Critical and different roles of DNA methylation in male germ cell development

Mathilde Dura

Sorbonne Université
Ecole doctorale Complexité du Vivant
ED515

Institut Curie
Génétique et Biologie du Développement
CNRS U934_INSERM U3215

CRITICAL AND DIFFERENT ROLES OF DNA METHYLATION IN MALE GERM CELL DEVELOPMENT

Thèse de Doctorat de Biologie de **SORBONNE UNIVERSITE**

présentée par
Mathilde DURA

et dirigée par **Deborah BOURC'HIS**

Présentée et soutenue publiquement le **22 septembre 2021**

Devant un jury composé de :

Pr. **Petra Hajkova**	*Rapportrice*
Pr. **Dónal O'Carroll**	*Rapporteur*
Dr. **Michael Weber**	*Examinateur*
Dr. **Céline Vallot**	*Examinatrice*
Dr. **Pierre Fouchet**	*Examinateur*
Dr**. Deborah Bourc'his**	*Directrice de thèse*

Institut Curie
Génétique et Biologie du Développement
CNRS U934_INSERM U3215

Sorbonne Université
Ecole doctorale Complexité du Vivant
ED515

CRITICAL AND DIFFERENT ROLES OF DNA METHYLATION IN MALE GERM CELL DEVELOPMENT

Thèse de Doctorat de Biologie de **SORBONNE UNIVERSITE**

présentée par
Mathilde DURA

et dirigée par **Deborah BOURC'HIS**

Présentée et soutenue publiquement le **22 septembre 2021**

Devant un jury composé de :

Pr. **Petra Hajkova**	*Rapportrice*
Pr. **Dónal O'Carroll**	*Rapporteur*
Dr. **Michael Weber**	*Examinateur*
Dr. **Céline Vallot**	*Examinatrice*
Dr. **Pierre Fouchet**	*Examinateur*
Dr**. Deborah Bourc'his**	*Directrice de thèse*

ACKNOWLEDGMENTS

I am grateful for the members of my jury who have agreed to evaluate my work: Pr. Petra Hajkova and Pr. Donal O'Carroll, *rapporteurs*; Dr. Michael Weber, Dr. Céline Vallot, and Dr. Pierre Fouchet *examinateurs*. Being evaluated by such inspiring scientists is a rare privilege and I thank them in advance for their time and their help.

I joined the lab of Deborah Bourc'his almost five years ago. When I arrived, I only knew that I wanted to study developmental processes. Déborah supported me to work alongside an amazing post-doc, Joan Barau, which is how I started to learn the developmental process of spermatogenesis with all its epigenetic implications. It was the beginning of an amazing journey! I want to thank Déborah for all the freedom and trust that she gave me. I want to thank her for always being supportive of my ideas and of the experiments that I wanted to perform (no matter the price..!). Also, after five years in this lab, I feel very grateful to have had the chance to evolve in such a stimulating environment. More particularly, Déborah always pushed me to present my results in international conferences or to speak with invited speakers in the institute and discuss my results with them – and I realize now how these moments were a true privilege, and how important they were to help me consider myself a real scientist!

As I mentioned, when I arrived in the lab, I started working with Joan Barau. I couldn't be more grateful to have had such an amazing mentor. Joan taught me so much: from the different waves of spermatogenesis, to having an extremely well-organized freezer full of colorful boxes, and most importantly, to always believe my data and that my experiments were at least as important as anyone else's. Joan left the lab two years ago to become a PI at IMB, but still, he was always there for me to discuss science, give precious advice and ideas, or to help me when I was texting "911" to save an experiment or to answer a chemical question. Joan, thank you so much.

Then, I want to thank Aurélie Teissandier, our bioinformatician, the "real" boss of the lab! I learned so much working in close collaboration with her. I want to thank her for being so patient with me, for being always positive and for cheering me up when I had the feeling that we were drowning in the quantity of data and infinite possibilities of analysis.

Finally, I want to thank Mélanie Armand, a master student that I mentored and is now an engineer in our lab. I was able to finish my paper so much faster thanks to her precious help with the microscopy experiments and analysis. Thank you, Mélanie, for your help.

More generally, I would like to thank all the members of the Bourc'his lab, past and present. Juliane for being my model as a PhD student; Raquel for her inconditional kindness and her

brilliant mind; Tomek for all the laughs, screams, true friendship and scientific advices; Max for his help and support and for being such an inspiration to be a focused and efficient scientist; Mattia for his joy and the way he enjoys life; Julian for his help with all the informatic stuff that I don't understand, Fatima for being such a kind and nice person; Giorgia for her help with the doctoral school; Charlotte P for all the nice discussion and for her encouragements; Elena for giving a really nice fresh air to the lab and for all the fun and laugh, and a special thank you to Charlotte D who was really supportive during the writing of my PhD manuscript and who helped me a lot by editing my English and for all of her many advices.

Then, I want to thank our little team of PhD students: Mathieu, Emeline and Lorraine. You guys gave me lot a strength, and support me in so many ways. I loved watching you evolve and grow in the lab. I felt so lucky to have you around me. Mathieu, thank you for your support and advice especially during the writing of this manuscript, thank you for being such a safe person. Emeline, thank you for being both calm and crazy at the same time, thank you most importantly for being such a balanced person and for always being able to give me some perspective. Finally, Lorraine, thank you for everything, for your daily support, for the laughs, for all the scientific projects that we have in mind and we loved to discuss, thank you for being such a good listener and an amazing friend.

Last but not least, I want to thanks my baby sister Jeanne, who helped me so much with the writing and page layout of this manuscript. I want to thank Boutayna, my girlfriend for being so understanding during the stressful period of the writing. I want to thank my two amazing parents for their unconditional love and support. And finally, I want to thank my grand-ma. She didn't do anything for the thesis, but I am pretty sure that she thinks she did! Anyways, she would tell you that she built me the way I am and without her existing on planet Earth, I would not be submitting this thesis right now, and that it is definitely enough to be thanked!

ABSTRACT / RÉSUMÉ

ABSTRACT

DNA methylation, in association with stable gene or transposable element (TE) repression, plays a key role in spermatogenesis. During germline development, the methylome of the future gametes is extensively reprogrammed: somatic DNA methylation patterns are first erased and germ cell-specific patterns are then established *de novo*. Three *de novo* DNA methyltransferases (DNMTs) are essential for shaping male germ cell DNA methylation in mice: the DNMT3C and DNMT3A enzymes and the DNMT3L co-factor. Mutation in any of these genes leads to male sterility. DNMT3C was recently shown to selectively methylate evolutionarily young TEs. However, the precise targets and developmental function of DNMT3A was still unknown. During my PhD, I investigated the interplay between DNMT3A and DNMT3C in the epigenetic regulation of male germline development. First (***project 1***), I reported a striking division of labor between these enzymes: while DNMT3C prevents TEs from interfering with meiosis, DNMT3A broadly methylates the genome—except DNMT3C-dependent TEs—and controls spermatogonial stem cell (SSC) plasticity. By reconstructing developmental trajectories through single-cell RNA-seq and by profiling chromatin states, I found that *Dnmt3A* mutant SSCs can only self-renew and no longer differentiate due to spurious enhancer activation that enforces an irreversible stem cell gene program. I therefore demonstrated a novel function for DNA methylation in male fertility: the epigenetic programming of SSC commitment to differentiation and to life-long spermatogenesis supply. Second (***project 2***), I investigated the chromatin determinants of DNMT3C specificity towards young TEs. I found that these sequences present unique chromatin dynamics: first a bivalent H3K4me3-H3K9me3 enrichment, followed by a switch to H3K9me3-only. Consistently, H3K9me3-enrichment was also a hallmark of the sequences that undergo DNA methylation upon ectopic DNMT3C expression in cultured embryonic stem cells. As a whole, my work provided novel insights into the complexity of DNA methylation-based control of reproduction.

RÉSUMÉ

La méthylation de l'ADN, associée à la répression stable des gènes et des éléments transposables (ET), joue un rôle essentiel dans la spermatogenèse. Au cours du développement de la lignée germinale, le méthylome des futurs gamètes est extensivement reprogrammé : après effacement des profils de méthylation somatiques, des profils spécifiques des cellules germinales sont établis *de novo*. Trois *de novo* ADN méthyltransférases (DNMT) sont essentielles à l'acquisition de la méthylation de l'ADN des cellules germinales mâles chez la souris : les enzymes DNMT3C et DNMT3A et leur cofacteur DNMT3L. Toute mutation dans l'un de ces gènes entraîne une stérilité mâle. Il a été récemment démontré que DNMT3C est l'enzyme qui méthyle sélectivement les ET les plus jeunes évolutivement. Cependant, les cibles et la fonction développementale de DNMT3A étaient encore inconnues. Au cours de ma thèse, je me suis intéressée aux rôles respectifs et complémentaires de DNMT3A et DNMT3C dans la régulation épigénétique du développement germinal mâle. J'ai d'abord démontré (***projet 1***) une division de travail remarquable entre ces enzymes: alors que DNMT3C empêche les ET d'interférer avec la méiose, DNMT3A méthyle largement le génome -à l'exception des ET dépendants de DNMT3C- et contrôle la plasticité des cellules souches spermatogoniales (CSS). Par une reconstruction de trajectoire développementale par scRNA-seq, j'ai découvert que les CSS mutantes pour *Dnmt3A* ne peuvent que se renouveler à l'identique, ayant perdu leur potentiel de différentiation. Par une analyse des profils chromatiniens, j'ai relié ce phénotype à l'activation erronée d'enhancers qui imposent un programme génétique irréversible de cellules souches. Ce travail révèle une nouvelle fonction de la méthylation de l'ADN dans la fertilité mâle : la programmation épigénétique de la capacité des CSS à se différencier et à alimenter la spermatogenèse tout au long de la vie. En parallèle (***projet 2***), j'ai étudié la nature chromatinienne de la spécificité de reconnaissance des jeunes ET par DNMT3C. J'ai trouvé que ces séquences présentent une dynamique chromatinienne unique: d'abord un profil bivalent de type H3K4me3-H3K9me3 qui évolue vers un enrichissement H3K9me3 exclusif. Mon travail a ainsi fourni des éléments originaux et nouveaux pour comprendre le rôle complexe de la méthylation de l'ADN en reproduction.

RESUME EN FRANÇAIS

La méthylation de l'ADN joue un rôle essentiel dans la spermatogenèse, en association avec la répression stable des gènes et des éléments transposables (ET). Au cours du développement de la lignée germinale, le méthylome des futurs gamètes est extensivement reprogrammé : après effacement des profils de méthylation somatiques, des profils spécifiques des cellules germinales sont établis *de novo*. Trois *de novo* ADN méthyltransférases (DNMT) sont essentielles à l'acquisition de la méthylation de l'ADN des cellules germinales mâles chez la souris : les enzymes DNMT3C et DNMT3A et leur cofacteur DNMT3L. Toute mutation dans l'un de ces gènes entraîne une stérilité mâle. Il a été récemment démontré que DNMT3C est l'enzyme qui méthyle sélectivement les ET les plus jeunes évolutivement, c'est-à-dire ceux qui sont exprimés ou actifs. Cependant, les cibles et la fonction développementale de DNMT3A étaient encore inconnues.

Au cours de ma thèse, je me suis intéressée aux rôles respectifs et complémentaires de DNMT3A et DNMT3C dans la régulation épigénétique du développement germinal mâle. Mes travaux de thèse ont permis d'apporter des résultats nouveaux et importants dans le domaine de la reprogrammation épigénétique de la lignée germinale mâle, en adressant deux points majeurs : *1)* la fonction et les cibles génomiques de DNMT3A et *2)* la signature chromatinienne des cibles génomiques de DNMT3C.

La fonction et les cibles génomiques de DNMT3A (projet 1) :

Par une caractérisation des profils génomiques de méthylation par Whole Genome Bisulfite Sequencing (WGBS), j'ai pu démontrer que les enzymes DNMT3A et DNMT3C ont une activité complémentaire et non redondante dans la méthylation du génome des cellules germinales mâles : alors que DNMT3C méthyle les promoteurs des éléments transposables (ET) actifs, DNMT3A méthyle tout le reste du génome. De plus, alors que DNMT3C empêche les ET d'interférer avec la méiose, j'ai découvert que DNMT3A ne joue aucun rôle critique pour la méiose, mais s'avère essentielle à la différentiation et à la plasticité des cellules souches spermatogoniales (CSS). Plus précisément, en absence de *Dnmt3A*, les CSS s'accumulent dans les testicules tout au long de la vie du mâle, avec incapacité d'alimenter les vagues de spermatogenèse successives. Par une reconstruction de trajectoire développementale et d'étude de la vélocité ARN par scRNA-seq, j'ai révélé que les CSS mutantes pour *Dnmt3A* ont perdu

leur plasticité et leur potentiel de différentiation. Par une analyse des profils chromatiniens, j'ai relié ce phénotype à l'activation erronée d'enhancers qui imposent une surexpression du programme de cellules souches, de manière irréversible. En plus, j'ai montré que ces enhancers erronés contiennent des motifs pour des facteurs de transcription (FT) sensibles à la méthylation de l'ADN, provenant des familles FOXO et ETS, notamment. D'une manière générale, en absence de méthylation de l'ADN, ce qui est observé dans le contexte *Dnmt3A*^{KO}, les FT des familles FOXO et ETS se lieraient plus stablement et fortement à l'ADN. Ceci induit probablement un déséquilibre entre l'expression des facteurs de cellules souches et l'expression des facteurs de différentiation, empêchant les CSS mutantes pour *Dnmt3A* de se différencier.

Ces travaux ont révélé une nouvelle fonction de la méthylation de l'ADN dans la fertilité mâle : la programmation épigénétique de la capacité des CSS à se différencier et à alimenter la spermatogenèse tout au long de la vie.

La signature chromatinienne des cibles génomiques de DNMT3C (projet 2) :

Comme je l'ai déjà énoncé, il a été montré que DNMT3C méthyle sélectivement les promoteurs des éléments transposables évolutivement jeunes et actifs. Ces séquences génomiques ne représentent pas plus de 1% du génome. Ainsi, dans le domaine de la programmation épigénétique de la lignée germinale mâle, une question très intrigante et importante reste non résolue : comment l'enzyme DNMT3C reconnaît si spécifiquement ses cibles génomiques ?

Afin d'apporter des nouveaux éléments de réponse à cette question, j'ai d'abord précisément établi quels ET et quelles parties d'ET sont dépendantes de DNMT3C ou de DNMT3A au cours du développement germinal mâle. Puis, j'ai observé que la dynamique d'action de DNMT3A et DNMT3C était différente : alors que DNMT3A est à son pic d'activité entre 13,5 et 16,5 de développement embryonnaire, l'enzyme DNMT3C est la plus active entre E16,5 et E18,5. Ce délai dans l'activité de DNMT3C suggère l'implication d'une signature chromatinienne particulière. En étudiant la nature chromatinienne des cibles génomiques de DNMT3C par Cut&Run, j'ai effectivement observé que ces séquences présentent une dynamique unique: d'abord un profil bivalent combinant marques actives de type H3K4me3 et marques répressives de type H3K9me3, qui évolue ensuite vers un enrichissement H3K9me3 exclusif au moment de l'activité de méthylation de DNMT3C. La bivalence H3K4me3-H3K9me3 est peu

commune, et suggère une régulation paradoxale et complexe. Enfin, par l'utilisation d'un système d'expression ectopique de DNMT3C dans des cellules en culture, il s'est avéré que DNMT3C cible aussi spécifiquement les ET actifs et enrichis en H3K9me3 dans ce système non-germinal. En conclusion, la méthylation de l'ADN dépendante de DNMT3C semble être étroitement liée à la marque H3K9me3.

Ces travaux ont apporté de nouveaux éléments importants pour comprendre le mécanisme complexe de recrutement de l'enzyme DNMT3C sur ces cibles génomiques et ouvrent de nouvelles directions de recherche sur le rôle de marques bivalentes de chromatine dans la reconnaissance des ET.

TABLE OF CONTENTS

Table of contents

INTRODUCTION

1 PREFACE

The process of development refers to a series of biological changes, mainly differentiation and growth, during an organism's lifespan. Even single cell organisms experience developmental processes to some extent. However, multicellular organisms require more complex and refined developmental processes to be orchestrated. Development is linked to cellular differentiation, where unspecialized cells become specialized cells in order to support a specific role and proper functioning of the organism. Therefore, perhaps the central question in developmental biology is to understand how a specific pattern of gene expression is chosen among infinite possibilities to correctly specify cell types. In my thesis, I investigated the multiple functions of DNA methylation in male-specific germline development and the regulation of spermatogenesis, which refers to the differentiation of mature spermatozoa from fetal germ cells. More specifically, I have studied the involvement of *de novo* DNA methyltransferases in male fertility in mammals, using the mouse model. My project was organized in two aims: *1)* identifying which genomic sequences DNMT3A methylates in male germ cells, and the role of such methylation in spermatogenesis, and *2)* probing the chromatin states that characterize young transposable elements during male germline development, as an attempt to understand how DNMT3C selectively methylates and controls these sequences against their deleterious effects on male fertility.

In this Introduction chapter, I will firstly provide an overview of spermatogenesis and the differentiation steps related to this unique developmental program. Then, I will focus on the biology of spermatogonial stem cells, the essential cell type that gives rise to continuous waves of spermatogenesis during the male lifespan. The second part of the Introduction relates to DNA methylation. I will detail the different DNA methyltransferases and their interplay with various chromatin states. Importantly and contrary to somatic cells, DNA methylation is highly dynamic in the germline and this dynamic behavior is key to germline identity, function and integrity. I will provide emphasis on the DNA methylation changes that accompany male germline development, and their implication for the control of transposable elements.

2 MALE GERM CELL DEVELOPEMENT

2.1 An overview of germ cell development, from commitment to maturation

2.1.1 Fetal germ cells

- *Primordial germ cells*

In animals, two modes of germ cell specification exist: the preformation and the epigenesis specification. Preformation involves the maternal inheritance of a localized determinant—often called germ plasm—present in the egg, which is endowed with a specialized cytoplasm that contains proteins and RNAs needed for primordial germ cell (PGC) formation. In that case, germline formation is secured at the very onset of development, even before embryogenesis begins. The preformation mode is found in many species, including *Drosophila melanogaster*, *Danio rerio* and *Xenopus laevis*. The second mode, referred to as epigenesis or induction mode, is prevalent in animals and more typically found in mammals. It occurs during embryogenesis, whereby PGCs are specified from somatic cells by cell-to-cell signaling pathways and undergo identity reprogramming (Whittle and Extavour, 2017).

In the mouse, PGCs first become identifiable as a cluster of approximately 40 cells in the epiblast around embryonic day (E) 6.5 – 7.25. PGCs first migrate to the developing hindgut endoderm, then into the mesentery before colonizing the developing gonadal ridges at E10.5. Migration continues towards the genital ridges and is complete by E13.5. Cells that migrate to the wrong place or do not arrive in time in the genital ridges, are eliminated by apoptosis. During the migration process, PGCs are highly proliferative, and reach a number of several thousands at E13.5 (Fig.1) (Saitou and Yamaji, 2012).

Germ cell fate acquisition requires both cell intrinsic and extrinsic factors. At E6.5, PGCs express *Blimp1* (or *Prdm1*), a key regulator of germ cell specification (Ohinata et al., 2005). BLIMP1 is a transcription factor (TF) that induces simultaneously the repression of the somatic program and the activation of the germline program. Germ cell specification relies also on signalization from nearby somatic cells that

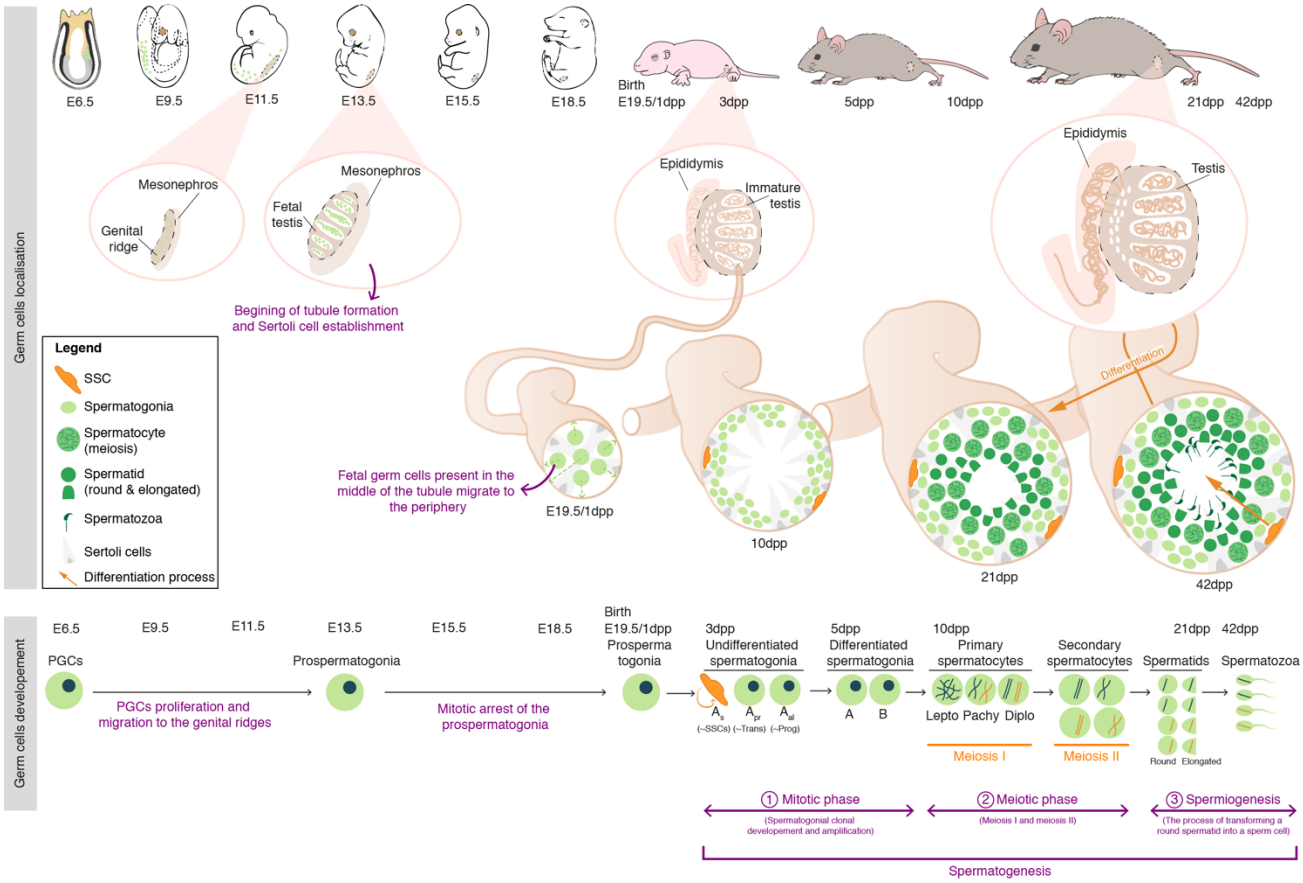


Figure 1 : Graphical summary of fetal germ cell development and spermatogenesis
Germ cells are located since E10.5 inside the seminiferous cords of the genital ridges, which give rise to the future seminiferous tubules and testes, respectively. The differentiation of male germ cells (spermatogenesis) occurs from the basal membrane of the seminiferous tubule to the lumen, as well as along the seminiferous tubule. Spermatogenesis is divided into three phases: mitotic phase, meiotic phase, spermiogenesis (purple).

release bone morphogenetic proteins (BMPs). Loss of *Prdm1* (expressed by PGCs) or *Bmp4* (expressed by the extraembryonic ectoderm) in the mouse embryo results in a dramatic loss of PGC number (Jan et al., 2012).

During migration and proliferation, PGCs undergo profound epigenetic reprogramming, characterized by the almost complete erasure of DNA methylation profiles. The details and determinants of this DNA methylation loss phase will be discussed in Chapter 3 “Male germline reprogramming”.

- *Prospermatogonia*

At E11.5, the sex determination occurs: the *Sry* gene (sex-determining region of Y) is expressed from the Y chromosome in a subset of somatic cells. *Sry* encode a regulatory protein that activates transcription factors, including *Sox9*, and drive differentiation of the somatic cell to Sertoli cells. The Sertoli cells direct, then, sexual development along a male pathway and the PGCs that have successfully reached the

genital ridges become prospermatogonia (or gonocytes). In absence of *Sry* or *Sox9*, the genital ridge develops by default into an ovary (Koopman et al., 1991). Around the same developmental timing, seminiferous cords that will develop into seminiferous tubules are also formed (Jan et al., 2012).

At E13.5, prospermatogonia enter mitotic arrest (G0 phase) and will stay mitotically inactive until a few days after birth. During this mitotically quiescent period, prospermatogonia acquire male and germ cell-specific DNA methylation patterns, as presented in Chapter 3.

A few days after birth, the prospermatogonia—localized in the center of the tubule—start moving towards the basement membrane (Fig. 1). At the same time, they resume proliferation and differentiation, indicating that the first step of spermatogenesis begins.

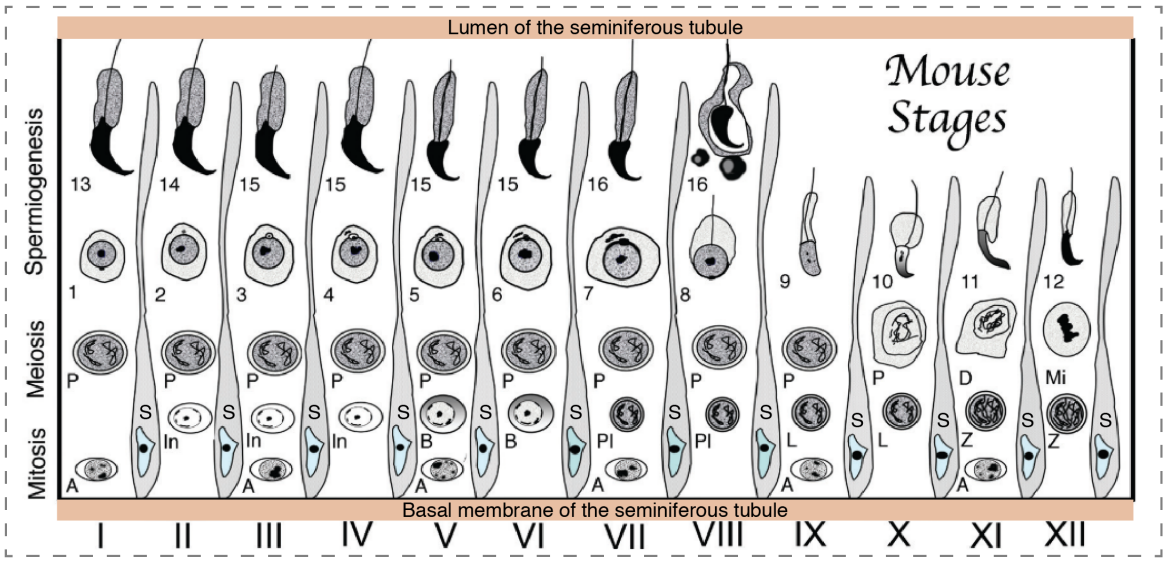
2.1.2 Spermatogenesis

Spermatogenesis starts after birth and is a highly organized developmental process. The goal is the uninterrupted production of spermatozoa to ensure maximum reproductive success throughout the male lifespan. Constant spermatozoa production relies on asynchronous repeating waves of spermatogenesis. Spermatogenesis takes place in seminiferous tubules where germ cells are surrounded by the somatic Sertoli cells. The Sertoli cells play a major role in spermatogenesis by providing the support and suitable microenvironment for germ cell development and maturation. Germ cell differentiation (or spermatogenesis progression) occurs from the basal lamina of the seminiferous epithelium to the lumen of the tubule, and runs throughout the length of the tubule (Fig. 1).

As a result of the spatial and temporal dimension of the spermatogenesis process, the seminiferous tubule at a given position contains a mixed population of germ cells at various states of differentiation (Hess and de Franca, 2009). This phenomenon was first described in rats in 1952 by Leblond and Clermont, who referred to a specific association of germ cell types at a given moment or position along the tubule as a “stage” (Leblond and Clermont, 1952). Each animal species holds a specific number of stages. The mouse possesses 12 stages (stage I to stage XII) (Fig.

2) that can be grouped into three categories: early (stage I-V), middle (stage VI-VIII) and late (stage VI-XIII). A complete series of the different stages is called "the cycle of the seminiferous epithelium" (or cycle). Later, in 1972, Clermont described that one entire cycle of the seminiferous epithelium lasts 8.6 days in the mouse (Clermont, 1972); namely, a particular area of the epithelium always goes through a similar sequence of events during this period of time. This observation also inferred that a new batch of spermatogonia/spermatocytes is being produced at regular intervals, i.e. every 8.6 days. Additionally, 4.5 cycles are necessary for the completion of spermatogenesis (Hess and de Franca, 2009). Therefore, it can be concluded that it

A



Adapted from: 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. http://link.springer.com/10.1007/978-0-387-09597-4_1.

B

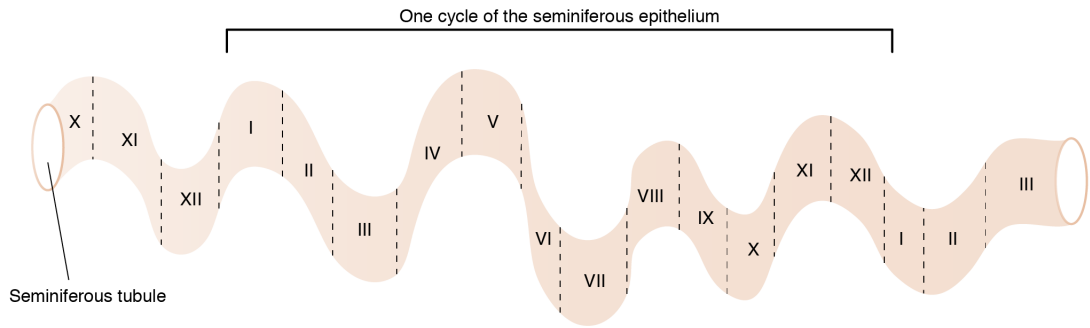


Figure 2 : Mouse cycle of seminiferous epithelium

A. Mouse stages in the cycle of the seminiferous epithelium (I-XII). Sertoli cells (S) separate each association of germ cells (stage). As an example, stage III typically contains: spermatogonia type A, spermatogonia intermediate, pachytene spermatocytes, round spermatids (3) and elongated spermatids (15). Abbreviations: Spermatogonia (A, In, B); spermatocytes (Pl: preleptotene, L: leptotene, Z: zygotene, P: pachytene, D: diakinesis, Mi: meiotic division); round spermatids (1-8); elongated spermatids (9-16). Adapted from Hess and de Franca 2009.

B. Seminiferous tubule divided into stages, one full cycle of the seminiferous epithelium is delineated.

takes around 40 days to produce a mature spermatozoon from a stem cell at the starting point of the spermatogenesis, namely the spermatogonial stem cells, described in details in the following section.

The divisions that germ cells undergo during spermatogenesis are different from any other cell type. Germ cell division is achieved by incomplete cytokinesis leading to cells that are interconnected by bridges to form chains of 2 to 4, 8,16 and sometimes 32 cells. These chains can be considered as interconnected clones. This process is extremely conserved from invertebrates (*D. Melanogaster*) to humans (Greenbaum et al., 2011). Germ cells can be interconnected from undifferentiated spermatogonia to spermatid, or potentially even later stages (Green et al., 2018). The germ cell bridges are essential for fertility, with TEX14 being a major component of the bridge structure. In the absence of *Tex14,* mice are sterile and germ cells are not able to complete meiosis (Greenbaum et al., 2006). It is thought that the bridges allow the sharing of cytoplasmic content, such as RNAs and proteins, and ensure synchronous development of a set of clonally related germ cells (Greenbaum et al., 2011). However, the precise function of cellular bridges is still unknown.

Spermatogenesis is divided into three phases, which are described further below: the mitotic phase (containing spermatogonial stem cells (SSCs), undifferentiated spermatogonia and differentiated spermatogonia), the meiotic phase (primary and secondary spermatocytes) and spermiogenesis (round and elongated spermatids and spermatozoa).

- *Mitotic phase*

The mitotic phase of spermatogenesis is the first event of spermatogenesis and is largely characterized by extensive proliferation. At around 3 days post-partum (dpp) the previously fetal prospermatogonia differentiate into spermatogonial stem cells (SSCs). SSCs are the foundation of spermatogenesis and are essential for male fertility. Indeed, the continuous and unlimited germ cell production across the male reproductive lifespan relies on SSCs. SSCs are rare, comprising only 0.01 to 0.03% of total germ cells in the mouse testis and divide once every three days (Fayomi and Orwig, 2018). The model to produce a new SSC or the discrimination of a “true” SSC from its differentiated progenies is not clear, and the different views/models will be discussed in the following sub-chapter. An accepted theory is that A_{single} (A_s)

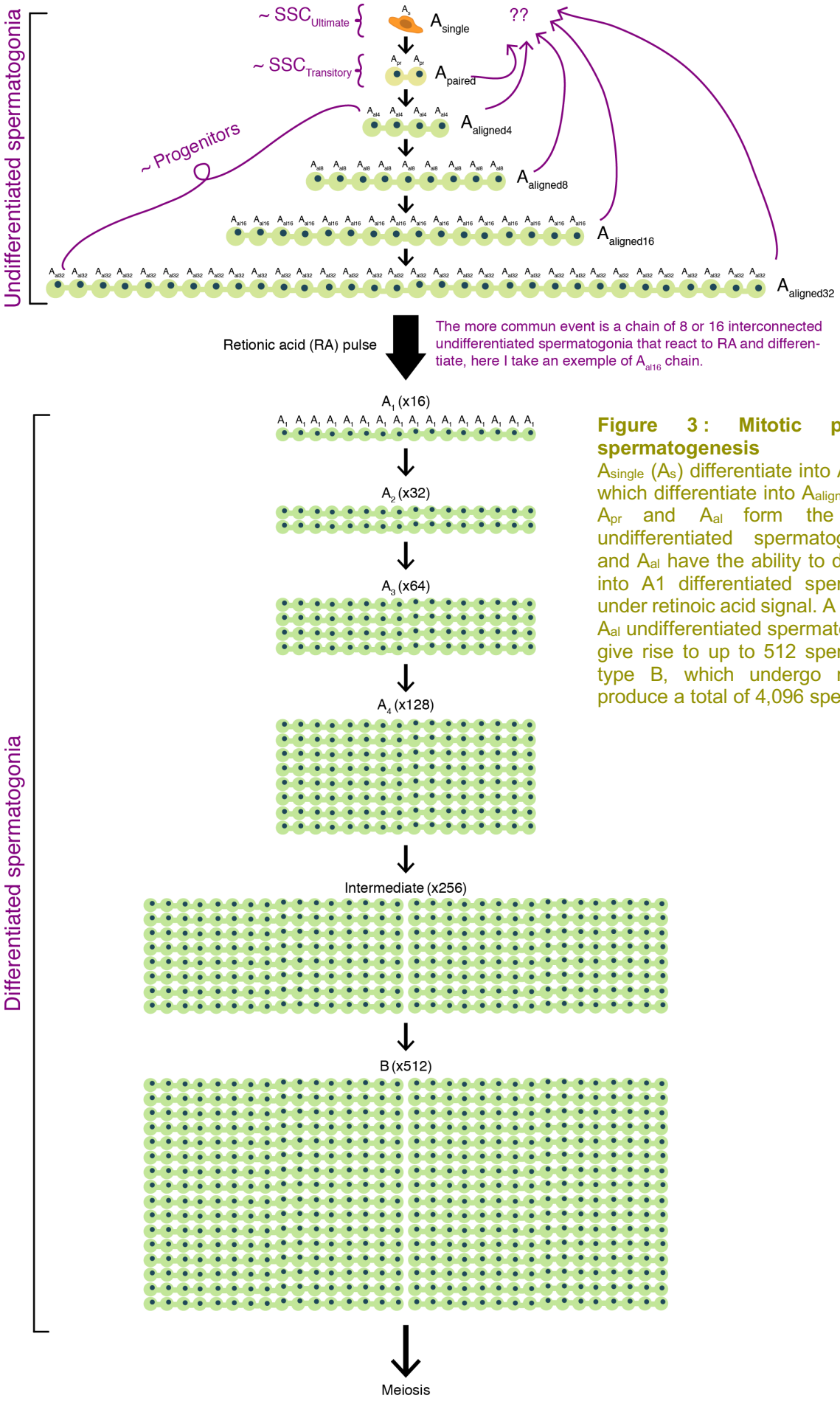


Figure 3 : Mitotic phase of spermatogenesis

A_{single} (A_s) differentiate into A_{paired} (A_{pr}), which differentiate into $A_{aligned}$ (A_{al}). A_s, A_{pr} and A_{al} form the pool of undifferentiated spermatogonia. A_{pr} and A_{al} have the ability to differentiate into A1 differentiated spermatogonia under retinoic acid signal. A chain of 16 A_{al} undifferentiated spermatogonia can give rise to up to 512 spermatogonia type B, which undergo meiosis to produce a total of 4,096 spermatozoa.

spermatogonia are situated at the outset of the undifferentiated spermatogonial pool.

A_s divide and produce a pair of undifferentiated spermatogonia A_{paired} (A_{pr}) that are connected by the intracytoplasmic bridge. These A_{pr} divide again to produce a chain of four undifferentiated spermatogonia $A_{aligned}$ (A_{al4}). A_{al4} can undergo one or more mitotic division to produce chains of interconnected cells from A_{al8} to A_{al16} until A_{al32} (Fayomi and Orwig, 2018; Jan et al., 2012). Collectively, the A_s, A_{pr}, A_{al} make up the population of undifferentiated spermatogonia. A_s can also be considered as SSCs, A_{pr} as transitory cells and A_{al} as progenitor cells (Fig. 3).

The chain of A_{al4}, A_{al8} or A_{al16} undifferentiated spermatogonia have the ability to form differentiated spermatogonia A1 as a simple differentiation without division upon reception of a retinoic acid (RA) signal, the active form of vitamin A. In contrast, SSCs and some small chains (2 to 4 cells) are resistant to this signal. It is believed that the longer the chain, the greater ability the cells have to differentiate under the RA pulse signal. It is likely that the most common differentiation occurs on clones of A_{al8} or A_{al16} (Lord and Oatley, 2017; Tagelenbosch and de Rooij, 1993). The RA is synthesized by the Sertoli cells, and is generated in a pulse (Gewiss et al., 2020). In absence of vitamin A, mice show an arrest in the differentiation into A1 spermatogonia, demonstrating the key function of RA in spermatogenesis. However, the retinoic acid target genes involved in this differentiation process are still largely unknown. On the other hand, it is known that the differentiation into A1 is also characterized by the acquisition of cKIT, which is the receptor for the KIT ligand (KITL), also known as the stem cell factor (SCF). cKIT is essential for differentiation: an heterozygous mutation of the *cKit* locus or mutation of its ligand SCF causes a block in the differentiation of Aal into A1 spermatogonia (Jan et al., 2012; Song and Wilkinson, 2014). There are multiple other genes that also follow a similar expression pattern to *cKit*, including *Sohlh1*, *Sohlh2*, *Cyclin D2*, *Stra8*, *Dmrtb1* and are therefore considered to be markers of differentiated spermatogonia (Green et al., 2018; Hermann et al., 2018; Song and Wilkinson, 2014).

Following A1 specification, the A1 spermatogonia undergo sequential mitotic divisions that produce A2, A3, A4, Intermediate and B spermatogonia. Due to these extensive rounds of mitosis and the formation of the cellular chain, one unique SSC can give rise to a tremendous amount of differentiated type B spermatogonia. These numbers can range between 128 to 1024 type B, depending of the size of the clone

that responded to RA signaling (Fig. 3). Altogether, A1 to B spermatogonia shape the pool of differentiated spermatogonia.

- *Meiotic phase*

Following the mitotic phase, differentiated type B spermatogonia are ready to divide into primary spermatocytes. These cells then enter the meiotic phase to produce haploid cells. Meiosis is achieved by one round of DNA duplication and two consecutive rounds of chromosome segregation: meiosis I (segregation of the homologous chromosomes) and meiosis II (segregation of the sister chromatids) (Jan et al., 2012).

Meiotic prophase I is an extended and crucial step of meiosis that can be divided into four cytological stages: leptotene, zygotene, pachytene and diplotene (Fig. 4).

Leptotene: The first stage of meiosis I, leptotene, occurs immediately after DNA replication. DNA damages in the form of programmed double-strand breaks (DSBs) are catalysed by the meiotic topoisomerase-like SPO11 protein. These DSB are essential to trigger the initiation of homologous recombination between homologous chromosomes. Indeed *Spo11* mutant mice present defects in the formation of DSBs, resulting in the arrest of spermatogenesis at the zygotene/pachytene stage. Meiotic *Spo11-/-* cells are eliminated by apoptosis, due to chromosome structural defects or synaptic failure (Baudat et al., 2000). Simultaneously with DSB formation, the chromatin starts to condensate.

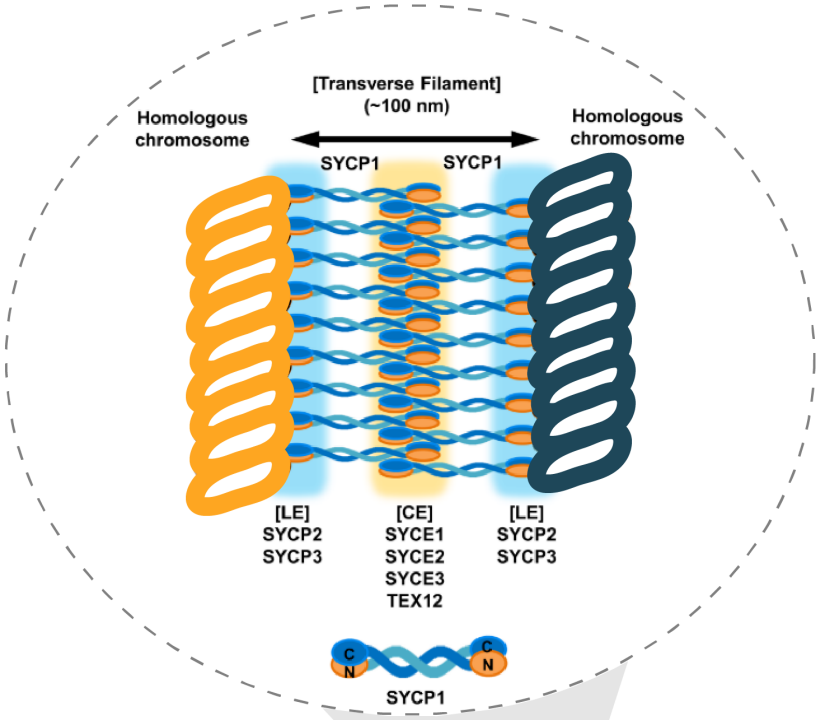
Zygotene: At zygotene, the chromatin is more condensed, and synapsis is initiated, meaning that homologous chromosomes start pairing. Synapsis is helped by the synaptonemal complex (SC), a protein structure composed of SYCP1, SYCP2, SYCP3 and TEX12 (Fig. 4) (Seo et al., 2016). The SC can be compared to a zipper structure, it facilitates the synapsis by holding the aligned chromosomes together.

Pachytene: At pachytene, the homologous chromosomes are fully aligned, and the synapsis is complete. The development of the recombination sites can begin, leading to at least one crossing over (CO) per homologous chromosome pair.

Diplotene: The homologous chromosomes migrate apart one from another by creating bubbles where the SC is disintegrated, in a process called de-synapsis.

A. Synaptonemal complex

Adapted from Seo, Eun Kyung et al. 2016. "Crystal Structure of C-Terminal Coiled-Coil Domain of SYCP1 Reveals Non-Canonical Anti-Parallel Dimeric Structure of Transverse Filament at the Synaptonemal Complex." PLoS ONE 11(8): 1–13.



B. Meiosis

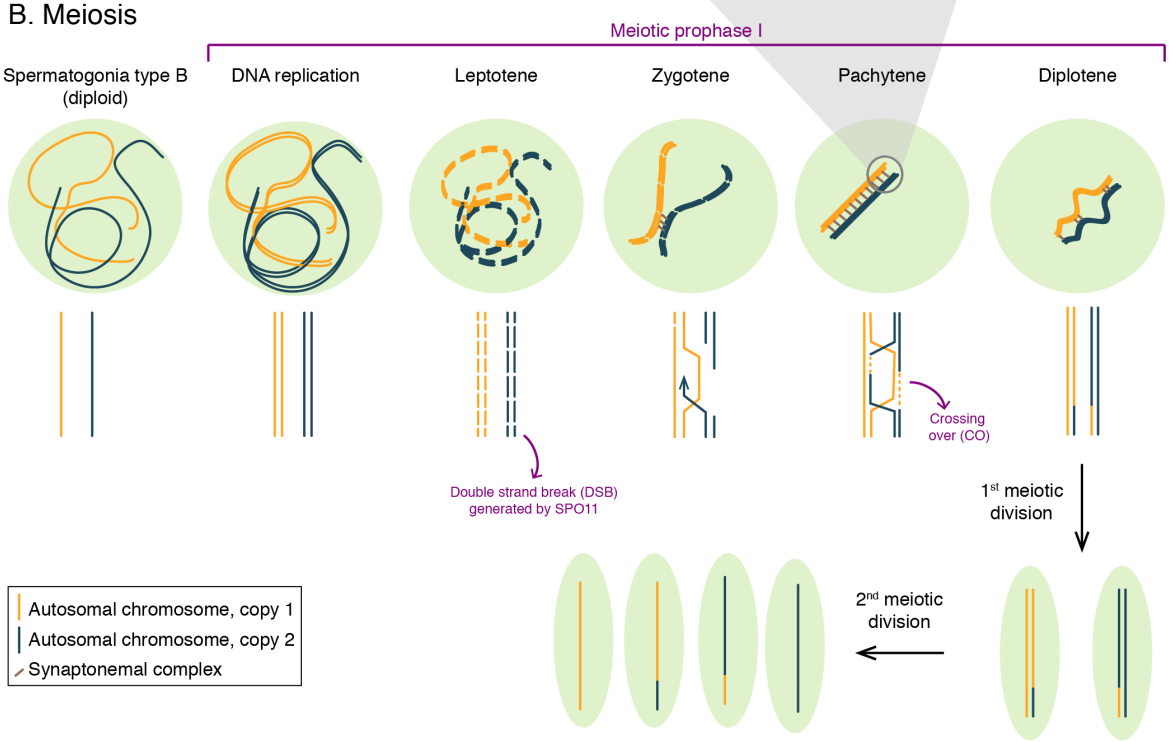


Figure 4 : Meiotic phase of spermatogenesis
A. Protein composition of the synaptonemal complex. SYCP1 and SYCP3 are commonly stained by immunofluorescence to reveal meiotic chromosomes and substages of prophase of meiosis I.
B. Detail of the first meiotic prophase followed by the two meiotic divisions. Representation of a pair of autosomal chromosomes (one chromosome in orange, the other one in blue).

When the meiotic prophase I is complete, and the homologous chromosomes are at least connected to each other by one chiasma, they can segregate together. Then, the two sister chromatids segregate in turn to eventually generate haploid round spermatids (Fig. 4).

Meiosis is often considered as the most crucial phase of spermatogenesis, probably due the existence of male-specific meiotic arrest at pachytene in various meiotic mutants. The pachytene checkpoint detects a failure in the pairing and alignment of homologous chromosomes, or in the meiotic recombination (Jan et al., 2012). However, this chapter shows the complexity and the equal importance of the pre-meiotic and post-meiotic phases as well.

- *Spermiogenesis*

Once meiosis is complete, the haploid cells no longer divide and begin to progress through drastic morphological and cytological changes in a process called spermiogenesis. More precisely, spermiogenesis corresponds to the formation of a sperm cell from a round spermatid through two main changes: 1) morphologically the spermatids develop a distinct head, midpiece and tail region and 2) cytologically they undergo chromatin remodeling, develop an acrosome and lose almost all of their cytoplasm (Fig. 5) (Jan et al., 2012).

Nuclear condensation & chromatin remodeling: The nucleus becomes smaller and DNA is packed more tightly. Histones are replaced by protamines, which are small basic proteins that have the ability to highly condensate the DNA. In the sperm nucleus, the DNA is six-fold more compact than in metaphase chromosomes (McLay and Clarke, 2003).

Acrosome formation: The acrosome is a vesicle located over the anterior half of the sperm head. It contains hydrolytic enzymes required to penetrate the oocyte during fertilization (Jan et al., 2012). Acrosomal biogenesis is a progressive process that starts in round spermatids and is complete in spermatozoa (Fig. 5). A powerful tool was developed in 2015 to visualize acrosomes, by fusing lectin (a peanut protein) with a fluorescent molecule, which has the ability to agglutinate on the acrosome (Nakata et al., 2015).

Flagellum development: The development of the flagellum, or sperm tail, is also a continuous process that is completed near the end of spermiogenesis. The flagellum starts to develop from a centriole at one pole of the round spermatid and is composed of a microtubular structure known as the axoneme. The energy required for flagellum movement originates from mitochondria packed into the midpiece of the spermatozoa.

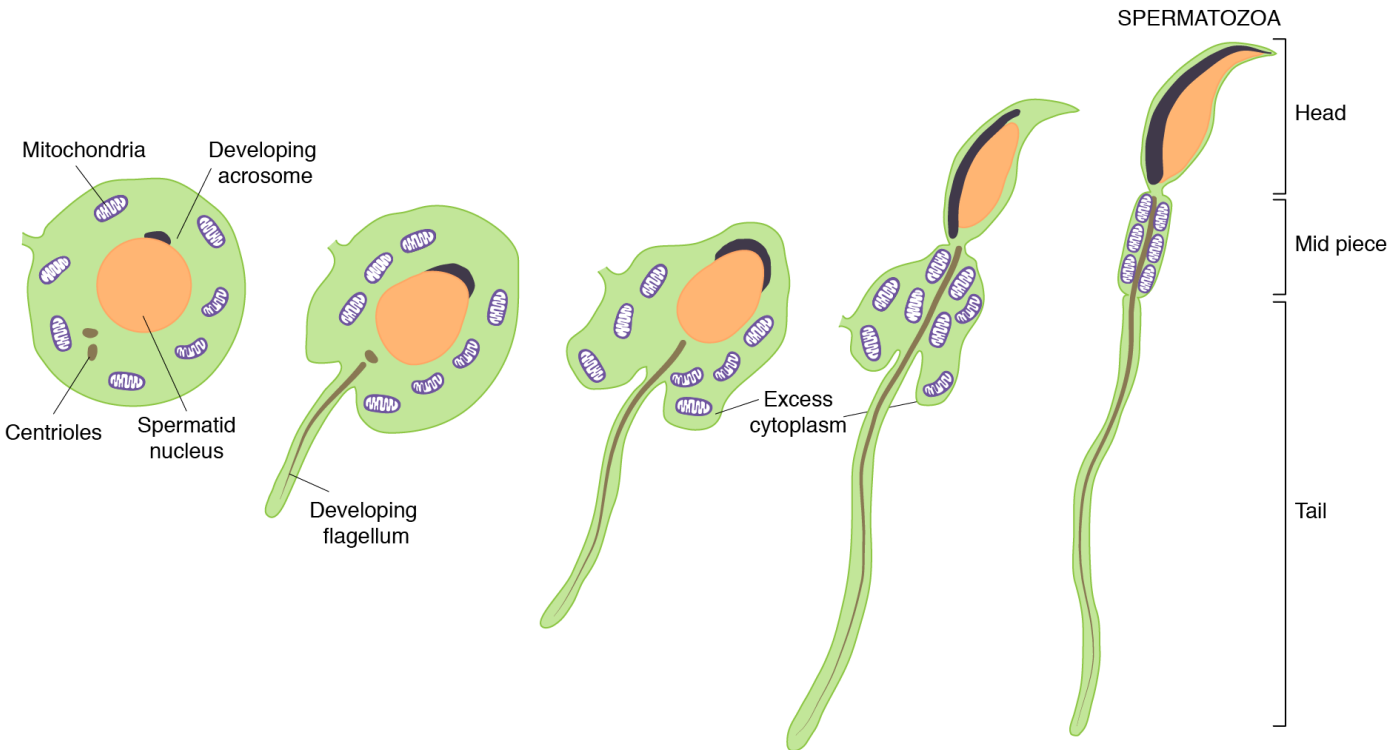


Figure 5: Spermiogenesis
Spermiogenesis is the process by which a mature spermatozoa evolves from an haploid cell (round spermatid). It requires many steps including nuclear condensation & chromatin remodeling, acrosome formation, development of a flagellum and cytoplasmic removal.

Cytoplasmic reduction: Cytoplasmic components are no longer needed in the sperm cell. Therefore, cytoplasmic removal is an important process to ensure the development of a compact and streamlined spermatozoa. Most of the cytoplasm is discarded and phagocytized by Sertoli cells.

At this point of spermatogenesis, ~30-40 days after the SSC stage, sperm cells are produced. However, they do not yet have the ability to swim. Spermatozoa must be released into the lumen of the seminiferous tubule to reach the epididymis where they will there gain motility.

2.2 Spermatogonial stem cell (SSC) biology

SSCs are defined by their dual potential: 1) self-renewal to maintain the stem cell pool and 2) differentiation to maintain continuous sperm production for the male lifespan.

2.2.1 Tools to study SSCs

Regulation and dynamics of SSCs where, for a long time, an enigma of spermatogenesis, for the reason that tools available to study them were limited. The gold standard method in claiming that an undifferentiated spermatogonia is a stem cell is by transplantation. Testicular cells of interest are first marked by a visible transgene (historically LacZ), and isolated usually by fluorescence-activated cell sorting (FACS) or magnetic-activated cell sorting (MACS). Sorted cells are then microinjected into donor testes that lack endogenous spermatogenesis (Brinster and Avarbock, 1994; Brinster and Zimmermann, 1994). If the microinjected cell is able to colonize the testis and to give rise to complete spermatogenesis, then the cell is judged to be a stem cell. On the contrary, absence of testis colonization suggests a lack of the stemness signature. This method has limitations: in the condition of a testis completely empty of germ cells, the availability of the niche will be increased compared to a steady-state condition. Some germ cells could acquire the stemness state, while it would not be the case in normal conditions. For example, it is known that in the presence of injuries, some cells that were not previously classified as SSCs gain the ability—by expressing anew SSC markers—to give rise to new spermatogenic waves (Lord and Oatley, 2017; Zhang et al., 2016).

The second method for monitoring SSC dynamics is the combinatorial use of a reporter transgene and lineage tracing. This method is advantageous in facilitating the observation of stem cells in steady-state conditions (Nakagawa et al., 2007, 2010).

Finally, in order to study SSCs, it is crucial to have markers that accurately identify and distinguish stem cells from their differentiating progenies, especially because the stem cells and their progenies lay down in the same epithelial area. These markers need to be validated by transplantation experiments or lineage tracing. Unfortunately, no consensus of SSC markers has ever been reached in the field. This lack of consensus may reveal the high heterogeneity of the undifferentiated spermatogonia population together with the fact that the expression of markers is more a continuum among the undifferentiated spermatogonia rather than a stringent expression in each undifferentiated spermatogonia subtype. Despite this lack of consensus, several markers are commonly used. GFRA1, ID4, PAX7, BCL6B, ETV5 are known to be more expressed in A_{single} (~SSCs) whereas NGN3, PLZF, SALL4, FOXO are up-regulated in $A_{aligned}$ (~progenitors) (Fayomi and Orwig, 2018; Jan et al., 2012; Lord and Oatley, 2017; de Rooij, 2017; Song and Wilkinson, 2014). It is not clear whether these factors are essential for SSC function and/or behavior. For example,

mice deficient for *Id4* show only a reduction of sperm concentration at 8 months (Oatley et al., 2011). No effect was observed in germline conditional mutant of *Pax7* (Aloisio et al., 2014). The mild effect upon deletion of these genes could be explained by redundancy. In contrast, some factors such as *Etv5* and *Bcl6b* are essential, and show complete sterility and progressive loss of spermatogenesis upon knock-out, respectively (Chen et al., 2005; Oatley et al., 2006).

2.2.2 First wave/round of spermatogenesis

It has been proposed that shortly after birth, at ~3dpp, some prospermatogonia can directly form differentiating spermatogonia (cKIT-positive), without entering a state of undifferentiated spermatogonia. More specifically, in 2006, Yoshida et al. observed through the application of a *Ngn3*-Cre system and a β-gal reporter that a subpopulation of A1 differentiated spermatogonia never transitioned through undifferentiated spermatogonia, which are NGN3-positive. However, this observation was never found in the adult testis, where differentiated spermatogonia always progress through undifferentiated spermatogonia. The authors concluded that fetal prospermatogonia have different fates at birth: 1) differentiate into undifferentiated spermatogonia (expressing NGN3) to contribute to the establishment of the SSC reservoir and perpetuate continuous waves of spermatogenesis during the male lifespan, or 2) differentiate directly to A1 differentiated spermatogonia (expressing cKIT), to enter what is referred to as the first wave/round of spermatogenesis (Fig. 6) (Yoshida et al., 2006).

The term “wave” in this context is easily confused with the “spermatogenic waves” described earlier, and the term “first round” is potentially more appropriate to refer to this process. As a consequence, the first wave of spermatogenesis (the first pool of produced spermatozoa) is composed of *1)* the classical wave of spermatogenesis dependent on SSCs and *2)* the first round of spermatogenesis independent of SSCs. It is important to precise that germ cells that embark into this SSC-independent spermatogenetic path are able to undergo meiosis and to produce mature spermatozoa competent to fertilize an oocyte (Yoshida et al., 2006).

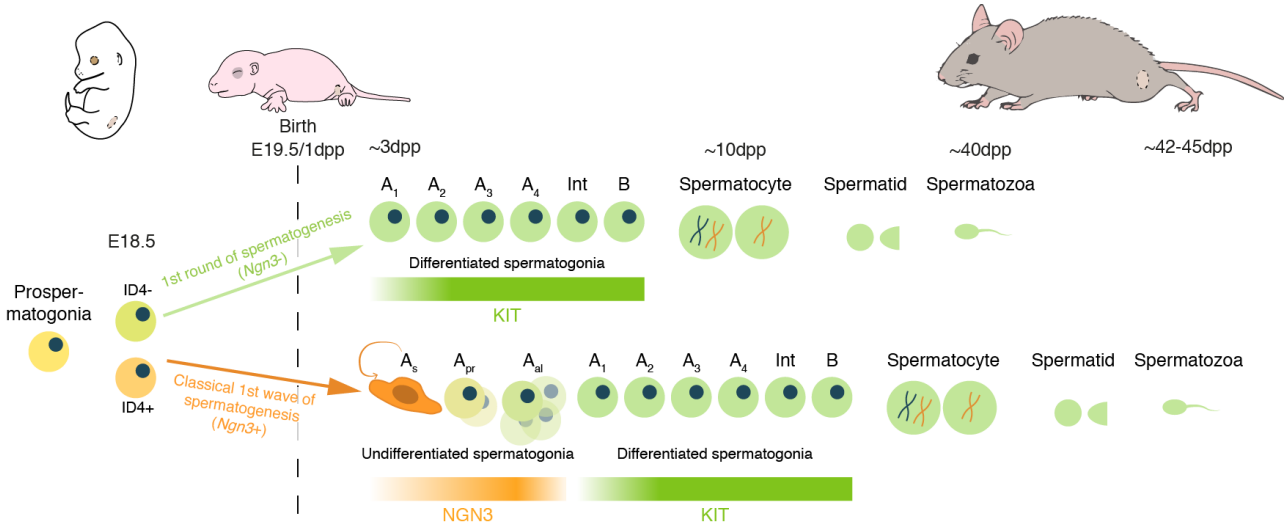


Figure 6 : First wave of spermatogenesis
Different germ cell fates during the first wave of spermatogenesis. Fetal prospermatogonia are preprogrammed to form either differentiated spermatogonia (cKIT-positive) that enter the first "round" of spermatogenesis (no ID4 expression) or to establish the spermatogonial stem cell pool (high ID4 expression).

The duality of the first wave of spermatogenesis is largely accepted. However, these data leave a remaining question. On one hand, the prospermatogonia could be a homogenous population in which each cell would have the potential to become either undifferentiated or differentiated spermatogonia in response to an external control, for example, a factor secreted by Sertoli cells. On the other hand, the prospermatogonia population could be heterogenous: a particular subpopulation may commit to differentiated spermatogonia fate, whereas the other population would contribute to the self-renewing component. Interestingly, in 2019, Oatley and colleagues, revealed by single-cell RNA-seq that fetal prospermatogonia are heterogenous. At E18.5, some prospermatogonia express ID4 (a marker of SSC) whereas some others do not (Fig. 6). Using transplantation assays, they discovered that ID4+ prospermatogonia have the ability to self-renew and colonize empty testis, while the ID4- prospermatogonia could not sustain spermatogenesis (Law et al., 2019). This indicates the the fate that prospermatogonia will adopt is predetermined before birth. It is not known whether the two fetal prospermatogonia subpopulations differ by other features (chromatin states etc..) than differential levels of ID4 expression and other markers of SSCs.

2.2.3 SSC establishment

SSCs are established at ~3dpp directly from prospermatogonia. However, the mechanism that drive SSC formation is not completely understood. It is known that at 3dpp, the diversity within the neonate male germline is considerable in terms of cellular

morphology and gene expression (Law and Oatley, 2020; Yoshida et al., 2006). Three different populations co-exist: population 1 is made up of cells expressing high levels of *Id4* and *Pax7* (Law et al., 2019), and will generate the pool of future SSCs. Population 2 is contributed by cells expressing *cKit* and have the ability to respond to the first RA pulse to form the first differentiated spermatogonia (referred as the "first round of spermatogenesis") (Yoshida et al., 2006). Finally, population 3 is a cell population in between the two others. These cells do not express *cKit*, neither a high level of *Id4* (Law et al., 2019). This population likely represent progenitors (undifferentiated spermatogonia) that do not respond to the first pulse of RA and will commit to differentiation shortly after (with the second or third pulse of RA). I hypothesize that this intermediate population could have the ability to become SSCs, due to their potential plasticity (see later sub-chapter "Maintenance of the SSC reservoir"). The most intriguing question is why among the pool of undifferentiated spermatogonia in neonate testes, some will become SSCs and some will not.

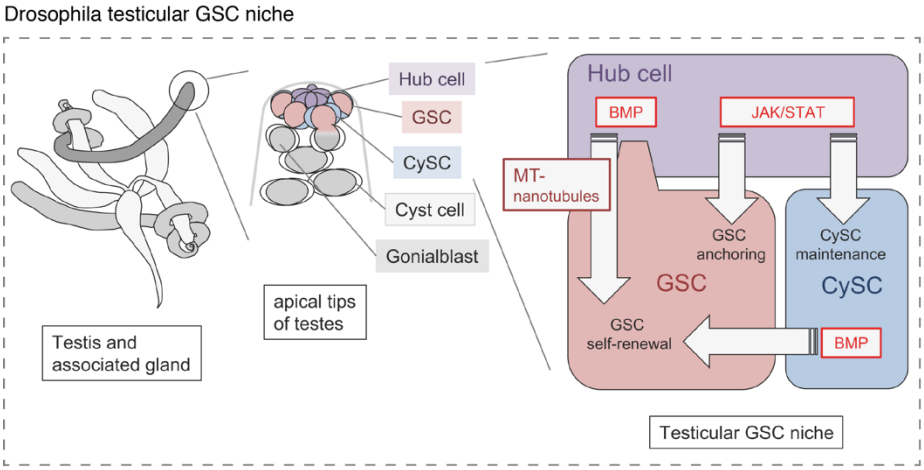
To become an SSC, the cells need to migrate to the basal lamina of the seminiferous epithelium. They will also acquire pseudopodia and create intimate cell-to-cell contacts with adjacent Sertoli cells (Leblond and Clermont, 1952; Orth et al., 1997). About half of the neonatal prospermatogonia form pseudopodia. Contrary to their round counterparts, the pseudopod-shaped cells have greater ability to colonize an empty testis upon transplantation experiments, indicating that migration is essential in SSC formation (Orwig et al., 2002).

