (A) & (B) Heatmaps showing H3K4me3 and H3K9me3 occupancy at DNMT3C DMRs (A) and mm10 genes (B). (C) & (D) CpG methylation levels at H3K4me3-marked regions (C) and H3K9me3-marked regions (D). This result shows that DNMT3C DMR that is H3K4me3-marked is not methylated, but DNMT3C DMR that is H3K9me3-marked is methylated.

6.5 DNMT3C RELATIONSHIP WITH LSD1

6.5.1 DNMT3C interacts with LSD1, a Lysine-specific histone demethylase

I wanted to know how DNMT3C enables H3K4me3 removal from its targets. I analyzed the DNMT3C IP-MS dataset and noticed that one of the highest-enriched proteins is LSD1 (Figure 6.33A). LSD1, or Lysine Specific Histone Demethylase 1, is a protein that removes H3K4me and H3K9me specifically (Metzger et al., 2005; Shi et al., 2004). Through this, it can act as a transcriptional repressor or activator, respectively. LSD1 has been shown to participate in the *Drosophila* piRNA pathway, where it interacts with Panoramix and removes H3K4me2 from the active transposon loci (Yu et al., 2015). LSD1 has been shown to interact with MIWI2 when tagged versions of each protein were expressed in 293T cells (Nagamori et al., 2018) . It is also significantly enriched in the SPOCD1 and MIWI2 Prospermatogonia IP-MS datasets (Zoch et al., 2020), increasing confidence in its participation in the mouse piRNA pathway. Additionally, it has been shown that DNMT3A interacts with LSD1 in mESCs. The LSD1-NuRD complex silences pluripotency gene enhancers during mESC differentiation by promoting DNMT3A-mediated DNA methylation through the removal of H3K4me2 (Petell et al., 2016). Interestingly, LSD1 and many of the NURD complex proteins are also enriched in the DNMT3C IP-MS (Figure 6.15B) and SPOCD1 IP-MS (Zoch et al., 2020). This suggests a role for this chromatin remodeling complex in the piRNA pathway. By MycTrap immunoprecipitation-western blot, I confirmed that DNMT3C interacts with LSD1 in E17.5/E18.5 testes (Figure 6.33B). Through its interaction with DNMT3C, LSD1 could be guided to the young transposon loci and remove the active histone mark, thereby enabling DNA methylation.

LSD1 has been shown to act only on H3K4me2 and H3K4me1 but not on H3K4me3 (Shi et al., 2004). Our CUT&TAG data shows that H3K4me3 marks the DNMT3C targets. This could mean there is another enzyme that removes this mark. Among the Histone demethylases specific for H3K4, only the JARID proteins demethylate H3K4me3 and H3K4me2, and LSD demethylases act on H3K4me2 and H3K4me1 (Kooistra and Helin, 2012). However, there have been conflicting reports of JARID1B's ability to demethylate K4me3/me2/me1.

Overexpression of JARID1B caused H3K4me1 demethylation in U2OS and HeLa cells. However, *in vitro* studies showed marginal or no activity for JARID1B in H3K4me1 (Christensen et al., 2007; Kristensen et al., 2012; Metzger et al., 2010; Xiang et al., 2007). The *in vivo* result could be an indirect consequence of overexpression. Therefore, complete demethylation of H3K4me3 would need the participation of both the JARID demethylases (me3/me2) and the LSD demethylases (me2/me1). Interestingly, none of the JARID proteins were significantly enriched in either the DNMT3C IP-MS or the IP-MS of SPOCD1 or MIWI2. An interaction between JARID proteins and the nuclear piRNA complex could be transient. A recent publication from Kristian Helin's lab showed that the global H3K4me3 is rapidly removed in the absence of core subunits of the H3K4 methyltransferase SET1/COMPASS complex (Wang, H. et al., 2023). They showed that the JARID1A and JARID1B proteins bring about this rapid removal. Even though JARID proteins can demethylate H3K4me2, the turnover of H3K4me2 is not as fast as H3K4me3. This suggests that JARID proteins continually turnover H3K4me3, which is then re-established through the SET1/COMPASS complex recruited by active transcription from the promoter. I hypothesize that JARID proteins could demethylate H3K4me3 at the DNMT3C targets in Prospermatogonia without being recruited by the nuclear piRNA pathway. This could explain their absence in the IP-MS datasets. Instead, DNMT3C could stabilize LSD1 at its target loci, which removes the remnant H3K4me2 and H3K4me1 marks and, in turn, enables accessibility for DNA methylation.

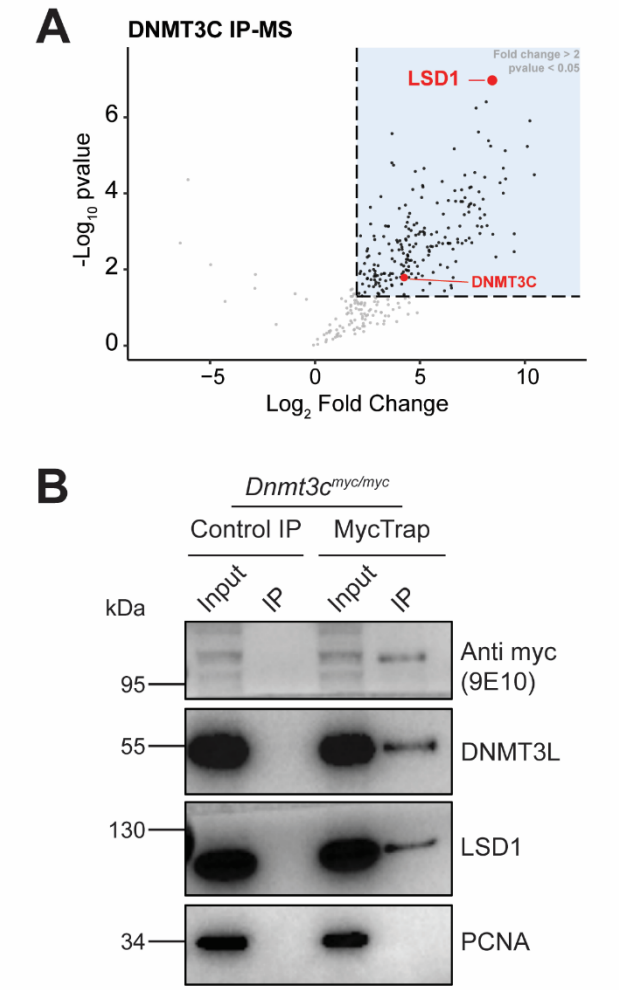


Figure 6.33: **DNMT3C interacts with LSD1**. (A) Volcano plots showing enrichment of LSD1 in the DNMT3C IP-MS. (B) Immunoprecipitation-western blot validation of DNMT3C-LSD1 interaction.

6.5.2 LSD1 Profiling in Prospermatogonia

To study this hypothesis, I profiled LSD1 in FACS-sorted Prospermatogonia from *Dnmt3c*$^{KO/KO}$ and *Dnmt3c*$^{KO/WT}$ littermates. Firstly, I optimized LSD1 CUT&RUN in mESCs. ChIP-seq was performed from LSD1 in mESCs (Cao et al., 2018), and this was used as a reference when designing the qPCR primers. I performed standard CUT&RUN (CNR) for LSD1 with 100,000 mESC cells per sample. The qPCR showed that there is high enrichment for "positive targets" (bound by LSD1 in the ChIP-seq) and not for the negative controls (Figure 6.34A), proving that the CUT&RUN has worked.

I performed CUT&RUN with sorted Prospermatogonia from stage E18.5 of *Dnmt3c*$^{KO/KO}$ and *Dnmt3c*$^{KO/WT}$ littermates. Contrary to our hypothesis, LSD1 can still bind DNMT3C DMRs in the *Dnmt3c*$^{KO/KO}$ mutant, although the binding seems slightly reduced (Figure 6.34B). This suggests that DNMT3C does not dictate the binding of LSD1 to the DMRs. CUT&RUN uses Micrococcal nuclease to cleave fragments bound by the protein of interest. These fragments float out of the cell and are collected from the supernatant. Since the nuclear piRNA pathway complex binds DNMT3C DMR regions and turns them into heterochromatin, these fragments might not readily float into the supernatant. Therefore, I modified the CUT&RUN protocol; after MNase digestion, I isolated the entire genomic DNA, not just the fragments from the supernatant. As MNase-generated fragments are short (maximum around 1000bp), I included a two-tailed size selection step in the library purification to exclude higher molecular weight DNA and adapter dimers. Profiling for LSD1 from E17.5 Prospermatogonia with this modified CUT&RUN protocol yielded the same result as we saw with the native CUT&RUN. LSD1 binding to DNMT3C DMRs is unaffected in *Dnmt3c*$^{KO/KO}$ (Figure 6.34C). I also modified the protocol of CUT&TAG (CNT), where I used NaCl at 150mM concentration in all the buffers, including the pA-Tn5 digestion and tagmentation buffers (instead of 300mM). I used this low NaCl concentration to preserve the interaction LSD1 with DNMT3C DMRs. Profiling LSD1 with this modified CUT&TAG protocol also showed that LSD1 binding to DNMT3C DMRs does not depend on DNMT3C (Figure 6.34D)

LSD1 has been shown to have functions independent of its catalytic activity. Catalytic inactivation of LSD1 has a mild impact on gene expression and cellular differentiation, whereas the deletion of LSD1 de-represses enhancers globally and impairs cell fate transition (Zeng et al., 2023). Additionally, LSD1 and LSD2 repress heterochromatic transcripts in *Schizosaccharomyces pombe* through mechanisms both dependent on and independent of their catalytic activities (Marayati et al., 2020). Therefore, DNMT3C may trigger the catalytic activity of LSD1 at DNMT3C DMRs but not its binding. To study this better, I created a germline conditional knockout of LSD1, as complete knockout of LSD1 is embryonically lethal in mice (Ancelin et al., 2016; Wang et al., 2007).

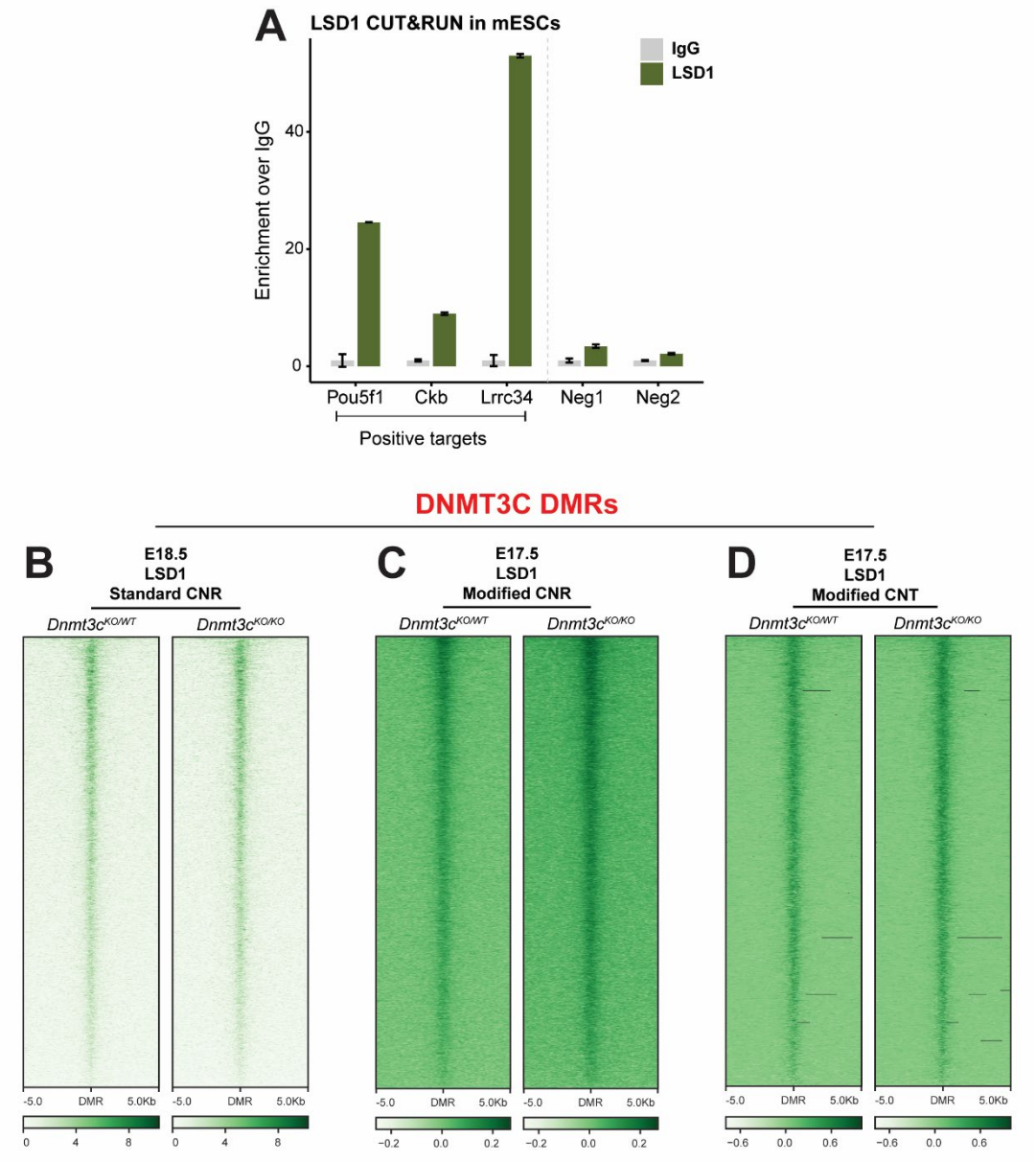


Figure 6.34: **LSD1 Profiling in Prospermatogonia**. (A) qPCR result from LSD1 CNR from mESC showing high enrichment for LSD1 over IgG for all positive targets. (B), (C) and (D) Heatmaps of LSD1 profiling centered at DNMT3C DMRs using standard CNR (B), modified CUT&RUN (C), or modified CUT&TAG (D). The heatmaps show that there is no change in LSD1 binding to DNMT3C targets, even in the absence of DNMT3C.

6.5.3 Germline Conditional Knockout of *Lsd1*

Germline conditional knockout (cKO) of LSD1 has already shown that it is essential for Spermatogenesis (Lambrot, Lafleur and Kimmins, 2015; Myrick et al., 2017). However, in both these studies, they drove the cKO using a *Vasa-Cre* driver. *Vasa-Cre* activity is strongly induced between E15.5 and E18.5, even though *Vasa (Ddx4)* is expressed much earlier in the male germ cells (Gallardo et al., 2007). Therefore, the *Vasa-Cre* driver would only create an LSD1 germline cKO from late Prospermatogonia/Spermatogonial stem cells, which is too late to study its role in the nuclear piRNA pathway. Therefore, I planned to create a cKO using the *Prdm1-Cre* driver (Ohinata et al., 2005). The *Prdm1-Cre* driver line (Ohinata et al., 2005) was a gift from the lab of Déborah Bourc'his in France. Mice carrying the LoxP alleles for conditional knockout (cKO) of *Lsd1* ($Lsd1^{fl/fl}$) or conditional catalytic knock-in (cKI) of *Lsd1* ($Lsd1^{fKI/fKI}$) were obtained from the lab of Roland Schüle in Germany (Duteil et al., 2016; Duteil et al., 2017). The $Lsd1^{fl/fl}$ allele loses the first exon of *Lsd1* after Cre-mediated recombination and becomes $Lsd1^{KO/KO}$, resulting in the ablation of the LSD1 protein (Figure 6.35A). The $Lsd1^{fKI/fKI}$ allele becomes $Lsd1^{KI/KI}$ after Cre-mediated recombination and expresses a full-length catalytically inactive LSD1 protein with a triple point mutation encoding K662A, W752A, and Y762S from the endogenous *Lsd1* promoter (Figure 6.35B).

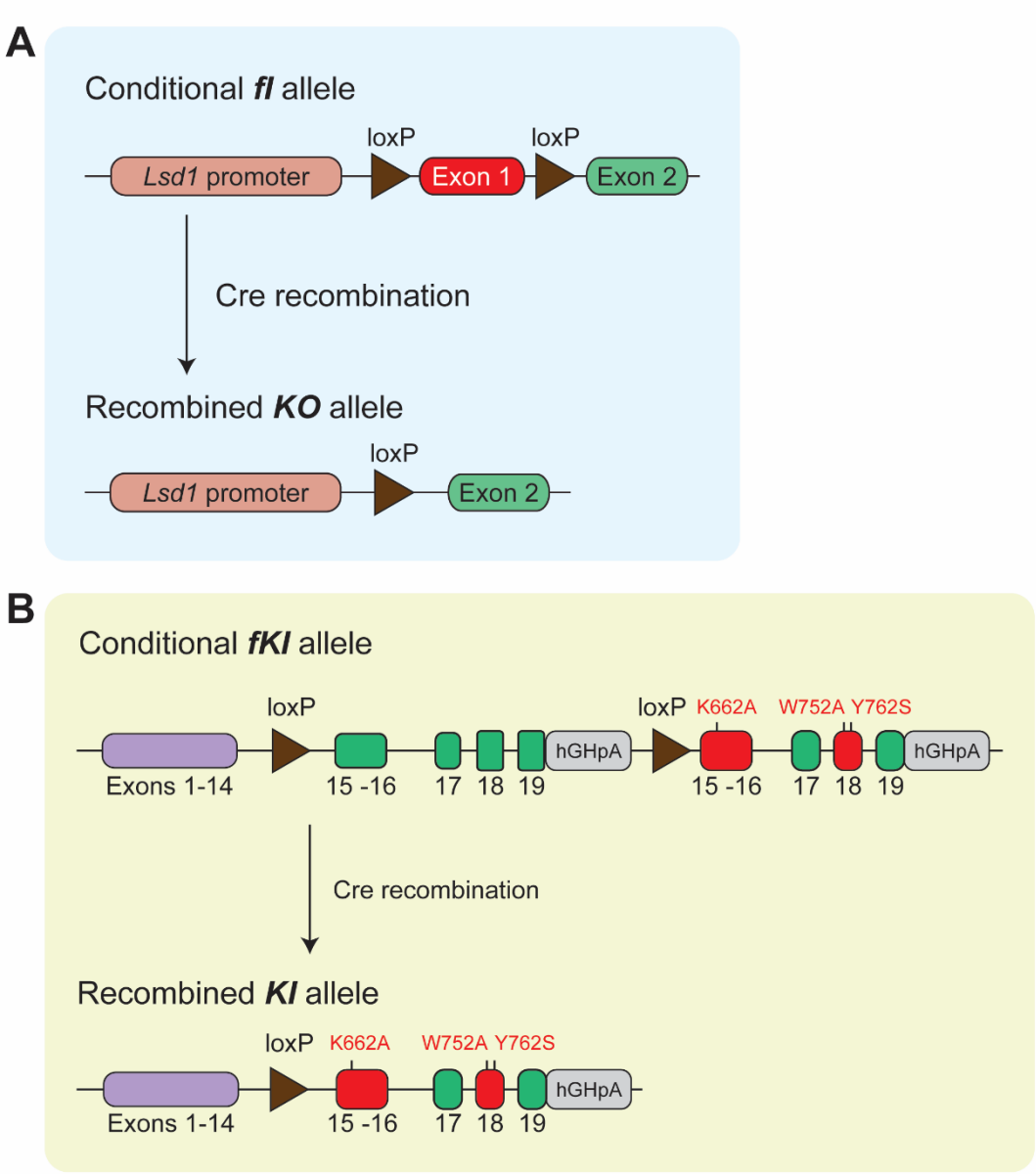


Figure 6.35: **Schematic illustration of generation of *Lsd1* conditional alleles**. (A) Illustration depicting the *Lsd1* conditional knockout [adapted from (Duteil et al., 2017)]. (B) Illustration depicting the *Lsd1* conditional catalytic knock-in. [adapted from (Duteil et al., 2016)]

I crossed the $Lsd1^{fl/fl}$ male mice with *Vasa-Cre* female mice. The *Vasa-Cre* female was obtained from the lab of Yuan Wang in China (Wang et al., 2017). This resulted in all the progeny becoming $Lsd1^{KO/WT}$ genotype, as the Cre protein deposited in the egg of the *Vasa-Cre* female results in global recombination even in progeny that do not inherit *Vasa-Cre* (Gallardo et al., 2007). I crossed $Lsd1^{KO/WT}$; $Vasa^{WT/WT}$ progeny female mice with $Prdm1^{Cre/WT}$ male mice to get progeny with the genotype $Lsd1^{KO/WT}$; $Prdm1^{Cre/WT}$. Next, I crossed this

Lsd1$^{KO/WT}$; *Prdm1*$^{Cre/WT}$ male mice with *Lsd1*$^{fl/fl}$ female mice. After multiple breeding crosses, I could not obtain the pups with the desired knockout genotype *Lsd1*$^{KO/fl}$; *Prdm1*$^{Cre/WT}$ (Figure 6.36A). I observed the *Lsd1*$^{KO/fl}$ genotype in pups that did not carry *Prdm1-Cre* (Pup #1; Figure 6.36B). At the same time, I also saw that the pups with the *Prdm1-Cre* also have all three alleles of *Lsd1* (*KO*, *WT*, and *fl*) (Pup #3; Figure 6.36B). This is possible only if the WT allele came from the father and the *fl* allele is converted to the KO allele in the somatic cells of the progeny. This meant the expression of *Prdm1-Cre* was not limited to the germline but was also seen in other cell types. It has been reported that *Prdm1* is expressed in PGCs and somatic cells (Heffner et al., 2012; Mikedis and Downs, 2017; Mould et al., 2012). In essence, it created a partial knockout of *Lsd1*, which could be lethal, resulting in no knockout progeny.

