



HAL
open science

Dynamique transcriptionnelle et remodelage chromatinien lors de la spermatogenèse : rôle de la marque H3K79me2 et de sa méthyltransférase DOT1L

Manon Coulée

► **To cite this version:**

Manon Coulée. Dynamique transcriptionnelle et remodelage chromatinien lors de la spermatogenèse : rôle de la marque H3K79me2 et de sa méthyltransférase DOT1L. Génétique. Université Paris Cité, 2023. Français. NNT : 2023UNIP5143 . tel-04888617

HAL Id: tel-04888617

<https://theses.hal.science/tel-04888617v1>

Submitted on 15 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université
Paris Cité



Université Paris Cité

École doctorale Bio Sorbonne Paris Cité (ED 562)

Laboratoire Institut Cochin (INSERM U1016/CNRS UMR8104)

Équipe : Des Gamètes à la Naissance

Dynamique transcriptionnelle et remodelage chromatinien lors de la spermatogenèse : rôle de la marque H3K79me2 et de sa méthyltransférase DOT1L

Par Manon COULÉE

Thèse de doctorat de Reproduction

Dirigée par Julie COCQUET

et co-encadrée par Laila EL KHATTABI

Présentée et soutenue publiquement le 20/10/2023

Rapportrices : - **Amélie BONNET-GARNIER**, Chargée de Recherche
Université Paris-Saclay – INRAE de Jouy-en-Josas
- **Patricia FAUQUE**, Professeur d'Université
Université de Bourgogne – CHU de Dijon

Examineurs : - **Frédéric CHALMEL**, Directeur de Recherche
Université de Rennes – IRSET
- **Pierre FOUCHET**, Directeur de Recherche
Université Paris-Saclay – CEA de Fontenay-aux-Roses

Directrice de thèse : **Julie COCQUET**, Chargée de Recherche
Université Paris Cité – Institut Cochin

Co-encadrante et membre invitée :

Laila EL KHATTABI, Maître de Conférence Universitaire
Sorbonne Université – Institut Cochin

Résumé

Dynamique transcriptionnelle et remodelage chromatinien lors de la spermatogenèse : rôle de la marque H3K79me2 et de sa méthyltransférase DOT1L

La spermatogenèse est le processus de différenciation des cellules germinales mâles en spermatozoïdes. Elle est divisée en trois phases – la prolifération des spermatogonies, la méiose et la spermiogenèse – associées à une très grande dynamique transcriptionnelle et chromatinienne. Les études de transcriptomiques (RNA-seq) et épigénomiques (ChIP-seq) de ces processus ont permis de mettre en évidence un enrichissement de la diméthylation de la lysine 79 de l’histone H3 (marque épigénétique appelée H3K79me2) et une forte expression de sa seule méthyltransférase DOT1L aux stades méiotiques et post-méiotiques. Notre équipe a récemment produit un modèle de souris knock-out (KO) de *Dot1l* dans les cellules germinales mâles et montré qu’il entraîne une hypofertilité masculine, caractérisée par des malformations des spermatozoïdes ainsi que des défauts de compaction de la chromatine.

L’objectif de ma thèse a été de caractériser, par des analyses bioinformatiques, le rôle moléculaire de DOT1L et de corrélérer la dynamique de la marque H3K79me2 avec les changements transcriptionnels et le remodelage chromatinien au cours de la spermatogenèse. Pour cela, j’ai analysé des données de ChIP-seq, RNA-seq et ATAC-seq, produites par l’équipe et issues de la littérature.

Pour caractériser la dynamique de H3K79me2 au cours de la spermatogenèse, j’ai analysé des données de ChIP-seq dans 4 stades de développement des cellules germinales mâles (des spermatogonies indifférenciées aux spermatides post-méiotiques et allongées). Un pipeline d’analyse a été développé en intégrant les outils bowtie2 pour l’alignement, MACS2 pour l’enrichissement de la marque suivi de sa caractérisation par ChIPseeker et ChromHMM. ChromHMM permet de prédire les états chromatinien des cellules germinales mâles à partir de données provenant de 6 marques d’histones différentes. Cette approche m’a conduit à développer un package R (ChromENVEE) permettant d’associer les gènes à leur région régulatrice à distance (*enhancer*) la plus probable. Mes analyses montrent que H3K79me2 est corrélée positivement avec la transcription, et enrichie au niveau du promoteur et du corps du gène, mais aussi au niveau des *enhancers*. J’ai également trouvé que H3K79me2 est impliquée dans la régulation des gènes portés par les chromosomes sexuels lors de la transition de la méiose vers les spermatides.

Parallèlement, j’ai étudié l’impact du KO-*Dot1l* sur la transcription à différents stades de la spermatogenèse par des analyses de RNA-seq. J’ai développé un pipeline d’analyse en utilisant l’outil STAR pour l’alignement suivi par des analyses d’expression différentielle utilisant DESeq2 et edgeR. Ces analyses ont également été réalisées en utilisant une normalisation externe par « *spike-in* », c’est-à-dire via des séquences exogènes ajoutées aux échantillons dans une concentration connue. En intégrant nos données de RNA-seq et de ChIP-seq, nous avons montré que le KO-*Dot1l* conduit à la dérégulation de centaines de gènes. Nous avons trouvé que la majorité des gènes sous-exprimés dans les cellules KO correspondent à des gènes cibles « directs », c’est-à-dire enrichis en H3K79me2 au niveau du corps du gène et/ou de l’*enhancer*. Contre-intuitivement, nous avons trouvé beaucoup plus de gènes sur-exprimés que sous-exprimés dans les cellules KO. Ce phénomène semble indépendant de H3K79me2, et ces gènes sont principalement situés dans des états chromatinien caractérisés par des marques bivalentes et répressives, en particulier la marque répressive H3K27me3.

En conclusion, nos résultats montrent que DOT1L et H3K79me2 sont indispensables à la régulation des gènes sur l'ensemble de la spermatogenèse. En plus de son rôle d'activateur, ces résultats présentent également un nouveau rôle de répresseur de la transcription pour DOT1L en fonction de l'environnement chromatinien.

Mots clés : RNA-seq, ChIP-seq, spermatogenèse, H3K79me2, DOT1L, états chromatinien

Abstract

Transcriptional dynamic and chromatin remodeling during spermatogenesis: role of the H3K79me2 mark and its methyltransferase DOT1L

Spermatogenesis is the process by which male germ cells differentiate into spermatozoa. It is divided into three phases – spermatogonial proliferation, meiosis and spermiogenesis – associated with elevated transcriptional and chromatin dynamics. Transcriptomic and ChIP-Seq studies of these processes have revealed an enrichment in dimethylation of histone H3 lysine 79 (an epigenetic mark known as H3K79me2) and strong expression of its sole methyltransferase DOT1L at meiotic and post-meiotic stages. Our team has recently produced a *Dot1l* knock-out (KO) mouse model in male germ cells and shown that it leads to male hypofertility, characterized by sperm malformations and chromatin compaction defects.

The aim of my thesis was to characterize, through bioinformatic analyses, the molecular role of DOT1L and to correlate the dynamics of H3K79me2 mark with transcriptional changes and chromatin remodeling during spermatogenesis. To this end, I analyzed ChIP-seq, RNA-seq and ATAC-seq data produced by the team and from the literature.

To characterize the dynamics of H3K79me2 during spermatogenesis, I analyzed ChIP-seq data in 4 stages of male germ cell development (from undifferentiated spermatogonia to post-meiotic, elongated spermatids). A pipeline was developed using bowtie2 for alignment, MACS2 for mark enrichment and followed by characterization using ChIPseeker and ChromHMM. ChromHMM predicts male germ cell chromatin states from data derived from 6 different histone marks. This approach led me to develop an R package (ChromENVEE) to associate genes with their most likely distal regulatory region (enhancer). My analyses show that H3K79me2 is positively correlated with transcription, and enriched at promoters and gene bodies, but also at enhancers. I also found that H3K79me2 is involved in the regulation of sex chromosomes-encoded genes during the transition from meiosis to spermatids.

In parallel, I studied the impact of *Dot1l*-KO on transcription at different stages of spermatogenesis using RNA-seq analyses. I developed a pipeline using STAR tool for alignment followed by differential expression analyses using DESeq2 and edgeR. These analyses were also carried out using external normalization by "spike-in", i.e. via exogenous sequences added to the samples in a known concentration. By integrating our RNA-seq and ChIP-seq data, we showed that *Dot1l*-KO leads to the deregulation of hundreds of genes. We found that the majority of downregulated genes in KO cells correspond to "direct" target genes, i.e. enriched in H3K79me2 at the gene body and/or enhancer. Counter-intuitively, we found many more genes upregulated than downregulated in KO cells. This phenomenon appears to be independent of H3K79me2, and these genes are mainly located in chromatin states characterized by bivalent and repressive marks, in particular the repressive H3K27me3 mark.

In conclusion, these results show that DOT1L and H3K79me2 are indispensable for gene regulation throughout spermatogenesis. In addition to its activator role, these results also present a novel transcription repressor role for DOT1L according to the chromatin environment.

Keywords: RNA-seq, ChIP-seq, spermatogenesis, H3K79me2, DOT1L, chromatin states

Remerciements

Dans un premier temps, je souhaite remercier les membres du jury, Amélie Bonnet-Garnier, Patricia Fauque, Pierre Fouchet et Frédéric Chalmel, qui ont accepté d'évaluer mon travail.

Je souhaite également remercier Julie et Laila qui ont supervisé mon travail pendant ces trois années et précédemment durant mon stage. Merci pour votre patience, vos conseils et pour m'avoir poussé à donner le meilleur de moi-même.

À Guillaume et Romain de MOABI, merci pour m'avoir accueilli dans votre équipe pendant mon stage et d'avoir continué de me suivre durant ma thèse. Merci pour votre expertise et votre aide face aux questions de bioinformatiques.

Maintenant je souhaite remercier les membres de l'équipe *Des gamètes à la naissance* qui m'ont accueilli durant ces trois ans. Merci à Daniel, Ahmed, Céline, Sandrine, Patrick, Laurence et Côme pour votre bonne humeur et vos conseils. Merci pour m'avoir aussi bien intégré dans l'équipe.

À Clara, Mélina, Olivier et Alberto du groupe *Régulation (épi)génétique au cours de la spermatogenèse*, Hayden et Giulia qui y ont séjourné, merci pour votre présence quotidienne, vos encouragements et vos conseils. Merci également pour vos travaux sans lesquels ma thèse ne serait pas la même.

À Sophie D., ma coéquipière de thèse, merci pour ta présence et ton soutien durant toute cette thèse. Merci d'avoir partagé toutes les étapes et galères de la thèse avec moi.

À Mitra, ma compère bioinfo de cette équipe, merci pour ton soutien, tes encouragements et tes conseils. Merci de m'avoir accompagné dans le projet « Apprentis Chercheurs » qui a été une si belle expérience.

À Marie-Sophie, Marine, Caroline, Audrey, Léa, Sophie F. et Lorenzo, merci d'avoir été de si belles personnes, pour votre bonne humeur et votre soutien durant ces trois ans. Merci pour les bons moments passés que ce soit au labo durant les pauses du midi ou après lors de nos afterwork.

À Ikram, Juliette, Nastia qui ont intégré l'équipe plus récemment, merci pour votre bonne humeur et vos encouragements.

Je tiens également à sincèrement remercier ma famille qui m'ont toujours soutenu dans mes études bien que le domaine leur est totalement inconnu. Merci également pour avoir eu confiance en moi toutes ces années.

À Maël, merci de m'avoir soutenu et encouragé durant ces trois années. Merci d'avoir été présent dans les bons et les mauvais moments. Merci d'avoir été toujours présent pour écouter mes présentations.

Sommaire

Résumé.....	3
Abstract.....	5
Remerciements.....	6
Sommaire.....	7
Liste des abréviations.....	9
Liste des outils.....	10
Introduction.....	11
1 Généralités sur l'organisation de l'information génétique et épigénétique chez les mammifères.....	11
1.1 Informations génétiques.....	12
1.1.1 La génomique.....	12
1.1.2 La transcriptomique.....	14
1.1.3 Analyses bioinformatiques des données de séquençage.....	16
1.2 Informations épigénétiques.....	17
1.2.1 La méthylation de l'ADN.....	17
1.2.2 La conformation ouverte ou fermée de la chromatine.....	18
1.2.3 Les modifications post-traductionnelles des histones.....	19
1.2.4 L'organisation spatiale de la chromatine.....	23
2 Dynamique transcriptionnelle et chromatinienne au cours de la spermatogenèse.....	26
2.1 Au cours de la phase pré-méiotique.....	28
2.2 Lors de la transition de la mitose à la méiose.....	31
2.3 Lors de la méiose.....	32
2.3.1 Dynamique transcriptionnelle et chromatinienne des autosomes.....	33
2.3.2 Dynamique transcriptionnelle et chromatinienne des chromosomes sexuels.....	34
2.4 Lors de la spermiogenèse.....	37
2.4.1 Déroulement de la spermiogenèse.....	37
2.4.2 Régulation chromatinienne de la transcription dans les spermatides rondes.....	38
2.4.3 Remplacement des histones par les protamines à partir du stade spermatides allongées.....	39
2.5 L'épigénome du spermatozoïde.....	41
2.5.1 La transcription des ARN non-codants.....	42
2.5.2 L'accessibilité de la chromatine et les modifications post-traductionnelles.....	44
2.5.3 La méthylation de l'ADN.....	45
3 DOT1L & la méthylation de la marque H3K79.....	46
3.1 DOT1/DOT1L, méthyltransférase de H3K79.....	46
3.1.1 Rôle biologique de DOT1L et de H3K79me.....	47
3.1.2 DOT1L et H3K79 dans l'embryogenèse.....	48
3.1.3 DOT1L et H3K79 durant la spermatogenèse.....	49
3.2 Méthylation de H3K79 et transcription des gènes.....	50
3.2.1 Rôle dans l'élongation de la transcription.....	51
3.2.2 H3K79me2 et l'épissage alternatif.....	52
3.3 Interaction (« <i>cross-talk</i> ») entre H3K79me et d'autres modifications post-traductionnelles.....	53
3.3.1 H3K79me2/3 et l'ubiquitination de l'histone H2B.....	54
3.3.2 H3K79me2/3 et l'acétylation de l'histone H4.....	54
3.3.3 H3K79me2/3 et les marques d'hétérochromatine.....	54
Résultats.....	56
1 Introduction et contexte de l'étude.....	56
2 Approches développées pour le projet.....	57
2.1 Données biologiques.....	57
2.1.1 Données de transcriptomique KO- <i>Dot1l</i> et contrôle.....	57
2.1.2 Données de ChIP-seq pour H3K79me2.....	58

2.2	Mise en place d'une analyse d'expression différentielle.....	58
2.2.1	Génération du modèle d'analyse.....	58
2.2.2	Estimation des gènes dérégulés.....	59
2.2.3	Analyse par <i>spike-in</i>	60
2.3	Caractérisation des états chromatinien.....	60
2.3.1	Association des gènes aux <i>enhancers</i>	62
2.3.2	Caractérisation de l'environnement chromatinien.....	64
3	DOT1L régule l'organisation de la chromatine et l'expression des gènes durant la différenciation spermatique.....	64
3.1	Résultats de l'étude.....	64
3.2	Article 1.....	66
4	DOT1L a des effets antagonistes sur la régulation de l'expression génique liés aux chromosomes sexuels et autosomiques durant la spermatogenèse.....	85
4.1	Résultats de l'étude.....	85
4.2	Article 2.....	87
Discussion.....		121
1	DOT1L dans la spermatogenèse.....	121
2	H3K79me2 est associée à une transcription active au niveau des gènes et des <i>enhancers</i>.....	122
3	DOT1L régule la transcription des gènes par plusieurs mécanismes.....	124
3.1	Analyse différentielle d'expression.....	124
3.2	DOT1L active la transcription de plusieurs centaines de gènes via la présence de H3K79me2.....	124
3.3	DOT1L est nécessaire à la réactivation des chromosomes sexuels après la méiose.....	126
3.4	DOT1L réprime la transcription de plusieurs centaines de gènes.....	128
4	H3K79me2 et développement embryonnaire précoce.....	130
Conclusion.....		131
Bibliographie.....		132
Annexes.....		148
	Vignette ChromENVEE.....	148
	Figures supplémentaires – Article 1.....	163
	Figures supplémentaires – Article 2.....	191

Liste des abréviations

3C	: <i>Chromosome Conformation Capture</i>
4C	: <i>Circular Chromosome Conformation Capture</i>
acH4	: acétylation de l'histone H4
ADN	: Acide désoxyribonucléique
ADNc	: ADN complémentaire
ARN	: Acide ribonucléique
ARNInc	: ARN long non codant
ARNm	: ARN messenger
ARNnc	: ARN non codant
ARNpolIII	: ARN polymérase II
ARNr	: ARN ribosomal
ARNsnc	: petit ARN non codant
ARNt	: ARN de transfert
ATAC-seq	: <i>Assay for Transposase-Accessible Chromatin using sequencing</i>
BH	: Test de Benjamini-Hochberg
BS-seq	: <i>Bisulfite sequencing</i>
CGP	: Cellule Germinale Primordiale
ChIP-seq	: <i>Chromatin Immunoprecipitation Sequencing</i>
CpG	: Cytosine-phosphate-Guanine
CPM	: <i>Count Per Million</i>
CUT&RUN	: <i>Cleavage Under Target & Release Using Nuclease</i>
CUT&TAG	: <i>Cleavage Under Target & Tagmentation</i>
DNMT	: DNA méthyltransferase
DDR	: <i>DNA Damage response</i>
DOT1L	: <i>Disruptor of telomeric silencing-1 like</i>
DotCom	: <i>Dot1-containing complex</i>
DSB	: <i>Double strand break</i>
ENCODE	: <i>Encyclopedia of DNA Elements</i>
ES	: Spermatide allongée
ESC	: Cellule Souche Embryonnaire
FACS	: <i>Fluorescence Activated Cell Sorting</i>
FAIRE-seq	: <i>Formaldehyde-Assisted Isolation of Regulatory Elements</i>
FC	: <i>Fold-Change</i>
FDR	: <i>False Discovery Rate</i>
GRC	: <i>Genome Reference Consortium</i>
GSC	: Cellule Souche Germinale
Hi-C	: <i>High-throughput chromatin conformation capture</i>
ICSI	: <i>Intracytoplasmic Sperm Injection</i>
Kit-	: Spermatogonie indifférenciée
Kit+	: Spermatogonie différenciée
KO	: <i>Knock-out</i>
MeDIP-seq	: <i>Methylated DNA immunoprecipitation sequencing</i>
MERVL	: <i>Murine Endogenous Retrovirus-L</i>
miARN	: microARN
MLL	: <i>Mixed Lineage Leukemia</i>
MSCI	: <i>Meiotic Sex Chromosome Inactivation</i>
NGS	: <i>New Generation Sequencing</i>
PCR	: <i>Polymerase Chain Reaction</i>
piARN	: ARN interagissant avec Piwi

PMSC	: <i>Post Meiotic Sex Chromatin</i>
PTM	: <i>Modification Post-Traductionnelle</i>
RNA-seq	: <i>RNA sequencing</i>
ROSI	: <i>Round Spermatid Injection</i>
RPKM	: <i>Reads Per Kilobase Million</i>
RRBS	: <i>Reduced Representation Bisulfite Sequencing</i>
RT-PCR	: <i>Reverse Transcriptase PCR</i>
RS	: <i>Spermatide Ronde</i>
scATAC-seq	: <i>Single-cell ATAC-seq</i>
SCI	: <i>Spermatocyte Primaire</i>
SCII	: <i>Spermatocyte Secondaire</i>
scRNA-seq	: <i>Single-cell RNA-seq</i>
SE	: <i>Super-Enhancer</i>
SEC	: <i>Super Elongation Complex</i>
SNP	: <i>Single Nucleotide Polymorphism</i>
SSR	: <i>Simple Sequence Repeat</i>
T2T	: <i>Telomere-to-telomere</i>
TAD	: <i>Domaine Topologique Associé</i>
TES	: <i>Transcription End Site</i>
TET	: <i>Ten-Eleven Translocation</i>
TPM	: <i>Transcript Per Million</i>
tsRNA	: <i>ARNt – dérivé des petits ARN</i>
TSS	: <i>Transcription Start Site</i>
UTR	: <i>Untranslated Transcribed Region</i>
WGBS	: <i>Whole Genome Bisulfite sequencing</i>

Liste des outils

BSPAT	: <i>Bisulfite Sequencing Pattern Analysis Tool</i>
BWA	: <i>Burrow-Wheeler Aligner</i>
ChromENVEE	: <i>Chromatin Environment and Enhancer-dependent Expression</i>
GSEA	: <i>Gene Set Enrichment Analysis</i>
HOMER	: <i>Hypergeometric Optimization of Motif EnRichment</i>
IGV	: <i>Integrative Genomics Viewer</i>
MACS2	: <i>Model-based Analysis of ChIP-seq version 2</i>
RGV	: <i>ReproGenomics Viewer</i>
ROSE	: <i>Rank Ordering of Super-Enhancers</i>
RUV	: <i>Remove Unwanted Variation</i>
SICER	: <i>Spatial Clustering for Identification of ChIP-Enriched Regions</i>
STAR	: <i>Spliced Transcript Alignment to a Reference</i>

Introduction

Ce projet de thèse a consisté en l'étude de la dynamique transcriptionnelle et du remodelage chromatinien lors de la spermatogenèse et, plus particulièrement, nous nous sommes intéressés au rôle de H3K79me2 et de sa méthyltransférase DOT1L. Ainsi, je commencerai par introduire l'organisation de l'information génétique et épigénétique chez les mammifères de façon générale puis je me concentrerai plus précisément sur son implication dans la spermatogenèse. Finalement, je décrirai le rôle chromatinien et transcriptomique de H3K79me2 et de DOT1L dans différents contextes cellulaires.

1 Généralités sur l'organisation de l'information génétique et épigénétique chez les mammifères

L'ADN contient l'ensemble de l'information génétique nécessaire au développement et au fonctionnement des cellules des mammifères. Cette information génétique s'accompagne également d'informations épigénétiques, qui désignent l'ensemble des processus transmissibles permettant de moduler l'expression des gènes sans pour autant altérer la séquence de l'ADN (Holliday, 1994). Ces processus épigénétiques sont indispensables pendant le développement et d'une manière générale à tout processus de différenciation cellulaire, comme la spermatogenèse qui sera abordée plus loin (cf. 2 *Dynamique transcriptionnelle et chromatinienne au cours de la spermatogenèse*).

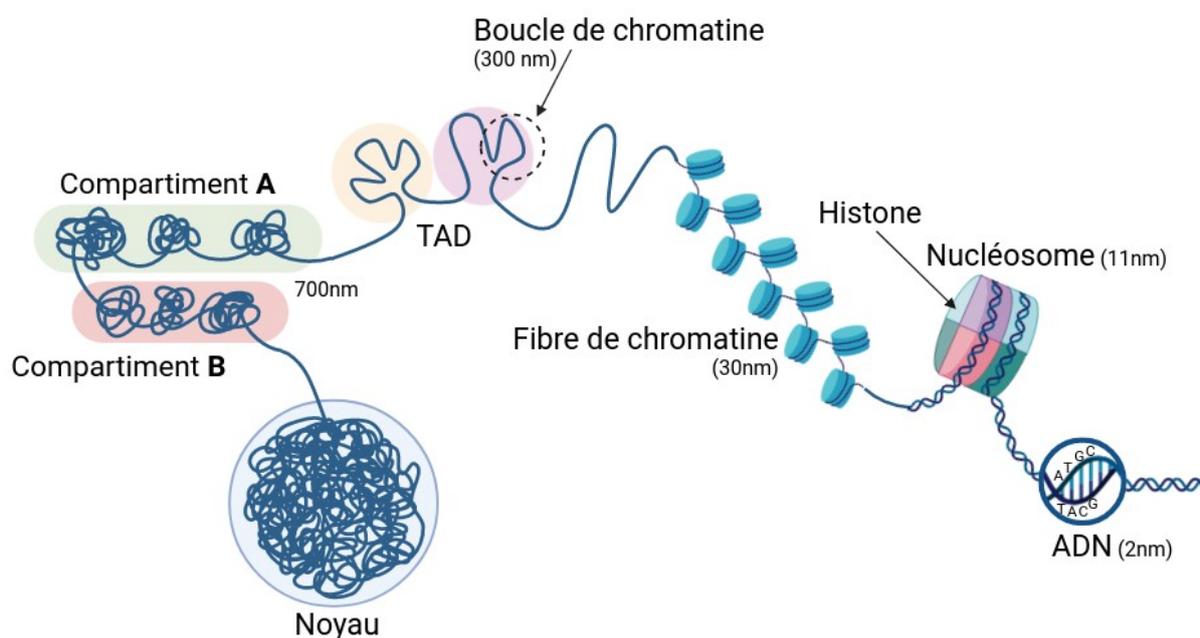


Figure 1 : Organisation de la chromatine chez les mammifères

Dans le noyau, l'ADN est associé à diverses protéines, et en particulier les histones pour former la chromatine. Celle-ci est organisée en une structure complexe comportant plusieurs niveaux. Les éléments principaux de cette structure sont, dans l'ordre croissant de taille, le nucléosome (octamère d'histones autour duquel s'enroule l'ADN), les fibres de chromatine, les boucles, les domaines topologiques et les compartiments (Fig. 1). Le

nucléosome est composé de deux exemplaires de chacune des quatre histones suivantes : H2A, H2B, H3 et H4 (Kornberg, 1974) reliées entre elles par une autre histone, H1, et forment ensemble la fibre chromatinienne. Au sein de cette fibre, les nucléosomes sont plus ou moins condensés, ce qui fait que la chromatine est présente sous deux conformations, l'euchromatine qui correspond à une conformation ouverte et l'hétérochromatine qui correspond à une conformation fermée. La chromatine s'enroule sur elle-même pour former des boucles de chromatine qui peuvent être de taille variable allant de quelques kilobases à plusieurs mégabases de longueurs (Misteli, 2020). Ces boucles permettent la régulation à distance de l'expression des gènes par le biais de régions appelées *enhancers* et sont regroupées dans des domaines topologiques, appelés des domaines d'association topologique (TAD pour *Topologically Associating Domain*) (Lieberman-Aiden et al., 2009; Rao et al., 2014; Farré et al., 2015). Ces TADs sont regroupés entre eux dans des structures appelées les compartiments. Ces compartiments existent sous deux formes, le compartiment A, localisé dans le centre du noyau, qui correspond à des régions riches en gènes et transcriptionnellement actives, et le compartiment B, localisé majoritairement dans la périphérie du noyau et qui correspond à des régions plus pauvres en gènes et transcriptionnellement inactives (Lieberman-Aiden et al., 2009; Rao et al., 2014; Hildebrand and Dekker, 2020).

1.1 Informations génétiques

1.1.1 La génomique

L'étude de la séquence de l'ADN, c'est-à-dire la génomique, a été rendue possible par la mise en place de techniques variées du domaine de la biologie moléculaire comme le transfert d'ADN (*southern blot*, Southern, 1975) et l'amplification par PCR (*Polymerase Chain Reaction*, Mullis and Faloona, 1987), mais c'est le développement des techniques de séquençage qui a permis l'essor de la génomique. Depuis le séquençage de Sanger en 1977 (Sanger et al., 1977), plusieurs méthodes, appelées séquençage de nouvelle génération (NGS), ont vu le jour, comme le « *shotgun* » (technologie développée par Illumina®), ou le pyroséquençage et plus récemment le séquençage de 3^e génération, également nommé le séquençage « *long-read* » (technologie développée par Nanopore® ou Pacific Biosciences®). Ces trois générations de séquençage sont assez différentes dans leur fonctionnement (Fig. 2).

Au fur et à mesure de leur évolution, ces méthodes sont devenues de plus en plus performantes et permettent d'avoir une résolution plus élevée (Fig. 2). Cependant, ces améliorations ont un coût et, de ce fait, le séquençage « *short read* » reste le plus couramment utilisé dans le cas d'analyses (épi)génomique et transcriptomique. En définitive, l'apparition de ces technologies a permis la production de grandes quantités de données, le développement de protocoles bioinformatiques permettant l'analyse de ces données.

	Première génération	Seconde génération	Troisième génération
Technologie fondamentale	Discrimination par la taille de brin fragments d'ADN labellisés, produit par SPS ou dégradation	"Wash-and-scan" SPS	SPS, par dégradation, par inspection directe de la molécule d'ADN
Résolution	Moyenne en fonction du nombre de copies de l'ADN séquencé	Moyenne en fonction du nombre de copies de l'ADN séquencé	Elevé (une molécule)
Précision de la séquence lue	Haute	Haute	Modérée
Taille de la séquence (read) lue	800/1000 pb (paires de bases)	10/50 pb	1000+ pb
Volume de données en sortie	Bas	Haut	Modéré
Coût	Elevé par base, bas par séquençage	Bas par base, élevé par séquençage	Bas/Modéré par base, bas par séquençage
Temps de séquençage (début de la première réaction jusqu'au résultat)	~ Heures	~ Jours	~ Heures
Préparation des échantillons	Modérément complexe, PCR non requise	Complexe, PCR requise	Simple/Complexe suivant la technologie employée
Analyse des données	Facile	Complexe à cause du nombre important des données et de la petite taille des reads qui compliquent l'assemblage des génomes et des algorithmes d'alignement.	Complexe : nombre important des données, ces nouvelles technologies émettent de nouveaux types d'informations et donc de nouvelles analyses doivent être implémentées
Résultats primaire	Identification des bases ("base calling") avec des scores de qualité	Identification des bases avec des scores de qualité	Identification des bases avec des scores de qualité, et d'autres informations comme la cinétique, la modifications des bases.

SPS = Séquençage Par Synthèse (SBS en anglais) : Détermination de la séquence d'ADN par synthèse de l'ADN complémentaire

Figure 2 : Tableau comparant les trois méthodes de séquençage. La première génération correspondant au séquençage de Sanger, la deuxième au séquençage « short-read » et la troisième au séquençage « long-read ». Figure provenant de <https://bioinfo-fr.net/le-sequençage> et issue de Schadt et al., 2010.

Le séquençage d'ADN peut être réalisé suivant plusieurs approches, le séquençage du génome entier (*whole genome sequencing*) ou le séquençage d'exome (*exome sequencing*). Ces techniques de séquençage ont permis d'établir la séquence génomique de nombreux organismes et individus, et ainsi de détecter les variations génétiques entre les individus d'une même espèce comme des événements de polymorphismes (simple nucléotide (SNP), séquences répétées (SSR)...). Ainsi, depuis le séquençage de ϕ X174, un virus bactériophage (Sanger et al., 1977), plus de 3000 autres organismes ont été séquencés et continuent de l'être, notamment *via* le projet zoonomia qui, en 2023, a séquencé 240 nouvelles espèces de mammifères (Vignieri, 2023). Le séquençage du premier génome humain a été achevé en 2001 (Lander et al., 2001). Depuis, de nouvelles versions sont régulièrement mises à jour notamment grâce au projet international *Genome Reference Consortium* (GRC) qui a pour but de séquencer l'entièreté du génome humain (Genome Reference Consortium, n.d.). Débuté il y a plus de 20 ans, ce projet a permis de séquencer $\approx 92\%$ du génome humain et

c'est l'apparition du projet T2T (*Telomere-to-Telomere*) en 2022 qui a permis quant à lui de compléter une grande partie des régions manquantes du génome humain (Mao and Zhang, 2022). En parallèle, le projet 1000 Génomes a été mis en place (1000 Genomes Project Consortium et al., 2010; Fairley et al., 2020). Initié en 2008, ce projet avait pour objectif de cataloguer les variations génétiques humaines en séquençant un millier d'individus. Depuis, les génomes de dizaines de milliers d'individus ont été séquencés.

1.1.2 La transcriptomique

Le produit de la transcription de l'ADN, les ARN, peut être divisé en deux catégories : les ARN codants (ou ARN messagers, ARNm) et les ARN non-codants (ARNnc). Les ARN non-codants englobent plusieurs types d'ARN, les ARN ribosomiaux (ARNr), les ARN de transfert (ARNt), les micros ARN (miARN), ou les autres catégories d'ARN non-codants (ARN long non-codant, ARNlnc ou ARN petit non-codant, ARNsnc).

Comme pour l'étude de l'ADN, l'étude de l'ensemble des ARNs, la transcriptomique, a été facilitée par l'apparition de techniques de biologie moléculaire comme le transfert d'ARN (*northern blot*, Kevil et al., 1997) et la transcription inversée de l'ARN en ADN complémentaire (ADNc) suivie de son amplification par PCR – technique appelée RT-PCR (*Reverse Transcriptase PCR*, Cavé et al., 2003). Mais c'est l'apparition du séquençage de nouvelle génération appliqué aux molécules d'ARN rétro-transcrites en ADNc (Nagalakshmi et al., 2008; Wilhelm et al., 2008), RNA-seq ou RNA-seq « *bulk* », qui a permis l'essor de la transcriptomique, et ainsi de caractériser le transcriptome des organes et populations cellulaires. Cette approche a permis une description globale des profils transcriptionnels de tissus ou cellules et de changements du niveau d'expression de gènes ou de l'épissage alternatif entre différentes conditions (Wang et al., 2009). Les étapes du RNA-seq sont (i) la phase de génération des fragments (appelée également librairie), (ii) la phase de séquençage, (iii) le retrait des adaptateurs et enfin, (iv) la phase de l'alignement (Fig. 3). De nombreux outils et protocoles ont été mis en place pour faciliter les analyses bioinformatiques de données RNA-seq (cf. 1.1.3 *Analyses bioinformatiques des données de séquençage*). Le RNA-seq peut être réalisé sur tous les types d'ARN mais se fait le plus souvent à partir des ARNm. Les techniques d'extraction des ARNs et de préparation des librairies ne sont pas les mêmes pour étudier les différentes catégories d'ARNs. Une extraction des ARNs totaux est le plus souvent réalisée, mais certains protocoles permettent d'enrichir en petits ARNs. Dans le cas des ARNm, ils sont, après extraction, sélectionnés par leur queue polyA au moment de la préparation de la librairie. Les protocoles d'analyses sont également différents en fonction du type d'ARN que l'on souhaite analyser. Dans le cadre de ma thèse, je me suis intéressée aux ARNm, de ce fait, je ne détaillerai pas les méthodes d'analyse pour les autres types d'ARNs.

L'alignement des fragments pour les analyses RNA-seq est réalisé couramment à l'aide d'outils spécialisés tel que STAR (Dobin et al., 2013) ou TopHat2 (Kim et al., 2013), qui permettent ainsi d'obtenir la quantification du nombre de fragments alignés par gène. Ces deux outils sont aujourd'hui parmi les plus cités dans le domaine de la transcriptomique mais, par comparaison avec TopHat2, STAR est vraisemblablement plus utilisé pour

l'analyse des variations de séquence de l'ADN et de l'épissage (Engström et al., 2013). L'analyse des changements de l'expression génique fait suite à l'étape d'alignement et se fait généralement à travers une comparaison entre plusieurs conditions biologiques pour définir, par exemple, l'impact d'un traitement biologique ou d'un processus de différenciation cellulaire (Fig. 3). Pour étudier ces changements plusieurs types d'analyses existent comme les analyses différentielles d'expression utilisant des packages R comme edgeR (Robinson et al., 2010) et DESeq2 (Love et al., 2014) ou les analyses d'enrichissement en certaines voies cellulaires ou processus biologiques comme GSEA (Subramanian et al., 2005). Ces analyses peuvent être utilisées de manière indépendante ou combinées, car donnant des informations complémentaires. En effet, les analyses différentielles d'expression permettent d'avoir une information sur les gènes dérégulés entre les deux conditions alors que l'analyse d'ontologie permet d'obtenir la liste des ontologies dérégulées en mettant en évidence des liens fonctionnels entre différents gènes dérégulés.

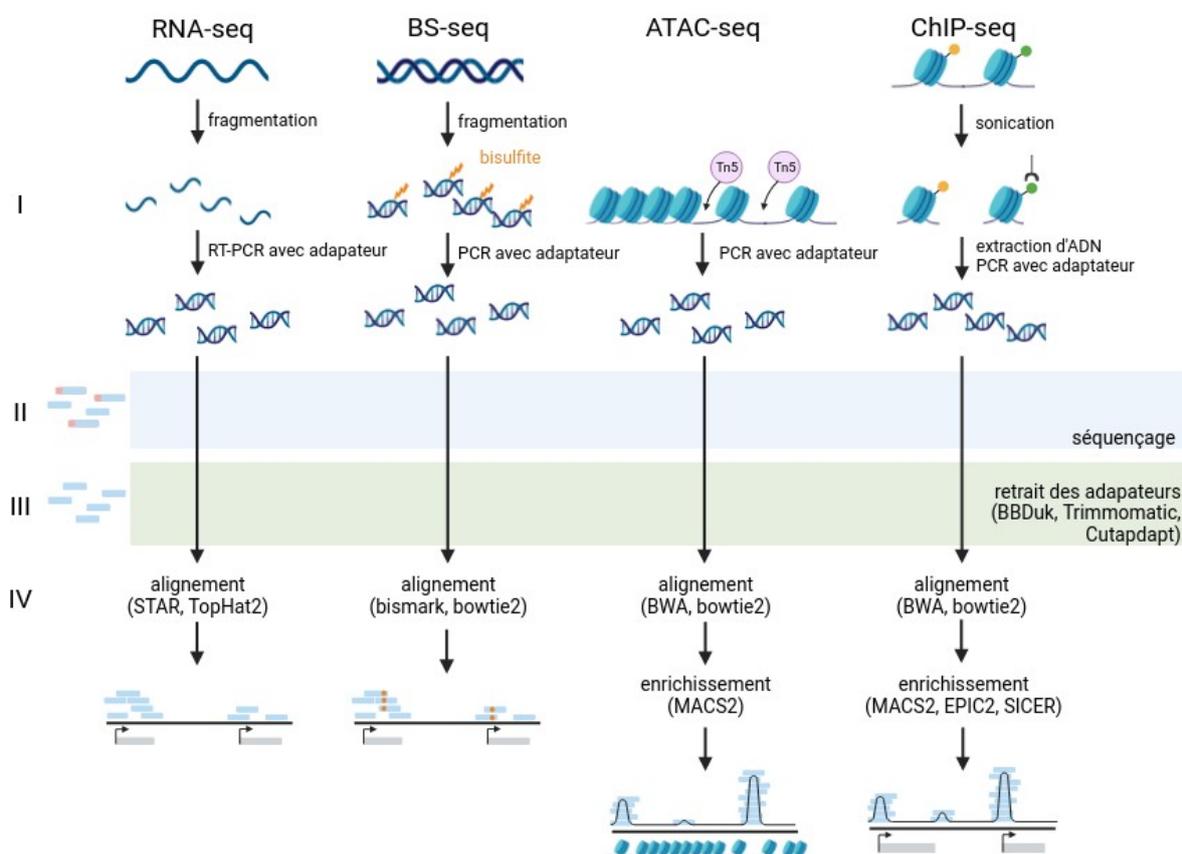


Figure 3 : Protocole d'alignement des données de séquençage sur un génome de référence pour des analyses de RNA-seq, de BS-seq, d'ATAC-seq et de ChIP-seq. En bleu, l'étape de séquençage (séquençage par Illumina©) et en vert l'étape de retrait des adaptateurs. Pour le BS-seq, en orange sont représentées les régions méthylées. Pour le ChIP-seq et l'ATAC-seq, un pic signifie un enrichissement dans la marque d'histone d'intérêt ou une région ouverte de la chromatine respectivement.

Depuis quelques années, une autre méthode de RNA-seq est apparue, le *single-cell* RNA-seq (scRNA-seq), qui permet, tout comme la méthode classique, d'analyser le transcriptome (Eberwine et al., 2014). Cependant cette méthode ne se fait pas à la même échelle, le RNA-seq donnant pour sa part des informations sur l'expression génique dans un

groupe de cellules plus ou moins complexe selon le matériel de départ, alors que le scRNA-seq permet d'avoir l'expression génique de chaque cellule au sein d'un tissu, organe ou groupe de cellules. Cette méthode est aujourd'hui très utilisée pour comprendre l'hétérogénéité transcriptomique, pour caractériser les différents types cellulaires au sein d'un tissu/organe voire en découvrir de nouveaux, ou pour étudier la dynamique transcriptionnelle à l'échelle de la cellule unique au cours du développement. Cependant, le scRNA-seq ne permet pas d'avoir une sensibilité aussi importante que le RNA-seq sur l'analyse de l'expression des gènes due à une détection limitée des gènes exprimés (seulement 10-45 % des ARN exprimés dans une cellule sont captés par ces méthodes, Mereu et al., 2020). Cette nouvelle méthode ne permet pas non plus d'obtenir des informations fiables sur l'épissage alternatif et sur les ARNs non codants (Li and Wang, 2021).

1.1.3 Analyses bioinformatiques des données de séquençage

Comme précisé, il existe trois types de séquençage (Sanger, « *short read* » et « *long-read* »), et de ce fait, la méthode d'analyse des données de séquençage dépend de la technique employée. Dans ce manuscrit, je me concentrerai sur les méthodes d'analyse de données provenant de la technologie « *short read* » par Illumina®, car c'est le type de séquençage qui a été utilisé pour générer les données que j'ai étudiées. Quel que soit le type d'analyse (RNA-seq, BS-seq, ChIP-seq ou ATAC-seq), le protocole d'analyse des données de séquençage est similaire et comprend les mêmes grandes étapes (Fig. 3). Les outils ne sont cependant pas toujours les mêmes, car ils sont développés spécifiquement pour chaque type de données. Le protocole de génération des fragments (ou préparation de la librairie de séquençage) est propre à chaque type d'analyse et est suivi d'un séquençage (Illumina® dans ce cas), puis les séquences sont traitées dans le but de retirer les séquences d'ADN rajoutées aux extrémités des fragments lors de l'étape de préparation de la librairie (adaptateurs), et qui permettent d'une part d'identifier l'échantillon pour permettre d'analyser plusieurs échantillons à la fois et, d'autre part, de disposer d'amorces nécessaires au séquençage. Plusieurs outils permettant le retrait des adaptateurs existent comme Cutadapt (Martin, 2011), BBDuk (Bushnell, 2014) ou Trimmomatic (Bolger et al., 2014) et sont communs aux différents types d'analyses. Ces trois outils ont des performances similaires et le choix d'utiliser l'un par rapport aux autres est en fonction de la préférence de l'utilisateur (Guzman and D'Orso, 2017). Le retrait des adaptateurs est suivi par une étape de filtration en fonction de la qualité des fragments puis par une étape d'alignement des fragments contre le génome de référence.

L'alignement des fragments est réalisé à l'aide d'outils spécialement développés en fonction du type d'analyse réalisée, par exemple, l'alignement de données de RNA-seq se fait en utilisant, entre autres, les outils STAR ou TopHat2, tandis que l'alignement des données de ChIP-seq et d'ATAC-seq est couramment fait à l'aide des outils bowtie2 (Langmead and Salzberg, 2012) ou BWA (Li and Durbin, 2009). L'alignement de données de RNA-seq permet d'obtenir le niveau d'expression de chacun des gènes, cette information se traduit par une quantification du nombre de fragments alignés sur chaque gène (Fig. 3).

L'alignement des données de ChIP-seq et d'ATAC-seq va permettre d'obtenir le niveau d'enrichissement d'une région génomique, se traduisant par un pic d'enrichissement obtenu à partir de la quantification du nombre de fragments alignés sur un intervalle donné du génome en utilisant des outils tels que MACS2 (Zhang et al., 2008).

1.2 Informations épigénétiques

1.2.1 La méthylation de l'ADN

L'ADN contient de régions riches en dinucléotides CG que l'on nomme des îlots CpG (Cytosine-phosphate-Guanine). Chez l'humain, les îlots CpG représentent en moyenne 41 % du génome (Piovesan et al., 2019) et 42 % du génome chez la souris (Zhao and Zhang, 2006). Plusieurs types de modifications associées aux cytosines existent comme la méthylation, l'hydroxy-méthylation ou la carboxylation, cependant, la méthylation est celle qui est la plus étudiée et décrite dans la littérature (Moore et al., 2013). La méthylation de l'ADN est également très représentée dans le génome où 70-80 % de ces îlots sont enrichies en cytosines méthylées (Li and Zhang, 2014; Lismer and Kimmins, 2023). Ces régions méthylées sont notamment localisées dans les satellites péri-centromériques ou dans le corps de gènes ou les régions promotrices riches en dinucléotides CG (Lister et al., 2009). La méthylation de l'ADN est impliquée dans plusieurs processus fonctionnels essentiels tels que la régulation de l'expression génique, la structure de l'ADN, l'inactivation de l'X chez les femelles ou encore le phénomène d'empreinte parentale, et tout particulièrement durant le développement embryonnaire (Greenberg and Bourc'his, 2019; Al Adhami et al., 2022). La méthylation de l'ADN est régulée à trois niveaux : l'établissement, la maintenance et la déméthylation, régis par des enzymes spécifiques, permettant d'apposer la méthylation de l'ADN, nommées les DNMTs pour *DNA methyltransferases* (Lister et al., 2009). Pendant longtemps, la méthylation de l'ADN a été décrite comme ayant un rôle de répresseur de la transcription (Weber et al., 2007), cependant les mécanismes sous-jacents sont multiples et restent encore peu compris, car la méthylation de l'ADN seule ne semble pas permettre la répression génique (Greenberg and Bourc'his, 2019). Bien que la méthylation de l'ADN semble être principalement localisée dans les régions ouvertes de la chromatine (Stadler et al., 2011), elle semble favoriser la formation d'hétérochromatine notamment en recrutant des enzymes permettant la modification des histones comme l'acétylation ou la méthylation de H3K9 (Table. I, cf. 1.2.3 *Les modifications post-traductionnelles des histones*) (Greenberg and Bourc'his, 2019).

L'étude de la méthylation de l'ADN peut se faire à l'aide de la technique de séquençage NGS couplée à un traitement au bisulfite (BS-seq ou Bisulfite-seq ou WGBS pour *whole-genome bisulfite sequencing*) (Feil et al., 1994; Lister et al., 2009; Berman et al., 2011) qui permet d'identifier les cytosines méthylées dans le génome du matériel biologique étudié. La technique consiste à traiter les échantillons d'ADN au bisulfite ce qui a pour conséquence de convertir l'ensemble des cytosines non méthylées en uracile (C→U), les uraciles sont ensuite convertis en thymine par amplification PCR (U→T). Les séquences d'ADN obtenues sont ensuite comparées avec le génome de référence pour avoir le nombre d'évènements de conversion (Lister et al., 2009). L'analyse des données de méthylation de

l'ADN suit le protocole classique d'analyse des données de séquençage (Fig. 3). L'alignement des données est réalisé majoritairement en utilisant l'outil Bismark (Krueger and Andrews, 2011) ou occasionnellement *via* bowtie2 (Langmead and Salzberg, 2012), cependant contrairement aux autres techniques (ATAC-seq, ChIP-seq et RNA-seq), l'alignement de données de BS-seq est plus compliqué dû à la conversion des cytosines en thymines. Ainsi d'autres techniques d'analyses ont été développées pour étudier la méthylation de l'ADN comme le RRBS (*Reduced Representation Bisulfite Sequencing*), le MeDIP-seq (*Methylated DNA immunoprecipitation sequencing*) ou encore l'analyse par hybridation sur puces. Ces méthodes, sauf la dernière où il n'y a pas d'étape d'alignement, utilisent les mêmes outils d'alignement que le BS-seq et sont suivies d'une étape d'analyse à l'aide d'outil tel que SeqMonk (<https://github.com/s-andrews/SeqMonk>) ou BSPAT (Hu et al., 2015).

1.2.2 La conformation ouverte ou fermée de la chromatine

Comme la chromatine peut se présenter sous deux conformations différentes qui sont l'euchromatine et l'hétérochromatine (Fig. 1). L'euchromatine correspond à l'état décondensé de la chromatine, également appelé état ouvert. L'ouverture de la chromatine permet l'accessibilité des facteurs de transcription et de la machinerie transcriptionnelle permettant ainsi l'expression des gènes (Buenrostro et al., 2013). L'hétérochromatine quant à elle correspond à l'état condensé de la chromatine, également nommé état fermé.

La caractérisation des régions accessibles de la chromatine se fait *via* les méthodes de FAIRE-seq, de DNase-seq ou d'ATAC-seq, qui permettent de détecter un enrichissement en séquences d'ADN correspondant aux régions ouvertes. Les régions génomiques considérées comme ouvertes sont donc caractérisées par un pic d'enrichissement alors que les régions fermées ne sont pas enrichies (Fig. 3). L'ATAC-seq pour *Assay for Transposase-Accessible Chromatin using sequencing* (Buenrostro et al., 2013) est aujourd'hui la méthode la plus utilisée, car elle est appréciée pour la rapidité d'exécution de son protocole et la faible quantité de matériel biologique nécessaire à sa mise en place. L'ATAC-seq est basée sur l'utilisation d'une enzyme spécifique, la transposase Tn5, qui se lie aux régions non condensées de l'ADN. Le traitement des données d'ATAC-seq suit le modèle classique d'analyse (Fig. 3) dont l'alignement est réalisé majoritairement grâce aux outils bowtie2 ou BWA. Les deux outils présentent des performances et fonctionnalités similaires bien que bowtie2 soit plus rapide (Yan et al., 2020). La méthode d'analyse des résultats provenant du séquençage de bibliothèques ATAC-seq est similaire à celle utilisée pour les données de ChIP-seq que ce soit au niveau des outils d'alignement ou pour l'analyse des pics d'enrichissement *via* l'outil MACS2 (Zhang et al., 2008). L'ATAC-seq présente très peu de bruit de fond, il n'est donc pas nécessaire d'avoir un fichier de contrôle (*input*) comme pour les données de ChIP-seq (fichier *input* correspondant au résultat du séquençage de l'ADN traité de façon identique au test mais sans immunoprécipitation par l'anticorps d'intérêt). Comme pour le RNA-seq, une méthode de *single-cell* a été développée pour l'ATAC-seq, le scATAC-seq (Buenrostro et al., 2015). Le principe est le même que pour l'ATAC-seq mais se fait à l'échelle de la cellule unique. Cependant, comme pour le scRNA-seq, le scATAC-seq ne

donne pas une information aussi précise que l'ATAC-seq « *bulk* » (sur population cellulaire) sur l'accessibilité de la chromatine due à une détection limitée des régions ouvertes par cellule (1-10 % des pics détectés) (Chen et al., 2019a).

