Orchestrating transposon regulation in the male germline – The role of DNA methylation, histone marks and transcription factors

Dissertation
Zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Biologie
Der Johannes Gutenberg-Universität Mainz

Jessica Leismann
geboren am 09.06.1995 in Rüsselsheim am Main

Mainz, 2025

Dekan: Prof. Dr. Eckhard Thines

1. Berichterstatter: Prof. Dr. Rene Ketting

2. Berichterstatter: Dr. Claudia Keller-Valsecchi

Tag der mündlichen Prüfung: 08. Mai 2025

Table of Contents

Zusammenfassung

Transposons sind mobile DNA-Sequenzen, die in allen eukaryotischen Genomen vorkommen und einen erheblichen Anteil der Säugetiergenome ausmachen. Während sie die evolutionäre Innovation vorantreiben, stellt ihre Fähigkeit zur Replikation eine ernsthafte Bedrohung für die genomische Stabilität dar. In somatischen Geweben werden Transposons effektiv durch DNA-Methylierung stillgelegt. Während der Entwicklung von Säugetieren ist die DNA-Methylierung jedoch dynamisch und durchläuft eine Reprogrammierung. Eine Reprogrammierungswelle in der Keimbahnentwicklung löscht die somatischen Methylierungsmuster, was die Aktivierung von Keimbahn-spezifischen Genen ermöglicht, die für die Entwicklung essenziell sind. Gleichzeitig öffnet sich dadurch jedoch ein mögliches Zeitfenster, für Transposons aktiv zu werden.

In männlichen Keimzellen wird dies durch den piRNA Pathway entgegengewirkt. Hierbei werden transposon-abgeleitete mRNAs in piRNAs verarbeitet, die die *de novo* DNA-Methylierung an den Promotoren junger, aktiver Transposons steuern. Dieser Prozess wird durch die männlich-spezifische Methyltransferase DNMT3C und ihren Kofaktor DNMT3L vermittelt. Zwischen dem piRNA Pathway und den Transposons besteht ein evolutionäres Wettrüsten: Während die Stilllegungs-Maschinerie ständig weiterentwickelt wird, um neu aktive Tranpsosons zu erkennen und zu unterdrücken, müssen sich Transposons dieser Kontrolle entziehen, um sich zu vermehren. Ein Versagen des piRNA Pathways führt zu einer Hochregulierung von Transposons, was zu fehlerhafter Chromosomenpaarung, meiotischem Arrest und schließlich männlicher Sterilität führt. Interessanterweise treten diese Spermatogenesedefekte erst mehrere Tage nach dem Fehlschlagen der *de novo* Methylierung in fetalen Keimzellen auf, was darauf hindeutet, dass zusätzliche Regulationsmechanismen Transposons in Abwesenheit von DNA-Methylierung kontrollieren.

Um die Regulation von Transposons in Abwesenheit von DNA-Methylierung zu untersuchen, analysierten wir zunächst die Transposon-Expression in *Dnmt3C*$^{KO/KO}$ Mutanten über sortierte Keimzellstadien hinweg. Entgegen früherer Berichte fanden wir, dass die jungen, aktiven Tranposons LINE1s und IAPs bereits vor der Meiose aktiv sind, wenn DNA-Methylierung fehlt und die LINE1-Expression in der Meiose weiter ansteigt. Anschließend untersuchten wir die mit der Transposon-Regulation assoziierten Histonmodifikationen und stellten dynamische Histonmodifikations-Signaturen an Transposons in Abwesenheit von DNA-Methylierung während der Spermatogenese fest. Insbesondere identifizierten wir bivalente Histonmodifikationen: H3K4me3-H3K9me3 an LINE1-Elementen in fetalen Keimzellen vor der piRNA-vermittelten DNA Methylierung und H3K4me3-H3K27me3 an Transposons vor der Meiose in *Dnmt3C*$^{KO/KO}$ Mutanten. Die bivalente H3K4me3-H3K27me3 Markierung könnte die Transposon-Expression bevor der

Meiose in einem zur Expression-bereitem Zustand halten und gleichzeitig vor DNA-Methylierung schützen. Zu Beginn der Meiose ging diese Markierung in einen offeneren Chromatinzustand über, was mit einer starken Transposon-Expression in der Abwesenheit von DNA-Methylierung korrelierte.

Um die Transposon-Regulation weiter zu erforschen, führten wir einen unvoreingenommenen proteomischen Pulldown-Screen durch, um Faktoren zu identifizieren, die Transposon-Promotoren in einem DNA-Methylierungs-abhängigen Kontext binden. Dabei identifizierten wir NRF1, einen DNA-methylierungssensitiven Transkriptionsfaktor, der für die Regulation von Keimbahn-Genen bekannt ist, als spezifischen Binder von unmethylierten Transposon-Promotoren in männlichen Keimzellen. Um NRF1 und andere an Transposon-bindende Faktoren zu charakterisieren, optimierten wir Chromatin-Profiling-Methoden in niedrigem Input. Diese Expertise ermöglichte uns Kooperationen mit anderen Laboren, in denen wir zeigten, dass der Chromatinleser SPIN1 die H3K4me3-H3K9me3-Signatur an LINE1-Elementen in fetalen Keimzellen erkennt und damit die SPOCD1-Rekrutierung sowie die piRNA-Weg-vermittelte Stilllegung von Transposons unterstützt. Zudem konnten wir zeigen, dass NRF1 in postnatalen Keimzellen an unmethylierte LINE1- und IAP-Elemente bindet, und die Bindung in der Meiose verstärkt ist.

Schließlich konnten wir mit einem konditionalen NRF1-Knockout in *Dnmt3C*$^{KO/KO}$ Mäusen nachweisen, dass NRF1 unmethylierte IAP-Elemente vor der Meiose aktiviert, während unsere Daten auf eine mögliche zusätzliche Rolle bei der LINE1-Regulation in der Meiose hindeuten.

Insgesamt zeigen wir, dass eine bivalente Chromatin-Signatur sowie NRF1 als ein DNA-methylierungssensitiver Transkriptionsfaktor die Transposon-Regulation in männlichen Keimzellen dirigieren, wenn DNA-Methylierung an ihren Promotoren fehlt. Unsere Erkenntnisse liefern neue Einblicke, wie Tranposons sich in Genomen ausbreiten mit Implikationen auf Keimzell-genomische Instabilität, und tragen zu einem besseren Verständnis der Interaktion zwischen epigenetischen Mechanismen und Transkriptionsfaktoren in der Genregulation bei.

Abstract

Transposable elements (TEs) are mobile DNA sequences, which can proliferate and change their position in the genome. They are part of all eukaryotic genomes and make up a significant portion of mammalian genomes. While they drive evolutionary innovation, their ability to replicate poses a major threat to genome stability. In somatic tissues, transposons are effectively silenced by DNA methylation. However, during mammalian development, DNA methylation is dynamic and undergoes reprogramming. The germline reprogramming wave erases somatic methylation patterns, allowing the activation of germline genes that are essential for development. However, this also creates a window of opportunity for TEs to become active.

In male germ cells, this is counteracted by the piRNA pathway, which processes TE-derived mRNA into piRNAs that guide *de novo* DNA methylation to the promoters of young, active TEs. This process is mediated by the male rodent-specific methyltransferase DNMT3C and its cofactor DNMT3L. The piRNA pathway and TEs are engaged in a continuous evolutionary arms race: while the silencing machinery evolves to suppress newly active TEs, TEs must evade this suppression to proliferate. Failure of the silencing machinery leads to TE upregulation, resulting in defective chromosome pairing, meiotic arrest, and ultimately male sterility. Notably, these spermatogenic defects arise several days after the failure of *de novo* methylation in fetal germ cells, suggesting that additional regulatory mechanisms govern TEs in the absence of DNA methylation.

To investigate TE regulation in the absence of DNA methylation, we first analyzed TE expression in *Dnmt3C*$^{KO/KO}$ mutants across sorted germ cell stages. Contrary to previous reports, we found that the young, active TEs, LINE1s and IAPs, become already active before meiosis in the absence of DNA methylation, while LINE1 expression further increases at meiosis. Next, we examined histone modifications associated with TE regulation in the absence of DNA methylation, revealing dynamic chromatin signatures during spermatogenesis. Specifically, we identified bivalent chromatin marks at unmethylated TE promoters: H3K4me3-H3K9me3 at LINE1 elements in fetal germ cells before DNMT3C-mediated methylation, and H3K4me3-H3K27me3 at TEs before meiosis in *Dnmt3C*$^{KO/KO}$ mutants. The bivalent H3K4me3-H3K27me3 mark may poise TE expression before meiosis while protecting them from DNA methylation. At the onset of meiosis, this mark transitioned to a more open chromatin state, correlating with high TE expression when DNA methylation is not in place.

To further explore TE regulation, we conducted an unbiased proteomic pulldown screen to identify factors binding TE promoters in a DNA methylation-dependent manner. Among other interesting candidates, we found NRF1, a DNA methylation-sensitive transcription

factor, known to regulate germline genes, to specifically bind unmethylated TE promoters in male germ cells. To characterize NRF1 and other factors binding at TEs, we explored and optimized low-input chromatin profiling methods. This expertise enabled collaborations with other laboratories, where we demonstrated that the chromatin reader SPIN1 binds the bivalent H3K4me3-H3K9me3 signature at LINE1 elements in fetal germ cells, facilitating SPOCD1 recruitment and DNMT3C-mediated silencing. Furthermore, we showed that NRF1 binds LINE1s and IAPs in postnatal germ cells when DNA methylation is absent at their promoters, with increased binding at meiosis.

Finally, we tested a causative correlation, using a conditional NRF1 knockout in *Dnmt3C*$^{KO/KO}$ mice, and demonstrated that NRF1 activates unmethylated IAP elements before meiosis, while our data suggest a potential additional role in LINE1 regulation at meiosis.

Overall, we show that a bivalent chromatin signature and NRF1 as a DNA methylation-sensitive transcription factor orchestrate TE regulation in male germ cells when DNA methylation is absent at TE promoters. Our findings shed light on how TEs expand in genomes threatening germline genomic integrity and provide broader insights into how epigenetic mechanisms and transcription factors interact to regulate gene expression.

List of Figures

List of Tables

List of Abbreviations

ABC	Ammonium Bicarbonate
ac	Acetylation
ACN	Acetonitrile
AME	Analysis of Motif Enrichment
ADD	ATRX-DNMT3-DNMT3L domain
ATAC-Seq	Accessible Chromatin with high-throughput sequencing
ChIP	Chromatin Immunoprecipitation
CUT&Run	Cleavage Under Targets & Release
CUT&Tag	Cleavage Under Targets & Tagmentation
DAP-Seq	DNA affinity purified sequencing
DBD	DNA binding domain
dKO	double knock-out: *Nrf1*$^{cKO/KO}$*; Dnmt3C*$^{KO/KO}$
DNMTs	DNA methyl transferases
DNMT3L	DNMT3-like enzyme
DSBs	Double strand breaks
E	Embryonic day
ERVs	Endogenous retroviruses
ERVKs	ERV-K promoter
ERVLs	ERV-L promoter
F	Offspring Generation
FBS	Fetal Bovine Serum
FC	Fold Change

H	Histone
H3K4	Histone H3 Lysine 4
H3K9	Histone H3 Lysine 9
H3K27	Histone H3 Lysine 27
HP1	Heterochromatin Protein 1
HUSH	Human Silencing Complex
IAP	Intracisternal A-type Particle
IP	Immunoprecipitation
IP-MS	Immunoprecipitated Pulldown followed by Mass-Spectrometry
KRAB-ZFPs	Krueppel-associated box domain containing zinc finger proteins
L1s	LINE1s
LINEs	Long Interspersed Nuclear particle
Log	Logarithmic transformation to the base of the mathematical constant e
LTRs	Long terminal repeats
MI	Meiosis I
MII	Meiosis II
MaLR	Mammalian-apparent LTR retrotransposons
MBD	Methyl binding domain proteins
mESCs	mouse Embryonic Stem Cells
me2	dimethylation
me3	trimethylation
MLV	Mouse Leukemia Virus
MNase	Micrococcal Nuclease

NGS	Next Generation Sequencing
NRF1	Nuclear Respiratory Factor 1
ORFs	Open Reading Frames
ORF1p	Open Reading Frame 1 protein
P	Postnatal day
PCNA	Proliferating cell nuclear antigen
PFA	Paraformaldehyde
PGCs	Primordial Germ Cells
PIC	Protease Inhibitor Cocktail
piRNAs	PIWI interacting RNAs
PIWI	PIWI-clade Argonaute
POI	Protein of Interest
Pol II	RNA polymerase II
Pol III	RNA polymerase III
PRC2	Polycomb Repressive Complex 2
PTGS	Post-transcriptional Gene Silencing
PWWP	Pro-Trp-Trp-Pro
qPCR	quantitative PCR
ROI	Region of Interest
RT	Room Temperature
SAM	S-Adenosylmethionine
SINEs	Short Interspersed Nuclear Elements
Spc	Spermatocytes

Spg	Spermatogonia
SSCs	Spermatogonial Stem Cells
SVA	SINE-VNTR-Alu
TAD	Topologically-Associated Domain
TEs	Transposable Elements
TET	Ten Eleven Translocation Enzymes
TFs	Transcription Factors
TGS	Transcriptional Gene Silencing
TKO	*Dnmt3A*$^{KO/KO}$*; Dnmt3B*$^{KO/KO}$*; Dnmt1*$^{KO/KO}$ mESCs
TIRs	Terminal Inverted Repeats
TPM	Transcripts per Million
TSDs	Tandem Site Duplications
TSS	Transcription Start Site
UHRF1	Ubiquitin-like containing PHD and RING finger domains, 1
ULI-ChIP	Ultra Low Input ChIP
UMIs	Unique Molecular Identifiers
UTR	Untranslated Region
WT	Wild-type
x	Crossed

1. Introduction

1.1. Transposable Elements (TEs)

Transposable Elements (TEs) or transposons, were first discovered and characterized in maize as ‘jumping genes’ by Barbara McClintock (McCLINTOCK 1950). Contrary to the prevailing scientific belief at the time, that genes and genetic elements were stably ordered on chromosomes (The mechanism of Mendelian heredity 1916), McClintock found that certain loci in the maize genome could change their position. She further found that this mobility could regulate the expression of pigment genes, resulting in mosaic coloration of maize kernels. Her groundbreaking finding was long met with skepticism and only several decades later McClintock was awarded the Nobel prize for this discovery in 1983, after similar findings were reported in other species such as bacteria and *Drosophila melanogaster* (Shapiro 1969; Kidwell et al. 1977).

TEs are mobile genetic elements, which can excise, create a new copy of themselves and change their genomic position. They comprise a big portion of almost all eukaryotic genomes (Wells and Feschotte 2020) with some species having up to 85% of their genome occupied by these elements (Wegrzyn et al. 2013). In mammals, TEs account for approximately half of the genome (Fueyo et al. 2022).

1.1.1. Biology of TEs

Most TEs are autonomous, meaning they encode all proteins, which are necessary for their own mobilization. However, non-autonomous elements, such as *Alu* elements, exist in all classes. These elements rely on the transposition machinery of autonomous elements to facilitate their movement (Wells and Feschotte 2020).

Based on the way they mobilize, called their transposition mechanisms, TEs are divided into two major classes: Class I elements, also known as RNA transposons or retrotransposons use the so-called ‘copy-and-paste mechanism’ by which they create an RNA copy of themself, which is retro-transcribed into a new location in the genome. Class II elements, also referred to as DNA transposons, use the so-called ‘cut-and-paste’ or ‘peel-and-paste’ mechanism without an RNA intermediate (Finnegan 1989). These classes are further subdivided into subclasses according to the mechanism of their genomic integration. Retrotransposons are subdivided into three main subclasses: Long Terminal Repeat (LTR) retrotransposons, Non-LTR retrotransposons and Tyrosine Recombinase mobilized elements. These subclasses are further classified into subgroups based on the genetic organization and into families (**Figure 1.1**) (Bourque et al. 2018). The following chapters will briefly describe the classes, subclasses and families of TEs in

mice and humans with a specific focus on retrotransposons, which are the only active class of TEs in humans and mice. More specifically it focuses on the most active and evolutionary young TEs in mice, namely Intracisternal A-type Particle elements, which are LTR retrotransposons, and Long interspersed nuclear element 1 elements.

Figure *1.1*

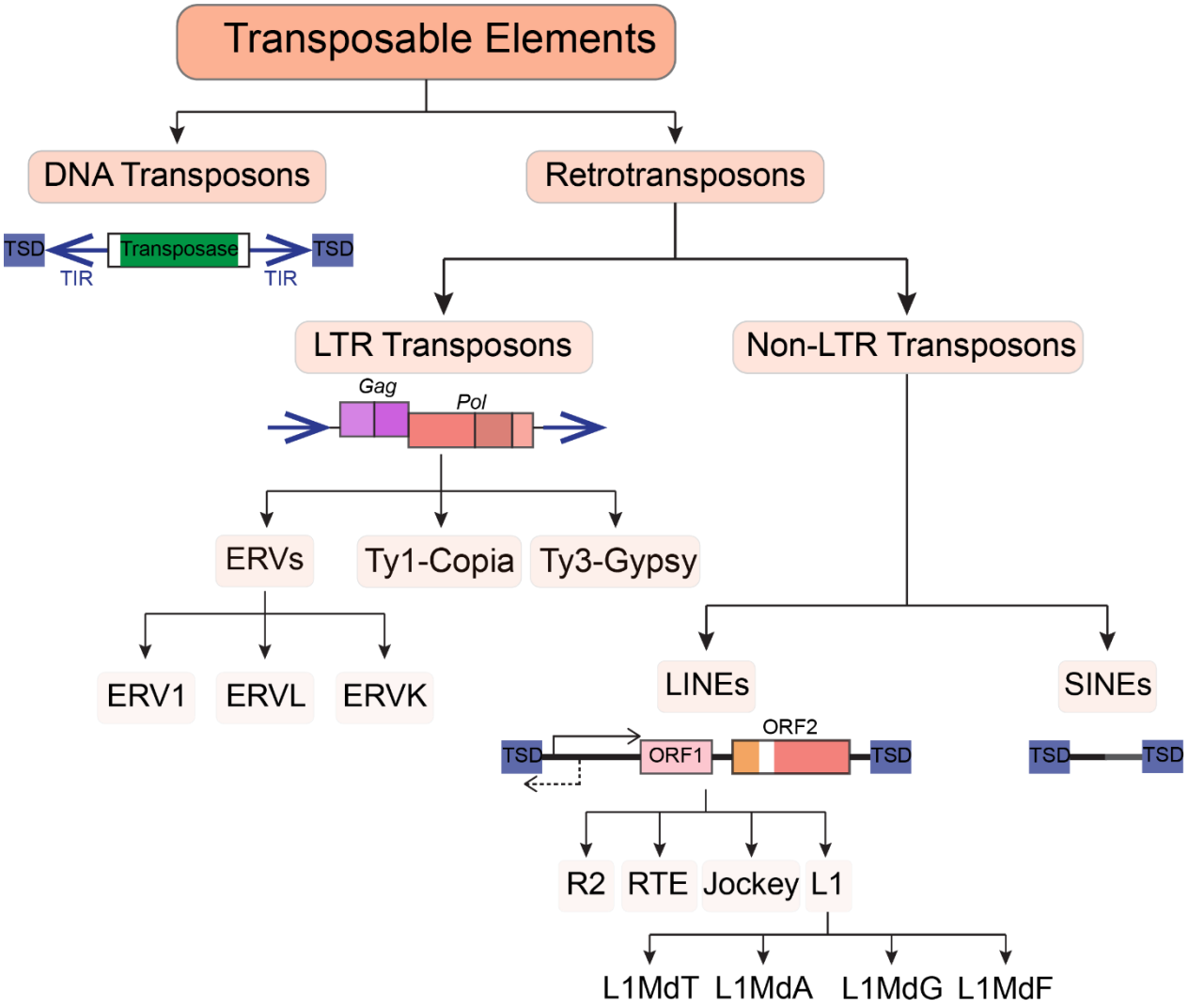

Figure 1.1: Classification and Structure of Transposable Elements. Different classes, subclasses and families of retro-transposons. Structure of DNA Transposons, LTR retroransposons and the Non-LTR retrotransposons, LINEs and SINEs.

1.1.1.1. DNA Transposons

DNA transposons account for approximately 3% of mammalian genomes. They move without an RNA-intermediate, most of them moving through a non-replicative 'cut-and-

paste' mechanism, excising from one location and inserting into another, via their encoded transposase (Greenblatt and Alexander Brink 1963). The transposase recognizes two terminal inverted repeats (TIRs) flanking its domain, enabling the excision of the TE. Upon integration into a new locus, the target site is duplicated, creating target site duplications (TSDs), a hallmark feature of DNA transposons and their remnants (**Figure 1.1**). However, some DNA transposons, such as Helitrons, use a different transposition mechanism called 'cut-and-peel', involving replicated DNA as an intermediate step (Grabundzija et al. 2016). In mice and human genomes, DNA transposons are inactive (Lander et al. 2001), therefore the following chapters focus on retrotransposons.

1.1.1.2 LTR retrotransposons

LTR Transposons are a subclass of retrotransposons that constitute approximately 10% of mouse and human genomes (Jern and Coffin 2008). They vary in length between 100 Bp to 25kB. Many evidences suggest that LTR elements originated from ancient infections by exogenous retroviruses that became endogenized, such as the Koala retrovirus, which is currently invading animal germlines and is incorporated into the genome (Magiorkinis et al. 2012; Löber et al. 2018). However, most LTR elements have lost the ability to form infectious particles and can therefore only replicate within the host genome. LTR Transposons are the most mutagenic TEs in mice, where their transposition events account for 10-15% of all inherited mutant phenotypes (Maksakova et al. 2006), while they remain mainly inactive in humans (Mills et al. 2007). Nevertheless, many LTR retrotransposons lack a functional open reading frame (ORF) and therefore depend on mobilization in *trans*. In addition, regulatory LTR sequences often exist independently as "solo LTRs" without any coding region, likely arising from homologous recombination between two LTRs of a TE (Akopov et al. 1998). Functional LTR retrotransposons are characterized by the presence of two LTR regulatory sequences, flanking the coding region, which are typically between 300 and 100 Bp in size. The coding region of autonomous LTR retrotransposons encompasses at least two genes important for their retrotransposition: *gag* and *pol* (**Figure 1.1**), ranging in size from 6,000 to 9,000 bp. *Gag* encodes a virus-like particle that is formed in the cytoplasm while *Pol* encodes for a reverse transcriptase, RNaseH and integrase (Sandmeyer and Clemens 2010; Wicker et al. 2007).

To retrotranspose, first the LTR elements get transcribed by the host RNA polymerase II, then the transcript gets 3' polyadenylated and 5' capped and exported to the cytoplasm. Subsequently, their bicistronic transcript functions as a template for both novel TE insertions as well as translation. The RNA gets reverse-transcribed into DNA by its encoded POL protein inside virus-like particles, which are assembled by its encoded GAG-protein. The reverse-transcribed DNA is then transported back into the nucleus and

gets eventually inserted in a new genomic location with the help of its encoded integrase protein (Hughes 2015).

The three main superfamilies of LTR retrotransposons are *Gypsy*, *Copia* and Endogenous Retroviruses (ERVs) (**Figure 1.1**) (Wicker et al. 2007). In mammals, only ERVs are present, comprising around 12% of the genome (Stocking and Kozak 2008). ERVs are evolutionarily closely related to retroviruses resembling their structure and replication mechanisms and often carry the *Env* gene that encodes for the viral envelope protein. Due to their origin from exogenous retroviruses, ERVs are highly diverse and are further categorized into three classes based on their ancestral retrovirus family (**Figure 1.1**).

Class I ERVs, also referred to as ERV1, represent around 1% of the mouse genome with only a few full-length copies. They cluster with Epsilon and Gamma-retroviruses. Prominent examples include the mouse leukemia virus (MLV) family or RLTR6 (Gifford et al. 2005; Stocking and Kozak 2008). Class II ERVs or ERV-K promoter (ERVKs) constitute around 3% of the mouse genome and cluster with lentivirus, alpha-, beta- and delta-retroviruses (Gifford et al. 2005). A well-characterized clade of ERVKs is the Intracisternal A-type Particle (IAP), which has around 1000 copies in mice and is the most active mutagenic class (Dewannieux et al. 2004; Kuff and Lueders 1988). Class III ERVs, include two subclasses, the ERV-L promoter (ERVLs) and the non-autonomous Mammalian-apparent LTR retrotransposons (MaLRs) and cluster with spuma viruses. Together, they form the largest ERV class, accounting for ~5% of the mouse genome. MaLRs lack ORFs and consist only of repetitive, non-coding DNA, whereas ERVLs retain retrotransposition potential. Despite their ancient origins, ERVLs remain active in the mouse genome (Bénit et al. 1999).

Overall, the human and mouse genomes differ significantly in ERV distribution and family diversity. While in humans, LTR retrotransposons are nearly extinct with the potential exception of HERVK-HML2 (Medstrand and Mager 1998), LTR elements remain the most active TEs in mice.

1.1.1.3 Non-LTR Retrotransposons

Non-LTR retrotransposons are defined by the lack of LTRs and their retrotransposition mechanism, called target primed reverse transcription. In this process, an endonuclease nicks one strand of the DNA target site leaving a 3’ hydroxyl end, which serves as a primer for reverse transcription of the TE mRNA directly onto the DNA target (Luan et al. 1993). Non-LTR retrotransposons include two subtypes, the autonomous Long Interspersed Nuclear Elements (LINEs) and the non-autonomous Short Interspersed Elements (SINEs) (Roy-Engel 2012) (**Error! Reference source not found.**), which are explained in more detail in the following chapters.

1.1.1.3.1. Long Interspersed Nuclear Elements

LINEs are autonomous non-LTR retrotransposons of 4-7kB length that encode 20% of mammalian genomes (Lander et al. 2001). They harbour at least an ORF2p encoding a Reverse Transcriptase and an Endonuclease domain. LINEs comprise five superfamilies, R2, RTE, Jockey, LINE1 (L1) and I based on the divergence of the Reverse Transcriptase domain. Apart from R2 and RTE superfamilies, LINEs encode for another protein called open reading frame 1 protein (ORF1p) that has nucleic acid binding activity and is important for their retrotransposition. (**Figure 1.1**) (Kapitonov et al. 2009; Malik et al. 1999). After LINEs are transcribed by the host RNA polymerase II, the mRNA is exported to the cytoplasm and gets translated into ORF1p and ORF2p (Goodier et al. 2007). ORF1p and ORF2p together with the L1 mRNA form a Ribonucleoprotein complex in *cis* that is transported back into the nucleus for target primed reverse transcription and genomic integration (Babushok and Kazazian 2007).

L1s represent 20% of the mouse genome and 18% of the human genome, with approximately 500,000-600,000 copies of which only 3,000 copies and 100 copies are estimated to be mobile in mice and humans, respectively (Goodier et al. 2001). In humans, LINE1s are the only active autonomous retrotransposons. Full-length L1s consist of a 5' Untranslated Region (UTR), ORF1p, ORF2p, a potential ORF0 and a 3' UTR and span around 7kB in size. Interestingly, the 5' UTR in mice unlike in humans consists of CG-rich tandemly repeated monomers that vary in number. The 5'UTR includes an internal promoter driving L1 mRNA transcription and an antisense promoter that can regulate the expression of neighbouring genes (Li et al. 2014). ORF0 has only been shown in primates and while it was shown to stimulate retrotransposition its exact function remains unclear (Denli et al. 2015). The families are classified into groups based on their promoter type: L1MdA and L1MdF groups are evolutionary young groups, which were recently active, therefore mostly intact and L1MdV groups are older or extinct. Elements from L1MdA groups gave further rise to new promoters as L1MdTf and L1MdGf elements (Goodier et al. 2001, Sookdeo et al. 2013).

1.1.1.3.2. Short Interspersed Nuclear Elements

SINEs are 100-500 bp long non-autonomous retrotransposons that do not encode proteins (**Figure 1.1**). Although SINEs are found across eukaryotes, they are particularly successful in mammals, occupying 5-15% of their genomes. In contrast to other TEs, which are transcribed by RNA polymerase II, SINEs bear an RNA-polymerase III (Pol III) promoter that allows for their autonomous transcription. Thus, the origin of SINEs is believed to stem from accidental reverse transcription of Pol III transcripts (Kramerov and Vassetzky 2005). For replication and mobility, SINEs hijack the machinery of autonomous LINEs using the same retrotransposition mechanism as LINEs (Dewannieux et al. 2003).

Well known examples of SINEs are SINE-VNTR-Alu (SVA), Alu, B1 and B2 elements. The SVA elements are among the most active and youngest TEs in humans, which are only present in hominoid genomes (Ostertag et al. 2003). In mice, B1 and B2 are the most abundant elements of the SINE family (Kass and Jamison 2007).

1.1.2 Implications and Roles for TEs

The activity of TEs poses an immediate threat to genome integrity and stability due to their mutagenic potential. However, TEs can also drive genetic diversity, contributing to long-term evolutionary innovation (**Figure 1.2**). To mitigate the risks, eukaryotes have developed various defense mechanisms, engaging in a constant evolutionary arms race with TEs, where TEs adapt to be expressed and defense mechanisms adapt to this to repress TEs (Slotkin and Martienssen 2007).

When TEs are inactive and not under positive selection, they accumulate point mutations, leading to their inactivation over time (Lynch 2007). Consequently, most TEs in genomes are no longer functional. For instance, only about 100 out of 500,000 L1 copies in the human genome remain potentially active (Brouha et al. 2003). Despite this, horizontal transfer of TEs can occur across all major TE types, enhancing their evolutionary success (Gilbert and Feschotte 2018; Wallau et al. 2018). Thus, in order to be successful, TEs have to balance their expression to proliferate without severely compromising the host fitness (Bourque et al. 2018).

1.1.2.1. Implications of TEs in Evolution

Previously, TEs have only been considered junk and selfish genetic parasites, although McClintock had already hypothesized their role in modulating gene expression early on (Mcclintock 1956). Recent studies more and more highlight the beneficial contributions of TEs to biological innovation.

TE insertions can for instance drive protein-coding gene evolution through transposon domestication or alternative splicing events (Naville et al. 2016). A notable example of transposon domestication is the ERV *env* gene, which evolved into syncytins facilitating cell fusion in placentation across mammalian lineages (Cornelis et al. 2015; Cornelis et al. 2017). Similarly, *Rag1* and *Rag2*, which are important for the adaptive immune system, are derived from domesticated transposases (Kapitonov and Jurka 2005).

TEs also impact alternative splicing, as intronic Alu elements can serve as alternative exons via cryptic splice sites (Lev-Maor et al. 2008). Additionally, TEs contribute cis-regulatory elements, acting as promoters, enhancers, insulators, or transcription factor binding sites (Bejerano et al. 2006; Bourque et al. 2008; Schmidt et al. 2012; Sundaram

et al. 2014). A well-known example is the Agouti mouse coat color, where the methylation of an upstream IAP promoter, forming a metastable epiallele, influences the expression of the *Agouti* gene resulting in different coat colors depending on the methylation level of the IAP promoter (Waterland and Jirtle 2003; Jirtle and Skinner 2007) (**Figure 1.2A**). Moreover, the transposition of a DNA transposon into the *cortex* gene of the peppered moth turned the originally light-colored moths into dark-colored moths, which provided an evolutionary advantage during the Industrial Revolution by making them less visible to predators and driving this evolutionary adaptation (Hof et al. 2016) (**Figure 1.2B**).

Lastly, in some tissues TE expression is essential. For instance, in the preimplantation embryo at the two-cell-stage MERVL upregulation is required to regulate the expression of zygotic genome activation (ZGA) genes (Sakashita et al 2023). Moreover, the expression of LINE1 elements is also important during ZGA regulating chromatin accessibility and organization, as well as active in the healthy human brain (Goodier 2014, Jachowicz et al. 2017).

Figure *1.2*

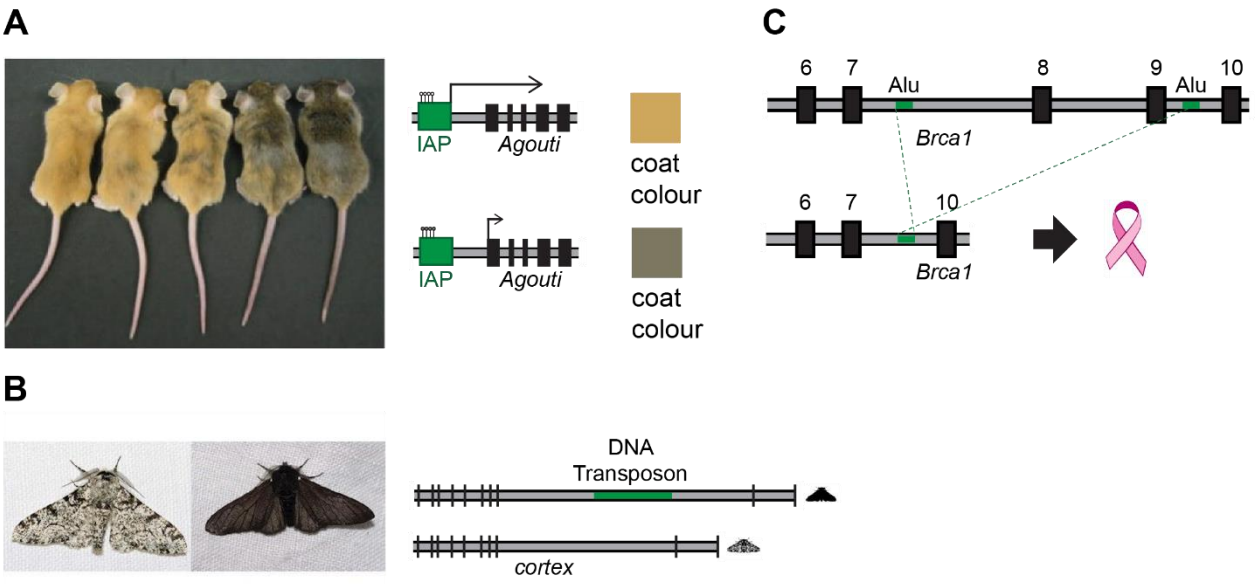

Figure 1.2: Examples of TE Implications on gene expression, evolution and disease. A) Example of the control of gene expression by a TE. The methylation status of an IAP promoter affects expression of the *Agouti* gene, resulting in different coat colours in mice (Jirtle and Skinner, 2007). B) Example of a TE driving evolution. Insertion of a DNA transposon into the *cortex* gene in the British peppered moth, which is responsible for the usual sprinkled-coloured phenotype, leads to transition to a dark-coloured phenotype, which was an evolutionary advantage in the industrial times (Hof et al., 2016). C) Example of TEs leading to a disease phenotype. Recombination of two Alu Elements in the *Brca1* gene locus leads to mutagenesis resulting in breast cancer in humans (Rohlfs et al., 2000).

1.1.2.2. Implications of TEs in Disease

To propagate into future generations, it would be of advantage for TEs to be expressed in the germline, where new TE insertions would be inherited (Haig 2016). A quite fascinating example for mastering this challenge is in ciliates, where TEs are excised from somatic macronuclei and are only retained in germline micronuclei (Arnaiz et al. 2012). However, in mammals, TEs are not only active in germ cells, but also exhibit activity in somatic cells. Their activity in somatic cells was found to be beneficial in early embryos, stem cells, and the brain, where L1 activity may enhance neuronal plasticity or harmful such as in ageing or cancer (Garcia-Perez et al. 2007; Baillie et al. 2011; Klawitter et al. 2016; Kazazian and Moran 2017).

Thus, while a balanced somatic TE activity is required and beneficial in some cases, aberrant TE expression is linked to several diseases. High L1 activity and *de novo* insertions into tumor suppressor or oncogene regions are associated with cancer, alongside TE promoter derepression that activates oncogenes (Iskow et al. 2010; Lee et al. 2012; Babaian and Mager 2016). A well-known example is the mutation of *Brca1* due to recombination of two Alu elements leading to the deletion of several exons and resulting in breast cancer in humans (Rohlfs et al. 2000) (**Figure 1.2C**). Additionally, LINE-1 ORF1p expression has emerged as a biomarker in many epithelial cancers, including ovarian carcinoma (Ardeljan et al. 2020). In addition to implications in cancer, cytotoxic ENV protein overexpression has been implicated in neurodegenerative diseases such as multiple sclerosis and amyotrophic lateral sclerosis, while upregulated LINE1 expression in the brain drives the Alzheimer disease (Perron et al. 2001; Li et al. 2015, Roy et al 2024). Failure to silence TEs and their resulting upregulation has also been shown to lead to male infertility, both in mice and humans (Barau et al. 2016, Stallmeyer et al. 2024). Finally, TE activity can also cause chromosomal rearrangements, including deletions, duplications, and inversions, through homologous recombination (Ade et al. 2013; Bennetzen and Wang 2014). Therefore, robust defense mechanisms are crucial to controlling TE activity in host genomes.

1.2. TE defense and control mechanisms in mammals

Defense mechanisms against TE expression and activity are not only important for functional TEs that could mobilize, but also for mutated TEs that could perturb regulation of other important genes. A variety of defense mechanisms has evolved to act on a transcriptional and post-transcriptional level. This chapter serves as an overview of the existing defense mechanisms in mammals, or more specifically in mice, with a focus on DNA methylation, histone marks and the piRNA pathway.

Many targeted mechanisms are involved in a constant arms race against TE invasion, where the mechanism adapt to recognize TEs and TEs in turn adapt to evade these mechanisms, which I will highlight in this chapter. I will briefly mention other mechanisms that help to protect the genome integrity potentially serving as a form of last resort: The innate immune system, which recognizes viral RNA, gets triggered by retrotransposon mRNA in the cytoplasm activating the type I interferon response and ultimately leading to cell death (Gazquez-Gutierrez et al. 2021). Moreover, non-sense mediated decay and degradation of TE mRNA has been shown in stress granules for L1s and in processing bodies for IAPs (Goodier et al. 2007; Lu et al. 2011; Mendell et al. 2004).

1.2.1. DNA methylation

The DNA sequence can be directly modified at the nucleotide level by the addition of a methyl group at the fifth position of the cytosine aromatic ring. This epigenetic modification is called 5-Methylcytosine, or hereafter referred to as DNA methylation (Johnson and Coghill 1925; Wyatt 1950). DNA methylation at enhancers and promoters is associated with transcriptional silencing, representing a “locked” state of transcription. Conversely, DNA methylation within gene bodies is often associated with transcriptional activation (Zemach et al. 2010).

Repression mediated by DNA methylation is crucial for various molecular processes, including X-chromosome inactivation (Mohandas et al. 1981), germline gene inactivation in somatic cells (Borgel et al. 2010) and genomic imprinting (Li et al. 1993). However, perhaps its most ancient and critical function is the control and silencing of TEs. It has even been suggested that the evolution of DNA methylation was driven by the need to suppress TEs (Yoder et al. 1997).

In plants and other organisms, DNA methylation occurs in three sequence contexts - CG (or CpG), CHG and CHH (where H refers to A, T or C). In mammals, however, DNA methylation is almost exclusively found in the CpG dinucleotide context, with symmetric methylation on the cytosines of both strands. DNA methylation appears to occur by default at all CpG sites unless specific mechanisms exclude it (Stadler et al. 2011). Approximately 70-80% of CpGs are methylated in somatic cells (Lister et al. 2009). However, high levels of DNA methylation come at a cost, increasing the likelihood of spontaneous mutations: Cytosine deamination frequently converts cytosine to uracil, which is recognized and removed by base-excision repair mechanisms. In contrast, deamination of 5-methylcytosine produces a thymine, which is a standard DNA nucleotide and therefore not recognized by repair machinery and retained instead (Walsh and Xu 2006; Bird 1980). As a result, CpG dinucleotides are globally depleted in mammalian genomes with the exception of CpG-rich islands, which are often located in gene promoter regions and TEs (Bird 1986; Illingworth et al. 2010; Jeltsch et al. 2018).

1.2.1.1. DNA methyltransferases

The DNA methyltransferase (DNMT) family of enzymes are primarily responsible for establishing DNA methylation by transferring a methyl group from an S-adenosylmethionine (SAM) donor to cytosine (**Figure 1.3Figure *1.3***

A).

Figure *1.3*

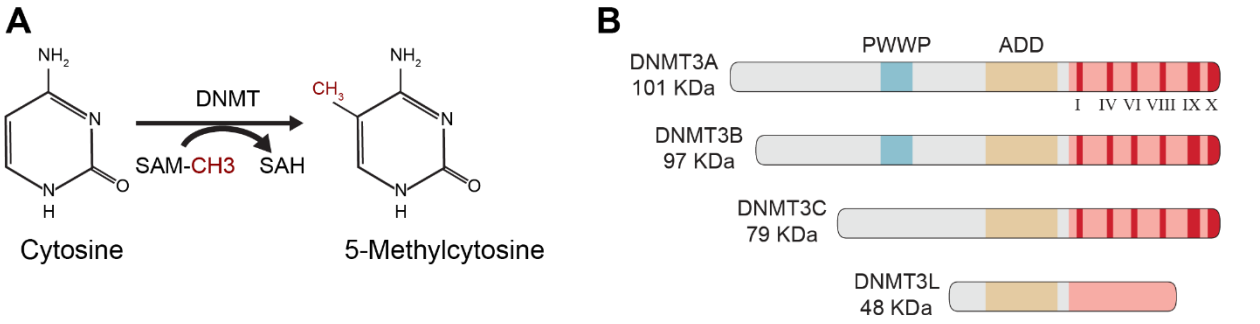

Figure 1.3: DNA methylation catalyzed by DNA methyltransferases (DNMTs). A) Cytosine and its conversion to 5-Methylcytosine. B) *de novo* methyltransferase proteins and their domain architectures. Image from Barau et al., 2016.

When DNA gets replicated, the symmetric CpG pattern results in hemimethylated DNA, where only one strand is methylated. DNMT1, also referred to as the maintenance DNA methyltransferase, recognizes these hemimethylated sites and faithfully restores DNA methylation patterns to the daughter strand. DNMT1 achieves this by interacting with Proliferating cell nuclear antigen (PCNA) and Ubiquitin-like containing PHD and RING finger domains, 1 (UHRF1) at the replication fork (Bestor et al. 1988; Leonhardt et al. 1992; Chuang et al. 1997).

The methyltransferases DNMT3A and DNMT3B perform *de novo* DNA methylation in vertebrates and have distinct, non-redundant functions (Okano et al. 1999). Both contain a Pro-Trp-Trp-Pro (PWWP) domain that binds to the histone mark H3K36me3, associated with transcription elongation, and an ATRX-DNMT3-DNMT3L (ADD) domain that binds to histone H3 tails while being repelled by the active H3K4me3 mark (Chédin 2011) (**Figure 1.3B**). DNMT3A is primarily responsible for DNA methylation at satellite repeats and imprinting genes, playing a critical role in embryogenesis, neural differentiation and spermatogenesis. DNMT3B targets gene bodies and intragenic regions, playing an important role in differentiation and placental development (Chédin 2011; Lister et al. 2013; Chen et al. 2003; Neri et al. 2017; Lauria et al. 2023; Borgel et al. 2010; Andrews et al. 2023).

The DNMT3-like enzyme (DNMT3L) lacks the PWWP domain (**Figure 1.3**

B) and does not bind DNA. Instead, it acts as a catalytically inactive cofactor that enhances the activity of *de novo* methyltransferases through physical interaction with them (Aapola et al. 2000; Suetake et al. 2004).

More recently, a third *de novo* methyltransferase, DNMT3C, was discovered in mice. DNMT3C is present in all muroids and is expressed in the male embryonic germline from Embryonic day (E)15.5 until Postnatal day (P)1 (Barau et al. 2016; Jain et al. 2017). In mouse germ cells DNMT3C, but not DNMT3A, methylates promoters of evolutionary young retrotransposon promoters from the L1 and ERV families, accounting for approximately 7,000 TE promoters, which makes up less than 1% of the mouse genome (Barau et al. 2016; Dura et al. 2022; Molaro et al. 2014). *Dnmt3C* originated from a tandem-duplication of *Dnmt3B* retaining a high sequence similarity but lacking the PWWP domain that binds H3K36me3 (**Figure 1.3**

B) (Jeltsch and Jurkowska 2016).

In mammals, DNA methylation is essential for survival in differentiated tissues, as loss of any of the three DNMTs, DNMT1, DNMT3A, DNMT3B, leads to embryonic or postnatal lethality (Li et al. 1992; Okano et al. 1999). While this lethality underscores its importance, primarily for TE silencing, direct evidence of TE upregulation upon DNMT depletion is limited due to this. Notable examples include TE upregulation in DNMT1-depleted mice embryos (Walsh et al. 1998) and in DNMT3L- or DNMT3C-depleted male germ cells (Bourc'his and Bestor 2004; Barau et al. 2016) as well as in various cancer types (Schulz 2006).

1.2.1.2. DNA methylation reprogramming in mammals

While DNA methylation is stably maintained in healthy somatic tissues, securing repression of TEs, mammalian genomes undergo two waves of DNA methylation reprogramming during embryonic development to reset DNA methylation patterns for embryonic and germline development (**Figure 1.4**). The first wave occurs post-fertilization to erase the methylation patterns inherited from the parents, reaching its lowest point at the blastocyst stage and allowing embryonic development. The second wave occurs during germline specification and erases the DNA methylation patterns of somatic cells to allow the expression of germline genes, which reaches the lowest DNA methylation levels in primordial germ cells at E13.5 (Sanford et al. 1987; Monk et al. 1987). The erasure of DNA methylation can occur in two ways: Passive demethylation by exclusion of DNMT1 or active demethylation mediated by enzymes from the Ten-Eleven Translocation (TET) family, which oxidize methylated cytosines that can be further actively or passively reverted to unmethylated cytosine (Kohli and Zhang 2013).

During post-fertilization reprogramming, active DNA demethylation is initiated by TET3, followed by passive DNA demethylation (Wossidlo et al. 2011; Gu et al. 2011). This demethylation continues until the blastocyst stage at E4.5, where global CpG methylation levels drop to approximately 20%. DNA methylation is retained at specific regions, including imprinting control regions and certain TEs, particularly IAPs (Lane et al. 2003; Wang et al. 2014). Global DNA methylation is reestablished in the epiblast by E6.5 (Wang et al. 2014).

Figure *1.4*

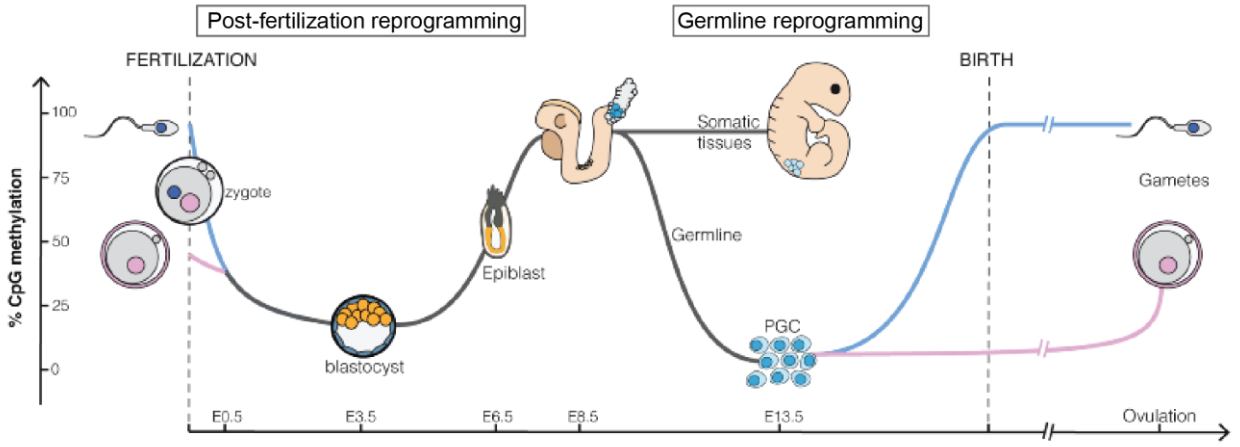

Figure 1.4: DNA methylation dynamics and reprogramming in mouse embryonic development. Graph illustration the percentage of global DNA methylation dynamics during mouse embryonic development of males (blue) and females (pink). DNA methylation reprogramming happens in two waves: The post-fertilization reprogramming reaching a minimum at the blastocyst stage and the germline reprogramming reaching a minimum at Primordial germ cell (PGC) stages. Image from PhD thesis Marius Walter 2015.

During germline reprogramming, a passive DNA demethylation of the majority of the genome is followed by an active TET1- and TET2-mediated demethylation of imprinted loci and germline-specific genes. By E13.5, CpG methylation levels in primordial germ cells reach a physiological minimum of approximately 3%-6% (Yamaguchi et al. 2012; Yamaguchi et al. 2013; Yamaguchi et al. 2013). However, DNA methylation persists at certain TE copies, including copies from IAP and ERV1 families (Guibert et al. 2012). Sex-specific DNA methylation patterns are subsequently established during germline development. In males, DNA methylation levels rise to 80% by birth, while in females, DNA methylation patterns are established after birth, reaching around 50% in oocytes (Smallwood et al. 2011; Wang et al. 2014). These DNA methylation levels are re-established at the majority of the genome by DNMT3A and DNMT3B, while evolutionary young TEs depend specificially on *de novo* DNA methylation by DNMT3C together with DNMT3L occurring between E16.5 and E19.5 (Barau et al., 2016). However, some regions, including IAP and L1 elements, resist *de novo* DNA methylation (Seisenberger et al. 2012; Lane et al. 2003).

This DNA methylation reprogramming creates a window of opportunity for TEs, which are no longer stably repressed, to become active and generate new copies. While during these time windows, a controlled TE expression seems beneficial for gene regulation and orchestration of development, early embryonic development and germline development are particularly vulnerable to aberrant TE activity (Seisenberger et al. 2012; (Jachowicz et al. 2017; Garcia-Perez et al. 2016, Maupetit-Mehouas and Vaury 2020). To counteract this, various other targeted defense mechanisms have evolved to repress TE activity specifically during these critical developmental stages. These mechanisms involve other transcriptional silencing, which repress TE expression on the chromatin level using epigenetic modifications.

1.2.2 Histone modifications

The genome of eukaryotes is vast, around 2 meters of DNA in humans, and must be highly compacted and organized as chromatin to fit within a nucleus of approximately 6 µm size in diameter. This compaction is achieved by wrapping DNA strands around nucleosomes, with 150 bp of DNA wrapping 1.7 times around each nucleosome. Nucleosomes are composed of an octamer of two sets of core histones, H2A, H2B, H3 and H4, and are linked by histone H1 (**Figure 1.5**). The chromatin structure is further folded into secondary and tertiary configurations (Lander et al. 2001; Luger et al. 1997).

Histones have flexible tails consisting of amino acids that can undergo various post-translational modifications, including acetylation, phosphorylation, methylation, ubiquitination, citrullination, SUMOylation and others (Kouzarides 2007). These histone modifications regulate chromatin compaction and thereby transcription by mediating either an open active structure, called euchromatin, or a compact silenced structure, called heterochromatin (**Figure 1.5**). Histone modifications create these structures either by directly altering the chromatin structure or by recruiting reader proteins (Kornberg 1974). For example, acetylation alters the chromatin structure directly by neutralizing the positive charge of lysines and thereby loosening the interaction of histones with the negatively charged DNA, while methylation can correlate with activation (e.g., H3K4, H3K36, H3K79) or repression of transcription (e.g., H3K9, H3K27, H4K20) dependent on the recruited reader proteins (Barski et al. 2007).

Histone modifications regulate gene expression epigenetically in a more dynamic manner than the relatively stable state of DNA methylation. During mammalian development, histone modifications also undergo reprogramming in early embryogenesis and germline development (Hajkova et al. 2008; Li 2002).

1.2.2.1. Key Histone Modifications

Among the diverse histone modifications that exist, this project focuses on H3K4me3, H3K27ac, H3K27me3, H3K9me2, and H3K9me3, which are some of the most well-studied histone marks (**Figure 1.5**).

H3K4me3, a trimethylation mark on lysine 4 of histone H3, is commonly found at promoters and transcription start sites of active genes (Howe et al. 2017). It is strongly associated with transcriptional activation and is catalyzed by the methyltransferases SET1A, SET1B, and the MLL1-MLL4 complexes (Miller et al. 2001; Yokoyama et al. 2004). Reader or effector proteins carrying a chromo, PWWP, PHD domains or other domains, such as chromatin remodelers CHD1 and NURF complex or transcription factors like IID, recognize H3K4me3 and promote transcription of the genes marked by this modification (Sims et al. 2005; Wysocka et al. 2006; Vermeulen et al. 2007).