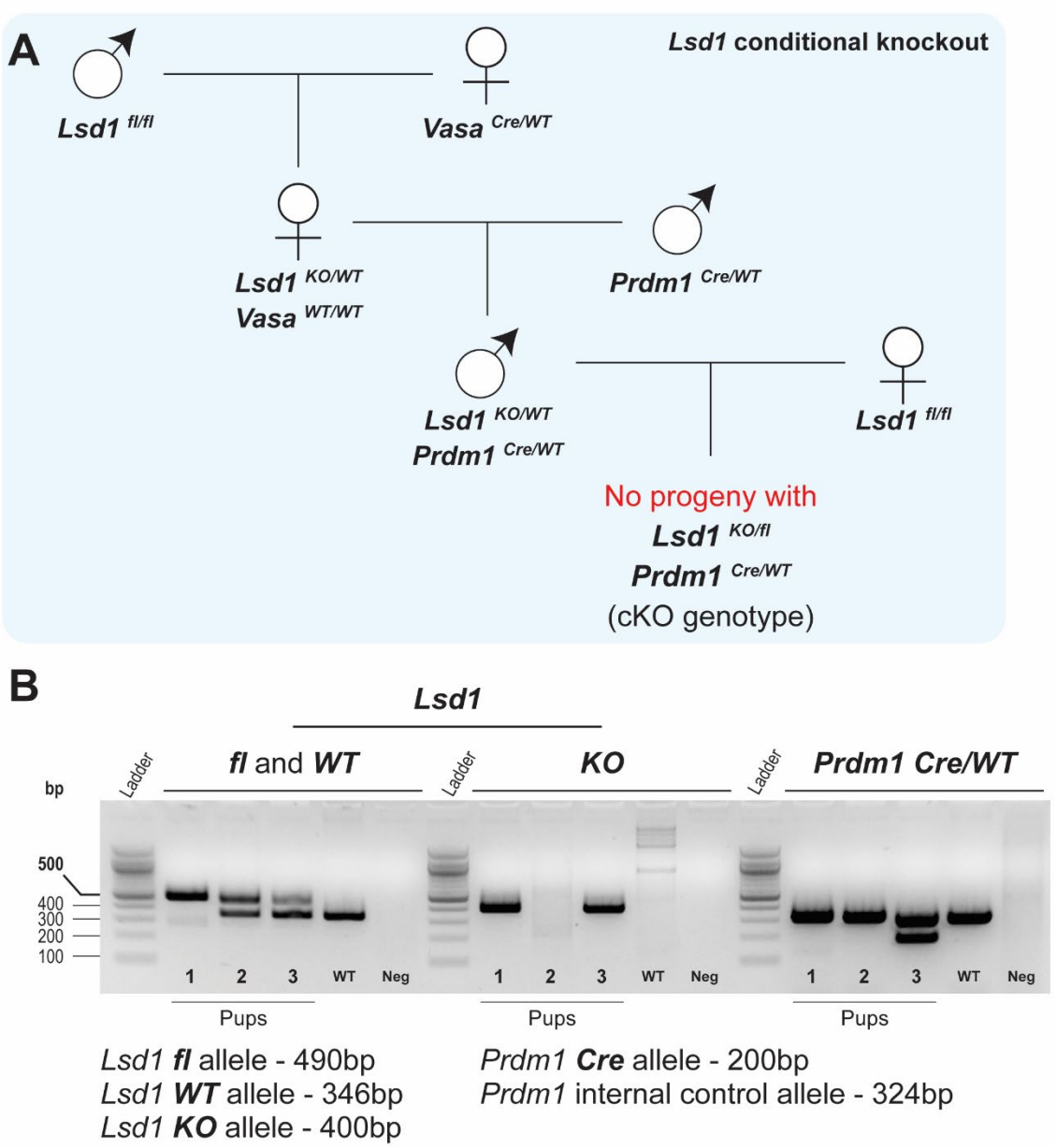


Figure 6.36: ***Lsd1* conditional knockout**. (A) Breeding strategy for obtaining the *Lsd1* conditional knockout using the *Prdm1 Cre*. (B) Genotyping PCR result of the final breeding cross. In the pups that carry *Prdm1-Cre*, we can see Cre-mediated recombination and creation of the *Lsd1 KO* allele (Pup #3), whereas we do not see this in pups that do not carry *Prdm1-Cre* (#1 & #2). Pup #3 is the conditional knockout.

6.5.4 Germline conditional catalytic knock-in of *Lsd1*

To circumvent this lethality, I began crossing the *Lsd1*$^{KO/WT}$; *Prdm1*$^{Cre/WT}$ male mice with *Lsd1*$^{fKI/fKI}$ females to obtain pups whose germline possesses one *Lsd1* allele as complete KO and the other allele being catalytically inactivated. Even from the first breeding cross, progeny with the desired genotype of *Lsd1*$^{KO/fKI}$; *Prdm1*$^{Cre/WT}$ were born in the Mendelian ratio and developed normally (Figure 6.37A). As seen in the cKO breeding, I also observed the conversion of the ***fKI*** allele to the ***KI*** allele in phalange/ear clip tissue in pups carrying *Prdm1-Cre* (Figure 6.37B). This shows that expressing even a half-dosage of a catalytically inactive LSD1 protein could rescue the lethal phenotype, providing more credence to the fact that LSD1 has both catalytic and non-catalytic functions.

I FACS-sorted Spermatogonial stem cells and Spermatogonia from stage P8 of *Lsd1*$^{KO/fKI}$; *Prdm1*$^{Cre/WT}$ (conditional catalytic knock-in; L01), *Lsd1*$^{KO/fKI}$; *Prdm1*$^{WT/WT}$ (L02) and *Lsd1*$^{WT/fKI}$; *Prdm1*$^{Cre/WT}$ (L03) genotypes. The cKI embryo L01 had a very low number of germ cells sorted compared to the controls (Figure 6.38A). This indicates that *Lsd1* cKI succeeded as knocking out *Lsd1* in Spermatogonia leads to a severe reduction of germ cells and male infertility (Lambrot, Lafleur and Kimmins, 2015; Myrick et al., 2017).

Direct bisulfite conversion and Pyrosequencing showed that the DNMT3C DMRs [transposon promoters (LINE1A, LINE1T, and IAP)] are methylated normally in the L01 Spermatogonia (Figure 6.38B). Thus, DNMT3C can methylate its targets even when LSD1 is catalytically inactive. Such a result could be explained in two ways.

(1) Incomplete Cre-mediated recombination has been reported in the literature (Bao et al., 2013; Vooijs, Jonkers and Berns, 2001; jax.org, 2024). So, some Spermatogonial cells can still express the *Lsd1* WT allele even in the conditional knock-in mutant. An incomplete excision could explain the phenotype seen in Figure 6.38B. (2) It is also possible for LSD1 to share a redundant role in the nuclear piRNA pathway with LSD2. Reports show redundant and non-redundant functions for both (Kim and Liu, 2024; Zhang, X. et al., 2021). Both LSD1 and LSD2 were enriched in the IP-MS of DNMT3C and SPOCD1. *Lsd2* is required to establish maternal

genomic imprints. Conditional knockout of *Lsd2* in males does not affect their fertility, but in females, it causes a maternal-effect lethal phenotype (Ciccone et al., 2009). The redundancy between LSD1 and LSD2 could also cause this. Knocking out both from the male germ cells should provide more information.

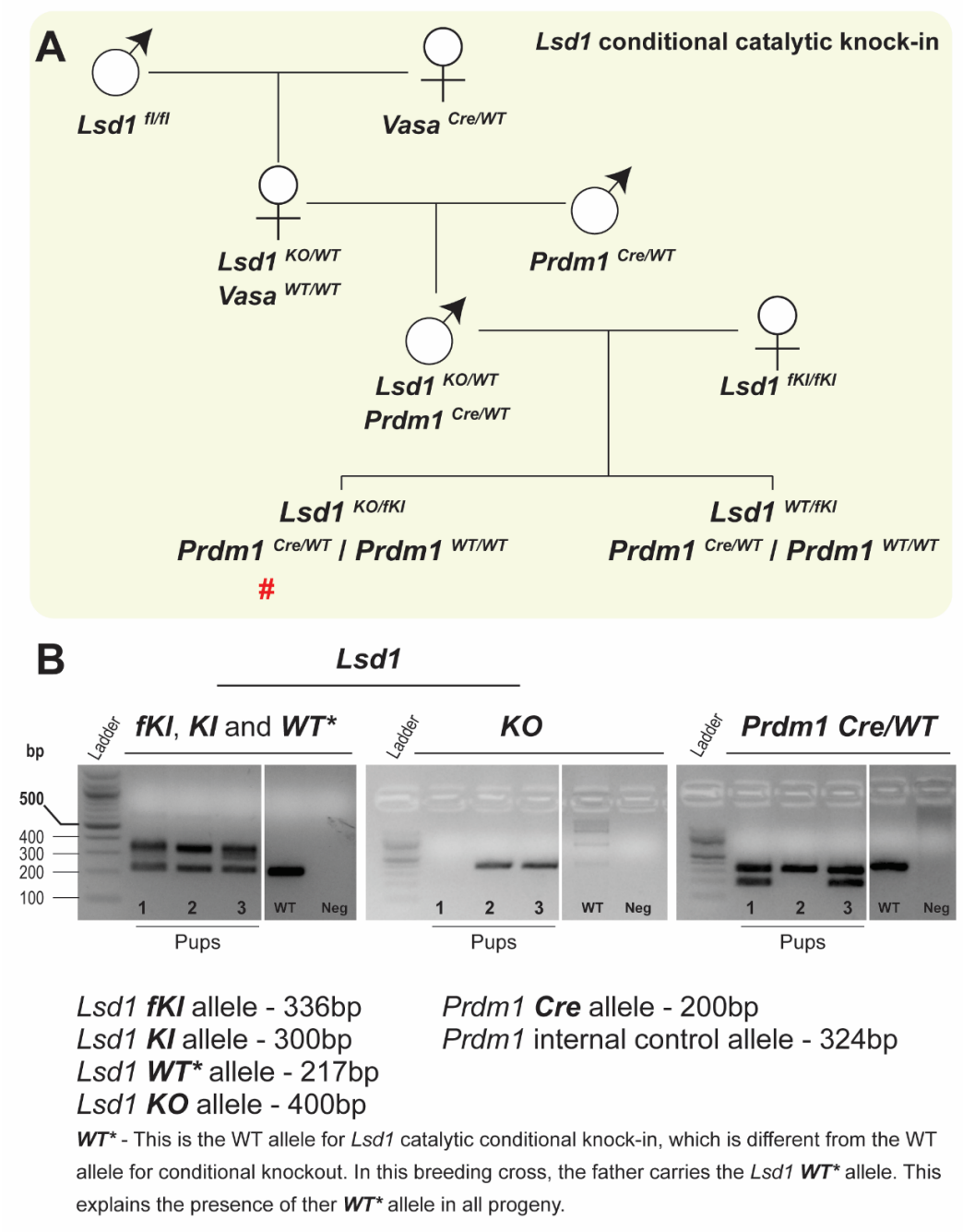


Figure 6.37: ***Lsd1* conditional catalytic knock-in**. (A) Breeding strategy for obtaining the *Lsd1* conditional catalytic knock-in using the *Prdm1 Cre*. # shows the genotype of the conditional catalytic knock-in (B) Genotyping PCR of progeny from the final breeding cross. In the pups that carry *Prdm1-Cre*, we can see Cre-mediated recombination and creation of the *Lsd1 KI* allele (Pup #1 & #3), whereas we do not see this in pups that do not carry *Prdm1-Cre* (#2). Pup #3 is the desired conditional catalytic knock-in.

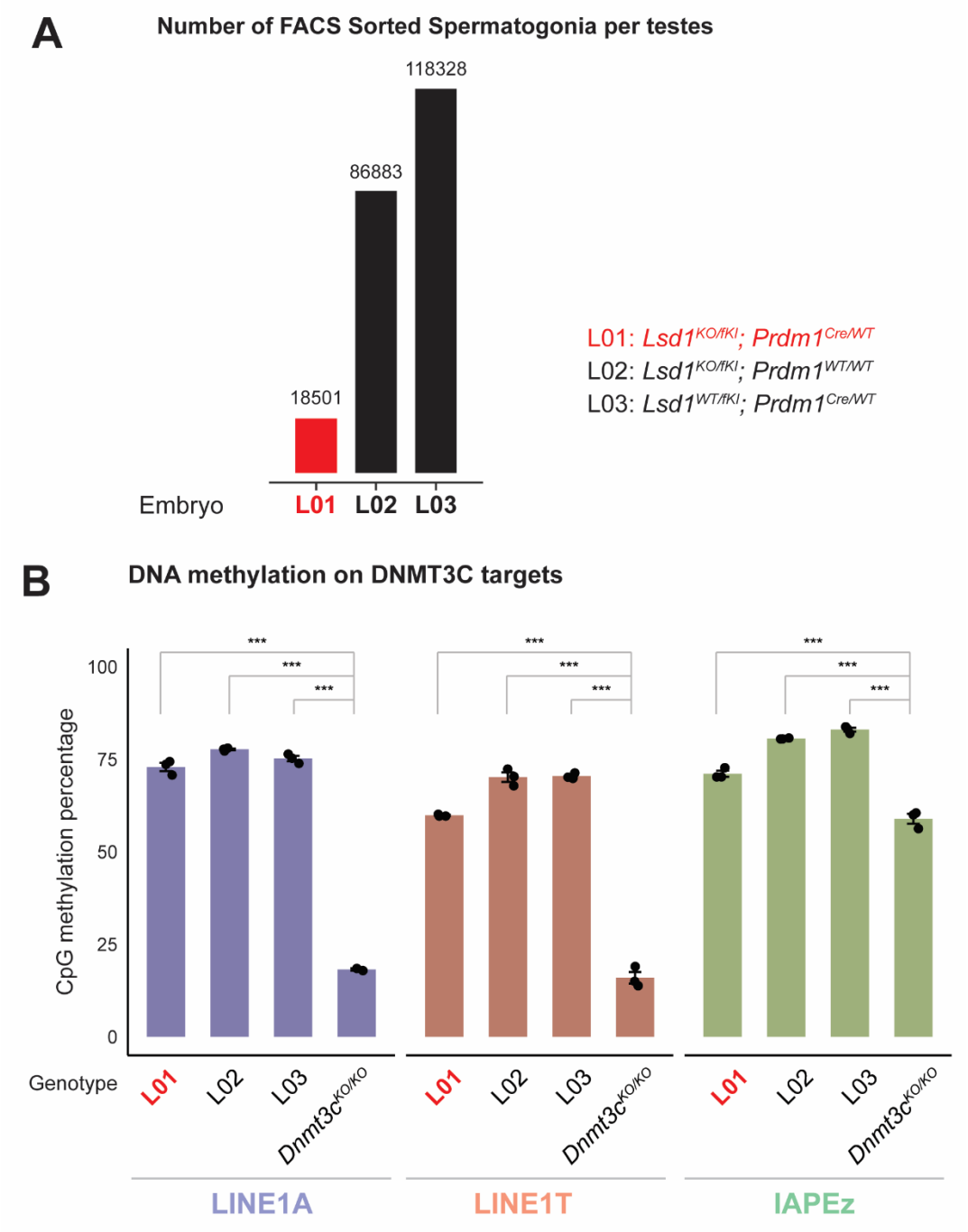


Figure 6.38: ***Lsd1* cKI shows reduced germ cell numbers but normal methylation at DNMT3C DMRs**. (A) The number of FACS-sorted Spermatogonia cells in L01 is significantly lower than in the controls. (B) Graph showing DNA methylation at promoter regions of L1-A, L1-T, and IAPEz elements in sorted testicular germ cells at P15 in different genotypes. *Lsd1* conditional catalytic knock-in (L01) does not affect the DNA methylation at these DNMT3C DMRs. *Dnmt3c*$^{KO/KO}$ is used here as a control. (*** represents t-test p-value < 0.001). Number of technical replicates = 3

6.5.5 H3K4me3 is present at DNMT3C DMRs in *Lsd1* cKI

The significantly reduced germ cell numbers indicate that the Cre-mediated recombination could be fully or near complete. Therefore, I wanted to test for the redundancy between LSD1 and LSD2. While most sorted Spermatogonia were used for the Pyrosequencing experiment, I also performed CUT&TAG for H3K4me3 from a small fraction (~4000 cells/sample for L01 and 10000 cells/sample for L02). Previously, it was shown that *Lsd1* cKO leads to higher enrichment of H3K4me2 at the *Pou5f1* (OCT4) promoter (Myrick et al., 2017). I also saw H3K4me3 heavily retained at the *Pou5f1* promoter in the L01 but not in the L02 embryo (Figure 6.39A), confirming that the L01 embryo indeed lacks a catalytically active LSD1.