Additionally, there is some topographical evidence that the environment, also called the 'niche', could play a role in SSC establishment. Yoshida et al. described that in early born mice, the NGN3 signal (expressed in undifferentiated spermatogonia, therefore potential SSCs) and the cKIT signal (expressed in differentiated spermatogonia) are spatially separated in different seminiferous tubule segments. This result indicates that undifferentiated spermatogonia and differentiated spermatogonia are generated in a spatially separated manner along the length of the seminiferous tubules, suggesting a potential role of somatic cells in creating an area suitable for the formation of SSCs (Yoshida et al., 2006). Moreover, it has been reported that between E18.5 and 3dpp, germ cells with similar *Id4* expression levels form nests, in the form of groups of ~15 cells. After 3dpp, the nest breaks, leading to the establishment of a SSC population. This observation suggests that the cells within a nest adopt a common

fate potentially through the signal of the somatic component (Law and Oatley, 2020; Law et al., 2019).

2.2.4 The SSC niche

A stem cell niche can be defined as an area within a tissue where the stem cells are physically anchored, and kept apart from the rest of the tissue and their differentiating progeny. The niche gives mitogenic and anti-differentiation signals to the stem cells to promote self-renewal. At the moment where a stem cell leaves the cellular environment of the niche, stemness is lost and the cell is able to differentiate. Such an organized germ cell niche is found, for example, in *D. melanogaster*, where the testicular niche is composed of different type of somatic cells (Fig. 7). However, the structure of an SSC niche has never been really found in mammals (Hayashi and Kobayashi, 2018; Morrison and Spradling, 2008).



Extracted from Hayashi, Y. & Kobayashi, S. Regulatory Mechanisms of the Germline Stem Cell Niche in Drosophila melanogaster. 19–35 (2018). doi:10.1007/978-4-431-56609-0_2

Figure 7 : Drosophila testicular germinal stem cell (GSC) niche
Morphology and molecular components of the testicular GSC niche in flies. The testicular GSC niche is located in the most anterior region of each testis (left). The testicular GSC niche consists of GSCs and two types of somatic gonadal cells, hub cells and CySCs (middle). In the testicular GSC niche, hub cells are thought to serve as the master signaling center. Hub cells maintain GSCs by providing JAK/STAT and BMP signals. Hub cells also maintain CySCs by providing JAK/ STAT signal, and CySCs provide BMP signal to associated GSCs, thereby contributing to their maintenance (right). Extracted from Hayashi, Y. & Kobayashi, S. Regulatory Mechanisms of the Germline Stem Cell Niche in *Drosophila melanogaster*. 19–35 (2018).

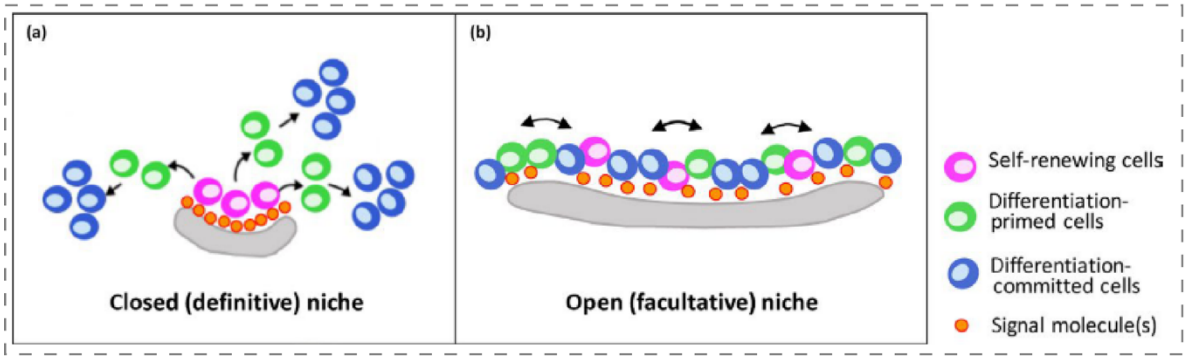
Among the somatic cells of the testis, the Sertoli cells provide first an architectural support but also certainly play a role in SSC behavior. Sertoli cells produce the major source of growth factors, including GDNF (Glial cell line-derived neurotrophic factor) and FGF2 (Fibroblast growth factor) which are known to be signals for self-renewal (Lord and Oatley, 2017; Meng, 2000; De Rooij, 2009). Firstly, these factors were shown to be both required for SSC growth *in vitro* (Song and Wilkinson, 2014) and secondly, it was shown that an excess of GDNF led to undifferentiated spermatogonia accumulation in the tubules, while a reduction (heterozygous mutant) resulted in a depletion of germ cells, thereby suggesting stem cell loss (Meng, 2000; De Rooij, 2009). Unsurprisingly, the SSCs express receptors for both GNDF and FGF2, named GFRA1 (or cRET) and FGFR2, respectively. Therefore, an appealing hypothesis is that Sertoli cells could be part of a potential SSC niche. However, it is also known that Sertoli cells are responsible for releasing differentiation factors towards the SSCs, such as BMP4 (bone morphogenic protein 4) (De Rooij, 2009). Data supporting the hypothesis that Sertoli cells may in principle have the same capacity to stimulate self-renewal and differentiation of SSCs, oppose the hypothesis that Sertoli cells would be part of a "strict" niche as described before. Therefore, we can make two interpretations. Firstly, the Sertoli cell population that is considered homogenous is in reality heterogenous. Different Sertoli cells could have different functions. Accordingly, recent scRNA-seq analysis of testicular cells from different groups—including ours—have revealed the existence of different Sertoli cell subtypes (from two to five subtypes, in neonatal and adult testes) (Dura et al., 2021; Green et al., 2018; Hermann et al., 2018). Secondly, it may be the factors and signals that are received by the Sertoli cells that determine whether they produce factors to induce self-renewal or differentiation of SSCs. In that case, the Sertoli cells would have the role of a messenger rather than a protective role within the niche.

Conversely, it has been described by several groups that GFRa1+ cells (~SSCs) were localized with a bias toward the vasculature (arterioles and venules) and surrounding interstitium (Chiarini-Garcia et al., 2001; De Rooij, 2009; Yoshida et al., 2006). The hypothesis that somatic cells in those areas could play a role in the SSC maintenance was appealing. In 2019, Yoshida and colleagues, provided strong evidence to validate this theory. They demonstrated that GFRA1+ cell (~SSC) fate is regulated by mitogenic and anti-differentiation effects of FGFs released from a subset of lymphatic endothelial and interstitial cells present proximally to the vasculature

network. They proposed that a competition for FGF regulates the SSC pool. In this mechanism, the FGF supply is a limiting factor: stem cells receiving a large amount of FGF will proliferate/self-renew rather than differentiate, leading to a greater stem cell population but less FGF availability for each cell. In contrast, cells receiving a small amount of FGF become primed toward differentiation, engendering a decrease in the stem cell pool. In summary, feedback through mitogen (FGF) consumption plays a major role in density regulation of SSCs (Givelet et al., 2019; Kitadate et al., 2019). This new data gives power to the concept that the SSC niche might be an "open" niche where the stem cells are motile and dispersed among their differentiating progeny, rather than a "strict" niche as found in *D. melanogaster*. Moreover, this mechanism of regulation could be involve other factors, like GDNF.

Finally, it can be concluded that decades of studies on the SSC niche have proved that a "strict" niche likely does not exist in mammals. I believe that the use of scRNA-seq analysis on testicular cells will bring new answers and deeper analyses could expand on the currently existing datasets. For example, for the first time, it was

A. «Open» niche



Extracted from Yoshida, Shosei. 2018. "Open Niche Regulation of Mouse Spermatogenic Stem Cells." Development, Growth & Differentiation 60(9): 542–52. https://onlinelibrary.wiley.com/doi/abs/10.1111/dgd.12574.

B. Pseudopod

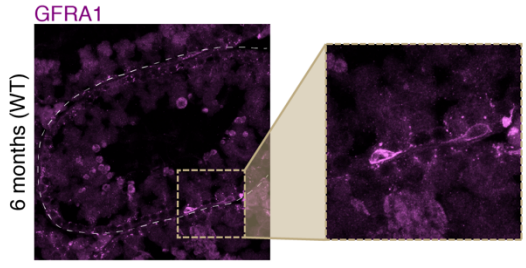


Figure 8: Hypothesis of an "open" spermatogonial stem cell niche
A. The concept of an "open" niche in opposition to a closed/definitive niche. In the close/strict niche, stem cells are conserved apart of the rest of the cells in a confined area. On the contrary, in an "open" niche, stem cells are constantly moving and sit around their differentiating progenies with no differential compartment.
B. Spermatogonial stem cell in adult WT mice, stained with anti-GFRA1antibody (my work). The pseudopods of the stem cells are oriented facing one another, suggesting a movement of the two cells toward the same area.

confirmed that different subtypes of Sertoli cells co-exist within the testis. Moreover, scRNA-seq analysis, including the one I have conducted as part of my PhD work, allowed the discovery of a previously-unknown somatic population of cells (Dura et al., 2021; Green et al., 2018; Hermann et al., 2018). It would be extremely interesting to investigate their function and the relations with the other testicular cells, as well as the role of the Sertoli cell subtypes.

I consider the concept of an "open" niche to be a robust proposition (Fig. 8). Additionally, one of the major cytological characteristics of the SSCs are their pseudopod (Fig. 8), and it is well established that a pseudopod is a signature for motile cells. Therefore, the concept of the open niche where the stem cells always move to find essential factors is in agreement with such a cytological shape.

2.2.5 Maintenance of the SSC reservoir

The lack of knowledge on the SSC niche is likely exacerbated by the fact that there is no consensus model to explain the maintenance of the SSC reservoir. Indeed, this mechanism is intimately linked to the SSC niche. Currently, three models have been proposed : the « classical model » or « Traditional A_{single} model », the « fragmented model », and « hierarchical model ».

On a molecular basis, transcription factors required for SSC maintenance could be induced by somatic cells through GNDF. Possible candidates include ETV5, BCL6b, LHX1, POU3F1, BRACHYURY or ID4. Alternatively, factors could be intrinsically expressed by the SSC (like PLZF, or FOXO1) (Song and Wilkinson, 2014).

- *A "Classical model" or "Traditional A_{single} model"*

In 1971, Claire Huckins proposed a model for spermatogonial stem cell maintenance that was adopted for years in the field (Huckins, 1971). She used whole-mounts of rat seminiferous tubules (Clermont and Bustos-Obregon, 1968) instead of sections, which allowed her to observe the topography and the organization of the stem cells closed to the tubule basal lamina.

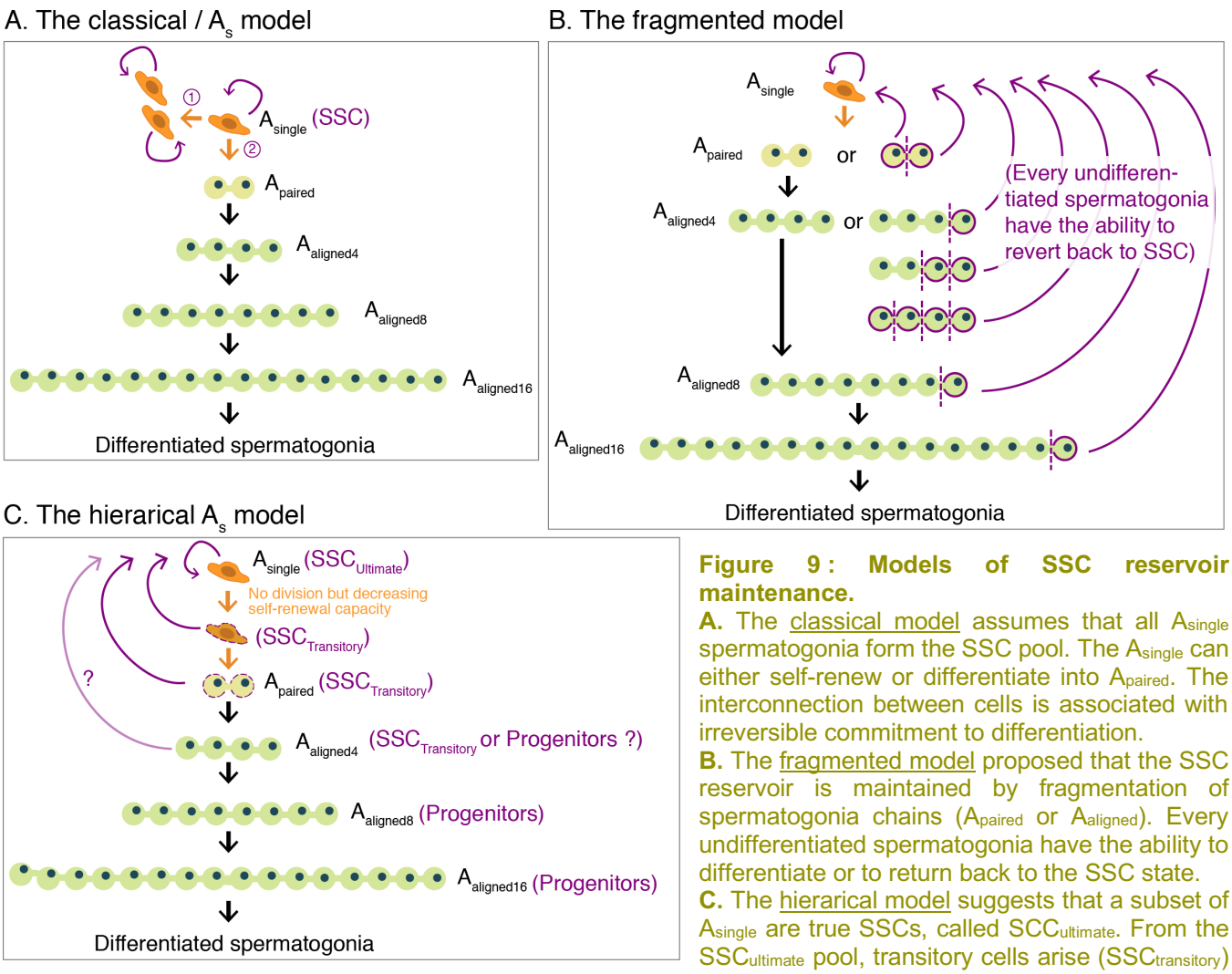


Figure 9 : Models of SSC reservoir maintenance.
A. The classical model assumes that all A_{single} spermatogonia form the SSC pool. The A_{single} can either self-renew or differentiate into A_{paired}. The interconnection between cells is associated with irreversible commitment to differentiation.
B. The fragmented model proposed that the SSC reservoir is maintained by fragmentation of spermatogonia chains (A_{paired} or $A_{aligned}$). Every undifferentiated spermatogonia have the ability to differentiate or to return back to the SSC state.
C. The hierarical model suggests that a subset of A_{single} are true SSCs, called $SCC_{ultimate}$. From the $SSC_{ultimate}$ pool, transitory cells arise ($SSC_{transitory}$) that can either continue to differentiate or revert back to the stem cell state. On the contrary, the progenitors lose their plasticity and are irremediable committed to differentiation.

She proposed that SSCs were exclusively A_{single} undifferentiated spermatogonia. The A_{single} (= SSC) would have two fates: *1)* dividing into two daughters cells disconnected from each other, that will lead to two new A_{single} (= SSCs) able to keep their self-renewing properties; or *2)* the A_{single} divides into A_{paired} connected by an intracellular bridge due to incomplete cytokinesis. This situation would indicate that the cells engage into the differentiation process and lose their stem cell activity, without any possibility of return. The A_{paired} would then differentiate into $A_{aligned4, 8}$ etc… (Fig. 9).

- *Fragmented model*

The fragmented model was proposed by the Yoshida group, over the last 10-15 years (Hara et al., 2014; Nakagawa et al., 2007, 2010). In this model, the stem cell potential

is not restricted to the A_{single} spermatogonia. It was proposed that a single spermatogonia could break off from a chain of A_{paired} or $A_{aligned4, 8, 16}$ and could give rise to a new A_{single} spermatogonia acquiring back the stem cell potential (Fig. 9). This model was in contradiction with the classical model where the commitment to differentiation was irreversible.

The fragmented model was predicted from live imaging and lineage tracing experiments using GFRA1 (~SSC marker) and NGN3 (~progenitor marker) reporter lines. This model has been controversial mainly due to technical issues of the live imaging. The tested mice were kept in long-term anesthesia that might be damaging to the spermatogonia and their properties. For example, it could disrupt the intercellular bridges and artificially cause the “fragmentation”. Additionally, the tubules followed by live imaging were the ones at the periphery of the testis, adjacent to the tunica albuginea. These tubules may not be representative of the entire population of undifferentiated spermatogonia. In particular, we know that SSC localization is biased toward the vasculature of the testis (Chiarini-Garcia et al., 2001; De Rooij, 2009; Yoshida et al., 2006), which is located centrally in the testis.

Even though the experimental procedure can be controversial, and that the fragmented model probably cannot fully explain the general behavior of the undifferentiated spermatogonia in the mouse seminiferous epithelium, this model is still valid and appealing. Importantly, this model suggests that the commitment to differentiation could be reversible. Additionally, due to the potential localization of the A_{al} chain far from the basal membrane, this model of SSC renewal would not be related to the function of the a “strict” SSC niche. Therefore, it brings questions about the role and organization of the SCC niche. Would it be possible that an undifferentiated spermatogonia would be able to revert back to a more naïve state far from the basal membrane? This hypothesis would be in favor of an “open” SSC niche and agree with the proposition that the SSC pool is regulated by competition for FGF (Kitadate et al., 2019).

- *Hierarchical model or A classical revisited*

The hierarchical model was proposed first by the Oatley group in 2017 (Helsel et al., 2017). This model relies on different studies and observations (Lord and Oatley, 2017; de Rooij, 2017):

- First, using an ID4-GFP transgenic line, *Id4* expression was described as a continuum from highly expressed in A_{single} spermatogonia ($ID4^{bright}$) to lowly expressed in $A_{aligned}$ spermatogonia ($ID4^{dim}$) (Helsel et al., 2017). In addition, only ~6,000 A_{single} spermatogonia presented a pattern of $ID4^{bright}$ among the ~35,000 presented in the testis (Chan et al., 2014). The self-renewable capacity of $ID4^{bright}$ and $ID4^{dim}$ cells was tested by a transplantation experiment, into which the $ID4^{bright}$ cells contributed to the formation of significantly more colonies than their $ID4^{dim}$ counterpart (5.5-fold difference). These data indicated firstly a heterogeneity among the undifferentiated spermatogonia and secondly a hierarchy among the SSCs.
- Second, several groups also revealed the heterogeneity among the undifferentiated spermatogonia and the expression continuum of self-renewing factors, including ID4 (Chan et al., 2014; Hermann et al., 2015; Oatley et al., 2011), PAX7 (Aloisio et al., 2014), GFRA1 (Grisanti et al., 2009), NANOS2 (Sada et al., 2009), NGN3 (Yoshida et al., 2004), and OCT4 (Ohmura et al., 2004).
- Finally, other studies (Barroca et al., 2009; Carrieri et al., 2017; Zhang et al., 2016) revealed that in critical conditions, such as upon injury, the NGN3+ cells and the subpopulation NGN3+ and MIWI2+, which were supposedly irreversibly engaged in differentiation where actually able to revert back and restore stem cell capacity. This last result suggests plasticity among the undifferentiated spermatogonia.

To conclude, this model proposed that in steady-state conditions, only a subset of A_{single} spermatogonia are « true » stem cells with self-renewable capacity, called $SSC_{ultimate}$. These $SSC_{ultimate}$ can differentiate into $SSC_{transitory}$, which are probably the A_{paired} and maybe $A_{aligned\ 4}$ spermatogonia, that would be plastic, having the ability to revert back into $SSC_{ultimate}$. Then $SSC_{transitory}$ differentiate into progenitors (likely $A_{aligned}$

$_{8,\ 16,\ or\ 32}$ spermatogonia) that become committed to differentiation, losing permanently the ability to self-renew (Fig. 9).

This model is in greater agreement with the classical model than the fragmented model. However, the fact that SSCs could be plastic and $SSC_{transitory}$ would have the ability to revert back into $SSC_{ultimate}$ is against the classical model and closer to the fragmented model that integrated the dimension of the reversibility.

To conclude on this part, there is strong evidence for the close link between the SSC maintenance and the SSC niche. New data on the SSC niche will provide insights onto the mechanisms of SSC maintenance and vice versa.

I believe that the reality of the SCC maintenance is probably somewhere between the fragmented model and the hierarchical model and I do believe that the the SSC pool is heterogenous and plastic and that the differentiation process is reversible (Barroca et al., 2009). It is possible that the system might be even more heterogenous and plastic that we currently understand. Recent scRNA-seq data, including the ones presented later here as part of my PhD work, validates these two emerging properties of the SSC pool (Dura et al., 2021; Green et al., 2018; Guo et al., 2018; Hermann et al., 2018). Both tSNE and UMAP analyses highlight a true continuum of expression of key factors among undifferentiated spermatogonia, without delineated cell types. In addition, RNA velocity analyses demonstrate a veritable plasticity with the potential capacity of SSCs to revert back (Dura et al., 2021; Guo et al., 2018). These data could eventually support a “plastic” model for SSC maintenance rather than a hierarchical model.

3 DNA METHYLATION

3.1 DNA methylation generalities: distribution and functions

DNA methylation is a well-characterized epigenetic modification by which a methyl group (-CH3) is covalently linked to the fifth carbon of cytosines (5mC) (Fig. 10). Cytosine methylation is a highly conserved mark across vertebrates, plants, fungi and bacteria (Goll and Bestor, 2005).

In mammals, the DNA methylation landscape is collectively shaped by DNA methyltransferases (DNMTs) and DNA demethylation enzymes of the Ten-Eleven Translocation family (TETs). The DNA methylation process is divided into two parts. Firstly, *de novo* establishment is under the control of the DNMT3 family, which contains four members in mice, DNMT3A, 3B, 3C and their catalytically inactive co-factor DNMT3L, and are competent to methylate previously unmethylated cytosines. Secondly, maintenance is ensured by DNMT1 that has the ability to faithfully copy DNA methylation patterns on the newly synthetized DNA strand during DNA replication (Fig. 10) (Chen et al., 2020).

DNA methylation is present genome wide in mammals. In the majority of instances, 5mC is found in a context of CpG dinucleotides. Indeed, although *de novo* DNA methylation can occur in any sequence context (CpG, CpA, CpC or CpT), only symmetrical CpG methylation can be maintained by DNMT1 through cell division. As

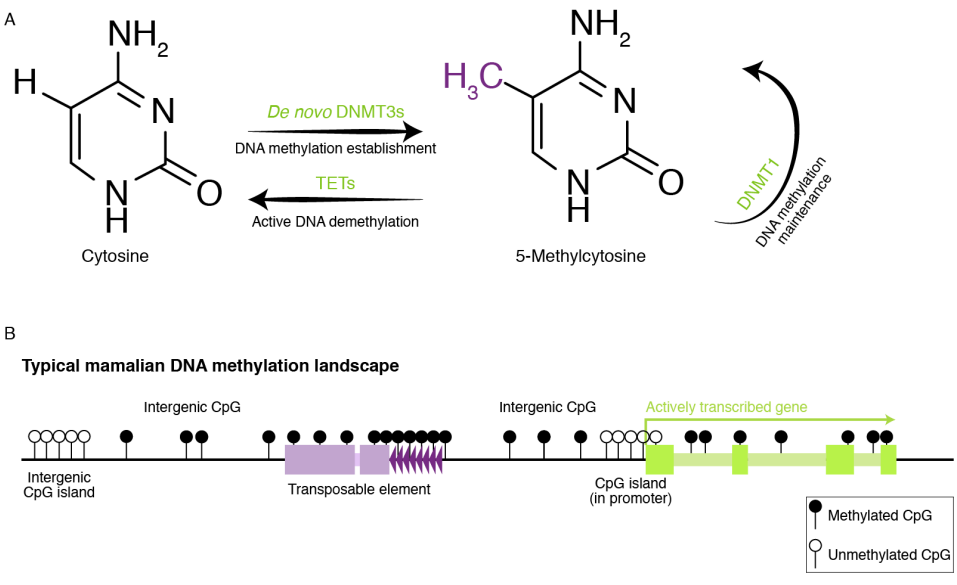


Figure 10 : DNA methylation in mammals
A. Topological formula of unmethylated and methylated cytosine. The addition of the methyl group onto the fifth carbon is catalyzed by *de novo* DNMT3s. DNMT1 maintains 5mC. And, TETs can induce the removale of the 5mC.
B. Typical mammalian DNA methylation landscape. CpG islands are commonly unmethylated, whereas CpGs mapping to transposable elements, and intergenic and intragenic sequences are usually methylated.

a result, while only CpG methylation is found in somatic tissues, 5mC can also be found in a CpA or CpG context in tissues where *de novo* DNMTs are highly expressed, such as mouse embryonic stem cells (ESCs), the brain, oocytes or in prospermatogonia (Greenberg and Bourc'his, 2019).

In mammals, around 4-6% of cytosines are typically methylated in somatic cells, and 60-80% of the ~28 million CpGs are methylated (Ehrlich et al., 1982; Lister et al., 2009). CpGs are not randomly distributed over the genome. Some genomic locations are more enriched in CpGs, these include gene bodies, repetitive elements and CpG islands (CGIs) (Ehrlich et al., 1982; Lister et al., 2009). CGIs are defined as regions with length greater than 200bp, and a CG content greater than 50% or a ratio of observed to expected CpG greater than 0.6 (Bibikova, 2016).

DNA methylation comes with an evolutionary cost. The presence of DNA methylation increases the frequency of spontaneous mutations, through a deamination process leading to a C to T transition. As a result, the genome is generally CpG poor compared to expected, with 5-fold fewer CpG dinucleotides than expected in mammals (Bird, 1980). However, CGIs are the exception. They are resistant to erosion by deamination because they are globally unmethylated, notably in the germline (Greenberg and Bourc'his, 2019). There are around 25,000 CGIs in the human genome: 50% are located in gene promoters and 25% lie in gene bodies, and can often serve as alternative promoters. Reciprocally, two-thirds of mammalian promoters contain a CGI, including all housekeeping genes (Lander et al., 2001). Most CGIs are silenced by polycomb repressive complex 2 (PRC)-mediated methylation of the lysine 27 of the histone H3 (H3K27), a mark that gives more plasticity to rapidly express or repress a gene, compared to DNA methylation (Marasca et al., 2018). Nonetheless, around 10% of CGIs present in some gene classes, such as genes present on the inactive X chromosome, imprinted genes and germline-specific genes, are methylated in somatic cells, providing durable silencing. The mechanism of DNMT recruitment to CGIs is detailed later in the subchapter "Interplay between DNA methylation and histone modifications".

The function of DNA methylation varies according to the genomic location of the mark. Historically, it is well-known to be a stable repressive regulator of promoters. Although how DNA methylation confers transcriptional silencing is not fully understood, promoter DNA methylation is strongly associated with stable silencing of transposable elements (TEs), germline-specific genes (in somatic cells), genes of the inactive X

chromosome and imprinted genes (Chen et al., 2020). On enhancers, DNA methylation is also described to be a repressive regulator. However, enhancer DNA methylation is more dynamic than promoter DNA methylation. As a matter of fact, the vast majority of cell-type specific DNA methylation changes happen at distal regulatory elements (Luo et al., 2018; Song et al., 2019; Stelzer et al., 2015). Additionally, DNA methylation is enriched in the body of highly transcribed genes. There, the DNA methylation mark is not associated with gene silencing (Li et al., 2021), and the precise function of gene body methylation is still unclear. It has been proposed that it could prevent cryptic transcription or facilitate transcription elongation and/or co-transcriptional splicing (Gelfman et al., 2013; Greenberg and Bourc'his, 2019; Shayevitch et al., 2018).

None of the DNMTs have sequence specificity, leading to an important question: how they are recruited to specific genomic locations to establish specific DNA methylation patterns? A strong correlation exists between the genome-wide distribution of DNA methylation and patterns of histone modifications (H3K4me3, H3K9me3, H3K27me3) suggesting a close relationship between these chromatin marks.

3.2 DNA methylation in development

During mouse embryonic development, DNA methylation patterns are erased on a global scale before being reestablished. This process of erasure and re-establishment is called epigenetic reprogramming, and occurs twice during development (Fig. 11) (Chen et al., 2020). The first reprogramming event occurs in the embryo, just after fertilization. CpG methylation levels drop from 70% to 20% during the period from fertilization/E0.5 to blastocyst/E3.5. Embryonic patterns are established *de novo* after implantation past E4.5 and reach somatic levels of 80% of CpG methylation by E8.5. This process is dependent of *Dnmt1*, *Dnmt3A* and *Dnmt3B* (Dahlet et al., 2020).

Deletion of the *de novo* DNA methylation enzyme gene *Dnmt3B* in the mouse leads to embryonic lethality just after E9.5, linked to growth impairment and neural tube defects. Deletion of *Dnmt3A* results in postnatal lethality at 25dpp (Okano et al., 1999; Ueda et al., 2006) (Fig. 11), linked to reduced growth and multiple organ failure. This indicates that the wave of DNA methylation establishment in the embryo is prominently

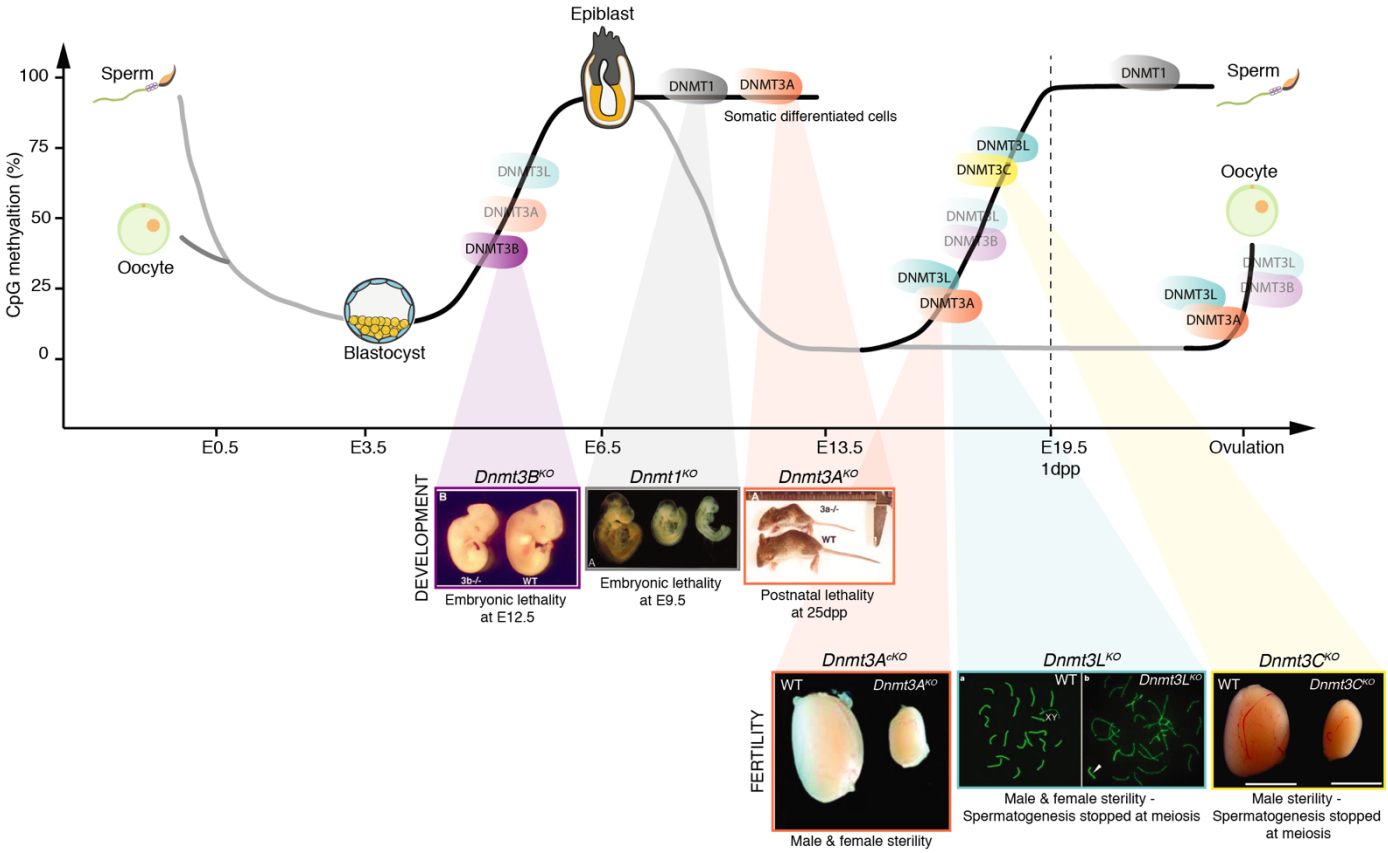


Figure 11: DNA methylation dynamics during mouse development
Two waves of DNA methylation erasure/re-establishment occur during mouse development. Maternal and paternal genomes have different kinetics and mechanism of demethylation and remethylation. DNMT3 proteins that are expressed but not essential for DNA methylation re-establishment at a given stage are represented in lighter color.

dependent on DNMT3B, while DNMT3A is required for postnatal development. DNMT3A has notably known functions in hematopoiesis and neurogenesis (Challen et al., 2012; Wu et al., 2010)

The second DNA methylation reprogramming occurs in developing germ cells, and will be will be discussed in more detail in the following chapter. Briefly, somatic DNA methylation patterns are erased in primordial germ cells (PGCs). In males, DNA methylation is re-established *de novo* in prospermatogonia, collectively by DNMT3A, DNMT3C and their cofactor DNMT3L (Fig. 11). In mice, germline conditional deletion of DNMT3A or constitutive deletion of DNMT3C or DNMT3L leads to a complete sterility phenotype in males (Barau et al., 2016; Bourc'his and Bestor, 2004; Kaneda et al., 2004). Interestingly, DNMT3C and DNMT3L have no function outside of the germline, highlighting the selective pressure imposed by reproduction for the evolution of the DNA methylation machinery in mammals. Following embryonic and germline epigenetic reprogramming, DNA methylation patterns are globally stable for the rest of

development, maintained by DNMT1. Deletion of *Dnmt1* in mice results in embryonic lethality at ~E9.5-10.5 (Fig. 11) (Li et al., 1992).

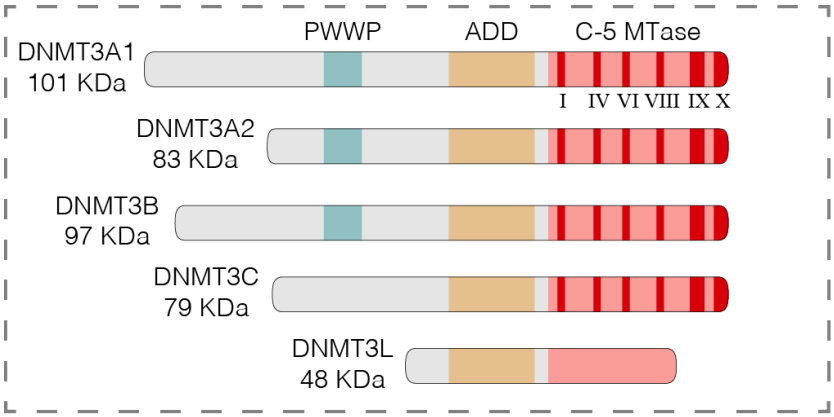
Mouse models of total loss of function have highlighted the essential role of DNA methylation in mammalian development. In humans, deregulation of DNA methylation has also been associated to congenital syndromes of immunodeficiency, growth phenotypes or haematological cancers (Greenberg and Bourc'his, 2019).

3.3 The *de novo* methylation machinery

My work focusing on the targeting of DNA methylation during male germ cell development, I will describe here the *de novo* DNA methylation machinery in mammals. For a comprehensive description of DNMT1, its domain organization and its interplay with the E3 ubiquitin ligase UHRF1 to maintain DNA methylation during replication, detailed information can be found in the following reviews: (Chen et al., 2020; Greenberg and Bourc'his, 2019; Li et al., 2021).

3.3.1 DNMT3 protein domains and their functions

The DNMT3A and DNMT3B enzymes are present in all mammals, while the DNMT3L co-factor has occurred by duplication of the *Dnmt3A gene* in eutherian mammals, likely some 110 million years ago (Yokomine et al., 2006). DNMT3C has emerged even more recently in Muroidea rodents specifically, by tandem duplication of *Dnmt3B,* some 60 millions years ago (Barau et al., 2016; Molaro et al., 2020). Of the four mammalian *de*



Adaptated from Barau et al. 2016. "The DNA Methyltransferase DNMT3C Protects Male Germ Cells from Transposon Activity." Science 354(6314): 909–12.

Figure 12: Domain structure of mouse *de novo* DNMTs
The two mouse isoforms of DNMT3A are represented, the long A1 isoform and the small A2 isoform that is truncated of 219 amino acids in N terminus.

novo DNMTs, DNMT3A and DNMT3B share the most similar domain structure. Beside a C-terminal catalytic methyltransferase (MTase) domain that is responsible for the reaction, they contain an ATRX-DNMT3L-DNMT3A (ADD) domain and a Pro-Trp-Trp-Pro (PWWP) domain (Fig. 12).

The ADD domain has the ability to recognize unmethylated lysine 4 of histone 3 (H3K4), but is repelled by methylation of this residue, with the greatest repulsion exerted by trimethylation (H3K4me3) (Ooi et al., 2007). This characteristic is likely to explain the strong anti-correlation of 5mC and H3K4me3 at CGIs. Another function has been recently attributed to the ADD domain: in the absence of an H3 tail to bind, the ADD domain is able to bind to the MTase domain and auto-inhibit the activity of the DNMT3s (Guo et al., 2015). Moreover, it had been reported that MeCP2, a DNA methylation binding protein, can bind DNMT3A via its ADD domain, *in vitro*. This binding stabilizes the auto-inhibitory conformation of DNMT3A. However, this inhibitory interaction with MeCP2 is relieved when DNMT3A bind to unmethylated H3K4. These observations, provide a complex regulation and competition between the ADD domain and MeCP2 to modulate the activity of DNMT3A (Rajavelu et al., 2018).

The PWWP domain can recognize the di- and tri-methylation of the lysine 36 of histone 3 (H3K36me2/3) (Dhayalan et al., 2010). The H3K36me3 mark (catalyzed by SETD2) is enriched in actively transcribed gene bodies, whereas H3K36me2 (catalyzed by NSD1-3) is present in a more dispersed distribution that includes both gene bodies and intergenic regions. On a closer examination, Weinberg et al. showed that H3K36me2 is enriched downstream of the transcriptional start site (TSS) until the first intron of actively transcribed genes, followed by a chromatin change to H3K36me3 on the remaining gene body (Weinberg et al., 2019). It is unclear why, but it has been reported that the DNMT3B-PWWP domain has better affinity to H3K36me3, whereas the DNMT3A-PWWP domain has a better affinity to H3K36me2 *in vitro* (Weinberg et al., 2019). However, in absence of NSD1, H3K36me2 is depleted from the genome and DNMT3A is redirected towards H3K36me3 (Weinberg et al., 2019).

While the DNMT3L co-factor has an ADD domain, it lacks the PWWP domain and its MTase domain is poorly conserved, explaining its catalytically inactive status (Fig. 12). Nonetheless, DNMT3L is essential for *de novo* DNA methylation in germ cells in both sexes, and has been shown to stimulate the activity of the DNMT3 enzymes through the formation of tetrameric complex that has been resolved by crystallography

(Jia et al., 2007). Moreover, it has been demonstrated, *in vitro*, that DNMT3L directly binds DNMT3A and DNMT3B via its carboxyl-terminal half (Suetake et al., 2004).

Being the result of a recent duplication of *Dnmt3B*, DNMT3C exhibits 70% identity with DNMT3B, while DNMT3A and DNMT3B are only 46% identical (Barau et al., 2016). Nevertheless, DNMT3C and DNMT3B do not share the same domain structure. DNMT3C has an ADD and an MTase domains but lacks the PWWP domain (Barau et al., 2016) (Fig. 12). Furthermore, in contrast to DNMT3B, DNMT3C N-terminus evolved under positive selective pressure, likely responsible for function diversification between DNMT3C and DNMT3B (Molaro et al., 2020). This observation will be discussed in more details in the chapters “Male germline reprogramming” and “Discussion”.

3.3.2 Different isoforms of DNMT3A

Two isoforms of DNMT3A can be produced by the usage of alternative promoters: the long isoform DNMT3A or DNMT3A1, and the short isoform DNMT3A2. DNMT3A2 is truncated in the N-terminal part by 219 amino acids in mice and by 223 amino acids in humans, but still contains the ADD and PWWP domains (Chen et al., 2002) (Fig. 12). DNMT3A2 is more highly expressed in mouse embryonic stem cells (mESCs) and in fetal male germ cells (prospermatogonia) compared to DNMT3A1 (Sakai et al., 2004). In contrast, postnatally, DNMT3A1 is more highly expressed than DNMT3A2, especially in brain, kidney and liver, where DNMT3A2 is barely detectable. In the neonate and adult testis, the two isoforms are present, although there is a predominance for DNMT3A1 expression (data from Margaret A. Goodell laboratory, currently unpublished). It has been recently shown that DNMT3A1 is preferentially localized to bivalent CGIs enriched in both H3K4me3 and H3K27me3 marks. At these locations, DNMT3A1, but not DNMT3A2, competes with TET enzymes to regulate methylcytosine turnover (Manzo et al., 2017).

It would be interesting to generate specific mutants of DNMT3A1 and DNMT3A2 to understand their relative function in development and in germline development.

3.4 Interplay between DNA methylation and histone modifications

3.4.1 Structural link with H3K4me3 and H3K36me3

As presented above, DNMT3s have the ability to sense H3K4 methylation negatively and H3K36 methylation positively through specialized domains, ADD and PWWP, respectively. Apart from these two chromatin marks, H3K27me3 and H3K9me3 are also implicated in an interplay with DNA methylation.

3.4.2 Antagonistic relationship with H3K27me3

Patternsof H3K27 methylation are established by the polycomb repressive complex 2 (PRC2) and catalyzed by the activity of enzymatic subunits of the complex, EZH1 or EZH2 (Enhancer Of Zeste 1/2 Polycomb Repressive Complex 2 Subunit). Enrichment in H3K27me3 characterizes a facultative type of heterochromatin and is associated with gene repression. The interplay between H3K27me3 and DNA methylation is complex and can be dynamic. Within the genome, the vast majority H3K27me3 blocks are at CGIs. CGIs are prevalently unmethylated—as earlier described—and can harbor different chromatin states according to the genomic localization and cell type. They can be decorated either by H3K4me3, H3K27me3, or both in a scenario called bivalent CGIs (Harikumar and Meshorer, 2015). As mentioned earlier, H3K4me3 is a strong antagonist of DNA methylation because of the structure of *de novo* DNMTs that are repulsed by H3K4me3 *via* their ADD domain. Therefore, bivalent CGIs or H3K4me3-CGIs are protected from DNA methylation through direct repulsion (Ooi et al., 2007). At this location, there is currently no evident co-localization of the DNA methylation and H3K27me3 at a single same genomic region, suggesting their mutual exclusivity (Li et al., 2021). It has been reported that CGIs initially regulated by H3K27me3 could undergo an epigenetic switch to transition to a DNA methylation state, during development or during cancer transitions (Mohn et al., 2008).

This mutual exclusivity can be interpreted by the fact that DNA methylation has an antagonistic role on the positioning of H3K27me3. Indeed, in the absence of DNA methylation, or when DNA methylation is particularly low at a transitioning cell state, it

is common to observe H3K27me3 spreading at previously methylated regions (Brinkman et al., 2012; Reddington et al., 2013). There is also evidence for a reverse relationship, where H3K27me3 could antagonize DNA methylation. In the context of CGIs enriched in H3K27me3, proteins containing a CXXC domain have the ability to recognize unmethylated CpG dinucleotides and prevent DNA methylation. For example, FBXL10 is a well-known CGI binding protein that can bind almost all CGIs. In the absence of FBLX10, CGIs that are enriched only in H3K27me3, but not in H3K4me3, undergo *de novo* DNA methylation (Boulard et al., 2015)

Despite the well-documented antagonism between DNA methylation and H3K27me3 at CGIs, they can also co-exist without being incompatible elsewhere. It is the case for CpG-poor regions, or in specific cell types, like cancer cells or highly differentiated somatic cells, where the pattern of H3K27me3 is broader compared to embryonic stem cells (Brinkman et al., 2012; Statham et al., 2012). Additionally, it is possible to observe DNA methylation at CGIs enriched by H3K27me3 when DNMT3A lacks of functional PWWP domain. DNMT3A-PWWP mutation in mice causes postnatal growth deficits due to DNA hypermethylation. This data suggested a potential mechanism of DNA methylation recruitment to H3K27me3 in absence of the PWWP domain (Sendžikaitė et al., 2019). Moreover, a recent paper described a molecular insight of this phenomenon. Weinberg *et al.* showed that a DNMT3A-PWWP mutant could interact with H2Aub modified nucleosomes, which are established by PRC1, via a putative site in the N-terminal region of DNMT3A1. Interestingly, DNMT3A2 or DNMT3B are not able to recognize H2Aub marks. The co-functionality between PRC1 and PRC2 is complex, and the two marks (H3K27me3 and H2Aub) commonly co-localize. In the case of the DNMT3A-PWWP mutant, the co-localization of H3K27me3 and *de novo* DNA methylation is not due to direct interaction between the components but is instead via the modified nucleosome H2Aub. Indeed, ablating PRC1 abolishes the localization of DNMT3A-PWWP mutant at CGIs enriched in H3K27me3 and avoids aberrant hypermethylation. In addition, this data implicates a new role for the PWWP domain in protecting the genome from hypermethylation (Weinberg et al., 2021). In this particular study, it was not mentioned whether the two marks were permanently co-localized over time, or if the *de novo* hypermethylated regions experienced elimination of H3K27me3.

3.4.3 Positive correlation with H3K9 methylation

H3K9 methylation, referring to the modification of lysine 9 on the tail of histone H3 (H3K9me1, H3K9me2, H3K9me3), is involved in the formation of constitutive heterochromatin. In mammals, there are five types of H3K9 methyltransferases (MTases), divided into three groups: GL and G9a that catalyze H3K9me1 and H3K9me2, and SUV39H1/2 and SETDB1 that catalyze H3K9me2 and H3K9me3 (Becker et al., 2016). The Heterochromatin Protein 1 (HP1) has the ability to recognize H3K9me2 and H3K9me3 through its chromodomain. It has been well described that H3K9 methylation and DNA methylation have a strong positive correlation genome-wide (Fu et al., 2020; Meissner et al., 2008). In certain species, including the plant *Arabidopsis thaliana*, DNA methylation is always guided by H3K9 methylation (Jackson et al., 2002). In mammals, the interplays between the two repressive marks is more complex.

In some cases, DNA methylation can recruit chromatin remodelers and modifiers to contribute to the formation of heterochromatin, via H3K9 MTases. The recruitment of the H3K9MTases can also occur through the intermediate of methyl-CpG-binding domain (MBD) proteins that have the ability to bind to CpG methylation and to interact with nucleosome remodeling and histone deacetylase complexes. For example, MeCP2, one of the MBD protein family members, is known to be associated with SUV39H1/2 to generate H3K9me3 at DNA methylated regions (Fuks et al., 2003).

On the other hand, the inverse interaction is prevalent, where H3K9 methylation recruits *de novo* DNMTs, to provide more stable repression at a genomic region. It was suggested that CGIs of meiotic gene promoters acquire DNA methylation after the actionof G9a that deposits H3K9me1/2 to silence the germline genes. In absence of G9a, mouse embryos show a decrease in DNA methylation at germline genes, coinciding with H3K9me2 reduction. However, the level of H3K9 methylation at the same genomic locations was not affected in the *Dnmt3B* mutant embryos: epigenetic silencing of germline genes seems initiated with the deposition of H3K9me2 and then recruitment of DNA methylation (Auclair et al., 2016).

Therefore, it is considered that DNMT recruitment by H3K9 methylation can occur through both direct or indirect interactions, and there are evidence for these two processes:

- Direct interactions: Previous studies have shown that DNMT3A/B physically interacts with SUVH391 and SETDB1 via their ADD domain (Fuks, 2003; Li et al., 2006). In the case of SETDB1, this association was essential for the silencing by DNA methylation of several gene promoters in cancer cells.

- Indirect interactions: In the case of endogenous retroviruses (ERVs) (described in details in Chapter 3), multi-layered modes of DNMT recruitment may occur. ERV sequences can be recognized in a sequence-specific manner by a set a rapidly evolving zinc finger proteins with a Krüppel-associated-box domain (KRAB-ZFPs). These proteins associate with the KAP1 repressor (also known as TRIM28), which is itself involved in a complex with SETDB1 and DNMTs (Groner et al., 2010; Rowe et al., 2010). Moreover, the HP1 protein that recognizes H3K9me3 also has the ability to recruit DNMT3A or DNMT3B to pericentric satellite repeats (Lehnertz et al., 2003).

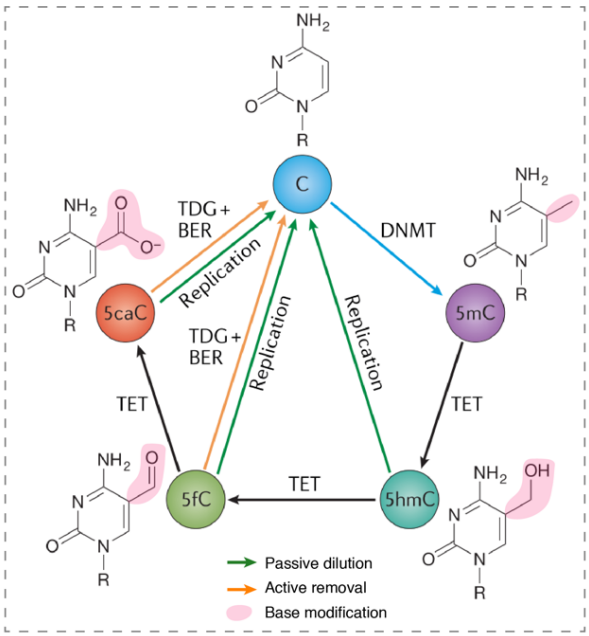
In summary, in mammals, the interplay between H3K9me3 and DNA methylation is highly complex and is context dependent.

3.5 Erasure by the TETs

DNA demethylation being an essential mechanism in germline development, I will describe briefly, here, the different mechanisms that erase DNA methylation.

5mC is both chemically and genetically stable. However, despite this stability, 5mC can still return to unmodified C. The reversion of 5mC can occur in two different forms: *1)* passive DNA demethylation via dilution during DNA replication, which results from the lack of functional DNA methylation maintenance and *2)* active DNA demethylation, dependent on TET enzymes, thymine DNA glycosylase (TDG) and, in certain circumstances, base excision repair (BER) (Wu and Zhang, 2017).

Three different TET enzymes have been described in mammals: TET1, TET2 and TET3. Two and three different isoforms have been reported for TET1 and TET3 respectively (Wu and Zhang, 2017). TETs have the ability to catalyze firstly the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), and secondly they catalyze the oxidation of 5hmC to 5-formylcytosine (5fC) and 5fC to 5-carboxylcytosine (5caC) (He et al., 2011; Ito et al., 2011) (Fig. 13).



Adaptated from Wu, Xiaoji, and Yi Zhang. 2017. "TET-Mediated Active DNA Demethylation: Mechanism, Function and Beyond." Nature Reviews Genetics 18(9): 517–34.

Figure 13: Active DNA demethylation process DNMT convert a cytosine to a methylated cytosine. 5mC can be converted back to unmethylated cytosine by the TET, which mediate oxidation of 5mC, or 5hmC, or 5fC or 5caC. 5fC and 5caC can be excised by thymine DNA glycosylase (TDG) coupled with base excision repair (BER).

The active DNA demethylation by TETs can occur through two different pathways, « active modification-passive dilution » or « active modification-active removal ». The active modification-passive dilution pathway is characterized by the action of TET enzymes that induce oxidation of 5mC to 5hmC (active modification), followed by dilution of 5hmC during DNA replication (passive dilution). Indeed, it has been shown, *in vitro*, that UHRF1 has a lower affinity for 5hmC than non-oxidized hemi-methylated CpG (Hashimoto et al., 2012a) and that DNMT1 is less efficient in maintaining DNA methylation at 5hmC or 5fC or, 5caC hemi-oxidized-methylated CpG (Hashimoto et al., 2012a). Therefore, through several rounds of DNA replication, the modified cytosine will be lost. Secondly, the active modification-active removal is defined by the unique action of TETs. Similarly to the first step of active modification-passive dilution, where 5mC is oxidized to 5hmC, 5hmC is here further oxidized to 5fC and eventually to 5caC. TDG recognizes specifically 5fC and 5caC (but not 5hmC) and induces

excision of the oxidized base from the genome (Hashimoto, Hong, et al. 2012; Zhang et al. 2012). The BER pathway restores the base but in its unmethylated version, leading to active demethylation (Fig. 13).

During development, TETs are required at different time points and different cell types. For example, TET3 is required just after fertilization to actively demethylate predominantly the paternal genome and also, in a minor way, the maternal genome (Chen et al. 2020). The role of TETs in germline reprogramming will be discussed in the next Chapter.

Unlike promoters, enhancers do not coincide with CGIs, therefore they are not subject to the same mechanisms of protection against DNA methylation. TETs play there a role in maintaining an unmethylated state. TET1 possesses a CXXC domain, that, when bound to unmethylated CpGs, helps refraining DNA methylation deposition. Triple knock-out of *Tet1/2/3* in mESCs results in ectopic DNA methylation at enhancers, influencing the transcription of associated genes (Dai et al., 2016).

4 MALE GERMLINE REPROGRAMMING

4.1 Removing somatic DNA methylation patterns in primordial germ cells

4.1.1 Cooperation between DNA demethylation mechanisms

As described in the first Chapter, approximately 40 primordial germ cells (PGCs) are specified from the E6.5-7.25 epiblast. After their emergence, PGCs extensively proliferate and migrate towards the genital ridges. In the course of this process, they undergo DNA methylation reprogramming, leading to the disappearance of epiblast-transmitted patterns that reflect a somatic identity. At E6.5, the level of DNA methylation in PGCs is comparable to the adjacent somatic cells (~70-80% CpG methylation), while at E13.5 DNA methylation is almost completely erased (~5-7% CpG methylation) (Lee et al., 2014; Saitou et al., 2012). The germline DNA demethylation process involves two consecutive steps: passive dilution of methylation from E6.5 to E9.5 followed by TET-dependnet active demethylation from E9.5 to E13.5 (Hill et al., 2014; Wu and Zhang, 2017).

The passive dilution of DNA methylation in PGCs is the main mechanism of DNA methylation erasure, and coincides with the downregulation of both maintenance and *de novo* machineries of DNA methylation (UHRF1, DNMT3A and DNMTB) from E6.5 (Kurimoto et al., 2008). The high proliferation rate of PGCs may also play a role in the passive dilution of DNA methylation. At E9.5, the global DNA methylation level is already down to 30%, with most of the genomic regions being demethylated (Seisenberger et al., 2012). However, some regions, including imprinting genes, meiotic gene promoters and endogenous retroviruses of the Intracisternal A Particle (IAP) family are resistant to the first stage of passive DNA demethylation (Seisenberger et al., 2012). Interestingly, those regions are not resistant on their own, but are protected by the maintenance enzyme DNMT1. Indeed, the conditional knock out (cKO) of *Dnmt1* in early PGCs (E6.5-7.5) by a *Prdm1*-Cre driver results in hypomethylation of imprinting genes and meiotic gene promoters at E10.5. The hypomethylation induces a decreased PGC number, and most importantly, precocious differentiation, as showed by premature entry into meiosis of female *Dnmt1*-cKO PGCs (Hargan-Calvopina et al., 2016).

The second stage of demethylation is an active phase that mainly depends on TET1, although possibly also TET2, and that occurs from E9.5 to E13.5. In this time window, the largest isoform of TET1 is expressed, with no N-terminal truncation (while it is commonly the case in somatic cells) (Wu and Zhang, 2017). The 5hmC mark—the oxidized form of 5mC by TET enzymes—becomes present at E9.5 and stay relatively constant until E13.5 (Hill et al., 2018). Bisulfite sequencing confirmed that specific regions that were resistant to passive demethylation, including imprints and meiotic gene promoters, were actually demethylated at E13.5 (Hackett et al., 2013; Seisenberger et al., 2012). More precisely, it has been shown that TET1 is responsible for protecting germline reprogramming-responsive genes when they are already in an unmethylated state. Intriguingly though, at E11.5-12.5, aberrant residual and/or *de novo* DNA methylation could occur on the promoters of these genes. Atypical DNA methylation occurring on these germline reprogramming-responsive promoters will induce gene repression, resulting in germ cell developmental defects. Therefore, TET1 has a crucial role in ensuring gene demethylation (Hill et al., 2018). Moreover, it was suggested that TET1 could have other functions than DNA methylation oxidation. TET1 may act as an activating factor to ensure the expression of these specific type of genes (Hill et al., 2018). On the contrary, IAPs are still resistant to this second stage of active demethylation and their average level of DNA methylation at E13.5 does not decrease below 40% (Seisenberger et al., 2012). Even though TETs are responsible for actively demethylating a minor portion of the genome, their function is essential. In absence of TET1, female PGCs have insufficient DNA demethylation and cannot enter meiosis due to meiotic gene repression by DNA methylation (Yamaguchi et al., 2012). In males, *Tet* mutant spermatogenesis can progress until spermatozoa. However, progeny from sperm carrying *Tet1* mutations present aberrant imprinted methylation and associated developmental phenotypes (Yamaguchi et al., 2013).

4.1.2 Resistance to demethylation: the case of the IAPs

As mentioned in the above sub-chapter, IAP retrotransposons exhibit a different behavior than the rest of the genome upon germline reprogramming. There are two types of genomic regions that are resistant to the passive DNA demethylation: *1)* the imprinting genes and promoters of meiotic genes that will eventually get demethylated during the second and active stage of erasure, and *2)* the IAPs (and some other

repetitive sequences) but these will remain resistant to the active demethylation, and will never show lower CpG methylation than 40% at the end of the process (Guibert et al., 2012; Hajkova et al., 2002; Seisenberger et al., 2012). This average level indicates that some IAP copies undergo demethylation, while other others are resistant. The question is how some IAPs are resistant to both passive and active demethylation?

The first insights into answering this question came from the use of genetic tools. The generation of the *Dnmt1* germline conditional mutant, driven by *Prdm1*-Cre, showed that DNA methylation was lost on IAPs at E10.5, thereby indicating that DNMT1 protects these regions from passive demethylation (Hargan-Calvopina et al., 2016). However, IAP elements were not de-repressed in *Dnmt1* mutant PGCs, probably due to their enrichment in H3K9me3, which was unaffected upon DNMT1 removal.

On this matter, it was revealed that IAPs are highly enriched in both H3K9me3 and H3K27me3 in PGCs at E13.5 (Liu et al. 2014), although it could not be assessed whether the two marks co-occur on the same copy or different ones. This is in contrast to imprinting genes and meiotic gene promoters that show low H3K9me3 levels. By generating a *Setdb1* germline conditional mutant, driven by *Tnap*-Cre, H3K9me3 was dramatically reduced in PGCs at IAP sequences, resulting in reactivation of these elements. It is important to note, however, that H3K9me3 was not fully absent because of incomplete efficiency of *Tnap*-Cre. Interestingly, loss of H3K9me3 on IAPs was accompanied by DNA methylation loss, and surprisingly, H3K27me3 decreased too. More importantly, H3K9me3 loss reduced the number of PGCs, resulting in germ cell depletion in neonatal and adult testis (Liu et al., 2014). All together this data suggests that the resistance to active DNA demethylation of IAPs can be due to the special retention of H3K9me3. It has been shown, in ESCs, that H3K9me3 was maintained through DNA replication by the help of KAP1 and SMARCAD1 (Jang et al., 2018; Sachs et al., 2019). We can speculate that a similar mechanism also occurs in PGCs, where KRAB-ZFPs could also recognize IAPs in a sequence specific manner, allowing them to maintain their H3K9me3 richness via KAP1 action (Groner et al., 2010). The study of a PGC-conditional mutant of KAP1 would probably bring some answers.

4.2 Acquiring male germline DNA methylation in prospermatogonia

Following the erasure of somatic DNA methylation occurring in PGCs, male-specific germline methylation patterns are established in fetal prospermatogonia, from E13.5 to birth. After birth, the level of CpG methylation remains stable at around 80%, maintained by DNMT1 activity (Kubo et al., 2015). This CpG methylation level persists throughout the production of spermatozoa, despite the presence of *de novo* DNMT3s that can be detected in later germ cell types, such as differentiated spermatogonia (Guo et al. 2017; Hermann et al. 2018; Law, Oatley, and Oatley 2019). A recent study confirmed this, showing that that very few, if any, 5mC changes could be detected during spermatozoa maturation in the epididymis (Galan et al., 2021).

Three *de novo* DNMT3s are required to establish the male germline methylome in fetal stages: DNMT3A, DNMT3C and their co-factor DNMT3L. Mutation in any of these three genes leads to male sterility (Barau et al., 2016; Bourc'his and Bestor, 2004; Kaneda et al., 2004). DNA methylation occurs in two successive waves in prospermatogonia. The first wave methylates most of the genome from E13.5-E14.5 to E16.5, except evolutionarily young transposable elements (TEs, defined in the next sub-chapter), which still present less than 40% CpG methylation at E16.5, while the rest of the genome is already methylatedat around 70% at that time (Molaro et al., 2014). The first round of methylation is suggested to be DNMT3A/DNMT3L-dependent, while the second "minor" wave—which is mostly devoted to young TEs—is thought to be DNMT3C/DNMT3L-dependent and requires the action of PIWI-interacting small RNAs (piRNAs), as described in more detail below.

4.2.1 Dedicated mechanisms for de novo DNA methylation of transposable elements

- *Transposable elements*

Transposable elements (TEs) are genetic elements that are present in genomes of most species. These elements are able to mobilize and replicate themselves in the host genome in a process called transposition. In mammals there are two classes of TEs:

- Class II: DNA transposons contribute ~3% of the mouse genome

- Class I: Retrotransposons contribute ~40-50% of the mouse genome. They duplicate via a « copy-paste » mechanism, involving an RNA intermediate. Thousands of copies have invaded the mouse genome over evolutionary times, representing around 40-50% of the mouse genome today. They are divided into two families depending on whether or not they possess long terminal repeats (LTRs), a feature that has promoter properties (Fig. 14):

- The LTR retrotransposons, also called endogenous retroviruses (ERVs, divided into ERV1, ERVL, ERVK), contribute ~8-10% of the mouse genome. These are derived from infectious retroviruses that have invaded the germline and became permanent residents of the host genome. It is important to note that the earlier-described IAPs that retain DNA methylation during the PGC erasure phase belong to the ERVK sub-family (Guibert et al., 2012; Hajkova et al., 2002; Seisenberger et al., 2012). This subfamily is one the most famous because is still very « aggressive » in the mouse genome, by retaining the ability to generate *de novo* insertions at a frequency largely dependent on the mouse strain (Rebollo et al., 2020).

- The non LTR retrotransposons are the most abundant category in the mouse genome, contributing up to 20% of the total DNA mass. In this family, there is the dominant LINE (long interspersed nucleotides elements) category, and the SINEs (short interspersed nucleotides elements) that are non-autonomous and use LINE-encoded proteins to replicate. LINEs do not have an LTR but instead, utilize their 5'UTR region as a promoter, which consists in tandem monomer arrays in the mouse (Fig. 14) (Sookdeo et al., 2013).