Lysine 27 of histone H3 (H3K27) can undergo acetylation (ac) or trimethylation (me3), resulting in distinct functional outcomes. H3K27ac is typically enriched at active promoters and enhancers and is correlated with active transcription. This modification is mediated by the acetyltransferases p300 and CBP (Heintzman et al. 2009; Ogryzko et al. 1996). In contrast, H3K27me3 is associated with heterochromatin and transcriptional repression. This mark is often found at the promoters of developmental genes and the Hox cluster or the inactive X-chromosome, forming broad repressive domains (Boyer et al. 2006). H3K27me3 is deposited by the Polycomb Repressive Complex 2 (PRC2), which includes key components such as EZH1 or EZH2, SUZ12, EED, and RpAp46/48 (Aloia et al. 2013; Mendenhall et al. 2010; Cao and Zhang 2004). EZH2 serves as the primary catalytic enzyme, while SUZ12 and RpAp46/48 interact with unmodified histone H3 (Schmitges et al. 2011). PRC1 is subsequently recruited to H3K27me3-marked regions, further compacting chromatin and reinforcing repression (Wang et al. 2004).

Methylation of histone H3 Lysine 9 (H3K9) is another key histone modification associated with repressive chromatin. Mono (H3K9me1), di- (H3K9me2) and trimethylation (H3K9me3) are commonly found in heterochromatin and are enriched at repetitive elements, such as transposons and satellite repeats (Tachibana et al. 2002). Several histone methyltransferases, including SUV39H1, SUV39H2, SETDB1, G9a/EHMT2, and GLP/EHMT1, are responsible for establishing these marks (Peters et al. 2001; Strepkos et al. 2021; Montavon et al. 2021; Tachibana et al. 2005). Monomethylation is primarily catalyzed by SETDB1. The G9a-GLP complex, which is stabilized by the multi-zinc-finger-containing Wiz protein, mono- and di-methylates H3K9 (Ueda et al. 2006). Whereas SUV39H1 and SUV39H2 are mainly responsible for H3K9 di- and tri-methylation. Together, H3K9me2 and H3K9me3 play crucial roles in maintaining genome stability and repressing TEs. Heterochromatin protein 1 (HP1) proteins bind to H3K9me3 through their

chromodomains and spread heterochromatin formation together with SUV39H proteins by interacting with histone deacetylases, transcriptional repressors and chromatin remodelers (Bannister et al. 2001; Hall et al. 2002).

Figure *1.5*

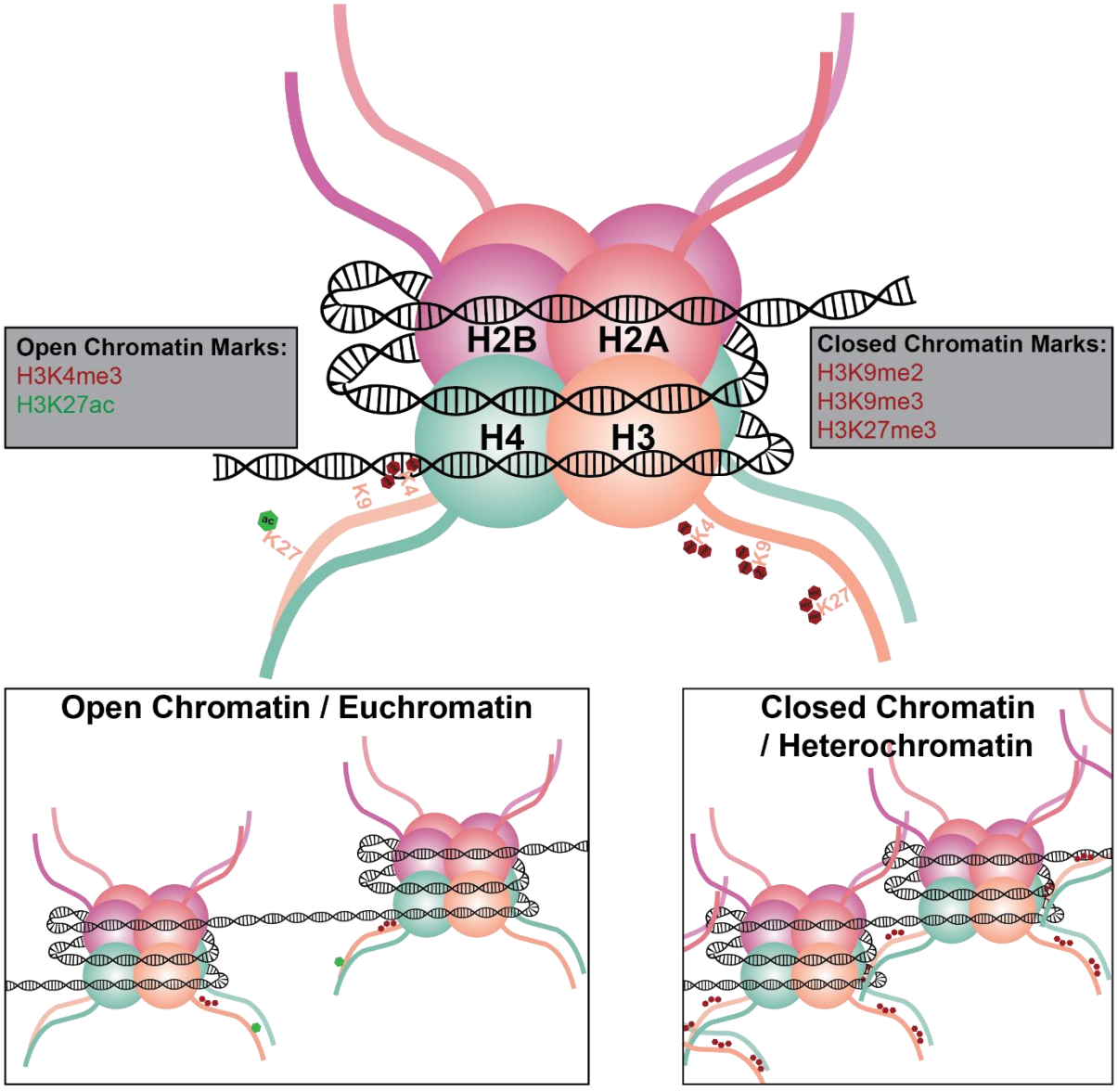

Figure 1.5: Illustration of nucleosome structure and chromatin marks. Nucleosome composition with an Octamer of two H2A, H2B, H3 and H4 histones and DNA wraped 1.5 times around nucleosome. Histone marks discussed in this thesis are depicted as hexamer structure in red for methylation and green for acetylation. Illustration of open chromatin and closed chromatin and its associated histone marks.

1.2.2.2. Crosstalk between histone modifications and DNA methylation

The various number of histone modifications and the regulatory effect they mediate through their reader proteins creates numerous possibilities for epigenetic regulation. These are even more complex due to crosstalk of histone modifications with DNA methylation and with each other.

DNA methylation can for example recruit chromatin remodelers, H3K9 methyltransferases and histone deacetylases through the methyl-CpG-binding domain (MBD) proteins, such as MBD1- MBD4 and MeCP2 (Dennis et al. 2001; Myant et al. 2011; Fuks et al. 2000; Baubec et al. 2013; Ng et al. 1999). Additionally, DNMTs bind regions marked by H3K36me3 via their PWWP domain but are excluded from H3K4me3-marked promoters through their ADD domain (Chédin 2011). While repressive H3K9me3 frequently co-occurs with DNA methylation, repressive H3K27me3 protects CpG-rich promoters from DNA methylation (Boulard et al. 2015). In line with this, H3K9me3 and H3K27me3 are almost never observed together, probably due to their crosstalk with DNA methylation (Mikkelsen et al. 2007).

H3K4me3 and H3K27me3 were shown to occur *in vivo* at the same nucleosome in a so-called bivalent chromatin structure (Bernstein et al. 2006; Schmitges et al. 2011; Voigt et al. 2012). This bivalency is prevalent at promoters in embryonic stem cells and thought to poise genes for future activation (Bernstein et al. 2006; Sachs et al. 2013) and to protect promoters from aberrant DNA methylation (Boulard et al. 2015). Another repressive chromatin mark, H3K9me3, was also suggested to co-occur with H3K4me3 in bivalent chromatin structure domains in adipose tissues (Matsumura et al. 2015).

1.2.2.3. Histone modifications regulating TEs

Histone modifications are critical for repressing and fine-tuning TE expression. In mouse embryonic stem cells, H3K9me3 and H3K27me3 establish heterochromatin at TEs, complementing DNA methylation for transcriptional repression (Schultz et al. 2002; Leeb et al. 2010). Strikingly, under conditions of hypomethylation, these marks compensate for the absence of DNA methylation by accumulating at TEs (Walter et al. 2016). Silencing by histone marks was suggested to be the original silencing mode for TEs, as it is conserved across many species, including *Drosophila* (Sienski et al. 2012), Chlamydomonas (Shaver et al. 2010), fungi (Dumesic et al. 2015) or ciliates (Miró-Pina et al. 2022). Similarly, in the hypomethylated genome during oogenesis of mice, TE repression relies on repressive H3K27me3 at their promoters (Huang et al. 2021).

In male germ cells, failure of the piRNA pathway, leading to the loss of DNA methylation at TE promoters, was shown to lead to TE upregulation only at the onset of meiosis. In

this context, their repression before meiosis in the absence of DNA methylation, was shown to rely on H3K9me2, while they were marked by H3K4me3 in meiotic stages (Di Giacomo et al. 2014; Zamudio et al. 2015). However, other histone marks decorating TE promoters in germline development in the absence of DNA methylation remain largely unexplored, leading us to explore the chromatin landscape at unmethylated TEs through spermatogenesis in this project.

1.2.3. Krueppel-associated box domain containing Zinc Finger Proteins

Krueppel-associated box domain containing zinc finger proteins (KRAB-ZFPs) are transcriptional repressors that bind genomic DNA and recruit KAP1/TRIM28 (Rowe et al. 2010). KAP1/TRIM28 serves as a scaffold for heterochromatin inducing factors, including SETDB1 and HP1 catalyzing H3K9me3, as well as nucleosome remodeling complexes and DNA methyltransferases (Schultz et al. 2002; Turelli et al. 2014).

The human genome encodes around 350 KRAB-ZFPs, many of which have unknown functions. However, the majority with more than 60% of the characterized KRAB-ZFPs are involved in the repression of TEs (Imbeault et al. 2017). Although the targeting of KRAB-ZFPs occurs sequence-dependent, a KRAB-ZFPs can target multiple TEs. The KRAB-ZFPs primarily target evolutionary older TEs, with most having lost their transposition potential. However, repression by KRAB-ZFPs is crucial during early embryogenesis, where depletion of KAP1/TRIM28 leads to the upregulation of a broad range of TEs (Turelli et al. 2014; Haring et al. 2021).

1.2.4. The human silencing hub (HUSH) Complex

The HUSH complex is a repressive complex that targets long intronless RNAs independently of their sequence (Tchasovnikarova et al. 2015; Seczynska et al. 2022). It recruits the effector proteins MORC2 and SETDB1 to establish heterochromatin via H3K9me3 deposition at its targets (Tchasovnikarova et al. 2015; Moissiard et al. 2012). Given that intronless RNAs are a hallmark of retrotransposition, the HUSH complex represses retrotransposons, with a particular preference for young full-length L1 (Liu et al. 2018).

This defense mechanism is essential in hypomethylated pluripotent stem cells and early embryogenesis (Soehn et al. 2009; Müller et al. 2021). It has been suggested that the HUSH complex could act as a defense against evolutionary younger TE activity during embryogenesis until KRAB-ZFPs evolve to silence the newly expanded TEs (Almeida et al. 2022). However, the role of the HUSH complex in germline development remains

unclear. While SETDB1 is required for germ cell development, its depletion has effects on TE expression (Liu et al. 2014).

1.2.5. Endo-siRNAs

RNA interference is a well-established defense mechanism against TE activity in plants, insects and nematodes (Slotkin and Martienssen 2007; Fire et al. 1998). This post-transcriptional defense mechanism has also been observed in mammals in specific cell types. In this pathway, DICER processes mRNA of retrotransposons into endo-siRNAs, which are subsequently bound by ARGONAUTE 2 (Flemr et al. 2013). Endo-siRNAs are present in mouse pluripotent stem cells and primordial germ cells with low DNA methylation. DICER depletion results in mildly increased TE expression in mouse embryonic stem cells and oocytes (Flemr et al. 2013; Berrens et al. 2017). Previous work proposed that endo-siRNAs provide an immediate response to TE activity during embryogenesis, preceding that of heterochromatin-mediated repression by the HUSH complex or KRAB-ZNFs (Berrens et al. 2017).

1.2.6 The PIWI interacting RNA pathway

PIWI-interacting RNAs (piRNAs) are small, non-coding RNAs, 24–35 nucleotides in length, that form complexes with PIWI-clade Argonaute (PIWI) proteins, guiding them to silence active retrotransposons in the germline of mammals (Czech and Hannon 2016). In mice, piRNAs are produced exclusively in male germ cells, whereas in hamsters and potentially other mammal species, they are produced in germ cells of both sexes (Zhang et al. 2021). piRNAs are categorized into three types based on the developmental stage at which they are produced: fetal piRNAs, pre-pachytene piRNAs, and pachytene piRNAs. Except for pachytene piRNAs, which are derived from long non-coding RNAs in intergenic regions and of largely unknown function, all piRNAs show high sequence complementarity to TEs and play an important role in TE silencing (Gainetdinov et al. 2018).

The mouse genome encodes three PIWI proteins: MILI/PIWIL1, MIWI/PIWIL2, and MIWI2/PIWIL3. Other mammals, such as humans and hamsters, encode a fourth PIWI protein, PIWIL3 (Zhang et al. 2021; Ishino et al. 2021). PIWI proteins and their associated piRNAs are essential for male fertility in mice and humans but are dispensable for female fertility: In other mammals such as hamsters, PIWI proteins are crucial for fertility in both sexes (Carmell et al. 2007; Deng and Lin 2002; Stallmeyer et al. 2024, Kuramochi-Miyagawa et al. 2004; Zhang et al. 2021).

1.2.6.1. The post-transcriptional piRNA pathway

The piRNA pathway counteracts the activity of TEs in mouse male germline reprogramming, when TEs are exempted from DNA methylation and it establishes their life-long silencing by DNA methylation. The piRNA pathway includes a post-transcriptional gene silencing (PTGS) that cleaves the produced mRNA from active TEs and a transcriptional gene silencing (TGS) that targets the promoters of active TEs for DNA methylation and heterochromatin formation (**Figure 1.6**).

During fetal stages, the piRNA production involves the ping-pong amplification cycle maximizing piRNA efficiency (Beyret et al. 2012). In the ping-pong cycle briefly, unistrand piRNA clusters are transcribed by the A-MYB transcription factor and the canonical RNA polymerase II machinery. The pre-piRNA precursors undergo processing steps such as capping, splicing, and polyadenylation. During maturation, a 5’ uridine and 2′-O-methylation at the 3’ end is generated (Sun et al. 2022). Primary piRNAs are processed from these long precursor transcripts by trimming and are loaded onto MILI. MILI recognizes newly generated TE mRNAs through base complementarity of its piRNAs and cleaves it at the 10th position into secondary piRNAs, which now have a bias towards the adenine nucleotide at the 10th position, and are loaded onto MIWI2. These secondary piRNAs, in turn, recognize additional TE mRNAs, ensuring further degradation (De Fazio et al. 2011) (**Figure 1.6**). This PTGS mechanism, which resembles an adaptive immune system, silences TEs by degrading TE mRNA in germ cell granules in the cytoplasm, called nuage, beginning at E14.5 and presumably lasting until P3, where MIWI2 is still expressed.

1.2.6.2. The transcriptional piRNA pathway

Loaded MIWI2 additionally translocates to the nucleus, starting at E16.5 in mice, to mediate TGS of TEs (Aravin et al. 2008) (**Figure 1.6**). This occurs through base complementarity, and involves a cascade of other proteins, ultimately leading to the establishment of DNA methylation and heterochromatin formation at TE promoters. This transcriptional silencing involves several enzymes, with DNMT3C being the most downstream effector responsible for *de novo* methylation of TE promoters (Barau et al. 2016). In addition to MIWI2 and DNMT3C, several other proteins, including TDRD9 (Shoji et al. 2009), TEX15 (Schöpp et al. 2020), SPOCD1 (Zoch et al. 2020), SPIN1(Dias Mirandela et al. 2024), C19ORF84 (Zoch et al. 2024) and DNMT3L (Bourc'his and Bestor 2004) have been shown to participate in the TGS pathway.

Failure of the piRNA pathway, and the resulting absence of DNA methylation at TE promoters following germline reprogramming, leads to TE upregulation at the onset of meiosis in males and meiotic arrest at the pachytene stage (**Figure 1.6**). Depletion of any

components involved in the piRNA pathway, including *Miwi2* (Aravin et al. 2008), *Dnmt3C* (Barau et al. 2016), *Dnmt3L* (Bourc'his and Bestor 2004), *Mael* (Soper et al. 2008), *C19orf84* (Zoch et al. 2024), *Tdrd9* (Shoji et al. 2009), and *Spocd1* (Zoch et al. 2020) results in a consistent phenotype in male mice characterized by these defects, ultimately causing hypogonadism and male sterility (**Figure 1.7**).

Figure *1.6*

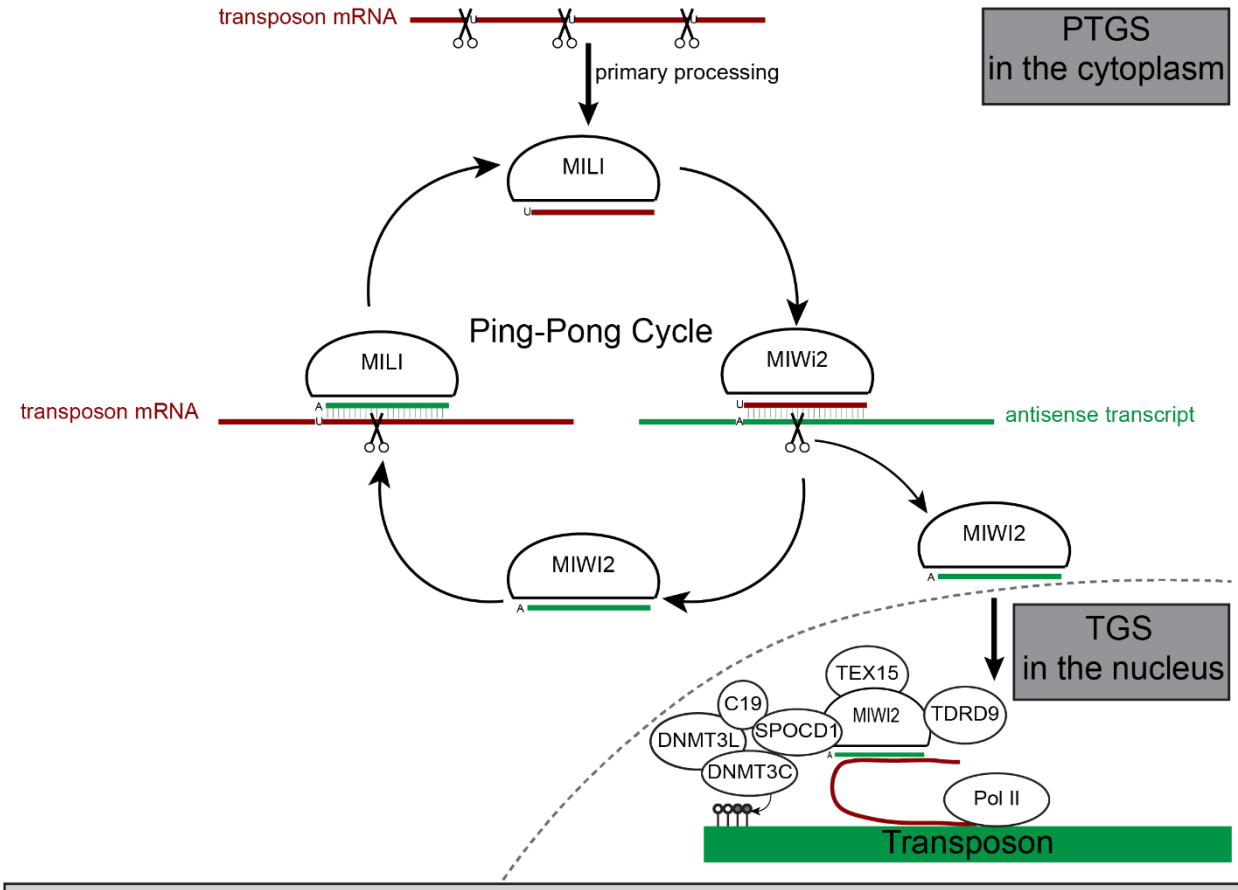

Figure 1.6: The current view of the piRNA pathway in mice. The PTGS in the cytoplasm and the TGS in the nucleus are illustrated according to the current model of the piRNA pathway in mice.

Co-immunoprecipitation of MIWI2 with members of the *de novo* methyltransferase family has so far been unsuccessful (Aravin et al. 2008), suggesting that additional enzymes might mediate the interaction between MIWI2 and the *de novo* methylation machinery. Many efforts have been made to co-immunoprecipitate all interactors and establish a molecular cascade of the TGS piRNA pathway. However, the complete cascade remains unresolved, while many involved proteins have been characterized to date: TRDR9 is an essential binding partner of MIWI2 in the nuage and in the nucleus (Shoji et al. 2009). TEX15 mediates double strand breaks during meiosis, but its role in the piRNA pathway remains unclear. It is mainly nuclear, interacts with MIWI2 and MILI, and remains nuclear

when MIWI2 is depleted (Yang et al. 2020; Schöpp et al. 2020). SPOCD1 also interacts with MIWI2 without affecting its nuclear localization upon depletion. SPOCD1 was shown to interact with chromatin remodeling complexes and the *de novo* methylation machinery suggesting that it may help heterochromatin formation at TE promoters (Zoch et al. 2020). The testis-specific nuclear C19ORF84 protein interacts with SPOCD1. C19ORF84 stabilizes the interaction of SPOCD1 with chromatin remodeling complexes and DNMT3L, but its exact function in the piRNA pathway remains elusive (Zoch et al. 2024). This leads to the current model of the TGS pathway, where MIWI2 interacts with TDRD9 and SPOCD1, which interacts with DNMT3C-DNMT3L stablilized by C19ORF84 (C19) (**Figure 1.6**).

Figure *1.7*

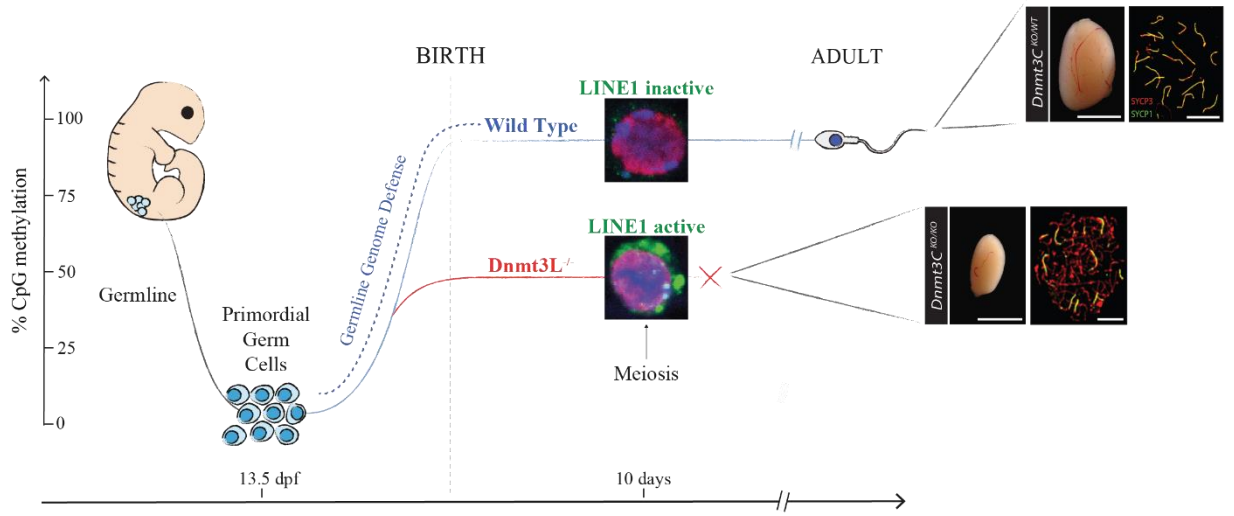

Figure 1.7: DNA methylation dynamics during male germline reprogramming with a defective piRNA pathway. Graph illustrating the percentage of global DNA methylation dynamics during male germline reprogramming in mice for Wild-type and *Dnmt3L* mutants. The resulting phenotype is illustrated using images from representative immunofluorescence images of LINE1 (green) in germ cells (pink), showing LINE1 upregulation in the mutant and the testis size showing hypogonadism in the absence of DNA methylation together with meiotic spreads, showing failed chromosome synapsis (Images adapted from PhD thesis Marius Walter 2015 and Barau et al 2016).

1.3 Transcription Factors

Transcription Factors (TFs) are proteins that bind to DNA by recognizing specific short DNA sequence motifs, thereby influencing gene expression (Maston et al. 2006). They can influence gene expression either by binding to promoter sequences to initiate transcription or by binding regulatory sequences to stimulate or repress the expression of the related genes. TFs are classified by their DNA binding domain (DBD), with approximately 100 DBDs identified in eukaryotes. Major TF families in eukaryotes include C2H2-zinc finger, Homeodomain, basic helix-loop-helix, basic leucine zipper and nuclear hormone receptor (Johnson and McKnight 1989).

1.3.1. The impact of epigenetic modifications on TF binding and activity

Around 2000 TFs are known in humans (Ravasi et al. 2010) and despite the vast number of predicted binding sites in the genome identified through motif analysis, only a subset of those are bound by TFs *in vivo*, as revealed by profiling techniques (Cusanovich et al. 2014). This demonstrates that factors beyond sequence specificity determine TF binding.

One such factor is TF cooperativity, which can either occur directly through protein-protein interactions or multimer formation, or indirectly, through collaborative competition for DNA access. TFs must compete with nucleosomes or bind nucleosomes to access their binding site, otherwise they require a cofactor to do so (Long et al. 2016). Pioneer TFs, can bind their target sites even in closed chromatin, leading to nucleosome displacement or repositioning (Iwafuchi-Doi and Zaret 2014; Soufi et al. 2015). Closed chromatin in this context is defined by lacking DNase hypersensitivity.

TF binding can be further determined by other epigenetic mechanisms, including histone modifications and DNA methylation. Even pioneer TFs may be blocked from binding in the presence of certain repressive histone modifications (Iwafuchi-Doi and Zaret 2014). The pluripotency transcription factors OCT4, SOX2 and KLF4 for example, are defined as pioneer TFs, but can be inhibited at their target sites by the presence of the repressive H3K9me3 mark (Soufi et al. 2012). Moreover, nucleosomes marked symmetrically by H3K4me3, H3K27me3 or asymmetrically by H3K4me3-H3K27me3 and displaying bivalency were reported to attract a different set of TFs in vitro (Bryan et al. 2025).

DNA methylation at CpG dinucleotides can additionally inhibit binding of some TFs, either through directly hindering the binding due to alteration of the binding site or through recruitment of methyl-binding proteins that block the target site (Dantas Machado et al. 2015). Other TFs, such as CTCF, can also bind methylated regions and some TFs bind their methylated targets and even promote its demethylation, such as REST (Tate and Bird 1993; Feldmann et al. 2013; Stadler et al. 2011). Interestingly, TF binding has been associated with preserving the unmethylated state in some cases, even at CpG islands (Brandeis et al. 1994; Krebs et al. 2014).

In addition, TF activity is modulated by post-translational modifications, such as phosphorylation, acetylation, methylation and glycosylation of the protein. These modifications can alter TF subcellular localization, stability and interaction with other factors (Everett et al. 2010; Tootle and Rebay 2005).

1.3.2 TFs regulating TEs

TFs as the main regulators of transcription can also influence TE expression. In fact, TE promoters are hubs of potential TF binding sites encompassing a huge variety of motifs (Hermant and Torres-Padilla 2021; Sundaram et al. 2014). Several TFs have been implicated in TE regulation. For example, CTCF, a TF involved in maintaining topologically-associated domain (TAD) boundaries, binds a significant number of TEs, primarily B2 elements in mice (Bourque et al. 2008). Other TFs affect TE expression indirectly, such as MYBL1 or also known as A-MYB, which regulates piRNA pathway genes and binds to enhancer LTR elements regulating germline gene expression (Sakashita et al. 2020, Tan and Wilkinson 2023).

TFs that bind TEs have been identified across various cell types but TFs rather exhibit cell-type specificity and no master regulator for TFs has been discovered so far (Karttunen et al. 2023; Hermant and Torres-Padilla 2021; Sundaram et al. 2014). For instance, CREB1 has been shown to bind unmethylated IAPs in differentiated neurons (Kaluscha et al. 2022) and YY1 was shown to activate L1Tfs during early embryogenesis (Saha et al. 2024). Another TF important for zygotic genome transcription, called DUX, activates MERVLs in preimplantation embryos, which is essential for ZGA (Sakashita et al. 2023).

Identifying TFs that bind TEs, sheds light on how TEs can escape defense mechanisms and explains their successful colonization of mammalian genomes. Germ cell development is the most critical time for this due to DNA methylation reprogramming and the necessity for TEs to be expressed in order to generate new copies. However, as examples of TFs that directly bind to TE promoters in the germline are highly limited, where only KRAB-ZFPs have been identified, and remain poorly studied (Otsuka et al. 2023).

1.4 Male germ cell development

Germ cell development refers to the process by which reproductive cells, called gametes, are produced: sperm in males and oocytes in females. This chapter focuses on male germ cell development, particularly in mice. Male germ cell development takes place in the seminiferous tubules of testis (**Figure 1.8A)** and can be roughly divided into three stages: The germ cell specification encompassing primordial germ cells, spermatogenesis, which includes the mitotic and meiotic phases, and spermiogenesis where the transformation of spermatids into mature spermatozoa occurs.

1.4.1. Germ cell specification

During embryogenesis in mammals, a subset of somatic cells is specified to become primordial germ cells (PGCs) through signaling pathways, a process known as epigenesis (Whittle and Extavour 2017) and subsequently undergo germline epigenetic reprogramming, including resetting of DNA methylation and histone modifications, as discussed in **chapter 1.2.1.2.** DNA methylation reprogramming in mammals. In mice, PGCs appear around E6.5 and migrate to colonize the gonadal ridges between E10.5 and E13.5. During this migration, PGCs undergo extensive proliferation, increasing in number from approximately 100 to 25,000 cells (Tam and Snow 1981).

There is no obvious difference between sexes in the initial specification and proliferation of PGCs until sex determination at E11.5, where sex-specific differentiation pathways are initiated. In males, the *Sry* gene on the Y chromosome, together with the TF SOX9, drives the differentiation in a subset of somatic cells into Sertoli cells (Jan et al. 2012). At E13.5, female PGCs enter meiosis and arrest at prophase I until birth, while male PGCs enter mitotic arrest (McLaren 1984). Male PGCs that reach the genital ridges differentiate into prospermatogonia (ProSpg), characterized by large, round nuclei. ProSpg remain mitotically inactive, until a few days after birth (Vergouwen et al. 1991).

Due to the epigenetic reprogramming in PGCs and the piRNA pathway along with the establishment of *de novo* chromatin marks, in ProSpg before birth, these stages are characterized by overall high TE expression, with LINE1 expression peaking at E16.5 (Seisenberger et al. 2012; Maupetit-Mehouas and Vaury 2020). Thus, we investigated TE regulation in the ProSpg stages at E14.5 and E15.5 before and at E18.5 after the establishment of DNA methylation at TEs, which we sorted using GFP expression under the control of the ProSpg marker gene, *Oct4*.

1.4.2. Spermatogenesis

Spermatogenesis begins after birth and continues throughout the lifespan of male mice, relying on asynchronous waves of germ cell development within the seminiferous tubules. As a result, these tubules contain a mixed population of germ cells at various stages of differentiation, arranged from the basal lamina toward the lumen (Hess and de Franca 2009) (**Figure 1.8**). Spermatogenesis starts with a mitotic differentiation phase, followed by a meiotic phase that ultimately produces haploid germ cells.

Figure *1.8*

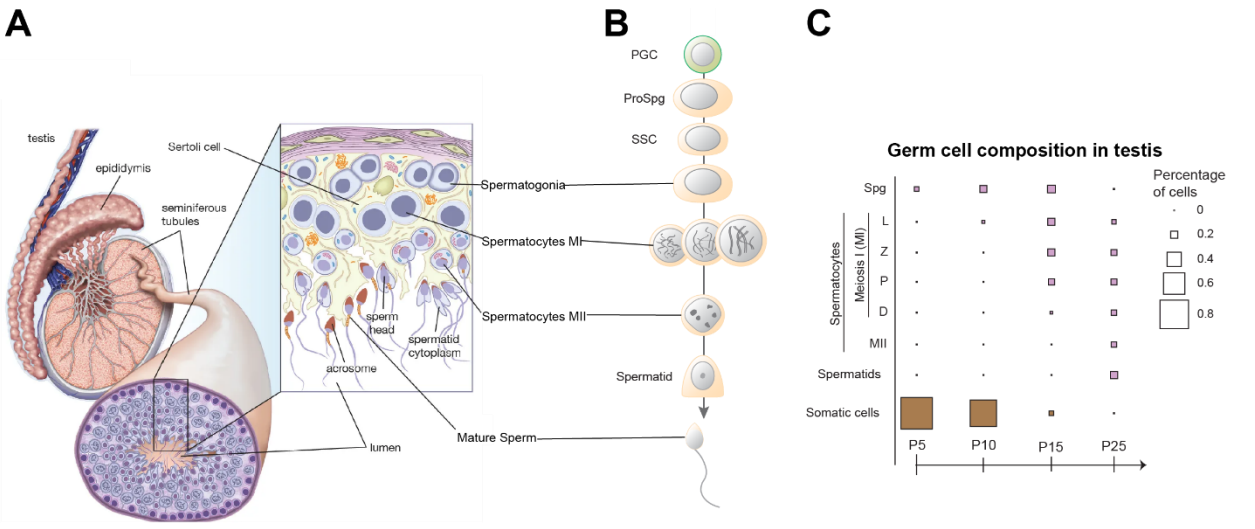

Figure 1.8: Mouse testis structure and male germ cell development. A) Cartoon of a cross-section of a mouse testis and a cross-section of a seminiferous tubule (The Editors of Encyclopedia Britannica, 2024). B) Progression of male germ cell development. C) Estimation of the germ cell stage composition in mouse testes at specific mouse ages in percentage. This image was modified from Ernst et al., 2019.

1.4.2.1. The mitotic phase

The mitotic phase of spermatogenesis encompasses spermatogonial stem cells (SSCs), undifferentiated spermatogonia (Spg), and differentiated Spg.

Around postnatal day (P)3, ProSpg migrate to the basal lamina of the seminiferous tubules and begin proliferating, giving rise to SSCs. SSCs are a rare cell population, comprising less than 0.05% of all germ cells, but are crucial for maintaining the continuous production of sperm through self-renewal and differentiation (Fayomi and Orwig 2018).

SSCs can produce pairs of undifferentiated Spg, which undergo two or more further mitotic divisions (de Rooij and Russell 2000). Upon perception of retinoic acid signaling from Sertoli cells, undifferentiated Spg differentiate into differentiated Spg, marked by the expression of cKIT (Gewiss et al. 2020; Green et al. 2018). Differentiated Spg undergo several mitotic divisions, producing type A and type B Spg. One SSC is capable of producing approximately up to 1,000 type B Spg.

Interestingly, during the first round of spermatogenesis, some ProSpg that lack the SSC marker ID4 can directly differentiate into differentiating Spg, bypassing the previous stages. This allows for the faster production of mature sperm (Chan et al. 2014; Yoshida et al. 2006).

After DNA methylation is established in PGC and ProSpg stages, it remains stable during spermatogenesis, ensuring persistent TE repression (Soper et al. 2008, Fanourgakis et al. 2025). Previous studies suggest that TE repression is maintained in the mitotic phase even in the absence of DNA methylation at TE promoters (Zamudio et al. 2015). To explore potential alternative mechanisms of TE repression, we investigated TE regulation in these germ cell stages. To isolate germ cells from the mitotic phase, we used the cell surface marker EpCAM, which does not distinguish between SSCs, undifferentiated Spg, and differentiated Spg (Kanatsu-Shinohara et al. 2011). Hereafter, we refer to these sorted cells collectively as spermatogonia.

1.4.2.2 The meiotic phase

Type B Spg enter meiosis, a specialized form of cell division that produces four haploid gametes and become spermatocytes (Spc). Meiosis involves one round of DNA replication followed by two rounds of chromosome segregation: Meiosis I, where homologous chromosomes segregate and Meiosis II, where sister chromatids segregate. First, the DNA is replicated, then two rounds of chromosome segregation follow, which include four stages, prophase, metaphase, anaphase and telophase (Jan et al. 2012).

Prophase I, the most extended and important phase of meiosis, is further divided into four stages: Leptotene, Zygotene, Pachytene and Diplotene. In Leptotene, double strand breaks are initiated by SPO11 triggering homologous recombination. Zygotene is characterized by condensed chromatin and the starting of synopsis, which is the pairing process of the homologous chromosomes via the synaptonemal complex. The synopsis is completed in Pachytene stage, where homologous recombination sites with crossing-overs are established. In Diplotene, homologous chromosomes separate and the synaptonemal complex is disassembled (Zickler and Kleckner 2015). Defects in alignment, pairing, or recombination during prophase I can trigger checkpoints, leading to meiotic arrest at the pachytene stage (Jan et al. 2012).

Following prophase I, metaphase I begins as homologous chromosomes align along the metaphase plate. During anaphase I, homologous chromosomes are pulled apart toward opposite poles of the cell. In telophase I, the cell undergoes cytokinesis, resulting in two haploid daughter cells. Meiosis II resembles mitosis and involves the separation of sister chromatids ultimately resulting in the formation of four haploid spermatids (Handel and Schimenti 2010).

While DNA methylation remains stable during meiosis, histone modifications such as H3K4me3 are dynamic and and some TEs, such as LINE1s, were reported to be transiently expressed in wild-type early meiotic stages (Branciforte and Martin 1994; Soper et al. 2008; Lam et al. 2019; Fanourgakis et al. 2025). LINE1 expression boosts in

meiosis when DNA methylation is absent at their promoters leading to meiotic arrest, making these germ cell stages particularly interesting for studying TE regulation.

1.4.3. Spermiogenesis

After meiosis is completed, the haploid cells stop dividing and undergo significant morphological and cytological changes to produce sperm. This process is called spermiogenesis, which involves several key steps. First, the acrosome is formed in the head of the sperm. Next, chromatin remodeling occurs. During this process, most nucleosomes are replaced with protamines, resulting in a highly compact and condensed chromatin structure. However, a recent study showed that a small fraction of nucleosomes is retained, on regions that are characterized by high CpG levels and low DNA methylation levels (Fanourgakis et al. 2025). A flagellum, or sperm tail, which is responsible for the motility of the sperm, also develops during spermiogenesis. Finally, upon completing these transformations, the mature spermatozoa are released into the lumen of the seminiferous tubules. Then, they are transported to the epididymis, where they undergo further maturation (O'Donnell and O'Bryan 2014).

In the context of this study, we did not explore any round or elongating spermatids or sperm, as germ cells lacking DNA methylation at TE promoters do not develop past meiotic stages (Bourc'his and Bestor 2004).

1.5 Aim and Scope of the thesis

Taken together, we currently understand various silencing mechanisms for TEs and the consequences of their failure leading to aberrant TE expression and various diseases. To further expand in the genomes, TEs must therefore strike a balance between their silencing, to not harm the fitness of their host, and escaping the silencing, which enables the generation of new copies in the genome. How TEs achieve this balance and how some TE copies escape the silencing remains elusive.

While the piRNA pathway is known to faithfully target TEs for DNA methylation in the male germline, thereby silencing them, TEs can form metastable epialleles with variegated DNA methylation patterns and become active. Key questions in the field include why TEs become active only at meiosis and what mechanisms silence TEs before meiosis in the absence of DNA methylation. Studying DNMT3C mutants thus provides insight into how TEs behave when DNA methylation is absent at their promoters. Additionally, it remains unknown what causes the meiotic catastrophe when the piRNA pathway fails.

For my PhD project, I addressed these questions by investigating how TEs are regulated in *Dnmt3C*$^{KO/KO}$ spermatogenesis. First, we explored how evolutionary young and active

TEs are expressed in the absence of DNA methylation in male germ cells in greater developmental detail: Before DNMT3C-mediated DNA methylation, after failure of DNA methylation prior to meiosis and at meiosis itself. Next, we examined chromatin modifications at these developmental stages that could further epigenetically regulate their expression. Finally, we sought to identify proteins, which could bind and regulate TEs depending on the presence or absence of DNA methylation using a DNA-pull down screen. We filtered the identified candidate regulators, by the presence of their motif on TEs and their expression in postnatal germ cells and finally characterized the binding of a methylation-sensitive TF to TEs during spermatogenesis. Beyond further characterization, we ultimately tested whether depleting our candidate TF in the absence of DNA methylation could rescue TE silencing.

2. Material and Methods

Some sections of the text in this thesis were refined and proofread using GPT-4o mini.

2.1. Mouse Strains and Housing

2.1.1. Mice Housing

Animal experiments were conducted in compliance with local and European animal welfare regulations and were approved by the relevant authorities (ethical committees on animal care and use of the federal state of Rhineland-Palatinate, Germany; covered under LUA license G 23-5-049). Mice, aged 7–24 weeks, were housed in ventilated cages with enrichment materials under controlled conditions (humidity 40–70%, temperature 22 ± 2 °C) on a 12-hour light/dark cycle and provided food and water ad libitum. For tissue and embryo collection, mice were euthanized by cervical dislocation. Prenatal testis samples were collected from embryos following timed pregnancies, with the first day postcoitum designated as E0.5. Postnatal time points were recorded starting at birth, considered day P1. All mouse strains used in this study had been previously described and were bred on a C57Bl6/J background (*Dnmt3C*$^{KO/KO}$ (Barau et al. 2016) *Nrf1*$^{fl/fl}$, *Ddx4*^{Cre}/*Vasa*^{Cre} (Wang et al. 2017) and *Oct4-eGFP* (Jackson Laboratories, stock 008214).

2.1.2. Mice breeding

To generate a germ cell-specific conditional *Nrf1* knockout in spermatogonia, *Nrf1*$^{fl/fl}$ animals were crossed with *Ddx4*^{Cre} animals. Specifically, *Nrf1*$^{fl/KO}$ female animals were bred with *Nrf1*$^{fl/WT}$; *Ddx4*^{Cre} male animals. *Nrf1*$^{fl/KO}$; *Ddx4*^{Cre}; *Dnmt3C*$^{KO/KO}$ mice were obtained by crossing *Nrf1*$^{fl/KO}$; *Dnmt3C*$^{KO/KO}$ female animals with *Nrf1*$^{fl/WT}$; *Ddx4*^{Cre}; *Dnmt3C*$^{KO/WT}$ males (**Figure 2.2.1**).

Additionally, *Dnmt3C*$^{KO/KO}$ mice were crossed with *Oct4-eGFP* mice to establish a combined *Dnmt3C*$^{KO/KO}$; *Oct4-eGFP* mouse line.

Figure 2.*2.1*

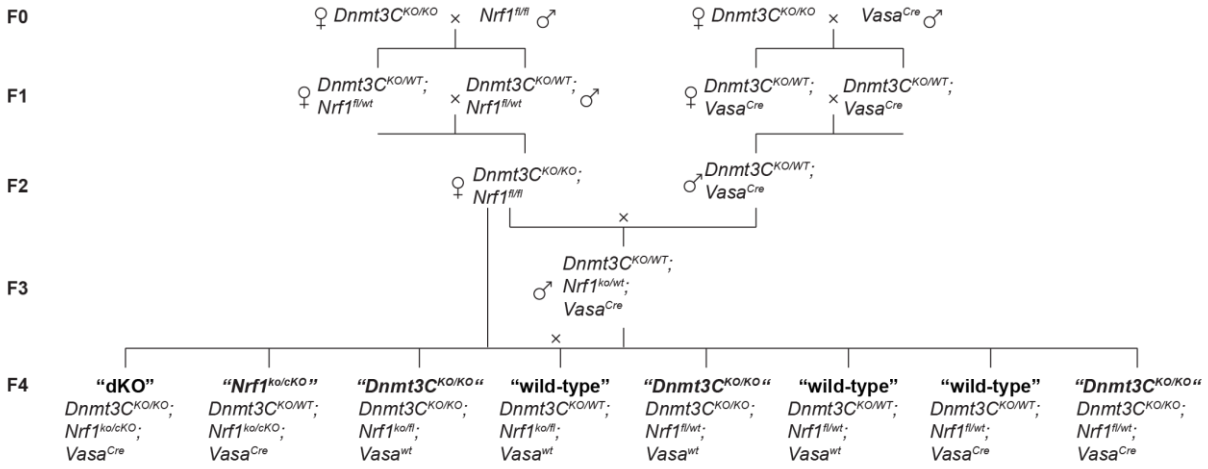

Figure 2.1: Mouse breeding scheme. Breeding Scheme to obtain the double-KO (dKO) animal *Dnmt3C$^{KO/KO}$; Nrf1$^{ko/cKO}$; Vasa^{Cre}* and littermate controls. Crosses (x) are indicated for four different generations (F). F4 shows all possible genotypes, which can occur from the final cross F3 and how they are referred to in the experiments (in bold) . For simplicity, for the other generations only the relevant genotypes are illustrated.

2.1.3. Mice genotyping

Genomic DNA for PCR genotyping was extracted by adding 50 µL of extraction solution (100 mM NaOH, 2.5 mM EDTA) to an ear clip or 30 µL to a neonatal phalange, followed by tissue dissolution at 95°C for 10 minutes. The solution was then spun down, cooled on ice, and 1 µL was used for a 25 µL PCR reaction. All offspring were genotyped using primers listed in **Table 1**.

2.2. mouse Embryonic Stem Cell (mESC) culture

E14 mESCs were seeded in feeder-free conditions on gelatin-coated plates and grown in serum-based medium composed of DMEM (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) and titrated 2i condions using 200 nM MEK inhibitor PD0325901 (Axon Medchem 1408), 3 mM Gsk3 inhibitor CHIR99021 (Axon Medchem 1386). Media was changed daily, and cells were passaged at ~70% confluence every other day using TrypLE Express (Gibco).

2.3. Histology and Immunofluorescence

Testes of mice were dissected and collected at different time points to assess TE expression at specific germ cell developmental stages for immunofluorescence.

2.3.1. Collection and cryosections of testis

Testes were collected at embryonic stages E15.5, E18.5, P2, P5, P8, P10, P15, P20, P30 from *Dnmt3C*$^{KO/KO}$ and *Dnmt3C*$^{KO/WT}$ males. Embryonic testes were fixed for six hours, and postnatal testes overnight, in 4% paraformaldehyde (PFA) at 4°C. Following fixation, the samples were washed twice in 1X PBS for 5 minutes each. They were then incubated sequentially in 15% and 30% sucrose solutions overnight at 4°C. Testes were equilibrated in O.C.T. compound for 1 hour at room temperature (RT) and frozen at -80 degree following embedding in O.C.T. compound (FSC 22 Frozen Section Compound, Leica). Cryoblocks were sectioned at 10 µm thickness, and mounted onto Superfrost Plus slides (Epredia).

2.3.2. Generation of IAP-POL antibody

As there is no commercially available antibody against IAP that we found to work well, we collaborated with the Boulard's lab to receive an IAP-Pol antibody they had produced:

Briefly, Polyclonal anti-IAP-POL antibodies were generated using a recombinant murine IAP-POL protein (Uniprot P12894) produced in *E. coli*. The IAP-POL cDNA was subcloned into a pETM11-SUMO3 expression vector and expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells grown in TB-FB medium. After induction with lactose at 25°C, cells were harvested, lysed, and the inclusion bodies were solubilized in guanidinium hydrochloride. The His6-SUMO3-IAP-POL fusion protein was purified using Ni-NTA chromatography, cleaved with SenP2 protease to remove the tag, and further purified via ion-exchange chromatography. The final IAP-POL protein (~1.1 mg/mL) was validated by mass spectrometry and stored at -80°C.

Polyclonal antibodies were raised in a New Zealand White rabbit immunized with the purified IAP-POL protein. After final bleeding, serum was collected, and IAP-POL-specific antibodies were affinity-purified using IAP-POL-coupled agarose beads. Eluted antibodies were analyzed by SDS-PAGE, pooled, and stored in 50% glycerol at -70°C.

2.3.3. Immunofluorescence Staining

Immunofluorescence on cryosections were performed by Styliani-Eirini Kanta using the following procedure: Slides were equilibrated to RT, outlined with a hydrophobic pen

(ImmEdge Pen, VectorLabs), and washed in 1X PBS for 10 minutes. Blocking and permeabilization were performed for 1 hour at RT using a solution of 10% donkey serum (Sigma), 3% BSA, 0.3% Triton X-100, and 1X Protease Inhibitor Cocktail (PIC) in PBS. Sections were incubated overnight at 4°C with primary antibodies (**Table 2**) diluted in blocking buffer. After three washes in 1X PBST (0.1% Tween-20 in PBS), sections were incubated for 1 hour at RT with secondary antibodies (donkey anti-rabbit Alexa Fluor 647, donkey anti-rat Alexa Fluor 488; ThermoFisher) at 1:800 dilution and DAPI (2 µg/µL) for nuclear staining. Following three additional 1X PBST washes, sections were mounted using ProLong Gold Antifade Mountant (Life Technologies).

2.3.4. Immunofluorescence Analysis

Quantification of TRA98 and L1ORF1 signals in the nucleus and cytoplasm was performed by Microscopy Core Facility using 256 x 256-pixel images of manually cropped cells from a larger field of view. Nuclei were segmented using the DAPI channel. A Gaussian filter (sigma = 6) was applied, and thresholds were set using the OTSU algorithm (Otsu, 1979). Segmentation was refined by filling holes and using a watershed operation to separate touching nuclei, creating the region of interest (ROI) for the nucleus. The cytoplasm ROI was defined by creating a ring around the nucleus ROI, dilating it 10 times, and subtracting the nucleus ROI. Mean TRA98 and L1ORF1 intensities were measured within the nucleus and cytoplasm ROIs. The total number of nuclei in whole sections was calculated by Styliani-Eirini Kanta usimg the Analyze Particles tool in ImageJ after applying a threshold. GraphPad Prism v10 was used to generate the bar plot.

2.4. Protein Extraction & Western Blotting

Protein extraction and western blotting was performed by Styliani-Eirini Kanta. Protein samples for western blotting were extracted by homogenizing testes in RIPA buffer supplemented with PIC (Roche, 11836170001). Samples were sonicated for 5 minutes (30 sec on, 30 sec off) and centrifuged at 16,000g for 30 minutes at 4°C. Protein concentration was measured using the Bradford assay. 15 µg of protein per sample were resolved on SDS-PAGE gels (GeneScript, M00653, M00654) and transferred to nitrocellulose membranes (ThermoFisher, 88018). Membranes were blocked for 40 minutes - 1 hour with 5% BSA in TBST, incubated overnight at 4°C with primary antibodies (**Table 2**), washed three times with TBST, and incubated for 1 hour at room temperature with HRP-conjugated secondary antibodies (**Table 2**). Protein bands were visualized using chemiluminescence reagents (SuperSignal West Pico and Femto, ThermoFisher, 1863096).

2.5. Cell Sorting by Flow Cytometry

To isolate germ cells at specific developmental stages, Fluorescent-Activated Cell sorting with specific procedures for ProSpg, Spg and Spc was used to sort cells using flow cytometry. The protocols to prepare the cells for Flow Cytometry and the Gating strategies are described in the following chapters.