Fascinatingly, I saw the same profiling characteristic for H3K4me3 in *Lsd1* cKI as seen in *Dnmt3c*$^{KO/KO}$. At DNMT3C DMRs, H3K4me3 is present in the cKI L01 embryo, but it is absent in the control L02 embryo, even though they are both carrying the WT allele of *Dnmt3c* (Figure 6.39 B and C). This phenomenon was specific for DNMT3C DMRs as we do not see the same over mm10 genes (Figure 6.39D). In fact, a marginal higher enrichment of H3K4me3 can be seen over mm10 genes in L01, consistent with the lack of LSD1's catalytic activity over genic promoters. This is highly intriguing and puzzling because these DNMT3C targets are usually methylated, as in WT (Figure 6.38B). Even though the methylation data comes from a Pyrosequencing experiment and not WGBS, this tells us that DNMT3C could methylate DNA marked by H3K4me3, making non-canonical DNA methylation possible. This result contrasts with the CUT&TAG-BS result, which showed canonical methylation by DNMT3C. Importantly, this H3K4me3 CUT&TAG result is still preliminary data as there was only one biological replicate, but it points toward a novel and complex mechanism for the nuclear piRNA pathway and DNMT3C. Repeating this CUT&TAG with more biological replicates and further experiments like H3K4me3-CUT&TAG-BS or WGBS of cKI Spermatogonia should shed more light on this phenomenon. It is also possible that H3K4me3 is removed at DNMT3C DMRs by another enzyme, LSD2, leading to canonical methylation in the L01 embryo, but in the absence of catalytically active LSD1, the DNMT3C DMRs gain H3K4me3 after birth. H3K4me3

CUT&TAG in cKI Prospermatogonia at stage E18.5 and P1/P2 should be able to test this hypothesis. Additionally, the ability of DNMT3C to methylate H3K4me3-marked DNA should be tested with recombinant modified nucleosomes and/or tethering DNMT3C to H3K4me3-marked reporter in mESCs, which could recapitulate MIWI2-mediated recruitment in the nuclear piRNA pathway.

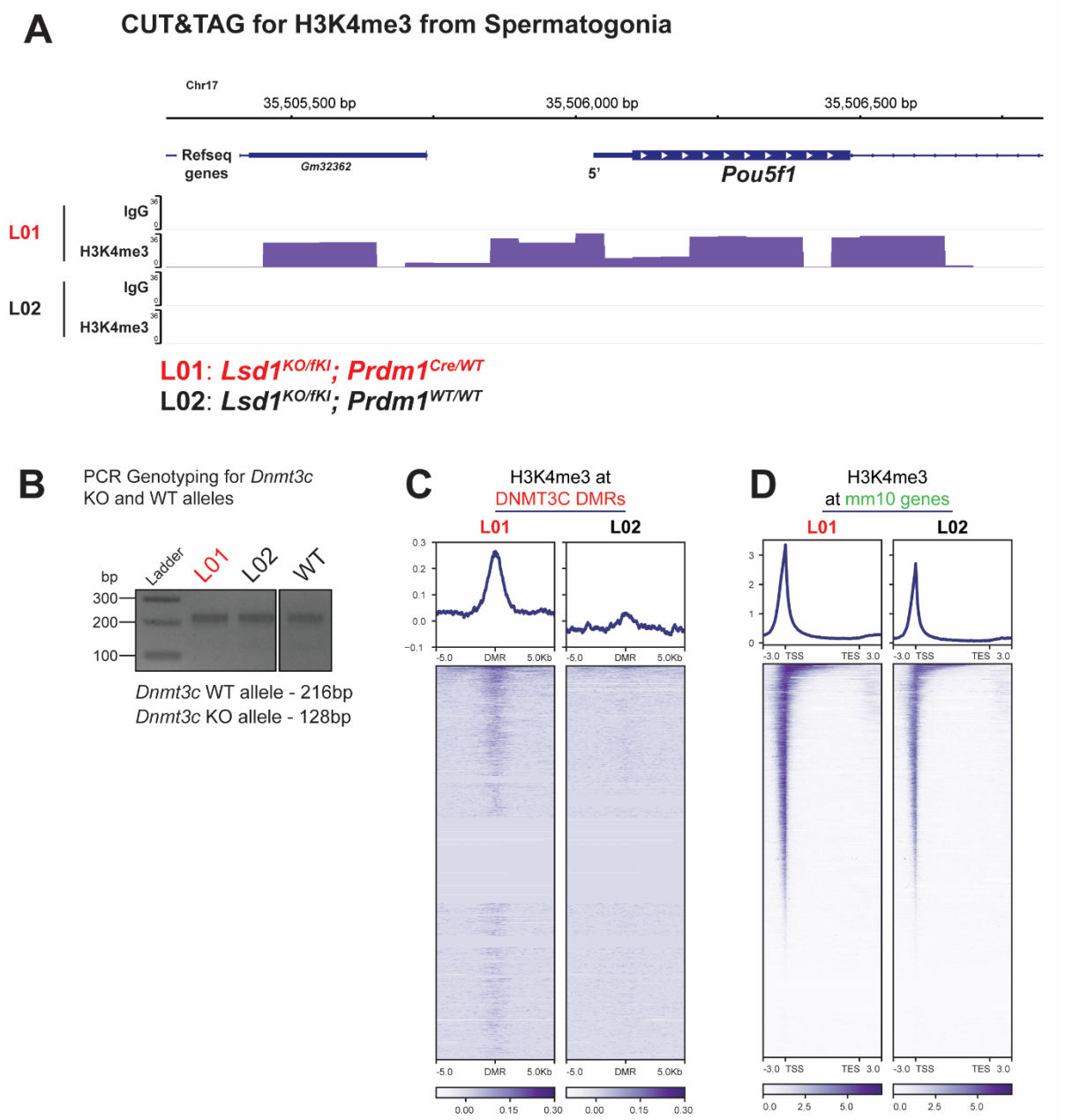


Figure 6.39: **Profiling H3K4me3 in *Lsd1* cKI Spermatogonia**. (A) IGV screenshot showing H3K4me3 occupancy at the *Pou5f1* promoter in L01 and L02 embryos. (B) PCR genotyping for Dnmt3c alleles confirms that L01 is Dnmt3cWT/WT. (C) & (D) Heatmaps showing H3K4me3 occupancy at DNMT3C DMRs (C) and mm10 genes (D) in L01 and L02. H3K4me3 is retained at DNMT3C DMRs in the L01 cKI mutant but not in the control L02.

DISCUSSION

7.1 General Summary

DNMT3C is the most recently discovered *de novo* DNA methyltransferase. When I started my PhD, we only knew it methylates a tiny portion of the mouse genome and is essential for male fertility. My questions were, “How?” and " Why only DNMT3C? "

“*Show me your friends, and I will show you who you are*.” This is a quote attributed to the *Homo sapiens* that walk the Earth. I believe this can also be applied more broadly, for example, to proteins. This is the approach I took to understand more about DNMT3C: Define its interactome and, through that information, appropriately place it in the cellular landscape. DNMT3C is a very intriguing protein. It is a protein that is lowly expressed in a particular cell type during a very specific developmental period and acts on a small proportion of the genome. Nevertheless, a major catastrophe is caused by its absence.

Due to the challenges in studying DNMT3C in its natural habitat, I wanted to try if a more malleable system could be used as a model. As I was starting my PhD in a lab that was also starting up, studying DNMT3C in mESCs was more in focus as the Mouse room was being set up. In mESCs, DNMT3C interacts with many interesting proteins, many of which are also expressed in Prospermatogonia and DNMT3C. Many of these proteins belong to the nucleolar organization and ribosomal biogenesis. However, as the testes DNMT3C IP-MS results came up, the particularity that makes DNMT3C special also made mESC unsuitable for studying it.

DNMT3C’s interactome in Prospermatogonia was another story. There were even more striking proteins in it. Specifically, members of the nuclear piRNA pathway, namely, MIWI2, SPOCD1 (and SPIN1), and TEX15, were significantly enriched in the dataset. My study validated the DNMT3C-MIWI2 and -SPOCD1 interactions, while another report validated the DNMT3C - SPOCD1 and -C19ORF84H interaction. Whole genome DNA methylation analysis showed that DNMT3C, SPOCD1, and MIWI2 have the same genomic targets. This placed DNMT3C in the effector complex of the nuclear piRNA pathway. This answered my first question: “How does DNMT3C methylate its targets?”

Knowing the relationship of DNMT3 proteins with histone modifications and MIWI2 with chromatin remodeling at piRNA targets prompted me to study this for DNMT3C. In doing so, I made a novel discovery that DNMT3C is essential for the removal of H3K4me3 at its targets. The presence of this antagonistic mark explains the lack of DNA methylation we observe in the absence of DNMT3C. This answers my second question: “Why can only DNMT3C do this job and not others?” DNMT3C was found to interact with LSD1, a lysine-specific histone demethylase. LSD1 can demethylate H3K4me2 and could provide the link between DNMT3C and H3K4me demethylation. Profiling for LSD1 binding and *Lsd1* conditional catalytic knock-in experiments showed that LSD1 might be redundant with LSD2. It is unclear if this redundancy exists in reality, and further experiments are needed to provide more insight into this. Interestingly, *Lsd1* catalytic conditional knock-in led to the retainment of H3K4me3 at DNMT3C DMRs, as seen in *Dnmt3c*$^{KO/KO}$, even though DNMT3C DMRs are methylated. This indicates the presence of a non-canonical mechanism of DNA methylation by DNMT3C. It also indicates that there might not be a redundancy between LSD1 and LSD2. If the non-canonical methylation hypothesis holds true, the role of “H3K4me3 removal” could be mainly to ensure proper double-strand breaks and recombination during meiosis and not to repel DNMT3A/3B.

In conclusion, my work has provided insight into the function of DNMT3C in Prospermatogonia and revealed another layer in the biology of piRNA-mediated transcriptional inactivation. This process has led to new questions and, most probably, more questions that we are yet to realize. Working on this project was challenging as it required different Mouse lines, new techniques to be optimized for low-input cells, and, most importantly, the unique intersection of transposon biology, Mouse developmental biology, Epigenetics, and piRNA biology. It turned out to be an exciting study as it constantly demanded me to look at the small and the big picture simultaneously.

7.2 DNMT3C IN MOUSE EMBRYONIC STEM CELLS

DNMT3C is involved in the nuclear piRNA pathway. This pathway is not present in mESCs, making it not an ideal model to study DNMT3C’s functions. However, mESC served

as an excellent model for studying the characteristics of DNMT3C's solubilization, through which we could learn and confirm its affinity to heterochromatin. Reconstituting the nuclear piRNA pathway in mESC by overexpressing its members and using a MIWI2-Zinc Finger fusion could make mESCs an attractive model for studying this pathway.

Interestingly, DNMT1 has been shown to have *de novo* DNA methylation activity in mESCs and preimplantation embryos. More interestingly, this *de novo* methylation has been demonstrated towards transposable elements (Haggerty et al., 2021). They show DNMT1 has *de novo* activity by discovering the presence of DNA methylation in DKO (*Dnmt3a*$^{KO/KO}$; *Dnmt3b*$^{KO/KO}$) mESCs *in vitro* in mESC and *in vivo* in the epiblast of the DKO embryos. Intriguingly, *Dnmt3c* is expressed in the preimplantation embryo at 2C-8C stages (Figure 7.1) (Sakashita et al., 2023). This expression is probably driven by the ERVL-MaLR element at the 5' of the *Dnmt3c gene,* as these elements are known to become transcriptionally active during the zygotic genome activation (Zhang et al., 2019). The role of DNMT3C in preimplantation embryos has not been studied. In *Dnmt*-tKO: *Dnmt3C* mESCs, DNMT3C methylated regions are also enriched for H3K9me3 (Dura, 2021). This could also indicate the recruitment of DNA methylation by this histone modification. DNMT3C could be methylating regions marked by H3K9me3 in preimplantation embryos, which would be maintained by DNMT1. Studying the DNA methylation loss in 3KO (*Dnmt3a*$^{KO/KO}$; *Dnmt3b*$^{KO/KO}$; *Dnmt3c*$^{KO/KO}$) epiblasts or measuring the DNA methylation in the somatic cells from *Dnmt3c*$^{KO/KO}$ mice should shed light on this hypothesis.

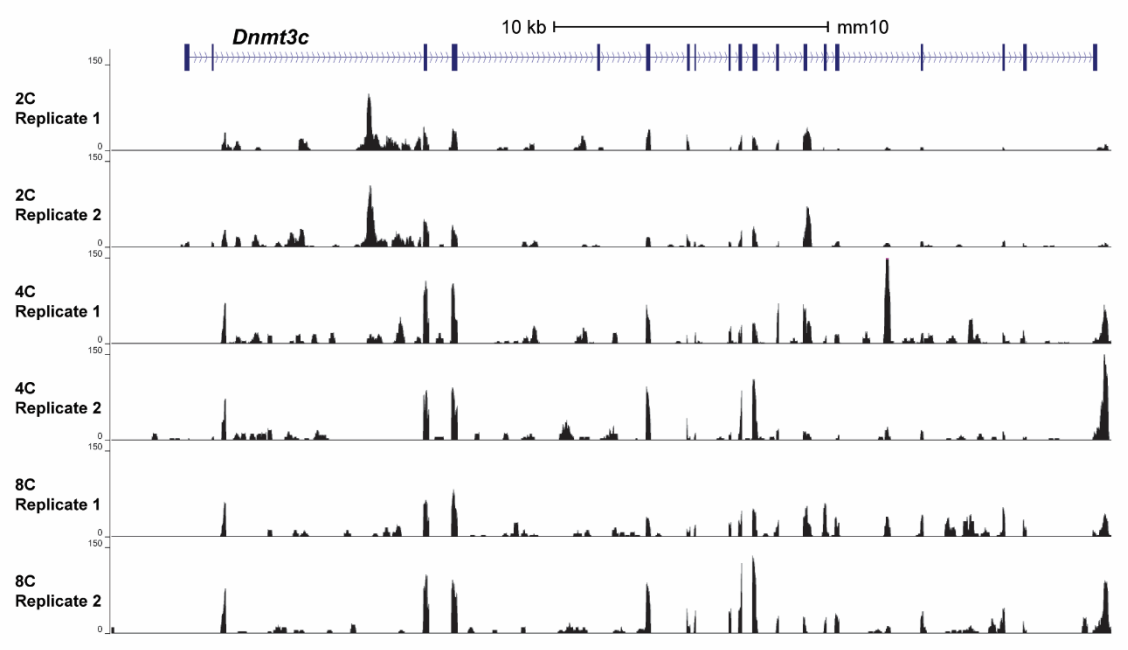


Figure 7.1: **mRNA expression of DNMT3C in preimplantation embryos (2C, 4C and 8C stages).**

A population of cells in mESC culture exists that resembles the 2C embryo. These cells are called 2C-like cells (2CLC). They are a transient cell population with features of 2-cell embryo blastomeres and expanded cell fate potential. Nearly all ES cells cycle in and out of this transient state. Transcriptome sequencing and bioinformatic analyses showed that many 2C-stage transcripts are initiated from long terminal repeats derived from endogenous retroviruses like MERVLs (Genet and Torres-Padilla, 2020; Zhu et al., 2021). This could mean that *Dnmt3c* is also expressed in these cells. This expression could lead to *de novo* DNA methylation, which DNMT1 would subsequently maintain in the DKO mESC. This could explain the small amount of DNA methylation seen in DKO cells and the virtually unmethylated state in TKO (Li et al., 2015). Comparing the regions that possess DNA methylation in DKO mESCs with those that gain DNA methylation in *Dnmt*-tKO: *Dnmt3C* mESCs should help us better study this.

7.3 DNMT3C IN ITS NATURAL HABITAT (PROSPERMATOGONIA)

Testes from 409 embryos were used for the IP-MS experiment. Since fewer WT embryos were collected and collecting more embryos would take more time, effort, and mice, I performed the control IP using magnetic agarose beads. On data analysis, this turned out to be the second-best control. Nevertheless, an unbiased method for filtering gave us a workable number of enriched proteins. Recent studies have shown that piRNA-mediated transcriptional silencing in *Drosophila* involves a condensate formation in nuclei (Eastwood et al., 2021; Schnabl et al., 2021). This could have led to the problematic purification of DNMT3C. Another approach could be to try Proximity Biotinylation in testes, not through creating a transgenic mouse with DNMT3C fusion protein but by antibody recognition (Bar et al., 2018). This approach would mark all the proteins, even within the potential condensate, with Biotin, making it easier to purify them. Accessibility for the antibody would also be a limitation in this approach. GO analysis gave a glimpse into the positioning of DNMT3C in the complex web of pathways inside the cell. One of the most interesting GO classes enriched was the piRNA pathway, and I have explored this connection extensively in the Results section. In short, my work has conclusively provided the biochemical link between DNMT3C and the piRNA pathway.

7.3.1 The Nuclear Localization Conundrum

I touched upon the exciting connection of DNMT3C with the nuclear pore complex. A recent study has shown that the Piwi–piRNA pathway can induce stepwise changes in nuclear architecture and chromatin state at its targets for transcriptional silencing in *Drosophila*. Interestingly, this occurs in a stepwise manner, where localization and removal of active histone marks are followed by repressive histone marks and chromatin conformational change. Piwi-piRNA system may send target regions to specific positions within the nucleus to regulate their expression efficiently (Iwasaki et al., 2021). The identification of nuclear pore complex proteins in DNMT3C IP-MS (and in SPOCD1 IP-MS) could point to such a mechanism occurring in male mouse germ cells as well. By removing the active histone marks (like H3K4me3) and establishing repressive histone marks (like H3K9me3), the nuclear piRNA complex could tether

the target loci to the nuclear periphery through its interaction with the Nuclear Pore complex. Then, the piRNA targets could be heterochromatinized at the nuclear periphery. Interestingly, the nuclear piRNA targets (a.k.a DNMT3C DMRs) are distributed throughout the mouse genome (Figure 7.2). As this pathway takes place over four days at the late gestation period over thousands of loci in thousands of cells, and because of the non-synchronous nature of Prospermatogonia, many of these targets could be tethered to the nuclear periphery at different times. This makes this phenomenon more challenging to investigate. A reporter locus would be the best way to study this. It is also possible for these loci to come together and be collected in foci, which could then be transported to the nuclear periphery instead of individually transporting each locus. This could explain why the nuclear localization of DNMT3C (and SPOCD1) is dispersed at the beginning of the nuclear piRNA pathway and forms foci as it progresses.