TEs are powerful for evolution because they are a source of genomic innovation (Enriquez-Gasca et al., 2020). On the other hand, selfish elements moving around the genome can be mutagenic and deleterious for the host organism. Therefore, host cells have evolved different layers of regulation to silence them. DNA methylation is one of the most stable forms of TE repression (Yoder et al., 1997). Interestingly, although TE regulatory mechanisms are commonly conserved across species, the TE themselves are species-specific (according to the TE invasion that each species encountered) (Deniz et al., 2019). For example, two species of *Drosophila* exhibit different TEs (Parhad et al., 2017). It should be noted that the regulation of TEs is much more complex than a permanent and strict repression. TEs can be co-opted by the host and utilized for essential functions, therefore requiring the need for TEs to be expressed, but in a controlled spatio-temporal manner. For example, in the mouse, the ERVL

family are highly expressed at the two-cell stage, driving the expression of two-cell stage related genes through their promoters (Evsikov et al., 2004; Hermant and Torres-Padilla, 2021). MaLRs, old elements belonging to the ERVL family, are expressed in the oocyte where they are essential to generate alternative transcription start sites (Brind'Amour et al., 2018).

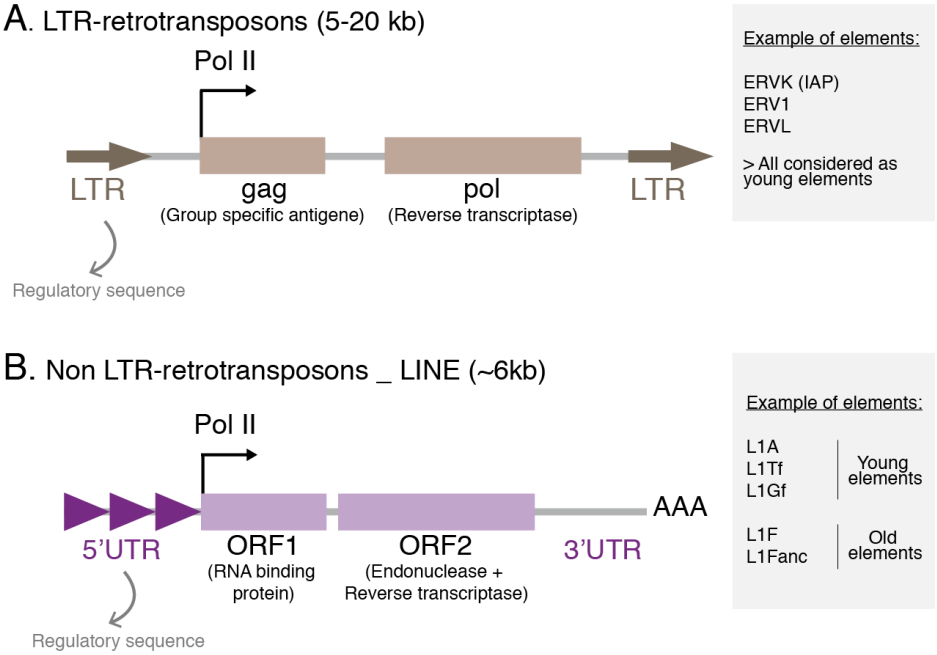


Figure 14 : Representation of mouse retrotransposons
A. LTR-retrotransposon representation. The element is flanked by two LTR sequences. **B.** Non LTR-retrotransposon representation (example of a LINE1). The 5'UTR that have promoter properties is formed of several monomers in a raw.

Evolutionary age matters when considering TEs. According to their time of invasion into the host genome, TEs can be divided into new or young elements, and old or ancient elements (Enriquez-Gasca et al., 2020). Young TEs are usually lineage- or species-specific, whereas old TEs can be common to ancestral lineages (Sookdeo et al., 2013). Young TEs are the most dangerous for the host genome, because they contain intact promoters that confer them with the ability to be expressed and thereby retrotranspose. Young elements are therefore the ones that are tightly regulated by the host through transcriptional and post-transcriptional mechanisms (Deniz et al., 2019). On the contrary, old TEs accumulated mutations over evolutionary time and lost their transcriptional capacity. The old elements harbor a behavior more closely resembling the host genome on average.

For the purpose of the next sub-chapter, it is important to note that the mouse genome was invaded first by LINE1 elements (L1). Some of these L1s appeared ~15 millions years ago. The extensive L1 family of L1s is divided between young and old subfamilies. The younger subfamiliess—L1A, L1Tf and L1Gf—appeared less than 2

million years ago. The remaining L1s, *e.g.* L1F and L1Fanc, are considered as old (Sookdeo et al., 2013). In the case of ERVs, they also had an ancient invasion. Some ERVs are dated from 70 million years ago. Contrarily to the LINEs, despite their age, some elements are still maintained in a transcriptionally active state (Stocking and Kozak, 2008). Therefore, even though all LTR copies are not young in the context of evolutionary time, it is prudent to consider that all LTR element have the potential to be expressed in specific mouse cell types or developmental stages.

- *The piRNA pathway*

In the germline of all animals, specific small RNAs play a key role in regulating TE abundance and expression, these are the PIWI-interacting RNAs (piRNAs). piRNAs are 24-30nt-long RNA species that are loaded onto PIWI proteins, a subgroup of Argonaute proteins. In mice, there are three PIWI proteins: MIWI (PIWIL1), MILI (PIWIL2) and MIWI2 (PIWIL4) (Chuma and Nakano, 2013). During male germline development, MIWI2 is specifically expressed in fetal prospermatogonia, while MILI is expressed in prospermatogonia and also later in postnatal spermatocytes and spermatids, along with MIWI (Chuma and Nakano, 2013). The PIWI proteins are assembled with various accessory proteins, forming the “nuage” structure in the cytoplasm of germ cells (Chuma and Nakano, 2013; Parhad and Theurkauf, 2019).

piRNAs are essential for germ cell development and mutations of the piRNA pathway lead to male sterility. *Mili* and *Miwi2* mutants display a spermatogenesis arrest at the pachytene stage of meiosis I (Carmell et al., 2007; Kuramochi-Miyagawa et al., 2004). The *Miw*i mutant demonstrates a later spermatogenesis arrest, in round spermatids (Deng and Lin, 2002; Reuter et al., 2011). This difference in phenotype reflects the existence of two classes of piRNAs: the pre-pachytene or fetal piRNAs, synthetized by MILI and MIWI2, and the pachytene piRNAs synthetized by MILI and MIWI (Parhad and Theurkauf, 2019). Pachytene piRNAs play a minor role in silencing TEs, and are mainly required to regulate genes (Rojas-Rıós and Simonelig, 2018). Therefore, for the purpose of the following chapter, I will focus on pre-pachytene piRNAs, whose function is to silence young TEs. The piRNA defense against TEs is exerted at two levels: post-transcriptional, via cleavage of TE transcripts, and transcriptional via the recruitment of repressive chromatin.

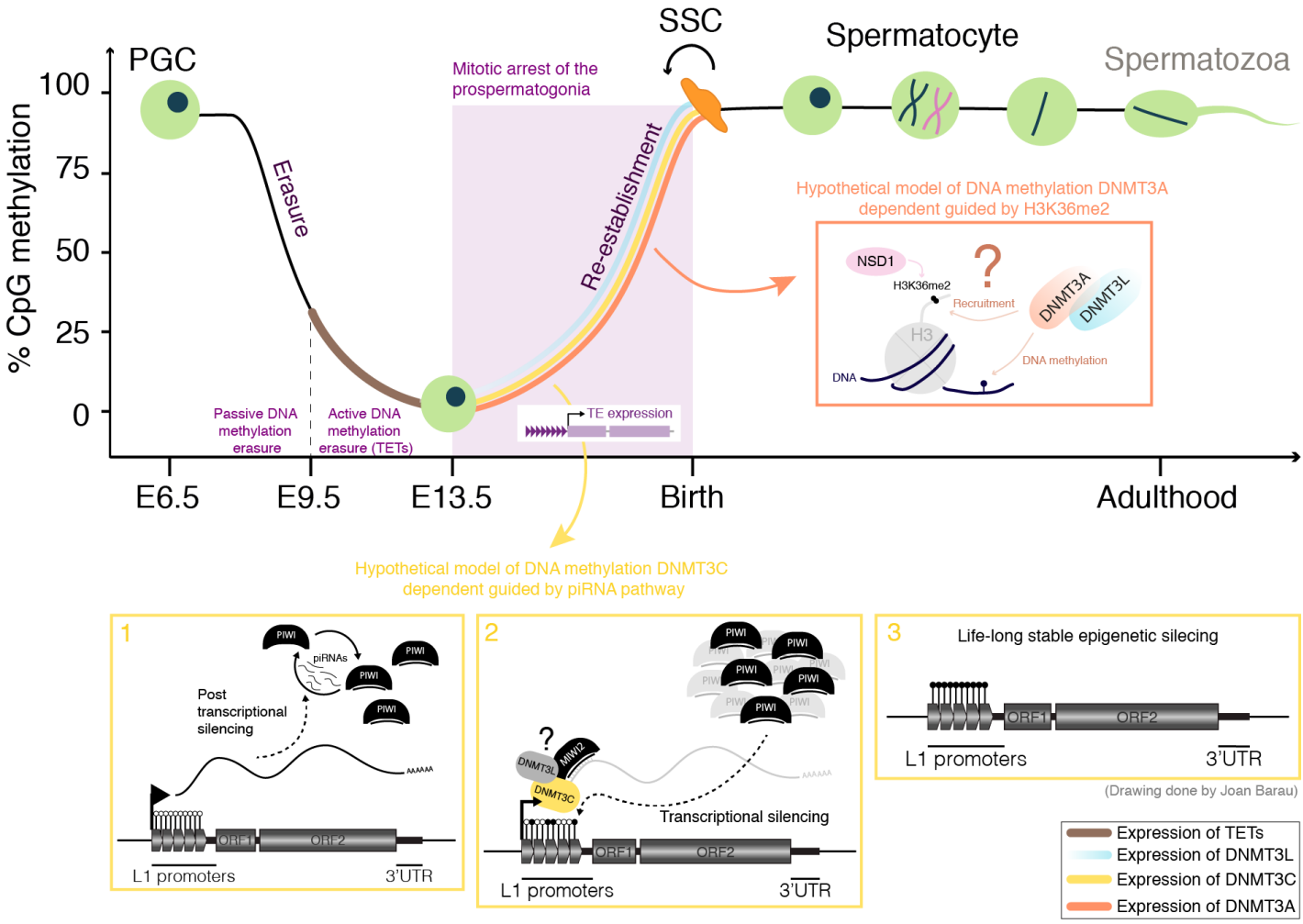


Figure 15 : Drawing of male germline reprograming which indicates TET enzymes expression patterns and essential *de novo* DNMT3s. Hypothetical model of DNMT3C (orange) and DNMT3A (yellow) guidance.

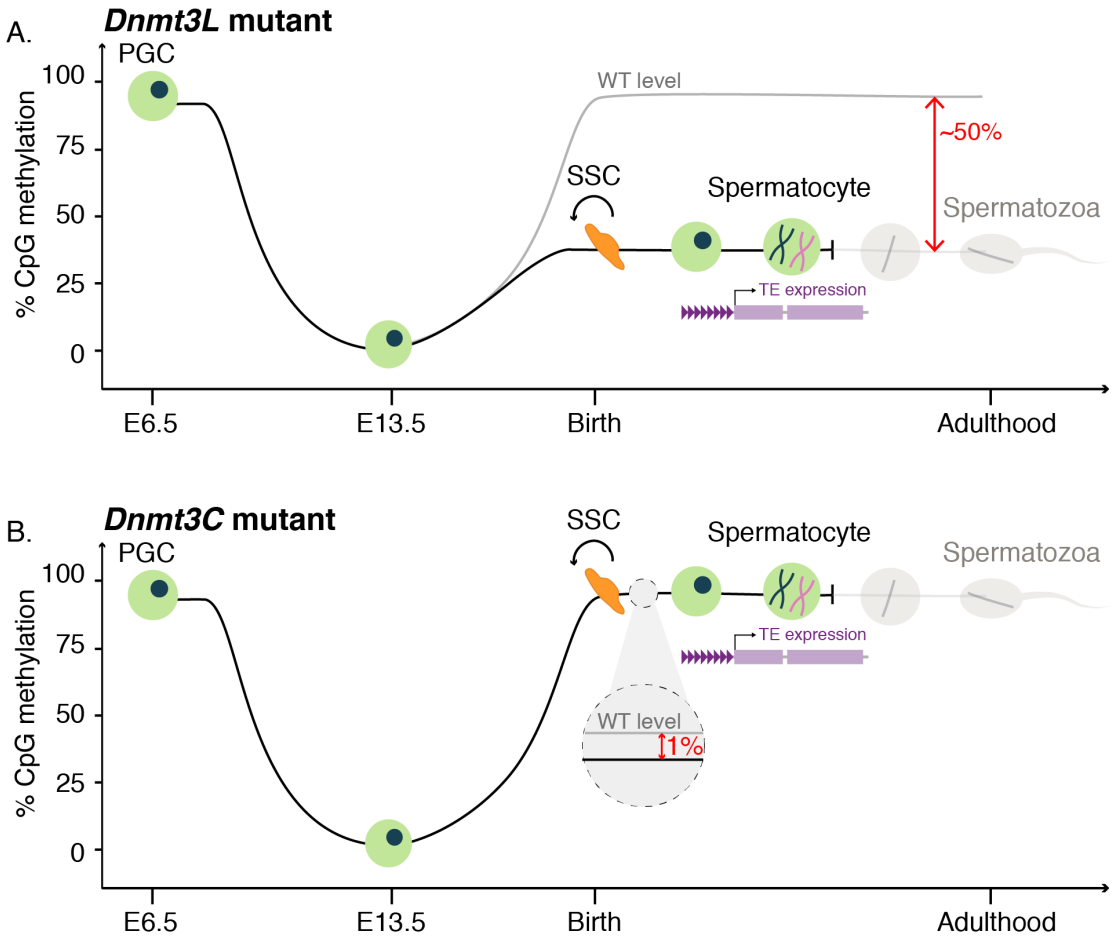
In mice, pre-pachytene piRNAs are required to guide *de novo* DNA methylation on young TEs, as shown from the lack of DNA methylation at young TE promoters in *Mili* or *Miwi2* mutants (Aravin et al., 2008; Carmell et al., 2007). Upon a successful wave of DNA methylation erasure in PGCs, transcriptionally-competent TEs can get reactivated and expressed in prospermatogonia. The production of piRNAs is a complex process that relies on two mechanisms: the “ping-pong” amplification cycle in which TE transcripts are cleaved by PIWI proteins in the cytoplasm upon targeting by a complementary piRNA, thus producing piRNAs that can in turn target piRNA precursor transcripts to produce more piRNAs in an amplification loop; and phasing in which piRNA precursors are cleaved repeatedly by an endonuclease, producing piRNAs that are loaded into nuclear PIWI proteins (Ozata et al., 2019). Phased piRNAs loaded into MIWI2 feed back into the nucleus where they act as recognition devices against complementarity sequences, likely acting on nascent TE transcripts. Then, MIWI2-loaded TE-derived piRNAs somehow drives the DNA methylation machinery to the promoter of TEs (Chuma and Nakano, 2013) (Fig. 15). However, the precise

interaction and relationship between the piRNA pathway and DNA methylation is not fully understood and will be discussed in the next sub-chapter.

- *DNMT3C*

DNMT3C was discovered in 2016 by two independent mouse genetic screens, including one from our lab (Barau et al., 2016; Jain et al., 2017). Previously annotated as a pseudogene, *Dnmt3C* originated by tandem duplication of *Dnmt3B* in the Muroidea lineage. DNMT3C is *de novo* DNMT required to target and repress TEs in the male germline of these species. DNMT3C is expressed exclusively in fetal male germ cell from E14.5 to birth, with a peak between E16.5 and E18.5. It is essential for male germ cell development, the deletion of DNMT3C inducing complete male sterility in mice. More precisely, spermatogenesis is arrested at the pachytene stage of meiosis, in association with extensive reactivation of young TEs, including L1A, L1Tf and IAPs, which is reminiscent of mutants of the piRNA pathway (Aravin et al., 2008; Barau et al., 2016; Carmell et al., 2007). Interestingly, DNMT3C is not essential for the female germline and in absence of DNMT3C, females are fertile (Barau et al., 2016). This is also the case for mutants of the piRNA pathway (Carmell et al., 2007; Kuramochi-Miyagawa et al., 2004).

Genetic evidence suggests that DNMT3C is stimulated by the co-factor DNMT3L: the deletion of DNMT3L leads to the same developmental phenotype as DNMT3C deletion, namely pachytene arrest during spermatogenesis with reactivation of the same young TE families (Barau et al., 2016; Bourc'his and Bestor, 2004) (Fig. 15). However, whole genome bisulfite sequencing (WGBS) of *Dnmt3C* and *Dnmt3L* mutant spermatogonia at 10dpp—after the wave of *de novo* DNA methylation—



C. Male germ cell methylomes (WGBS)

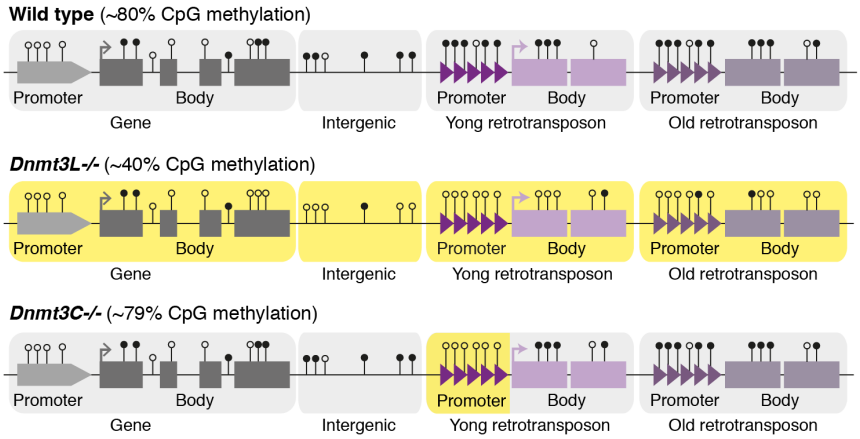


Figure 16
A and **B**. Developmental and molecular phenotype upon *Dnmt3L* and *Dnmt3C* deletion compare to WT condition (light grey).
C. Drawing showing the precise DNA methylation defect on the different compartment of the genome upon *Dnmt3L* and *Dnmt3C* deletion compare to WT condition (top raw).

revealed distinct molecular phenotypes. Whilst the *Dnmt3L* mutant showed ~40% of CpG methylation genome-wide compared to ~80% in age-matched wildtype germ cells, the *Dnmt3C* mutant demonstrated ~79% of CpG methylation (Fig. 16). In the *Dnmt3L* mutant, all compartments of the genome were affected by the DNA methylation loss, including TEs, old and young, their body and their promoters. Conversely, the *Dnmt3C* mutant revealed a loss of DNA methylation exclusively on the

promoters of young TEs. Overall, DNMT3C is considered to be essential for male germ cell development, despite methylating only 1% of the genome (Barau et al., 2016) (Fig. 16).

The question of how DNMT3C can be so specific to the promoters of young TEs is not yet resolved. Many groups are aiming to understand the underlying mechanism of DNMT3C-dependent targeting of DNA methylation in mice, and this is a project I have myself been following during my PhD. Two main directions are under exploration, and these are not mutually exclusive: the role of piRNAs in guiding DNMT3C to young TEs and the role of the intrinsic protein structure of DNMT3C.

- *Mechanism 1: DNMT3C is guided by the piRNA pathway*

MIWI2 being both nuclear and cytoplasmic, this is the known piRNA component that is the epistatically closest to DNMT3C. However, the biochemical nature of the recruitment of DNMT3C by MIWI2 is largely unknown. The interaction between DNMT3C and MIWI2 could be either direct or indirect via a protein complex, or via a chromatin remodeler able to confer a chromatin state suitable to recruit DNMT3C. Nevertheless, some important genetic insights exist: mutations of several genes encoding nuclear proteins have been shown to exactly phenocopy the absence of DNMT3C or MIWI2 proteins: pachytene interruption, DNA methylation defect at young TEs promoters and reactivation of these elements. The DNA methylation defect also includes the paternally imprinted *Rasgrf1* locus, which is known to be dependent on the piRNA pathway and DNMT3C, because of the presence of a TE in the locus (Watanabe et al. 2011). Contrary to *Miwi2* mutants but similarly to *Dnmt3C* mutants, none of these new mutants of TE reactivation have a defect in piRNA biogenesis. They therefore operate downstream of piRNAs /MIWI2 but upstream of DNA methylation/DNMT3C. Among these proteins, I will describe three interesting candidates, TEX15 (Testis Expressed 15) , SPOCD1 (SPOC Domain Containing 1) and MORC1 (MORC Family CW-Type Zinc Finger 1) (Pastor et al., 2014; Schöpp et al., 2020; Zoch et al., 2020)

SPOCD1 was recently identified by immunoprecipitation of MIWI2 followed by mass spectrometry (IP-MS) in male fetal gonads (Zoch et al., 2020), indicating that MIWI2 and SPOCD1 are part of a complex, with a direct or indirect interactions.

Curiously, IP-MS of SPOCD1 itself revealed interactions with DNMT3L and DNM3A, but not DNMT3C. Based on the genetic evidence that suggests SPOCD1 is connected to TE methylation in prospermatogonia, it is perhaps surprising to see no association with DNMT3C. This lack of association could be explained by the interaction being indirect, and not measurable in the IP-MS. Also, the study may have limitations in detecting DNMT3C, which is lowly expressed compared to other DNMT3s in prospermatogonia. However, upon expression of DNMT3C and SPOCD1 in HEK cells, co-immunoprecipitation of SPOCD1 revealed an interaction with DNMT3C. The molecular function of SPOCD1 is still unknow but it bears an interesting domain, the SPOC domain. The SPOC domain is known to have the potential to interact with the transcriptional co-repressors NCoR-SMRT, a known component of the histone deacetylation complex (Mikami et al., 2014). Overall, SPOCD1 may be part of a large complex in prospermatogonia, with MIWI2, chromatin remodelers and potentially, *de novo* DNMTs.

TEX15, a protein of unknown function, was also identified through a MIWI2-centered IP-MS in male fetal gonads (Schöpp et al., 2020). The association between TEX15 and MIWI2 seems to occur in the context of chromatin only. Moreover, TEX15 does not interact with SPOCD1, raising the question of a potential parallel function or a distant position within the same pathway.

Finally, MORC1 is another interesting protein to speculate about the link between the piRNA pathway and DNA methylation (Pastor et al., 2014). The MORC family of proteins are described as epigenetic regulators and remodelers. MORC1 has a PHD-X/ZF-CW domain whose function is to bind H3K4me3 and an ATPase domain able to compact the chromatin (Li, Nair, and Kumar 2013). At E14.5 to E16.5 stages of development, TEs are expressed and therefore enriched in H3K4me3 (Yamanaka et al., 2019), a mark that is not appealing to the ADD domain of *de novo* DNMT3s (Ooi et al., 2007). Via its PHD-X/ZF-CW domain, MORC1 could bind to H3K4me3-decorated TEs and at same time reduce H3K4me3 via its ATPase activity or help recruiting on site histone modifiers such as SETDB1, to catalyze H3K9me3. This scenario could induce a chromatin switch on TEs from H3K4me3 “activated” to H3K9me3 “repressed”, which may in turn allow the recruitment of *de novo* DNMT3s (see previous Chapter “Positive correlation of DNA methylation and H3K9me3). In *Drosophila*, it is well established the piRNA pathway guides H3K9me3 towards TEs and this mark is sufficient for permanent repression in this DNA methylation-free

organism (Sienski et al., 2012; Wang and Elgin, 2011). In mice, it is still not clear whether H3K9me3 is guided by the piRNA pathway or not. Alternatively, DNMT3C could also directly interact with MORC, which could force its recruitment to H3K4me3-enriched TEs, without H3K9me3 intermediate.

It is not known how the piRNA pathway and DNA methylation are linked through MORC1, but it is probable that the mechanism underlying this interaction requires a large complex, including many proteins with different functions in parallel.

In summary, more and more proteins have been identified as potential players in the crosstalk between piRNAs and DNMT3C. However, it is still unclear how and in which order they assemble, and whether the recruitment of DNMT3C by MIWI2 is direct or through the intermediate establishment of a favorable chromatin state, or both.

- *Mechanism 2: DNMT3C structure*

As described earlier, while DNMT3C is similar by 70% with DNMT3B, it strikingly lacks a PWWP domain. This domain recognizes H3K36me2/3, and is used by DNMT3B and DNMT3A to methylate intragenic and intergenic regions. We could speculate that by not binding to H3K36me2/3 (which covers an important part of the genome), DNMT3C would have more availability to focus on young TE promoters. Adding an artificial PWWP domain to DNMT3C would allow addressing this hypothesis, a direction that will be discussed in more detail in the Discussion part of this thesis manuscript.

Moreover, the N-terminal domain of DNMT3C also has properties that could explain its specificity. It has been well-described that TEs and the host repression mechanisms that evolved are in constant competition, either to escape repression (for the TEs) or to create new repression mechanisms (for the host). This battle is referred to as “evolutionary arms race” and the genes implied in this battle can be detected because they are under positive evolutionary selection (Molaro et al., 2020). Interestingly, it has been shown that four amino acid of the N-terminus of DNMT3C, but not DNMT3B, underwent strong diversifying selection in rodent species, thereby indicating that DNMT3C N-terminus is likely involved in an ongoing genetic conflict against TEs (Molaro et al., 2020). This last part, will be discuss as well in the Discussion part of the manuscript.

4.2.2 DNA methylation establishment on the rest of the genome

Establishment of DNA methylation in male germ cells was originally suggested in 2004 to be dependent on DNMT3A and DNMT3L, at all genomic compartments. At this time, DNMT3C had not been discovered yet (Bourc'his and Bestor 2004; Kaneda et al. 2004). As a matter of fact, a *Dnmt3A* germline conditional mutant, driven by *Tnap*-Cre, revealed male complete sterility (Kaneda et al., 2004), while the germline conditional deletion of *Dnmt3B* had not effect on male germ cell development and fertility (Kaneda et al., 2004). This data suggested no function for DNMT3B in re-shaping the male germline methylome, despite its high expression in prospermatogonia.

However, the precise arrest of spermatogenesis in the *Dnmt3A* mutant was controversial. Some studies claimed that *Dnmt3A* deletion phenocopied *Dnmt3L* deletion—interruption of spermatogenesis at the pachytene stage of meiosis—implying a potential role in TE methylation (Kaneda et al., 2004; Kato et al., 2007). However, another study reported that *Dnmt3A* mutant spermatogenesis could progress at least until the round spermatid stage, therefore past meiosis (Yaman and Grandjean, 2006). In summary, the developmental arrest of spermatogenesis was uncertain and left room to investigate the exact cause of *Dnmt3A* mutant mice sterility.

Additionally, at the molecular level, the precise targets of DNMT3A in male germ cells were not fully known. Using targeted methods, DNMT3A was found to be required for methylating paternally imprinted genes and SINE elements (Kaneda et al., 2004; Kato et al., 2007). However, when considering young TEs (L1s and IAPs), results were inconclusive and the lack of consistent DNA methylation loss at these sequences in absence of DNMT3A was interpreted as a sign of functional redundancy between DNMT3A and DNMT3B (Kaneda et al., 2004; Kato et al., 2007). Finally, there was no information on the particular case of the paternal *Rasgrf1* imprint, which behaves as young TEs in terms of DNA methylation targeting. Overall, the molecular targets of DNMT3A have never been meticulously described and thoroughly assessed by genome-wide techniques. In particular, concerning TEs, the discovery of DNMT3C opened the question as to whether DNMT3A participates or not in TE re-methylation during male germline development.

Whatever enzyme is responsible for methylating the genomic bulk in prospermatogonia, the mechanism seems considerably simpler compared to TE promoters. A recent paper demonstrated that DNA methylation is guided genome-wide

by H3K36me2 in prospermatogonia (Shirane et al., 2020). Germline conditional deletion (using *Tnap* -Cre) of NSD1, an enzyme that catalyzes H3K36me2, induces an extensive decrease of H3K36me2 in prospermatogonia, and in turn, DNA hypomethylation at all genomic compartments, including paternal imprints, but not young TE promoters and the imprinted *Rasgrf1* locus (Shirane et al., 2020). This result shows, one more time, that DNA methylation recruitment occurs through distinct mechanisms on young TE promoters versus the rest of the genome.

Interestingly, while *Nsd1* and *Dnmt3L* mutants share similar DNA methylation defects and a male sterility phenotype, the developmental details of this sterility are more severe in *Nsd1* mutants than in *Dnmt3L* mutants, with the number of fetal prospermatogonia being specifically reduced in *Nsd1* mutants. Molecularly, the absence of a genome-wide H3K36me2 blanket in *Nsd1* mutants leads to H3K27me3 spreading genome-wide, while this is not the case in *Dnmt3L* mutants. This ectopic spreading of H3K27me3 induces repression of important genes and interferes with proper transcriptional regulation in *Nsd1* mutant prospermatogonia (Shirane et al., 2020). As we already discuss that DNMT3A has a greater affinity *in vitro* for H3K36me2 than H3K36me3 (Weinberg et al., 2019), therefore we could hypothesize that DNMT3A could be the enzyme responsible to establish DNA methylation down stream of NSD1 (Fig. 15).

RESULTS

1 DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis

In the first part of my PhD work, I investigated the genomic targets of DNMT3A and its associated developmental function in spermatogenesis. I demonstrated that DNMT3A methylates the entire genome, except the promoter of the young transposable elements that are selectively targeted by DNMT3C. Moreover, I showed that DNMT3A was essential for SSC differentiation. *Dnmt3A* mutant SSCs can only self-renew and no longer differentiate due to spurious enhancer activation that enforces an irreversible stem cell gene program.

The presented data were compiled mostly by myself, Joan Barau (ex-postdoctoral fellow in the lab, now PI at the IMB Mainz) produced the whole-genome DNA methylation maps in the prospermatogonia of various *Dnmt3* mutants. I have also oriented and supervised all the bioinformatic analyses carried out by our bioinformatician Aurélie Teissandier.

This study is presented in the manuscript format that was submitted and is currently in revision in the journal *Nature Genetics*.

DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis

Mathilde Dura[1], Aurélie Teissandier[1], Mélanie Armand[1], Joan Barau[2], Lorraine Bonneville[1], Michael Weber[3], Laura G. Baudrin[4], Sonia Lameiras[4] and Deborah Bourc'his[1*]

[1] Genetics and Developmental Biology Department, Institut Curie, PSL Research University, INSERM, CNRS, Paris, France

[2] Institute of Molecular Biology (IMB), Mainz, Germany

[3] Biotechnology and Cell Signaling, University of Strasbourg, CNRS, Illkirch Cedex, France

[4] ICGex Next-Generation Sequencing platform, Institut Curie, PSL Research University, 75005 Paris, France

*corresponding author: deborah.bourchis@curie.fr

Abstract

DNA methylation plays a critical role in spermatogenesis, as evidenced by the male sterility of DNA methyltransferase (DNMT) mutant mice. Here, we report a striking division of labor in the establishment of the methylation landscape of male germ cells and its functions in spermatogenesis: while DNMT3C is essential for preventing retrotransposons from interfering with meiosis, DNMT3A broadly methylates the genome—at the exception of DNMT3C-dependent retrotransposons—and controls spermatogonial stem cell (SSC) plasticity. By reconstructing developmental trajectories through single-cell RNA-seq and by profiling chromatin states, we found that *Dnmt3A* mutant SSCs can only self-renew and no longer differentiate due to spurious enhancer activation that enforces an irreversible stem cell gene program. We therefore provide a novel function for DNA methylation in male fertility: the epigenetic programming of SSC commitment to differentiation and to life-long spermatogenesis supply.

Introduction

Cytosine DNA methylation is an epigenetic mark that is crucial for proper mammalian development. Promoter methylation provides stable and long-term repression, with little variation in patterns across somatic tissues. This regulatory mode mostly applies to retrotransposon control, and to the constitutive repression of a small subset of genes, such as germline genes, and genes subject to genomic imprinting or X chromosome inactivation[1,2]. By contrast, enhancer methylation is prevalent and dynamic, across tissues and developmental stages[3–5]. The identification of methyl-sensitive transcription factors (TFs) in biochemical assays has provided a conceptual frame for the function of enhancer DNA methylation[6,7]. However, there is still limited *in vivo* evidence that DNA methylation modulates gene programs by limiting TF-enhancer functional interactions during differentiation processes.

Male germline differentiation provides a highly relevant context to study the breadth of DNA methylation distribution and function. Germline epigenetic reprogramming produces an extensively hypomethylated genome, onto which male germ cell-specific DNA methylation is established during fetal life, and impacts all genomic compartments: genes, intergenic sequences and retrotransposons. Remethylation occurs prior to the formation of spermatogonial stem cells (SSCs), which sustain life-long spermatogenesis through their dual capacity to self-renew and differentiate[8]. Incidentally, DNA methylation patterns established in fetal germ cells are mostly unaltered in post-natal life, and are propagated from SSCs to the successive differentiating types that lead to spermatozoa production[9].

Male germline methylation requires two *de novo* methyltransferases in mice, DNMT3A and DNMT3C, and a catalytically inactive co-factor, DNMT3L[1]. Individual mutations in these genes consistently lead to male sterility, highlighting the key role of DNA methylation for spermatogenesis[10–12]. The recently discovered DNMT3C enzyme selectively methylates the promoters of young retrotransposon lineages, which represents only 1% of the mouse genome[12]. Nevertheless, this restricted function is absolutely essential for meiosis: failure to

60 establish retrotransposon methylation in fetal stages results in their post-natal activation at
61 meiosis, subsequent perturbation of the meiotic chromatin landscape and spermatogenic
62 interruption by apoptosis, as observed in *Dnmt3C* and *Dnmt3L* mutants[12–14]. Prior to the
63 identification of DNMT3C, DNMT3A was regarded as the sole enzyme responsible for male
64 germline methylation, including at retrotransposons. This idea is still persistent although
65 DNMT3A targets have never been profiled genome-wide during germline reprogramming and
66 the nature of the spermatogenetic impairment of *Dnmt3A* mutants remains uncertain[11,15,16].
67 Here, we demonstrate that DNMT3A function in male germline development is not related to
68 retrotransposon control and meiosis protection. Rather, DNMT3A-dependent DNA
69 methylation pre-emptively programs the capacity of spermatogonial stem cells to commit to
70 spermatogenetic differentiation after birth, by limiting aberrant enhancer activity associated
71 with stem cell identity.

Results

DNMT3A broadly methylates the genome in fetal male germ cells

DNMT3A was previously shown to methylate paternally imprinted genes in male germ cells[11,16]. However, contrary to the DNMT3L co-factor and the retrotransposon-specific DNMT3C enzyme, there is currently no genome-wide maps of DNMT3A targets during male germ cell development. We therefore performed Whole-Genome Bisulfite Sequencing (WGBS) on DNA extracted from wildtype (WT) and *Dnmt3A* mutant (*Dnmt3A* knockout, *Dnmt3A*^{KO}) prospermatogonia at embryonic day 18.5 (E18.5) (**Fig. 1a** and **Supplementary Table 1**). Prospermatogonia were isolated by FACS using the *Oct4-eGFP* transgenic line[17] and libraries were generated by post-bisulfite adaptor tagging (PBAT)[18]. We included age-matched *Dnmt3L*^{KO} and *Dnmt3C*^{KO} prospermatogonia to allow for direct comparison, as previous whole-genome methylation profiles for these mutants were generated in postnatal germ cells[12]. We chose the E18.5 time-point because *Dnmt3A, 3L* and *3C* genes are upregulated[12,19] and *de novo* DNA methylation of male germ cells is still ongoing yet close to completion (**Fig. 1a**), avoiding potential compensatory or secondary effects that could occur later, in postnatal life.

Global CpG methylation levels were dramatically reduced in *Dnmt3A*^{KO} compared to WT prospermatogonia, with mean values dropping from 67.7% to 20.2%, in an extent similar to *Dnmt3L*^{KO} (18.7%) (**Fig. 1b**). We identified 555,893 differentially methylated regions (DMRs), all reflecting hypomethylation in *Dnmt3A*^{KO}. By contrast, levels were barely diminished in *Dnmt3C*^{KO} (64.5%)—matching previous reports in postnatal germ cells[12]—with a number of 7,620 DMRs (all hypomethylated compared to WT) that did not overlap with *Dnmt3A*^{KO} DMRs. DNMT3A appeared necessary for methylating all genomic compartments, including genes, intergenic sequences, and transposable elements, akin to DNMT3L (**Fig. 1b** and **Extended data Fig. 1a**). Both LTR (ERV1, ERVK and ERVL) and non-LTR (LINE1 and SINE) retrotransposons were globally hypomethylated in *Dnmt3A*^{KO}, with a reduction 2- to 3-fold greater than observed in *Dnmt3C*^{KO}, even at ERVK and LINE1 which are preferential

Figure 1

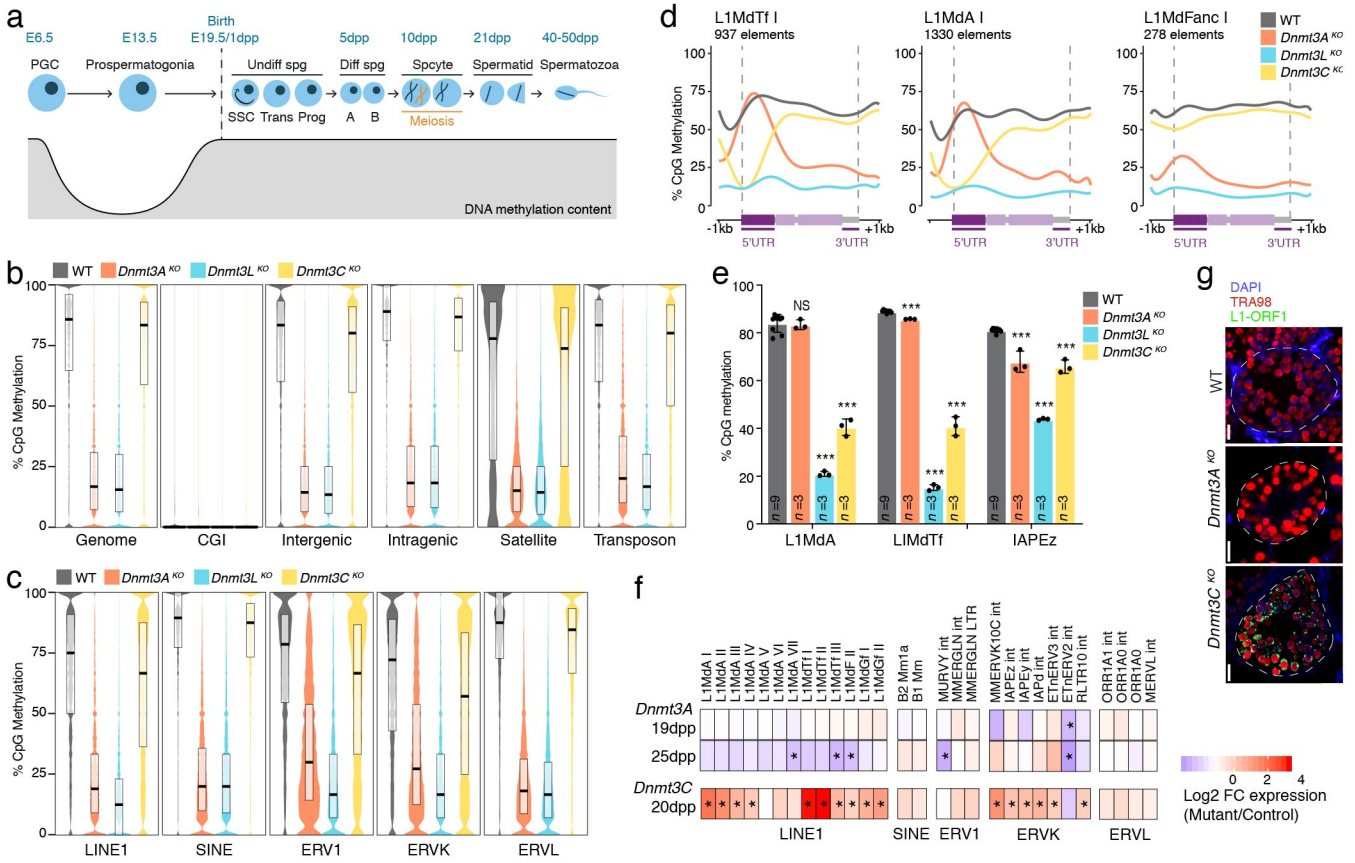


Fig. 1 | DNMT3A broadly methylates the male germ cell genome but is dispensable for retrotransposon silencing. a, Schematic representation of developmental and DNA methylation dynamics of the male germline in the mouse. Postnatal ages indicate predominant germ cell types present in the testis during the first wave of spermatogenesis. PGC: primordial germ cell, SSC: spermatogonial stem cell, Spg: spermatogonia, Trans: transitory, Prog: progenitors. **b**, **c**, Violin plot representation of CpG methylation content over the whole genome and different genomic compartments (b) and over retrotransposon families (c) in WT (grey), *Dnmt3A*^{KO} (orange), *Dnmt3L*^{KO} (blue) and *Dnmt3C*^{KO} (yellow) fetal germ cells (E18.5), as determined by WGBS. Black horizontal bars represent the median, upper and lower hinges correspond to 75 and 25% quantile. **d**, Metaplots of DNA methylation levels over uniquely assigned full-length elements (>5kb) from three L1 families. **e**, CpG methylation levels assessed by bisulfite-pyrosequencing of the promoters of indicated retrotransposon families in 10dpp sorted germ cells. Data are mean ± SD (black bar) from biological replicates, black dots represent biological replicates, n= number of animals (student t-test over WT, *p<0.05, **p<0.005, ***p<0.0005). **f**, RNA-seq heatmap shows log2-fold change (FC) in retrotransposon mRNA levels in *Dnmt3A*^{KO} versus WT littermates at 19 and 25dpp (top and middle rows), and previously published changes in *Dnmt3C* mutants versus heterozygous littermates at 20dpp (bottom row)[12]. Annotations are from RepeatMasker. Two biological replicates were sequenced per age and genotype. **g**, L1-ORF1 immunostaining on testis sections at 19dpp in WT, *Dnmt3A*^{KO} and *Dnmt3C*^{KO} germ cells (TRA98-positive). White dotted lines: tubule delineation. Scale, 20 µm.

targets of DNMT3C[12] (**Fig. 1c**). However, metaplot analysis of uniquely mapped copies from individual retrotransposon families revealed a striking inverted pattern of defective DNA methylation in *Dnmt3A*^{KO} and *Dnmt3C*^{KO} germ cells. When considering evolutionarily young L1 families (L1Md-Tfl and L1Md-Al), methylation of the promoters of these elements was confirmed to be DNMT3C-dependent uniquely, while DNMT3A was required for methylating the entire body of these same elements (**Fig. 1d** and **Extended data Fig. 1a**). In evolutionarily old L1s (L1Md-FancI), DNMT3A was responsible for methylating the full element length, while DNMT3C was generally dispensable. This situation was similar for ERVL elements (**Extended data Fig. 1b**), while uniquely assigned ERVK elements belonging to the young IAPEz family relied on DNMT3C only (**Extended data Fig. 1c**). In contrast with these specificities, DNMT3L had a global role in assisting *de novo* DNA methylation in fetal male germ cells: *Dnmt3L*^{KO} prospermatogonia displayed hypomethylation over both DNMT3A and DNMT3C genomic targets (**Fig. 1d** and **Extended Fig. 1a-c**). Overall, we reveal here that DNMT3A is a largely indiscriminate enzyme that *de novo* methylates the whole genome of prenatal male germ cells, with the notable exception of young retrotransposon promoters (**Extended data Fig. 1d**).

DNMT3A does not silence retrotransposons during spermatogenesis

Dnmt3A^{KO} fetal germ cells display normal methylation at young retrotransposon promoters, which implies proper silencing. However, *Dnmt3A* mutants were previously reported to phenocopy *Dnmt3L* mutants[11], in which retrotransposon reactivation culminates after birth in meiotic cells, in association with apoptosis and spermatogenesis interruption[13,14]. We therefore went on to verify whether defective retrotransposon methylation and silencing could occur postnatally in *Dnmt3A*^{KO} germ cells, around meiosis (**Fig. 1a**). Using targeted bisulfite pyrosequencing, we measured promoter DNA methylation of young retrotransposons (L1MdA, L1MdTf and IAPEz) in FACS-sorted germ cells (EpCAM positive; β2-Microglobulin negative) from males at 10 days post-partum (10dpp). At young L1 promoters—similarly to fetal stages—*Dnmt3A*^{KO} displayed normal methylation, while *Dnmt3C*^{KO} and *Dnmt3L*^{KO}

showed decreased CpG methylation (**Fig. 1e**). We observed a slight decrease on IAPEz promoters in *Dnmt3A*^{KO} (67.7% versus 81.1% in WT), indicating that DNA methylation of a subset of IAP copies may require DNMT3A.

However, no IAP or L1 reactivation was detected in testes of *Dnmt3A*^{KO} males (between 19 and 25dpp), at the RNA level using RNA-seq (**Fig. 1f**) and RT-qPCR (**Extended data Fig. 1e**), and at the protein level by immunodetection of L1-encoded ORF1 proteins (**Fig. 1g**), in striking contrast to *Dnmt3C*^{KO} and *Dnmt3L*^{KO} males. Moreover, SINE and ERVL, which are exclusively methylated by DNMT3A (**Fig. 1c**), also maintained repression in *Dnmt3A*^{KO} testes (**Fig. 1f** and **Extended data Fig. 1e**). Lack of retrotransposon reactivation did not reflect a lack of meiotic cells: meiotic genes showed similar mRNA levels in *Dnmt3A* mutants compared to *Dnmt3C* mutants (**Extended data Fig. 1f**). As a whole, these findings definitely exclude a role for DNMT3A in silencing retrotransposons during spermatogenesis. They also allude to the possibility that although they all share a sterility phenotype, the etiology of this sterility may be different in *Dnmt3A* mutants compared to *Dnmt3C* and *Dnmt3L* mutants.

Dnmt3A mutants only progress through the first wave of spermatogenesis

To uncover the function of DNMT3A-dependent DNA methylation in spermatogenesis, we phenotyped *Dnmt3A* mutant testes across ages. Constitutive *Dnmt3A*^{KO} animals are developmentally delayed after birth and die around 25dpp[20] (**Extended data Fig. 2a**). All analyses past 25dpp were therefore performed on germ-cell conditional *Dnmt3A*^{KO} by crossing the *Dnmt3A2lox* line with the *Prdm1-Cre* line, which promotes recombination at E9.5[21] (*Prdm1-Dnmt3A*^{cKO}) (**Extended data Fig. 2b**).

From 10dpp to 6 months, *Dnmt3A* mutant males showed significant and increasing reduction in testis weight and seminiferous tubule surface in comparison to their WT littermates (**Extended data Fig. 2c,d**). We then performed histological assessment of testis sections at 19dpp, 6 weeks, 8-9 weeks and 6 months, and compared to age-matched *Dnmt3L*^{KO} and *Dnmt3C*^{KO} males (**Fig. 2a** and **Extended data Fig. 2e**). As previously

Figure 2

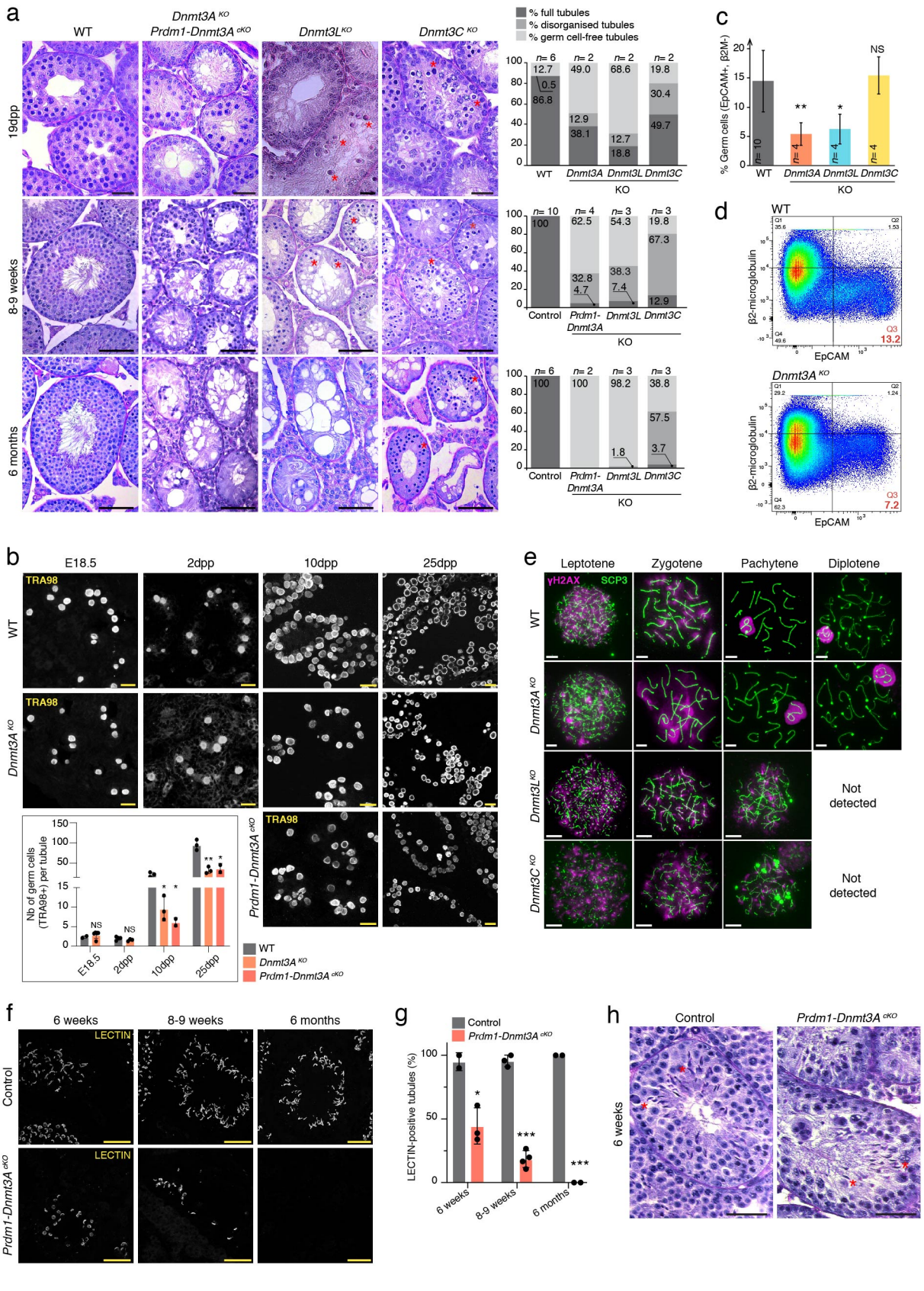


Fig. 2 | Dnmt3A mutants males complete the first wave of spermatogenesis with reduced germ cell numbers. a, (Left) Representative images of Periodic Acid Shift (PAS)-stained testis sections of different genotypes at 19dpp (scale, 20 µm), 8-9 weeks and 6 months (scale, 50 µm). Red stars (*) point apoptotic cells. (Right) Quantification of the percentage of different classes of tubules per genotype. n= number of animals. Control genotypes at 8-9 weeks were *Dnmt3A*$^{2lox/KO}$; *Prdm1-Cre*$^{0/0}$ (n= 2), *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$, *Dnmt3A*$^{2lox/WT}$; *Prdm1-Cre*$^{0/0}$, *Dnmt3L*$^{KO/WT}$ (n= 3) and *Dnmt3C*$^{KO/WT}$ (n= 3). Control genotype at 6 months were *Dnmt3A*^{WT} (n= 4), *Dnmt3A*$^{2lox/WT}$; *Prdm1-Cre*$^{0/0}$ and *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$. **b**, (Top) Representative image of TRA98 staining on testis sections from WT, *Dnmt3A*^{KO} and *Prdm1-Dnmt3A*^{cKO} males at E18.5, 2dpp, 10dpp and 25dpp (scale, 20 µm). (Bottom) Quantification of TRA98-positive cells normalized per tubule. Data are mean ± SD (black bar) and individual points represent biological replicates (student t-test over WT *p<0.05, **p<0.005). **c**, Percentage of live germ cells (DAPI-Neg, EpCAM-Pos, β2M-Neg) per testis of different genotypes at 10dpp, assessed by FACS. Data are mean ± SD (black bar) from at least four biological replicates, n= number of animals (student t-test over WT, NS= nonsignificant, *p<0.05, **p<0.005). **d**, Representative FACS plots of live EpCAM-Pos, β2M-Neg gated testicular cells.

from 10dpp *Dnmt3A*^{KO} and WT males. Cell percentages are indicated in each quarter, with percentage of EpCAM-Pos, β2M-Neg cells in red. **e**, Representative microscopy images of SCP3 (green) and γH2AX (pink) immunodetection on WT, *Dnmt3A*^{KO}, *Dnmt3L*^{KO} and *Dnmt3C*^{KO} meiotic spreads (16-25dpp). Diplotene stages were absent in *Dnmt3L*^{KO} and *Dnmt3C*^{KO}. Scale, 5 µm. **f**, Representative image of LECTIN (acrosome/haploid marker) staining on testis sections from *Prdm1-Dnmt3A*^{cKO} and controls at indicated ages. Scale, 50 µm. **g**, Quantification of tubule percentage with LECTIN-stained cells. Control genotypes at 6 weeks: *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$ and *Dnmt3A*$^{2lox/KO}$; *Prmd1-Cre*$^{0/0}$ - at 8-9 weeks: *Dnmt3A*$^{2lox/KO}$; *Prmd1-Cre*$^{0/0}$ (n= 2) and *Dnmt3A*$^{2lox/WT}$; *Prdm1-Cre*$^{0/0}$ - at 6 months: *Dnmt3A*$^{KO/WT}$; *Prmd1-Cre*$^{Tg/0}$ and *Dnmt3A*$^{2lox/WT}$; *Prmd1-Cre*$^{0/0}$. Data are mean ± SD (black bar) and individual points represent biological replicates (student t-test over WT *p<0.05, **p<0.005, ***p<0.0005). **h**, Representative images of PAS-stained testis sections of 6 week-old *Prdm1-Dnmt3A*^{cKO} male and control littermate (*Dnmt3A*$^{2lox/KO}$; *Prdm1-Cre*$^{0/0}$). Red stars (*) point spermatozoa heads. Scale, 30 µm.

reported[12,13], *Dnmt3L*^{KO} and *Dnmt3C*^{KO} tubules exhibited pyknotic nuclei from apoptotic meiotic cells, and lack of subsequent post-meiotic stages. In contrast, we did not detect apoptotic cells in *Dnmt3A* mutant testes, indicating that spermatogenesis might not be interrupted at meiosis. Then, while *Dnmt3C*^{KO} tubules continuously produced pre-meiotic cells, *Dnmt3L*^{KO} showed progressive spermatogenic decline with age, with only spermatogenesis-free tubules remaining at 6 months. Overall, *Dnmt3A*^{KO} and *Prdm1-Dnmt3A*^{cKO} mutants displayed a *Dnmt3L*^{KO}-like phenotype of progressive loss of spermatogenic ability (**Fig. 2a**).

To pinpoint the onset of spermatogenic failure, *Dnmt3A* mutant germ cells were counted across the first wave of spermatogenesis (around birth time to 6 weeks) (**Fig. 1a**), by immunodetection of the pan germ cell marker TRA98 (**Fig. 2b**). No difference was scored compared to WT immediately before (E18.5) and after birth (2dpp): *Dnmt3A*^{KO} males were therefore born with appropriate numbers of prospermatogonia, from which the postnatal spermatogonial contingent emerges. However, starting at 10dpp, germ cell depletion became significant, with a 2- to 3-fold reduction in average numbers per tubule compared to age-matched WT. This was confirmed by FACS analysis (**Fig. 2c, d**): 5% of testicular cells were scored as “EpCAM-pos; β2M-neg” germ cells in 10dpp-old *Dnmt3A*^{KO} males against 15% in WT. This reduction rate was similar in *Dnmt3L*^{KO} testes while *Dnmt3C*^{KO} had normal germ cell counts. Importantly, germ cell depletion was an intrinsic cell defect, being also observed in germ cell-specific *Prdm1-Dnmt3A*^{cKO} mutants (at 10dpp, **Fig. 2b**).

Consistent with the lack of apoptotic cells in histological sections (**Fig. 2a**), prepuberal *Dnmt3A*^{KO} males completed the first prophase of meiosis: diplotene figures were readily detectable upon SCP3 and γH2AX labeling of meiotic chromosome spreads around 20dpp (**Fig. 2e** and **Extended data Fig. 2f, g**). This was in stark contrast to *Dnmt3C*^{KO} and *Dnmt3L*^{KO} males where spermatogenesis was arrested prior to pachytene (**Fig. 2e**). RT-PCR detection of haploid cell markers further indicated that spermatids were specified in constitutive *Dnmt3A*^{KO} males at 25dpp (**Extended data Fig. 2h**). Lectin-mediated staining of acrosome structures confirmed the presence of round and elongated spermatids in testes of

187 *Prdm1-Dnmt3A*^{cKO} males at 6 weeks—with half numbers of lectin-positive tubules compared
188 to controls—(**Fig. 2f, g**) and morphologically mature spermatozoa were detected (**Fig. 2h**
189 and **Extended data Fig. 2i**). However, as described above (**Fig. 2a**), pre- and post-meiotic
190 germ cells eventually disappeared from *Prdm1-Dnmt3A*^{cKO} testes, upon initiation of
191 subsequent rounds of spermatogenesis, and were visibly absent at 6 months (**Fig. 2f,g** and
192 **Extended data Fig. 2f**).

193 In sum, contrary to *Dnmt3C* and *Dnmt3L* mutants, *Dnmt3A* mutants—both
194 constitutive and germ cell-conditional—can progress past meiosis. This likely relates to the
195 lack of retrotransposon reactivation in *Dnmt3A* mutants. However, *Dnmt3A* mutants can only
196 fulfill the first postnatal wave of spermatogenesis, after which they lose spermatogenic
197 potential. Interestingly, the first cells that enter spermatogenesis—3 to 5 days after birth—
198 emanate from a pool of fetal prospermatogonia that bypass the spermatogonial stem cell
199 (SSC) state and directly transition to differentiated spermatogonia[22] (**Fig. 1a**). Meanwhile, the
200 remainder of prospermatogonia generate the foundational SSC compartment, from which
201 spermatogenesis initiates throughout the reproductive lifespan[23]. The phenotypic features of
202 *Dnmt3A* mutants suggests that SSC-dependent spermatogenesis is compromised.
203
204 **SSCs accumulate in *Dnmt3A* mutants**
205 We next wanted to determine whether SSC establishment, self-renewal or commitment to
206 differentiation was altered in *Dnmt3A* mutants. By counting GFRA1-positive cells (a general
207 SSC marker) on testis sections, we excluded a problem in the initial establishment of the
208 SSC pool in *Dnmt3A*^{KO} neonates: at 10dpp, similar numbers of GFRA1-positive cells were
209 observed compared to WT littermates (**Fig. 3a**). Unexpectedly, at all subsequent ages
210 examined, we did not observe progressive exhaustion of the SSC pool, but rather increased
211 GFRA1-positive cells, in both constitutive and conditional *Dnmt3A* mutants compared to
212 control mice (**Fig. 3a**), while somatic Sertoli cells (SOX9-positive) remained in normal
213 amounts (**Extended data Fig. 3a,b**). In control mice, the number of GFRA1-positive cells per
214 tubule mm^2 declined by 30-fold from 10dpp to 6 weeks; comparatively, it dropped by 5-fold

Figure 3

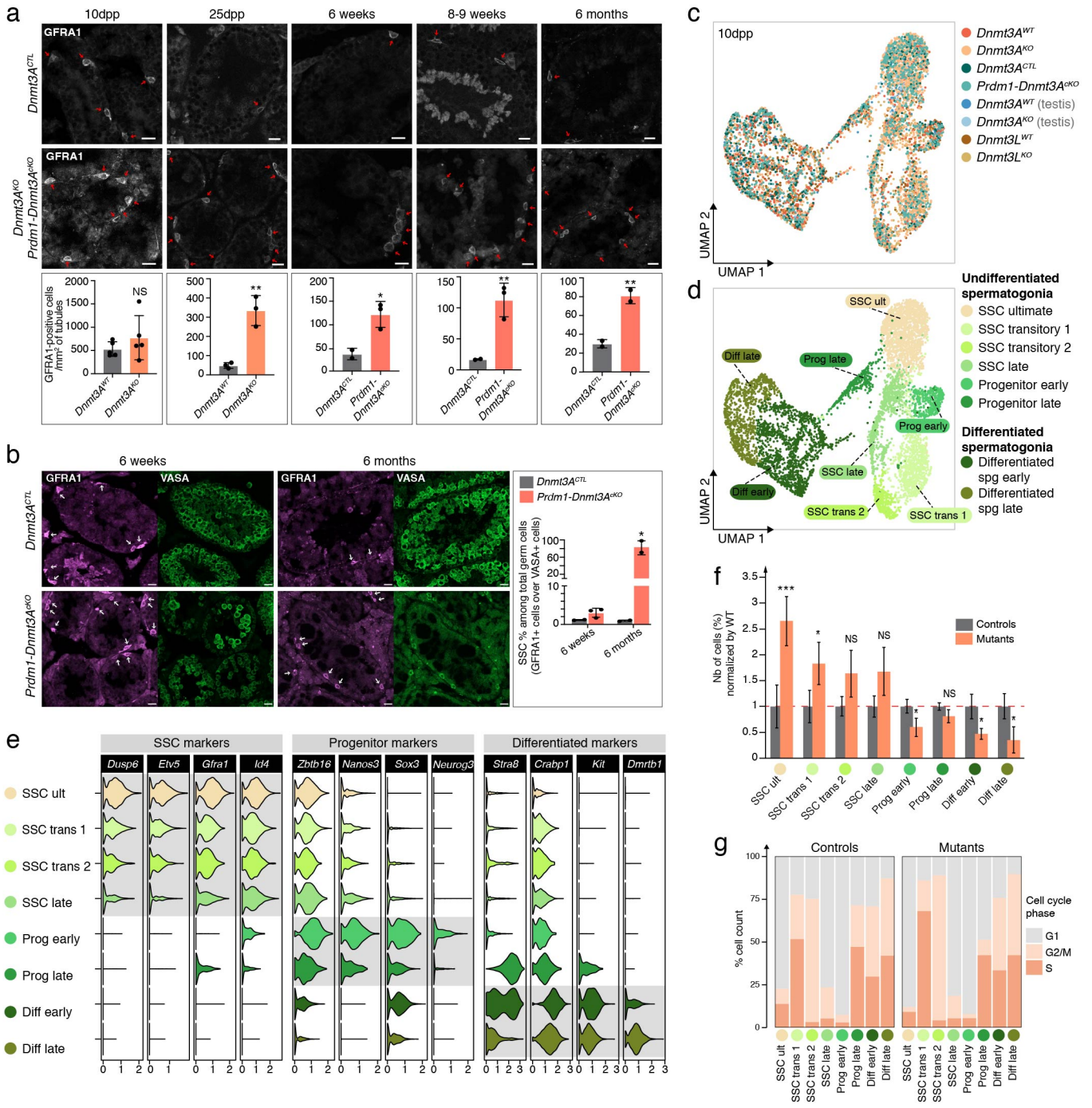


Fig. 3 | SSCs accumulate in absence of DNMT3A-dependent DNA methylation . **a**, (Top) Representative image of GFRA1 (SSC marker) staining on testis sections from *Dnmt3A*^{KO}, *Prdm1-Dnmt3A*^{cKO} and controls at indicated ages. Scale, 20 µm. (Bottom) Quantification of GFRA1-positive cells per mm2 of tubule. Control genotypes at 6 weeks: *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$ and *Dnmt3A*$^{2lox/KO}$; *Prmd1-Cre*$^{0/0}$ - at 8-9 weeks: *Dnmt3A*$^{2lox/KO}$; *Prmd1-Cre*$^{0/0}$ (n=2) and *Dnmt3A*$^{2lox/WT}$; *Prdm1-Cre*$^{0/0}$ - at 6 months: *Dnmt3A*$^{KO/WT}$; *Prmd1-Cre*$^{Tg/0}$ and *Dnmt3A*$^{2lox/WT}$; *Prmd1-Cre*$^{0/0}$. Data are mean ± SD (black bar) and individual points represent biological replicates (student t-test over WT, NS: non-significant, *p<0.05, **p<0.005). **b**, (Left) Representative microscopy images of GFRA1 (purple) and VASA (green) immunodetection at 6 weeks and 6 months. (Right) Quantification of percentage of GFRA1-positive cells (SSCs) among VASA-positive cells (germ cells). Control genotypes at 6 weeks: *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$ and *Dnmt3A2*$^{lox/KO}$; *Prmd1-Cre*$^{0/0}$- at 6 months: *Dnmt3A*$^{KO/WT}$; *Prmd1-Cre*$^{Tg/0}$ and *Dnmt3A*$^{2lox/WT}$; *Prmd1-Cre*$^{0/0}$. Data are mean ± SD (black bar) and individual points represent biological replicates (student t-test over WT *p<0.05). **c**, UMAP dimensionality reduction representation of scRNA-seq-integrated data from 10dpp germ cells (n= 6,638). Colors represent 8 genotypes and conditions. All conditions correspond to FACS-enriched germ cells except for "(testis)", which indicates that all testicular cells were analyzed. **d**, Unbiased cell clustering onto an UMAP representation of scRNA-seq data demonstrated eight germ cell clusters (from cluster 0 to 13 of unbiased cell clustering in Extended data Fig. 5a). **e**, Violin plots show mRNA level variation of different markers of SSCs, progenitors and differentiated spermatogonia among the eight germ cell clusters. **f**, Bar plot showing changes in germ cell cluster frequencies for pooled mutants normalized by controls. Data are mean ± SD (black bar) from four mutant samples (*Dnmt3A*^{KO}, *Prdm1-Dnmt3A*^{cKO}, *Dnmt3L*^{KO} FACS-enriched germ cells and *Dnmt3A*^{KO} all testicular cells) and four control samples (*Dnmt3A*^{WT}, *Dnmt3A*^{CTL}, *Dnmt3L*^{WT} FACS-enriched germ cells and *Dnmt3A*^{WT} all testicular cells). Statistical comparison was performed by student t-test over WT, NS= non-significant, *p<0.05, **p<0.005, ***p<0.005. **g**, Comparison of cell cycle signature for controls and mutants for each germ cell cluster.

only in *Dnmt3A* mutants (**Fig. 3a**). In 6 week-old *Prdm1-Dnmt3A*^{cKO} males, GFRA1-stained cells represented 3.5% of all germ cells (VASA positive), versus 1.2% in control mice (**Fig. 3b**). At 6 months, although *Prdm1-Dnmt3A*^{cKO} tubules appeared as germ cell-free by histological examination (**Fig. 2a**), GFRA1-positive cells were actually still present, at least 3-times more abundantly than in tubules of control mice (**Fig. 3a**), and were the quasi-exclusive germ cell type to remain (85%) (**Fig. 3b**). Comparatively, GFRA1-positive cells are rare in control males at this age, representing 0.9% of all germ cells. To better define the origin of this potential SSC increase, we utilized the *Id4-eGFP* reporter transgenic line[24,25], which we crossed onto the *Dnmt3A*^{KO} background (**Extended data Fig. 3b**). The *Id4-eGFP* transgene allows sorting undifferentiated spermatogonia sub-populations along a continuum that ranges from "ID4-GFP^{bright}" enriched in regenerative SSCs at the top of the SSC hierarchy to "ID4-GFP^{dim}" enriched in non-regenerative spermatogonial progenitors; "ID4-GFP^{medium}" denotes various intermediate and transitory states[25] (**Fig.1a**). FACS analysis at 10dpp revealed a 10-fold excess in ID4-GFP^{bright} SSCs in *Dnmt3A*^{KO} compared to WT testes (19.45% versus 2.27% of all ID4-GFP-positive cells) and a 3-fold excess in ID4-GFP^{medium}, while ID4-GFP^{dim} progenitors were relatively reduced by half (**Extended data Fig. 3c**). Therefore, despite the progressive spermatogenic decline of *Dnmt3A* mutants, we found that their SSCs are specified, and can further self-renew and maintain their most naive state, although excessively. We started considering that commitment to differentiation may be defective in *Dnmt3A* mutant SSCs.