2.5.1. Isolation of ProSpg

ProSpg were isolated from E14.5 or E15.5 Oct4-eGFP male mice. Testes were dissected, collected in 1X PBS and decapsulated. The decapsulated testes were transferred to tubes containing 100 µl collagenase solution in HBSS (Collagenase A, 2× AAs, 2× Na-pyruvate, and 25 mM HEPES-KOH, pH 7.5) and dissociated at 37 °C for 8 minutes with flicking. After adding 200 µl TrypLE Express, the cell suspension was incubated for 5 minutes at 37 °C and pipetted to obtain a single-cell suspension. TrypLE was quenched with 70 µl pre-warmed fetal bovine serum (FBS). Cells were washed once with FACS buffer (PBS, 2 mM EDTA, 25 mM HEPES-KOH, 1.5% BSA, 10% FBS) supplemented with 10% FBS and once with FACS buffer after centrifugation at 600 g for 4 minutes. Cells were resuspended in DAPI-supplemented (2 µg/mL) FACS buffer, filtered through 40µm cell strainers, and sorted using an Invitrogen Bigfoot with a 100-µm nozzle. DAPI-negative and OCT4-eGFP-positive cells were sorted into low-binding tubes with 100 µl PBS (**Figure 2.2.2A**).

2.5.2. Isolation of Spg

For SSCs and Spg, testes from Wild-type and *Dnmt3C*$^{KO/KO}$ 14-day-old mice were processed similarly as described in 2.5.1. Single-cell suspensions were stained with 8.25 mg per 5 million cells of CD326 (Ep-CAM)-Alexa Fluor® 647 (BioLegened, 118212) and β-2-Microglobulin-PE (Santa Cruz Biotechnology, SC-32241) on ice for 20 minutes, washed with FACS Buffer twice and resuspended in DAPI-supplemented (1 µg/mL) FACS Buffer, then sorted using the same conditions. Single stainings of Alexa Fluor 647 and PE, were used to draw gates for DAPI-negative, β-2-Microglobulin-negative, and Ep-CAM-positive cells, which were collected in low-binding tubes with 100 µl PBS (**Figure 2.2.2B**).

2.5.3. Isolation of Spc

Spc were isolated from testes of wild type and *Dnmt3C*$^{KO/KO}$ mice between P20 and P30. Testes were decapsulated and dissociated in 2 mL collagenase solution (Collagenase A, 1.5× AAs, 1.5× Na-pyruvate, and 25 mM HEPES-KOH, pH 7.5) at 35 °C. The cell

suspension was sedimented for 2 minutes, the supernatant was aspirated and resuspended in another 2mL collagenase solution. After treating with 250 µl TrypLE and filtering, 10 million cells were stained with 0.5 µL Hoechst 3342 (Thermo Scientific, 62249) for 8 minutes. Cells were centrifuged at 4 °C 6,000g for 4 min, resuspended in incubation buffer (PBS, 1.5× AAs, 1.5× Na-pyruvate, and 25 mM HEPES-KOH, pH 7.5) supplemented with 1µL Hoechst 3342 per 10 million cells at 35 °C for another 8 minutes, filtered, and stained with Propidium Iodide (Invitrogen, P3566). Cells were sorted using the same conditions by excluding Propidium-Iodide positive cells to sort live cells and Hoechst 3342 staining that indicates cell size and DNA content to sort the cells having 2N DNA content. An additional gate was drawn for round spermatids serving as a negative control for *Dnmt3C*$^{KO/KO}$ sorting, where round spermatids are absent due to meiotic arrest (**Figure 2.2.2C, D**).

Figure 2.2.2

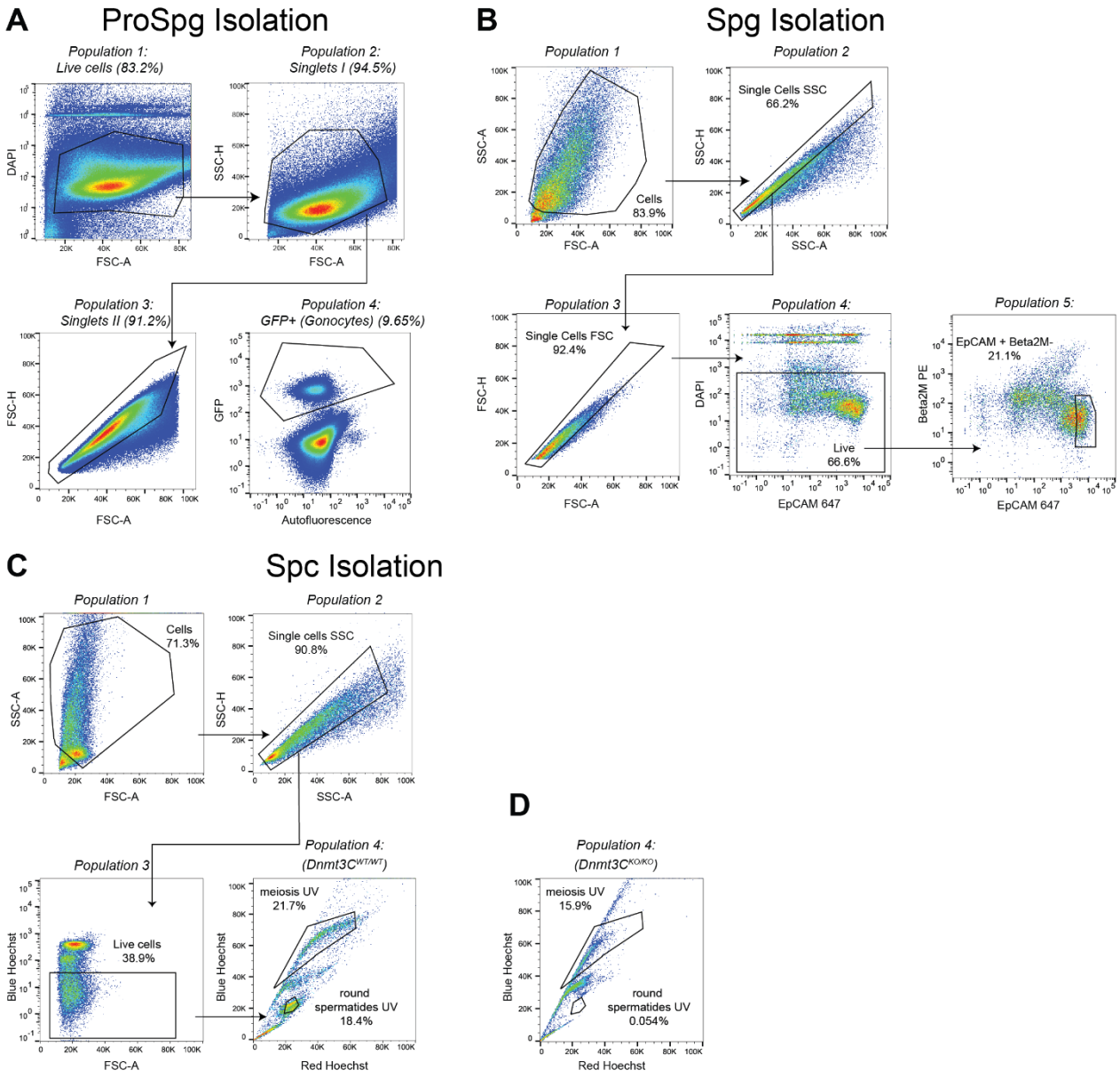

Figure 2.2: Gating strategies for Cell Sorting by Flow Cytometry. A) Gating strategy used for ProSpg sorting. In the first gate, live cells that are DAPI-negative are sorted. Next, singlets are sorted by side (population 2) and forward scattering (population 3). In the final gating GFP-positive cells are sorted. B) Gating strategy used for Spg sorting. In the first three gates, healthy singlet cells are sorted. The fourth gate excludes dead cells that are DAPI-positive. The final gating strategy sorts cells, which are β2-Microglobulin(Beta2M)-PE negative and EpCAM-AF647 positive. (C) and (D) Gating strategy used for Spc sorting. Population 1 and 2 (C) sort singlet healthy cells. The third population of (C) excludes dead cells stained by Propidium Iodide. The fourth population sorts Spc dependent on their DNA content and cell size stained by Hoechst3342 in (C) for wild-type and (D) for *Dnmt3C*$^{KO/KO}$ Spc, where not all meiotic stages are present and round spermatids are absent.

2.6. RNA analysis

In order to measure TE expression in germ cells of different mutants, RNA was extracted from sorted germ cells or whole testis. To obtain an overview of TE expression in *Dnmt3C*$^{KO/KO}$ versus wild type germ cell development, RNA from sorted germ cells of ProSpg, Spg and Spc stages were sequenced. TE expression in conditional knockout mutants of *Nrf1* was assessed by extracting RNA from whole testes and performing Reverse transcription quantitative PCR (qPCR) and RNA sequencing.

2.6.1. RNA extraction

Sorted germ cells or whole testes of littermates were collected and frozen in Trizol until triplicates of each genotype were collected. RNA extraction was performed using the DirectZol RNA MicroPrep kit (Zymo, R2061). The extracted RNA was either used directly for Library preparation performed by the Genomics Core Facility or split into half, one for processing with Reverse-Transcription and quantitative PCR (qPCR) and one for Library preparation.

2.6.2. Library preparation & Sequencing

Next Generation Sequencing (NGS) libraries were prepared using Illumina's Stranded mRNA Prep Ligation Kit, following the Stranded mRNA Prep Ligation Reference Guide (April 2021, Document 1000000124518 v02). For RNA extracted from sorted Spg and Spc, 10 ng of input RNA was used, while 29 ng was used for RNA from whole testes mutants. Libraries were amplified through 13 and 12 PCR cycles, respectively. Two post-PCR purification steps were performed to remove residual primers and adapter dimers.

The libraries were analyzed on a 2100 Bioanalyzer using the DNA High Sensitivity chip (Agilent Technologies) and quantified with the Qubit 1x dsDNA HS Assay Kit on a Qubit 4.0 Fluorometer (Invitrogen, Thermo Fisher Scientific). A total of 16 samples from sorted Spg and Spc and 17 samples from whole testes mutants were pooled in equimolar ratios, targeting a minimum of 50 million reads per sample. Sequencing was conducted on an Illumina NextSeq 2000 system as 2×108-nt paired-end reads with one dark cycle at the start of R1 and R2. Each sample yielded 50–60 million read pairs.

2.6.3. Bioinformatic Analysis of RNA sequencing

The RNA-sequencing data was analyzed by Bioinformatics Core Facility. Reads were aligned to the Mus musculus genome assembly GRCm38 (GENCODE release M25) using STAR (v2.7.3a) with the following parameters: --outMultimapperOrder Random, --

outSAMattributes NH HI AS nM MD, --outSJfilterReads All, --outSAMunmapped Within, --outFilterMismatchNoverReadLmax 0.04, --outFilterMismatchNmax 999, --winAnchorMultimapNmax 100, and --outFilterMultimapNmax 100 (Dobin et al. 2013). The TEtranscripts tool (v2.2.3) was employed to annotate reads to both genes and transposable elements using the parameters --mode multi, --sortByPos, and --stranded reverse (Jin et al. 2015).

Differential expression analysis was performed with DESeq2 (v1.44.0), applying a corrected p-value threshold (Benjamini-Hochberg FDR) of <0.01 (Love et al. 2014). To quantify TE expression at the locus level, TElocal (v1.1.1) was used with the parameters --mode multi, --sortByPos, and --stranded reverse, and the counts were normalized to Transcripts Per Million (TPM) values.

2.7. DNA pulldown & Mass Spectrometry

To identify potential regulators of TEs in the presence and absence of DNA methylation in germ cell development, a DNA pulldown approach of L1MdTf and LTR1a promoters was coupled to Liquid Chromatography- mass spectrometry (LC-MS).

Briefly, the bait DNA of TE promoters and scrambled controls in methylated and unmethylated flavors were incubated in nuclear protein extracts from mouse testis, pulled down and the bound proteins were measured by LC-MS.

2.7.1. Extraction of nuclear proteins

Proteins were extracted from pooled testes of 20 P15 male mice in triplicates using nuclear extraction. Testes were homogenized in Cytoplasmic Buffer (10 mM Hepes-KOH pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 1X PIC) and incubated on ice for 30 minutes, a sample was taken as total extracts. After centrifugation at 1100 g for 10 minutes at 4°C, the supernatant was stored as the cytoplasmic fraction, the nuclear pellet was resuspended in Nuclear Soluble Buffer (20 mM Hepes-KOH pH 7.9, 420 mM NaCl, 2 mM $MgCl_2$, 20% glycerol, 1 mM DTT, 0.2% IGEPAL CA-630, 1X PIC), vortexed, and incubated for 30 minutes at 4°C. The insoluble fraction was pelleted by centrifugation at 14,000 g for 20 minutes at 4°C, the supernatant was kept as the nuclear soluble fraction. The pellet was resuspended in Nuclear Soluble Buffer, sonicated (10 cycles, 10s on/50s off), and centrifuged again. The solubilized fraction was combined with the soluble nuclear extract and the total protein concentration was quantified by Bradford assay.

2.7.2. Bait Preparation

The sequences of the TE promoters from IAP LTR1a and the first two monomers from L1MdT were amplified from pre-made plasmids carrying a full-length TE sequence, cloned into blunt-end vectors, and PCR-amplified using the primers listed in **Table 1**. To generate sequence-unspecific scrambled controls, the CpG sites of TE promoters were preserved, but the rest of the sequences were shuffled, synthesized (IDT DNA gBlocks), cloned into a blunt-end vector, amplified, and biotinylated similarly (**Figure 2.2.3A, B**). Biotinylated probes were *in vitro* methylated using the MssI CpG methyltransferase kit (NEB, M0226M) with a ratio of 0.8 µL enzyme per 2 µg DNA.

Figure 2.2.3

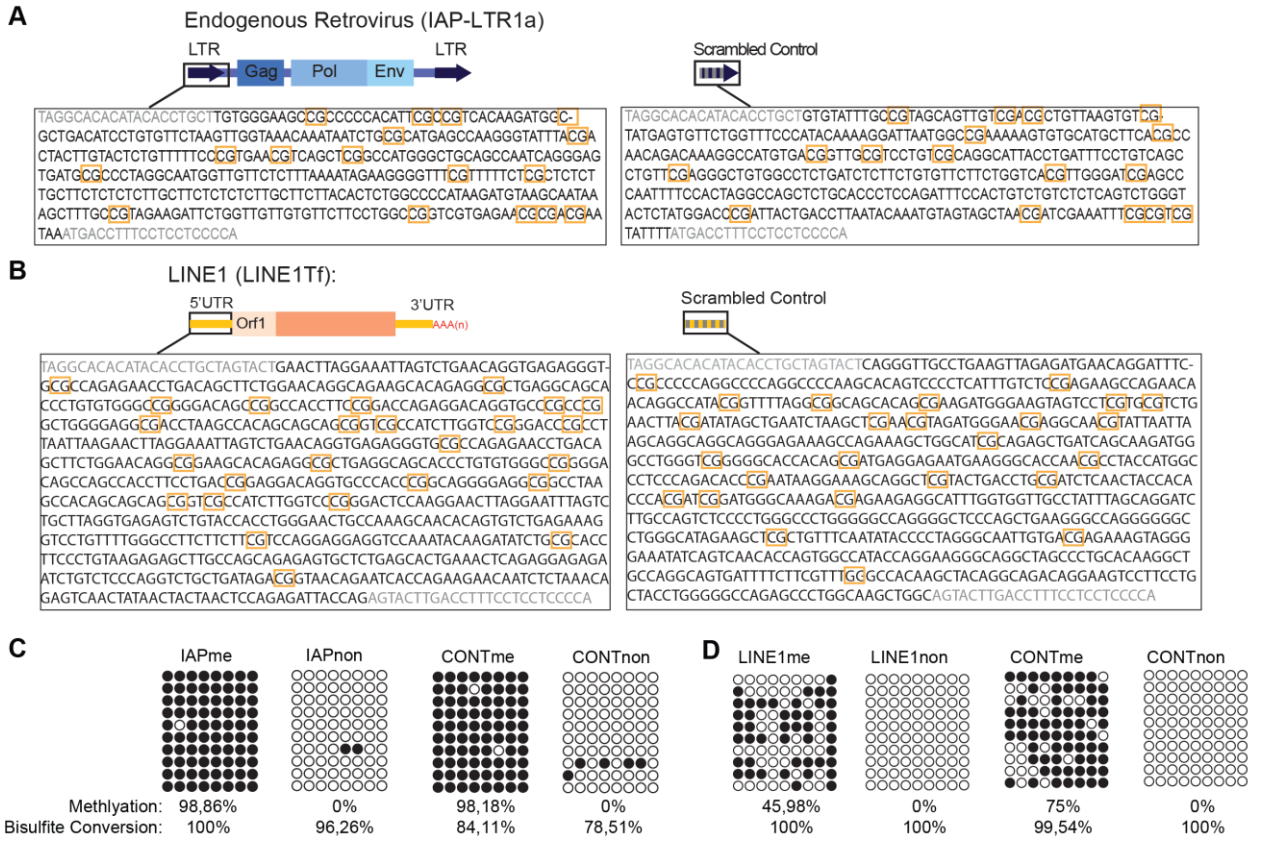

Figure 2.3: Bait preparation for IP-MS approach. A) Illustration of the bait sequence used from the IAP LTR1a promoter and the corresponding scrambled control used in the DNA-pulldown experiment. CpG sites, which are identical between IAP LTR1a and the scrambled control, are highlighted with a yellow frame. The sequences corresponding to random primers used for amplification and biotinylation, are shown in gray. B) as in (A) for the LINE1MdT promoter and its scrambled control. C) Bisulfite cloning results for IAPme, IAPnon, CONTme, and CONTnon, represented as black/white circle diagrams, with black circles indicating methylated cytosines and white indicating unmethylated. Each column is a CpG site, and each row represents a DNA clone. Average methylation percentages and bisulfite conversion rates are indicated at the bottom. D) as in (C) for LINE1me, LINE1non, CONTme and CONTnon.

2.7.3. Bisulfite Treatment & Analysis

The Methylation levels were checked by bisulfite cloning. To this end, methylated and unmethylated probes were bisulfite converted (EZ DNA Methylation Quick kit, D5005), amplified using the primers listed in **Table 1**, cloned into pMiniT™ 2.0 Vectors (NEB, E1203S), followed by transformation and Sanger sequencing of 10 clones per probe. The transformation of cytosines into uracils, if unmethylated, or thymines if methylated, was reported to measure the Bisulfite conversion rate (**Figure 2.2.3C, D**).

2.7.4. Immunoprecipitation (IP)

Following the protocol outlined by Harris et al. (2018), 60 µg of bait DNA was bound to Streptavidin magnetic beads in triplicates (Invitrogen Dynabeads MyOne, 650032) using a coupling buffer containing 20 mM Tris-HCl (pH 8.0), 2M NaCl, 1 mM EDTA, and 0.1% Tween-20. The beads were rotated at RT for 1 hour, then resuspended in Pull-Down Buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM MgCl2, 0.5% IGEPAL CA-630, 1 mM DTT, 1X PIC). Subsequently, they were incubated with 250 µg of protein extracts for 3 hours at 4°C. After three washes, the bound proteins were eluted in LDS buffer with DTT, heated at 80°C for 10 minutes, and prepared for LC-MS.

2.7.5. Mass Spectrometry Preparation and analysis

LC-MS preparation and analysis was performed in collaboration with the Butter lab. Briefly, protein samples were separated on a 4%–12% NuPAGE Novex Bis-Tris gel (Thermo) at 180 V for 8 minutes in 1x MES buffer, fixed (7% acetic acid, 40% methanol), stained with Coomassie G250, and destained with water. Gel lanes were excised, minced, destained with 50% EtOH/25 mM ammonium bicarbonate (ABC), dehydrated in acetonitrile (ACN), reduced with 10 mM DTT in 50 mM ABC at 56°C for 1 hour, and alkylated with 50 mM iodoacetamide in 50 mM ABC for 45 minutes at room temperature in the dark. Gel pieces were dried, rehydrated with trypsin (1 µg/sample), and incubated overnight at 37°C. Peptides were extracted with 30% ACN, dried, desalted on StageTips, and separated on a Reprosil C18 column using an Easy nLC 1000 system with a 90-minute gradient (2–60% ACN, 0.1% formic acid) at 225 nL/min. Eluted peptides were analyzed on a Q Exactive Plus mass spectrometer using HCD fragmentation with a Top10 acquisition scheme. Data were processed with MaxQuant (v1.6.5.0) using a UniProt mouse database (54,220 entries), enabling match-between-runs and LFQ quantitation. Contaminants, reverse hits, and single-peptide identifications were excluded. Final data analysis and visualization were performed in R.

2.8. TF Motif Analysis

TF motif analysis was performed by Ishita Amar. High-confidence TE annotations, excluding those overlapping with genes, were generated using the pipeline from Sakashita et al. (2020). LINE, IAP, and ERV families were isolated using the grep tool and intersected with DNMT3C differentially methylated regions (Barau et al. 2016) to identify DNMT3C targets. Motif enrichment analysis was performed with the AME tool from MEME Suite (McLeay and Bailey 2010), using the HOCOMOCOv11 database (HOCOMOCOv11_full_MOUSE_mono_meme_format.meme). TFs enriched for DNMT3C targets relative to all TEs were identified, and these hits were cross-referenced with proteins enriched in the IP-MS results using their UniProt identifiers.

2.9. Fluorescence Polarization (FP)

To test the binding of NRF1 to methylated and unmethylated baits in vitro, we collaborated with the IMB Protein Production Core Facility to express and purify full-length NRF1 and perform a Fluorescence Polarization Assay.

2.9.1. Protein purification

The pET-MGSH6-MBP-3C-mmNRF1(fl)-Twinstrep construct was expressed in E. coli BL21 (DE3) codon+ cells. Protein expression was induced in TB media with 0.5 mM IPTG at 18°C overnight.

Cell lysates were prepared in lysis buffer containing 500 mM NaCl, followed by purification using Ni-NTA resin. Elution fractions from the Ni-NTA pull-down were subjected to His-3C protease digestion during incubation with Streptactin beads. Digestion was carried out with 20 µg of His-3C protease for 1 hour at 4°C to cleave the His-MBP tag. The purified protein was subsequently eluted and prepared for downstream analyses.

2.9.2. FP Assay

Oligos containing the known binding motif and oligos with the peak region of the *Asz1* promoter were synthesized and ordered from IDT Cy5-labeled methylated and unmethylated (**Table 1**). Using a 386-well plate (Corning™, Low-Volume, Polystyrene, black), Cy5-labeled double-stranded oligonucleotide substrates (10 nM) were incubated with serial dilutions of purified mmNRF1 (17.9µM stock) in a total volume of 20 µL of FP buffer (10 mM Na-Hepes pH 7.4, 150 mM NaCl, 1 mM TCEP, 5% glycerol). Following a 10-minute incubation at 20°C, FP of the Cy5-labeled oligonucleotides was measured

using a Tecan Spark 20M plate reader at 20°C (excitation wavelength: 625 nm, emission wavelength: 665 nm, gain: 120, flashes: 15, integration time: 40 µs).

Normalized FP values were calculated by subtracting the FP measurement of the oligonucleotide-only control from all conditions containing varying amounts of the recombinant protein. The normalized FP values, along with standard deviations from three independent experiments, were plotted using GraphPad Prism 8.

EC50 values, serving as proxies for the dissociation constant (Kd), were determined using a four-parameter [Agonist] vs. Response fit with a variable slope in GraphPad Prism 8, when applicable.

2.10. Chromatin Immunoprecipitation

Profiling of NRF1 was tested using Chromatin immunoprecipitation (ChIP) in mESCs and whole mouse testis. Testis from P16 mice were dissected and dissociated into a single-cell suspension as described in chapter **2.5.0**. After filtering and washing, 4.5 million cells were resuspended in 10 mL PBS. WT and *Dnmt3A$^{KO/KO}$; Dnmt3B$^{KO/KO}$; Dnmt1$^{KO/KO}$* (TKO) mESCs were cultured as described in chapter **2.2.** mouse Embryonic Stem Cell (mESC) culture at 70% confluency.

2.10.1. Crosslinking

The cells were fixed by adding Formaldehyde to mESCs directly in culture or testis single cell suspension to a final concentration of 1% and incubating for 10 minutes at RT. Glycine was added to quench fixation to a final concentration of 0.25 M and incubated shaking at RT for 5 minutes. The cells were washed, and mESCs were collected in ice-cold PBS. After centrifugation at 3,000 rpm for 5 minutes at 4°C, the cell pellet was resuspended in 1 mL of ice-cold Lysis Buffer A (100 mM Tris-HCl pH 8, 10 mM DTT, PIC in water) and incubated on ice for 15 minutes. Following an additional pelleting step, the cells were resuspended in 1 mL of ice-cold Lysis Buffer B (100 mM HEPES, 10 mM EDTA, 0.5 mM EGTA, 0.5% Triton X-100, PIC in water) and incubated for another 15 minutes on ice. After pelleting, the cells were incubated in 1 mL of ice-cold Lysis Buffer C (10 mM HEPES, 10 mM EDTA, 0.5 mM EGTA, 200 mM NaCl, PIC in water) for 5 minutes on ice, followed by pelleting and resuspension in 400 µL of Lysis Buffer D (50 mM Tris-HCl pH 8, 10 mM EDTA, 0.75% SDS, PIC in water).

2.10.2. Chromatin shearing

The samples were split into two sonication tubes, and the chromatin was digested by sonication using the Bioruptor Pico (Diagenode, B01080010) at the highest setting, with

30 sec on and 30 sec off. The optimal shearing cycle was determined to be 8 cycles by collecting samples every two cycles from 2 to 12 cycles and extracting the chromatin to assess fragmentation on the Bioanalyzer (**Figure 2.2.4**). After sonication, the chromatin was pelleted by centrifugation at 12,000 g for 10 minutes at 4°C. The chromatin was then resuspended in 1.5 mL ChIP buffer (15 mM Tris-HCl pH 8, 180 mM NaCl, 1.2 mM EDTA, 1.2% Triton X-100, and PIC in water) and 10% of input was collected.

Figure 2.*2.4*

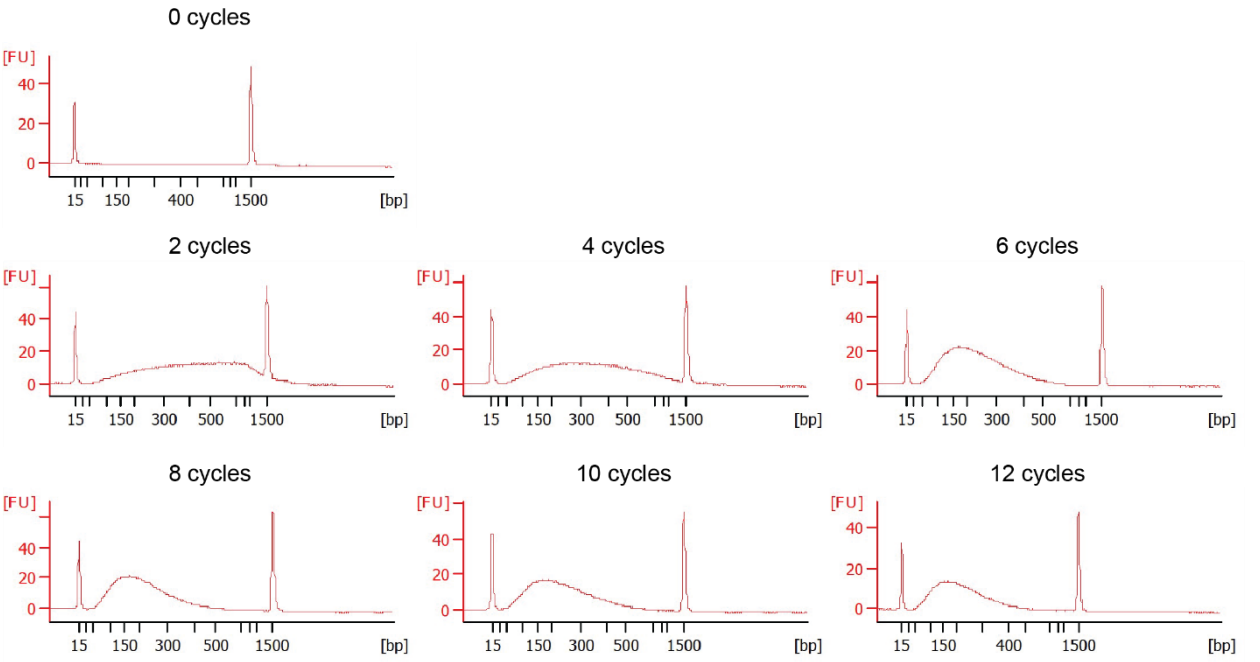

Figure 2.4: Analysis of sheared chromatin for ChIP. Electropherograms showing fragment sizes of chromatin after sonication with 0-12 cycles to identify the optimal cycle number.

2.10.3. Pulldown

Following chromatin digestion, the samples were split into three tubes for incubation with NRF1 antibody and an IgG control, listed in **Table 2**. A final concentration of 1 µg per 4.5 million cells was added, and the samples were incubated overnight at 4°C with rotation. Subsequently, each sample was incubated with 30 µL of Dynabeads protein A (Invitrogen, 10002D) at 4°C for 1 hour. The samples were washed on a magnetic stand with 1 mL of High Salt Buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100, in water) and shaken for 5 minutes at room temperature, followed by a wash with 1 mL of LiCl Buffer (10 mM Tris-HCl, 250 mM LiCl, 1 mM EDTA, 1% sodium deoxycholate, 1% NP-40) and shaken for 5 minutes at RT. This was followed by a wash in TE Buffer (10 mM Tris-HCL pH 8, 2 mM EDTA, in water), incubating for 5 minutes with shaking at RT.

Finally, the samples were eluted twice in 50 µL of Elution Buffer (100 mM NaHCO3, 1% SDS, in water) by incubation at 30°C for 15 minutes with shaking.

The immunoprecipitated samples and input samples were supplemented with NaCl to a final concentration of 200 mM, SDS to a final concentration of 1%, and $NaHCO_3$ to a final concentration of 100 mM. Samples were reverse crosslinked by shaking at 65°C overnight, followed by treatment with RNase A at 37°C for 1 hour and 20 µg of Proteinase K at 55°C for 2 hours. The isolated chromatin was extracted using the DNA Clean & Concentrator kit (Zymo, D4014) and eluted in 30 µL of water.

2.10.4. Library preparation and sequencing

Libraries were prepared using the xGEN ssDNA & Low-Input Library Preparation Kit (IDT, 10009817) following the manufacturer's protocol. Indexing primers from the "xGen UDI Primer Plate 2, 8nt, 96rxn" kit (Integrated DNA Technologies 10009816) was used for unique dual indexing. The eluted DNA was then cleaned of primer-dimers using 1.3X AMPure beads and washed with ethanol. The quality of the libraries was assessed using the DNA High Sensitivity Kit (Bioanalyzer, 5067-4626). The libraries were subsequently used for qPCR.

2.10.5 qPCR

To evaluate the suitability of this technique for profiling NRF1, enrichment at known positive and negative control targets, as well as TEs as our targets of interest, was tested by qPCR using Power SYBR Green PCR Master Mix prior to NGS. Input samples were diluted 1:10 to achieve 1% input, and 0.5 µL of each sample was used per reaction. The primers used for this analysis are listed in **Table 1**.

2.11. Ultra-Low Input ChIP

To find a suitable protocol for low input material we tested a modified protocol of Ultra-Low Input ChIP (ULI ChIP), which we compared to the standard ChIP protocol (Brind'Amour et al. 2015).

2.11.1. ULI ChIP Protocol

Testis from P16 mice were dissociated into a single-cell suspension, as described in chapter 2.5., and were crosslinked with 1% formaldehyde at RT for 10 minutes with gentle rotation. Crosslinking was quenched by adding glycine to a final concentration of 125 mM and incubating for 2 minutes at RT, followed by washing the cells in PBS.

The cells were then resuspended in ice-cold ChIP dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl) and treated with 5 µL Micrococcal Nuclease (MNase) for 15 minutes at 37°C to digest the DNA. Digestion was stopped by adding 10 mM EDTA and 20 mM EGTA. A fraction of the DNA was extracted to assess fragmentation using a Bioanalyzer.

Following centrifugation, the soluble extract was used for Input and IP, as described in Chapter 2.10.3. The samples were washed twice with Low Salt Buffer (20 mM Tris-HCl, 2 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 0.1% SDS) and High Salt Buffer (20 mM Tris-HCl, 2 mM EDTA, 500 mM NaCl, 0.1% Triton X-100, 1% SDS).

2.11.2. Library Preparation & Analysis

Libraries were prepared as described in chapter **2.10.4.0**.

2.11.3. qPCR

To assess the suitability of this method, a qPCR was performed before size selection with AMPure beads and after size selection. The qPCR was performed as described in chapter **2.10.5.0**.

2.12. Cleavage under Targets and Release

Cleavage under Targets and Release (CUT&Run) was tested to profile NRF1 in WT and TKO mESCs as well as WT and *Dnmt3C*$^{KO/KO}$ sorted germ cells, where Spg's were used as a reference, with 20,000- 40,000 cells per sample.

2.12.1. CUT&Run Protocol

First, cells were incubated in Wash Buffer and bound to 10 µL of activated concanavalin A-coated beads for 10 minutes (Polysciences, 86057-10). After binding to the beads, the treatment proceeded either without fixation or with fixation using 0.2% formaldehyde for 2 minutes, followed by quenching with 125 mM glycine. The samples were washed on a magnet in Wash Buffer. The remaining CUT&Run protocol was performed as previously described (Janssens and Henikoff 2019). The used primary antibodies are listed in **Table 2**. Protein A-MNase fusion protein was diluted at 1:100 (produced by the Recombination Protein Platform of the Institut Curie, 0.785 mg/mL). The DNA was finally extracted using the DNA Clean & Concentrator kit.

2.12.2. Library Preparation & Sequencing

Library preparation for CUT&Run was tested in two different ways: Firstly, library preparation was performed with the standard NEBNext® Ultra™ DNA Library Prep kit for Illumina® With Size Selection (E7370) using the adapted protocol for CUT&Run samples from Liu 2019 with NEB adapters. Secondly, libraries were prepared as described in **chapter 2.11.2.** Library Preparation & Analysis using the xGEN ssDNA & Low-Input Library Preparation Kit (IDT, 10009817) following the manufacturer's protocol. Pooled libraries were sequenced on an Illumina NextSeq 2000 sequencing device as 2×155-nt paired-end reads, yielding 10-15 million read pairs per sample.

2.12.3 Bioinformatic Analysis of CUT&Run data

Paired-end reads were quality-trimmed using Trim Galore (v.0.6.10) with a minimum read length of 10 bp, and quality reports were generated using FastQC. Trimmed reads were aligned to the mm10 reference genome using Bowtie2 (v.2.5.3) in local alignment mode allowing a minimum fragment length of 10 bp and a maximum of 1000 bp, while excluding mixed and discordant alignments. Multimapping reads were assigned randomly as previously described (Teissandier et al. 2019). Since CUT&Run fragments can share identical starting and ending positions as the MNase has an inherent cutting preference, duplicates were retained. BAM files were first indexed using Samtools (v.1.10). Then, BigWig files were generated from the BAM files using bamCoverage (v. 2.27.1), with normalization set to CPM. DNMT3C targets were identified using published WGBS data (Barau et al. 2016). Downstream visualization of reads as heatmaps or metaplots were generated using deepTools (v.3.5.5). Peaks were called using MACS2 –narrowPeak using the IgG sample as a control file. The analysis of the CUT&Run in mESCs was performed by the Bioinformatics core facility.

2.13. Cleavage under Targets and Tagmentation

Cleavage under Targets and Tagmentation (CUT&Tag) was used to profile histone modifications and to profile NRF1 and SPIN1 in sorted germ cells with two replicates per genotype. The protocol essentially followed the published protocol from (Kaya-Okur et al. 2019) with a few modifications.

2.13.1. Transposome Preparation

For histone mark profiling, transposomes consisting of the in-house produced pA-Tn5 prepared by Protein Production core facility was used at a dilution of 1:200.

To this end, Mosaic end-adapters (ME-A, ME-B) and Mosaic end-reverse oligonucleotide (ME-Reverse) (**Table 1**) were diluted to 200 µM in annealing buffer (10 mM Tris pH 8, 50 mM NaCl, 1 mM EDTA).

Each oligo pair (ME-A + ME-Reverse and ME-B + ME-Reverse) was mixed to yield 100 µM annealed products. Samples were annealed in a PCR machine with the program: 95°C for 5 min, 71 cycles of 95°C (1 min, decreasing by 1°C/cycle), and 25°C for 5 min, followed by a 4°C hold. Annealed products (8 µL each) were combined with 100 µL of 5.5 µM Protein A-Tn5 fusion protein and gently mixed. The mixture was rotated at RT for 1 hour and stored at -20°C.

The efficiency of transposome loading and tagmentation was assessed by incubating 5 µL of the loaded and unloaded Tn5 enzyme with 150 ng of plasmid DNA in tagmentation buffer for 7 minutes at 55°C. After this, 0.1% SDS was added, and the mixture was incubated for an additional 7 minutes at 55°C, followed by proteinase K digestion for 7 minutes at 55°C. The reaction was then loaded onto a 0.7% agarose gel to visualize a smear representing the tagmented DNA and a band corresponding to the adapters.

For CUT&Tag profiling of other proteins, pre-loaded pA–Tn5 was used at a 1:400 dilution (Diagenode, C01070001).

2.13.2. CUT&Tag Protocol

Around 20,000 cells were bound to beads for native or fixed CUT&Tag as described in chapter **2.12.1.** CUT&Run Protocol. An optimal Digitonin concentration was identified, by incubating the cells for 30 min in different concentrations of Digitonin in Wash Buffer, followed by staining with Trypan Blue and counting live versus dead cells. The percentage of Digitonin, in which around 50% of the cell types were stained by Trypan Blue, was used as Dig-Wash-Buffer in CUT&Tag. All cells were incubated with Dig-Wash Buffer containing 0.05% Digitonin, with the exception of Spc's, which were supplemented with 0.025% Digitonin. The primary and secondary antibodies used for this method are listed in **Table 2**. Finally, DNA was extracted using the DNA Clean & Concentrator kit (Zymo, D4014).

2.13.3. Library Preparation & Sequencing

The libraries were amplified by standard PCR reaction with Q5 High-Fidelity DNA Polymerase (M0491S) using 15 cycles. Primer-dimer clean-up was performed using AMPure beads with a double-sided size selection by 0.3X and 1.3X ratio. Pooled libraries were sequenced on an Illumina NextSeq 2000 sequencing device as 2×155-nt paired-end reads, yielding 7-15 million read pairs per sample.

2.13.4. Bioinformatic Analysis of CUT&Tag data

CUT&Tag data was analyzed as described in **chapter 2.12.3** Bioinformatic Analysis of CUT&Run databut using RPGC for normalization for histone marks at sorted E15.5 ProSpg, Spg, Spc and for NRF1 libraries. With the following additional analysis steps:

Peak calling was performed using MACS2 (v.2.1.2) on BAM files, specifying the corresponding IgG sample as control for the chromatin marks (**Figure 3.0.3, Figure 3.0.5A,B**) and no control for NRF1 (**Figure 3.0.11D-H)**. The analysis was conducted using broad peak calling mode, with genome size set to 'mm' (for mouse) and otherwise default settings. To identify likely-bivalent, H3K4me3-marked, H3K27me3-marked and Uncategorized clusters of DNMT3C targets, called peaks from H3K4me3 and H3K27me3 CUT&Tag in *Dnmt3C*$^{KO/KO}$ were overlapped with DNMT3C targets using bedtools (v.2.29.2). Peaks commonly called in H3K4me3 and H3K27me3 datasets overlapping with DNMT3C targets were called likely-bivalent, peaks overlapping with only one of the two datasets and DNMT3C targets were called H3K4me3-marked or H3K27me3-marked, respectively. DNMT3C targets not overlapping with any peaks from both datasets were called Uncategorized.

Venn diagrams and peak overlaps from NRF1 CUT&Tag datasets with DNMT3C targets were analyzed and visualized by Ishita Amar using ChIPpeakAnno (v.3.34.1) using maxgap of 750 bp. Repeat family analysis was annotated using the RepeatMasker annotation (open-4.0.5-Repeat Library 20140131).

The remaining data (E14.5 ProSpg) was analyzed in collaboration. The data was analyzed as explained above with some differences: The reads from SPIN1 CUT&Tag were deduplicated using Picard MarkDuplicates v.2.24.0 –REMOVE_DUPLICATES after assessment of high duplication rate. Log2 enrichment over IgG controls were generated using deepTools bamCompare -bs 1 –normalizeUsing CPM –exactScaling –ignoreForNormalization MT –scaleFactorsMethod None. Positions of repetitive elements were extracted from a table of mouse mm10 repeatMasker annotations downloaded from the UCSC table browser and filtered for elements greater than 5 kb in length. Peak calling was performed using MACS2 callpeak on individual replicates as well as on the combined replicates, with IgG samples used as controls. The --keep-dup all parameter was applied to include duplicate reads in the peak calling model when present. To generate a set of high-confidence peaks, we filtered for peaks with a minimum coverage of 20 reads in the CUT&Tag sample and a peak score exceeding the mean peak score.

2.13 Supplemental Material

Table 1: Primers and Oligonucleotides

Name	Usage	Full Sequence (5' to 3')	Annealing
IAPez-LTR1 FWD	cloning	TAGGCACACATACACCTGCTTGTGGGAAGCCGCCCC	TGTGGGAAGCCGCCCC
IAPez-LTR1 REV	cloning	GGGGAGGAGGAAAGGTCATTTATTCGTCGCGTTCTCACGAT	TTATTCGTCGCGTTCTCACGA
L1spa FWD	cloning	TAGGCACACATACACCTGCTCGGAACTTAGGAAATTAGTCTGAACA	CGGAACTTAGGAAATTAGTCTGAACA
L1spa REV	cloning	TGGGGAGGAGGAAAGGTCATGGACGAAGAAGAAGGCCCAA	GGACGAAGAAGAAGGCCCAA
Extension FWD	cloning	TAGGCACACATACACCTGCT	TAGGCACACATACACCTGCT
Biotinylated extension REV	cloning	/5Biosg/TGGGGAGGAGGAAAGGTCAT	TGGGGAGGAGGAAAGGTCAT
Bisulfite converted IAP FWD	cloning	AGGTATATATATATTTGTTTGTGGGAAG	AGGTATATATATATTTGTTTGTGGGAAG
Bisulfite converted IAP REV	cloning	CGACAAAACTTTATTACTTACATCTTA	CGACAAAACTTTATTACTTACATCTTA
Bisulfite converted scrambled control IAP FWD	cloning	AGGTATATATATATTTGTTGTGTATTTG	AGGTATATATATATTTGTTGTGTATTTG
Bisulfite converted scrambled control IAP REV	cloning	AAAATTAAACTCGATCCCAA	AAAATTAAACTCGATCCCAA
Bisulfite converted L1T FWD	cloning	GTATTGAATTTAGGAAATTAGTTTGAA	GTATTGAATTTAGGAAATTAGTTTGAA

Bisulfite converted L1T REV	cloning	AAAATCAAATACTCTAATAATCTCTAAA	AAAATCAAATACTCTAATAATCTCTAAA
Bisulfite converted L1 scrambled contr FWD	cloning	TATTTAGGGTTGTTTGAAGTTAGA	TATTTAGGGTTGTTTGAAGTTAGA
Bisulfite converted L1 scrambled contr REV	cloning	AAAACTTCCTATCTACCTATAACTT	AAAACTTCCTATCTACCTATAACTT
Vasa-Cre FWD	genotyping	CACGTGCAGCCGTTTAAGCCGCGT	CACGTGCAGCCGTTTAAGCCGCGT
Vasa-Cre REV	genotyping	TTCCCATTCTAAACAACACCCTGAA	TTCCCATTCTAAACAACACCCTGAA
Nrf1-flox FWD	genotyping	GACACTGGCATGGAGTATTTG	GACACTGGCATGGAGTATTTG
Nrf1-flox REV	genotyping	GTCGTTAGACATCTCCCTTTTC	GTCGTTAGACATCTCCCTTTTC
Nrf1-flox-cKO FWD	genotyping	AACCGAAGTTCCTATTCCGAAGTTCC	AACCGAAGTTCCTATTCCGAAGTTCC
Nrf1-flox-cKO REV	genotyping	AACATAGCAGGTGCAGGTTTCC	AACATAGCAGGTGCAGGTTTCC
Dnmt3C-WT-FWD	genotyping	TGGCTCCCACAGATCTTACC	TGGCTCCCACAGATCTTACC
Dnmt3C-KO-FWD	genotyping	AACCCTCCAAATCCTTAGGC	AACCCTCCAAATCCTTAGGC
Dnmt3C-WT-KO-REV	genotyping	ACTATAGCCCATCAGTAAAGCC	ACTATAGCCCATCAGTAAAGCC
Asz1 FWD	qPCR	TCACTATCGCTGCTCTCGC	TCACTATCGCTGCTCTCGC
Asz1 REV	qPCR	AGCTCTGATAGTGCGCAGG	AGCTCTGATAGTGCGCAGG
IAP FWD	qPCR	CACATTCGCCGTCACAAGAT	CACATTCGCCGTCACAAGAT
IAP REV	qPCR	TGGCTCATGCGCAGATTATT	TGGCTCATGCGCAGATTATT
L1A_I-II-III FWD	qPCR	CTGCCCCAATCCAATCGCG	CTGCCCCAATCCAATCGCG

L1A_I-II-III REV	qPCR	GAGCGGAAGGGACTTGTGC	GAGCGGAAGGGACTTGTGC
L1T_I-II-III FWD	qPCR	AGAGGGTGCGCCAGAGAA	AGAGGGTGCGCCAGAGAA
L1T_I-II-III REV	qPCR	CCTGTCCTCTGGTCCGGA	CCTGTCCTCTGGTCCGGA
NRF1 TKO target FWD	qPCR	AAATAGGTATGCAGCCGGAC	AAATAGGTATGCAGCCGGAC
NRF1 TKO target REV	qPCR	AGGATCTACGGTAGGGATGG	AGGATCTACGGTAGGGATGG
ME-rev adapter	Transposome preparation	[PHO]CTGTCTCTTATACACATCT	
ME-A adapter	Transposome preparation	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	
ME-B adapter	Transposome preparation	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	
NRF1 FP unmethylated ctrl oligo	FP	/5Cy5/ATATCTGCGCATGCGCATAA	
NRF1 FP unmethylated ctrl oligo	FP	TTATGCGCATGCGCAGATAT	
NRF1 FP methylated ctrl oligo	FP	/5Cy5/ATATCTGCCCATGGGCATAA	
NRF1 FP methylated ctrl oligo	FP	TTATGCCCATGGGCAGATAT	
NRF1 FP Asz1	FP	/5Cy5/TTGCGCAAGCGCCTGGGACCTCTGGCGC	

promoter oligo		
NRF1 FP Asz1 promoter oligo	FP	GCGCCAGAGGTCCCAGGCGCTTGCGCAA
NRF1 FP Asz1 promoter oligo methylated	FP	/5Cy5/GCTCTG/iMe-dC/GCCTG/iMe-dC/GCCAGAGGTCCCAGG/iMe-dC/GCTTG/iMe-dC/GCAAG
NRF1 FP Asz1 promoter oligo methylated	FP	CTTG/iMe-dC/GCAAG/iMe-dC/GCCTGGGACCTCTGG/iMe-dC/GCAGG/iMe-dC/GCAGAGC

Table 2: Antibodies

Antibody target	**Species**	**Source**	**Cat. No.**	**Application**	**Dilution**
IgG control	Rabbit	Abcam	ab37415	all profiling methods	1:50
Rabbit IgG	Guinea pig	Antibodies	ABIN101961	CUT&Tag	1:100
Mouse IgG	Rabbit	Antibodies	ABIN101785	CUT&Tag	1:100
H3K27me3 (C36B11)	Rabbit	Cell Signalling	9733S	CUT&Tag	1:50
H3K27ac	Rabbit	Abcam	ab4729	CUT&Tag	1:50
H3K9me2	Mouse	Abcam	ab1220	CUT&Tag	1:50
H3K4me3	Rabbit	Merck Millipore	07-473	CUT&Tag	1:50
H3K9me3	Rabbit	Abcam	ab8898	CUT&Tag	1:50
NRF1	Rabbit	Abcam	ab34682	CUT&Tag, CUT&Run	1:50
NRF1	Rabbit	Abcam	ab175932	Immunofluorescense, ChIP, ULI-ChIP	1:600

L1ORF1p	Rabbit	Abcam	ab216324	Immunofluorescense, Western Blot	1:400, 1:1000
TRA98	Rat	Abcam	ab82527	Immunofluorescense, Western Blot	1:500
POL II (phospho S5)	Rabbit	Abcam	ab5131	Western Blot	1:100
α-TUBULIN	Mouse	Sigma	T5168	Western Blot	1:10000
IAP-POL	Rabbit	produced by Boulard lab		Immunofluorescense, Western Blot	1:400, 1:1000
SPIN1	Rabbit	Cell Signaling	89139S	CUT&Tag	1:50

3. Results

Silencing young retrotransposons through cytosine DNA methylation is crucial for spermatogenesis, as defects in methylating their promoters leads to their reactivation, meiotic failure, and infertility. However, the mechanisms driving retrotransposon reactivation in the absence of DNA methylation remain poorly understood.
In my PhD project, we addressed these fundamental questions by investigating TE expression in the absence of DNA methylation during male germ cell development, profiling the associated chromatin landscape, and identifying key regulators of TE expression. To overcome challenges inherent to the low input of germ cells and the repetitive nature of TEs, I optimized various chromatin profiling techniques, analysis of TE regulation.
This work not only facilitated a collaborative study with the O'Carroll lab, culminating in a Nature publication that characterized a piRNA pathway candidate and its interaction with L1 elements underlying a bivalent H3K4me3-H3K9me3 chromatin signature, but also led to the discovery of NRF1 as a regulator of unmethylated TEs and its interplay with H3K4me3-H3K27me3 marks, now under revisions in EMBO Reports. Importantly, the second project largely benefited from contributions by my colleague Styliani Eirini-Kanta, who performed key immunofluorescence and Western blot analyses, as well as our Bioinformatics Core Facility team, who supported me in the extensive genomic data analysis.

3.1. TE expression during male germ cell development

In this project, we utilized the *Dnmt3C* knockout mouse model (Barau et al. 2016; Jain et al. 2017) to investigate the expression of retrotransposons upon defective germline DNA methylation specifically on TE promoters. DNMT3C, guided by the piRNA pathway, selectively methylates young retrotransposon promoters active in germ cells (Barau et al. 2016; Zoch et al. 2024). Importantly, *Dnmt3C* mutant germ cells retain a nearly normal global DNA methylation profile, with defects confined to promoters of evolutionary young TEs (Barau et al. 2016). Leveraging this model, we examined TE activity through immunostaining, western blotting, and RNA sequencing, revealing distinct expression patterns aligned with the stages of spermatogenic development.

3.1.1. TE Protein Expression during male germ cell development

To study the progression of spermatogenesis in mice, we utilized the synchronized nature of the first wave of this process, enabling us to sample testes at precise embryonic and

postnatal stages (**Figure 1.8**). At embryonic day E15.5, germ cells are in the ProSpg stage, prior to the nuclear enrichment of MIWI2 and the initiation of piRNA-directed DNA methylation (Aravin et al. 2008; Zoch et al. 2020). To assess a later ProSpg stage when DNMT3C-mediated DNA methylation is more advanced, we collected testes at E18.5 (Barau et al. 2016). By P5, DNA methylation patterns are established in wild-type germ cells, which have transitioned into SSCs and differentiating Spg. At P10, meiosis begins, with the majority of the cells progressing to meiosis by P15. At P20, meiotic cells have surpassed the stage where *Dnmt3C*$^{KO/KO}$ germ cells typically arrest, and by P30, germ cells have completed the first wave of meiosis (**Figure 1.8**).