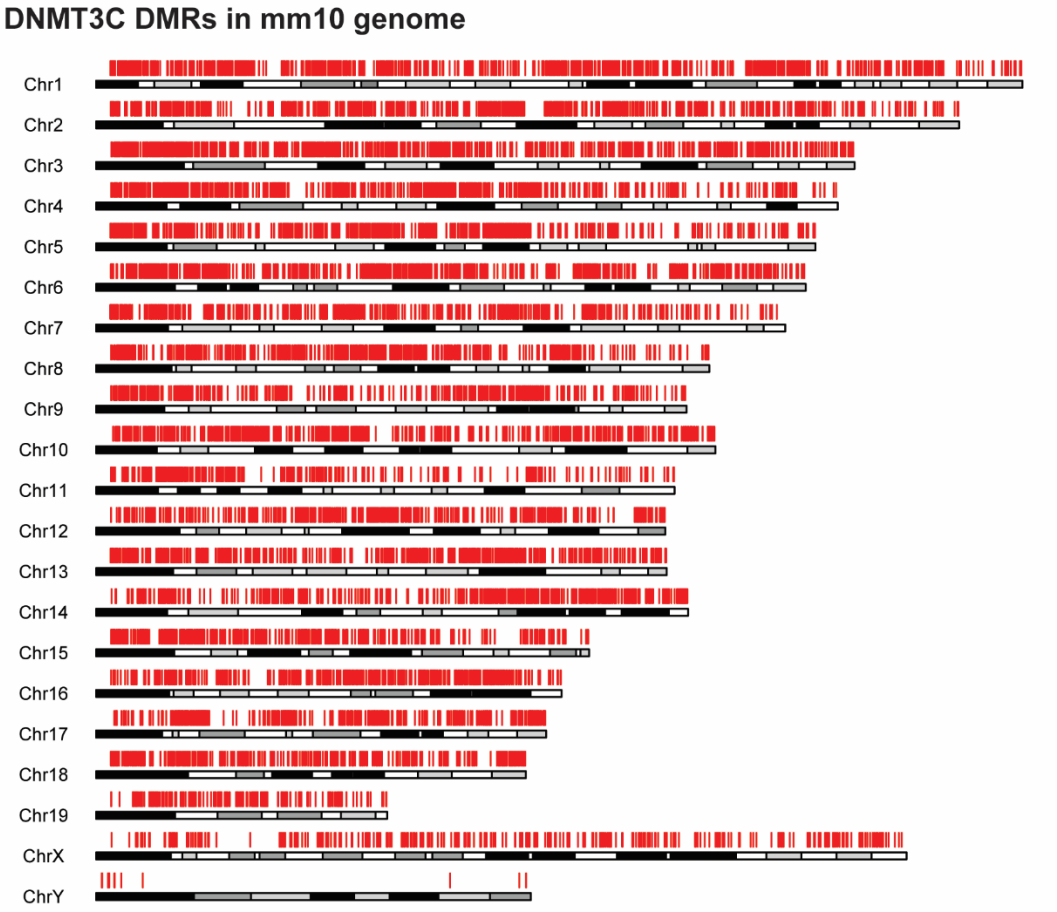


Figure 7.2: **Distribution of DNMT3C DMRs in mm10 genome**

7.3.2 The case of the N-terminal positive selection

As the major difference between the different DNMT3 proteins, the N-terminal is hypothesized to play an essential role in DNMT3C's function. The N terminal of DNMT3C is fast evolving under positive selection (Molaro, Malik and Bourc'his, 2020). In particular, four amino acids rapidly evolve among rodent species (Figure 7.3A). Engineering point mutations at these AAs and introgression of the mutated *Dnmt3c* in *Dnmt3c*$^{KO/KO}$ mice would give us insight into these amino acids' roles. Interestingly, like in DNMT3C, two amino acids located in the N-terminus domain of DNMT3A have been shown to evolve under positive selection in primates, while it was not the case for DNMT3B. It would also be fascinating to see if expressing a chimeric version of mouse DNMT3C carrying the N terminal of Human DNMT3A could rescue *Dnmt3c*$^{KO/KO}$.

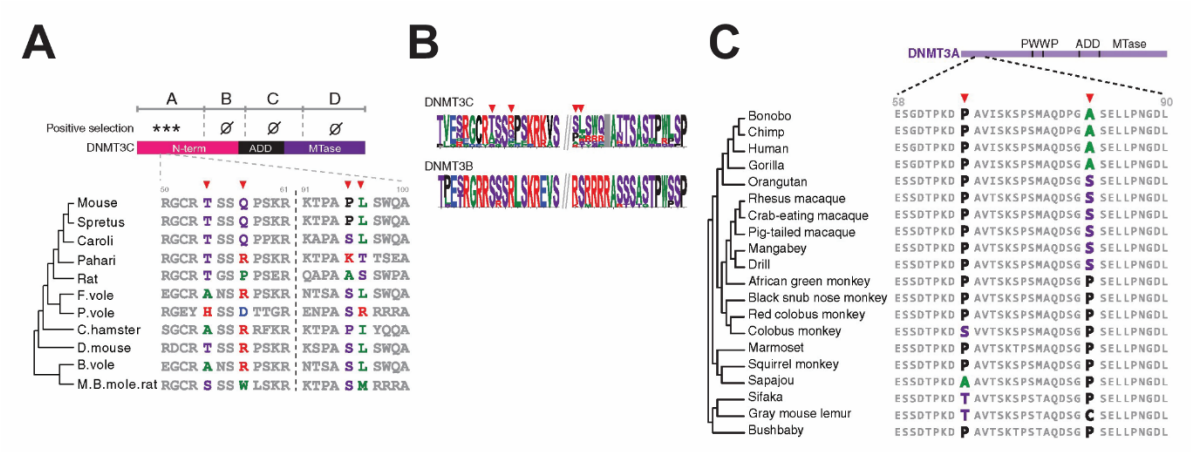


Figure 7.3: **Comparative genomics of the N-terminus of DNMT3s in mammals**. (A) and (B) Positive selection of the DNMT3C N-terminus in rodents, compared to DNMT3B. (C) Positive selection of the DNMT3A N-terminus in primates

7.3.3 One complex to rule them all

MIWI2 translocates to the nucleus after being loaded with antisense piRNAs. It engages the nuclear piRNA pathway effector complex for co-transcriptional silencing of transposons. The mechanics of this complex assembly are still unknown. There are two possibilities. MIWI2 marks the loci of nascent transposon transcription and assembles the complex on site. This is the most probable as there are thousands of loci distributed throughout the genome. Another possibility is that certain members of the complex are already at the site of piRNA targets. This possibility becomes more entertaining because of the participation of proteins like SPIN1 and SPOCD1. SPIN1 is expressed before MIWI2 becomes nuclear (Figure 7.4A). SPIN1 is a chromatin reader which reads the bivalent mark H3K4me3 H3K9me3 (Du et al., 2021). This bivalent mark is present at DNMT3C DMRs (MIWI2 targets) as they are demethylated (at E13.5) and before the piRNA pathway is active (Figure 7.4B) (Yamanaka et al., 2019). Therefore, SPIN1 could be present at DNMT3C DMRs by binding to this specific bivalent histone mark. SPOCD1 is a strong interactor with SPIN1, so the latter could tether the former to these DNMT3C DMRs. MIWI2-mediated recognition could commit the complex to co-transcriptional silencing.

If the first possibility holds, the effector complex may never form without one of its members. If the second possibility holds, parts of the effector complex could assemble and carry out their function even in a mutant like *Dnmt3c*$^{KO/KO}$. All the steps until DNA methylation (halting the transcriptional process, removal of active histone marks H3K4me, and histone acetylation) could occur without DNMT3C. However, without DNA methylation, all of this could be reversed, allowing the H3K4me3 mark to be re-established. In the absence of DNA methylation, an epigenetic switch, as seen in mESC (Walter et al., 2016), could take over, leading to transcriptional repression of transposons. Assembly of even the partial effector complex could explain why LSD1 can bind DNMT3C DMRs in *Dnmt3c*$^{KO/KO}$.

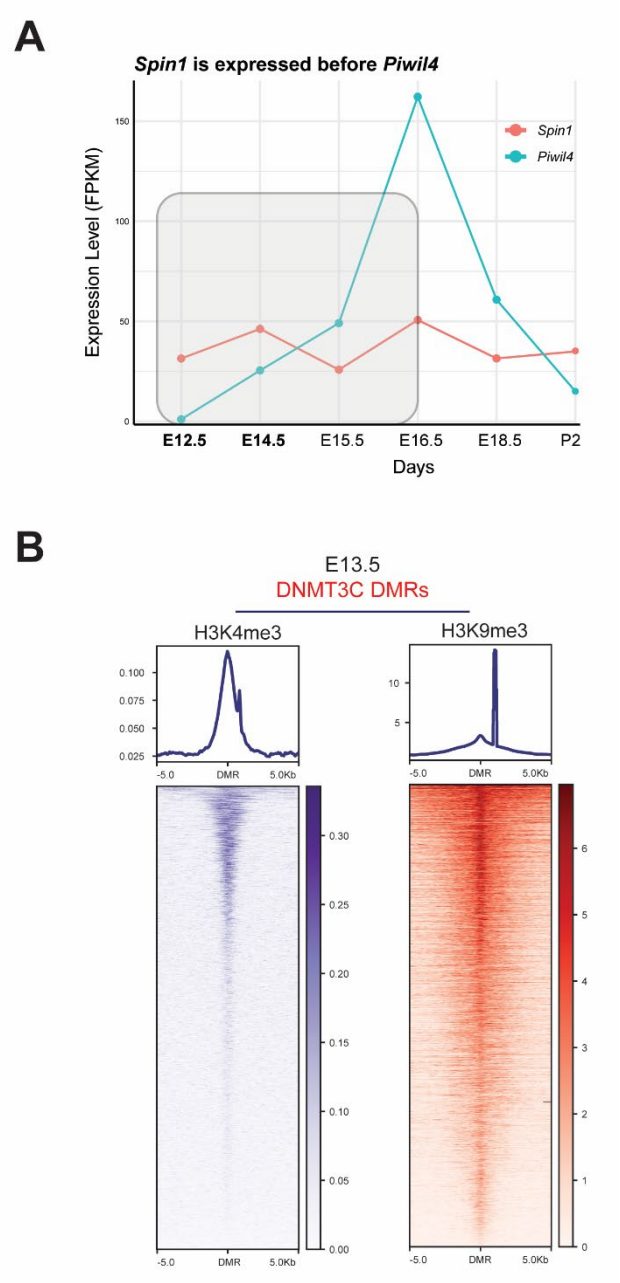


Figure 7.4: **Partial piRNA complex could be formed before MIWI2 becomes nuclear**. (A) mRNA expression in Prospermatogonia shows that *Spin1* expression starts before *Piwil4*. (B) Many of the DNMT3C DMRs are marked by the bivalent H3K4me3 H3K9me3 at stage E13.5, which could recruit SPIN1 and by extension, SPOCD1 [Data from (Yamanaka et al., 2019)].

7.3.4 Hypothetical Model of Nuclear piRNA Pathway

This pathway has proven to be very complex, with many intricacies. With the data I generated and the literature search, I propose the following model for the nuclear piRNA pathway (Figure 7.5). MILI and MIWI2 generate primary and secondary piRNAs in the cytoplasm. Anti-sense piRNA-loaded MIWI2 translocates to the nucleus. It recognizes the sites of nascent transposon transcription by piRNA-mediated complementarity. It then assembles the effector complex *de novo* or licenses the activity of the already present members. Either way, the complex assembles particularly on the transposon promoter, possibly due to the binding of SPIN1 to H3K4me3 H3K9me3, a bivalent mark present at the young transposon promoters. SPOCD1 could interface with the RNA polymerase II, halting active transcription. DNMT3C and DNMT3L are recruited to this location through its interaction with MIWI2 and SPOCD1. LSD1 interacts with DNMT3C and SPOCD1 and thus removes the H3K4me3 in conjunction with JARID proteins. DNMT3C-DNMT3L binds to the histone tail and methylates the transposon promoter DNA. This methylation could still be non-canonical, where DNMT3C binds to H3K4me3 marked regions and methylates the DNA. Such a non-canonical mechanism could have evolved to deal with the specificity of transposon promoter methylation.

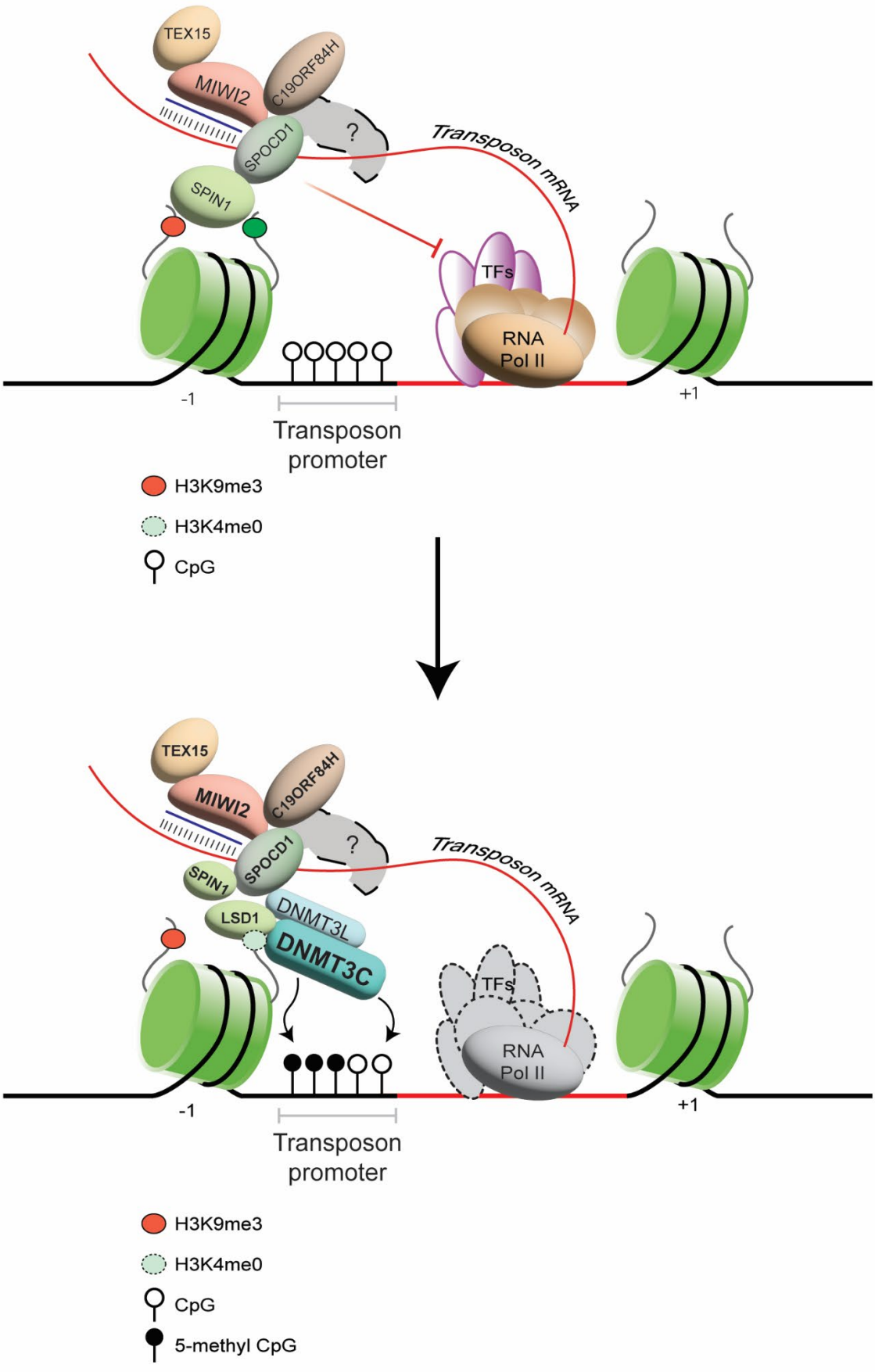


Figure 7.5: **Hypothetical model of co-transcriptional silencing by the nuclear piRNA pathway effector complex.**

REFERENCES

https://www.jax.org/news-and-insights/jax-blog/2013/september/a-dozen-facts-you-didnt-know-about-cre-lox (2024), 23 August (Accessed: 23 August 2024).

Adams, I.R. and McLaren, A. (2002) 'Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis', *Development*, 129(5), pp. 1155–1164.

Adhikari, S. and Curtis, P.D. (2016) 'DNA methyltransferases and epigenetic regulation in bacteria', *FEMS Microbiology Reviews*, 40(5), pp. 575–591. doi: 10.1093/femsre/fuw023

Ancelin, K. *et al.* (2016) 'Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation', *Elife*, 5. doi: 10.7554/eLife.08851

Anderson, R. *et al.* (2000) 'The onset of germ cell migration in the mouse embryo', *Mechanisms of Development*, 91(1-2), pp. 61–68. doi: 10.1016/s0925-4773(99)00271-3

Andrews, S. *et al.* (2023) 'Mechanisms and function of de novo DNA methylation in placental development reveals an essential role for DNMT3B', *Nature Communications*, 14(1), p. 371. doi: 10.1038/s41467-023-36019-9

Appel, L.-M. *et al.* (2021) 'PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC', *Nature Communications*, 12(1), p. 6078. doi: 10.1038/s41467-021-26360-2

Appel, L.-M. *et al.* (2023) 'SPOC domain proteins in health and disease', *Genes Dev*, 37(5-6), pp. 140–170. doi: 10.1101/gad.350314.122

Appel, L.-M. *et al.* (2023) 'The SPOC domain is a phosphoserine binding module that bridges transcription machinery with co- and post-transcriptional regulators', *Nature Communications*, 14(1), p. 166. doi: 10.1038/s41467-023-35853-1

Ara, T. *et al.* (2003) 'Impaired colonization of the gonads by primordial germ cells in mice lacking a chemokine, stromal cell-derived factor-1 (SDF-1)', *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), pp. 5319–5323. doi: 10.1073/pnas.0730719100

Aravin, A. *et al.* (2006) 'A novel class of small RNAs bind to MILI protein in mouse testes', *Nature*, 442(7099), pp. 203–207.

Aravin, A.A. *et al.* (2008) 'A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice', *Mol. Cell*, 31(6), pp. 785–799.

Aravin, A.A. *et al.* (2009) 'Cytoplasmic Compartmentalization of the Fetal piRNA Pathway in Mice', *PLoS Genet*, 5(12).

Ballow, D. *et al.* (2006) 'Sohlh1 is essential for spermatogonial differentiation', *Developmental Biology*, 294(1), pp. 161–167. doi: 10.1016/j.ydbio.2006.02.027

Bao, J. *et al.* (2013) 'Incomplete cre-mediated excision leads to phenotypic differences between Stra8-iCre; Mov10l1(lox/lox) and Stra8-iCre; Mov10l1(lox/Δ) mice', *Genesis*, 51(7), pp. 481–490. doi: 10.1002/dvg.22389

Bar, D.Z. *et al.* (2018) 'Biotinylation by antibody recognition-a method for proximity labeling', *Nature Methods*, 15(2), pp. 127–133. doi: 10.1038/nmeth.4533

Barau, J. *et al.* (2016) 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity', *Science*, 354(6314), pp. 909–912.

Bartosovic, M., Kabbe, M. and Castelo-Branco, G. (2021) 'Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues', *Nature Biotechnology*, 39(7), pp. 825–835. doi: 10.1038/s41587-021-00869-9

Batista, P.J. *et al.* (2008) 'PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans', *Molecular Cell*, 31(1), pp. 67–78. doi: 10.1016/j.molcel.2008.06.002

Batzer, M.A. and Deininger, P.L. (2002) 'Alu repeats and human genomic diversity', *Nature Reviews Genetics*, 3(5), pp. 370–379. doi: 10.1038/nrg798

Baubec, T. *et al.* (2015) 'Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation', *Nature*, 520(7546), pp. 243–247. doi: 10.1038/nature14176

Baubec, T. and Schübeler, D. (2014) 'Genomic patterns and context specific interpretation of DNA methylation', *Current Opinion in Genetics & Development*, 25, pp. 85–92. doi: 10.1016/j.gde.2013.11.015

Bekker-Jensen, D.B. *et al.* (2020) 'A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients', *Mol. Cell. Proteomics*, 19(4), pp. 716–729.

Bera, T.K. *et al.* (2001) 'Cse1l is essential for early embryonic growth and development', *Mol. Cell. Biol.*, 21(20), pp. 7020–7024.