SSCs cannot exit the stem cell state in absence of DNMT3A-dependent DNA methylation

To resolve the origin of the SSC phenotype, we generated unbiased droplet-based single-cell RNA-seq (10X Genomics Chromium) at 10dpp. Eight samples were sequenced, from FACS-enriched EpCAM-pos germ cells (*Dnmt3A*^{KO}, *Prdm1-Dnmt3A*^{cKO}, *Dnmt3L*^{KO} and respective littermate controls) and from whole testis cell suspension (*Dnmt3A*^{KO} and WT) (**Supplementary Table 1**). A total of 41,582 cells were mixed, integrated (to reduce noise

and variability between conditions) and analyzed together as biological replicates to gain statistical confidence (**Extended data Fig. 4a**). Based on the expression of known markers, we identified germ cells and testicular somatic cells (**Extended data Fig. 4b,c** and **Supplementary Table 2**) and a stringent filtering pipeline allowed us to retain 6,638 high quality single germ cell transcriptomes for analysis (See Methods) (**Fig. 3c** and **Extended data Figs. 4d-f and 5b-d**). Unbiased cell clustering revealed the existence of eight germ cell populations projected into UMAP representation (**Fig. 3d**), whose identities were derived from published classification[25,26], known cell-specific markers[8,27,28] and hierarchical clustering: SSC ultimate (SSC ult), SSC transitory 1 and transitory 2 (SSC trans 1 and SSC trans 2), SSC late (SSC late), Progenitor early, Progenitor late, Differentiated spermatogonia early and late (Diff early and Diff late) (**Fig. 3e** and **Extended data Fig. 5e-g**).

The relative frequency of germ cell clusters was affected in mutant samples, with increased SSC populations while progenitors and differentiated spermatogonia were under-represented (**Fig. 3f**). According to our defined criteria, SSC ultimate—the stem cell category at the foundation of spermatogenesis[8,25]—was the most strongly expanded in mutants, by 3-fold more abundant than in controls (2.7% versus 1%). Importantly, this cellular phenotype was similarly observed when constitutive *Dnmt3A*^{KO} and germ cell-conditional *Prdm1-Dnmt3A*^{cKO} were considered individually (**Extended data Fig. 5h**), again confirming that the phenotype is germ-cell autonomous. This phenotype could result from failure to acquire DNMT3A-dependent DNA methylation patterns in fetal prospermatogonia and/or from an intrinsic function of DNMT3A in post-natal SSCs, as *Dnmt3A* is expressed in SSCs and their derivatives (**Extended data Fig. 5i**). However, the latter option was disproved by the fact that *Dnmt3L*^{KO} mutants also showed SSC expansion by scRNA-seq (**Extended data Fig. 5h**), while *Dnmt3L* is exclusively expressed in fetal prospermatogonia but ceases expression at birth[19,29] (**Extended data Fig. 5h,i**). What *Dnmt3A* and *Dnmt3L* mutants have in common is a lack of DNMT3A-dependent DNA methylation inherited from earlier stages, the fetal prospermatogonia, when DNMT3A and DNMT3L are both expressed and establish male germline DNA methylation.

As a first attempt to explain this abnormal cellular distribution, we performed cell cycle analysis[30]. The categories “SSC ultimate”, “SSC late” and “Progenitors early” were the most enriched in G1 phase, suggesting a quiescent state, while “SSC trans”, “Progenitors late” and subsequent categories were mostly cycling cells, as described recently[31] (**Fig. 3g**). However, mutant and control samples showed similar cell cycle distribution, excluding that mutant SSC expansion—and “SSC ultimate” notably—results from higher mitotic rate and proliferation. Focusing on *Dnmt3A* mutants specifically (constitutive and germ-cell specific), we then analyzed cellular trajectories in pseudotime using Monocle. Spermatogenesis being a unidirectional differentiation process, control germ cells were arranged into a linear trajectory (**Fig. 4a**). By contrast, *Dnmt3A* mutant germ cells adopted different branches, as it would be the case for a complex, multi-path differentiation process. The four SSCs subtypes contributed to the pseudotime branches (**Fig. 4b**), likely indicative of a blockage. The same trend was observed in *Dnmt3L*^{KO} germ cells (**Extended data Fig. 6a,b**).

To further understand the dynamics of *Dnmt3A* mutant SSCs, we relied on RNA velocity analysis, which predicts the future state of a cell based on spliced versus unspliced mRNA balance[32]. With this tool, transcriptional trajectories are denoted as vectors of different amplitudes and directions. When applied to control cells, this analysis clearly illustrated SSC plasticity: SSC ultimate, transitory 1, 2 and late were all organized into a cycling process (**Fig. 4c**). Forward and backward movements between SSC subtypes likely relate to their uncommitted behaviors. In contrast, progenitors and differentiated spermatogonia were organized into a unidirectional movement, suggesting irreversible commitment towards differentiation. Strikingly, in *Dnmt3A* mutants, SSC ultimate cells no longer cycled with the other SSC subtypes. Instead, they were arranged on a unidirectional direction, opposite to spermatogenesis progression. This can be interpreted as an inability to change state and differentiate.

These findings indicate that SSCs proliferate normally but are unable to exit the stem cell pool in *Dnmt3A* mutants, which explains the spermatogenic failure. Importantly, we observed the same aberrant developmental program in *Dnmt3L* mutants: lack of DNA

Figure 4

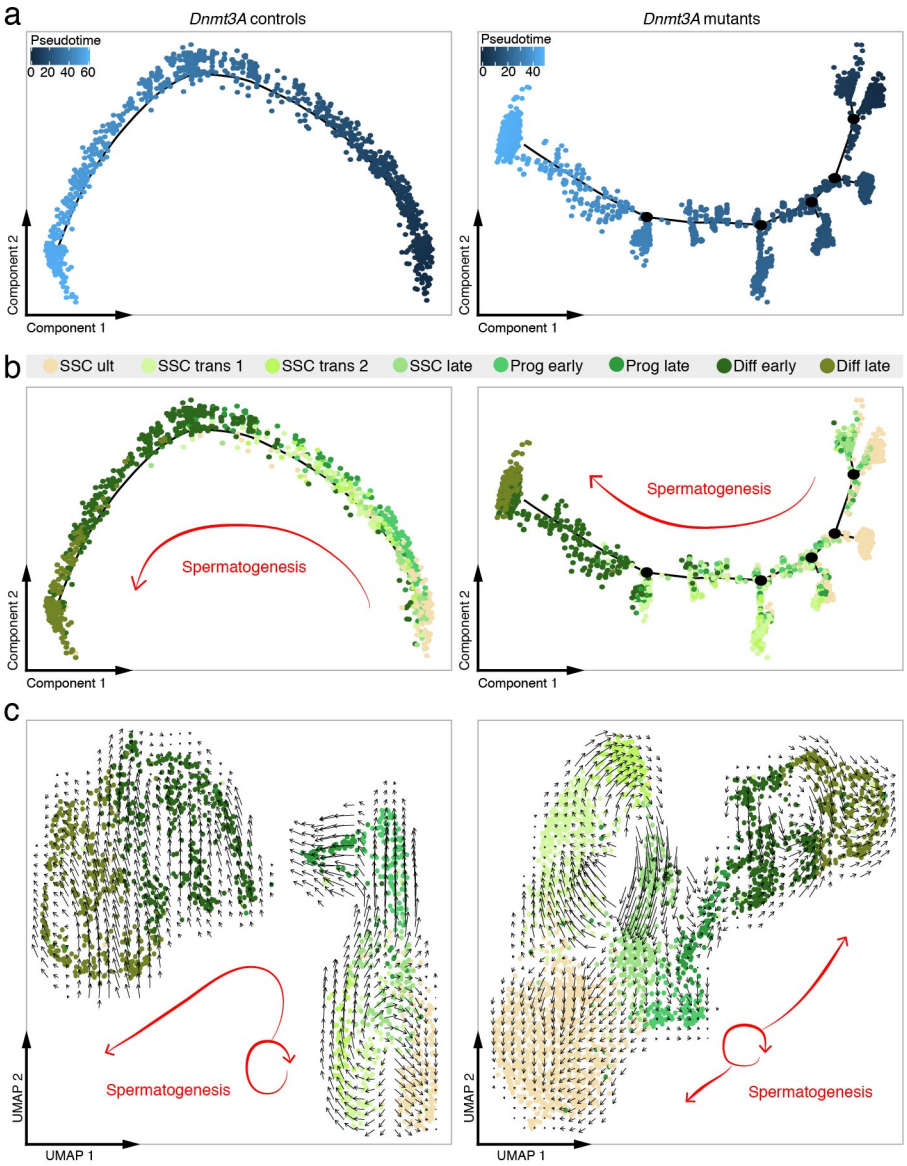


Fig. 4 | DNMT3A is essential to SSC plasticity. **a**, **b**, Pseudotime trajectory of germ cells from 10dpp testes. (Left) *Dnmt3A* control germ cells (FACS-sorted $Dnmt3A^{WT}$ and $Dnmt3A^{CTL}$). (Right) *Dnmt3A* mutant germ cells (FACS-sorted $Dnmt3A^{KO}$ and $Prdm1\text{-}Dnmt3A^{cKO}$). In (a), cells were ordered from beginning (dark blue) to the end (light blue).In (b), cells were colored following cluster allocation, see key above. **c**, Visualization of RNA velocity analysis on UMAP plot of control germ cells (Left) and *Dnmt3A* mutant germ cells (Right). Direction and size of arrows predict the future cellular state.

methylation marks established by DNMT3A prior to birth—rather than constitutive lack of DNMT3A *per se*—likely underpins the loss of SSC plasticity in *Dnmt3A* mutants.

DNA methylation regulates SSC plasticity by limiting enhancer activity

To identify the molecular pathways that lock *Dnmt3A*^{KO} SSCs into a stem cell identity, we performed bulk RNA-seq on pooled ID4-GFP^{bright} and ID4-GFP^{med} populations (ID4-GFP$^{bright+med}$) at 10dpp, to recover ultimate SSCs and their early transitory derivatives (**Extended data Fig. 3b** and **Supplementary Table 1**). We found 726 significantly misregulated genes in *Dnmt3A*^{KO} compared to WT cells (FDR < 5% and log2FC >1): 440 upregulated and 286 downregulated (**Fig. 5a** and **Supplementary Table 3**). Upregulated genes were mostly enriched in genes required for SSC homeostasis and maintenance (*Gfra1*, *Id4*, *Pax7*), while downregulated genes were involved in SSC differentiation (*Stra8*, *Kit, Sox3*) (**Fig. 5b**). Variable composition of the ID4-GFP$^{bright+med}$ pool between *Dnmt3A*^{KO} and WT could underlie the observed transcriptional changes. However, using our scRNA-seq data, we confirmed that SSC-specific markers were upregulated in individual *Dnmt3A*^{KO} SSCs, from ultimate to late subtypes (**Fig. 5c**). Transcriptional features linked to SSC identity are therefore intrinsically exaggerated in *Dnmt3A*^{KO} SSCs.

We then investigated whether this enhanced SSC transcriptome was reflected at the chromatin level. We focused on two histone marks whose occupancy at promoters and enhancers is anticorrelated with DNA methylation, H3K27me3 and H3K27ac, which are themselves antagonist to each other[33–35]. In *Dnmt3A*^{KO} SSCs, genome-wide reduction in DNA methylation could lead to H3K27me3 spreading and aberrant gene repression, and/or to H3K27ac gain and enhancer activation. Optimized Cleavage Under Targets and Release Using Nuclease (Cut&Run) was performed to profile these two marks in SSC-enriched ID4-GFP$^{bright+med}$ at 10dpp (**Supplementary Table 1**). We did not observe reciprocal gain and loss in H3K27me3 and H3K27ac in *Dnmt3A*^{KO} compared to WT and therefore analyzed these two marks independently (**Extended data Fig. 7a**). With regards to H3K27me3, we identified 7,139 peaks in *Dnmt3A*^{KO} and WT cells, among which 15.5% (*n*= 1,108) were present in

Dnmt3A^{KO} only, and were referred to as "up-enriched" peaks (**Extended data Fig. 7b,c** and **Supplementary Table 4)**. When associated to the nearest genes (<5kb from the TSS), 33 of these genes were downregulated and conjointly lost DNA methylation in *Dnmt3A*^{KO} cells (>30% compared to WT), at the same regions where H3K27me3 was gained (**Extended data Fig. 7d**). However, most *Dnmt3A*^{KO}-gained H3K27me3 peaks (>75%) were intragenic (**Extended data Fig. 7e**): rather than directly repressing genes, these H3K27me3 peaks likely arose in *Dnmt3A*^{KO} SSCs as a consequence of transcriptional silencing.

When considering enhancer-enriched H3K27ac marks, we counted a total of 36,431 peaks, among which 6.7% (*n*= 2,438) were present in *Dnmt3A*^{KO} only (**Extended data Fig. 8a,b** and **Supplementary Table 4)**. Investigating public ChIP-seq data from ID4-GFP^{bright} and ID4-GFP^{dim} cells from prepuberal testes[36], H3K27ac "up-enriched" peaks were neither referenced, nor decorated by H3K4me1 in one or the other cell type (**Fig. 5d**). This indicates that *Dnmt3A*^{KO}-gained H3K27ac peaks in ID4-GFP$^{bright+med}$ cells did not relate to active or primed enhancers that are normally present in ID4-GFP^{bright}/SSC ultimate or ID4-GFP^{dim}/progenitor cells, but were a specific feature of the *Dnmt3A*^{KO} chromatin landscape. Association of the *de novo* peaks with nearest genes revealed a total of 142 upregulated genes located in the vicinity of 284 regions that gained H3K27ac enrichment when losing DNA methylation in *Dnmt3A*^{KO} germ cells (**Fig. 5e** and **Supplementary Table 4)**. This suggests that a cohort of enhancers may ectopically form in hypomethylated *Dnmt3A*^{KO} SSCs. Interestingly, about half (50.7%) of the 284 *Dnmt3A*^{KO}-specific and DNA methylation-restricted H3K27ac peaks overlapped with one or more ENCODE candidate *cis*-Regulatory Elements (cCREs), the vast majority being annotated as distal enhancers in other tissue types (**Fig. 5f** and **Supplementary Table 4**). This was exemplified by *Tal1*, a transcription factor involved in hematopoietic development. While not substantially expressed in normal SSCs, *Tal1* was strongly activated in *Dnmt3A*^{KO} SSCs (6-fold) in association with a region that transitioned from DNA methylated to H3K27ac-occupied and that encompassed two known hematopoietic stem cell enhancers[37] (**Fig. 5g**). Interestingly, a second H3K27ac peak was gained nearby, but in a DNA methylation-independent manner—this position was not

methylated in WT SSCs—indicating a secondary recruitment, as also observed at the *Gata2* gene, another hematopoietic transcription factor (**Extended data Fig. 8c**). Genes expressed in SSCs could also further increase expression upon gaining DNA methylation-restricted H3K27ac peaks in *Dnmt3A*^{KO}: this was the case for *Fzd2*, a Wnt receptor important for SSC specification and proliferation[23] (**Fig. 5g**).

Finally, we searched for transcriptional motif signatures underlying *Dnmt3A*^{KO}-specific enhancers, considering the whole set of gained H3K27ac peaks to reach enough confidence (*n*= 2,438). Using HOMER, we identified four motifs that were significantly over-represented ($p < 1e^{-14}$) and all had potential CpG dinucleotides in their reconstructed consensus sequence. These motifs were predicted to be targeted by 35 transcription factors (TFs) that mainly belonged to ETV, ETS, ELK and FOXO families (**Fig. 5h**). Importantly, none of these TFs were significantly overexpressed in single *Dnmt3A*^{KO} cells (scRNA-seq data), whether they were normally expressed in SSCs or not (**Fig. 5h** and **Extended data Fig. 8d)**. The gain of enhancers in *Dnmt3A* mutants is therefore not linked to increased TF availability but rather to increased accessibility of the underlying binding motifs, likely because of the lack of DNA methylation**.** Incidentally, ETV, ETS, ELK and FOXO-type TFs have been reported to be DNA methylation-sensitive and notably gain novel binding sites in DNA methylation-deficient embryonic stem cells[6,38,39]. A number of 116 putative gene targets were then defined by their proximity to one or more motif-containing *Dnmt3A*^{KO}-specific H3K27ac peaks. These were more expressed in SSCs than in progenitors and differentiated cells in scRNA-seq data, and they were clearly upregulated in *Dnmt3A*^{KO} SSC subtypes—and the most acutely in SSC ultimate—compared to WT (**Fig. 5i**). Most saliently, all but one putative gene targets belonged to the list of 142 upregulated genes directly impacted by the lack of DNMT3A-dependent DNA methylation (**Extended data Fig. 8e** and **Supplementary Table 4**).

Taken all together, these results reveal that lack of DNMT3A-dependent DNA methylation is associated with ectopic recruitment of enhancers in SSCs, which activate or enhance expression of genes that buttress SSC identity.

Figure 5

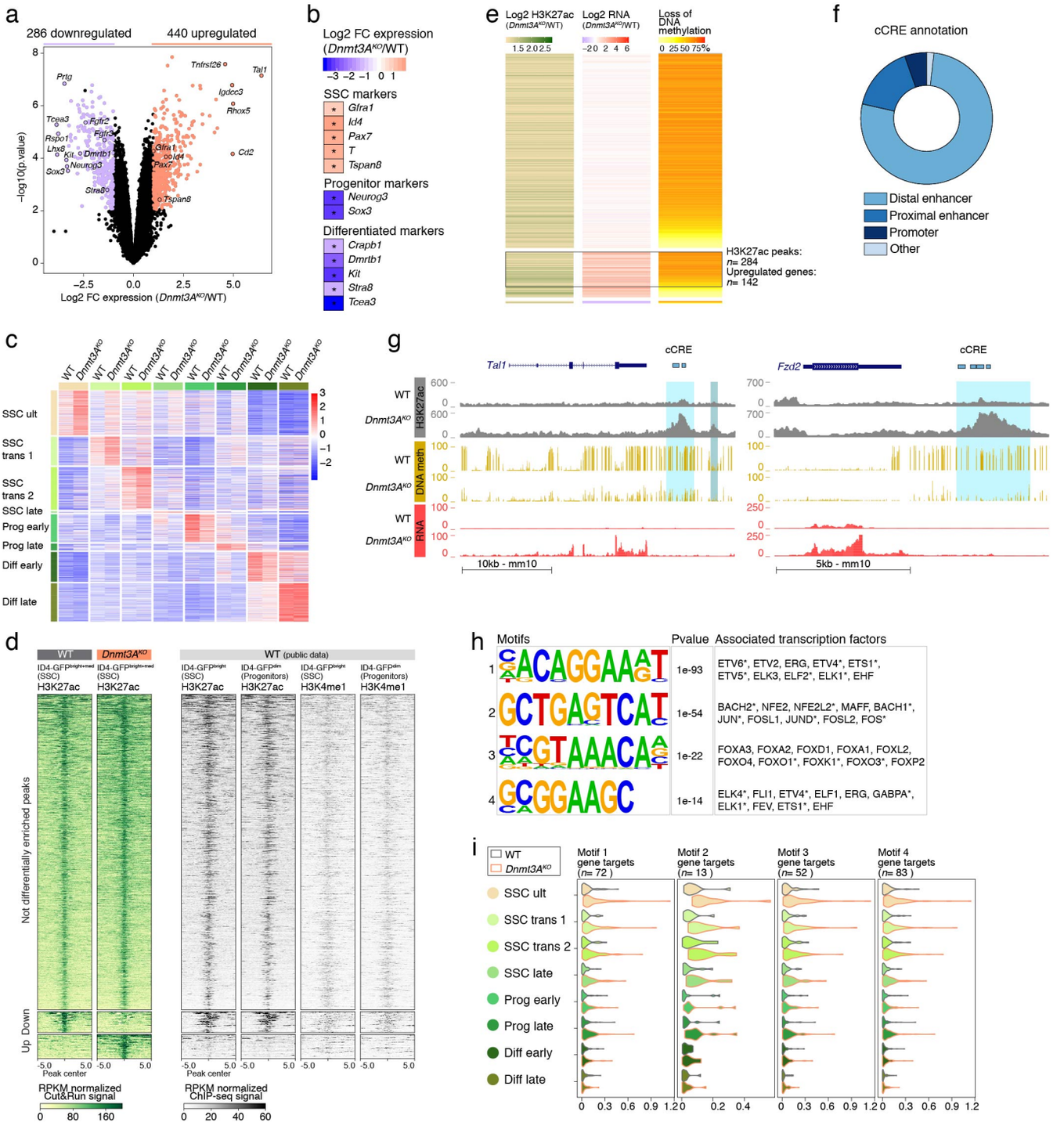


Fig. 5 | DNA methylation controls stem cell gene expression by limiting enhancer activity in SSCs. a, Volcano plots showing differential RNA levels in ID4-eGFP$^{bright+med}$ germ cells between *Dnmt3A^{KO}* and WT (10dpp) in log2 FC versus -log10 of the P values. Upregulated genes (440) are highlighted in red, downregulated genes (286) in purple (FDR<5% and FC>2, n= 3 biological replicates/genotype). **b**, RNA-seq heatmap representation of selected germ cell type-markers that are differentially expressed between *Dnmt3A^{KO}* and WT (FDR<5% and FC>2). **c**, scRNA-seq heatmap shows normalized expression value of markers characterizing each germ cell population in corresponding *Dnmt3A^{KO}* and WT germ cell types. **d**, (Left) Heatmap showing levels of H3K27ac peak enrichment in RPKM-normalized Cut&Run signal from *Dnmt3A^{KO}* and WT ID4-GFP$^{bright+med}$ germ cells (10dpp). Enrichment is assessed +/- 5kb from the center of the peak. Peaks are divided into three categories: not differentially enriched (*Dnmt3A^{KO}* = WT), down-enriched (*Dnmt3A^{KO}* < WT), up-enriched (*Dnmt3A^{KO}* > WT) (FDR<5% and FC>1). Number of biological replicates: *Dnmt3A^{KO}*= 3, WT= 4. (Right) Enrichment in H3K27ac and H3K4me1 from public ID4-eGFP^{bright} and ID4-eGFP^{dim} ChIP-seq data[36] (RPKM-normalized signal) at H3K27ac peaks described here in *Dnmt3A^{KO}* and WT. **e**, Heatmap focusing on i) *Dnmt3A^{KO}*-gained H3K27ac peaks (log2 FC) compared to WT (left), ii) differential expression (log2 FC) of genes associated with H3K27ac peaks (TSS +/- 5kb from the peak) between *Dnmt3A^{KO}* and WT—from ID4-eGFP$^{bright+med}$ RNA-seq—(center), and iii) percentage of DNA methylation loss in *Dnmt3A^{KO}* versus WT on H3K27ac peak genomic location—from E18.5 WGBS—(right). Rows are ordered according to gene expression changes: top, not differentially expressed; middle, upregulated; bottom, downregulated. **f**, Pie chart of genomic annotations of ENCODE candidate cis-Regulatory Elements (cCREs) overlapping with *Dnmt3A^{KO}*-specific H3K27ac peaks associated with >30% DNA methylation loss and upregulated genes. **g**, Genome browser representations of H3K27ac enrichment (grey), DNA methylation (yellow) and RNA expression (red) for *Dnmt3A^{KO}* and WT. Regions showing Dnmt3AKO de novo H3K27ac peaks associated with >30% DNA methylation loss are shaded in light blue;

regions showing de novo H3K27ac peaks at regions not controlled by DNA methylation are shaded in grey-blue. **h**, (Left) Enriched motifs 1 to 4 found at *Dnmt3A*^{KO}-specific H3K27ac peaks (n= 2,438). (Center) P values of motif 1 to 4. (Right) Transcription factors (TFs) that have the best match with the enriched motif, *indicates expression in SSCs, based on 10dpp scRNA-seq data (see Extended Fig. 8d). **i**, Violin plots showing expression levels of TF gene targets in germ cell population subtypes discriminated by scRNA-seq for *Dnmt3A*^{KO} and WT. TF gene targets were defined by association with *Dnmt3A*^{KO}-specific H3K27ac peaks (TSS +/- 5kb from the peak) containing one or more of the four motifs.

Discussion

The process of *de novo* DNA methylation is an early event of spermatogenesis, occurring prior to birth, before SSC emergence and meiosis. We show here that DNMT3A and DNMT3C jointly shape the fetal male germline methylome with perfect complementarity and striking lack of redundancy in the genomic sequences they target, and most importantly, in the spermatogenetic phase they control. While it was known that DNMT3C is crucial for meiosis, by selectively methylating and silencing retrotransposons[12], we found that DNMT3A-dependent DNA methylation is widespread and regulates SSC homeostasis. More precisely, the hypomethylated genomic landscape of *Dnmt3A* mutant SSCs results in appearance of ectopic enhancers that reinforce the robustness of the stem cell gene expression program, which very likely locks SSCs into self-renewal and prevents commitment to differentiation. Collectively, we provide evidence for a novel reproductive function for DNA methylation, beside meiosis protection: the programming of life-long spermatogenesis fueling (**Fig. 6**).

This study resolves several questions and controversies regarding DNMT3A role in spermatogenesis. First, conflicting reports existed as to whether *Dnmt3A*^{KO} germ cells can undergo meiosis or not[11,15]. We demonstrate that *Dnmt3A* mutant males complete meiosis during the first round of spermatogenesis—which bypasses the SSC state—and can generate haploid cells, including spermatozoa. However, they produce post-meiotic cells only once in their life, after which their germ cell contingent exclusively consists in differentiation-deficient SSCs. Second, we definitely exclude a role for DNMT3A in retrotransposon control during spermatogenesis; DNMT3C is solely dedicated to this pathway in mice. Coincidently, we prove that meiotic recombination remarkably tolerates low genomic methylation levels, provided that retrotransposon promoters are methylated. Finally, NSD1-dependent H3K36me2 deposition was recently shown to drive *de novo* DNA methylation of euchromatic regions in fetal male germ cells[40]. We reveal that *Nsd1*^{KO} and *Dnmt3A*^{KO} prospermatogonia share a genome-wide pattern of hypomethylation that excludes young retrotransposon promoters, highlighting that DNMT3A is the enzyme recruited to H3K36me2-marked regions,

Figure 6

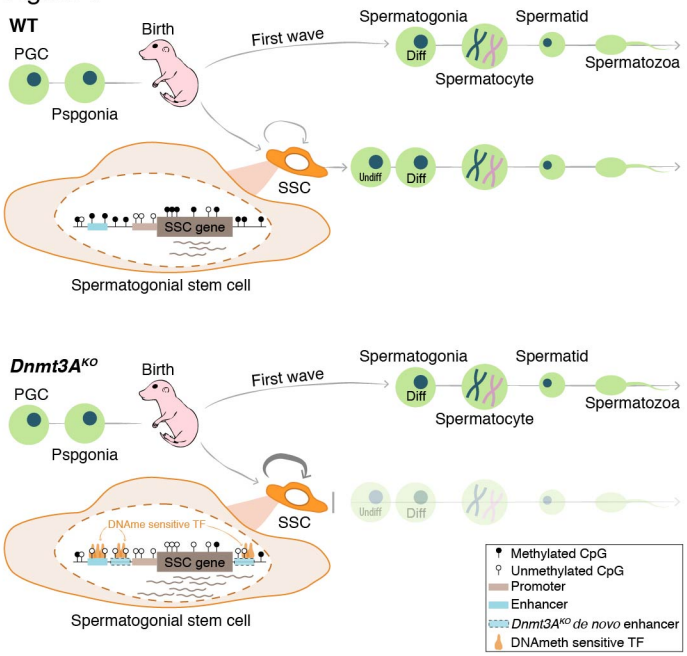


Fig. 6 | Model of DNMT3A-dependent DNA methylation in spermatogenesis. Top, Developmental options of fetal prospermatogonia (Pspgonia) at birth in WT males: i) directly transition to differentiated spermatogonia (Diff) to initiate the first wave of spermatogenesis, or ii) form the foundational SSC pool, which must balance self-renewability and differentiation to sustain spermatogenesis across male reproductive lifespan. Bottom, lifelong reproductive consequences of failure to establish DNA methylation in *Dnmt3A*^{KO} Pspgonia: i) the first wave of SSC-independent spermatogenesis can proceed, however ii) once prospermatogonia develop into SSCs, these can only self-renew and are unable to differentiate, preventing following waves of spermatogenesis. The hypomethylated genome uncovers ectopic enhancers (dashed lines) that become accessible to methyl-sensitive TFs (orange) and enforce irreversible expression of SSC-promoting genes.

likely through its PWWP reading domain. However, despite convergent DNA methylation defects, we do not expect a convergent phenotype in *Dnmt3A* and *Nsd1* mutants. Accumulation of H3K27me3 leads to specific gene dysregulation in *Nsd1*^{KO} germ cells, a feature that was not observed in *Dnmt3L*^{KO} germ cells[40], and which we did not observe in *Dnmt3A*^{KO} SSCs. Moreover, *Nsd1* mutants exhibit early spermatogonia loss[40]: this contrasts with the expanding SSC phenotype we uncovered in *Dnmt3A* and *Dnmt3L* mutants. This discrepancy underscores the complexity of the epigenetic interplay that regulates male germ cell development.

The spermatogenetic failure of *Dnmt3A* mutant males is quite unique among other SSC-related infertility phenotypes. Disrupted SSC homeostasis and subsequent spermatogenetic decline have been previously linked to defects in generating the foundational SSC compartment, to impaired self-renewal and precocious exhaustion of the SSC pool, or to defective differentiation priming[41,42]. Remarkably, *Dnmt3A*^{KO} SSCs lose their plasticity by continuously self-renewing and ignoring the differentiation route. Through chromatin profiling combined with bulk and single-cell transcriptomics, we were able to provide detailed mechanistic insights into this phenotype: SSC expansion is not linked to hyperproliferation but rather to an intrinsic exacerbation of a stem cell program. A cohort of ectopic H3K27ac-enriched enhancers arise in *Dnmt3A*^{KO} SSCs, and this is linked to upregulation of nearby SSC-specific genes but also, notably, to the activation of stem cell genes that are normally relevant to other tissue-specific stem cells, such as hematopoietic stem cells. We found that these enhancers are largely methylated by DNMT3A during fetal development, prior to the formation of the SSC pool, and most importantly, they possess binding motifs for methyl-sensitive TFs, such as the ETV- or ETS-type TF families[38,43]. Interestingly, in proximity to these DNA methylation-restricted enhancers that gained H3K27ac in *Dnmt3A*^{KO} SSCs, we frequently observed additional enhancers that emerged at regions that were not controlled by DNA methylation, indicating secondary reinforcement of the SSC gene program through additive and synergistic TF binding. We propose that DNMT3A-dependent methylation, assisted by DNMT3L in fetal germ cells, programs SSC

dual capacity to both self-renew and differentiate after birth, by limiting the accessibility of stem cell-related enhancers to methyl-sensitive TFs that are abundantly present in SSCs.

The essential role of DNMT3A has been previously described in other stem cell systems, although mechanisms at play may be different and/or not fully resolved. In neural stem cells (NSCs), DNMT3A-dependent DNA methylation does not limit but rather activate enhancers linked to neurogenesis, by antagonizing H3K27me3[44]. Moreover, *Dnmt3A*^{KO} NSCs are normal but undergo skewed differentiation towards a glial rather than neuronal fate. In hematopoietic stem cells (HSCs), *Dnmt3A* deficiency is also not impactful in steady-state-conditions but skewed trajectory occurs upon differentiation, in association with increased accessibility of CpG-rich TF binding motifs linked to the myeloid lineage[45,46]. Interestingly, when challenged by serial transplantation, the *Dnmt3A*^{KO} HSC compartment increasingly expands towards unlimited self-replication, while differentiation potential inversely declines[47,48]. This is reminiscent to the effect we found in *Dnmt3A*^{KO} SSCs, suggesting that the molecular mechanisms we evidenced could also apply to transplanted *Dnmt3A*^{KO} HSCs. Overall, differences in DNMT3A role in various stem cell-supplied systems may be attributable to distinct physiology and epigenetic dynamics. More precisely, spermatogenesis is a unidirectional process, while hematopoietic or neural differentiation are multilineage paths. Then, DNA methylation is dramatically reduced in *Dnmt3A*^{KO} SSCs (by 3 to 4 fold compared to WT), while changes are relatively sparse and focal in *Dnmt3A*^{KO} HSCs or NSCs. As demonstrated here, DNMT3A is the main *de novo* methylation enzyme in embryonic germ cells, while its activity is less important in embryonic somatic cells, whose genome mostly acquires DNA methylation in a DNMT3B-dependent manner[49].

To conclude, these results provide novel and significant advances towards understanding the regulation of stem cell function by DNMT3A and the relationship between male infertility and epigenetic anomalies.

Acknowledgements

We are grateful for support and feedback from members of the Bourc'his lab, to N. Fayaubost for animal care, M. Greenberg for critical reading of the manuscript and N. Servant for bioinformatic assistance. We thank A. Clark for the *Prdm1*-Cre and *Oct4*-eGFP mice, J. Oatley for the *Id4*-GFP mouse and K. Laband and J. Dumont for the anti-SCP3 antibody. We acknowledge the ICGex NGS platform of Institut Curie (supported by grants ANR-10-EQPX-03, Equipex and ANR-10-INBS-09-08, France Génomique)- and the Cell and Tissue Imaging Platform-PICT-IBiSA (member of France-Bioimaging, ANR-10-INBS-04) of the Genetics and Developmental Biology Dpt (UMR3215/U934) of Institut Curie. The laboratory of D.B. is part of the LABEX DEEP (ANR-11-LABX-0044, ANR-10-IDEX-0001-02). This work was supported by a grant from the Agence Nationale pour la Recherche (ANR-17-CE12-00013-01), the Fondation Bettencourt Schueller and the Fondation pour la Recherche Médicale (FRM Team Label). M.D. was supported by PhD fellowships from Région Ile-de-France and Fondation pour la Recherche Médicale. L.B. is the recipient of a PhD Boehringer Ingelheim Fonds fellowship.

Author contributions

M.D. and D.B. designed and conceived the study. Most experiments were performed and analyzed by M.D. M.A. assisted in the histology, immunofluorescence and microscopy. J.B. performed the WGBS experiment. L.B. and M.A. assisted in the Cut&Run experiment. L.G.B and S.L. performed NGS experiments and supervised generation of single cell and NGS data. A.T. performed the bioinformatic analyses. M.W. provided the *Dnmt3A* mouse strain. D.B. and M.D. interpreted data and wrote the paper. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Methods

Mice. Mice were hosted on a 12h light/12h dark cycle with free access to food and water in the pathogen-free Animal Care Facility of the Institut Curie (agreement C 75-05-18). All experimentations were approved by the Institut Curie Animal Care and Use Committee and adhered to European and national regulation for the protection of vertebrate animals used for experimental and other scientific purposes (directives 86/609 and 2010/63). For tissue and embryo collection, mice were euthanized by cervical dislocation.

Mouse strains, all bred onto C57Bl6/J background, were previously described: *Dnmt3A*^{2lox} [11], *Dnmt3L*^{KO} [10], *Dnmt3C*^{KO} [12], *Prdm1-Cre* [21] (gift of A. Clark, UCLA), *Id4-eGFP* [24] (gift of J. Oatley, Washington State University) and *Oct4-eGFP* (Jackson Laboratories, stock no. 008214)[17] (gift of A. Clark, UCLA). The *Dnmt3A*^{KO} strain was obtained by constitutive defloxing of the *Dnmt3A*$^{2lox/2lox}$ strain using the *Zp3-Cre* line (Jackson Laboratories, stock no. 003394)[50]. Germ-cell conditional *Dnmt3A* mutants were obtained by crossing *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$ males with *Dnmt3A*$^{2lox/2lox}$ females. The mutant genotype of interest was *Dnmt3A*$^{KO/2Lox}$; *Prdm1-Cre*$^{Tg/0}$ (referred to as *Prdm1-Dnmt3A*^{cKO} in figures and text); control genotypes were *Dnmt3A*$^{2lox/WT}$; *Prdm1-Cre*$^{Tg/0}$, *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$, *Dnmt3A*$^{2lox/KO}$; *Prdm1-Cre*$^{0/0}$, or *Dnmt3A*$^{2lox/WT}$; *Prdm1-Cre*$^{0/0}$, as indicated in legends of each experiment (referred to as *Dnmt3A* controls).

Prenatal time points were obtained by following timed pregnancies where the following day post-coitum was considered as embryonic day E0.5. Postnatal time points were measured starting from birth, considering the first postnatal day as 1 day post-partum (1dpp).

RT-qPCR and RT-PCR. Testes (one testis per biological replicate) were homogenized in Trizol (Life Technologies) using TissueLyser LT (Qiagen). Total RNA was extracted according to the manufacturer's recommendations, and DNase-treated with the Qiagen RNase-Free DNase Set. RNA was quantified using Qubit Fluorometric Quantitation (Life Technologies), and checked for integrity using the TapeStation (Agilent). For RT-PCR and

RT-qPCR, 2µg of RNA was reverse transcribed using random priming with SuperScript III (Life Technologies). RT-PCR and qRT-PCR primers are listed in **Supplementary Table 5**.

Quantitative PCR (RT-qPCR) was conducted using Power SYBR Green PCR Master Mix (Life Technologies) on a ViiA7 Real-Time PCR System (Applied Biosystems). Data were normalized to *beta-actin* using the delta-delta Ct approach. Minus RT controls were included for each collected data point.

PCR (RT-PCR) was conducted using GoTaq DNA polymerase (Promega). PCR programs were the following, *Sp10*: 30 cycles (94°C for 30 sec; 60°C for 30 sec; 72°C for 1 min) and *Tp2*: 30 cycles (94°C for 30 sec; 58°C for 30 sec; 72°C for 30 sec). Minus RT controls were included for each collected data point.

Histological sections. Testes were dissected without disrupting the tunica albuginea, fixed over-night (ON) in fresh Davidson's-modified 4M formaldehyde fixative (4M formaldehyde, 15% ethanol, 5% glacial acetic acid), washed (2x10min and 1x30min) in 1XPBS with 0.25M Tris-HCl pH7.2 and then stored at 4°C in 1xPBS-70% ethanol with 0.25M Tris-HCl pH7.2. Alternatively, testes were fixed ON in Bouin's fixative solution (Sigma), washed and stored in 70% ethanol at 4°C.

Testes were then paraffin-embedded, sectioned (8 µm) and stained with periodic-acid-Schiff (PAS) using standard protocols. Slides were imaged on a Leica upright epifluorescence microscope using color camera. For quantification of tubule types (**Fig. 2a** and **Extended data Fig. 2e**), tubules from one testis section per animal were counted. This represents above 100 tubules for all ages and genotypes, except at 6 months for *Dnmt3A*^{KO} and *Dnmt3L*^{KO} genotypes (due to severe testis size reduction), where around 60 tubules were counted per section.

Immunofluorescence microscopy. For immunofluorescence on testis sections, testes were collected, fixed at 4°C in 4% paraformaldehyde (2dpp: 1h; 10dpp: 2h; 25dpp: 2h30; adult: ON), washed 2x10min and 1x30min in 1XPBS with 0.25M Tris-HCl pH7.2 and 3x5min in

554 1XPBS. Then testes were soaked in 15% and 30% sucrose solution in 1XPBS with 50mM
555 Tris-HCl pH7.4 for two consecutive ON incubations, embedded in OCT medium (Tissue-Tek)
556 and frozen at -80°C. Frozen sections were cut (6-10 µm) and spotted onto positively-charged
557 slides (Superfrost Plus slides- Thermo Fisher Scientific) and stored at -20°C prior to use. For
558 immunofluorescence detection, slides were thawed at room temperature (RT), washed in
559 1XPBS, and then blocked and permeabilized for 1h at RT in blocking buffer (10% donkey
560 serum, 3% BSA and 0.2% Triton in 1XPBS). Sections were incubated ON at 4°C with
561 primary antibody diluted in blocking buffer, followed by washes (3x5min) with 1XPBS
562 supplemented with 0.1% Tween-20, and then incubated for 1-3h at RT with Alexa Fluor-
563 conjugated secondary antibodies diluted in blocking buffer. After final washes, sections were
564 mounted in DTG (Glycerol, DABCO 2.5%, 50mM Tris pH8.6) with DAPI. Images were
565 obtained using an inverted laser-scanning confocal microscope LSM 700 or 900 (Zeiss).
566 Each staining was performed on three different testis sections from at least three different
567 animals. Primary antibodies were omitted as a negative control. For quantification, the
568 number of cells were normalized by the mean of the tubule surface.
569 For immunofluorescence on meiotic spreads[51], testes (16 to 25dpp) were collected,
570 seminiferous tubules were released from the tunica albuginea, rinsed in 1XPBS and
571 incubated for 1h in hypotonic buffer (30mM Tris, 50mM sucrose, 17mM trisodium citrate
572 dihydrate, 5mM EDTA, 0.5mM DTT, pH 8.2) containing 0.5mM protease inhibitors (Roche).
573 Seminiferous tubules were then placed in a drop of sucrose solution (100mM sucrose pH
574 8.2) and tubules were dissociated until single cell suspension. Cells were slowly dispersed
575 on slides pre-covered of a fresh fixative solution (2% formaldehyde, 0.05% triton X-100,
576 0.02% SDS). Slides were air-dried, washed (3x5min) in 0.2% Photo-Flo (Kodak) and stored
577 at -80°C until use. Immunofluorescence detection was performed as described above, except
578 that slides were blocked for 20min at RT and incubated with primary and secondary antibody
579 for 1h at RT. Slides were imaged on an upright epifluorescence microscope. Antibody
580 references and working dilutions are available in **Supplementary Table 6**.
581

Cell sorting and analysis by fluorescent activated cell sorting (FACS). Testes were dissected from E18.5 or 10dpp males, decapsulated and transferred in low-binding Eppendorf tube containing 100-150µL collagenase solution in HBSS (Collagenase type IV (Gibco), 2X AAs (Gibco), 2X Na-pyruvate (Gibco), 25mM HEPES-KOH pH7.5). Dissociation was achieved at 37°C for 5-7min by gently flicking the tube. Then 400-600µL of TrypLE Express (Gibco) was added to the cell suspension and incubated for 5-7min at 37°C. Single-cell suspension was obtained by up-and-down pipetting. TrypLE Express was quenched with 140-210µL fetal bovine serum (FBS) and the cell suspension was centrifuged at 600g/4min/4°C, washed with FACS buffer (1XPBS, 2mM EDTA, 1%BSA) and filtered through a 35 µm strainer.

For sorting prospermatogonia or spermatogonial stem cells (SSCs), *Oct4-eGFP* and *Id4-eGFP* transgenic mice were used, respectively. Washed single-cell suspension was centrifuged and resuspended in FACS buffer supplemented with DAPI (2ng/ml), in order to discriminate dead cells. Cell sorting was performed on a BD FACSaria II (BD Biosciences) using 100 µm nozzle. Prospermatogonia were isolated by gating DAPI-neg, OCT4-GFP-pos cells. SSCs were isolated by gating DAPI-neg, ID4-GFP-pos cells. ID4-GFP positive cells were further separated into three size-matched compartments (bright, medium and low); only ID4-GFP^{bright} and ID4-GFP^{med} were sorted and considered as SSCs (ID4-GFP$^{bright+med}$)(see gates **Extended data Fig. 3c**).

For spermatogonia sorting and analysis, single-cell suspensions were incubated 20min on ice in 500µL FACS buffer containing 3.5µg/mL of rat anti-mouse EpCAM antibody conjugated with Alexafluor 647 (Biolegend) and 8µg/mL mouse anti-β2-Microglobulin monoclonal antibody conjugated with PE (Santa Cruz). After antibody labeling, cells were washed in FACS buffer and resuspended in DAPI-FACS buffer. Compensation controls were applied for each experiment. Cell sorting and data acquisition were performed on FACSaria II (BD Biosciences) using 100 µm nozzle. Spermatogonia were isolated by gating DAPI-neg, EpCAM-pos, β2M-neg cells. Further analyses were performed with FlowJo software.

DNA methylation analyses. WGBS was performed on prospermatogonia sorted by FACS (with the *Oct4-eGFP* transgenic line) using direct post-bisulfite adaptor tagging (PBAT)[18]. Prospermatogonia of three embryos from each genotype were pooled for genomic DNA extraction using the QIAamp DNA Micro Kit (Qiagen) and spiked-in with 0.3% of unmethylated Lambda DNA (Promega). Bisulfite conversion was performed using EZ DNA Methylation-Gold Kit (Zymo Research) and the whole converted DNA was used as input for library preparation using the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) according to the recommended protocol and indexing PCR using 2X KAPA HiFi HotStart Uracil+ ReadyMix (KAPA Biosystems). Indexed, cleaned-up libraries were analyzed and quantified on a 2200 Tapestation instrument using a D5000 screen tape (Agilent). Library pools were sequenced on an Illumina NovaSeq in 100bp paired-end (PE) reads run.

Targeted analysis of DNA methylation at retrotransposon promoters (IAPEz, L1MdA and L1MdTf) was performed on FACS-sorted spermatogonia (gating DAPI-neg, EpCAM-pos, β2M-neg cells) at 10dpp. DNA was extracted by incubating one volume of lysis buffer (200mM Tris-HCl pH8, 10mM EDTA, 0.4% SDS, 400mM NaCl, 80µg/mL linear polyacrylamide, 0.4mg/mL proteinase K) at 55°C for 2-4h, precipitated with isopropanol and cleaned with 70% ethanol. DNA was then bisulfite-converted by using EZ DNA Methylation-Lightning™ Kit (Zymo Research). CpG methylation was quantified on at least three individual animals of each genotype by pyrosequencing on a PyroMark Q48 (Qiagen) using the Q48 Software. Primers are available in **Supplementary Table 5**.

RNA sequencing. For bulk RNA-seq of whole testes, testes were collected from *Dnmt3A*^{KO} and WT littermate animals at 19dpp and 25dpp (*n*=2 per genotype and age) and homogenized in Trizol (Life Technologies) using TissueLyser LT (Qiagen). Total RNA was extracted according to the manufacturer's recommendations, and DNase-treated with the RNase-Free DNase Set (Qiagen). RNA was quantified using Qubit Fluorometric Quantitation (Life Technologies), and checked for integrity using TapeStation (Agilent). RNA-seq libraries were performed with TruSeq Stranded mRNA Kit (Illumina) and sequenced in 100bp PE

reads run on NovaSeq 6000 (Illumina). Published data for *Dnmt3C* mutants (*Dnmt3C*$^{IAP/IAP}$) at 20dpp were including in the analysis[12].

For bulk RNA-seq of SSCs, ID4-GFP$^{brigth+med}$ germ cells from 10dpp mice were FACS-sorted as described above and 5,000-10,000 cells were collected for each animal (*n*= 3 per genotype) in low binding Eppendorf tubes pre-filled with 100µL of Extraction Buffer from Arcturus PicoPure RNA isolation Kit (Applied Biosystems). Samples were frozen and stored at -80°C for later use. RNA extraction was performed following detailed protocol of the same kit and DNase-treated with RNase-Free DNase Set (Qiagen). RNA concentration and quality were assessed by Bioanalyzer RNA 6000 Pico Assay (Agilent). RNA-seq libraries were performed with SMARTer® Stranded Total RNA-Seq Kit v2 (Takara - rRNA depletion) and sequenced in 100bp PE reads run on NovaSeq 6000 (Illumina), with three biological replicates per genotype.

For single cell RNA-seq (scRNA-seq), germ cells were enriched by FACS from 10dpp mice as described above (gating DAPI-neg, EpCAM-pos). Cell suspensions were loaded into Chromium Single Cell 3' Kit (V2 or V3) and used to isolate single-cell in Gel Bead-in-Emulsion (GEMs) using the Chromium controller (10X Genomics). In all cases, cells were loaded in the proper concentration on the instrument with the expectation of collecting up to 6,000 GEMs containing cells per samples. Single-cell libraries were performed according to manufacturer's recommendations of Chromium (10X Genomics) and sequenced in 26-91bp paired-end reads run for V2 samples (*Dnmt3L*^{WT} and *Dnmt3L*^{KO}) and 28-91bp paired-end reads run for V3 samples (all other samples) on NovaSeq 6000 (Illumina).

Cut&Run. The Cut&Run protocol was modified from Skene et al. (2017)[52]. In brief, 20µL of Concanavalin A beads (Polysciences) per sample were resuspended in 1mL of Binding Buffer (20mM HEPES-KOH, 10mM KCl, 1mM CaCl2, 1mM MnCl2). Beads were washed twice in 1mL Binding Buffer and resuspended in 20µL of Binding Buffer/ sample.

ID4-eGFP$^{bright+med}$ germ cells (between 5,000-10,000 cells) were sorted by FACS into 1X PBS (~50µL), as described before. Samples were split into aliquots according to the

number of antibodies profile required and 1mL of Wash buffer (20mM HEPES-KOH, 150mM NaCl, 0.5mM spermidine (Sigma) and 1X Complete™ EDTA-free protease inhibitor cocktail (Roche) was gently added to the cell solution and 20µL of pre-washed beads. Cells and beads were incubated for 10min at RT on a rotating wheel. Cells were then collected on magnetic beads, the supernatant was discarded and cells were resuspended in 400µl of Antibody buffer (Wash buffer with 2mM EDTA, 0.05% digitonin (Millipore) and 1:200 antibody). Cells were incubated with antibodies for 1h at RT on a rotating wheel and washed twice with 1ml of Dig wash buffer (Wash buffer with 0.05% digitonin). Antibody references are available in **Supplementary Table 6**.

Samples were then incubated with 1:400 ProteinA-MNase fusion protein (produced by the Recombination Protein Platform of the Institut Curie, 0.785 mg/mL) for 15 min at RT followed by two washes with 1ml Dig wash buffer. Cells were then resuspended in 150µl Dig wash buffer and cooled down on ice for 5min before addition of CaCl2 to a final concentration of 2mM. Targeted digestion was performed for 30 min on ice, reaction was stopped by adding 150µL of 2X STOP buffer (340mM NaCl, 20mM EDTA, 4mM EGTA, 0.02% Digitonin, 0.125mg/mL RNase A, 0.25mg/mL Glycogen). Samples were then incubated at 37°C for 20min to release cleaved chromatin fragments into the supernatant. After centrifugation at 15,000rpm for 5min, supernatants were transferred to new low-binding tubes. Following addition of 0.1% SDS and 0.17mg/mL Proteinase K, samples were mixed by inversion and incubated at 70° C for 30 minutes. DNA was purified using phenol/chloroform followed by chloroform extraction and precipitated with 10µg of glycogen and 3 volumes of 100% ethanol for at least 20 minutes on ice. DNA was pelleted at 14,000 rpm at 4°C for 20 minutes. The DNA pellet was washed in 85% ethanol, spin down and resuspended in 40µL low Tris-EDTA (10mM Tris, 0.1mM EDTA) after complete evaporation of the ethanol.

Library preparation was performed according to manufacturer's instructions (Accel-NGS 2S DNA library kit, Swift biosciences) with a modified library amplification program: 98°C for 45sec, (98°C for 10sec, 60°C for 15sec, 68°C for 1min)x15 cycles, hold at 4°C.

Average library size was tested on Agilent 4200 Tapestation using a DNA5000 screentape and quantification was performed on Invitrogen QUBIT 4 using high sensitivity DNA kit. Cut&Run libraries were sequenced on a NovaSeq (Illumina) using PE 100bp run, with four biological replicates for *Dnmt3A*^{WT} and three for *Dnmt3A*^{KO}.

WGBS analysis. Adapters were trimmed using Atropos (v1.1.16). Trimmed reads were cleaned by removal of 5bp in 5' end of read1 and 12bp in 5' end of read2 using Cutadapt v1.12. Reads shorter than 15bp were discarded. Cleaned reads were aligned onto the Mouse Reference Genome (GRCm38/mm10) using Bismark v0.18.2 with Bowtie2-2.2.9 allowing one mismatch in a seed alignment. Only reads mapping uniquely to the genome were kept, and methylation calls were extracted after duplicate removal by considering only CpG dinucleotides covered by a minimum of five reads. CpG methylation levels over different genomic compartments were calculated by extracting methylation calls with positional overlap with coordinates for Gencode vM16 gene annotations ("Intragenic") and the RepeatMasker database ("Transposons"). CpG islands ("CGIs") were defined previously[54]. "Intergenic" partitions were defined as genomic regions that did not overlap with CGIs, intragenic regions or transposable elements. Differentially Methylated Region (DMR) calling was performed using the bioconductor package DSS[55] with the following parameters: a CpG methylation level difference of at least 25%, at least five CpGs called, minimum length of 200 bp, and at least 500 bp between two DMRs. CpG methylation data metaplots were processed using deepTools v2.5.3; only individual element annotations with size greater than 5 kb were analyzed for L1s, and greater than 500bp for MERVL and IAPEz (to account for solo LTR elements).

RNA-seq analysis. Adapters were trimmed using Atropos v1.1.16. Paired-end read alignment was performed onto the Mouse Reference Genome (mm10) with STAR (v2.7.0a) reporting randomly one position and allowing 6% mismatches (*--outFilterMultimapNmax 5000 --outSAMmultNmax 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax*

0.06). Repeat annotation was downloaded from RepeatMasker (http://www.repeatmasker.org/). To reconstruct full-length LTR copies, we used the perl tool "*One code to find them all*"**.** Reconstructed transposon annotation and basic gene annotation from GENCODE v18 were merged and used as input for quantification with FeatureCounts v1.5.1. Differential expression analysis was performed using edgeR's normalisation combined with voom transformation from limma R package. P-values were computed using limma and adjusted with the Benjamini-Hochberg correction. Genes and transposon families were declared as differentially expressed if FDR < 5% and log2FC >1.

Single-cell RNA-seq analysis. Raw count matrices were generated using Cell Ranger v3.0.2 and were imported to Seurat v3.1.1. To filter out low-quality cells, cells with at least 1000 genes detected and less than 15% of mitochondrial reads were kept for the following analyses. Log-normalization procedure (*NormalizeData* function) and detection of variable genes (*FindVariableFeatures* function) were performed for each sample independently. Samples were integrated by identifying anchors (*FindIntegrationAnchors and IntegrateData* functions) with 30 dimensions. Integrated data were scaled. The non-linear dimensional reduction technique UMAP was run with the 30 principal components stored after PCA analysis on the integrated data. Clusters were defined using a resolution parameter of 0.8.

Marker genes for each cluster were determined using edgeR package for the normalization and limma package for differential expression analysis if FDR<5% and log2 Fold-change > 0.25. Cell clusters were defined as germ cell clusters if they were *Ddx4* positive (universal germ cell marker) and *Gata4* negative (general somatic marker)[57]. Some clusters had a poor-quality control with low number of UMIs and detected genes, or barely no marker gene defined or too much mitochondrial genes. These clusters were excluded from the following analyses. A second round of cell selection was performed using the same method. The final analyses were done on 6,638 retained germ cells.

After identification of germ cell clusters, raw count matrices without testicular somatic cells were re-analyzed with Seurat using the same procedure as before for normalization,

sample integration, clustering (resolution parameter of 0.3) and detection of markers. The average expression (per germ cell cluster) of highly variable genes (*n*=2,000 genes) was used as input to calculate Spearman correlation between germ cell clusters.

Cell cycle analysis was performed using the *CellCycleScoring* function in Seurat. Genes specific to S and G2M phases were defined using the following criteria: expressed in spermatogonia and specific to one of the two phases based on a cell-cycle gene expression analysis[30].

Raw count matrices without testicular somatic cells were imported to Monocle2 (v2.12.0) and only expressed genes above threshold (0.1) were used for analyses. When two samples were included together in a pseudotime analysis, sample batch effects were removed via a model formula[58].

Velocyto (v0.17.16) was used to processed raw data to count spliced and unspliced reads for each gene and generated a loom file for each sample. Loom files were imported in R (v3.6.0) using SeuratWrappers (v0.1.0) R package. RNA velocity was estimated using *RunVelocity* function and Velocyto.R package (v0.6) (using gene-relative model with k=20 cell kNN pooling and using top/bottom 2% quantiles for gamma fit). The RNA velocity map was projected onto the UMAP plot with a neighborhood size of *n*= 200 cells[32]. RNA velocity could not be performed on *Dnmt3L*^{KO}, due to insufficient germ cells (1,031 WT and 1,182 *Dnmt3L*^{KO}) and the use of anterior version of the scRNA-seq kit (10X Genomics V2).

Cut&Run analysis. Paired-end reads were trimmed using Trim Galore v0.4.4. The alignment was performed onto a concatenated genome using the Mouse Reference Genome (mm10) and the *Escherichia coli* genome (str. K-12 substr. MG1655, Genbank: NC_000913) with STAR (v2.7.0e) reporting randomly one position, allowing 4% of mismatches (*--outFilterMultimapNmax 5000 --outSAMmultNmax 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --alignIntronMax 1 --alignMatesGapMax 2000*). PCR duplicates were removed using Picard v2.6.0 (http://broadinstitute.github.io/picard/). Peaks were called using MACS2 v2.1.1 using IgG sample as control. The broad option was used for

H3K27me3 samples whereas narrow peaks were detected for H3K27ac samples. Heatmap using peak centers as windows was performed using Deeptools v2.5.3. *De novo* motifs were called using HOMER v4.11 with the HOCOMOCO mouse motif database v11 as known motifs. Genomic coordinates of ENCODE candidate cis-Regulatory Elements (cCREs) were downloaded from the SCREEN website (https://screen.encodeproject.org/), and intersected with 284 *Dnmt3A*^{KO}-gained H3K27ac peaks (associated with nearby gene upregulation and loss of DNA methylation in *Dnmt3A*^{KO} SSCs) using bedTools (v2.27.1). Only peaks that fully overlapped with cCRE regions were considered.

References

1. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
2. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
3. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
4. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
5. Gifford, C. A. *et al.* Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013).
6. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
7. Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: In the right place at the right time. **1340**, 1336–1340 (2018).
8. Lord, T. & Oatley, J. M. Regulation of spermatogonial stem cell maintenance and self-renewal. in *The Biology of Mammalian Spermatogonia* 91–129 (Springer New York, 2017). doi:10.1007/978-1-4939-7505-1_5
9. Kubo, N. *et al.* DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. *BMC Genomics* **16**, 624 (2015).
10. Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539 (2001).
11. Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900–903 (2004).
12. Barau, J. *et al.* The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**, 909–912 (2016).
13. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in

male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).

14. Zamudio, N. *et al.* DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* **29**, 1256–1270 (2015).

15. Yaman, R. & Grandjean, V. Timing of entry of meiosis depends on a mark generated by DNA methyltransferase 3a in testis. *Mol. Reprod. Dev.* **73**, 390–397 (2006).

16. Kato, Y. *et al.* Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* **16**, 2272–2280 (2007).

17. Lengner, C. J. *et al.* Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell* **1**, 403–415 (2007).

18. Miura, F. *et al.* Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **47**, e85–e85 (2019).

19. Shovlin, T. C. *et al.* Sex-specific promoters regulate Dnmt3L expression in mouse germ cells. *Hum. Reprod.* **22**, 457–467 (2007).

20. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**, 247–257 (1999).

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

22. Yoshida, S. *et al.* The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development* **133**, 1495–1505 (2006).

23. Law, N. C., Oatley, M. J. & Oatley, J. M. Developmental kinetics and transcriptome dynamics of stem cell specification in the spermatogenic lineage. *Nat. Commun.* **10**, 2787 (2019).