1.2.3 Les modifications post-traductionnelles des histones

Le nucléosome est composé de quatre histones : H2A, H2B, H3 et H4, chacune présente en deux exemplaires. Ces histones peuvent subir des modifications post-traductionnelles (PTMs) et/ou peuvent être remplacées par des variants d'histones. Les variants diffèrent des formes canoniques des histones sur quelques résidus qui peuvent ainsi leur conférer des propriétés différentes (Jang et al., 2015; Shaytan et al., 2015). Certains variants ont été impliqués dans divers processus biologiques tel que la spermatogenèse (Wang et al., 2019a) mais je ne parlerai ici que des modifications post-traductionnelles des histones, objet principal de mon travail.

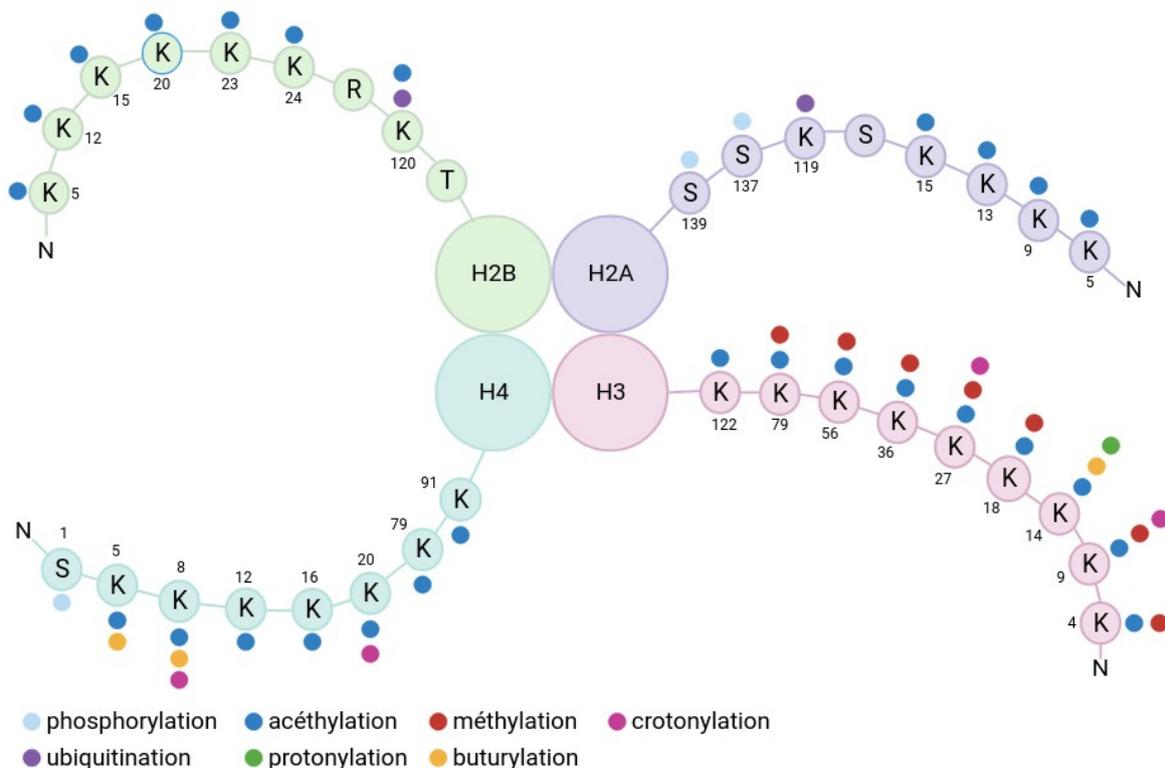


Figure 4 : Position des résidus d'histones et exemples de modifications post-traductionnelles. Les modifications présentées ne sont pas exhaustives. Le N correspondant à la position N-terminale

Les modifications post-traductionnelles des histones, comme les acétylations (ac), méthylations (me), phosphorylations (ph) ou ubiquitinations (ub) des lysines (K) (Fig. 4) sont essentielles à la régulation épigénétique. Ces modifications sont nommées suivant une nomenclature précise qui sera utilisée dans le reste de ce manuscrit. À titre d'exemple, la marque H3K27ac correspond à l'acétylation de la lysine en position 27 de l'histone H3. Ces modifications sont régulées par des enzymes qui permettent leur mise en place (*writers*) ou leur retrait (*erasers*). Il est important de préciser que les enzymes impliquées dans les PTMs ne sont pas nécessairement spécifiques d'une marque et plusieurs enzymes peuvent apposer une même modification (Table. I).

Protéine	Rôle	Marque	Référence
DOT1L / KMT4	méthyltransférase	H3K79me1, 2 et 3	van Leeuwen et al., 2002
SETDB1 / ESET	méthyltransférase	H3K9me3	Wang et al., 2003
EZH2	méthyltransférase	H3K27me1, 2 et 3	Cao et al., 2002
GCN5 / KAT2A	acétyltransférase	H3K9/14ac H4K5/8/12/16ac	Brownell et al., 1996
SETD2	méthyltransférase	H3K36me3	Yoh et al., 2008
PRDM9	méthyltransférase	H3K4me1, 2 et 3	Mihola et al., 2009

Table 1 : Liste non exhaustive des marques citées dans ce manuscrit (cf. 2 Dynamique transcriptionnelle et chromatinienne au cours de la spermatogenèse) et des enzymes les modifiant.

Ces modifications organisent la structure de la chromatine en affectant le recrutement des protéines et/ou l'interaction entre les nucléosomes. Elles sont également associées lors de la différenciation cellulaire à la régulation de l'expression génique (activation ou inhibition), à la répression des éléments répétés, et/ou à la réparation de l'ADN (Kouzarides, 2007). Plusieurs de ces modifications ont été étudiées de manière approfondie au cours de ces deux dernières décennies, et sont désormais connues pour être associées à une région fonctionnelle du génome et/ou un état transcriptionnel particulier (Fig. 5). Certaines marques comme H3K79me2, H3K36me3, H3K27me1 et H4K20me1 sont associées à une transcription active contrairement à H3K27me3, H3K9me3 et H4K20me3 qui sont caractéristiques de l'état d'hétérochromatine et de la répression des gènes (Peters et al., 2003; Squazzo et al., 2006). En utilisant la méthode de ChIP-chip, l'article de Bernstein et al, fait partie des premiers articles décrivant la position de l'enrichissement de la marque H3K4me3 dans le promoteur des gènes (Bernstein et al., 2002). Depuis, la position de l'enrichissement de nombreuses marques a été décrite (Fig. 5). Une marque peut être enrichie au niveau des gènes, que ce soit dans le promoteur pour H3K4me3 et H3K27me3, au niveau du site d'initiation de la transcription TSS (*Transcription Start Site*) et dans le corps des gènes (*gene body*) pour H3K79me2 ou dans des régions distales comme les *enhancers* pour H3K4me1. Contrairement aux autres marques d'histones, H3K36me3 est localisée dans le corps du gène et vers le site de terminaison de la transcription TES (*Transcription End Site*). En plus de leur localisation différente, les marques sont classées en fonction de leur type d'enrichissement : les marques peuvent être étalées (*broad*) comme H3K79me2 ou étroites (*narrow*) comme H3K4me3. Ces différences sont à prendre en compte lors des analyses de données de séquençage.

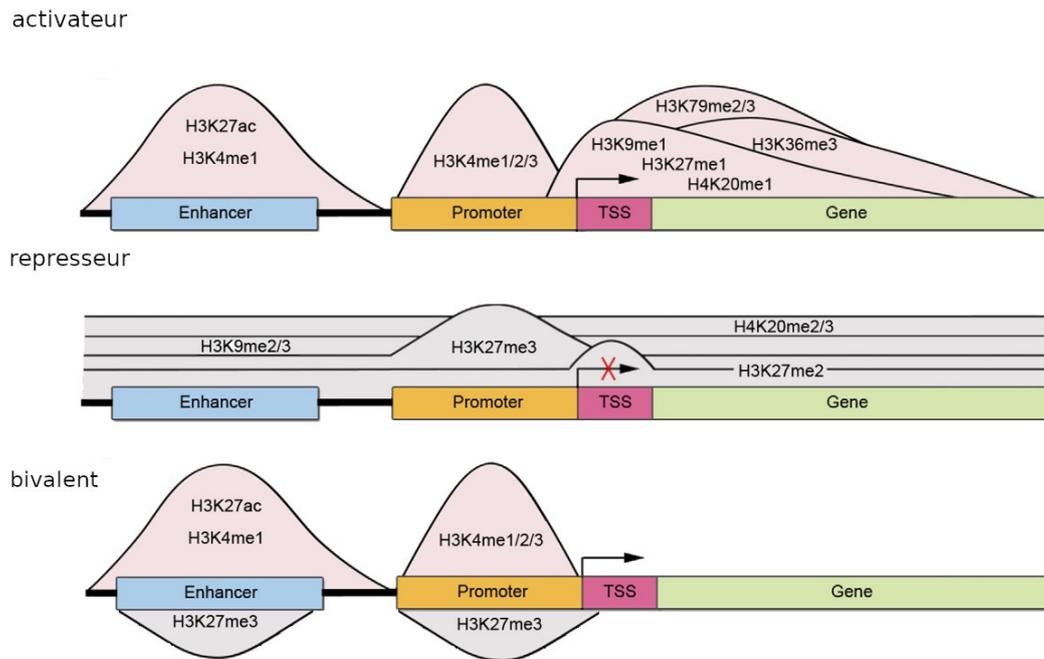


Figure 5 : Schéma représentant les principales modifications d'histones et leur position par rapport aux gènes. Les marques sont classées en fonction de leur rôle : activateur, répresseur ou bivalent. Figure adaptée de Barbero et al., 2023

L'étude de ces modifications peut être réalisée suivant plusieurs techniques de quantification comme la quantification absolue via la spectrométrie de masse ou la quantification relative *via* le *western blot* ou l'immunofluorescence. Elles peuvent également être analysées en immunoprécipitant la marque d'intérêt grâce à un anticorps spécifique, suivi d'une quantification par qPCR ou d'une analyse par séquençage (ChIP-seq), ce qui permet d'identifier les régions génomiques dans lesquelles la marque est enrichie.

Analyse par séquençage à haut débit des marques d'histone

Le ChIP-seq est une méthode de séquençage NGS permettant de capturer la localisation génomique et le niveau d'enrichissement des marques d'histones dans tout le génome grâce à l'utilisation d'anticorps dirigés spécifiquement contre les marques d'histones d'intérêt. Son analyse suit le même protocole que les autres analyses de NGS type Illumina (Fig. 3). Comme pour l'ATAC-seq, l'alignement des fragments sur le génome se fait couramment en utilisant l'outil BWA (Li and Durbin, 2009) ou bowtie2 (Langmead and Salzberg, 2012). Cette étape est suivie par une étape d'analyse grâce à des outils tels que MACS2 (Zhang et al., 2008), SICER (Xu et al., 2014) ou epic2 (Stovner and Sætrom, 2019) qui permettent d'estimer la position et le niveau d'enrichissement de la marque d'intérêt. Initialement développé pour analyser les marques *narrow*, MACS2 a depuis été mis à jour pour prendre en compte les marques *broad* et permet ainsi d'analyser un large panel de marques. *A contrario*, SICER et plus récemment epic2 sont spécialisés dans l'analyse de marques *broad*. Le niveau d'enrichissement de la marque d'histone se traduit par un pic, appelé pic d'enrichissement qui sera plus ou moins haut proportionnellement à l'enrichissement de la région génomique en la marque. Ces pics sont visualisables à l'aide de logiciel tel que IGV (Robinson et al., 2011) et l'utilisation de packages R tel que

ChIPseeker (Yu et al., 2015) permet leur caractérisation, leur annotation, la comparaison et la visualisation des données.

La qualité des données générées diffère selon les marques d'histones. En effet, les marques les plus éloignées de la queue N-terminale de l'histone, plus proche du centre, sont plus difficiles d'accès et donc fournissent des résultats de moins bonne qualité (Fig. 4). À titre d'exemple, il est plus aisé d'avoir une bonne qualité pour une modification sur H3K4 que sur H3K79. De plus, pour l'analyse de ChIP-seq, il est nécessaire d'avoir un grand nombre de cellules pour avoir une résolution suffisante. De ce fait, pour des populations cellulaires de faible nombre (comme celles correspondant aux premières étapes du développement embryonnaire (Hainer et al., 2019; Zhou et al., 2023) ou dans lesquelles la chromatine est difficile d'accès car hautement compactée, comme c'est le cas des spermatozoïdes, de nouvelles techniques ont été développées demandant moins de matériel biologique : le CUT&RUN (Skene and Henikoff, 2017) et le CUT&TAG (Kaya-Okur et al., 2020). Comme pour le ChIP-seq, le CUT&RUN et le CUT&TAG utilisent un anticorps spécifique de la marque d'histone d'intérêt qui, dans le cas de ces deux techniques, est couplé à une enzyme, MNase pour le CUT&RUN, et Tn5 pour le CUT&TAG (Fig. 6). Une fois la génération des fragments réalisée, l'analyse des données suit le même protocole et utilise les mêmes outils que pour l'analyse de ChIP-seq (Fig. 3).

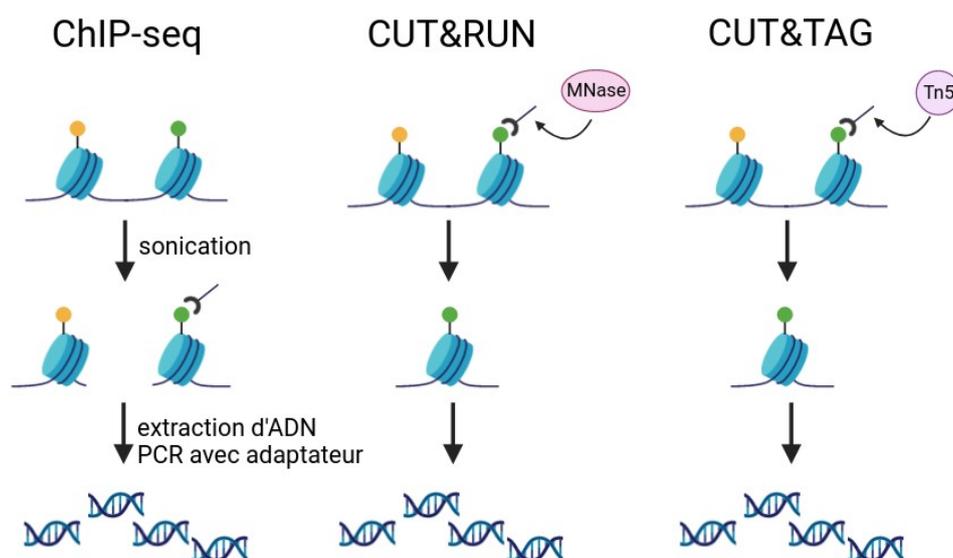


Figure 6 : Comparaison de la génération des fragments entre les méthodes de ChIP-seq, CUT&RUN et de CUT&TAG

L'analyse bioinformatique de ces données permet d'avoir des informations concernant la localisation génomique de ces marques et, de ce fait, a montré que plusieurs marques peuvent être présentes au sein d'une même région. Cette co-présence définit des états chromatinien spécifiques. L'utilisation d'outils bioinformatiques tel que ChromHMM (Ernst and Kellis, 2012) conjointement avec les données du projet ENCODE (ENCODE Project Consortium, 2012) a permis l'identification et la caractérisation de ces états. En effet, ChromHMM utilise un modèle HMM (Modèle Caché de Markov, Schuster-Böckler and Bateman, 2007) pour prédire des états chromatinien à une région donnée en fonction

de l'enrichissement en une ou plusieurs marques d'histones (Fig. 7). Cet outil met en évidence, entre autres, plusieurs catégories d'*enhancers*, chacune caractérisée par la présence de H3K4me1 (Barski et al., 2007). La présence d'autres marques d'histones permet de définir des *enhancers* actifs (H3K27ac), inactifs (H3K27me3) ou bivalents (H3K27me3, H3K4me3 et H3K27ac à un faible niveau). Cette dernière catégorie étant la plus représentée dans les cellules embryonnaires (Bernstein et al., 2005; Creighton et al., 2010; Zentner et al., 2011).

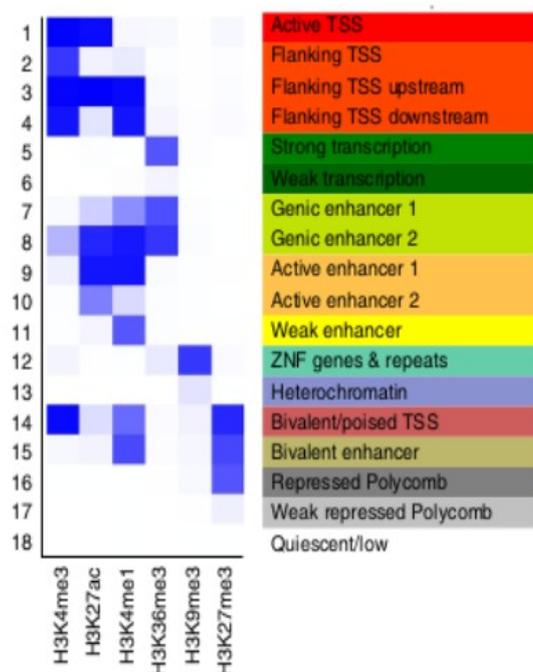


Figure 7 : Modèle des états chromatinien caractérisés par une combinaison de marques d'histones en utilisant l'outil ChromHMM. À gauche, le niveau d'enrichissement des marques d'histones et à droite le nom des états chromatinien associés. Figure adaptée de Ernst and Kellis, 2017

1.2.4 L'organisation spatiale de la chromatine

L'expression génique est étroitement liée à l'organisation tridimensionnelle de la chromatine. L'analyse de la conformation chromatinienne se fait généralement en utilisant la méthode appelée « Capture de la Conformation de la Chromatine » (ou 3C) qui se décline sous plusieurs formes : 4C (*Circular Chromosome Conformation Capture*), Hi-C (*High-throughput Chromatin Conformation Capture*), etc. L'ensemble de ces méthodes reprend le même principe qui consiste à identifier les zones de contact entre régions génomiques distantes (Fig. 8). La méthode 3C (Dekker et al., 2002) servant de base à toutes les autres méthodes consiste à figer la conformation tridimensionnelle de la chromatine en utilisant un agent fixateur. Dans un premier temps, les régions d'ADN spatialement proches les unes des autres dans le noyau sont liées entre elles (*crosslink*), cette étape est suivie d'une digestion enzymatique de l'ADN puis d'une ligation des séquences d'ADN se retrouvant à proximité pour obtenir un fragment d'ADN hybride composé de fragments distants sur une représentation linéaire mais proche dans une organisation tridimensionnelle (Fig. 8).

Dans le cadre d'une analyse Hi-C, tous ces fragments d'ADN peuvent ensuite être séquencés puis alignés contre le génome de référence en utilisant les mêmes outils que pour l'alignement des données de CHIP-seq (Fig. 3) (Lieberman-Aiden et al., 2009). Après alignement et différentes étapes de curation des données, les résultats peuvent être représentés sous forme d'une carte de fréquences de contact entre des régions génomiques à l'aide par exemple de l'outil HiCExplorer (Ramírez et al., 2018; Wolff et al., 2018). Les TADs y sont représentés sous la forme de triangles dans une matrice de corrélation correspondant à la conformation dépliée de l'ADN (Fig. 8).

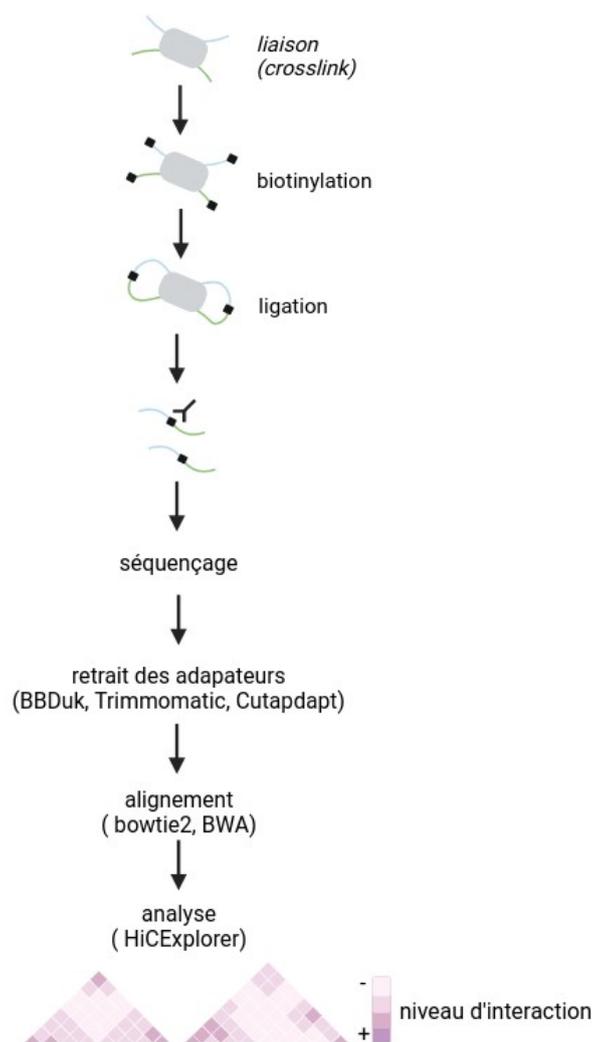


Figure 8 : Processus d'analyse des données de Hi-C. La représentation graphique correspond à une représentation linéaire des TADs.

Les TADs sont étroitement liés à la régulation de l'expression génique notamment en agissant sur le repliement de la chromatine ou en agissant en tant que barrière (Boltsis et al., 2021) qui empêche la propagation de l'hétérochromatine vers les régions actives et inversement (Narendra et al., 2015). La régulation de l'expression génique passe par l'interaction entre les *enhancers* et les promoteurs des gènes (Dixon et al., 2012; Schoenfelder et al., 2015; Schoenfelder and Fraser, 2019), en effet une altération des TADs peut entraîner des modifications au niveau de l'interaction *enhancer*-promoteur, en

permettant entre autres de nouvelles interactions ou en supprimant celles existantes (Flavahan et al., 2016). Les *enhancers* sont des régions génomiques, de 150 à 1500bp, qui permettent de réguler l'expression génique à distance en formant des boucles d'ADN avec les promoteurs des gènes (Fig. 9) (Levine, 2010; Long et al., 2016). Ces régions régulatrices peuvent être localisées en amont ou en aval d'un gène (*enhancers* intergéniques) ou dans les introns d'un gène (intragéniques) (Schoenfelder and Fraser, 2019). L'utilisation d'analyse par ChIP-seq sur H3K27ac et d'Hi-C a permis d'identifier plusieurs centaines de milliers d'*enhancers actifs* (entre 400 000 et 1,4 millions selon l'espèce) (ENCODE Project Consortium, 2012; Shen et al., 2012; Plank and Dean, 2014). Notons qu'un *enhancer* peut réguler plusieurs gènes et qu'un gène n'est, en retour, pas nécessairement régulé par un unique *enhancer* (Karnuta and Scacheri, 2018). Certaines analyses dérivées de l'Hi-C permettent d'étudier de façon précise la régulation des gènes par les *enhancers* et notamment en définissant les couples *enhancers*-promoteurs (Lieberman-Aiden et al., 2009). Ces méthodes restent, cependant, complexes à mettre en place et très coûteuses si l'on souhaite atteindre une résolution suffisante pour identifier les interactions *enhancers*-promoteurs. Certaines études ont donc développé des méthodes mathématiques et statistiques pour prédire ces couples *enhancers*-promoteurs en s'affranchissant d'analyse Hi-C. À la place, des données de ChIP-seq pour H3K4me1 et H3K27ac sont utilisées pour identifier les *enhancers*, et le gène le plus proche de l'*enhancer* dans un intervalle inférieur à 100kb lui est associé (Rada-Iglesias et al., 2011). Cette méthode a néanmoins ses limites car, dans certains cas, les *enhancers* ne régulent pas le gène le plus proche et peuvent réguler des gènes à plusieurs centaines de kilobases de distance (Amano et al., 2009; Sanyal et al., 2012; Schoenfelder et al., 2015; Schoenfelder and Fraser, 2019).

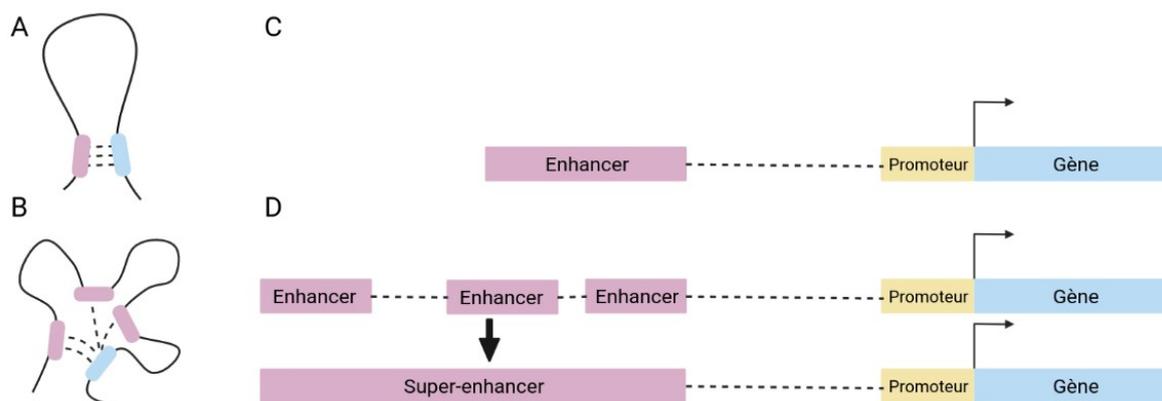


Figure 9 : Schéma des *enhancers* et des *super-enhancers*. À gauche, la représentation du repliement de l'ADN formant une boucle d'ADN avec un *enhancer* (A) et un *super-enhancer* (B) et à droite, la représentation linéaire du génome pour un *enhancer* (C) et un *super-enhancer* (D). Dans ce schéma les *super-enhancers* sont définis suivant le modèle de Whyte et al, 2013

Dans la littérature, certains articles font mention de régions régulatrices à distances nommées les *super-enhancers* (SE). Dans leur article, Whyte et al., définissent les *super-enhancers* comme des agrégats d'*enhancers* (Fig. 9) plus larges que les *enhancers* dit « classiques » (Whyte et al., 2013). Ces *super-enhancers* sont également caractérisés par un enrichissement plus important en certains facteurs de transcription pionniers comme SOX2

ou NANOG. En se basant sur cette définition, l'outil ROSE (<https://github.com/stjude/ROSE>) a été développé pour permettre de différencier les *enhancers* des *super-enhancers* en se basant sur des données de ChIP-seq pour H3K27ac ou H3K4me1 (Lovén et al., 2013; Whyte et al., 2013). Les études utilisant ROSE montrent que le nombre de *super-enhancers* définis varie en fonction du type cellulaire allant d'environ 200 *super-enhancers* dans les cellules souches embryonnaires (ESC) jusqu'à 1000 dans les spermatozoïdes (Whyte et al., 2013; Maezawa et al., 2020; Cao et al., 2023). De plus, de récentes études utilisant la technique Hi-C montrent que les *super-enhancers* permettent aussi des interactions longue distance entre eux alors même que les interactions inter-TADs sont quasi inexistantes (Cao et al., 2023; Kessler et al., 2023).

Chez les mammifères et dans tous les types cellulaires, les *super-enhancers* régulent les gènes qui sont essentiels pour l'identité cellulaire, notamment dans les processus de développement et de différenciation cellulaire, comme la spermatogenèse (Whyte et al., 2013; Sin et al., 2015; Maezawa et al., 2020).

2 Dynamique transcriptionnelle et chromatinienne au cours de la spermatogenèse

Le testicule est l'organe mâle dans lequel se déroule la spermatogenèse, c'est-à-dire, le processus de différenciation des cellules germinales mâles nécessaire à la procréation. La spermatogenèse est très dynamique d'un point de vue transcriptionnel et protéomique (Soumillon et al., 2013), ce qui en fait un excellent modèle pour étudier les mécanismes épigénétiques qui s'opèrent durant la différenciation cellulaire. C'est un processus qui est extrêmement bien conservé chez les mammifères, et qui est organisé en trois phases successives qui sont (i) la prolifération des spermatogonies, (ii) la méiose et (iii) la spermiogenèse (Roosen-Runge, 1962; Hess and Renato de Franca, 2008) (Fig. 10). Chacune de ces phases sera développée dans la suite de cette introduction.

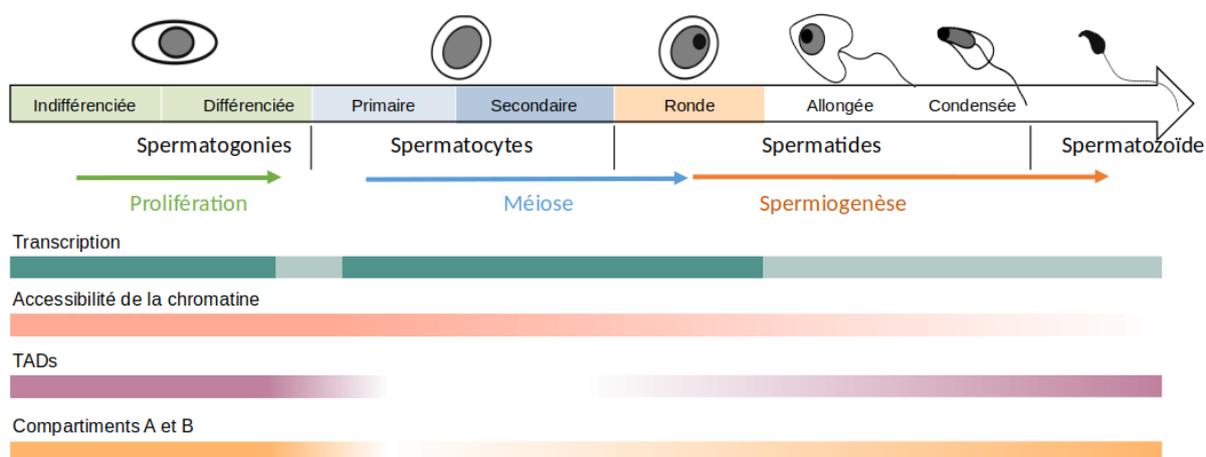


Figure 10 : Les trois grandes étapes de la spermatogenèse avec les stades cellulaires associées. Figure basée sur les données de Soumillon et al, 2013, Luo et al, 2019, Wang et al, 2019b et les articles citées dans cette partie

L'étude du transcriptome des différents types cellulaires de la spermatogenèse via RNA-seq (Gan et al., 2013; Soumillon et al., 2013; Hasegawa et al., 2015; Sin et al., 2015) a permis d'étudier la dynamique transcriptionnelle au cours de la spermatogenèse depuis plus de 10 ans (Fig. 10). Les cellules germinales mâles ayant une taille et/ou un contenu en ADN distincts à chacune des 3 étapes majeures de la spermatogenèse, il est, en effet, possible de collecter des fractions enrichies en cellules germinales pré-méiotiques, méiotiques et post-méiotiques par cytométrie en flux ou FACS (méthode la plus utilisée) ou par élutriation – qui permet un meilleur rendement cellulaire mais avec une pureté inférieure qu'avec la méthode FACS (Meistrich, 1977; Bastos et al., 2005) – et ainsi produire des données RNA-seq « *bulk* ». Avec le FACS, l'utilisation de marqueurs permet de distinguer certains sous-types de cellules mais à partir de 2018, l'utilisation de scRNA-seq a permis de caractériser avec précision les différents types/sous-types cellulaires existants au sein du testicule (Chen et al., 2018b; Green et al., 2018; Guo et al., 2018; Hermann et al., 2018; Lukassen et al., 2018; Wang et al., 2018; Ernst et al., 2019). L'utilisation conjointe du RNA-seq « *bulk* » et du scRNA-seq montre que la spermatogenèse se caractérise par d'importants changements transcriptionnels faisant intervenir plusieurs milliers de gènes (Fig. 11). En effet, durant son développement la spermatogenèse alterne entre des phases de transcription active (durant la prolifération et le début de la spermiogenèse) et de transcription inactive (une brève inactivation au tout début de la méiose, puis l'extinction transcriptionnelle à la fin de la spermiogenèse) (Fig. 10). La spermatogenèse est également caractérisée par des modifications de l'activité transcriptionnelle des ARNInc, des éléments transposables, d'éléments non géniques comme les pseudogènes ou par des événements d'épissage alternatif, ce qui fait du testicule l'organe avec la plus grande diversité transcriptomique (Soumillon et al., 2013).

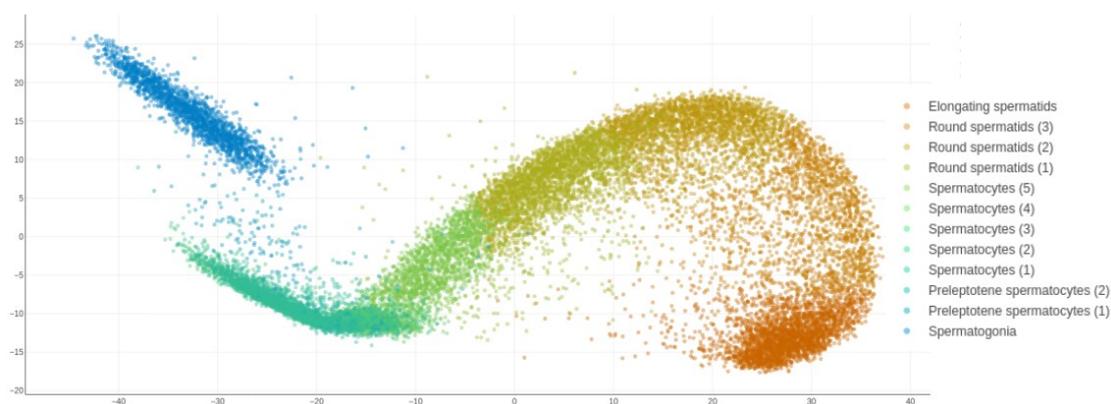


Figure 11 : Clusters d'expression au cours de la spermatogenèse en utilisant une approche de scRNA-seq. Figure basée sur les données de Green et al, 2018 en utilisant RGV (<https://rgv.genouest.org/>, Darde et al., 2015, 2019)

Cette diversité transcriptomique est associée à des mécanismes de régulation épigénétique comme la régulation par des ARN non codants, la méthylation de l'ADN et les modifications post-traductionnelles et variants d'histones (An et al., 2014; de Mateo and Sassone-Corsi, 2014; Rathke et al., 2014; Stewart et al., 2016; Lin et al., 2017; Zuo et al., 2018). De la même façon que pour les études transcriptomiques, les modifications de l'ADN (méthylation, hydroxylation, carboxylation...) (Gan et al., 2013; Blythe et al., 2021), les PTMs

et les variants d'histones ont pu être étudiés dans des populations de cellules germinales triées (Tan et al., 2011; Hammoud et al., 2014; Bryant et al., 2015; Zuo et al., 2018; Liu et al., 2019; Chen et al., 2020b). Ces études, réalisées en grande majorité sur le modèle murin, ont permis de montrer que la spermatogenèse se caractérise par des changements importants en termes d'architecture chromatinienne (TAD et compartiments chromatinien) et de modifications d'histones (Alavattam et al., 2019; Patel et al., 2019; Vara et al., 2019; Wang et al., 2019b; Luo et al., 2020) (Fig. 10).

2.1 Au cours de la phase pré-méiotique

Chez les mammifères, les spermatogonies sont une population hétérogène composée de plusieurs sous-types classés en fonction de leur niveau de différenciation. Actuellement, la classification des spermatogonies se fait via des marqueurs comme Thy1 ou c-kit pour la différenciation et MKI67 ou $\alpha 6$ pour la prolifération (Lord and Oatley, 2017; Caldeira-Brant et al., 2020).

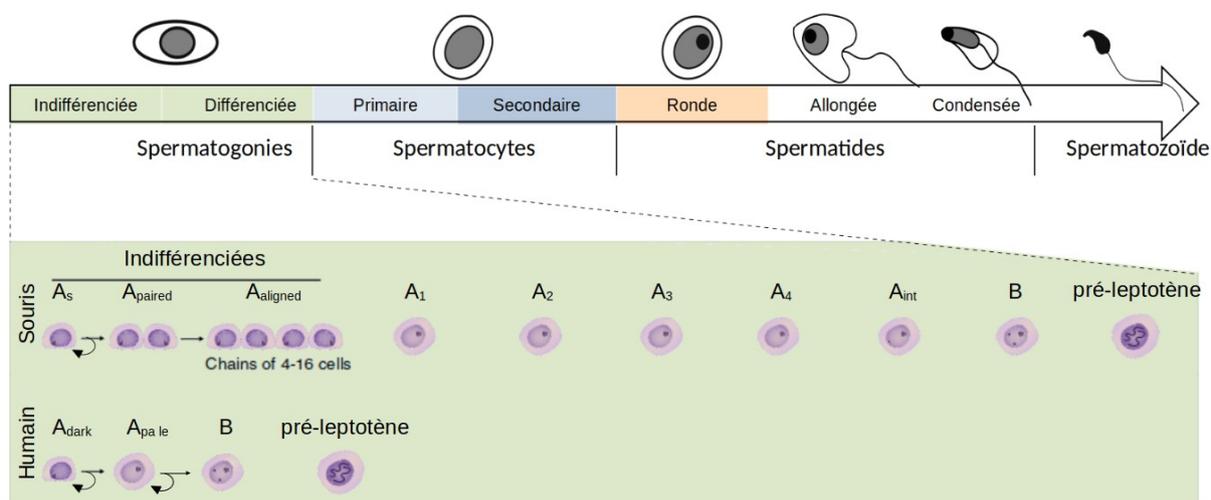


Figure 12 : Étapes de différenciation des spermatogonies chez la souris, en haut et chez l'homme, en bas. Figure adaptée de Ernst et al, 2019

Depuis l'article de Clermont en 1966, basé sur des observations histologiques, il est établi que chez l'homme les spermatogonies sont divisées en trois catégories. La première catégorie dans l'ordre de différenciation correspond aux spermatogonies indifférenciées qui sont les plus proches des cellules souches ; elles sont nommées A_{dark} (Fig. 12). Parmi ces cellules A_{dark} , certaines se différencient en spermatogonies indifférenciées actives A_{pale} . Les deux stades A_{dark} et A_{pale} ont la capacité de renouvellement, permettant ainsi un maintien du stock de cellules spermatogoniales. Les spermatogonies A_{pale} deviennent, pour une partie, des spermatogonies différenciées B (Clermont, 1966; Caldeira-Brant et al., 2020).

Chez la souris, le processus de différenciation comprend plus d'étapes de divisions que chez l'homme (Fig. 12). La première catégorie correspond aux spermatogonies indifférenciées A_{single} ou A_{s} , certaines se renouvellent pour constituer le stock de cellules spermatogoniales. Les spermatogonies A_{s} subissent ensuite une phase de mitose pour obtenir les spermatogonies A_{paired} (ou A_{pr}), suivie de plusieurs divisions mitotiques

successives pour obtenir le groupe des spermatogonies $A_{aligned}$: A_{al-4} , A_{al-8} et A_{al-16} . Les spermatogonies indifférenciées se différencient via six divisions mitotiques successives pour obtenir les spermatogonies A_1 , puis A_2 , A_3 et A_4 , puis en spermatogonies intermédiaires (A_{int}) et finalement en spermatogonies B (de Rooij, 2017; Caldeira-Brant et al., 2020). Les spermatogonies A_s , A_p et A_{al} sont regroupées dans cette introduction sous le nom de spermatogonies indifférenciées et les spermatogonies du stade A_1 jusqu'au stade B sont catégorisées comme spermatogonies différenciées. À la suite de cela, les spermatogonies B se divisent pour obtenir les spermatocytes pré-leptotènes qui engagent l'étape suivante de la spermatogenèse qui est la phase de méiose.

Plusieurs études ont analysé la dynamique transcriptionnelle et chromatinienne de deux types cellulaires collectés par FACS : i) les spermatogonies dites indifférenciées, qui expriment plusieurs marqueurs (≈ 10 marqueurs connus) avec, entre autres, le marqueur Thy1 ou le marqueur SSEA4 (Hammoud et al., 2014; Guo et al., 2017; Maezawa et al., 2018b) ; ii) les spermatogonies différenciées, qui expriment le marqueur c-kit (Guo et al., 2017; Maezawa et al., 2018b). En utilisant des données de scRNA-seq, Guo et al., ont montré que plusieurs milliers de gènes sont différentiellement exprimés entre les spermatogonies indifférenciées et les spermatogonies différenciées humaines (Guo et al., 2017). Cette hétérogénéité cellulaire est également retrouvée chez la souris (Ernst et al., 2019). Chez l'homme et la souris, les gènes spécifiques des spermatogonies indifférenciées sont associés à des facteurs de transcription des cellules souches ou des récepteurs membranaires. *A contrario*, les gènes spécifiques des spermatogonies différenciées sont nécessaires au cycle cellulaire et à la prolifération, indiquant que ce changement transcriptionnel pourrait être essentiel pour la différenciation des spermatogonies (Guo et al., 2017, 2018; Ernst et al., 2019). De plus, l'analyse RNA-seq de ces deux stades cellulaires montre que, parmi les gènes différentiellement exprimés, plusieurs sont des enzymes modifiant les marques d'histones (Hammoud et al., 2014; Barbero et al., 2023), montrant ainsi que modifications chromatinienne sont régies par des modifications transcriptionnelles.

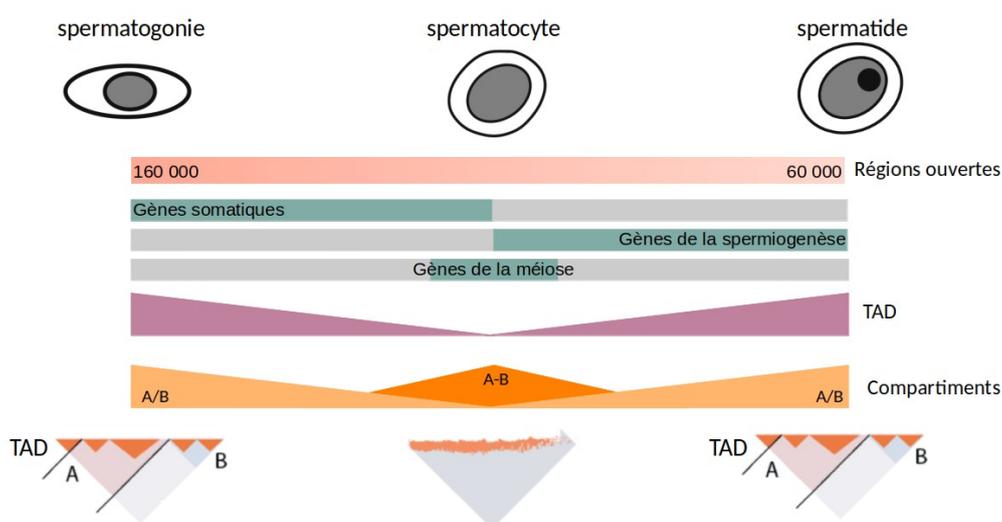


Figure 13 : Dynamique transcriptionnelle, de l'accessibilité et de la conformation de la chromatine au cours de la spermatogenèse. Figure basée sur les informations de Maezawa et al, 2018a (ATAC-seq), Luo et al, 2020, Alvarez-Gonzalez et al, 2022 et de Wang et al, 2019 (Hi-C)

Du côté de la chromatine, les spermatogonies indifférenciées sont dépourvues d'hétérochromatine visible au microscope. Elle se forme au fur et à mesure de la différenciation des spermatogonies. L'étude de la chromatine par ATAC-seq ne montre pas de changements importants entre les spermatogonies indifférenciées et différenciées chez la souris comme chez l'homme, avec une grande majorité de pics d'ATAC-seq en commun dans les spermatogonies indifférenciées et différenciées (Guo et al., 2017; Maezawa et al., 2018b). La comparaison avec les stades méiotiques et post-méiotiques, chez la souris, montre une disparition progressive d'une centaine de milliers de régions « ouvertes » au fur et à mesure de la différenciation, passant de 160 000 pics dans les spermatogonies à 60 000 dans les spermatides rondes (Maezawa et al., 2018b) (Fig. 13). La différenciation des spermatogonies s'accompagne, cependant, de changements au niveau des histones PTMs comme une augmentation de l'enrichissement de la méthylation et de l'acétylation des histones H3 et H4 dans les spermatogonies différenciées. En accord avec le processus de formation d'hétérochromatine à la fin de la phase proliférative pré-méiotique, la marque H3K9me3 augmente dans les spermatogonies différenciées par rapport au stade indifférencié (Liu et al., 2019). De la même façon, l'utilisation d'une approche ChIP-seq pour H3K4me3 et d'analyse par immunomarquage montre que H3K4me3 augmente au cours de la différenciation des spermatogonies (Godmann et al., 2007; Song et al., 2011; Hammoud et al., 2014). Cette observation corrèle avec l'augmentation de l'enzyme PRDM9 (méthyltransférase de H3K4me1, 2 et 3) dans les spermatogonies différenciées par comparaison avec les spermatogonies indifférenciées (Barbero et al., 2023). La présence de PRDM9 est importante pour la spermatogenèse, car elle permet la formation des cassures doubles brins (DSB pour *Double Strand Breaks*) survenant au cours de la méiose (Chen et al., 2020b; Sun et al., 2022). La présence de la marque H3K4me3 est également corrélée à la présence de sa méthyltransférase KMT2B qui est fortement exprimée. Tomizawa et al., ont montré que l'enzyme KMT2B, qui est nécessaire au passage des cellules souches à des cellules progénitrices, catalyse l'acquisition de la marque H3K4me3 aux promoteurs de deux catégories de gènes : ceux dits « bivalents » qui portent H3K27me3 et qui ne seront exprimés que pendant le développement embryonnaire, et ceux dits « monovalents » (qui n'ont pas H3K27me3) (Fig. 5) et qui sont, en majorité, exprimés lors de la spermiogenèse (Tomizawa et al., 2018).

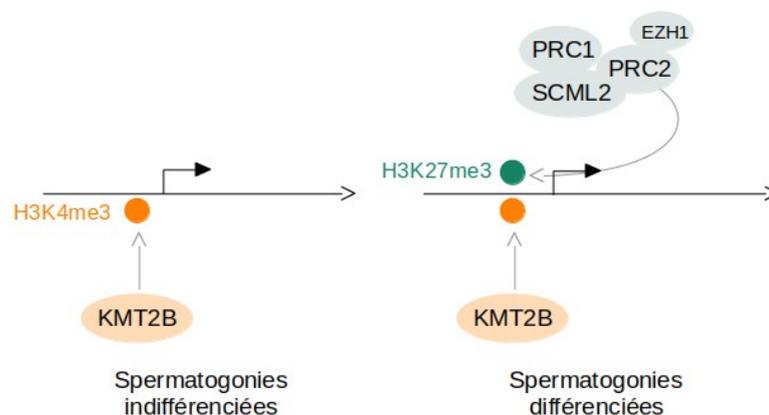


Figure 14 : Dynamique de la régulation des états bivalents durant la spermatogenèse. Figure basée sur les données de Tomizawa et al, 2018 et de Maezawa et al, 2018a

En relation avec cette étude, Maezawa et al., ont montré qu'il existe deux catégories de gènes « bivalents » dans la spermatogenèse : ceux qui portent H3K27me3 dès le stade spermatogonies indifférenciées (≈ 2000 gènes), et ceux qui acquièrent H3K27me3 au moment de l'entrée en méiose (≈ 700 gènes) et dont l'acquisition de la marque H3K27me3 est catalysée par SCML2 (Maezawa et al., 2018a). En effet, SCML2 est une enzyme permettant l'établissement de H3K27me3 par le complexe PRC2 (Fig. 14) et qui se retrouve fortement exprimée dans les spermatogonies (Maezawa et al., 2018a). Cette observation corrèle avec les données transcriptomiques qui montrent une sur-expression de *Ezh1* dans les spermatogonies indifférenciées (Barbero et al., 2023), supposant la mise en place d'un état bivalent lors de la méiose (Maezawa et al., 2018a; Barbero et al., 2023).

2.2 Lors de la transition de la mitose à la méiose

La transition de la phase mitotique vers la phase méiotique est une étape caractérisée par des changements chromatinien et transcriptomiques drastiques. L'analyse du transcriptome des cellules germinales mâles par RNA-seq puis, plus récemment, par scRNA-seq permet de mettre en évidence un « *switch* » du programme génétique dans les spermatocytes au stade pré-leptotène. Lors de cette transition, la transcription s'arrête progressivement, atteignant un niveau extrêmement bas au stade pré-leptotène (Monesi, 1964; Ernst et al., 2019). Bien qu'observé dans plusieurs études, le ou les mécanismes responsables de ce phénomène restent encore méconnus. Cet arrêt transcriptionnel pourrait permettre la reprogrammation génétique lors de la transition mitose-méiose. Celle-ci se caractérise en effet par un changement drastique des gènes exprimés : les gènes somatiques s'éteignent au profit des gènes nécessaires à la méiose et à la spermiogenèse (Soumillon et al., 2013; Sin et al., 2015; da Cruz et al., 2016; Chen et al., 2018a; Maezawa et al., 2018b). En effet, les spermatocytes au stade leptotène expriment fortement les gènes méiotiques qui sont, pour la majorité, spécifiques de ce stade cellulaire (Soumillon et al., 2013; da Cruz et al., 2016; Maezawa et al., 2018b). C'est dans les stades plus tardifs de la méiose (stades pachytènes) que les gènes nécessaires à la spermiogenèse s'expriment (Sin et al., 2015; Maezawa et al., 2018b).