Bestor, T. *et al.* (1988) 'Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial

restriction methyltransferases', *Journal of Molecular Biology*, 203(4), pp. 971–983. doi: 10.1016/0022-2836(88)90122-2

Bestor, T.H. (1992) 'Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain', *EMBO J*, 11(7), pp. 2611–2617. doi: 10.1002/j.1460-2075.1992.tb05326.x

Bestor, T.H. and Ingram, V.M. (1983) 'Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 80(18), pp. 5559–5563. doi: 10.1073/pnas.80.18.5559

Bird, A. (2002) 'DNA methylation patterns and epigenetic memory', *Genes Dev*, 16(1), pp. 6–21. doi: 10.1101/gad.947102

Bird, A.P. (1986) 'CpG-rich islands and the function of DNA methylation', *Nature*, 321(6067), pp. 209–213. doi: 10.1038/321209a0

Bloom, J.C. and Schimenti, J.C. (2020) 'Sexually dimorphic DNA damage responses and mutation avoidance in the mouse germline', *Genes & Development*, 34(23-24), pp. 1637–1649. doi: 10.1101/gad.341602.120

Borgel, J. *et al.* (2010) 'Targets and dynamics of promoter DNA methylation during early mouse development', *Nat. Genet.*, 42(12), pp. 1093–1100.

Bostick, M. *et al.* (2007) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science (New York, N.Y.)*, 317(5845), pp. 1760–1764. doi: 10.1126/science.1147939

Bouras, E. *et al.* (2019) 'Gene promoter methylation and cancer: An umbrella review', *Gene*, 710, pp. 333–340. doi: 10.1016/j.gene.2019.06.023

Bourc'his, D. *et al.* (2001) 'Dnmt3L and the establishment of maternal genomic imprints', *Science*, 294(5551), pp. 2536–2539.

Bourc'his, D. and Bestor, T.H. (2004) 'Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L', *Nature*, 431(September), pp. 2–5.

Bourque, G. *et al.* (2008) 'Evolution of the mammalian transcription factor binding repertoire via transposable elements', *Genome Res*, 18(11), pp. 1752–1762. doi: 10.1101/gr.080663.108

Bourque, G. *et al.* (2018) 'Ten things you should know about transposable elements', *Genome Biol*, 19(1), p. 199.

Boyko, K. *et al.* (2022) 'Structure of the DNMT3B ADD domain suggests the absence of a DNMT3A-like autoinhibitory mechanism', *Biochem. Biophys. Res. Commun.*, 619, pp. 124–129.

Boyko, K. M. *et al.* (2022) *Crystal structure of ADD domain of the human DNMT3B methyltransferase*.

Braydich-Stolle, L. *et al.* (2005) 'Role of glial cell line-derived neurotrophic factor in germ-line stem cell fate', *Annals of the New York Academy of Sciences*, 1061, pp. 94–99. doi: 10.1196/annals.1336.010

Brennecke, J. *et al.* (2007) 'Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila', *Cell*, 128(6), pp. 1089–1103. doi: 10.1016/j.cell.2007.01.043

Bronkhorst, A.W. and Ketting, R.F. (2018) 'Trimming it short: PNLDC1 is required for piRNA maturation during mouse spermatogenesis', *EMBO Reports*, 19(3). doi: 10.15252/embr.201845824

Bulut-Karslioglu, A. *et al.* (2014) 'Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells', *Molecular Cell*, 55(2), pp. 277–290. doi: 10.1016/j.molcel.2014.05.029

Cao, K. *et al.* (2018) 'An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells', *Sci. Adv.*, 4(1), eaap8747.

Carmell, M.A. *et al.* (2007) 'MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline', *Dev. Cell*, 12(4), pp. 503–514.

Carter, A.C. *et al.* (2020) 'Spen links rna-mediated endogenous retrovirus silencing and x chromosome inactivation', *Elife*, 9, pp. 1–58.

Chakrabarti, P. and Chakravarty, D. (2022) 'Intrinsically disordered proteins/regions and insight into their biomolecular interactions', *Biophysical Chemistry*, 283, p. 106769. doi: 10.1016/j.bpc.2022.106769

Chedin, F., Lieber, M.R. and Hsieh, C.-L. (2002) 'The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a', *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), pp. 16916–16921. doi: 10.1073/pnas.262443999

Chen, T. *et al.* (2002) 'A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation', *Journal of Biological Chemistry*, 277(41), pp. 38746–38754. doi: 10.1074/jbc.M205312200

Cheng, X. and Blumenthal, R.M. (2008) 'Mammalian DNA methyltransferases: a structural perspective', *Structure (London, England : 1993)*, 16(3), pp. 341–350. doi: 10.1016/j.str.2008.01.004

Christensen, J. *et al.* (2007) 'RBP2 Belongs to a Family of Demethylases, Specific for Tri-and Dimethylated Lysine 4 on Histone 3', *Cell*, 128(6), pp. 1063–1076.

Chuma, S. *et al.* (2006) 'Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice', *Proceedings of the National Academy of Sciences of the United States of America*, 103(43), pp. 15894–15899. doi: 10.1073/pnas.0601878103

Chuma, S. and Nakano, T. (2013) 'piRNA and spermatogenesis in mice', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1609), p. 20110338. doi: 10.1098/rstb.2011.0338

Ciccarone, F. *et al.* (2012) 'Poly(ADP-ribosyl)ation acts in the DNA demethylation of mouse primordial germ cells also with DNA damage-independent roles', *PLoS One*, 7(10), e46927. doi: 10.1371/journal.pone.0046927

Ciccone, D.N. *et al.* (2009) 'KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints', *Nature*, 461(7262), pp. 415–418.

CLERMONT, Y. (1972) 'Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal', *Physiological Reviews*, 52(1), pp. 198–236. doi: 10.1152/physrev.1972.52.1.198

Cobrinik, D. (2005) 'Pocket proteins and cell cycle control', *Oncogene*, 24(17), pp. 2796–2809. doi: 10.1038/sj.onc.1208619

Cooper, G.M. (2000) *The Nucleolus*: Sinauer Associates.

Correll, C.C., Bartek, J. and Dundr, M. (2019) 'The Nucleolus: A Multiphase Condensate Balancing Ribosome Synthesis and Translational Capacity in Health, Aging and Ribosomopathies', *Cells*, 8(8).

Cost, G.J. *et al.* (2002) 'Human L1 element target-primed reverse transcription in vitro', *EMBO J*, 21(21), pp. 5899–5910. doi: 10.1093/emboj/cdf592

Cox, J. *et al.* (2011) 'Andromeda: a peptide search engine integrated into the MaxQuant environment', *J. Proteome Res.*, 10(4), pp. 1794–1805.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nat. Biotechnol.*, 26(12), pp. 1367–1372.

Czech, B. *et al.* (2013) 'A transcriptome-wide RNAi screen in the drosophila ovary reveals factors of the germline piRNA pathway', *Mol. Cell*, 50(5), pp. 749–761.

Czech, B. *et al.* (2018) 'piRNA-Guided Genome Defense: From Biogenesis to Silencing', *Annu. Rev. Genet.*, 52(1), pp. 131–157.

Dadoune, J.-P. (2003) 'Expression of mammalian spermatozoal nucleoproteins', *Microscopy Research and Technique*, 61(1), pp. 56–75. doi: 10.1002/jemt.10317

Danecek, P. *et al.* (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2). doi: 10.1093/gigascience/giab008

Davis, K.N. *et al.* (2023) 'Mutations in human DNA methyltransferase DNMT1 induce specific genome-wide epigenomic and transcriptomic changes in neurodevelopment', *Human Molecular Genetics*, 32(21), pp. 3105–3120. doi: 10.1093/hmg/ddad123

DeBerardinis, R.J. *et al.* (1998) 'Rapid amplification of a retrotransposon subfamily is evolving the mouse genome', *Nature Genetics*, 20(3), pp. 288–290. doi: 10.1038/3104

Deininger, P.L. *et al.* (2003) 'Mobile elements and mammalian genome evolution', *Current Opinion in Genetics & Development*, 13(6), pp. 651–658. doi: 10.1016/j.gde.2003.10.013

Des Marais, D.L. and Rausher, M.D. (2008) 'Escape from adaptive conflict after duplication in an anthocyanin pathway gene', *Nature*, 454(7205), pp. 762–765. doi: 10.1038/nature07092

Dhayalan, A. *et al.* (2010) 'The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation', *The Journal of Biological Chemistry*, 285(34), pp. 26114–26120. doi: 10.1074/jbc.M109.089433

Ding, D. *et al.* (2017) 'PNLDC1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice', *Nature Communications*, 8(1), p. 819. doi: 10.1038/s41467-017-00854-4

Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635

Dong, Q. *et al.* (2018) 'Roles of the CSE1L-mediated nuclear import pathway in epigenetic silencing', *Proc. Natl. Acad. Sci. U. S. A.*, 115(17), E4013-E4022.

Dossin, F.c. *et al.* (2020) 'SPEN integrates transcriptional and epigenetic control of X-inactivation', *Nature*, 578(7795), pp. 455–460.

Dossmann, L. *et al.* (2024) 'Specific DNMT3C flanking sequence preferences facilitate methylation of young murine retrotransposons', *Communications Biology*, 7(1), p. 582. doi: 10.1038/s42003-024-06252-z

Du, Y. *et al.* (2021) 'Structural mechanism of bivalent histone H3K4me3K9me3 recognition by the Spindlin1/C11orf84 complex in rRNA transcription activation', *Nature Communications*, 12. doi: 10.1038/s41467-021-21236-x

Duncan, C.G. *et al.* (2018) 'Dosage compensation and DNA methylation landscape of the X chromosome in mouse liver', *Scientific Reports*, 8(1), p. 10138. doi: 10.1038/s41598-018-28356-3

Dupressoir, A., Lavialle, C. and Heidmann, T. (2012) 'From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation', *Placenta*, 33(9), pp. 663–671. doi: 10.1016/j.placenta.2012.05.005

Dura, M. (2021) *Critical and different roles of DNA methylation in male germ cell development.* Sorbonne université.

Dura, M. *et al.* (2022) 'DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis', *Nat. Genet.*, 54(4), pp. 469–480.

Duteil, D. *et al.* (2016) 'Lsd1 Ablation Triggers Metabolic Reprogramming of Brown Adipose Tissue', *Cell Reports*, 17(4), pp. 1008–1021. doi: 10.1016/j.celrep.2016.09.053

Duteil, D. *et al.* (2017) 'Lsd1 prevents age-programed loss of beige adipocytes', *Proceedings of the National Academy of Sciences of the United States of America*, 114(20), pp. 5265–5270. doi: 10.1073/pnas.1702641114

Duymich, C.E. *et al.* (2016) 'DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells', *Nature Communications*, 7, p. 11453. doi: 10.1038/ncomms11453

Eastwood, E.L. *et al.* (2021) 'Dimerisation of the PICTS complex via LC8/Cut-up drives co-transcriptional transposon silencing in Drosophila', *Elife*, 10. doi: 10.7554/eLife.65557

Ehrlich, M. *et al.* (2001) 'DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes', *Human Molecular Genetics*, 10(25), pp. 2917–2931. doi: 10.1093/hmg/10.25.2917

Elgin, S.C.R. and Reuter, G. (2013) 'Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila', *Cold Spring Harbor Perspectives in Biology*, 5(8), a017780. doi: 10.1101/cshperspect.a017780

Elisaphenko, E.A. *et al.* (2008) 'A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements', *PLoS One*, 3(6), pp. 1–11.

Emeline Roger (2023) *Novel chromatin pathway controlling transposable elements and impact on genome stability.* Université Paris sciences et lettres. Available at: https://in2p3.hal.science /FNCLCC/tel-04315843v1.

Epsztejn-Litman, S. *et al.* (2008) 'De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes', *Nature Structural & Molecular Biology*, 15(11), pp. 1176–1183. doi: 10.1038/nsmb.1476

Esteller, M. (2008) 'Epigenetics in cancer', *The New England Journal of Medicine*, 358(11), pp. 1148–1159. doi: 10.1056/NEJMra072067

Estève, P.-O. *et al.* (2006) 'Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication', *Genes Dev*, 20(22), pp. 3089–3103. doi: 10.1101/gad.1463706

Estève, P.-O. *et al.* (2009) 'Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells', *Proceedings of the National Academy of Sciences of the United States of America*, 106(13), pp. 5076–5081. doi: 10.1073/pnas.0810362106

Evan, G.I. *et al.* (1985) 'Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product', *Molecular and Cellular Biology*, 5(12), pp. 3610–3616. doi: 10.1128/mcb.5.12.3610-3616.1985

Faulkner, G.J. *et al.* (2009) 'The regulated retrotransposon transcriptome of mammalian cells', *Nature Genetics*, 41(5), pp. 563–571. doi: 10.1038/ng.368

Fayomi, A.P. and Orwig, K.E. (2018) 'Spermatogonial stem cells and spermatogenesis in mice, monkeys and men', *Stem Cell Res*, 29, pp. 207–214.

Fazio, S. de *et al.* (2011) 'The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements', *Nature*, 480(7376), pp. 259–263.

Feng, Q. *et al.* (1996) 'Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition', *Cell*, 87(5), pp. 905–916. doi: 10.1016/s0092-8674(00)81997-2

Feschotte, C. and Pritham, E.J. (2007) 'DNA transposons and the evolution of eukaryotic genomes', *Annu. Rev. Genet.*, 41, pp. 331–368.

Frankish, A. *et al.* (2019) 'GENCODE reference annotation for the human and mouse genomes', *Nucleic Acids Res*, 47(D1), D766-D773. doi: 10.1093/nar/gky955

Fraune, J. *et al.* (2012) 'The mammalian synaptonemal complex: protein components, assembly and role in meiotic recombination', *Experimental Cell Research*, 318(12), pp. 1340–1346. doi: 10.1016/j.yexcr.2012.02.018

Fuks, F. *et al.* (2000) 'DNA methyltransferase Dnmt1 associates with histone deacetylase activity', *Nature Genetics*, 24(1), pp. 88–91. doi: 10.1038/71750

Fuks, F. *et al.* (2001) 'Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription', *EMBO J*, 20(10), pp. 2536–2544. doi: 10.1093/emboj/20.10.2536

Fuks, F. *et al.* (2003) 'The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation', *Journal of Biological Chemistry*, 278(6), pp. 4035–4040. doi: 10.1074/jbc.M210256200

Gallardo, T. *et al.* (2007) 'Generation of a germ cell-specific mouse transgenic Cre line, Vasa-Cre', *Genesis*, 45(6), pp. 413–417. doi: 10.1002/dvg.20310

Gallenberger, M. *et al.* (2011) 'Lack of WDR36 leads to preimplantation embryonic lethality in mice and delays the formation of small subunit ribosomal RNA in human cells in vitro', *Hum. Mol. Genet.*, 20(3), pp. 422–435.

Gaudet, F. *et al.* (2003) 'Induction of tumors in mice by genomic hypomethylation', *Science (New York, N.Y.)*, 300(5618), pp. 489–492. doi: 10.1126/science.1083558

Ge, Y.-Z. *et al.* (2004) 'Chromatin targeting of de novo DNA methyltransferases by the PWWP domain', *Journal of Biological Chemistry*, 279(24), pp. 25447–25454. doi: 10.1074/jbc.M312296200

Geiman, T.M. *et al.* (2004) 'DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system', *Biochemical and Biophysical Research Communications*, 318(2), pp. 544–555. doi: 10.1016/j.bbrc.2004.04.058

Genet, M. and Torres-Padilla, M.-E. (2020) 'The molecular and cellular features of 2-cell-like cells: a reference guide', *Development*, 147(16). doi: 10.1242/dev.189688

Gewiss, R., Topping, T. and Griswold, M.D. (2019) 'Cycles, waves, and pulses: Retinoic acid and the organization of spermatogenesis', *Andrology*, 8(4), pp. 892–897. doi: 10.1111/andr.12722

Gifford, R.J. *et al.* (2018) 'Nomenclature for endogenous retrovirus (ERV) loci', *Retrovirology*, 15(1), p. 59. doi: 10.1186/s12977-018-0442-1

Gkountela, S. *et al.* (2015) 'DNA Demethylation Dynamics in the Human Prenatal Germline', *Cell*, 161(6), pp. 1425–1436. doi: 10.1016/j.cell.2015.05.012

Godin, I. and Wylie, C.C. (1991) 'TGF beta 1 inhibits proliferation and has a chemotropic effect on mouse primordial germ cells in culture', *Development*, 113(4), pp. 1451–1457. doi: 10.1242/dev.113.4.1451

Goll, M.G. and Bestor, T.H. (2005) 'Eukaryotic cytosine methyltransferases', *Annual Review of Biochemistry*, 74, pp. 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Gordon, C.A., Hartono, S.R. and Chédin, F. (2013) 'Inactive DNMT3B splice variants modulate de novo DNA methylation', *PLoS One*, 8(7), e69486. doi: 10.1371/journal.pone.0069486

Green, C.D. *et al.* (2018) 'A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq', *Dev. Cell*, 46(5), 651-667.e10.

Greenberg, M.V.C. and Bourc'his, D. (2019) 'The diverse roles of DNA methylation in mammalian development and disease', *Nat. Rev. Mol. Cell Biol.*, 20(10), pp. 590–607.

Griswold, M.D. (2016) 'Spermatogenesis: The Commitment to Meiosis', *Physiological Reviews*, 96(1), pp. 1–17. doi: 10.1152/physrev.00013.2015

Gruenbaum, Y. *et al.* (1981) 'Sequence specificity of methylation in higher plant DNA', *Nature*, 292(5826), pp. 860–862. doi: 10.1038/292860a0

Guibert, S., Forné, T. and Weber, M. (2012) 'Global profiling of DNA methylation erasure in mouse primordial germ cells', *Genome Res*, 22(4), pp. 633–641.

Gujar, H., Weisenberger, D.J. and Liang, G. (2019) 'The roles of human DNA methyltransferases and their isoforms in shaping the epigenome', *Genes*, 10(2).

Guo, F. *et al.* (2015) 'The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells', *Cell*, 161(6), pp. 1437–1452. doi: 10.1016/j.cell.2015.05.015

Guo, X. *et al.* (2015) 'Structural insight into autoinhibition and histone H3-induced activation of DNMT3A', *Nature*, 517(7536), pp. 640–644. doi: 10.1038/nature13899

Hackett, J.A. *et al.* (2013) 'Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine', *Science*, 339(6118), pp. 448–452. doi: 10.1126/science.1229277

Haggerty, C. *et al.* (2021) 'Dnmt1 has de novo activity targeted to transposable elements', *Nature Structural & Molecular Biology*, 28(7), pp. 594–603. doi: 10.1038/s41594-021-00603-8

Hajkova, P. *et al.* (2010) 'Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway', *Science (New York, N.Y.)*, 329(5987), pp. 78–82. doi: 10.1126/science.1187945

Handel, M.A. and Schimenti, J.C. (2010) 'Genetics of mammalian meiosis: regulation, dynamics and impact on fertility', *Nature Reviews Genetics*, 11(2), pp. 124–136. doi: 10.1038/nrg2723

Hargan-Calvopina, J. *et al.* (2016) 'Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation', *Dev. Cell*, 39(1), pp. 75–86.

Haynes, C. *et al.* (2006) 'Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes', *PLoS Computational Biology*, 2(8), e100. doi: 10.1371/journal.pcbi.0020100

Heffner, C.S. *et al.* (2012) 'Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource', *Nature Communications*, 3, p. 1218. doi: 10.1038/ncomms2186

Henikoff, S. *et al.* (2020) 'Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation', *Elife*, 9.

Hermann, B.P. *et al.* (2018) 'The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids', *Cell Reports*, 25(6), 1650-1667.e8. doi: 10.1016/j.celrep.2018.10.026

Heyn, H. *et al.* (2012) 'Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient', *Epigenetics*, 7(6), pp. 542–550. doi: 10.4161/epi.20523

Hill, P.W.S. *et al.* (2018) 'Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte', *Nature*, 555(7696), pp. 392–396.