Figure 3.*0.1*

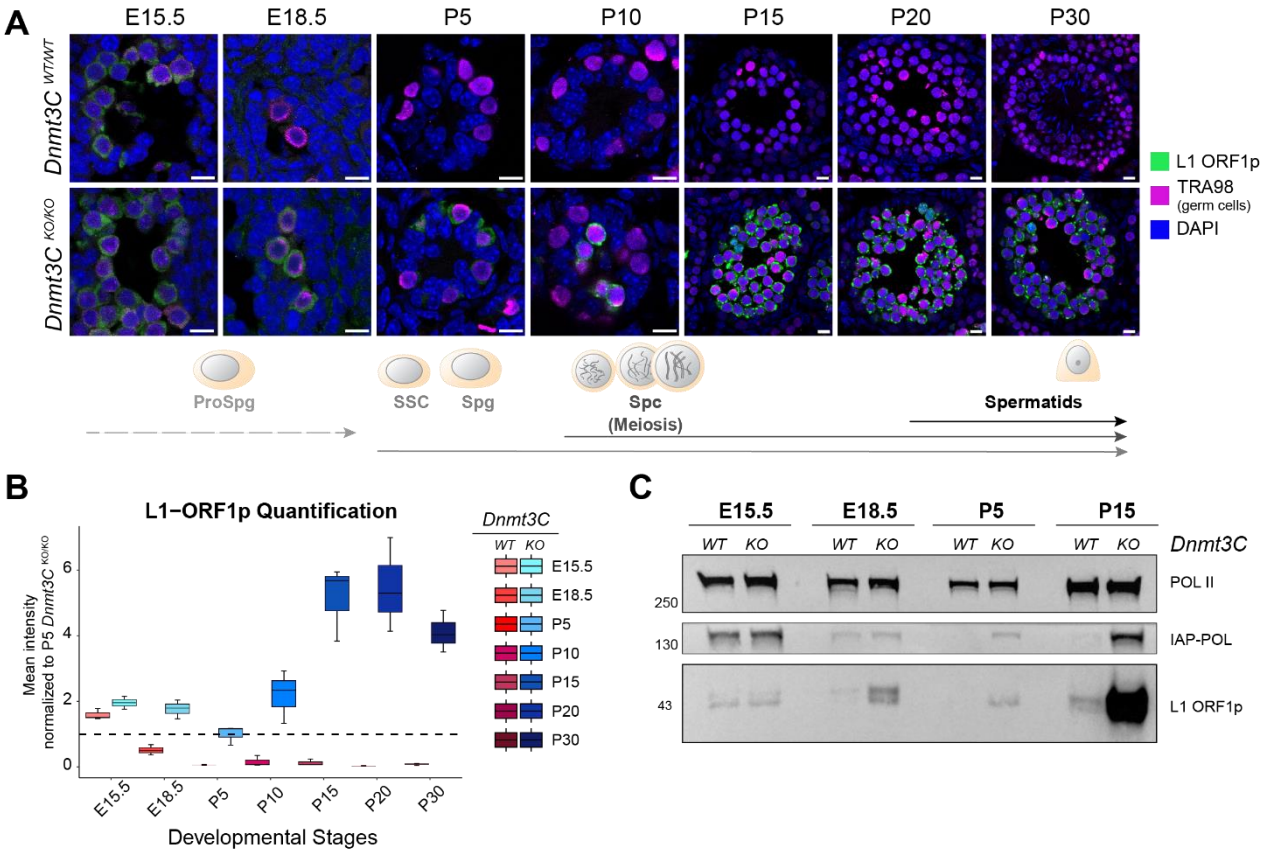

Figure 3.1: Protein Expression of TEs in wild-type and *Dnmt3C*$^{KO/KO}$ male germ cell development. A) Immunofluorescence staining of L1-ORF1p and the germ cell marker TRA98 in testis cryosections from wild-type and mice at different developmental time points. B) Box plots displaying mean intensity quantification of LINE1-ORF1p signal from (A). C) Representative Western blot analysis of LINE1-ORF1p, IAP-POL and RNA polymerase 2 (POL II) as a loading control from wild-type and *Dnmt3C*$^{KO/KO}$ testes at the indicated developmental time points. The blot was developed using Pico chemistry. The experiment was performed in biological duplicates.

To investigate TE activity, we focused on L1 ORF1 protein expression across these stages using immunofluorescence on testis cryosections (**Figure 3.0.1A, B**). As expected, at E15.5, L1-ORF1 was detectable in both wild-type and *Dnmt3C*$^{KO/KO}$ germ cells, consistent with the absence of *de novo* methylation at this stage (Molaro et al. 2014; Barau et al.

2016). In wild-type cells, L1 silencing was evident by E18.5. In contrast, *Dnmt3C*$^{KO/KO}$ cells showed persistent L1 ORF1 expression, reflecting failure to silence retrotransposons transcriptionally due to the loss of DNMT3C-mediated methylation (Barau et al. 2016). Interestingly, in *Dnmt3C*$^{KO/KO}$ spermatogonia at P5, L1-ORF1 protein levels were comparable to those at E15.5 (**Figure 3.0.1A-B**). This contrasts with the observations in other piRNA pathway mutants, such as *Mael*$^{KO/KO}$, *Dnmt3L*$^{KO/KO}$, and *Mili*$^{KO/KO}$, where L1-ORF1p was undetectable before meiosis (Soper et al. 2008; Zamudio et al. 2015; Di Giacomo et al. 2013). This difference likely reflects the specificity of our *Dnmt3C*$^{KO/KO}$ model, which affects only *de novo* methylation at evolutionary young retrotransposon promoters. However, consistent with findings from other models, L1 activity in *Dnmt3C*$^{KO/KO}$ germ cells increased significantly upon the onset of meiosis at P10 (**Figure 3.0.1A-B**).

Western blot analysis further validated these observations, showing L1-ORF1 protein in *Dnmt3C*$^{KO/KO}$ testes at all stages examined (**Figure 3.0.1C**). Differences in L1-ORF1 expression between wild-type and *Dnmt3C*$^{KO/KO}$ became apparent by E18.5 consistent with the progression of the piRNA pathway. L1-ORF1 protein being detectable in E15.5 and E18.5 *Dnmt3C*$^{KO/KO}$ germ cells, suggests that some L1 mRNAs are translated despite piRNA pathway-mediated post-transcriptional silencing (Aravin et al. 2008). At P5, L1-ORF1 levels were slightly reduced compared to E18.5 but increased markedly at P15, especially in *Dnmt3C*$^{KO/KO}$ cells. We also assessed IAP expression by western blot using an IAP-POL antibody in the same stages and observed a similar trend, showing high IAP upregulation at P15. Interestingly, IAP-POL levels were higher in E15.5 than in E18.5 and P5 *Dnmt3C*$^{KO/KO}$. However, a clear difference in IAP-POL levels between wild-type and *Dnmt3C*$^{KO/KO}$ cells emerged only by P5, with a pronounced increase by P15 (**Figure 3.0.1C**). This may suggest that IAPs could be silenced later by DNMT3C than L1s and therefore IAPs remain expressed to the same level in E18.5 *Dnmt3C*$^{KO/KO}$ as in wild-type.

3.1.2. TE mRNA Expression during male germ cell development

Next, we performed RNA sequencing on FACS-isolated germ cells from E15.5 prospermatogonia (ProSpg), postnatal premeiotic spermatogonia (Spg), and meiotic spermatocytes (Spc) from wild-type and *Dnmt3C*$^{KO/KO}$ mice to characterize retrotransposon transcriptional activity in response to the loss of promoter DNA methylation (**Figure 2.2.2A-C**).

At E15.5, we anticipated no significant differences between wild-type and *Dnmt3C*$^{KO/KO}$ germ cells, as the TGS piRNA pathway and DNMT3C-mediated DNA methylation initiates only at E16.5 when MIWI2 becomes nuclear (Aravin et al. 2008). Consistent with this, we observed only minor differences in TE expression at this stage (**Figure 3.0.2A**). However, in *Dnmt3C*$^{KO/KO}$ spermatogonia, we detected a pronounced transcriptional upregulation of

the most active ERVK subclasses (Dewannieux et al. 2004), particularly IAPs and MMERVK10C, as well as the L1 families L1MdTf, L1MdA, and L1MdGf (**Figure 3.0.2B**). While IAP upregulation due to the loss of DNA methylation in piRNA pathway-defective germ cells has been previously reported (Vasiliauskaitė et al. 2018), our novel finding of L1 transcriptional upregulation in spermatogonia provides orthogonal validation of the L1-ORF1 protein expression we detected by western blot and immunofluorescence before meiosis (**Figure 2.2.2**).

Of note, in wild-type spermatocytes, we observed upregulation of several ancient TE families (**Figure 3.0.2C**). However, this likely only reflects differences in germ cell composition compared to *Dnmt3C*$^{KO/KO}$, where germ cells arrest at the pachytene stage of meiosis (Barau et al. 2016). This germ cell arrest resulted in sorting later meiotic stages in Spc wild-type with a distinct transcriptional profile, showing that overall TEs get more transcriptionally active at meiosis (**Figure 2.2.2C-D**). In *Dnmt3C*$^{KO/KO}$ spermatocytes, we confirmed further transcriptional upregulation of L1 elements during meiosis, while ERVs maintained consistent expression levels (**Figure 3.0.2C**).

To systematically compare the transcriptional responses of ERVs and L1s to the loss of promoter DNA methylation across spermatogenesis, we plotted the average Z-scores and the logarithmic transformed (Log) Transcripts per Million (TPM) values from TE expression in ProSpg, Spg, and Spc (**Figure 3.0.2D**). Surprisingly, L1 expression did not differ significantly between ProSpg and wild-type postnatal stages, contrasting with the protein expression patterns we observed in the immunofluorescence staining (**Figure 3.0.2D, Figure 3.0.1A**). In contrast, IAP-POL protein expression aligned with ERVK transcription in ProSpg **(Figure 3.0.2D, Figure 3.0.1A)**. Strikingly, this comparison revealed that both ERVK and L1 are upregulated in all postnatal stages upon loss of promoter DNA methylation and while ERVK expression remained relatively stably expressed, L1 transcription increased substantially in *Dnmt3C*$^{KO/KO}$ spermatocytes.

Figure 3.*0.2*

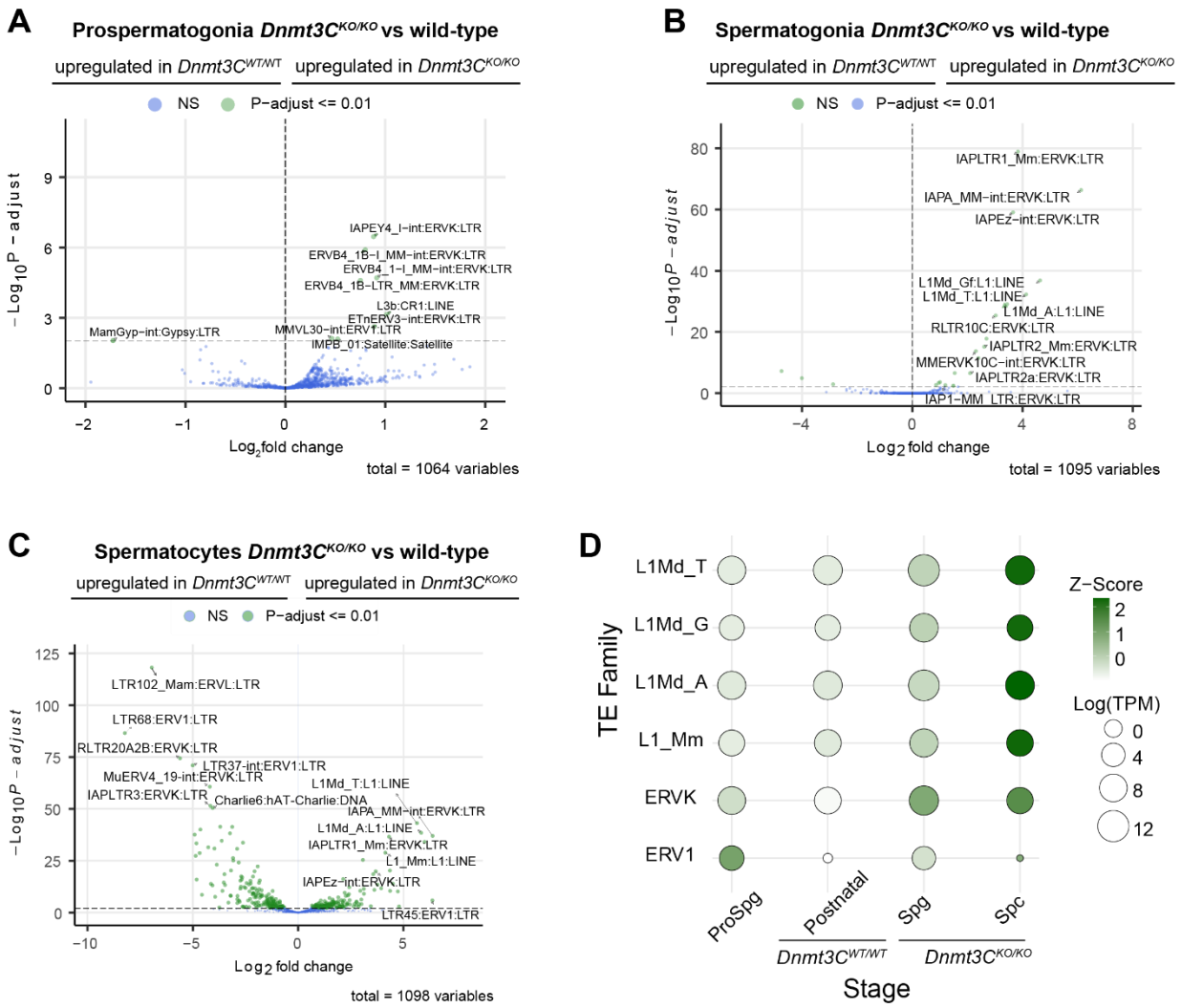

Figure 3.2: Transcriptional Expression of TEs in wild-type and *Dnmt3C*$^{KO/KO}$ male germ cell development. A) Volcano plots showing -log10P-adjusted values and log2 fold changes from three biological replicates of differentially expressed TE-transcripts between wild-type and *Dnmt3C*$^{KO/KO}$ sorted germ cells at prospermatogonia stages. The display of TE names follows the pattern TEname:Family:Class. B) as (A) at spermatogonia stages. C) as (A) at spermatocyte stages. D) Bubble chart displaying average Z-score and log(TPM) of expression values from three biological replicates of ProSpg (wild-type and *Dnmt3C*$^{KO/KO}$), wild-type and *Dnmt3C*$^{KO/KO}$ (Spg and Spc) for TE families that were differentially expressed in Spg and Spc.

Taken together, our data reveal a dynamic regulation of retrotransposon expression in germ cells lacking DNA methylation at their promoters. While ERVK expression in *Dnmt3C*$^{KO/KO}$ mutants remains consistently expressed, L1 transcription markedly increases at meiosis. This expression pattern suggests that the activating cues for demethylated L1s and ERVKs differ and shift during spermatogenesis.

3.2. TE regulation before silencing by the piRNA pathway

To understand how TEs become expressed and active in the absence of DNA methylation, it is important to examine the regulatory mechanisms and chromatin modifications underlying before the piRNA pathway establishes DNA methylation to silence TEs. Our observations indicate that TEs exhibit distinct expression patterns already in embryonic germ cells. Specifically, L1 elements appear to be silenced earlier, showing a reduction in expression as early as E15.5, whereas ERVs seem to remain expressed until E18.5 (**Figure 3.0.1, Figure 3.0.2D)**. However, how different TE families are regulated before silencing by DNMT3C-mediated DNA methylation remains poorly understood. To address this, we sought to characterize the chromatin landscape and identify the factors regulating TEs in ProSpg prior to the onset of the TGS piRNA pathway at E16.5.

3.2.1. SPIN1 & SPOCD1 interact with H3K4me3-H3K9me3-marked young LINE1s

To uncover additional components of the TGS piRNA pathway, a previous study that characterized SPOCD1 as a member of this pathway performed IP-MS to identify SPOCD1 interactors. Alongside other important interactors such as DNMT3L, DNMT3A, C19ORF48, and components of chromatin remodeling complexes, the study identified SPINDLIN1 (SPIN1) as a novel SPOCD1 interactor (Zoch et al. 2020).

SPIN1 is a chromatin reader protein containing three Tudor-like domains, enabling it to bind to the chromatin marks H3K4me3 and H3K9me3 (Yang et al. 2012; Lachner et al. 2001). In a collaborative study with the O'Carroll lab, we demonstrated that SPOCD1 and SPIN1 interact directly and that this protein complex specifically associates with nucleosomes modified by H3K4me3-H3K9me3 in vitro (Dias Mirandela et al. 2024). Building on this, after optimizing TF profiling methods for low-input material (detailed in **Chapter 3.3.2.3**), I performed CUT&Tag profiling in E14.5 sorted ProSpg (**Figure 2.2.2A**) to investigate the presence of the histone marks H3K4me3, H3K9me3, and SPIN1 at young TEs. We seleceted E14.5 as the time point for our experiments because SPIN1 and SPOCD1 are already expressed, but MIWI2 is not yet nuclear.

Our profiling revealed that the repressive chromatin mark H3K9me3 is highly enriched at IAPs and L1 families at E14.5. Interestingly, at younger L1 families (e.g., L1Md_T, L1Md_A, L1Md_Gf), H3K9me3 co-occurs with the active histone mark H3K4me3 (**Figure 3.0.3A-C**), suggesting that these young L1 families may recruit the SPIN1-SPOCD1 complex. Consistently, SPIN1 profiling showed its occupancy specifically at these young

L1 families, with no binding detected at older L1 elements or IAPs that were not marked by H3K4me3-H3K9me3 (**Figure 3.0.3**) (Dias Mirandela et al. 2024).

Together, these findings demonstrate that young L1 families, but not older copies or IAPs, are marked by the bivalent chromatin mark H3K4me3-H3K9me3. This specific chromatin signature recruits the SPIN1-SPOCD1 complex, potentially initiating the TGS piRNA pathway. Furthermore, by characterizing testes from mice with disrupted SPOCD1-SPIN1 interaction, we demonstrated that this interaction is essential for normal germ cell development and that its loss leads to DNA methylation deficits and upregulation of young L1 families (Dias Mirandela et al. 2024).

Figure 3.*0.3*

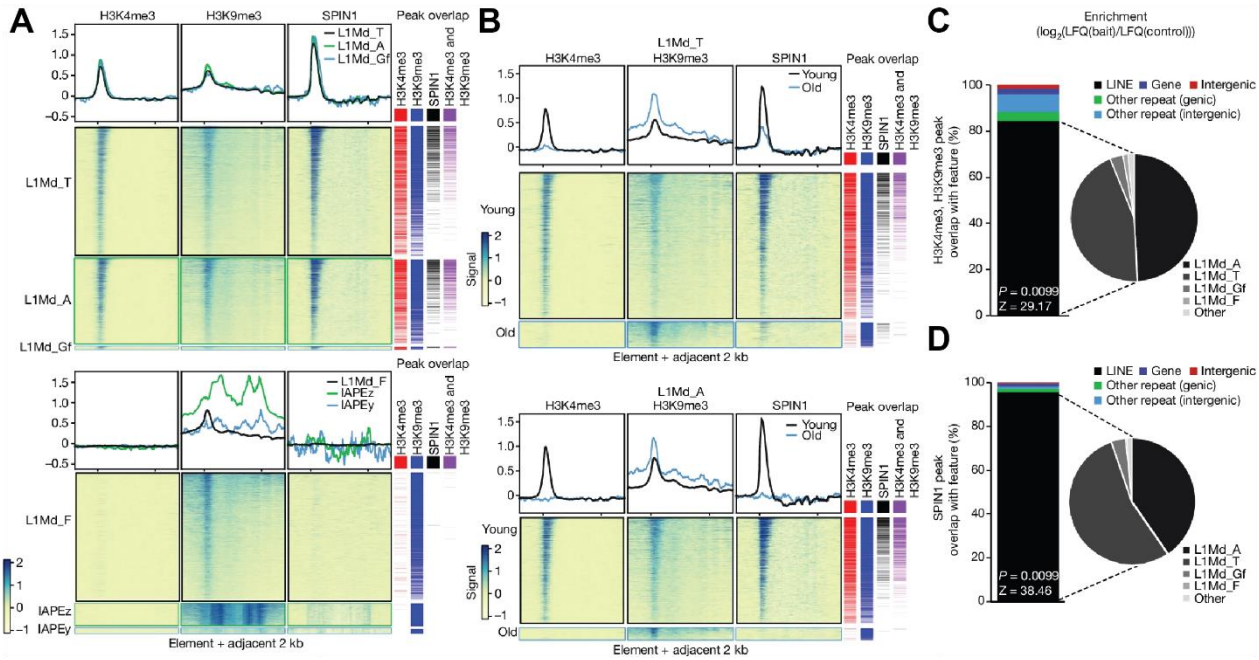

Figure 3.3: H3K4me3, H3K9me3 and SPIN1 mark young L1 elements prior to MIWI2-directed DNA methylation. A) Metaplots, Heatmaps and statistically significant peak overlaps for L1Md_T, L1Md_A, L1Md_Gf, L1Md_F, IAPEz, IAPEy TE families from CUT&Tag data for H3K4me3, H3K9me3 and SPIN1 on E14.5 ProSpg. Data are merged from two (H3K4me3, H3K9me3) and three (SPIN1) biological replicates. B) as in (A) for young and old L1Md_T and L1Md_A copies. C) Column charts show statistically signifcant overlaps of H3K4me3 and H3K9me3 peaks and their distribution in the genome. D) as in (C) for SPIN1 peaks. P-values and Z-scores from one-tailed permutation tests for statistical significance of overlaps of CUT&Tag peaks with LINE1 elements are shown. This image is taken from Dias Mirandela et al. 2024.

3.2.2. The chromatin landscape at TEs in E15.5 ProSpg

Having profiled the presence of the specific histone marks H3K4me3 and H3K9me3 in E14,5, we set out to identify the distribution of additional key histone marks, H3K4me3, H3K27me3, H3K9me2, and H3K9me3, in sorted E15.5 ProSpg using CUT&Tag. These histone marks were chosen as key histone marks to characterize chromatin states in the absence of DNA methylation during male germ cell development.

We mapped these chromatin marks at the evolutionary young TEs that have defective DNA methylation at their promoters in the absence of DNMT3C, and are therefore direct targets of the piRNA pathway. We focused most of our analysis throughout this project on these TEs and refer hereafter to these as DNMT3C targets (Barau et al. 2016). This analysis revealed that both active chromatin marks, H3K4me3 and H3K27ac, and repressive marks, H3K27me3, H3K9me2, and H3K9me3, mark DNMT3C targets prior DNMT3C-mediated DNA methylation (**Figure 3.0.4A**).

Mapping these marks on TE subfamilies separately, we found that the IAP and MMERVK10C families, which belong to the ERVK subclass, are marked by the repressive histone marks H3K27me3 and H3K9me3, as well as the activating enhancer mark H3K27ac at this stage (**Figure 3.0.4**). In contrast, the evolutionarily young L1 families, L1MdA and L1MdT exhibit all analyzed histone marks, including H3K4me3, H3K27me3, H3K9me2, H3K9me3, and H3K27ac, with the highest enrichment observed for H3K4me3 and H3K27me3. On the other hand, the older L1 family, L1MdF, shows weak enrichment for H3K4me3 but high enrichment for H3K27me3 (**Figure 3.0.4B)**.

The observation that younger L1 families are marked by more open chromatin marks compared to ERVK families at E15.5 ProSpg aligns with our E14.5 CUT&Tag profiling data (**Figure 3.0.3**). However, this chromatin landscape contrasts with the expression patterns we observed at this stage (**Figure 3.0.2D**), where ERVK families showed higher expression levels than L1s. This discrepancy suggests that the differential expression of these TE families is likely regulated by mechanisms other than the chromatin state alone.

Figure 3.*0.4*

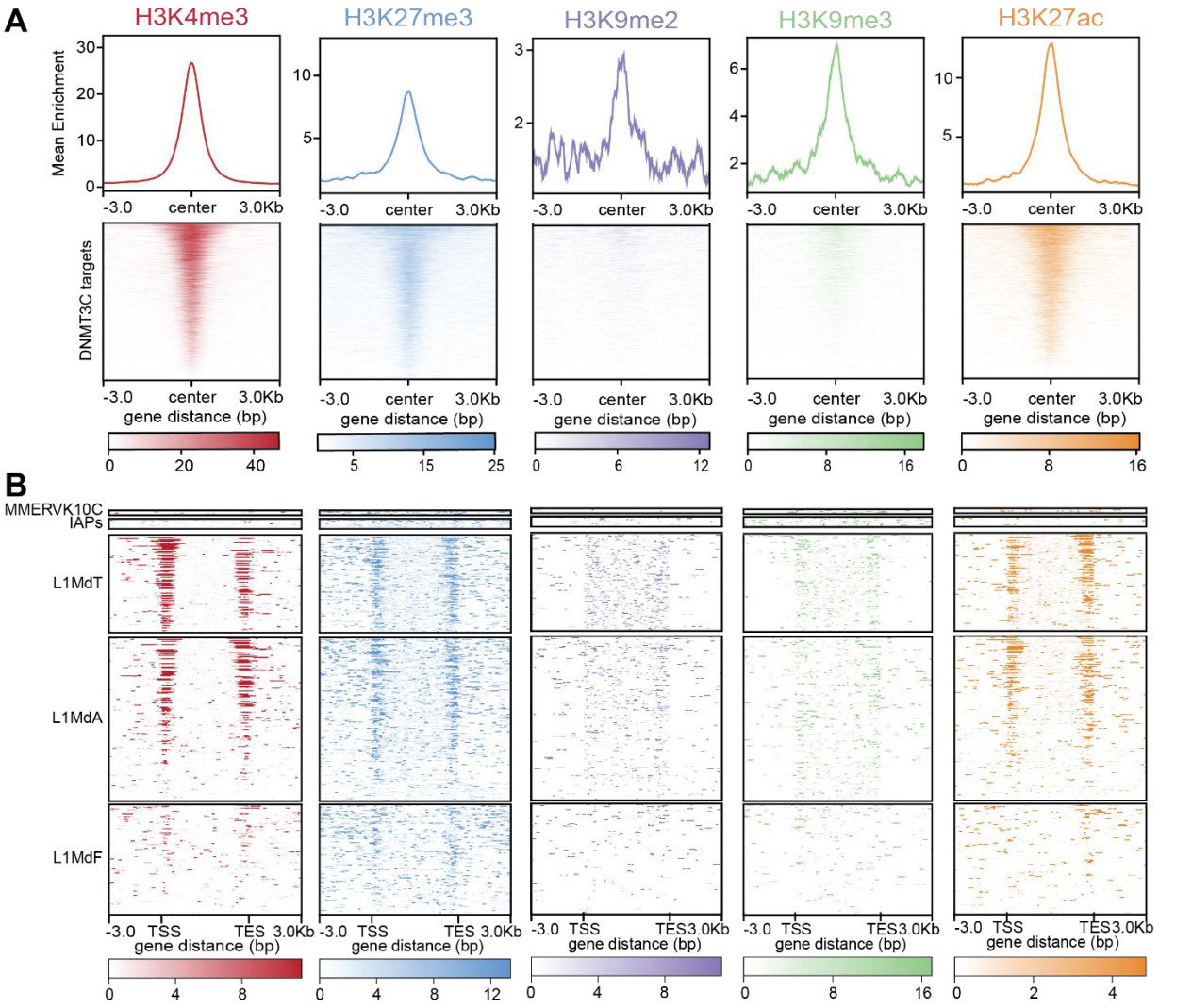

Figure 3.4: Chromatin landscape at retrotransposons at ProSpg E15.5. A) Metaplots and Heatmaps from CUT&Tag proflining of H3K4me3 (red), H3K27me3 (blue), H3K9me2 (purple), H3K9me3 (green) and H3K27ac (orange) on DNMT3C targets centers from E15.5 ProSpg. B) Heatmpas from CUT&Tag profiling of H3K4me3 (red), H3K27me3 (blue), H3K9me2 (purple), H3K9me3 (green) and H3K27ac (orange) on Transcription start site (TSS) and Transcription end site (TES) of MMERVK10C, IAPs, L1MdT, L1MdA, and L1MdF in E15.5 ProSpg. Data are merged from two biological replicates.

3.3. TE regulation in postnatal stages upon defective silencing by DNA methylation

As shown before, the failure to methylate young TE promoters after germline reprogramming, leads to IAP upregulation together with mild L1 upregulation prior to meiosis and substantial L1 upregulation at the onset of meiosis, ultimately resulting in meiotic arrest in male mice. To investigate why DNA methylation is essential for silencing

TEs and what leads to this differential TE expression pattern, we aimed to understand how TEs are regulated during postnatal spermatogenesis in the absence of DNA methylation and how this regulation correlates with the observed TE expression patterns. To address this, we sought to characterize the chromatin landscape at TEs in postnatal stages and test the regulatory roles of potential TE-binding candidates.

3.3.1. A switch from a bivalent to an active chromatin state underpins TE expression at meiosis

A potential mechanism regulating the differential expression pattern of unmethylated retrotransposons could involve the local presence of repressive or active chromatin marks at TE promoters. Previous studies in mESCs showed for example that hypomethylated retrotransposons become enriched with repressive H3K9me3 and H3K27me3 histone modifications (Walter et al. 2016). Additionally, there is evidence that repressive H3K9me2 can maintain TE repression independently of DNA methylation in *Dnmt3L* and *Miwi2* mutants before meiosis (Di Giacomo et al. 2013; Zamudio et al. 2015).

To investigate whether the differential TE expression pattern observed in *Dnmt3C* mutants correlates with specific chromatin dynamics during spermatogenesis, we profiled the histone modifications H3K4me3, H3K27me3, H3K9me2, H3K9me3, and H3K27ac using CUT&Tag at the spermatogonia (Spg) and spermatocyte (Spc) stages.

As expected, in postnatal wild-type stages (Spg and Spc), where DNMT3C targets are stably repressed by DNA methylation, only the repressive marks H3K9me3 and H3K9me2 were detected, which often correlate with the presence of DNA methylation (**Figure 3.0.5A**). In contrast, in *Dnmt3C*$^{KO/KO}$ postnatal germ cells lacking DNA methylation, H3K9me3 levels were significantly reduced, and other chromatin marks became more prominent. In Spg, H3K4me3, H3K27me3, and H3K9me2 were highly enriched but decreased slightly in Spc, while H3K27ac increased over DNMT3C targets. This is consistent with earlier findings showing that H3K9me2 represses retrotransposons in DNA methylation-deficient settings before meiosis (Zamudio et al. 2015; Di Giacomo et al. 2014). Additionally, the presence of the active H3K4me3 and H3K27ac marks at DNMT3C targets (**Figure 3.0.5A**), aligns with their transcriptional activity in *Dnmt3C* mutants (**Figure 3.0.2**). The simultaneous presence of repressive H3K27me3 and active H3K4me3 especially in Spg, might indicate a bivalent H3K4me3-H3K27me3 signature and could silence the DNMT3C targets. Whereas, the reduction of H3K27me3 and gain of H3K27ac at Spc stages, along with the persistence of H3K4me3, suggests a shift from a bivalent H3K4me3-H3K27me3 state to an active chromatin configuration.

Figure 3.*0.5*

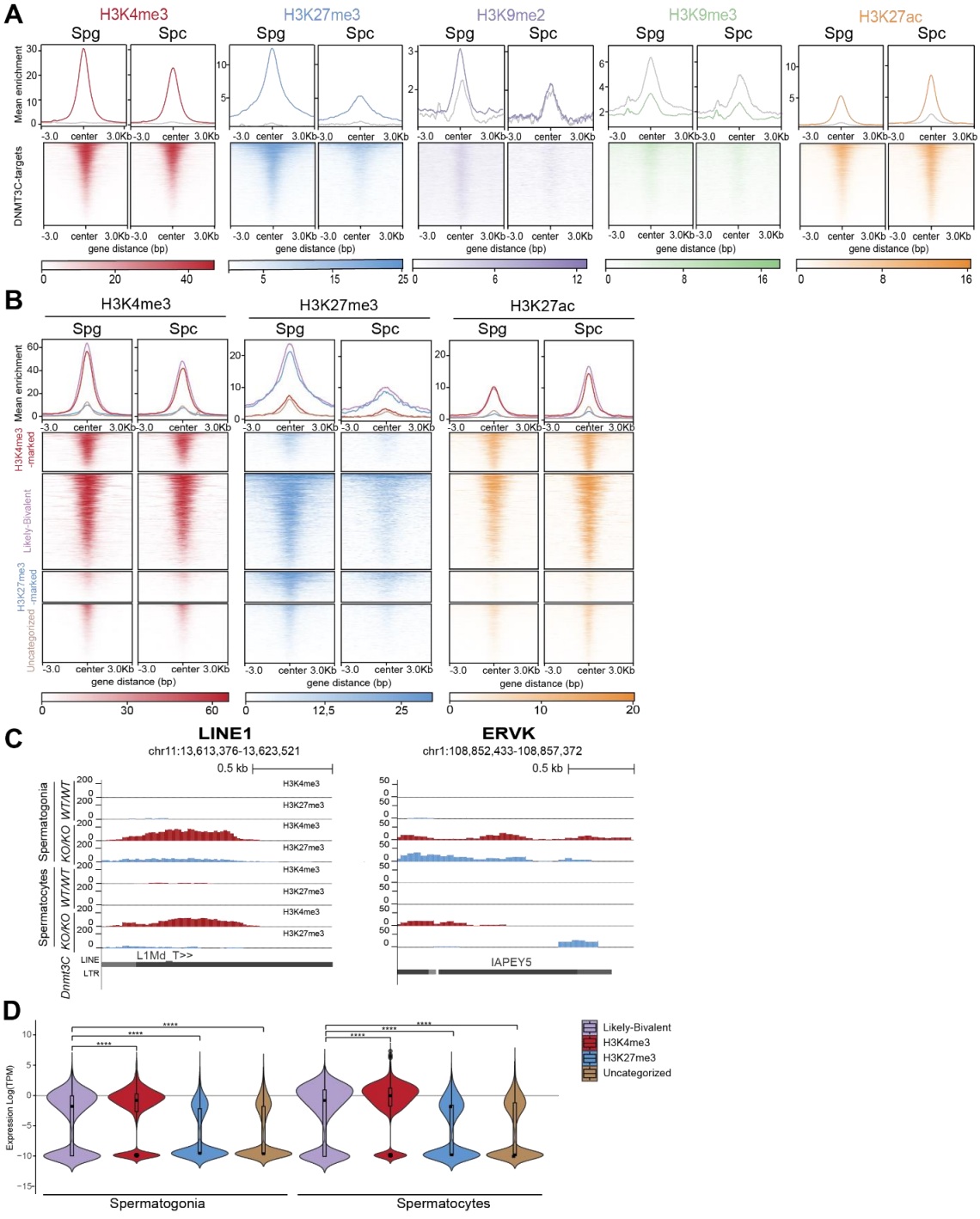

Figure 3.5: A transition from bivalent to an active chromatin state marks DNMT3C targets in spermatogenesis in the absence of DNA methylation A) Metaplots and Heatmaps from CUT&Tag profiling of H3K4me3 (red), H3K27me3 (blue), H3K9me2 (purple), H3K9me3 (green) and H3K27ac (orange) on DNMT3C targets centers in Spg and Spc. B) Heatmpas and metaplots from H3K4me3, H3K27me3 and H3K27ac CUT&Tag on H3K4me3-marked, likley-bivalent, H3K27me3-marked categories of DNMT3C targets in *Dnmt3C*$^{KO/KO}$ Spg and Spc. C) Track examples of bivalently-marked LINE1 and ERVK copies in wild-type and *Dnmt3C*$^{KO/KO}$ Spg and Spc. D) Violine plots illustration expression from TE copies clustered as Bivalent, H3-K4me3-marked, H3K27me3-marked or Uncategorized copies. The asterics depict statistical significance by t-test using the Bonferroni correction.

Bivalent chromatin is thought to regulate developmental gene expression by poising genes for activation (Bernstein et al. 2006; Boulard et al. 2015). In embryonic germ cells for instance, bivalent marks maintain germline-specific genes in a poised state, allowing stage-specific activation (Sachs et al. 2013). Bivalency may also protect promoters from aberrant DNA methylation (Bernstein et al. 2006). To explore whether the bivalency observed at TEs in *Dnmt3C* mutant Spg reflects a similar poising mechanism for later expression, we analyzed the dynamics of bivalent marks during germline development.

Using peak calling on H3K4me3 and H3K27me3 CUT&Tag data from *Dnmt3C* mutant Spg, we categorized DNMT3C targets into four chromatin states: likely-bivalent (overlapping H3K4me3 and H3K27me3), H3K4me3-marked, H3K27me3-marked, and uncategorized (no peaks) (**Figure 3.0.5B, C**). In Spg, DNMT3C targets were predominantly likely-bivalent, marked by H3K4me3 and H3K27me3 (**Figure 3.0.5B**). By Spc, likely-bivalent targets exhibited widespread H3K27me3 reduction and an increase in active H3K27ac (**Figure 3.0.5B**). H3K4me3-marked targets followed a similar pattern but displayed low H3K27me3 levels even in Spg. In contrast, H3K27me3-marked targets showed reduced H3K27me3 in Spc without gaining active marks like H3K4me3 or H3K27ac (**Figure 3.0.5B**). These results indicate that likely bivalent H3K4me3-H3K27me3 DNMT3C targets in Spg transition to a H3K4me3-H3K27ac configuration in meiosis.

We next examined the relationship between bivalency and TE transcription. Comparing TE expression between Spg and Spc stages across the four chromatin categories revealed that bivalent and H3K4me3-marked TEs exhibited the highest expression levels in *Dnmt3C* mutants (**Figure 3.0.5D**). H3K4me3-marked targets showed consistently high expression, while bivalent targets displayed a more bimodal distribution, with some copies being expressed at high levels comparable to H3K4me3-marked targets and others being expressed at low levels. This suggests that bivalent TEs follow distinct expression patterns (**Figure 3.0.5D**).

To determine whether chromatin signatures correlated with specific TE families, we compared these signatures between L1 and ERV families. Among L1 families, the percentage of H3K27me3-marked copies remained relatively stable, while for H3K4me3-marked copies it increased progressively in younger families: L1MdT > L1MdA > L1MdGf > L1MdF (Figure 3.6A). This suggests a direct relationship between retrotransposition potential and active chromatin marks at promoters of younger L1 families in the absence of DNA methylation. For ERVs, ERVL showed the highest percentage of H3K4me3-marked copies, followed by ERV1 and ERVK. Interestingly, despite IAP and MMERVK10C, which belong to the ERVK subclass, being the most upregulated ERVs in *Dnmt3C* mutants **(Figure 3.0.2D)**, ERVK displayed fewer H3K4me3-marked and more

H3K27me3-marked copies (**Figure 3.0.6A**). This implies that IAPs and MMERVK10C may rely on the expression from fewer elements with stronger promoters (Rebollo et al. 2020).

Figure 3.*0.6*

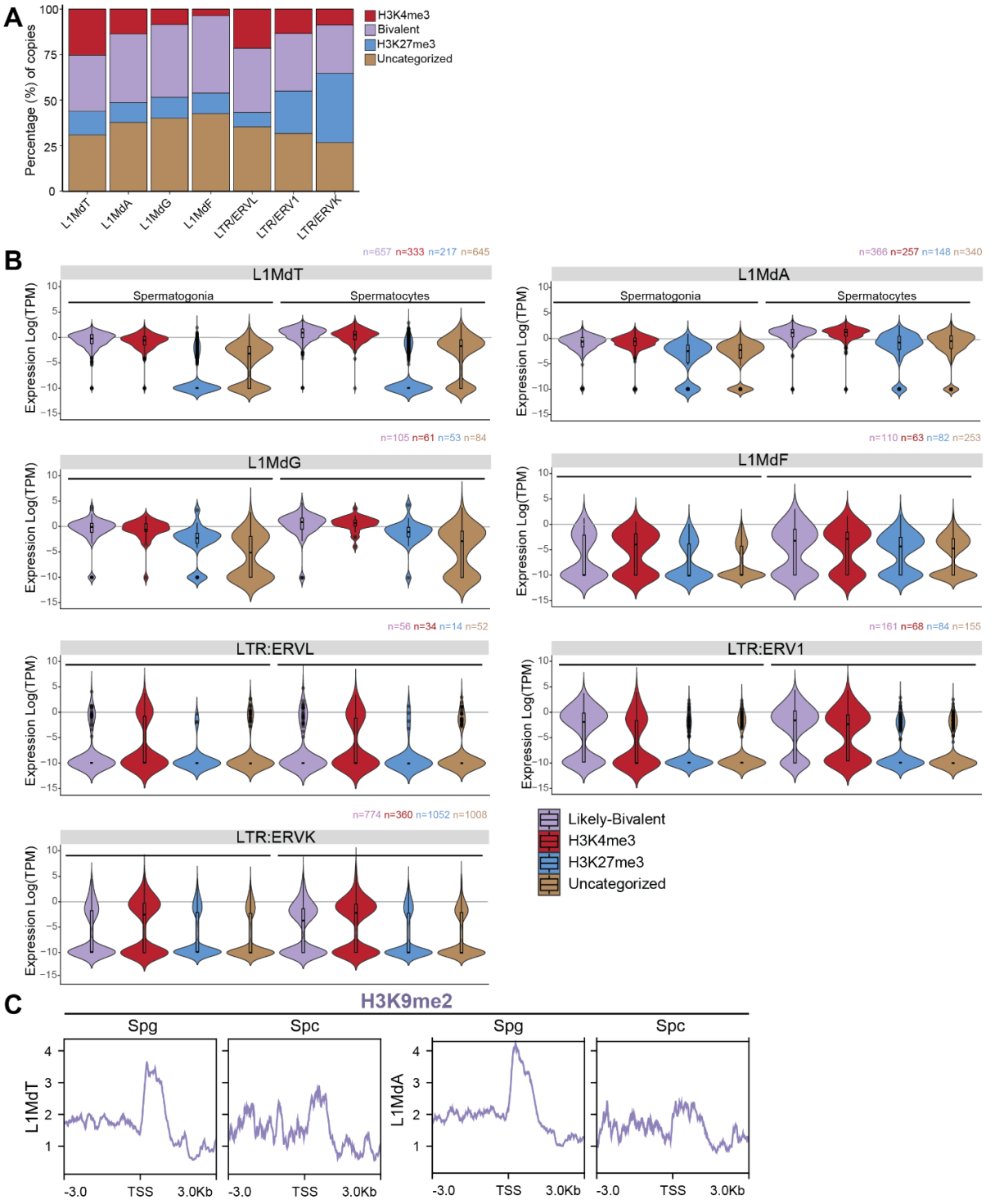

Figure 3.6: TE families differ in enrichment of bivalent chromatin marks and its correlation with expression. A) Column chart showing the percentage of TE family copies classified into likely-bivalent, H3K4me3-marked, H3K27me3-marked, uncategorized DNMT3C targets in *Dnmt3C*$^{KO/KO}$ Spg. B) Violin plots showing log(TPM) expression values obtained from three biological replicates using TElocal of the TE subfamilies L1MdT, L1MDA, L1MdG, L1MdF, ERVL, ERV1, ERVK categorized into likely-bivalent, H3K4me3-marked, H3K27me3-marked and Uncategorized. C) Metaplots showing mean enrichment of H3K9me2 CUT&Tag at L1MdT and L1MdA TE families in *Dnmt3C*$^{KO/KO}$ Spg and Spc.

To assess whether the observed variable TE expression was linked to the different chromatin signatures within families, we compared expression across these signatures at different TE families (**Figure 3.0.6B**). H3K4me3-marked copies of most L1 families, except L1MdF, did not exhibit higher expression than likely-bivalent copies. H3K27me3-marked L1MdT copies appeared silenced, while other L1 families were less effectively repressed by H3K27me3 (**Figure 3.0.6B**). This suggests that while L1MdT copies primarily relied on H3K27me3 for silencing in Spg, L1MdA elements might depend on a different chromatin mark. Given that L1 elements have previously been shown to be repressed by H3K9me2 in Spg (Di Giacomo et al. 2014; Zamudio et al. 2015), we compared H3K9me2 enrichment between L1MdT and L1MdA. Our analysis revealed a stronger accumulation of H3K9me2 at L1MdA elements in the Spg of *Dnmt3C* mutants, indicating that these elements may rely more on H3K9me2, rather than H3K27me3, for silencing (**Figure 3.0.6C**). In contrast, ERVL and ERVK were more highly expressed when H3K4me3-marked compared to likely-bivalent or H3K27me3-marked copies, suggesting that LTR elements are more effectively repressed by H3K27me3 (**Figure 3.0.6B**).

Overall, our findings indicate that in the absence of DNA methylation, bivalent and H3K4me3-marked DNMT3C targets are more susceptible to transcriptional reactivation than H3K27me3-marked or uncategorized TEs. Additionally, H3K27me3 is more effective at repressing ERVs than LINE1 elements, while variable expression across chromatin signatures points to additional regulatory mechanisms modulating TE activity in the absence of DNA methylation.

3.3.2. Identifying candidate regulators of TEs

Having characterized the impact of DNA methylation on TE transcription and its correlation with histone methylation changes during spermatogenesis, we sought to identify DNA methylation-dependent binding proteins potentially responsible for TE regulation. Previous studies in cell culture models proposed two potential mechanisms how DNA methylation can mediate TE silencing: recruitment of repressive complexes via TFs that bind specifically to methylated cytosines (Boulard et al. 2020), or the inhibition of DNA methylation–sensitive TF binding (Kaluscha et al. 2022). However, the recruitment of repressive complexes to TEs in germ cells has not been shown yet and the identity of methyl-sensitive TFs regulating TEs in germ cells remains unknown, leaving the precise silencing mechanism elusive.

To identify TFs that bind retrotransposon promoters in a DNA methylation-dependent manner, we performed an unbiased proteomic screen using biotin/streptavidin-based DNA bait pulldown experiments with nuclear extracts from mouse germ cells, followed by mass spectrometry. DNA baits were derived from promoters of retrotransposons that highly reactivate in *Dnmt3C*$^{KO/KO}$ mutants: a consensus IAP LTR1a promoter and L1MdT

monomers (repeats 1 and 2) (**Figure 2.2.3**). Scrambled DNA sequences with matched length, nucleotide composition, and CpG positions served as controls. Half the baits were in vitro methylated at CpG sites, with methylation confirmed via bisulfite sequencing (**Figure 2.2.3**).

Among interesting interactors, which we found in our IP-MS, the widely interspaced zinc finger protein (WIZ), which stabilizes the G9a-GLP complex mediating the repressive histone mark H3K9me2, is an interesting repressor candidate for methylated IAPs (Ueda et al. 2006). Additionally, other proteins were reported to be critical for germline development, making them interesting candidates that could be hijacked by TEs in the germline. This includes the dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX1) protein encoded by *NR0B1*, which we found to bind unmethylated IAPs, the CAMP-responsive element binding protein 1 (CREB1) (Martianov et al. 2010), which was enriched on unmethylated IAPs and the nuclear respiratory factor 1 (NRF1), enriched on both unmethylated IAPs and LINE1s. Mutations in DAX1 and NRF1 were reported to lead to hypogonadism (Gupta et al. 2022, Wang et al. 2017). Furthermore, other proteins such as phosphatidylinositol-4-phosphate 3-kinase C2 domain-containing alpha polypeptide (PIK3C2A), which was enriched on unmethylated L1s and CAMP-responsive element binding protein 1 (CREB1) have been implicated in TE regulation already (Mita et al. 2020, Kaluscha et al. 2022) (**Figure 3.0.7A, B**). Strikingly, we also found TFs enriched on unmethylated TE probes, which were shown to be sensitive to DNA methylation: NRF1 was previously shown to be a DNA methylation-sensitive TF that is outcompeted at most of its motifs when the CpG sites are methylated (Domcke et al. 2015) and the DNA methylation-sensitive TF CREB1 was previously shown to bind IAPs in DNA-methylation-deficient neurons (Kaluscha et al. 2022). Although we detected CREB1 in the unmethylated IAP pulldowns, it was not enriched in a sequence-specific manner (**Figure 3.0.7A, Table *5***). Among the proteins consistently detected in pulldowns with unmethylated baits, NRF1 was enriched as a sequence-specific binder of both IAP and L1 baits (**Figure 3.0.7A, B**).

To further investigate TFs, which could bind TE promoters, we performed *in silico* Analysis of Motif Enrichment (AME) (McLeay and Bailey 2010) on IAP and L1 fractions of DNMT3C targets, using ERVK and LINE1 promoters as background. From this analysis, we retained only TFs expressed at levels above a 5 TPM cut-off in our RNA-seq data to exclude proteins that are not expressed. NRF1 was among the top hits, showing motif enrichment at IAPs and L1s in addition to being highly expressed in germ cells (**Table 3**,

Table 4). Integrating the filtered hits with our IP-MS data (**Table 5**, **Table 6**), we identified common enrichment of the candidate TFs NRF1, NR0B1, and CREB1 at IAPs and NRF1 and RBPJ at LINE1s (**Figure 3.0.7C**).

Given its role as a DNA methylation-sensitive TF (Domcke et al. 2015), and key regulator of germline genes (Wang et al. 2017), we tested whether NRF1 binding to germline gene promoters also depends on DNA methylation in vitro. To address this, we performed a Fluorescent Polarization Assay with purified full-length NRF1 protein and incubated it at varying concentrations with fluorescently-labeled oligonucleotides containing its binding motif (ctrl oligo) and the promoter region of a known germline gene target, *Asz1*, in both methylated and unmethylated states. Our results revealed that NRF1 binds the *Asz1* promoter more strongly than the control oligo, likely because the selected region of the *Asz1* oligo contains two NRF1 motifs (**Figure 3.0.7D**, **Table 1**). Furthermore, we observed that while NRF1 can bind its motif in the *Asz1* promoter even when methylated, its binding is significantly enhanced in the absence of DNA methylation (**Figure 3.0.7D**).

Figure 3.*0.7*

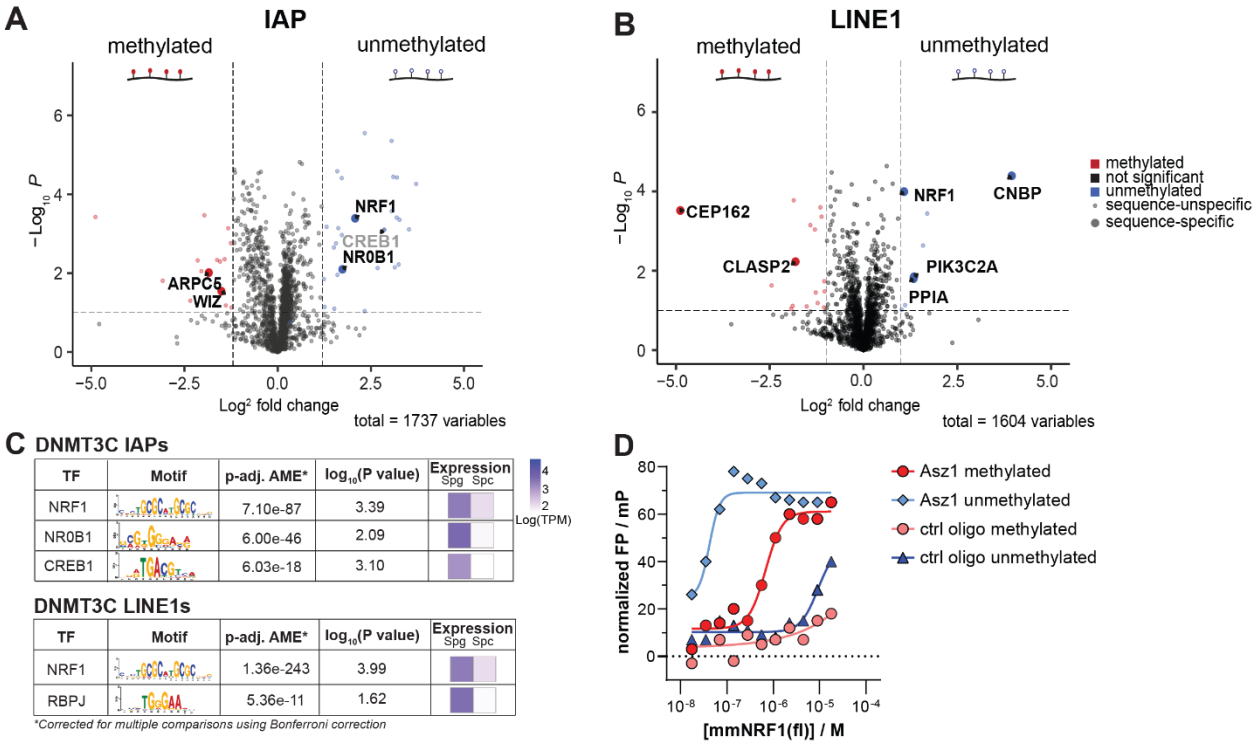

Figure 3.7: Identifying potential TE regulators in male germ cells. A) Volcano plot showing -log10P-adjusted values and log2 fold changes of proteins enriched on methylated and unmethylated IAP LTR1a from three biological replicates detected by IP-MS experiments. Sequence-unspecific proteins that were not detected as enriched on the scrambled controls are represented as small points, while sequence-specific proteins are highlighted by larger, filled points. Labeled proteins refer to sequence-specifically enriched proteins apart from CREB1 labeled in gray. B) as in (A) for methylated and unmethylated LINE1MdT. C) Table of TFs detected in (A) alongside AME motif enrichment on the sequence of the bait (IAP-LTR1a and L1MdT) using all TEs (all IAPs and all L1s) as background. Transcriptional expression from sorted wild-type Spg and Spc is illustrated as a heatmap of log(TPM) values. D) Graph showing binding of purified NRF1 to differentially methylated oligos measured by Fluorescent Polarization.