24. Chan, F. *et al.* Functional and molecular features of the Id4+ germline stem cell population in mouse testes. *Genes Dev.* **28**, 1351–1362 (2014).

842 25. Helsel, A. R. *et al.* ID4 levels dictate the stem cell state in mouse spermatogonia.
843 *Development* **144**, 624–634 (2017).
844 26. Lord, T. & Oatley, J. M. A revised Asingle model to explain stem cell dynamics in the
845 mouse male germline. *Reproduction* **131**, 1796–1803 (2017).
846 27. Hermann, B. P. *et al.* The mammalian spermatogenesis single-Cell transcriptome,
847 from spermatogonial stem cells to spermatids. *Cell Rep.* **25**, 1650-1667.e8 (2018).
848 28. Green, C. D. *et al.* A comprehensive roadmap of murine spermatogenesis defined by
849 single-cell RNA-seq. *Dev. Cell* **46**, 651-667.e10 (2018).
850 29. Vasiliauskaitė, L. *et al.* Defective germline reprogramming rewires the spermatogonial
851 transcriptome. *Nat. Struct. Mol. Biol.* **25**, 394–404 (2018).
852 30. Roy Choudhury, D. *et al.* Microarray-Based Analysis of Cell-Cycle Gene Expression
853 During Spermatogenesis in the Mouse. *Biol. Reprod.* **83**, 663–675 (2010).
854 31. Suzuki, S., McCarrey, J. R. & Hermann, B. P. An mTORC1-dependent switch
855 orchestrates the transition between mouse spermatogonial stem cells and clones of
856 progenitor spermatogonia. *Cell Rep.* **34**, 108752 (2021).
857 32. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
858 33. King, A. D. *et al.* Reversible regulation of promoter and enhancer histone landscape
859 by DNA methylation in mouse embryonic stem cells. *Cell Rep.* **17**, 289–302 (2016).
860 34. Jermann, P., Hoerner, L., Burger, L. & Schubeler, D. Short sequences can efficiently
861 recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA
862 methylation. *Proc. Natl. Acad. Sci.* **111**, E3415–E3421 (2014).
863 35. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-
864 scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**,
865 1128–1138 (2012).
866 36. Cheng, K. *et al.* Unique epigenetic programming distinguishes regenerative
867 spermatogonial stem cells in the developing mouse testis. *iScience* **23**, 101596
868 (2020).
869 37. Smith, A. M. *et al.* A novel mode of enhancer evolution: The Tal1 stem cell enhancer

recruited a MIR element to specifically boost its activity. *Genome Res.* **18**, 1422–1432 (2008).

38. Domcke, S. *et al.* Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015).

39. Lea, A. J. *et al.* Genome-wide quantification of the effects of DNA methylation on human gene regulation. *Elife* **7**, 1–27 (2018).

40. Shirane, K., Miura, F., Ito, T. & Lorincz, M. C. NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. *Nat. Genet.* **52**, 1088–1098 (2020).

41. Song, H.-W. & Wilkinson, M. F. Transcriptional control of spermatogonial maintenance and differentiation. *Semin. Cell Dev. Biol.* **30**, 14–26 (2014).

42. Zhang, T., Oatley, J., Bardwell, V. J. & Zarkower, D. DMRT1 Is Required for Mouse Spermatogonial Stem Cell Maintenance and Replenishment. *PLOS Genet.* **12**, e1006293 (2016).

43. Stephens, D. C. & Poon, G. M. K. Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains. *Nucleic Acids Res.* **44**, 8671–8681 (2016).

44. Wu, H. *et al.* Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444–448 (2010).

45. Izzo, F. *et al.* DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* **52**, 378–387 (2020).

46. Ketkar, S. *et al.* Remethylation of Dnmt3a −/− hematopoietic cells is associated with partial correction of gene dysregulation and reduced myeloid skewing. *Proc. Natl. Acad. Sci.* **117**, 3123–3134 (2020).

47. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2012).

48. Jeong, M. *et al.* Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Rep.* **23**, 1–10 (2018).

898 49. Dahlet, T. *et al.* Genome-wide analysis in the mouse embryo reveals the importance of
899 DNA methylation for transcription integrity. *Nat. Commun.* **11**, 3153 (2020).
900 50. Lewandoski, M., Wassarman, K. M. & Martin, G. R. Zp3–cre, a transgenic mouse line
901 for the activation or inactivation of loxP-flanked target genes specifically in the female
902 germ line. *Curr. Biol.* **7**, 148–151 (1997).
903 51. Peters, A. H. F. ., Plug, A. W., van Vugt, M. J. & de Boer, P. A drying-down technique
904 for the spreading of mammalian meiocytes from the male and female germline.
905 *Chromosom. Res.* **5**, 66–71 (1997).
906 52. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution
907 mapping of DNA binding sites. *Elife* **6**, 1–35 (2017).
908 53. Didion, J. P., Martin, M. & Collins, F. S. Atropos: specific, sensitive, and speedy
909 trimming of sequencing reads. *PeerJ* **5**, e3720 (2017).
910 54. Illingworth, R. S. *et al.* Orphan CpG Islands Identify Numerous Conserved Promoters
911 in the Mammalian Genome. *PLoS Genet.* **6**, e1001134 (2010).
912 55. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect
913 differentially methylated loci from single nucleotide resolution sequencing data.
914 *Nucleic Acids Res.* **42**, e69–e69 (2014).
915 56. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
916 (2013).
917 57. Velte, E. K. *et al.* Differential RA responsiveness directs formation of functionally
918 distinct spermatogonial populations at the initiation of spermatogenesis in the mouse.
919 *Development* **146**, dev173088 (2019).
920 58. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories.
921 *Nat. Methods* **14**, 979–982 (2017).
922 59. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a
923 mechanism for epigenetic inheritance in the mouse. *genesis* **35**, 88–93 (2003).
924
925

EXTENDED DATA FIGURES

DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis

---

Mathilde Dura[1], Aurélie Teissandier[1], Mélanie Armand[1], Joan Barau[2], Lorraine Bonneville[1], Michael Weber[3], Laura G. Baudrin[4], Sonia Lameiras[4] and Deborah Bourc'his[1*]

[1] Genetics and Developmental Biology Department, Institut Curie, PSL Research University, INSERM, CNRS, Paris, France

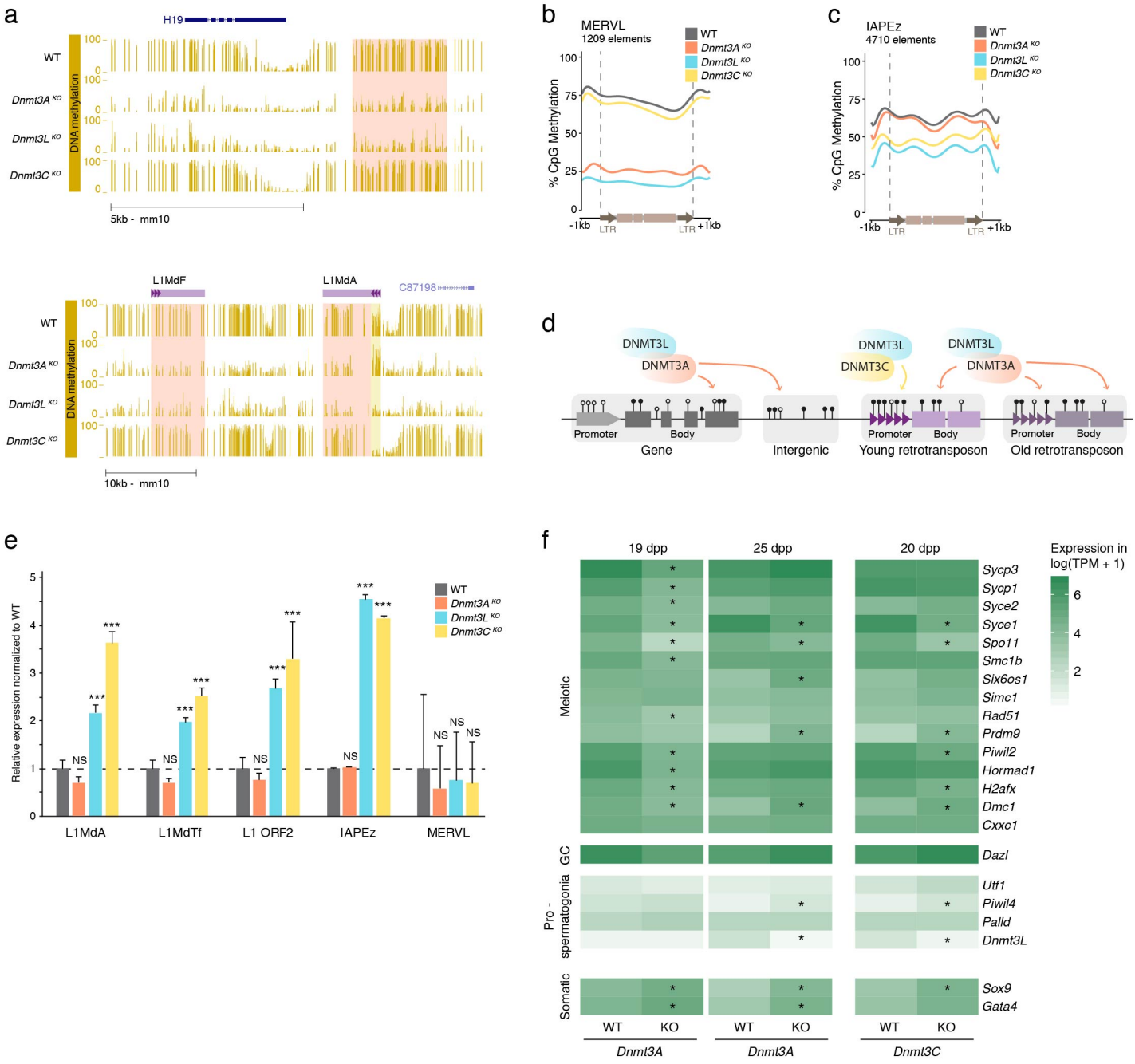
[2] Institute of Molecular Biology (IMB), Mainz, Germany

[3] Biotechnology and Cell Signaling, University of Strasbourg, CNRS, Illkirch Cedex, France

[4] ICGex Next-Generation Sequencing platform, Institut Curie, PSL Research University, 75005 Paris, France

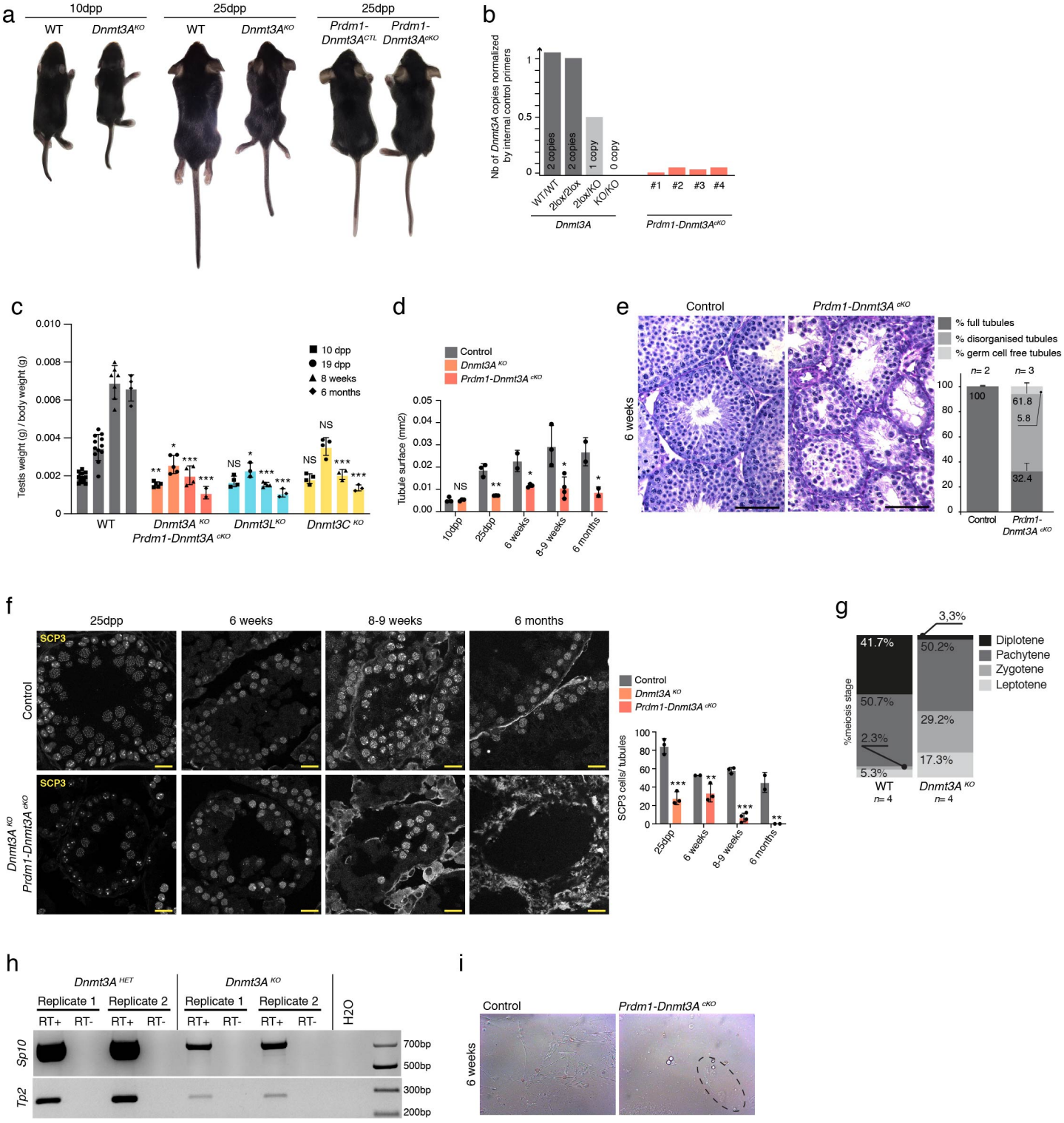
*corresponding author: deborah.bourchis@curie.fr

Extended data figure 1



Extended data Fig. 1 | DNMT3A is not required for retrotransposon silencing. **a**, Genome browser representation showing DNA methylation levels assessed by WGBS at E18.5 in WT, *Dnmt3A*^{KO}, *Dnmt3L*^{KO} and *Dnmt3C*^{KO} prospermatogonia. Top panel: the highlighted region (pink) coincides with the imprinting control region of the H19-Igf2 locus, with decreased methylation in *Dnmt3A*^{KO} and *Dnmt3L*^{KO}. Bottom panel: an evolutionary ancient L1MdF element is hypomethylated throughout its length in *Dnmt3A*^{KO} (pink), while a recent L1MdA element shows hypomethylation over its promoter in *Dnmt3C*^{KO} (yellow) and over its body in *Dnmt3A*^{KO} (pink). **b and c**, Metaplots of DNA methylation levels over full-length MERVL (b) and IAPEz (c) comparing WT, *Dnmt3A*^{KO}, *Dnmt3L*^{KO}, and *Dnmt3C*^{KO}. IAP methylation levels are less decreased compared to other retrotransposon families, in agreement with a subset of IAPEz copies being resistant to DNA methylation reprogramming in PGCs[59]. **d**, Model of division of function for de novo DNA methylation in fetal male germ cells. DNMT3C selectively methylates the promoters of evolutionarily young retrotransposons. DNMT3A methylates the rest of the genome, including the body of young retrotransposons and the entire length of old retrotransposons. DNMT3L is a universal co-factor that stimulates both activities. **e**, Fold change expression (KO/WT) of different retrotransposon families in 19dpp whole testes, assessed by RT-qPCR normalized by beta-actin expression. Data are mean ± normalized SD (black bar) from at least three biological replicates, (student t-test over WT, *p<0.05, **p<0.005, ***p<0.0005). **f**, RNA-seq expression in log(TPM+1) (TPM, transcript counts per million) of genes specifically expressed during meiosis in prospermatogonia in 19dpp and 25dpp *Dnmt3A*^{KO} and WT testes and in public datasets from 20dpp *Dnmt3C*-^{KO} and WT[12].

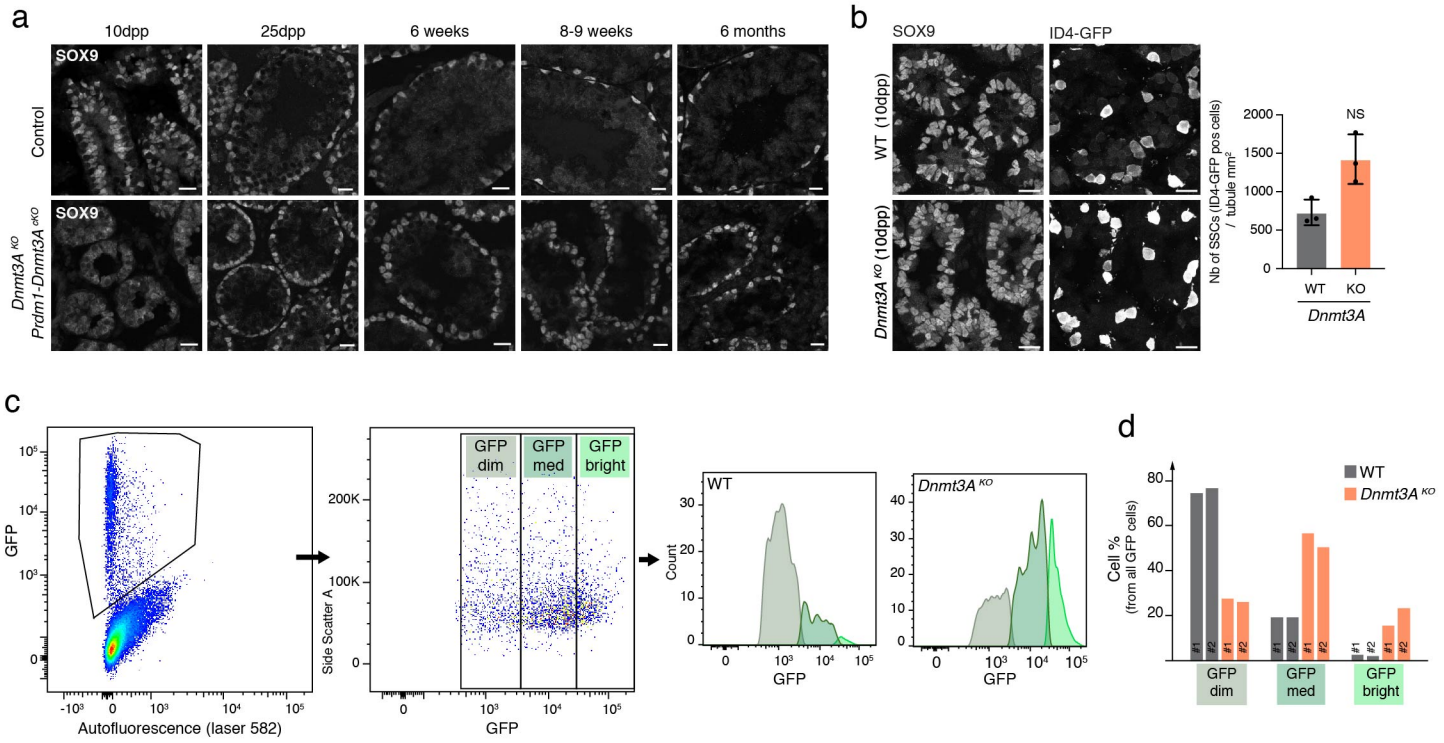
Extended data figure 2



Extended data Fig. 2 | Dnmt3A mutants complete the first wave of spermatogenesis until spermatozoa with a reduced germ cell numbers. **a**, Representative photograph of a smaller constitutive *Dnmt3A*^{KO} animal with its WT littermate at 10ddp (left panel) and 25dpp (middle panel), and a germ cell-conditional *Prdm1-Dnmt3A*^{cKO} animal with its control littermate at 25dpp (right panel). All animals are males. **b**, Genomic qPCR of *Dnmt3A* copy number. *Dnmt3A*^{WT}, *Dnmt3A*^{2lox}, *Dnmt3A*$^{2lox/KO}$, *Dnmt3A*^{KO} correspond to liver DNA. Genomic DNA from *Prdm1-Dnmt3A*^{cKO} (# indicates individual replicates) was extracted from EpCAM-pos, β2M-neg FACS-sorted germ cells (10dpp). Germ cell sorting purity assessed by TRA98 staining was the following: #2: 79%, #3: 96.8%, #4: 93.3%. **c**, Testicular weight normalized by body weight at 10dpp, 19dpp, 8 weeks and 6 months. Data are mean ± normalized SD (black bar) from biological replicates, and individual points represent biological replicates, (student t-test over WT, NS = non significant, *p<0.05, **p<0.005, ***p<0.0005). **d**, Mean of tubule surface in mm2 in *Dnmt3A*^{KO}, *Prdm1-Dnmt3A*^{cKO} and controls at 10dpp, 25dpp, 6weeks, 8-9weeks and 6 months. Data are mean ± normalized SD (black bar) from biological replicates, and individual points represent biological replicates, (student t-test with p values as in c). **e**, (Left) Representative images of Periodic Acid Shift (PAS)-stained testis sections from germ cell-conditional *Prdm1-Dnmt3A*^{cKO} mutants and controls at 6 weeks (scale, 50 µm). (Right) Quantification of the percentage of different classes of tubules per genotype. Data are mean ± normalized SD (black bar) from biological replicates, n = number of animals. Control genotypes: *Dnmt3A*$^{2lox/KO}$; *Prdm1-Cre*$^{0/0}$ and *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$. **f**, (Right) Representative image of SCP3 (meiotic marker)

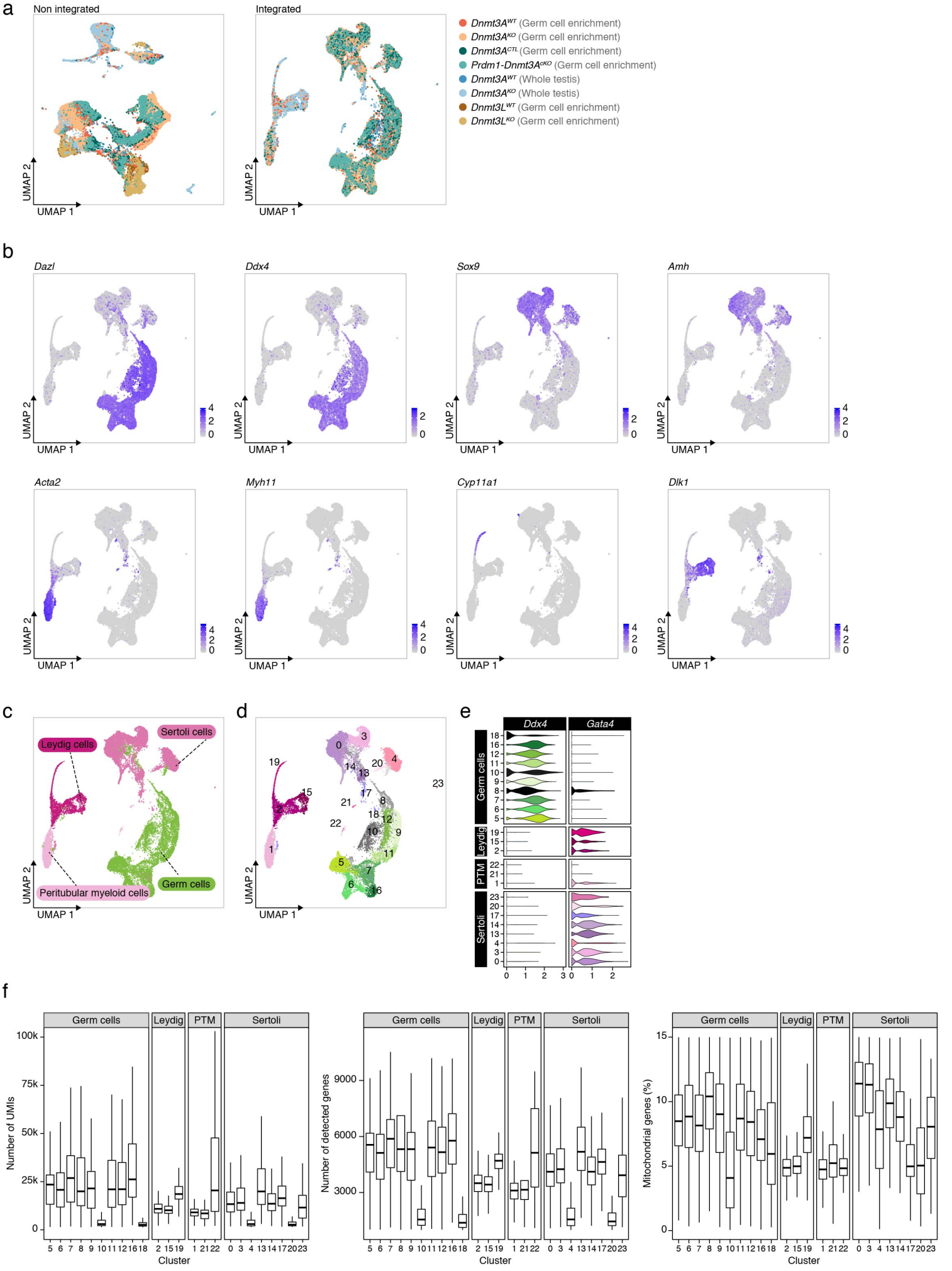
staining on testis sections from *Dnmt3A*^{KO}, *Prdm1-Dnmt3A*^{cKO} and controls at 25dpp, 6 weeks, 8-9weeks and 6 months (scale, 20 µm). (Left) Quantification of SCP3-positive cells normalized per tubule. Control genotypes at 6 weeks: *Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$ and *Dnmt3A*$^{2lox/KO}$; *Prmd1-Cre*$^{0/0}$ - at 8-9 weeks: *Dnmt3A*$^{2lox/KO}$; *Prmd1-Cre*$^{0/0}$ (n= 2) and *Dnmt3A*$^{2lox/WT}$; *Prdm1-Cre*$^{0/0}$ - at 6 months: *Dnmt3A*$^{KO/WT}$; *Prmd1-Cre*$^{Tg/0}$ and *Dnmt3A*$^{2lox/WT}$; *Prmd1-Cre*$^{0/0}$. Data are mean ± SD (black bar), individual points represent biological replicates (student t-test over WT *p<0.05, **p<0.005, ***p<0.0005). **g**, Percentage of the different stages of the prophase of first meiosis assessed by double-immunofluorescence detection of SCP3 and γH2AX on WT and *Dnmt3A*^{KO} (25dpp) meiotic spreads (representative image on Fig. 2e). Data are mean for biological replicates, n = number of animals, at least 100 cells were counted per replicate. **h**, RT-PCR detection of spermatid-specific markers (Sp10 and Tp2) in *Dnmt3A*$^{KO/WT}$ and *Dnmt3A*$^{KO/KO}$ whole testes (25dpp). Two biological replicates for each genotype. **i**, Representative photograph of spermatozoa collected after epididymis squeezing in a conditional *Dnmt3A*^{cKO} mutant and control littermate (*Dnmt3A*$^{KO/WT}$; *Prdm1-Cre*$^{Tg/0}$) at 6 weeks.

Extended data figure 3



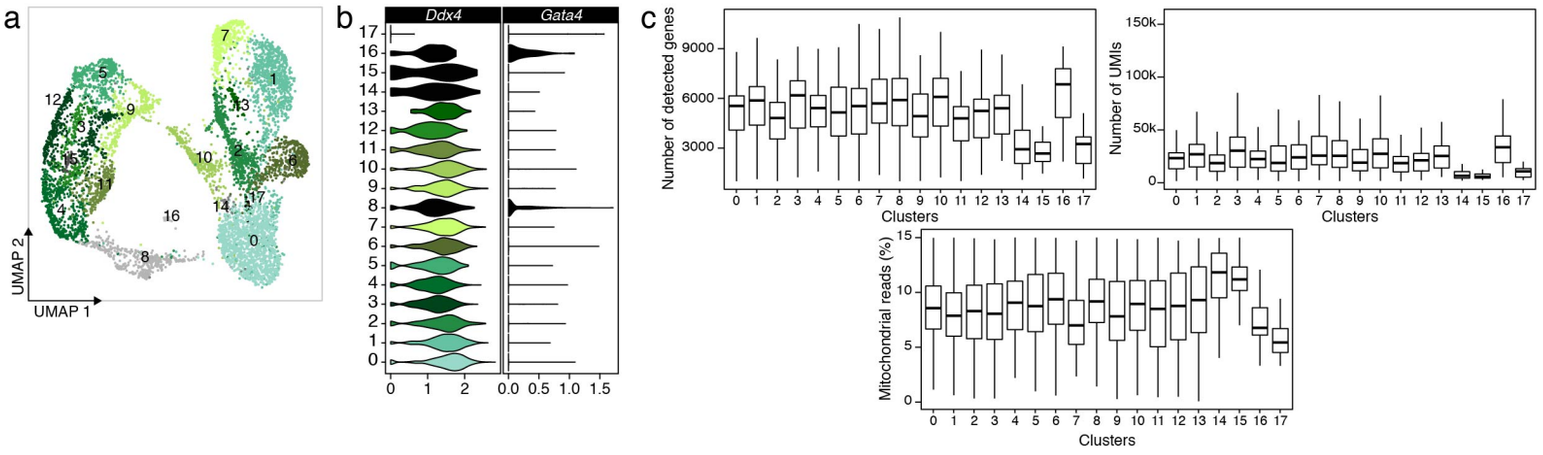
Extended data Fig. 3 | Accumulation of ID4-eGFP positive germ cells in Dnmt3A^{KO} testes. **a**, Representative image of SOX9 (Sertoli cell marker) staining on testis sections from *Dnmt3A^{KO}*, *Prdm1-Dnmt3A^{cKO}* and controls at 10dpp, 25dpp, 6 weeks, 8-9weeks and 6 months. Scale, 20 µm. **b**, (Left) Representative microscopy images of double-immunofluorescence detection of SOX9 (Sertoli cell marker) and ID4-eGFP (SSC) at 10dpp. Scale, 20 µm. (Right) Quantification of the number of ID4-eGFP positive cells per mm2 of tubule in *Dnmt3A^{KO}* and WT. Data are mean ± SD (black bar) and individual points represent biological replicates (student t-test over WT, NS: non-significant). **c**, (Left) Representative FACS analysis plots of live ID4-eGFP-positive gated testicular cells from 10dpp mice. ID4-eGFP cells were divided into three classes: GFP dim, GFP med and GFP bright. (Right) Representative count of cell numbers in each GFP subclass in *Dnmt3A^{KO}* and WT. **d**, Cell count in percentage for each GFP sub class in WT and *Dnmt3A^{KO}* for two biological replicates. #1 et #2 represent individual replicates.

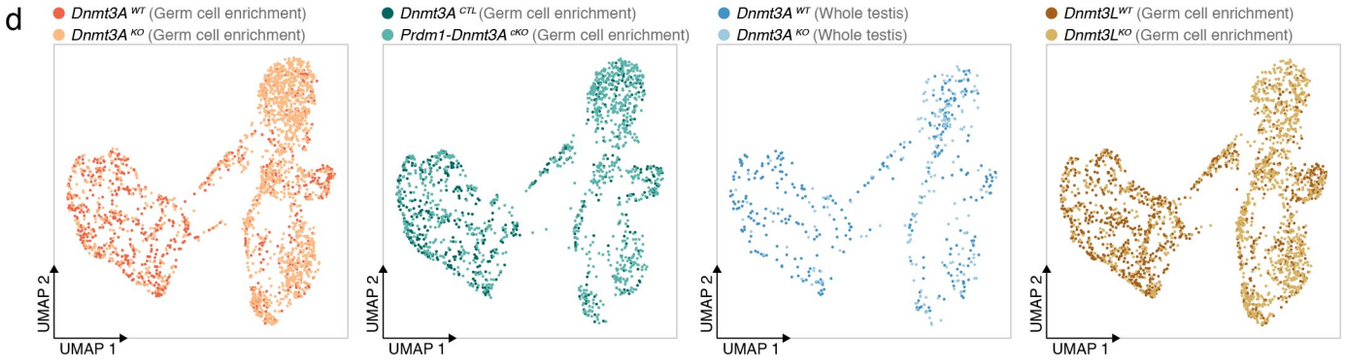
Extended data figure 4

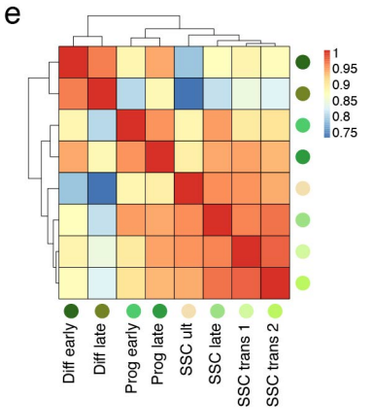


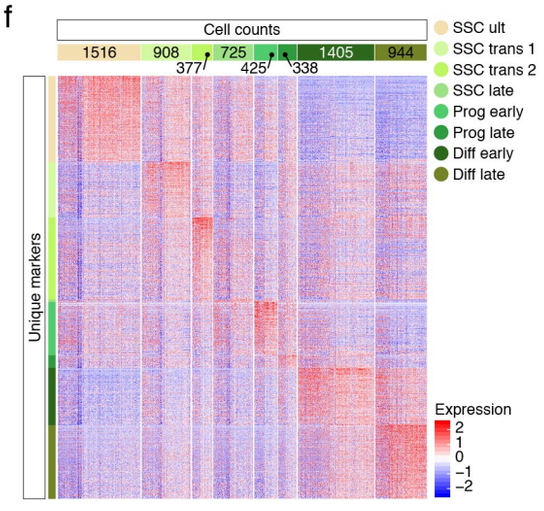
Extended data Fig. 4 | scRNA-seq quality controls and cluster annotation of germ and somatic cells. a, (Left) UMAP dimensionality reduction representation of scRNA-seq data from 10dpp testicular cells before batch effect correction of eight samples, each sample is one animal of the genotype of interest (41,582 cells). Colors represent different conditions, see key. (Right) UMAP dimensionality reduction representation of the same sample after batch effect correction (see Methods, integrated data). **b**, UMAP representation of scRNA-seq data from 10dpp testicular cells (n= 8 animals from different genotypes) with a color gradient representing the expression of known markers of germ cells (Dazl and Ddx4), Sertoli cells (Sox9 and Amh), peritubular myeloid cells (PTM) (Myh11 and Acta2) and Leydig cells (Dlk1 and Cyp11a1). **c**, UMAP representation of scRNA-seq data from 10dpp testicular cells, colors represent cellular types based on marker expression of b. **d**, Unbiased cell clustering onto an UMAP representation of scRNA-seq data from 10dpp testicular cells demonstrated 24 clusters, each color represents a cluster annotated by a number (from 0 to 23). **e**, Violin plots show expression of key cell-type specific markers among cell clusters to distinguish those containing germ cells (Ddx4-positive) from testicular somatic cells (Gata4-positive). Cluster 8 presented expression of the two markers: this germ cell cluster was considered as somatically contaminated—probably due to cellular doublets—and was removed from further analysis. Cluster in black was removed from analysis because of poor quality control (see Extended data Fig. 4f). **f**, Quality control of scRNA-seq integrated data split by cellular cluster, presenting the number of UMIs (left panel), the number of detected genes (middle panel) and the percentage of mitochondrial reads (right panel).

Extended data figure 5



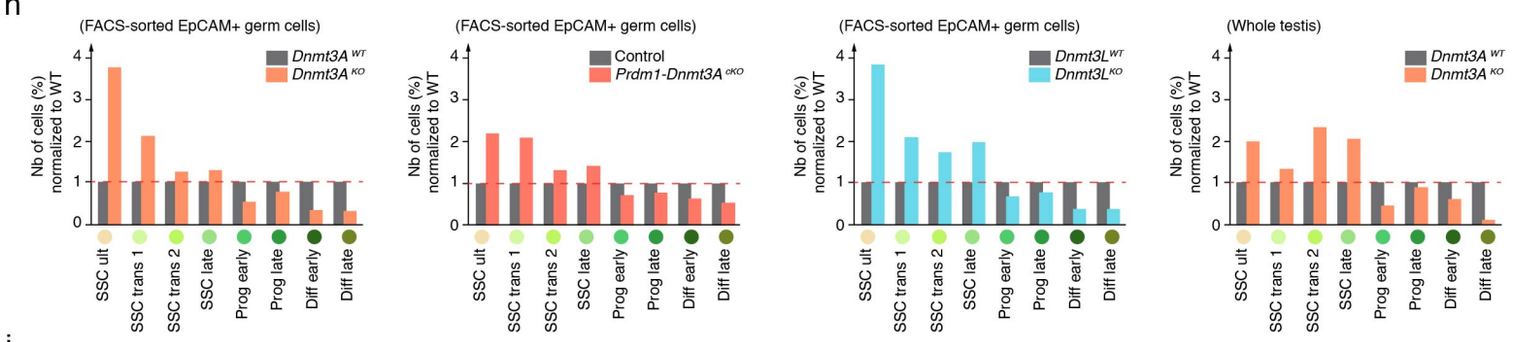


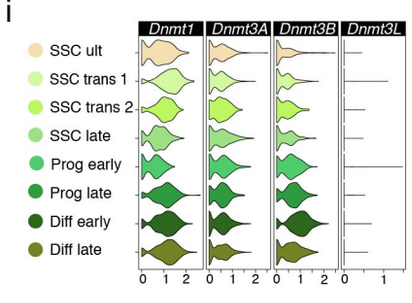




g

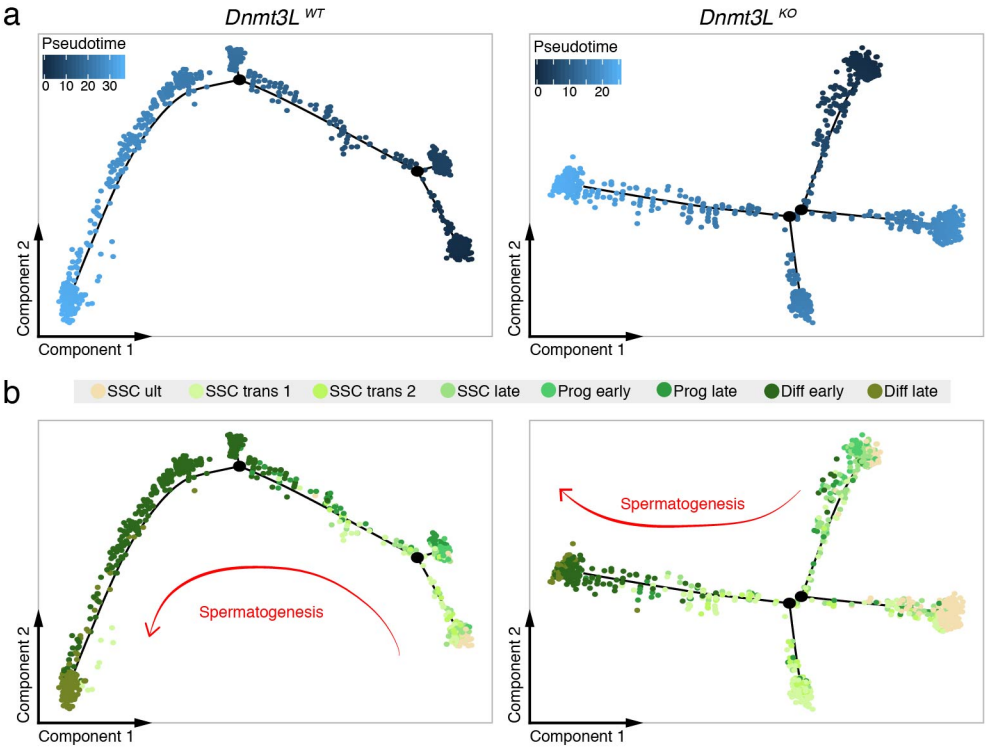
Cell cluster	Marker numbers	Top 10 markers (expression) Top 10 markers (p value)
SSC ult	329	Ret, Rpl22l1, Rasgrp2, Gpsm3, Malat1, Ccnd2, Rps4x, Tnrc6c, Chd4, Sdc4 Gfra1, Ret, Tcl1, Foxc2, Etv5, Dusp6, Dppa4, Mcam, Egr2, Lhx1
SSC trans 1	279	Pclaf, 1810037I17Rik, Lig1, Atad2, Tubb5, Top2a, CCnd2, Sdc4, Malat1, Rpl22l1 Pclaf, hist1h2ae, Prim1, Hist1h1b, Pbk, Lig1, Atad2, 1810037I17, Fam111a, Ranbp1
SSC trans 2	468	Ube2c, Ccnb1, Cdc20, Cks2, Arl6ip1, Cenpa, Ube2s, Tubb5, Top2a, Cenpf Ube2c, Nusap1, Cdc20, Ccnb1, Cks2, Cenpa, Arl6ip1, Cdkn3, Ube2s, Ccnb2
SSC late	62	Rbp1, Stra8, Calm2, Crapb1, Kit, Uchl1, Marcksl1, Il11, Stmn1, AC160336.1 Psmd7, Ptgr1, Rragd, Glrx2, Sdc4, Egr4, H2afy2, Ube2q2, Sox4, Aldoa
Prog early	293	Crapb1, Dynll2, Marcksl1, Ppp1r14b, Stmn1, P3h3, Hmgn2, Cenpf, Tubb4b, AC160336.1 Lgfbp6, Upp1, Ddit4, Foxf1, Egr4, Onecut2, Six4, Neurog3, Rin2, Nanos3
Prog late	135	Ly6k, Prss50, Meioc, 1700013H16Rik, Tex101, Defb19, Aard, Tmsb4x, Sycp3, Stra8 Stra8, Sall4, Pim1, Siva1, H2-Q7, Cebpb, Lgals2, Rhox10, Pou5f1, Jund
Diff early	293	Esx1, Rps27l, Cited1, Hmgn2, Atp5f1, Tubb4b, H1fx, Carpb1, Hist1h2ap, Nasp Crapb1, Pcsk1n, Kit, Lmo1, Marcksl1, Rbp1, Rnd2, Cdk6, Tbx1, Bcl7c
Diff late	460	Egra, Psmd7, Rragd, Sdc4, Sys1, Sox4, Serf1, Tubb5, Rps4x, Rpl22l1 Esx1, Dmrtb1, Meioc, Rhox13, Prss50, Prss50, Nt5dc2, Cited1, 1700013H16Rik, Baz1a, Ssr2





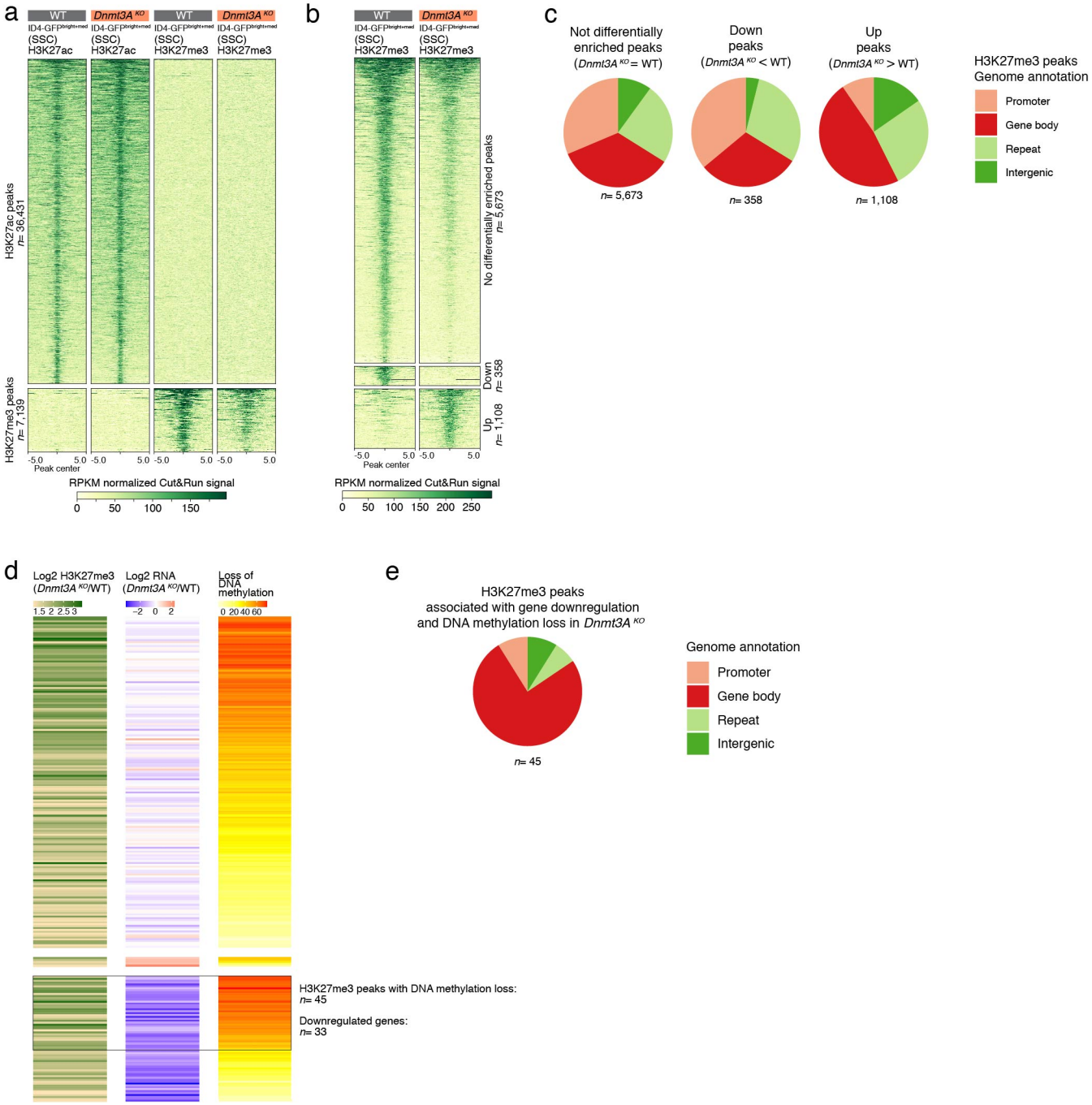
Extended data Fig. 5 | Supporting evidence to name final germ cell clusters. **a**, Unbiased cell clustering onto an UMAP representation of scRNA-seq data from 10dpp germ cells (n = 8 independent animals from different genotypes). Germ cells were selected from clusters 5, 6, 7, 9, 11, 12, 16 based on Dxd4 expression and quality controls. Identification of 18 clusters (from 0 to 17), colors represent the different clusters, cluster in grey (#8) was removed from final analysis (see Extended data Fig. 5b,c). **b**, Violin plots show expression of key cell-type specific markers among germ cell clusters to distinguish those containing germ cells (Ddx4-positive) from testicular somatic cells (Gata4-positive). The five clusters in black were removed from further analysis: clusters 14, 15 and 17 because of poor sequencing quality (Extended data Fig. 5c), and clusters 8 and 16 because of somatic contamination probably due to cellular doublets. **c**, Quality control of scRNA-seq integrated data on germ cell clusters. Cluster 14, 15 and 17 were removed from further analysis for poor sequencing quality. **d**, UMAP dimensionality reduction representation of scRNA-seq integrated data from 10dpp germ cells (6,638). Colors represent the different samples and conditions, red: *Dnmt3A*^{WT} FACS-enriched germ cells, orange: *Dnmt3A*^{KO} FACS-enriched germ cells, dark green: *Dnmt3A*^{CTL} (*Dnmt3A*$^{2lox/KO}$; *Prdm1-Cre*$^{0/0}$) FACS-enriched germ cells, light green: *Prdm1-Dnmt3A*^{cKO} (*Dnmt3A*$^{2lox/KO}$; *Prdm1-Cre*$^{Tg/0}$) FACS-enriched germ cells, dark blue: *Dnmt3A*^{WT} all testicular cells, light blue: *Dnmt3A*^{KO} all testicular cells, brown: *Dnmt3L*^{WT} FACS-enriched germ cells, yellow: *Dnmt3L*^{KO} FACS-enriched germ cells. **e**, Heatmap of pairwise Spearman correlation matrix among the 8 germ cell clusters. **f**, Heatmap of single cell expression profile of markers defined in only one cluster for each germ cell cluster. Numbers on the upper row represent the number of cells in each germ cell cluster. **g**, Table of top 10 markers for each germ cell cluster. Marker numbers correspond to the total number of identified markers that give a signature for each cluster. **h**, Bar plot showing the percentage of cells normalized to WT per cell type for each mutant and condition. **i**, Violin plots show expression of Dnmt3 genes across germ cell populations.

Extended data figure 6



Extended data Fig. 6 | DNMT3L is essential to SSC plasticity. a, **b**, Pseudotime trajectory of germ cells from 10dpp testes. (Left) *Dnmt3L*^{WT} germ cells (FACS-sorted). (Right) *Dnmt3L*^{KO} germ cells (FACS-sorted). In (a), cells were ordered from beginning (dark blue) to the end (light blue). In (b), cells were colored following germ cell cluster allocation, see key above. Unfortunately, RNA velocity could not be performed on *Dnmt3L*^{KO}, due to insufficient germ cells (1,031 WT and 1,182 *Dnmt3L*^{KO}) and the use of anterior version of the scRNA-seq kit (10X Genomics V2).

Extended data figure 7



Extended data Fig.7 | Relationship between H3K27me3 distribution, gene expression and DNA methylation in *Dnmt3A^{KO}* SSCs (ID4-eGFP$^{bright+med}$). a, Heat map showing levels of H3K27ac and H3K27me3 peak enrichment in RPKM normalized Cut&Run signal for ID4-eGFP$^{bright+med}$ germ cells (SSCs) from 10dpp *Dnmt3A^{KO}* and WT males. Enrichment is assessed +/- 5kb from the center of the peak. Number of biological replicates: *Dnmt3A^{KO}*= 3, WT= 4. There is no reciprocal gain or loss of one lark at the expense of the other *Dnmt3A^{KO}* germ cells. b, Heatmap showing levels of H3K27me3 peak enrichment in RPKM normalized Cut&Run signal for ID4-eGFP$^{bright+med}$ germ cells (SSCs) in 10dpp *Dnmt3A^{KO}* and WT males. Peaks are divided into three categories: not differentially enriched (*Dnmt3A^{KO}* = WT), down-enriched (*Dnmt3A^{KO}* < WT), up-enriched (*Dnmt3A^{KO}* > WT) (FDR<5% and FC>1). c, Genomic annotation of each category of H3K27me3 peaks. d, Heatmap focusing on i) *Dnmt3A^{KO}*-gained H3K27me3 peaks (log2 FC) compared to WT (left), ii) differential expression (log2 FC) of genes associated with H3K27me3 peaks (TSS +/- 5kb from the peak) between *Dnmt3A^{KO}* and WT—from ID4-eGFP$^{bright+med}$ RNA-seq—(center), and iii) percentage of DNA methylation loss in *Dnmt3A^{KO}* versus WT on H3K27me3 peak genomic location—from E18.5 WGBS—(right). Rows are ordered according to gene expression changes: top, not differentially expressed; middle, upregulated; bottom, downregulated. e, Genomic annotation of *Dnmt3A^{KO}*-gained H3K27me3 peaks, associated with downregulated genes and >30% DNA methylation loss on H3K27me3 peak location.

Extended data figure 8

Extended data Fig. 8 | Relationship between H3K27ac distribution, gene expression and DNA methylation in Dnmt3A^{KO} SSCs (ID4-eGFP$^{bright+med}$). a, Genomic annotation of each category of H3K27ac peaks. **b**, Quantification of intersection between H3K27ac peaks from *Dnmt3A^{KO}*, *Dnmt3A^{WT}*, WT SSC (ID4-eGFP^{bright}) and progenitors (ID4-eGFP^{dim}) (from published data[36]), down-enriched peaks and up-enriched peaks. Intersections lower than 200 were removed from the figure. **c**, Genome browser representation of the Gata2 locus, showing H3K27ac enrichment (grey), DNA methylation (yellow) and RNA expression (red) for *Dnmt3A^{KO}* and WT. Regions showing *Dnmt3A^{KO}* de novo H3K27ac peaks associated with >30% DNA methylation loss are shaded in light blue; regions showing de novo H3K27ac peaks at regions not controlled by DNA methylation are shaded in grey-blue. **d**, Violin plots showing expression levels of TF factors that have the best match with the motifs present in *Dnmt3A^{KO}*-specific H3K27ac peaks. Expression is represented across SSCs and spermatogonia subtypes discriminated by scRNA-seq for *Dnmt3A^{KO}* and WT. Upper panel: 17 TFs with SSC/spermatogonia expression, lower panel: 18 TFs not expressed in SSC/spermatogonia. **e**, Venn diagram representation of the overlap between 142 upregulation genes associated with *Dnmt3A^{KO}*-gained H3K27ac peaks that lose DNA methylation (see Fig. 5e), and 116 upregulated genes defined as putative targets of TFs with binding motif sites within *Dnmt3A^{KO}*-gained H3K27ac peaks (see Fig. 5i).

2 Chromatin determinants of DNMT3C targets

In this second part of my PhD work, I investigated the specificity of DNMT3C targeting towards young TE promoters. First, I established a classification of all TEs according to their subfamilies and segments (promoter 5'UTR or 5'LTR, internal, 3'LTR, full length elements versus solo LTRs) and their DNA methylation dynamics during prospermatogonia development, in particular whether they acquired DNA methylation in a DNM3A- or/and DNMT3C-dependent manner. Then, I tried to address the following questions: are there any chromatin signatures that distinguish DNMT3C-targeted TEs in prospermatogonia? How do these chromatin marks evolve before, during and after DNMT3C methylating function, and are they altered in the absence of DNMT3C? Can DNMT3C also recognize young TE promoters outside of the male fetal germline/prospermatogonia, in a cellular system of ectopic DNMT3C expression where the piRNA pathway is not effective?

The presented data and analyses are the results of a joint work between Joan Barau (ex-postdoctoral fellow in the lab, now PI at the IMB Mainz) and myself. More specifically, Joan Barau generated transgenic DNMT3C- and DNMT3B-expressing mouse embryonic stem cell lines and produced whole-genome DNA methylation maps in these cells and in the prospermatogonia of various *Dnmt3* mutants. I have carried out chromatin mark profiling by Cut&Run at different timepoints of prospermatogonia development and I have oriented and supervised all the bioinformatic analyses carried out by our bioinformatician Aurélie Teissandier, on *in vivo* material and ES cells, from datasets we generated and datasets that were publicly available.

This study is close to completion and will be compiled in a manuscript with the tentative title:

DNMT3C-targeted retrotransposons demonstrate a switch from bivalent H3K4me3/H3K9me3 to H3K9me3-only chromatin signature.

Mathilde Dura*, Joan Barau*, Aurélie Teissandier, Maxim V.C. Greenberg, and Deborah Bourc'his.

*contributed equally to this work

INTRODUCTION

Transposable elements (TEs) contribute a tremendous portion of mammalian genomes, making up to approximately 50% of the DNA mass in the mouse. These elements have the ability to move and to insert at a new position within the genome. Over evolutionary time, TEs can be beneficial, providing genetic innovations and diversity (Chuong et al., 2017; Jangam et al., 2017). However, in the short term, TE activity has a deleterious impact on genomic architecture and function by creating DNA damage and insertional mutagenesis, by promoting chromosomal rearrangements or by interfering with gene expression patterns (Hancks and Kazazian, 2016).

In the mouse, retrotransposons are divided into two classes: the LTR (long terminal repeat) retrotransposons, comprising the endogenous retroviruses family (including ERVK, ERV1 and ERVL) and the non-LTR retrotransposons, comprising the LINE (Long INterspersed Elements) and SINE (Short INterspersed Elements) (Rodriguez-Terrones and Torres-Padilla, 2018). The retention of TE activity, particularly for LINEs, is largely dependent on their evolutionary age, i.e., the time at which a specific element invaded the host genome. Contrary to ancient elements that are mostly transcriptionally inert by mutation accumulation throughout evolution (Sookdeo et al., 2013), young elements possess an intact promoter that confers transcriptional capacity and therefore, retrotransposition competency. There are several repressive pathways that function to repress young TEs, acting at the transcriptional level, through DNA methylation and repressive histone marks, or at the post-transcriptional level, through RNA methylation, editing or interference (Deniz et al., 2019).

Strict TE control is particularly essential in the germline to protect the hereditary material, and even more particularly in the male germline, whereby stem cell-driven spermatogenesis relies on continuous cell divisions, which provide opportunities for TE integration and amplification. The life cycle of TEs is tightly linked to DNA methylation reprogramming during male germ cell development. In primordial germ cells (PGCs), DNA methylation is firstly erased on a global scale to remove somatic patterns. Genome-wide reprogramming results in extremely low DNA methylation levels (around 5-7% of CpG methylation) in fetal PGCs at embryonic day E12.5 (Kobayashi et al., 2013; Seisenberger et al., 2012), which allows transcription-

competent TEs to temporarily escape silencing and to acquire active chromatin marks, including trimethylation of lysine 4 of histone H3 (H3K4me3) (Yamanaka et al., 2019). A notable exception are some members of the Intracisternal A Particle (IAP) family of ERVK elements, which are innately resistant to PGC methylation reprogramming (Hajkova et al., 2002). Male germline-specific DNA methylation is then established from E13.5 to birth by the *de novo* DNA methyltransferases DNMT3 enzymes assisted by the DNMT3L co-factor, in mitotically-arrested prospermatogonia. Ancient TEs and the body of young TEs are remethylated as a default pathway, like the rest of the genome, through the action of DNMT3A (Dura et al., 2021). However, the remethylation of the promoters of young TEs requires DNMT3C, a highly TE-selective enzyme that has occurred some 60 million years ago in rodent genomes, by tandem duplication of *Dnmt3B* (Barau et al., 2016; Molaro et al., 2020). In the absence of DNMT3C, lack of DNA methylation at young TE promoters is associated with their reactivation at meiosis, subsequent interruption of spermatogenesis at the pachytene stage and complete male sterility (Barau et al., 2016).

The determinants of DNMT3C selectivity towards the promoters of young TEs—that represent only 1% of the total genome—are not understood. DNMT3C shows some rapidly evolving residues in its most N-terminal part, which suggests participation in a genetic conflict, may be through direct interactions with TE sequences (Molaro et al., 2020). Specific relationships with underlying chromatin marks may also play a role. Although 70% similar at the amino-acid level, DNMT3C demarcates from DNMT3B by a lack of a PWWP domain, which is required for DNMT3B preferential targeting to H3K36me3, a chromatin hallmark found in the bodies of transcriptionally active genes (Baubec et al., 2015). Finally, genetic evidence suggests that PIWI-interacting RNAs (piRNAs) may act upstream of DNMT3C: genetic deletion of *Mili* or *Miwi2* results in the same selective lack of methylation at young TE promoters and to identical phenotypes as *Dnmt3C* male mutants (Barau et al., 2016; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2004). It is likely that, following the demethylation of young TEs in PGCs, *i)* MILI and MIWI2 process TE transcripts into piRNAs in the cytoplasm of fetal prospermatogonia, *ii)* TE-derived piRNAs are transferred into the nucleus by MIWI2, *iii)* MIWI2-loaded piRNAs recognize nascent transcripts emanating from young TE promoters with homologous sequences and *iv)* and this promotes the establishment DNMT3C-dependent DNA methylation of young TE promoters and long-term epigenetic silencing. However, the nature of the interaction between MIWI2 and

DNMT3C is still an enigma. Although some chromatin players have recently been identified as acting downstream of the MIWI2/piRNA pathway (Pastor et al., 2014; Schöpp et al., 2020; Zoch et al., 2020), whether DNMT3C physically interacts with these or whether it recognizes some chromatin states shaped by these components is unknown.

In *Drosophila melanogaster,* the piRNA pathway is known to be instrumental for guiding H3K9me3—a chromatin mark associated with heterochromatin formation and transcriptional repression—at the promoters of TEs (H3K9me3) (Aravin et al., 2008). In mammals, H3K9me3 and DNA methylation have a positive correlation (Fu et al., 2020; Meissner et al., 2008) and it was even shown that H3K9me3 can recruit the DNA methylation machinery in certain contexts (Auclair et al., 2016). In the mouse germline, the piRNA pathway was also suggested to recruit H3K9me3 at L1 promoters, based on the observation of a H3K9me3 decrease in post-natal germ cells of *Miwi2*^{KO} males (Pezic et al., 2014). However, the ChIP-seq assay was performed at 10dpp, at the time germ cells enter meiosis and TEs get reactivated in piRNA mutants. It is therefore unclear as to whether the H3K9me3 defect is primary or secondary to L1 reactivation in these mutants. We finally do not know whether H3K9me3 could be an intermediate between the piRNA and the DNA methylation pathways and more specifically, whether this mark may be instructive for recruiting DNMT3C on young TE promoters.

Here, we investigated whether the chromatin features of young TE promoters could elucidate the selective mechanism of DNMT3C recruitment. Using whole genome DNA methylation profiling of fetal prospermatogonia of wildtype and various *Dnmt3* mutants, we first categorized in a precise and exhaustive manner TEs that are resistant to PGC DNA methylation reprogramming, DNMT3C-dependent TEs and DNMT3A-dependent TEs. Then, we revealed that DNMT3C-targeted TE sequences are specifically marked by bivalent chromatin marks prior to *de novo* DNA methylation, consisting in the association of active H3K4me3 and repressive H3K9me3 marks, and then undergo a switch in H3K9me3-only enrichment. Finally, we showed that outside of the male germline, upon ectopic expression in embryonic stem cells, DNMT3C also recognized young TEs, and these also displayed H3K9me3 enrichment.

RESULTS

ERVKs are resistant to DNA demethylation, while L1s are demethylated anew

During male germ cell development, DNA methylation is erased globally through genome-wide reprogramming, and then re-installed in two waves in fetal prospermatogonia: the genome bulk (genes, intergenic regions) is remethylated first, while young TEs show delayed remethylation timing (Molaro et al., 2014). However, the precise kinetics of TE DNA methylation has never been precisely monitored in peri-natal prospermatogonia, along with information as to which DNMT3 enzyme is involved.

To address this, we sorted prospermatogonia by fluorescence-activated cell sorting (FACS), every day from ages E12.5 to E21.5 (equivalent to 2 days post-partum, dpp), using an *Oct4*-eGFP mouse transgenic line (**Extended data Fig. 1A**). Using pyrosequencing, we quantified CpG methylation levels on promoters of IAPs (ERVK), and young representatives of the L1 family, namely L1A and L1Tf. First, this timeline confirmed that IAPs showed resistance to PGC demethylation and retained relatively high CpG methylation content at E12.5, above 40% on average (**Fig. 1A**), compared to the genome bulk that reaches ~5-7% DNA methylation at this age (Kobayashi et al., 2013; Seisenberger et al., 2012). Second, we found that, similarly to IAPs, young L1 promoters were also quite resistant to PGC demethylation and retained ~40% of CpG methylation on average for L1A, and ~20% for L1Tf, at E12.5. However, in contrast to IAPs, young L1 methylation further declined and reached the lowest point three days later, in E15.5 prospermatogonia, with 24.9% for L1A and 16.5% for L1Tf. This indicates that L1 demethylation occurs with an extended timeline, likely via an active mechanism as prospermatogonia enter mitotic arrest at E13.5. After E15.5, both IAPs and young L1s acquired *de novo* DNA methylation progressively until birth, where they reached 80% of CpG methylation, a level that is globally maintained in adult spermatozoa (**Fig. 1A**).