Cette reprogrammation transcriptionnelle entre les stades pré-méiotique et méiotique s'accompagne d'une réorganisation de l'accessibilité de la chromatine et de sa conformation (Sin et al., 2015; Maezawa et al., 2018b; Alavattam et al., 2019; Vara et al., 2019; Wang et al., 2019b). L'analyse de données d'ATAC-seq montre que, lors de cette transition, la majorité des régions ouvertes de la chromatine dans les spermatogonies différenciées sont fermées dans les spermatocytes pachytènes (Fig. 13) (Maezawa et al., 2018b). En parallèle, des régions ouvertes apparaissent spécifiquement dans les spermatocytes. Ainsi, deux types de chromatines ouvertes s'opposent durant la spermatogenèse : les régions ouvertes présentant un programme génétique spécifique des spermatogonies (qui comprend de nombreux gènes exprimés dans les cellules somatiques) et les régions ouvertes caractérisées par un programme génétique spécifique de la spermiogenèse (Maezawa et al., 2018b). La dynamique de certaines modifications d'histones est également intéressante à regarder. L'analyse de données de ChIP-seq

obtenues pour différentes marques d'histone montre une apparition des marques H3K4me3, H3K27ac et H3K36me3 dans les spermatocytes à partir du stade leptotène (cf. 2.3 *Lors de la méiose*) (Maezawa et al., 2018b; Chen et al., 2020b). L'augmentation de ces marques d'histone, et principalement de H3K4me3, corrèle positivement avec l'expression des gènes nécessaires à la formation des DSB se déroulant durant la méiose. Comme présenté dans la partie précédente, de nombreux gènes spécifiques de la spermiogenèse sont prêts à être activés (« *poised* ») dès le stade spermatogonie (Sin et al., 2015). La diminution de H3K27me3 dans les régions bivalentes lors de la transition vers les spermatocytes est associée à l'expression de ces gènes dans les spermatocytes et les spermatides (Maezawa et al., 2018a). Cette diminution serait une conséquence de la régulation par le complexe PRC2 durant la transition de la phase de mitose vers la phase de méiose (Fig. 14). À l'inverse, l'augmentation de H3K27ac corrèle positivement avec le nombre d'*enhancers* et de *super-enhancers* dans les spermatocytes (≈ 200 *super-enhancers* définis par la présence de H3K27ac dans les spermatogonies différenciées et ≈ 500 dans les spermatocytes pachytènes) (Maezawa et al., 2020). Il est intéressant de noter que peu de *super-enhancers* sont communs entre les spermatogonies et les spermatocytes, indiquant que la majorité des *super-enhancers* retrouvés dans les spermatocytes pachytènes apparaissent lors de la méiose. Ces *super-enhancers* sont pour la plupart associés à des gènes nécessaires pour la spermiogenèse (Maezawa et al., 2020).

Pour finir, les données de Hi-C ont montré qu'à partir des stades spermatogonies indifférenciées, les compartiments A et B disparaissent au profit de l'apparition de structures A-B, structures où les compartiments A et B ne sont pas clairement distincts (Fig. 13) (Alavattam et al., 2019; Patel et al., 2019; Vara et al., 2019; Wang et al., 2019b; Luo et al., 2020; Álvarez-González et al., 2022). Cette diminution s'accompagne également d'une diminution des TADs et donc des interactions chromatiniennes intra-TADs (Luo et al., 2020).

2.3 Lors de la méiose

La seconde phase de la spermatogenèse est la phase de méiose qui fait suite à la différenciation des spermatogonies (Fig. 15). Elle consiste en une phase de réplication de l'ADN suivie de deux divisions successives : la division réductionnelle (i) à partir des spermatocytes primaires (SCI) pour obtenir deux spermatocytes secondaires (SCII) et la division équationnelle (ii), aboutissant à la formation de quatre spermatides rondes (RS). Chacune de ces divisions est composée de quatre étapes (prophase, métaphase, anaphase et télophase) dont la prophase I est, de loin, la plus longue et la plus dynamique côté transcriptionnel et chromatinien (Cobb and Handel, 1998; Ernst et al., 2019; Patel et al., 2019; Wang et al., 2019b). Elle commence par la phase leptotène qui correspond à la condensation de la chromatine pour former les chromosomes et au début de l'appariement des chromosomes homologues. Cette phase est suivie par la phase zygotène au cours de laquelle se poursuit l'appariement des chromosomes. Lors des phases leptotène et zygotène, les DSB sont mises en place. Une fois tous les chromosomes appariés, lors de la phase pachytène, la réparation des DSB par recombinaison induit des événements de *crossing-*

over. La phase diplotène est caractérisée par le retrait des synapses entre les chromosomes. Durant le processus de méiose, la séparation et l'appariement des chromosomes sont étroitement liés à la ré-organisation de la chromatine et notamment à une compaction importante de la chromatine (Fig. 13).

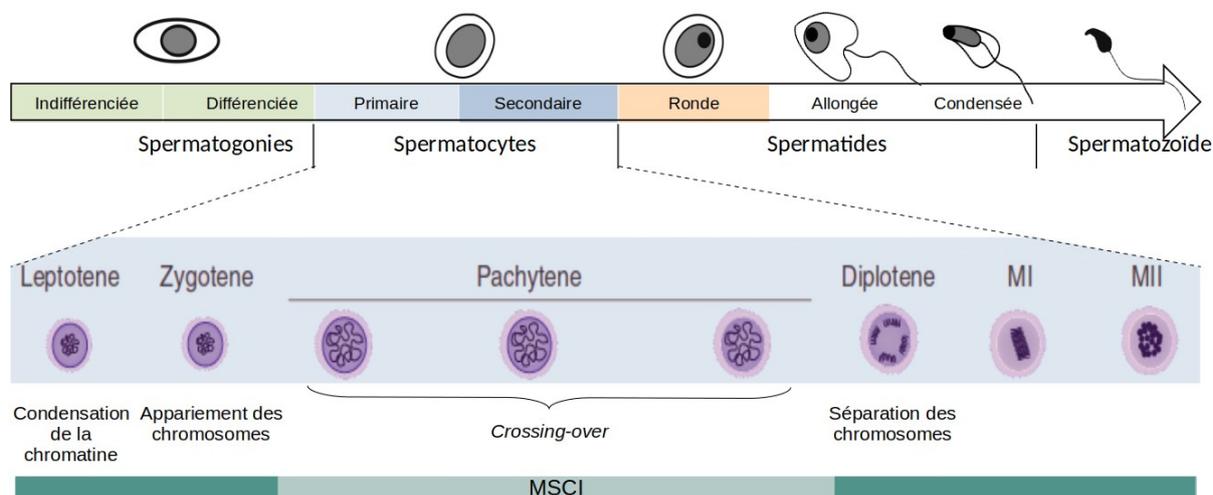


Figure 15 : Étapes de la méiose chez la souris. Figure adaptée de Ernst et al, 2019

La dynamique transcriptionnelle et épigénétique lors de la méiose est très conservée entre les différentes espèces mammifères étudiées comme les rongeurs et les primates (Murat et al., 2022; Álvarez-González et al., 2022). Cependant, elle se déroule différemment entre les autosomes et les chromosomes sexuels. Je traiterai donc d'abord des autosomes au cours de la méiose, puis des chromosomes sexuels au cours de la méiose et la spermiogenèse.

2.3.1 Dynamique transcriptionnelle et chromatinienne des autosomes

La méiose s'accompagne de modifications transcriptionnelles et post-traductionnelles, de changements de l'accessibilité de la chromatine et de la conformation tridimensionnelle de la chromatine plus importants que dans les stades pré-méiotiques (Kimmins and Sassone-Corsi, 2005; Hammoud et al., 2009; Sin et al., 2015; Maezawa et al., 2018b; Wang et al., 2019b). Pour la plupart, ces modifications résultent de la transition des spermatogonies vers les spermatocytes (cf. 2.2 *Lors de la transition de la mitose à la méiose*). L'ensemble de ces modifications permet la mise en place d'une configuration chromatinienne unique dans la spermatogenèse qui se caractérise par le lien entre transcription des gènes et organisation chromatinienne (Wang et al., 2019b).

Les analyses scRNA-seq de la spermatogenèse mettent en évidence des différences transcriptionnelles au sein des spermatocytes et notamment entre les stades zygotène et pachytène (Guo et al., 2018; Wang et al., 2018; Ernst et al., 2019) : les gènes exprimés au stade zygotène sont impliqués dans l'organisation de la chromatine et la recombinaison de l'ADN alors que les gènes exprimés au stade pachytène sont spécifiques de la spermiogenèse (Wang et al., 2018).

Comme précisé plus haut, les données d’Hi-C montrent une atténuation voire une déplétion des TADs et des compartiments chromatinien A et B durant la méiose et particulièrement dans les spermatocytes pachytènes ; ces structures sont rétablies par la suite dans les spermatides rondes (Fig. 13) (Alavattam et al., 2019; Patel et al., 2019; Vara et al., 2019; Wang et al., 2019b; Luo et al., 2020; Álvarez-González et al., 2022). La déplétion des compartiments A et B dans les spermatocytes pachytènes se fait au profit de l’émergence d’une structure A-B qui semble liée au processus de compaction de la chromatine essentielle pour l’appariement des chromosomes (Patel et al., 2019; Vara et al., 2019; Wang et al., 2019b; Álvarez-González et al., 2022). Malgré la réorganisation importante qui survient dans les spermatocytes, des analyses ATAC-seq montrent quelques régions ouvertes mais ayant une grande stabilité au niveau des régions promotrices (Fig. 13) (Maezawa et al., 2018b). La fermeture de nombreuses régions de la chromatine semble liée à la présence de H3K9me3 médiée par l’enzyme SETDB1. En effet, le *knock-out* (KO) conditionnel du gène codant pour cette protéine entraîne une augmentation du nombre de régions ouvertes dans les spermatocytes (Cheng et al., 2021). Il entraîne également des changements dans la méthylation de l’ADN des éléments transposables et des altérations dans la dynamique épigénétique durant la spermatogenèse, résultant en une activation prématurée des gènes spécifiques des spermatides (Hirota et al., 2018; Cheng et al., 2021). La réorganisation de la chromatine durant la méiose s’accompagne d’une apparition d’*enhancers* actifs caractérisés par la présence de H3K27ac, qui sont associés à des gènes fortement exprimés dans les spermatides (Sin et al., 2015). Il est intéressant de noter que les *enhancers* et notamment les *super-enhancers* sont importants pour la production de piARN – essentiels à la méiose – dans les spermatocytes pachytènes (Maezawa et al., 2020).

2.3.2 Dynamique transcriptionnelle et chromatinienne des chromosomes sexuels

Chez les mammifères mâles, les chromosomes sexuels sont dits hétérologues, c’est-à-dire qu’ils sont caractérisés par la co-présence d’un chromosome X et d’un chromosome Y, contrairement aux autosomes qui sont homologues. Lors de la division méiotique, les chromosomes sexuels s’apparient et recombinent entre eux au niveau de leurs régions pseudo autosomales, processus qui est indispensable pour la ségrégation méiotique des chromosomes sexuels. Les régions hétérologues quant à elles, ne s’apparient pas et ne recombinent pas. Cette absence de recombinaison lors du passage du stade zygotène à pachytène dans la prophase I, s’accompagne également d’une condensation des chromosomes sexuels et la formation d’un compartiment chromatinien spécifique à la périphérie du noyau, appelé le corpuscule sexuel, « *sex body* » ou « *XY body* » (Fig. 16) [voir Turner, 2007 pour revue]. Ces observations sont confirmées via des approches d’Hi-C qui montre une réorganisation spécifique et isolée des chromosomes sexuels par rapport aux autosomes (Patel et al., 2019; Wang et al., 2019b; Álvarez-González et al., 2022). Cette condensation entraîne, dans le *sex body*, une absence d’expression des gènes des chromosomes X et Y (Fig. 16). Ce phénomène d’inactivation transcriptionnelle des gènes et de remodelage des chromosomes sexuels est appelé le MSChI, pour *Meiotic Sex Chromosome Inactivation* (Handel, 2004; Namekawa et al., 2006; Turner, 2007). C’est un processus

extrêmement bien conservé entre les espèces de mammifères (Murat et al., 2022) qui se caractérise par un arrêt de la transcription des gènes portés par les chromosomes sexuels suivi d'une réactivation dans les spermatozoïdes à la fin de la méiose.

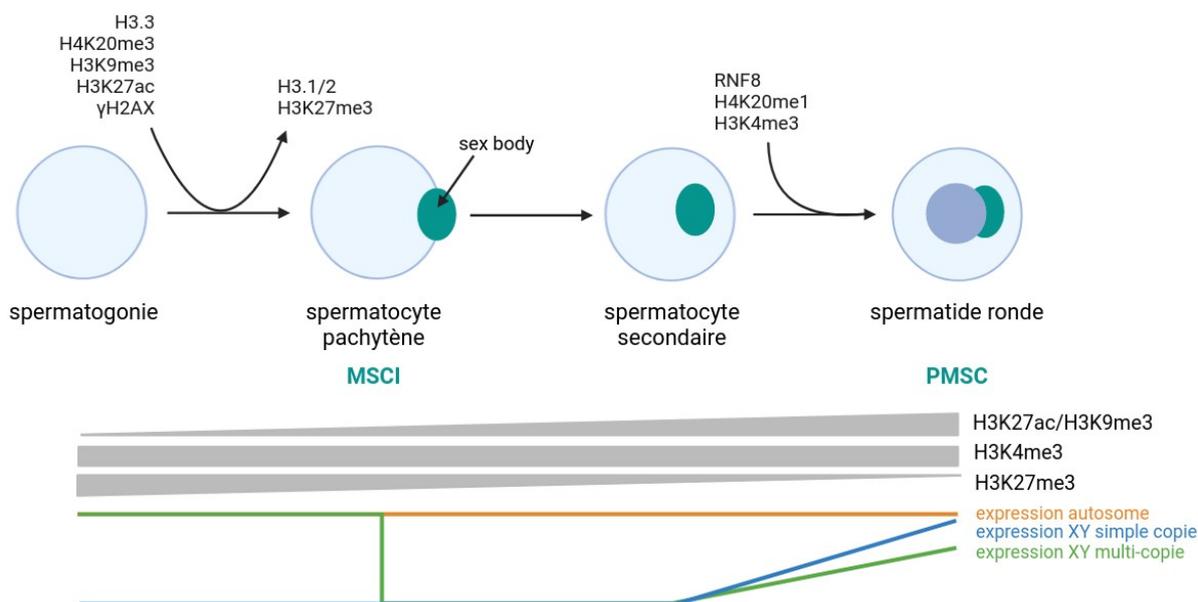


Figure 16 : Modifications transcriptionnelles et chromatiniennees intervenant pendant le processus de MSCI et continuant dans les spermatozoïdes. En gris, le niveau d'enrichissement des PTMs dans les chromosomes sexuels en comparaison des autosomes. Les lignes représentent le niveau d'expression des gènes dans les autosomes (en orange), dans les chromosomes sexuels simples copies (en bleu) et multi-copies (en vert). Figure basée sur les informations de Sin et al, 2015, Moretti et al, 2016 et Ernst et al, 2019 et des articles cités dans cette partie.

Dans la littérature, les changements moléculaires accompagnant le MSCI sont largement étudiés, cependant son rôle dans la lignée germinale mâle des mammifères reste encore mal compris (Turner, 2015). Plusieurs hypothèses ont été émises sur le rôle potentiel du MSCI. Ce processus pourrait permettre la prévention des DSB au niveau des régions non appariés mais également dans la réponse au point de contrôle de la prophase I. À noter que ces deux hypothèses ne sont pas forcément indépendantes l'une de l'autre (Turner, 2015). Un des points qui reste méconnu est de savoir si ces modifications épigénétiques mises en place lors du MSCI sont conservées partiellement dans les spermatozoïdes (cf. sous-partie *Réactivation des chromosomes sexuels après la méiose*) et si elles sont transmises des gamètes mâles à la descendance (Hammoud et al., 2009; Brykczynska et al., 2010; Turner, 2015).

Inactivation des chromosomes sexuels durant la méiose

Le MSCI est initié par la phosphorylation du variant d'histone H2AX (γ H2AX) et s'accompagne d'un remodelage chromatinien important caractérisé notamment par le remplacement des variants d'histones H3.1/2 par le variant H3.3 (Fig. 16). Ce remodelage s'accompagne de la diminution de plusieurs modifications post-traductionnelles comme H3K27me3 et de l'augmentation de H3K27ac et des marques d'hétérochromatine

H3K9me2/3 et H4K20me3 (Black et al., 2012; Sin et al., 2015; Maezawa et al., 2018b) ; H3K9me3 est particulièrement abondante sur les chromosomes sexuels lors du MSCI (Ernst et al., 2019) (Fig 16). L'enrichissement en H3K9me3 au niveau des gènes du chromosome X augmente à partir du stade pachytène et reste élevé dans les spermatides rondes. Il est important de noter que, dans ces études, le chromosome Y n'est pas regardé (Wang et al., 2018; Ernst et al., 2019; Liu et al., 2019). L'enrichissement en H3K9me3 corrèle avec la sur-expression de sa méthyltransférase SETDB1 au stade pachytène. Le rôle de la marque H3K9me3 dans le MSCI reste encore à déterminer et deux modèles s'opposent sur son implication dans l'initiation du MSCI. Dans leur article, Hirota et al., décrivent que les protéines de la voie de réponse au dommage à l'ADN (DDR pour *DNA damage response*), dont γ H2AX, recrutent la méthyltransférase SETDB1 sur les chromosomes sexuels ce qui entraîne l'acquisition de la marque H3K9me3 et l'inactivation transcriptionnelle des gènes de l'X et de l'Y (Hirota et al., 2018). La délétion de *Setdb1*, par KO conditionnel chez la souris, entraînant une perturbation du processus du MSCI (Hirota et al., 2018), les auteurs concluent que H3K9me3 est nécessaire au processus d'initiation du MSCI (Fig. 17). De leur côté, en utilisant un modèle ex-vivo de culture de cellules germinales pendant un temps limité, Abe et al., montrent que l'initiation du MSCI n'est pas médiée par H3K9me3 mais par les facteurs de la DDR (Abe et al., 2022).

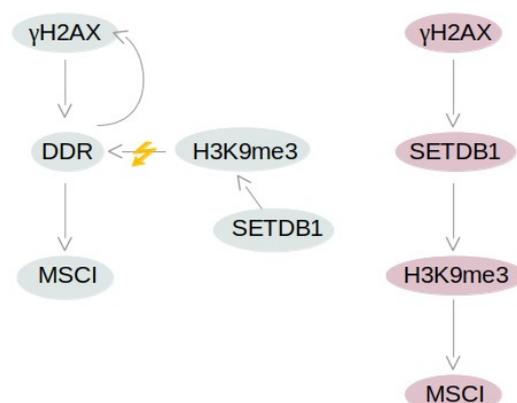


Figure 17 : Schéma du rôle de H3K9me3 dans l'initiation du MSCI. A gauche (en vert) l'hypothèse émise par Abe et al, 2022 et à droite (en rose) l'hypothèse émise par Hirota et al, 2018

Réactivation des chromosomes sexuels après la méiose

Après la méiose, de nombreux gènes portés par les chromosomes sexuels sont ré-activés, ou exprimés *de novo*, représentant 60 % des gènes de l'X et la quasi-totalité des gènes de l'Y chez la souris (Moretti et al., 2016). Cependant, la chromatine de l'X et de l'Y se distingue toujours de celle des autosomes et est appelée PMSC, pour *Post Meiotic Sex Chromatin* (Namekawa et al., 2006; Turner, 2007; Sin et al., 2015; Moretti et al., 2016; Ernst et al., 2019). Après la méiose, la chromatine sexuelle se caractérise par la conservation de certaines marques d'histones mises en place lors du MSCI (Fig. 16). Par comparaison avec le niveau dans les autosomes, les marques H3K27ac et H3K9me3 sont fortement enrichies dans les chromosomes sexuels et plus particulièrement sur l'X pour H3K9me3 (Ernst et al., 2019; Liu et al., 2019) et sur l'Y pour H3K27ac (Moretti et al., 2016). Au contraire, la marque

H3K27me3 reste très faible dans les chromosomes sexuels par rapport aux autosomes (Sin et al., 2015).

Après la méiose, l'expression des gènes portés par les chromosomes sexuels est également co-régulée par un processus spécifique, qui implique différentes protéines comme RNF8 ou SLY (Cocquet et al., 2009; Sin et al., 2012). En effet, RNF8 est une protéine permettant l'établissement de H3K4me3 et de H4K20me1 dans les chromosomes XY durant la méiose et qui permet l'activation des gènes de la PMSC (Sin et al., 2012). La sortie de méiose s'accompagne en effet d'une (ré)activation de certains gènes portés par les chromosomes sexuels et, en particulier, des gènes présents en multi-copies (Fig. 16) (Moretti et al., 2016, 2020). Ces gènes sont caractérisés par la présence de modifications d'histones, comme H3K4me3 ou H3K79me2, ainsi que l'acétylation et la crotonylation d'H3K27 (Moretti et al., 2017; Crespo et al., 2020). Ces gènes multi-copies sont difficiles à analyser du fait de la nature répétée de leur séquence. De plus, ils sont généralement faiblement exprimés individuellement, mais collectivement les copies d'un même gène ont un niveau d'expression élevé (Moretti et al., 2016). Cette accumulation (amplification) de gènes multi-copies sur les chromosomes sexuels est présumée être une conséquence du MSC1 : afin d'échapper à l'environnement chromatinien répressif mis en place lors du MSC1, des gènes essentiels à la spermiogenèse (qui sont portés par l'X et l'Y) deviennent multi-copies (Mueller et al., 2008). Ce phénomène est conservé au cours de l'évolution des mammifères bien que les gènes portés par les chromosomes sexuels (et en particulier par l'Y) évoluent rapidement (Murat et al., 2022). Parmi les gènes multi-copies portés par l'Y chez la souris, le gène *Sly* est exprimé spécifiquement dans les spermatides rondes et est indispensable à la spermatogenèse (Cocquet et al., 2009; Moretti et al., 2017). En effet, l'inhibition de *Sly* entraîne une dérégulation de plusieurs centaines de gènes portés par les chromosomes sexuels (Moretti et al., 2017). Elle entraîne également la dérégulation de quelques gènes autosomaux, parmi lesquels le gène *Dot1l* qui code pour une histone méthyltransférase de H3K79 (Moretti et al., 2017). Le rôle complet de DOT1L durant la spermatogenèse sera précisément décrit ultérieurement (cf. 3.2.2 *DOT1L dans la spermatogenèse*).

2.4 Lors de la spermiogenèse

2.4.1 Déroulement de la spermiogenèse

La spermiogenèse (qui dure ≈ 23 jours chez l'homme et 12 jours chez la souris) est la dernière phase de la spermatogenèse au cours de laquelle les cellules haploïdes rondes que sont les spermatides se différencient en spermatozoïdes (Fig. 18). La spermiogenèse est un processus de spécialisation cellulaire « extrême » caractérisée par l'acquisition de nouvelles structures (flagelle et acrosome) et l'élimination de la majorité du cytoplasme. Ce processus est également caractérisé par une dynamique transcriptionnelle et chromatinienne très importantes faisant intervenir plusieurs centaines de gènes et de nombreuses PTMs, aboutissant ainsi à une réorganisation importante de la chromatine puis à sa compaction extrême. Tous ces événements font que la spermiogenèse a un programme génétique et épigénétique très dynamique.

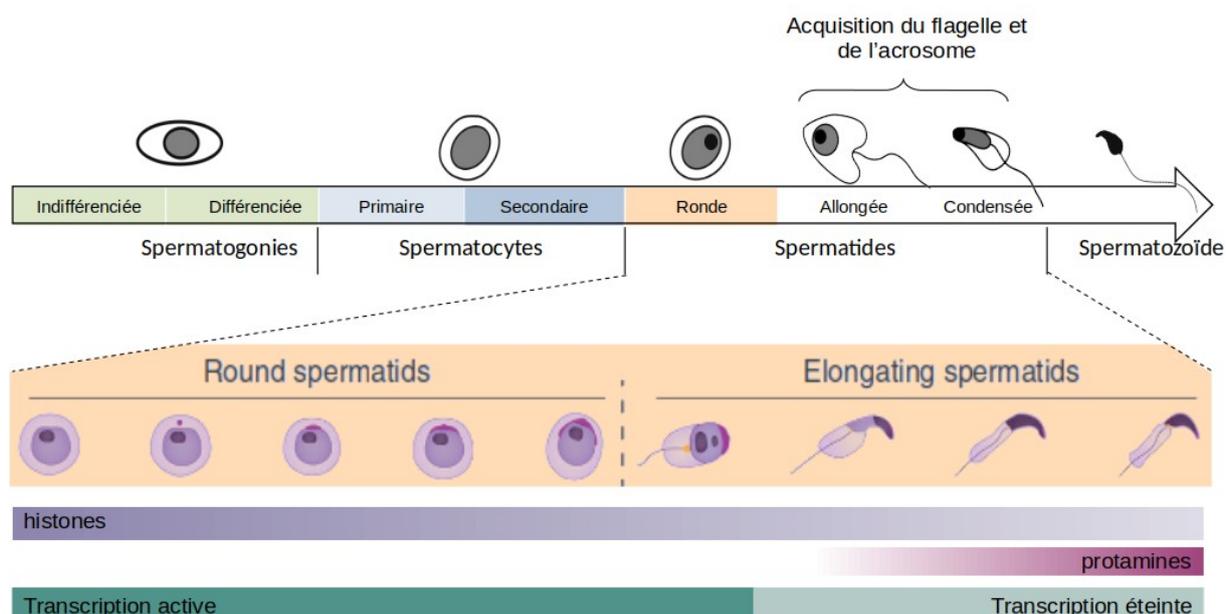


Figure 18 : Étapes de la spermiogenèse chez la souris et sa dynamique transcriptionnelle et chromatinienne. Figure adaptée d'Ernst et al, 2019

2.4.2 Régulation chromatinienne de la transcription dans les spermatides ronds

Les gènes exprimés spécifiquement dans les spermatides ronds contribuent fortement à la grande diversité/complexité transcriptionnelle observée dans le testicule (Soumillon et al., 2013). Une étude de l'évolution du transcriptome des cellules germinales dans 10 espèces de mammifères et une espèce aviaire a montré récemment que, contrairement aux spermatogonies et aux spermatocytes dont les gènes exprimés sont très conservés entre les espèces, les gènes exprimés spécifiquement dans les spermatides ronds évoluent rapidement (Murat et al., 2022). Les auteurs suggèrent que la compétition entre spermatozoïdes est à l'origine de ce phénomène.

Dans les spermatides ronds, la configuration de la chromatine revient à sa configuration initiale (Wang et al., 2019b; Luo et al., 2020; Álvarez-González et al., 2022). La structure A-B présente dans les spermatocytes disparaît pour reformer les structures initiales que sont les compartiments A et B (Wang et al., 2019b) ; de la même façon, les TADs qui avaient disparu pendant la méiose, réapparaissent (Wang et al., 2019b; Luo et al., 2020; Álvarez-González et al., 2022) (Fig. 13). La reformation des TADs corrèle avec l'augmentation du nombre d'*enhancers* et de *super-enhancers* dans les spermatides ronds en comparaison avec les spermatocytes (≈ 500 dans les spermatocytes et ≈ 1100 dans les spermatides ronds) (Maezawa et al., 2020) ; ceux-ci sont maintenus jusque dans les spermatozoïdes (Jung et al., 2017). L'identification de ces *enhancers* est faite par la présence de H3K27ac (Maezawa et al., 2020) et ces *enhancers* sont le plus souvent co-marqués par H3K27cr dans les spermatides ronds (Crespo et al., 2020). Les analyses protéomiques montrent que H3K27cr est fortement associée à H3K36me2 (Crespo et al., 2020)c, précédemment décrite comme caractéristique des *enhancers* et des *super-enhancers* (Soldi et al., 2017). L'augmentation de H3K27cr et majoritairement de H3K27ac au niveau des

enhancers et des promoteurs est proposée comme étant essentielle à la différenciation des spermatides lors de la spermiogenèse. En effet, l'augmentation du nombre de *super-enhancers* et d'*enhancers* induit une augmentation de la transcription dans les spermatides (Maezawa et al., 2020; Sakashita et al., 2020). Deux types d'*enhancers* sont mis en évidence : les *super-enhancers* qui activent des gènes nécessaires pour la spermatogenèse et des *enhancers* dérivant de certaines séquences de rétrovirus endogènes qui acquièrent une fonction d'*enhancers* activant la transcription de gènes qui régulent la spermatogenèse (Maezawa et al., 2020; Sakashita et al., 2020).

La première moitié de la spermiogenèse s'accompagne également d'une régulation post-transcriptionnelle, des piRNA et des structures localisées dans le cytoplasme des cellules germinales mâles appelées « *chromatoid body* », dont les détails ne seront pas développés dans ce manuscrit.

2.4.3 Remplacement des histones par les protamines à partir du stade spermatides allongées

À partir de la 2^e moitié de la spermiogenèse, les spermatides allongées subissent un remodelage important de leur chromatine au cours duquel la quasi-totalité des histones canoniques est remplacée par des variants d'histones et des protéines de transition, qui vont guider l'incorporation des protamines (Barral et al., 2017). Ceci entraîne une réorganisation totale de la chromatine sous la forme de toroïdes, avec seuls quelques solénoïdes d'histones maintenus, et une hyper-compaction de la chromatine. Ce processus est accompagné par une extinction de la transcription. Il est indispensable au maintien de l'intégrité génomique puisqu'il permet de protéger le génome paternel d'éventuels stress extérieurs ; il contribue également à la forme hydrodynamique de la tête des spermatozoïdes (Hazzouri et al., 2000; Govin et al., 2007; Rathke et al., 2014; Bao and Bedford, 2016). Un mauvais remplacement des histones par les protamines entraîne une infertilité masculine (Blanco and Cocquet, 2019).

Côté transcriptionnel, le transcriptome des spermatides allongées est très proche de celui des spermatides rondes – les transcrits sont conservés (Moretti et al., 2016; Chen et al., 2018a), jusqu'à extinction de la transcription et élimination d'une grande partie du cytoplasme dans les spermatides condensées.

Les variants d'histones et les modifications post-traductionnelles nécessaires lors de la transition histones-protamines sont très nombreux. Plusieurs revues décrivent l'ensemble des événements et le rôle de chacun dans cette étape de transition (Fig. 19) (Bao and Bedford, 2016; Wang et al., 2020). Le rôle de nombreuses d'entre elles – en particulier l'hyperacétylation des histones – est de faciliter l'ouverture de la chromatine et l'enlèvement des histones. Cette chromatine « permissive » est supposée contribuer à la diversité transcriptomique observée lors de la spermiogenèse (Soumillon et al., 2013; Murat et al., 2022). L'hyperacétylation des histones est le changement le plus caractéristique et le plus étudié de ce processus (Hazzouri et al., 2000; Govin et al., 2007; Rathke et al., 2014; Luense et al., 2019) et intervient sur chacune des histones : H2AK5/9/13/15,

H2BK5/12/15/20/23/24, H3K4/9/14/18/27/36/56/79/122 et H4K5/8/12/16/20/79/91 (Fig. 19) (Barbero et al., 2023). L'hyperacétylation des histones est facilitée par la présence de nombreux facteurs établis tout au long de la spermatogenèse et ceux spécifiques des spermatides comme GCN5, Nut ou CDH5 (Li et al., 2014; Shiota et al., 2018; Luense et al., 2019).

En 2014, Dottermusch-Heidel et al., décrivent que l'hyperacétylation de l'histone H4 s'accompagne d'une augmentation de la méthylation de H3K79, au moment de la transition histones-protamines (Dottermusch-Heidel et al., 2014a, 2014b). Cette observation corrèle avec l'augmentation de l'expression de DOT1L, méthyltransferase de H3K79, à la fin de la méiose et dans les spermatides (Moretti et al., 2017). Comme décrit plus haut, *Dot1l* est l'un des gènes cibles de *Sly*, ainsi la délétion de *SLY* entraîne une diminution de l'expression de *Dot1l* et de la méthylation de H3K79 ainsi qu'une compaction anormale du noyau entraînant une rétention des histones (Moretti et al., 2017). Le rôle complet de DOT1L durant la spermatogenèse sera décrit ultérieurement (cf. 3.2.2 *DOT1L dans la spermatogenèse*).

Durant la spermiogenèse, le retrait des histones n'est pas total. En effet, chez la souris $\approx 5\%$ des histones sont retenus dans les spermatozoïdes, et 15% chez l'humain (Yamaguchi et al., 2018). De ce fait, beaucoup d'histones acétylées disparaissent vers la fin de la spermiogenèse (états tardifs des spermatides condensées), excepté les marques acétylées H4K5/8/12ac et H3K9ac (Hazzouri et al., 2000; Govin et al., 2007; Steilmann et al., 2011; Bell et al., 2014; Luense et al., 2016). Une perte de l'acétylation dans les spermatides allongées entraîne un défaut de condensation dans les spermatozoïdes (Bell et al., 2014). La crotonylation de H3K27 fait également partie des marques qui sont conservées dans les spermatozoïdes (Crespo et al., 2020).

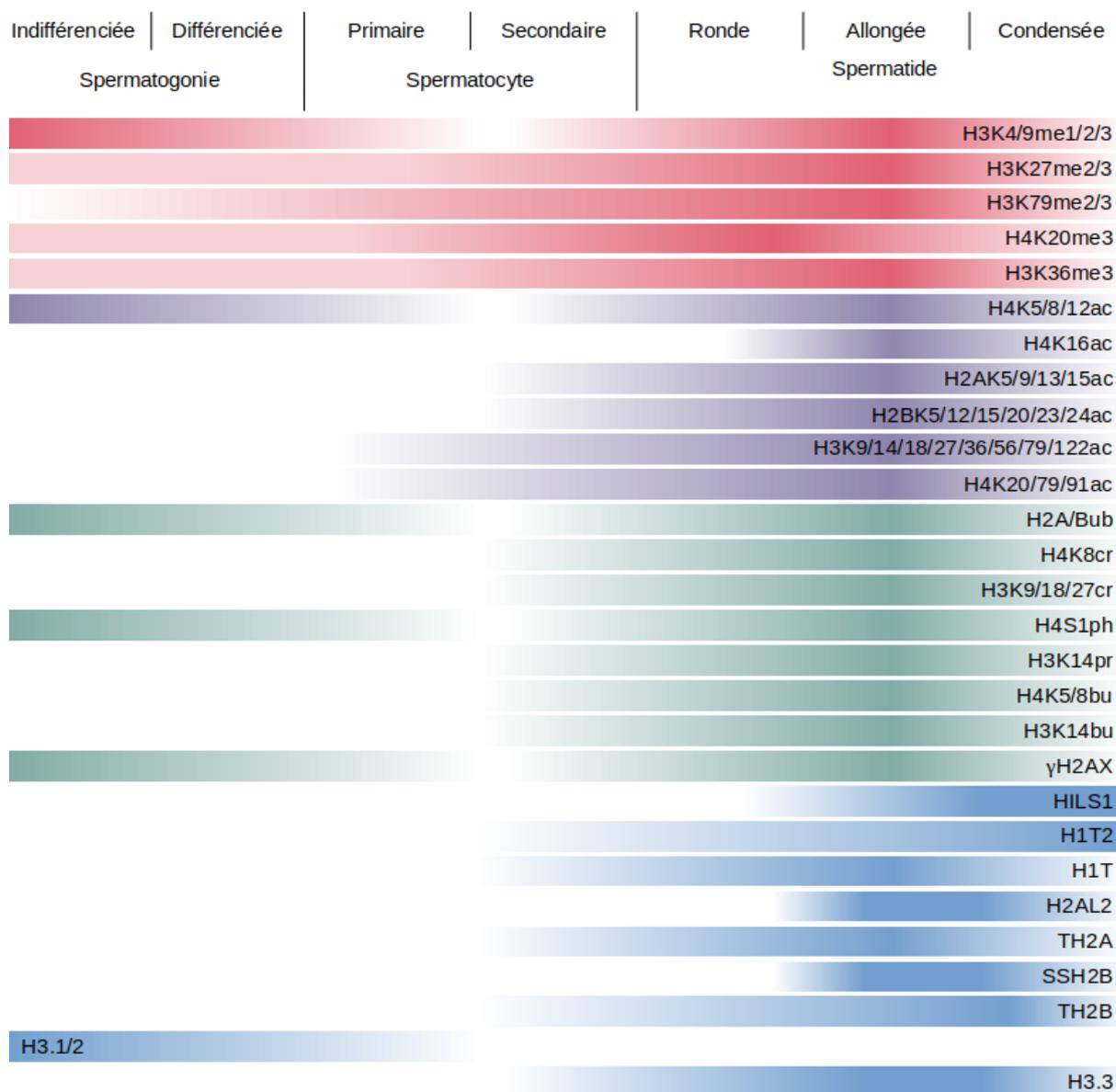


Figure 19 : Schéma représentant les modifications d'histones s'opérant au cours de la spermatogenèse. En rose sont représentés les méthylation d'histone, en violet les acétylation d'histone, en vert les autres modifications d'histone et en bleu les variants d'histones. Figure produite à partir des données de Wang et al, 2019a et des autres publications citées dans cette partie.

2.5 L'épigénome du spermatozoïde

Le spermatozoïde, le produit final de la spermatogenèse, véhicule l'information génétique (son ADN) mais également épigénétique, transmises à la descendance. Le programme épigénétique des spermatozoïdes, également appelée « épigénome », est établi tout au long de la spermatogenèse et son bon déroulement, en particulier la compaction de la chromatine lors de la spermiogenèse, est indispensable pour le maintien de l'intégrité du matériel génétique et sa transmission après la fécondation (Hazzouri et al., 2000; Govin et al., 2007; Rathke et al., 2014; Bao and Bedford, 2016). Par conséquent, des altérations de la mise en place de l'épigénome des spermatozoïdes sont souvent associées à une infertilité

masculine [voir Blanco and Cocquet, 2019 et Moritz and Hammoud, 2022 pour revue]. Par exemple, Luense et al., décrivent que la perte de l'histone acétyltransférase GNC5 dans les cellules germinales mâles entraîne un développement spermatique anormal et une infertilité masculine (Luense et al., 2019).

Bien qu'il soit retrouvé chez tous les vertébrés, le spermatozoïde montre des divergences évolutives importantes caractérisées notamment par des différences morphologiques (Temple-Smith et al., 2018; Murat et al., 2022). La structure générale du spermatozoïde est cependant extrêmement bien conservée, composée de la tête qui contient l'information génétique compactée dans le noyau, du flagelle permettant la mobilité du spermatozoïde et d'une pièce de connexion permettant d'attacher la tête à la queue du spermatozoïde (Temple-Smith et al., 2018) (Fig. 20). Bien que l'information génétique soit retrouvée exclusivement dans le noyau, l'information épigénétique, en particulier l'expression des ARNs non codants, peut être contenue dans l'ensemble du spermatozoïde (Santiago et al., 2021).

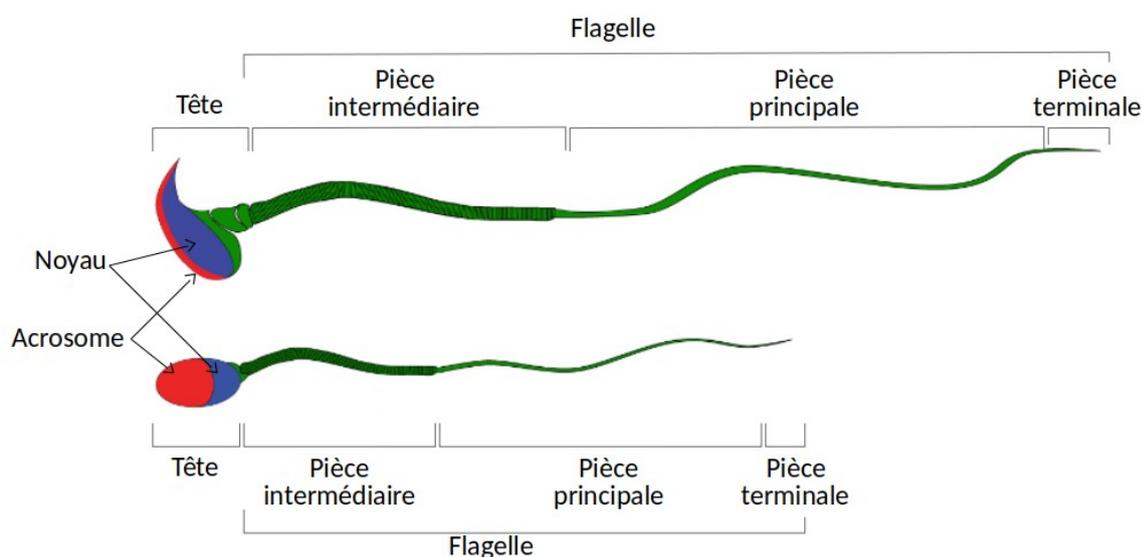


Figure 20 : Structure du spermatozoïde de la souris (en haut) et de l'humain (en bas). Figure adaptée de Linn et al., 2021

L'épigénome du spermatozoïde consiste en trois grands types de modifications : la méthylation de l'ADN, les modifications post-traductionnelles des histones et les ARN non-codants (Champroux et al., 2018; Lismer and Kimmins, 2023). Ces modifications ne sont pas nécessairement indépendantes puisqu'il existe des interactions entre ces éléments (Lambrot et al., 2021; Lismer and Kimmins, 2023). Ces trois grands types de modifications sont retrouvés dans toutes les cellules de l'organisme, cependant l'épigénome du spermatozoïde est unique et très différent des autres cellules.

2.5.1 La transcription des ARN non-codants

La transcription étant quasiment éteinte dans les spermatozoïdes, le transcriptome détecté dans les spermatozoïdes est en majorité le résidu du transcriptome des spermatides

rondes et allongées (Johnson et al., 2011) (Fig. 10). Malgré la faible quantité d'ARNm détectée dans les spermatozoïdes, il est observé une abondance de petits ARNs non codants (sncRNA), suggérant que ceux-ci jouent un rôle régulateur y compris après la fécondation (Erkek et al., 2013; Hammoud et al., 2014). Plusieurs récentes études ont, en effet, démontré que certains sncRNA issus des spermatozoïdes peuvent avoir un impact sur la fécondation, le développement embryonnaire et/ou la santé de la descendance [pour revues voir Zhang et al., 2019, Sharma, 2019 ou Santiago et al., 2021]. Parmi les sncRNA qui sont présents dans les spermatozoïdes, les principales catégories mises en évidence sont les miRNA, les piRNA et les tsRNA (Table. II) (Sharma et al., 2016; Santiago et al., 2021). L'utilisation de données de RNA-seq permet de montrer que les miRNA sont hautement représentés dans les spermatozoïdes au début de l'épididyme et décroissent durant le trajet dans l'épididyme contrairement aux piRNA qui ont un profil opposé (Pantano et al., 2015; Hutcheon et al., 2017). Les études des miRNA chez la souris montrent que les miRNA régulent la spermatogenèse, en effet, la dérégulation de l'expression des miRNA entraîne un arrêt de la spermatogenèse et des défauts des spermatozoïdes (Yao et al., 2015).

	ARNm	ARNlnc	miARN	piARN	tsARN
Taille	Variable	200-10000nt	20-24nt	24-30nt	18-40nt
Fonction	remodelage chromatinien ; traduction <i>de novo</i>	formation d'hétérochromatine ; méthylation de l'ADN ; régulation transcriptionnelle et post-transcriptionnelle	contrôle de la traduction ; dégradation de l'ARN	répression des rétrotransposons ; contrôle de la traduction ; dégradation de l'ARN	régulation, inactivation et stabilité de l'ARNm ; modification de l'ARN ; liaison des protéines
Expression	Ubiquitaire	Ubiquitaire	Exprimé dans la majorité des tissus	Majoritairement exprimé dans la spermatogenèse (spermatocyte, spermatide)	Très exprimé dans le spermatozoïde
Localisation dans le spermatozoïde	Tête, pièce intermédiaire et queue	Tête (compartiment nucléaire et péri-nucléaire)	Tête (compartiment nucléaire et péri-nucléaire) et queue	Queue	Tête
Fonction dans le spermatozoïde	fécondation ; transmission de l'ARN du spermatozoïde aux stades embryonnaires précoces	héritage transgénérationnel ; régulation de l'expression génique ; modification d'histone ; développement embryonnaire précoce	remplacement des histones dans les stades embryonnaires ; embryogenèse précoce ; modification épigénétique ; maturation des spermatozoïdes	protection des transposons ; modification épigénétique	modification épigénétique ; héritage transgénérationnel ; développement embryonnaire précoce

Table II : Résumé des types, caractéristiques et rôles des types d'ARN les plus étudiés dans le spermatozoïde. Traduction d'un tableau publié dans Santiago et al., 2021

Le croisement de données de CHIP-seq de plusieurs marques d'histones avec différents types de sncRNA indiquent que l'expression de ces sncRNA est régulée au niveau chromatinien (Lismer and Kimmins, 2023). Parmi les sncRNA, les tsRNA (ARNt dérivé de petits ARNs) semblent jouer un rôle très important dans le spermatozoïde, car ils

représentent 65 % des sncRNA présents dans les spermatozoïdes, ils sont plus enrichis que les miRNA et sont plus abondants dans la tête que dans la queue des spermatozoïdes (Chen et al., 2016; Hua et al., 2019). De plus, le RNA-seq des tsRNAs montre que ceux-ci ont un rôle dans des fonctions essentielles pour le spermatozoïde comme la stabilité, la régulation et l'inactivation de l'ARNm ou la prolifération cellulaire (Table. II) (Chen et al., 2020a). L'analyse des tsRNA dans le spermatozoïde montre que ceux-ci sont nécessaires pour le développement embryonnaire précoce et la transmission transgénérationnelle (Chen et al., 2016; Sharma et al., 2016).

Des études par microfluidique à l'échelle de la cellule unique sur plusieurs individus montrent une hétérogénéité des miARN exprimés dans les spermatozoïdes entre les différents individus (Yu et al., 2017). Ces divergences entre individus et le long de l'épididyme suggèrent un haut niveau de complexité et rendent difficile leur analyse.

2.5.2 L'accessibilité de la chromatine et les modifications post-traductionnelles

Dans les spermatozoïdes, $\approx 5\%$ et 15% des histones ne sont pas remplacées par les protamines durant la spermiogenèse respectivement chez la souris et chez l'humain, et sont donc conservées dans les spermatozoïdes (Yoshida et al., 2018). Cette rétention des histones dans les spermatozoïdes permet la régulation de gènes important pour le développement embryonnaire (Hammoud et al., 2009; Lesch et al., 2016; Lismer et al., 2021). (La Spina et al., 2014)

La localisation génomique de ces histones retenues a été discutée dans une dizaine d'études publiées depuis 2009 et fait encore débat. Plusieurs études ont observé une rétention des histones dans les régions hétérochromatiniennes (régions pauvres en gènes, et riches en éléments répétés) (Carone et al., 2014; Yamaguchi et al., 2018) en accord avec les observations faites par immunofluorescence (Govin et al., 2007). Certaines histones modifiées comme H3K9me3 semblent avoir une localisation assez similaire (enrichies dans les régions pauvres en gènes, en particulier les régions centromériques), tandis que d'autres marques, comme H3K4me3 et H3K27me3, sont retenues au promoteur de gènes impliqués dans la spermatogenèse et dans le développement embryonnaire précoce. Les analyses par ChIP-seq pour H3K4me3 dans les spermatozoïdes montrent que les histones sont particulièrement enrichies au niveau des promoteurs des gènes (Hammoud et al., 2009; Brykczynska et al., 2010; Erkek et al., 2013). Dans leur article, Yamaguchi et al., supposent que l'enrichissement de H3K4me3 dans les promoteurs est un phénomène universel dans les spermatozoïdes (Yamaguchi et al., 2018). En effet, ces résultats sont similaires à ceux trouvés dans d'autres études (Hammoud et al., 2009; Brykczynska et al., 2010; Erkek et al., 2013). *A contrario*, Carone et al., montrent, par analyse ChIP-seq, que les histones des spermatozoïdes sont généralement appauvries au niveau des promoteurs, y compris ceux des gènes du développement (Carone et al., 2014).

De plus, la co-présence de H3K4me3 et H3K4ac dans les spermatozoïdes est supposée permettre le maintien de l'état ouvert de la chromatine pour quelques régions (Godmann et al., 2007; Rathke et al., 2014; Sin et al., 2015; Maezawa et al., 2018b). Ces

observations s'expliquent par un changement de la structure chromatinienne du spermatozoïde durant sa maturation lors de son passage dans l'épididyme (Bedi et al., 2022). Ces modifications structurelles se localisent au niveau des promoteurs des gènes ou au niveau des *enhancers* (Bedi et al., 2022). Au niveau des *enhancers*, les *super-enhancers* présents dans les spermatozoïdes sont initialement présents dans les spermatides rondes (Crespo et al., 2020).

En raison de la compaction très importante de la chromatine dans le spermatozoïde et en particulier chez la souris, il est très difficile de mettre en place un protocole de ChIP-seq permettant d'obtenir le niveau d'enrichissement des marques d'histones. Très récemment, une étude pré-publiée, a comparé les quatre méthodes les plus couramment utilisées – ChIP-seq (lyse par onde sonore (sonication) ou traitement par MNase), CUT&RUN et CUT&TAG – dans le spermatozoïde entier et retrouvent les résultats des études précédentes (Yin et al., 2023). Ces observations supposent que les résultats des études précédentes (Erkek et al., 2013; Jung et al., 2017) pourraient être contaminés par des cellules somatiques et donc que ces résultats ne refléteraient pas le niveau d'enrichissement des histones dans le noyau du spermatozoïde.

2.5.3 La méthylation de l'ADN

La méthylation de l'ADN est nécessaire à la fécondation et au développement embryonnaire. En effet, des changements de méthylation de l'ADN dans les spermatozoïdes sont associés à une dérégulation transcriptionnelle des gènes de développement embryonnaire (Milekic et al., 2015; Greenberg and Bourc'his, 2019). Les études du méthylome par WGBS et BS-seq montrent que le génome du spermatozoïde est hautement différenciellement méthylé avec environ 70 % des îlots CpG qui sont méthylés (Molaro et al., 2011; Shea et al., 2015; Lismer and Kimmins, 2023). Dans le spermatozoïde, la méthylation de l'ADN est unique par rapport aux cellules somatiques, en effet, les îlots CpG sont principalement méthylés dans le corps des gènes et au niveau des éléments répétés contrairement aux autres types cellulaires qui sont méthylés dans les promoteurs (Lambrot et al., 2021; Lismer and Kimmins, 2023).