Based on this and our previous findings, we focused on NRF1 as a potential regulator of unmethylated TEs in germ cells. NRF1 is a ubiquitously expressed transactivator, most

highly expressed in testis and was initially identified for its role in binding to the promoter of cytochrome c and regulating mitochondrial biogenesis (Escrivá et al. 1999, Evans and Scarpulla 1989; Scarpulla 2002). However, more recent studies have revealed a broader scope of NRF1's functions, identifying thousands of binding sites that increase upon hypomethylation and demonstrating its involvement in regulating genes associated with DNA stability and repair, the cell cycle, and germ cell development (Cam et al. 2004; Domcke et al. 2015; Wang et al. 2017; Wang et al. 2024; Palmer et al. 2019). NRF1 has a novel unique DNA-binding domain and does not belong to any of the major TF families (Schaefer et al. 2000).

3.3.2.2. Exploring profiling methods to study NRF1 binding to TEs in male germ cells

To investigate NRF1 binding to TEs in male germ cells, we encountered several challenges. One is that tracking TE regulation before and after meiosis requires either collecting whole testes at pre-meiotic (P5) and meiotic (P20) stages or isolating Spg and Spc via flow cytometry. However, P5 testes yield approximately 200,000 cells, a low input unsuitable for some TF profiling methods, and cell sorting reduces this further to around 20,000 cells. Despite advancements in low-input profiling, challenges persist due to the limited input material especially for profiling TFs. Another challenge is the repetitive nature of our targets, the TEs, which complicates read mapping especially for short DNA fragments.

To identify a reliable method for profiling NRF1 in our model system, I compared and tested various approaches, optimizing the protocols of the most promising ones. The successful optimization of these techniques led to the collaboration with the O'Carroll laboratory, as summarized in Chapter 3.2.1.

ChIP-sequencing (ChIP-seq) is a widely used method for TF and chromatin profiling, involving crosslinking, chromatin fragmentation, and immunoprecipitation of protein-DNA complexes (Solomon et al. 1988) (**Figure 3.0.8A**). While robust and well-established, ChIP-seq faces limitations, including high background, artefacts and epitope-masking from crosslinking, challenges lie in optimizing chromatin fragmentation by sonication, and inefficiency with low-input samples with less than 1 million cells (Baranello et al. 2016). Low-input adaptations, such as ULI-ChIP (Brind'Amour et al. 2015), iChIP (Lara-Astiaso et al. 2014), MINT-ChIP (van Galen et al. 2016), and ChIPmentation (Schmidl et al. 2015), address some of these limitations. Notably, Tn5-transpoase based approaches like iChIP have achieved successful profiling with as few as 10,000 cells (Lara-Astiaso et al. 2014).

As a low-input adaptation, we tested ULI-ChIP, which uses Micrococcal Nuclease (MNase) digestion to generate small chromatin fragments (50-150 bp) and allows for

immunoprecipitation under native or mildly crosslinked conditions (Brind'Amour et al. 2015) (**Figure 3.0.8B**). Building on this, CUT&Run was developed for low-input samples (100,000–500,000 cells), leveraging pA-MNase digestion to selectively release chromatin bound by the target protein, enhancing signal-to-noise ratios and reducing sequencing depth requirements (Janssens and Henikoff 2019) (**Figure 3.0.8C**).

CUT&Tag further improves this by using pA-Tn5 preloaded with sequencing adapters, simplifying library preparation and increasing specificity (**Figure 3.0.8D**). However, CUT&Tag requires high-salt conditions (300 mM NaCl) for Tn5 specificity, which is unproblematic for profiling histone marks but may displace proteins that loosely bind to chromatin (Kaya-Okur et al. 2019). Moreover, both CUT&Run and CUT&Tag protocols are highly dependent on antibody quality of the target. While antibodies for chromatin marks are well-established, TF-specific antibodies vary in quality, and CUT&Run-grade antibodies are still emerging (Wang et al. 2018). Thus, profiling TFs like NRF1 requires optimization based on antibody quality, protein structure, and chromatin interactions.

Consequently, I tested NRF1 profiling in whole testes using standard ChIP and ULI-ChIP protocols, as well as in sorted cells using CUT&Run and CUT&Tag.

Figure 3.*0.8*

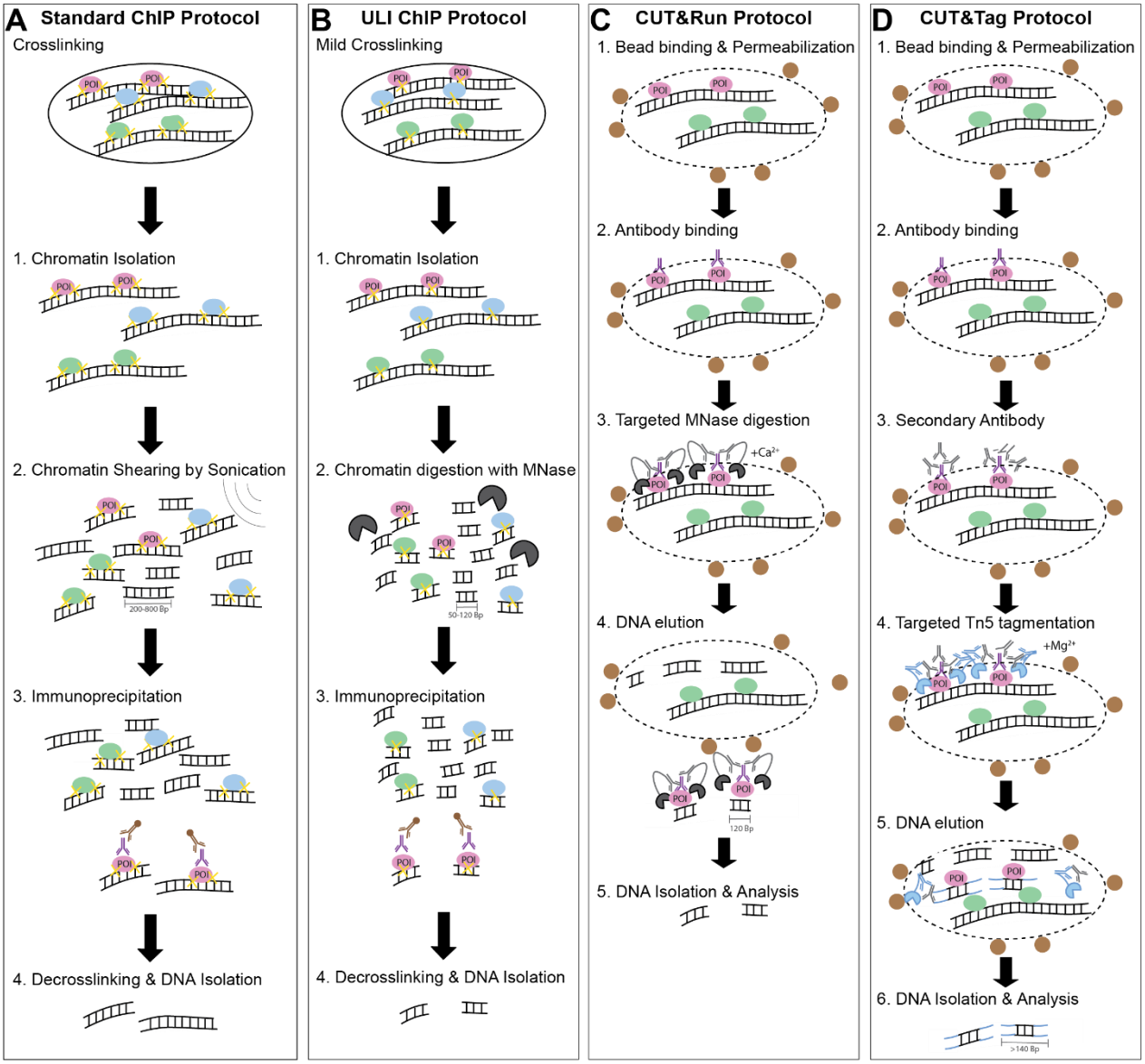

Figure 3.8: Comparison of profiling methods to study a protein of interest (POI). A) Protocol flow for standard ChIP assay. The yellow crosses indicate crosslinks between proteins and DNA. B) Protocol flow for ULI-ChIP assay. Dark grey objects in the second step illustrate MNases. C) Protocol flow for CUT&Run. Brown circles illustrate Concanavalin beads and the grey antibody with MnAse illustrates protein A (pA) binding to the primary antibody. D) Protocol Flow for CUT&Tag, as in (C). Blue objects illustrate the pA-Tn5, the blue DNA extensions illustrate the illumina adapters integrated by Tn5.

3.3.2.2.1 Exploring ChIP and ULI-ChIP for NRF1 profiling

ChIP-qPCR in whole testes from P15 mice revealed NRF1-enrichment at the reported germline gene target, *Asz1*, indicating the functionality of our ChIP protocol (Wang et al. 2017). Additionally, we compared enrichment levels at the TE targets IAP, L1MdA, and L1MdT and found that NRF1 is enriched at these targets in wild-type, with its enrichment

increasing drastically in testes from *Dnmt3C*$^{KO/KO}$ mice (**Figure 3.0.9A**). This suggests that NRF1 might bind unmethylated TEs in *Dnmt3C*$^{KO/KO}$ germ cells *in vivo*.

Applying the ULI-ChIP protocol followed by qPCR similarly showed NRF1 enrichment at TEs, with increased binding at TEs in *Dnmt3C*$^{KO/KO}$ germ cells (**Figure 3.0.9B**). However, following the size selection procedure of the libraries using AMPure beads, which is commonly used to eliminate primer-dimers from library preparation, we found that NRF1 enrichment was lost at TE targets when performing qPCR (**Figure 3.0.9B**). This may be attributed to the small size of DNA fragments released after MNase treatment (**Figure 3.0.8B**), which are difficult to distinguish from primer dimers by size exclusion. Thus, we did not further optimize ULI-ChIP or sequence the libraries from these experiments.

As the qPCR on NRF1 ChIP indicated a successful enrichment of NRF1, we performed NRF1 ChIP-seq from P5 and P15 whole testes. However, visualization of this data revealed only a very mild enrichment of NRF1 at DNMT3C targets in P5 *Dnmt3C*$^{KO/KO}$ mice. At P15, the mild enrichment levels for NRF1 were highly similar with the enrichment observed in the IgG control at DNMT3C targets, indicating low IP enrichment of NRF1 over background (**Figure 3.0.9C**). Several factors could explain this, including suboptimal DNA fragmentation caused by under- or over-sonication, low DNA concentration, an improper ratio of antibody to chromatin depending on the target abundance, or an ineffective antibody. As a result, optimizing a ChIP protocol for NRF1 in testis would be highly time-consuming. Therefore, we decided to explore alternative low-input chromatin profiling techniques instead.

Figure 3.*0.9*

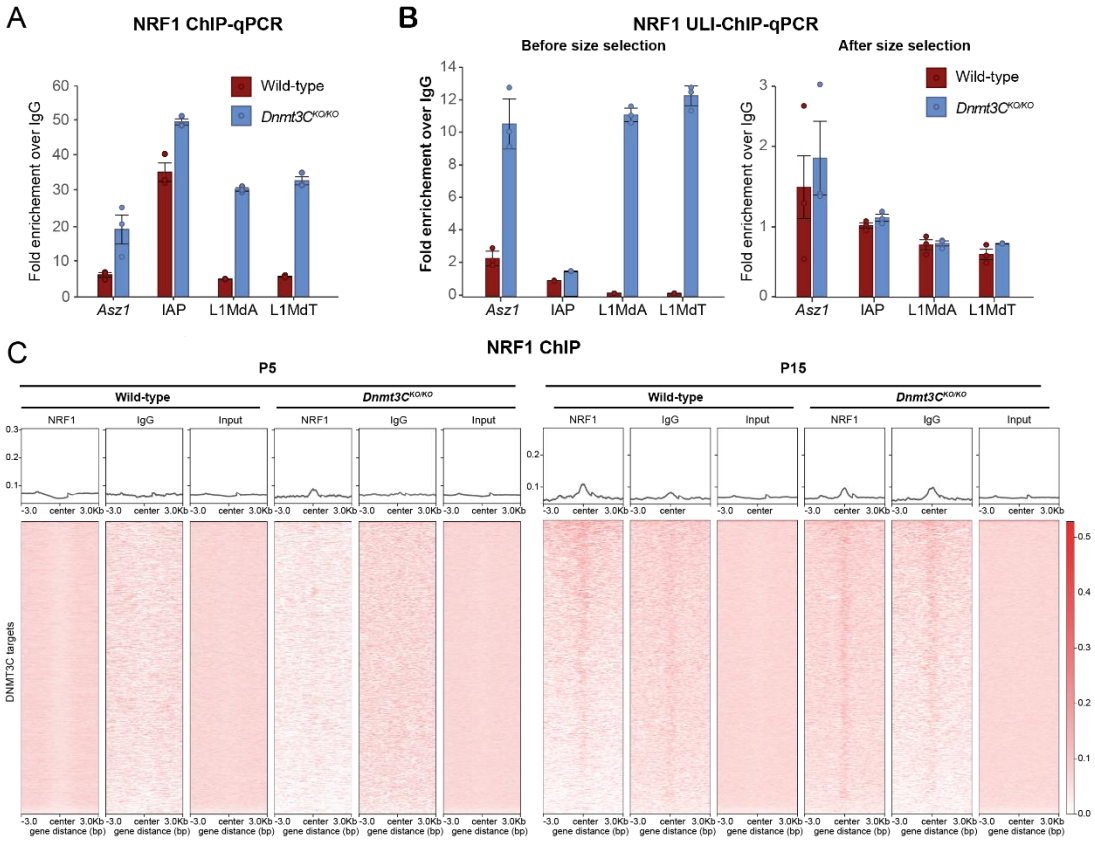

Figure 3.9: NRF1 ChIP and ULI-ChIP in whole testes. A) Barplots showing mean fold enrichment of NRF1 ChIP-qPCR over IgG in whole testes from P15 wild-type and *Dnmt3C*$^{KO/KO}$ mice testing for enrichment at the germline gene *Asz1* and the TEs IAP, L1MdA, L1MdT. The error bar shows the standard deviation from three technical replicates. B) Barplots showing mean fold enrichment of NRF1 ULI-ChIP over IgG in whole testes from P15 wild-type and *Dnmt3C*$^{KO/KO}$ mice at the same targets as in (A) before and after size selection. The error bars indicate the standard deviation from three technical replicates. C) Metaplots and heatmaps illustrating enrichment of reads from NRF1 ChIP in whole testes from P5 and P15 wild-type and *Dnmt3C*$^{KO/KO}$ mice centered on DNMT3C targets for NRF1, IgG and Input samples.

3.3.2.2.2 Exploring CUT&Run and CUT&Tag for NRF1 profiling

We tested CUT&RUN profiling of NRF1 in wild-type and hypomethylated TKO mESCs, using 50,000 cells per condition to validate the method in low input and compare it to published NRF1-ChIP data produced in wild-type and TKO mESCs (Domcke et al. 2015). Following NRF1 CUT&Run, we tested NRF1 enrichment by qPCR on a target, which was reported to gain NRF1 binding in the hypomethylated TKO mESCs (Domcke et al., 2015). We observed a drastic enrichment of NRF1 at this target in TKO compared to wild-type mESCs, indicating the functionality of the NRF1 CUT&Run procedure (**Figure 3.0.10A**). However, we did not detect NRF1 binding at IAPs in TKO mESCs by qPCR, consistent with the study of Domcke et al. (2015) not reporting any enrichment at TEs in TKO mESCS, suggesting that NRF1 does not bind unmethylated TEs in mESCs.

Figure 3.*0.10*

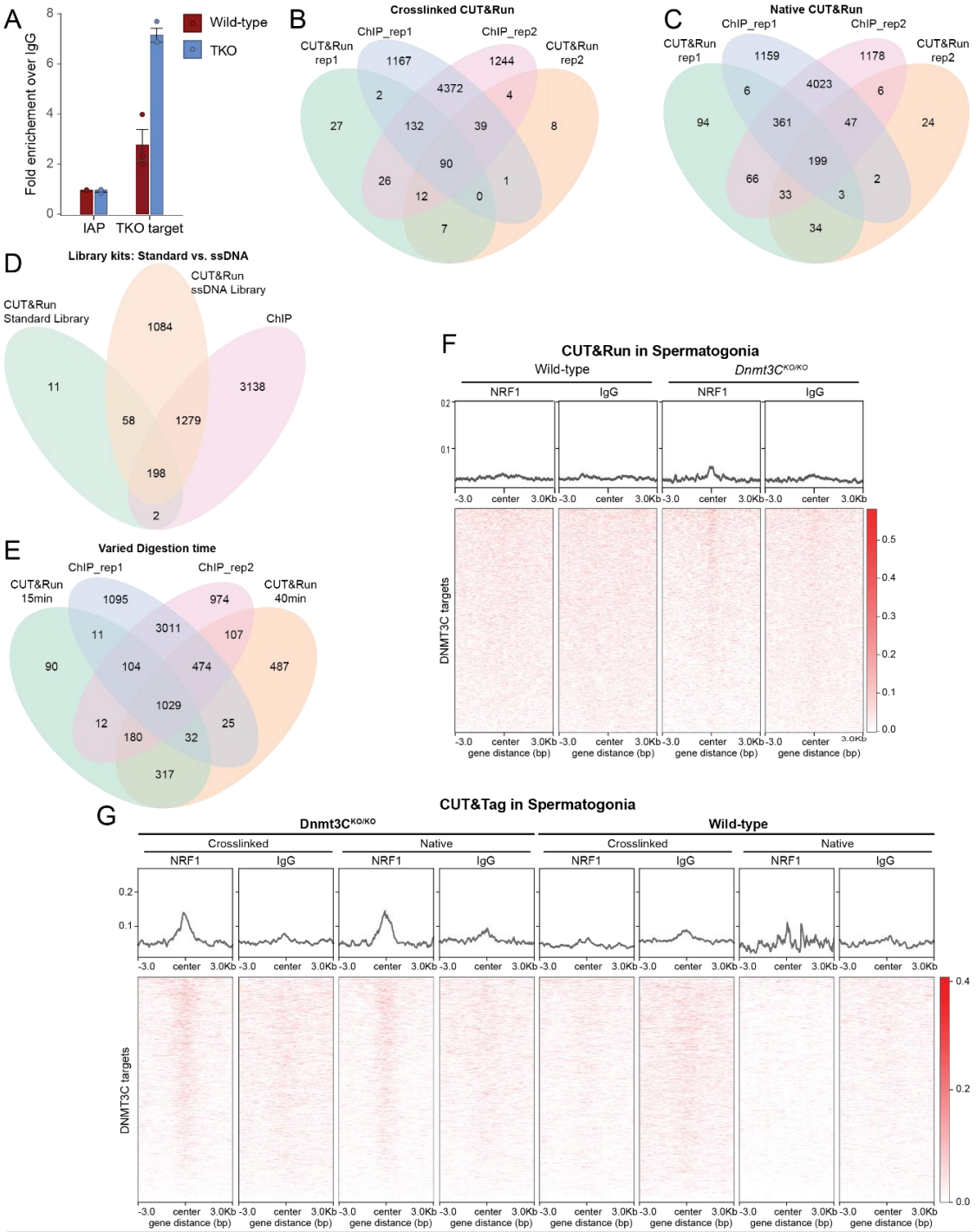

Figure 3.10: NRF1 CUT&Run and CUT&Tag. A) Barplots showing mean fold enrichment of NRF1 CUT&Run over IgG in WT and TKO mESCs on IAPs and a known TKO target. B) Venn Diagram illustrating peak overlaps from NRF1 CUT&Run with mild crosslinking in TKO mESCs compared to peaks from publically available NRF1-ChIP dataset in TKO mESCs (Domcke et al., 2015). C) As in (B) from NRF1 native CUT&Run. D) As in (A) on libraries prepared with a standard illumina library kit (NEB) and a ssDNA library kit (IDT) from native CUT&Run. E) As in (A) on libraries prepared with the ssDNA library kit (IDT) from native CUT&Run with 15 min and 40 min digestion time with MNase. F) Metaplot and heatmap showing mean enrichment of reads from NRF1 CUT&Tag in sorted Spg from wild-type and *Dnmt3C*$^{KO/KO}$ micecentered on DNMT3C targets. G) As in (F) from crosslinked and native NRF1 CUT&Tag.

To further optimize NRF1 profiling by CUT&Run, we tested variations of the protocol in TKO mESCs and compared the peaks identified from our CUT&Run data to the publicly available ChIP-seq data from TKO mESCs, which we downsampled to 8 million reads to match the read count of our CUT&Run data (Domcke et al. 2015) (**Figure 3.0.10B-E**). Performing NRF1 CUT&Run on mildly crosslinked (**Figure 3.0.10B**) and native (**Figure 3.0.10C**) mESCs, revealed that the native condition yielded 723 common peaks with the ChIP-seq dataset, compared to only 306 common peaks from the crosslinked condition. Further improvement was achieved by performing native CUT&Run and using a single-stranded DNA (ssDNA) library preparation kit (IDT) instead of the standard illumina library preparation kit (NEB), increasing the overlap to 1,477 peaks. Testing the influence of MNase digestion time revealed that a 40-minute digestion resulted in 1,667 overlapping peaks, whereas a 15-minute digestion yielded slightly fewer overlaps with 1,368 peaks. Despite these optimizations, the overlap between CUT&Run peaks and the ChIP-seq peaks remained lower than the peaks overlapping between the two ChIP-seq replicates, having 4,630 peaks common peaks. However, the ChIP-seq replicates also showed more than 1,000 unique peaks each, suggesting that while CUT&Run may not capture all true NRF1 binding sites, it may also be less prone to background signals or crosslinking artifacts **(Figure 3.0.10E)**. Additionally, native CUT&Run might also only profile tight interactions of NRF1 with chromatin, while ChIP-seq could identify additional low-affinity binding sites due to extensive crosslinking. Moreover, using a different antibody in our experiment and different cell culture conditions might add up to the observed differences. Considering this and the low input of 50,000 cells, which we used for the CUT&Run compared to the millions of cells needed for ChIP-seq, we concluded that CUT&Run is a suitable technique for estimating NRF1 binding in sorted germ cells yielding low input material.

Using our optimized protocol, we tested NRF1 CUT&Run profiling in sorted Spg from wild-type and *Dnmt3C*$^{KO/KO}$ germ cells. Similar to our previous observations, visualization of the data at DNMT3C targets showed no NRF1 enrichment at DNMT3C-targeted TEs in wild-type Spg, while a slight enrichment was observed in *Dnmt3C*$^{KO/KO}$ Spg (**Figure 3.0.10F**). However, this enrichment was very mild and insufficient for further analysis, and profiling less abundant germ cell types proved even more challenging (data not shown). This could be due to the reliance of CUT&Run on MNase digestion, creating smaller fragments, which may pose a challenge for AMPure bead size selection, similar to our observations from ULI-ChIP qPCR (**Figure 3.0.9B**), and for mapping to repetitive loci. One factor determining fragment size in CUT&Run is wether the bound targets are occupied and protected from fragmentation by nucleosomes, which would increase the fragment size to larger than 150 bp, which is the size of the DNA that is wrapped around a nucleosome. However, most TEs, particularly L1s, are depleted of nucleosomes, which

might result in short fragmentation of TEs by MNase that could get removed by size-exclusion using this method (Sultana et al. 2019).

To address this issue, we tested the CUT&Tag method, which uses the Tn5 transposase to integrate adapters directly into the target sites, thereby increasing fragment size and library specificity. As we previously observed that proofing a profiling method to be reliable for non-TE targets in mESCs does not necessarily result in the successful profiling at TEs in germ cells, we directly tested the CUT&Tag protocol in sorted Spg. Profiling NRF1 in Spg from wild-type and *Dnmt3C*$^{KO/KO}$ mice under both mildly crosslinked and native conditions showed higher NRF1 enrichment at DNMT3C targets in *Dnmt3C*$^{KO/KO}$ Spg than wild-type and this enrichment was higher than we had observed with CUT&Run or ChIP-seq (**Figure 3.0.10G**). These results demonstrate that NRF1 does bind unmethylated TEs in *Dnmt3C*$^{KO/KO}$ germ cells *in vivo*. Since both CUT&Tag conditions performed equally well, we opted to use the native CUT&Tag protocol for further characterization of NRF1 binding in spermatogenesis. Additionally, having optimized TF profiling in low-input material facilitated collaborations with other laboratories facing similar challenges, enabled the profiling of other proteins, such as SPIN1, in low-input material as shown in chapter 3.2.1. SPIN1 & SPOCD1 interact with H3K4me3-H3K9me3-marked young LINE1s(**Figure 3.0.3**).

3.3.2.3. NRF1 binds unmethylated TE promoters in spermatogenesis

After confirming NRF1 expression and nuclear localization by immunofluorescence in E15.5 prospermatogonia, postnatal P5 pre-meiotic and P15 meiotic germ cells (**Figure 3.0.11A**), we employed the previously optimized CUT&Tag protocol to profile NRF1 binding across spermatogenesis from wild-type and *Dnmt3C*$^{KO/KO}$ sorted germ cells. NRF1 occupancy at DNMT3C targets was minimally enriched in wild-type ProSpg. NRF1 binding appeared increased in *Dnmt3C*$^{KO/KO}$ over the IgG control in Spg and Spc compared to wild-type (**Figure 3.0.11B**). Overall, NRF1 binding at DNMT3C targets seemed even enhanced in *Dnmt3C*$^{KO/KO}$ spermatocytes. However, when inspecting track examples at specific copies, some unmethylated LTR copies revealed higher NRF1 binding in Spg than Spc, whereas other TEs, such as the illustrated L1MdA copy showed binding of NRF1 in both, *Dnmt3C*$^{KO/KO}$ spermatogonia as well as spermatocytes, but was increased in Spc (**Figure 3.0.11C**). Interestingly, the location of the NRF1 peak center seemed shifted between spermatogonia and spermatocyte stages at this track example (**Figure 3.0.11C**).

Peak analysis of NRF1 CUT&Tag data showed a notable overlap with DNMT3C targets (**Figure 3.0.11D-F**). Moreover, it confirmed the increased NRF1 binding to unmethylated TEs in *Dnmt3C*$^{KO/KO}$ spermatocytes, where the overlap with DNMT3C targets increased from approximately 3.5% in *Dnmt3C*$^{KO/KO}$ spermatogonia to 9% in *Dnmt3C*$^{KO/KO}$

spermatocytes (**Figure 3.0.11G**). NRF1 peaks gained in *Dnmt3C*$^{KO/KO}$ spermatocytes were predominantly associated with L1MdA and L1MdT elements, consistent with their elevated expression observed in RNA-sequencing (**Figure 3.0.11H, Figure 3.0.2C-D**). Interestingly, the total NRF1 peak numbers were overall higher in wild-type samples than in *Dnmt3C*$^{KO/KO}$ samples at the sites that did not overlap with DNMT3C targets. This might indicate that NRF1 might be more dispersed from its other binding sites due to increased binding at TEs upon loss of DNA methylation (**Figure 3.0.11F**). Of note, we observed additional binding sites for ERVKs in wild-type spermatocytes, while *Dnmt3C*$^{KO/KO}$ spermatocytes showed a similar number of ERVK peaks as *Dnmt3C*$^{KO/KO}$ spermatogonia. However, we did not distinguish between full-length LTRs and solo-LTRs in our analysis, which predominantly serve as regulatory elements for other genes andmight explain some of the peaks gained at ERVKs in wild-type meiosis.

Together, these findings indicate that NRF1 binds to evolutionarily young retrotransposon promoters during spermatogenesis in a TE family-specific pattern, where it predominantly binds unmethylated ERVKs and L1s before meiosis and gains binding at unmethylated L1s in meiosis.

Figure 3.*0.11*

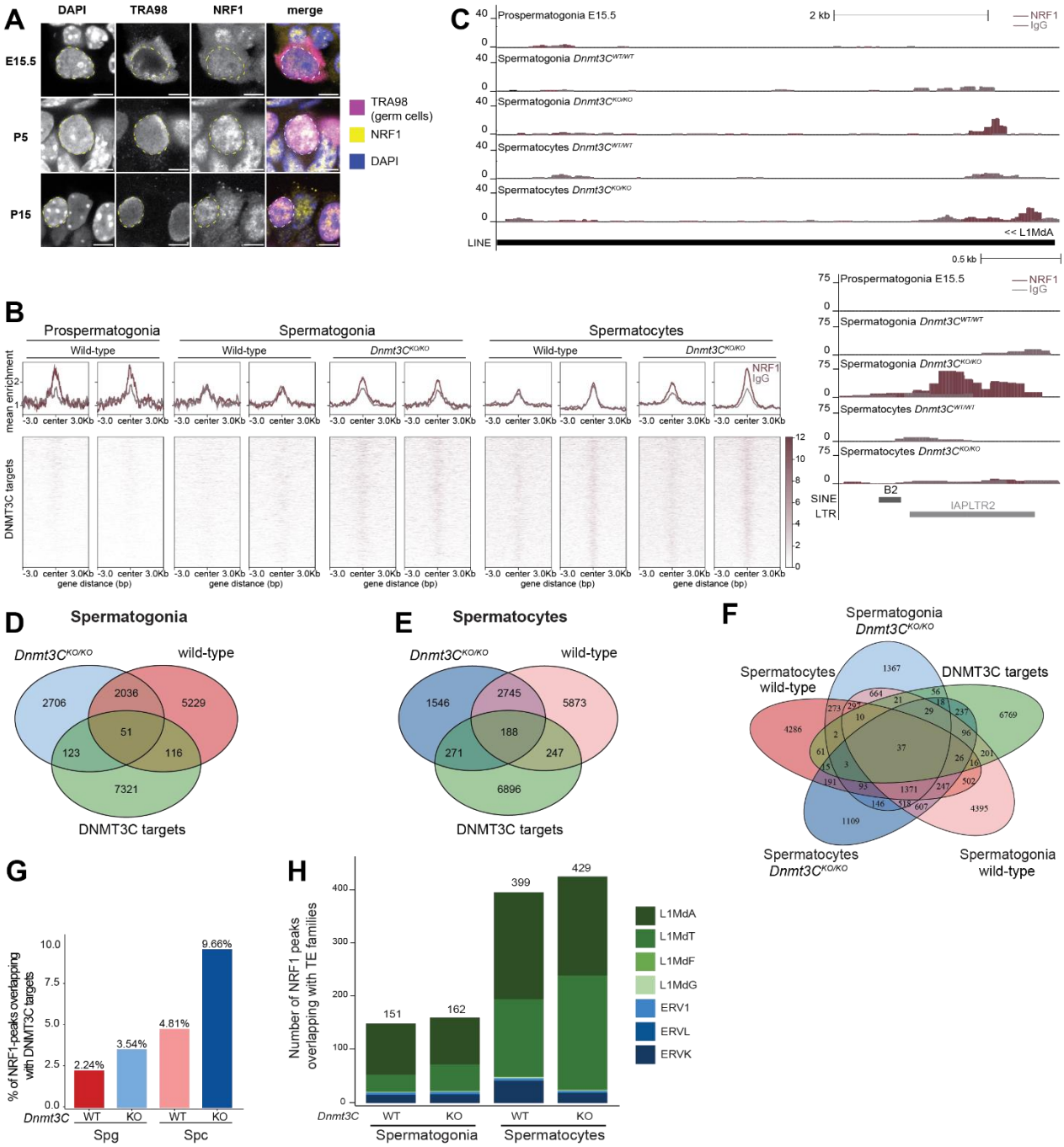

Figure 3.11: The DNA methylation-sensitive TF NRF1 binds unmethylated TE promoters in male mouse germ cells. A) Immunofluorescence staining showing NRF1 expression in testis cryosections from E15.5, P5 and P15 mice. B) Heatmaps and metaplots showing NRF1 (pink-red) and IgG (gray) enrichment by CUT&Tag in sorted ProSpg, Spg and Spc from wild-type and *Dnmt3C*$^{KO/KO}$ mice. Two biological replicates. C) Track examples of NRF1 (pink-red) and IgG (gray) CUT&Tag on an IAPLTR2 and L1MdA element in sorted ProSpg, Spg and Spc from wild-type and *Dnmt3C*$^{KO/KO}$ mice. D) Venn Diagram showing peak overlaps from NRF1 CUT&Tag in Spg from wild-type and *Dnmt3C*$^{KO/KO}$ mice with DNMT3C targets. E) As in (D) in Spc. F) as in (D) for Spg and Spc. G) Barplots showing percentage of peaks from NRF1 CUT&Tag in Spg and Spc from wild-type and *Dnmt3C*$^{KO/KO}$ mice overlapping with DNMT3C targets. H) Column charts showing the number of peaks from NRF1 CUT&Tag Spg and Spc from wild-type and *Dnmt3C*$^{KO/KO}$ mice overlapping with TE families.

3.3.3. Testing a causative role for NRF1 in TE expression

To investigate the potential causal role of NRF1 in retrotransposon activation in the absence of DNA methylation, we analyzed TE expression in the absence of DNA methylation upon germ cell depletion of NRF1.

3.3.3.1. Conditional ablation of *Nrf1* in germ cells defective of DNA methylation

Given that *Nrf1* is essential for viability, with its global knockout resulting in embryonic lethality between E3.5 and E6.5 (Huo and Scarpulla 2001), we focused on conditional knockouts in germ cells from the spermatogonia stage onward, driven by *Vasa-Cre*. Mice carrying a floxed *Nrf1* allele (*Nrf1*$^{fl/fl}$) (Wang et al. 2017) were crossed with *Ddx4-Cre* (Vasa-Cre) driver mice, which induce germ cell-specific knockout starting at E18.5 (Gallardo et al. 2007). The resulting progeny was then bred into the *Dnmt3C*$^{KO/KO}$ background (Barau et al., 2016) (**Figure 2.2.1**). Genotypes were confirmed by PCR, with four separate reactions per animal testing for the presence of *Vasa-Cre*, a floxed *Nrf1* allele, a knockout *Nrf1* allele, and the *Dnmt3C* knockout allele (**Figure 3.0.12A**). The probability of obtaining a male double knockout (dKO; *Nrf1*$^{cKO/KO}$*, Dnmt3C*$^{KO/KO}$) by Mendelian inheritance was 1 in 16 animals (**Figure 2.2.1**). However, an observed breeding lethality rate of ~40% further reduced the chances of producing dKO males.

After eventually obtaining the desired genotypes, immunofluorescence staining for NRF1 and TRA98, a germ cell marker, in the testes was used to assess successful CRE-mediated recombination and allele conversion in germ cells. The presence of nuclear NRF1 in wild-type and *Dnmt3C*$^{KO/KO}$ germ cells and the absence in *Nrf1*$^{cKO/KO}$ and dKO germ cells, confirmed efficient knockout of *Nrf1* (**Figure 3.0.12B**). However, full allele conversion to *Nrf1*$^{cKO/KO}$ from *Nrf1*$^{F/KO}$*; Vasa* occurred only in 55.56% of mice (data not shown), requiring the collection of additional animals. To ensure sufficient material for experiments, we collected dKO animals along with littermate controls, aiming for quadruplicate biological replicates.

In order to determine an appropriate stage for investigating the effects of *Nrf1* depletion on TE expression, we analyzed testes from wild-type, *Dnmt3C*$^{KO/KO}$, *Nrf1*$^{cKO/KO}$, and dKO animals at P5, P8, and P10 by immunofluorescence staining. By P10, germ cells were nearly absent in dKO testes, with a similar but slightly less pronounced germ cell depletion observed at P8 (**Figure 3.0.12C-D**). This was consistent with previous studies on germ-cell ablation of *Nrf1* using the Vasa-Cre driver (Wang et al. 2017). This depletion was further exacerbated in adult testes, where hypogonadism was more severe in dKO compared to *Dnmt3C*$^{KO/KO}$ alone (**Figure 3.0.12E**), suggesting that the germ cell arrest in

dKO occurred earlier than the pachytene stage of meiosis, where *Dnmt3C*$^{KO/KO}$ germ cells typically arrest (Barau et al. 2016). In contrast, at P5, a measurable population of germ cells remained across all genotypes, although *Nrf1*$^{cKO/KO}$ and dKO testes still contained only about half of the germ cell number observed in wild-type testes (**Figure 3.0.12C-D**).

Figure 3.*0.12*

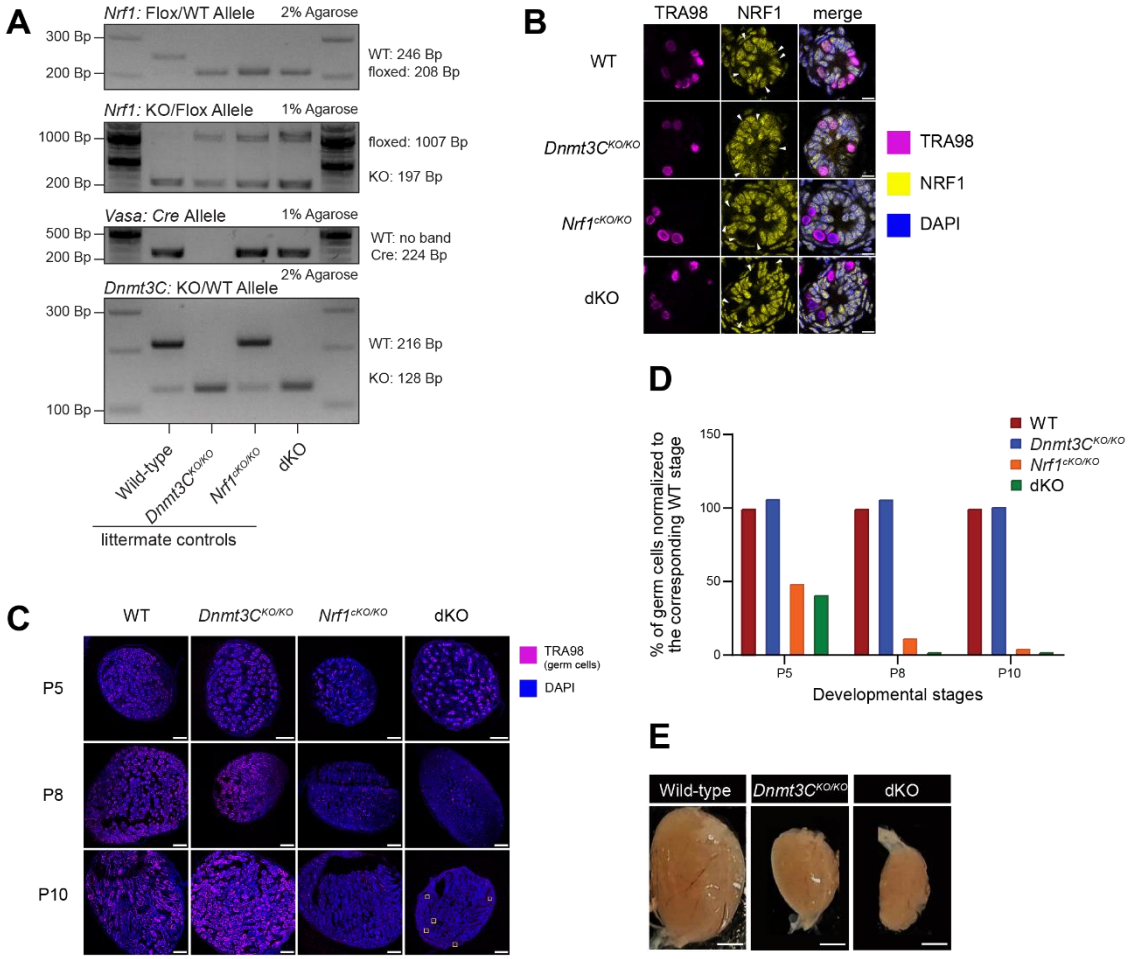

Figure 3.12: Germline conditional knockouts of NRF1 show spermatogenic defects. A) Representative agarose gels showing genotyping results for *Vasa-Cre*, *Dnmt3C*, and *Nrf1* floxed (Flox) and KO alleles in wild-type, *Dnmt3C*$^{KO/KO}$, *Nrf1*$^{cKO/KO}$ and dKO mutants. B) Representative immunostaining of NRF1, TRA98 (germ cell marker) and DAPI (nuclear staining) on cryosections of wild-type, *Dnmt3C*$^{KO/KO}$, *Nrf1*$^{cKO/KO}$ and dKO testes at P5. Yellow arrowheads indicate germ cells, Scale bar, 10µm. The staining was performed in biological triplicates. C) Immunostaining of whole testis cryosections showing TRA98 (germ cell marker) and DAPI at P5, P8 and P10 for wild-type, *Dnmt3C*$^{KO/KO}$, *Nrf1*$^{cKO/KO}$ and dKO. Scale bar 200µm. TRA98-marked cells in P10 dKO are marked with yellow boxes for illustration. D) TRA98/DAPI counts generated by Fiji ImageJ automated counting from one stage were normalized to TRA98/DAPI counts in the corresponding WT stage from one biological replicate. E) Representative image of testes from wild-type, *Dnmt3C*$^{KO/KO}$ and dKO of 5-week-old mice showing hypogonadism phenotype. Scale bar 2mm.

Thus, to minimize confounding effects of germ cell loss in the *Nrf1*$^{cKO/KO}$ and dKO backgrounds, we focused our experiments on the P5 stage. At this time point, spermatogonia and spermatogonial stem cells are the sole germ cell populations (**Figure**

1.8), providing a controlled context for analyzing the effects of NRF1 depletion on TE transcription. For all subsequent analyses, we compared dKO to *Nrf1*$^{cKO/KO}$ and *Dnmt3C*$^{KO/KO}$ to wild-type levels, which had comparable numbers of germ cells, to account for the observed germ cell loss in the mutant backgrounds (**Figure 3.0.12C, D**).

3.3.3.2. IAP silencing is rescued in *Nrf1* and DNA methylation deficient germ cells

For all downstream analyses, we collected total RNA, total protein, and fixed tissue for immunofluorescence from the same mouse, matching littermate controls wherever possible. To investigate the impact of *Nrf1* knockout on retrotransposon expression, we performed poly-A-enriched RNA sequencing on whole testes at P5 and analyzed differentially expressed TE families of mutant genotypes.

As expected, the comparison of the transcriptomes of *Dnmt3C*$^{KO/KO}$ and wild-type controls revealed a significant upregulation of three young ERVK subfamilies: IAPEz-int, IAPA_Mm-int, and MMERVK10C-int, alongside their associated LTRs (Error! Reference source not found.**A**). Notably, no significant upregulation of young *LINE1* subfamilies was detected as we had detected in sorted Spg (**Figure 3.0.2Figure 3*.0.1***), likely due to background RNA from whole testes rather than sorted Spg. A similar finding of exclusive ERVK upregulation was previously reported in RNA-seq from *Dnmt3L*$^{KO/KO}$ whole testes at P10 (Zamudio et al. 2015). Comparison of *Nrf1*$^{cKO/KO}$ and wild-type testes showed a trend toward downregulation of most young retrotransposons, with only IAP1_Mm-int surpassing the significance cutoff. This shows that the germ cell loss due to NRF1 depletion in the mutants does not strongly bias the P5 transcriptomic comparisons of TEs.

To further examine the role of NRF1 in the transcription of unmethylated TEs, we conducted pairwise comparisons using *Dnmt3C*$^{KO/KO}$ versus wild-type as a reference when comparing dKO (*Nrf1*$^{cKO/KO}$*; Dnmt3C*$^{KO/KO}$) versus control genotypes (Error! Reference source not found.**A**) to compensate for any bias due to germ cell loss. Strikingly, in dKO vs. *Nrf1*$^{cKO/KO}$, the upregulation of several IAP subfamilies (IAPLTR2_Mm, IAPA_MM-int, IAPLTR2a, IAPLTR1_Mm, and IAPEz-int) was markedly reduced compared to *Dnmt3C*$^{KO/KO}$ vs. wild-type (Error! Reference source not found.**A, B, F, G**). This indicates that NRF1 is essential for full IAP activation in the absence of DNA methylation. Surprisingly, while LINE1 subfamilies were not upregulated in *Dnmt3C*$^{KO/KO}$ compared to wild-type (Error! Reference source not found.**A, F, G**), L1MdA exhibited increased expression in dKO but not in *Nrf1*$^{cKO/KO}$ alone (Error! Reference source not found.**B, C, F, G**), suggesting that an additive effect between DNA methylation loss and NRF1 knockout leads to this upregulation. When comparing dKO to either wild-type or *Dnmt3C*$^{KO/KO}$, IAP expression was further reduced, though some of these differences may stem from the differential germ cell loss (Error! Reference source not found.**D, E**). In

contrast, MMERVK10C-int and associated LTRs (RLTR10C, RLTR10B, and RLTR10B2) maintained similar expression levels in dKO and *Dnmt3C*$^{KO/KO}$, suggesting that these elements do not depend on NRF1 for activation (Error! Reference source not found.**F, G**).

At the protein level, the immunofluorescence analysis of testis cryosections showed LINE1-ORF1p expression in *Dnmt3C*$^{KO/KO}$ germ cells but not in wild-type or *Nrf1*$^{cKO/KO}$ testes, consistent with transcriptomic data. In dKO, LINE1-ORF1 protein levels were comparable to *Dnmt3C*$^{KO/KO}$, showing no rescue of L1 silencing (Error! Reference source not found.**H**). Similarly, IAP-POL protein was detected in *Dnmt3C*$^{KO/KO}$ germ cells but not in wild-type or *Nrf1*$^{cKO/KO}$. Strikingly, we did not detect IAP-POL in dKO germ cells, which exhibited similar IAP-POL levels as wild-type and *Nrf1*$^{cKO/KO}$ germ cells, supporting that NRF1 is required for IAP protein expression in the absence of DNA methylation (Error! Reference source not found.**I**).

Western blot analysis corroborated these findings. LINE1-ORF1 protein levels were similar between dKO and *Dnmt3C*$^{KO/KO}$ at P5 (Error! Reference source not found.**J**), though a reduction in LINE1-ORF1 expression was observed in dKO by P8, likely reflecting progressive germ cell loss (**Figure 3.0.13K)**. Additionally, the active IAP-POL fusion protein seen in *Dnmt3C*$^{KO/KO}$ testes was significantly reduced in dKO to levels comparable as in wild-type and *Nrf1*$^{cKO/KO}$ (**Figure 3.0.13L**).

Together, these findings demonstrate that NRF1 is essential for the activation of unmethylated IAP elements in premeiotic stages but not for the transcriptional upregulation of LINE1 in male germ cells before meiosis.

Figure 3.*0.13*

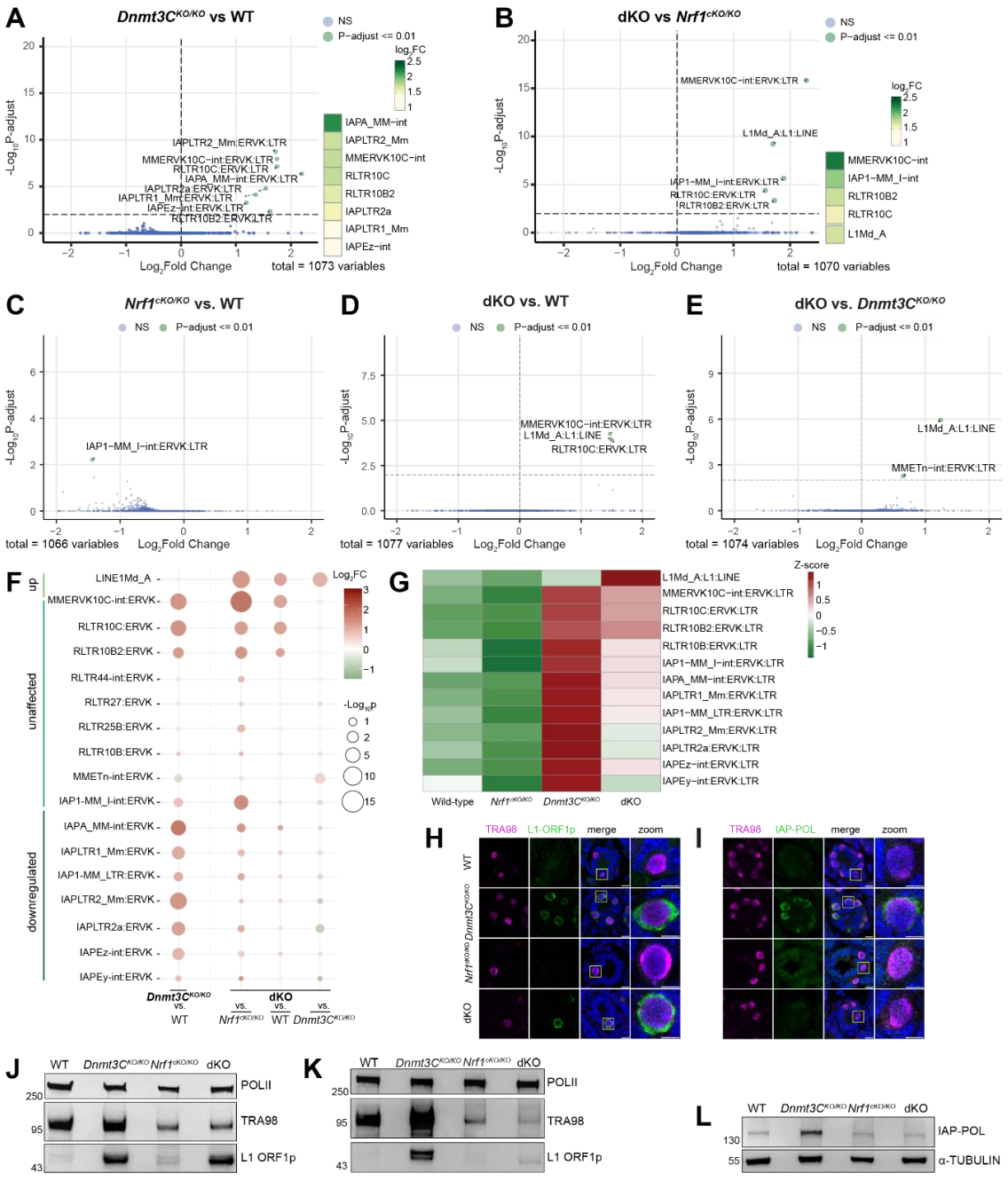

Figure 3.13: NFR1 is required for the expression of unmethylated TEs in the male germline. A) Volcano plot showing -log10P-adjusted and log2-fold change (FC) values for differentially expressed TEs from biological triplicates in *Dnmt3C*$^{KO/KO}$ vs. WT displayed as TEname:family:TEclass. Non-significant (NS) data points depicted in blue, significant values in green. The heatmap displays log2FC values for the same comparison. B) as in (A) for dKO vs *Nrf1*$^{cKO/KO}$. C) as in (A) for *Nrf1*$^{cKO/KO}$ vs WT. D) as in (A) for dKO vs WT. E) as in (A) for dKO vs *Dnmt3C*$^{KO/KO}$. F) Bubble chart showing log2FC as a heatmap, with -log10P-adjusted values represented as circle size, for TE subfamilies differentially enriched in pairwise mutant comparisons. G) Heatmaps showing Z-score of expression values from differentially expressed TE families in all mutants. H) showing a representative immunostaining of LINE1-ORF1p and TRA98 from biological triplicates. I) as in (H) of IAP-POL and TRA98. Scale bar, 10µm (5µ m in cropped images). J) Representative Western blot analysis of POL II, TRA98 and LINE1-ORF1p on protein extracts from WT, *Dnmt3C*$^{KO/KO}$, *Nrf1*$^{cKO/KO}$ and dKO testes at P5, performed in biological duplicates. K) as in (J) from P8 testes. L)as in (J) for α-TUBULIN and IAP-POL. L1-ORF1p and IAP-POL blots were developed using femto chemistry.

3.3.3.3. NRF1-regulated TEs show specific features

To characterize the NRF1-targeted TEs in dKO testes at P5, we analyzed NRF1 binding and chromatin signatures across three categories of DNMT3C targets that we identified in our RNA-seq experiments: downregulated, upregulated, and unaffected in dKO vs. *Nrf1*$^{cKO/KO}$, relative to their expression changes in *Dnmt3C*$^{KO/KO}$ vs. WT (**Figure 3.0.13F**).