To better assess the regions that are resistant to PGC demethylation, we searched for remaining methylated regions (RMRs) using public whole-genome bisulfite sequencing (WGBS) datasets (Kobayashi et al., 2013) from E13.5 prospermatogonia (**Extended data Fig. 1B**). Out of 8,189 identified RMRs, 98.3% were annotated as repeats among which RMRs were significantly enriched—against randomized annotations across the genome—for L1, ERVK and ERV1 but not ERVL

Figure 1

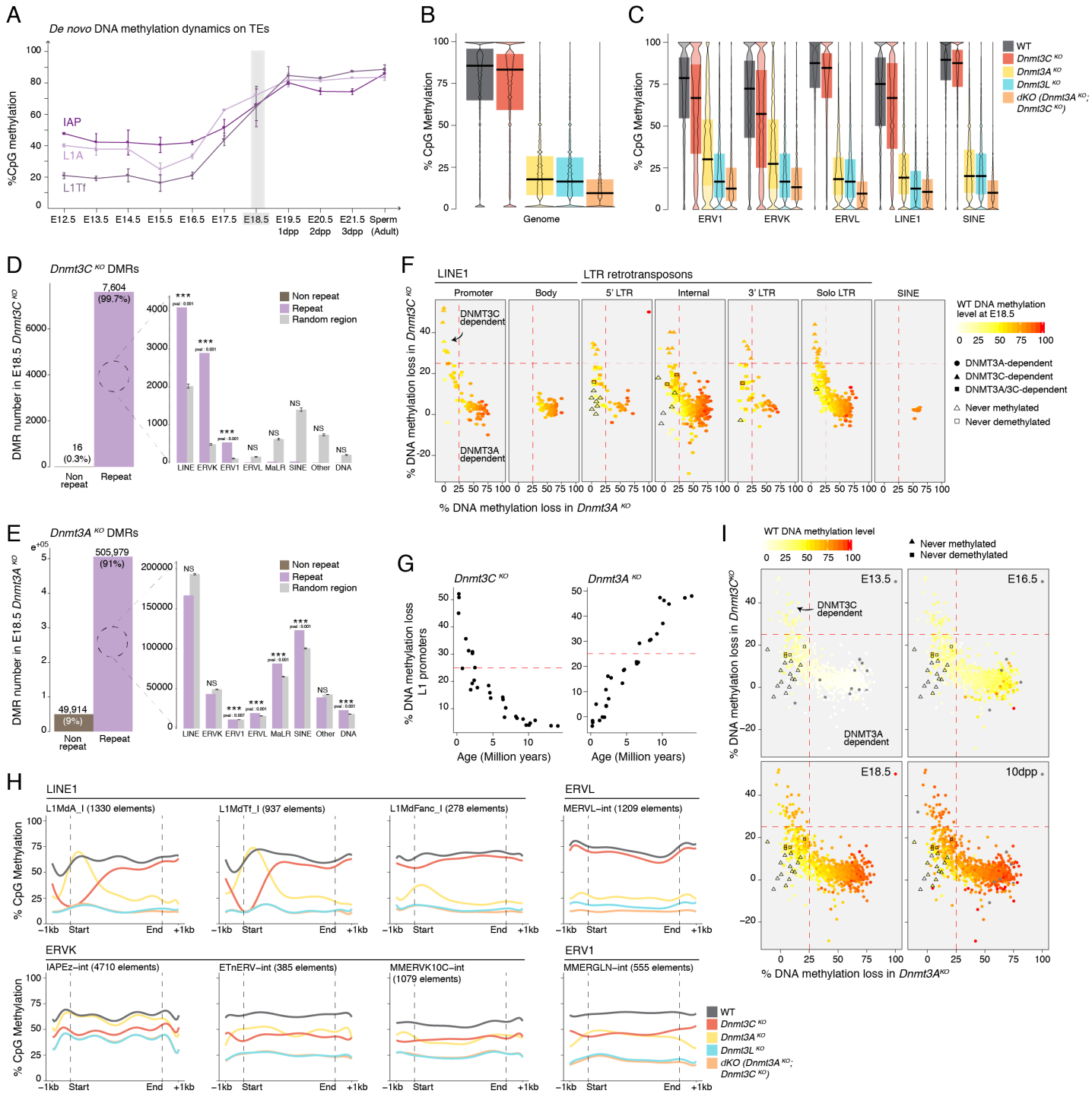


Figure 1 I DNMT3C genomic targets map to promoters of young L1, ERVK and ERV1
A. Timeline of *de novo* DNA methylation dynamics of IAP, L1A and L1Tf in OCT4-GFP-positive perinatal prospermatogonia and adult spermatozoa. DNA methylation levels were assessed by pyrosequencing. Dots and error bars represent the mean and the standard deviation (SD), respectively, between two or three different animals (*n*=2 for E12.5, E13.5, E17.5, E19.5, E20.5, positive control sperm and *n*=3 for E15.5, E15.5, E16.5, E18.5, E21.5). Subsequent WGBS was performed at E18.5 (grey). **B**, **C**. Violin plot representation of CpG methylation content over the whole genome (**B**), and TE families (**C**) in fetal prospermatogonia (E18.5) from WT (grey), *Dnmt3C* mutant (red), *Dnmt3A* mutant (yellow), *Dnmt3L* mutant (blue) and *dKO Dnmt3A*; *Dnmt3C* mutant (orange) as determined by WGBS. Black horizontal bars represent the median, upper and lower hinges correspond to 75 and 25% quantile respectively. **D**, **E**. (Left) Bar plot representing the number of differentially methylated regions (DMRs) mapping to non-repeat (brown) and repeat (purple) regions in *Dnmt3C* mutant (**D**) and *Dnmt3A* mutant (**E**) over WT as determined by WGBS at E18.5. (Right) In purple, distribution of repeat-related DMRs over individual TE

families. In grey, distribution of randomly picked genomic regions (7,604 and 505,979 total, for *Dnmt3C*^{KO} and *Dnmt3A*^{KO}, respectively). Permutation test over random regions: *p<0.05, **p<0.005, ***p<0.0005). **F**. Scatter plot showing in abscissa the percentage of DNA methylation loss in *Dnmt3A*^{KO}/WT and in ordinate the percentage DNA methylation loss in *Dnmt3C*^{KO} /WT at E18.5. Each scatter plot shows a functional TE segment and each dot represents a family. WT CpG methylation level at E18.5 is represented using white to red gradient. For simple reading: dots present in the bottom-right corner represent TE segments that acquire DNA methylation in a DNMT3A-dependent manner, triangles in the top-left relate to DNMT3C-dependent TE segments and squares in the bottom-left corner to DNMT3A and DNMT3C-dependent TE segments. Uncolored squares represent TE segments resistant to DNA methylation erasure (DNA methylation level > 45% at E13.5, datasets from Kobayashi et al. 2013), namely IAPs (LTRs and internal part), as seen in Fig. 1I. Uncolored triangles represent TE segments never re-methylated (DNA methylation level < 45% at 10dpp, from Barau *et al*. 2016). **G**. Scatter plot showing DNA methylation loss at L1 promoters in *Dnmt3C*^{KO} (up) and *Dnmt3A*^{KO} (bottom) according to the evolutionary age of L1 elements (in million years) (Sookdeo et al., 2013). **H**, Metaplots of DNA methylation levels over uniquely assigned full-length elements from LINE1 (>5kb), ERVL, ERVK and ERV1 (>500bp) families in E18.5 prospermatogonia of different *Dnmt3* genotypes. The number of unique elements is indicated per family. **I.** Scatter plot showing DNA methylation loss in *Dnmt3A*^{KO} in abscissa and DNA methylation loss in *Dnmt3C*^{KO} as measured at E18.5. From top to bottom, the evolution of level of WT CpG methylation at equivalent regions across male germ cell development is represented at E13.5, E16.5, E18.5 and 10dpp with white to red gradient, calculated from available WGBS datasets (Barau et al., 2016; Dura et al., 2021; Kobayashi et al., 2013). Dots represent TE families divided into functional segments: L1 promoter, L1 body, 5'LTR, LTR Body, 3'LTR, Solo LTR and SINE. Reading should be performed as in F. Grey dots relate to TE segments for which DNA methylation levels could not be calculated.

or SINEs, with the greatest RMR number mapping to ERVK families (*n*=4,292) (**Extended data Fig. 1B**). Metaplot analysis over ERV1, ERVK, ERVL (solo LTR included) and all L1s confirmed retention of CpG methylation at IAPs but also at some L1s (maximum of 20% CpG methylation, all L1 families included) at E13.5 (**Extended data Fig. 1C**). Moreover, "resistant" DNA methylation was the most prominent around the transcription start sites of these TE families, meaning around the promoter region. Finally, further integration of WGBS datasets at E16.5 (Kobayashi et al., 2013) and E18.5 (Dura et al., 2021) showed progressive TE remethylation across prospermatogonia development, showing similar methylation levels across promoters and internal sequences of TEs at the end of the process (E18.5) and relatively lower DNA methylation levels of L1s compared to other full-length TEs at E16.5 (**Extended data Fig. 1C**), in agreement with our pyrosequencing-based results. Some interesting TE family-specific features were also observed regarding DNA methylation distribution in regions adjacent to TEs: ERVKs showed high methylation retention extending over regions 5' of their transcription start sites at E13.5 (for both solo LTRs and full-length elements), while it was not the case for L1s.

In sum, our analysis demonstrated that a fraction of both IAP and L1 elements are resistant to PGC DNA methylation erasure. However, contrary to IAPs, L1s further undergo demethylation in arrested prospermatogonia: the lowest level of L1 promoter methylation is not attained at E12.5- E13.5 as for the rest of the genome, but at E15.5.

DNMT3C genomic targets are promoters of young L1, ERVK and ERV1

In male germ cells, *de novo* DNA methylation is dependent on DNMT3A, DNMT3C and the DNMT3L co-factor. Point mutations in any of the associated genes confer male sterility (Barau et al., 2016; Bourc'his and Bestor, 2004; Dura et al., 2021; Kaneda et al., 2004). Although highly expressed in fetal prospermatogonia at the time of *de novo* DNA methylation, DNMT3B seems dispensable for this process: germ cell-conditional *Dnmt3B* mutant males (constitutive mutants die around E9.5) show normal germ cell DNA methylation and fertility, although incomplete excision of the *Dnmt3B* gene using the *Tnap-Cre* driver was reported (Kaneda et al., 2004).

To precisely identify the complete set of genomic targets—and in particular TE targets—of all DNMT3 members, we used WGBS datasets we previously generated at E18.5 in wildtype (WT) and constitutive $Dnmt3C^{KO}$, $Dnmt3A^{KO}$ and $Dnmt3L^{KO}$ mutant prospermatogonia sorted using the *Oct4-eGFP* transgene (Dura et al., 2021). To

exclude potential inefficiency of the *Tnap-Cre*, we relied on the *Prdm1-Cre* line to generate germ cell-conditional *Dnmt3B* mutants, with our collaborator Mickael Weber (Strasbourg University). However, despite our efforts, we failed to obtain *Oct4-eGFP*-positive; *Prdm1-Cre*-positive; *Dnmt3B*$^{KO/2lox}$ embryos and their necessary controls. As a proxy for DNMT3B targets, we therefore used double *Dnmt3A*^{KO};*Dnmt3C*^{KO} (*dKO*) mutants (**Extended data Fig. 1D**).

As previously shown (Dura et al., 2021), *Dnmt3C*^{KO} had minor loss of DNA methylation genome wide (70.74% of mean methylation *versus* 73.77% in WT), while *Dnmt3A*^{KO} and *Dnmt3L*^{KO} presented profound CpG methylation decrease, reaching 21.35% and 20.32% of mean methylation, respectively (**Fig. 1E and Extended data Fig. 1E**). The *dKO* mutants showed the lowest CpG methylation content genome-wide (10.58%) and across individual genomic compartments, even compared to *Dnmt3L*^{KO}. Higher hypomethylation may be expected in enzymatic mutants (*dKO*) compared to mutants of a co-factor (*Dnmt3L*^{KO}). To identify potential DNMT3B targets, we compared DMRs between *dKO* mutants that lack both DNMT3A and DNMT3C and have intact DNMT3B, and *Dnmt3L*^{KO} mutants, whereby DNMT3A, DNMT3C and DNMT3B activity should be compromised. A single region—a CpG island linked to the *Tmem267* gene known to code for a transmembrane protein—was less methylated in *Dnmt3L*^{KO} compared to *dKO* (**Extended data Fig. 1G, 1H**). This observation suggests a negligible role, if any, for DNMT3B in methylating the prospermatogonia genome, including TEs.

Having excluded a role for DNMT3B in methylating TEs, we focused on the respective TE targets of DNMT3C and DNMT3A. Upon calling differentially methylated regions (DMRs) compared to WT, 99.7% of the 7,620 *Dnmt3C*^{KO} DMRs were annotated as repeats and among them, significant enrichment was observed at L1, ERVK and ERV1 specifically but not ERVL or SINEs, compared to the proportion of these various families in the genome (permutation test over random regions) (**Fig. 1D**). Comparatively, the 91% of the 505,893 *Dnmt3A*^{KO} DMRs that mapped to repeats followed a random distribution (**Fig. 1E**), indicating that DNMT3A methylates TEs in an unspecific manner.

As a first attempt to understand DNMT3C and DNMT3A specificity towards distinct TE features, we divided TEs into functional segments (*n*=1,637): L1s were divided into promoters and bodies, and ERVs were divided into 5'LTR, bodies, 3'LTR (for the full length elements) and solo LTRs (<500 bp). Each TE functional segment

was plotted according to the loss of DNA methylation in *Dnmt3A* and in *Dnmt3C* mutants at E18.5 (**Fig. 1F**), to highlight TE fragments whose methylation was DNMT3A-dependent (bottom right rectangle) or DNMT3C-dependent (top left rectangle), with a threshold of 25% DNA methylation loss. L1 promoters appeared as two categories, one DNMT3C-dependent and one DNMT3A-dependent, while L1 bodies were exclusively DNMT3A-dependent (**Fig. 1F**). Upon a closer examination of L1 promoters, we observed that L1 families that are younger than 2.5 million year-old lost DNA methylation in *Dnmt3C*^{KO}*,* while L1s older than 6 million years globally lost DNA methylation in *Dnmt3A*^{KO} (**Fig. 1G**) (Sookdeo et al., 2013). This was confirmed by a metaplot analysis, separating uniquely mappable elements into young (L1MdA_I and L1MdTf_I) and old L1 subfamilies (L1MdFanc_I)(**Fig. 1H**). This result confirms previous reports (Barau et al., 2016) that DNMT3C targets young L1 promoters, whereas DNMT3A targets ancient L1 promoters and all (young plus ancient) L1 bodies.

Upon examination of ERV families, functional segments were less explicitly divided into DNMT3A-dependent versus DNMT3C-dependent categories. Both LTR and body fragments could rely on one enzyme or the other (**Fig. 1F**). Interestingly, a large number of ERV segments were neither dependent on DNMT3C nor DNMT3A for acquiring DNA methylation in prospermatogonia (< 25% DNA methylation loss in mutants) (bottom left rectangle, **Fig. 1F**). By compiling available WGBS datasets before E18.5 (at E13.5 (Kobayashi et al., 2013) and E16.5 (Kobayashi et al., 2013)) and after E18.5 (10dpp (Barau et al., 2016)) (**Fig. 1I**), we noticed that four of these seemingly DNMT3C&3A-independent segments corresponded to TEs that were never demethylated in male germ cells (>45% DNA methylation at E13.5, black square): these were all IAP-related segments mapping to LTR or internal regions of full-length elements but not solo LTRs (**Fig. 1F** and **Table 1**). Conversely, fourteen ERV segments were never remethylated (<45% DNA methylation at 10dpp, black triangle) (**Fig. 1I**), and these were mainly on the 5'LTR and internal part of ERVB4 and RLTR44 elements (ERVK). Finally, having excluded these specific cases of TE segments that never get demethylated or remethylated, metaplot analysis of individual ERV families revealed that among TE fragments that appeared as demethylated at E13.5 and remethylated at 10dpp but not significantly demethylated in E18.5 *Dnmt3A*^{KO} or *Dnmt3C*^{KO}, some were in fact partially dependent on both DNMT3A and DNMT3C. Namely, ETnERV-int and MMERVK10C-int (both ERVK) and MMERGLN-int (ERV1) families showed similarly decreased methylation across their length in both *Dnmt3C*^{KO}

and *Dnmt3A*^{KO} prospermatogonia, at an intermediate place between WT levels and *Dnmt3L*^{KO} or *dKO* levels (**Fig. 1H**). This intermediate level could reflect that half of the elements within a family is DNMT3C-dependent while the other half is DNMT3A-dependent, or that an element within a family would rely on both DNMT3C and DNMT3A to acquire full DNA methylation. A heatmap presenting DNA methylation at all individual ETnERV-int, MMERVK10C-int and MMERGLN-int annotations showed that both situations existed within a same subfamily (**Extended data Fig. 1I**). We are currently refining this analysis in a quantitative manner.

Finally, by monitoring the DNA methylation dynamics of DNMT3C- and DNMT3A-dependent TE segments during WT prospermatogonia development (**Fig. 1I**), we revealed that DNMT3C-dependent targets were more methylated than DNMT3A-dependent targets at E13.5 (average of 19.81% versus 3.31% CpG methylation, respectively) but reached similar DNA methylation levels at 10dpp (83.48% and 85.64%, respectively). In between, DNMT3A-targets reached higher DNA methylation levels earlier than DNMT3C-targets, as seen at E18.5. While it was known that young TEs tended to be delayed in acquiring DNA methylation in prospermatogonia (Molaro et al., 2014), we show here that this is linked to different kinetics of action of DNMT3C and DNMT3A.

This analysis provides the first extensive classification of TE subfamilies and fragments according to their DNA methylation specifics during male germline development: 1) refractory to PGC demethylation (never demethylated, RMRs), 2) refractory to prospermatogonia remethylation (never remethylated), 3) DNMT3A-dependent, 4) DNMT3C-dependent, and 5) mixed DNMT3A-DNMT3C-dependency (**Table 1**).

L1 promoters are marked by bivalent H3K4me3-H3K9me3 chromatin signatures

Before analyzing the chromatin features that may distinguish DNMT3C-dependent TEs, we firstly investigated the chromatin signatures of all TEs across prospermatogonia development, indistinctively of their targeting logics. We took advantage of publicly available ATAC-seq (Assay for Transposase-Accessible Chromatin) and ChIP-seq datasets of H3K4me3, H3K9me3 and H3K27me3 in *Vasa-Venus*-sorted male germ cells at E13.5, E17.5, E19.5 and 2dpp (Yamanaka et al., 2019).

Combined integration of chromatin accessibility (ATAC-seq) and active chromatin (H3K4me3 ChIP-seq) was used as a proxy for an active transcriptional state of TEs. These features may be a better predictor than RNA-seq, as TE transcripts are cleaved by the piRNA pathway during this period. Heat map and metaplot analyses demonstrated that L1 promoters were accessible and H3K4me3-enriched at all timepoints, with the highest peaks at E17.5 (**Fig. 2A, 2B**). In contrast, ERVs showed no visible ATAC-seq or H3K4me3 enrichment. The repressive H3K9me3 mark decorated both L1 and ERVK promoters, but also to some extent the internal part of these elements, but no H3K9me3 enrichment was detected on ERVL and ERV1 (**Fig. 2A, 2B**). Finally, H3K27me3 did not show any specific pattern over TE promoters in general (**Fig. 2A, 2B**). In sum, while ERVKs were exclusively decorated by H3K9me3, L1 promoters presented both H3K4me3 and H3K9me3 marks.

A closer examination of this intriguing chromatin bivalency of L1 promoters revealed two different patterns of H3K4-H3K9me3 decoration according to the evolutionary age of the L1s. The evolutionarily youngest L1s (L1Al_III and L1Tf_I and _II) exhibited high H3K4me3 levels (peaking at E17.5), while H3K9me3 was globally low with linear increase over developmental time (**Fig. 2C**). Although relatively older L1s showed similar H3K4me3 dynamics than young L1s, with a peak at E17.5, their average H3K4me3 levels were lower than young L1s. Most strikingly, their H3K9me3 dynamics was clearly distinct from the one of young L1s, with the highest enrichment at E13.5, a decrease to a lowest point at E19.5, before increasing again until 2dpp (**Fig. 2C**).

Taken all together, the dynamics of H3K4me3 and H3K9me3 enrichment at L1 promoters is related to the age of L1 elements. Moreover, we found evidence of the two marks being present on the promoter of the same L1 element, indicating a bivalent chromatin status between permissive H3K4me3 and repressive H3K9me3 chromatin. We cannot conclude at this stage whether these two marks simultaneously occupy the same L1 promoter within the same prospermatogonia, or whether they are present on the same L1 promoter but in different cells. In any case, our observation suggests a complex chromatin regulation of L1 promoters during prospermatogonia development.

Work to further underpin the TE chromatin landscape in prospermatogonia is in progress (**Extended data Fig. 2**). To better delineate the

Figure 2

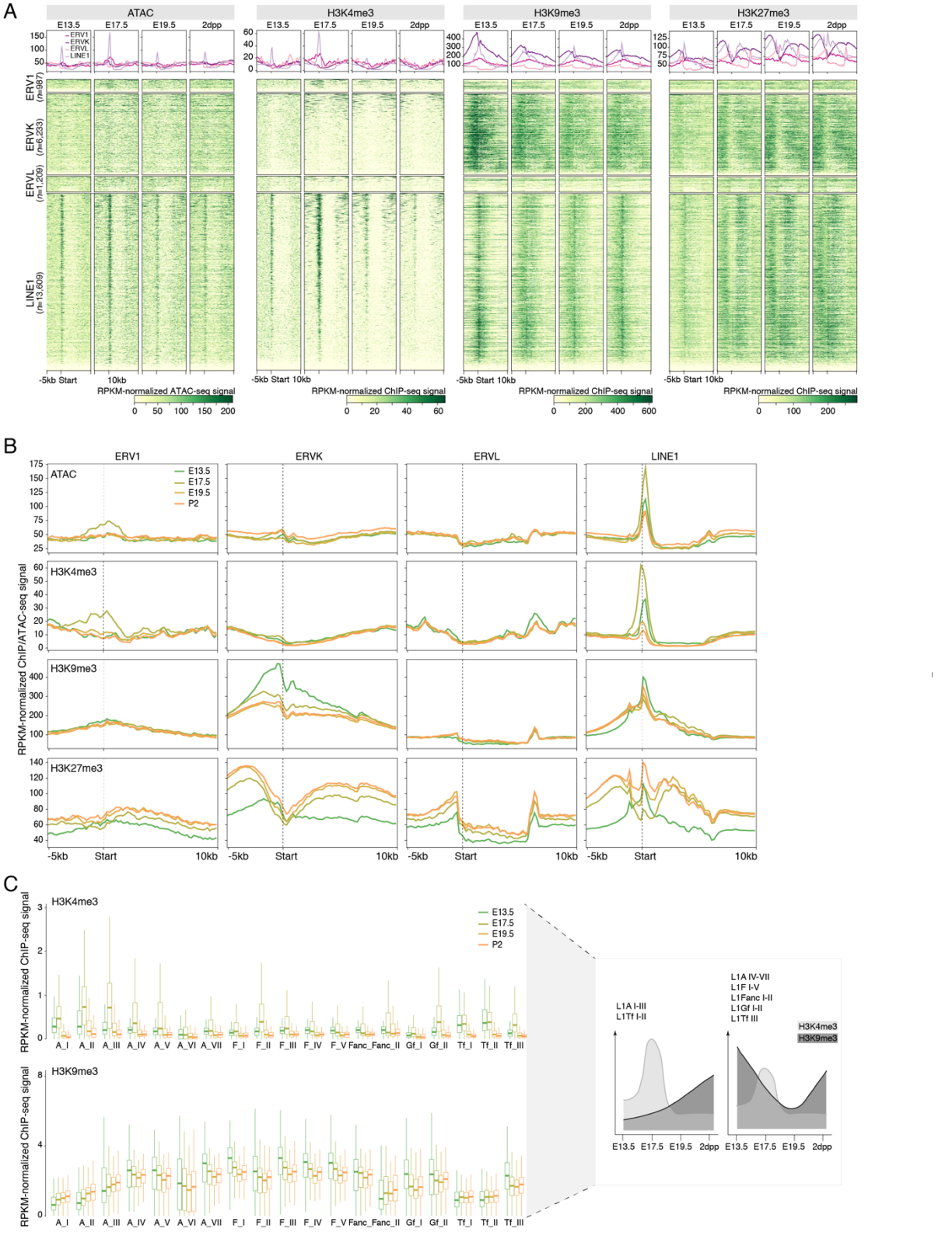


Figure 2 I L1 promoters harbor bivalent H3K4me3/H3K9m3 chromatin signatures
A, **B**. Heat maps (A) and metaplots (B) illustrating ATAC-seq and ChIP-seq (H3K4me3, H3K9me3 and H3K27me3) signals normalized by number of reads in prospermatogonia at E13.5, E17.5, E19.5 and 2dpp over TEs (datasets from (Kobayashi et al., 2013)). Uniquely mapped reads were considered only. For L1 annotations, "Md"-annotated full length (>5kb) elements were considered only; for ERVL, ERVK, ERV1, only elements >500bp, excluding solo LTRs. Heat maps and metaplots are centered over the start of the element. **C**. Bar plot showing enrichment of H3K4me3 (top) and H3K9me3 (bottom) as determined by ChIP-seq on individual L1Md families at E13.5, E17.5, E19.5, 2dpp, using uniquely mapped reads. (Right) Summary of the two categories of elements identified in this analysis as showing different dynamics of H3K4me3 and H3K9me3 signals in developing prospermatogonia.

temporal relationship between H3K4me3 and H3K9me3, we aim to perform Cut&Run in sorted prospermatogonia at E13.5, E15.5, E18.5, and 2dpp. Moreover, mapping these marks at 2dpp in both WT and *Dnmt3C*^{KO} may allow us understanding whether changes in these marks occur upstream or downstream of DNMT3C-dependent *de novo* DNA methylation. Our first data suggests that the E15.5 timepoint is particularly interesting (**Extended data Fig. 2**): of two replicates thus far sequenced, one is highly enriched in H3K4me3 on TE promoters and barely enriched in H3K9me3, whilst the second presents the inverse pattern, being lowly enriched in H3K4me3 and highly enriched in H3K9me3. This observation indicates that E15.5 could be a potential "switching" timepoint where TEs, especially L1s, could transition from H3K4me3 to H3K9me3. More sequencing and analyses will be performed in this direction.

Dnmt3C^{KO} DMRs switch from bivalent H3K4me3-H3K9me3 to H3K9me3-only chromatin

After describing the chromatin signature of TE promoters (22,038), we looked for the chromatin signatures that may specify DNMT3C-targeted TEs (which represent only 7,620 regions among the larger group of TE promoters).

We first relied on public ATAC-seq and ChIP-seq datasets from active chromatin marks (H2A.Zac, H3K4me3, H3K9ac) performed in E15.5 prospermatogonia (Watanabe et al., 2018). Using a random mapping approach and focusing on L1A subfamilies (I to VII), which show remarkable age-dependent classification, we found that the younger the L1A promoter is, the more active and accessible its chromatin is at E15.5 (**Fig. 3A**), and the greatest is the methylation defect in *Dnmt3C*^{KO} prospermatogonia (**Fig. 3B, 3C**). In contrast and as demonstrated before, older L1A subfamily types had a greater DNA methylation defect in *Dnmt3A*^{KO} (**Fig. 3B**), and this correlated with a less active chromatin status (**Fig. 3A**). We conclude that L1 promoters that are methylated by DNMT3C have an active chromatin pattern prior to *de novo* DNA methylation (E15.5), which suggests active transcription, according to their processing by the piRNA pathway.

Having previously found that L1 promoters were enriched in both H3K4me3 and H3K9me3, we assessed the correlation between their enrichment in these two marks at E17.5 (Yamanaka et al., 2019) and the dependency towards DNMT3C, *i.e.* the extent of DNA methylation loss in *Dnmt3C*^{KO} prospermatogonia. Co-occurrence of the two marks was observed on the promoter of single L1A elements, although levels were

Figure 3

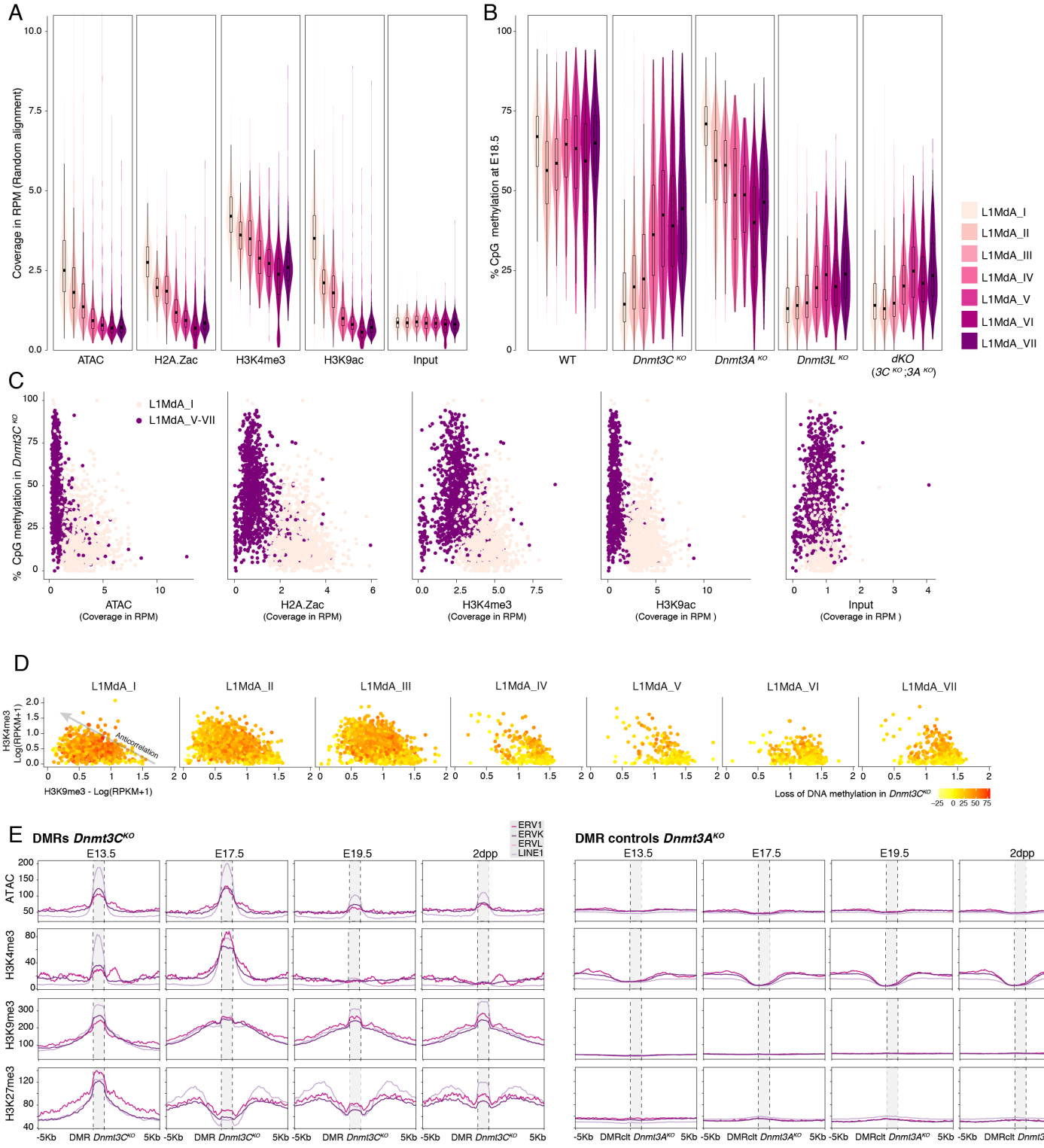


Figure 3 I *Dnmt3C*^{KO} DMRs switch from bivalent H3K4me3-H3K9me3 to H3K9me3-only chromatin signature

A. Violin plot showing the enrichment of ATAC-seq, ChIP-Seq signals for H2A.Zac, H3K4me3, K3K9ac (all active marks) and Input over L1MdA promoters (from I to VII) in E15.5 prospermatogonia. The window used to mean the enrichment is -1Kb from the TSS and +2Kb after the TSS. Datasets are from Watanabe et al. 2018, random alignment was performed. **B**. Violin plot showing the level of CpG methylation at E18.5 determined by WGBS in WT, *Dnmt3C*^{KO}, *Dnmt3A*^{KO}, *Dnmt3L*^{KO} and *dKO* (*Dnmt3A*^{KO} *; Dnmt3C*^{KO}) over L1A promoters (I to VII). The window used to calculate the methylation % corresponds to the promoter (1st kb from the beginning of the element). **C**. Scatter plot showing the correlation between DNA methylation level in *Dnmt3C*^{KO} and the enrichment of ATAC, H2A.Zac, H3K4me3, H3K9ac and Input over LI1A promoters (-1kb from the TSS and +2kb). **D**. Scatter plot indicating the correlation between the loss of DNA methylation in *Dnmt3C*^{KO} versus WT (yellow to red

gradient) and the enrichment in H3K4me3 (ordonate) and H3K9me3 (abscissa) determined by ChIP-seq at E17.5 (Yamanaka et al., 2019) over the promoter of individual L1 elements, classified into A-type subfamilies. **E**. Enrichment of ATAC-seq, and ChIP-seq for H3K4me3, H3K9me3 and H3K27me3 for *Dnmt3C*^{KO} DMRs (left) and *Dnmt3A*^{KO} DMRs as controls (right). The same number as *Dnmt3C*^{KO} DMRs were randomly selected (10 times) among ERV1, ERVL, ERVK and L1 sequences from *Dnmt3A*^{KO} DMRs to create the pool of *Dnmt3A*^{KO} control DMRs.

anticorrelated: the elements with the highest H3K9me3 showed relatively less H3K4me3, and vice versa. Moreover, we found L1A promoters with H3K9me3-only enrichment were the less demethylated in *Dnmt3C*^{KO} (< 25% methylation loss), indicating they were not DNMT3C targets (**Fig. 3D**). Therefore, the bivalent chromatin characteristic of L1 promoters can be nuanced when focusing on the correlation of the relative presence of H3K4me3 and H3K9me3 among L1 promoters.

We next focused on all *Dnmt3C*^{KO} DMRs (*n*= 7,620). *Dnmt3C*^{KO} DMRs were homogeneously distributed on all chromosomes, with the Y chromosome having the fewest (**Extended data Fig. 3**). *Dnmt3C*^{KO} DMRs were annotated as 79% intergenic and 21% intragenic. We classified *Dnmt3C*^{KO} DMRs by TE families, namely ERV1, ERVK, ERVL and LINE1, and analyzed their ATAC-seq profiles and ChIP-seq profiles for H3K4me3, H3K9me3, and H3K27me3, in WT prospermatogonia at E13.5, 17.5, 19.5, and 2dpp. Although coming from different TE families, *Dnmt3C*^{KO} DMRs behaved homogeneously (**Fig. 3E**). At E13.5, they showed high enrichment in all assayed chromatin features: ATAC, H3K4me3, H3K9me3, and H3K27me3. However, in later developmental stages, patterns became more dynamic and contrasted. From E17.5 until 2dpp, *Dnmt3C*^{KO} DMRs became depleted in H3K27me3 compared to surrounding regions and maintained an H3K27me3-poor status until 2dpp. Simultaneously, at E17.5, H3K4me3 and ATAC both peaked on *Dnmt3C*^{KO} DMRs, and then decreased at E19.5 and 2dpp. Meanwhile, H3K9me3 enrichment was maintained from E13.5 to 2dpp, and even increased at L1-related *Dnmt3C*^{KO} DMRs. The double decoration of H3K4me3-H3K9me3 here seen on *Dnmt3C*^{KO} DMRs at E17.5 is reminiscent to the double decoration we earlier described at young L1 promoters (**Fig. 2C**). Conversely, randomized *Dnmt3A*^{KO} DMR controls (from TEs) were not enriched in any of the assayed chromatin features at any stage, indicating a specific chromatin signature for *Dnmt3C*^{KO} DMRs.

Taken together, we revealed that *Dnmt3C*^{KO} DMRs are enriched in active marks such as H3K4me3 in early to mid-prospermatogonia, suggesting active transcription of DNMT3C genomic targets prior to (E13.5) and at the beginning of *de novo* DNA methylation (E17.5). Moreover, the finding that these DMRs carry H3K9me3 at the same time points (E13.5 and E17.5), indicated that *Dnmt3C*^{KO} DMRs have bivalent chromatin signatures as *de novo* DNA methylation starts. Closer examination of both H3K4me3 and H3K9me3 dynamics during prospermatogonia development and the

anticorrelated nature of their level on individual L1 promoters at E17.5 (**Fig. 3D**) indicates that the bivalency is likely an ephemeral status for *Dnmt3C*^{KO} DMRs, which rapidly undergo a chromatin switch from bivalent to H3K9me3-only. Whether this occurs prior to DNMT3C-dependent DNA methylation, simultaneously or right after is not known yet.

DNMT3C is recruited by H3K9me3 in an *in vitro* system

To further assess the involvement of specific chromatin signatures in DNMT3C targeting, we transitioned to an *in cellula* system of ectopic DNMT3C expression. We relied on mouse embryonic stem cells (mESCs), which do not significantly express endogenous *Dnmt3C* (Barau et al., 2016) and further used mESCs that were triple knockout for *Dnmt1*, *Dnmt3A*, and *Dnmt3B* (referred to as *Dnmt*-tKO) to profile *de novo* DNA methylation targets in this system. In the *Dnmt*-tKO background mESCs, we introduced DNMT3C or DNMT3B expression by producing stable-integration transgenic mESCs at the *Rosa26* safe-harbor locus (*Dnmt*-tKO::*RosaDnmt3C* and *Dnmt*-tKO::*RosaDnmt3B* lines, hereon referred to as *Dnmt*-tKO::*Dnmt3C* and *Dnmt*-tKO::*Dnmt3B*) (**Fig. 4A and Extended data Fig. 4A**). DNMT3B was used as the control line, since *Dnmt3C* evolved by tandem duplication of *Dnmt3B* (Barau et al., 2016).

Quantification by luminometric methylation assay (LUMA) revealed similar gains of CpG methylation in *Dnmt*-tKO::*Dnmt3C* and *Dnmt*-tKO::*Dnmt3B* lines (30.4% and 33.4%, respectively), over the 0% methylated *Dnmt*-tKO genome (**Extended data Fig. 4B**). Liquid chromatography-mass spectrometry (LC-MS) confirmed the DNA methylation gain, with 0.79% and 0.81% of methylated cytosines for *Dnmt*-tKO::*Dnmt3C* and *Dnmt*-tKO::*Dnmt3B*, respectively (**Extended data Fig. 4C**). Genome-wide profiling of DNA methylation by enzymatic methyl-seq (EM-seq) in *Dnmt*-tKO*::Dnmt3C* and *Dnmt*-tKO*::Dnmt3B* revealed a mean of CpG methylation of 8.0% and 13.8%, respectively (**Fig. 4B**); levels were comparable across all genomic compartments (**Fig. 4C**). Lower levels quantified by EM-seq versus LUMA may come from the bias of LUMA towards CG-rich regions, due to the use of the *HpaII/MspI* isoschizomers. Only CpG methylation was detectable in transgenic *Dnmt*-tKO::*Dnmt3C* and *Dnmt*-tKO::*Dnmt3B* mESCs, not CpC, CpA or CpT methylation (**Fig. 4D**).

Figure 4

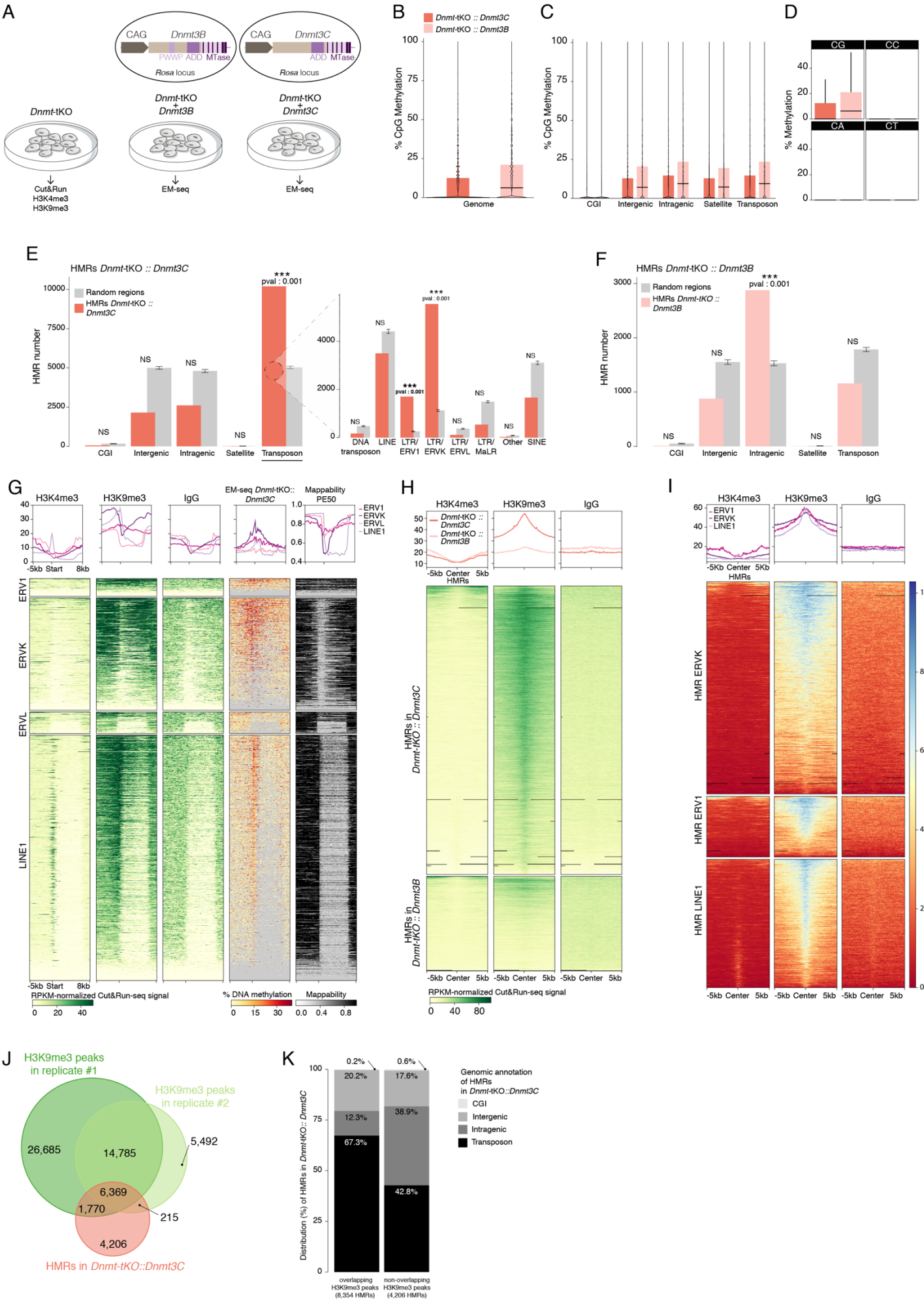

A
CAG
Dnmt3B
PWWP ADD MTase
Rosa locus
CAG
Dnmt3C
ADD MTase
Rosa locus
Dnmt-tKO
Dnmt-tKO + Dnmt3B
Dnmt-tKO + Dnmt3C
Cut&Run H3K4me3 H3K9me3
EM-seq
EM-seq
B
Dnmt-tKO :: Dnmt3C
Dnmt-tKO :: Dnmt3B
% CpG Methylation
Genome
C
% CpG Methylation
CGI
Intergenic
Intragenic
Satellite
Transposon
D
CG
CC
CA
CT
% Methylation
E
HMRs Dnmt-tKO :: Dnmt3C
Random regions
HMRs Dnmt-tKO :: Dnmt3C
HMR number
pval : 0.001
CGI
Intergenic
Intragenic
Satellite
Transposon
DNA transposon
LINE
LTR/ ERV1
LTR/ ERVK
LTR/ ERVL
LTR/ MaLR
Other
SINE
F
HMRs Dnmt-tKO :: Dnmt3B
Random regions
HMRs Dnmt-tKO :: Dnmt3B
HMR number
pval : 0.001
CGI
Intergenic
Intragenic
Satellite
Transposon
G
H3K4me3
H3K9me3
IgG
EM-seq Dnmt-tKO:: Dnmt3C
Mappability PE50
ERV1
ERVK
ERVL
LINE1
-5kb Start 8kb
RPKM-normalized Cut&Run-seq signal
% DNA methylation
Mappability
H
H3K4me3
H3K9me3
IgG
Dnmt-tKO :: Dnmt3C
Dnmt-tKO :: Dnmt3B
-5kb Center 5kb
HMRs
HMRs in Dnmt-tKO :: Dnmt3C
HMRs in Dnmt-tKO :: Dnmt3B
RPKM-normalized Cut&Run-seq signal
I
H3K4me3
H3K9me3
IgG
ERV1
ERVK
LINE1
-5kb Center 5kb
HMRs
HMR ERVK
HMR ERV1
HMR LINE1
J
H3K9me3 peaks in replicate #1
H3K9me3 peaks in replicate #2
26,685
14,785
5,492
6,369
1,770
215
4,206
HMRs in Dnmt-tKO::Dnmt3C
K
Distribution (%) of HMRs in Dnmt-tKO:: Dnmt3C
0.2%
20.2%
12.3%
67.3%
0.6%
17.6%
38.9%
42.8%
overlapping H3K9me3 peaks (8,354 HMRs)
non-overlapping H3K9me3 peaks (4,206 HMRs)
Genomic annotation of HMRs in Dnmt-tKO::Dnmt3C
CGI
Intergenic
Intragenic
Transposon

Figure 4 I DNMT3C is recruited to H3K9me3-enriched TEs in *Dnmt*-tKO mESCs
A. Schematic of the *Dnmt3* transgenes inserted at the *Rosa26* locus in *Dnmt*-tKO mESCs and experimental rationale. **B**. Violin plot showing CpG methylation level genome-wide *in Dnmt*-tKO*::Dnmt3C* (red) and in *Dnmt*-tKO*::Dnmt3B* (pink), as determined by EM-seq. Black horizontal bars represent the median, upper and lower hinges correspond to 75 and 25% quantile, respectively. **C**. Violin plot showing CpG methylation level (EM-seq) on the different compartments of the genome *in Dnmt*-tKO*::Dnmt3C* (red) and in *Dnmt*-tKO*::Dnmt3B* (pink). Black horizontal bars represent the median, upper and lower hinges correspond to 75 and 25% quantile respectively. D. Box plot showing CpG, CpC, CpA, CpT methylation (EM-seq) in *Dnmt*-tKO*::Dnmt3C* (red) and in *Dnmt*-tKO*::Dnmt3B* (pink). Black horizontal bars represent the median, upper and lower hinges correspond to 75 and 25% quantile, respectively. **E**, **F**. (Left) Bar plot showing the number of hypermethylated regions (HMRs) on the different compartments of the genome, for *Dnmt*-tKO*::Dnmt3C* (red, E) and *Dnmt*-tKO*::Dnmt3B* (pink, F). In grey, number of random regions. (Right) In red, number of HMRs in each TE family, annotated previously as “transposons” (red). In grey, number of random regions picked in the whole genome. Permutation test over random regions, *p<0.05, **p<0.005, ***p<0.0005. **G**. Heat map showing the enrichment in H3K4me3, H3K9me3, IgG (determined by Cut&Run, signal normalized by read number), then the level of DNA methylation in *Dnmt*-tKO*::Dnmt3C* (EM-seq), and the mappability for reads of 50bp on all elements from ERV1, ERVK, ERVL, and LINE1 families. TEs are aligned on their transcriptional start site. **H**. Heat map exhibiting the enrichment in H3K4me3 and H3K9me3 (Cut&Run signal normalized by read number) in HMRs of *Dnmt*-tKO*::Dnmt3C* and *Dnmt*-tKO*::Dnmt3B*. I. Heat map exhibiting the enrichment in H3K4me3 and H3K9me3 (Cut&Run signal normalized by read number) in HMRs of *Dnmt*-tKO*::Dnmt3C* annotated on ERVK, ERV1 and LINE1. **J**. Venn diagram showing the intersection between H3K9me3 peaks and *Dnmt*-tKO*::Dnmt3C* HMRs. **K**. Bar plot showing the genomic annotation (in percentage) of the *Dnmt*-tKO*::Dnmt3C* HMRs overlapping or not with H3K9me3 peaks.

We next went on identifying the sequences that acquire DNA methylation (defined as hypermethylated regions, HMRs) in *Dnmt*-tKO:: *Dnmt3C* and *Dnmt*-tKO::*Dnmt3B* lines. We found 15,057 HMRs in *Dnmt*-tKO::*Dnmt3C* and 4,928 HMRs in *Dnmt-tKO*::*Dnmt3B*. Strikingly, *Dnmt-tKO*::*Dnmt3C* HMRs were significantly more enriched in TEs compared to random regions (**Fig. 4E**), while *Dnmt*-tKO::*Dnmt3B* HMRs preferentially mapped to intragenic regions (**Fig. 4F**), consistent with previous reports (Baubec et al., 2015). Classifying TEs by family revealed that *Dnmt-tKO*::*Dnmt3C* HMRs were the most highly significantly enriched in ERVK (33.4% of HMRs) and ERV1 (9.54%). Although L1s represented a high fraction of *Dnmt-tKO*::*Dnmt3C* HMRs (20.2%), they were not significantly enriched compared to randomly picked regions. These observations reveal that in a non-germ cell context, here embryonic stem cells, DNMT3C also preferentially methylates TEs.

To investigate the link between *Dnmt*-tKO::*Dnmt3C* HMRs and pre-existing chromatin features, we performed Cut&Run in *Dnmt*-tKO mESCs for H3K4me3 and H3K9me3 and determined their level of enrichment on full-length TEs (ERV1, ERVK, ERVL and L1). All TE families appeared enriched in H3K9me3 on both their promoter and body. However, only L1 promoters were enriched in H3K4me3 (**Fig. 4G**). This result indicates that, as in prospermatogonia, L1 promoters could have a bivalent H3K4-H3K9me3 chromatin signature. Moreover, the level of DNA methylation at TEs in *Dnmt-tKO*::*Dnmt3C* correlated with H3K9me3 enrichment in *Dnmt*-tKO mESCs, and this was restricted to promoter regions. This observation was particularly prominent at L1s (**Fig. 4G**).

We next concentrated on the chromatin features of *Dnmt-tKO*::*Dnmt3C* HMRs specifically. We observed that *Dnmt-tKO*::*Dnmt3C* HMRs were highly enriched in H3K9me3, whilst no enrichment in H3K4me3 was detected (**Fig. 4H**). Focusing more precisely on *Dnmt-tKO*::*Dnmt3C* HMRs that are annotated in ERVK, ERV1 or L1, the level of H3K9me3 remained high (**Fig. 4I**). In contrast, in *Dnmt-tKO*::*Dnmt3B* HMRs, no enrichment of either H3K9me3 or H3K4me3 was detected (**Fig. 4H**).

Finally, we performed H3K9me3 peak calling in *Dnmt*-tKO mESCs. We found that two-thirds (*n*=8,354) of *Dnmt-tKO*::*Dnmt3C* HMRs overlapped with a H3K9me3 peak. Among these, 67.3% were annotated as TEs. In contrast, *Dnmt-tKO*::*Dnmt3C* HMRs that did not overlap with H3K9me3 peaks were not significantly enriched in TEs: 42.8% mapped to TEs, which correspond to the TE percentage in the genome. In

conclusion, our data supports a correlation between H3K9me3 enrichment and DNMT3C recruitment on TEs in mESCs.

DISCUSSION

In this study, we provided an extensive and precise description of the genomic targets of DNMT3C, using WGBS in E18.5 prospermatogonia. We confirmed that DNMT3C singularly methylates young L1 promoters and IAPs. Some classes of ERVK and ERV1 also rely on DNMT3C, but collectively, they also require DNMT3A to reach full methylation. In all scenarios, both DNMT3C and DNMT3A require the cofactor DNMT3L for methylating TEs. By performing WGBS on *dKO* (*Dnmt3A*^{KO}; *Dnmt3C*^{KO}), we also excluded any function for the remaining expressed DNMT3B in male germ cell remethylation.

For the first time, we described a H3K4me3-H3K9me3 bivalent chromatin signature at L1 promoters in both prospermatogonia and *Dnmt*-tKO mESCs (to a lesser extent). Regarding ESCs, it would be interesting to verify whether this feature is a specificity of the DNA methylation-free background of the *Dnmt*-tKO cells or whether it is also present in normal ESCs. Repeated TEs represent a challenge in bioinformatic analyses, because of uncertain genomic annotation for genetically similar elements. However, by focusing on uniquely mappable elements only, we can be confident that the two chromatin marks were present at the same genomic locus at the same time. Nonetheless, because Cut&Run was performed on a population of prospermatogonia, we cannot know whether the two opposite marks are present simultaneously in the same cell, or segregate in different cells. Single-cell Cut&Run or Cut&Tag approaches have been recently developed (Kaya-Okur et al., 2019; Skene and Henikoff, 2017), although it would remain challenging to address this question for reasons of limited coverage and TE mappability. Bivalency between active and repressive chromatin marks has been widely described between H3K4me3 and H3K27me3, a mark associated with facultative heterochromatin. This signature may have a role for keeping developmental genes lowly expressed in embryonic stem cells and poised for later activation upon differentiation (Harikumar and Meshorer, 2015). However, chromatin domains that are enriched in both H3K4me3 and H3K9me3—rather associated with long-term repression of constitutive heterochromatin—have been sparsely described so far (Matsumura et al., 2015; Rugg-Gunn et al., 2010), and they were not investigated at TEs. Further analyses will be required to elucidate the nature of this chromatin bivalency in prospermatogonia and ES cells, which anyway suggests a complex regulation of L1 promoters.

In prospermatogonia, DNMT3C targets (promoters of young L1s and of other TEs too) harbor this double H3K4me3 and H3K9me3 decoration at E13.5 and E17.5, and then switch to an H3K9me3-only status. DNMT3C target enrichment in H3K4me3 (and other active chromatin features) is not surprising: the TEs that are methylated by DNMT3C need to be first expressed to be processed into piRNAs, which themselves will guide DNMT3C to *de novo* methylate corresponding genomic loci. What is more puzzling is how young TEs, that are enriched in H3K9me3 at E13.5, subsequently gain H3K4me3 and become expressed, even in presence of H3K9me3?

An essential consideration is the development timepoints that are studied. At E17.5, when DNMT3C-dependent DNA methylation already started, our data showed that DNMT3C genomic targets have bivalent H3K4me3-K9me3 chromatin features. This may provide an important instruction for being targeted by DNMT3C. However, H3K4me3 likely repels DNMT3C, which is endowed with an ADD domain, as it was shown for DNMT3L or DNMT3A (Ooi et al., 2007). Some proteins have been shown to recognize both H3K4me3 and H3K9me3 marks simultaneously and such proteins could grant access of DNMT3C to its targets, despite the presence of H3K4me3. Such proteins may also help removing the H3K4me3 marl. Accordingly, DNMT3C targets switch to H3K9me3 only later during prospermatogonia development, DNMT3C targets become enriched in H3K9me3 only, at the same time they acquire DNA methylation. These two repressive marks seem therefore to follow similar dynamics. Is the H3K9me3-only status essential for recruiting DNMT3C, or is DNA methylation necessary to operate the H3K9me3-only switch? Both scenarios are mechanistically possible, as H3K9me3 can recruit the DNA methylation machinery (Auclair et al., 2016), and in other contexts, DNA methylation can recruit H3K9me3 (Fuks et al., 2003). Performing H3K9me3 and H3K4me3 Cut&Run at 2dpp in *Dnmt3C*^{KO} male germ cells will be essential for addressing the causal relationship between H3K9me3 and DNA methylation. If H3K9me3-only is present at DNMT3C targets despite the lack of DNA methylation, this would imply that H3K9me3 comes first. Conversely, decreased H3K9me3 — and possibly H3K4me3 retention — would indicate that DNMT3C-dependent DNA methylation is necessary for operating the H3K9me3-only switch at young TEs. Similar experiments were previously performed by Pezic *et al.* (2014) (Pezic et al., 2014) on *Miwi2* mutants, showing significant decrease of H3K9me3 on L1A promoters. One caveat is that their experiment was performed on 10dpp germ cells, quite late after *de novo* DNA methylation and at the time young TEs become

reactivated. Moreover, the analysis was not specifically focused on MIWI2 or DNMT3C TE targets.

To further address the importance of the chromatin states associated with DNMT3C target specificity, we turned to embryonic stem cells which are naturally devoid of the piRNA pathway and of DNMT3C. There as well, genomic targets that acquire DNA methylation upon DNMT3C expression showed H3K9me3 enrichment. However, only 20% of all H3K9me3 peaks were *de novo* methylated by DNMT3C, showing that H3K9me3 is not sufficient. This observation opens further avenues of research: are there other factors besides H3K9me3 that could recruit DNMT3C to TEs? We hypothesize that, independently of H3K9me3, the genetic sequence of TEs may matter for being a DNMT3C target, directly or through proteins with sequence-specific binding ability, such as transcription factors. By focusing on single mappers, we noticed a high level of heterogeneity in levels of DNA methylation loss in *Dnmt3C*^{KO} prospermatogonia for rather similar elements belonging to the same young TE family and with all high H3K4me3 enrichment. Motif search could allow us addressing this question, focusing on members from the same L1 subfamily for example, and looking for nucleotide differences between the ones that acquire DNA methylation in a DNMT3C-dependent manner, and the others that use the DNMT3A-dependent default route.

MATERIAL AND METHODS

Mice. Mice were hosted on a 12h light/12h dark cycle with free access to food and water in the pathogen-free Animal Care Facility of the Institut Curie (agreement C 75-05-18). All experimentations were approved by the Institut Curie Animal Care and Use Committee and adhered to European and national regulation for the protection of vertebrate animals used for experimental and other scientific purposes (directives 86/609 and 2010/63). For tissue and embryo collection, mice were euthanized by cervical dislocation. Prenatal and early postnatal time points were obtained by following timed pregnancies where the following day post-coitum was considered as embryonic day E0.5.

Mouse strains, all bred onto C57Bl6/J background, were previously described: *Dnmt3L*^{KO} (Bourc'his et al., 2001), *Dnmt3C*^{KO} (Barau et al., 2016), and *Oct4-eGFP* (Jackson Laboratories, stock no. 008214) (Lengner et al., 2007) (gift of A. Clark, UCLA). The *Dnmt3A*^{KO} strain was obtained by constitutive defloxing of the *Dnmt3A*$^{2lox/2lox}$ (Kaneda et al., 2004) strain using the *Zp3-Cre* line (Jackson Laboratories, stock no. 003394) (Lewandoski et al., 1997). Double knock-out of *Dnmt3A* and *Dnmt3C* (referred as *dKO* (*Dnmt3A*^{KO}; *Dnmt3C*^{KO})) were obtained by intercrossing *Dnmt3A*$^{KO/WT}$; *Dnmt3C*$^{KO/WT}$ animals.

Cell culture and cell transfection. Mouse embryonic stem cells (E14 line, derived from 1291*0la* strain*)* were cultured in feeder-free conditions on 0.2% gelatin-coated plates and grown in standard serum- based medium: Glasgow medium (Sigma), 15% ES-approved FBS (Gibco), 2mM L-Glutamine, 0.1mM MEM non-essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), 0.1mM β-mercaptoethanol, 1000U/mL leukemia inhibitory factor (LIF, Miltenyi), with daily changes and passages at 70% confluence. Mycoplasma-free status was assessed using the VenorGeM Classic Mycoplasma Detection Kit (Minerva Biolabs). E14 triple DNA methyltransferase knockout (*Dnmt1* -/-;*Dnmt3A* -/-;*Dnmt3B* -/-, *Dnmt*-tKO) ESCs were obtained by CRISPR-Cas9-mediated knockout upon nucleofection (P3 Primary Cell 4D-Nucleofector, Lonza) as described (Domcke et al., 2015). E14 *Dnmt*-tKO expressing *Dnmt3B*- and *Dnmt3C*-derived transgenes were obtained by nucleofection (P3 Primary Cell 4D-Nucleofector, Lonza) using CRISPR-Cas9 to knock-in donor plasmids pR26-CAG- 2xMYC-DNMT3B/C (modified from Barau et al., 2016) (Barau et al., 2016) into

the *Rosa26* locus by homologous recombination and clone selection in ES culture medium with hygromycin. Homozygous *Dnmt*-tKO*::RosaDnmt3B* and *Dnmt*-tKO*::RosaDnmt3C* knock-in lines were selected by PCR genotyping, and referred to as *Dnmt*-tKO*::Dnmt3B and Dnmt*-tKO*::Dnmt3C* for simplicity. Transgenic DNMT3 protein expression was confirmed by SDS-PAGE and western blot detection of MYC-tagged DNMT3s.

Cell sorting (prospermatogonia) by fluorescent activated cell sorting (FACS). For sorting prospermatogonia, the *Oct4-eGFP* transgenic mouse line was used. Testes were dissected from E12.5 to E21.5 males, decapsulated and transferred in low-binding Eppendorf tubes containing 100-150µL collagenase solution in HBSS (Collagenase type IV (Gibco), 2X AAs (Gibco), 2X Na-pyruvate (Gibco), 25mM HEPES-KOH pH7.5). Dissociation was achieved at 37°C for 5-7min by gently flicking the tube. Then 400-600µL of TrypLE Express (Gibco) were added to the cell suspension and incubated for 5-7min at 37°C. Single-cell suspension was obtained by up-and-down pipetting. TrypLE Express was quenched with 140-210µL fetal bovine serum (FBS) and the cell suspension was centrifuged at 600g/4min/4°C and washed with FACS buffer (1XPBS, 2mM EDTA, 1%BSA). Just before sorting, cells were resuspended in FACS buffer supplemented with DAPI (2ng/ml), in order to discriminate dead cells, and filtered through a 35 µm strainer. Cell sorting was performed on a BD FACSaria II (BD Biosciences) using 100 µm nozzle. Prospermatogonia were isolated by gating DAPI-neg, OCT4-EGFP-pos cells (see gates Extended data Fig. 1B).

DNA methylation assays.

Pyrosequencing. Targeted analysis of DNA methylation at retrotransposon promoters (IAPEz, L1MdA and L1MdTf) was performed on FACS-sorted prospermatogonia. DNA was extracted by incubating one volume of lysis buffer (200mM Tris-HCl pH8, 10mM EDTA, 0.4% SDS, 400mM NaCl, 80µg/mL linear polyacrylamide, 0.4mg/mL proteinase K) at 55°C for 2-4h, precipitated with isopropanol and cleaned with 70% ethanol. DNA was then bisulfite-converted using the EZ DNA Methylation-Lightning™ Kit (Zymo Research). CpG methylation was quantified on at least two individual animals by pyrosequencing on a PyroMark Q24 and PyroMark Q48 (Qiagen) using the Q24 and Q48 Softwares. Primers are available in Supplementary Table.

WGBS. Whole genome bisulfite sequencing was performed on DNA extracted from E18.5 prospermatogonia sorted by FACS (with the *Oct4-eGFP* transgenic line) using direct post-bisulfite adaptor tagging (PBAT) (Miura et al., 2019). Prospermatogonia of three embryos from each genotype (WT, *Dnmt3A*$^{KO/KO}$, *Dnmt3C*$^{KO/KO}$, *Dnmt3L*$^{KO/KO}$ and double *Dnmt3A*$^{KO/KO}$; *Dnmt3C*$^{KO/KO}$) were pooled for genomic DNA extraction using the QIAamp DNA Micro Kit (Qiagen) and spiked-in with 0.3% of unmethylated Lambda DNA (Promega). Bisulfite conversion was performed using EZ DNA Methylation-Gold Kit (Zymo Research) and the whole converted DNA was used as input for library preparation using the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) according to the recommended protocol and indexing PCR using 2X KAPA HiFi HotStart Uracil+ ReadyMix (KAPA Biosystems). Indexed, cleaned-up libraries were analyzed and quantified on a 2200 Tapestation instrument using a D5000 screen tape (Agilent). Library pools were sequenced on an Illumina NovaSeq in 100bp paired-end (PE) reads run.