Hill, P.W.S., Amouroux, R. and Hajkova, P. (2014) 'DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story', *Genomics*, 104(5), pp. 324–333. doi: 10.1016/j.ygeno.2014.08.012

Hirai, Y. *et al.* (2013) 'Nucleolar scaffold protein, WDR46, determines the granular compartmental localization of nucleolin and DDX21', *Genes Cells*, 18(9), pp. 780–797.

Hiramatsu, R. *et al.* (2009) 'A critical time window of Sry action in gonadal sex determination in mice', *Development*, 136(1), pp. 129–138. doi: 10.1242/dev.029587

Hirano, T., Hasuwa, H. and Siomi, H. (2017) 'Identification of Mouse piRNA Pathway Components Using Anti-MIWI2 Antibodies', in Buszczak, M. (ed.) *Germline Stem Cells.* New York, NY: Springer New York, pp. 205–216.

Hittinger, C.T. and Carroll, S.B. (2007) 'Gene duplication and the adaptive evolution of a classic genetic switch', *Nature*, 449(7163), pp. 677–681. doi: 10.1038/nature06151

Holliday, R. and Pugh, J.E. (1975) 'DNA Modification Mechanisms and Gene Activity During Development', *Science (New York, N.Y.)*, 187(4173), pp. 226–232. doi: 10.1126/science.187.4173.226

Hotchkiss, R.D. (1948) 'THE QUANTITATIVE SEPARATION OF PURINES, PYRIMIDINES, AND NUCLEOSIDES BY PAPER CHROMATOGRAPHY', *Journal of Biological Chemistry*, 175(1), pp. 315–332. doi: 10.1016/S0021-9258(18)57261-6

Huang, H. *et al.* (2011) 'piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling', *Developmental Cell*, 20(3), pp. 376–387. doi: 10.1016/j.devcel.2011.01.004

Iakoucheva, L.M. *et al.* (2002) 'Intrinsic disorder in cell-signaling and cancer-associated proteins', *Journal of Molecular Biology*, 323(3), pp. 573–584. doi: 10.1016/s0022-2836(02)00969-5

Ikami, K. *et al.* (2015) 'Hierarchical differentiation competence in response to retinoic acid ensures stem cell maintenance during mouse spermatogenesis', *Development*, 142(9), pp. 1582–1592. doi: 10.1242/dev.118695

Ipsaro, J.J. *et al.* (2012) 'The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis', *Nature*, 491(7423), pp. 279–283. doi: 10.1038/nature11502

Ishino, K. *et al.* (2021) 'Hamster PIWI proteins bind to piRNAs with stage-specific size variations during oocyte maturation', *Nucleic Acids Res*, 49(5), pp. 2700–2720. doi: 10.1093/nar/gkab059

Ito, S. *et al.* (2011) 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science (New York, N.Y.)*, 333(6047), pp. 1300–1303. doi: 10.1126/science.1210597

Ivancevic, A.M. *et al.* (2016) 'LINEs between Species: Evolutionary Dynamics of LINE-1 Retrotransposons across the Eukaryotic Tree of Life', *Genome Biology and Evolution*, 8(11), pp. 3301–3322. doi: 10.1093/gbe/evw243

Iwanami, N. *et al.* (2008) 'WDR55 is a nucleolar modulator of ribosomal RNA synthesis, cell cycle progression, and teleost organ development', *PLoS Genet*, 4(8), e1000171.

Iwasaki, Y.W. *et al.* (2021) 'Piwi-piRNA complexes induce stepwise changes in nuclear architecture at target loci', *The EMBO Journal*, 40(18), e108345. doi: 10.15252/embj.2021108345

Jain, D. *et al.* (2017) 'Rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline', *rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline*, p. 121822.

Jan, S.Z. *et al.* (2012) 'Molecular control of rodent spermatogenesis', *Biochimica Et Biophysica Acta*, 1822(12), pp. 1838–1850. doi: 10.1016/j.bbadis.2012.02.008

Jeanpierre, M. *et al.* (1993) 'An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome', *Human Molecular Genetics*, 2(6), pp. 731–735. doi: 10.1093/hmg/2.6.731

Jeltsch, A. (2006) 'On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme', *Epigenetics*, 1(2), pp. 63–66. doi: 10.4161/epi.1.2.2767

Jensen, S. *et al.* (2020) 'Conserved Small Nucleotidic Elements at the Origin of Concerted piRNA Biogenesis from Genes and lncRNAs', *Cells*, 9(6). doi: 10.3390/cells9061491

Jern, P. and Coffin, J.M. (2008) 'Effects of retroviruses on host genome function', *Annual Review of Genetics*, 42, pp. 709–732. doi: 10.1146/annurev.genet.42.110807.091501

Johnson, T.B. and Coghill, R.D. (1925) 'RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS 1', *Journal of the American Chemical Society*, 47(11), pp. 2838–2844. doi: 10.1021/ja01688a030

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2

Jurkowska, R.Z., Jurkowski, T.P. and Jeltsch, A. (2011) 'Structure and function of mammalian DNA methyltransferases', *Chembiochem : a European Journal of Chemical Biology*, 12(2), pp. 206–222. doi: 10.1002/cbic.201000195

Kagiwada, S. *et al.* (2013) 'Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice', *The EMBO Journal*, 32(3), pp. 340–353. doi: 10.1038/emboj.2012.331

Kamesaki, H. *et al.* (1998) 'TGF-β1 Induces the Cyclin-Dependent Kinase Inhibitor p27 Kip1 mRNA and Protein in Murine B Cells', *The Journal of Immunology*, 160(2), pp. 770–777. doi: 10.4049/jimmunol.160.2.770

Kaneda, M. *et al.* (2004) 'Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting', *Nature*, 429(6994), pp. 900–903.

Kapitonov, V.V. and Jurka, J. (2001) 'Rolling-circle transposons in eukaryotes', *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp. 8714–8719. doi: 10.1073/pnas.151269298

Kapitonov, V.V. and Jurka, J. (2005) 'RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons', *PLoS Biol*, 3(6), e181. doi: 10.1371/journal.pbio.0030181

Kashimada, K. and Koopman, P. (2010) 'Sry: the master switch in mammalian sex determination', *Development*, 137(23), pp. 3921–3930. doi: 10.1242/dev.048983

Kaufmann, C. and Wutz, A. (2023) 'IndiSPENsable for X Chromosome Inactivation and Gene Silencing', *Epigenomes*, 7(4). doi: 10.3390/epigenomes7040028

Kaya-Okur, H.S. *et al.* (2019) 'CUT&Tag for efficient epigenomic profiling of small samples and single cells', *Nat. Commun.*, 10(1), pp. 1–10.

Kazazian, H.H. (2004) 'Mobile elements: drivers of genome evolution', *Science (New York, N.Y.)*, 303(5664), pp. 1626–1632. doi: 10.1126/science.1089670

Kazazian, H.H. and Moran, J.V. (2017) 'Mobile DNA in Health and Disease', *The New England Journal of Medicine*, 377(4), pp. 361–370. doi: 10.1056/NEJMra1510092

Kehler, J. *et al.* (2004) 'Oct4 is required for primordial germ cell survival', *EMBO Rep*, 5(11), pp. 1078–1083. doi: 10.1038/sj.embor.7400279

Khazina, E. and Weichenrieder, O. (2009) 'Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame', *Proceedings of the National Academy of*

Sciences of the United States of America, 106(3), pp. 731–736. doi: 10.1073/pnas.0809964106

Kigami, D. *et al.* (2003) 'MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos', *Biology of Reproduction*, 68(2), pp. 651–654. doi: 10.1095/biolreprod.102.007906

Kim, H.-M. and Liu, Z. (2024) 'LSD2 Is an Epigenetic Player in Multiple Types of Cancer and Beyond', *Biomolecules*, 14(5). doi: 10.3390/biom14050553

Kim, P.M. *et al.* (2008) 'The role of disorder in interaction networks: a structural analysis', *Molecular Systems Biology*, 4, p. 179. doi: 10.1038/msb.2008.16

Kimura, H. and Sato, Y. (2022) 'Imaging transcription elongation dynamics by new technologies unveils the organization of initiation and elongation in transcription factories', *Current Opinion in Cell Biology*, 74, pp. 71–79. doi: 10.1016/j.ceb.2022.01.002

Kleckner, N. (2006) 'Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex', *Chromosoma*, 115(3), pp. 175–194. doi: 10.1007/s00412-006-0055-7

Klein, C.J. *et al.* (2011) 'Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss', *Nature Genetics*, 43(6), pp. 595–600. doi: 10.1038/ng.830

Kluin, P.M. and Rooij, D.G. de (1981) 'A comparison between the morphology and cell kinetics of gonocytes and adult type undifferentiated spermatogonia in the mouse', *International Journal of Andrology*, 4(4), pp. 475–493. doi: 10.1111/j.1365-2605.1981.tb00732.x

Kobayashi, H. *et al.* (2013) 'High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice', *Genome Research*, 23(4), pp. 616–627. doi: 10.1101/gr.148023.112

Kojima-Kita, K. *et al.* (2016) 'MIWI2 as an Effector of DNA Methylation and Gene Silencing in Embryonic Male Germ Cells', *Cell Rep*, 16(11), pp. 2819–2828.

Konieczny, G. (2022) *Role of C19ORF84H in MIWI2-mediated silencing of retrotransposons in the male germline*.

Kooistra, S.M. and Helin, K. (2012) 'Molecular mechanisms and potential functions of histone demethylases', *Nat. Rev. Mol. Cell Biol.*, 13(5), pp. 297–311.

Koubova, J. *et al.* (2006) 'Retinoic acid regulates sex-specific timing of meiotic initiation in mice', *Proceedings of the National Academy of Sciences of the United States of America*, 103(8), pp. 2474–2479. doi: 10.1073/pnas.0510813103

Krieg, A.M., Gourley, M.F. and Perl, A. (1992) 'Endogenous retroviruses: potential etiologic agents in autoimmunity', *FASEB J*, 6(8), pp. 2537–2544. doi: 10.1096/fasebj.6.8.1592206

Kristensen, L.H. *et al.* (2012) 'Studies of H3K4me3 demethylation by KDM5B/Jarid1B/PLU1 reveals strong substrate recognition in vitro and identifies 2,4-pyridine-dicarboxylic acid as an in vitro and in cell inhibitor', *The FEBS Journal*, 279(11), pp. 1905–1914. doi: 10.1111/j.1742-4658.2012.08567.x

Krueger, F. and Andrews, S.R. (2011) 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics*, 27(11), pp. 1571–1572.

Kubo, N. *et al.* (2015) 'DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis', *BMC Genomics*, 16(1), p. 624. doi: 10.1186/s12864-015-1833-5

Kubo, N. *et al.* (2024) 'Combined and differential roles of ADD domains of DNMT3A and DNMT3L on DNA methylation landscapes in mouse germ cells', *Nat. Commun.*, 15(1), p. 3266.

Kulis, M. and Esteller, M. (2010) 'DNA methylation and cancer', *Advances in Genetics*, 70, pp. 27–56. doi: 10.1016/B978-0-12-380866-0.60002-2

Kuramochi-Miyagawa, S. (2008) 'DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.pdf', *Genes and Development*, 1, pp. 908–917.

Kuramochi-Miyagawa, S. *et al.* (2010) 'MVH in piRNA processing and gene silencing of retrotransposons', *Genes & Development*, 24(9), pp. 887–892. doi: 10.1101/gad.1902110

Kurimoto, K. *et al.* (2008) 'Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice', *Genes Dev*, 22(12), pp. 1617–1635. doi: 10.1101/gad.1649908

Kustatscher, G. *et al.* (2014) 'Chromatin enrichment for proteomics', *Nature Protocols*, 9(9), pp. 2090–2099. doi: 10.1038/nprot.2014.142

Lambrot, R., Lafleur, C. and Kimmins, S. (2015) 'The histone demethylase KDM1A is essential for the maintenance and differentiation of spermatogonial stem cells and progenitors', *FASEB J*, 29(11), pp. 4402–4416. doi: 10.1096/fj.14-267328

Lander, E.S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. doi: 10.1038/35057062

Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nat. Methods*, 9(4), pp. 357–359.

Lauria, A. *et al.* (2023) 'DNMT3B supports meso-endoderm differentiation from mouse embryonic stem cells', *Nature Communications*, 14(1), p. 367. doi: 10.1038/s41467-023-35938-x

Lawson, K.A. *et al.* (1999) 'Bmp4 is required for the generation of primordial germ cells in the mouse embryo', *Genes Dev*, 13(4), pp. 424–436. doi: 10.1101/GAD.13.4.424

Lees-Murdock, D.J. and Walsh, C.P. (2008) 'DNA methylation reprogramming in the germ line', *Epigenetics*, 3(1), pp. 5–13. doi: 10.4161/epi.3.1.5553

Lehnertz, B. *et al.* (2003) 'Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin', *Current Biology : CB*, 13(14), pp. 1192–1200. doi: 10.1016/s0960-9822(03)00432-9

Ley, T.J. *et al.* (2010) 'DNMT3A mutations in acute myeloid leukemia', *The New England Journal of Medicine*, 363(25), pp. 2424–2433. doi: 10.1056/NEJMoa1005143

Li, E., Beard, C. and Jaenisch, R. (1993) 'Role for DNA methylation in genomic imprinting', *Nature*, 366(6453), pp. 362–365. doi: 10.1038/366362a0

Li, E., Bestor, T.H. and Jaenisch, R. (1992) 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell*, 69(6), pp. 915–926. doi: 10.1016/0092-8674(92)90611-F

Li, J. *et al.* (2014) 'An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition', *Nucleic Acids Res*, 42(7), pp. 4546–4562.

Li, J. and Tibshirani, R. (2013) 'Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data', *Stat. Methods Med. Res.*, 22(5), pp. 519–536.

Li, R., Grimm, S.A. and Wade, P.A. (2021) 'CUT&Tag-BS for simultaneous profiling of histone modification and DNA methylation with high efficiency and low cost', *Cell Rep Methods*, 1(8).

Li, Y., Zhang, Y. and Liu, M. (2021) 'Knockout Gene-Based Evidence for PIWI-Interacting RNA Pathway in Mammals', *Front Cell Dev Biol*, 9, p. 681188. doi: 10.3389/fcell.2021.681188

Li, Z. *et al.* (2015) 'Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells', *Genome Biol*, 16(1), p. 115. doi: 10.1186/s13059-015-0685-2

Liao, Y., Smyth, G.K. and Shi, W. (2014) 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30(7), pp. 923–930. doi: 10.1093/bioinformatics/btt656

Lim, S.L. *et al.* (2015) 'HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse', *PLoS Genetics*, 11(10), e1005620. doi: 10.1371/journal.pgen.1005620

Lister, R. *et al.* (2009) 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature*, 462(7271), pp. 315–322. doi: 10.1038/nature08514

Lister, R. *et al.* (2013) 'Global epigenomic reconfiguration during mammalian brain development', *Science (New York, N.Y.)*, 341(6146), p. 1237905. doi: 10.1126/science.1237905

Liu, S. *et al.* (2014) 'Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells', *Genes Dev*, 28(18), pp. 2041–2055.

Liu, Y. *et al.* (2019) 'Maternal DCAF13 Regulates Chromatin Tightness to Contribute to Embryonic Development', *Sci. Rep.*, 9(1), p. 6278.

Lord, T. and Oatley, J.M. (2017) 'A revised Asingle model to explain stem cell dynamics in the mouse male germline', *Reproduction*, 154(2), R55-64. doi: 10.1530/REP-17-0034

Luan, D.D. *et al.* (1993) 'Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition', *Cell*, 72(4), pp. 595–605. doi: 10.1016/0092-8674(93)90078-5

Mager, D.L. and Stoye, J.P. (2015) 'Mammalian Endogenous Retroviruses', *Microbiology Spectrum*, 3(1), MDNA3-0009-2014. doi: 10.1128/microbiolspec.mdna3-0009-2014

Magnúsdóttir, E. *et al.* (2013) 'A tripartite transcription factor network regulates primordial germ cell specification in mice', *Nature Cell Biology*, 15(8), pp. 905–915. doi: 10.1038/ncb2798

Maksakova, I.A. *et al.* (2006) 'Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line', *PLoS Genet*, 2(1), e2.

Mallona, I. *et al.* (2020) 'Flanking sequence preference modulates de novo DNA methylation in the mouse genome', *Nucleic Acids Res*, 49(1), pp. 145–157. doi: 10.1093/nar/gkaa1168

Manakov, S.A. *et al.* (2015) 'MIWI2 and MILI Have Differential Effects on piRNA Biogenesis and DNA Methylation', *Cell Rep*, 12(8), pp. 1234–1243.

Marayati, B.F. *et al.* (2020) 'The Catalytic-Dependent and -Independent Roles of Lsd1 and Lsd2 Lysine Demethylases in Heterochromatin Formation in Schizosaccharomyces pombe', *Cells*, 9(4). doi: 10.3390/cells9040955

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12.

Martin, S.L. (1991) 'Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells', *Molecular and Cellular Biology*, 11(9), pp. 4804–4807. doi: 10.1128/mcb.11.9.4804

Martin, S.L. and Bushman, F.D. (2001) 'Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon', *Molecular and Cellular Biology*, 21(2), pp. 467–475. doi: 10.1128/MCB.21.2.467-475.2001

Maserati, M. *et al.* (2011) 'Wdr74 is required for blastocyst formation in the mouse', *PLoS One*, 6(7).

Matson, C.K. *et al.* (2010) 'The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells', *Developmental Cell*, 19(4), pp. 612–624. doi: 10.1016/j.devcel.2010.09.010

Mattei, A.L., Bailly, N. and Meissner, A. (2022) 'DNA methylation: a historical perspective', *Trends Genet*, 38(7), pp. 676–707.

Maurer-Stroh, S. *et al.* (2003) 'The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains', *Trends in Biochemical Sciences*, 28(2), pp. 69–74. doi: 10.1016/S0968-0004(03)00004-5

MCCLINTOCK, B. (1950) 'The origin and behavior of mutable loci in maize', *Proceedings of the National Academy of Sciences of the United States of America*, 36(6), pp. 344–355. doi: 10.1073/pnas.36.6.344

MCCLINTOCK, B. (1956) 'Controlling elements and the gene', *Cold Spring Harbor Symposia on Quantitative Biology*, 21, pp. 197–216. doi: 10.1101/SQB.1956.021.01.017

Metzger, E. *et al.* (2005) 'LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription', *Nature*, 437(7057), pp. 436–439. doi: 10.1038/nature04020

Metzger, E. *et al.* (2010) 'Phosphorylation of histone H3T6 by PKCbeta(I) controls demethylation at histone H3K4', *Nature*, 464(7289), pp. 792–796. doi: 10.1038/nature08839

Michalski, A. *et al.* (2011) 'Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer', *Mol. Cell. Proteomics*, 10(9), M111.011015.

Mikedis, M.M. and Downs, K.M. (2017) 'PRDM1/BLIMP1 is widely distributed to the nascent fetal-placental interface in the mouse gastrula', *Developmental Dynamics : an Official Publication of the American Association of Anatomists*, 246(1), pp. 50–71. doi: 10.1002/dvdy.24461

Mita, P. *et al.* (2018) 'LINE-1 protein localization and functional dynamics during the cell cycle', *eLife Sciences Publications, Ltd,* 8 January. Available at: https://elifesciences.org/articles/30058 (Accessed: 27 June 2024).

Molaro, A. *et al.* (2014) 'Two waves of de novo methylation during mouse germ cell development', *Genes Dev*, 28(14), pp. 1544–1549.