In wild-type spermatogonia and spermatocytes, NRF1 binding did not show a distinct enrichment among these categories (**Figure 3.0.14A**). However, in *Dnmt3C*$^{KO/KO}$ spermatogonia, NRF1 binding was more enriched at DNMT3C targets that were downregulated in dKO at P5 (**Figure 3.0.14A**). This trend reversed in *Dnmt3C*$^{KO/KO}$ spermatocytes, where NRF1 preferentially bound DNMT3C targets that were upregulated in dKO at P5, suggesting NRF1-mediated regulation of these TEs may become prominent at the onset of meiosis. This could not be tested in our Vasa-Cre driven dKO due to germ cell defects prior to meiosis (**Figure 3.0.14A**).

To further explore the chromatin context, we profiled H3K4me3 and H3K27me3 at these NRF1 target categories. In *Dnmt3C*$^{KO/KO}$ spermatogonia, DNMT3C targets belonging to the category of downregulated TEs in dKO exhibited enrichment for the active mark H3K4me3 but lacked the repressive H3K27me3 mark, whereas the other categories displayed bivalency with both marks (**Figure 3.0.14B**). In contrast, in *Dnmt3C*$^{KO/KO}$ spermatocytes, H3K27me3 was specifically reduced at DNMT3C targets that were upregulated in dKO, coinciding with NRF1 binding (**Figure 3.0.14A-B**). This pattern suggests that the reduction of H3K27me3 at unmethylated TEs may facilitate NRF1-binding at these targets.

Overall, this analysis revealed that unmethylated DNMT3C targets, which are less upregulated when NRF1 is depleted, are characterized by an active chromatin signature and NRF1 binding. In contrast, other TEs display a bivalent chromatin signature with lower NRF1 enrichment before meiosis. NRF1 may potentially become responsible for their activation at meiosis, when these TEs adopt an open chromatin signature and gain NRF1 binding.

Figure 3.*0.14*

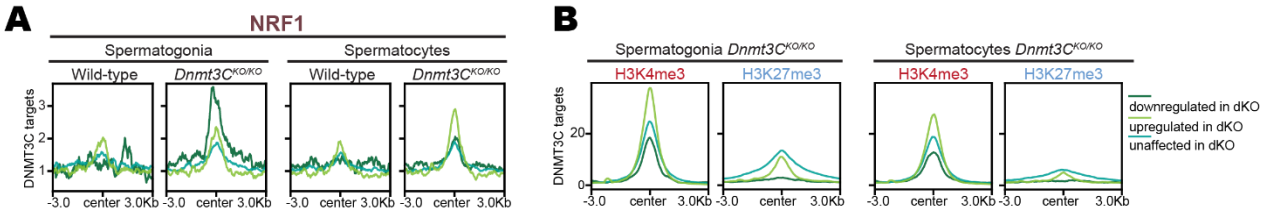

Figure 3.14: Unmethylated TEs downregulated in the absence of NRF1 have a distinct chromatin landscape. A) Metaplots showing mean NRF1 enrichment on DNMT3C targets categorized into down-, upregulated and unaffected in dKO by CUT&Tag in sorted Spg and Spc from wild-type and *Dnmt3C*$^{KO/KO}$ mice. Two biological replicates were merged. B) Metaplots showing mean H3K4me3 and H3K27me3 enrichment on DNMT3C targets categorized into down-, upregulated and unaffected in dKO by CUT&Tag in sorted Spg and Spc from *Dnmt3C*$^{KO/KO}$ mice.

3.4 Supplemental Results

Table 3: Top 100 enriched motifs of TFs in DNMT3C-targeted IAP promoters against all IAPs filtered for 5 TPM expression cut-off

AME_Rank	motif_ID	Uniprot	consensus	p.value	adj_p.value	gene_name
1	NFIC_MOUSE.H11MO.1.A	P70255	YYTGGCWBNNNKCCM	8.99e-178	1.55e-173	Nfic
4	ETV4_MOUSE.H11MO.0.B	P28322	SAGGAAGT	1.38e-145	2.54e-141	Etv4
5	NFIB_MOUSE.H11MO.0.C	P97863	CYTGGCWSV	1.45e-139	2.16e-135	Nfib
6	PBX3_MOUSE.H11MO.0.A	O35317	YBYTGATTGRYTGR	3.68e-132	5.28e-128	Pbx3
7	PBX1_MOUSE.H11MO.0.A	P41778	TGAKTGACAGS	1.91e-131	2.63e-127	Pbx1
8	NFYB_MOUSE.H11MO.0.A	P63139	YRRCCAATSAGMR	8.60e-125	1.02e-120	Nfyb
10	NFYA_MOUSE.H11MO.0.A	P23708	BYRRCCAATCAGAR	3.61e-122	4.76e-118	Nfya
11	NFYC_MOUSE.H11MO.0.B	P70353	BYARCCAATCAGAR	4.33e-122	5.21e-118	Nfyc
14	FOXJ2_MOUSE.H11MO.0.C	Q9ES18	TGTTTRTTTW	2.45e-110	2.93e-106	Foxj2
15	PRRX1_MOUSE.H11MO.0.D	P63013	TAAYCTG	6.87e-104	8.45e-100	Prrx1
16	ATF7_MOUSE.H11MO.0.D	Q8R0S1	RRTGABGTCAB	6.56e-104	8.47e-100	Atf7
17	PRDM9_MOUSE.H11MO.0.C	Q96EQ9	AADADGGYAGYAGCWBCTB	3.54e-103	4.15e-99	Prdm9
18	ETS1_MOUSE.H11MO.0.A	P27577	VRRRCMGGAAGTGG	7.08e-100	5.96e-96	Ets1

21	CLOCK_MOUSE.H11MO.0.C	O08785	CWGHCACGTGMVVM	3.03e-97	4.19e-93	Clock
22	ZFP57_MOUSE.H11MO.0.B	Q8C6P8	CTGCGGCARR	2.16e-95	2.80e-91	Zfp57
23	CREM_MOUSE.H11MO.0.C	P27699	CRVTGACGTCA	1.04e-94	8.38e-91	Crem
24	HXD9_MOUSE.H11MO.0.D	P28357	ARWTTNATKD	1.00e-94	1.47e-90	Hoxd9
25	BHE40_MOUSE.H11MO.0.A	O35185	RCACGTGAS	5.58e-94	4.49e-90	Bhlhe40
26	SP5_MOUSE.H11MO.1.C	Q9JHX2	GRRGDGGGAGGRG	6.00e-93	8.26e-89	Sp5
28	NRF1_MOUSE.H11MO.0.A	Q9WU00	CWSTGCGCATGCGCVRS	7.13e-91	7.10e-87	Nrf1
29	PKNX1_MOUSE.H11MO.0.A	O70477	TGABTGRCAGS	1.31e-89	1.88e-85	Pknox1
30	CDX2_MOUSE.H11MO.0.A	P43241	TTTTATKRCYYT	2.15e-89	3.36e-85	Cdx2
32	FOXI1_MOUSE.H11MO.0.B	Q922I5	YTBTGATTGGYY	4.75e-85	5.17e-81	Foxi1
33	CUX2_MOUSE.H11MO.0.C	P70298	WRATCAATRAA	3.50e-85	6.33e-81	Cux2
34	MAX_MOUSE.H11MO.0.A	P28574	SVSCACGTGGC	1.01e-84	9.78e-81	Max
35	FOXP3_MOUSE.H11MO.0.D	Q99JB6	AATKTGWTT	1.68e-83	2.11e-79	Foxp3
36	MXI1_MOUSE.H11MO.0.A	P50540	VCACGTGGSNGSBGS	3.03e-80	1.93e-76	Mxi1
37	LHX6_MOUSE.H11MO.0.C	Q9R1R0	VCTGATTRV	1.38e-80	2.28e-76	Lhx6
38	TYY1_MOUSE.H11MO.0.A	Q00899	CAAVATGGCGGC	2.19e-75	2.70e-71	Yy1
39	NFAT5_MOUSE.H11MO.0.D	Q9WV30	GYGGAAAANTYCMTK	2.70e-71	1.55e-67	Nfat5
40	PAX5_MOUSE.H11MO.0.A	Q02650	VDRVKCAGYSRAGCGTGAC	4.78e-70	3.13e-66	Pax5
41	FOXF1_MOUSE.H11MO.0.D	Q61080	WTGTTTATTTW	3.39e-67	4.46e-63	Foxf1
42	PO3F1_MOUSE.H11MO.0.C	P21952	CTCATGMATATKYAWD	1.71e-66	2.32e-62	Pou3f1
43	ETV2_MOUSE.H11MO.0.A	P41163	RRARRCAGGAARYRGS	2.70e-63	2.74e-59	Etv2
44	ELK3_MOUSE.H11MO.0.D	P41971	VMCHGGAARTSC	1.33e-62	9.30e-59	Elk3
45	XBP1_MOUSE.H11MO.0.C	O35426	GACGTGKCMTWW	7.72e-61	3.81e-57	Xbp1

46	FOXO3_MOUSE.H11MO.0.B	Q9WVH4	SYTGTTTACH	3.52e-60	6.16e-56	Foxo3
47	FOXO1_MOUSE.H11MO.0.A	Q9R1E0	KBBTGTTTAC	1.17e-58	2.37e-54	Foxo1
48	FOXP2_MOUSE.H11MO.0.A	P58463	BTGTTTACN	1.29e-58	2.53e-54	Foxp2
49	E2F5_MOUSE.H11MO.0.B	Q61502	SGCGCSAAAH	7.53e-56	6.48e-52	E2f5
50	NKX31_MOUSE.H11MO.0.C	P97436	NTWAAGTGBTTD	3.06e-55	4.43e-51	Nkx3-1
51	FOXJ3_MOUSE.H11MO.1.B	Q8BUR3	TKKTTWWTGTTTA	6.69e-55	5.39e-51	Foxj3
52	HXC9_MOUSE.H11MO.0.A	P09633	TGATTTATKGCY	1.26e-54	1.81e-50	Hoxc9
53	MITF_MOUSE.H11MO.0.A	Q08874	YCWYGTGACY	4.59e-53	4.72e-49	Mitf
54	NR0B1_MOUSE.H11MO.0.D	Q61066	GMGKGGGASR	8.13e-50	6.00e-46	Nr0b1
55	ERG_MOUSE.H11MO.0.A	P81270	VVRCMGGAAGYRVV	1.46e-48	2.26e-44	Erg
56	RARA_MOUSE.H11MO.1.A	P11416	GGRRGRTSMRRAGKTCAAGG	2.93e-47	2.60e-43	Rara
57	NKX32_MOUSE.H11MO.0.B	P97503	TTAAGTGBWT	5.20e-46	5.26e-42	Nkx3-2
58	FOXO4_MOUSE.H11MO.0.C	Q9WVH3	TTGTTTAYK	6.13e-46	9.51e-42	Foxo4
59	SMAD4_MOUSE.H11MO.0.A	P97471	CTGTCTGNCACY	6.61e-44	1.04e-39	Smad4
60	KLF8_MOUSE.H11MO.0.C	Q8BLM0	CAGGGKGTG	1.34e-42	3.20e-39	Klf8
61	NANOG_MOUSE.H11MO.1.A	Q80Z64	VRGCCATYARV	1.36e-42	2.26e-38	Nanog
62	NR4A1_MOUSE.H11MO.0.B	P12813	RAAADGKCA	8.25e-41	1.89e-36	Nr4a1
64	MYCN_MOUSE.H11MO.0.A	P03966	SCACGTGG	7.98e-39	7.22e-35	Mycn
67	KAISO_MOUSE.H11MO.0.B	Q8BN78	SARRYCTCGCGAGAV	2.35e-35	4.07e-32	Zbtb33
68	E2F7_MOUSE.H11MO.0.C	Q6S7F2	GDGGCGGGAARDR	7.09e-36	6.63e-32	E2f7
71	HIF1A_MOUSE.H11MO.0.C	Q61221	GDACGTGM	6.48e-32	4.62e-28	Hif1a
73	EPAS1_MOUSE.H11MO.0.C	P97481	VTACGTGMC	4.99e-31	3.32e-27	Epas1
74	FOXK1_MOUSE.H11MO.0.A	P42128	SBTGTTTACNY	7.39e-27	1.42e-22	Foxk1

77	MSX2_MOUSE.H11MO.0.D	Q03358	TAATTDK	1.28e-23	1.63e-19	Msx2
78	ETV5_MOUSE.H11MO.0.D	Q9CXC9	RRRSAGGAARDGRV	2.00e-23	4.20e-19	Etv5
79	CREB1_MOUSE.H11MO.0.A	Q01147	NRRTGACGTMA	8.95e-22	6.03e-18	Creb1
80	SRF_MOUSE.H11MO.0.A	Q9JM73	TTBCCWTATWTGGCHAY	7.23e-22	6.08e-18	Srf
82	SMAD3_MOUSE.H11MO.0.B	Q8BUN5	SYCTSHCWSCWS	6.34e-22	1.42e-17	Smad3
84	HSF1_MOUSE.H11MO.1.A	P38532	BTTCTGGAA	2.62e-19	1.49e-15	Hsf1
85	HLTF_MOUSE.H11MO.0.D	Q6PCN7	KANKGCTGSMAM	9.90e-20	1.63e-15	Hltf
86	COT2_MOUSE.H11MO.2.B	P43135	RGRGYCARAGGTCARVR	1.72e-19	2.02e-15	Nr2f2
87	CEBPB_MOUSE.H11MO.0.A	P28033	DRTTGYGCAAY	1.65e-17	7.27e-14	Cebpb
88	ETV6_MOUSE.H11MO.0.C	P97360	RCAGGAARK	9.73e-18	2.13e-13	Etv6
89	HINFP_MOUSE.H11MO.0.C	Q8K1K9	DMSHHMGCGGACGTTV	7.15e-16	2.82e-13	Hinfp
90	HAND1_MOUSE.H11MO.1.C	Q64279	KSYCTGG	1.48e-16	2.28e-12	Hand1
91	TBX20_MOUSE.H11MO.0.C	Q9ES03	DRGTGHTGACAGSH	4.20e-16	6.15e-12	Tbx20
92	SOX10_MOUSE.H11MO.1.A	Q04888	BYCTTTGTBYB	4.35e-16	7.92e-12	Sox10
93	SOX13_MOUSE.H11MO.0.D	Q04891	YATTGTTY	7.52e-16	1.32e-11	Sox13
94	MECP2_MOUSE.H11MO.0.C	Q9Z2D6	SCCGGRR	2.58e-15	1.35e-11	Mecp2
95	RFX1_MOUSE.H11MO.0.A	P48377	SSSSCBGTTGCCAKGGVRACSS	3.87e-14	3.37e-11	Rfx1
96	CXXC1_MOUSE.H11MO.0.D	Q9CWW7	CGKTGKY	1.35e-14	4.45e-11	Cxxc1
97	RELB_MOUSE.H11MO.0.C	Q04863	RGGGRMTTTCCM	6.89e-14	6.43e-10	Relb
99	SOX9_MOUSE.H11MO.1.A	Q04887	YCTTTGTTCYB	1.48e-13	3.19e-9	Sox9

Table 4: Top 100 enriched motifs of TFs in DNMT3C-targeted L1 promoters against all L1s filtered for 5 TPM expression cut-off

AME_rank	motif_ID	Uniprot	consensus	p.value	adj_p.value	gene_name

1	MBD2_MOUSE.H11MO.0.B	**Q9Z2E1**	SSGKCCGGMGR	5.79e-2646	2.56e-2642	Mbd2
2	TYY1_MOUSE.H11MO.0.A	**Q00899**	CAAVATGGCGGC	5.20e-2446	5.93e-2441	Yy1
3	MECP2_MOUSE.H11MO.0.C	**Q9Z2D6**	SCCGGRR	5.49e-2288	1.63e-2283	Mecp2
4	AP2D_MOUSE.H11MO.0.D	**Q91ZK0**	CGCCYGVGGCSCGT	1.10e-2193	6.89e-2190	Tfap2d
5	SNAI2_MOUSE.H11MO.0.A	**P97469**	RRCAGGTGCNB	3.09e-1686	1.07e-1681	Snai2
7	AP2B_MOUSE.H11MO.0.B	**Q61313**	GCCYGVGGGS	1.02e-1560	1.85e-1556	Tfap2b
8	ZEB1_MOUSE.H11MO.0.B	**Q64318**	BVCAGGTGWG	4.58e-1452	8.21e-1447	Zeb1
10	ITF2_MOUSE.H11MO.0.B	**Q60722**	VCAGGTGCD	7.90e-1432	2.50e-1427	Tcf4
11	MAFG_MOUSE.H11MO.1.A	**O54790**	TGCTGAGT	7.64e-1406	2.03e-1400	Mafg
12	ZFX_MOUSE.H11MO.0.B	**P17012**	GSCBVGGCCTGSSSSSCBS	5.18e-1389	9.85e-1385	Zfx
13	MAFK_MOUSE.H11MO.1.A	**Q61827**	TGCTGAGTCAGCA	8.75e-1338	8.65e-1333	Mafk
14	SNAI1_MOUSE.H11MO.0.C	**Q02085**	SCAGGTGK	5.50e-1315	3.78e-1310	Snai1
15	MYF5_MOUSE.H11MO.0.D	**P24699**	SCAGCTGTYHY	2.47e-1254	3.75e-1249	Myf5
16	MAFA_MOUSE.H11MO.0.D	**Q8CF90**	BTGCTGASTCWGB	6.10e-1210	1.99e-1204	Mafa
17	MYOD1_MOUSE.H11MO.1.A	**P10085**	SCAGCTGYYSC	6.38e-1166	2.86e-1161	Myod1
19	MXI1_MOUSE.H11MO.0.A	**P50540**	VCACGTGGSNGSBGS	1.42e-1103	3.18e-1099	Mxi1
20	GLI1_MOUSE.H11MO.0.C	**P47806**	YSTGGGTGGTCY	4.35e-1076	4.99e-1071	Gli1
21	GLI3_MOUSE.H11MO.0.D	**Q61602**	BSTGGGTGGYCY	2.38e-1040	3.73e-1035	Gli3
22	MYOG_MOUSE.H11MO.0.A	**P12979**	CHSCAGCTGYYYB	2.84e-985	1.34e-980	Myog
23	SMAD3_MOUSE.H11MO.0.B	**Q8BUN5**	SYCTSHCWSCWS	2.15e-984	8.86e-979	Smad3
24	BHA15_MOUSE.H11MO.0.A	**Q9QYC3**	SSAGCAGCTGG	2.49e-940	3.17e-935	Bhlha15
25	MYF6_MOUSE.H11MO.0.C	**P15375**	CASCTGC	6.09e-906	4.75e-901	Myf6
26	E2F2_MOUSE.H11MO.0.B	**P56931**	GGCGCGAAAC	3.05e-893	1.54e-889	E2f2

28	MAFF_MOUSE.H11MO.1.A	**O54791**	TGCTGASTCA	1.14e-832	9.19e-828	Maff
29	HTF4_MOUSE.H11MO.0.A	**Q61286**	RGCAGGTGKG	3.76e-798	5.59e-793	Tcf12
30	HSF1_MOUSE.H11MO.1.A	**P38532**	BTTCTGGAA	2.57e-797	4.64e-792	Hsf1
31	AP2C_MOUSE.H11MO.0.A	**Q61312**	YRSCCYCAGGSYN	1.04e-794	3.97e-790	Tfap2c
32	ZN335_MOUSE.H11MO.1.A	**A2A5K6**	YGCCTGW	3.88e-785	4.59e-780	Zfp335
33	TBX5_MOUSE.H11MO.0.D	**P70326**	AGGTGTGA	1.71e-779	3.67e-774	Tbx5
34	NKX25_MOUSE.H11MO.0.A	**P42582**	DSTBRAGTGS	1.00e-773	1.71e-768	Nkx2-5
35	ASCL2_MOUSE.H11MO.0.C	**O35885**	VCASCTGCT	2.21e-772	2.24e-767	Ascl2
36	P63_MOUSE.H11MO.0.B	**O88898**	DRCAWGYCCAGRCWTGYYB	2.70e-763	8.17e-759	Trp63
37	GABPA_MOUSE.H11MO.0.A	**Q00422**	GGVRCCGGAAGTGV	2.00e-755	1.05e-750	Gabpa
38	NR0B1_MOUSE.H11MO.0.D	**Q61066**	GMGKGGGASR	2.69e-720	3.57e-715	Nr0b1
41	TGIF1_MOUSE.H11MO.0.A	**P70284**	TGACAGCTG	4.17e-705	6.16e-700	Tgif1
42	PTF1A_MOUSE.H11MO.1.A	**Q9QX98**	RGCAGCTGG	2.02e-688	3.86e-683	Ptf1a
43	ELK4_MOUSE.H11MO.0.B	**P41158**	SRRCCGGAAGYRG	1.36e-686	2.50e-681	Elk4
44	FLI1_MOUSE.H11MO.0.A	**P26323**	GGVRCCGGAAGYGS	7.83e-673	5.98e-668	Fli1
45	RFX6_MOUSE.H11MO.0.C	**Q8C7R7**	SKGTTGCYNDG	4.97e-618	9.31e-613	Rfx6
46	TFE2_MOUSE.H11MO.0.A	**P15806**	SMCAGCTGCWB	3.16e-609	1.73e-604	Tcf3
47	ESR2_MOUSE.H11MO.1.A	**O08537**	RGGTCAGV	1.37e-596	1.47e-591	Esr2
48	ELK1_MOUSE.H11MO.0.B	**P41969**	RCCGGAAGTGV	1.63e-550	8.43e-546	Elk1
49	SMAD2_MOUSE.H11MO.0.A	**Q62432**	TCTGYCTGB	5.19e-524	1.38e-518	Smad2
50	VDR_MOUSE.H11MO.0.A	**P48281**	RRGGTCABYGRGKTCA	2.37e-513	2.63e-508	Vdr
51	TF2L1_MOUSE.H11MO.0.C	**Q3UNW5**	RDCYRGTTYHRDCYRGYTY	4.04e-483	2.49e-478	Tfcp2l1
52	PAX6_MOUSE.H11MO.0.C	**P63015**	TYMYGCTTSABT	5.10e-471	1.50e-465	Pax6

53	ZFP57_MOUSE.H11MO.0.B	**Q8C6P8**	CTGCGGCARR	4.81e-463	6.50e-459	Zfp57
54	HES1_MOUSE.H11MO.0.D	**P35428**	VMGVCACGMGMCM	2.28e-455	1.66e-450	Hes1
55	HAND1_MOUSE.H11MO.1.C	**Q64279**	KSYCTGG	1.07e-441	2.43e-436	Hand1
56	RFX1_MOUSE.H11MO.1.A	**P48377**	BGTTGCYAKGG	4.55e-393	2.80e-388	Rfx1
57	RARG_MOUSE.H11MO.1.C	**P18911**	VAGGTCAS	2.20e-330	2.36e-325	Rarg
58	NFIB_MOUSE.H11MO.0.C	**P97863**	CYTGGCWSV	1.62e-322	2.09e-317	Nfib
59	TBX21_MOUSE.H11MO.0.A	**Q9JKD8**	RRAGGTGWGARD	1.24e-301	5.11e-296	Tbx21
60	P73_MOUSE.H11MO.0.B	**Q9JJP2**	RRRCAWGYCCARRCWTGYC	5.40e-297	5.81e-292	Trp73
61	PRGR_MOUSE.H11MO.1.A	**Q00175**	VRGRACAK	1.70e-280	6.89e-275	Pgr
62	RFX2_MOUSE.H11MO.1.A	**P48379**	SGTTGCCATGG	5.91e-262	1.88e-257	Rfx2
63	NRF1_MOUSE.H11MO.0.A	**Q9WU00**	CWSTGCGCATGCGCVRS	1.48e-246	1.36e-243	Nrf1
65	E2F3_MOUSE.H11MO.0.A	**O35261**	GGGCGGGARS	1.23e-232	3.87e-228	E2f3
66	ZBT7A_MOUSE.H11MO.0.B	**O88939**	VRGGGKCKY	2.01e-225	1.37e-220	Zbtb7a
67	MAF_MOUSE.H11MO.1.A	**P54843**	YTGCTGASW	9.19e-223	2.27e-217	Maf
69	CEBPD_MOUSE.H11MO.0.B	**Q00322**	RRTKDBGCAAT	2.69e-220	9.68e-215	Cebpd
71	CEBPA_MOUSE.H11MO.0.A	**P53566**	DRTTGTGCAAY	2.70e-211	4.49e-206	Cebpa
72	COE1_MOUSE.H11MO.0.A	**Q07802**	RGYCCCWGGGGAS	5.66e-202	3.17e-197	Ebf1
73	TBX3_MOUSE.H11MO.0.B	**P70324**	RAGGTGYBA	1.70e-201	2.47e-196	Tbx3
74	RFX3_MOUSE.H11MO.0.C	**P48381**	CSGTTGCCATGGMRAC	1.21e-197	1.02e-193	Rfx3
76	CEBPE_MOUSE.H11MO.0.A	**Q6PZD9**	TTKCGHAATMTB	2.94e-191	8.71e-186	Cebpe
78	E2F4_MOUSE.H11MO.1.A	**Q8R0K9**	GGCGGGAADTTBNR	2.93e-163	4.36e-159	E2f4
80	BHE40_MOUSE.H11MO.0.A	**O35185**	RCACGTGAS	7.92e-157	1.31e-152	Bhlhe40
82	NFIC_MOUSE.H11MO.0.A	**P70255**	TYYTGGCWS	4.91e-151	1.25e-145	Nfic

83	NDF1_MOUSE.H11MO.0.A	**Q60867**	RRCAGATGGY	8.47e-150	1.47e-144	Neurod1
84	ERG_MOUSE.H11MO.0.A	**P81270**	VVRCMGGAAGYRVV	6.34e-145	1.30e-139	Erg
85	MITF_MOUSE.H11MO.0.A	**Q08874**	YCWYGTGACY	2.11e-138	2.26e-133	Mitf
86	ZN322_MOUSE.H11MO.0.B	**Q8BZ89**	GAGCCTGSTACWSWGCCTGR	4.77e-138	3.07e-133	Zfp322a
87	COT2_MOUSE.H11MO.0.A	**P43135**	BMARAGGTCARRR	2.71e-131	5.31e-126	Nr2f2
88	CEBPB_MOUSE.H11MO.0.A	**P28033**	DRTTGYGCAAY	2.40e-129	2.84e-124	Cebpb
89	RUNX3_MOUSE.H11MO.0.A	**Q64131**	KKCTGTGGTTWS	5.50e-123	7.13e-118	Runx3
91	PAX8_MOUSE.H11MO.0.D	**Q00288**	BTVAYTSRMGYRKR	5.68e-123	1.34e-117	Pax8
92	BACH2_MOUSE.H11MO.0.A	**P97303**	TGCTGAGTCAY	6.83e-119	4.55e-114	Bach2
93	SIX4_MOUSE.H11MO.0.C	**Q61321**	WGWAACCTGASMSY	1.89e-117	8.13e-112	Six4
94	ESR1_MOUSE.H11MO.1.A	**P19785**	CHRGGTCASV	9.82e-117	9.78e-112	Esr1
95	ASCL1_MOUSE.H11MO.0.A	**Q02067**	CWSCASCTGCYBCY	1.52e-115	1.07e-110	Ascl1
96	MAFB_MOUSE.H11MO.0.B	**P54841**	TGCTGASTYAD	3.61e-115	1.29e-109	Mafb
98	NDF2_MOUSE.H11MO.0.A	**Q62414**	RRCAGATGGY	4.65e-110	7.28e-105	Neurod2
99	TFEB_MOUSE.H11MO.0.C	**Q9R210**	RGTCACGTG	1.93e-105	8.87e-102	Tfeb

Table 5: Filtered significant raw log10p and differences of mean values from the IP-MS Pulldown of IAP promoter between IAPnon and IAPme

	Log10 pValue of						**Difference between mean of**					
Gene Names	**CONTnon_CONTme**	**IAPme_CONTme**	**IAPnon_CONTme**	**IAPme_CONTnon**	**IAPnon_CONTnon**	**IAPnon_IAPme**	**CONTnon_CONTme**	**IAPme_CONTme**	**IAPnon_CONTme**	**IAPme_CONTnon**	**IAPnon_CONTnon**	**IAPnon_IAPme**
Lin28a	3.76	1.27	6.52	3.08	0.45	4.26	3.75	0.17	3.88	-3.59	0.13	3.71

Polr 3f	2.67	2.49	0.84	2.84	2.17	3.11	0.82	-3.33	0.19	-4.15	-0.63	3.52
Polr 3b	3.00	6.16	0.21	4.70	2.17	3.35	0.77	-3.21	0.06	-3.97	-0.71	3.27
Polr 3c	1.03	2.34	0.18	2.55	3.25	2.22	1.16	-3.05	0.20	-4.21	-0.96	3.25
Rbms2	0.21	0.33	1.64	0.16	2.39	4.43	-0.29	-0.41	2.79	-0.12	3.07	3.20
Lin28b	3.42	0.05	4.13	2.94	3.33	3.41	2.04	-0.02	3.17	-2.06	1.13	3.19
Hnrnpab	2.91	0.10	2.40	2.43	0.29	2.15	3.31	0.10	3.24	-3.21	-0.07	3.14
Polr 3d	3.07	3.43	0.11	4.27	2.46	4.44	0.81	-3.13	-0.03	-3.94	-0.85	3.10
Hnrnpd	3.48	0.75	2.95	4.62	2.49	3.57	3.24	-0.22	2.83	-3.46	-0.41	3.06
Polr 3a	3.00	5.30	0.67	5.68	2.78	5.36	0.78	-2.94	0.12	-3.72	-0.66	3.06
Creb1;Crem; Atf1	3.27	0.33	2.73	3.90	0.15	3.11	3.07	0.17	3.03	-2.90	-0.04	2.86
Polr 3e	2.75	3.47	1.62	3.25	3.58	3.10	1.49	-2.33	0.53	-3.82	-0.96	2.86
Polr 3k	2.04	2.17	0.32	2.26	2.10	2.13	0.73	-2.56	0.12	-3.29	-0.61	2.68
Msh 2	2.50	2.57	1.90	6.15	3.38	5.55	1.39	-1.58	0.76	-2.97	-0.63	2.34
Nrf1	1.70	0.85	3.62	2.75	2.84	3.39	0.39	-0.18	1.90	-0.57	1.51	2.08
Nr0b 1	0.67	0.06	3.29	0.71	0.58	2.10	0.91	-0.04	1.69	-0.95	0.78	1.74
Fbxo18	4.73	2.59	3.08	3.04	3.27	2.97	0.71	-1.38	0.34	-2.09	-0.37	1.72
Xrcc 5	2.35	2.72	1.94	5.21	1.60	4.41	0.80	-1.18	0.54	-1.98	-0.26	1.72

Xrcc6	2.92	2.44	2.09	2.92	1.73	3.14	0.82	-1.16	0.47	-1.98	-0.35	1.63
Polr1d	1.54	2.31	0.07	4.82	1.18	1.96	0.77	-1.55	0.06	-2.31	-0.71	1.60
Hnrnpu;Gm28062	4.02	2.49	4.29	4.08	3.69	4.59	2.01	-0.41	1.19	-2.43	-0.83	1.60
Polr1c	3.55	3.79	0.90	3.59	2.29	2.76	0.70	-1.40	0.16	-2.10	-0.54	1.56
Agpat4	2.06	0.14	1.96	2.84	0.29	2.65	1.54	-0.11	1.40	-1.65	-0.14	1.51
Eprs	3.22	2.44	1.67	2.88	2.06	3.18	1.21	-0.88	0.43	-2.09	-0.78	1.31
Asph	4.94	0.56	1.74	2.78	0.94	1.85	2.00	0.19	1.43	-1.81	-0.57	1.24
Polr2e	3.15	3.31	1.05	4.42	1.88	3.69	0.56	-0.99	0.19	-1.55	-0.38	1.18
Srebf2	2.77	1.79	2.83	3.30	1.02	2.72	0.83	-0.51	0.61	-1.34	-0.22	1.12
Recql	1.47	1.72	2.63	1.79	1.00	3.02	1.48	-0.46	0.64	-1.94	-0.84	1.10
Ace	0.01	0.68	0.30	0.98	0.43	3.42	-0.07	3.38	-1.51	3.45	-1.44	-4.89
Rsrp1	0.32	0.80	0.64	3.85	0.61	1.81	-0.69	1.74	-1.34	2.43	-0.65	-3.09
Ubl5;Gm16381	1.09	1.06	2.25	1.34	0.63	2.32	-2.88	0.79	-1.36	3.66	1.51	-2.15
Mbnl1;Mbnl2;Mbnl3	0.10	0.78	0.49	1.85	0.98	2.05	-0.19	1.27	-0.78	1.46	-0.59	-2.05
Txnl4a	2.83	1.54	2.37	3.15	1.27	3.47	-1.88	0.70	-1.27	2.58	0.61	-1.97

Arpc5	0.54	0.25	0.89	2.40	1.11	2.01	-0.86	0.42	-1.43	1.27	-0.58	-1.85
Cpsf6	1.98	1.07	1.41	2.05	1.00	2.33	-1.64	0.70	-0.92	2.34	0.71	-1.63
Srsf9	2.29	1.56	1.42	2.43	1.88	2.30	-2.60	0.73	-0.81	3.33	1.78	-1.55
Wiz	0.07	0.36	0.39	0.42	0.82	1.54	0.17	0.73	-0.78	0.56	-0.95	-1.51
Thoc3	3.51	0.77	3.17	2.54	2.71	2.35	-1.87	0.32	-1.11	2.18	0.76	-1.42
Mapre1	2.14	0.83	3.95	1.82	0.13	3.14	-1.40	-0.14	-1.47	1.26	-0.07	-1.33
Clasrp	2.23	1.01	1.53	2.12	1.38	2.76	-1.70	0.49	-0.77	2.18	0.93	-1.26
Thoc7	2.69	3.32	2.29	2.96	2.63	3.89	-2.92	0.74	-0.44	3.67	2.48	-1.18
Casc3	0.61	0.61	0.81	2.08	0.28	1.83	-0.49	0.50	-0.68	0.99	-0.18	-1.17
Bclaf1	3.30	2.11	2.68	5.52	4.23	4.62	-1.79	0.40	-0.76	2.19	1.03	-1.16
Pabpn1; Gm20521	3.61	1.89	2.31	3.99	2.54	2.77	-2.37	0.48	-0.67	2.86	1.70	-1.16
Srsf1	2.90	2.95	3.71	3.11	2.29	4.55	-1.92	0.44	-0.70	2.36	1.21	-1.15
Thrap3	5.53	1.97	3.98	3.74	4.09	3.33	-1.42	0.43	-0.70	1.85	0.72	-1.13
Srrt	4.04	1.84	2.03	4.41	3.23	2.98	-2.03	0.50	-0.61	2.53	1.42	-1.12
Srsf4	3.77	1.95	1.88	4.38	2.18	2.58	-1.44	0.46	-0.64	1.91	0.81	-1.10
Clk2	5.69	1.68	2.11	3.08	2.23	3.32	-1.52	0.40	-0.67	1.92	0.85	-1.07
Srsf2	3.62	2.40	2.32	4.24	3.09	3.40	-1.55	0.52	-0.53	2.07	1.02	-1.06

Lsm4	4.04	1.27	1.57	4.28	2.52	1.91	-2.16	0.34	-0.69	2.50	1.46	-1.03
Srsf6	2.81	2.02	4.12	3.74	2.25	2.85	-1.77	0.34	-0.67	2.12	1.10	-1.01
Elavl1	2.94	1.33	1.30	3.34	3.93	2.07	-1.88	0.51	-0.50	2.39	1.39	-1.01
Pnn	2.95	1.74	3.09	3.77	2.04	2.61	-1.56	0.33	-0.67	1.89	0.89	-1.00

Table 6: Filtered significant raw log10p and differences of mean values from the IP-MS Pulldown of L1 promoter between L1non and L1me

	Log10 pValue of						**Difference between mean of**					
Gene Names	**CONTnon_CONTme**	**LINE1me_CONTme**	**LINE1non_CONTme**	**LINE1me_CONTnon**	**LINE1non_CONTnon**	**LINE1non_LINE1me**	**CONTnon_CONTme**	**LINE1me_CONTme**	**LINE1non_CONTme**	**LINE1me_CONTnon**	**LINE1non_CONTnon**	**LINE1non_LINE1me**
Cnbp	1.58	0.08	3.86	1.90	4.59	4.39	-0.88	0.05	4.00	0.93	4.88	3.95
Ywhaq	0.16	2.65	1.06	0.69	0.37	3.44	-0.28	-1.38	0.32	-1.10	0.60	1.70
Pccb	0.31	4.28	1.55	2.23	1.29	2.64	0.10	-1.05	0.54	-1.15	0.44	1.59
Pik3c2a	0.97	1.64	1.35	2.32	0.43	1.85	0.50	-1.07	0.28	-1.56	-0.22	1.35
Ppia	0.68	0.82	2.43	0.37	0.69	1.80	0.85	0.38	1.72	-0.47	0.87	1.33
Nrf1	3.73	3.03	3.12	4.86	0.79	3.99	0.74	-0.38	0.70	-1.12	-0.04	1.08
Hmgb3	0.68	0.27	1.83	0.49	1.44	1.74	-0.39	-0.17	-1.22	0.22	-0.83	-1.05
Mageb4	2.55	0.03	2.68	3.19	0.51	3.36	-1.02	-0.01	-1.13	1.01	-0.11	-1.12
Ascc3	0.32	2.69	3.59	0.79	2.43	3.60	-0.13	-0.43	-1.58	-0.30	-1.45	-1.15

Ncl	1.56	2.90	4.62	2.12	3.84	3.04	-0.41	-1.08	-2.37	-0.67	-1.97	-1.30
Polrmt	1.12	0.20	1.76	3.01	1.71	3.15	-0.75	0.13	-1.28	0.88	-0.54	-1.42
Clasp2	0.28	0.26	0.74	0.75	0.46	2.23	-0.63	0.49	-1.33	1.12	-0.70	-1.81
Uhrf1	2.81	0.79	4.75	4.13	0.40	3.78	-1.82	0.13	-1.73	1.95	0.09	-1.86
Rbpj	0.83	0.42	1.12	3.30	0.41	1.63	-1.33	0.66	-1.79	1.99	-0.47	-2.45
Cep162	0.91	0.39	0.90	5.12	0.15	3.52	-3.55	1.43	-3.45	4.99	0.10	-4.89

4. Discussion

Active TEs are both drivers of evolutionary innovation and threats to germline integrity. In mammals, their activity in male germ cells is tightly controlled by the piRNA pathway, which identifies active TEs and directs their long-term silencing through cytosine DNA methylation. This establishes an ongoing evolutionary arms race: while the silencing machinery constantly adapts to recognize and repress newly active TEs, TEs evolve mechanisms to evade suppression and propagate. When the piRNA pathway fails, loss of DNA methylation allows TEs to escape repression, leading to meiotic failure, germ cell death, and male sterility. Thus, a balance of silencing and expression needs to be struck for TEs to remain active and propagate.

In this project, I aimed to understand why TEs become active specifically at meiosis when DNA methylation is not deposited at their promoters during embryonic germ cell development. More specifically, I asked which mechanisms silence TEs before meiosis or what activates them at meiosis in the absence of DNA methylation and why meiosis is particularly vulnerable to the loss of DNA methylation. By examining the expression patterns of TEs before the piRNA pathway is established and in piRNA pathway-deficient *Dnmt3C*$^{KO/KO}$ mice both before and in meiosis, we found that the widely accepted view of unmethylated TEs becoming active only at meiosis does not entirely hold true. Instead, we observed a differential expression pattern: IAPs remain active throughout postnatal germ cell development, while L1 elements remain active in the absence of DNA methylation but exhibit a boost in activity at meiosis. Moreover, we identified distinct histone modifications regulating IAPs and L1s both before their silencing by the piRNA pathway and at postnatal stages when DNA methylation is absent. We further demonstrated how silencing by DNA methylation is achieved and identified NRF1 as a TF that contributes to TE activation when DNA methylation-based repression is lost. Finally, our findings shed light on the causes for meiotic catastrophe following the failure of the piRNA pathway.

Taken together, our findings provide new insights into how retrotransposons establish the germline expression patterns necessary for their proliferation in mammalian genomes, a process that poses a threat to genome instability but is also central to evolution.

4.1 Differential regulation of L1 and IAP retrotransposons in embryonic germ cells

As part of understanding how evolutionarily young TEs are regulated in the absence of DNA methylation, we sought to determine how they are regulated before the

establishment of DNA methylation by the piRNA pathway. Thus, we examined the chromatin landscape before DNMT3C-mediated DNA methylation at TEs and their expression before and after the establishment of DNA methylation in both wild-type and fetal *Dnmt3C*$^{KO/KO}$ germ cells

4.1.1 L1s and IAPs are marked by distinct histone modifications before *de novo* methylation

To better understand how TEs are regulated in the absence of DNA methylation, we profiled key histone modifications in E15.5 ProSpg before the establishment of DNA methylation mediated by the piRNA pathway at TEs. We found that, at this stage, L1s are not only marked by H3K4me3 and H3K9me3 but also by the repressive H3K27me3 and H3K9me2 marks, which may further repress these elements. In contrast, IAPs showed enrichment only for H3K27me3 (**Figure 3.0.4B**). Furthermore, we detected another activating histone mark, H3K27ac, at L1 elements in E15.5 ProSpg, but not at IAPs (**Figure 3.0.4B**). A notable feature of L1s is the co-occurrence of H3K4me3 and H3K9me3 as a bivalent chromatin mark before DNMT3C-mediated *de novo* methylation (**Figure 3.0.3**), which is recognized by SPIN1-SPOCD1. This raises the question of how these marks are deposited at L1s. The *de novo* methylation machinery is excluded from H3K4me3 marked promoters due to its ADD domain, while H3K9me3 promotes binding via the same domain (Chédin 2011). This suggests that when L1s acquire the active H3K4me3 mark, H3K9me3 may be deposited to facilitate efficient recruitment of DNMT3C-DNMT3L by SPIN1-SPOCD1. The HUSH complex, which targets TEs by recognizing their long intronless RNA, may for example recruit SETDB1 to deposit H3K9me3 marks at L1s. While the HUSH complex has not yet been implicated in silencing TEs in germ cell development, SETDB1 is essential for germ cell development and TE silencing in primordial germ cells (Liu et al. 2014). Specifically, SETDB1 depletion in E13.5 germ cells resulted in reduced H3K9me3 levels at L1s and a loss of H3K9me3 at IAPs, leading to decreased DNA methylation and derepression of IAPs (Liu et al. 2014). Thus, while the HUSH complex and SETDB1 may contribute to H3K9me3 deposition at both IAPs and L1s in embryonic gonocytes, this pathway does not appear to be required for L1 silencing. This suggests that either H3K9me2 might be sufficient to silence L1 or an alternative mechanism may deposit H3K9me3 at H3K4me3-marked L1 promoters, potentially allowing SPIN1-SPOCD1 to remain bound even in the absence of SETDB1.

However, a key question remains: how do L1s acquire the active H3K4me3 mark at their promoters in embryonic germ cell development, whereas IAPs do not? The role of the active H3K4me3 mark in gene expression remains highly debated: whether it directly mediates transcription or is merely a consequence of expression. While several loss-of-function studies have shown that H3K4me3 is not required for the expression of all genes

(Clouaire et al. 2012; Howe et al. 2017), a recent study demonstrated that artificially installing H3K4me3 at closed chromatin sites led to further gain of active H3K27ac marks, chromatin remodeling to an open state, and transcriptional upregulation (Policarpi et al. 2024). Collectively, these studies suggest that H3K4me3 may function both as a cause and a consequence of active transcription. In the context of L1s, H3K4me3 could therefore be deposited as a result of transcription or actively acquired to facilitate transcription. However, since IAPs are also active at the same stage but lack H3K4me3 marking, this finding suggests that either H3K4me3 is actively acquired at L1s rather than being a byproduct of transcription or that other factors hinder IAPs from acquiring H3K4me3. While our Pulldown-Screen did not identify proteins that could facilitate the deposition of H3K4me3 at L1s or hinder H3K4me3 acquisition at IAPs, it may reveal such proteins when performed using promoter sequences on in-vitro assembled nucleosomes rather than on naked DNA in adult and fetal germ cells.

4.1.2 L1s might be earlier silenced by DNMT3C-mediated *de novo* methylation than IAPs

Our TE expression analysis by Western blot in fetal germ cells showed an upregulation of L1s detected as early as E18.5 in *Dnmt3C*$^{KO/KO}$ germ cells compared to wild-type. In contrast, IAP expression remained unchanged between mutants and wild-type at this stage. IAPs were only upregulated in *Dnmt3C*$^{KO/KO}$ by P5, though we did not examine time points between E18.5 and P5, meaning that differential IAP expression between *Dnmt3C*$^{KO/KO}$ and wild-type may become detectable earlier within this time frame (**Figure 3.0.1C**). These findings indicate that L1 elements respond more quickly to the absence of DNA methylation at their promoters in fetal germ cells than IAPs.

The time difference in responding to the absence of DNMT3C could be attributed to the distinct regulation of TE families through chromatin marks or TFs. However, our observation that SPIN1 and SPOCD1 recruit DNMT3C-mediated *de novo* DNA methylation specifically to H3K4me3-H3K9me3-marked L1s but not IAPs (**Figure 3.0.3**), might suggest that this pre-licensing mechanism may lead to an earlier *de novo* DNA methylation of L1s compared to IAPs, resulting in two different *de novo* methylation waves mediated by DNMT3C. Future work tracking the timing of DNA methylation at TE families during *Dnmt3C* expression, starting at E13.5, which represents the lowest DNA methylation levels of TEs, through P1, would provide insights on this hypothesis (Barau et al. 2016). While DNA methylation is unlikely to happen at the same time at all TE loci, long-read enzymatic methyl sequencing (Sun et al. 2021), which would leave TE copy information intact, could give an idea if DNA methylation starts at different timepoints for copies of different TE families.

Overall, we found that IAPs and L1s are differentially regulated both in terms of chromatin modifications and binding proteins before the piRNA pathway established DNA methylation at their promoters and that IAPs and L1s show different expression patterns when DNMT3C-mediated *de novo* methylation fails in prenatal stages.

4.2. Differential upregulation of IAPs and L1s in the absence of DNA methylation in postnatal stages

The prevailing question at beginning of my PhD was to understand why TEs only become active at meiosis when DNA methylation mediated by the piRNA pathway fails, although the marks are lost much earlier in embryonic germ cell development. While previous work in *Dnmt3L* and *Miwi2* mutants reported that DNA methylation is largely dispensable before meiosis due to silencing by H3K9me2 in spermatogonia and is only essential for L1 silencing at meiosis (Di Giacomo et al. 2014; Zamudio et al. 2015), our findings surprisingly highlighted a role for DNA methylation in silencing L1s already before meiosis.

We observed that the loss of DNA methylation at TE promoters activates IAPs after birth, as previously reported (Vasiliauskaitė et al. 2018), maintaining their activity throughout germ cell development. Meanwhile, L1s also remain active in the absence of DNA methylation but are highly upregulated at meiosis (**Figure 3.0.1**, **Figure 3.0.2**). The differences in L1 expression observed in our study compared to others could be attributed to the *Dnmt3C*$^{KO/KO}$ mouse model, which only affects *de novo* methylation at young TE promoters without confounding effects (Barau et al. 2016). For instance, *Dnmt3L* mutants show a global loss of DNA methylation as DNMT3L is also a cofactor of DNMT3A (Suetake et al. 2004). A recent study showed that the depletion of DNMT3A in germ cells leads to a different germ cell defect occurring at the differentiation from spermatogonial stem cells to spermatogonia (Dura et al. 2022). This difference in germ cell composition between *Dnmt3L* and *Dnmt3C* mutants could for example explain the observed TE expression differences. Thus, while undifferentiated spermatogonia may not express unmethylated L1 as previously reported (Vasiliauskaitė et al. 2018), differentiated spermatogonia, which were largely overlooked in *Dnmt3L* mutants, may express unmethylated L1s, as observed in our data. Thus, in future work TE expression in *Dnmt3C*$^{KO/KO}$ could be assessed from different pre-meiotic germ cell populations by sorting SSCs, undifferentiated and differentiated Spg separately, using specific markers such as ID4, PLZF, and cKit for SSCs, undifferentiated and differentiated Spg, respectively.

4.2.1. A transition from a bivalent to an open chromatin chromatin state may explain upregulation of unmethylated retrotransposons at meiosis

Having identified a distinct expression pattern of IAPs and L1s in the absence of DNMT3C-mediated DNA methylation, we asked what could mediate this expression pattern and whether this correlates with the presence of specific histone marks at these elements during spermatogenesis.

Using CUT&Tag in sorted germ cell stages, we observed that the H3K4me3 mark, in contrast to the absence at IAPs before DNMT3C-mediated DNA methylation, was gained at all DNMT3C targeted TEs in DNA methylation-deficient spermatogonia (**Figure 3.0.3A, Figure 3.0.4B, Figure 3.0.5B, Figure 3.0.6A, B, Figure 3.0.14B**). Notably, this active mark was more pronounced at younger L1s, suggesting a correlation between H3K4me3 and the potential of TEs to become transcriptionally active and retrotranspose (**Figure 3.0.6A**). However, we also detected the repressive histone mark H3K9me2 at L1s in *Dnmt3C*$^{KO/KO}$ spermatogonia, which was reduced at meiosis, consistent with previous studies, suggesting that H3K9me2 might suppress full L1 activation before meiosis (Di Giacomo et al. 2014; Zamudio et al. 2015) (**Figure 3.0.5**, **Figure 3.0.6**). At the same time, H3K4me33-marked TEs often carried the repressive H3K27me3 mark in the absence of DNA methylation. The substantial overlap of H3K4me3 and H3K27me3 peaks at DNMT3C-targeted TEs in *Dnmt3C*$^{KO/KO}$ spermatogonia indicates that these elements are likely bivalently marked (**Figure 3.0.5B)**.

Previous reports have suggested that bivalent chromatin plays a crucial role in poising developmental gene expression (Bernstein et al. 2006). In embryonic germ cells, bivalent H3K4me3 and H3K27me3 marks at promoters keep germline-specific genes in a poised state, allowing them to be activated at the appropriate developmental stage (Sachs et al. 2013). Other studies have proposed that bivalent marks may protect promoters from aberrant DNA methylation (Boulard et al. 2015). Interestingly, a recent study showed that bivalent marks were not only found at developmental gene but also at unmethylated full-length L1 promoters in pluripotent mouse embryonic stem cells, where L1 expression is required for the activation of genes from the pluripotency network (Jachowicz et al. 2017; Gretarsson and Hackett 2020). DPPA2 and DPPA4 bind and prime bivalently marked promoters and the depletion of one of them led to the acquisition of DNA methylation at L1s and their downregulation (Eckersley-Maslin et al. 2020; Gretarsson and Hackett 2020).

If bivalency at TEs in spermatogenesis serves a similar function to that at developmental genes and L1s in mouse embryonic stem cells, this chromatin state may contribute both

to the poising of TE expression before meiosis and to protecting these elements from DNA methylation later in development: Getting active at meiosis would provide TEs with an opportunity to retrotranspose at a critical stage of germline development, where transposition events could be transmitted to daughter cells.

At the same time, H3K27me3-mediated repression may act as a safeguard, preventing premature activation and allowing TEs to evade bulk DNMT3A-mediated *de novo* methylation. The presence of H3K27me3 as a silencing factor in various hypomethylated contexts, such as oocytes (Huang et al. 2021), hypomethylated mESCs (Walter et al. 2016), and in other eukaryotic organisms like *Arabidopsis* (Hure et al. 2024), *Paramecium* (Miró-Pina et al. 2022) and *flagellates* (Gahan et al. 2024), further suggests that this mechanism may represent an ancestral fallback strategy for TE repression when DNA methylation is compromised. Indeed, in the absence of DNA methylation in *Dnmt3C*$^{KO/KO}$ mutants, retrotransposons classified as likely-bivalent or H3K4me3-marked exhibited significantly higher expression levels in both spermatogonia and spermatocytes than H3K27me3-marked (**Figure 3.0.5D**).

Overall, our data demonstrate that, in the absence of DNA methylation, bivalent and H3K4me3-marked TEs are more prone to transcriptional reactivation than H3K27me3-marked or uncategorized TEs. This further suggests that active chromatin marks play a key role in regulating retrotransposon expression, particularly in younger LINE1 and ERV families, highlighting the dynamic interplay between chromatin states and TE regulation in *Dnmt3C*-deficient germ cells. However, to confirm true bivalency at these targets, the co-occupancy of both marks at the same nucleosome would need to be tested using simultaneous profiling techniques such as Multi-CUT&Tag (Gopalan et al. 2021).