Enzymatic methyl-sequencing (EM-seq). Genomic DNA from E14 *Dnmt*-tKO*::Dnmt3B* and *Dnmt*-tKO*::Dnmt3C* lines was extracted using GenElute Mammalian Genomic DNA Miniprep (Sigma) and fragmented on a Covaris S2 instrument to an average size of 500 bp. Enzymatic conversion and library preparation was performed using NEBNext® Enzymatic Methyl-seq Kit (NEB) following the recommended protocol for use with Large Insert Libraries (470–520 bp). Quality and size distribution of libraries were checked on a Tapestation 4200 system (Agilent).

LUMA . Global CpG methylation levels of *Dnmt*-tKO lines expressing *Dnmt3* transgenes was assessed using the LUminometric Methylation Assay (LUMA) (Karimi et al., 2006) Briefly, 500ng of genomic DNA was digested with *MspI*/*EcoRI* and *HpaII*/*EcoRI* (NEB) in parallel reactions. *EcoRI* is included as an internal reference. Filling of the genome-wide protruding ends of the restriction digestions where quantified in a pyrosequencing reaction (PyroMark Q48 autoprep, Qiagen, dispensation order: ACTCGA). Global CpG methylation levels were then calculated from the *HpaII*/*MspI* signal intensity ratio.

Liquid chromatography-mass spectrometry (LC-MS). LC-MS-based quantification of methylcytosines was performed on 1 µg of DNA degraded to nucleosides with nuclease P1 (Roche), snake venom phosphodiesterase (Worthington) and alkaline phosphatase (Fermentas). An equal volume of isotopic standard mixture ($^{15}N_3$-C (Silantes), $^{2}H_3$-5mC (TRC) and self-synthesized $^{15}N_3$-5hmC,

$^{15}N_3$-5fC and $^{15}N_3$-5caC were added to the nucleoside mixture and injected for LC-MS/MS analysis as described previously (Schomacher et al., 2016). Quantitative analysis was performed on an Agilent 1290 Infinity Binary LC system (Agilent technologies) using a ReproSil 100 C18 column (Jasco) coupled to an Agilent 6490 triple quadrupole mass spectrometer (Agilent technologies).

Western Blot. MYC-tagged DNMT3B/C expression was detected by western blot. After quantification using Bradford reagent (BioRad), 5 µg of total proteins were mixed with 4x Laemmli Sample Buffer, denatured at 90°C for 5 minutes and loaded into 4 to 12% gradient polyacrylamide gels (Thermo Fischer). Run was performed using the NuPAGE system with recommended buffers (Thermo Fischer) and proteins were transferred to a 0.45 µm nitrocellulose membrane (Amersham). Membranes were blocked with 5% skim milk solution in Tris-buffered saline containing 0.1% Tween 20 (TBST) and incubated with mouse anti-MYC (clone 9E10) at 4°C overnight, followed by five washes in TBST, 1h incubation with anti-mouse HRP conjugated antibody (Thermo Fischer) and five additional washes in TBST. Detection and documentation were done using luminogenic substrate (SuperSignal™ West Pico PLUS Chemiluminescent Substrate, Thermo Fischer) and the iBright gel documentation system (Thermo Fischer).

Cut&Run. Cut&Run was performed according Skene et al. (2017) (Skene and Henikoff, 2017). In brief, 20µL of Concanavalin A beads (Polysciences) per sample were resuspended in 1mL of Binding Buffer (20mM HEPES-KOH, 10mM KCl, 1mM CaCl2, 1mM MnCl2). Beads were washed twice in 1mL Binding Buffer and resuspended in 150µL of Binding Buffer/ sample.

For prospermatogonia, 5,000-10,000 cells/animal were sorted by FACS into 1X PBS (~50µL), as described before. Gonads from C57Bl6/J male embryos were used at E13.5, E15.5 and E18.5. For the E20.5 point, C57Bl6/J *Dnmt3C*$^{KO/KO}$ and WT littermates were used. For ES cells, 300,000 cells were used to performed Cut&Run. Samples were split into two or three aliquots according to the number of antibody profiles required. Cells were very gently resuspended in 300uL of cold NE buffer (20mM HEPES-KOH, 10mM KCl, 0,5mM Spermidine (Sigma), 0,1% Triton-X100, 20% Glycerol, 1X Complete™ EDTA-free protease inhibitor cocktail (Roche)) for exactly 5min. Then, 150uL of prepared beads were quickly added to the samples. Cells and

beads were incubated for precisely 10min at RT on a rotating wheel. Cells were collected on a magnetic stand (Ambion), the supernatant was discarded and cells were resuspended in 1mL of Blocking buffer (20mM HEPES-KOH, 150mM NaCl, 0.5mM spermidine (Sigma), 0.1% BSA, 1X Complete™ EDTA-free protease inhibitor cocktail (Roche) and 2mM EDTA) for 5min at RT on the bench.

Again, cells were collected on a magnetic stand, the supernatant was discarded and cells were resuspended in 400µl of Wash buffer + 1:200 primary Antibody (Wash buffer: 20mM HEPES-KOH, 150mM NaCl, 0.5mM spermidine (Sigma), 0.1% BSA and 1X Complete™ EDTA-free protease inhibitor cocktail (Roche)). Cells were incubated with antibodies for 2:30h at 4°C on a rotating wheel and washed twice with 1ml of Wash buffer. Antibody references are available in Supplementary Table.

Samples were then incubated with 1:400 ProteinA-MNase fusion protein (produced by the Recombination Protein Platform of the Institut Curie, 0.785 mg/mL) for 1h at 4°C on the rotating wheel followed by two washes with 1ml Wash buffer. Cells were then resuspended in 150µl Wash buffer and cooled down on ice for 5min before addition of CaCl2 to a final concentration of 2mM. Targeted digestion was performed for 30 min on ice, reaction was stopped by adding 150µL of 2X STOP buffer (200mM NaCl, 20mM EDTA, 4mM EGTA, 50µg/mL RNase A, 40µg/mL Glycogen, 0.1% NP40). Samples were then incubated at 37°C for 20min to release cleaved chromatin fragments into the supernatant. After centrifugation at 15,000rpm for 5min, supernatants were transferred to new low-binding tubes. Following addition of 0.1% SDS and 0.17mg/mL Proteinase K, samples were mixed by inversion and incubated at 70° C for 30 minutes. DNA was purified using phenol/chloroform followed by chloroform extraction and precipitated with 10µg of glycogen and 3 volumes of 100% ethanol for at least 20 minutes on ice. DNA was pelleted at 14,000 rpm at 4°C for 20 minutes. The DNA pellet was washed in 85% ethanol, spun down and resuspended in 40µL low Tris-EDTA (10mM Tris, 0.1mM EDTA) after complete evaporation of the ethanol.

Library preparation was performed according to manufacturer's instructions (Accel-NGS 2S DNA library kit, Swift biosciences) with a modified library amplification program: 98°C for 45sec, (98°C for 10sec, 60°C for 15sec, 68°C for 1min)x12 cycles, hold at 4°C. Average library size was tested on Agilent 4200 Tapestation using a DNA5000 screentape and quantification was performed on Invitrogen QUBIT 4 using high sensitivity DNA kit. Cut&Run libraries were sequenced on a NovaSeq (Illumina)

using PE 100bp or 50bp run, with two biological replicates for each time point or condition.

Publicly available dataset. To performed analysis, we also used the following public data: ATAC-Seq and Chip-Seq (H3K4me3, H3K9me3 and H3K27me3) data on WT prospermatogonia (FACS sorted using a *Vasa*-Venus transgene) at E13.5, E17.5, E19.5 and E21.5 (Yamanaka et al., 2019). ATAC-Seq and ChIP-Seq (H2A.Zac, H3K4me3, H3K9ac data on WT prospermatogonia (FACS sorted using O*ct4*-eGFP transgene) at E15.5 (Watanabe et al., 2018). WGBS at E13.6 and E16.5 on WT prospermatogonia (FACS sorted using O*ct4*-eGFP transgene) (Kobayashi et al., 2013). And finally, WGBS at E18.5 on WT, *Dnmt3C*, *Dnmt3A* and *Dnmt3L* mutant prospermatogonia (FACS sorted using O*ct4*-eGFP transgene) (Dura et al., 2021). The public data were process in the similar manner than our data, see below.

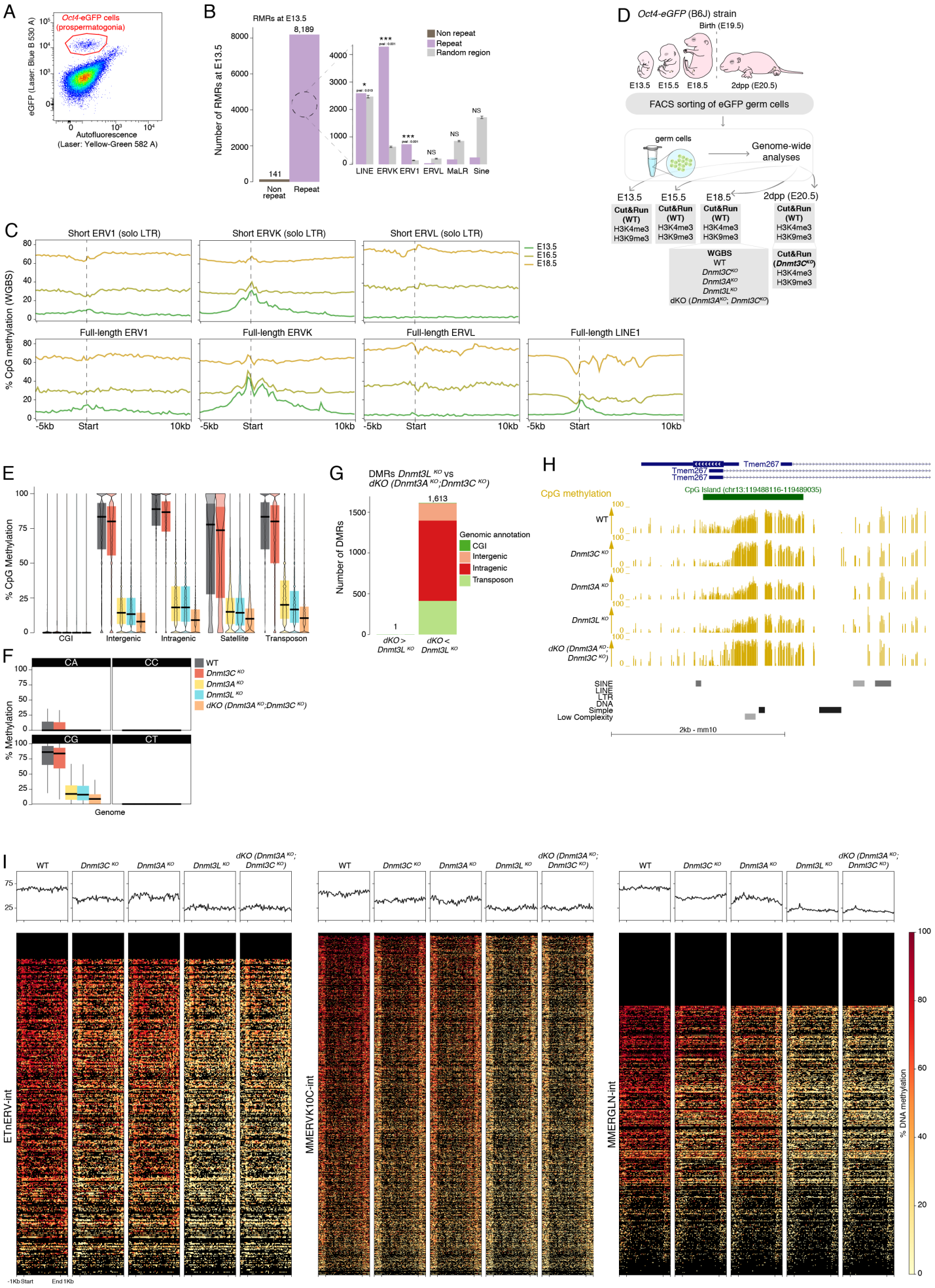
WGBS and EM-seq analysis. Adapters were trimmed using Atropos v1.1.16 (Didion et al., 2017). Trimmed reads were cleaned by removal of 5bp in 5' end of read1 and 12bp in 5' end of read2 using Cutadapt v1.12 (Martin, 2011) . Reads shorter than 15bp were discarded. Cleaned reads were aligned onto the Mouse Reference Genome (GRCm38/mm10) using Bismark v0.18.2 (Krueger and Andrews, 2011) with Bowtie2-2.2.9 allowing one mismatch in a seed alignment. Only reads mapping uniquely to the genome were kept, and methylation calls were extracted after duplicate removal by considering only CpG dinucleotides covered by a minimum of five reads. CpG methylation levels over different genomic compartments were calculated by extracting methylation calls with positional overlap with coordinates for Gencode vM16 gene annotations (“Intragenic”) and the RepeatMasker database (“Transposons” and “Satellite”). CpG islands (“CGIs”) were defined previously (Illingworth et al., 2010). “Intergenic” partitions were defined as genomic regions that did not overlap with CGIs, intragenic regions, satellite or transposable elements. Differentially Methylated Region (DMR) calling was performed using the DSS bioconductor package (Feng et al., 2014) with the following parameters: >25% of CpG methylation level difference, >5 CpGs called, regions >200 bp, and >500 bp distance between two DMRs. DMRs were annotated using the repeatMasker database. Random region sets were created 1000 times and were annotated. Permutation tests using regioneR package were computated (Gel et al., 2015). CpG methylation metaplots were processed using

deepTools v2.5.3 (Ramírez et al., 2016); only individual element annotations with size greater than 5 kb were analyzed for L1s, and greater than 500bp for ERVs. TE elements were divided into functional segments : full-length L1 elements (>5kb) were split into L1 promoter (1st kb) and L1 body (rest of the element); truncated L1 elements were considered as L1 body; short LTR (<500bp) as solo LTR; full-length LTR elements (>5kb) were split into 5'LTR (500 first bp), LTR body (from +500 to -500bp) and 3'LTR (last 500bp); truncated LTR (>500bp and <5kb) were considered as LTR body. Hypermethylated regions (HMR) in transgenic Dnmt-tKO ES samples (EM-seq) were identified using the MethPipe pipeline (Song et al., 2013) with default parameters.

ChIP-seq & Cut&Run analyses. Paired-end reads were trimmed using Trim Galore v0.4.4. The alignment was performed onto a concatenated genome using the Mouse Reference Genome (mm10) and the *Escherichia coli* genome (str. K-12 substr. MG1655, Genbank: NC_000913) for the Cut&Run samples. For the ChIP-seq samples (Watanabe et al., 2018; Yamanaka et al., 2019), the Mouse Reference Genome (mm10) was used for the alignment. In both cases, STAR (v2.7.0e) was used (Dobin et al., 2013) reporting randomly one position, allowing 4% of mismatches (*--outFilterMultimapNmax 5000 --outSAMmultNmax 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --alignIntronMax 1 --alignMatesGapMax 2000*) (Dobin et al., 2013). PCR duplicates were removed using Picard v2.6.0 (http://broadinstitute.github.io/picard/). Heatmaps and profiles were performed using Deeptools v2.5.3. 50-mers track mappability was used (Karimzadeh et al., 2018). Peaks were called using MACS2 v2.1.1 with the broad option and a q-value threshold of 0.05 (Zhang et al., 2008).

EXTENDED DATA FIGURES AND SUPPLEMENTAL TABLE

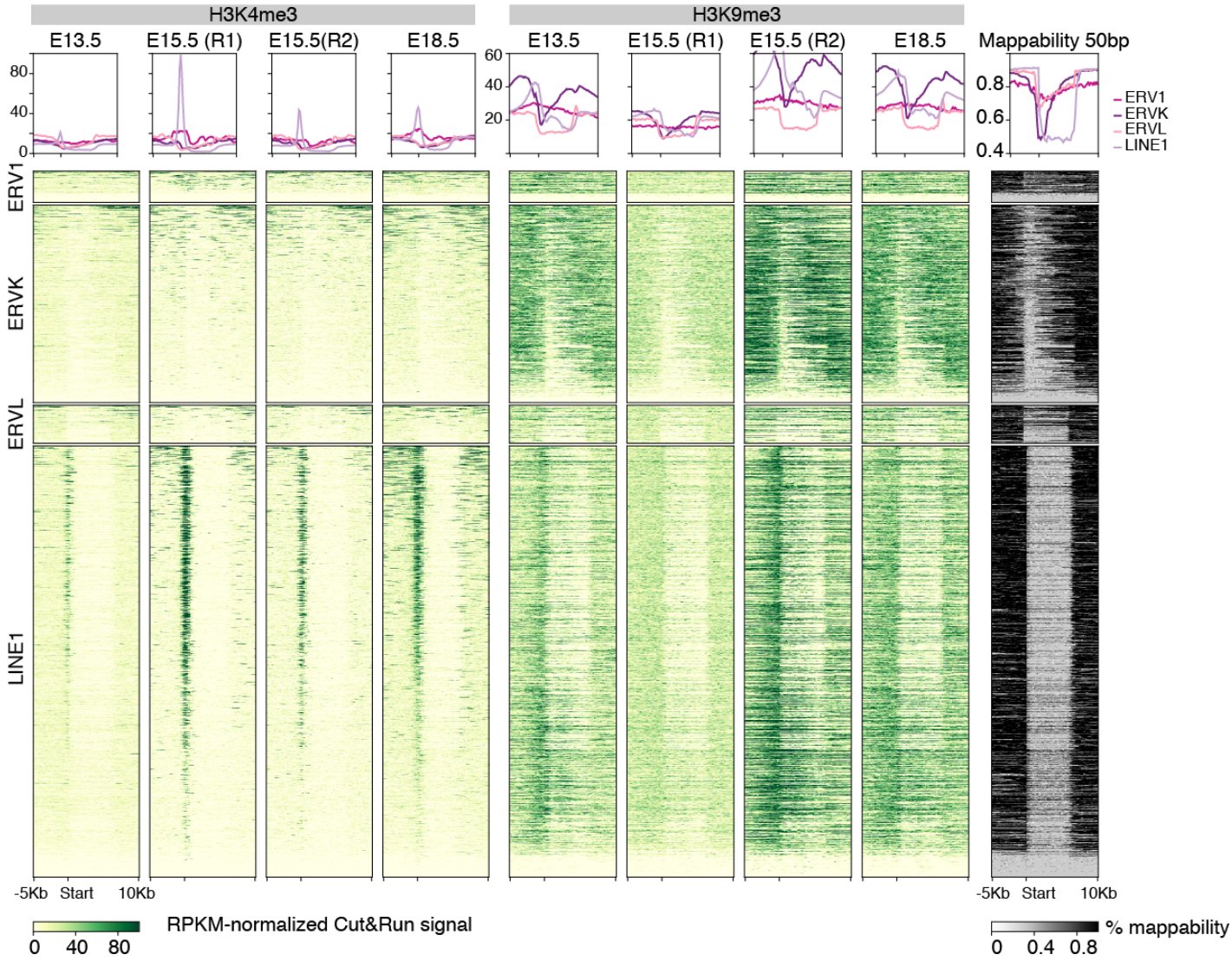
Extended data figure 1



Extended data figure 1 I DNA methylation distribution in prospermatogonia of various *Dnmt3* mutants

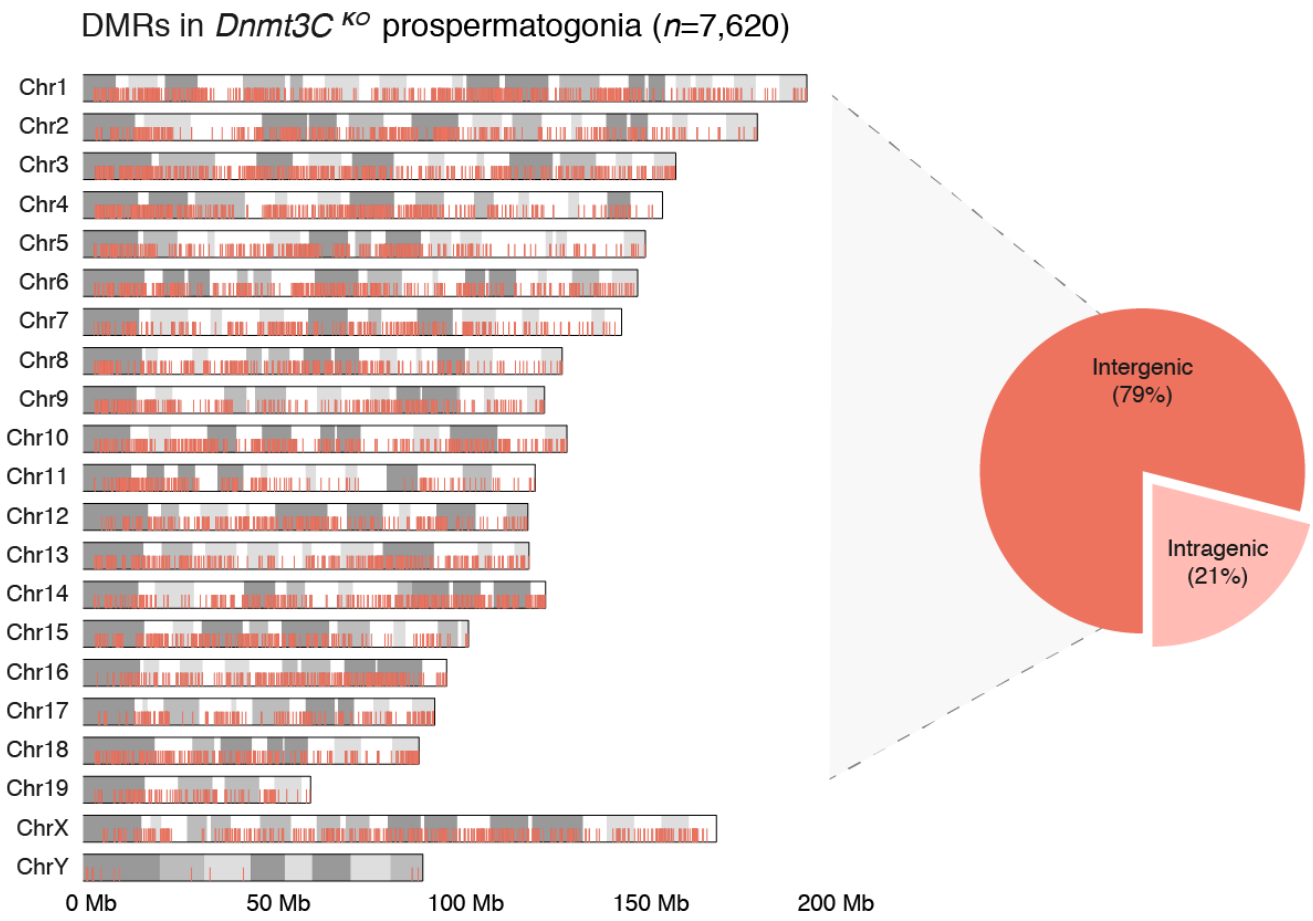
A. Representative FACS plot. Gating for sorting live OCT4-eGFP-positive cells (fetal prospermatogonia) is circled in red. **B**. (Left) Bar plot showing the number of remaining methylation regions (RMRs) at E13.5, mapping to non-repeat (brown) and repeat (purple) regions, using whole genome bisulfite sequencing (WGBS, public data, (Kobayashi et al., 2013)). (Right) In purple, distribution of repeat-related RMRs over individual TE families. In grey, distribution of randomly picked genomic regions (8,189 total). Permutation test over random regions, *p<0.05, **p<0.005, ***p<0.0005). **C**. Metaplot indicating the percentage of CpG methylation on solo or full-length LTR and L1 elements (all families) at E13.5, E16.5 and E18.5 as determined by WGBS (Barau et al., 2016; Dura et al., 2021; Kobayashi et al., 2013). **D**. Schematic of the datasets generated in the current study, using the *Oct4-eGFP* transgenic line for sorting prospermatogonia at indicated ages. **E**. Violin plot representation of CpG methylation content over genomic compartments in fetal prospermatogonia (E18.5) from WT (grey), *Dnmt3C* ^{KO} (red), *Dnmt3A* ^{KO} (yellow), *Dnmt3L* ^{KO} (blue) and *dKO Dnmt3A*^{KO}; *Dnmt3C*^{KO} (orange) as determined by WGBS. Black horizontal bars represent the median, upper and lower hinges correspond to 75 and 25% quantile, respectively. **F**. Box plot representing CpA, CpC, CpG and CpG methylation over the whole genome. No CpC and CpT methylation were found. **G**. Bar plot exhibiting the genomic annotation of DMRs that are more methylated in *dKO* (*Dnmt3A*^{KO}; *Dnmt3C*^{KO}) compared to *Dnmt3L*^{KO} (left, *n*=1), and DMRs less methylated in the *dKO* compared to *Dnmt3L*^{KO} (*n*= 1,613). DMRs more methylated in the *dKO* compared to *Dnmt3L*^{KO} can be interpreted as genomic targets of DNMT3B. **H**. Genome browser representation of the only DMR with higher methylation in *dKO* compared to *Dnmt3L*^{KO}. **I**. Heat maps showing DNA methylation levels in percentage at E18.5 (by WGBS) over individual elements of three families, taking into account full-length elements only (>500bp): ETnERV-int (ERVK), MMERVK10C-int (ERVK) and MMERGLN-int (ERV1). Black squares represent non-uniquely mappable sites.

Extended data figure 2



Extended data figure 2 I Cut&Run timeline in developing prospermatogonia (In progress)
(Left) Heat map illustrating H3K4me3 and H3K9me3 enrichment over TEs as determined by Cut&Run. The signal is normalized over the number of reads and only unique reads are considered. For L1 annotations, “Md”-annotated full length (>5kb) elements were considered only; for ERVL, ERVK, ERV1, only elements >500bp, excluding solo LTRs. E13.5 and E18.5 represented a merged of two replicates, while for E15.5 the two replicates are showed separately due to important variation. (Right) Heat map showing the mappability over TEs for 50bp-long reads.

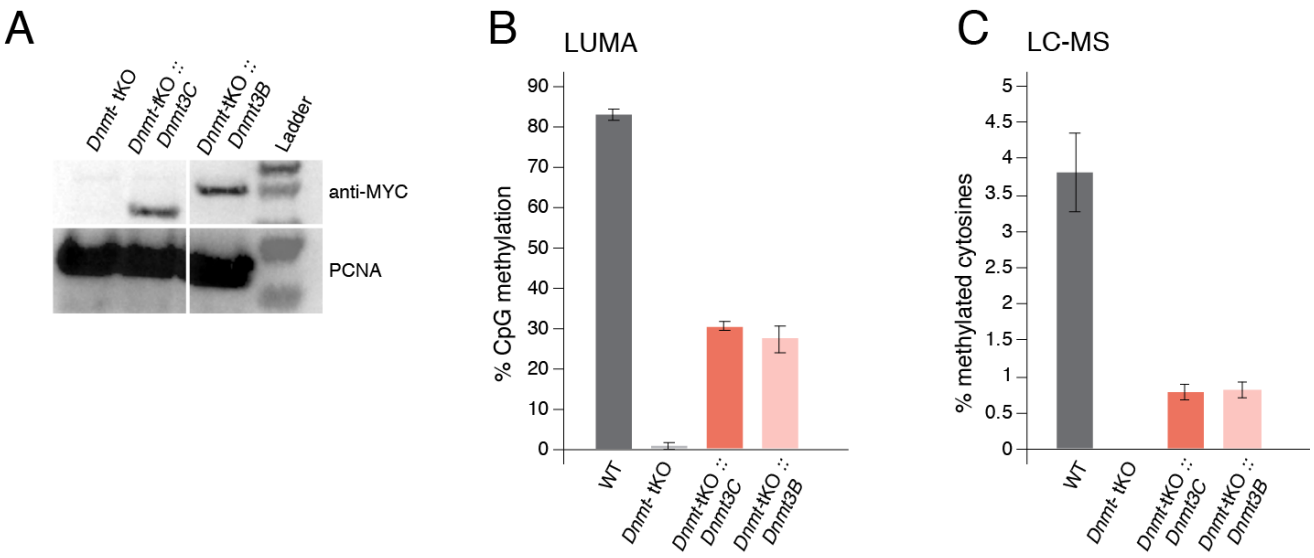
Extended data figure 3



Extended data Figure 3 I Chromosome distribution of *Dnmt3C*^{KO} DMRs in E18.5 prospermatogonia

A. (Left) Visualization of *Dnmt3C*^{KO} DMRs over WT prospermatogonia across the 21, XY mouse karyotype. (Right insert) Genomic distribution of *Dnmt3C*^{KO} DMRs between intergenic and intragenic compartments.

Extended data figure 4



Extended data figure 4 I Characterization of *Dnmt-tKO::RosaDnmt3C* and *Dnmt-tKO::RosaDnmt3B* mESCs.
A. Western blot showing the expression of DNMT3C and DNMT3B after stable integration at the *Rosa26* locus of *Dnmt*-tKO mESCs, using anti-MYC antibody. Transgenic DNMT3C and DNMT3B constructs possess a 2xMYC tag. **B**. Bar plot showing the level of CpG methylation (in %) in WT mESCs, *Dnmt*-tKO, *Dnmt*-tKO*::Dnmt3C*, and *Dnmt*-tKO*::Dnmt3B,* as determined by LUMA. **C**. Bar plot showing the level of methylated cytosines (in %) in WT mESCs, *Dnmt*-tKO, *Dnmt*-tKO*::Dnmt3C*, and *Dnmt*-tKO*::Dnmt3B,* as determined by LC-MS.

Table 1: DNA methylation status or dependency of every TE function segments

Category	TE functional segments
1- Never demethylated (*n*= 4)	- **LTR5**: IAPEz-int - **LTR body**: IAPEz-int - **LTR3'**: IAPEz-int, IAPLTR2a
2- Never remethylated (*n*= 14)	- **LTR5'**: ERVB3_1-I_MM, ERVB4_3-I_MM, MMERVK10D3_I-int, MMERVK9C_I-int, MMTV-int, MURVY-int, RLTR44-int - **LTR body**: ERVB4_3-I_MM, ERVB7_2-LTR_MM, LTR16D1, RLTR44B, RLTR9C - **LTR3'**: ERVB3_1-I_MM - **Solo LTR**: IAPEY4_LTR
3- DNMT3C dependent (*n*= 35)	- **L1 promoter :** L1MdA_I, L1MdA_II, L1MdA_III, L1MdF_II, L1MdGf_II, L1MdTf_I, L1MdTf_II, L1MdTf_III - **LTR5' :** ETnERV-int, MMERGLN-int, MMETn-int, RLTR10-int, RLTR4_MM-int - **LTR body :** IAPEY2_LTR, IAPLTR2_Mm, MMERGLN_LTR, RLTR10-int, RLTR10C, RLTR4_Mm - **LTR3' :** MMERGLN-int, , MMETn-int, RLTR10-int, RLTR4_MM-int - **Solo LTR** : ERVB7_1-LTR_MM, ETnERV-int, IAPA_MM-int, IAPLTR1_Mm, IAPLTR1a_Mm, IAPLTR2_Mm, IAPLTR2a, IAPLTR2b, MMERGLN_LTR, MMETn-int, RLTR10-int, RLTR10B2, RLTR10C
4- DNMT3A dependent (*n*=1,424)	- **L1 promoter:** L1_Mur1, L1_Mur2, L1_Mur3, L1_Mus3, L1_Rod, L1Lx_I, L1Lx_II, L1Lx_III, L1Lx_IV, L1M2, L1M2a, L1M3c, L1M3e, L1M4, L1M4b, L1M4c, L1MA4, L1MA4A, L1MA6, L1MA7, L1MA8, L1MA9, L1MB1, L1MB2, L1MB3, L1MB5, L1MB7, L1MB8, L1MC1, L1MC2, L1MC3, L1MC5a, L1MCa, L1MD1, L1MD2, L1MDa, L1MdFanc_I, L1MdFanc_II, L1MdMus_II, L1MdV_I, L1MdV_II, L1MdV_III, L1ME1, L1ME2, L1ME3Cz, L1Mec, L1Meg, L1VL1, Lx, Lx10, Lx2, Lx2A, Lx2A1, Lx2B, Lx2B2, Lx3_Mus, Lx3A, Lx3B, Lx3C, Lx4A, Lx4B, Lx5, Lx5b, Lx5c, Lx6, Lx7, Lx8, Lx8b, Lx9, MusHAL1 - **L1 body:** CR1_Mam, CR1-1_Amn, CR1-11_Crp, CR1-12_Ami, CR1-13_Ami, CR1-16_Ami, CR1-3_Croc, CR1-L3A_Croc, CR1-L3B_Croc, HAL1, HAL1b, HAL1M8, HAL1ME, L1_Mur1, L1_Mur2, L1_Mur3, L1_Mus3, L1_Rod, L1Lx_I, L1Lx_II, L1Lx_III, L1Lx_IV, L1M, L1M1, L1M2, L1M2a, L1M2a1, L1M2b, L1M2c, L1M3, L1M3a, L1M3b, L1M3c, L1M3d, L1M3de, L1M3e, L1M3f, L1M4, L1M4a1, L1M4a2, L1M4b, L1M4c, L1M5, L1M6, L1M6B, L1M7, L1M8, L1MA10, L1MA4, L1MA4A, L1MA5, L1MA5A, L1MA6, L1MA7, L1MA8, L1MA9, L1MB1, L1MB2, L1MB3, L1MB4, L1MB5, L1MB7, L1MB8, L1MC, L1MC1, L1MC2, L1MC3, L1MC4, L1MC4a, L1MC5, L1MC5a, L1MCa, L1MCb, L1MCc, L1MD, L1MD1, L1MD2, L1MD3, L1MDa, L1MdA_I, L1MdA_II, L1MdA_III, L1MdA_IV, L1MdA_V, L1MdA_VI, L1MdA_VII, L1MDb, L1MdF_I, L1MdF_II, L1MdF_III, L1MdF_IV, L1MdF_V, L1MdFanc_I, L1MdFanc_II, L1MdGf_I, L1MdGf_II, L1MdMus_I, L1MdMus_II, L1MdN_I, L1MdTf_I, L1MdTf_II, L1MdTf_III, L1MdV_I, L1MdV_II, L1MdV_III, L1ME1, L1ME2, L1ME2z, L1ME3, L1ME3A, L1ME3B, L1ME3C, L1ME3Cz, L1ME3D, L1ME3E, L1ME3F, L1ME3G, L1ME4a, L1ME4b, L1ME4c, L1ME5, L1MEa, L1MEb, L1MEc, L1Med, L1MEf, L1Meg, L1MEg1, L1MEg2, L1Meh, L1Mei, L1MEj, L1P5, L1PB4, L1VL1, L2, L2-1_Ami, L2-1_Crp, L2-2_Mam, L2-3_Crp, L2a, L2b, L2c, L2d, L2d2, L3, L3b, L4_A_Mam, L4_B_Mam, L4_C_Mam, L5, Lx, Lx10, Lx2, Lx2A, Lx2A1, Lx2B, Lx2B2, LX3, Lx3_Mus, Lx3A, Lx3B, Lx3C, Lx4A, Lx4B, Lx5, Lx5b, Lx5c, Lx6, Lx7, Lx8, Lx8b, Lx9, Mam_R4, MamRTE1, MamRTE2, MARE6, MusHAL1, Penelope1_Vert, Plat_L3, UCON49, UCON86, X1_LINE, X12_LINE, X13_LINE, X15_LINE, X17_LINE, X2_LINE, X20_LINE, X21_LINE, X24_LINE, X5B_LINE, X6A_LINE, X6B_LINE, X7A_LINE, X7B_LINE, X7C_LINE, X7D_LINE, X8_LINE, X9_LINE - **SINE:**, AmnSINE1, AmnSINE2, B1_Mm, B1_Mur1, B1_Mur2, B1_Mur3, B1_Mur4, B1_Mus1, B1_Mus2, B1F, B1F1, B1F2, B2_Mm1a, B2_Mm1t, B2_Mm2, B3, B3A, B4, B4A, CRP1, FAM, ID, ID_B1, ID2, ID4, ID4_, LFSINE_Vert, MamSINE1, MIR, MIR1_Amn, MIR3, MIRb, MIRc, PB1, PB1D10, PB1D11, PB1D7, B1D9, RSINE1 - **LTR5' :** ERVB2_1-I_MM, ERVB2_1A-I_MM, ERVB3_1-I_MM, ERVB5_1-I_MM, ERVL-B4-int, ERVL-int, MERV1_I-int, MERVK26-int, MERVL_2A-int, MERVL-int, MurERV4_19-int, MurERV4-int, MYSERV-int, MYSERV6-int, RLTR14-int, RLTR18-int, RLTR19-int, RLTR42-int, RMER15-int, RMER16-int, RMER17C-int, RodERV21-int, SRV_MM-int - **LTR Internal :** BGLII, BGLII_A, BGLII_B, BGLII_B2, BGLII_Mur, BGLII_Mus, ERV3-16A3_I-int, ERVB2_1-I_MM, ERVB2_1A-I_MM, ERVB3_1-I_MM, ERVB3_1-LTR_MM, ERVB4_1C-LTR_Mm, ERVB4_3-LTR_MM, ERVB5_1-I_MM, ERVL-B4-int, ERVL-E-int, ERVL-int, ETnERV2-int, Eutr10, Eutr5, HERV16-int, HERVL40-int, HERVL74-int, LTR10_RN, LTR101_Mam, LTR102_Mam, LTR104_Mam, LTR105_Mam, LTR106_Mam, LTR107_Mam, LTR108e_Mam, LTR16A2, LTR16C, LTR16D, LTR16D2, LTR16E1, LTR16E2, LTR33, LTR33A, LTR33B, LTR33C, LTR37-int, LTR40a, LTR40A1, LTR41, LTR41B, LTR41C, LTR50, LTR52-int, LTR53, LTR53-int, LTR53B, LTR55, LTR65, LTR67B, LTR68, LTR69, LTR72_RN, LTR73, LTR75, LTR75B, LTR78, LTR78B, LTR80A, LTR80B, LTR81, LTR81A, LTR81B, LTR81C, LTR82A, LTR82B, LTR83, LTR84a, LTR84b, LTR85a, LTR85b, LTR85c, LTR86C, LTR87, LTR89, LTR89B, LTR90B, LTR91, LTR91A, LTRIS_Mus, LTRIS2, LTRIS3, LTRIS4, LTRIS4A, MamGyp-int, MamGypLTR1a, MamGypLTR1c, MamGypLTR1d, MamGypLTR2b, MamGypLTR3, MamGypLTR3a, MamGypLTR4, MamGypsy2-I, MamRep1151, MamRep1527, MamRep605, MamRep605b, MER110-int, MER21-int, MER21B, MER21C-int, MER31A, MER34, MER34-int, MER34A, MER34A1, MER34B, MER34B-int, MER34C, MER4B-int, MER52-int, MER54A, MER54B, MER57-int, MER65-int, MER67A, MER67B, MER67C, MER67D, MER68, MER68-int, MER68B, MER70-int, MER70B, MER70C, MER73, MER74A, MER74B, MER76, MER76-int, MER77, MER77B, MER89, MER89-int, MER90, MER90a, MER92-int, MER92B, MER92C, MERV1_I-int, MERV1_LTR, MERVK26-int, MERVL_2A-int, MERVL-int, MLT-int, MLT1-int, MLT1A-int, MLT1A0-int, MLT1A1-int, MLT1B, MLT1B-int, MLT1C, MLT1C-int, MLT1C2, MLT1C2-int, MLT1D, MLT1D-int, MLT1E, MLT1E-int, MLT1E1, MLT1E1-int, MLT1E1A, MLT1E1A-int, MLT1E2, MLT1E2-int, MLT1E3, MLT1E3-int, MLT1F, MLT1F-int, MLT1F1, MLT1F1-int, MLT1F2, MLT1F2-int, MLT1G, MLT1G-int, MLT1G1, MLT1G1-int, MLT1G3, MLT1G3-int, MLT1H, MLT1H-int, MLT1H1, MLT1H1-int, MLT1H2, MLT1H2-int, MLT1I-int, MLT1J, MLT1J-int, MLT1J1-int, MLT1J2-int, MLT1K, MLT1K-int, MLT1L, MLT1L-int, MLT1M, MLT1N2, MLT1N2-int, MLT1O, MLT1O-int, MLT2B1, MLT2B2, MLT2B3, MLT2B4, MLT2B5, MLT2D, MLT2E, MLT2F, MLTR11A, MLTR11B, MLTR12, MLTR13, MLTR14, MLTR18_MM, MLTR18A_MM, MLTR18B_MM, MLTR18C_MM, MLTR18D_MM, MLTR25A, MLTR25C, MLTR31A_MM, MLTR31C_MM, MLTR31D_MM, MLTR31E_MM, MLTR31F_MM, MLTR31FA_MM, MLTR32C_MM, MLTR73, MMERVK9E_I-int, MMTV-int, MMVL30-int, MRLTR33, MT-int, MT2_MmMT2A, MT2B, MT2B1, MT2B2, MT2C_Mm, MTA_Mm-int, MTB, MTB_Mm, MTB_Mm-int, MTB-int, MTC, MTC-int, MTD, MTD-int, MTE-int, MTE2a, MTE2a-int, MTE2b, MTE2b-int, MTEa, MTEa-int, MTEb, MTEb-int, MuLV-int, MurERV4_19-int, MurERV4-int, MYSERV-int, MYSERV16_I-int, MYSERV6-int, ORR1A0, ORR1A0-int, ORR1A1, ORR1A1-int, ORR1A2, ORR1A2-int, ORR1A3, ORR1A3-int, ORR1A4, ORR1A4-int, ORR1B1, ORR1B1-int, ORR1B2, ORR1B2-int, ORR1C1, ORR1C1-int, ORR1C2, ORR1C2-int, ORR1D-int, ORR1D1, ORR1D1-int, ORR1D2, ORR1D2-int, ORR1E, ORR1E-int, ORR1F, ORR1F-int, ORR1G-int, RLTR11A, RLTR11A2, RLTR11B, RLTR11C_MM, RLTR11D, RLTR12A, RLTR12B, RLTR12B2, RLTR12BD_Mm, RLTR12C, RLTR12D, RLTR12E, RLTR12F, RLTR12G, RLTR12H, RLTR13A, RLTR13A1, RLTR13A2, RLTR13A3, RLTR13B1, RLTR13B2, RLTR13B3, RLTR13B4, RLTR13C1, RLTR13C2, RLTR13C3, RLTR13D, RLTR13D1, RLTR13D2, RLTR13D3, RLTR13D3A, RLTR13D3A1, RLTR13D4, RLTR13D5, RLTR13D6, RLTR13E, RLTR13F, RLTR13G, RLTR14_RN, RLTR14-int, RLTR16, RLTR16B_MM, RLTR16C_MM, RLTR17, RLTR17B_Mm, RLTR17C_Mm, RLTR17D_Mm, RLTR18, RLTR18-int, RLTR18B, RLTR19, RLTR19-int, RLTR19A, RLTR19A2, RLTR19B, RLTR19C, RLTR1A2_MM, RLTR1C, RLTR20A, RLTR20A1, RLTR20A2, RLTR20A2B_MM, RLTR20A3_MM, RLTR20A4, RLTR20B1, RLTR20B2, RLTR20B3, RLTR20B3A_MM, RLTR20B4_MM, RLTR20B5_MM, RLTR20C, RLTR20C1_MM, RLTR20C2_MM, RLTR20D, RLTR21, RLTR22_Mur, RLTR22_Mus, RLTR23, RLTR24, RLTR25A, RLTR25B, RLTR26, RLTR26B_MM, RLTR26C_MM, RLTR26D_MM, RLTR28, RLTR28B, RLTR3_Mm, RLTR30B_MM, RLTR30C_MM, RLTR30D_MM, RLTR30D2_MM, RLTR30E_MM, RLTR31_Mm, RLTR31_Mur, RLTR31A_Mm, RLTR31B_Mm, RLTR31B2, RLTR31C_MM, RLTR31D_MM, RLTR31M, RLTR33, RLTR34B_MM, RLTR34C_MM, RLTR34D_MM, RLTR35B_MM, RLTR4_MM-int, RLTR40, RLTR41, RLTR41A2, RLTR41B, RLTR41C, RLTR42-int, RLTR45-int, RLTR48A, RLTR48B, RLTR48C, RLTR49, RLTR50B, RMER10A, RMER10B, RMER12, RMER12B, RMER12C, RMER13A, RMER13A1, RMER13A2, RMER13B, RMER15, RMER15-int, RMER16, RMER16_Mm, RMER16-int, RMER16A2, RMER16A3, RMER16C, RMER17A, RMER17A-int, RMER17C, RMER17C-int, RMER17C2, RMER17D, RMER17D2, RMER19A, RMER19B, RMER19B2, RMER19C, RMER2, RMER20A, RMER20B, RMER20C_Mm, RMER21A, RMER21B, RMER3D-int, RMER4A, RMER4B, RMER5, RMER6-int, RMER6A, RMER6B, RMER6BA, RMER6C, RMER6D, RNERVK23-int, RodERV21-int, SRV_MM-int - **LTR3' :** ERVB2_1-I_MM, ERVB2_1A-I_MM, ERVB4_1-I_MM, ERVB5_1-I_MM, ERVL-B4-int, ERVL-int, ETnERV2-int, MERV1_I-int, MERVK26-int, MERVL_2A-int, MERVL-int, MMERVK9C_I-int, MMERVK9E_I-int, MMTV-int, MMVL30-int, MuLV-int, MurERV4_19-int, MurERV4-int, MuRRS-int, MuRRS4-int, MYSERV-int, MYSERV6-int, RLTR14-int, RLTR18-int, RLTR19-int, RLTR42-int, RMER15-int, RMER16-int, RMER17C-int, RodERV21-int, SRV_MM-int - **Solo LTR :** BGLII, BGLII_A, BGLII_B, BGLII_B2, BGLII_C, BGLII_Mur, BGLII_Mus, ERV3-16A3_I-int, ERV3-16A3_LTR, ERVB2_1-I_MM, ERVB2_1A-I_MM, ERVB3_1-I_MM, ERVB3_1-LTR_MM, ERVB4_1-LTR_MM, ERVB4_1B-LTR_MM, ERVB4_2-I_MM, ERVB4_3-I_MM, ERVB4_3-LTR_MM, ERVB5_1-I_MM, ERVB5_1-LTR_MM, ERVB5_2-LTR_MM, ERVB7_3-LTR_MM, ERVB7_4-LTR_MM, ERVL-B4-int, ERVL-E-int, ERVL-int, ETnERV2-int, Eutr10, Eutr18, Eutr5, EUTREP15, EUTREP16, EUTREP7, EUTREP8, HERV16-int, HERVL40-int, HERVL74-int, LTR10_RN, LTR101_Mam, LTR102_Mam, LTR103_Mam, LTR103b_Mam, LTR104_Mam, LTR105_Mam, LTR106_Mam, LTR107_Mam, LTR108a_Mam, LTR108b_Mam, LTR108c_Mam, LTR108d_Mam, LTR108e_Mam, LTR109A2, LTR16, LTR16A, LTR16A1,

	LTR16A2, LTR16B, LTR16B1, LTR16B2, LTR16C, LTR16D, LTR16D1, LTR16D2, LTR16E1, LTR16E2, LTR33, LTR33A, LTR33A_, LTR33B, LTR33C, LTR37-int, LTR37A, LTR37B, LTR40a, LTR40A1, LTR40b, LTR40c, LTR41, LTR41B, LTR41C, LTR45, LTR45C, LTR50, LTR52, LTR52-int, LTR53, LTR53-int, LTR53B, LTR55, LTR58, LTR65, LTR67B, LTR68, LTR69, LTR72_RN, LTR73, LTR75, LTR75_1, LTR75B, LTR77, LTR78, LTR78B, LTR79, LTR80A, LTR80B, LTR81, LTR81A, LTR81AB, LTR81B, LTR81C, LTR82A, LTR82B, LTR83, LTR84a, LTR84b, LTR85a, LTR85b, LTR85c, LTR86A1, LTR86A2, LTR86B1, LTR86B2, LTR86C, LTR87, LTR88a, LTR88b, LTR88c, LTR89, LTR89B, LTR90A, LTR90B, LTR91, LTR91A, LTRIS_Mm, LTRIS2, LTRIS3, LTRIS4, LTRIS4A, LTRIS5, LTRIS6, MamGyp-int, MamGypLTR1a, MamGypLTR1b, MamGypLTR1c, MamGypLTR1d, MamGypLTR2b, MamGypLTR2c, MamGypLTR3, MamGypLTR3a, MamGypLTR4, MamGypsy2-I, MamGypsy2-LTR, MamRep1151, MamRep1527, MamRep605, MamRep605b, MER101B, MER110, MER110-int, MER110A, MER129, MER21-int, MER21B, MER21C, MER31-int, MER31A, MER31B, MER34, MER34-int, MER34A, MER34A1, MER34B, MER34B-int, MER34C, MER34C_, MER4B-int, MER52-int, MER54A, MER54B, MER57-int, MER57C2, MER57D, MER57E1, MER57E2, MER65-int, MER65D, MER67A, MER67B, MER67C, MER67D, MER68, MER68-int, MER68B, MER68C, MER70-int, MER70A, MER70B, MER70C, MER73, MER74A, MER74B, MER74C, MER76, MER76-int, MER77, MER77B, MER83A-int, MER83B-int, MER88, MER89, MER89-int, MER90, MER90a, MER92-int, MER92A, MER92B, MER92C, MER92D, MER95, MERV1_I-int, MERV1_LTR, MERVK26-int, MERVL_2A-int, MERVL-int, MLT-int, MLT1-int, MLT1A, MLT1A-int, MLT1A0, MLT1A0-int, MLT1A1, MLT1A1-int, MLT1B, MLT1B-int, MLT1C, MLT1C-int, MLT1C2, MLT1C2-int, MLT1D, MLT1D-int, MLT1E, MLT1E1, MLT1E1-int, MLT1E1A, MLT1E1A-int, MLT1E2, MLT1E2-int, MLT1E3, MLT1E3-int, MLT1F, MLT1F-int, MLT1F1, MLT1F1-int, MLT1F2, MLT1F2-int, MLT1G, MLT1G-int, MLT1G1, MLT1G1-int, MLT1G3, MLT1G3-int, MLT1H, MLT1H-int, MLT1H1, MLT1H1-int, MLT1H2, MLT1H2-int, MLT1I, MLT1I-int, MLT1J, MLT1J-int, MLT1J1, MLT1J1-int, MLT1J2, MLT1J2-int, MLT1K, MLT1K-int, MLT1L, MLT1L-int, MLT1M, MLT1N2, MLT1O, MLT2B1, MLT2B2, MLT2B3, MLT2B4, MLT2B5, MLT2C1, MLT2C2, MLT2D, MLT2E, MLT2F, MLTR11A, MLTR11B, MLTR12, MLTR13, MLTR14, MLTR18_MM, MLTR18A_MM, MLTR18B_MM, MLTR18C_MM, MLTR18D_MM, MLTR25A, MLTR25C, MLTR31A_MM, MLTR31C_MM, MLTR31D_MM, MLTR31E_MM, MLTR31F_MM, MLTR31FA_MM, MLTR32C_MM, MLTR73, MMERGLN-int, MMERVK9E_I-int, MMTV-int, MMVL30-int, MRLTR33, MT-int, MT2_Mm, MT2A, MT2B, MT2B1, MT2B2, MT2C_Mm, MTA_Mm, MTA_Mm-int, MTB, MTB_Mm, MTB_Mm-int, MTB-int, MTC, MTC-int, MTD, MTD-int, MTE-int, MTE2a, MTE2a-int, MTE2b-int, MTEa, MTEa-int, MTEb, MTEb-int, MuLV-int, MurERV4_19-int, MurERV4-int, MuRRS-int, MuRRS4-int, MURVY-int, MURVY-LTR, MYSERV-int, MYSERV16_I-int, MYSERV6-int, ORR1A0, ORR1A0-int, ORR1A1, ORR1A1-int, ORR1A2, ORR1A2-int, ORR1A3, ORR1A3-int, ORR1A4, ORR1A4-int, ORR1B1, ORR1B1-int, ORR1B2, ORR1B2-int, ORR1C1, ORR1C1-int, ORR1C2, ORR1C2-int, ORR1D-int, ORR1D1, ORR1D1-int, ORR1D2, ORR1D2-int, ORR1E, ORR1E-int, ORR1F, ORR1F-int, ORR1G, ORR1G-int, RLTR1, RLTR11A, RLTR11A2, RLTR11B, RLTR11C_MM, RLTR11D, RLTR12A, RLTR12B, RLTR12B2, RLTR12BD_Mm, RLTR12C, RLTR12D, RLTR12E, RLTR12F, RLTR12G, RLTR12H, RLTR13A, RLTR13A1, RLTR13A2, RLTR13A3, RLTR13B1, RLTR13B2, RLTR13B3, RLTR13B4, RLTR13C1, RLTR13C2, RLTR13C3, RLTR13D, RLTR13D1, RLTR13D2, RLTR13D3, RLTR13D3A, RLTR13D3A1, RLTR13D4, RLTR13D5, RLTR13D6, RLTR13E, RLTR13F, RLTR14, RLTR14_RN, RLTR14-int, RLTR15, RLTR16, RLTR16B_MM, RLTR16C_MM, RLTR17, RLTR17B_Mm, RLTR17C_Mm, RLTR17D_Mm, RLTR18, RLTR18-int, RLTR18B, RLTR19, RLTR19-int, RLTR19A, RLTR19A2, RLTR19B, RLTR19C, RLTR19D, RLTR1A2_MM, RLTR1C, RLTR1D, RLTR1D2_MM, RLTR1E_MM, RLTR20A, RLTR20A1, RLTR20A2, RLTR20A2B_MM, RLTR20A3_MM, RLTR20A4, RLTR20B1, RLTR20B2, RLTR20B3, RLTR20B3A_MM, RLTR20B4_MM, RLTR20B5_MM, RLTR20C, RLTR20C1_MM, RLTR20C2_MM, RLTR20D, RLTR21, RLTR22_Mur, RLTR22_Mus, RLTR23, RLTR24, RLTR24B_MM, RLTR25A, RLTR25B, RLTR26, RLTR26_Mus, RLTR26B_MM, RLTR26C_MM, RLTR26D_MM, RLTR28, RLTR28B, RLTR3_Mm, RLTR30, RLTR30B_MM, RLTR30C_MM, RLTR30D_MM, RLTR30D_RN, RLTR30D2_MM, RLTR30E_MM, RLTR31_Mm, RLTR31_Mur, RLTR31A_Mm, RLTR31B_Mm, RLTR31B2, RLTR31C_MM, RLTR31D_MM, RLTR31M, RLTR33, RLTR34B_MM, RLTR34C_M, RLTR34D_MM, RLTR35B_MM, RLTR4_Mm, RLTR4_MM-int, RLTR40, RLTR41, RLTR41A2, RLTR41B, RLTR41C, RLTR42, RLTR42-int, RLTR43A, RLTR43C, RLTR45, RLTR45-int, RLTR47, RLTR47_MM, RLTR48A, RLTR48B, RLTR48C, RLTR49, RLTR50A, RLTR51A_Mm, RLTR51B_Mm, RLTR53_Mm, RLTR53B_M, RLTR6-int, RLTR6B_Mm, RLTR8, RLTR9A, RLTR9A2, RLTR9A3, RLTR9A3A, RLTR9A3B, RLTR9A4, RLTR9B, RLTR9B2, RLTR9C, RLTR9D, RLTR9D2, RLTR9E, RLTR9F, RMER10A, RMER10B, RMER12, RMER12B, RMER12C, RMER13A, RMER13A1, RMER13A2, RMER13B, RMER15, RMER15-int, RMER16, RMER16_Mm, RMER16-int, RMER16A2, RMER16A3, RMER16B, RMER16B2, RMER16B3, RMER16C, RMER17A, RMER17A-int, RMER17A2, RMER17B, RMER17B2, RMER17C, RMER17C-int, RMER17C2, RMER17D, RMER17D2, RMER19A, RMER19B, RMER19B2, RMER19C, RMER2, RMER20A, RMER20B, RMER20C_Mm, RMER21A, RMER21B, RMER3D-int, RMER3D1, RMER3D2, RMER3D3, RMER3D4, RMER4A, RMER4B, RMER5, RMER6-int, RMER6A, RMER6B, RMER6BA, RMER6C, RMER6D, RNERVK23-int, RNLTR23, RodERV21-int, SRV_MM-int, UCON51, X1_LR
5- Mixed DNMT3A-DNMT3C-dependent (*n*=160)	- **L1 promoter :** L1MA5, L1MdA_IV, L1MdA_V, L1MdA_VI, L1MdA_VII, L1MdF_I, L1MdF_III, L1MdF_IV, L1MdF_V, L1MdGf_I, L1MdMus_I, L1MdN_I, L1ME3G - **LTR5':** ERVB4_1B-I_MM, ERVB4_2-I_MM, ETnERV2-int, ETnERV3-int, IAP-d-int, IAP1-MM_I-int, IAPEy-int, IAPEY3-int, IAPEY4_I-int, IAPLTR3-int, MMERVK10C-int, MMERVK9E_I-int, MMVL30-int, MuLV-int, MuRRS-int, MuRRS4-int, RLTR1B-int, RLTR45-int, RLTR6-int - **LTR body:** ERVB4_1-I_MM, ERVB4_1B-I_MM, ERVB4_1B-LTR_MM, ERVB4_2-I_MM, ERVB4_2-LTR_MM, ERVB5_1-LTR_MM, ERVB5_2-LTR_MM, ETnERV-int, ETnERV3-int, IAP-d-int, IAP1-MM_I-int, IAPEy-int, IAPEY3-int, IAPEY4_I-int, IAPEY5_I-int, IAPLTR2a2_Mm, IAPLTR3-int, IAPLTR4_I, LTRIS_Mm, MamGypLTR1b, MMERGLN-int, MMERVK10C-int, MMERVK10D3_I-int, MMERVK9C_I-int, MMETn-int, MuRRS-int, MuRRS4-int, MURVY-int, MURVY-LTR, RLTR1, RLTR10, RLTR1B, RLTR1B-int, RLTR1D, RLTR1D2_MM, RLTR1E_MM, RLTR1F_Mm, RLTR26_Mus, RLTR44-int, RLTR44C, RLTR44D, RLTR45, RLTR5_Mm, RLTR6_Mm, RLTR6-int, RLTR6B_M, RLTR6C_Mm, RLTR9B, RMER17A2, RMER17B, RMER17B2 - **LTR3':** ERVB4_1B-I_MM, ERVB4_2-I_MM, ERVB4_3-I_MM, ETnERV-int, ETnERV3-int, IAP-d-int, IAP1-MM_I-int, IAPEy-int, IAPEY3-int, IAPEY4_I-int, IAPLTR3-int, MMERVK10C-int, MMERVK10D3_I-int, MURVY-int, RLTR1B-int, RLTR44-int, RLTR45-int, RLTR6-int - **Solo LTR:** ERVB4_1-I_MM, ERVB4_1B-I_MM, ERVB4_1C-LTR_Mm, ERVB4_2-LTR_MM, ERVB7_2-LTR_MM, ERVB7_2B-LTR_MM, ETnERV3-int, IAP-d-int, IAP1-MM_I-int, IAP1-MM_LTR, IAPEY_LTR, IAPEy-int, IAPEY2_LTR, IAPEY3_LTR, IAPEY3-int, IAPEY3C_LTR, IAPEY4_I-int, IAPEY5_I-int, IAPEY5_LTR, IAPEz-int, IAPLTR2a2_Mm, IAPLTR3, IAPLTR3-int, IAPLTR4, IAPLTR4_I, LTRIS_Mus, LTRIS4B, MMERVK10C-int, MMERVK10D3_I-int, MMERVK10D3_LTR, MMERVK9C_I-int, MMETn-int, RLTR10, RLTR10A, RLTR10B, RLTR10D, RLTR10D2, RLTR10E, RLTR10F, RLTR13G, RLTR1B, RLTR1B-int, RLTR1F_Mm, RLTR27, RLTR44-int, RLTR44A, RLTR44B, RLTR44C, RLTR44D, RLTR44E, RLTR46, RLTR46A, RLTR46A2, RLTR46B, RLTR5_Mm, RLTR50B, RLTR6_Mm, LTR6C_Mm, RLTRETN_Mm

DISCUSSION

1 General conclusion and personal impression

When I started my thesis, the exact function of DNMT3A in male germ cell development was both unclear and controversial, especially as DNMT3C—a novel DNA methyltransferase with male germ cell-specific functions—had just been identified by our lab. My main work therefore aimed at deciphering what sequences DNMT3A methylates and what purpose it serves for spermatogenesis. Using genome-wide methylation profiling performed by a postdoc researcher in the lab, I identified for the first time DNMT3A's precise genomic targets, namely the whole genome except for young TE promoters. Furthermore, I revealed DNMT3A's developmental functions: DNMT3A is required for SSCs to differentiate, likely by preventing DNA methylation-sensitive TFs to bind to enhancers of stem cell-related genes. As a second part of my PhD research, I participated in deciphering DNMT3C's precise genomic targets and their relationship with underlying chromatin marks. What we found was particularly thrilling for me: prior to their methylation, DNMT3C targets show a bivalent H3K4me3-K9me3 chromatin signature, which brings new knowledge to the understanding of TE regulation during male germ cell development.

Working on these two subjects was a rich experience, as they allowed me to experience many different aspects of biological research: developmental biology, mouse work and management, molecular biology and genomics.

DNMT3A developmental characterization was certainly the most exciting part of my PhD, because I had to use groundbreaking techniques that had not been used in the lab before, like 10X scRNA-seq for instance. However, deciphering the exact developmental function of DNMT3A did not come without its challenges. As a matter of fact, I rapidly found out that, in *Dnmt3A* mutants, the developmental phenotype was linked to SSCs, implying that I would have to work on this type of cells which were not, prior to my thesis studies, a field of expertise of the laboratory. Moreover, this project required a high number of different mutant mouse lines that were quite challenging to manage, mainly due to their extremely low reproductive performance. Finally, to complete the DNMT3A project, I had to adapt Cut&Run to SSCs, as I spent some time realizing that the protocol that was routinely used in the lab failed to work on SSCs, although it worked on prospermatogonia.

Of course, as such is the way with research in biology, many questions stay unresolved. During the course of my work, I made a number of intriguing observations, which I would like to further discuss. I categorized them in four themes: germline epigenetics, transposable element biology, germline development and spermatogenesis, and DNMT3A-dependent DNA methylation in stem cells. I will focus on discussing the implications of this work, the perspectives it opens for future research, and the remaining questions that were raised during my PhD studies.

2 Germline epigenetics

2.1 "You only had one job!"

In this part, we basically solved the basis and function(s) of the epigenetic programming of male germ cells by DNA methylation (Dura et al., 2021). This started with mapping DNA methylation in various *Dnmt3* mutants by WGBS at E18.5, therefore at the end of the establishment of *de novo* DNA methylation. Published methylomes for *Dnmt3C*^{KO} and *Dnm3L*^{KO} where already available (Barau et al., 2016). However, they were performed in post-natal germ cells (pool of undifferentiated and differentiated spermatogonia) at 10dpp, more than a week after the end of *de novo* DNA methylation in prospermatogonia. Until then, one could not exclude some minor DNA methylation changes happening after birth (Kubo et al., 2015) or secondary/compensatory effects that would obscure the real breadth of DNMT3C targets during fetal germ cell development. By carrying out methylation profiling directly in the germ cells that undergo *de novo* DNA methylation, we were able to identify in a precise and exhaustive manner the genomic targets of every DNMT3s expressed in prospermatogonia. We could prove the specificity of DNMT3C towards young TE promoters and the broad action of DNMT3A in methylating the rest of the genome. Additionally, using a constitutive double-mutant for *Dnmt3A* and *Dnmt3C*, we disproved DNMT3B as having significant role in methylating male germ cells.