Molaro, A. and Malik, H.S. (2016) 'Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline', *Current Opinion in Genetics & Development*, 37, pp. 51–58. doi: 10.1016/j.gde.2015.12.001

Molaro, A., Malik, H.S. and Bourc'his, D. (2020) 'Dynamic Evolution of De Novo DNA Methyltransferases in Rodent and Primate Genomes', *Mol. Biol. Evol.*, 37(7), pp. 1882–1892.

Molyneaux, K.A. *et al.* (2003) 'The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival', *Development*, 130(18), pp. 4279–4286. doi: 10.1242/dev.00640

Montgomery, B.E. *et al.* (2021) 'Dual roles for piRNAs in promoting and preventing gene silencing in C. elegans', *Cell Reports*, 37(10), p. 110101. doi: 10.1016/j.celrep.2021.110101

Moreno, S.G. *et al.* (2010) 'TGFbeta signaling in male germ cells regulates gonocyte quiescence and fertility in mice', *Developmental Biology*, 342(1), pp. 74–84. doi: 10.1016/j.ydbio.2010.03.007

Morgan, M. *et al.* (2019) 'A programmed wave of uridylation-primed mRNA degradation is essential for meiotic progression and mammalian spermatogenesis', *Cell Research*, 29(3), pp. 221–232. doi: 10.1038/s41422-018-0128-1

Morgante, M. *et al.* (2005) 'Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize', *Nature Genetics*, 37(9), pp. 997–1002. doi: 10.1038/ng1615

Morselli, M. *et al.* (2015) 'In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse', *Elife*, 4, e06205.

Mortusewicz, O. *et al.* (2005) 'Recruitment of DNA methyltransferase I to DNA repair sites', *Proceedings of the National Academy of Sciences of the United States of America*, 102(25), pp. 8905–8909. doi: 10.1073/pnas.0501034102

Mould, A. *et al.* (2012) 'Blimp1/Prdm1 governs terminal differentiation of endovascular trophoblast giant cells and defines multipotent progenitors in the developing placenta', *Genes & Development*, 26(18), pp. 2063–2074. doi: 10.1101/gad.199828.112

Müller, I. and Helin, K. (2024) 'Keep quiet: the HUSH complex in transcriptional silencing and disease', *Nature Structural & Molecular Biology*, 31(1), pp. 11–22. doi: 10.1038/s41594-023-01173-7

Munafò, M. *et al.* (2021) 'Channel nuclear pore complex subunits are required for transposon silencing in Drosophila', *Elife*, 10, e66321.

Myrick, D.A. *et al.* (2017) 'KDM1A/LSD1 regulates the differentiation and maintenance of spermatogonia in mice', *PLoS One*, 12(5), e0177473. doi: 10.1371/journal.pone.0177473

Naas, T.P. *et al.* (1998) 'An actively retrotransposing, novel subfamily of mouse L1 elements', *EMBO J*, 17(2), pp. 590–597. doi: 10.1093/emboj/17.2.590

Nagamori, I. *et al.* (2018) 'Relationship between PIWIL4-Mediated H3K4me2 Demethylation and piRNA-Dependent DNA Methylation', *Cell Rep*, 25(2), pp. 350–356.

Neri, F. *et al.* (2017) 'Intragenic DNA methylation prevents spurious transcription initiation', *Nature*, 543(7643), pp. 72–77. doi: 10.1038/nature21373

Nishimura, T. *et al.* (2018) 'PNLDC1, mouse pre-piRNA Trimmer, is required for meiotic and post-meiotic male germ cell development', *EMBO Reports*, 19(3). doi: 10.15252/embr.201744957

Nowialis, P. *et al.* (2019) 'Catalytically inactive Dnmt3b rescues mouse embryonic development by accessory and repressive functions', *Nat. Commun.*, 10(1).

O'Donnell, L. (2014) 'Mechanisms of spermiogenesis and spermiation and how they are disturbed', *Spermatogenesis*, 4(2), e979623. doi: 10.4161/21565562.2014.979623

O'Hare, K. and Rubin, G.M. (1983) 'Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome', *Cell*, 34(1), pp. 25–35. doi: 10.1016/0092-8674(83)90133-2

Ohinata, Y. *et al.* (2005) 'Blimp1 is a critical determinant of the germ cell lineage in mice', *Nature*, 436(7048), pp. 207–213.

Ohno, R. *et al.* (2013) 'A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells', *Development*, 140(14), pp. 2892–2903. doi: 10.1242/dev.093229

Okano, M. *et al.* (1999) 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development', *Cell*, 99(3), pp. 247–257. doi: 10.1016/S0092-8674(00)81656-6

Okano, M., Xie, S. and Li, E. (1998) 'Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases', *Nature Genetics*, 19(3), pp. 219–220. doi: 10.1038/890

Okutman, O. *et al.* (2015) 'Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family', *Human Molecular Genetics*, 24(19), pp. 5581–5588. doi: 10.1093/hmg/ddv290

Olsen, J.V. *et al.* (2007) 'Higher-energy C-trap dissociation for peptide modification analysis', *Nat. Methods*, 4(9), pp. 709–712.

Ooi, S.K.T. *et al.* (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature*, 448(7154), pp. 714–717.

Otani, J. *et al.* (2009) 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain', *EMBO Rep*, 10(11), pp. 1235–1241.

Ozata, D.M. *et al.* (2019) 'PIWI-interacting RNAs: small RNAs with big functions', *Nat. Rev. Genet.*, 20(2), pp. 89–108.

Ozturk, S. (2023) 'Genetic variants underlying spermatogenic arrests in men with non-obstructive azoospermia', *Cell Cycle*, 00(00), pp. 1–41.

Pandey, R.R. *et al.* (2013) 'Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice', *Proceedings of the National Academy of Sciences of the United States of America*, 110(41), pp. 16492–16497. doi: 10.1073/pnas.1316316110

Parikh, R.Y., Lin, H. and Gangaraju, V.K. (2018) 'A critical role for nucleoporin 358 (Nup358) in transposon silencing and piRNA biogenesis in Drosophila', *J. Biol. Chem.*, 293(24), pp. 9140–9147.

Pastore, B. *et al.* (2021) 'pre-piRNA trimming and 2'-O-methylation protect piRNAs from 3' tailing and degradation in C. elegans', *Cell Reports*, 36(9), p. 109640. doi: 10.1016/j.celrep.2021.109640

Percharde, M. *et al.* (2018) 'A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity', *Cell*, 174(2), 391-405.e19.

Petell, C.J. *et al.* (2016) 'An epigenetic switch regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation', *Nucleic Acids Res*, 44(16), pp. 7605–7617.

Pisacane, P. and Halic, M. (2017) 'Tailing and degradation of Argonaute-bound small RNAs protect the genome from uncontrolled RNAi', *Nature Communications*, 8, p. 15332. doi: 10.1038/ncomms15332

Platt, R.N., Vandewege, M.W. and Ray, D.A. (2018) 'Mammalian transposable elements and their impacts on genome evolution', *Chromosome Res*, 26(1-2), pp. 25–43.

Pradhan, S. *et al.* (1999) 'Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation', *Journal of Biological Chemistry*, 274(46), pp. 33002–33010. doi: 10.1074/jbc.274.46.33002

Pritham, E.J. and Feschotte, C. (2007) 'Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus', *Proceedings of the National Academy of Sciences of the United States of America*, 104(6), pp. 1895–1900. doi: 10.1073/pnas.0609601104

Ramakrishna, N.B. *et al.* (2022) 'Mouse primordial germ-cell-like cells lack piRNAs', *Dev. Cell*, 57(23), 2661-2668.e5.

Ramírez, F. *et al.* (2016) 'deepTools2: a next generation web server for deep-sequencing data analysis', *Nucleic Acids Res*, 44(W1), W160-5. doi: 10.1093/nar/gkw257

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) 'Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips', *Nat. Protoc.*, 2(8), pp. 1896–1906.

Ratnam, S. *et al.* (2002) 'Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development', *Developmental Biology*, 245(2), pp. 304–314. doi: 10.1006/dbio.2002.0628

Ravindran, S. (2012) ‘Barbara McClintock and the discovery of jumping genes’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp. 20198–20199. doi: 10.1073/pnas.1219372109

Reale, A. *et al.* (2005) ‘Modulation of DNMT1 activity by ADP-ribose polymers’, *Oncogene*, 24(1), pp. 13–19. doi: 10.1038/sj.onc.1208005

Reik, W., Dean, W. and Walter, J. (2001) ‘Epigenetic Reprogramming in Mammalian Development’, *Science*, 293(5532), pp. 1089–1093.

Ribet, D. *et al.* (2007) ‘Murine MusD retrotransposon: structure and molecular evolution of an "intracellularized" retrovirus’, *Journal of Virology*, 81(4), pp. 1888–1898. doi: 10.1128/jvi.02051-06

Ribet, D. *et al.* (2008) ‘An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus’, *Genome Res*, 18(4), pp. 597–609. doi: 10.1101/gr.073486.107

Richards, A.J., Enders, G.C. and Resnick, J.L. (1999) ‘Activin and TGFbeta limit murine primordial germ cell proliferation’, *Developmental Biology*, 207(2), pp. 470–475. doi: 10.1006/dbio.1998.9174

Riggs, A.D. (1975) ‘X inactivation, differentiation, and DNA methylation’, *Cytogenetics and Cell Genetics*, 14(1), pp. 9–25. doi: 10.1159/000130315

Rooij, D.G. de (2017) ‘The nature and dynamics of spermatogonial stem cells’, *Development*, 144(17), pp. 3022–3030. doi: 10.1242/dev.146571

Rooij, D.G. de and Russell, L.D. (2000) ‘All you wanted to know about spermatogonia but were afraid to ask’, *Journal of Andrology*, 21(6), pp. 776–798.

Rose, N.R. and Klose, R.J. (2014) ‘Understanding the relationship between DNA methylation and histone lysine methylation’, *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(12), pp. 1362–1372.

Ross, S.R. (2010) ‘Mouse mammary tumor virus molecular biology and oncogenesis’, *Viruses*, 2(9), pp. 2000–2012. doi: 10.3390/v2092000

Rowe, H.M. *et al.* (2010) ‘KAP1 controls endogenous retroviruses in embryonic stem cells’, *Nature*, 463(7278), pp. 237–240. doi: 10.1038/nature08674

Rubin, G.M., Kidwell, M.G. and Bingham, P.M. (1982) 'The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations', *Cell*, 29(3), pp. 987–994. doi: 10.1016/0092-8674(82)90462-7

Rubin, G.M. and Spradling, A.C. (1982) 'Genetic transformation of Drosophila with transposable element vectors', *Science (New York, N.Y.)*, 218(4570), pp. 348–353. doi: 10.1126/science.6289436

Runyan, C. *et al.* (2006) 'Steel factor controls midline cell death of primordial germ cells and is essential for their normal proliferation and migration', *Development*, 133(24), pp. 4861–4869. doi: 10.1242/dev.02688

Russell, L.D. *et al.* (1993) 'Histological and Histopathological Evaluation of the Testis', *International journal of andrology*, 16(1), p. 83. doi: 10.1111/j.1365-2605.1993.tb01156.x

Sager, R. and Kitchin, R. (1975) 'Selective Silencing of Eukaryotic DNA', *Science (New York, N.Y.)*, 189(4201), pp. 426–433. doi: 10.1126/science.189.4201.426

Saini, S.K. *et al.* (2017) 'DNA Methyltransferase1 (DNMT1) Isoform3 methylates mitochondrial genome and modulates its biology', *Scientific Reports*, 7(1), p. 1525. doi: 10.1038/s41598-017-01743-y

Saitou, M., Barton, S.C. and Surani, M.A. (2002) 'A molecular programme for the specification of germ cell fate in mice', *Nature*, 418(6895), pp. 293–300. doi: 10.1038/nature00927

Saitou, M., Kagiwada, S. and Kurimoto, K. (2012) 'Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells', *Development*, 139(1), pp. 15–31. doi: 10.1242/dev.050849

Saitou, M. and Yamaji, M. (2012) 'Primordial germ cells in mice', *Cold Spring Harb. Perspect. Biol.*, 4(11), pp. 1–20.

Sakashita, A. *et al.* (2023) 'Transcription of MERVL retrotransposons is required for preimplantation embryo development', *Nat. Genet.*, 55(3), pp. 484–495.

Sandhu, C. *et al.* (1997) 'Transforming growth factor beta stabilizes p15INK4B protein, increases p15INK4B-cdk4 complexes, and inhibits cyclin D1-cdk4 association in human mammary epithelial cells', *Molecular and Cellular Biology*, 17(5), pp. 2458–2467. doi: 10.1128/MCB.17.5.2458

Sangrithi, M.N. *et al.* (2017) 'Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline', *Dev. Cell*, 40(3), 289-301.e3.

Sato, K. and Siomi, M.C. (2020) 'The piRNA pathway in Drosophila ovarian germ and somatic cells', *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, 96(1), pp. 32–42. doi: 10.2183/pjab.96.003

Saxe, J.P. *et al.* (2013) 'Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline', *The EMBO Journal*, 32(13), pp. 1869–1885. doi: 10.1038/emboj.2013.121

Schapira, M. *et al.* (2017) 'WD40 repeat domain proteins: a novel target class?' *Nat. Rev. Drug Discov.*, 16(11), pp. 773–786.

Schnabl, J. *et al.* (2021) 'Molecular principles of Piwi-mediated cotranscriptional silencing through the dimeric SFiNX complex', *Genes & Development*, 35(5-6), pp. 392–409. doi: 10.1101/gad.347989.120

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*, 9(7), pp. 671–675. doi: 10.1038/nmeth.2089

Schöpp, T. *et al.* (2020) 'TEX15 is an essential executor of MIWI2-directed transposon DNA methylation and silencing', *Nat. Commun.*, 11(1), p. 3739.

Schöpp, T. *et al.* (2023) 'The DUF3715 domain has a conserved role in RNA-directed transposon silencing', *RNA*, 29(10), pp. 1471–1480. doi: 10.1261/rna.079693.123

Seisenberger, S. *et al.* (2012) 'The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells', *Mol. Cell*, 48(6), pp. 849–862.

Sendzikaite, G. (2021) *The role of the PWWP domain in the DNA methyltransferase 3A targeting to the genome.* Apollo - University of Cambridge Repository.

Senft, A.D. and Macfarlan, T.S. (2021) 'Transposable elements shape the evolution of mammalian development', *Nat. Rev. Genet.*, 22(11), pp. 691–711.

Shah, P. *et al.* (2024) 'Primordial germ cell DNA demethylation and development require DNA translesion synthesis', *Nature Communications*, 15(1), p. 3734. doi: 10.1038/s41467-024-47219-2

Sharp, A.J. *et al.* (2011) 'DNA methylation profiles of human active and inactive X chromosomes', *Genome Res*, 21(10), pp. 1592–1600. doi: 10.1101/gr.112680.110

Shevelyov, Y.Y. (2023) 'Interactions of Chromatin with the Nuclear Lamina and Nuclear Pore Complexes', *International Journal of Molecular Sciences*, 24(21). doi: 10.3390/ijms242115771

Shi, Y. *et al.* (2004) 'Histone demethylation mediated by the nuclear amine oxidase homolog LSD1', *Cell*, 119(7), pp. 941–953.

Shirane, K. *et al.* (2020) 'NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing', *Nat. Genet.*, 52(10), pp. 1088–1098.

Shoji, M. *et al.* (2009) 'The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline', *Dev. Cell*, 17(6), pp. 775–787.

Sijen, T. and Plasterk, R.H.A. (2003) 'Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi', *Nature*, 426(6964), pp. 310–314. doi: 10.1038/nature02107

Siomi, M.C. *et al.* (2011) 'PIWI-interacting small RNAs: the vanguard of genome defence', *Nature Reviews. Molecular Cell Biology*, 12(4), pp. 246–258. doi: 10.1038/nrm3089

Skene, P.J. and Henikoff, S. (2017) 'An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites', *Elife*, 6.

Slotkin, R.K. and Martienssen, R. (2007) 'Transposable elements and the epigenetic regulation of the genome', *Nature Reviews Genetics*, 8(4), pp. 272–285. doi: 10.1038/nrg2072

Smit, A.F. *et al.* (1995) 'Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences', *Journal of Molecular Biology*, 246(3), pp. 401–417. doi: 10.1006/jmbi.1994.0095

Smith, A.M. *et al.* (2021) 'Functional and epigenetic phenotypes of humans and mice with DNMT3A Overgrowth Syndrome', *Nature Communications*, 12(1), p. 4549. doi: 10.1038/s41467-021-24800-7

Smith, Z.D. and Meissner, A. (2013) 'DNA methylation: roles in mammalian development', *Nature Reviews Genetics*, 14(3), pp. 204–220. doi: 10.1038/nrg3354

Song, H.-W. and Wilkinson, M.F. (2014) 'Transcriptional control of spermatogonial maintenance and differentiation', *Seminars in Cell & Developmental Biology*, 30, pp. 14–26. doi: 10.1016/j.semcdb.2014.02.005

Sookdeo, A. *et al.* (2013) 'Revisiting the evolution of mouse LINE-1 in the genomic era', *Mobile DNA*, 4(1), p. 3. doi: 10.1186/1759-8753-4-3

Soper, S.F. *et al.* (2008) 'Mouse Maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis', *Developmental Cell*, 15(2), pp. 285–297. doi: 10.1016/j.devcel.2008.05.015

Spada, F. *et al.* (2006) 'Regulation of DNA methyltransferase 1', *Advances in Enzyme Regulation*, 46, pp. 224–234. doi: 10.1016/j.advenzreg.2006.01.011

Speek, M. (2001) 'Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes', *Molecular and Cellular Biology*, 21(6), pp. 1973–1985. doi: 10.1128/MCB.21.6.1973-1985.2001

Spradling, A.C. and Rubin, G.M. (1982) 'Transposition of cloned P elements into Drosophila germ line chromosomes', *Science (New York, N.Y.)*, 218(4570), pp. 341–347. doi: 10.1126/science.6289435

Stallmeyer, B. *et al.* (2024) 'Inherited defects of piRNA biogenesis cause transposon de-repression, impaired spermatogenesis, and human male infertility', *Nature Communications*, 15(1), p. 6637. doi: 10.1038/s41467-024-50930-9

Stewart-Morgan, K.R. *et al.* (2023) 'Quantifying propagation of DNA methylation and hydroxymethylation with iDEMS', *Nature Cell Biology*, 25(1), pp. 183–193. doi: 10.1038/s41556-022-01048-x

Stoye, J.P. (2012) 'Studies of endogenous retroviruses reveal a continuing evolutionary saga', *Nature Reviews. Microbiology*, 10(6), pp. 395–406. doi: 10.1038/nrmicro2783

Suarez, N.A., Macia, A. and Muotri, A.R. (2018) 'LINE-1 retrotransposons in healthy and diseased human brain', *Developmental Neurobiology*, 78(5), pp. 434–455. doi: 10.1002/dneu.22567

Suetake, I. *et al.* (2004) 'DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction', *J. Biol. Chem.*, 279(26), pp. 27816–27823.

Sun, Y.H. *et al.* (2020) 'Ribosomes guide pachytene piRNA formation on long intergenic piRNA precursors', *Nat. Cell Biol.*, 22(2), pp. 200–212.

Sun, Y.H., Lee, B. and Li, X.Z. (2022) 'The birth of piRNAs: how mammalian piRNAs are produced, originated, and evolved', *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, 33(2), pp. 293–311. doi: 10.1007/s00335-021-09927-8

Taglini, F. *et al.* (2024) 'DNMT3B PWWP mutations cause hypermethylation of heterochromatin', *EMBO Rep*, 25(3), pp. 1130–1155.