La comparaison de la méthylation entre 7 espèces de mammifères montre une forte divergence des régions méthylées au cours de l'évolution, principalement pour les régions qui ne sont pas localisées dans les promoteurs. En effet, le nombre et la taille des régions hypométhylées dans le spermatozoïde augmente au cours de l'évolution (Qu et al., 2018) (Fig. 21). Les régions hypométhylées conservées entre l'homme et la souris sont présumées importantes pour les processus biologiques intervenant durant le développement embryonnaire (Fang et al., 2019) car elles sont localisées au niveau des promoteurs et des *enhancers* de gènes importants pour ces processus (Jones, 2012). Les régions hyperméthylées qui sont conservées entre les espèces sont associées à des gènes permettant l'extinction de l'immunité après la fécondation permettant l'adhérence et l'implantation de l'embryon dans l'utérus (Fang et al., 2019).

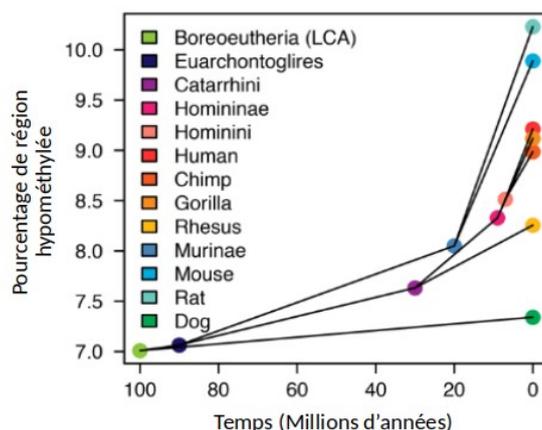


Figure 21 : Évolution du pourcentage de régions hypométhylées dans le génome des mammifères au cours de l'évolution. Figure adaptée de Qu et al., 2018

L'analyse conjointe de données de WGBS et de ChIP-seq pour la marque H3K4me3 montre que les régions enrichies en H3K4me3 pourraient être associées à un niveau de méthylation de l'ADN intermédiaire voire élevé (Lambrot et al., 2021). Ces observations supposent que la méthylation de l'ADN permettrait la régulation du niveau de H3K4me1/3 au niveau des promoteurs et les *enhancers*, pour potentiellement réguler l'expression des gènes permettant le développement embryonnaire (Lismer and Kimmins, 2023).

3 DOT1L & la méthylation de la marque H3K79

3.1 DOT1/DOT1L, méthyltransférase de H3K79

DOT1 (*Disruptor of telomeric silencing-1*) est une protéine initialement découverte chez la levure *Saccharomyces cerevisiae* (Singer et al., 1998) connue sous le nom de DOT1L chez les mammifères. Elle est très conservée dans les espèces eucaryotes comme *Arabidopsis thaliana*, plusieurs espèces de mycètes, la drosophile et les mammifères (van Leeuwen et al., 2002; Zhang et al., 2007; Frederiks et al., 2008; Dottermusch-Heidel et al., 2014b; Kim et al., 2014).

DOT1L est la seule méthyltransférase capable de métyler la lysine en position 79 de l'histone 3 (H3K79). En effet, la délétion du gène *Dot1l* entraîne une perte de la méthylation de H3K79 chez la levure, la drosophile, la souris et l'humain (van Leeuwen et al., 2002; Shanower et al., 2005; Jones et al., 2008; Deshpande et al., 2013; Vlaming and van Leeuwen, 2016; Wood et al., 2018; Feng et al., 2022). Cette modification post-traductionnelle existe sous trois formes : mono-, di- et tri-méthylée, catalysées par DOT1 chez la levure / DOT1L chez les mammifères (van Leeuwen et al., 2002; Frederiks et al., 2008; Kim et al., 2014; Feng et al., 2022). Chez la levure, H3K79 est retrouvée à 90 % sous une forme méthylée avec H3K79me3 comme forme principale (van Leeuwen et al., 2002), contrairement aux mammifères où H3K79 est principalement sous une forme non méthylée, avec H3K79me1 comme forme méthylée la plus abondante (6 à 30 %) (Vlaming and van Leeuwen, 2016). Les trois formes méthylées de H3K79 sont des marques étalées caractérisées par un pic

d'enrichissement juste après le site d'initiation de la transcription TSS qui décroît progressivement le long du corps du gène (Marson et al., 2008; Huff et al., 2010).

3.1.1 Rôle biologique de DOT1L et de H3K79me

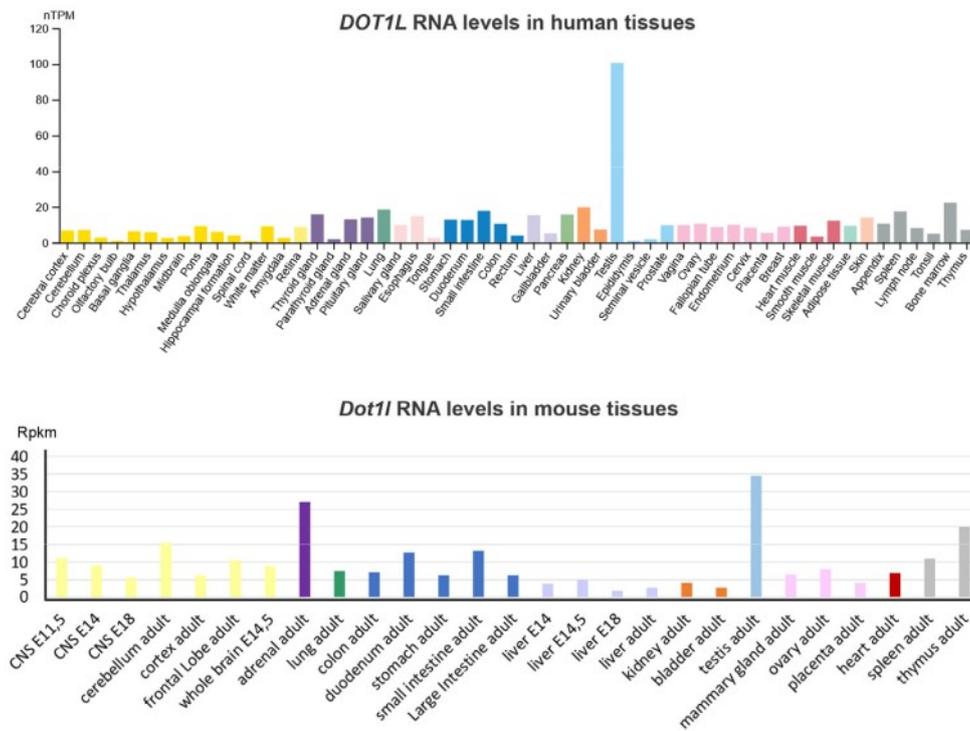


Figure 22 : En haut, expression en ARN (TPM) de *Dot1l* dans 55 tissus humains classés par groupes de tissus. Figure provenant de Human Protein Atlas. En bas, expression en ARN (RPKM) de *Dot1l* chez la souris dans plusieurs tissus différents. En bleu clair sont présentés les tissus de l'appareil reproducteur mâle (Blanco et al., 2023)

DOT1L est essentielle pour de nombreuses fonctions biologiques comme la régulation de la transcription, la réponse aux dommages à l'ADN, le cycle ou la prolifération cellulaire (Jones et al., 2008; Nguyen et al., 2011; Wood et al., 2018; Feng et al., 2022; Wille and Sridharan, 2022). De par son implication dans plusieurs fonctions biologiques, DOT1L est une protéine ubiquitaire exprimée dans une majorité de types cellulaires chez l'humain et chez la souris (Fig. 22). L'analyse de l'expression de *Dot1l* indique qu'à l'âge adulte, il est fortement exprimé dans le testicule et plus particulièrement dans les spermatoocytes et les spermatoïdes (Dottermusch-Heidel et al., 2014a, 2014b; Moretti et al., 2017). Ces résultats suggèrent un rôle de DOT1L dans le processus de spermatogenèse, rôle qui sera décrit plus précisément ultérieurement.

Chez la souris, le KO constitutif de *Dot1l* entraîne une létalité de l'embryon entre 9,5 et 10,5 jours après fécondation dû à des défauts de formation du sac vitellin, de croissance et du cœur (Jones et al., 2008). Pour contourner la létalité embryonnaire, de nombreuses études ont caractérisé le rôle de *Dot1l* en produisant un KO conditionnel induit dans un type cellulaire ou un tissu particulier (Fig. 23) [voir Wille and Sridharan, 2022 pour revue]. Ces études montrent l'implication de DOT1L dans une multitude de fonctions biologiques,

en particulier dans la différenciation cellulaire, comme celle des lymphocytes T et B, des cardiomyocytes ou des cellules germinales mâles. À titre d'exemples, lors de la différenciation des lymphocytes le KO de *Dot1l* entraîne une différenciation précoce des cellules T (Kwesi-Maliepaard et al., 2020) et un arrêt dans les cellules B (Aslam et al., 2021).

De nombreuses équipes ont également étudié le rôle de DOT1L dans les leucémies type MLL (*Mixed Lineage Leukemia*). Leurs travaux montrent que DOT1L est une potentielle cible thérapeutique pour le traitement de ces leucémies (Mueller et al., 2007; Biswas et al., 2011; Nguyen et al., 2011). Le mécanisme par lequel DOT1L est impliqué dans les leucémies MLL est développé plus bas (cf. 3.3.1 *Rôle dans l'élongation de la transcription*).

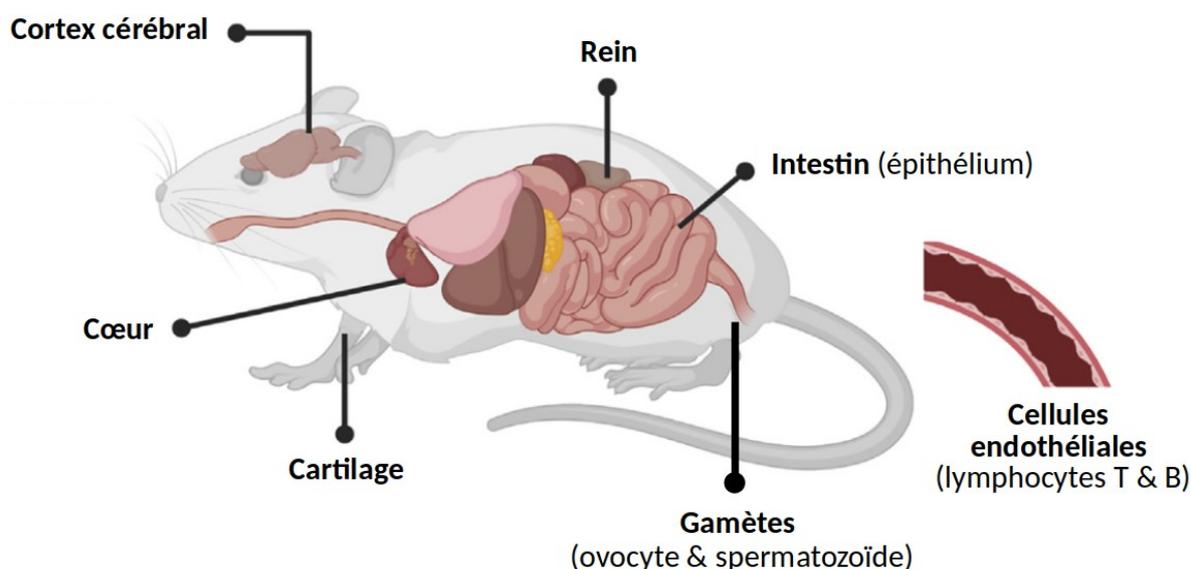


Figure 23 : Schéma des organes dans lesquels le KO de *Dot1l* a été étudié. Figure adaptée de Wille and Sridharan, 2022

3.1.2 DOT1L et H3K79 dans l'embryogenèse

Durant le développement pré-implantatoire des mammifères, le transcrit *Dot1l* est détecté au stade 1-cellule, avant l'activation du génome zygotique (Fig. 24). Puis son expression est diminuée aux stades 2 et 4 cellules avant d'augmenter progressivement jusqu'au stade blastocyste (Ooga et al., 2013). H3K79me2 disparaît complètement après la fécondation pour rester à un niveau indétectable jusqu'au stade 4-cellules de l'embryogenèse (Ooga et al., 2008). Le niveau de méthylation augmente ensuite progressivement jusqu'au stade blastocyste (Ooga et al., 2013). Ces observations suggèrent que le transcrit *Dot1l* – hérité de l'ovocyte – n'impacte pas H3K79 au stade 1-cellule (Ooga et al., 2013). En effet, le KO de *Dot1l* dans les ovocytes n'affecte pas le développement embryonnaire ni la fertilité de la descendance (Liao and Szabó, 2020) démontrant que la contribution paternelle suffit.

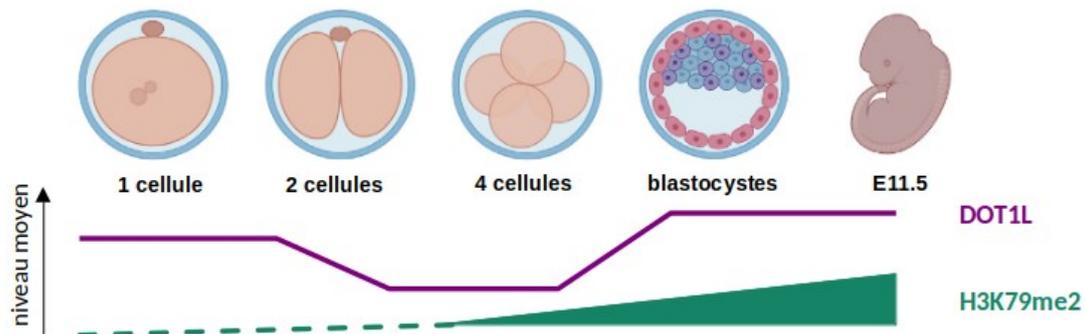


Figure 24 : Expression de *DOT1L* et de *H3K79me2* au cours du développement embryonnaire chez les mammifères. Figure basée sur les données d'immunofluorescence de Ooga et al., 2008, 2013

3.1.3 DOT1L et H3K79 durant la spermatogénèse

Comme présenté précédemment, l'analyse par RNA-seq montre que le gène *Dot1l* est fortement exprimé dans les cellules germinales mâles (Fig. 22) avec un pic d'expression dans les cellules méiotiques et post-méiotiques (Dottermusch-Heidel et al., 2014a, 2014b; Ontoso et al., 2014; Moretti et al., 2017). Ces observations ont été confirmées par des analyses scRNA-seq sur l'ensemble de la spermatogénèse qui montrent une augmentation importante du niveau d'expression de *Dot1l* à ces stades (Green et al., 2018). Il est intéressant de noter que la protéine DOT1L ainsi que la méthylation de H3K79 suivent la même dynamique (Dottermusch-Heidel et al., 2014a, 2014b; Moretti et al., 2017).

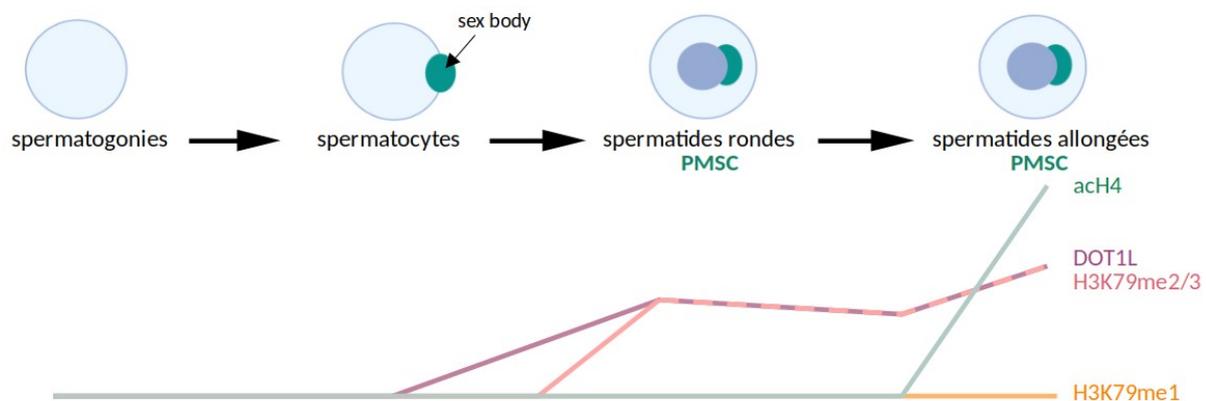


Figure 25 : Figure représentant le niveau d'expression de *Dot1l* et l'enrichissement de *H3K79me1/2/3* et de *acH4* dans le noyau et le PMSC au cours de la spermatogénèse. Figure basée sur les données de Moretti et al., 2016 ; Green et al., 2018 ; Ontoso et al., 2014 et de Dottermusch-Heidel et al., 2014a, 2014b

Dans les spermatocytes, la présence de la protéine DOT1L et de H3K79me a été étudiée dans le noyau et au niveau du *sex body*. L'approche par immunofluorescence ne montre pas de changement pour la marque H3K79me1 au cours de la méiose que ce soit dans le noyau ou dans le *sex body* (Fig. 25) (Ontoso et al., 2014). Au niveau du *sex body*, la protéine DOT1L est visible à partir du stade diplotène ce qui correspond à la phase de réactivation de l'expression des gènes des chromosomes X et Y (Ontoso et al., 2014). Après la méiose, la protéine DOT1L est toujours très visible dans le noyau des spermatides avec un signal particulièrement intense au niveau de la chromatine de l'X et de l'Y (PMSC)

(Moretti et al., 2017). L'intensité du signal DOT1L diminue ensuite au fur et à mesure de la phase « spermatides rondes » avant d'augmenter à nouveau juste avant l'élongation des spermatides. Cette augmentation précède de peu l'augmentation de H3K79me2/3 qui accompagne l'hyperacétylation de H4 dans les spermatides allongées (Dottermusch-Heidel et al., 2014a, 2014b; Moretti et al., 2017). Ces observations suggèrent que DOT1L et H3K79me pourraient être essentielles à la différenciation des cellules germinales mâles.

Récemment, plusieurs équipes – dont la nôtre – ont caractérisé le rôle de DOT1L au cours de la spermatogenèse et plus particulièrement dans les spermatogonies et durant la spermiogenèse (Lin et al., 2022, 2023; Blanco et al., 2023; Malla et al., 2023a). Ces études ont été réalisées par différents KO conditionnels dans les cellules germinales. Les résultats de ces études seront présentés dans la partie résultat et dans la discussion de mon manuscrit de thèse.

3.2 Méthylation de H3K79 et transcription des gènes

De la levure aux mammifères, il a été observé que les formes mono-, di- et triméthylées sont associées à l'activation de la transcription (Steger et al., 2008; Vlaming and van Leeuwen, 2016; Wood et al., 2018). En effet, l'enrichissement en la marque est positivement corrélé avec le niveau d'expression des gènes, c'est-à-dire qu'un fort enrichissement est associé à un haut niveau d'expression des gènes (Fig. 26).

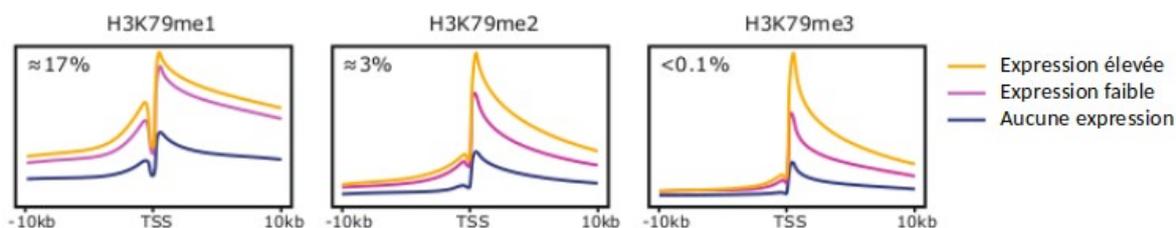


Figure 26 : Profil d'enrichissement de la marque H3K79 dans sa forme mono-, di- et triméthylée en fonction du niveau d'expression des gènes. Figure traduite de Vlaming et al., 2016

Chez la drosophile et les mammifères, les gènes enrichis en H3K79me2 sont associés à un fort taux d'élongation de l'ARN polymérase II et sont également enrichis pour le variant H3.3 (McKittrick et al., 2004; Jonkers et al., 2014; Veloso et al., 2014). Ce variant est décrit comme étant associé à la transcription active dans ces organismes (McKittrick et al., 2004). De plus, l'utilisation de ChIP-seq permet de mettre en évidence que H3K79me2 est enrichie au niveau des gènes codants pour des miARNs et pour des protéines (Suzuki et al., 2011). Toutes ces informations, corrélées avec l'augmentation de l'enrichissement de la marque en fonction du niveau d'expression des gènes, indiquent que DOT1L et la méthylation de H3K79me2 sont associées à l'activation de la transcription.

Cependant, depuis 2019, plusieurs études RNA-seq – dont les nôtres – utilisant différents modèles murins/cellulaires de KO pour *Dot1l*, suggèrent un autre rôle de DOT1L dans la régulation de la transcription. Ces études seront présentées et discutées ultérieurement dans les parties résultats et discussion.

3.2.1 Rôle dans l'élongation de la transcription

Depuis plusieurs années, la marque H3K79me2 est décrite comme étant associée à l'élongation de la transcription médiée par l'ARN polymérase II (Krogan et al., 2003; Bitoun et al., 2007; Steger et al., 2008; Mohan et al., 2010; Nguyen et al., 2011; Kim et al., 2012; Cao et al., 2020). Chez les mammifères, DOT1L est retrouvée dans plusieurs complexes permettant l'élongation de la transcription par l'ARN polymérase II (ARN polII) [voir Wood et al., 2018 pour revue], en particulier les complexes DotCom pour *Dot1-containing complex* (Mohan et al., 2010) et SEC, pour *Super Elongation Complex* (Luo et al., 2012). Le complexe SEC, contient, entre autres, la protéine AF9 (MLLT3) et est impliqué dans le développement embryonnaire (Lin et al., 2011) et dans certaines formes de cancer leucémiques et notamment celles par translocation impliquant le gène *MLL* (Luo et al., 2012). L'élongation de l'ARN polII par le complexe SEC est associé à une transcription active des gènes et notamment les gènes de contrôles du développement (Luo et al., 2012). DotCom comme son nom l'indique est un complexe composé de plusieurs sous-unités dont AF10 (MLLT10) et DOT1L (Mohan et al., 2010). Au sein de ces complexes, les enzymes AF9 et AF10 sont des cofacteurs de DOT1L, car ils permettent de moduler son activité de méthyltransférase. En effet une modification (sur-expression ou inhibition) de AF10 et AF9 corrèle positivement avec des changements (augmentation ou diminution) de la méthylation de H3K79me2/3 (Mohan et al., 2010; Wood et al., 2018).

La présence de la méthylation de H3K79 au niveau du promoteur et du corps du gène est un indicateur du processus d'élongation de la transcription par l'ARN polII (Steger et al., 2008). Pour comprendre le rôle de DotCom, Cao et al., ont réalisé un KO de *Dot1l* dans les cellules ESC (Cao et al., 2020). L'analyse d'expression différentielle entre les individus contrôles et KO-*Dot1l* montre peu d'effet sur l'élongation de la transcription par l'ARN polII. Cependant, le KO de *Dot1l* précédé d'une inhibition du complexe SEC par un traitement au KL-2 montre que la délétion de *Dot1l* renforce les défauts d'élongation causés par la perte de SEC (Fig. 27) (Cao et al., 2020). L'analyse par CHIP-seq des trois formes méthylées de H3K79 confirme que le KO de *Dot1l* entraîne une inhibition de son activité catalytique notamment en diminuant l'enrichissement de H3K79me (Cao et al., 2020).

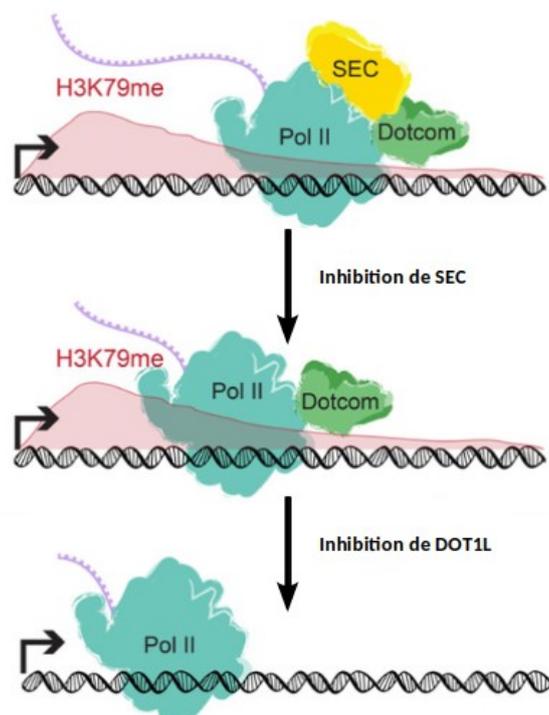


Figure 27 : Schéma de la fonction de DOT1L dans l'élongation de la transcription. Figure traduite de Cao et al., 2020

Comme évoqué précédemment, plusieurs études montrent également l'implication de DOT1L dans certains types de cancers et notamment les leucémies causées par des translocations MLL (Nguyen et al., 2011). Dans le cas des leucémies MLL-AF9, cette translocation est caractérisée par la fusion de AF9 avec MLL, ce qui entraîne une stabilisation de la protéine MLL. DOT1L, qui interagit directement avec AF9 au travers du complexe SEC, est associée à une activation inappropriée de gènes dans ces cellules (Mueller et al., 2007; Biswas et al., 2011). Cette dérégulation est associée à une augmentation locale (au niveau des gènes dérégulés) de H3K79me_{2/3} et H3K36me₃ (Krivtsov et al., 2008; Bernt et al., 2011; Kerry et al., 2017). Ces études montrent que DOT1L et ses partenaires (AF9, AF10) sont des cibles thérapeutiques potentielles pour le traitement de cancer MLL (Okuda et al., 2017; Grigsby et al., 2021).

3.2.2 H3K79me₂ et l'épissage alternatif

En 2018, l'article de Li et al., est le premier décrivant une corrélation entre DOT1L/H3K79me₂ et les événements d'épissages alternatifs dans différents types cellulaires (34 types cellulaires sains et cancéreux) chez l'Homme (Li et al., 2018). Les auteurs s'intéressent à l'impact de DOT1L sur le niveau d'expression des gènes associés à des sites d'épissage alternatif et particulièrement aux événements de saut d'exon (*exon skipping*) car c'est l'évènement le plus représenté avec jusqu'à 3035 gènes impactés (Li et al., 2018). L'*exon skipping* apparaît lors de l'épissage des ARNm, au cours duquel, un ou plusieurs exons peuvent être « sautés » entraînant un nouveau transcrit, et donc, potentiellement, une nouvelle isoforme de la protéine (Fig. 28).

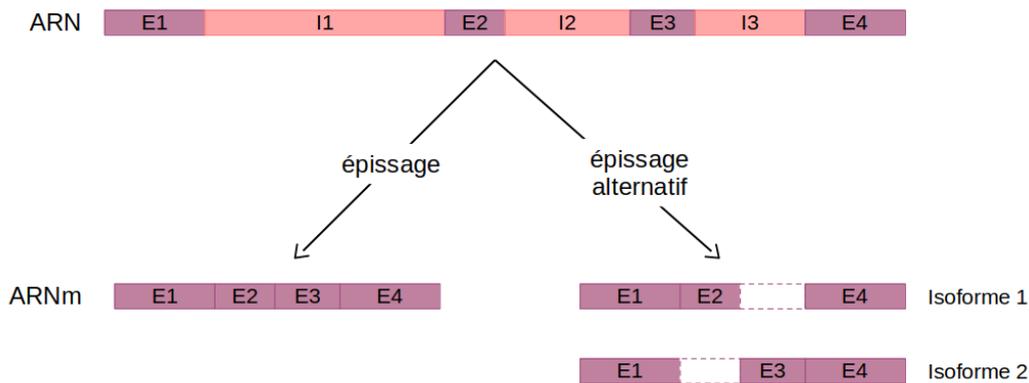


Figure 28 : Schéma du saut d'exon lors de la transcription

Dans cette étude, les auteurs montrent que le KO de *Dot1l* entraîne une modification des isoformes, obtenues par épissage alternatif, dans deux types de MLL, notamment en incluant des exons au lieu de les exclure (Li et al., 2018). L'analyse d'expression par RNA-seq, des cellules où DOT1L est inhibée, ne montre pas de différence significative d'expression entre les transcrits associés à un événement d'*exon skipping* et ceux qui ne le sont pas. Cependant, ils observent que de nombreux événements d'*exon skipping* sont caractérisés par des pics H3K79me₂, indiquant que la corrélation entre H3K79me₂ et l'épissage alternatif est indépendante du niveau d'expression des gènes (Li et al., 2018). Ces résultats montrent que H3K79me₂ joue un rôle dans le processus d'épissage alternatif et que ces modifications peuvent influencer la transformation et la progression de la maladie dans le cas des leucémies MLL.

3.3 Interaction (« *cross-talk* ») entre H3K79me et d'autres modifications post-traductionnelles

En parallèle de leur rôle sur la transcription des gènes, DOT1L et la méthylation de H3K79 semblent interagir avec d'autres modifications post-traductionnelles comme l'ubiquitination de H2B (H2Bub) ou l'acétylation de H4 (acH4).

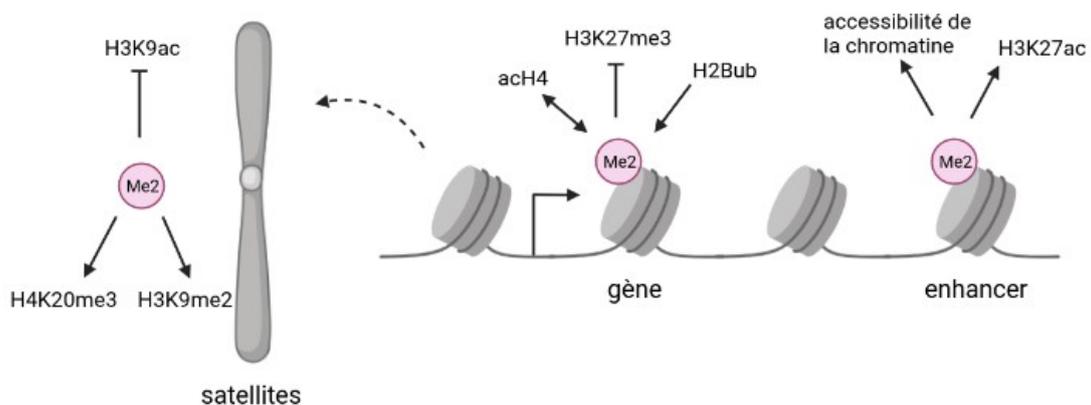


Figure 29 : Schéma des interactions entre H3K79me₂ représenté par Me₂ (en rose) et d'autres modifications post-traductionnelles. Figure adaptée de Wille and Sridharan, 2022.

3.3.1 H3K79me2/3 et l'ubiquitination de l'histone H2B

Une des modifications post-traductionnelles impactant H3K79me est l'ubiquitination de H2B (Lee et al., 2018; van Welsem et al., 2018). Chez plusieurs espèces et notamment la levure et les mammifères, il a été montré que l'ubiquitination de H2B est nécessaire à la mise en place de la di- et tri-méthylation de H3K79 dans le nucléosome (Davie and Murphy, 1990; Sun and Allis, 2002; Zhu et al., 2005; Fuchs et al., 2014). En effet, la délétion de l'enzyme nécessaire à l'ubiquitination de H2B entraîne une baisse de la méthylation H3K79 (Sun and Allis, 2002; Zhu et al., 2005). Des études récentes de la structure des protéines ont démontré que H2Bub augmente l'activité catalytique de DOT1L (Anderson et al., 2019; Jang et al., 2019; Worden et al., 2019; Yao et al., 2019). Ainsi l'augmentation de l'ubiquitination de H2B induit une augmentation de la méthylation de H3K79 (Fig. 29) (Valencia-Sánchez et al., 2021; Sharma et al., 2022).

Conjointement avec le rôle de DOT1L dans l'élongation de la transcription, il a été montré que le niveau de H2Bub corrèle avec le niveau d'expression et le taux d'élongation de l'ARN polymérase II (Fuchs et al., 2014; Vlaming and van Leeuwen, 2016). Cette observation évoque une coopération entre la méthylation de H3K79me et H2Bub pour promouvoir le processus d'élongation de la transcription dirigé par H3K79me2 (Vlaming and van Leeuwen, 2016).

3.3.2 H3K79me2/3 et l'acétylation de l'histone H4

Une autre modification post-traductionnelle impactant H3K79me2 est l'acétylation de l'histone H4 (acH4). Comme décrit précédemment, l'augmentation de la méthylation de H3K79 accompagne l'hyperacétylation de l'histone H4 durant la spermiogenèse (Dottermusch-Heidel et al., 2014a, 2014b). Parmi les marques acétylées de H4, DOT1L interagit particulièrement avec H4K16ac (Fig. 29). En effet, l'étude de la structure protéique indique que l'interaction entre DOT1L et la marque H4K16ac permet d'augmenter l'activité méthyltransférase de DOT1L et que l'association conjointe des deux marques H2Bub et H4K16ac permet d'augmenter davantage la méthylation de H3K79 par DOT1L (Worden et al., 2019; Valencia-Sánchez et al., 2021). La présence conjointe de ces marques permet donc à DOT1L de déposer plusieurs résidus méthyl successivement sur H3K79 améliorant ainsi sa processivité (Worden et al., 2019). Le « *cross-talk* » entre H3K79 et acH4 semble bi-directionnel puisque dans leur étude, Gilan et al., observent que DOT1L via H3K79me2 facilite l'acétylation de H4 et permet de recruter l'enzyme BRD4 qui interagit avec les histones acétylées (Gilan et al., 2016).

Il a également été montré qu'une diminution de DOT1L induit une diminution de la méthylation de H3K79 et de l'hyperacétylation dans les spermatides allongées (Moretti et al., 2017).

3.3.3 H3K79me2/3 et les marques d'hétérochromatine

Pang et al., ont décrit que des marques activatrices comme H3K36me3 et H3K79me2 peuvent être retrouvées dans des régions d'hétérochromatine en association avec des

marques répressives comme H3K27me3 (Pang and Snyder, 2020). Des expériences d'immunofluorescence ont d'ailleurs montré que H3K79me3 est localisée dans l'hétérochromatine péricentromérique (Ooga et al., 2008; Abe et al., 2011). Dans des ESC, il a été observé que le KO de *Dot1l* entraîne la perte de la méthylation de H3K79, comme attendu, et s'accompagne d'une augmentation de l'acétylation de H3K9 et d'une réduction des marques d'histones H3K9me2 et H4K20me3 dans les centromères et les télomères (Fig. 29), confirmant la relation entre DOT1L et la régulation des régions hétérochromatiniennes (Jones et al., 2008). De manière intéressante, l'expression forcée de DOT1L dans les embryons pré-implantatoires perturbe l'organisation de cette même hétérochromatine indépendamment de son activité méthyltransférase (Ooga et al., 2013). Ces résultats suggèrent que DOT1L est nécessaire pour la régulation de l'activité transcriptionnelle au niveau des régions hétérochromatiniennes. Récemment, des articles se sont donc intéressés à décrire le rôle transcriptionnel de DOT1L sur l'hétérochromatine durant le développement embryonnaire précoce et la spermatogenèse (Malla et al., 2023a, 2023b). Les résultats de ces études seront abordés plus loin dans la partie discussion.

Résultats

Comme décrit ci-dessus, la spermatogenèse est un processus biologique caractérisé par une dynamique transcriptionnelle et chromatinienne importante, ce qui fait de la spermatogenèse, un très bon modèle pour étudier l'interaction entre la chromatine et la transcription. Mon projet de thèse a consisté à caractériser la marque H3K79me2 et sa méthyltransférase, DOT1L, au cours de la spermatogenèse, afin de comprendre le rôle moléculaire de DOT1L, en particulier son impact sur la régulation de l'expression génique, lors de ce processus.

1 Introduction et contexte de l'étude

Avant mon arrivée, de précédents travaux de l'équipe avaient mis en avant le rôle du gène *Sly* lors de la réactivation des chromosomes sexuels dans les spermatides rondes, et dans la transition histone-protamine lors de la spermiogenèse (Cocquet et al., 2009; Moretti et al., 2017). Ces études avaient montré par ChIP et analyse du transcriptome que la protéine SLY régule l'expression du gène *Dot1l*. Le *knock-down* de *Sly* dans les spermatides rondes induit, en effet, une diminution de l'expression du gène *Dot1l* ainsi qu'une diminution de l'enrichissement en H3K79me2, juste avant la transition histone-protamine. Cette observation avait permis de proposer un modèle selon lequel la diminution de *Dot1l* et H3K79me2 contribuerait aux défauts de remplacement des histones observés dans les souris déficientes en *Sly* (Moretti et al., 2017). Il avait également été observé par immunofluorescence que la protéine DOT1L est fortement exprimée dans les spermatides rondes, et présente au niveau de la chromatine sexuelle (PMSC) supposant un rôle de DOT1L dans la régulation de cette structure (Moretti et al., 2017).

Toutes ces données indiquaient que DOT1L est impliquée dans la spermiogenèse, l'équipe a donc produit un modèle de souris KO conditionnel afin d'étudier le rôle de DOT1L dans ce processus. Dans ce cadre, l'objectif de mon projet de thèse a été, dans un premier temps, d'analyser les données de RNA-seq issues des spermatides contrôles et KO-*Dot1l* afin de comprendre le rôle moléculaire de DOT1L et le lien entre la dérégulation génique lors de la spermiogenèse, et le phénotype des souris KO. Ces résultats sont présentés sous la forme d'un article publié cette année (cf. 3.2 Article 1).

Dans un deuxième temps, j'ai intégré la totalité des données de RNA-seq générées à partir de ce modèle avec des données de ChIP-seq pour la marque H3K79me2 obtenues à partir de souris contrôle, dans le but de caractériser de façon globale la dynamique de la marque tout au long de la spermatogenèse, et de mieux comprendre l'impact de DOT1L sur la régulation de l'expression génique. Ces analyses m'ont permis de mettre en évidence un rôle de répresseur de la transcription en plus du rôle d'activateur déjà décrit pour DOT1L. En parallèle de ces analyses, j'ai étudié l'implication de H3K79me2 et de DOT1L dans le processus de régulation de l'expression des gènes localisés dans les chromosomes sexuels. Ces résultats font l'objet d'un second article qui sera bientôt soumis et qui est présenté dans la partie suivante de ce manuscrit (cf. 4.2 Article 2).

D'autres études visant à caractériser le phénotype des souris avec un KO de *Dot1l* dans la lignée germinale, au stade spermatogonies (Lin et al., 2022) ou dans les spermatides (Lin et al., 2023; Malla et al., 2023a), ont été publiées en parallèle de la nôtre. Cependant, à ce jour, aucune étude ne décrit le rôle antagoniste de DOT1L dans la régulation de la transcription, en particulier celle des chromosomes sexuels lors de la spermatogenèse, ni la dynamique de la marque H3K79me2 tout au long de ce processus particulier de différenciation. Ces résultats seront discutés à la fin de ce manuscrit.

2 Approches développées pour le projet

2.1 Données biologiques

2.1.1 Données de transcriptomique KO-*Dot1l* et contrôle

Pour réaliser ce projet, l'équipe a produit un modèle murin KO conditionnel pour *Dot1l* (KO-*Dot1l*), qui correspond à une délétion de l'exon 2 de *Dot1l* induite par la *Stra8-Cre* dans la lignée germinale et plus particulièrement à partir du stade spermatogonies. À partir de ce modèle, l'équipe a généré des données de RNA-seq sur des individus contrôles (CTL) et des individus KO-*Dot1l* dans plusieurs stades cellulaires de la spermatogenèse comme les spermatocytes primaires (SCI) et secondaires (SCII) et les spermatides rondes (RS) (Fig. 30). Ces données ont été complétées avec des données de spermatogonies indifférenciées (Kit-) et différenciées (Kit+) provenant de nos collaborateurs (équipe de Pierre Fouchet, Laboratoire Cellules Souches Germinales, CEA de Fontenay-aux-Roses). Les différentes populations cellulaires ont été isolées par tri (FACS) selon leur taille et leur contenu en ADN.

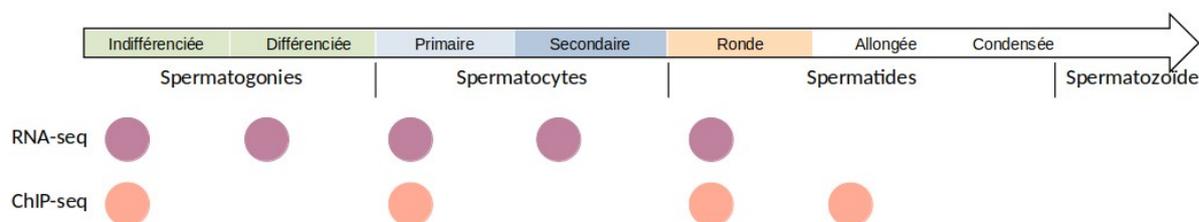


Figure 30 : Liste des données transcriptomiques et épignétiques utilisées dans notre étude. En orange, les données de ChIP-seq pour H3K79me2 et en rose, les données de RNA-seq pour des individus KO-*Dot1l* et contrôles

Les données de transcriptomiques ont été réalisées selon deux méthodes, la méthode de RNA-seq « classique », dont la normalisation consiste à l'utilisation d'une méthode (CPM pour *Count per Million*, TPM pour *Transcrit per Million* ou RPKM pour *Reads Per Kilobase Million*) permettant de normaliser le niveau d'expression des gènes et la méthode de RNA-seq *spike-in*. Cette deuxième consiste en un protocole d'analyse de RNA-seq standard dans lequel des séquences exogènes, appelées *spike-in*, sont ajoutées lors de la préparation des bibliothèques. Le *spike-in* est un ensemble de séquences exogènes ayant une composition et une concentration connues qui sont ajoutées à l'échantillon afin de normaliser les résultats, soit au sein d'un même échantillon, soit entre différents

échantillons. Un pipeline d'alignement des données de RNA-seq a été mis en place suivant le modèle classique utilisé pour les analyses de transcriptomique (Fig. 3) en utilisant les outils BBDuk pour le retrait des adaptateurs et STAR pour l'alignement. Les paramètres utilisés seront décrits plus précisément dans la partie « Matériels et Méthodes » des articles de cette étude.

2.1.2 Données de ChIP-seq pour H3K79me2

Nous avons également généré des données de ChIP-seq pour la marque H3K79me2 dans les spermatogonies indifférenciées (GSC), les spermatocytes primaires et les spermatides rondes et allongées (ES) (Fig. 30). La perte de DOT1L entraînant une forte diminution de la marque sur tous les sites testés par ChIP-qPCR, il n'a pas semblé pertinent de réaliser un ChIP-seq H3K79me2 sur des échantillons KO-*Dot1l*. Contrairement aux données de transcriptomiques dont les spermatogonies indifférenciées sont directement issues du tri grâce à un marqueur KIT, les spermatogonies indifférenciées obtenues pour les données de ChIP-seq sont des spermatogonies indifférenciées triées selon la même méthode puis mises en culture pour augmenter la quantité de matériel – des expériences de ChIP-seq H3K79me2 sur faible quantité de matériel (*low input*) n'ayant pas fonctionné. Un pipeline d'alignement des données de ChIP-seq a été mis en place suivant le modèle classique utilisé pour les analyses de ChIP-seq (Fig. 3) en utilisant les outils BBDuk pour le retrait des adaptateurs, bowtie2 pour l'alignement et MACS2 pour déterminer les pics de H3K79me2. Comme pour les données de transcriptomiques, les paramètres utilisés par chacun de ces outils seront décrits plus précisément dans la partie « Matériels et Méthodes » des articles de cette étude.

2.2 Mise en place d'une analyse d'expression différentielle

L'un des objectifs de mon projet a été d'analyser l'impact de DOT1L sur la transcription tout au long de la spermatogenèse. Pour étudier cette dynamique, j'ai réalisé une analyse différentielle de l'expression des gènes entre 5 stades de cellules germinales contrôles et KO-*Dot1l*. L'analyse de l'expression génique a été faite à l'aide des packages R DESeq2 (Love et al., 2014) et edgeR (Robinson et al., 2010). Pour chacun des échantillons, le nombre de séquences alignées par gène est reporté en CPM, le nombre de séquences par gène est ainsi normalisé en fonction du nombre de séquences identifiés tous gènes confondus.

2.2.1 Génération du modèle d'analyse

L'une des difficultés du projet a été de mettre en place un modèle pour réaliser l'analyse différentielle d'expression en parallèle dans les 5 types cellulaires, sachant qu'il existe une très grande diversité transcriptionnelle entre les stades pré-méiotiques et méiotiques/post-méiotiques. Le modèle qui a été développé pour réaliser l'analyse différentielle d'expression prend en compte trois paramètres qui sont le type cellulaire (Kit-, Kit+, SCI, SCII et RS), le génotype (KO-*Dot1l* et contrôle) et l'interaction entre les deux premiers paramètres (Formule. n°1). Ce troisième paramètre a été ajouté pour prendre en

compte l'interaction entre le type cellulaire et le génotype étudié, et ainsi « compenser » cette hétérogénéité transcriptionnelle.

modèle = ~ type cellulaire + génotype + type cellulaire : génotype

Formule n°1 : Formule permettant la caractérisation du modèle mis en place pour les analyses différentielles d'expression

2.2.2 Estimation des gènes dérégulés

Une fois le modèle établi, l'objectif est d'identifier les gènes dérégulés entre nos deux conditions KO-*Dot1l* et contrôle. Les gènes dérégulés sont estimés en fonction de la valeur de deux paramètres : un seuil de *fold-change* (FC) et une valeur de significativité. Le *fold-change* correspond à la différence d'expression entre deux conditions A et B (Formule n°2). Ainsi une valeur de *fold-change* de 2 entre une condition A et B signifie que la condition B est deux fois plus exprimée que la condition A. Dans le cas de notre étude, nous avons établi notre valeur de *fold-change* à 1,5 c'est-à-dire que l'on estime que les gènes sont dérégulés si la différence d'expression entre les individus KO-*Dot1l* et les individus contrôles est de minimum 1,5.

$$FC = \left(\frac{B}{A}\right) - 1 \quad \text{ou} \quad FC = \frac{B - A}{A}$$

Formule n°2 : Formule pour calculer le fold-change entre deux conditions avec A représentant l'état initial et B représentant l'état final

Dans le cas de notre étude, nous estimons un gène comme dérégulé significativement s'il a une valeur de *p-value* inférieure à 5 % (*p-value* < 0,05 ou *p-value* < 5 %). La *p-value* permet d'estimer la significativité statistique d'une valeur par rapport à une hypothèse nulle H_0 . Ainsi, une *p-value* inférieure à 5 % signifie que l'on a moins de 5 % de chance que l'hypothèse H_0 soit vraie. Lorsqu'un grand nombre de *p-value* est calculé, comme c'est le cas pour une analyse différentielle d'expression où l'on a une valeur de *p-value* par gène exprimé, il est nécessaire de corriger les valeurs de *p-value* en leur appliquant une correction permettant ainsi de déterminer leur valeur de *p-value* ajustée, également nommée *q-value*. Cette correction se fait grâce à plusieurs méthodes comme la correction de Benjamini-Hochberg (BH ou FDR (*False Discovery Rate*, Benjamini and Hochberg, 1995)) ou la correction de Bonferroni (Dunn, 1961) et permet d'estimer la significativité de la valeur de la *p-value*. Ainsi, une valeur de *q-value* inférieure à 5 % signifie que l'on a moins de 5 % de se tromper sur la significativité de la *p-value*.

Dans notre étude, nous estimons deux niveaux de dérégulations, les gènes dérégulés significativement (FC > 1,5 et *p-value* < 0,05) et les gènes dérégulés avec un haut niveau de significativité (FC > 1,5 et *q-value* < 0,05). Pour la dernière catégorie, la *q-value* a été calculée en utilisant la correction de Benjamini-Hochberg et sera donc nommé FDR dans la suite de ce manuscrit. La correction de la *p-value* prend en compte l'ensemble des gènes exprimés,

de ce fait le nombre de gènes très significativement dérégulés ($q\text{-value} < 5\%$) est inférieur au nombre de gènes significativement dérégulés ($p\text{-value} < 5\%$). Une approche supervisée de la correction de la $p\text{-value}$ a donc été faite dans le but de valider les gènes significativement dérégulés. Cette approche montre ainsi que l'intégralité des gènes dérégulés ($p\text{-value} < 5\%$) a une $q\text{-value}$ inférieure à 5 %, validant ainsi la liste de gènes dérégulés.

2.2.3 Analyse par *spike-in*

Comme précisé, nos échantillons pour l'analyse transcriptomique ont été couplés avec des séquences *spike-in*, séquences exogènes utilisées pour normaliser les échantillons par une approche externe. L'objectif de cette normalisation est d'ajuster l'expression des gènes en fonction de données connues qui sont les valeurs d'expression des *spike-in* et permettre d'estimer si les dérégulations observées dans une approche classique sont le reflet de la réalité biologique ou si c'est un biais de normalisation. Cette approche est particulièrement intéressante à mettre en place dans le cas de dérégulation globale ou pour comparer des types cellulaires ayant un niveau d'expression très différents.

Pour cela, nous avons utilisé la méthode de normalisation décrite par Risso et al., appelée RUV pour *Remove Unwanted Variation* (Risso et al., 2014). Cette méthode consiste à utiliser un set de gènes comme contrôle négatif, c'est-à-dire qui ne sont pas différentiellement exprimés entre les deux conditions (ce qui va être le cas des *spike-in*). Dans le cas où les *spike-in* ne montrent pas de différences d'expression, la $p\text{-value}$ doit suivre une loi normale. Une fois normalisées, les données sont analysées par une analyse différentielle d'expression en suivant le même schéma que celui présenté dans la partie précédente.

2.3 Caractérisation des états chromatinien

Pour comprendre et déterminer l'interaction de H3K79me2 avec d'autres marques épigénétiques, il a été nécessaire de mettre en place une analyse permettant la caractérisation de H3K79me2 en fonction des états chromatinien.