The exact mechanism leading to the observed differential regulation of TEs and robust upregulation of L1s during meiosis remains unresolved. However, our findings, which reveal a specific chromatin landscape at unmethylated TEs before and after meiosis, may provide a partial explanation to this. The transition from a rather closed chromatin state with H3K27me3, H3K9me2 and H3K4me3 primarily marking unmethylated TEs before meiosis to the more open chromatin state marked by the presence of H3K4me3 and H3K27ac in meiosis, might partially explain the boost of TE expression observed in *Dnmt3C*$^{KO/KO}$ animals at meiosis. Furthermore, our findings indicate that ERVs are more effectively repressed by H3K27me3 marks than L1s, suggesting a differential chromatin-mediated regulation of these TE families.

4.3 How does DNA methylation achieve the silencing of TEs?

After having examined the histone modifications at TEs in the absence of DNA methylation during spermatogenesis, we set out to identify proteins, which could bind TEs and regulate their expression additionally.

Since the discovery of DNMT3L's role in transposon silencing and the subsequent elucidation of the piRNA pathway's role in guiding DNA methylation, the field has largely focused on how TEs are methylated (Bourc'his and Bestor 2004; Barau et al. 2016). While DNA methylation is well established as a silencing mechanism for TEs, the underlying processes of how it achieves silencing and which factors drive TE reactivation in hypomethylated conditions remain unclear. Readers of DNA methylation that bind methylated DNA and recruit chromatin remodelers and writers of repressive histone marks have been identified in some cell types (Filion et al. 2006). However, a recent study demonstrated that at least some of the DNA methylation readers, more specifically the four methyl-binding proteins, MBD1, MBD2, MBD4 and MeCP2 are dispensable, individually and in combination, for TE silencing in mESCs (Kaluscha et al. 2022). Still, it remains possible that other repressive methyl-binding proteins play a role or might compensate for the absence of these MBD proteins. The same study showed that the depletion of DNA methylation in differentiated neurons, leads to IAP upregulation mediated by the DNA methylation-sensitive TF CREB1, suggesting that the inhibition of DNA methylation-sensitive TFs is the main silencing mechanism of DNA methylation (Kaluscha et al. 2022). Nevertheless, this mechanism of silencing may differ in other tissues and the proteins responsible for TE regulation during germ cell development remain unknown.

In our project, we took a different approach and compared which protein binds to TE promoters when they are methylated versus unmethylated. Using our pulldown strategy with methylated versus unmethylated promoters from L1s and IAPs, we found no methyl-binding proteins significantly enriched on the methylated baits, except for UHRF1, which binds hemimethylated DNA and recruits DNMT1 (**Figure 3.0.7A,B, Table 5, Table 6**) (Spruijt et al. 2013). UHRF1 was also reported to bind to methylated retrotransposons in spermatogenesis, validating our finding (Dong et al. 2019). On the other hand, we identified the methylation-sensitive TFs NRF1 and CREB1 binding to the unmethylated sequences. This suggests that, even in germ cells, TE repression by DNA methylation might be mediated by the inhibition of DNA methylation-sensitive TFs.

4.4 TFs regulating TEs: Insights, Challenges, and Future Directions

4.4.1 The Importance of Identifying TFs in TE Regulation

Studies like ours and the work by Kaluscha et al. 2022 highlight the importance of identifying TFs that bind and regulate TEs. This is not only crucial for understanding the mechanisms of TE silencing but also for elucidating how TEs become active when DNA methylation is absent. TEs often exist as metastable epialleles that have variegating DNA methylation patterns (Bertozzi et al. 2021; Kazachenka et al. 2018). Thus, our insights into TFs binding TEs in *Dnmt3C*$^{KO/KO}$ mice, can help explain how TEs retrotranspose, generate new copies, and expand within the genome dependent on their methylation status. At the same time, this knowledge is critical for understanding diseases associated with aberrant TE activity upon hypomethylation, such as neurological disorders, infertility and cancer (Howard et al. 2008; Wolff et al. 2010; Solyom and Kazazian 2012). Identifying TFs recruited to TEs in different tissues may therefore open avenues for gene-editing therapies to prevent binding of TFs to TEs.

As part of this project, we characterized the binding of NRF1 to TEs in germ cells. However, using our unbiased combination of DNA-pulldown IP-MS with motif enrichment analysis at TEs, alongside TF expression analysis in germ cells, we also identified other promising TF candidates, such as CREB1, NR0B1, or RBPJ that require further characterization (**Figure 3.0.7A, C**). This could lead to the discovery of additional TFs that regulate TE expression during germ cell development.

4.4.2 Challenges, Limitations and potential solutions for low-input TF profiling at TEs

Despite the advances in profiling methods, challenges remain in profiling TFs due to the repetitive nature of TEs and the limited input material available for some experiments, which highly limit the studies that have previously described TFs regulating TEs. In my project, I optimized reliable profiling methods to characterize NRF1 binding in sorted germ cells, finally using native CUT&Tag protocols (**Figure 3.0.10**). However, as previously mentioned, CUT&Tag might not be ideal for TF profiling due to the high salt conditions required for Tn5 transposase specificity. While this approach may work for certain TFs, it is not universally suitable (Kaya-Okur et al. 2019). Additionally, NRF1 profiling using CUT&Tag did not produce as clear binding landscape as expected from profiling of other highly abundant proteins such as CTCF (Kaya-Okur et al. 2019) or using methods like high-input ChIP-seq (Domcke et al. 2015), though it did show a subtle enrichment over

IgG controls (**Figure 3.0.10**, **Figure 3.0.11**). Thus, the profiling of NRF1 might benefit from adapted or other newly emerging methods.

Moreover, during my PhD, after I had completed profiling chromatin marks in different germ cell stages, the issue of CUT&Tag and CUT&Run normalization became more apparent (Chen et al. 2015; Patel et al. 2024). Normalization by sequencing depth (RPGC), which we used to normalize our experiments, proved suboptimal for accurate comparisons between samples (Dickson et al. 2020). The emerging standard for normalization now involves exogeneous spike-in of cells from a different species to minimalize the variability between replicates (Wooten et al. 2023). However, since many of our profiling experiments were completed before this became widely adopted, we maintained consistency in analysis by using our original normalization method. However, more recently, new methods for internal normalization of CUT&Run and CUT&Tag have been developed, offering potential advantages over spike-in normalization by eliminating pipetting errors and the potential bias from the assumption that the target does not vary in the spike-in material. For example, signal artifacts arising from MNase or Tn5 enzyme cutting bias, which are present in all profiling experiments regardless of the target, can be used to normalize for cell input and cutting efficiency, defining a so-called "greenlist" (de Mello et al. 2024). Normalization by greenlist could also be applied on the CUT&Tag data produced in this project to optimize the current normalization, but it would first require the definition of a suitable CUT&Tag greenlist in mouse germ cells.

Another limitation of CUT&Run and CUT&Tag is that, due to the enzyme's cutting preferences, true signals and PCR duplicates cannot be easily distinguished. This presents a trade-off in analysis: either retaining duplicates, which risks artificially amplifying the signal due to PCR bias, or removing duplicates, which eliminates true signals. In our analysis, we chose to retain PCR duplicates, as the removal of true signals could be particularly high due to the repetitive nature of TEs and the high copy number of the same elements in the genome. A potential solution to this issue is the incorporation of unique molecular identifiers (UMIs) into the CUT&Tag procedure. This approach has been successfully implemented in a similar technique, Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq), which also utilizes the Tn5 transposase for sequencing open chromatin regions (Zhu et al. 2020). In this method, identical reads carrying the same UMI sequence are identified as PCR duplicates, whereas identical reads with different UMIs are assumed to originate from distinct cells or genomic locations. This strategy enables bioinformatic filtering of PCR duplicates without the risk of removing true signal. To integrate UMIs into CUT&Tag, the Tn5 transposase would need to be loaded with Mosaic End (ME) adapters containing UMI sequences, allowing for precise identification and filtering of PCR duplicates during data analysis.

In general, working with highly repetitive TEs presents challenges in mappability, as previously mentioned. One reason is that the promoters sometimes consist of repeating

monomers. Additionally, their full-length sequences often span several thousand base pairs and therefore TEs are often not sequenced full-length, leaving positional information incomplete. This makes evolutionarily young TEs, which have not yet accumulated point mutations, particularly difficult to map and their copies mostly indistinguishable (Teissandier et al. 2019). The emergence of long-read sequencing technologies, such as Oxford Nanopore Technology, offers a promising solution, providing better tools for studying TEs (Gerdes et al. 2023). Profiling TFs, DNA methylation, and histone marks at TEs using long-read sequencing may unveil previously unknown correlations and shed light on how TEs evade germline silencing mechanisms.

4.5. NRF1 activates IAPs before meiosis and might activate L1s at the onset of meiosis

Using our DNA-pulldown approach, we identified NRF1 as a DNA methylation sensitive TF potentially regulating TEs in male germ cells when their promoters are unmethylated (**Figure 3.0.7**). Using CUT&Tag, we demonstrated that NRF1 binds unmethylated IAPs and L1s before meiosis and increasingly binds unmethylated L1s in meiosis (**Figure 3.0.11**).

4.5.1 IAP silencing is rescued in dKO germ cells depleted of NRF1 and DNMT3C but L1A expression is upregulated

Moreover, TE expression analysis in dKO germ cells depleted of NRF1 in the absence of DNMT3C-mediated DNA methylation showed that IAP silencing was rescued before meiosis. This suggests that NRF1 is responsible for IAP upregulation in *Dnmt3C*$^{KO/KO}$ Spermatogonia and that IAP activation relies on NRF1 binding to their unmethylated promoters (**Figure 3.0.13**, **Figure 4.1**).

Figure 4.1

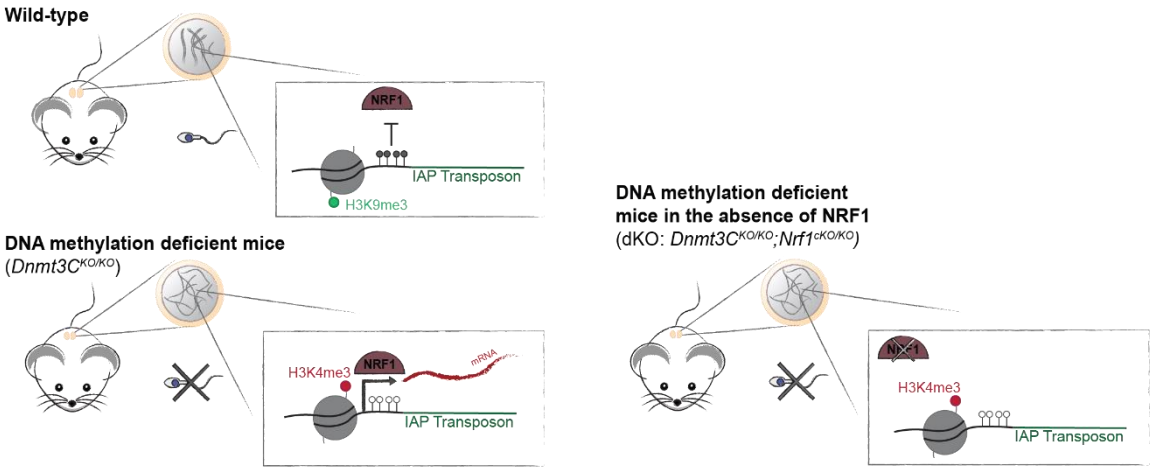

Figure 4.1: Model illustrating how NRF1 regulates IAP elements in male germ cells from wild-type, *Dnmt3C*$^{KO/KO}$ and dKO germ cells.

Although NRF1 CUT&Tag also indicated subtle binding to unmethylated L1s before meiosis, we observed no reduction in L1 expression in the dKO (*Nrf1*$^{cKO/KO}$*; Dnmt3C*$^{KO/KO}$). Surprisingly, L1MdA expression was upregulated instead **(Figure 3.0.11, Figure 3.0.13B, F, G)**. Interestingly, a recent study suggests that NRF1 can act as both a transcriptional activator and a repressor, depending on its binding position relative to the transcription start site (TSS). Specifically, NRF1 binding upstream of the TSS promotes transcriptional activation, while binding downstream of the TSS leads to repression (Duttke et al. 2024). Our examination of NRF1 profiling tracks in *Dnmt3C*$^{KO/KO}$ germ cells revealed that NRF1 peaks shift at some L1MdA loci before and after meiosis (**Figure 3.0.11C**). This might indicate that NRF1 binding to L1s downstream of the TSS before meiosis leads to L1 repression, whereas a shift to upstream binding could activate L1s at the onset of meiosis. The occlusion of NRF1 from its upstream binding site may correlate with the presence of H3K27me3 at L1s in spermatogonia, which is drastically reduced at meiosis (**Figure 3.0.14**).

Consistent with our observation of significant L1MdA upregulation before meiosis in dKO animals (**Figure 3.0.13B, F, G**), we observed a more severe failure of spermatogenesis in dKO testes compared to *Dnmt3C*$^{KO/KO}$ or *Nrf1*$^{cKO/KO}$ testes, suggesting that excessive L1 upregulation before meiosis may enhance the germ cell loss phenotype (**Figure 3.0.12C, D, E**). NRF1 acting as a transcriptional repressor of L1s before meiosis could explain this observation. Alternatively, L1 upregulation in dKO could simply result from downregulation of other genes by NRF1 depletion, such as the downregulation of EHMT1/GLP, a known NRF1 target (Palmer et al. 2019). GLP together with G9a catalyze H3K9me2, which we found a repressive mark silencing especially L1MdAs before meiosis in the absence of DNA methylation (**Figure 3.0.6C**). As we have found that L1MdAs

marked by H3K27me3 are not significantly less expressed than when marked by active H3K4me3 or bivalently, this further indicates that L1MdA might rely more on repression by H3K9me2 in the absence of DNA methylation than other L1 elements such as L1MdT (**Figure 3.0.6B**). Thus, NRF1 depletion could lead to a loss or reduction of repressive H3K9me2 at L1MdAs, resulting in their premature upregulation in dKO mice.

4.5.2 NRF1 might be responsible for L1 expression at meiosis

Additionally, our NRF1 profiling data indicate that L1s gain NRF1 binding during meiosis, which may activate their transcription at meiosis (**Figure 3.0.11H**). However, interpreting this correlation as the direct cause of L1 upregulation at meiosis should be approached with caution. Several studies have suggested that NRF1 binding to a target does not necessarily directly translate into transcriptional activation, as the study previously mentioned. Moreover, NRF1's role as a transcriptional activator at its binding sites may depend on additional mechanisms, such as its binding conformation as a monomer or homodimer. A recent study demonstrated for instance, that NRF1 can bind its half-site motif as a monomer, whereas full-site motifs are bound as a homodimer, which is an essential requirement for driving transcriptional activity (Liu et al. 2024).

Finally, we could not test a potential rescue of L1 silencing at meiosis in the dKO, as the observed germ cell arrest of dKO animals driven by *Vasa*^{Cre} occurred already before the onset of meiosis (**Figure 3.0.12C, D**). To overcome this, dKO meiotic germ cells could be obtained by breeding *Nrf1*$^{KO/Fl}$*; Dnmt3C*$^{KO/KO}$ in the *Stra8*^{Cre} background, which initiates allelic KO-conversion at P8 (Sadate-Ngatchou et al. 2008). Yet, this would require a careful characterization of the germ cell arrest phenotype in dKO driven by *Stra8*^{Cre}, performing meiotic spreads and counting the percentage of cells in specific meiotic stages. After characterizing this in detail, L1-ORF1p expression would need to be tested by immunofluorescence using meiotic spreads in combination with NRF1 staining in dKO, to confirm allelic KO conversion and to compare it to the same meiotic stage as the L1-ORF1 stainings in *Dnmt3C*$^{KO/KO}$. However, due to the remaining time left for our laboratory, we could not complete these experiments and test this hypothesis.

4.5.3. NRF1 binding to TEs might be restricted by other mechanisms

Interestingly, and similar to a previous study, we found that NRF1 does not bind unmethylated TEs in TKO mESCs (Domcke et al. 2015, **Figure 3.0.10A**), where TEs are transcriptionally silenced by the heterochromatic H3K27me3 and H3K9me3 marks (Walter et al. 2016). Combined with our observation that H3K27me3 marks the TEs, which are not differentially regulated in dKO before meiosis, in spermatogonia (**Figure 3.0.14**), this

suggests that H3K27me3 or other repressive histone marks may additionally prevent NRF1 from binding and activating TEs, even when they are unmethylated. In concordance with this, NRF1 was not found to bind all its high-confidence motifs in TKO mESCs when they were unmethylated and one third remained unbound, suggesting that additional mechanisms apart from DNA methylation might restrict NRF1 binding (Domcke et al. 2015). Moreover, the NRF1-bound sites in TKO mESCs often coincided with H3K27ac, the antagonistic activating mark of H3K27me3 (Domcke et al. 2015).

Another interesting observation in mammalian germline development is that the female germline remains unmethylated throughout development and therefore does not rely on DNA methylation to silence TEs (Smallwood et al. 2011; Wang et al. 2014). In mice, DNMT3C is not even expressed in the female germline and failure of the piRNA pathway does not affect female germ cells (Barau el al. 2016, Carmell et al. 2007, Shoji et al. 2009). This raises the question of why NRF1, which is expressed in the female germline, does not upregulate unmethylated TEs in this hypomethylated context. Strikingly, a recent study showed that TE silencing in oogenesis depends on H3K27me3, and the loss of this mark in *Ezh2* mutants leads to TE upregulation and germ cell loss (Huang et al. 2021). If NRF1 is blocked from binding to TEs not only by DNA methylation but also by heterochromatic marks, this could explain the observed sexual dimorphism in TE regulation.

A possible correlation of the absence of H3K27me3 and the binding of NRF1 could be bioinformatically tested in our CUT&Tag dataset. Additionally, future work could examine a causal consequence by performing in vitro pulldown assays of NRF1 with its binding sites in both methylated and unmethylated forms, as well as with assembled nucleosomes modified with H3K27me3, H3K9me3, or other heterochromatic marks. Additionally, it would be interesting to examine NRF1 binding at TEs in a hypomethylated mESC model, where H3K27me3 and H3K9me3 are depleted, using epigenome editing (Policarpi et al. 2024) or hypomethylated mESCs depleted of *Ezh2* and H3K9-methyl transferases.

Apart from histone marks occluding NRF1, NRF1 binding might also depend on indirect cooperativity with other pioneer TFs that open the heterochromatin, such as the pioneer TF REST, which has been implicated in opening and promoting demethylation of NRF1 binding sites (Domcke et al. 2015).

Overall, NRF1 acts as a DNA methylation-sensitive TF that activates IAPs before and potentially during meiosis while it may repress LINE1s before meiosis, either indirectly via the loss of H3K9me2 or directly through binding as a repressor. However, at meiosis, NRF1 may activate LINE1s in the absence of DNA methylation, as suggested by its increased binding, which may explain why DNA methylation at TEs becomes essential at this stage.

4.5.4 Transferring the mechanism of TE regulation by NRF1 to humans

NRF1 and its DNA binding domain are well-conserved from *Drosophila* to human (Fazio et al. 2001). Its sensitivity to DNA methylation has been demonstrated not only in mice but also in humans (Domcke et al. 2015) and even in a marine sponge family, which exhibits high DNA methylation patterns similar to those of mammalian genomes (de Mendoza et al. 2019). This suggests that NRF1's methylation dependency might be highly conserved across species as well.

Moreover, another study showed that a single-nucleotide polymorphism in the *Dazl* promoter, located within the NRF1 binding motif, results in reduced NRF1 binding and correlates with infertility in human (Teng et al. 2012). This suggests that NRF1 may function not only as a master regulator of germline genes in mouse spermatogenesis but also in humans. Thus, this raises the question of whether our finding that NRF1 regulates TEs in mice spermatogenesis, might also hold true for humans.

While IAPs are largely inactive in humans, L1 elements are still active and failure of the piRNA pathway in humans, was shown to also lead to L1 upregulation and infertility in males, similar to mice (Stallmeyer et al. 2024). However, as described in Chapter **1.1.1.3.1.** Long Interspersed Nuclear Elements, the L1 promoter sequences in humans differ from those in mice. Unlike murine L1 promoters, human L1 promoters do not contain highly repetitive CpG-rich monomers. Thus, the presence of NRF1 binding motifs of L1s may differ between the two species as well. Nevertheless, reanalysis of data from the ChIP-Atlas revealed that NRF1 is enriched at the 5' UTR of human L1 elements (Li et al. 2024). This enrichment might be further increased when DNA methylation is absent at L1 promoters. Methods such as DNA Affinity Purification Sequencing (DAP-Seq) could be exploited in the future to test such a hypothesis in the highly restricted material from testicular biopsy of men (Bartlett et al. 2017). In this method, the purified tagged NRF1 protein is incubated with the libraries of genomic DNA, where DNA methylation is left intact, from a certain tissue and TF-DNA complexes are purified.

Taken together, these evidences suggest that while our model of TE regulation in DNMT3C-deficient male germ cells is specific to mice, since DNMT3C evolved only in Muroidea, the regulation of TEs by NRF1 that we found in this model, particularly for L1 elements, may be conserved in other species, indicating possible broader implications beyond mice.

4.6 Unraveling the Causes of Meiotic Arrest upon failure of piRNA pathway-mediated DNA Methylation at TE promoters

Since the first discovery that defects in DNA methylation at TE promoters mediated by the piRNA pathway lead to meiotic catastrophe (Bourc'his and Bestor 2004), multiple studies have reported a similar phenotype in piRNA pathway mutants (Carmell et al. 2007; Soper et al. 2008; Reuter et al. 2009; Shoji et al. 2009; Zamudio et al. 2015; Barau et al. 2016; Zoch et al. 2020). These mutants are all characterized by defective chromosome pairing at the pachytene stage of meiosis, a process known as synapsis, coinciding with major L1 upregulation. Surprisingly, this germ cell arrest occurs in meiosis several days after the initial DNA methylation defects at TEs arise in fetal stages. This raises a key question in the field, which also caught our attention: what drives this phenotype, and why is meiosis particularly vulnerable to TE activity?

Several hypotheses have been proposed, including genomic instability caused by L1 retrotransposition (Soper et al. 2008), perturbation of gene expression (Bourc'his and Bestor 2004), and aberrant SPO11-induced meiotic double-strand breaks (DSBs) at H3K4me3-marked TE promoters (Zamudio et al. 2015). Surprisingly, the most intuitive explanation, that TE upregulation leads to genomic instability caused by massive retrotransposition events, does not seem to account for the phenotype: Depletion of *Dnmt3l* or *Miwi2* does not lead to increased DNA damage, which would be expected upon widespread retrotransposition, nor does it result in a significant number of new TE insertions (Bourc'his and Bestor 2004; Zamudio et al. 2015). While overall DNA damage does not seem to increase in piRNA pathway mutants, meiotic DSBs have been shown to shift toward TEs due to gain of H3K4me3 marks, turning them into meiotic hotspots, potentially creating challenges for DSB repair (Zamudio et al. 2015). Based on these studies and the results of my project, I propose additional explanations to the meiotic arrest phenotype while ruling out some others.

4.6.1 Potential Mechanisms Contributing to Meiotic Arrest and Future Directions

One possible cause is that germline gene deregulation contributes to meiotic failure. We identified NRF1 as a TF that regulates both germline genes and TEs upon DNA hypomethylation (Wang et al. 2017) (**Figure 3.0.11**, **Figure 3.0.12**). The creation of thousands of new NRF1 binding sites at unmethylated TEs that increase in meiosis could sequester NRF1 away from its usual germline gene targets, potentially disrupting meiosis (**Figure 3.0.11**). This hypothesis is supported by the fact that *Nrf1* depletion in the context

of different spermatogenesis stages leads to deregulation of key germline genes and germ cell arrest (Wang et al. 2017; Wang et al. 2024). Moreover, by analyzing called peaks from NRF1 CUT&Tag, we observed a reduced number of peaks in *Dnmt3C*$^{KO/KO}$ at loci that did not overlap with DNMT3C-targeted TEs, suggesting that while gaining peaks at DNMT3C targets, other NRF1 peaks may be lost (**Figure 3.0.11F**). However, if this were the primary reason for the meiotic failure, we would expect to observe widespread differential gene expression in *Dnmt3C*$^{KO/KO}$ mutants. Instead, our differential expression analysis between *Dnmt3C*$^{KO/KO}$ and wild-type germ cells at P5 revealed only a few differentially expressed genes, none of which were identified as NRF1 targets in germ cells, making this an unlikely primary cause (data not shown) (Wang et al. 2017). Nevertheless, this might differ when analyzing gene expression later in development, because NRF1 gains binding at unmethylated TEs at meiosis and NRF1 expression ceases in meiosis, restricting NRF1 protein availability further (**Figure 3.0.11C**).

Another interesting observation is that the meiotic arrest phenotype appears incomplete when not all young TE families are unmethylated, gain H3K4me3 marks, and are upregulated. Mice carrying mutations for the SPIN1-SPOCD1 interaction, exhibit specific L1 upregulation while IAPs remain unaffected, show a different germ cell phenotype than *Dnmt3C* or *Dnmt3L* mutants. SPIN1-SPOCD1 interaction mutants display a mixed phenotype, where some seminiferous tubules contain germ cells progressing past meiosis while others are arrested (Dias Mirandela et al. 2024). Similarly, a study on *Mili* endonuclease mutants found that L1 upregulation alone led to only partial germ cell arrest (De Fazio et al. 2011). This suggests that L1 hypomethylation and upregulation alone do not fully account for the meiotic arrest phenotype seen in *Dnmt3C* or *Dnmt3L* mutants. Likewise, IAP expression alone also appears insufficient to cause the sterility phenotype, as *Setdb1* KO animals exhibit IAP upregulation specifically and show only partial germ cell arrest (Liu et al. 2014).

Thus, two potential explanations remain likely: (1) aberrant meiotic DSB formation at H3K4me3-marked TE promoters and (2) cytotoxicity resulting from the combined upregulation of L1 and IAP elements. To prove one of these explanations true, further investigation is needed to disentangle the effects of H3K4me3-marked TE promoters from those of TE upregulation. One possible approach for this would be to deplete the TF responsible for TE upregulation at meiosis, potentially NRF1, while retaining H3K4me3 marks. If TE silencing were rescued without affecting the presence of H3K4me3 marks, a rescue of the meiotic arrest phenotype would provide key insights into whether aberrant DSB formation or TE expression itself is the primary driver of meiotic arrest. However, testing this in the case of NRF1 is largely limited by the germ cell arrest occurring when depleting NRF1 due to germline genes being deregulated. While other TFs that could potentially regulate TEs in the absence of DNA methylation might not cause germ cell

arrest when depleted, this relies on the discovery of these TFs. Another possibility would be to remove only the H3K4me3 mark, by depleting the factor that is responsible for its deposition at TEs, if such a specific factor can be identified. H3K4me3-mark formation at unmethylated TEs was previously suggested to rely on PRDM9, the enzyme being responsible for the deposition of H3K4me3 at meiotic hotspots, but if the deposition at TEs depends on PRDM9 still remains to be tested (Parvanov et al. 2010; Zamudio et al. 2015).

References

Aapola, U., K. Kawasaki, H. S. Scott, J. Ollila, M. Vihinen, M. Heino, A. Shintani, et al. 2000. “Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5-Methyltransferase 3 Gene Family.” *Genomics* 65 (3): 293–98.

Ade, Catherine, Astrid M. Roy-Engel, and Prescott L. Deininger. 2013. “Alu Elements: An Intrinsic Source of Human Genome Instability.” *Current Opinion in Virology* 3 (6): 639–45.

Akopov, S. B., L. G. Nikolaev, P. P. Khil, Y. B. Lebedev, and E. D. Sverdlov. 1998. “Long Terminal Repeats of Human Endogenous Retrovirus K Family (HERV-K) Specifically Bind Host Cell Nuclear Proteins.” *FEBS Letters* 421 (3): 229–33.

Almeida, Miguel Vasconcelos, Grégoire Vernaz, Audrey L. K. Putman, and Eric A. Miska. 2022. “Taming Transposable Elements in Vertebrates: From Epigenetic Silencing to Domestication.” *Trends in Genetics: TIG* 38 (6): 529–53.

Aloia, Luigi, Bruno Di Stefano, and Luciano Di Croce. 2013. “Polycomb Complexes in Stem Cells and Embryonic Development.” *Development (Cambridge, England)* 140 (12): 2525–34.

Andrews, Simon, Christel Krueger, Maravillas Mellado-Lopez, Myriam Hemberger, Wendy Dean, Vicente Perez-Garcia, and Courtney W. Hanna. 2023. “Mechanisms and Function of de Novo DNA Methylation in Placental Development Reveals an Essential Role for DNMT3B.” *Nature Communications* 14 (1): 371.

Aravin, Alexei A., Ravi Sachidanandam, Deborah Bourc’his, Christopher Schaefer, Dubravka Pezic, Katalin Fejes Toth, Timothy Bestor, and Gregory J. Hannon. 2008. “A piRNA Pathway Primed by Individual Transposons Is Linked to de Novo DNA Methylation in Mice.” *Molecular Cell* 31 (6): 785–99.

Arnaiz, Olivier, Nathalie Mathy, Céline Baudry, Sophie Malinsky, Jean-Marc Aury, Cyril Denby Wilkes, Olivier Garnier, et al. 2012. “The Paramecium Germline Genome Provides a Niche for Intragenic Parasitic DNA: Evolutionary Dynamics of Internal Eliminated Sequences.” *PLoS Genetics* 8 (10): e1002984.

Babaian, Artem, and Dixie L. Mager. 2016. “Endogenous Retroviral Promoter Exaptation in Human Cancer.” *Mobile DNA* 7 (1): 24.

Babushok, Daria V., and Haig H. Kazazian Jr. 2007. “Progress in Understanding the Biology of the Human Mutagen LINE-1.” *Human Mutation* 28 (6): 527–39.

Baillie, J. Kenneth, Mark W. Barnett, Kyle R. Upton, Daniel J. Gerhardt, Todd A. Richmond, Fioravante De Sapio, Paul M. Brennan, et al. 2011. “Somatic Retrotransposition Alters the Genetic Landscape of the Human Brain.” *Nature* 479 (7374): 534–37.

Bannister, A. J., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides. 2001. “Selective Recognition of Methylated Lysine 9 on Histone H3 by the HP1 Chromo Domain.” *Nature* 410 (6824): 120–24.

Baranello, Laura, Fedor Kouzine, Suzanne Sanford, and David Levens. 2016. “ChIP Bias as a Function of Cross-Linking Time.” *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 24 (2): 175–81.

Barau, Joan, Aurélie Teissandier, Natasha Zamudio, Stéphanie Roy, Valérie Nalesso, Yann Hérault, Florian Guillou, and Déborah Bourc’his. 2016. “The DNA Methyltransferase DNMT3C Protects Male Germ Cells from Transposon Activity.” *Science (New York, N.Y.)* 354 (6314): 909–12.

Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. 2007. “High-Resolution Profiling of Histone Methylations in the Human Genome.” *Cell* 129 (4): 823–37.

Baubec, Tuncay, Robert Ivánek, Florian Lienert, and Dirk Schübeler. 2013. “Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family.” *Cell* 153 (2): 480–92.

Bejerano, Gill, Craig B. Lowe, Nadav Ahituv, Bryan King, Adam Siepel, Sofie R. Salama, Edward M. Rubin, W. James Kent, and David Haussler. 2006. “A Distal Enhancer and an Ultraconserved Exon Are Derived from a Novel Retroposon.” *Nature* 441 (7089): 87–90.

Bénit, L., J. B. Lallemand, J. F. Casella, H. Philippe, and T. Heidmann. 1999. “ERV-L Elements: A Family of Endogenous Retrovirus-like Elements Active throughout the Evolution of Mammals.” *Journal of Virology* 73 (4): 3301–8.

Bennetzen, Jeffrey L., and Hao Wang. 2014. “The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes.” *Annual Review of Plant Biology* 65 (1): 505–30.

Bernstein, Bradley E., Tarjei S. Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J. Huebert, James Cuff, Ben Fry, et al. 2006. “A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells.” *Cell* 125 (2): 315–26.

Berrens, Rebecca V., Simon Andrews, Dominik Spensberger, Fátima Santos, Wendy Dean, Poppy Gould, Jafar Sharif, et al. 2017. “An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells.” *Cell Stem Cell* 21 (5): 694–703.e7.

Bertozzi, Tessa M., Nozomi Takahashi, Geula Hanin, Anastasiya Kazachenka, and Anne C. Ferguson-Smith. 2021. “A Spontaneous Genetically Induced Epiallele at a Retrotransposon Shapes Host Genome Function.” *eLife* 10 (March). https://doi.org/10.7554/eLife.65233.

Bestor, Timothy, Andrew Laudano, Robert Mattaliano, and Vernon Ingram. 1988. “Cloning and Sequencing of a cDNA Encoding DNA Methyltransferase of Mouse Cells.” *Journal of Molecular Biology* 203 (4): 971–83.

Beyret, Ergin, Na Liu, and Haifan Lin. 2012. “piRNA Biogenesis during Adult Spermatogenesis in Mice Is Independent of the Ping-Pong Mechanism.” *Cell Research* 22 (10): 1429–39.

Bird, A. P. 1980. “DNA Methylation and the Frequency of CpG in Animal DNA.” *Nucleic Acids Research* 8 (7): 1499–1504.

———. 1986. “CpG-Rich Islands and the Function of DNA Methylation.” *Nature* 321 (6067): 209–13.

Borgel, Julie, Sylvain Guibert, Yufeng Li, Hatsune Chiba, Dirk Schübeler, Hiroyuki Sasaki, Thierry Forné, and Michael Weber. 2010. “Targets and Dynamics of Promoter DNA Methylation during Early Mouse Development.” *Nature Genetics* 42 (12): 1093–1100.

Boulard, Mathieu, John R. Edwards, and Timothy H. Bestor. 2015. “FBXL10 Protects Polycomb-Bound Genes from Hypermethylation.” *Nature Genetics* 47 (5): 479–85.

Boulard, Mathieu, Sofia Rucli, John R. Edwards, and Timothy H. Bestor. 2020. “Methylation-Directed Glycosylation of Chromatin Factors Represses Retrotransposon Promoters.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (25): 14292–98.

Bourc’his, Déborah, and Timothy H. Bestor. 2004. “Meiotic Catastrophe and Retrotransposon Reactivation in Male Germ Cells Lacking Dnmt3L.” *Nature* 431 (7004): 96–99.

Bourque, Guillaume, Kathleen H. Burns, Mary Gehring, Vera Gorbunova, Andrei Seluanov, Molly Hammell, Michaël Imbeault, et al. 2018. “Ten Things You Should Know about Transposable Elements.” *Genome Biology* 19 (1): 199.

Bourque, Guillaume, Bernard Leong, Vinsensius B. Vega, Xi Chen, Yen Ling Lee, Kandhadayar G. Srinivasan, Joon-Lin Chew, et al. 2008. “Evolution of the Mammalian Transcription Factor Binding Repertoire via Transposable Elements.” *Genome Research* 18 (11): 1752–62.

Brandeis, M., D. Frank, I. Keshet, Z. Siegfried, M. Mendelsohn, A. Nemes, V. Temper, A. Razin, and H. Cedar. 1994. “Sp1 Elements Protect a CpG Island from de Novo Methylation.” *Nature* 371 (6496): 435–38.

Brind’Amour, Julie, Sheng Liu, Matthew Hudson, Carol Chen, Mohammad M. Karimi, and Matthew C. Lorincz. 2015. “An Ultra-Low-Input Native ChIP-Seq Protocol for Genome-Wide Profiling of Rare Cell Populations.” *Nature Communications* 6 (1): 6033.

Brouha, Brook, Joshua Schustak, Richard M. Badge, Sheila Lutz-Prigge, Alexander H. Farley, John V. Moran, and Haig H. Kazazian Jr. 2003. “Hot L1s Account for the Bulk of Retrotransposition in the Human Population.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (9): 5280–85.

Cam, Hugh, Egle Balciunaite, Alexandre Blais, Alexander Spektor, Richard C. Scarpulla, Richard Young, Yuval Kluger, and Brian David Dynlacht. 2004. “A Common Set of Gene Regulatory Networks Links Metabolism and Growth Inhibition.” *Molecular Cell* 16 (3): 399–411.

Cao, Ru, and Yi Zhang. 2004. “The Functions of E(Z)/EZH2-Mediated Methylation of Lysine 27 in Histone H3.” *Current Opinion in Genetics & Development* 14 (2): 155–64.

Carmell, Michelle A., Angélique Girard, Henk J. G. van de Kant, Deborah Bourc’his, Timothy H. Bestor, Dirk G. de Rooij, and Gregory J. Hannon. 2007. “MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline.” *Developmental Cell* 12 (4): 503–14.

Chan, Frieda, Melissa J. Oatley, Amy V. Kaucher, Qi-En Yang, Charles J. Bieberich, Cooduvalli S. Shashikant, and Jon M. Oatley. 2014. “Functional and Molecular Features of the Id4+ Germline Stem Cell Population in Mouse Testes.” *Genes & Development* 28 (12): 1351–62.

Chédin, Frédéric. 2011. “The DNMT3 Family of Mammalian DE Novo DNA Methyltransferases.” In *Progress in Molecular Biology and Translational Science*, 255–85. Progress in Molecular Biology and Translational Science. Elsevier.

Chen, Kaifu, Zheng Hu, Zheng Xia, Dongyu Zhao, Wei Li, and Jessica K. Tyler. 2015. “The Overlooked Fact: Fundamental Need for Spike-in Control for Virtually All Genome-Wide Analyses.” *Molecular and Cellular Biology* 36 (5): 662–67.

Chen, Taiping, Yoshihide Ueda, Jonathan E. Dodge, Zhenjuan Wang, and En Li. 2003. “Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b.” *Molecular and Cellular Biology* 23 (16): 5594–5605.

Chuang, L. S., H. I. Ian, T. W. Koh, H. H. Ng, G. Xu, and B. F. Li. 1997. “Human DNA-(cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1.” *Science (New York, N.Y.)* 277 (5334): 1996–2000.

Clouaire, Thomas, Shaun Webb, Pete Skene, Robert Illingworth, Alastair Kerr, Robert Andrews, Jeong-Heon Lee, David Skalnik, and Adrian Bird. 2012. “Cfp1 Integrates Both CpG Content and Gene Activity for Accurate H3K4me3 Deposition in Embryonic Stem Cells.” *Genes & Development* 26 (15): 1714–28.

Cornelis, Guillaume, Mathis Funk, Cécile Vernochet, Francisca Leal, Oscar Alejandro Tarazona, Guillaume Meurice, Odile Heidmann, et al. 2017. “An Endogenous Retroviral Envelope Syncytin and Its Cognate Receptor Identified in the Viviparous Placental *Mabuya* Lizard.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (51): E10991–0.

Cornelis, Guillaume, Cécile Vernochet, Quentin Carradec, Sylvie Souquere, Baptiste Mulot, François Catzeflis, Maria A. Nilsson, et al. 2015. “Retroviral Envelope Gene Captures and Syncytin Exaptation for Placentation in Marsupials.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (5): E487–96.

Cusanovich, Darren A., Bryan Pavlovic, Jonathan K. Pritchard, and Yoav Gilad. 2014. “The Functional Consequences of Variation in Transcription Factor Binding.” *PLoS Genetics* 10 (3): e1004226.

Czech, Benjamin, and Gregory J. Hannon. 2016. “One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing.” *Trends in Biochemical Sciences* 41 (4): 324–37.

De Fazio, Serena, Nenad Bartonicek, Monica Di Giacomo, Cei Abreu-Goodger, Aditya Sankar, Charlotta Funaya, Claude Antony, Pedro N. Moreira, Anton J. Enright, and Dónal O’Carroll. 2011. “The Endonuclease Activity of Mili Fuels piRNA Amplification That Silences LINE1 Elements.” *Nature* 480 (7376): 259–63.

Deng, Wei, and Haifan Lin. 2002. “Miwi, a Murine Homolog of Piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis.” *Developmental Cell* 2 (6): 819–30.

Denli, Ahmet M., Iñigo Narvaiza, Bilal E. Kerman, Monique Pena, Christopher Benner, Maria C. N. Marchetto, Jolene K. Diedrich, et al. 2015. “Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity.” *Cell* 163 (3): 583–93.

Dennis, K., T. Fan, T. Geiman, Q. Yan, and K. Muegge. 2001. “Lsh, a Member of the SNF2 Family, Is Required for Genome-Wide Methylation.” *Genes & Development* 15 (22): 2940–44.

Dewannieux, Marie, Anne Dupressoir, Francis Harper, Gérard Pierron, and Thierry Heidmann. 2004. “Identification of Autonomous IAP LTR Retrotransposons Mobile in Mammalian Cells.” *Nature Genetics* 36 (5): 534–39.

Dewannieux, Marie, Cécile Esnault, and Thierry Heidmann. 2003. “LINE-Mediated Retrotransposition of Marked Alu Sequences.” *Nature Genetics* 35 (1): 41–48.

Dias Mirandela, Madeleine, Ansgar Zoch, Jessica Leismann, Shaun Webb, Rebecca V. Berrens, Devisree Valsakumar, Yuka Kabayama, et al. 2024. “Two-Factor Authentication Underpins the Precision of the piRNA Pathway.” *Nature*, September, 1–7.

Dickson, Bradley M., Rochelle L. Tiedemann, Alison A. Chomiak, Evan M. Cornett, Robert M. Vaughan, and Scott B. Rothbart. 2020. “A Physical Basis for Quantitative ChIP-Sequencing.” *The Journal of Biological Chemistry* 295 (47): 15826–37.

Di Giacomo, Monica, Stefano Comazzetto, Harpreet Saini, Serena De Fazio, Claudia Carrieri, Marcos Morgan, Lina Vasiliauskaite, Vladimir Benes, Anton J. Enright, and Dónal O’Carroll. 2013. “Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis.” *Molecular Cell* 50 (4): 601–8.

Di Giacomo, Monica, Stefano Comazzetto, Srihari C. Sampath, Srinath C. Sampath, and Dónal O’Carroll. 2014. “G9a Co-Suppresses LINE1 Elements in Spermatogonia.” *Epigenetics & Chromatin* 7 (1): 24.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics (Oxford, England)* 29 (1): 15–21.

Domcke, Silvia, Anaïs Flore Bardet, Paul Adrian Ginno, Dominik Hartl, Lukas Burger, and Dirk Schübeler. 2015. “Competition between DNA Methylation and Transcription Factors Determines Binding of NRF1.” *Nature* 528 (7583): 575–79.

Dong, Juan, Xiaoli Wang, Congcong Cao, Yujiao Wen, Akihiko Sakashita, Si Chen, Jin Zhang, et al. 2019. “UHRF1 Suppresses Retrotransposons and Cooperates with PRMT5 and PIWI Proteins in Male Germ Cells.” *Nature Communications* 10 (1): 4705.

Dumesic, Phillip A., Christina M. Homer, James J. Moresco, Lindsey R. Pack, Erin K. Shanle, Scott M. Coyle, Brian D. Strahl, Danica G. Fujimori, John R. Yates III, and Hiten D. Madhani. 2015. “Product Binding Enforces the Genomic Specificity of a Yeast Polycomb Repressive Complex.” *Cell* 160 (1-2): 204–18.

Dura, Mathilde, Aurélie Teissandier, Mélanie Armand, Joan Barau, Clémentine Lapoujade, Pierre Fouchet, Lorraine Bonneville, et al. 2022. “DNMT3A-Dependent DNA Methylation Is Required for Spermatogonial Stem Cells to Commit to Spermatogenesis.” *Nature Genetics* 54 (4): 469–80.

Duttke, Sascha H., Carlos Guzman, Max Chang, Nathaniel P. Delos Santos, Bayley R. McDonald, Jialei Xie, Aaron F. Carlin, Sven Heinz, and Christopher Benner. 2024. “Position-Dependent Function of Human Sequence-Specific Transcription Factors.” *Nature* 631 (8022): 891–98.

Eickbush, Thomas H., and Harmit S. Malik. 2002. “Origins and Evolution of Retrotransposons.” In *Mobile DNA II*, 1111–44. American Society of Microbiology.

Escrivá, H., A. Rodríguez-Peña, and C. G. Vallejo. 1999. “Expression of Mitochondrial Genes and of the Transcription Factors Involved in the Biogenesis of Mitochondria Tfam, NRF-1 and NRF-2, in Rat Liver, Testis and Brain.” *Biochimie* 81 (10): 965–71.

Evans, M. J., and R. C. Scarpulla. 1989. “Interaction of Nuclear Factors with Multiple Sites in the Somatic Cytochrome c Promoter.” *The Journal of Biological Chemistry* 264 (24): 14361–68.

Everett, Logan, Matthew Hansen, and Sridhar Hannenhalli. 2010. “Regulating the Regulators: Modulators of Transcription Factor Activity.” In *Methods in Molecular Biology*, 297–312. Methods in Molecular Biology (Clifton, N.J.). Totowa, NJ: Humana Press.

Fayomi, Adetunji P., and Kyle E. Orwig. 2018. “Spermatogonial Stem Cells and Spermatogenesis in Mice, Monkeys and Men.” *Stem Cell Research* 29 (May):207–14.

Feldmann, Angelika, Robert Ivanek, Rabih Murr, Dimos Gaidatzis, Lukas Burger, and Dirk Schübeler. 2013. “Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions.” *PLoS Genetics* 9 (12): e1003994.

Filion, Guillaume J. P., Svetlana Zhenilo, Sergey Salozhin, Daisuke Yamada, Egor Prokhortchouk, and Pierre-Antoine Defossez. 2006. “A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription.” *Molecular and Cellular Biology* 26 (1): 169–81.

Finnegan, David J. 1989. “Eukaryotic Transposable Elements and Genome Evolution.” *Trends in Genetics: TIG* 5:103–7.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. 1998. “Potent and Specific Genetic Interference by Double-Stranded RNA in Caenorhabditis Elegans.” *Nature* 391 (6669): 806–11.

Flemr, Matyas, Radek Malik, Vedran Franke, Jana Nejepinska, Radislav Sedlacek, Kristian Vlahovicek, and Petr Svoboda. 2013. “A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes.” *Cell* 155 (4): 807–16.

Fueyo, Raquel, Julius Judd, Cedric Feschotte, and Joanna Wysocka. 2022. “Roles of Transposable Elements in the Regulation of Mammalian Transcription.” *Nature Reviews. Molecular Cell Biology* 23 (7): 481–97.

Fuks, F., W. A. Burgers, A. Brehm, L. Hughes-Davies, and T. Kouzarides. 2000. “DNA Methyltransferase Dnmt1 Associates with Histone Deacetylase Activity.” *Nature Genetics* 24 (1): 88–91.

Gahan, James M., Lily W. Helfrich, Laura A. Wetzel, Natarajan V. Bhanu, Zuo-Fei Yuan, Benjamin A. Garcia, Rob Klose, and David S. Booth. 2024. “Chromatin Profiling Identifies Putative Dual Roles for H3K27me3 in Regulating Transposons and Cell Type-Specific Genes in Choanoflagellates.” *bioRxiv.org: The Preprint Server for Biology*, May, 2024.05.28.596151.

Gainetdinov, Ildar, Cansu Colpan, Amena Arif, Katharine Cecchini, and Phillip D. Zamore. 2018. “A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals.” *Molecular Cell* 71 (5): 775–90.e5.

Galen, Peter van, Aaron D. Viny, Oren Ram, Russell J. H. Ryan, Matthew J. Cotton, Laura Donohue, Cem Sievers, et al. 2016. “A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes.” *Molecular Cell* 61 (1): 170–80.

Gallardo, Teresa, Lane Shirley, George B. John, and Diego H. Castrillon. 2007. “Generation of a Germ Cell-Specific Mouse Transgenic Cre Line, Vasa-Cre.” *Genesis (New York, N.Y.: 2000)* 45 (6): 413–17.

Garcia-Perez, Jose L., Maria C. N. Marchetto, Alysson R. Muotri, Nicole G. Coufal, Fred H. Gage, K. Sue O’Shea, and John V. Moran. 2007. “LINE-1 Retrotransposition in Human Embryonic Stem Cells.” *Human Molecular Genetics* 16 (13): 1569–77.

Gazquez-Gutierrez, Ana, Jeroen Witteveldt, Sara R Heras, and Sara Macias. 2021. “Sensing of Transposable Elements by the Antiviral Innate Immune System.” *RNA (New York, N.Y.)* 27 (7): 735–52.

Gerdes, Patricia, Dorothy Chan, Mischa Lundberg, Francisco J. Sanchez-Luque, Gabriela O. Bodea, Adam D. Ewing, Geoffrey J. Faulkner, and Sandra R. Richardson. 2023. “Locus-Resolution Analysis of L1 Regulation and Retrotransposition Potential in Mouse Embryonic Development.” *Genome Research* 33 (9): 1465–81.

Gewiss, Rachel, Traci Topping, and Michael D. Griswold. 2020. “Cycles, Waves, and Pulses: Retinoic Acid and the Organization of Spermatogenesis.” *Andrology* 8 (4): 892–97.

Gifford, Robert, Peter Kabat, Joanne Martin, Clare Lynch, and Michael Tristem. 2005. “Evolution and Distribution of Class II-Related Endogenous Retroviruses.” *Journal of Virology* 79 (10): 6478–86.

Gilbert, Clément, and Cédric Feschotte. 2018. “Horizontal Acquisition of Transposable Elements and Viral Sequences: Patterns and Consequences.” *Current Opinion in Genetics & Development* 49 (April):15–24.

Goodier, J. L., E. M. Ostertag, K. Du, and H. H. Kazazian Jr. 2001. “A Novel Active L1 Retrotransposon Subfamily in the Mouse.” *Genome Research* 11 (10): 1677–85.

Goodier, John L., Lili Zhang, Melissa R. Vetter, and Haig H. Kazazian Jr. 2007. “LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced Silencing Complex.” *Molecular and Cellular Biology* 27 (18): 6469–83.

Gopalan, Sneha, Yuqing Wang, Nicholas W. Harper, Manuel Garber, and Thomas G. Fazzio. 2021. “Simultaneous Profiling of Multiple Chromatin Proteins in the Same Cells.” *Molecular Cell* 81 (22): 4736–46.e5.

Grabundzija, Ivana, Simon A. Messing, Jainy Thomas, Rachel L. Cosby, Ilija Bilic, Csaba Miskey, Andreas Gogol-Döring, et al. 2016. “A Helitron Transposon Reconstructed from Bats Reveals a Novel Mechanism of Genome Shuffling in Eukaryotes.” *Nature Communications* 7 (1): 10716.

Greenblatt, Irwin M., and R. Alexander Brink. 1963. “Transpositions of Modulator in Maize into Divided and Undivided Chromosome Segments.” *Nature* 197 (4865): 412–13.

Green, Christopher Daniel, Qianyi Ma, Gabriel L. Manske, Adrienne Niederriter Shami, Xianing Zheng, Simone Marini, Lindsay Moritz, et al. 2018. “A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq.” *Developmental Cell* 46 (5): 651–67.e10.

Guibert, Sylvain, Thierry Forné, and Michael Weber. 2012. “Global Profiling of DNA Methylation Erasure in Mouse Primordial Germ Cells.” *Genome Research* 22 (4): 633–41.

Gupta, Priyanka, Rajni Sharma, and Vandana Jain. 2022. “Adrenal Hypoplasia Congenita-Hypogonadotropic Hypogonadism Syndrome due to NR0B1 Gene Mutations.” *Indian Journal of Pediatrics* 89 (6): 587–90.

Gu, Tian-Peng, Fan Guo, Hui Yang, Hai-Ping Wu, Gui-Fang Xu, Wei Liu, Zhi-Guo Xie, et al. 2011. “The Role of Tet3 DNA Dioxygenase in Epigenetic Reprogramming by Oocytes.” *Nature* 477 (7366): 606–10.

Haig, David. 2016. “Transposable Elements: Self-Seekers of the Germline, Team-Players of the Soma.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 38 (11): 1158–66.

Hall, Ira M., Gurumurthy D. Shankaranarayana, Ken-Ichi Noma, Nabieh Ayoub, Amikam Cohen, and Shiv I. S. Grewal. 2002. “Establishment and Maintenance of a Heterochromatin Domain.” *Science (New York, N.Y.)* 297 (5590): 2232–37.

Handel, Mary Ann, and John C. Schimenti. 2010. “Genetics of Mammalian Meiosis: Regulation, Dynamics and Impact on Fertility.” *Nature Reviews. Genetics* 11 (2): 124–36.