Focusing on TEs, we demonstrated that some TE families were methylated collaboratively by DNMT3A and DNMT3C. However, although we found that DNMT3A methylates more TE-derived sequences than DNMT3C does (more than 70 times

more), we excluded DNMT3A-dependent methylation as having a role in silencing TEs. This means that methylation of the body of young TEs is not required for silencing, and more importantly, that methylation of certain classes of young TEs, such as ERVL and SINEs, is not necessary for their silencing during spermatogenesis. This last conclusion is an important step for the field, because DNMT3A was until now still referred to as the enzyme responsible for TE methylation and silencing during spermatogenesis.

In conclusion, my work demonstrated that DNMT3A and DNMT3C, stimulated by their cofactor DNMT3L, non-redundantly methylate the male germ cell genome, and they control different steps of spermatogenesis through the genomic sequences they specifically methylate: stem cell-based supply of spermatogenesis for DNMT3A or meiosis for DNMT3C (Fig. 17).

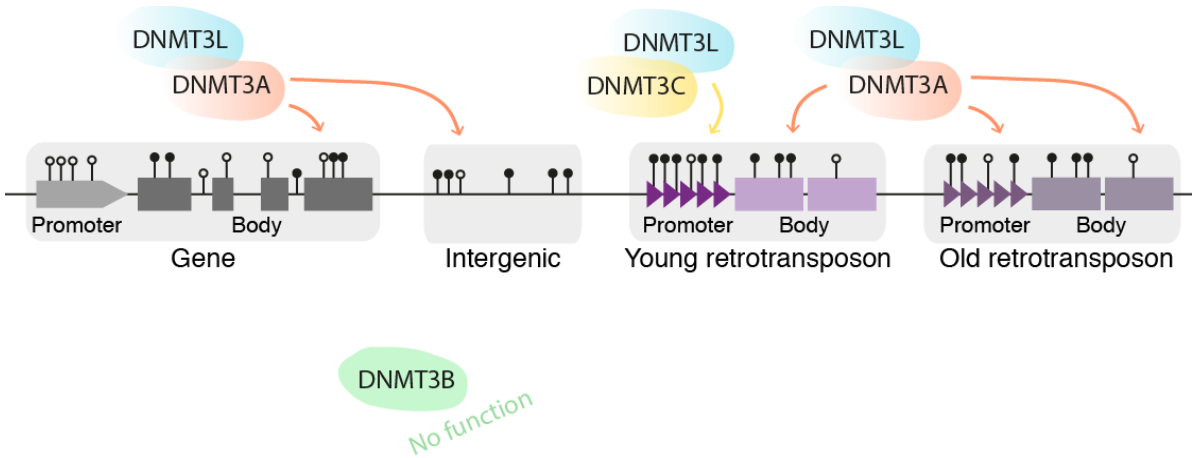


Figure 17 Drawing representing the specific methylation roles of DNMT3A and DNMT3C, associated with DNMT3L.

2.2 Recruitment modes of DNMT3A and DNMT3C to their genomic targets

DNMT3A and DNMT3C shape the male germ cell genome in a collaborative and non-redundant manner. However, the recruitment mechanisms for each enzyme, which direct them towards their specific genomic targets, are very different.

- DNMT3A recruitment

A recent paper demonstrated that H3K36me2 covers the genome of prospermatogonia, resulting from the action of the H3K36 dimethylase NSD1, and that this mark is necessary for establishing DNA methylation genome-wide (Shirane et al., 2020). However, the link with DNMT3A had not been done, as the targets of DNMT3A in prospermatogonia were not known at that time. By revealing that DNMT3A broadly methylates the prospermatogonia genome, we wanted to further investigate the genetic relationship between NSD1 and DNMT3A by performing a side analysis of the overlap between the DMRs found in *Nsd1*^{KO} and *Dnmt3A*^{KO} relative to their WT controls. As shown in Fig. 18, we found that all *Nsd1*^{KO} DMRs overlapped with *Dnmt3A*^{KO}, while none overlapped with *Dnmt3C*^{KO}. However, we noticed that only a small fraction of *Dnmt3A*^{KO} DMRs overlapped with *Nsd1*^{KO} DMRs. Different timepoints were used between *Nsd1*^{KO} (E16.5, (Shirane et al., 2020)) and *Dnmt3A*^{KO} (E18.5, our study) methylation datasets, which may explain some of the non-overlap. However, giving the large difference, it is also likely that DNMT3A may be recruited by other modes. We could hypothesize that other members of the NSD family, namely NSD2 and NSD3, could also participate to H3K36me2 deposition and DNMT3A recruitment.

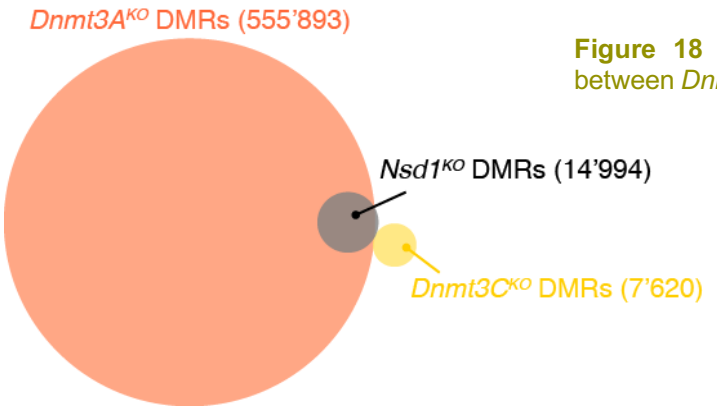


Figure 18 Venn diagram showing the overlap between *Dnmt3A*^{KO}, *Nsd1*^{KO} and *Dnmt3C*^{KO} DMRs.

- DNMT3C recruitment

As of today, how DNMT3C recognizes young TE promoters is still an enigma. As it was presented in the Introduction of this thesis, *de novo* DNA methylation of young TEs is dependent on proteins of the piRNA pathway and others proteins like SPOCD1, TEX15, which may act as intermediates. This observation indicates that DNMT3C may be guided by a massive protein complex that I believe will be more understood in a few years. By revealing that DNMT3C genomic targets bear bivalent H3K4me3/H3K9me3

chromatin prior to being methylated by DNMT3C may indicate some function of this peculiar chromatin in DNMT3C targeting. A protein with the ability to recognize these two marks could recognize these peculiar genomic regions and direct DNMT3C towards them.

Another aspect that has been less studied is the role that the domain structure of DNMT3C may play in its specificity towards young TE promoters. Three hypotheses will be discussed below, none of them being exclusive to the others:

1) Can DNMT3C recognize a specific genomic sequence?

We artificially forced the expression of DNMT3C in embryonic stem cells to compare DNMT3C targeting in male germ cells *versus* cells that lack effective piRNA function. We found that only part of DNMT3C targets observed in prospermatogonia were also methylated by DNMT3C in embryonic stem cells: ERV1 and ERVK were significantly methylated in *Dnmt3C* transgenic ESCs, while L1 promoters were not.

We could envision that DNMT3C recruitment may occur according to two modes, depending on the TE family: it would need the piRNA pathway for L1 methylation, but not for ERV methylation. This bimodal recruitment mode could be tested by assessing the DMR overlap between *Dnmt3C* mutants and *Miwi2/Mili* mutants. If this hypothesis is correct, we would expect a biased defect in DNA methylation on L1s in piRNA mutants, while a general DNA methylation defect on both L1s and ERVs is observed in *Dnmt3C* mutants.

Using this ectopic cellular system, we also revealed that DNMT3C preferentially methylates H3K9me3-enriched ERVs. However, our results suggest that H3K9me3 enrichment is not sufficient nor absolutely necessary for being a DNMT3C target in this system: *i)* H3K9me3 is also present on L1s that are not targeted by DNMT3C, and elsewhere in the genome outside of TEs, and *ii)* DNMT3C HMRs only overlapped with one-third of H3K9me3 peaks. This observation might suggest that DNMT3C may recognize a specific flavor of H3K9me3 chromatin, determined by a specific H3K9 trimethylase among SETDB1, SUV39H1 and SUV39H2. On this matter, SETDB1 together with KAP1 was shown to mainly control ERVK, and some ERV1 (Karimi et al. 2011), while the SUV39H system can catalyze H3K9me3 at both young ERVs and L1s, although the effect may be the strongest for young L1s (Bulut-Karslioglu et al., 2014). Anyway, causality of H3K9me3 for recruiting DNMT3C could be tested by knocking out

Setdb1 in transgenic *Dnmt3C* ESCs, and look whether DNA methylation still occurs at ERVK and ERV1. Another possibility is that DNMT3C could recognize, directly or indirectly, specific DNA sequences that happen to be enriched in H3K9me3. This hypothesis could be tested by searching for motifs among the DNMT3C targets or by performing an *in vitro* assay testing the interaction between DNMT3C and different TE sequences (like an EMSA assay), provided that the recognition is direct in this case.

2) Could the absence of the PWWP domain give DNMT3C its specificity towards young TEs?

DNMT3C diverges from DNMT3A and DNMT3B by the lack of a PWWP chromatin-reading domain. The absence of PWWP implies that DNMT3C may not recognize H3K36me2/3, which is present in abundance in highly transcribed genes and over the genome, and could therefore focus on young TE promoters. Moreover, it was recently shown that the PWWP domain prevents DNMT3A from methylating domains occupied by H2AUb, a mark catalyzed by the polycomb repressive complex 1 (PRC1) (Weinberg et al., 2021). There is no available H2Aub maps in prospermatogonia, so we do not know whether DNMT3C targets are enriched for this mark. However, H2Aub and H3K27me3 are usually colocalized and we revealed that at E17.5, DNMT3C targets were not enriched in H3K27me3. DNMT3A was shown to recognize H2Aub via a potential UDR domain (Ubiquitin Domain Recognition); we could search for the existence of such domain in the N-terminus domain of DNMT3C.

An *in vivo* experiment testing the role of the lack of PWWP for DNMT3C function is currently in progress in the lab. I started this project as a Master student, and then trained another Master student, Mélanie Armand, who is now an engineer in the lab,

Figure 19 Schematic of the different construct tested by *in vivo* additive transgenesis

working under my supervision on this project. Briefly, the project consists in integrating in *Dnmt3C*^{KO} mice various DNMT3C-related constructs inserted by homologous recombination at the *Tigre* safe harbor locus, to perform an *in vivo* complementation by additive transgenesis. Namely, we plan on inserting a WT DNMT3C construct (rescue control), a DNMT3C construct with an artificial PWWP domain (DNMT3C::PWWP), and a DNMT3B construct without a PWWP domain (DNMT3BΔPWWP), to resemble DNMT3C (Fig. 19). All are under the control of the *Miwi2* promoter that was previously shown to properly drive expression in prospermatogonia (Itou et al., 2015), and they carry a 2xMYC tag. Conveniently, the PWWP domain is encoded by two exons whose length is an exact multiple of 3 (symmetrical exons) and can therefore be inserted or removed without altering protein frames.

The functionality of the integrated protein is tested by its ability to rescue the *Dnmt3C*^{KO} phenotype. More precisely, we measure male fertility, DNA methylation level at TE promoters (young L1s and IAPs, by pyrosequencing) and TE expression around the time of meiosis (RT-qPCR). As a proof of principle, we were able to show that two copies of the WT *Dnmt3C* transgene (integrated at the *Tigre* locus) fully rescued the *Dnmt3C*^{KO} phenotype: males are fertile and TEs are properly methylated and silenced. Concerning the *Dnmt3C::PWWP* transgene, we were never able to recover homozygous transgenic animals, males or females, and independently of the *Dnmt3C*^{KO} background, suggesting embryonic lethality (Fig. 20). This result might already suggest a different function for DNMT3C::PWWP and DNMT3C. To avoid embryonic lethality, we consider inserting the PWWP-encoding exons directly at the endogenous *Dnmt3C* locus.

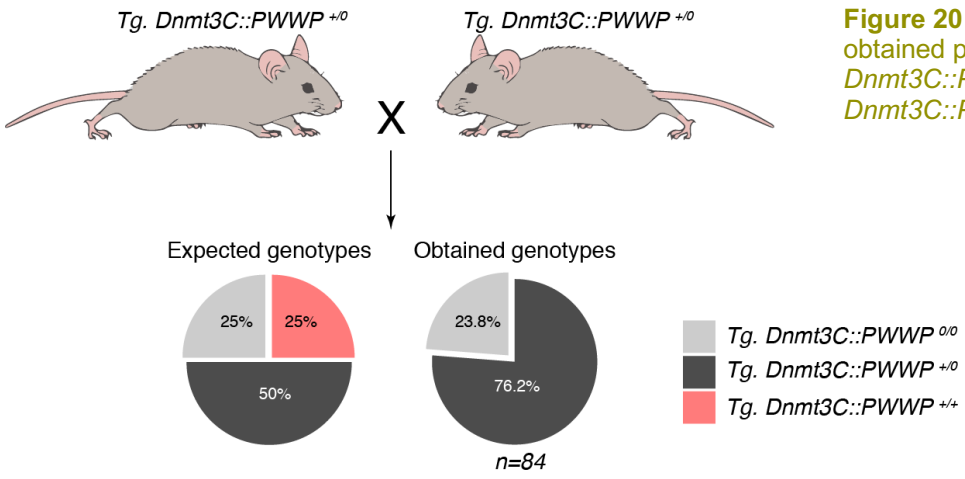
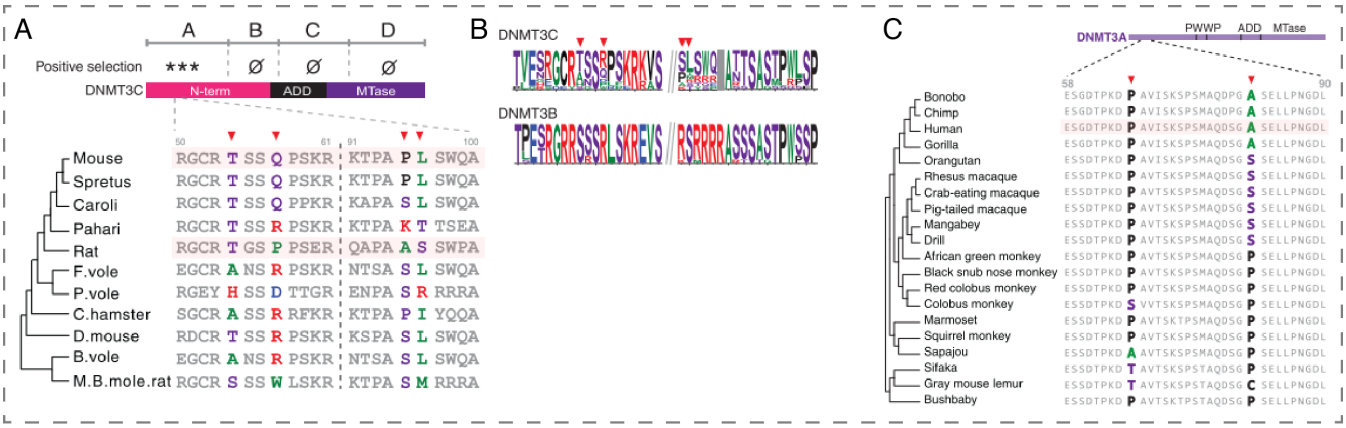


Figure 20 Schematic of expected and obtained pup genotypes for (*Tg. Dnmt3C::PWWP*$^{+/0}$ by *Tg. Dnmt3C::PWWP*$^{+/0}$) crosses

3) Does DNMT3C N-terminus play a role in its specificity towards TE promoters?

The N-terminus of DNMT3C is also intriguing, as discussed in the Introduction: it evolved under positive selection, which implies involvement in a genetic arms race, and potential direct interaction with TE sequences or proteins that recognize TE sequences. Four amino acids (AA) rapidly evolve among rodent species (Fig. 21) (Molaro et al., 2020). To probe the role of these AAs in TE recognition/silencing, we could engineer point mutations at these AAs and test the mutated DNMT3C in our *in vivo* additive transgenesis: introgression of the mutated *Dnmt3C* in *Dnmt3C*^{KO} mice, and measurement of its ability to rescue the phenotype. Alternatively, we could test, in the same *in vivo* system, the ability of the rat DNMT3C to recognize mouse TEs. As a matter of fact, the rat DNMT3C is very similar to the mouse DNMT3C except at these rapidly evolving N-terminal AAs (Fig. 21). If the N-terminus plays a role in TE silencing,



Adaptated from: Molaro, Antoine, Harmit S. Malik, Deborah Bourc'his, and Yoko Satta. 2020. "Dynamic Evolution of de Novo DNA Methyltransferases in Rodent and Primate Genomes." Molecular Biology and Evolution 37(7): 1882–92.

Figure 21 Comparative genomics of the N-terminus of DNMT3s in mammals
A and B. Positive selection of the DNMT3C N-terminus in rodents, compared to DNMT3B
C. Positive selection of the DNMT3A N-terminus in primates

we could assume that the rat DNMT3C will not be able to rescue the lack of DNMT3C in mice. Indeed, even though mechanisms to silence TEs can be conserved between species, the TEs themselves are largely different, inducing the inability to repress TEs in an inter-species manner, as demonstrated in *Drosophila* (Parhad et al., 2017).

2.3 How do humans manage to be fertile without DNMT3C?

As detailed in the Introduction, *Dnmt3C* appeared by tandem duplication of *Dnmt3B* only in Muroidea rodents, between 46 to 60 million years ago. In rodents,

DNMT3C function is essential for male fertility because it silences young TEs. Therefore, it raises the question as to how DNMT3C-less mammals, including men, manage to silence their TEs and to protect their fertility?

DNMT3C function must therefore be carried out by DNMT3A or DNMT3B in humans. Interestingly, like in DNMT3C, two amino acids located in the N-terminus domain of DNMT3A have been shown to evolve under positive selection in primates, while it was not the case for DNMT3B (Fig. 21) (Molaro et al., 2020). This observation indicates that DNMT3A might be the enzyme involved in the battle against TEs in primates. Moreover, if the absence of the PWWP domain matters for targeting young TEs—as hypothesized above—the production of a DNMT3A isoform lacking the PWWP domain would be possible by alternative splicing. As explained above, the PWWP domain is coded by two exons, and the number of base pairs in these two exons is a multiple of three, so that the exclusion of these exons would not alter the protein frame.

2.4 The (not so) sensitive case of male meiosis

In males, the first prophase of meiosis is a critical step of spermatogenesis, and failure to methylate TEs seems particularly damageable for meiosis. This was first suggested in male *Dnmt3L* mutant germ cells, who have a genome-wide DNA methylation defect, and cannot progress past the pachytene stage of meiosis as a result of apoptosis (Bourc'his and Bestor, 2004). Then, a paper reported that *Dnmt3A* mutants phenocopied *Dnmt3L* mutants, molecularly and developmentally, although spermatogenesis characterization was rather rudimentary (Kaneda et al., 2004). More recently, a number of studies reported that a DNA methylation defect that is specifically localized at TE promoters, as observed in *Dnmt3C*, *Miwi2*, or *Spocd1* mutants, was enough to reproduce the developmental phenotype of *Dnmt3L* mutants, namely a pachytene arrest (Barau et al., 2016; Carmell et al., 2007; Zoch et al., 2020).

This shows that a specific lack of DNA methylation at TE promoters is enough to interfere with the meiotic process (as observed in $Dnmt3C^{KO}$). However, it still unclear whether additional methylation defects elsewhere in the genome may be also be detrimental to meiosis (as observed in $Dnmt3L^{KO}$). On this matter, the appearance of female-specific meiotic hotspots were reported in male $Dnmt3L^{KO}$ spermatocytes

(Brick et al., 2018). Showing that *Dnmt3A*^{KO} cells can undergo meiosis, despite a heavily hypomethylated genome, was therefore a provocative finding, regarding the importance of chromatin-based mechanisms in meiosis regulation. In other words, while meiotic recombination occurs genome-wide, we proved that only TE promoter methylation (1% of the genome) is essential for meiosis.

The question as to why meiosis is sensitive to TE reactivation is still open. Strikingly, in all mutants linked to TE derepression in male germ cells, while the DNA methylation defect occurs early, in fetal prospermatogonia, the developmental arrest occurs only weeks later, after birth, in meiotic cells (Aravin et al., 2008; Barau et al., 2016; Carmell et al., 2007). It may come from the time of TE reactivation in these mutants. For L1s, it seems that H3K9me2 takes an important role for maintaining them repressed until the entry of meiosis, when this mark is developmentally programmed to disappear (Di Giacomo et al., 2013). If L1s fail to be properly methylated during fetal development, meiotic cells will therefore be left with neither H3K9me2, nor DNA methylation to silence these elements, as it was shown in *Dnmt3L* mutants (Zamudio et al. 2015). The situation for IAPs is different though: in *Miwi2* and *Dnmt3L* mutants, IAPs start being re-expressed as soon as in spermatogonia after birth (Zamudio et al. 2015; Vasiliauskaitė et al. 2018). However, spermatogonia seem to tolerate such expression, even though IAP elements could create chimeric transcripts with neighboring genes (Vasiliauskaitė et al. 2018). Now, as to why the absence of DNA methylation on TE promoters is deleterious for meiosis, some answers have started being provided. In *Dnmt3L* mutants, our lab previously showed that IAPs and L1s acquired abnormal H3K4me3 enrichment, related to their transcriptional activity (Zamudio et al. 2015). This may perturb the meiotic chromatin landscape, the distribution of meiotic hot spots, and then chromosome recognition and pairing. However, these observations were done using targeted approaches (ChIP-qPCR), and more importantly, in *Dnmt3L* mutants, in which the DNA methylation occurs genome-wide. To really understand the impact of TEs—specifically—on the meiotic process, the *Dnmt3C* mutant is a better model, and chromatin states and meiotic double-strand break distribution should now be assessed genome-wide, to provide complete and unbiased information.

2.5 Perspective for paternal epigenetic inheritance in the early embryo

Finally, I would like to mention one exciting outcome of our findings related to germline epigenetic inheritance. By demonstrating that young *Dnmt3A*^{KO} males can produce spermatozoa (even they do so only once in their life, from the SSC-independent first round of spermatogenesis), we opened the path for studying the role of paternal DNA methylation inheritance in embryonic development. This was previously not conceivable, as *Dnmt3L*^{KO} and *Dnmt3C*^{KO} males never produce spermatozoa.

3 Transposon biology

We revealed for the first time that bivalent H3K4me3/H3K9me3 chromatin decorates DNMT3C-dependent TE targets. This observation suggests a complex regulation of young TEs during male germline development, and raises important mechanistic questions. More particularly, how can these doubly decorated TEs be transcriptionally active, knowing that H3K9me3 characterizes constitutive heterochromatin? They are some important precedents related to active transcription within H3K9me3-enriched domains: in *Drosophila*, the piRNA clusters (transposon-rich genomic loci from where piRNAs are produced) are enriched in H3K9me3, yet they are transcriptionally active. Transcription in this context depends on a protein complex that recognizes H3K9me3 specifically at these genomic loci, RHINO, a paralog of HP1 (Ozata et al., 2019). A similar mechanism could be present in the mouse, allowing TEs to escape from transcriptional silencing. Evolution of such a complex mechanism could be interpreted two ways. If other genomic loci are also regulated in the same manner, TEs could have highjacked this mechanism for their own interest in proliferating. Alternatively, evolution of a such mechanism (and related protein) for the specific purpose to serve TE expression may imply that TEs exert an important cellular function in prospermatogonia. If the latter is true, a parallel could be established with piRNA clusters in *Drosophila*, which have a clear cellular function (repressing TEs).

This thought is related to a second intriguing result that I observed: L1 promoters may undergo a specific demethylation at E15.5. At this developmental timepoint, germline reprogramming is achieved, in the sense that somatic DNA methylation patterns have been erased and DNA methylation re-establishment on the genome

(except TEs) is ongoing. If prospermatogonia have no obvious advantage in TE expression, why would L1s become more demethylated at E15.5 than at E13.5? L1 transcription is needed to feed the piRNA pathway and re-install DNA methylation at L1 loci, but if L1s were not expressed in the first place, the piRNA-DNA methylation pathway would not be required. IAPs also retain high DNA methylation levels at E13.5 but contrary to L1s, they keep high DNA methylation throughout spermatogonia development. Their specific protection against demethylation may imply that reactivation of these IAP copies is highly dangerous. Alternatively, lack of further demethylation of IAPs may suggest that contrary to L1s, they do not exert essential function for the prospermatogonia.

Before further speculating on the subject, we would first need to confirm this result—obtained by targeted pyrosequencing— using genome-wide DNA methylation profiling at E15.5. Secondly, it would be interesting to assess whether late L1 demethylation is dependent on TET proteins, and if yes, which one. The experiment would be complicated to carry on: germline conditional *Tet* deletion would need to be precisely triggered at E13.5-E14.5 and not before, to avoid interfering with TET function in PGCs from E9.5 to E13.5. Then the potential function of L1 expression could be assessed: use of a dead Cas9 fused to transcriptional repression factors (likely a KRAB domain) targeted to L1A promoters to force their repression in a stable way in prospermatogonia and assess the developmental impact on the male germline.

4 Germline development and spermatogonial stem cells

The study of stem cell biology is important for regenerative therapies. However, the SSC biology field has for a long time lagged behind other stem cell research fields, because of the lack of appropriate tools to distinguish and study them. Using recent state-of-the-art methods (scRNA-seq and Cut&Run) and tools (*Id4-GFP* transgenic mouse), we were able to generate new important insights into SSC biology. Furthermore, there are increasing efforts to classify SSCs according to a precise hierarchy, related to their plasticity and relative potential for self-renewal and differentiation. Several models have emerged, suggesting that their organization may not be unidirectional but rather involves continuous identity exchanges between SSC types. Using RNA velocity, we definitely and for the first time illustrated the plastic

potential of the SSC population, revealing the involvement of the different SSC types in a cycling trajectory. Moreover, the *Dnmt3A*^{KO} model proves that uncommitted SSC behaviour is essential for allowing them to differentiate and sustain spermatogenesis.

At a certain point, I tried to derive and culture *Dnmt3A*^{KO} SSCs *in vitro*, with the help of Pierre Fouchet and Clémentine Lapoujade (CEA). While WT SSCs were growing properly and forming colonies, *Dnmt3A*^{KO} SSCs failed to grow *in vitro*. The *in vitro* protocol for long-term SSC cultures is the following: postnatal germ cells are FACS-sorted using general germ cell-markers and plated on petri dishes. The theory is that, from all the germ cells present in the testis, only SSCs can proliferate on the petri dishes through their high self-renewing potential, while more committed germ cells die. The *Dnmt3A*^{KO} mice possess SSCs, in higher numbers than their WT littermates, and with a skewing towards self-renewing. We were expecting to see higher potential for *in vitro* SSC derivation in these mutants and were surprised to observe their inability to form colonies.

However, this result may not be that counterintuitive. In a normal situation, the most naïve SSCs are very rare, and are likely not the ones responsible for establishing SSC colonies *in vitro*. The colonies are more likely to derive from the more abundant transitory and late SSCs, which could be able to revert to a naïve state. This mechanism has been observed *in vivo*, in case of injury: a pool of SSC progenitors (NGN3 positive) can revert back, express the naïve marker ID4 and give rise to new spermatogenic waves (Zhang et al., 2016). The lack of plasticity that we observed in *Dnmt3A*^{KO} SSCs might explain the absence of SCC colonies in culture conditions.

Furthermore, the *Dnmt3A*^{KO} phenotype is unheard of: these mutants remarkably produce spermatozoa during the first wave of spermatogenesis, but not subsequently, due to loss of SSC plasticity. The first wave of spermatogenesis, which intriguingly occurs in an SSC-independent manner, has been difficult to study. The *Dnmt3A*^{KO} model provides the opportunity to specifically isolate the cellular and molecular details of SSC-independent spermatogenesis, as it is the only source of differentiating germ cells in this mutant.

Finally, germline conditional *Dnmt3A* mutants could provide a useful model to study testicular cancers. As a matter of fact, we showed that *Dnmt3A*^{KO} males suffer from a lack of SSC differentiation with high SSC proliferation, which is a characteristic

required for cancer development. I have carried out some aging experiments, but did not observe cancers or hyperplasia in the testis of 6 month-old or older *Dnmt3A*^{KO} males, in the C57BL/6J background. Changing the strain background may have a different outcome: FVB mice for example has a 50-60% chance to develop cancer compared to 1-7% in C57BL/6J (ref: The Jackson laboratory). This may be interesting to test whether backcrosses onto a FVB background may promote testicular cancers in *Dnmt3A*^{KO} males.

5 DNMT3A and stem cell biology

It is known that DNMT3A-dependent DNA methylation is particularly important in tissue stem cells. In particular, this was demonstrated in hematopoietic stem cells, neuronal stem cells, and epidermal stem cells (Challen et al., 2012; Rinaldi et al., 2016).

Interestingly, in spermatogenesis, lack of DNMT3A-dependent DNA methylation affects SSCs only and no other germ cell types: even in absence of DNA methylation genome wide, *Dnmt3A*^{KO} germ cells can complete the first wave of spermatogenesis (SSC-independent), and therefore pass through spermatogonia proliferation, meiosis, and spermatid maturation. This observation indicates that globally, DNA methylation is not required for the spermatogenesis program and the step-wise progress of male germ cell differentiation. However, the SSC phenotype was clear, and upon deletion of *Dnmt3A*, SSCs lost their ability to differentiate. This observation opens an interesting question: why SSCs and more generally stem cells are more sensitive to DNA methylation loss than non-stem cell types?

Our study revealed the emergence of *de novo* enhancers in *Dnmt3A*^{KO} SSCs, which have motifs for methyl-sensitive transcription factors (TFs). It has been shown that while the vast majority of TFs are DNA methylation insensitive, a small fraction of TFs have a greater affinity for unmethylated DNA (Domcke et al., 2015; Ginno et al., 2020). We could therefore hypothesize that stem cells are more sensitive to the loss of DNMT3A-dependent DNA methylation compared to other cell types because stem cell programs tend to be regulated by DNA methylation sensitive TFs preferentially.

REFERENCES

Aloisio, G.M., Nakada, Y., Saatcioglu, H.D., Peña, C.G., Baker, M.D., Tarnawa, E.D., Mukherjee, J., Manjunath, H., Bugde, A., Sengupta, A.L., et al. (2014). PAX7 expression defines germline stem cells in the adult testis. J. Clin. Invest. *124*, 3929–3944.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. Mol. Cell *31*, 785–799.

Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné, T., Feil, R., et al. (2016). EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. Genome Res. *26*, 192–202.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science (80-.). *354*, 909–912.

Barroca, V., Lassalle, B., Coureuil, M., Louis, J.P., Le Page, F., Testart, J., Allemand, I., Riou, L., and Fouchet, P. (2009). Mouse differentiating spermatogonia can generate germinal stem cells in vivo. Nat. Cell Biol. *11*, 190–196.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243–247.

Baudat, F., Manova, K., Yuen, J.P., Jasin, M., and Keeney, S. (2000). Chromosome Synapsis Defects and Sexually Dimorphic Meiotic Progression in Mice Lacking Spo11. Mol. Cell *6*, 989–998.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. *32*, 29–41.

Bibikova, M. (2016). DNA Methylation Microarrays (Elsevier Inc.).

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. *8*, 1499–1504.

Boulard, M., Edwards, J.R., and Bestor, T.H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. Nat. Genet. *47*, 479–485.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96–99.

Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science (80-.). *294*, 2536–2539.

Brick, K., Thibault-Sennett, S., Smagulova, F., Lam, K.-W.G., Pu, Y., Pratto, F., Camerini-Otero, R.D., and Petukhova, G. V. (2018). Extensive sex differences at the initiation of genetic recombination. Nature *561*, 338–342.

Brind'Amour, J., Kobayashi, H., Richard Albert, J., Shirane, K., Sakashita, A., Kamio, A., Bogutz, A., Koike, T., Karimi, M.M., Lefebvre, L., et al. (2018). LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. Nat. Commun. *9*, 3331.

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale

investigation of chromatin and DNA methylation cross-talk. Genome Res. *22*, 1128–1138.

Brinster, R.L., and Avarbock, M.R. (1994). Germline transmission of donor haplotype following spermatogonial transplantation. Proc. Natl. Acad. Sci. *91*, 11303–11307.

Brinster, R.L., and Zimmermann, J.W. (1994). Spermatogenesis following male germ-cell transplantation. Proc. Natl. Acad. Sci. *91*, 11298–11302.

Bulut-Karslioglu, A., De La Rosa-Velázquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, G.E., Engist, B., Gerle, B., et al. (2014). Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. Mol. Cell *55*, 277–290.

Carmell, M.A., Girard, A., van de Kant, H.J.G., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. Dev. Cell *12*, 503–514.

Carrieri, C., Comazzetto, S., Grover, A., Morgan, M., Buness, A., Nerlov, C., and O'Carroll, D. (2017). A transit-amplifying population underpins the efficient regenerative capacity of the testis. J. Exp. Med. *214*, 1631–1641.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat. Genet. *44*, 23–31.

Chan, F., Oatley, M.J., Kaucher, A. V, Yang, Q.-E., Bieberich, C.J., Shashikant, C.S., and Oatley, J.M. (2014). Functional and molecular features of the Id4+ germline stem cell population in mouse testes. Genes Dev. *28*, 1351–1362.

Chen, C., Ouyang, W., Grigura, V., Zhou, Q., Carnes, K., Lim, H., Zhao, G.-Q., Arber, S., Kurpios, N., Murphy, T.L., et al. (2005). ERM is required for transcriptional control of the spermatogonial stem cell niche. Nature *436*, 1030–1034.

Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A Novel Dnmt3a Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Activede Novo Methylation. J. Biol. Chem. *277*, 38746–38754.

Chen, Z., Chen, Z., Chen, Z., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., and Zhang, Y. (2020). Role of Mammalian DNA Methyltransferases in Development. Annu. Rev. Biochem. *89*, 135–158.

Chiarini-Garcia, H., Hornick, J.R., Griswold, M.D., and Russell, L.D. (2001). Distribution of Type A Spermatogonia in the Mouse Is Not Random1. Biol. Reprod. *65*, 1179–1185.

Chuma, S., and Nakano, T. (2013). piRNA and spermatogenesis in mice. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *368*, 20110338.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.

Clermont, Y. (1972). Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. Physiol. Rev. *52*, 198–236.

Clermont, Y., and Bustos-Obregon, E. (1968). Re-examination of spermatogonial renewal in the rat by means of seminiferous tubules mounted "in toto." Am. J. Anat. *122*, 237–247.

Dahlet, T., Argüeso Lleida, A., Al Adhami, H., Dumas, M., Bender, A., Ngondo, R.P., Tanguy, M., Vallet,

J., Auclair, G., Bardet, A.F., et al. (2020). Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. Nat. Commun. *11*, 3153.

Dai, H.-Q., Wang, B.-A., Yang, L., Chen, J.-J., Zhu, G.-C., Sun, M.-L., Ge, H., Wang, R., Chapman, D.L., Tang, F., et al. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. Nature *538*, 528–532.

Deng, W., and Lin, H. (2002). miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. Dev. Cell *2*, 819–830.

Deniz, Ö., Frost, J.M., and Branco, M.R. (2019). Regulation of transposable elements by DNA modifications. Nat. Rev. Genet.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. J. Biol. Chem. *285*, 26114–26120.

Didion, J.P., Martin, M., and Collins, F.S. (2017). Atropos: specific, sensitive, and speedy trimming of sequencing reads. PeerJ *5*, e3720.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature *528*, 575–579.

Dura, M., Teissandier, A., Armand, M., Barau, J., Bonneville, L., Weber, M., Baudrin, L.G., Lameiras, S., and Bourc'his, D. (2021). DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis. BioRxiv 1–60.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res. *10*, 2709–2721.

Enriquez-Gasca, R., Gould, P.A., and Rowe, H.M. (2020). Host gene regulation by transposable elements: The new, the old and the ugly. Viruses *12*.

Evsikov, A.V., de Vries, W.N., Peaston, A.E., Radford, E.E., Fancher, K.S., Chen, F.H., Blake, J.A., Bult, C.J., Latham, K.E., Solter, D., et al. (2004). Systems biology of the 2-cell mouse embryo. Cytogenet. Genome Res. *105*, 240–250.

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

Feng, H., Conneely, K.N., and Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic Acids Res. *42*, e69–e69.

Fu, K., Bonora, G., and Pellegrini, M. (2020). Interactions between core histone marks and DNA methyltransferases predict DNA methylation patterns observed in human cells and tissues. Epigenetics *15*, 272–282.

Fuks, F. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. *31*, 2305–2312.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. (2003). The Methyl-CpG-binding Protein MeCP2 Links DNA Methylation to Histone Methylation. J. Biol. Chem. *278*, 4035–4040.

Galan, C., Serra, R.W., Sun, F., Rinaldi, V.D., Conine, C.C., and Rando, O.J. (2021). Stability of the cytosine methylome during post-testicular sperm maturation in mouse. PLOS Genet. *17*, e1009416.

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2015). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics *32*, btv562.

Gelfman, S., Cohen, N., Yearim, A., and Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. Genome Res. *23*, 789–799.

Gewiss, R., Topping, T., and Griswold, M.D. (2020). Cycles, waves, and pulses: Retinoic acid and the organization of spermatogenesis. Andrology *8*, 892–897.

Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliauskaite, L., Benes, V., Enright, A.J., and O'Carroll, D. (2013). Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis. Mol. Cell *50*, 601–608.

Ginno, P.A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., Zilbermann, F., Peters, A.H.F.M., Edenhofer, F., Smallwood, S.A., et al. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. Nat. Commun. *11*.

Givelet, M., Lapoujade, C., and Fouchet, P. (2019). Competition for Food Drives Stem Cell Fate in Facultative Niches (Preview). Cell Stem Cell *24*, 1–2.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic Cytosine Methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. Dev. Cell *46*, 651-667.e10.

Greenbaum, M.P., Yan, W., Wu, M.-H., Lin, Y.-N., Agno, J.E., Sharma, M., Braun, R.E., Rajkovic, A., and Matzuk, M.M. (2006). TEX14 is essential for intercellular bridges and fertility in male mice. Proc. Natl. Acad. Sci. *103*, 4982–4987.

Greenbaum, M.P., Iwamori, T., Buchold, G.M., and Matzuk, M.M. (2011). Germ Cell Intercellular Bridges. Cold Spring Harb. Perspect. Biol. *3*, a005850–a005850.

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

Grisanti, L., Falciatori, I., Grasso, M., Dovere, L., Fera, S., Muciaccia, B., Fuso, A., Berno, V., Boitani, C., Stefanini, M., et al. (2009). Identification of Spermatogonial Stem Cell Subsets by Morphological Analysis and Prospective Isolation. Stem Cells *27*, N/A-N/A.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet. *6*.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res. *22*, 633–641.

Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., et al. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. Cell Stem Cell *21*,

533-546.e6.

Guo, J., Grow, E.J., Mlcochova, H., Maher, G.J., Lindskog, C., Nie, X., Guo, Y., Takei, Y., Yun, J., Cai, L., et al. (2018). The adult human testis transcriptional cell atlas. Cell Res. *28*, 1141–1157.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature *517*, 640–644.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. Science (80-.). *339*, 448–452.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech. Dev. *117*, 15–23.

Hancks, D.C., and Kazazian, H.H. (2016). Roles for retrotransposon insertions in human disease. Mob. DNA *7*.

Hara, K., Nakagawa, T., Enomoto, H., Suzuki, M., Yamamoto, M., Simons, B.D., and Yoshida, S. (2014). Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. Cell Stem Cell *14*, 658–672.

Hargan-Calvopina, J., Taylor, S., Cook, H., Hu, Z., Lee, S.A., Yen, M.R., Chiang, Y.S., Chen, P.Y., and Clark, A.T. (2016). Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. Dev. Cell *39*, 75–86.

Harikumar, A., and Meshorer, E. (2015). Chromatin remodeling and bivalent histone modifications in embryonic stem cells. EMBO Rep. *16*, 1609–1619.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012a). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

Hashimoto, H., Hong, S., Bhagwat, A.S., Zhang, X., and Cheng, X. (2012b). Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. Nucleic Acids Res. *40*, 10203–10214.

Hayashi, Y., and Kobayashi, S. (2018). Regulatory Mechanisms of the Germline Stem Cell Niche in Drosophila melanogaster. pp. 19–35.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science (80-.). *333*, 1303–1307.

Helsel, A.R., Yang, Q.-E., Oatley, M.J., Lord, T., Sablitzky, F., and Oatley, J.M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. Development *144*, 624–634.

Hermann, B.P., Mutoji, K.N., Velte, E.K., Ko, D., Oatley, J.M., Geyer, C.B., and McCarrey, J.R. (2015). Transcriptional and Translational Heterogeneity among Neonatal Mouse Spermatogonia1. Biol. Reprod. *92*, 1–12.

Hermann, B.P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K.N., Chen, I.-C., Gildersleeve, H., Lehle, J.D., Mayo, M., Westernströer, B., et al. (2018). The mammalian spermatogenesis single-Cell transcriptome, from spermatogonial stem cells to spermatids. Cell Rep. *25*, 1650-1667.e8.

Hermant, C., and Torres-Padilla, M.E. (2021). TFs for TEs: The transcription factor repertoire of mammalian transposable elements. Genes Dev. *35*, 22–39.

Hess, R.A., and de Franca, L.R. (2009). Spermatogenesis and Cycle of the Seminiferous Epithelium. In Advances in Experimental Medicine and Biology, pp. 1–15.

Hill, P.W.S., Amouroux, R., and Hajkova, P. (2014). DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: An emerging complex story. Genomics *104*, 324–333.

Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. Nature *555*, 392–396.

Huckins, C. (1971). The spermatogonial stem cell population in adult rats. I. Their morphology, proliferation and maturation. Anat. Rec. *169*, 533–557.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genet. *6*, e1001134.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science (80-.). *333*, 1300–1303.

Itou, D., Shiromoto, Y., Shin-ya, Y., Ishii, C., Nishimura, T., Ogonuki, N., Ogura, A., Hasuwa, H., Fujihara, Y., Kuramochi-Miyagawa, S., et al. (2015). Induction of DNA methylation by artificial piRNA production in male germ cells. Curr. Biol. *25*, 901–906.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556–560.

Jain, D., Meydan, C., Lange, J., Bouuaert, C.C., Mason, C.E., Anderson, K. V., and Keeney, S. (2017). Rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. BioRxiv.

Jan, S.Z., Hamer, G., Repping, S., de Rooij, D.G., van Pelt, A.M.M., and Vormer, T.L. (2012). Molecular control of rodent spermatogenesis. Biochim. Biophys. Acta - Mol. Basis Dis. *1822*, 1838–1850.

Jang, S.M., Kauzlaric, A., Quivy, J.-P., Pontis, J., Rauwel, B., Coluccio, A., Offner, S., Duc, J., Turelli, P., Almouzni, G., et al. (2018). KAP1 facilitates reinstatement of heterochromatin after DNA replication. Nucleic Acids Res. *46*, 8788–8802.

Jangam, D., Feschotte, C., and Betrán, E. (2017). Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends Genet. *33*, 817–831.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248–251.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900–903.

Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grandér, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C., and Ekström, T.J. (2006). LUMA (LUminometric Methylation Assay)—A high throughput method to the analysis of genomic DNA methylation. Exp. Cell Res. *312*, 1989–1995.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. Cell Stem Cell *8*, 676–687.

Karimzadeh, M., Ernst, C., Kundaje, A., and Hoffman, M.M. (2018). Umap and Bismap: quantifying genome and methylome mappability. Nucleic Acids Res. *46*, 1–13.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum. Mol. Genet. *16*, 2272–2280.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat. Commun. *10*, 1–10.

Kitadate, Y., Jo, D.J., Tokue, M., Takahashi, S., Simons, B.D., Yoshida, S., Tokue, M., Maruyama, A., Ichikawa, R., Tsuchiya, S., et al. (2019). Competition for Mitogens Regulates Spermatogenic Stem Cell Homeostasis in an Open Niche. Cell Stem Cell 79–92.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T., et al. (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res. *23*, 616–627.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. Nature *351*, 117–121.

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

Kubo, N., Toh, H., Shirane, K., Shirakawa, T., Kobayashi, H., Sato, T., Sone, H., Sato, Y., Tomizawa, S., Tsurusaki, Y., et al. (2015). DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. BMC Genomics *16*, 624.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development *131*, 839–849.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. *22*, 1617–1635.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., et al. (2001). Correction: Initial sequencing and analysis of the human genome. Nature *412*, 565–566.

Law, N.C., and Oatley, J.M. (2020). Developmental underpinnings of spermatogonial stem cell establishment. Andrology *8*, 852–861.

Law, N.C., Oatley, M.J., and Oatley, J.M. (2019). Developmental kinetics and transcriptome dynamics of stem cell specification in the spermatogenic lineage. Nat. Commun. *10*, 2787.

Leblond, C.P., and Clermont, Y. (1952). DEFINITION OF THE STAGES OF THE CYCLE OF THE SEMINIFEROUS EPITHELIUM IN THE RAT. Ann. N. Y. Acad. Sci. *55*, 548–573.

Lee, H.J., Hore, T.A., and Reik, W. (2014). Reprogramming the Methylome: Erasing Memory and

Creating Diversity. Cell Stem Cell *14*, 710–719.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. Curr. Biol. *13*, 1192–1200.

Lengner, C.J., Camargo, F.D., Hochedlinger, K., Welstead, G.G., Zaidi, S., Gokhale, S., Scholer, H.R., Tomilin, A., and Jaenisch, R. (2007). Oct4 expression is not required for mouse somatic stem cell self-renewal. Cell Stem Cell *1*, 403–415.

Lewandoski, M., Wassarman, K.M., and Martin, G.R. (1997). Zp3–cre, a transgenic mouse line for the activation or inactivation of loxP-flanked target genes specifically in the female germ line. Curr. Biol. *7*, 148–151.

Li, D.-Q., Nair, S.S., and Kumar, R. (2013). The MORC family. Epigenetics *8*, 685–693.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, H., Rauch, T., Chen, Z.X., Szabó, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J. Biol. Chem. *281*, 19489–19500.

Li, Y., Chen, X., and Lu, C. (2021). The interplay between DNA and histone methylation: molecular mechanisms and disease implications. EMBO Rep. *22*, 1–21.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev. *28*, 2041–2055.

Lord, T., and Oatley, J.M. (2017). A revised Asingle model to explain stem cell dynamics in the mouse male germline. Reproduction *131*, 1796–1803.

Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: In the right place at the right time. Science (80-.). *361*, 1336–1340.

Manzo, M., Wirz, J., Ambrosi, C., Villaseñor, R., Roschitzki, B., and Baubec, T. (2017). Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. EMBO J. e201797038.

Marasca, F., Bodega, B., and Orlando, V. (2018). How Polycomb-Mediated Cell Memory Deals With a Changing Environment. BioEssays *40*, 1700137.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10.

Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., et al. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. Mol. Cell *60*, 584–596.

McLay, D.W., and Clarke, H.J. (2003). Remodelling the paternal chromatin at fertilization in mammals.

Reproduction *125*, 625–633.
Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.
Meng, X. (2000). Regulation of Cell Fate Decision of Undifferentiated Spermatogonia by GDNF. Science (80-.). *287*, 1489–1493.
Mikami, S., Kanaba, T., Takizawa, N., Kobayashi, A., Maesaki, R., Fujiwara, T., Ito, Y., and Mishima, M. (2014). Structural Insights into the Recruitment of SMRT by the Corepressor SHARP under Phosphorylative Regulation. Structure *22*, 35–46.
Miura, F., Shibata, Y., Miura, M., Sangatsuda, Y., Hisano, O., Araki, H., and Ito, T. (2019). Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res. *47*, e85–e85.
Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. Mol. Cell *30*, 755–766.
Molaro, A., Falciatori, I., Hodges, E., Aravin, A.A., Marran, K., Rafii, S., Richard McCombie, W., Smith, A.D., Hannon, G.J., Molaro, A., et al. (2014). Two waves of de novo methylation during mouse germ cell development Two waves of de novo methylation during mouse germ cell development. Genes Dev. *28*, 1544–1549.
Molaro, A., Malik, H.S., Bourc'his, D., and Satta, Y. (2020). Dynamic Evolution of de Novo DNA Methyltransferases in Rodent and Primate Genomes. Mol. Biol. Evol. *37*, 1882–1892.
Morrison, S.J., and Spradling, A.C. (2008). Stem Cells and Niches: Mechanisms That Promote Stem Cell Maintenance throughout Life. Cell *132*, 598–611.
Nakagawa, T., Nabeshima, Y., and Yoshida, S. (2007). Functional Identification of the Actual and Potential Stem Cell Compartments in Mouse Spermatogenesis. Dev. Cell *12*, 195–206.
Nakagawa, T., Sharma, M., Nabeshima, Y.I., Braun, R.E., and Yoshida, S. (2010). Functional Hierarchy and Reversibility Within the Murine Spermatogenic Stem Cell Compartment. Science (80-.). *328*, 62–67.
Nakata, H., Wakayama, T., Takai, Y., and Iseki, S. (2015). Quantitative Analysis of the Cellular Composition in Seminiferous Tubules in Normal and Genetically Modified Infertile Mice. J. Histochem. Cytochem. *63*, 99–113.
Oatley, J.M., Avarbock, M.R., Telaranta, A.I., Fearon, D.T., and Brinster, R.L. (2006). Identifying genes important for spermatogonial stem cell self-renewal and survival. Proc. Natl. Acad. Sci. *103*, 9524–9529.
Oatley, M.J., Kaucher, A. V., Racicot, K.E., and Oatley, J.M. (2011). Inhibitor of DNA Binding 4 Is Expressed Selectively by Single Spermatogonia in the Male Germline and Regulates the Self-Renewal of Spermatogonial Stem Cells in Mice1. Biol. Reprod. *85*, 347–356.
Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207–213.
Ohmura, M., Yoshida, S., Ide, Y., Nagamatsu, G., Suda, T., and Ohbo, K. (2004). Spatial analysis of

germ stem cell development in Oct-4/EGFP transgenic mice. Arch. Histol. Cytol. *67*, 285–296.
Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell *99*, 247–257.
Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.
Orth, J.M., Qiu, J., Jester, W.F., and Pilder, S. (1997). Expression of the c-kit Gene is Critical for Migration of Neonatal Rat Gonocytes in Vitro1. Biol. Reprod. *57*, 676–683.
Orwig, K.E., Ryu, B.-Y., Avarbock, M.R., and Brinster, R.L. (2002). Male germ-line stem cell potential is predicted by morphology of cells in neonatal rat testes. Proc. Natl. Acad. Sci. *99*, 11706–11711.
Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2019). PIWI-interacting RNAs: small RNAs with big functions. Nat. Rev. Genet. *20*, 89–108.
Parhad, S.S., and Theurkauf, W.E. (2019). Rapid evolution and conserved function of the piRNA pathway. Open Biol. *9*.
Parhad, S.S., Tu, S., Weng, Z., and Theurkauf, W.E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev. Cell *43*, 60-70.e5.
Pastor, W.A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S.A., Moissiard, G., Zamudio, N., Bourc'his, D., et al. (2014). MORC1 represses transposable elements in the mouse male germline. Nat. Commun. *5*, 5795.
Pezic, D., Manakov, S.A., Sachidanandam, R., and Aravin, A.A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes Dev. *28*, 1410–1428.
Rajavelu, A., Lungu, C., Emperle, M., Dukatz, M., Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., et al. (2018). Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. Nucleic Acids Res. *46*, 9044–9056.
Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.
Rebollo, R., Galvão-Ferrarini, M., Gagnier, L., Zhang, Y., Ferraj, A., Beck, C.R., Lorincz, M.C., and Mager, D.L. (2020). Inter-Strain Epigenomic Profiling Reveals a Candidate IAP Master Copy in C3H Mice. Viruses *12*, 783.
Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H., et al. (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol. *14*.
Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., and Pillai, R.S. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. Nature *480*, 264–267.
Rinaldi, L., Datta, D., Serrat, J., Morey, L., Solanas, G., Avgustinova, A., Blanco, E., Pons, J.I., Matallanas, D., Von Kriegsheim, A., et al. (2016). Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. Cell Stem Cell *19*, 491–501.
Rodriguez-Terrones, D., and Torres-Padilla, M.E. (2018). Nimble and Ready to Mingle: Transposon

Outbursts of Early Development. Trends Genet. *34*, 806–820.
Rojas-Rıós, P., and Simonelig, M. (2018). piRNAs and PIWI proteins: Regulators of gene expression in development and stem cells. Dev. *145*.
de Rooij, D.G. (2017). The nature and dynamics of spermatogonial stem cells. Development *144*, 3022–3030.
De Rooij, D.G. (2009). The spermatogonial stem cell niche. Microsc. Res. Tech. *72*, 580–585.
Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.
Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc. Natl. Acad. Sci. U. S. A. *107*, 10783–10790.
Sachs, P., Ding, D., Bergmaier, P., Lamp, B., Schlagheck, C., Finkernagel, F., Nist, A., Stiewe, T., and Mermoud, J.E. (2019). SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. Nat. Commun. *10*, 1335.
Sada, A., Suzuki, A., Suzuki, H., and Saga, Y. (2009). The RNA-Binding Protein NANOS2 Is Required to Maintain Murine Spermatogonial Stem Cells. Science (80-.). *325*, 1394–1398.
Saitou, M., and Yamaji, M. (2012). Primordial Germ Cells in Mice. Cold Spring Harb. Perspect. Biol. *4*, a008375–a008375.
Saitou, M., Kagiwada, S., and Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development *139*, 15–31.
Sakai, Y., Suetake, I., Shinozaki, F., Yamashina, S., and Tajima, S. (2004). Co-expression of de novo DNA methyltransferases Dnmt3a2 and Dnmt3L in gonocytes of mouse embryos. Gene Expr. Patterns *5*, 231–237.
Schomacher, L., Han, D., Musheev, M.U., Arab, K., Kienhöfer, S., von Seggern, A., and Niehrs, C. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. Nat. Struct. Mol. Biol. *23*, 116–124.
Schöpp, T., Zoch, A., Berrens, R. V., Auchynnikava, T., Kabayama, Y., Vasiliauskaitė, L., Rappsilber, J., Allshire, R.C., and O'Carroll, D. (2020). TEX15 is an essential executor of MIWI2-directed transposon DNA methylation and silencing. Nat. Commun. *11*.
Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. Mol. Cell *48*, 849–862.
Sendžikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., and Kelsey, G. (2019). A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nat. Commun. *10*, 1884.
Seo, E.K., Choi, Y., Jeong, J.H., Kim, Y.G., and Park, H.H. (2016). Crystal structure of C-terminal coiled-coil domain of SYCP1 reveals non-canonical anti-parallel dimeric structure of transverse filament at the synaptonemal complex. PLoS One *11*, 1–13.
Shayevitch, R., Askayo, D., Keydar, I., and Ast, G. (2018). The importance of DNA methylation of exons

on alternative splicing. RNA *24*, 1351–1362.
Shirane, K., Miura, F., Ito, T., and Lorincz, M.C. (2020). NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. Nat. Genet. *52*, 1088–1098.
Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. Cell *151*, 964–980.
Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*, 1–35.
Song, H.-W., and Wilkinson, M.F. (2014). Transcriptional control of spermatogonial maintenance and differentiation. Semin. Cell Dev. Biol. *30*, 14–26.
Song, Q., Decato, B., Hong, E.E., Zhou, M., Fang, F., Qu, J., Garvin, T., Kessler, M., Zhou, J., and Smith, A.D. (2013). A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS One *8*.
Song, Y., van den Berg, P.R., Markoulaki, S., Soldner, F., Dall'Agnese, A., Henninger, J.E., Drotar, J., Rosenau, N., Cohen, M.A., Young, R.A., et al. (2019). Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs. Mol. Cell *75*, 905-920.e6.
Sookdeo, A., Hepp, C.M., McClure, M.A., and Boissinot, S. (2013). Revisiting the evolution of mouse LINE-1 in the genomic era. Mob. DNA *4*, 3.
Statham, A.L., Robinson, M.D., Song, J.Z., Coolen, M.W., Stirzaker, C., and Clark, S.J. (2012). Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res. *22*, 1120–1127.
Stelzer, Y., Shivalila, C.S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution. Cell *163*, 218–229.
Stocking, C., and Kozak, C.A. (2008). Murine endogenous retroviruses. Cell. Mol. Life Sci. *65*, 3383–3398.
Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J. Biol. Chem. *279*, 27816–27823.
Tagelenbosch, R.A.J., and de Rooij, D.G. (1993). A quantitative study of spermatogonial multiplication and stem cell renewal in the C3H/101 F1 hybrid mouse. Mutat. Res. Mol. Mech. Mutagen. *290*, 193–200.
Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., and Li, E. (2006). Roles for Dnmt3b in mammalian development: A mouse model for the ICF syndrome. Development *133*, 1183–1192.
Wang, S.H., and Elgin, S.C.R. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc. Natl. Acad. Sci. *108*, 21164–21169.
Watanabe, T., Tomizawa, S. -i., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A., et al. (2011). Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus. Science (80-.). *332*, 848–852.
Watanabe, T., Cui, X., Yuan, Z., Qi, H., and Lin, H. (2018). MIWI2 targets RNAs transcribed from piRNA-

dependent regions to drive DNA methylation in mouse prospermatogonia. EMBO J. e95329.
Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature *573*, 281–286.
Weinberg, D.N., Rosenbaum, P., Chen, X., Barrows, D., Horth, C., Marunde, M.R., Popova, I.K., Gillespie, Z.B., Keogh, M.-C., Lu, C., et al. (2021). Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. Nat. Genet. *53*, 794–800.
Whittle, C.A., and Extavour, C.G. (2017). Causes and evolutionary consequences of primordial germ-cell specification mode in metazoans. Proc. Natl. Acad. Sci. *114*, 5784–5791.
Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. *18*, 517–534.
Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science (80-.). *329*, 444–448.
Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443–447.
Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. Nature *504*, 460–464.
Yaman, R., and Grandjean, V. (2006). Timing of entry of meiosis depends on a mark generated by DNA methyltransferase 3a in testis. Mol. Reprod. Dev. *73*, 390–397.
Yamanaka, S., Nishihara, H., Toh, H., Eijy Nagai, L.A., Hashimoto, K., Park, S.-J., Shibuya, A., Suzuki, A.M., Tanaka, Y., Nakai, K., et al. (2019). Broad Heterochromatic Domains Open in Gonocyte Development Prior to De Novo DNA Methylation. Dev. Cell 1–14.
Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. *13*, 335–340.
Yokomine, T., Hata, K., Tsudzuki, M., and Sasaki, H. (2006). Evolution of the vertebrate DNMT3 gene family: a possible link between existence of *DNMT3L* and genomic imprinting. Cytogenet. Genome Res. *113*, 75–80.
Yoshida, S., Takakura, A., Ohbo, K., Abe, K., Wakabayashi, J., Yamamoto, M., Suda, T., and Nabeshima, Y. (2004). Neurogenin3 delineates the earliest stages of spermatogenesis in the mouse testis. Dev. Biol. *269*, 447–458.
Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T., and Nabeshima, Y. (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development *133*, 1495–1505.
Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., and Bourc'his, D. (2015). DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev. *29*, 1256–1270.
Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.-L., Luo, C., Jiang, H., and He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat. Chem. Biol. *8*, 328–

330.

Zhang, T., Oatley, J., Bardwell, V.J., and Zarkower, D. (2016). DMRT1 Is Required for Mouse Spermatogonial Stem Cell Maintenance and Replenishment. PLOS Genet. *12*, e1006293.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zoch, A., Auchynnikava, T., Berrens, R. V., Kabayama, Y., Schöpp, T., Heep, M., Vasiliauskaitė, L., Pérez-Rico, Y.A., Cook, A.G., Shkumatava, A., et al. (2020). SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. Nature *584*, 635–639.

ABSTRACT

DNA methylation, in association with stable gene or transposable element (TE) repression, plays a key role in spermatogenesis. During germline development, the methylome of the future gametes is extensively reprogrammed: somatic DNA methylation patterns are first erased and germ cell-specific patterns are then established *de novo*. Three *de novo* DNA methyltransferases (DNMTs) are essential for shaping male germ cell DNA methylation in mice: the DNMT3C and DNMT3A enzymes and the DNMT3L co-factor. Mutation in any of these genes leads to male sterility. DNMT3C was recently shown to selectively methylate evolutionarily young TEs. However, the precise targets and developmental function of DNMT3A was still unknown. During my PhD, I investigated the interplay between DNMT3A and DNMT3C in the epigenetic regulation of male germline development. First (***project 1***), I reported a striking division of labor between these enzymes: while DNMT3C prevents TEs from interfering with meiosis, DNMT3A broadly methylates the genome—except DNMT3C-dependent TEs—and controls spermatogonial stem cell (SSC) plasticity. By reconstructing developmental trajectories through single-cell RNA-seq and by profiling chromatin states, I found that *Dnmt3A* mutant SSCs can only self-renew and no longer differentiate due to spurious enhancer activation that enforces an irreversible stem cell gene program. I therefore demonstrated a novel function for DNA methylation in male fertility: the epigenetic programming of SSC commitment to differentiation and to life-long spermatogenesis supply. Second (***project 2***), I investigated the chromatin determinants of DNMT3C specificity towards young TEs. I found that these sequences present unique chromatin dynamics: first a bivalent H3K4me3-H3K9me3 enrichment, followed by a switch to H3K9me3-only. Consistently, H3K9me3-enrichment was also a hallmark of the sequences that undergo DNA methylation upon ectopic DNMT3C expression in cultured embryonic stem cells. As a whole, my work provided novel insights into the complexity of DNA methylation-based control of reproduction.

RÉSUMÉ

La méthylation de l'ADN, associée à la répression stable des gènes et des éléments transposables (ET), joue un rôle essentiel dans la spermatogenèse. Au cours du développement de la lignée germinale, le méthylome des futurs gamètes est extensivement reprogrammé : après effacement des profils de méthylation somatiques, des profils spécifiques des cellules germinales sont établis *de novo*. Trois *de novo* ADN méthyltransférases (DNMT) sont essentielles à l'acquisition de la méthylation de l'ADN des cellules germinales mâles chez la souris : les enzymes DNMT3C et DNMT3A et leur cofacteur DNMT3L. Toute mutation dans l'un de ces gènes entraîne une stérilité mâle. Il a été récemment démontré que DNMT3C est l'enzyme qui méthyle sélectivement les ET les plus jeunes évolutivement. Cependant, les cibles et la fonction développementale de DNMT3A étaient encore inconnues. Au cours de ma thèse, je me suis intéressée aux rôles respectifs et complémentaires de DNMT3A et DNMT3C dans la régulation épigénétique du développement germinal mâle. J'ai d'abord démontré (***projet 1***) une division de travail remarquable entre ces enzymes: alors que DNMT3C empêche les ET d'interférer avec la méiose, DNMT3A méthyle largement le génome -à l'exception des ET dépendants de DNMT3C- et contrôle la plasticité des cellules souches spermatogoniales (CSS). Par une reconstruction de trajectoire développementale par scRNA-seq, j'ai découvert que les CSS mutantes pour *Dnmt3A* ne peuvent que se renouveler à l'identique, ayant perdu leur potentiel de différentiation. Par une analyse des profils chromatiniens, j'ai relié ce phénotype à l'activation erronée d'enhancers qui imposent un programme génétique irréversible de cellules souches. Ce travail révèle une nouvelle fonction de la méthylation de l'ADN dans la fertilité mâle : la programmation épigénétique de la capacité des CSS à se différencier et à alimenter la spermatogenèse tout au long de la vie. En parallèle (***projet 2***), j'ai étudié la nature chromatinienne de la spécificité de reconnaissance des jeunes ET par DNMT3C. J'ai trouvé que ces séquences présentent une dynamique chromatinienne unique: d'abord un profil bivalent de type H3K4me3-H3K9me3 qui évolue vers un enrichissement H3K9me3 exclusif. Mon travail a ainsi fourni des éléments originaux et nouveaux pour comprendre le rôle complexe de la méthylation de l'ADN en reproduction.