ChromHMM est un outil développé par Ernst et al., qui permet de séparer le génome d'intérêt en différentes régions ou « états » chromatinien (Ernst and Kellis, 2012). Ces régions sont définies par la présence d'une ou de plusieurs marques épigénétiques. Le nombre et les marques utilisées influencent le type d'état chromatinien défini par ChromHMM. Pour définir ces états chromatinien, l'outil utilise un modèle caché de Markov (HMM pour *Hidden Markov Model*). ChromHMM est un outil de prédiction, il est donc indispensable d'avoir un grand jeu de données pour avoir une meilleure efficacité de prédiction des états chromatinien. Dans le cas de notre étude, nous avons utilisé six marques d'histone : H3K4me1, H3K4me3, H3K36me3, H3K9me3, H3K27ac et H3K27me3 pour obtenir 18 états chromatinien différents répartis en plusieurs catégories comme les *enhancers*, les états répressifs, les états actifs ou les états quiescents (Fig. 31). L'état quiescent représentant la majorité du génome quel que soit le type cellulaire étudié.

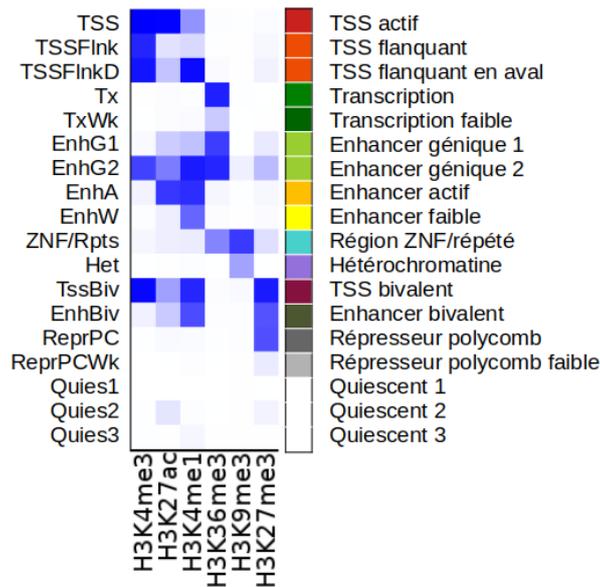


Figure 31 : Modèle chromatinien obtenu en utilisant ChromHMM

Pour augmenter la précision de notre modèle, nous avons récupéré dans la littérature les marques épigénétiques de 17 tissus ou types cellulaires différents parmi lesquels des stades de la spermatogenèse (spermatogonies indifférenciées et différenciées, spermatocytes primaires et spermatides rondes) (Fig. 32). L'objectif de ce modèle est d'intégrer les états chromatinien obtenus avec nos données de transcriptomique et de H3K79me2 pour différents stades de la spermatogenèse.

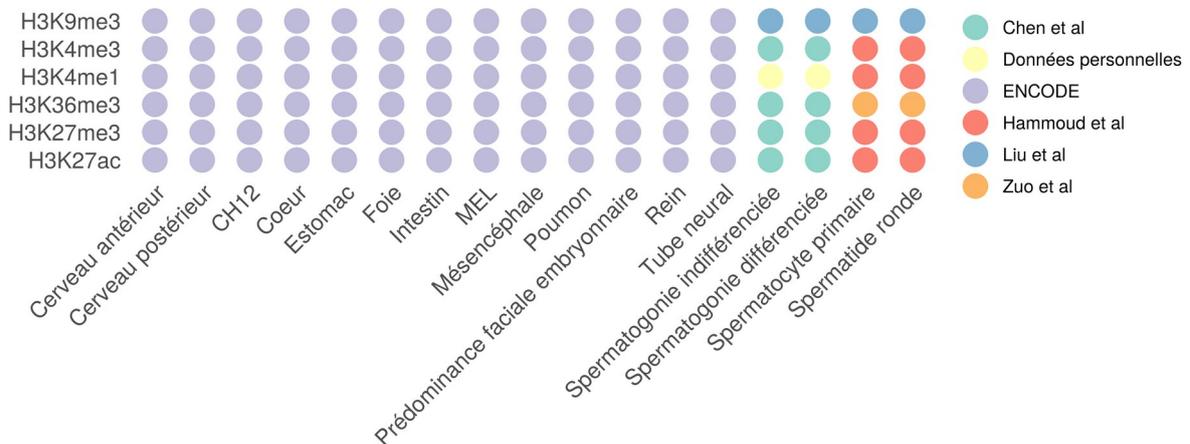


Figure 32 : Liste des données utilisées pour la mise en place du modèle (B). MEL correspondant à une lignée cellulaire dérivant d'érythroleucémie et CH12 à une lignée cellulaire dérivant de lymphome. Les données de ChIPseq proviennent du projet ENCODE et de la littérature (ENCODE Project Consortium, 2012; Hammoud et al., 2014; Zuo et al., 2018; Liu et al., 2019; Chen et al., 2020b)

À partir du modèle ChromHMM qui a été obtenu (Fig. 31), plusieurs analyses ont été réalisées. La première a consisté à associer H3K79me2 à une expression génique lorsque cette marque est positionnée dans les *enhancers*, pour cela, une méthode d'association des gènes aux *enhancers* a été mise en place. La deuxième analyse a consisté à analyser l'environnement chromatinien des gènes dérégulés dans le KO-*Dot1l*.

Ces deux volets d'analyses ont été compilés dans un package R, nommé ChromENVEE pour *Chromatin Environment and Enhancer-dependent Expression*. Ce package est actuellement disponible sur GitHub (Annexe. Vignette ChromENVEE, <https://github.com/ManonCoulee/ChromENVEE>).

2.3.1 Association des gènes aux *enhancers*

Les *enhancers* sont des régions permettant de réguler l'expression des gènes à distance. L'attribution des gènes aux *enhancers* se fait couramment en utilisant la méthode d'Hi-C (cf. 1.2.4 *L'organisation spatiale de la chromatine*). Cependant cette méthode est relativement coûteuse et difficile à mettre en place. De ce fait, des articles ont utilisé une méthode permettant l'estimation du couple gène-*enhancer* à partir de données de CHIP-seq (Rada-Iglesias et al., 2011). Cette méthode considère comme gène régulé par un *enhancer*, le gène situé le plus proche de l'*enhancer* à une distance maximale de 100kb (Fig. 33A). Cette méthode est, par exemple, reprise dans l'article de Godfrey et al., pour caractériser l'impact de H3K79me2/3 sur l'expression génique lorsque ces marques sont situées au niveau des *enhancers* (Godfrey et al., 2019). Dans le cas de Godfrey et al., seuls les gènes enrichis en H3K79me2/3 au niveau du corps du gène sont gardés dans l'analyse et ceux sans H3K79me2/3 sont exclus car ayant généralement un niveau d'expression plus faible (Fig. 33B).

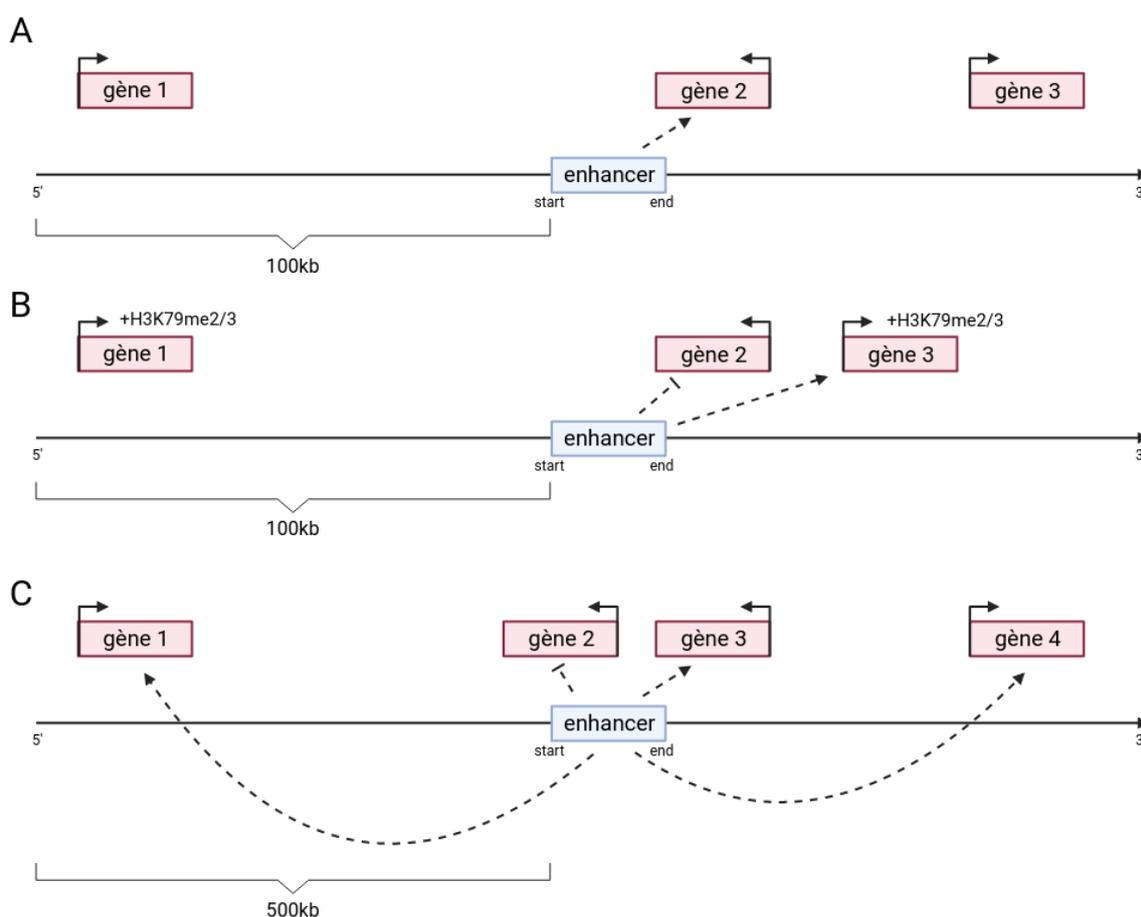


Figure 33 : Estimation des gènes régulés par des *enhancers*. A. méthode de Rada-Iglesias et al, 2011., B. méthode de Godfrey et al, 2019., et C. méthode utilisée dans le package ChromENVEE

Cette approche a ses limites, car les *enhancers* ne régulent pas systématiquement le gène le plus proche. Les gènes régulés par les *enhancers* peuvent parfois être très éloignés, allant jusqu'à un million de paires de base de distance (Lettice et al., 2003; Amano et al., 2009; Sanyal et al., 2012; Novo et al., 2018; Schoenfelder and Fraser, 2019). De plus, dans leur article, Karnuta et al., décrivent qu'un *enhancer* peut réguler plusieurs gènes et qu'inversement un gène peut être régulé par plusieurs *enhancers* (Karnuta and Scacheri, 2018). Ainsi mon objectif a été de développer une méthode permettant d'estimer quels sont les gènes régulés par les *enhancers* en prenant en compte l'ensemble des informations spécifiques des *enhancers*. Pour cela, je me suis basée sur les catégories d'*enhancer* défini par notre modèle ChromHMM (Fig. 31). Contrairement à la méthode présentée dans l'article de Rada-Iglesias et al., j'ai tout d'abord considéré comme gènes associés à l'*enhancer*, l'ensemble des gènes dont le promoteur est situé à une distance maximale de 500kb de part et d'autre de l'*enhancer* (Rada-Iglesias et al., 2011). Les gènes dont le promoteur est situé dans l'*enhancer* ont été exclus comme c'est le cas du gène 2 dans l'exemple (Fig. 33C).

La méthode que j'ai développée permet d'avoir des informations supplémentaires concernant les gènes associés aux *enhancers* comme :

- le nombre de gène associé aux *enhancers*,
- la distance entre l'*enhancer* et le promoteur du gène,
- le niveau d'expression des gènes associé aux *enhancers* à partir de donnée de transcriptomique.

En comparant le niveau d'expression des gènes en fonction de leur distance avec l'*enhancer*, nous avons observé que la présence de H3K79me2 au niveau des *enhancers* affecte l'expression des gènes proches (ayant une distance avec l'*enhancer* inférieure à 100kb) (Fig. 34). Ainsi pour la suite des analyses utilisant les couples gènes-*enhancers*, nous ne nous intéresserons qu'à ceux ayant une distance de moins de 100kb.

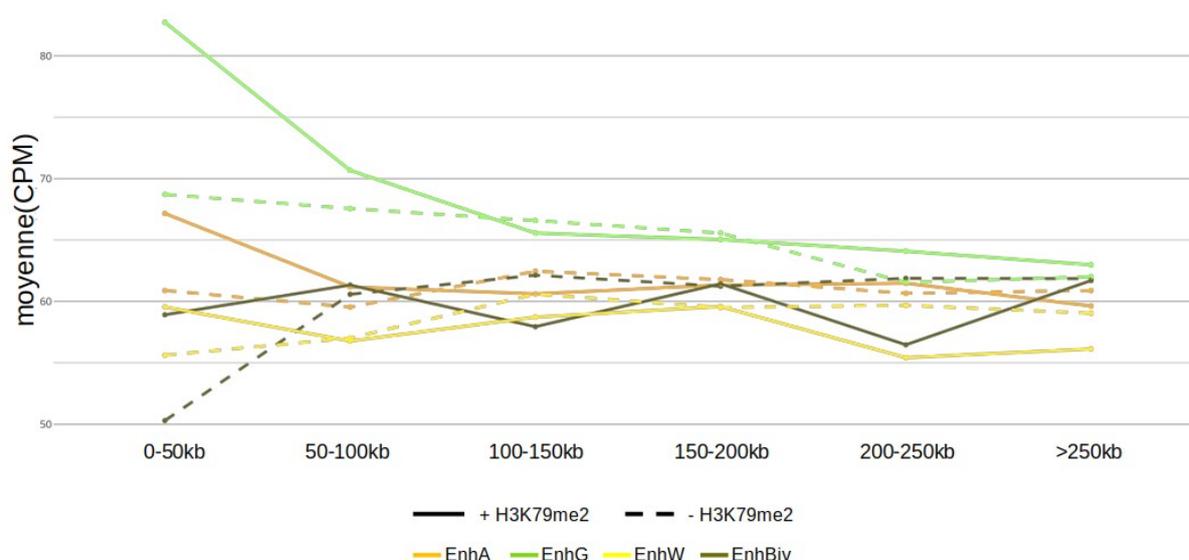


Figure 34 : Impact de la présence de H3K79me2 au niveau des *enhancers* sur l'expression génique en fonction de la distance gène-*enhancer* dans les spermatozoïdes ronds

2.3.2 Caractérisation de l'environnement chromatinien

À partir des états chromatinien définis par le modèle ChromHMM, l'objectif a été de déterminer l'environnement chromatinien des gènes dérégulés dans notre analyse transcriptomique.

En se basant sur la méthode décrite dans l'article de Ferrari et al., j'ai considéré comme environnement chromatinien l'ensemble des états chromatinien qui sont situés dans l'environnement d'un gène (Ferrari et al., 2020). Dans le cas de notre étude, nous avons estimé que l'environnement d'un gène est de 3kb autour du TSS de ce gène, soit un environnement total de 6kb (Fig. 35), paramètre similaire à celui utilisé pour l'analyse via ChIPseeker (Yu et al., 2015) pour déterminer l'annotation génomique des pics de H3K79me2. L'objectif de la méthode que j'ai développée dans le package ChromENVEE est de déterminer les états chromatinien présents dans l'environnement du gène à partir de leur localisation génomique. Une fois l'environnement déterminé, nous cherchons à estimer l'état chromatinien majoritaire de chaque gène. Pour cela, nous considérons comme état chromatinien majoritaire, l'état ayant la plus grande couverture dans l'environnement.

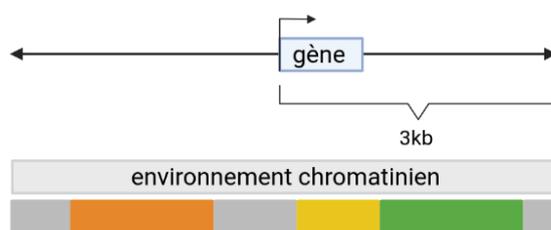


Figure 35 : Caractérisation de l'environnement chromatinien d'un gène par ChromENVEE. Chaque couleur correspond à un état chromatinien différent.

3 DOT1L régule l'organisation de la chromatine et l'expression des gènes durant la différenciation spermatique

3.1 Résultats de l'étude

L'analyse phénotypique du modèle murin KO-*Dot1l* (réalisée par d'autres membres de l'équipe) a permis de mettre en évidence le rôle de DOT1L dans la spermatogenèse et en particulier lors de la spermiogenèse en montrant que les mâles KO présentent des défauts de compaction de la chromatine. Ces observations sont confirmées par l'analyse par spectrométrie de masse qui montre que l'absence de DOT1L induit une dérégulation de plusieurs marques d'histones, en particulier une diminution drastique de l'hyperacétylation de H4 dans les spermatides allongées. Cette marque permet l'ouverture de la chromatine et est indispensable au retrait des histones avant leur remplacement par les protamines. Cette dérégulation entraîne, dans le spermatozoïde, une rétention d'histones plus importante chez les individus KO-*Dot1l* par rapport aux individus contrôles. Les analyses par *western blot* montrent également que le noyau des spermatozoïdes KO-*Dot1l* contient des protéines de transitions et des formes immatures de la protamine 2 (*Prm2*).

En plus des défauts de compaction de la chromatine, les souris KO-*Dot1l*, montrent des défauts de formation du flagelle et de mobilité des spermatozoïdes. Ces multiples défauts des spermatozoïdes induisent une hypofertilité. Pour comprendre l'origine des défauts de la spermiogenèse observée, j'ai comparé les données transcriptomiques des stades méiotiques et post-méiotiques pour des individus KO-*Dot1l* et contrôles par une analyse différentielle « classique » (Fig. 31). L'analyse d'expression des gènes montre que DOT1L régule l'expression de plusieurs centaines de gènes avec plus de gènes sur-exprimés que de gènes sous-exprimés. En analysant l'ontologie des cellules KO par rapport aux cellules contrôles, nous observons que DOT1L régule l'expression des gènes impliqués dans l'organisation de la chromatine, la formation du flagelle, le cycle cellulaire, la régulation de la transcription et le métabolisme, ce qui corrèle avec les observations faites lors de l'analyse phénotypique. En intégrant des données de ChIP-seq pour H3K79me2 dans les spermatides rondes, nous observons que les gènes sous-exprimés sont associés à un enrichissement de la marque H3K79me2 au niveau du gène alors que les gènes sur-exprimés ne sont pas enrichis en H3K79me2 au niveau du corps du gène. Ces données indiquent que les gènes sous-exprimés dans KO sont des cibles « directes » de DOT1L alors que les gènes sur-exprimés seraient des cibles « indirectes ».

L'ensemble de ces résultats ont fait l'objet d'un article publié dans EMBO Reports en 2023 qui est présenté dans la partie suivante (Blanco et al., 2023).

3.2 Article 1

DOT1L regulates chromatin reorganization and gene expression during sperm differentiation

Publié dans EMBO Reports le 26 avril 2023

Mélina Blanco, Laila El Khattabi, Clara Gobé, Marion Crespo, Manon Coulée, Côme Ialy-Radio, Clémentine Lapoujade, Maëlle Givelet, Alberto de la Iglesia, Marion Delessard, Ivan Seller-Corona, Kosuke Yamaguchi, Nadège Vernet, Fred Van Leeuwen, Alban Lermine, Yuki Okada, Romain Daveau, Rafael Oliva, Pierre Fouchet, Ahmed Ziyat, Delphine Pflieger & Julie Cocquet

DOI : [10.15252/embr.202256316](https://doi.org/10.15252/embr.202256316)

Les figures supplémentaires sont disponibles en annexe de ce manuscrit



Article

SOURCE
DATATRANSPARENT
PROCESSOPEN
ACCESSEMBO
reports

DOT1L regulates chromatin reorganization and gene expression during sperm differentiation

Mélina Blanco¹, Laila El Khattabi^{1,2,†} , Clara Gobé^{1,†}, Marion Crespo^{3,4}, Manon Coulée¹, Alberto de la Iglesia^{1,5} , Côme Ialy-Radio¹, Clementine Lapoujadé⁶, Maëlle Givelet⁶, Marion Delessard¹, Ivan Seller-Corona¹ , Kosuke Yamaguchi⁷, Nadège Vernet⁸ , Fred Van Leeuwen⁹ , Alban Lermine¹⁰, Yuki Okada⁷, Romain Daveau¹⁰, Rafael Oliva^{5,11} , Pierre Fouchet⁶, Ahmed Ziyat^{1,12} , Delphine Pflieger³ & Julie Cocquet^{1,*}

Abstract

Spermatozoa have a unique genome organization. Their chromatin is almost completely devoid of histones and is formed instead of protamines, which confer a high level of compaction and preserve paternal genome integrity until fertilization. Histone-to-protamine transition takes place in spermatids and is indispensable for the production of functional sperm. Here, we show that the H3K79-methyltransferase DOT1L controls spermatid chromatin remodeling and subsequent reorganization and compaction of the spermatozoon genome. Using a mouse model in which *Dot1l* is knocked-out (KO) in postnatal male germ cells, we found that *Dot1l*-KO sperm chromatin is less compact and has an abnormal content, characterized by the presence of transition proteins, immature protamine 2 forms and a higher level of histones. Proteomic and transcriptomic analyses performed on spermatids reveal that *Dot1l*-KO modifies the chromatin prior to histone removal and leads to the deregulation of genes involved in flagellum formation and apoptosis during spermatid differentiation. As a consequence of these chromatin and gene expression defects, *Dot1l*-KO spermatozoa have less compact heads and are less motile, which results in impaired fertility.

Keywords flagellum development; gene regulation; H3K79 methylation; histone-to-protamine transition; spermatogenesis

Subject Categories Chromatin, Transcription, & Genomics; Development; Post-translational Modifications & Proteolysis

DOI 10.15252/embr.202256316 | Received 17 October 2022 | Revised 31 March 2023 | Accepted 13 April 2023 | Published online 26 April 2023

EMBO Reports (2023) 24: e56316

Introduction

During the postmeiotic phase of spermatogenesis, male germ cells known as spermatids undergo profound morphological and functional changes to differentiate in spermatozoa. This process is driven by a rich genetic program characterized by the expression of thousands of genes in round spermatids (Soumillon *et al.*, 2013; da Cruz *et al.*, 2016; Chen *et al.*, 2018; Green *et al.*, 2018; Ernst *et al.*, 2019). Then, when spermatids elongate, their chromatin is extensively remodeled which ultimately results in the eviction of ~85–99% of nuclear histones and their replacement by more basic, smaller proteins, called protamines (for reviews, see Rathke *et al.*, 2014; Bao & Bedford, 2016). This unique chromatin reorganization induces a 6–10 times higher level of compaction of the spermatozoon chromatin compared to the canonical nucleosome-based chromatin (Ward & Coffey, 1991). Non-histone packaging of male germ cell genome is conserved throughout evolution and expected to be important to protect paternal DNA from damages and prepare

1 Université Paris Cité, INSERM, CNRS, Institut Cochin, Paris, France

2 Chromosomal Genomics Unit, Medical Genetics Department, Sorbonne Université et APHP, Hôpital Armand Trousseau, Paris, France

3 University Grenoble Alpes, CEA, INSERM, UA13 BGE, CNRS, CEA, FR2048, Grenoble, France

4 ADLIN Science, Pépinière « Genopole Entreprises », Evry, France

5 Molecular Biology of Reproduction and Development Research Group, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Fundació Clínica per a la Recerca Biomèdica, Universitat de Barcelona (UB), Barcelona, Spain

6 Université de Paris and Université Paris-Saclay, IRCM/IBF] CEA, UMR Stabilité Génétique Cellules Souches et Radiations, Laboratoire des Cellules Souches Germinales, Fontenay-aux-Roses, France

7 Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan

8 Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Département de Génétique Fonctionnelle et Cancer, CNRS, INSERM, Université de Strasbourg, Illkirch, France

9 Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands

10 MOABI-APHP Bioinformatics Platform-I&D-DSI, Assistance Publique-Hôpitaux de Paris, Paris, France

11 Biochemistry and Molecular Genetics Service, Clinic Barcelona, Barcelona, Spain

12 Service d'Histologie, d'Embryologie, Biologie de la Reproduction, AP-HP, Hôpital Cochin, Paris, France

*Corresponding author. Tel: +33 1 44 41 23 10; E-mail: julie.cocquet@inserm.fr

†These authors contributed equally to this work

its reprogramming in the zygote, if fertilization occurs (Rathke et al, 2014). A compact genome is also an advantage for the motile spermatozoa as it is compatible with a smaller and more hydrodynamic head (Braun, 2001).

In the past decades, the molecular mechanisms driving histone eviction and subsequent compaction of the sperm genome have been the focus of several studies which have demonstrated the essential role of transition proteins (Zhao et al, 2004a,b), of histone variants (Montellier et al, 2013; Barral et al, 2017), of histone post-translational modifications (PTMs), in particular hyperacetylation (Oliva et al, 1987; Goudarzi et al, 2016), and of writers and readers of histone acetylation (Shang et al, 2007; Gaucher et al, 2012; Dong et al, 2017; Shiota et al, 2018; Luense et al, 2019). Interestingly, high levels of histone H3 Lysine 79 di- and tri-methylation (H3K79me2 and me3) have been observed at the same time as histone hyperacetylation, just prior to histone removal (Dottermusch-Heidel et al, 2014a,b; Moretti et al, 2017; Fig 1A), but their biological significance remains to date unknown. H3K79 methylation is mediated by one enzyme, encoded by the gene *Dot1l*, of which pattern of expression and sequence are conserved from *Drosophila* to mammals. *Dot1l* has been implicated in development, cell reprogramming, differentiation, and proliferation (for review, see Kim et al, 2014; Vlaming & van Leeuwen, 2016). It has recently been found to be essential for spermatogonial stem cell self-renewal (Lin et al, 2022) but its role in postmeiotic male germ cells, where it is the most highly expressed (Dottermusch-Heidel et al, 2014a,b; Moretti et al, 2017), has not been addressed.

In the present paper, we investigated *Dot1l* role in postmeiotic cells by a comprehensive analysis of the phenotypic and molecular consequences of its knockout in mouse male germ cells. We found that *Dot1l* is required for gene regulation and chromatin reorganization in spermatids and that its knockout leads to the production of abnormally shaped nonfunctional spermatozoa with a deficient chromatin and nucleus compaction.

Results

Dot1l is essential for spermatogenesis and male fertility

In both mouse and human, *Dot1l/DOT1L* gene is highly expressed in the testis (Appendix Fig S1A). More precisely, it is expressed in all testicular cells with a strong expression at postnatal stage in meiotic cells (spermatocytes) and postmeiotic cells (spermatids; Appendix Fig S1A–C; Dottermusch-Heidel et al, 2014a,b; Moretti et al, 2017; Green et al, 2018). To investigate the role of *Dot1l* in male germ cells, we generated a conditional knockout using *Stra8-Cre* recombinase (Sadate-Ngatchou et al, 2008) and a mouse transgenic line in which *Dot1l* exon 2 is flanked by LoxP sites (Jo et al, 2011; Appendix Fig S1D). Previous studies have reported that *Stra8-Cre* is specifically expressed in the male germline where it is expressed from postnatal day 3 and CRE recombinase reaches maximum efficiency in pachytene spermatocytes (Sadate-Ngatchou et al, 2008; Bao et al, 2013). Bao et al (2013) have also observed that *Stra8-Cre* is more efficient at cutting one allele rather than two. We, therefore, generated mice in which one allele of *Dot1l* is floxed and one allele is already deleted (Δ) to increase efficiency of floxed exon excision upon *Cre* recombinase expression (*Dot1l^{flox}/ Δ* ; *Stra8-Cre*

mice, hereafter called *Dot1l*-KO or KO). First, to estimate the efficiency of *Dot1l* knockout, we detected DOT1L protein by western blot and immunocytochemistry on adult testes from *Dot1l*-KO (*Dot1l^{flox}/ Δ* ; *Stra8-Cre*) and controls (*Dot1l^{flox}/flox*; Fig 1B and C and Appendix Fig S1E–G). We found a reduction of > 85% of DOT1L signal but not complete abolishment/knockout of DOT1L level, despite strong activity of *Stra8-Cre* in most adult germ cells (Appendix Fig S1H). Analyses of purified germ cell fractions confirmed reduced level but not complete absence of DOT1L protein, along with a reduction of H3K79me2 level (Figs 1D and EV1 and Appendix Fig S1I). Multiple protein isoforms of DOT1L have been described in the literature and several of them were observed in our western blots, including the canonical (~ 165 kDa) and testis-specific (~ 185 kDa) isoforms (Zhang et al, 2004; Dottermusch-Heidel et al, 2014b), as well as less studied isoforms such as Q679P5 (~ 122kDa) or Q6XZL7 (~ 68kDa; <https://www.uniprot.org>). All of them were markedly reduced in *Dot1l*-KO testicular extracts (Appendix Fig S1E and I).

To investigate the impact of *Dot1l* knockout on spermatogenesis and male fertility, the reproductive parameters of adult *Dot1l*-KO males were investigated: significant decreases in testis weight and sperm count were observed (independently of body weight, Fig 2A and Appendix Fig S2A and B). Histological analyses revealed the presence of “empty” tubules without germ cells or missing germ cell layers (Fig 2B and Appendix Fig S2C) highlighting a spermatogonial defect as described by Lin et al (2022). However, there was no homogenous blockage/arrest at a specific stage of *Dot1l*-KO spermatogenesis. These observations were confirmed by cytometry analyses in which the number of all germ cell populations (i.e., spermatogonia with Side Population phenotype, spermatocytes, and spermatids) was found decreased in *Dot1l*-KO testes (Fig 2C), but their frequency was overall not altered (Appendix Fig S2F). Collectively, these data show that the *Dot1l*-KO we produced does not lead to a spermatogenesis arrest. When mated to WT females, *Dot1l*-KO males were found to be hypofertile with significant reductions in the number of litters and in litter size compared to that of control (CTL) males (Fig 2D and Appendix Fig S2D and E). We genotyped the few pups obtained and observed that the floxed allele was transmitted to 35% of babies (12 out of 34) and the deleted allele to 65% of them. This indicates that the deleted allele can be transmitted but less efficiently than what was expected based on our quantification of *Dot1l*-KO spermatids by immunofluorescence (i.e., only 65% of transmission of the deleted allele although > 85% of spermatids were devoid of DOT1L protein).

The reduction in litter number and size could result from impaired fertilizing ability of *Dot1l*-KO sperm or from a developmental defect. To address this question, we measured the fertilization rate 24 h after WT females were mated with *Dot1l*-KO or CTL males. The numbers of fertilized and non-fertilized oocytes collected in the oviduct ampulla were counted in both groups (a total of 113 oocytes were analyzed from 9 WT females mated with 4 CTL, and 73 oocytes from 6 WT females mated with 3 KO). A significant reduction in the fertilization rate was observed for *Dot1l*-KO indicating that *Dot1l*-KO sperm are less functional than CTL sperm (Fig 2E). In addition to *in vivo* fertility monitoring, we performed *in vitro* fertilization (IVF) assays using the same number of spermatozoa per experiment. Using spermatozoa from *Dot1l*-KO males, no oocyte was fertilized *in vitro*, while spermatozoa from control males

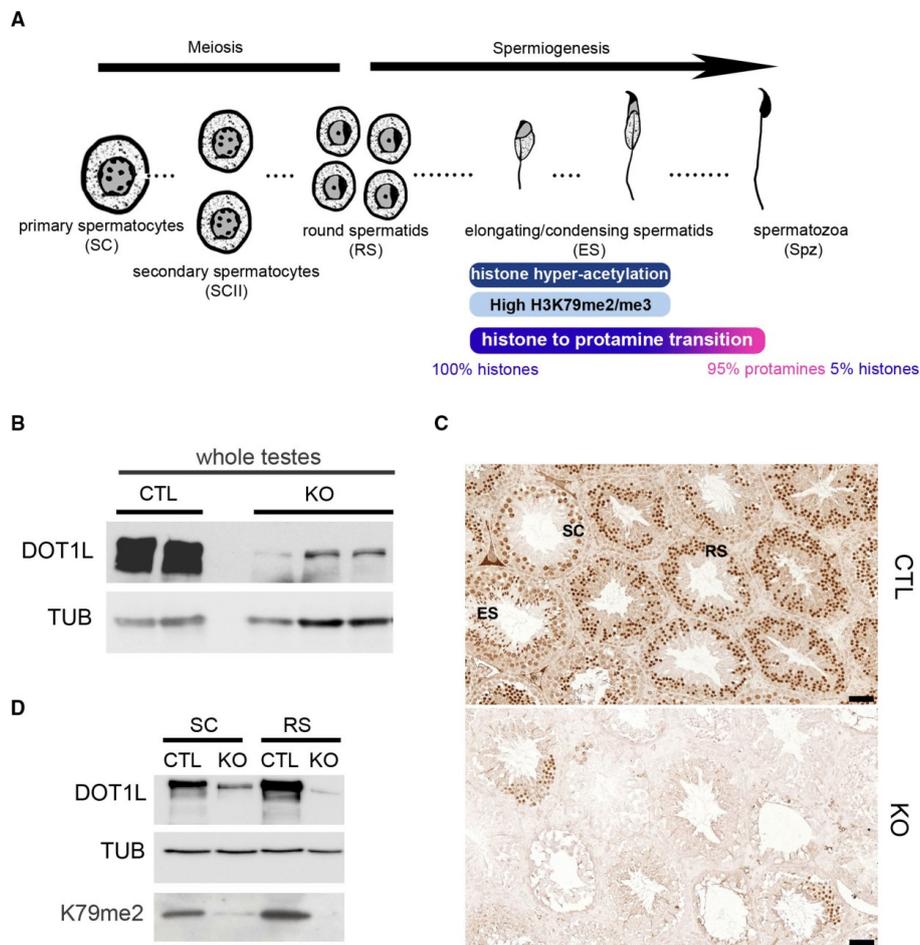


Figure 1. DOT1L expression in meiotic and postmeiotic germ cells of *Dot1l*-KO and CTL male mice.
 A Spermatogenesis scheme representing the dynamic of histone to protamine transition.
 B Western blot detection of DOT1L and TUBULIN (TUB) in whole testicular protein extracts from CTL and *Dot1l*-KO (KO) adult mice.
 C Immunohistochemistry detection of DOT1L in testicular sections from CTL and *Dot1l*-KO (KO) adult mice. ES, elongating spermatids; RS, round spermatids; SC, spermatocytes. Pictures were taken using the same parameters. Scale bars indicate 50 μ m.
 D Western blot detection of DOT1L and H3K79me2 in *Dot1l*-KO (KO) and CTL germ cells. Normalization was performed by detecting the membrane with anti-TUBULIN (TUB). SC = primary spermatocytes, RS = round spermatids.
 Data information: See also Fig EV1 and Appendix Fig S1.
 Source data are available online for this figure.

(including WT, CTL and HET males) led to the fertilization of ~ 37% of oocytes (Fig 2E, and see Appendix Fig S2G for individual values). When using oocytes devoid of zona pellucida, in order to bypass the zona pellucida crossing step, spermatozoa from *Dot1l*-KO males still performed very poorly compared to that of CTL males (~ 3% of fertilized oocytes using *Dot1l*-KO sperm compared to

~ 88% for CTL, Fig 2E), confirming that *Dot1l*-KO dramatically reduces sperm fertilizing ability. IVF is known to be less efficient and to lead to lower fertilization rates than *in vivo*. It is, therefore, more sensitive to sperm defects, as observed in other studies (Fujihara et al, 2010; Kawano et al, 2010; Barraud-Lange et al, 2020; Gadadhar et al, 2021).

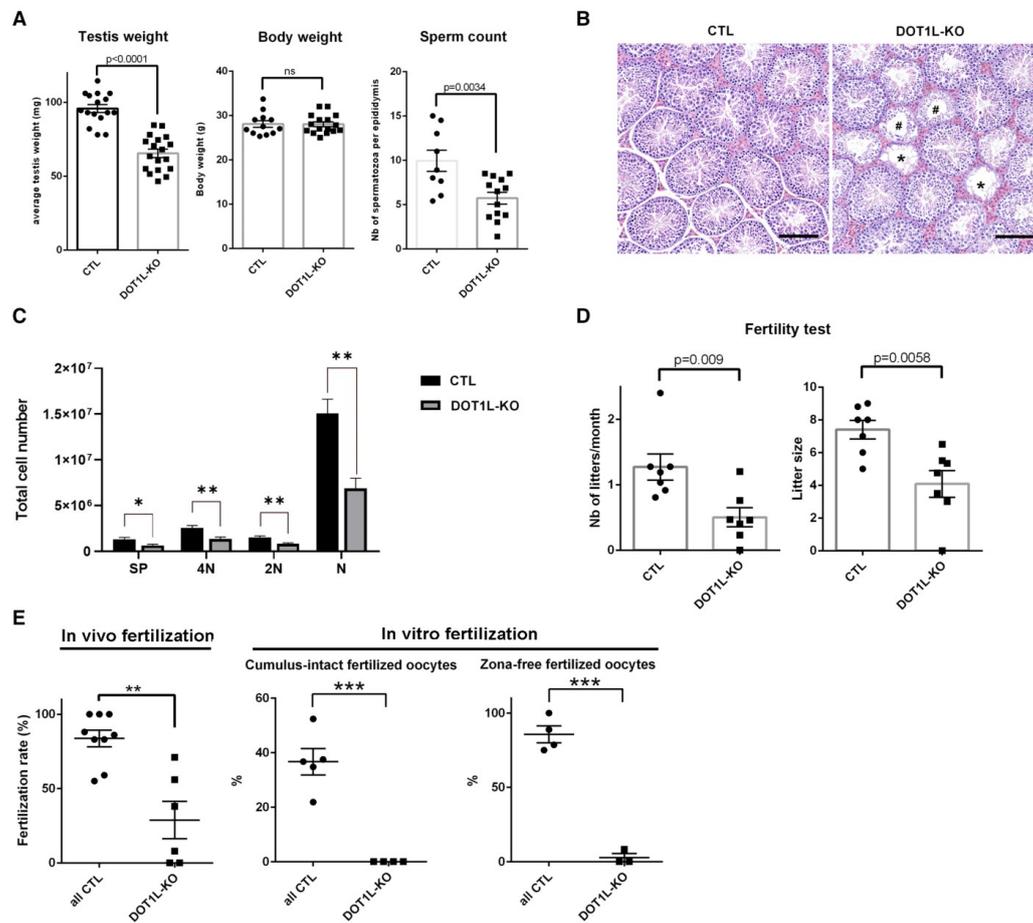


Figure 2. *Dot1l*-KO in male germ cells impairs spermatogenesis and male fertility.

A Scatter plots (mean ± standard error of the mean, SEM) showing the average testis weight, body weight and sperm count in *Dot1l*-KO and CTL adult mice (from ~ 2- to 4-month-old males, N > 10 biological replicates for CTL and *Dot1l*-KO).

B Histology of *Dot1l*-KO and CTL testes (periodic acid-Schiff staining). Scale bars indicate 200 µm.

C Schematic diagram representing total number of each cell type per testis in CTL and *Dot1l*-KO males, as calculated by FACS (mean ± SEM). 4N = Primary spermatocytes, 2N = Secondary spermatocytes, N = spermatids, SP = Side Population representing premeiotic germ cells (spermatogonia). Testes from 4 CTL and 3 *Dot1l*-KO males were analyzed (biological replicates). *P-value < 0.05 or **P-value < 0.01 (Student t-test).

D Results from the tests of fertility (natural mating) of *Dot1l*-KO and CTL males (seven males in each group mated for ~ 3 months with wild-type females). Schematic diagrams representing the number of litters and the litter size from progeny of *Dot1l*-KO or CTL males (mean ± SEM). The P-value obtained with a Student t-test is indicated on each graph.

E *In vivo* (left panel) and *in vitro* fertilization results. The percentages of fertilized oocytes are indicated for WT, CTL, and HET males ('all CTL') and for *Dot1l*-KO males. The left panel shows the mean (± SEM) fertilization rate of oocytes collected in superovulated females 24 h after natural mating. Nine and six females with a positive vaginal plug were analyzed after mating with CTL and *Dot1l*-KO males, respectively. This represents 113 and 73 oocytes analyzed after mating with CTL or KO males, respectively. The central and right panels show *in vitro* fertilization results (mean fertilization rate ± SEM) using oocytes with intact cumulus, or using oocytes of which the zona pellucida was removed beforehand. **P-value < 0.005 or ***P-value < 0.0005 (unpaired t-test performed after angular transformation of percentages).

Data information: See also Appendix Fig S2.
Source data are available online for this figure.

Dot1l-KO disrupts spermiogenesis and leads to the production of malformed nonfunctional spermatozoa

To determine the reason why *Dot1l*-KO spermatozoa are less functional than CTL ones, we investigated their morphology and motility. Spermatozoa are highly specialized cells characterized by a small head which encompasses their very compact nucleus, a long flagellum that confers motility and almost no cytoplasm. First, we observed that *Dot1l*-KO epididymal spermatozoa displayed malformed heads and abnormal flagella (characterized by thinning of some sections in their midpiece; Figs 3A and EV2A). Inside the flagellum, the axoneme normally contains a ring of nine outer microtubule doublets and two central microtubules (9 + 2 axoneme). In *Dot1l*-KO sperm, ultrastructure analyses revealed abnormal axonemes, characterized by disorganized microtubules (< 9 + 2 microtubules; Fig 3B). Nuclear compaction was also found impaired, observed as less contrasted nuclei in KO in comparison with CTL sperm. *Dot1l*-KO sperm also showed an increased cytoplasmic retention (Figs 3B and EV2A). Using CASA (computer-assisted sperm analyses), we observed that *Dot1l*-KO sperm motility was impaired, with a reduced proportion of motile and progressively motile (i.e., swimming in an overall straight line or in very large circles) spermatozoa (Figs 3C and EV2B). This was not due to cell death as sperm vitality was comparable between CTL and *Dot1l*-KO mice (Appendix Fig S3A). Axoneme/flagellum organization, nucleus compaction, and cytoplasm elimination occur during the differentiation of spermatids in spermatozoa (i.e., spermiogenesis). We measured apoptosis during this transition by TUNEL assay on testicular sections, and found a higher incidence of apoptotic elongating/condensing spermatids in *Dot1l*-KO than in CTL, another indication of a defective spermiogenesis (i.e., 6% of tubules containing apoptotic elongating/condensing spermatids in CTL vs. 40% of tubules in DOT1L-KO,

$P = 0.0007$, Fig 3D). Collectively, our data indicate that *Dot1l*-KO spermiogenesis is impaired at multiple levels and produces less functional spermatozoa.

Dot1l-KO modifies the chromatin of postmeiotic male germ cells, which results in abnormal chromatin reorganization and compaction in spermatozoa

During spermiogenesis, spermatid chromatin undergoes specific reorganization and compaction as most histones are removed and replaced with protamines (Rathke et al, 2014; Bao & Bedford, 2016). Histone variants (such as TH2B or H2AL2) and transition proteins (TNP) are highly expressed in spermatids during histone-to-protamine transition and are essential for the correct reorganization and compaction of the sperm genome with protamines (Montellier et al, 2011; Barral et al, 2017). In addition to the nuclear compaction defects observed by electron microscopy (Fig 3B), we noticed that *Dot1l*-KO sperm chromatin is more sensitive to nucleoplasmin-induced decompaction than CTL sperm chromatin. Nucleoplasmin, or NPM, is a histone chaperone known to facilitate paternal chromatin decompaction and reorganization after fertilization owing to its ability to bind histones and sperm nuclear basic proteins (Frehlick et al, 2007). Following NPM treatment, a higher proportion of soluble histones was found in *Dot1l*-KO compared to CTL sperm (Appendix Fig S3B) in agreement with a chromatin compaction defect. We next quantified TH2B and H3 histones in epididymal spermatozoa by western blot and, despite inter individual variability, we found significantly more residual histones in *Dot1l*-KO than in CTL sperm chromatin (Fig 3E and H and Appendix Fig S3C and D). We also quantified transition protein 2 (TNP2) and found a clear retention of this protein in *Dot1l*-KO spermatozoa while nothing could be detected in CTL sperm (Fig 3F and H and Appendix

Figure 3. *Dot1l*-KO in male germ cells leads to multiple spermiogenesis defects including flagellar abnormalities and incomplete sperm chromatin reorganization and compaction.

- A Representative pictures of *Dot1l*-KO and CTL epididymal spermatozoa. The black arrow indicates an abnormal sperm head (without apical hook) and white arrows indicate thinning of the flagellum. Scale bars represent 10 μ m.
- B Ultrastructure pictures (electron microscopy) from epididymal spermatozoa showing multiple abnormalities in *Dot1l*-KO compared to CTL: disorganized microtubules (blue arrow), impaired nuclear compaction (top panels) and increased cytoplasmic retention (bottom right panel). Scale bars represent 1 μ m.
- C Scatter plots (mean values following angular transformation \pm SEM) of the percentage of motile and progressively motile spermatozoa in *Dot1l*-KO and CTL adult mice ($N = 12$ biological replicates for CTL and 10 biological replicates for *Dot1l*-KO) obtained following CASA (computer-assisted sperm analysis). *** P -value < 0.0005 (Mann-Whitney test performed after angular transformation of the percentages).
- D Scatter plots showing the percentage of testicular tubules with TUNEL-positive elongating/condensing spermatids (ES) (mean per animal \pm SEM, $N = 5$ biological replicates for CTL and 6 biological replicates for *Dot1l*-KO). ** P -value < 0.005 (unpaired t -test performed after angular transformation of percentages). A representative picture of *Dot1l*-KO testicular sections following TUNEL assay is shown on the left. TUNEL positive elongating/condensing spermatids are visible inside the tubule (in green, white arrowhead). DAPI (blue) was used to stain nuclei. Scale bar indicates 20 μ m.
- E Western blot detection of histones H3 and TH2B in spermatozoa from *Dot1l*-KO and CTL males. The same quantity of material was loaded in each well (i.e., extracts from 2 million spermatozoa). NS indicates a non-specific band also observed on the membrane stained with Ponceau (see Appendix Fig S3C).
- F Western blot detection of transition protein 2 (TNP2) in spermatozoa from *Dot1l*-KO and CTL males. The same quantity of material was loaded in each well (i.e., extracts from 2 million spermatozoa).
- G Coomassie-stained protamine extracts from CTL and *Dot1l*-KO spermatozoa following acid urea gel electrophoresis (same gel, two different intensities). The same quantity of material has been loaded in each well (i.e., extracts from 1.4 million spermatozoa). Protamine 1 and 2 bands (PRM1 and PRM2, respectively) are detected at the bottom of the gel. The rectangle highlights the bands that are likely immature forms of Protamine 2 (i.e., non-cleaved precursor forms). The right panel shows a western blot detection using anti-PRM2 antibody which confirms that one of the high molecular weight band only observed in *Dot1l*-KO spermatozoa is an immature form of Protamine 2 (Pre-PRM2). The corresponding Coomassie-stained gel is also shown.
- H Quantification of histones TH2B and H3 ($N = 11$ and 10 biological replicates for CTL and KO, respectively), of TNP2 ($N = 8$ biological replicates for CTL and for KO) and of PRM1/PRM2 ratio ($N = 3$ biological replicates for CTL and for KO) in *Dot1l*-KO and CTL spermatozoa, related to Fig 3E-G. The data shown are normalized to control mean values (\pm SEM). * P -value < 0.05, *** P -value < 0.0005 (unpaired t -tests). A.U. = arbitrary units.

Data information: See also Fig EV2 and Appendix Fig S3.
Source data are available online for this figure.

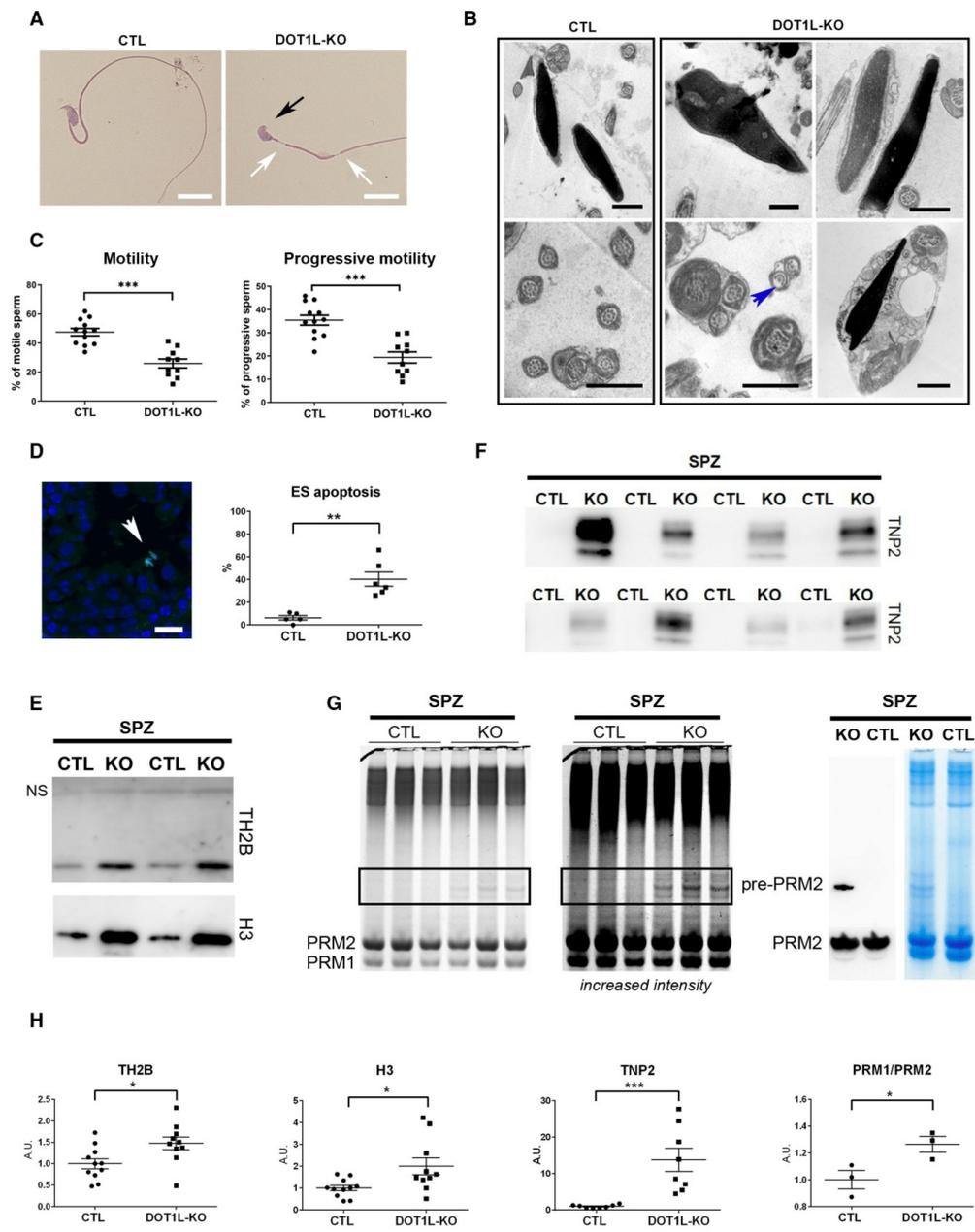


Figure 3.