Haring, Nina L., Elisabeth J. van Bree, Whitney S. Jordaan, Judith R. E. Roels, Gonzalo Congrains Sotomayor, Tiziana M. Hey, Fred T. G. White, Marc D. Galland, Marten P. Smidt, and Frank M. J. Jacobs. 2021. “ZNF91 Deletion in Human Embryonic Stem Cells Leads to Ectopic Activation of SVA Retrotransposons and up-Regulation of KRAB Zinc Finger Gene Clusters.” *Genome Research* 31 (4): 551–63.

Harris, C. Jake, Marion Scheibe, Somsakul Pop Wongpalee, Wanlu Liu, Evan M. Cornett, Robert M. Vaughan, Xueqin Li, et al. 2018. “A DNA Methylation Reader Complex That Enhances Gene Transcription.” *Science (New York, N.Y.)* 362 (6419): 1182–86.

Heintzman, Nathaniel D., Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, et al. 2009. “Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression.” *Nature* 459 (7243): 108–12.

Hermant, Clara, and Maria-Elena Torres-Padilla. 2021. “TFs for TEs: The Transcription Factor Repertoire of Mammalian Transposable Elements.” *Genes & Development* 35 (1-2): 22–39.

Hess, Rex A., and Luiz Renato de Franca. 2009. “Spermatogenesis and Cycle of the Seminiferous Epithelium.” In *Advances in Experimental Medicine and Biology*, 1–15. Advances in Experimental Medicine and Biology. New York, NY: Springer New York.

Hof, Arjen E. Van’t, Pascal Campagne, Daniel J. Rigden, Carl J. Yung, Jessica Lingley, Michael A. Quail, Neil Hall, Alistair C. Darby, and Ilik J. Saccheri. 2016. “The Industrial Melanism Mutation in British Peppered Moths Is a Transposable Element.” *Nature* 534 (7605): 102–5.

Howard, G., R. Eiges, F. Gaudet, R. Jaenisch, and A. Eden. 2008. “Activation and Transposition of Endogenous Retroviral Elements in Hypomethylation Induced Tumors in Mice.” *Oncogene* 27 (3): 404–8.

Howe, Françoise S., Harry Fischl, Struan C. Murray, and Jane Mellor. 2017. “Is H3K4me3 Instructive for Transcription Activation?” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 39 (1): 1–12.

Huang, Tien-Chi, Yi-Fang Wang, Eric Vazquez-Ferrer, Ina Theofel, Cristina E. Requena, Courtney W. Hanna, Gavin Kelsey, and Petra Hajkova. 2021. “Sex-Specific Chromatin Remodelling Safeguards Transcription in Germ Cells.” *Nature* 600 (7890): 737–42.

Huo, L., and R. C. Scarpulla. 2001. “Mitochondrial DNA Instability and Peri-Implantation Lethality Associated with Targeted Disruption of Nuclear Respiratory Factor 1 in Mice.” *Molecular and Cellular Biology* 21 (2): 644–54.

Hure, Valentin, Florence Piron-Prunier, Tamara Yehouessi, Clémentine Vitte, Aleksandra E. Kornienko, Gabrielle Adam, Magnus Nordborg, and Angélique Déléris. 2024. “Alternative Silencing States of Transposable Elements in Arabidopsis.” *bioRxiv*. https://doi.org/10.1101/2024.03.16.585326.

Illingworth, Robert S., Ulrike Gruenewald-Schneider, Shaun Webb, Alastair R. W. Kerr, Keith D. James, Daniel J. Turner, Colin Smith, David J. Harrison, Robert Andrews, and Adrian P. Bird. 2010. “Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome.” *PLoS Genetics* 6 (9): e1001134.

Imbeault, Michaël, Pierre-Yves Helleboid, and Didier Trono. 2017. “KRAB Zinc-Finger Proteins Contribute to the Evolution of Gene Regulatory Networks.” *Nature* 543 (7646): 550–54.

Ishino, Kyoko, Hidetoshi Hasuwa, Jun Yoshimura, Yuka W. Iwasaki, Hidenori Nishihara, Naomi M. Seki, Takamasa Hirano, et al. 2021. “Hamster PIWI Proteins Bind to piRNAs with Stage-Specific Size Variations during Oocyte Maturation.” *Nucleic Acids Research* 49 (5): 2700–2720.

Iskow, Rebecca C., Michael T. McCabe, Ryan E. Mills, Spencer Torene, W. Stephen Pittard, Andrew F. Neuwald, Erwin G. Van Meir, Paula M. Vertino, and Scott E. Devine. 2010. “Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons.” *Cell* 141 (7): 1253–61.

Iwafuchi-Doi, Makiko, and Kenneth S. Zaret. 2014. “Pioneer Transcription Factors in Cell Reprogramming.” *Genes & Development* 28 (24): 2679–92.

Jain, Devanshi, Cem Meydan, Julian Lange, Corentin Claeys Bouuaert, Nathalie Lailler, Christopher E. Mason, Kathryn V. Anderson, and Scott Keeney. 2017. “Rahu Is a Mutant Allele of Dnmt3c, Encoding a DNA Methyltransferase Homolog Required for Meiosis and Transposon Repression in the Mouse Male Germline.” *PLoS Genetics* 13 (8): e1006964.

Jan, Sabrina Z., Geert Hamer, Sjoerd Repping, Dirk G. de Rooij, Ans M. M. van Pelt, and Tinke L. Vormer. 2012. “Molecular Control of Rodent Spermatogenesis.” *Biochimica et Biophysica Acta* 1822 (12): 1838–50.

Janssens, Derek, and Steven Henikoff. 2019. “CUT&RUN: Targeted in Situ Genome-Wide Profiling with High Efficiency for Low Cell Numbers v3.” *Protocols.io*. ZappyLab, Inc. https://doi.org/10.17504/protocols.io.zcpf2vn.

Jeltsch, Albert, Julian Broche, and Pavel Bashtrykov. 2018. “Molecular Processes Connecting DNA Methylation Patterns with DNA Methyltransferases and Histone Modifications in Mammalian Genomes.” *Genes* 9 (11): 566.

Jeltsch, Albert, and Renata Z. Jurkowska. 2016. “Allosteric Control of Mammalian DNA Methyltransferases - a New Regulatory Paradigm.” *Nucleic Acids Research* 44 (18): 8556–75.

Jern, Patric, and John M. Coffin. 2008. “Effects of Retroviruses on Host Genome Function.” *Annual Review of Genetics* 42 (1): 709–32.

Jin, Ying, Oliver H. Tam, Eric Paniagua, and Molly Hammell. 2015. “TEtranscripts: A Package for Including Transposable Elements in Differential Expression Analysis of RNA-Seq Datasets.” *Bioinformatics (Oxford, England)* 31 (22): 3593–99.

Jirtle, Randy L., and Michael K. Skinner. 2007. “Environmental Epigenomics and Disease Susceptibility.” *Nature Reviews. Genetics* 8 (4): 253–62.

Johnson, P. F., and S. L. McKnight. 1989. “Eukaryotic Transcriptional Regulatory Proteins.” *Annual Review of Biochemistry* 58 (1): 799–839.

Johnson, Treat B., and Robert D. Coghill. 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): 2838–44.

Kaluscha, Sebastian, Silvia Domcke, Christiane Wirbelauer, Michael B. Stadler, Sevi Durdu, Lukas Burger, and Dirk Schübeler. 2022. “Evidence That Direct Inhibition of Transcription Factor Binding Is the Prevailing Mode of Gene and Repeat Repression by DNA Methylation.” *Nature Genetics* 54 (12): 1895–1906.

Kapitonov, Vladimir V., and Jerzy Jurka. 2005. “RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons.” *PLoS Biology* 3 (6): e181.

Kapitonov, Vladimir V., Sébastien Tempel, and Jerzy Jurka. 2009. “Simple and Fast Classification of Non-LTR Retrotransposons Based on Phylogeny of Their RT Domain Protein Sequences.” *Gene* 448 (2): 207–13.

Karttunen, Konsta, Divyesh Patel, Jihan Xia, Liangru Fei, Kimmo Palin, Lauri Aaltonen, and Biswajyoti Sahu. 2023. “Transposable Elements as Tissue-Specific Enhancers in Cancers of Endodermal Lineage.” *Nature Communications* 14 (1): 5313.

Kaya-Okur, Hatice S., Derek H. Janssens, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2020. “Efficient Low-Cost Chromatin Profiling with CUT&Tag.” *Nature Protocols* 15 (10): 3264–83.

Kaya-Okur, Hatice S., Steven J. Wu, Christine A. Codomo, Erica S. Pledger, Terri D. Bryson, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2019. “CUT&Tag for Efficient Epigenomic Profiling of Small Samples and Single Cells.” *Nature Communications* 10 (1): 1930.

Kazachenka, Anastasiya, Tessa M. Bertozzi, Marcela K. Sjoberg-Herrera, Nic Walker, Joseph Gardner, Richard Gunning, Elena Pahita, Sarah Adams, David Adams, and Anne C. Ferguson-Smith. 2018. “Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-Genetic Inheritance.” *Cell* 175 (5): 1259–71.e13.

Kazazian, Haig H., Jr, and John V. Moran. 2017. “Mobile DNA in Health and Disease.” *The New England Journal of Medicine* 377 (4): 361–70.

Kidwell, M. G., J. F. Kidwell, and J. A. Sved. 1977. “Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination.” *Genetics* 86 (4): 813–33.

Klawitter, Sabine, Nina V. Fuchs, Kyle R. Upton, Martin Muñoz-Lopez, Ruchi Shukla, Jichang Wang, Marta Garcia-Cañadas, et al. 2016. “Reprogramming Triggers Endogenous L1 and Alu Retrotransposition in Human Induced Pluripotent Stem Cells.” *Nature Communications* 7 (1): 10286.

Kohli, Rahul M., and Yi Zhang. 2013. “TET Enzymes, TDG and the Dynamics of DNA Demethylation.” *Nature* 502 (7472): 472–79.

Kornberg, R. D. 1974. “Chromatin Structure: A Repeating Unit of Histones and DNA.” *Science (New York, N.Y.)* 184 (4139): 868–71.

Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell* 128 (4): 693–705.

Kramerov, Dimitri A., and Nikita S. Vassetzky. 2005. “Short Retroposons in Eukaryotic Genomes.” *International Review of Cytology* 247:165–221.

Krebs, Wolfgang, Susanne V. Schmidt, Alon Goren, Dominic De Nardo, Larisa Labzin, Anton Bovier, Thomas Ulas, et al. 2014. “Optimization of Transcription Factor Binding Map Accuracy Utilizing Knockout-Mouse Models.” *Nucleic Acids Research* 42 (21): 13051–60.

Kuff, Edward L., and Kira K. Lueders. 1988. “The Intracisternal A-Particle Gene Family: Structure and Functional Aspects.” In *Advances in Cancer Research*, 183–276. Advances in Cancer Research. Elsevier.

Kuramochi-Miyagawa, Satomi, Tohru Kimura, Takashi W. Ijiri, Taku Isobe, Noriko Asada, Yukiko Fujita, Masahito Ikawa, et al. 2004. “Mili, a Mammalian Member of Piwi Family Gene, Is Essential for Spermatogenesis.” *Development (Cambridge, England)* 131 (4): 839–49.

Lachner, M., D. O’Carroll, S. Rea, K. Mechtler, and T. Jenuwein. 2001. “Methylation of Histone H3 Lysine 9 Creates a Binding Site for HP1 Proteins.” *Nature* 410 (6824): 116–20.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921.

Lane, Natasha, Wendy Dean, Sylvia Erhardt, Petra Hajkova, Azim Surani, Jörn Walter, and Wolf Reik. 2003. “Resistance of IAPs to Methylation Reprogramming May Provide a Mechanism for Epigenetic Inheritance in the Mouse.” *Genesis (New York, N.Y.: 2000)* 35 (2): 88–93.

Lara-Astiaso, David, Assaf Weiner, Erika Lorenzo-Vivas, Irina Zaretsky, Diego Adhemar Jaitin, Eyal David, Hadas Keren-Shaul, et al. 2014. “Immunogenetics. Chromatin State Dynamics during Blood Formation.” *Science (New York, N.Y.)* 345 (6199): 943–49.

Lauria, Andrea, Guohua Meng, Valentina Proserpio, Stefania Rapelli, Mara Maldotti, Isabelle Laurence Polignano, Francesca Anselmi, et al. 2023. “DNMT3B Supports Meso-Endoderm Differentiation from Mouse Embryonic Stem Cells.” *Nature Communications* 14 (1): 367.

Leeb, Martin, Diego Pasini, Maria Novatchkova, Markus Jaritz, Kristian Helin, and Anton Wutz. 2010. “Polycomb Complexes Act Redundantly to Repress Genomic Repeats and Genes.” *Genes & Development* 24 (3): 265–76.

Lee, Eunjung, Rebecca Iskow, Lixing Yang, Omer Gokcumen, Psalm Haseley, Lovelace J. Luquette 3rd, Jens G. Lohr, et al. 2012. “Landscape of Somatic Retrotransposition in Human Cancers.” *Science (New York, N.Y.)* 337 (6097): 967–71.

Leonhardt, H., A. W. Page, H. U. Weier, and T. H. Bestor. 1992. “A Targeting Sequence Directs DNA Methyltransferase to Sites of DNA Replication in Mammalian Nuclei.” *Cell* 71 (5): 865–73.

Lev-Maor, Galit, Oren Ram, Eddo Kim, Noa Sela, Amir Goren, Erez Y. Levanon, and Gil Ast. 2008. “Intronic Alus Influence Alternative Splicing.” *PLoS Genetics* 4 (9): e1000204.

Li, E., C. Beard, and R. Jaenisch. 1993. “Role for DNA Methylation in Genomic Imprinting.” *Nature* 366 (6453): 362–65.

Li, E., T. H. Bestor, and R. Jaenisch. 1992. “Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality.” *Cell* 69 (6): 915–26.

Li, Jingfeng, Manoj Kannan, Anna L. Trivett, Hongling Liao, Xiaolin Wu, Keiko Akagi, and David E. Symer. 2014. “An Antisense Promoter in Mouse L1 Retrotransposon Open Reading Frame-1

Initiates Expression of Diverse Fusion Transcripts and Limits Retrotransposition." *Nucleic Acids Research* 42 (7): 4546–62.

Lister, Ryan, Eran A. Mukamel, Joseph R. Nery, Mark Urich, Clare A. Puddifoot, Nicholas D. Johnson, Jacinta Lucero, et al. 2013. "Global Epigenomic Reconfiguration during Mammalian Brain Development." *Science (New York, N.Y.)* 341 (6146): 1237905.

Lister, Ryan, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, et al. 2009. "Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences." *Nature* 462 (7271): 315–22.

Liu, Ke, Weifang Li, Yuqing Xiao, Ming Lei, Ming Zhang, and Jinrong Min. 2024. "Molecular Mechanism of Specific DNA Sequence Recognition by NRF1." *Nucleic Acids Research* 52 (2): 953–66.

Liu, Nan. 2019. "Library Prep for CUT&RUN with NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (E7645) v1." *Protocols.io*. ZappyLab, Inc. https://doi.org/10.17504/protocols.io.wvgfe3w.

Liu, Nian, Cameron H. Lee, Tomek Swigut, Edward Grow, Bo Gu, Michael C. Bassik, and Joanna Wysocka. 2018. "Selective Silencing of Euchromatic L1s Revealed by Genome-Wide Screens for L1 Regulators." *Nature* 553 (7687): 228–32.

Liu, Sheng, Julie Brind'Amour, Mohammad M. Karimi, Kenjiro Shirane, Aaron Bogutz, Louis Lefebvre, Hiroyuki Sasaki, Yoichi Shinkai, and Matthew C. Lorincz. 2014. "Setdb1 Is Required for Germline Development and Silencing of H3K9me3-Marked Endogenous Retroviruses in Primordial Germ Cells." *Genes & Development* 28 (18): 2041–55.

Liu, Xiaoli, Qinqin Gao, Pishun Li, Qian Zhao, Jiqin Zhang, Jiwen Li, Haruhiko Koseki, and Jiemin Wong. 2013. "UHRF1 Targets DNMT1 for DNA Methylation through Cooperative Binding of Hemi-Methylated DNA and Methylated H3K9." *Nature Communications* 4 (1): 1563.

Li, Wenxue, Myoung-Hwa Lee, Lisa Henderson, Richa Tyagi, Muzna Bachani, Joseph Steiner, Emilie Campanac, et al. 2015. "Human Endogenous Retrovirus-K Contributes to Motor Neuron Disease." *Science Translational Medicine* 7 (307): 307ra153.

Löber, Ulrike, Matthew Hobbs, Anisha Dayaram, Kyriakos Tsangaras, Kiersten Jones, David E. Alquezar-Planas, Yasuko Ishida, et al. 2018. "Degradation and Remobilization of Endogenous Retroviruses by Recombination during the Earliest Stages of a Germ-Line Invasion." *Proceedings of the National Academy of Sciences of the United States of America* 115 (34): 8609–14.

Long, Hannah K., Sara L. Prescott, and Joanna Wysocka. 2016. "Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution." *Cell* 167 (5): 1170–87.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.

Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 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): 595–605.

Lu, Chunye, Xavier Contreras, and B. Matija Peterlin. 2011. "P Bodies Inhibit Retrotransposition of Endogenous Intracisternal a Particles." *Journal of Virology* 85 (13): 6244–51.

Luger, K., A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. 1997. "Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution." *Nature* 389 (6648): 251–60.

Lynch, M. 2007. *Sunderland: Sinauer Associates*.

Magiorkinis, Gkikas, Robert J. Gifford, Aris Katzourakis, Joris De Ranter, and Robert Belshaw. 2012. "Env-Less Endogenous Retroviruses Are Genomic Superspreaders." *Proceedings of the National Academy of Sciences of the United States of America* 109 (19): 7385–90.

Maksakova, Irina A., Mark T. Romanish, Liane Gagnier, Catherine A. Dunn, Louie N. van de Lagemaat, and Dixie L. Mager. 2006. “Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line.” *PLoS Genetics* 2 (1): e2.

Malik, H. S., W. D. Burke, and T. H. Eickbush. 1999. “The Age and Evolution of Non-LTR Retrotransposable Elements.” *Molecular Biology and Evolution* 16 (6): 793–805.

Martianov, Igor, Mohamed-Amin Choukrallah, Arnaud Krebs, Tao Ye, Stephanie Legras, Erikjan Rijkers, Wilfred Van Ijcken, Bernard Jost, Paolo Sassone-Corsi, and Irwin Davidson. 2010. “Cell-Specific Occupancy of an Extended Repertoire of CREM and CREB Binding Loci in Male Germ Cells.” *BMC Genomics* 11 (1): 530.

Maston, Glenn A., Sara K. Evans, and Michael R. Green. 2006. “Transcriptional Regulatory Elements in the Human Genome.” *Annual Review of Genomics and Human Genetics* 7 (1): 29–59.

Matsumura, Yoshihiro, Ryo Nakaki, Takeshi Inagaki, Ayano Yoshida, Yuka Kano, Hiroshi Kimura, Toshiya Tanaka, et al. 2015. “H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation.” *Molecular Cell* 60 (4): 584–96.

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): 344–55.

Mcclintock, B. 1956. “Controlling Elements and the Gene.” *Cold Spring Harbor Symposia on Quantitative Biology* 21 (0): 197–216.

McLaren, A. 1984. “Meiosis and Differentiation of Mouse Germ Cells.” *Symposia of the Society for Experimental Biology* 38:7–23.

McLeay, Robert C., and Timothy L. Bailey. 2010. “Motif Enrichment Analysis: A Unified Framework and an Evaluation on ChIP Data.” *BMC Bioinformatics* 11 (1): 165.

Medstrand, P., and D. L. Mager. 1998. “Human-Specific Integrations of the HERV-K Endogenous Retrovirus Family.” *Journal of Virology* 72 (12): 9782–87.

Mello, Fabio N. de, Ana C. Tahira, Maria Gabriela Berzoti-Coelho, and Sergio Verjovski-Almeida. 2024. “The CUT&RUN Greenlist: Genomic Regions of Consistent Noise Are Effective Normalizing Factors for Quantitative Epigenome Mapping.” *Briefings in Bioinformatics* 25 (2). https://doi.org/10.1093/bib/bbad538.

Mendell, Joshua T., Neda A. Sharifi, Jennifer L. Meyers, Francisco Martinez-Murillo, and Harry C. Dietz. 2004. “Nonsense Surveillance Regulates Expression of Diverse Classes of Mammalian Transcripts and Mutes Genomic Noise.” *Nature Genetics* 36 (10): 1073–78.

Mendenhall, Eric M., Richard P. Koche, Thanh Truong, Vicky W. Zhou, Biju Issac, Andrew S. Chi, Manching Ku, and Bradley E. Bernstein. 2010. “GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells.” *PLoS Genetics* 6 (12): e1001244.

Mikkelsen, Tarjei S., Manching Ku, David B. Jaffe, Biju Issac, Erez Lieberman, Georgia Giannoukos, Pablo Alvarez, et al. 2007. “Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells.” *Nature* 448 (7153): 553–60.

Miller, T., N. J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J. F. Greenblatt, and A. Shilatifard. 2001. “COMPASS: A Complex of Proteins Associated with a Trithorax-Related SET Domain Protein.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (23): 12902–7.

Miró-Pina, Caridad, Olivia Charmant, Takayuki Kawaguchi, Daniel Holoch, Audrey Michaud, Isadora Cohen, Adeline Humbert, et al. 2022. “Paramecium Polycomb Repressive Complex 2 Physically

Interacts with the Small RNA-Binding PIWI Protein to Repress Transposable Elements." *Developmental Cell* 57 (8): 1037–52.e8.

Mita, Paolo, Xiaoji Sun, David Fenyö, David J. Kahler, Donghui Li, Neta Agmon, Aleksandra Wudzinska, et al. 2020. "BRCA1 and S Phase DNA Repair Pathways Restrict LINE-1 Retrotransposition in Human Cells." *Nature Structural & Molecular Biology* 27 (2): 179–91.

Mohandas, T., R. S. Sparkes, and L. J. Shapiro. 1981. "Reactivation of an Inactive Human X Chromosome: Evidence for X Inactivation by DNA Methylation." *Science (New York, N.Y.)* 211 (4480): 393–96.

Moissiard, Guillaume, Shawn J. Cokus, Joshua Cary, Suhua Feng, Allison C. Billi, Hume Stroud, Dylan Husmann, et al. 2012. "MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing." *Science (New York, N.Y.)* 336 (6087): 1448–51.

Molaro, Antoine, Ilaria Falciatori, Emily Hodges, Alexei A. Aravin, Krista Marran, Shahin Rafii, W. Richard McCombie, Andrew D. Smith, and Gregory J. Hannon. 2014. "Two Waves of de Novo Methylation during Mouse Germ Cell Development." *Genes & Development* 28 (14): 1544–49.

Monk, M., M. Boubelik, and S. Lehnert. 1987. "Temporal and Regional Changes in DNA Methylation in the Embryonic, Extraembryonic and Germ Cell Lineages during Mouse Embryo Development." *Development (Cambridge, England)* 99 (3): 371–82.

Montavon, Thomas, Nicholas Shukeir, Galina Erikson, Bettina Engist, Megumi Onishi-Seebacher, Devon Ryan, Yaarub Musa, et al. 2021. "Complete Loss of H3K9 Methylation Dissolves Mouse Heterochromatin Organization." *Nature Communications* 12 (1): 4359.

Müller, Iris, Ann Sophie Moroni, Daria Shlyueva, Sudeep Sahadevan, Erwin M. Schoof, Aliaksandra Radzisheuskaya, Jonas W. Højfeldt, et al. 2021. "MPP8 Is Essential for Sustaining Self-Renewal of Ground-State Pluripotent Stem Cells." *Nature Communications* 12 (1): 3034.

Myant, Kevin, Ausma Termanis, Arvind Y. M. Sundaram, Tristin Boe, Chao Li, Cara Merusi, Joe Burrage, Jose I. de Las Heras, and Irina Stancheva. 2011. "LSH and G9a/GLP Complex Are Required for Developmentally Programmed DNA Methylation." *Genome Research* 21 (1): 83–94.

Naville, M., I. A. Warren, Z. Haftek-Terreau, D. Chalopin, F. Brunet, P. Levin, D. Galiana, and J-N Volff. 2016. "Not so Bad after All: Retroviruses and Long Terminal Repeat Retrotransposons as a Source of New Genes in Vertebrates." *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 22 (4): 312–23.

Neri, Francesco, Stefania Rapelli, Anna Krepelova, Danny Incarnato, Caterina Parlato, Giulia Basile, Mara Maldotti, Francesca Anselmi, and Salvatore Oliviero. 2017. "Intragenic DNA Methylation Prevents Spurious Transcription Initiation." *Nature* 543 (7643): 72–77.

Ng, H. H., Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg, and A. Bird. 1999. "MBD2 Is a Transcriptional Repressor Belonging to the MeCP1 Histone Deacetylase Complex." *Nature Genetics* 23 (1): 58–61.

O'Donnell, Liza, and Moira K. O'Bryan. 2014. "Microtubules and Spermatogenesis." *Seminars in Cell & Developmental Biology* 30 (June):45–54.

Ogryzko, V. V., R. L. Schiltz, V. Russanova, B. H. Howard, and Y. Nakatani. 1996. "The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases." *Cell* 87 (5): 953–59.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. "DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for de Novo Methylation and Mammalian Development." *Cell* 99 (3): 247–57.

Ostertag, Eric M., John L. Goodier, Yue Zhang, and Haig H. Kazazian Jr. 2003. “SVA Elements Are Nonautonomous Retrotransposons That Cause Disease in Humans.” *The American Journal of Human Genetics* 73 (6): 1444–51.

Otsuka, Kai, Akihiko Sakashita, So Maezawa, Richard M. Schultz, and Satoshi H. Namekawa. 2023. “KRAB-Zinc-Finger Proteins Regulate Endogenous Retroviruses to Sculpt Germline Transcriptomes and Genome Evolution.” https://doi.org/10.1101/2023.06.24.546405.

Palmer, Nathan, S. Zakiah A. Talib, Chandrahas Koumar Ratnacaram, Diana Low, Xavier Bisteau, Joanna Hui Si Lee, Elisabeth Pfeiffenberger, et al. 2019. “CDK2 Regulates the NRF1/Ehmt1 Axis during Meiotic Prophase I.” *The Journal of Cell Biology* 218 (9): 2896–2918.

Patel, Lauren A., Yuwei Cao, Eric M. Mendenhall, Christopher Benner, and Alon Goren. 2024. “The Wild West of Spike-in Normalization.” *Nature Biotechnology* 42 (9): 1343–49.

Perron, H., E. Jouvin-Marche, M. Michel, A. Ounanian-Paraz, S. Camelo, A. Dumon, C. Jolivet-Reynaud, et al. 2001. “Multiple Sclerosis Retrovirus Particles and Recombinant Envelope Trigger an Abnormal Immune Response in Vitro, by Inducing Polyclonal Vbeta16 T-Lymphocyte Activation.” *Virology* 287 (2): 321–32.

Peters, A. H., D. O’Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schöfer, K. Weipoltshammer, et al. 2001. “Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability.” *Cell* 107 (3): 323–37.

Policarpi, Cristina, Marzia Munafò, Stylianos Tsagkris, Valentina Carlini, and Jamie A. Hackett. 2024. “Systematic Epigenome Editing Captures the Context-Dependent Instructive Function of Chromatin Modifications.” *Nature Genetics* 56 (6): 1168–80.

Ravasi, Timothy, Harukazu Suzuki, Carlo Vittorio Cannistraci, Shintaro Katayama, Vladimir B. Bajic, Kai Tan, Altuna Akalin, et al. 2010. “An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man.” *Cell* 140 (5): 744–52.

Rebollo, Rita, Mariana Galvão-Ferrarini, Liane Gagnier, Ying Zhang, Ardian Ferraj, Christine R. Beck, Matthew C. Lorincz, and Dixie L. Mager. 2020. “Inter-Strain Epigenomic Profiling Reveals a Candidate IAP Master Copy in C3H Mice.” *Viruses* 12 (7): 783.

Reuter, Michael, Shinichiro Chuma, Takashi Tanaka, Thomas Franz, Alexander Stark, and Ramesh S. Pillai. 2009. “Loss of the Mili-Interacting Tudor Domain-Containing Protein-1 Activates Transposons and Alters the Mili-Associated Small RNA Profile.” *Nature Structural & Molecular Biology* 16 (6): 639–46.

Rohlfs, E. M., N. Puget, M. L. Graham, B. L. Weber, J. E. Garber, C. Skrzynia, J. L. Halperin, G. M. Lenoir, L. M. Silverman, and S. Mazoyer. 2000. “An Alu-Mediated 7.1 Kb Deletion of BRCA1 Exons 8 and 9 in Breast and Ovarian Cancer Families That Results in Alternative Splicing of Exon 10.” *Genes, Chromosomes & Cancer* 28 (3): 300–307.

Rooij, D. G. de, and L. D. Russell. 2000. “All You Wanted to Know about Spermatogonia but Were Afraid to Ask.” *Journal of Andrology* 21 (6): 776–98.

Rowe, Helen M., Johan Jakobsson, Daniel Mesnard, Jacques Rougemont, Séverine Reynard, Tugce Aktas, Pierre V. Maillard, et al. 2010. “KAP1 Controls Endogenous Retroviruses in Embryonic Stem Cells.” *Nature* 463 (7278): 237–40.

Roy-Engel, Astrid M. 2012. “A Tale of an A-Tail: The Lifeline of a SINE.” *Mobile Genetic Elements* 2 (6): 282–86.

Sachs, Michael, Courtney Onodera, Kathryn Blaschke, Kevin T. Ebata, Jun S. Song, and Miguel Ramalho-Santos. 2013. “Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline in Vivo.” *Cell Reports* 3 (6): 1777–84.

Sadate-Ngatchou, Patricia I., Christopher J. Payne, Andrea T. Dearth, and Robert E. Braun. 2008. “Cre Recombinase Activity Specific to Postnatal, Premeiotic Male Germ Cells in Transgenic Mice.” *Genesis (New York, N.Y.: 2000)* 46 (12): 738–42.

Saha, Karabi, Grace I. Nielsen, Raj Nandani, Yizi Zhang, Lingqi Kong, Ping Ye, and Wenfeng An. 2024. “YY1 Is a Transcriptional Activator of the Mouse LINE-1 Tf Subfamily.” *Nucleic Acids Research* 52 (21): 12878–94.

Sakashita, Akihiko, So Maezawa, Kazuki Takahashi, Kris G. Alavattam, Masashi Yukawa, Yueh-Chiang Hu, Shohei Kojima, et al. 2020. “Endogenous Retroviruses Drive Species-Specific Germline Transcriptomes in Mammals.” *Nature Structural & Molecular Biology* 27 (10): 967–77.

Sandmeyer, Suzanne B., and Kristina A. Clemens. 2010. “Function of a Retrotransposon Nucleocapsid Protein.” *RNA Biology* 7 (6): 642–54.

Sanford, J. P., H. J. Clark, V. M. Chapman, and J. Rossant. 1987. “Differences in DNA Methylation during Oogenesis and Spermatogenesis and Their Persistence during Early Embryogenesis in the Mouse.” *Genes & Development* 1 (10): 1039–46.

Scarpulla, Richard C. 2002. “Nuclear Activators and Coactivators in Mammalian Mitochondrial Biogenesis.” *Biochimica et Biophysica Acta - Gene Structure and Expression* 1576 (1-2): 1–14.

Schmidl, Christian, André F. Rendeiro, Nathan C. Sheffield, and Christoph Bock. 2015. “ChIPmentation: Fast, Robust, Low-Input ChIP-Seq for Histones and Transcription Factors.” *Nature Methods* 12 (10): 963–65.

Schmidt, Dominic, Petra C. Schwalie, Michael D. Wilson, Benoit Ballester, Ângela Gonçalves, Claudia Kutter, Gordon D. Brown, Aileen Marshall, Paul Flicek, and Duncan T. Odom. 2012. “Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages.” *Cell* 148 (1-2): 335–48.

Schmitges, Frank W., Archana B. Prusty, Mahamadou Faty, Alexandra Stützer, Gondichatnahalli M. Lingaraju, Jonathan Aiwazian, Ragna Sack, et al. 2011. “Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks.” *Molecular Cell* 42 (3): 330–41.

Schöpp, Theresa, Ansgar Zoch, Rebecca V. Berrens, Tania Auchynnikava, Yuka Kabayama, Lina Vasiliauskaitė, Juri Rappsilber, Robin C. Allshire, and Dónal O’Carroll. 2020. “TEX15 Is an Essential Executor of MIWI2-Directed Transposon DNA Methylation and Silencing.” *Nature Communications* 11 (1): 3739.

Schultz, David C., Kasirajan Ayyanathan, Dmitri Negorev, Gerd G. Maul, and Frank J. Rauscher 3rd. 2002. “SETDB1: A Novel KAP-1-Associated Histone H3, Lysine 9-Specific Methyltransferase That Contributes to HP1-Mediated Silencing of Euchromatic Genes by KRAB Zinc-Finger Proteins.” *Genes & Development* 16 (8): 919–32.

Schulz, Wolfgang A. 2006. “L1 Retrotransposons in Human Cancers.” *Journal of Biomedicine & Biotechnology* 2006 (1): 83672.

Seczynska, Marta, Stuart Bloor, Sergio Martinez Cuesta, and Paul J. Lehner. 2022. “Genome Surveillance by HUSH-Mediated Silencing of Intronless Mobile Elements.” *Nature* 601 (7893): 440–45.

Shapiro, J. A. 1969. “Mutations Caused by the Insertion of Genetic Material into the Galactose Operon of Escherichia Coli.” *Journal of Molecular Biology* 40 (1): 93–105.

Shaver, Scott, J. Armando Casas-Mollano, Ronald L. Cerny, and Heriberto Cerutti. 2010. “Origin of the Polycomb Repressive Complex 2 and Gene Silencing by an E(z) Homolog in the Unicellular Alga Chlamydomonas.” *Epigenetics: Official Journal of the DNA Methylation Society* 5 (4): 301–12.

Sheldon, Jyoti, Han Lu, Li Wang, and Xiao Jun Ma. 2020. “Incorporation of Spatial Mapping and Confirmation of Genes by Single Cell in Situ Hybridization Technology into Single Cell RNA Sequencing Workflows.” *Journal of Biomolecular Techniques: JBT* 31 (Suppl): S13.

Shoji, Masanobu, Takashi Tanaka, Mihoko Hosokawa, Michael Reuter, Alexander Stark, Yuzuru Kato, Gen Kondoh, et al. 2009. “The TDRD9-MIWI2 Complex Is Essential for piRNA-Mediated Retrotransposon Silencing in the Mouse Male Germline.” *Developmental Cell* 17 (6): 775–87.

Sienski, Grzegorz, Derya Dönertas, and Julius Brennecke. 2012. “Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression.” *Cell* 151 (5): 964–80.

Sims, Robert J., 3rd, Chi-Fu Chen, Helena Santos-Rosa, Tony Kouzarides, Smita S. Patel, and Danny Reinberg. 2005. “Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains.” *The Journal of Biological Chemistry* 280 (51): 41789–92.

Slotkin, R. Keith, and Robert Martienssen. 2007. “Transposable Elements and the Epigenetic Regulation of the Genome.” *Nature Reviews. Genetics* 8 (4): 272–85.

Smallwood, Sébastien A., Shin-Ichi Tomizawa, Felix Krueger, Nico Ruf, Natasha Carli, Anne Segonds-Pichon, Shun Sato, Kenichiro Hata, Simon R. Andrews, and Gavin Kelsey. 2011. “Dynamic CpG Island Methylation Landscape in Oocytes and Preimplantation Embryos.” *Nature Genetics* 43 (8): 811–14.

Soehn, Anne S., Thu-Trang Pham, Karin Schaeferhoff, Thomas Floss, Daniela M. Vogt Weisenhorn, Wolfgang Wurst, Michael Bonin, and Olaf Riess. 2009. “Periphilin Is Strongly Expressed in the Murine Nervous System and Is Indispensable for Murine Development.” *Genesis (New York, N.Y.: 2000)* 47 (10): 697–707.

Solomon, M. J., P. L. Larsen, and A. Varshavsky. 1988. “Mapping Protein-DNA Interactions in Vivo with Formaldehyde: Evidence That Histone H4 Is Retained on a Highly Transcribed Gene.” *Cell* 53 (6): 937–47.

Solyom, Szilvia, and Haig H. Kazazian Jr. 2012. “Mobile Elements in the Human Genome: Implications for Disease.” *Genome Medicine* 4 (2): 12.

Soper, Sarah F. C., Godfried W. van der Heijden, Tara C. Hardiman, Mary Goodheart, Sandra L. Martin, Peter de Boer, and Alex Bortvin. 2008. “Mouse Maelstrom, a Component of Nuage, Is Essential for Spermatogenesis and Transposon Repression in Meiosis.” *Developmental Cell* 15 (2): 285–97.

Soufi, Abdenour, Meilin Fernandez Garcia, Artur Jaroszewicz, Nebiyu Osman, Matteo Pellegrini, and Kenneth S. Zaret. 2015. “Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming.” *Cell* 161 (3): 555–68.

Spruijt, Cornelia G., Felix Gnerlich, Arne H. Smits, Toni Pfaffeneder, Pascal W. T. C. Jansen, Christina Bauer, Martin Münzel, et al. 2013. “Dynamic Readers for 5-(hydroxy)methylcytosine and Its Oxidized Derivatives.” *Cell* 152 (5): 1146–59.

Stadler, Michael B., Rabih Murr, Lukas Burger, Robert Ivanek, Florian Lienert, Anne Schöler, Erik van Nimwegen, et al. 2011a. “DNA-Binding Factors Shape the Mouse Methylome at Distal Regulatory Regions.” *Nature* 480 (7378): 490–95.

———. 2011b. “DNA-Binding Factors Shape the Mouse Methylome at Distal Regulatory Regions.” *Nature* 480 (7378): 490–95.

Stocking, C., and C. A. Kozak. 2008. “Murine Endogenous Retroviruses.” *Cellular and Molecular Life Sciences: CMLS* 65 (21): 3383–98.

Strepkos, Dimitrios, Mariam Markouli, Alexia Klonou, Athanasios G. Papavassiliou, and Christina Piperi. 2021. “Histone Methyltransferase SETDB1: A Common Denominator of Tumorigenesis with Therapeutic Potential.” *Cancer Research* 81 (3): 525–34.

Suetake, Isao, Fuminori Shinozaki, Junichi Miyagawa, Hideyuki Takeshima, and Shoji Tajima. 2004. “DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction.” *The Journal of Biological Chemistry* 279 (26): 27816–23.

Sultana, Tania, Dominic van Essen, Oliver Siol, Marc Bailly-Bechet, Claude Philippe, Amal Zine El Aabidine, Léo Pioger, et al. 2019. “The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-Insertion Sequence Biases and Post-Insertion Selection.” *Molecular Cell* 74 (3): 555–70.e7.

Sundaram, Vasavi, Yong Cheng, Zhihai Ma, Daofeng Li, Xiaoyun Xing, Peter Edge, Michael P. Snyder, and Ting Wang. 2014. “Widespread Contribution of Transposable Elements to the Innovation of Gene Regulatory Networks.” *Genome Research* 24 (12): 1963–76.

Sun, Yu H., Brent Lee, and Xin Zhiguo Li. 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): 293–311.

Tachibana, Makoto, Kenji Sugimoto, Masami Nozaki, Jun Ueda, Tsutomu Ohta, Misao Ohki, Mikiko Fukuda, et al. 2002. “G9a Histone Methyltransferase Plays a Dominant Role in Euchromatic Histone H3 Lysine 9 Methylation and Is Essential for Early Embryogenesis.” *Genes & Development* 16 (14): 1779–91.

Tachibana, Makoto, Jun Ueda, Mikiko Fukuda, Naoki Takeda, Tsutomu Ohta, Hiroko Iwanari, Toshiko Sakihama, Tatsuhiko Kodama, Takao Hamakubo, and Yoichi Shinkai. 2005. “Histone Methyltransferases G9a and GLP Form Heteromeric Complexes and Are Both Crucial for Methylation of Euchromatin at H3-K9.” *Genes & Development* 19 (7): 815–26.

Tam, P. P., and M. H. Snow. 1981. “Proliferation and Migration of Primordial Germ Cells during Compensatory Growth in Mouse Embryos.” *Journal of Embryology and Experimental Morphology* 64 (August):133–47.

Tan, Kun, and Miles F. Wilkinson. 2023. “Developmental Regulators Moonlighting as Transposons Defense Factors.” *Andrology* 11 (5): 891–903.

Tate, P. H., and A. P. Bird. 1993. “Effects of DNA Methylation on DNA-Binding Proteins and Gene Expression.” *Current Opinion in Genetics & Development* 3 (2): 226–31.

Tchasovnikarova, Iva A., Richard T. Timms, Nicholas J. Matheson, Kim Wals, Robin Antrobus, Berthold Göttgens, Gordon Dougan, Mark A. Dawson, and Paul J. Lehner. 2015. “GENE SILENCING. Epigenetic Silencing by the HUSH Complex Mediates Position-Effect Variegation in Human Cells.” *Science (New York, N.Y.)* 348 (6242): 1481–85.

Teissandier, Aurélie, Nicolas Servant, Emmanuel Barillot, and Deborah Bourc’his. 2019. “Tools and Best Practices for Retrotransposon Analysis Using High-Throughput Sequencing Data.” *Mobile DNA* 10 (1): 52.

“The Mechanism of Mendelian Heredity.” 1916. *Nature* 97 (2423): 117–18.

Tootle, Tina L., and Ilaria Rebay. 2005. “Post-Translational Modifications Influence Transcription Factor Activity: A View from the ETS Superfamily.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 27 (3): 285–98.

Turelli, Priscilla, Nathaly Castro-Diaz, Flavia Marzetta, Adamandia Kapopoulou, Charlène Raclot, Julien Duc, Vannary Tieng, Simon Quenneville, and Didier Trono. 2014. “Interplay of TRIM28 and DNA Methylation in Controlling Human Endogenous Retroelements.” *Genome Research* 24 (8): 1260–70.

Ueda, Jun, Makoto Tachibana, Tsuyoshi Ikura, and Yoichi Shinkai. 2006. “Zinc Finger Protein Wiz Links G9a/GLP Histone Methyltransferases to the Co-Repressor Molecule CtBP.” *The Journal of Biological Chemistry* 281 (29): 20120–28.

Vasiliauskaitė, Lina, Rebecca V. Berrens, Ivayla Ivanova, Claudia Carrieri, Wolf Reik, Anton J. Enright, and Dónal O’Carroll. 2018. “Defective Germline Reprogramming Rewires the Spermatogonial Transcriptome.” *Nature Structural & Molecular Biology* 25 (5): 394–404.

Vermeulen, Michiel, Klaas W. Mulder, Sergei Denissov, W. W. M. Pim Pijnappel, Frederik M. A. van Schaik, Radhika A. Varier, Marijke P. A. Baltissen, Henk G. Stunnenberg, Matthias Mann, and H. Th Marc Timmers. 2007. “Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4.” *Cell* 131 (1): 58–69.

Voigt, Philipp, Gary LeRoy, William J. Drury 3rd, Barry M. Zee, Jinsook Son, David B. Beck, Nicolas L. Young, Benjamin A. Garcia, and Danny Reinberg. 2012. “Asymmetrically Modified Nucleosomes.” *Cell* 151 (1): 181–93.

Wallau, Gabriel Luz, Cristina Vieira, and Élgion Lúcio Silva Loreto. 2018. “Genetic Exchange in Eukaryotes through Horizontal Transfer: Connected by the Mobilome.” *Mobile DNA* 9 (1). https://doi.org/10.1186/s13100-018-0112-9.

Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. “Transcription of IAP Endogenous Retroviruses Is Constrained by Cytosine Methylation.” *Nature Genetics* 20 (2): 116–17.

Walsh, C. P., and G. L. Xu. 2006. “Cytosine Methylation and DNA Repair.” In *Current Topics in Microbiology and Immunology*, 283–315. Berlin/Heidelberg: Springer-Verlag.

Walter, Marius, Aurélie Teissandier, Raquel Pérez-Palacios, and Déborah Bourc’his. 2016. “An Epigenetic Switch Ensures Transposon Repression upon Dynamic Loss of DNA Methylation in Embryonic Stem Cells.” *eLife* 5 (January):e11418.

Wang, Junpeng, Chao Tang, Qian Wang, Jun Su, Ting Ni, Wenjing Yang, Yongsheng Wang, et al. 2017. “NRF1 Coordinates with DNA Methylation to Regulate Spermatogenesis.” *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 31 (11): 4959–70.

Wang, Liangjun, J. Lesley Brown, Ru Cao, Yi Zhang, Judith A. Kassis, and Richard S. Jones. 2004. “Hierarchical Recruitment of Polycomb Group Silencing Complexes.” *Molecular Cell* 14 (5): 637–46.

Wang, Lu, Patrick A. Ozark, Edwin R. Smith, Zibo Zhao, Stacy A. Marshall, Emily J. Rendleman, Andrea Piunti, et al. 2018. “TET2 Coactivates Gene Expression through Demethylation of Enhancers.” *Science Advances* 4 (11): eaau6986.

Wang, Lu, Jun Zhang, Jialei Duan, Xinxing Gao, Wei Zhu, Xingyu Lu, Lu Yang, et al. 2014. “Programming and Inheritance of Parental DNA Methylomes in Mammals.” *Cell* 157 (4): 979–91.

Wang, Pengxiang, Jun Su, Junpeng Wang, Yilin Xie, Wei Chen, Jinhai Zhong, and Yuan Wang. 2024. “NRF1 Promotes Primordial Germ Cell Development, Proliferation and Survival.” *Cell Proliferation* 57 (1): e13533.

Waterland, Robert A., and Randy L. Jirtle. 2003. “Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation.” *Molecular and Cellular Biology* 23 (15): 5293–5300.

Wegrzyn, Jill L., Brian Y. Lin, Jacob J. Zieve, William M. Dougherty, Pedro J. Martínez-García, Maxim Koriabine, Ann Holtz-Morris, et al. 2013. “Insights into the Loblolly Pine Genome: Characterization of BAC and Fosmid Sequences.” *PloS One* 8 (9): e72439.

Wells, Jonathan N., and Cédric Feschotte. 2020. “A Field Guide to Eukaryotic Transposable Elements.” *Annual Review of Genetics* 54 (1): 539–61.

Whittle, Carrie A., and Cassandra G. Extavour. 2017. “Causes and Evolutionary Consequences of Primordial Germ-Cell Specification Mode in Metazoans.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (23): 5784–91.

Wicker, Thomas, François Sabot, Aurélie Hua-Van, Jeffrey L. Bennetzen, Pierre Capy, Boulos Chalhoub, Andrew Flavell, et al. 2007. “A Unified Classification System for Eukaryotic Transposable Elements.” *Nature Reviews. Genetics* 8 (12): 973–82.

Wolff, Erika M., Hyang-Min Byun, Han F. Han, Shikhar Sharma, Peter W. Nichols, Kimberly D. Siegmund, Allen S. Yang, Peter A. Jones, and Gangning Liang. 2010. “Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer.” *PLoS Genetics* 6 (4): e1000917.

Wooten, Matthew, Brittany Takushi, Kami Ahmad, and Steven Henikoff. 2023. “Aclarubicin Stimulates RNA Polymerase II Elongation at Closely Spaced Divergent Promoters.” *Science Advances* 9 (24): eadg3257.

Wossidlo, Mark, Toshinobu Nakamura, Konstantin Lepikhov, C. Joana Marques, Valeri Zakhartchenko, Michele Boiani, Julia Arand, Toru Nakano, Wolf Reik, and Jörn Walter. 2011. “5-Hydroxymethylcytosine in the Mammalian Zygote Is Linked with Epigenetic Reprogramming.” *Nature Communications* 2 (1): 241.

Wyatt, G. R. 1950. “Occurrence of 5-Methylcytosine in Nucleic Acids.” *Nature* 166 (4214): 237–38.

Wysocka, Joanna, Tomek Swigut, Hua Xiao, Thomas A. Milne, So Yeon Kwon, Joe Landry, Monika Kauer, et al. 2006. “A PHD Finger of NURF Couples Histone H3 Lysine 4 Trimethylation with Chromatin Remodelling.” *Nature* 442 (7098): 86–90.

Yamaguchi, Shinpei, Kwonho Hong, Rui Liu, Azusa Inoue, Li Shen, Kun Zhang, and Yi Zhang. 2013. “Dynamics of 5-Methylcytosine and 5-Hydroxymethylcytosine during Germ Cell Reprogramming.” *Cell Research* 23 (3): 329–39.

Yamaguchi, Shinpei, Kwonho Hong, Rui Liu, Li Shen, Azusa Inoue, Dinh Diep, Kun Zhang, and Yi Zhang. 2012. “Tet1 Controls Meiosis by Regulating Meiotic Gene Expression.” *Nature* 492 (7429): 443–47.

Yamaguchi, Shinpei, Li Shen, Yuting Liu, Damian Sendler, and Yi Zhang. 2013. “Role of Tet1 in Erasure of Genomic Imprinting.” *Nature* 504 (7480): 460–64.

Yang, Fang, Yemin Lan, Radha Raman Pandey, David Homolka, Shelley L. Berger, Ramesh S. Pillai, Marisa S. Bartolomei, and P. Jeremy Wang. 2020. “TEX15 Associates with MILI and Silences Transposable Elements in Male Germ Cells.” *Genes & Development* 34 (11-12): 745–50.

Yang, Na, Weixiang Wang, Yan Wang, Mingzhu Wang, Qiang Zhao, Zihe Rao, Bing Zhu, and Rui-Ming Xu. 2012. “Distinct Mode of Methylated Lysine-4 of Histone H3 Recognition by Tandem Tudor-like Domains of Spindlin1.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (44): 17954–59.

Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. “Cytosine Methylation and the Ecology of Intragenomic Parasites.” *Trends in Genetics: TIG* 13 (8): 335–40.

Yokoyama, Akihiko, Zhong Wang, Joanna Wysocka, Mrinmoy Sanyal, Deborah J. Aufiero, Issay Kitabayashi, Winship Herr, and Michael L. Cleary. 2004. “Leukemia Proto-Oncoprotein MLL Forms a SET1-like Histone Methyltransferase Complex with Menin to Regulate Hox Gene Expression.” *Molecular and Cellular Biology* 24 (13): 5639–49.

Yoshida, Shosei, Mamiko Sukeno, Toshinori Nakagawa, Kazuyuki Ohbo, Go Nagamatsu, Toshio Suda, and Yo-Ichi Nabeshima. 2006. “The First Round of Mouse Spermatogenesis Is a Distinctive Program That Lacks the Self-Renewing Spermatogonia Stage.” *Development (Cambridge, England)* 133 (8): 1495–1505.

Zamudio, Natasha, Joan Barau, Aurélie Teissandier, Marius Walter, Maté Borsos, Nicolas Servant, and Déborah Bourc’his. 2015. “DNA Methylation Restrains Transposons from Adopting a Chromatin Signature Permissive for Meiotic Recombination.” *Genes & Development* 29 (12): 1256–70.

Zemach, Assaf, Ivy E. McDaniel, Pedro Silva, and Daniel Zilberman. 2010. “Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation.” *Science (New York, N.Y.)* 328 (5980): 916–19.

Zhang, Hongdao, Fengjuan Zhang, Qinghua Chen, Mingzhe Li, Xiaolong Lv, Yali Xiao, Zhaozhen Zhang, et al. 2021. “The piRNA Pathway Is Essential for Generating Functional Oocytes in Golden Hamsters.” *Nature Cell Biology* 23 (9): 1013–22.

Zhu, Tao, Keyan Liao, Rongfang Zhou, Chunjiao Xia, and Weibo Xie. 2020. “ATAC-Seq with Unique Molecular Identifiers Improves Quantification and Footprinting.” *Communications Biology* 3 (1): 675.

Zickler, Denise, and Nancy Kleckner. 2015. “Recombination, Pairing, and Synapsis of Homologs during Meiosis.” *Cold Spring Harbor Perspectives in Biology* 7 (6): a016626.

Zoch, Ansgar, Tania Auchynnikava, Rebecca V. Berrens, Yuka Kabayama, Theresa Schöpp, Madeleine Heep, Lina Vasiliauskaitė, et al. 2020. “SPOCD1 Is an Essential Executor of piRNA-Directed de Novo DNA Methylation.” *Nature* 584 (7822): 635–39.

Zoch, Ansgar, Gabriela Konieczny, Tania Auchynnikava, Birgit Stallmeyer, Nadja Rotte, Madeleine Heep, Rebecca V. Berrens, et al. 2024. “C19ORF84 Connects piRNA and DNA Methylation Machineries to Defend the Mammalian Germ Line.” *Molecular Cell* 84 (6): 1021–35.e11.