Tam, P.P. and Snow, M.H. (1981) 'Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos', *Journal of Embryology and Experimental Morphology*, 64, pp. 133–147.

Tang, W.W.C. *et al.* (2015) 'A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development', *Cell*, 161(6), pp. 1453–1467. doi: 10.1016/j.cell.2015.04.053

Tang, W.W.C. *et al.* (2016) 'Specification and epigenetic programming of the human germ line', *Nature Reviews Genetics*, 17(10), pp. 585–600. doi: 10.1038/nrg.2016.88

Tatton-Brown, K. *et al.* (2014) 'Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability', *Nature Genetics*, 46(4), pp. 385–388. doi: 10.1038/ng.2917

Taylor, M.S. *et al.* (2018) 'Dissection of affinity captured LINE-1 macromolecular complexes', *Elife*, 7, pp. 1–3.

Thawani, A. *et al.* (2024) 'Template and target-site recognition by human LINE-1 in retrotransposition', *Nature*, 626(7997), pp. 186–193. doi: 10.1038/s41586-023-06933-5

'The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update' (2024), *Nucleic Acids Res*, 52(W1), W83-W94.

Thomas, J. and Pritham, E.J. (2015) 'Helitrons, the Eukaryotic Rolling-circle Transposable Elements', *Microbiology Spectrum*, 3(4). doi: 10.1128/microbiolspec.MDNA3-0049-2014

Tóth, K.F. *et al.* (2016) 'The piRNA Pathway Guards the Germline Genome Against Transposable Elements', *Advances in Experimental Medicine and Biology*, 886, pp. 51–77. doi: 10.1007/978-94-017-7417-8_4

Tseng, Y.-T. *et al.* (2015) 'Epigenetic factors in the regulation of prospermatogonia and spermatogonial stem cells', *Reproduction*, 150(3), R77-91.

Tubio, J.M.C. *et al.* (2014) 'Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes', *Science (New York, N.Y.)*, 345(6196), p. 1251343. doi: 10.1126/science.1251343

Tyanova, S. *et al.* (2016) 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nat. Methods*, 13(9), pp. 731–740.

Uehara, R. *et al.* (2023) 'The DNMT3A ADD domain is required for efficient de novo DNA methylation and maternal imprinting in mouse oocytes', *PLoS Genetics*, 19(8), e1010855. doi: 10.1371/journal.pgen.1010855

Uneme, Y. *et al.* (2024) 'Morc1 reestablishes H3K9me3 heterochromatin on piRNA-targeted transposons in gonocytes', *Proceedings of the National Academy of Sciences of the United States of America*, 121(13), e2317095121. doi: 10.1073/pnas.2317095121

Vergouwen, R.P. *et al.* (1991) 'Proliferative activity of gonocytes, Sertoli cells and interstitial cells during testicular development in mice', *Journal of Reproduction and Fertility*, 93(1), pp. 233–243. doi: 10.1530/jrf.0.0930233

Vlachogiannis, G. *et al.* (2015) 'The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis', *Cell Reports*, 10(6), pp. 944–956. doi: 10.1016/j.celrep.2015.01.021

Vooijs, M., Jonkers, J. and Berns, A. (2001) 'A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent', *EMBO Rep*, 2(4), pp. 292–297. doi: 10.1093/embo-reports/kve064

Walsh, C.P., Chaillet, J.R. and Bestor, T.H. (1998) 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation [4]', *Nat. Genet.*, 20(2), pp. 116–117.

Walter, M. *et al.* (2016) 'An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells', *Elife*, 5.

Wang, H. *et al.* (2023) 'H3K4me3 regulates RNA polymerase II promoter-proximal pause-release', *Nature*, 615(7951), pp. 339–348.

Wang, J. *et al.* (2007) 'Opposing LSD1 complexes function in developmental gene activation and repression programmes', *Nature*, 446(7138), pp. 882–887. doi: 10.1038/nature05671

Wang, J. *et al.* (2017) 'NRF1 coordinates with DNA methylation to regulate spermatogenesis', *FASEB J*, 31(11), pp. 4959–4970. doi: 10.1096/fj.201700093R

Wang, J.-Y. *et al.* (2023) 'Gene-wide significant association analyses of DNMT1 genetic variants with Parkinson's disease', *Frontiers in Genetics*, 14. doi: 10.3389/fgene.2023.1112388

Wang, X. *et al.* (2020) 'Mitochondria Associated Germinal Structures in Spermatogenesis: piRNA Pathway Regulation and Beyond', *Cells*, 9(2). doi: 10.3390/cells9020399

Watanabe, T. *et al.* (2011) 'MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline', *Developmental Cell*, 20(3), pp. 364–375. doi: 10.1016/j.devcel.2011.01.005

Watanabe, T. *et al.* (2018) 'MIWI2 targets RNAs transcribed from piRNA-dependent regions to drive DNA methylation in mouse prospermatogonia', *EMBO J*, 37(18).

Weinberg, D.N. *et al.* (2019) 'The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape', *Nature*, 573(7773), pp. 281–286. doi: 10.1038/s41586-019-1534-3

Weinberg, D.N. *et al.* (2021) 'Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands', *Nature Genetics*, 53(6), pp. 794–800. doi: 10.1038/s41588-021-00856-5

Weisenberger, D.J. *et al.* (2004) 'Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation', *Molecular Cancer Research : MCR*, 2(1), pp. 62–72.

Wenda, J.M. *et al.* (2017) 'Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis and Function', *Developmental Cell*, 41(6), 623-637.e9. doi: 10.1016/j.devcel.2017.05.021

Wensink, P.C. *et al.* (1974) 'A system for mapping DNA sequences in the chromosomes of Drosophila melanogaster', *Cell*, 3(4), pp. 315–325. doi: 10.1016/0092-8674(74)90045-2

Western, P.S. *et al.* (2008) 'Dynamic regulation of mitotic arrest in fetal male germ cells', *Stem Cells*, 26(2), pp. 339–347. doi: 10.1634/stemcells.2007-0622

Wu, X. and Zhang, Y. (2017) 'TET-mediated active DNA demethylation: mechanism, function and beyond', *Nature Reviews Genetics*, 18(9), pp. 517–534. doi: 10.1038/nrg.2017.33

Wyrwoll, M.J. *et al.* (2022) 'The piRNA-pathway factor FKBP6 is essential for spermatogenesis but dispensable for control of meiotic LINE-1 expression in humans', *The American Journal of Human Genetics*, 109(10), pp. 1850–1866. doi: 10.1016/j.ajhg.2022.09.002

Xiang, Y. *et al.* (2007) 'JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 104(49), pp. 19226–19231. doi: 10.1073/pnas.0700735104

Xiol, J. *et al.* (2012) 'A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing', *Molecular Cell*, 47(6), pp. 970–979. doi: 10.1016/j.molcel.2012.07.019

Xu, G.L. *et al.* (1999) 'Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene', *Nature*, 402(6758), pp. 187–191. doi: 10.1038/46052

Yamaguchi, S. *et al.* (2009) 'Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells', *Development*, 136(23), pp. 4011–4020. doi: 10.1242/dev.041160

Yamanaka, S. *et al.* (2019) 'Broad Heterochromatic Domains Open in Gonocyte Development Prior to De Novo DNA Methylation', *Dev. Cell*, 51(1), 21-34.e5.

Yang, F. *et al.* (2008) 'Mouse TEX15 is essential for DNA double-strand break repair and chromosomal synapsis during male meiosis', *The Journal of Cell Biology*, 180(4), pp. 673–679. doi: 10.1083/jcb.200709057

Yang, F. *et al.* (2020) 'TEX15 associates with MILI and silences transposable elements in male germ cells', *Genes Dev*, 34(11-12), pp. 745–750.

Yang, L. and Bennetzen, J.L. (2009) 'Structure-based discovery and description of plant and animal Helitrons', *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), pp. 12832–12837. doi: 10.1073/pnas.0905563106

Yang, Z. *et al.* (2016) 'PIWI Slicing and EXD1 Drive Biogenesis of Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway', *Mol. Cell*, 61(1), pp. 138–152.

Ying, Y. *et al.* (2000) 'Requirement of Bmp8b for the generation of primordial germ cells in the mouse', *Molecular Endocrinology (Baltimore, Md.)*, 14(7), pp. 1053–1063. doi: 10.1210/MEND.14.7.0479

Ying, Y. and Zhao, G.Q. (2001) 'Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse', *Developmental Biology*, 232(2), pp. 484–492. doi: 10.1006/DBIO.2001.0173

Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) 'Cytosine methylation and the ecology of intragenomic parasites', *Trends in Genetics*, 13(8), pp. 335–340. doi: 10.1016/s0168-9525(97)01181-5

Yoshida, S. *et al.* (2006) 'The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage', *Development*, 133(8), pp. 1495–1505. doi: 10.1242/dev.02316

Yoshimizu, T. *et al.* (1999) 'Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice', *Dev. Growth Differ.*, 41(6), pp. 675–684.

Yoshimura, T. *et al.* (2018) 'Mouse GTSF1 is an essential factor for secondary piRNA biogenesis', *EMBO Reports*, 19(4). doi: 10.15252/embr.201642054

Yu, Y. *et al.* (2015) 'Panoramix enforces piRNA-dependent cotranscriptional silencing', *Science*, 350(6258), pp. 339–342.

Yuki, H. *et al.* (1979) 'Determination of 6-methyladenine in DNA by high-performance liquid chromatography', *Journal of Chromatography*, 168(2), pp. 489–494. doi: 10.1016/0021-9673(79)80020-5

Zampieri, M. *et al.* (2009) 'Parp1 Localizes within the Dnmt1 Promoter and Protects Its Unmethylated State by Its Enzymatic Activity', *PLoS One*, 4(3). doi: 10.1371/journal.pone.0004717

Zamudio, N. *et al.* (2015) 'DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination', *Genes and Development*, 29(12), pp. 1256–1270.

Zeller, P. *et al.* (2016) 'Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability', *Nature Genetics*, 48(11), pp. 1385–1395. doi: 10.1038/ng.3672

Zemojtel, T. *et al.* (2007) 'Exonization of active mouse L1s: a driver of transcriptome evolution?' *BMC Genomics*, 8(1), p. 392. doi: 10.1186/1471-2164-8-392

Zeng, C. *et al.* (2023) 'Demethylase-independent roles of LSD1 in regulating enhancers and cell fate transition', *Nature Communications*, 14(1), p. 4944. doi: 10.1038/s41467-023-40606-1

Zhang, H. *et al.* (2021) 'The piRNA pathway is essential for generating functional oocytes in golden hamsters', *Nature Cell Biology*, 23(9), pp. 1013–1022. doi: 10.1038/s41556-021-00750-6

Zhang, H. *et al.* (2022) 'Whole Exome Sequencing Identifies Genes Associated With Non-Obstructive Azoospermia', *Frontiers in Genetics*, 13. doi: 10.3389/fgene.2022.872179

Zhang, H.-Q. *et al.* (2021) 'DNA Methyltransferase 1 Is Dysregulated in Parkinson's Disease via Mediation of miR-17', *Molecular Neurobiology*, 58(6), pp. 2620–2633. doi: 10.1007/s12035-021-02298-w

Zhang, W. *et al.* (2019) 'Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes', *Nucleic Acids Res*, 47(16), pp. 8485–8501. doi: 10.1093/nar/gkz594

Zhang, X. *et al.* (2021) 'Histone demethylase AMX-1 is necessary for proper sensitivity to interstrand crosslink DNA damage', *PLoS Genetics*, 17(7), e1009715. doi: 10.1371/journal.pgen.1009715

Zhang, Y. *et al.* (2017) 'An essential role for PNLDC1 in piRNA 3′ end trimming and male fertility in mice', *Cell Res*, 27(11), pp. 1392–1396. doi: 10.1038/cr.2017.125

Zhao, J. *et al.* (2021) 'Cell-fate transition and determination analysis of mouse male germ cells throughout development', *Nat. Commun.*, 12(1), pp. 1–20.

Zhao, M.-Z. *et al.* (2022) ‘piRNA 3' uridylation facilitates the assembly of MIWI/piRNA complex for efficient target regulation in mouse male germ cells’, *Cell Research*, 32(11), pp. 1030–1033. doi: 10.1038/s41422-022-00659-1

Zheng, K. *et al.* (2010) ‘Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway’, *Proceedings of the National Academy of Sciences of the United States of America*, 107(26), pp. 11841–11846. doi: 10.1073/pnas.1003953107

Zheng, Y., Ahmad, K. and Henikoff, S. (2020) ‘CUT&Tag Data Processing and Analysis Tutorial’.

Zhou, D. and Robertson, K.D. (2016) ‘Role of DNA Methylation in Genome Stability’, in *Genome Stability:* Elsevier, pp. 409–424.

Zhu, Y. *et al.* (2021) ‘Cell cycle heterogeneity directs spontaneous 2C state entry and exit in mouse embryonic stem cells’, *Stem Cell Reports*, 16(11), pp. 2659–2673.

Zickler, D. and Kleckner, N. (2015) ‘Recombination, Pairing, and Synapsis of Homologs during Meiosis’, *Cold Spring Harbor Perspectives in Biology*, 7(6). doi: 10.1101/cshperspect.a016626

ZINDER, N.D. and LEDERBERG, J. (1952) ‘Genetic exchange in Salmonella’, *Journal of Bacteriology*, 64(5), pp. 679–699. doi: 10.1128/jb.64.5.679-699.1952

Zoch, A. *et al.* (2020) ‘SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation’, *Nature*.

Zoch, A. *et al.* (2024) ‘C19ORF84 connects piRNA and DNA methylation machineries to defend the mammalian germ line’, *Molecular Cell*, 84(6), 1021-1035.e11. doi: 10.1016/j.molcel.2024.01.014

CURRICULUM VITAE

SRINIVASA ABISHEK PRAKASH

Doctoral Candidate
Institute of Molecular Biology
Ackermannweg 4
55128, Mainz

Work Experience & Education

PhD in Molecular Biology at the Institute of Molecular Biology, Mainz, Germany (2019-Present)

- Project management expertise, including *in vivo* Mouse work, Mammalian cell line engineering, NGS, Genomics, and Protein biochemistry.
- Robust troubleshooting expertise with an ability to independently design, execute, and evaluate experiments and analyze data, including bioinformatic data analysis of NGS and proteomics.
- Excellent communication skills and cross-functional collaboration experience with internal and international research groups.

Researcher at the Institute of Molecular Infection Biology, Wuerzburg, Germany (2017- 2018)

- RNA biology and regulation expertise through DNA sequence design, *in vitro* transcription, RNA purification and Radioactive labeling, Northern and Southern blot, and EMSA.
- Expertise in cloning, including electroporation, viral transfection, and infectious organism handling in an S2 safety-level lab.

Master's in Biotechnology and Biochemical Engineering & Teaching Assistant from the Indian Institute of Technology, Kharagpur, India (2015 b2017)

- Experience in cloning, recombinant protein expression, and puri^lcation using chromatography techniques (Af^lnity, SEC) and mammalian cell culture.
- Laboratory management and hands-on training experience on ƎCell culture and ImmunoTechnologyƐ and ƎRecombinant DNA technologyƐ topics gained as a Teaching Assistant.

Research Assistant (2014 b2015) at, and **Bachelorƨs in Industrial Biotechnology** (2010 b 2014) from Anna University, Chennai, India

- Bacterial and Yeast cloning experience, upstream process development in shake ^lasks, and lab-scale bioreactors for recombinant protein expression, including scFv antibodies, from *Pichia pastoris.*
- Mentorship experience through guiding junior researchers and high school students.

Technical Skills

Tissue Culture and Microbiology	Mammalian cell line engineering/maintenance, insect cell line maintenance, Transfection using Lipofectamine, CRISPR-Cas and piggy bac-based genome editing, Cell-based cytotoxicity assays, Insect cell culture, competent *E. coli* preparation, Bacterial cloning, Biofilm Assays
Next-Generation sequencing	NGS Library preparation (CUT&RUN, CUT&TAG) and purification, NGS library QC and library pooling, NGS Data analysis (using Galaxy) and visualization.
Analytical Molecular Biology	Pyrosequencing, DNA/RNA quantitation, PCR, quantitative PCR, *in vitro* transcription, Agarose and Polyacrylamide gel electrophoresis, DNA/RNA purification, Northern and Southern blot, Radioactive labelling of RNA, Electrophoretic mobility shift assay.
Proteomics and Protein biochemistry	Recombinant protein/antibody expression and purification using affinity and size exclusion chromatography (SEC), Lab-scale bioreactor handling and monitoring, Proteomic data analysis using MaxQuant and Perseus, AlphaFold, and Pymol.
Immunoassays and techniques	ELISA, Sandwich ELISA, Immunofluorescence, Immunoprecipitation, Multi-color Flow Cytometry and cell sorting (FACS), Western blot (WB).
Microscopy & Imaging	Tissue preparation for cryosection, cryostat sectioning, Microscopy (Light, Fluorescent, and Confocal), Image J analysis
Animal skills	Mouse handling and husbandry, intraperitoneal injection, mouse adult, and mouse embryo dissection, and FELASA Function A certificate.
Bioinformatics & IT skills	Geneious, Snapgene, Sanger sequencing, Basic R programming (data analysis and visualization), Microsoft MS Office Suite (Word, PowerPoint, Excel, Teams, Outlook), Adobe Illustrator, Zoom, Google Workspace.

Non-technical Skills

Communication	Scientific report preparation, Public Speaking, Presentation, Active Listening, Event planning and management.
Leadership	Independent learning, Attention to detail, Organization, Mentorship, Critical thinking, Cross-functional Teamwork, Work ethic, Solution-oriented mindset, Project and people management, Adaptability, Problem Solving, Perseverance, Time Management, and Creativity.

Publications

- **Prakash SA,** Barau J**.** Chromatin Profiling in Mouse Embryonic Germ Cells by CUT&RUN. *Methods Mol Biol.* 2021; 2214:163-176. DOI: 10.1007/978-1-0716-0958-3_17
- Schoenfelder SMK, Lange C, **Prakash SA,** et al. The Small Non-Coding RNA *RsaE* Influences Extracellular Matrix Composition in Staphylococcus epidermidis Biofilm Communities. Vol Is.; 2019. doi:10.1371/journal.ppat.1007618

Af^lliations, Awards & Hobbies

- Volleyball Champion at the Institute of Molecular Biology, Mainz, 2019.
- Recipient of the International PhD Program PhD Fellowship from the Institute of Molecular Biology, Johannes Gutenberg University, Mainz, Germany.
- Recipient of the Graduate School of Life Sciences PhD Fellowship from the University of Wuerzburg, funded by the German Research Foundation DFG Excellence Initiative.
- Institute Silver Medal awardee for the best outgoing postgraduate student from the Indian Institute of Technology, Kharagpur
- Ranked 45th in the All India qualifying exam ∃Graduate Aptitude Test in Engineering∈ (GATE) in 2015 and received the GATE scholarship from the Government of India for postgraduate studies.

Languages

- English: Advanced (C1/C2 level)
- German: Intermediate (A2 level)
- Marathi: Mother Tongue
- Tamil: Native level
- Hindi: Basic (A1 level)

Professional References

Dr. Joan Barau
Group Leader
Institute of Molecular Biology
55128 Mainz, Germany
J.Barau@imb-mainz.de

Prof. Dr. René Ketting
Scientific Director
Institute of Molecular Biology
55128 Mainz, Germany
R.Ketting@imb-mainz.de

Mainz, September 2024 / Srinivasa Abishek Prakash