Fig S3E). Finally, we quantified protamines on acid-urea gels and observed an accumulation of immature (non-cleaved) forms of Protamine 2 in *Dot1l*-KO spermatozoa versus CTL (Fig 3G). A slight but significant distorted Protamine1/Protamine2 (PRM1/PRM2) ratio was also observed in *Dot1l*-KO spermatozoa (Fig 3G and H and Appendix Fig S3F).

The transition from nucleosomes to protamine-based chromatin starts in elongating spermatids with the incorporation of histone post-translational modifications (PTMs) and histone variants (for reviews see Oliva, 2006; Rathke et al, 2014; Bao & Bedford, 2016). To better understand how DOT1L impacts on chromatin organization prior to histone removal, we quantified histone PTMs and variants by liquid chromatography–tandem mass spectrometry (LC–MS/MS) in whole testes and in elongating/condensing spermatids. In whole testis histone extracts, we observed, as expected, a clear decrease of H3K79 mono- and di-methylation in *Dot1l*-KO samples compared to CTL (Appendix Fig S4A). The tri-methylation of H3K79 was not detectable with this analysis, possibly due to its small stoichiometry. No other significant differences in histone H3 PTMs were detected between *Dot1l*-KO and CTL whole testicular histone extracts. However, a significant decrease in histone H4 hyperacetylated form (i.e., H4K5ac, K8ac, K12ac, K16ac) was observed (Appendix Fig S4A). In elongating/condensing spermatids (ES), the stage during which chromatin is extensively remodeled, a significant decrease in H3K79 mono- and di-methylation levels was also observed (Fig 4A). Here also, H3K79 tri-methylation was not detectable. Several other quantitative changes in histone PTMs were not only observed, such as modest increases in H3K23ac, H2AR88me1, and TH2BK35ac, but also a dramatic decrease of H4K20 or R23 mono-, di-, and tri-methylated forms (Fig 4 and Appendix Fig S4B). Quantification of histone variants showed significant changes in H1 variants and in canonical H2A (Appendix Fig S4C). The most remarkable change was, as in whole testes, a severe decrease in H4 acetylation in *Dot1l*-KO elongating/condensing spermatids, in particular, hyperacetylated forms such as H4K5ac, K12ac, K16ac, H4K8ac, K12ac, K16ac, and H4K5ac, K8ac, K12ac, K16ac (Fig 4B). In elongating spermatids, histone H4 hyperacetylation is an essential step of histone-to-protamine transition, as it facilitates chromatin loosening and nucleosome disassembly prior to histone removal (for review see Oliva, 2006). By western blot and immunofluorescence, we also observed a decrease in H4 acetylation in *Dot1l*-KO elongating/condensing spermatids compared to CTL using anti-H4K16ac or anti-poly H4ac antibody (Fig EV3 and Appendix Fig S4D).

We next quantified histone PTMs and variants at an earlier stage of spermiogenesis, in round spermatids (RS), and did not see any decrease in H4 acetylation or H4K20 methylation in KO versus CTL RS (Fig EV4) in contrast with what was observed in ES (Fig 4). Apart from the expected decrease in H3K79 methylation, the only quantitative changes that were detected in RS were increases in H3K18 crotonylation (H3K18cr), in H4K8cr, K12ac, K16ac and in H1 variants (Fig EV4 and Appendix Fig S4C and E). In spermatozoa, some H3K79 methylation has been detected on persistent histones (Luense et al, 2016; Moretti et al, 2017) and a significant decrease in H3K79me1 and H3K79me2 levels was visible in *Dot1l*-KO spermatozoa using LC–MS/MS. Apart from a modest increase in H3K18ac, no other changes were detected at this stage (Appendix Fig S4F).

All these observations indicate that *Dot1l* KO has a profound impact on postmeiotic chromatin in ES at the time of its

reorganization, which disrupts histone-to-protamine transition and leads to a higher proportion of retained histones, transition proteins, and unprocessed protamine 2 in spermatozoa.

DOT1L regulates the expression of genes involved in essential processes of spermatid differentiation

In order to gain insight into the gene expression changes associated with *Dot1l* deficiency in male germ cells, we performed RNA-Seq analyses on purified germ cells from three different spermatogenic stages, each in five replicates: primary spermatocytes (SC), secondary spermatocytes (SCII), and round spermatids (RS; Fig 5A). We did not include elongating/condensing spermatids in this analysis because their transcriptome is very similar to that of round spermatids, and, at this stage, transcription progressively shuts down as a consequence of histone-to-protamine transition and genome compaction. RNA-Seq differential analyses showed hundreds of significantly deregulated genes in *Dot1l*-KO versus CTL samples (DESeq2, with a Fold change > 1.5 and FDR < 0.05). In parallel, we performed a less stringent analysis (DESeq2, with a Fold change > 1.5 and $P < 0.05$, then *a posteriori* validation of FDR) and found between ~ 600 and 1,000 deregulated genes at each stage (Fig 5B and Appendix Fig S5A). With both analyses, the number of deregulated genes increased with progression of male germ cell differentiation and was therefore highest in RS (Fig 5B and Appendix Fig S5A), and approximately half of deregulated genes were found in common in at least two stages (Appendix Fig S5A). In the three cell types, more genes were found upregulated than downregulated in KO samples (Fig 5B) in agreement with results obtained recently in *Dot1l*-KO somatic cells (Aslam et al, 2021; Cattaneo et al, 2022). To link gene deregulation in *Dot1l*-KO mice with DOT1L methyltransferase activity, we compared deregulated genes with the genomic location of H3K79me2 that we obtained by ChIP-seq in RS. H3K79me2 is the best characterized H3K79 methylation mark and is known to mark the body of transcriptionally active genes. H3K79me1 and me3 have similar trends, though broader for the former, and more restricted for the latter (Vlaming & van Leeuwen, 2016). Here, in wild type RS, H3K79me2 was found enriched at the body (with the strongest signal found near the Transcriptional Start Site) of genes which were downregulated in the KO but not those found upregulated (Fig 5C and Appendix Fig S5B). Observation of ChIP peaks confirmed the presence of H3K79me2 mark at down but not upregulated genes (Fig 5D).

We next investigated deregulated pathways using GSEA (Gene Set Enrichment Analysis), a threshold-free method which considers all genes based on their differential expression rank and found a downregulation of biological pathways related to apoptosis, transcription/RNA regulation, chromatin/chromosome organization, and mitochondria activity in *Dot1l*-KO versus CTL round spermatids (Fig 5E). Prior to spermiogenesis, pathways related to cilium or flagellum motor activity were found downregulated in *Dot1l*-KO primary spermatocytes while, in *Dot1l*-KO secondary spermatocytes, pathways related to chromatin/chromosome organization were the most significantly deregulated (Fig 5E). When looking at gene ontology for “cellular components,” pathways related to ribosomal structure or to mitochondrial proteins were found deregulated in RS (Fig EV5). Strikingly, several *Slc* genes were found deregulated in *Dot1l*-KO RS (Appendix Fig S5C). *Slc* genes encode solute carriers,

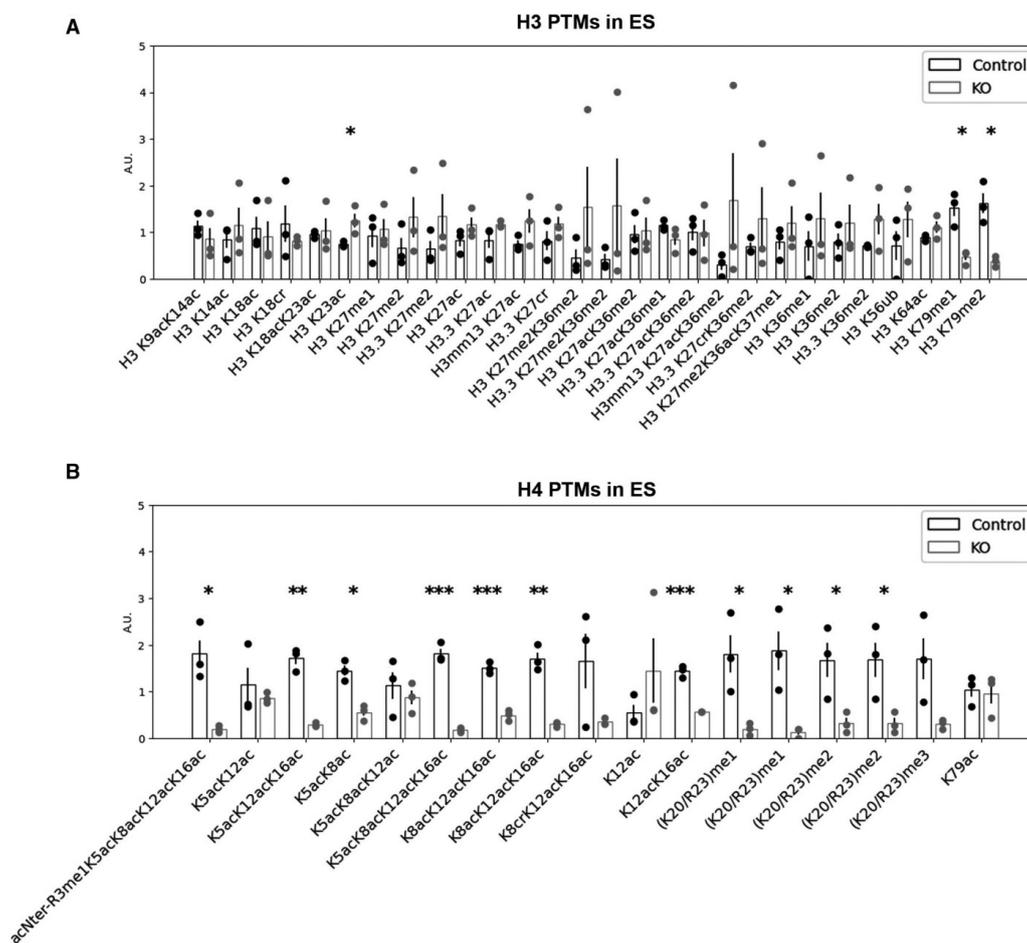


Figure 4. *Dot1l*-KO modifies the chromatin of elongating/condensing spermatids.
 A, B Bar plots showing the quantification of post-translational modifications (PTMs) in histones H3 and H4 in ES (elongating/condensing spermatids) by LC-MS/MS. After normalization to be at constant amount of histone H3 or H4 in each analyzed sample (see Materials and Methods), mass spectrometry signals were divided by the average signal in both conditions (CTL and KO), so as to be able to represent all peptides in the same figure, whatever their MS intensity. Quantitative data obtained on biological replicates ($N = 3$ for CTL and 3 for KO) were plotted as dots, and the height of the bars indicates the average value (mean \pm standard deviation, SD). T -tests were performed to estimate the possible significant difference of modified peptide abundances between the two genotypes. * P -value < 0.05 , ** P -value < 0.005 , and *** P -value < 0.0005 . When interpreting our mass spectrometry data against the database MS_histoneDB, we were able to identify various H3 variants, namely canonical H3 and variants H3.3, testis-specific H3t and H3mm13. "H3" indicates sequences shared among several variants; "H3.3," "H3t," or "H3mm13," each variant. A.U. = arbitrary units.
 Data information: See also Figs EV3 and 4 and Appendix Fig S4.
 Source data are available online for this figure.

and several of them have been implicated in sperm motility via their effects on flagellum differentiation and/or sperm energy production (Maruyama et al, 2016; Toure, 2019; Kuang et al, 2021).

In agreement with the downregulation of pathways related to apoptosis, *Bcl6*, *Jak3*, and *Tsc22d3* which encode proteins with anti-

apoptotic effects (Thomis et al, 1997; Kurosu et al, 2003; Aguilar et al, 2014) were found significantly downregulated in RS. *Bcl6* codes for a transcription repressor (Chang et al, 1996) and appears to be a direct target of DOT1L since it is downregulated upon *Dot1l* loss, and enriched in H3K79me2 in wild-type RS (Fig 5E).

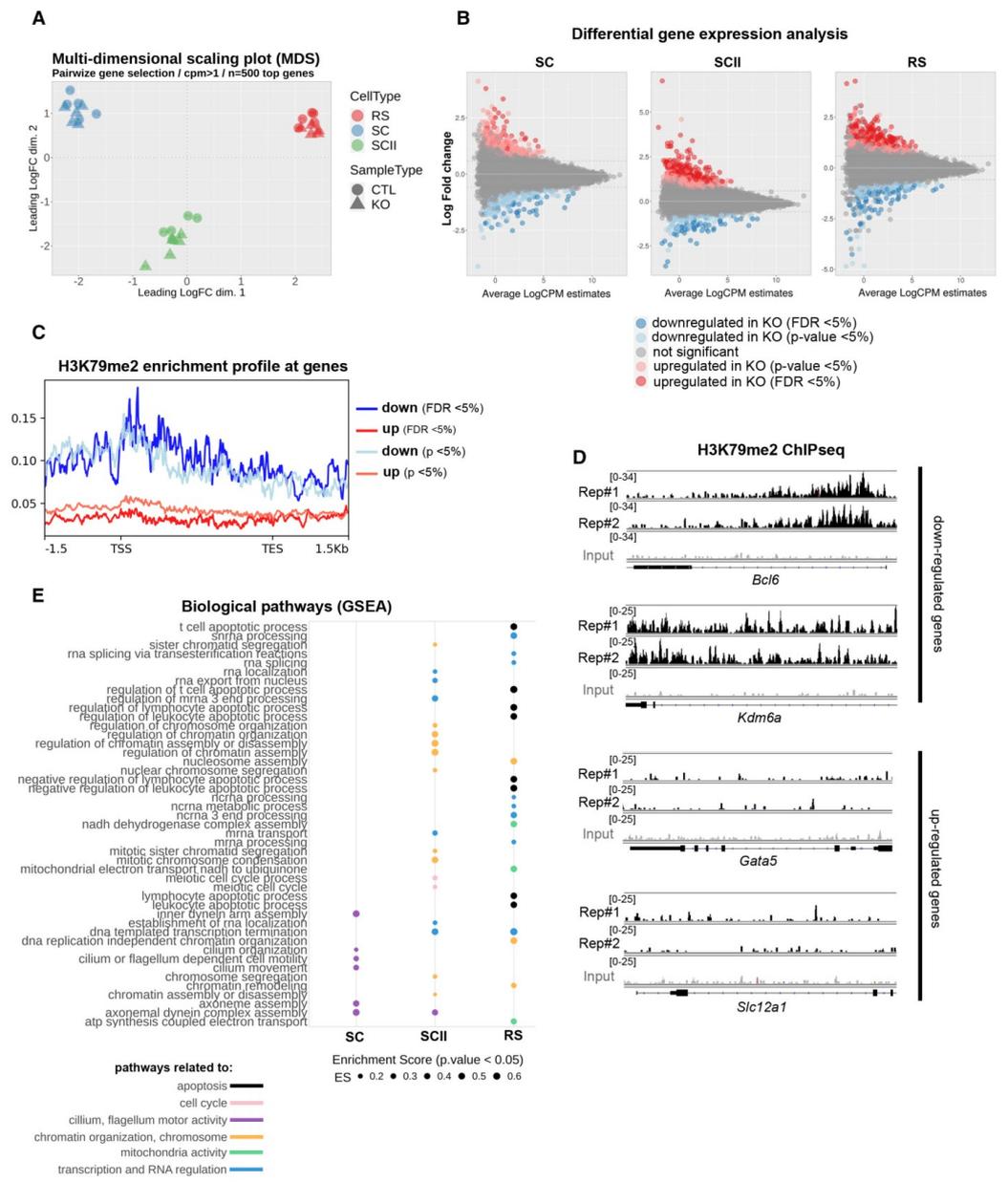


Figure 5.

Figure 5. *Dot1l*-KO leads to the deregulation of genes involved in flagellum function, transcription regulation, chromatin organization, mitochondria function and apoptosis.

A Multi-dimensional scaling plot on the 500 top genes in *Dot1l*-KO versus CTLs primary spermatocytes (SC), secondary spermatocytes (SCII), and round spermatids (RS) using pairwise gene selection.

B MD plot of the genes found deregulated in *Dot1l*-KO versus CTLs SC, SCII, and RS, using differential gene expression analyses with two different parameters (FDR < 0.05 and P-value < 0.05).

C H3K79me2 enrichment profile at deregulated genes (same parameters as in B).

D H3K79me2 enrichment profile at selected deregulated genes, in two biological replicates of ChIP-Seq performed in wild type RS (Rep#1 and 2). The input value is shown in gray.

E GSEA analysis of RNA-Seq data from *Dot1l*-KO versus CTL SC, SCII and RS. The figure shows all the biological pathways found significantly downregulated in *Dot1l*-KO primary spermatocytes (SC), secondary spermatocytes (SCII), and round spermatids (P < 0.05), ranked by their enrichment score (ES).

Data information: See also Fig EV5 and Appendix Fig S5.

Interestingly, an unsupervised search for DNA motifs identified BCL6/BCL6B DNA binding site as the most significantly enriched motif in the promoter of upregulated genes (15 upregulated genes, Appendix Fig S5D). This finding is in agreement with a model in which downregulated genes are direct H3K79me2/DOT1L targets while upregulated genes are indirect targets.

Discussion

In the present study, using a conditional knockout mouse model, we demonstrate that the H3K79 methyltransferase DOT1L is essential for gene regulation and chromatin remodeling during spermatid differentiation.

DOT1L is highly conserved among eukaryotes. In mammalian it has been extensively studied and found to be involved in many different biological processes such as mixed lineage leukemia (Okada et al, 2005; Nguyen et al, 2011), cell cycle (Kim et al, 2014), development (Jones et al, 2008), reprogramming (Onder et al, 2012), DNA damage repair (Lin et al, 2009; Zhu et al, 2018), or transcription activation (Steger et al, 2008; Wang et al, 2008). In the male germline, DOT1L is highly expressed at the RNA and protein levels, particularly in postmeiotic cells (Dottermusch-Heidel et al, 2014a,b; Moretti et al, 2017). Before meiosis, DOT1L has been found to be required for the self-renewal of adult stem cells. Indeed, with a *Dot1l* KO induced by a Cre recombinase expressed earlier than the one used in the present study, Lin et al (2022) observed a progressive (age-dependent) loss of all male germ cell types, ultimately resulting in a Sertoli cell only syndrome (i.e., absence of germ cells). In the *Dot1l*-KO males we characterized here, we also found empty tubules as well as a decrease of all germ cell types, but milder, which enabled us to address the role of DOT1L in adult postmeiotic germ cells (i.e., spermatids). Strikingly, we found that *Dot1l*-KO spermatozoa present multiple anomalies such as thinner and distorted flagella, cytoplasmic retention and impaired nuclear compaction. As a result, KO spermatozoa are less motile and their fertilizing ability is compromised. These defects are associated with the deregulation of hundreds of genes in *Dot1l*-KO meiotic and post-meiotic germ cells. Previous studies have shown that spermatid differentiation is associated with the specific expression (or highly enriched expression) of thousands of genes, and that this genetic program starts as early as in primary spermatocytes (Soumillon et al, 2013; da Cruz et al, 2016; Chen et al, 2018; Green et al, 2018; Ernst et al, 2019). Here, our analyses indicate that pathways related to “cilium/flagellum assembly or motility” are deregulated from the

spermatocyte stage, and those related to “nucleosome assembly, chromatin remodelling” are deregulated in secondary spermatocytes and round spermatids. A significant proportion of *Dot1l*-KO spermatids undergo apoptosis and this was also visible by RNA-Seq with the deregulation of “apoptosis” related pathways in RS, including the downregulation of genes with anti-apoptotic effects such as *Bcl6* or *Jak3* (Thomis et al, 1997; Kurosu et al, 2003). Interestingly, several potential BCL6 target genes were identified among the genes which are upregulated in *Dot1l*-KO spermatids. Like other upregulated genes, most of them are devoid of H3K79me2 in wild type spermatids and may not be directly regulated by DOT1L. In contrast, *Bcl6* and many other downregulated genes in *Dot1l*-KO spermatids are marked by H3K79me2 in wild-type cells and are likely directly regulated by DOT1L. Our findings are in agreement with recent studies in which *Dot1l* knockout was investigated in B cells, T cells or cardiomyocytes. Each time, H3K79me2 was found to be enriched at the body of downregulated genes but not of upregulated genes (Kwesi-Maliepaard et al, 2020; Aslam et al, 2021; Cattaneo et al, 2022).

Strikingly, many genes encoding members of the solute carrier family were deregulated in *Dot1l*-KO male germ cells. Several members of this family such as *Slc22a14* and *Slc26a8* have previously been implicated in sperm motility via their effects on flagellar differentiation and/or sperm energy production (Maruyama et al, 2016; Toure, 2019; Kuang et al, 2021); the deregulation of *Slc* genes with similar function could, therefore, contribute to the sperm flagellar and motility defects of *Dot1l*-KO males. Overall, these data indicate that *Dot1l*-KO-induced gene deregulation is responsible for the multiple defects of postmeiotic male germ cell differentiation.

In addition to being involved in the regulation of spermatid gene expression, we show that DOT1L is essential for spermatid chromatin remodeling and subsequent reorganization and compaction of spermatozoon genome. Indeed, the chromatin of *Dot1l*-KO spermatozoa is less compact, and characterized by a higher level of retained histones, of transition protein 2, and of immature forms of protamine 2, as well as altered protamine 1/2 ratio. Several other studies have observed similar anomalies in particular retained histones and immature forms of protamine 2, as a result of incomplete histone-to-protamine transition, such as *Trnp1*-KO (*transition protein 1*; Yu et al, 2000), *H2a2*-KO (*H2A histone family member L2A*; Barral et al, 2017), deletion of the mouse Y chromosome long arm (Yamauchi et al, 2010), or *Kat2a* -KO (a.k.a. *Ccn5*) which encodes the lysine acetyltransferase 2A (Luense et al, 2019). Similarly, a recent study has shown that aberrant pre-PRM2 cleavage leads to altered levels of histone variants and transition proteins in sperm

(Arevalo et al., 2022). In all those mouse models as well as in the *Dot1l*-KO presented here, abnormal sperm chromatin content is associated with fertility problems. This is also the case in humans, since impaired protamine 1/2 ratio and accumulation of pre-PRM2 are associated with male infertility (Torregrosa et al., 2006; de Mateo et al., 2009).

In elongating spermatids, histones—in particular histone H4—are hyperacetylated prior to their replacement. This phenomenon is a prerequisite for histone-to-protamine transition as it creates a permissive state for histone removal (for reviews see Oliva, 2006; Rathke et al., 2014; Bao & Bedford, 2016). H3K79me2/3 levels peak at the time of H4 hyperacetylation (Dottermusch-Heidel et al., 2014a, b; Moretti et al., 2017) and we show here that both H3K79 methylation and H4 hyperacetylation are drastically and specifically reduced in *Dot1l*-KO elongating spermatids. Several acetyltransferases are involved in H4 hyperacetylation such as KAT2A, CBP-P300, or NuA4 complex (Boussouar et al., 2014; Shiota et al., 2018; Luense et al., 2019) but none of them was found deregulated in *Dot1l*-KO spermatids. We, therefore, propose that the decrease in histone acetylation is a consequence of impaired H3K79 methylation. In other contexts than the male germline, H4 acetylation and H3K79 methylation have previously been shown to influence each other in different ways. In 2016, Gilan et al. proposed that DOT1L-mediated H3K79me2 leads to a more open chromatin which facilitates the recruitment of the histone acetyltransferase P300 which, in turn, increases histone H4 acetylation (in particular, H4K5ac), in a cell model of acute myeloid leukemia (Gilan et al., 2016). In contrast, H4K16ac has been shown to stabilize the binding of yeast Dot1 and human DOT1L on the nucleosome and consequently stimulates its methyltransferase activity *in vitro* (Valencia-Sanchez et al., 2021). In elongating spermatids, H3K79 methylation and H4 acetylation could therefore contribute to increase each other signal. It would be interesting to test if H3K79 methylation level is affected in spermatids in which histone acetylation is impaired, for instance in *Kat2a* knockout spermatids.

In our study, we also observed a decrease in the three methylation states of H4K20 in elongating spermatids. This is in line with the observation made by Jones et al. (2008) that H4K20 trimethylation level is reduced at centromeres and telomeres of *Dot1l*-KO embryonic stem cells. H3K9 di-methylation (but not trimethylation) was also found reduced in KO embryonic stem cells, but this histone PTM could not be detected in our analyses. Methylation of H4K20 has been shown to be associated with chromatin accessibility. H4K20me1, by changing the conformation of nucleosomes, increases accessibility for other factors involved in DNA processing and expression (Shoib et al., 2021). In contrast, H4K20me2 and H4K20me3 are associated with chromatin compaction (Everts et al., 2013). For the moment, the impact of these marks during spermatid differentiation is unknown, but H4K20me3 is one of the most dynamic histone modifications during spermatogenesis, with a peak in elongating spermatids; H4K20me2 follows the same pattern but is less abundant than H4K20me3 (Luense et al., 2016; Wang et al., 2021). One can, therefore, speculate that H4K20 methylation is also important for histone to protamine transition. Interestingly, Ho et al. (2021) have found that the H4K20me1 methyltransferase, SET8, and DOT1L (which is a non-SET domain methyl-transferase) bind to the nucleosomal acidic patch by a similar mechanism. Further work will be needed to understand the

mechanism by which DOT1L and/or H3K79 methylation may influence H4K20 methylation.

Other, more modest quantitative changes in histone PTMs and variants were observed in elongating spermatids. They could be compensatory mechanisms for H3K79 methylation or H4 acetylation decrease, as observed in *Th2b* knock-out spermatids where H2B was upregulated and several histone PTMs changed to compensate for the loss of H2B variant TH2B (Montellier et al., 2013). The deregulation of genes associated to “chromatin function” could also contribute to the changes in chromatin content observed in *Dot1l*-KO spermatids.

In conclusion, our data show that DOT1L is indispensable for the differentiation of postmeiotic germ cells into functional spermatozoa. Indeed, DOT1L regulates the expression of genes essential for spermatid differentiation and, DOT1L-mediated H3K79 methylation is necessary for proper histone H4 hyperacetylation in elongating spermatids and for an efficient histone-to-protamine transition and subsequent compaction of the sperm chromatin.

Materials and Methods

Mouse strains and genotyping

Conditional *Dot1l* knockout mice were obtained from the Knockout Mouse Project (CSD29070; to F.V.D.). This line has previously been described in (Jo et al., 2011). In these mice, *Dot1l* exon 2 is flanked by loxP sites (see Appendix Fig S1D). Recombination of loxP sites by Cre recombinase leads to the deletion of exon 2 and consequently a frameshift and premature stop codon in *Dot1l* Coding DNA sequence. Tg(*Stra8-iCre*)1Reb/J were obtained from the Jackson lab (JAX stock #017490; Sadate-Ngatchou et al., 2008). All animals analyzed in this study were from a C57BL6/J background. Genotypes of animals were determined by polymerase chain reaction (PCR) using *Dot1l* primers F + R1 + R2 (Appendix Fig S1D) or *iCre* primers as described on the Jackson lab website. The latter PCR was performed with *Ymtx* internal control (see Appendix Table S1 for primer list and sequences).

Unless specified, all experiments were performed on (adult) 3–5-month-old males. Reproductive parameters were analyzed on two types of *Dot1l* knocked-out animal models. First, on animals hosted in a conventional animal house (cah) and of the following genotypes: *Dot1l^{fl/fl}*; *Stra8-Cre* mice (DOT1Lcah) and *Dot1l^{fl/fl}* siblings (without *Stra8-Cre* transgene) as controls (CTLcah). Second, on animals hosted in a Specific Pathogen-Free (SPF) animal house and of the following genotypes: *Dot1l^{fl/Δ}*; *Stra8-Cre* (hereafter called DOT1L-KO) and *Dot1l^{fl/fl}* siblings as controls (hereafter called CTL). Analyses of the reproductive parameters of both types of KO males (*Dot1l^{fl/fl}*; *Stra8-Cre* or *Dot1l^{fl/Δ}*; *Stra8-Cre*) gave similar results (Appendix Fig S2A and B). Heterozygous *Dot1l^{fl/Δ}* males (HET) were also analyzed. They did not differ from controls (Appendix Fig S2A–E). *Dot1l^{fl/Δ}*; *Stra8-Cre* animals were generated because it had been shown that *Stra8-iCre* is not 100% efficient (Bao et al., 2013). In these animals, one allele of *Dot1l* is floxed and one allele is already deleted (Δ) to increase the efficiency of floxed exon excision upon Cre recombinase expression. Nevertheless, some remaining DOT1L protein signal was observed by western blot and Immunofluorescence (see Fig 1B and D and Appendix Fig S1E–I).

All other phenotypic analyses and all molecular analyses were performed on *Dot11^{Fl/Δ}*; *Stra8-Cre*, hereafter called *Dot11-KO*, and on *Dot11^{Fl/Fl}* siblings as CTL. Mice from mTmG line (Muzumdar et al, 2007) were crossed to *Stra8-Cre* line to verify *Stra8-Cre* recombinase location and activity.

The mice were fed ad libitum with a standard diet and maintained in a temperature- and light-controlled room. Animal procedures were approved by Université de Paris ethical committee (Comité d’Éthique pour l’Expérimentation Animale; registration number CEEA34.JC.114.12, APAFIS 14214-201707251044852v26).

Fertility tests

Dot11-KO and CTL males, aged from 8 to 10 weeks, were housed with two wild-type C57BL6/J females per cage (aged from 6 weeks, Janvier Labs, France) for up to 4 months. Vaginal plugs were checked and females separated when found. For each group, litter size and number were assessed.

Another series of test were performed to evaluate the fertilization rate following mating (*in vivo* fertilization). For this WT females were first superovulated by injection of 5 IU of PMSG (pregnant mare’s serum gonadotrophin; Intervet, France) followed by 5 IU of hCG (human chorionic gonadotrophin; Intervet) 48 h later. Females were then mated with WT or *Dot11-KO* males at night and the next day females with a vaginal plug were isolated. Oocytes were collected from ampullae of oviducts and freed from the cumulus cells by brief incubation at 37°C with hyaluronidase (Sigma, St. Louis, MO, USA) in M2 medium (Sigma). Oocytes were rinsed and the number of fertilized oocytes was evaluated using DAPI immunofluorescent staining to visualize the DNA.

IVF

Oocyte preparation

WT C57BL6/J female mice aged 6–8 weeks (Janvier Labs, France) were superovulated with 5 IU of pregnant mare serum gonadotropin (PMSG) and 5 IU human chorionic gonadotropin (hCG; Intervet, France) 48 h apart. About 14 h after hCG injection, animals were sacrificed by cervical dislocation. Cumulus oocyte complexes were collected by tearing the ampulla wall of the oviduct, placed in Fertilcult medium (FertiPro N.V, Belgium) supplemented with 3% BSA (Sigma-Aldrich), and maintained at 37°C under 5% CO₂ in air under mineral oil (FertiPro N.V, Belgium). When experiments were performed with Zona-free oocytes, cumulus cells were first removed by a brief exposure to hyaluronidase IV-S (1 mg/ml, Sigma-Aldrich). The zona pellucida was then dissolved with acidic Tyrode’s (AT) solution (pH 2.5; Sigma-Aldrich) under visual monitoring. Zona-free eggs were rapidly washed five times and kept at 37°C under 5% CO₂ in air for 2–3 h to recover their fertilization ability.

Capacitated sperm preparation: mouse spermatozoa were obtained from the cauda epididymides of DOT1L-KO, CTL, HET and WT C57BL6/J males (aged 8–12 weeks) and capacitated at 37°C under 5% CO₂ for 90 min in a 500 µl drop of Fertilcult medium supplemented with 3% BSA, under mineral oil.

In vitro fertilization: cumulus-intact and Zona-free eggs were inseminated with capacitated spermatozoa for 3 h in a 100 µl drop of Fertilcult medium, 3% BSA at a final concentration of 10⁶ or 10⁵ per ml, respectively. Then, they were washed and directly mounted

in Vectashield/DAPI (Vector laboratories, CA, USA) for observation under UV light (Nikon Eclipse E600 microscope). Only oocytes showing at least one fluorescent decondensed sperm head within their cytoplasm were considered fertilized.

Testicular germ cell purification by elutriation or FACS

Germ cells were purified from adult males using fluorescence-activated cell sorting (FACS) or elutriation as previously described in (Cocquet et al, 2009; Comptour et al, 2014) and more recently in (Crespo et al, 2020). Elutriated fractions with a purity ~ 95–99% for elongating/condensing spermatids and ~ 80% for round spermatids, were used for LC-MS/MS analyses. Highly enriched fractions of primary or secondary spermatocytes (> 90%) and round spermatids (~ 99%) were used for RNA-Seq analyses. Flow cytometric analysis of testicular cell suspension were performed as previously described (Corbineau et al, 2017; Ragazzini et al, 2019).

Sperm collection and purification with Percoll

Spermatozoa were extracted from cauda epididymis in pre-warmed (37°C) M2 medium (Sigma-Aldrich) by gentle pressure. Then, epididymides were perforated with a thin needle and incubated in M2 for 10 min at 37°C to allow remaining sperm to swim up. For molecular analyses, spermatozoa were next purified on a Percoll gradient. In brief, spermatozoa were centrifuged at 2,000 g for 10 min and pellets incubated in somatic cell lysis buffer (1× PBS, 0.1% SDS, 0.5% Triton X-100) for 10 min on ice. Sperm cells were washed with PBS-BSA (1× PBS, 0.5% BSA, 2 mM EDTA) and briefly sonicated to remove flagella (ON 5 s – OFF 30 s × 3 Cycles, bioruptor Pico, Diagenode). The samples were transferred in low retention tubes, loaded on 50% Percoll (Sigma-Aldrich) and centrifuged at 2,500 g for 5 min to remove somatic cells and flagella. This step was repeated once and sperm cells were washed in PBS-BSA twice and counted (minimum 1,000 cells). The purity was > 99.7%. Spermhead pellets were snap-frozen in liquid nitrogen and stored at –80°C prior to use.

Nucleoplasmin decompaction

Experiments were performed as previously described in (Yamaguchi et al, 2018). Following Percoll purification, sperm heads were washed in 1× PBS, 0.5% Tween, permeabilized with 0.02 U/M Streptolysin O (Sigma-Aldrich) in 1× PBS for 10 min on ice and washed once with 1× PBS, 0.5% Tween. Cells were then incubated with 1× PBS containing 1 mM DTT for 10 min at 37°C, washed with 1× PBS, 0.5% Tween, resuspended in KH buffer (100 mM KCl, 20 mM Hepes-KOH, pH 7.7) in low retention tubes and stored at –80°C for later use. Ten million of sperm heads were resuspended in 1 mL NPM treatment buffer (20 mM HEPES-KOH, pH 7.7, 100 mM KCl, 2.5 mM MgCl₂, 5 mM Na-But, 10 mM DTT, 1X complete EDTA-free protease inhibitors, 250 µM NPM) and incubated for 2 h at 37°C in a Thermomixer (Mixer HC, Starlab) at 1,000 rpm. After centrifugation at 20,000 g for 5 min at 4°C, cells were washed with 5 mM Na-But in 1× PBS, 0.5% Tween twice and sperm heads were fixed with 1% PFA for 10 min at RT, quenched with 250 mM glycine and washed in 5 mM Na-But in PBS/Tween 0.5% twice.

After fixation, sperm pellets were resuspended in NP40 Buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.5% NP40, 1× complete

EDTA-free protease inhibitors, Sigma Aldrich) and incubated on ice for 30 min. Samples were washed with 1× PBS, 0.5% Tween, 5 mM Na-But, and resuspended in 200 µl of sonication buffer (10 mM Tris-HCl, pH 8.0, 20% glycerol, 0.25% SDS, 5 mM Na-But, 1× complete EDTA-free protease inhibitors). Sperm cells were sonicated (ON 30s – OFF 30s x13 cycles, Bioruptor Pico, Diagenode) and the sheared chromatin was centrifuged at 20,000 g for 10 min at 4°C. Supernatants were transferred in new 1.5 ml tubes and pellets resuspended in 200 µl of KH buffer. Prior to western blot analyses, both supernatant and pellets were resuspended in 4× NuPage/10% β-mercapto-ethanol (ThermoFisher), heated 10 min at 95°C and briefly sonicated.

Histone extraction from spermatozoa

Experiments were performed as described in (Crespo *et al*, 2020) with minor modifications. In brief, 5 million of Percoll-purified and sonicated spermheads were incubated in 50 mM DTT for 30 min (at 4°C) then mixed with sulfuric acid (0.4 M final volume), sonicated and acid extracted with 20% trichloroacetic acid (TCA). Histone pellets were washed with cold acetone containing 0.05% HCl, dried at room temperature, and resuspended in 50 µl of SDS-PAGE loading buffer containing 10% β-mercaptoethanol.

Western blot

Electrophoresis was performed in polyacrylamide gels at 120 V in denaturing buffer containing 25 mM Tris Base, 190 mM glycine, 0.1% SDS, and proteins were transferred on nitrocellulose membranes (GE Healthcare). Membranes were then rinsed and incubated 3 min in Ponceau stain to visualize transfer efficiency. Then membranes were incubated in 1× PBS, 0.01% Tween, 5% milk. All primary antibodies were incubated over night at 4°C (see Appendix Table S2 for references and dilutions) and 2 h at room temperature for secondary antibodies. The revelation was performed with Super-Signal West Pico Plus[®] ECL from ThermoFisher (34580) and Immobilon ECL Ultra Western HRP substrate from Millipore (WBULS0100) on ImageQuant[™] LAS 4000 imager.

Extraction and analysis of protamines

PRM1/PRM2 protamine ratio was analyzed following procedures of protamine-rich fraction of sperm nuclear proteins extraction and analysis previously described, with minor modifications (Solvent-Ventura *et al*, 2018). Briefly, histones and other basic proteins loosely attached to DNA were extracted incubating sperm in 0.5 M HCl for 10 min at 37°C after vortexing, and centrifuged at 2,000 g 20 min at 4°C. This step was repeated three times. Resulting pellet was resuspended in 0.5% Triton X-100, 20 mM Tris-HCl (pH 8), and 2 mM MgCl₂. After centrifugation at 8,940 g 5 min at 4°C, the sediment was resuspended in milliQ H₂O with 1 mM PMSF, centrifuging again at the same conditions. Chromatin was then denaturated by resuspending the pellets in 20 mM EDTA, 1 mM PMSF, 100 mM Tris-HCl- pH8 and adding 1 volume of 575 mM DTT in 6 M GuHCl prior vortexing. The solution was incubated at 37°C with 0.8% 4-vinylpyridine to inhibit cysteine disulfide bonds for 30 min, vortexing every 5 min, in fumehood and in dark conditions. Chromatin was then precipitated with a minimum of 10 min incubation with cold ethanol at –20°C, followed by centrifugation

12,880 g for 15 min at 4°C. Basic nuclear proteins (protamine-enriched fraction) were extracted from DNA incubating with 0.5 M HCl at 37°C and recovered in the supernatant after centrifugation at 17,530 g for 10 min at 4°C. Precipitation was carried out with 20% TCA on ice and centrifugation using the same conditions. Protamines were washed twice with 1% β-mercaptoethanol in acetone and dried out at room temperature. For in-gel quantification, dried purified extracts were resuspended 5.5 M urea, 20% β-mercaptoethanol, 5% acetic acid, and separated using acetic acid urea gel electrophoresis (AU-PAGE). Gels were stained with EZBlue[™] Gel Staining Reagent (#G1041, Sigma Aldrich) and optic density of the bands corresponding to mouse PRM1 and PRM2 was quantified using Quantity One 1-D analysis software (BioRad, Hercules, CA, USA) to calculate the PRM1/PRM2 ratio. PRM1/PRM2 ratios were normalized against the mean value of the control group.

For western blot detection, basic nuclear protein extracts from 2.1 M spermatozoa were loaded into AU-PAGE. Proteins were then transferred towards the negative pole onto a 0.45-µm pore size nitrocellulose membrane (88018, ThermoFisher) for 4 h at 4°C using an acidic transfer buffer consisting of 0.9 mM acetic acid. The rest of the procedure (blocking, antibody incubation, and ECL detection) was similar as for other western blot experiment (see dedicated paragraph above).

Histological analyses

Mouse testes were fixed in 4% buffered paraformaldehyde (PFA) for minimum 4 h, before they were cut in halves and incubated overnight in Bouin (Sigma). Testes were then washed in 70% ethanol at room temperature for 30 min, dehydrated, and embedded in paraffin. Four-micrometer sections were stained with periodic acid–Schiff (PAS).

Immunohistochemistry and immunofluorescence

Mouse testes were fixed in 4% buffered paraformaldehyde (PFA) for minimum 4 h, before they were cut in halves and incubated overnight in 4% PFA. Testes were then washed in 70% ethanol at room temperature for 30 min, dehydrated, and embedded in paraffin. Immunohistochemistry or immunofluorescence experiments were performed on 4-µm sections using Novolink polymer detection system (7140-K, Leica Micro-systems) according to the manufacturer's instructions, or following the procedure described in (Comptour *et al*, 2014) with the following modifications. Permeabilization was performed for 15 min in 1× PBS, 0.5% Triton X-100. Blocking was performed for 30 min to 1 h at room temperature in 1× PBS, 0.1% Tween, 1% BSA. Primary antibody was incubated overnight at 4°C (dilution and reference of each antibody are available in Appendix Table S2). Some slides were counterstained with Hematoxylin. For immunofluorescence experiments, lectin (L-21409 or L-32459, ThermoFisher) was diluted at 1/500 in 1× PBS and incubated for 1 h at room temperature along with secondary antibodies (see Comptour *et al*, 2014). Lectin was used to stain the developing acrosome and determine the stage of testis tubules as described in (Ahmed & de Rooij, 2009). DAPI (in VECTASHIELD Mounting Medium, Vectorlab) was used to stain nuclei. TUNEL assay was performed on 4-µm paraffin-embedded testicular sections using In Situ Cell Death Detection Kit, Fluorescein as described by the manufacturer (Roche,

Sigma-Aldrich). GFP immunofluorescence pictures were taken with a NikonE600 microscope using the Nikon Digital Sight SD-V3 camera (DS-Fi1) with the NIS software. All other immunofluorescence pictures were taken with an Olympus BX63 microscope and analyzed using ImageJ 1.48v (<http://imagej.nih.gov/ij/>). Immunohistochemistry pictures were taken with Perkin Elmer Lamina slide scanner and analyzed using CaseViewer software.

Papanicolaou staining of spermatozoa

Spermatozoa were collected from cauda epididymides in M2 medium (Sigma-Aldrich) and spread onto a Superfrost Plus slide (ThermoFisher). Cells were fixed in 4% PFA for 10 min and stained using Papanicolaou staining. Briefly, the slides were washed in 95% ethanol, incubated in Harris hematoxylin for 3 min for nucleus counterstaining, washed and stained with OG-6 dye (RAL diagnostics 05.12013, Martillac, France) and with EA-50 (RAL diagnostics 05.12019, Martillac, France). The slides were then dehydrated and mounted with permanent mounting medium. Spermatozoa were then observed with NikonE600 optic microscope and pictures were taken with the 40X objective.

Electron microscopy

Ten to 20 million epididymal spermatozoa were collected in 2 ml of fresh M2 medium (Sigma-Aldrich) prewarmed at 37°C. After pelleting the sperm at room temperature (300 g, 10 min), the pellet was fixed in 3% glutaraldehyde for 1 h before washing twice in 1× PBS (1,000 g, 10 min). Samples were then dehydrated in ethanol baths and incubated in propylene oxide before inclusion in gelatin. Slices of 60 nm were cut with a Diatome and observed with a JEOL 1011 microscope at a magnification of 1,500× or 2,000×.

Computer assisted sperm motility analysis (CASA)

Immediately after collection of epididymal spermatozoa in M2 medium (Sigma-Aldrich), sperm motility was analyzed by computer-assisted sperm analysis (CASA) systems CASA CEROS II (Hamilton Thorne, Beverly, MA, USA), using a Zeiss microscope.

Sample preparation for RNA-Seq

Total RNA from 150,000 to 600,000 cells of sorted primary spermatocytes (SCI), secondary spermatocytes (SCII) and round spermatids (RS) was extracted using the Ambion RNAqueous micro kit (ThermoFisher) following manufacturers' instructions. For each cell type, five replicates were analyzed. Quantity and quality were assessed using Bioanalyzer chips (ThermoFisher). Libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs) according to supplier recommendations. Paired-end sequencing of 100-bp reads was then carried out on the Illumina HiSeq4000 system.

RNA-Seq analyses

Snakemake was used for RNA-seq analyses (Koster & Rahmann, 2018; v. 3.9.0). Adaptors were trimmed and reads with poor quality (quality < 20) were filtered out with BBduk from

BBTools (v. 38.23, <https://sourceforge.net/projects/bbmap/>). Alignment was performed on the mouse genome build mm10 (GRCm38.p6) using STAR (Dobin & Gingeras, 2015; v. 2.7.2d) and Gencode vM19 gene annotation GTF with following arguments: --outMultimapperOrder Random --quantMode GeneCounts --sjdbOverhang 74. For each sample, the number of aligned reads was converted in cpm (count per million). Genes with an expression level of at least > 1 cpm in a minimum of two samples were included in the analysis, to exclude genes with a very low expression. Differential expression analysis was carried out using DESeq2 (Love et al, 2014) and edgeR (Robinson et al, 2010) packages using default parameters and a design using genotype and cell type. Differentially expressed genes (FDR < 5%) were obtained using glmTreat edgeR's function which conducts a modified likelihood ratio test (LRT) against the fold change threshold (> 1.5). This method is based on TREAT method (McCarthy & Smyth, 2009). Two analyses were performed in parallel: one selecting the deregulated genes with a FDR < 5% (adjusted P-values), and one selecting deregulated genes with a P-value < 5% (non-adjusted P-values). On this second category of genes, the FDR was calculated *a posteriori* using Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) and was confirmed to be < 5%. Enrichment analysis was performed with GSEA (v. 4.0.3, c5.all.v7.1.symbols.gmt, GO cellular components, biological pathways, or all) on gene expression cpm values between KO and control samples (Subramanian et al, 2005; Appendix Fig S5E). Beforehand, mouse Ensembl gene ID were converted to human Ensembl gene ID (hg38) with the biomaRt R package (Durinck et al, 2005). The search for DNA motif in the promoter regions of deregulated genes was performed using HOMER (Heinz et al, 2010) with the findMotifs.pl function and the following parameters: -start -2000 -end 0.

To generate Appendix Fig S1B, we used <https://rgv.genouest.org/> (Darde et al, 2015, 2019).

ChIP-Seq analyses

Snakemake was used for ChIP-Seq analyses. Adaptor were trimmed and filtered using the same parameters than RNA-Seq analyses. Alignment was performed on the mouse genome (mm10) using bowtie2 (Langmead & Salzberg, 2012) with the following arguments: --local. Peak calling was process with MACS2 (Zhang et al, 2008) with default parameters against input file to obtains peaks enrichment score of H3K79me2. Mitochondrial peaks and peaks with FDR > 5% were filtered. Peaks were annotated with the R package ChIPseeker (Yu et al, 2015) using the annotatePeak() function with the following parameters tssRegion = c(-3000,3000).

Code availability

The fully reproducible and documented analysis for RNA-Seq and ChIP-Seq is available on github at https://github.com/ManonCoulee/RNaseq_DOT1L_Blanco_2022.

MS quantification of histone PTMs and of histone variants

To identify histone PTMs, samples were prepared and processed essentially as described in Crespo et al (2020). Briefly, histones were precipitated by acid extraction from whole testes which were

beforehand reduced into powder using a pestle and mortar on dry ice, or from elutriated elongating/condensing spermatids (ES). They were resuspended in loading gel buffer (LDS sample buffer reference 84788, supplemented with NuPAGE sample reducing agent reference NP0009, ThermoFisher) and separated on a 4–12% NuPAGE acrylamide gel (reference NP0321BOX, ThermoFisher). After blue staining (SimplyBlue SafeStain, ThermoFisher), gel bands corresponding to H3 (at about 17 kDa) and H4 (at about 14 kDa) and two gel slices in-between containing H2A and H2B, were cut and then reduced with dithiothreitol, alkylated with iodoacetamide and in-gel digested with 0.1 µg trypsin (V511, Promega) per slice using a Freedom EVO150 robotic platform (Tecan Trading AG, Switzerland). Alternatively, to analyze the relative abundance of histone variants between conditions, histone samples were simply loaded on a gel and migrated for a few millimeter. The whole zone was cut for robotic in-gel tryptic digestion. The resulting tryptic peptides were analyzed by a liquid chromatography–tandem mass spectrometry coupling made up of a C18 reversed-phase capillary column (75 µm i.d. × 25 cm ReproSil-Pur C18-AQ, 1.9 µm particles, Cluzeau, France) using the UltiMate™ 3000 RSLCnano system (ThermoFisher) coupled to a Q-Exactive HF mass spectrometer (ThermoFisher). The mobile phases consisted of water with 0.1% formic acid (A) and acetonitrile with 0.08% (v/v) formic acid (B). Peptides were eluted with a gradient consisting of an increase of solvent B from 2.8% to 7.5% for 7.5 min, then from 7.5% to 33.2% over 33.5 min and finally from 33.2% to 48% over 6.5 min. Mass spectrometry acquisitions were carried out by alternating one full MS scan with Orbitrap detection acquired over the mass range 300 to 1,300 *m/z* and data-dependent MS/MS spectra on the 10 most abundant precursor ions detected in MS. The peptides were isolated for fragmentation by higher energy collisional dissociation (HCD) with a collision energy of 27.

Identification of modified peptides was obtained using our in-house established database of histone sequences, MS_histoneDB (El Kennani et al, 2017), completed with a list of 500 common contaminants. MS/MS data interpretation was carried out with the program Mascot (<http://www.matrixscience.com/>) with the following search parameters. The precursor and fragment mass tolerances were 5 ppm and 25 mmu, respectively; enzyme specificity was trypsin; the maximum number of trypsin missed cleavages was set to 5, carbamidomethyl (Cys) was specified as a fixed modification. We indicated as variable PTMs in Mascot acetylation, crotonylation, and ubiquitination (more precisely the dipeptide “GlyGly”) of Lys residues, methylation and di-methylation of Lys/Arg and trimethylation of Lys, as well as N-terminal acetylation of proteins. Filtering of peptide identifications and quantification was performed by using the program Proline (Bouyssie et al, 2020). All peptide/spectrum matches of scores below 15 were filtered out; next, all identifications of modified peptides suggested by Mascot for H3 and H4 were visually validated by demanding that a continuous stretch of minimally 5 amino acids be identified in terms of b or y fragment ions, and by ascertaining PTM positioning on Lys/Arg residues. This score threshold is very low, yet it was relevant to use so as to be able to identify very short peptides, such as H3-K18modQLATK. When interpreting our mass spectrometry data against the database MS_histoneDB, we were able to identify and quantify various H3 variants, namely canonical H3 and variants H3.3, testis-specific H3t, and H3mm13. H3t differs from canonical H3 at residue 24, while

H3.3 and H3mm13 differ from H3 at residues 31 and 29. To estimate the relative abundance of modified histone peptides between samples, normalization to be at constant amount of total H2A, H2B, H3, or H4 was done by dividing the raw MS signals of modified peptides by the sum of the MS signals of modified and unmodified peptides for each histone. To estimate the relative abundance of histone variants between *Dot1l*-KO and CTL samples, variants were quantified by Proline by relying only on peptides specific of these sequences. Samples were normalized to be at constant amount of histone H4. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al, 2019) partner repository with the dataset identifier PXD030734.

Statistical analysis

Chi-square and unpaired *t*-test were used to analyze IVF data. Sperm motility parameters were analyzed using a Mann–Whitney tests following angular transformation of percentages. To compare the incidence of sperm head abnormalities, and apoptotic (TUNEL+) spermatids between *Dot1l*-KO and CTL, percentages were converted in angles prior to performing *t*-tests. Student *t*-test (corrected for multiple testing when appropriate) was used for all other analyses (fertility, sperm count, testis weight, histology, histone PTM, western blot, and normalized PRM1/2 ratio quantifications). Statistical analyses for RNA-Seq/ChIP-Seq are described in the dedicated paragraphs.

Data availability

RNA-Seq data have been submitted to ENA repository under project number PRJEB50887 (<https://www.ebi.ac.uk/ena/>). H3K79me2 ChIP-Seq project number is PRJNA643726. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al, 2019) partner repository with the dataset identifier PXD030734 (<https://www.ebi.ac.uk/pride/>).

Expanded View for this article is available online.

Acknowledgements

We would like to thank Cochin Institute (INSERM U1016, CNRS UMR8104, Université Paris Cité) core facilities, in particular, Alain Schmitt from the electron microscopy platform, and all the staff from the animal house, histology (HistIM), genomic (GENOM'IC), cytometry (CYBIO), and imaging (IMAC'IC) core facilities. We would also like to thank Guillaume Meurice for advices on bioinformatics analyses, and Aminata Touré and Marjorie Whitfield for advices and discussion on sperm morphology and motility analyses. M.Cr. and D.P. are grateful to their colleagues in EDyP for their support on LC–MS instruments and in informatics. This work was supported by the Agence Nationale de la Recherche (ANR-17-CE12-0004-01 to J.C., ANR-21-CE44-0035 to D.P.), the Fondation pour la Recherche Médicale (SPF20190909274 to C.G.) and grants from Institut Cochin (PIC-2018 to L.E.K.), “Ministerio de Economía e competitividad” F117/00224 to A.I., and “Ministerio de Ciencia e Innovación” PI20/00936 to R.O. and MV20/00026 to A.I. M.B. and M.Co. received a PhD funding from Université Paris Cité, M.Cr., from University Grenoble Alps (UGA). The proteomic experiments were partially supported by Agence Nationale de

la Recherche under projects ProFI (Proteomics French Infrastructure, ANR-10-INBS-08) and GRAL, a program from the Chemistry Biology Health (CBH) Graduate School of University Grenoble Alpes (ANR-17-EURE-0003). This publication is also based upon work from COST Action CA20119 (ANDRONET) supported by European Cooperation in Science and Technology (www.cost.eu).

Author contributions

Melina Blanco: Formal analysis; methodology; writing – original draft.
Laila El Khattabi: Conceptualization; formal analysis; writing – review and editing.
Clara Gobé: Formal analysis; methodology.
Marion Crespo: Formal analysis; visualization.
Manon Coulée: Formal analysis; visualization.
Alberto de la Iglesia: Formal analysis; methodology.
Côme Ialy-Radio: Formal analysis.
Marion Delessard: Formal analysis.
Ivan Seller-Corona: Formal analysis.
Kosuke Yamaguchi: Resources; methodology.
Nadège Vernet: Formal analysis.
Fred Van Leeuwen: Resources.
Alban Lermine: Resources.
Yuki Okada: Resources; methodology.
Romain Daveau: Resources; formal analysis.
Rafael Oliva: Resources.
Pierre Fouchet: Conceptualization; formal analysis.
Ahmed Ziyat: Conceptualization; formal analysis.
Delphine Pflieger: Conceptualization; formal analysis; funding acquisition; methodology; writing – review and editing.
Julie Cocquet: Conceptualization; formal analysis; supervision; funding acquisition; investigation; writing – original draft; project administration; writing – review and editing.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

Aguilar D, Strom J, Chen QM (2014) Glucocorticoid induced leucine zipper inhibits apoptosis of cardiomyocytes by doxorubicin. *Toxicol Appl Pharmacol* 276: 55–62

Ahmed EA, de Rooij DG (2009) Staging of mouse seminiferous tubule cross-sections. *Methods Mol Biol* 558: 263–277

Arevalo L, Merges GE, Schneider S, Oben FE, Neumann IS, Schorle H (2022) Loss of the cleaved-protamine 2 domain leads to incomplete histone-to-protamine exchange and infertility in mice. *PLoS Genet* 18: e1010272

Aslam MA, Alemdehy MF, Kwesi-Maliepaard EM, Muhaimin FI, Caganova M, Pardiack IN, van den Brand T, van Welsem T, de Rink I, Song JY et al (2021) Histone methyltransferase DOT1L controls state-specific identity during B cell differentiation. *EMBO Rep* 22: e51184

Bao J, Bedford MT (2016) Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction* 151: R55–R70

Bao J, Ma HY, Schuster A, Lin YM, Yan W (2013) Incomplete cre-mediated excision leads to phenotypic differences between Stra8-iCre; Mov10l1(lox/lox) and Stra8-iCre; Mov10l1(lox/Delta) mice. *Genesis* 51: 481–490

Barral S, Morozumi Y, Tanaka H, Montellier E, Govin J, de Dieuleveult M, Charbonnier C, Coute Y, Puthier D, Buchou T et al (2017) Histone variant H2A.L.2 guides transition protein-dependent protamine assembly in male germ cells. *Mol Cell* 66: 89–101.e8

Barraud-Lange V, Ialy-Radio C, Chalas C, Holtzmann I, Wolf JP, Barbaux S, Ziyat A (2020) Partial sperm beta1 integrin subunit deletion proves its involvement in mouse gamete adhesion/fusion. *Int J Mol Sci* 21: 8494

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* 57: 289–300

Boussouar F, Goudarzi A, Buchou T, Shiota H, Barral S, Debernardi A, Guardiola P, Brindle P, Martinez G, Arnoult C et al (2014) A specific CBP/p300-dependent gene expression programme drives the metabolic remodelling in late stages of spermatogenesis. *Andrology* 2: 351–359

Bouyssie D, Hesse AM, Mouton-Barbosa E, Rompais M, Macron C, Carapito C, Gonzalez de Peredo A, Coute Y, Dupierriis V, Burel A et al (2020) Proline: an efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* 36: 3148–3155

Braun RE (2001) Packaging paternal chromosomes with protamine. *Nat Genet* 28: 10–12

Cattaneo P, Hayes MGB, Baumgarten N, Hecker D, Peruzzo S, Aslan GS, Kunderfranco P, Larcher V, Zhang L, Contu R et al (2022) DOT1L regulates chamber-specific transcriptional networks during cardiogenesis and mediates postnatal cell cycle withdrawal. *Nat Commun* 13: 7444

Chang CC, Ye BH, Chaganti RS, Dalla-Favera R (1996) BCL-6, a POZ/zinc-finger protein, is a sequence-specific transcriptional repressor. *Proc Natl Acad Sci USA* 93: 6947–6952

Chen Y, Zheng Y, Gao Y, Lin Z, Yang S, Wang T, Wang Q, Xie N, Hua R, Liu M et al (2018) Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res* 28: 879–896

Cocquet J, Ellis PJ, Yamauchi Y, Mahadevaiah SK, Affara NA, Ward MA, Burgoyne PS (2009) The multicopy gene sly represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol* 7: e1000244

Comptour A, Moretti C, Serrentino ME, Auer J, Ialy-Radio C, Ward MA, Toure A, Vaiman D, Cocquet J (2014) SSTY proteins co-localize with the post-meiotic sex chromatin and interact with regulators of its expression. *FEBS J* 281: 1571–1584

Corbinea S, Lassalle B, Givélet M, Souissi-Sarahoui I, Firlej V, Romeo PH, Allemand I, Riou L, Fouchet P (2017) Spermatogonial stem cells and progenitors are refractory to reprogramming to pluripotency by the transcription factors Oct3/4, c-Myc, Sox2 and Klf4. *Oncotarget* 8: 10050–10063

Crespo M, Damont A, Blanco M, Lastrucci E, Kennani SE, Ialy-Radio C, Khattabi LE, Terrier S, Louwagie M, Kieffer-Jaquinod S et al (2020) Multi-omic analysis of gametogenesis reveals a novel signature at the promoters and distal enhancers of active genes. *Nucleic Acids Res* 48: 4115–4138

da Cruz I, Rodriguez-Casuriaga R, Santinaque FF, Farias J, Curti G, Capoano CA, Folle GA, Benavente R, Sotelo-Silveira JR, Geisinger A (2016) Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genomics* 17: 294

Darde TA, Sallou O, Becker E, Evrard B, Monjeaud C, Le Bras Y, Jegou B, Collin O, Rolland AD, Chalmel F (2015) The ReproGenomics viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Res* 43: W109–W116

Darde TA, Lecluze E, Lardenois A, Stevant I, Alary N, Tuttmann F, Collin O, Nef S, Jegou B, Rolland AD et al (2019) The ReproGenomics viewer: a multi-omics and cross-species resource compatible with single-cell studies for the reproductive science community. *Bioinformatics* 35: 3133–3139

Dobin A, Gingeras TR (2015) Mapping RNA-seq reads with STAR. *Curr Protoc Bioinformatics* 51: 11.14.1–11.14.19

Dong Y, Isono KI, Ohbo K, Endo TA, Ohara O, Maekawa M, Toyama Y, Ito C, Toshimori K, Helin K et al (2017) EPC1/TIP60-mediated histone acetylation facilitates spermiogenesis in mice. *Mol Cell Biol* 37: e00082-17

Dottermusch-Heidel C, Gartner SM, Tegeder I, Rathke C, Barckmann B, Bartkuhn M, Bhushan S, Steger K, Meinhardt A, Renkawitz-Pohl R (2014a)

H3K79 methylation: a new conserved mark that accompanies H4 hyperacetylation prior to histone-to-protamine transition in drosophila and rat. *Biol Open* 3: 444–452

Dottermusch-Heidel C, Klaus ES, Gonzalez NH, Bhushan S, Meinhardt A, Bergmann M, Renkawitz-Pohl R, Rathke C, Steger K (2014b) H3K79 methylation directly precedes the histone-to-protamine transition in mammalian spermatids and is sensitive to bacterial infections. *Andrology* 2: 655–665

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21: 3439–3440

El Kennani S, Adrait A, Shaytan AK, Khochbin S, Bruley C, Panchenko AR, Landsman D, Pflieger D, Govin J (2017) MS_HistoneDB, a manually curated resource for proteomic analysis of human and mouse histones. *Epigenetics Chromatin* 10: 2

Ernst C, Eling N, Martinez-Jimenez CP, Marioni JC, Odom DT (2019) Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. *Nat Commun* 10: 1251

Evertts AG, Manning AL, Wang X, Dyson NJ, Garcia BA, Coller HA (2013) H4K20 methylation regulates quiescence and chromatin compaction. *Mol Biol Cell* 24: 3025–3037

Frehlick LJ, Eirin-Lopez JM, Ausio J (2007) New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Bioessays* 29: 49–59

Fujihara Y, Murakami M, Inoue N, Satouh Y, Kaseda K, Ikawa M, Okabe M (2010) Sperm equatorial segment protein 1, SPESP1, is required for fully fertile sperm in mouse. *J Cell Sci* 123: 1531–1536

Gadadhar S, Alvarez Viar G, Hansen JN, Gong A, Kostarev A, Ialy-Radio C, Leboucher S, Whitfield M, Ziyat A, Toure A et al (2021) Tubulin glycylation controls axonemal dynein activity, flagellar beat, and male fertility. *Science* 371: eabd4914

Gaucher J, Boussouar F, Montellier E, Curtet S, Buchou T, Bertrand S, Hery P, Jounier S, Depaux A, Vitte AL et al (2012) Bromodomain-dependent stage-specific male genome programming by Brdt. *EMBO J* 31: 3809–3820

Gilan O, Lam EY, Becher I, Lugo D, Cannizzaro E, Joberty G, Ward A, Wiese M, Fong CY, Ftouni S et al (2016) Functional interdependence of BRD4 and DOT1L in MLL leukemia. *Nat Struct Mol Biol* 23: 673–681

Goudarzi A, Zhang D, Huang H, Barral S, Kwon OK, Qi S, Tang Z, Buchou T, Vitte AL, He T et al (2016) Dynamic competing histone H4 K5K8 acetylation and Butyrylation are hallmarks of highly active gene promoters. *Mol Cell* 62: 169–180

Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Guczyński SJ, Moore BB et al (2018) A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. *Dev Cell* 46: 651–667.e10

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589

Ho CH, Takizawa Y, Kobayashi W, Arimura Y, Kimura H, Kurumizaka H (2021) Structural basis of nucleosomal histone H4 lysine 20 methylation by SET8 methyltransferase. *Life Sci Alliance* 4: e20200919

Jo SY, Granowicz EM, Maillard I, Thomas D, Hess JL (2011) Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation. *Blood* 117: 4759–4768

Jones B, Su H, Bhat A, Lei H, Bajko J, Hevi S, Baltus GA, Kadam S, Zhai H, Valdez R et al (2008) The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS Genet* 4: e1000190

Kawano N, Kang W, Yamashita M, Koga Y, Yamazaki T, Hata T, Miyado K, Baba T (2010) Mice lacking two sperm serine proteases, ACR and PRSS21, are subfertile, but the mutant sperm are infertile in vitro. *Biol Reprod* 83: 359–369

Kim W, Choi M, Kim JE (2014) The histone methyltransferase Dot1/DOT1L as a critical regulator of the cell cycle. *Cell Cycle* 13: 726–738

Koster J, Rahmann S (2018) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 34: 3600

Kuang W, Zhang J, Lan Z, Deepak R, Liu C, Ma Z, Cheng L, Zhao X, Meng X, Wang W et al (2021) SLC22A14 is a mitochondrial riboflavin transporter required for sperm oxidative phosphorylation and male fertility. *Cell Rep* 35: 109025

Kurosu T, Fukuda T, Miki T, Miura O (2003) BCL6 overexpression prevents increase in reactive oxygen species and inhibits apoptosis induced by chemotherapeutic reagents in B-cell lymphoma cells. *Oncogene* 22: 4459–4468

Kwesi-Malilepaard EM, Aslam MA, Alemdehy MF, van den Brand T, McLean C, Vlaming H, van Welsem T, Korthout T, Lancini C, Hendriks S et al (2020) The histone methyltransferase DOT1L prevents antigen-independent differentiation and safeguards epigenetic identity of CD8(+) T cells. *Proc Natl Acad Sci USA* 117: 20706–20716

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* 9: 357–359

Lin YH, Kakadia PM, Chen Y, Li YQ, Deshpande AJ, Buske C, Zhang KL, Zhang Y, Xu GL, Bohlander SK (2009) Global reduction of the epigenetic H3K79 methylation mark and increased chromosomal instability in CALM-AF10-positive leukemias. *Blood* 114: 651–658

Lin H, Cheng K, Kubota H, Lan Y, Riedel SS, Kakiuchi K, Sasaki K, Bernt KM, Bartolomei MS, Luo M et al (2022) Histone methyltransferase DOT1L is essential for self-renewal of germline stem cells. *Genes Dev* 36: 752–763

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

Luense LJ, Wang X, Schon SB, Weller AH, Lin Shiao E, Bryant JM, Bartolomei MS, Coutifaris C, Garcia BA, Berger SL (2016) Comprehensive analysis of histone post-translational modifications in mouse and human male germ cells. *Epigenetics Chromatin* 9: 24

Luense LJ, Donahue G, Lin-Shiao E, Rangel R, Weller AH, Bartolomei MS, Berger SL (2019) Gcn5-mediated histone acetylation governs nucleosome dynamics in spermiogenesis. *Dev Cell* 51: 745–758.e6

Maruyama SY, Ito M, Ikami Y, Okitsu Y, Ito C, Toshimori K, Fujii W, Yogo K (2016) A critical role of solute carrier 22a14 in sperm motility and male fertility in mice. *Sci Rep* 6: 36468

de Mateo S, Gazquez C, Guimera M, Balasch J, Meistrich ML, Balleas JL, Oliva R (2009) Protamine 2 precursors (pre-P2), protamine 1 to protamine 2 ratio (P1/P2), and assisted reproduction outcome. *Fertil Steril* 91: 715–722

McCarthy DJ, Smyth GK (2009) Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25: 765–771

Montellier E, Rousseaux S, Zhao Y, Khochbin S (2011) Histone crotonylation specifically marks the haploid male germ cell gene expression program: post-meiotic male-specific gene expression. *Bioessays* 34: 187–193

Montellier E, Boussouar F, Rousseaux S, Zhang K, Buchou T, Fenaille F, Shiota H, Debernardi A, Hery P, Curtet S et al (2013) Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev* 27: 1680–1692

Moretti C, Serrentino ME, Ialy-Radio C, Delessard M, Soboleva TA, Tores F, Leduc M, Nitschke P, Drevet JR, Tremethick DJ et al (2017) SLY regulates genes involved in chromatin remodeling and interacts with TBL1XR1 during sperm differentiation. *Cell Death Differ* 24: 1029–1044

Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L (2007) A global double-fluorescent Cre reporter mouse. *Genesis* 45: 593–605

Nguyen AT, Taranova O, He J, Zhang Y (2011) DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis. *Blood* 117: 6912–6922

Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, Su L, Xu G, Zhang Y (2005) hDOT1L links histone methylation to leukemogenesis. *Cell* 121: 167–178

Oliva R (2006) Protamines and male infertility. *Hum Reprod Update* 12: 417–435

Oliva R, Bazett-Jones D, Mezquita C, Dixon GH (1987) Factors affecting nucleosome disassembly by protamines in vitro. Histone hyperacetylation and chromatin structure, time dependence, and the size of the sperm nuclear proteins. *J Biol Chem* 262: 17016–17025

Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcari BO, Unternaehrer J, Gupta PB et al (2012) Chromatin-modifying enzymes as modulators of reprogramming. *Nature* 483: 598–602

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M et al (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442–D450

Ragazzini R, Perez-Palacios R, Baymaz IH, Diop S, Ancelin K, Zielinski D, Michaud A, Givélet M, Borsos M, Aflaki S et al (2019) EZHIP constrains Polycomb repressive complex 2 activity in germ cells. *Nat Commun* 10: 3858

Rathke C, Baarends WM, Awe S, Renkawitz-Pohl R (2014) Chromatin dynamics during spermiogenesis. *Biochim Biophys Acta* 1839: 155–168

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140

Sadate-Ngatchou PI, Payne CJ, Dearth AT, Braun RE (2008) Cre recombinase activity specific to postnatal, premeiotic male germ cells in transgenic mice. *Genesis* 46: 738–742

Shang E, Nickerson HD, Wen D, Wang X, Wolgemuth DJ (2007) The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* 134: 3507–3515

Shiota H, Barral S, Buchou T, Tan M, Couté Y, Charbonnier G, Reynoird N, Boussouar F, Gerard M, Zhu M et al (2018) Nut directs p300-dependent, genome-wide H4 hyperacetylation in male germ cells. *Cell Rep* 24: 3477–3487.e6

Shoaib M, Chen Q, Shi X, Nair N, Prasanna C, Yang R, Walter D, Frederiksen KS, Einarsson S, Svensson JP et al (2021) Histone H4 lysine 20 monomethylation directly facilitates chromatin openness and promotes transcription of housekeeping genes. *Nat Commun* 12: 4800

Soler-Ventura A, Castillo J, de la Iglesia A, Jodar M, Barrachina F, Balleca JL, Oliva R (2018) Mammalian sperm protamine extraction and analysis: a step-by-step detailed protocol and brief review of protamine alterations. *Protein Pept Lett* 25: 424–433

Soumillon M, Necseula A, Weier M, Brawand D, Zhang X, Gu H, Barthes P, Kokkinaki M, Nef S, Gnirke A et al (2013) Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep* 3: 2179–2190

Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim JE, Chen J, Lazar MA et al (2008) DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol* 28: 2825–2839

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550

Thomis DC, Lee W, Berg LJ (1997) T cells from Jak3-deficient mice have intact TCR signaling, but increased apoptosis. *J Immunol* 159: 4708–4719

Torregrosa N, Dominguez-Fandos D, Camejo MI, Shirley CR, Meistrich ML, Balleca JL, Oliva R (2006) Protamine 2 precursors, protamine 1/protamine 2 ratio, DNA integrity and other sperm parameters in infertile patients. *Hum Reprod* 21: 2084–2089

Toure A (2019) Importance of SLC26 transmembrane anion exchangers in sperm post-testicular maturation and fertilization potential. *Front Cell Dev Biol* 7: 230

Valencia-Sanchez MI, De Ioannes P, Wang M, Truong DM, Lee R, Armache JP, Boeke JD, Armache KJ (2021) Regulation of the Dot1 histone H3K79 methyltransferase by histone H4K16 acetylation. *Science* 371: eabc6663

Vlaming H, van Leeuwen F (2016) The upstreams and downstreams of H3K79 methylation by DOT1L. *Chromosoma* 125: 593–605

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ et al (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40: 897–903

Wang Y, Iwamori T, Kaneko T, Iida H, Iwamori N (2021) Comparative distributions of RSBN1 and methylated histone H4 lysine 20 in the mouse spermatogenesis. *PLoS one* 16: e0253897

Ward WS, Coffey DS (1991) DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol Reprod* 44: 569–574

Yamaguchi K, Hada M, Fukuda Y, Inoue E, Makino Y, Katou Y, Shirahige K, Okada Y (2018) Re-evaluating the localization of sperm-retained histones revealed the modification-dependent accumulation in specific genome regions. *Cell Rep* 23: 3920–3932

Yamauchi Y, Riel JM, Stoytcheva Z, Burgoyne PS, Ward MA (2010) Deficiency in mouse Y chromosome long arm gene complement is associated with sperm DNA damage. *Genome Biol* 11: R66

Yu YE, Zhang Y, Unni E, Shirley CR, Deng JM, Russell LD, Weil MM, Behringer RR, Meistrich ML (2000) Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proc Natl Acad Sci USA* 97: 4683–4688

Yu G, Wang LG, He QY (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31: 2382–2383

Zhang W, Hayashizaki Y, Kone BC (2004) Structure and regulation of the mDot1 gene, a mouse histone H3 methyltransferase. *Biochem J* 377: 641–651

Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al (2008) Model-based analysis of ChIP-seq (MACS). *Genome Biol* 9: R137

Zhao M, Shirley CR, Hayashi S, Marcon L, Mohapatra B, Sukanuma R, Behringer RR, Boissonneault G, Yanagimachi R, Meistrich ML (2004a) Transition nuclear proteins are required for normal chromatin condensation and functional sperm development. *Genesis* 38: 200–213

Zhao M, Shirley CR, Mounsey S, Meistrich ML (2004b) Nucleoprotein transitions during spermiogenesis in mice with transition nuclear protein Tnp1 and Tnp2 mutations. *Biol Reprod* 71: 1016–1025

Zhu B, Chen S, Wang H, Yin C, Han C, Peng C, Liu Z, Wan L, Zhang X, Zhang J et al (2018) The protective role of DOT1L in UV-induced melanomagenesis. *Nat Commun* 9: 259



License: This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

4 DOT1L a des effets antagonistes sur la régulation de l'expression génique liés aux chromosomes sexuels et autosomiques durant la spermatogenèse

4.1 Résultats de l'étude

Nos analyses transcriptomiques ont permis de mettre en évidence que le KO conditionnel de *Dot1l* entraîne un nombre plus important de gènes sur-exprimés en comparaison du nombre de gènes sous-exprimés au cours de la spermiogenèse (cf. 3.2 *Article 1*). Cette observation va à l'encontre d'un rôle strictement activateur de la transcription de DOT1L et de la corrélation entre méthylation de H3K79 et expression génique, observés dans les études publiées jusqu'alors. Dans un premier temps, mon objectif a donc été de vérifier la validité de ces observations par une approche de normalisation externe via *spike-in* dans les stades étudiés précédemment et dans deux stades pré-méiotiques (Fig. 30). Ainsi, mes résultats montrent que, normalisé par *spike-in*, le KO de *Dot1l* entraîne la dérégulation de plusieurs centaines de gènes avec un nombre de gènes sur-exprimés plus important à tous les stades étudiés, confirmant que cette sur-expression n'est pas due à un biais de normalisation. L'analyse de la dynamique de la dérégulation par DOT1L sur l'expression des gènes apparaît majoritairement à partir du stade de spermatocyte secondaire et est conservée dans les spermatides rondes.

En parallèle, j'ai caractérisé la dynamique de la marque H3K79me2 dans trois stades de la spermatogenèse (Fig. 30). L'intégration de nos données de ChIP-seq avec les données de transcriptomiques permettent de mettre en évidence que l'augmentation de H3K79me2 corrèle positivement avec l'augmentation de l'expression génique lorsque la marque est située au niveau du corps du gène et au niveau des *enhancers*.

La comparaison de ces résultats avec les données de RNA-seq sur les cellules KO-*Dot1l* montre également que l'activation des gènes par DOT1L est associée à la présence de H3K79me2 au niveau du corps du gène avec ou sans présence de la marque aux *enhancers*. La localisation de cette marque semble être cruciale, puisqu'à l'inverse, la sur-expression observée dans les cellules KO est corrélée avec la présence de H3K79me2 seulement au niveau des *enhancers*. Ces analyses ont permis de mettre en évidence le rôle de répresseur de la transcription en plus du rôle d'activateur de la transcription déjà établi pour DOT1L. Au moment de l'initiation du projet de thèse, cette fonction de DOT1L n'avait pas encore été décrite dans la littérature mais a depuis été observée dans d'autres types cellulaires (cf. *Discussion*). J'ai donc ensuite cherché à identifier le mécanisme par lequel DOT1L agit en tant que répresseur, par l'environnement chromatinien ou par activation d'un répresseur. L'analyse de l'environnement chromatinien des gènes dérégulés permet de mettre en évidence que les gènes sur-exprimés sont associés à des états chromatiniens bivalents et répresseurs, caractérisé par un enrichissement de H3K27me3.

Pour finir, mes travaux ont permis de mettre en évidence que l'impact majeur de DOT1L sur l'activation de la transcription était associé à la présence de la marque H3K79me2 et concernait les gènes portés par les chromosomes sexuels. Ces résultats

Résultats

montrent ainsi que DOT1L semble jouer un rôle lors de la réactivation des gènes après le MSCI dans les spermatocytes secondaires et les spermatides rondes.

DOT1L a des effets antagonistes sur la régulation de l'expression génique liés aux chromosomes sexuels et autosomiques durant la spermatogenèse

4.2 Article 2

DOT1L has opposite effects on sex chromosome-linked and autosomal gene expression during male germ cell differentiation

Prêt à être soumis

Manon Coulée, Clara Gobé, Mélina Blanco, Clémentine Lapoujade, Côme laly-Radio, Guillaume Meurice, Pierre Fouchet, Julie Cocquet & Laila El Khattabi

Pour plus de simplicité, les figures sont incluses dans ce manuscrit

Les figures supplémentaires sont disponibles en annexe de ce manuscrit

1 **DOT1L has opposite effects on sex chromosome-linked and autosomal gene**
2 **expression during male germ cell differentiation**

3

4

5 Manon Coulée¹, Clara Gobé¹, Mélina Blanco^{1,2}, Clémentine Lapoujade², Côme laly-Radio¹,
6 Guillaume Meurice³, Pierre Fouchet², Julie Cocquet^{1*}, Laila El Khattabi^{1,4*}

7

8 ¹ *Institut Cochin, INSERM U1016, CNRS UMR8104, Université Paris Cité, Paris, France*

9 ² *Université Paris Cité et Université Paris-Saclay, iRCM/IBJF CEA, UMR Stabilité Génétique*
10 *Cellules Souches et Radiations, Laboratoire des Cellules Souches Germinales, Fontenay-aux-*
11 *Roses, France*

12 ³ *MOABI, plateforme de Bioinformatique de l'APHP, Paris, France*

13 ⁴ *APHP centre-université de Paris, service de génétique, Paris, France*

14 *These authors contributed equally to this work.

15 Correspondence to julie.cocquet@inserm.fr and laila.el-khattabi@aphp.fr

16

17

18 **Abstract**

19

20 The H3K79 methyl transferase DOT1L is essential for multiple biological processes where it has
21 been shown to regulate gene expression. H3K79 methylation is a mark of active transcription,
22 yet, recent studies have shown that DOT1L can induce gene repression as well as activation, in
23 several somatic cell differentiation systems. Here, we used male germline differentiation, a
24 highly dynamic process in terms of transcription and chromatin remodeling, as a biological
25 model to decipher the role of DOT1L.

26 By integrating epigenomic analyses with spike-in RNA-seq performed on *Dot1l* knocked-out
27 mouse male germ cells, we showed that DOT1L plays a dual role in gene regulation depending
28 on the chromatin environment. At all stages (premeiotic, meiotic and postmeiotic), DOT1L
29 inhibits more than activates gene expression, and H3K79me2 positively correlates with male
30 germ cell genetic program. After meiosis I, where DOT1L expression radically increases, a high
31 number of H3K79me2 regions are established, in particular on the X and Y chromosomes. In
32 agreement with this observation, DOT1L was found to activate the expression of many XY
33 genes enriched in H3K79me2 at their bodies and/or neighboring enhancers. In contrast,
34 repressed genes are located in an H3K27ac-poor/H3K27me3-rich environment, devoid of
35 H3K79me2 or restricted at their enhancers. Gene set enrichment analyses showed that, in
36 early stages, repressed genes are involved in processes required towards the end of
37 spermatogenesis, while, in later stages, repressed genes are involved in apoptosis.
38 Interestingly, we identified several repressors among the genes that are positively regulated
39 by DOT1L, in particular *Bcl6* and *Bcor1*, involved in the control of apoptosis, and the gene
40 coding for histone deacetylase HDAC8 ; all could contribute to the repressive effect of DOT1L.
41 Taken together, our results indicate that DOT1L plays a pivotal role in the transcriptional and
42 chromatin remodeling during spermatogenesis, on one hand, by directly regulating the
43 postmeiotic (re)activation of XY genes, following meiotic sex chromosome inactivation, and,
44 on the other hand, by indirectly controlling the timely expression of differentiation genes.

45

46

47

48

49

50 **Introduction**

51 DOT1L (*Disruptor of telomeric silencing 1-like*) is the sole H3K79 (lysine 79 of histone H3)
 52 methyltransferase (Feng *et al*, 2002) and is highly conserved in eukaryotes. It has been shown
 53 to be involved in multiple biological processes, such as embryonic development, cell cycle
 54 progression, cell reprogramming, differentiation and proliferation as well as in the emergence
 55 of mixed lineage leukemia (Jones *et al*, 2008; Kim *et al*, 2014; Nguyen *et al*, 2011; Onder *et al*,
 56 2012; Vlaming & van Leeuwen, 2016; Wille & Sridharan, 2022).

57 Mechanistically, DOT1L protein interacts with the active form of RNA polymerase II (Kim *et al*,
 58 2012) and with protein complexes that promote transcription elongation, such as EAP (for ENL-
 59 associated proteins) and SEC (for Super Elongation Complex)(Park *et al*, 2010; Smith *et al*,
 60 2011). These interactions are at the basis of DOT1L involvement in cancer development: In
 61 some types of mixed lineage leukemia, the fusion between MLL1 protein and DOT1L partners
 62 (such as ENL, AF9 or AF10) leads to the ectopic activation of MLL1 target genes by DOT1L
 63 (Bernt *et al*, 2011; Okada *et al*, 2005; Okuda *et al*, 2017). More recently, DOT1L has also been
 64 found to interact with the transcription initiation factor TFIID (Wu *et al*, 2021). All forms of
 65 H3K79 methylation (mono, di and trimethylation) are enriched at the promoter and gene body
 66 of transcriptionally active genes leading to the view that DOT1L acts as a transcriptional
 67 activator (Guenther *et al*, 2007; Pursani *et al*, 2018; Steger *et al*, 2008).

68 Yet, recent molecular studies have challenged this model by reporting that DOT1L loss of
 69 function does not lead to extensive gene downregulation, as one could have expected, but
 70 rather leads to the de-repression of hundreds to thousands of genes, depending on the cell
 71 type [for review see (Wille & Sridharan, 2022)].

72 DOT1L expression is ubiquitous, with a variable level across tissues and cell types, but is
 73 particularly high in the testis, and more particularly in meiotic and postmeiotic germ cells
 74 (Blanco *et al*; Dottermusch-Heidel *et al*; Dottermusch-Heidel *et al*; Moretti *et al*). We and
 75 others have recently shown that DOT1L is essential for the differentiation of spermatids into
 76 functional spermatozoa as well as for spermatogonia self-renewal (Blanco *et al*.; Lin *et al*; Malla
 77 *et al*).

78 Spermatogenesis is one of the biological processes during which transcription and chromatin
 79 remodeling are the most dynamic (Green *et al*, 2018; Maezawa *et al*, 2018; Soumillon *et al*,

80 2013). It consists in very specific cell stages (i.e. spermatogonial stem cell stage, spermatogonia
81 proliferation, meiosis and spermatid differentiation, see **Fig. 1a**) which are underlain by
82 distinctive genetic programs (da Cruz *et al*; Ernst *et al*; Green *et al*.; Sin *et al*.; Soumillon *et al*.,
83 2013). Spermatogenesis is also characterized by extensive chromatin reorganization, in
84 particular i) at meiosis, where homologous chromosomes compact, pair and recombine, while
85 heterologous sex chromosomes are temporary silent, and ii) towards the end of
86 spermatogenesis, where most histones are replaced by protamines resulting in a tight
87 compaction of the sperm chromatin (Alavattam & Namekawa, 2020; Bao & Bedford; Blanco &
88 Cocquet; Dottermusch-Heidel *et al*.; Patel *et al*, 2019; Vara *et al*, 2019). Here, we took
89 advantage of these features to in-depth characterize the role of DOT1L in gene expression and
90 the dynamic of H3K79me2 mark, throughout spermatogenesis.

91

92 **Results**

93 **H3K79me2 at gene body and enhancers correlates with active transcription during male** 94 **germ cell differentiation**

95 To investigate H3K79me2 dynamics throughout spermatogenesis, we performed ChIP-Seq
96 experiments using anti-H3K79me2 antibody at four different stages of male germ cell
97 development: before meiosis, in undifferentiated germinal stem cells (GSC), at meiotic stage,
98 in primary spermatocytes (SCI), and after meiosis, in round spermatids (RS) and in
99 elongating/condensed spermatids (ES) (**Fig 1a**). Experiments were performed in duplicates and
100 showed strong correlation for each cell stage (**Fig S1a**). The number of peaks between
101 duplicates was also very similar (**Fig. S1b**), except for ES, probably because these cells are in
102 transition between a nucleosomal structure and a protamine-organized chromatin which
103 increases cell population variability. We therefore decided to exclude ES from most of
104 subsequent analyses.

105 Genomic annotation showed that most H3K79me2 peaks locate at promoters and intragenic
106 regions, while a smaller number was found to map at distal intergenic regions (**Fig 1b**). The
107 distribution was significantly different between cell types. In particular, we observed a
108 remodeling of H3K79me2 genomic landscape, with a progressive increase of the proportion of
109 peaks at distal intergenic regions throughout differentiation [1% in GSC (n = 170), 5% in SCI (n
110 = 2122) and 15% of peaks in RS (n = 8788), **Fig. 1b and S1b**]. At the gene level, we observed
111 the strongest enrichment right after the transcriptional start site which slowly decreased in

Fig. 1

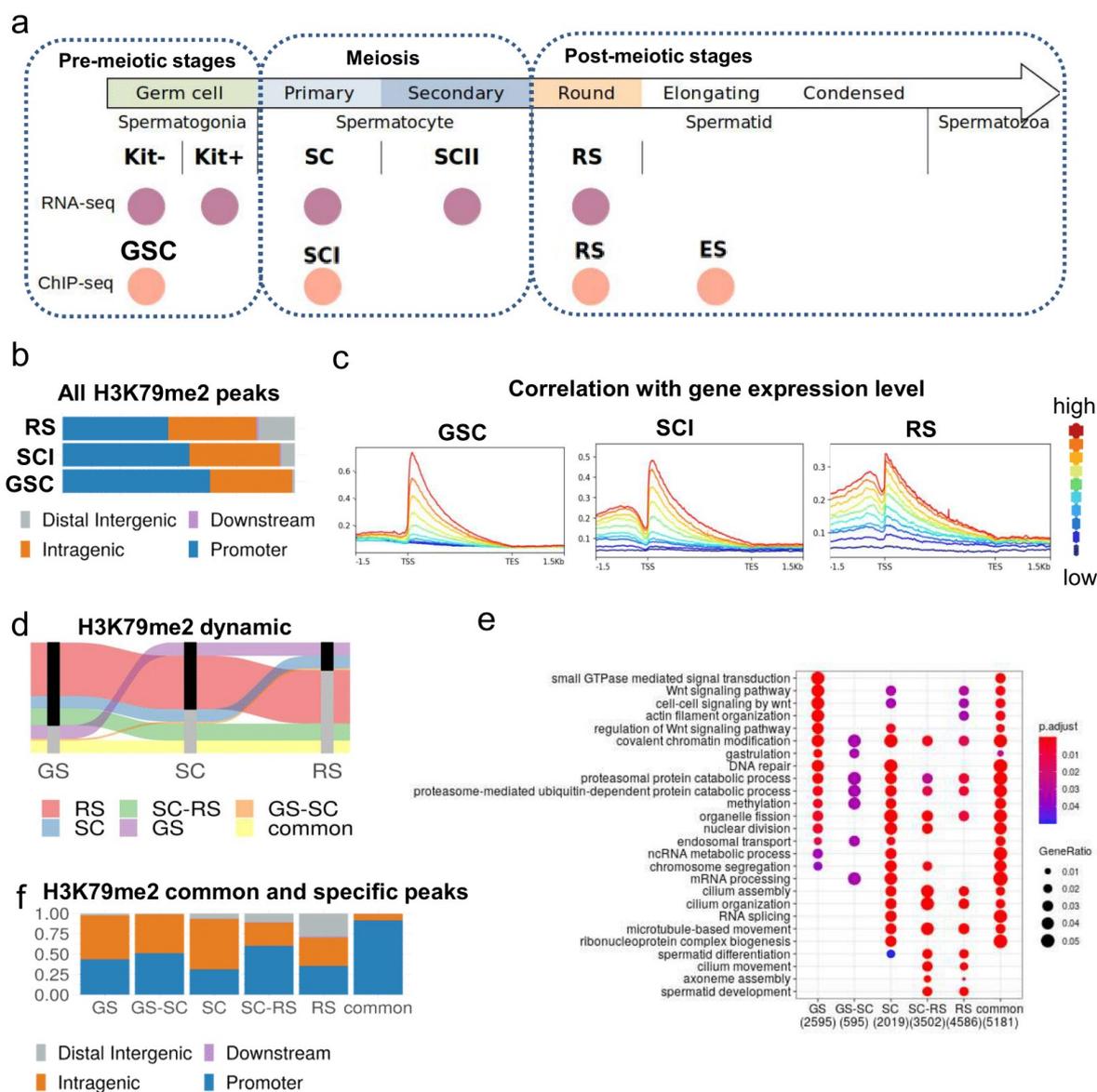


Figure 1. Characterization of H3K79me2 dynamic during male germ cell differentiation.

a. List of samples used in the present study. RNA-seq: Dot1l-KO and CTL in classic and spike-in RNA-seq studies. ChIP-seq: H3K79me2 ChIP-seq data. **b.** Genomic annotation of H3K79me2 ChIP-Seq peaks in germinal stem cells (GSC), primary spermatocytes (SCI), and round spermatids (RS), using CHIPseeker. **c.** H3K79me2 enrichment according to gene expression level in GSC, SCI and RS. (blue: low expressed genes, red: high expressed genes. Gene expression level is divided into percentile groups in each cell type). **d.** Alluvial plot showing the dynamic of H3K79me2 peaks throughout spermatogenesis. Regions enriched in H3K79me2 at the specific cell type are shown in grey, regions devoid of the mark in black. **e.** Ontology of the genes enriched in H3K79me2 at specific stages using clusterProfiler, p-value is adjusted according to Benjamini-Hochberg correction and

DOT1L a des effets antagonistes sur la régulation de l'expression génique liés aux chromosomes sexuels et autosomiques durant la spermatogenèse

GeneRatio corresponds to the ratio of deregulated genes among all genes in the pathway. **f.** Genomic annotation of H3K79me2 ChIP-Seq peaks that are stage-specific (GSC, SCI or RS), or common to two successive stages (GSC-SCI or SCI-RS) or those common to all stages (common).

112 intensity along the gene body, in all cell types (**Fig. 1c and 1c**), consistently with what was
 113 previously described in somatic cells (Vlaming & van Leeuwen, 2016). Comparison to gene
 114 expression levels showed that H3K79me2 enrichment positively correlates with the later; the
 115 more a gene is expressed the more it is enriched in H3K79me2 (**Fig 1c**).

116 A finest analysis of H3K79me2 dynamics between cell types showed a high conservation of
 117 peaks present in GSC throughout spermatogenesis, with ~54% of GSC peaks maintained in SCI
 118 and RS. In contrast, a high number of peaks (%) were *de novo* established at meiosis, in SCI,
 119 and even greater (~64%) after meiosis, in RS (**Fig. 1d**). Interestingly, H3K79me2 regions
 120 specifically acquired at meiotic or postmeiotic stages (“SCI-RS” and “RS-spe” peaks) were
 121 found enriched in genes involved in spermiogenesis such as spermatid development, and
 122 cilium assembly, organization and movement, while peaks conserved throughout
 123 spermatogenesis (“common” peaks) mapped to genes of various pathways, including Wnt
 124 signaling and actin filament organization pathways (**Fig. 1e**).

125 The analysis of H3K79me2 genomic distribution among cell types, revealed a high proportion
 126 of peaks mapping to distal intergenic regions among those that were specifically established
 127 at the postmeiotic stage (RS-spe), while peaks common to all stages were almost exclusively
 128 enriched at gene promoters (**Fig. 1f**).

129 To better characterize H3K79me2 peaks, we used ChromHMM tool (Ernst & Kellis, 2017) and
 130 defined 18 chromatin states characterized by distinct combinations of histone PTMs, which
 131 were then merged into 15 chromatin states (see **Fig 2a**). Four of them correspond to
 132 enhancers: active enhancers (EnhA) characterized by the presence of H3K4me1 and H3K27ac,
 133 weak enhancers (EnhW) characterized by H3K4me1 (with low H3K27ac level), bivalent
 134 enhancers (EnhBiv) enriched in H3K4me1, H3K27me3 and, to a lesser extent, H3K27ac, and
 135 genic enhancers (EnhG) which are enriched in H3K4me1, H3K36me3 and H3K27ac (and to a
 136 lesser extent in H3K27me3). We next compared the chromatin states of regions enriched in
 137 H3K79me2 (“H3K79me2+”) to those devoid of H3K79me2 (“H3K79me2-”) and found a
 138 dynamic H3K79me2 enrichment at enhancer regions throughout male germ cell
 139 differentiation: In GSC, H3K79me2+ regions were found enriched in active enhancers (2.8% of
 140 H3K79me2+ regions vs 0.4% of H3K79me2- ; Chi2, $p=5.8e-5$) and weak enhancers (12.3% vs
 141 6.2%, Chi2, $p = 0.0113$), compared to H3K79me- regions; while in SCI and RS, H3K79me2
 142 regions were highly enriched in genic enhancers [49.9% vs 4% in SCI (Chi2, $p = 3e-120$) and
 143 36% vs 3.7% in RS (Chi2, $p = 1.6e-64$)] (**Fig 2b**).

DOT1L a des effets antagonistes sur la régulation de l'expression génique liés aux chromosomes sexuels et autosomiques durant la spermatogenèse

Fig. 2

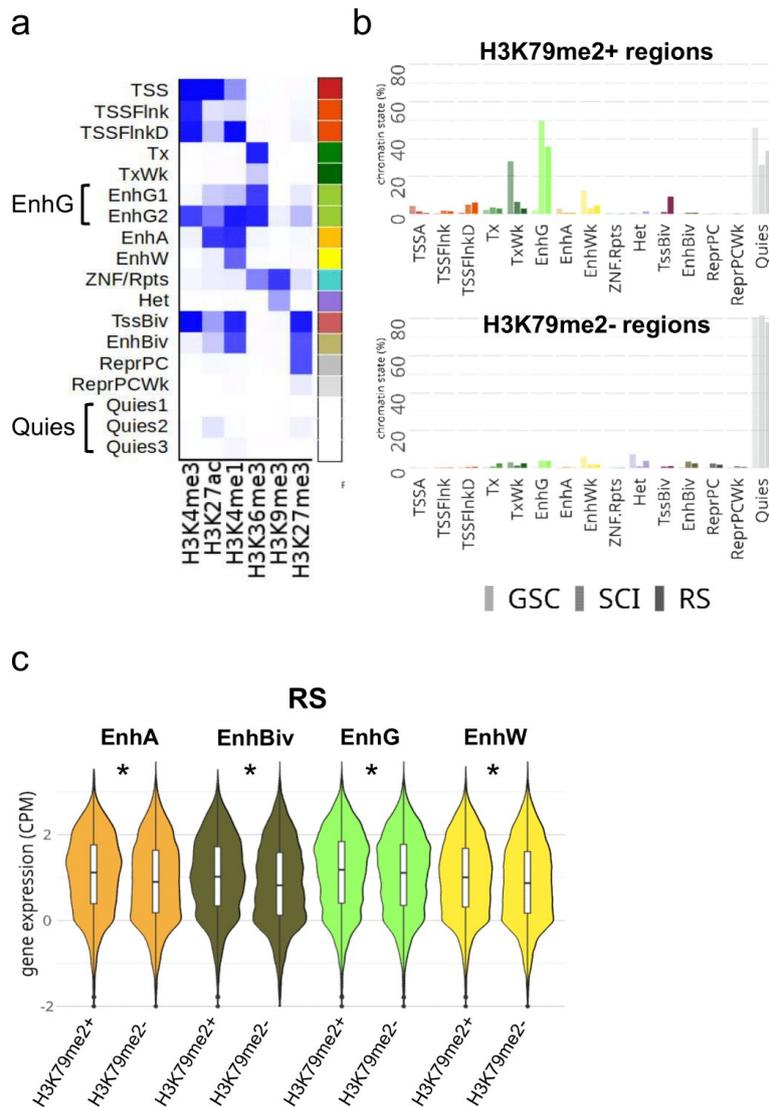


Figure 2. Chromatin environment at H3K79me2 regions in male germ cells.

a. The 18 chromatin state model obtained with ChromHMM. TSS: Transcription Start Site, TSSFlnk: Flanking TSS, TSSFlnkD: Flanking TSS downstream, TX: Strong transcription, TxWk: Weak transcription, EnhG1 and EnhG2: Genic enhancers (which were pooled in one state named EnhG), EnhA: Active enhancer, EnhWk: Weak enhancer, ZNF/Rpts: ZNF genes and repeats, Het: Heterochromatin, TSSBiv: Bivalent/Poised TSS, EnhBiv: Bivalent enhancer, ReprPC: Repressed PolyComb, ReprPCWk: Weak repressed polycomb, Quies1, Quies2 and Quies3: Quiescent/Low (which were pooled in one state called Quies). **b.** Distribution of chromatin states among H3K79me2 enriched regions (H3K79me2+) and H3K79me2 devoid regions (H3K79me2-). **c.** Mean gene expression level (log(cpm)) of genes associated with enhancers enriched (+) or not (-) in H3K79me2 in RS.

144 We also observed a dynamic enrichment of H3K79me2 at chromatin states associated with
 145 transcription: in particular at weak transcription-related chromatin state (TxWk) for GSC and
 146 at bivalent TSS for RS. At SCI and RS stages, H3K79me2+ regions were also over-represented in
 147 chromatin states flanking the TSS, compared to H3K79me2- regions.

148 When focusing on enhancers, we observed that, for all types of enhancers, an enrichment in
 149 H3K79me2 was associated with a significantly higher level of expression of neighboring genes
 150 compared to enhancers devoid of H3K79me2 (**Fig 2c** and **Fig S2**).

151 In sum, our analysis shows a switch of H3K79me2 landscape at the mitosis-to-meiosis
 152 transition, and a particularly high number of peaks acquired postmeiotically, especially at
 153 distal intergenic regions. H3K79me2 peaks mapping at enhancers are associated with a higher
 154 level of expression of neighboring genes, whichever the type of enhancer (genic, non-genic
 155 active, weak or bivalent). H3K79me2 at gene bodies and distal enhancers is therefore a mark
 156 of active transcription.

157

158 **Spike-in RNA-seq analyses reveal that DOT1L has a broad repressive rather than activating**
 159 **effect on gene expression**

160 In order to characterize the role of DOT1L in transcriptional regulation during male germ cell
 161 differentiation, we performed RNAseq analyses from a *Dot1l* male germline-conditional KO
 162 mouse model (described previously in (Blanco *et al.*)). Male germ cells from 5 different stages
 163 were investigated: undifferentiated spermatogonia (Kit-), differentiated spermatogonia (Kit+),
 164 primary spermatocytes (SCI), secondary spermatocytes (SCII) and round spermatids (RS) (**Fig.**
 165 **1a, 3a, b and S3**). Differential gene expression analyses showed that gene deregulation
 166 intensifies as spermatogenesis progresses, with very few deregulated genes in premeiotic
 167 spermatogonial stages and 377 deregulated genes in meiotic and postmeiotic cells ($p < 0.05$;
 168 $FDR < 0.05$; $FC > 1.5$) (**Fig. 3b** and **Fig. S3**). Given the positive correlation between H3K79
 169 methylation and transcription, we were surprised to observe an imbalance towards gene over-
 170 expression in SCII and RS. We therefore decided to investigate the existence of a global effect
 171 of *Dot1l*-KO on the transcriptome, and used spike-in sequences to normalize gene expression
 172 levels to the number of cells from which RNA was extracted (Risso *et al.*, 2014). The results
 173 obtained with both the “classic” and spike-in normalization showed that most deregulated
 174 genes found using the first method were also similarly deregulated using the second
 175 normalization method (94%, 96% and 77% for SCI, SCII and RS, respectively). A higher number

DOT1L a des effets antagonistes sur la régulation de l'expression génique liés aux chromosomes sexuels et autosomiques durant la spermatogenèse

Fig. 3

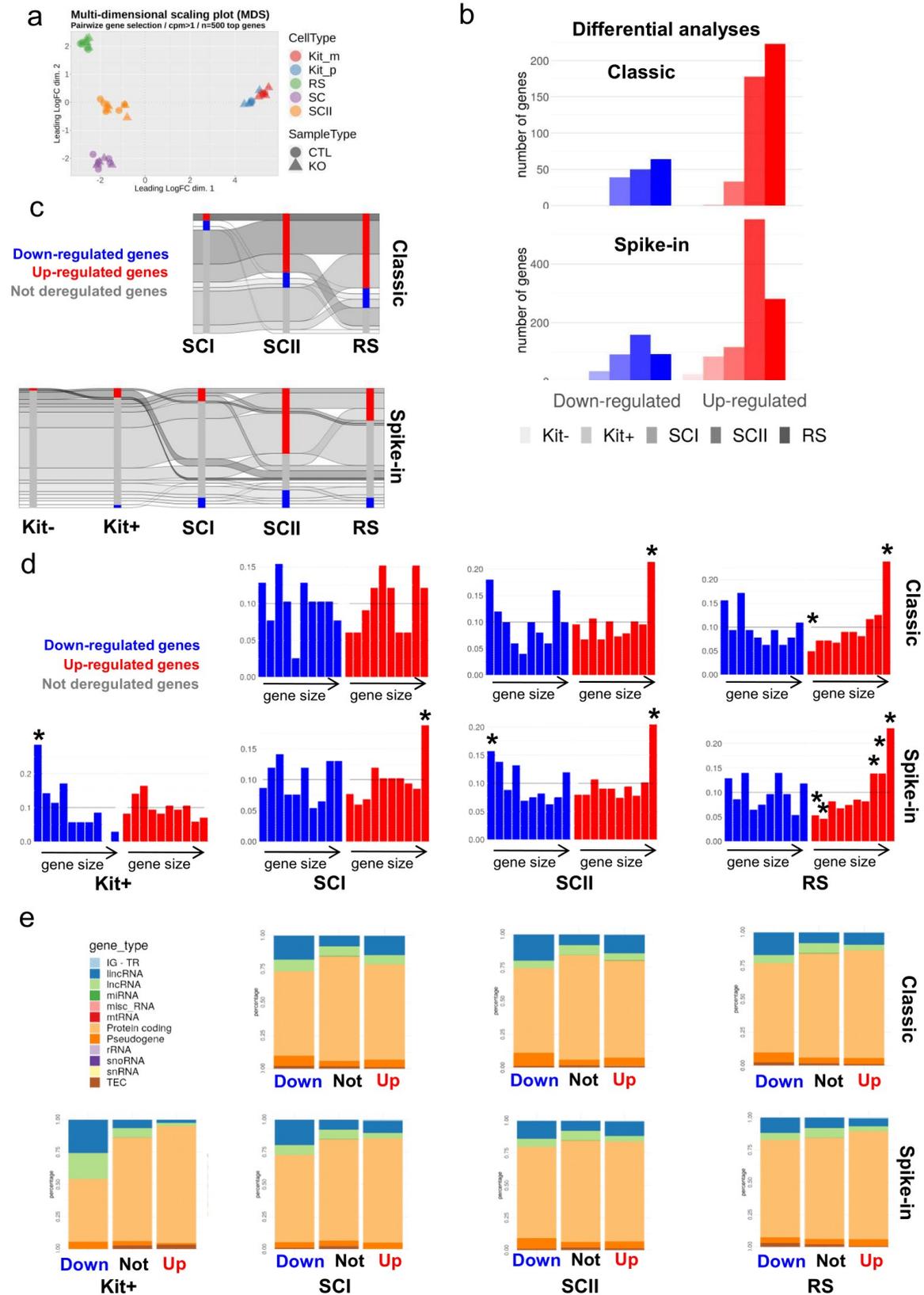


Figure 3. Transcriptional changes throughout spermatogenesis upon Dot1l knockout.

a. Multi-Dimensional Scaling plot showing RNA-Seq data from 2 genotypes (Dot1-KO and CTL cells) and 5 male germ cell types (N= 3 KO and 3 CTL for Kit⁻ and Kit⁺; N= 5 KO and 5 CTL for SCI, SCII and RS). Kit⁻: undifferentiated spermatogonia, Kit⁺: differentiated spermatogonia expressing the c-kit marker, SCI: primary spermatocytes, SCII: secondary spermatocytes and RS: round spermatids. **b.** Number of deregulated genes according to a classic differential expression analysis (top, n= three replicates for Kit⁻, Kit⁺ in each condition, and n=5 replicates for SCI, SCII and RS in each condition) or including a spike-in normalization (bottom, n= 3 replicates for all cell types and conditions) in 5 male germ cell types (p<0.05; FDR<0.05; FC > 1.5). **c.** Dynamics of the genes deregulated in Dot1l-KO vs CTL male germ cells. **d.** Barplot representing the average length of deregulated genes (up, down) and not deregulated genes. Gene size is divided into percentile groups in each cell type. A star indicates a pvalue<0.05 found by comparison with not deregulated genes (using a Chi2 test). **e.** Functional annotation of deregulated genes (up, down) and not deregulated genes. IG-TR: Immunoglobulin/T- receptor ; lincRNA: long intergenic non-coding RNA; lnc: long non-coding RNA; miRNA: microRNA ; misc_RNA: miscellaneous RNA (that are not in any other classification); mtRNA: mitochondrial RNA; Protein coding ; Pseudogene; rRNA: ribosomal RNA; snoRNA: small nucleolar RNA ; snRNA: small nuclear RNA; TEC: transcripts to be experimentally confirmed.

176 of deregulated genes was nevertheless found using the spike-in normalization, still with a bias
177 towards gene upregulation, but this time at all stages of male germ cell development (**Fig. 3b**).
178 We next compared the dynamic of deregulated genes, and observed a different program
179 before and after meiosis. On one hand, most deregulated genes in Kit⁺ were not deregulated
180 in meiotic and postmeiotic cells, and reciprocally. On the other hand, many deregulated genes
181 were common between meiotic and postmeiotic stages (**Fig. 3c**).

182 Further characterization of the length of deregulated genes revealed that upregulated genes
183 in SCI, SCII and RS are often large (Chi2, $p < 0.05$ between upregulated and all genes, for the
184 90th percentile of size in SCII and RS for “classic” analyses, and in SCI, SCII and RS for “spike-in”
185 analyses), while downregulated genes in Kit⁺ cells are small (Chi2, $p = 0.00025$ between
186 downregulated and all genes, for the 10th percentile of size in Kit⁺ for “spike-in” analyses) (**Fig.**
187 **3d**). The functional annotation of deregulated genes was overall similar to that of other genes,
188 with a majority of “protein coding” genes in all categories and cell types. There was
189 nevertheless a significant increase in the proportion of lincRNA and lncRNA among the
190 downregulated genes, particularly visible in Kit⁺ cells (Chi2 between downregulated and all
191 genes, $p = 3.61e-06$ for lincRNA; $p = 0.0029$ for lncRNA) (**Fig. 3e**).

192 In sum, the extent of gene deregulation induced by *Dot1l* loss increases progressively with
193 male germ cell development, in agreement with the timing of *Dot1l* conditional knockout [i.e.
194 *Stra8-Cre* expression starts in Kit⁻ and reaches its maximum efficiency in SCI (Sadate-Ngatchou
195 *et al*)]. Yet, gene deregulation in male germ cells of all stages is unexpectedly biased towards
196 upregulation, and often of large genes.

197
198

199 **DOT1L has opposite effects on gene expression depending on H3K79me2 but also H3K27me3** 200 **chromatin environment**

201 To address the link between gene expression deregulation following *Dot1l* inactivation and
202 H3K79 methylation, we compared H3K79me2 enrichment pattern at deregulated and not
203 deregulated genes. We observed that H3K79me2 is enriched at the body of the majority of
204 not deregulated genes, and of downregulated genes (in RS) while upregulated (UR) genes are
205 mainly devoid of the mark in both SCI and RS (Chi2: $p < 2.2e-16$ for down vs. not genes in SCI
206 and RS)(**Fig 4a**). Since we found a positive correlation between H3K79me2 enrichment at
207 enhancers and the expression level of neighboring genes (**Fig 2c** and **Fig S2**), we hence asked

Fig. 4

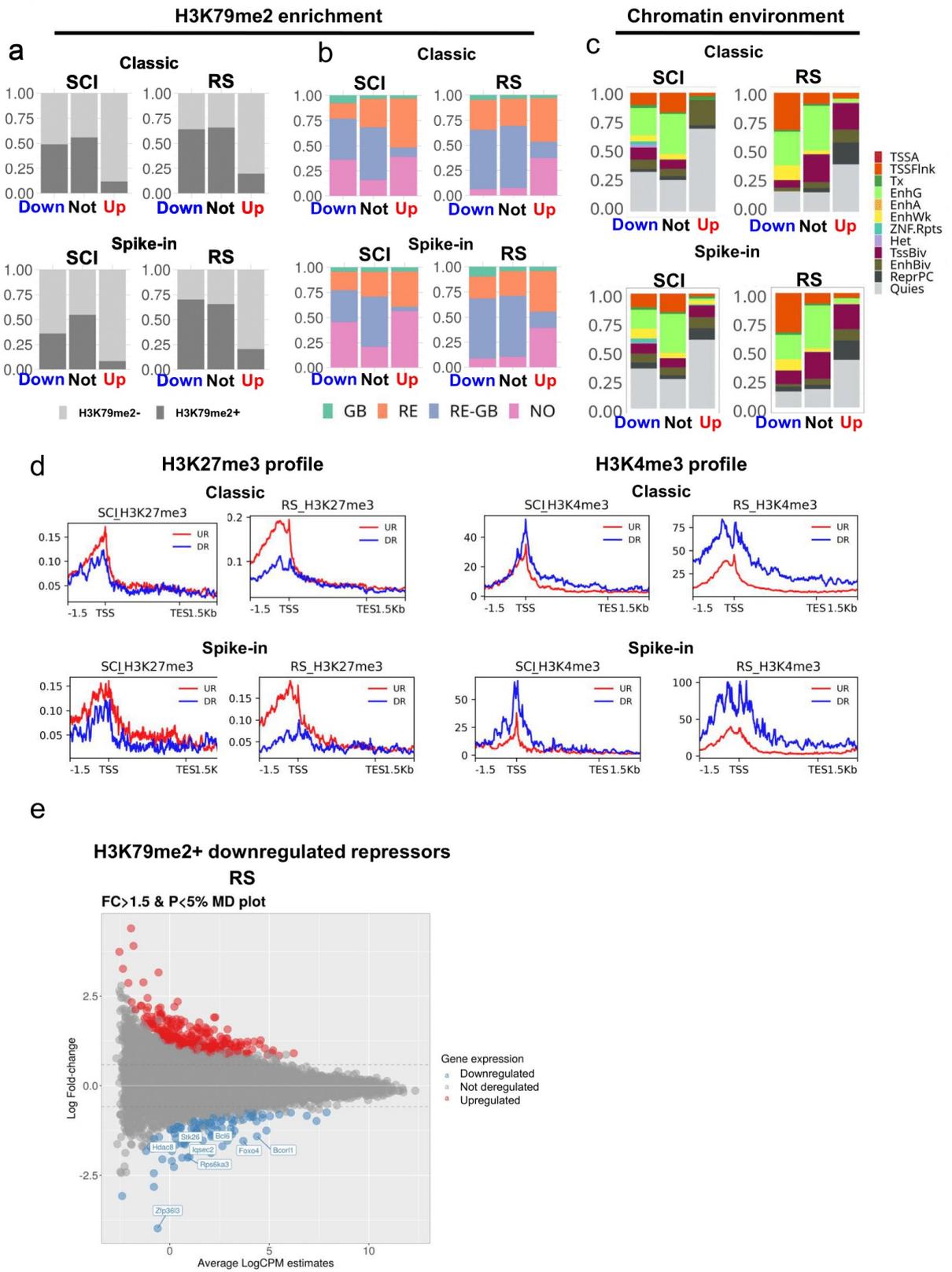


Figure 4. Correlation between differentially expressed genes and their chromatin environment.

a. Distribution of H3K79me2 enriched (H3K79me2+, dark grey) or not (H3K79me2-, light grey) genes among different groups of differentially expressed genes (Up: upregulated genes, Down : downregulated genes, Not: not deregulated genes) following Dot1l KO. H3K79me2 enrichment was evaluated at the gene body in SCI and RS. Kit-/GSC data was not included because only few genes were found deregulated at this stage. NB. “classic” data were already shown in (Blanco et al 2023) for RS. **b.** Distribution of H3K79me2 enrichment at or nearby genes (at the gene body (GB), at enhancers/regulatory elements (RE) as defined by ChromHMM analysis, at both regions (GB-RE) or in none of these (NO)), in different gene categories (up, down or not deregulated), in SCI and RS. **c.** Chromatin states (as defined in Fig. 2a) associated with the promoter region (+/-3Kb interval around the TSS) of deregulated genes or not, in SCI and RS. **d.** H3K27me3 and H3K4me3 enrichment profiles at differentially expressed genes. **e.** MD plot showing differential expression levels (Log fold change) according to gene expression level (LogCPM) in RS (spike-in RNA-seq analysis). Significantly downregulated genes that are enriched in H3K79me2 and known transcriptional repressors are indicated in the figure.

208 whether deregulated genes in *Dot1l*-KO compared to control (CTL) male germ cells are
 209 associated with H3K79me2+ enhancers (**Fig 4b**). The results show that, in SCI and RS, the
 210 enrichment profile between downregulated genes and unaffected genes is similar, with a
 211 majority of genes enriched in H3K79me2 at their body and/or nearby enhancers (presence of
 212 H3K79me2 at gene body and/or enhancers in ~79% of not deregulated genes and ~54% of
 213 downregulated genes with H3K79me2, for SCI; ~90% of not deregulated genes and ~91% of
 214 downregulated genes for RS – spike in analysis). In contrast, upregulated genes are significantly
 215 enriched in H3K79me2 enhancers only regions representing 35% of upregulated (vs. 24% of
 216 not deregulated) genes in SCI and 40% of upregulated (vs. 24% of not deregulated) genes in
 217 RS (Chi2: $p < 2.2 \times 10^{-16}$ for down vs. not genes in SCI and RS) (**Fig 4b**).

218 Upregulated genes are, therefore, mostly enriched in H3K79me2 strictly at their enhancers, or
 219 devoid of H3K79me2. These results indicate different mechanisms of control of gene
 220 expression by DOT1L, and that the repressive effect of DOT1L on gene expression is, at least
 221 partially, independent of H3K79me2.

222 To search what drives the observed gene deregulation in an H3K79me2-independent manner,
 223 we next investigated the chromatin environment of deregulated genes. To do so, we used
 224 ChromHMM (Fig. 2a) to assign the predominant chromatin state flanking the promoter of
 225 deregulated genes. First, we looked at downregulated genes and did not observe any
 226 chromatin signature, different from that of not deregulated genes, except for a decrease in the
 227 proportion of in the "genic enhancers" chromatin state in SCI and RS (EnhG, Chi2: $p = 0.0006$ for
 228 down vs. not genes in SCI and $p = 0.0025$ in RS) (**Fig. 4c**). In contrast, upregulated genes were
 229 highly associated to bivalent enhancers ("classic analysis" EnhBiv: $p = 0.0004$ for SCI and $p =$
 230 9.8×10^{-5} in RS; "spike-in" EnhBiv: $p = 0.25$ for SCI and $p = 0.00078$ for RS), repressed Polycomb
 231 chromatin state ("classic analysis" ReprPC: $p = 2.5 \times 10^{-30}$ in RS; "spike-in analysis": ReprPC: $p =$
 232 3.6×10^{-5} for SCI and 2.8×10^{-28} for RS), and to quiescent state ("spike-in analysis": Quiescent $p =$
 233 7.01×10^{-17} for SCI and $p = 7.27 \times 10^{-30}$ for RS) (**Fig 4c**). These results indicate that many upregulated
 234 genes are particular located in H3K27me3-rich regions. H3K27me3 enrichment profile over
 235 deregulated genes confirmed that upregulated genes are enriched in H3K27me3; conversely
 236 downregulated genes are enriched in H3K4me3 (**Fig 4c**).

237 When including all chromatin states enriched in H3K27me3 (i.e. EnhBiv, ReprPC and TSSBiv),
 238 we found that ~30% and ~48% of upregulated genes are in H3K27me3-rich regions in SCI and
 239 RS, respectively, compared to 18% and 32% for not deregulated genes. Most of the other

240 upregulated genes (i.e. ~60% and 42% of upregulated genes vs. 26% and 16% of not
241 deregulated genes) do not bear any specific chromatin signature since they are located in
242 regions annotated as quiescent, i.e. without any specific mark. We next sought if those genes
243 could be upregulated as a consequence of the downregulation of transcription repressors. For
244 this, we listed downregulated genes annotated as repressors and which are enriched in
245 H3K79me2, and therefore presumed to be direct DOT1L-target genes. In contrast with what
246 was observed in B and T cells (Aslam *et al*, 2021), we did not find a downregulation of the gene
247 encoding for the H3K27me3 methyltransferase EZH2 in *Dot1l*-KO male germ cells at any stage.
248 However, the “spike-in” RNA-seq analysis unveils a downregulation of *Hdac8*, the gene
249 encoding for a histone deacetylase, previously shown to deacetylate H3K27 residue (Chen *et al*,
250 2019; Wang *et al*, 2022)(**Fig. 4e**). We also confirmed the downregulation of *Bcl6* (previously
251 observed in (Blanco *et al.*)) in meiotic and postmeiotic cells, and additionally found the
252 downregulation of *Bcor1*, which encodes a BCL6 co-repressor, in SCII and RS. Other genes
253 classified as repressors were found in this list, such as *Foxo4*, *lqsec2* *Rps6ka3* *Stk26* *Zfp36l3* and
254 *Zkscan17* (**Fig 4e and S3c**). Interestingly, *Zkscan17* (also known as *Nizp1*) is highly expressed in
255 male germ cells from the meiotic stage and encodes a KRAB domain zinc finger protein of yet
256 unknown role in spermatogenesis. In premeiotic Kit- or Kit+ spermatogonial cells, none of
257 these repressor genes were downregulated.
258 Collectively, our data suggest that, in the male germline, DOT1L represses gene expression via
259 several mechanisms, linked to chromatin environment and nearby regulatory elements but
260 also through indirect repressive effects.

261

262 **DOT1L activates XY gene expression following meiotic sex chromosome inactivation, in**
263 **correlation with a specific H3K79me2 dynamic**

264 Among these repressor genes, most are encoded by the X chromosome, we therefore looked
265 at the chromosome location of deregulated genes and discovered a striking over-
266 representation of XY-encoded genes among downregulated genes in *Dot1l*-KO SCII and RS,
267 reaching ~45% and 48% of downregulated genes in RS, in “classic” and “spike in” RNA-seq,
268 respectively (**Fig. 5a and S4a**). The average Log2 fold change of autosomal vs. XY-encoded
269 deregulated genes was also significantly different in SCII and RS (**Fig. 5b**). When considering
270 “spike-in” analysis, the difference was also significant at earlier stages, though less striking
271 than in SCII and RS (**Fig. 5b**).

Fig. 5

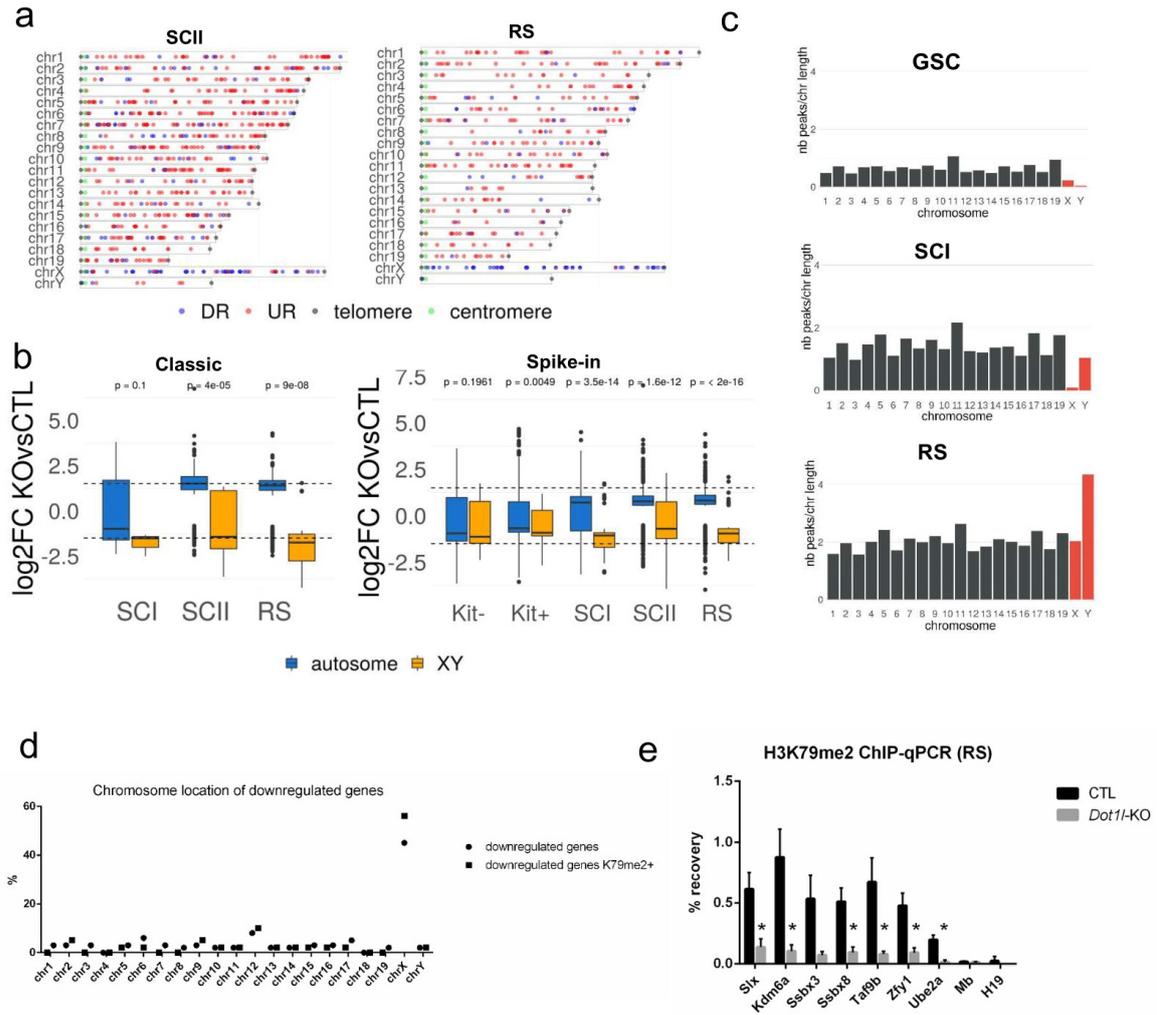


Figure 5. DOT1L regulates XY gene expression via H3K79me2.

a. Chromosome localization of deregulated genes (“spike-in” analysis) in SCII and RS. Blue: downregulated genes (DR), red: upregulated genes (UR), the green dot indicates the centromere, and the grey dot, the telomere. **b.** Comparison of gene expression levels between Dot1l-KO and CTL male germ cells. Boxplots showing gene expression log₂-fold change of Dot1l-KO relative to CTL. Box: 25th/75th percentiles. Bar in the box: median. Whiskers: 1.5 times the interquartile range from the 25th/75th percentiles. dashed lines: 1.5-fold change. ns: not significant. p-value calculated using Wilcoxon test are indicated above. **c.** Number of H3K79me2 peaks, normalized to the chromosome length, per chromosome in GSC, SCI and RS. Bars representing the results for sex chromosomes are colored in red. **d.** Chromosome location of downregulated genes in Dot1l-KO vs. CTL RS indicated in % (circle). The squares indicate the values obtained for downregulated genes which are marked by H3K79me2 at their gene body, at the TSS or at the closest distal regulatory region. **e.** Quantification of H3K79me2 level at the gene body of deregulated genes by ChIP-qPCR (n= 3 replicates for CTL and KO conditions). Bar plots indicate the mean

DOT1L a des effets antagonistes sur la régulation de l'expression génique liés aux chromosomes sexuels et autosomiques durant la spermatogenèse

enrichment value (IP/input recovery %) \pm SEM (the star indicates a $p < 0.05$ obtained with unpaired t-test).

272 The bias towards the downregulation of XY genes could be related to the specific regulation of
 273 XY gene expression during normal spermatogenesis. Indeed, in SCI at the pachytene stage, sex
 274 chromosome transcription is shut down following meiotic sex chromatin inactivation (MSCI)
 275 (Turner, 2015). This transcriptional shut down is associated with an enrichment in the
 276 heterochromatin mark H3K9me3 over the sex chromosomes (van der Heijden *et al*, 2007).
 277 When primary spermatocytes develop into secondary spermatocytes and then round
 278 spermatids, XY gene expression is activated as sex chromosomes acquire active histone marks,
 279 such as H3K4me3 and H3K27ac (Moretti *et al*; Sin *et al.*). With our ChIP-Seq data, we looked
 280 at H3K79me2 dynamic during wild-type spermatogenesis and observed a very different
 281 enrichment of the mark between sex chromosomes and autosomes (after normalization to
 282 the chromosome size). In particular, H3K79me2 enrichment on sex chromosomes increases
 283 dramatically from SCI to RS stage (**Fig. 5c**), and, indeed, many of peaks that are acquired at RS
 284 stage are located on these chromosomes.

285 This dynamic was directly linked with the observed downregulation of XY genes, since the
 286 majority of XY downregulated genes were found marked by H3K79me2 in wild-type RS (**Fig.**
 287 **5d**). By ChIP-qPCR we observed that H3K79me2 enrichment is lost at the body of several X and
 288 Y-encoded genes (**Fig. 5e**) indicating that DOT1L regulates XY gene activation following MSCI
 289 via H3K79me2. When compared to other histone marks known to be differentially enriched
 290 between sex chromosomes and autosomes in RS (Moretti *et al.*; Sin *et al.*), H3K79me2
 291 appeared to be the most dynamic mark between SCI and RS, with a 4-fold increase for
 292 chromosome Y and a 24-fold increase for chromosome X (**Fig. S4b, c and S5**), and therefore
 293 more enriched than H3K27ac, an active mark previously described to promote XY gene
 294 activation in RS (Adams *et al*; Moretti *et al.*).

295 When including GSC datasets, it appeared clear that chromatin remodeling dynamic is
 296 different between the X and Y chromosomes from an earlier stage. For instance, H3K79me2
 297 enrichment, normalized to the chromosome size, was higher on the Y than on the X, at the
 298 GSC to SCI transition; this difference was even more striking for H3K4me1, H3K27ac and
 299 H3K27me3 (**Fig. S5a and b**). As for H3K9me3, we previously showed that it is similarly enriched
 300 on the Y chromosome than chromosomes X and 14, based on Bryant *et al* RS dataset (Bryant
 301 *et al*) and using BWA aligner (Moretti *et al.*). Here, using bowtie2 aligner (which is more robust
 302 compared to BWA (Zhang *et al*, 2021)) and the recently published data from Liu *et al*,
 303 2017) which includes GSC and SCI datasets, we observed that H3K9me3 enrichment as

304 spermatocytes specialize into spermatids is higher on the X than on the Y chromosome, and
305 higher on chromosome 14 than other autosomes (**Fig. S4b**). We also observed that H3K9me3
306 is prominent on sex chromosomes compared to autosomes since the pre-meiotic cell stage
307 (**Fig. S5a**).

308 Overall, our data show that a high proportion of downregulated genes are enriched in
309 H3K79me2 and are located on sex chromosomes suggesting that DOT1L is required for XY gene
310 activation, following MSCI, likely through *de novo* H3K79 dimethylation. Our data also indicate
311 that the X and Y chromosomes have distinct chromatin remodeling dynamics following the
312 MSCI and that H3K79 methylation is the most dynamic histone mark involved in sex
313 chromosomes epigenetic remodeling during spermatogenesis.

314

315 **Discussion**

316 DOT1L is an essential regulator of development and differentiation, but its underlying
317 molecular mechanism remains often unclear. In the recent years, several studies have
318 investigated the link between DOT1L, H3K79 methylation and transcription regulation using
319 transcriptomic analyses (Appiah *et al.*, 2023; Aslam *et al.*, 2021; Cao *et al.*, 2020; Cattaneo *et al.*,
320 2022; Godfrey *et al.*, 2019; Kwesi-Maliepaard *et al.*, 2020); they indicate that DOT1L, initially
321 thought to be a transcription activator, has more diverse effects on gene expression, with
322 common and specific features depending on the studied cell type (Wille & Sridharan). In the
323 male germline, DOT1L is involved at several steps: before meiosis, in the self-renewal of
324 spermatogonial stem cells and, after meiosis, in the regulation of histone-to-protamine
325 transition in spermatids and their proper differentiation into functional spermatozoa (Blanco
326 *et al.*, 2023; Lin *et al.*, 2022; Malla *et al.*, 2023a). The transcriptional and epigenetic landscapes
327 are highly dynamic during male germ cell differentiation, and DOT1L and H3K79 methylation
328 levels are particularly high. Consequently, spermatogenesis constitutes a great model to
329 investigate the link between DOT1L, H3K79 methylation and transcription. To this aim, we
330 studied the transcriptional changes induced by *Dot1l*-KO throughout spermatogenesis,
331 profiled H3K79me2 at different cell stages (pre-, meiotic, and post-meiotic cells) using ChIP-
332 seq, and integrated these data with male germ cell chromatin environment inferred from 6
333 other histone marks.

334 **H3K79me2 dynamic correlates with male germ cell genetic program**

335 Our ChIP-seq results show that the number of H3K79me2+ regions increases with male germ
 336 cell progression, and that many peaks are established after meiosis (i.e. a ~3-fold increase in
 337 RS compared to GSC). This pattern correlates with the increase in DOT1L protein level observed
 338 at this stage (Blanco *et al.*; Dottermusch-Heidel *et al.*; Dottermusch-Heidel *et al.*; Moretti *et*
 339 *al.*) and indicates that, although DOT1L is presumed to have other functions (Wille & Sridharan,
 340 2022; Zhao *et al.*, 2023), its H3K79 methyltransferase activity is a major consequence of its
 341 expression during spermatogenesis.

342 Our data also show that cell stage-specific H3K79me2 peaks are located in genes with
 343 functions consistent with the corresponding cell-specific transcriptional program, such as Wnt
 344 signaling pathway for GSC, DNA repair, chromosome segregation and nuclear division for SCI,
 345 and cilium and spermatid development for RS. This demonstrates that H3K79 methylation
 346 genomic landscape is highly remodeled throughout spermatogenesis and that the
 347 redistribution and acquisition of H3K79me2 is linked to the ongoing developmental biological
 348 process.

349 It has been widely described that H3K79me2 is enriched at the gene body of transcriptionally
 350 active genes, but this pattern is not exclusive since H3K79 methylation has also been reported
 351 at enhancers, at replication origins and at repetitive elements (Cattaneo *et al.*, 2022; Godfrey
 352 *et al.*, 2019; Malla *et al.*, 2023b; Steger *et al.*, 2008). In male germ cells, we observed the typical
 353 H3K79me2 enrichment at gene body (including the promoter), but also at distal intergenic
 354 regions. Interestingly, as spermatogenesis advances, these regions reach ~15% of all peaks
 355 found at the RS stage, and ~25% of RS-specific peaks.

356 Using previously published ChIP-seq datasets to define male germ cells chromatin states, and
 357 in particular those corresponding to enhancers, we show that all types of H3K79me2+
 358 enhancers are associated with a higher expression of nearby genes than H3K79me2-
 359 enhancers, in line with Godfrey *et al.* findings in MLL cell lines (Godfrey *et al.*, 2019). When we
 360 investigated the link between H3K79me2 and *Dot1l*-KO induced gene deregulation, we
 361 observed that the majority of downregulated genes are enriched in H3K79me2 at their body
 362 and/or enhancer, in agreement with the positive correlation between H3K79me2 and gene
 363 expression. For upregulated genes, however, H3K79me2 was only enriched at enhancer
 364 regions, suggesting that H3K79me2 link with gene expression depends on its pattern and on
 365 its chromatin environment (see also below). These results agree with those obtained by
 366 Cattaneo *et al.* on cardiomyocytes, in which H3K79me2 at gene bodies and enhancers have

367 been found to promote gene expression, while H3K79me2 restricted to enhancers regions
368 represses it (Cattaneo *et al.*).

369 Overall our results, along with previous findings, confirm the role of H3K79 methylation in
370 defining cell specific transcriptional signatures during differentiation processes including
371 spermatogenesis. ChIP-seq in mouse spermatozoa is particularly challenging due to the high
372 level of compaction of their chromatin and the limited number of retained histones, but it
373 would be interesting to follow H3K79me2 mark through mature spermatozoa and after
374 fecundation to assess whether H3K79me2 enrichment plays any role in early embryonic
375 development.

376 **DOT1L repressive effect at enhancers is mediated by an H3K27me3-dependent process**

377 Using a classic and a spike-in normalization, we investigated the consequences of *Dot1l*-KO on
378 the transcriptome of 5 different male germ cell types. Strikingly, at all stages, the inactivation
379 of *Dot1l* led to a higher proportion of upregulated genes than downregulated ones. In somatic
380 cells, similar results were obtained by others (Aslam *et al.*, 2021; Borosha *et al.*, 2022; Cattaneo
381 *et al.*, 2022; Feng *et al.*, 2022; Ferrari *et al.*, 2020; Franz *et al.*, 2019; Kwesi-Maliepaard *et al.*,
382 2020; Wille & Sridharan) but, to the best of our knowledge, our study is the first one that uses
383 spike-in normalization to confirm the transcriptional shift towards upregulation.

384 Our combined analysis of the enrichment in 6 histone marks assessed by ChIP-seq uncovered
385 distinct epigenetic signatures at the promoters of differentially expressed genes:
386 downregulated genes were co-marked with H3K4me3 while upregulated ones were enriched
387 in H3K27me3, among which many bivalent H3K4me3/H3K27me3 regions. Therefore, during
388 male germ differentiation, H3K79 methylation could cooperate with bivalent sites, and lead to
389 a shift from H3K27me3 to H3K79me2 at key differentiation genes, as suggested for
390 cardiomyocytes differentiation (Cattaneo *et al.*, 2016) and recently observed for neural stem
391 cell differentiation (Appiah *et al.*, 2023).

392 With this model, at progenitor stages, DOT1L would repress the late germ cells genetic
393 program, and, indeed, our gene set enrichment analysis showed that, in premeiotic cells (i.e.
394 Kit- and Kit+ spermatogonia), deregulated genes are involved in processes associated with the
395 postmeiotic differentiation of spermatids in spermatozoa, i.e. sperm motility and capacitation,
396 acrosome assembly and reaction, flagellum and cilium organization and assembly, etc.

397 In addition, an interaction between H3K79me2 at enhancers and H3K27ac has been found to
398 contribute to gene deregulation in *Dot1l*-KO cardiomyocytes (Cattaneo *et al.*). Interestingly, we

399 found that the expression of *Hdac8*, a gene coding for an H3K27 deacetylase, is directly
 400 controlled by DOT1L. HDAC8 has been shown to repress the expression of certain genes
 401 through its capacity to deacetylate enhancers (Chen *et al.*, 2019). In the early embryo, it can
 402 prevent the premature expression of developmental genes (Wang *et al.*, 2022). We therefore
 403 propose that DOT1L prevents the expression hundreds of genes by impeding H3K27
 404 acetylation and promoting H3K27 trimethylation at their enhancers. The fact that the gene
 405 coding for the H3K27 methylase, *Kdm6a*, is, like *Hdac8*, downregulated in KO SCII and RS,
 406 [present study and (Blanco *et al.*)] suggest a complex crosstalk that will require further work
 407 to elucidate.

408 Finally, we identified other repressors among DOT1L direct targets, indicating that DOT1L
 409 repressive effect is likely mediated by several pathways. This is in agreement with the
 410 repressive function of DOT1L observed by others (Zhao *et al.*, 2023).

411

412 **DOT1L activates XY gene expression by an H3K79me2-dependent mechanism**

413 A specific epigenetic regulation of sex chromosomes takes place during mammalian
 414 spermatogenesis, where X and Y chromosomes are transcriptionally silenced during meiosis,
 415 known as the meiotic sex chromosome inactivation (MSCI) (Turner). After meiosis, a number
 416 of sex chromosome genes are reactivated or expressed *de novo* (Moretti *et al.*).

417 Our study provides novel insights into the mechanism of XY gene regulation: first, we identify
 418 H3K79me2 as a hallmark of post-MSCI state since it is highly enriched in sex chromosomes
 419 compared to autosomes at RS stage, more than any other studied histone mark (H3K27ac,
 420 H3K27me3, H3K4me1, H3K4me3 or H3K9me3). Furthermore, we show that downregulated
 421 genes following *Dot1l* inactivation are predominantly sex chromosome-encoded and enriched
 422 in H3K79me2, indicating that H3K79me2 is required for post-MSCI gene activation. Our results
 423 are in agreement with DOT1L pattern of expression which has been shown to co-localize with
 424 the sex chromosomes towards the end of MSCI (from the diplotene stage) in meiotic and
 425 postmeiotic cells (Lin *et al.*; Moretti *et al.*; Ontoso *et al.*).

426 In sum, we show that DOT1L and H3K79me2 play a pivotal role in XY gene activation, following
 427 MSCI. The several chromatin regulators, including *Kdm6a* and *Hdac8*, which are located on the
 428 X chromosome and downregulated in *Dot1l*-KO could contribute to the observed down and
 429 upregulation of autosomal genes.

430

431 **Acknowledgement/funding**

432 This work was supported by the Agence Nationale de la Recherche (ANR-17-CE12-0004-01 to
433 J.C.), the Fondation pour la Recherche Médicale (SPF201909009274 to C.G.) and the Institut
434 Cochin (PIC to L.E.K).

435

436 **Material and Methods**

437 **Sample collection**

438 *For ChIP-seq*

439 ChIP-seq analyses were performed on male germ cells of four different types: undifferentiated
440 spermatogonia grown in culture (GSC), primary spermatocytes (SCI), round spermatids (RS)
441 and elongating/condensed spermatids (ES) (see Fig. 1a). Cells were isolated from adult (3 to 5
442 month-old) wild-type C57/Bl6 mice by elutriation, as described in (Cocquet *et al.*). ChIP-seq
443 experiments were performed as described in (Moretti *et al.*) using anti-H3K79me2 antibody
444 (Diagenode C15410051, Lot A1193D). Two biological replicates and one input sample were
445 produced by cell type.

446 *For RNA-seq*

447 RNA-seq analyses were performed on male germ cells from *Dot1l* conditional knock-out (KO)
448 and control mice (CTL) as described in (Blanco *et al.*, 2023). Male germ cells of five different
449 types [undifferentiated spermatogonia (Kit-), differentiated spermatogonia (Kit+), primary
450 spermatocytes (SCI), secondary spermatocytes (SCII) and round spermatids (RS)] were isolated
451 from adult (3 to 5 month-old) males by flow cytometry sorting according to their size and DNA
452 content as described in (Bastos *et al.*, 2005). Total RNA was extracted using the RNAqueous
453 Ambion kit (Thermofisher) following the manufacturer's protocol. In total, three replicates
454 were generated for Kit- and Kit+ cells, and five replicates for SCI, SCII and RS. For "spike-in"
455 analyses, ERCC (External RNA Controls Consortium) was added to RNA samples according to
456 the quantity of cells used for RNA isolation, following the manufacturer's instructions
457 (Ambion). ERCC are exogenous synthetic polyadenylated transcripts of 25 to 2000nts which
458 can be added to samples before library preparation to serve as internal standards for RNA-seq.
459 The transcripts are designed to mimic natural eukaryotic mRNAs. ERCC was added in the three
460 replicates of each cell type. Libraries were prepared for all samples using the NEB Next Ultra II
461 Directional RNA Library Prep Kit (New England Biolabs) according to supplier

462 recommendations. Paired-end sequencing of 100-bp reads was then carried out on the
463 Illumina HiSeq4000 system.

464

465 **ChIP-seq analyses**

466 Snakemake was used for ChIP-seq analyses (Koster & Rahmann, 2018). Adaptors were trimmed
467 and reads with poor quality (quality < 20) were filtered out with BBduk from BBTools (v. 38.23,
468 <https://sourceforge.net/projects/bbmap/>). Alignment was performed on the mouse reference
469 genome (mm10) using bowtie2 (Langmead & Salzberg, 2012) with the following arguments: -
470 -local. Peak calling was processed against the input file using the MACS2 tool (Zhang *et al*,
471 2008) with a “broad pattern” setting to obtain peaks enrichment score of H3K79me2. The
472 parameter “broad” is also used for H3K36me3 and H3K27me3 histone marks. Mitochondrial
473 peaks and peaks with FDR > 5% were removed. Peaks were annotated with the R package
474 ChIPseeker (Yu *et al*, 2015) using the annotatePeak() function with the following parameters
475 tssRegion = c(-3000,3000), and divided in 4 categories: promoter (promoter’s annotations),
476 intragenic (exon, intro and UTR), downstream and distal intergenic.

477 Common peaks were defined as regions with a minimum overlap of 1bp coverage using
478 BEDtools merge. Ontology analyses for H3K79me2 genic peaks were performed with
479 compareCluster function from clusterProfiler package (Yu *et al*, 2012) using BP (biological
480 processes) database and Benjamini-Hochberg correction to calculate the adjusted p-value.

481

482 **Characterization of chromatin states using Hidden Markov Modeling**

483 ChromHMM (Ernst & Kellis, 2017) was used with default parameters to predict the chromatin
484 state associated with each genomic region (parameter bin size = 200bp). We created a model
485 of 18 chromatin states by integrating the data from 6 histone marks (H3K4me1, H3K4me3,
486 H3K27ac, H3K27me3, H3K36me3 and H3K9me3) originating from 17 different cell
487 types/stages, retrieved from the literature and the mouse ENCODE project (Consortium,
488 2012). ChIP-seq data from male germ cells were obtained from (Chen *et al*, 2020; Hammoud
489 *et al*, 2014; Liu *et al*, 2019; Zuo *et al*, 2018). Chromatin states were named according to
490 ChromHMM nomenclature (Ernst & Kellis, 2017; Ferrari *et al.*, 2020; Godfrey *et al.*, 2019)(van
491 der Velde *et al*, 2021). Four chromatin states were defined as enhancers: active enhancers
492 (EnhA), genic enhancers (EnhG), weak enhancers (EnhW) and bivalent enhancers (EnhBiv).
493 Each type was subdivided into H3K79me2+ or H3K79me2-, according to the enrichment score

494 from H3K79me2 ChIP-seq data. To assign genes to each category of enhancers, two method
495 were used: the first one, based on (Rada-Iglesias *et al*), assigns each enhancer to the nearest
496 gene within 100kb from each side of the enhancer boundaries.

497 The second one associates each enhancer with all genes found in the same interval,
498 considering that one enhancer may regulate the expression of several genes (Karnuta &
499 Scacheri, 2018)). We used ChromENVEE package
500 (<https://github.com/ManonCoulee/ChromENVEE>) to calculate the average expression level
501 (CPM) of genes within the interval of +/-100kb from the enhancer of interest with
502 enhancerExpression function. Statistical analysis are performed using wilcoxon test between
503 enhancer with (+H3K79me2) and without H3K79me2 (-H3K79me2). GeneEnvironment()
504 function with default parameters was used to identify the chromatin states at promoter
505 regions of genes deregulated in *Dot1l* conditional knock-out (KO) vs. control (CTL) samples (see
506 below, for detailed RNA-seq analyses). A predominant chromatin state was assigned to each
507 deregulated gene by determining the chromatin environment around the TSS (+/-3kb) and
508 selecting the state with the largest coverage in base pairs.

509

510 **RNA-seq analyses**

511 Five samples of each genotype (KO and CTL) for SCI, SCII and RS, and three samples for Kit- and
512 Kit+ were analyzed. Snakemake was used for RNA-seq analyses (Koster & Rahmann, 2018)(v.
513 3.9.0). Adaptors were trimmed and reads of low quality filtered out using the same parameters
514 than for ChIP-seq analyses. Alignment was performed on the mouse reference genome build
515 mm10 (GRCm38.p6) using STAR (Dobin & Gingeras, 2015) (v. 2.7.2d) and Gencode vM19 gene
516 annotation GTF with following arguments: --outMultimapperOrder Random --quantMode
517 Genecounts --sjdbOverhang 74. For each sample, the number of aligned reads was converted
518 to cpm (counts per million). Genes with an expression level of at least 1cpm in a minimum of
519 two samples were included in the analysis, to exclude genes with a very low expression level.
520 Differential expression analysis was carried out using DESeq2 (Love *et al*, 2014) and edgeR
521 (Robinson *et al*, 2010) packages with default parameters and a design accounting for genotype
522 and cell type induced variability. Differentially expressed genes (FDR<5%) were obtained using
523 glmTreat edgeR's function which performs a modified likelihood ratio test (LRT) against the
524 fold change threshold (> 1.5) based on TREAT method (McCarthy & Smyth, 2009). Deregulated
525 genes with FDR < 5% (adjusted p-value) and a Fold Change >1.5 were selected. Enrichment

526 analysis was performed with GSEA (v. 4.0.3, c5.all.v7.1.symbols.gmt, GO cellular components,
 527 biological pathways, or all) by processing gene expression cpm values from KO and control
 528 samples (Subramanian *et al*, 2005) and using same parameter than Blanco paper (Blanco *et al*,
 529 2023). Beforehand, mouse Ensembl gene IDs were converted to human Ensembl gene IDs
 530 (hg38) with the bioMart R package (Durinck *et al*, 2005).

531

532 **Identification of gene expression repressors**

533 We checked whether the list of downregulated genes included transcriptional repressors
 534 linked with one of the following pathways: “negative regulation”, “negative transcription of
 535 regulation”, “RNA polymerase II and repressor” and enriched in H3K79me2 at gene body.

536

537 **“spike-in” RNA-seq analysis**

538 Five samples of each genotype (KO and CTL) were analyzed including Kit-, Kit+, SCI, SCII and RS
 539 (in which ERCC was added prior to library preparation in all three replicates). RNA “spike-in”
 540 data were analyzed using the same alignment pipeline described for classic RNA-seq data.
 541 Normalization was realized using RUV method and RUVseq R package (Risso *et al.*, 2014) to
 542 remove unwanted variation, followed by differential gene expression analysis. It consists in
 543 using ERCC values as control to normalize the data. Differential expression analysis was then
 544 performed using the same parameters as described above.

545

546 **Gene length, functional annotation**

547 Gene length and functional annotation were performed comparing deregulated genes with
 548 GTF information from mm10. For gene length information, we divided gene length in ten
 549 categories according to percentile value. Functional annotation were divided in six groups of
 550 annotation: IG-TR, Protein coding, Pseudogene, lncRNA, mtRNA and snoRNA. Statistical
 551 analysis were performed using Chi2 test by comparison of deregulated genes with not
 552 deregulated genes.

553

554 **Data availability**

555 RNA-Seq data have been submitted to ENA repository under the project numbers PRJEB50887
 556 and PRJEB64263 (<https://www.ebi.ac.uk/ena/>). H3K79me2 ChIP-Seq project number is
 557 PRJNA643726 and PRJEB64263. The project identifiers for ChIP-seq data from spermatogonial

558 cells are PRJNA548107 and PRJEB64263, and from spermatocytes and round spermatids are
559 SRP028576 and PRJNA428362. H3K9me3 ChIP-seq project number is PRJNA566429. The other
560 ChIP-seq data that were used are available in the ENCODE project database
561 (<https://www.encodeproject.org/>).

562

563 **Code availability**

564 All scripts used in this paper can be found in the GitHub repository
565 (https://github.com/ManonCoulee/H3K79me2_Coulee_2023)

566

567

568

569 **References**

570

571 Adams SR, Maezawa S, Alavattam KG, Abe H, Sakashita A, Shroder M, Broering TJ, Sroga Rios
572 J, Thomas MA, Lin X *et al* (2018) RNF8 and SCML2 cooperate to regulate ubiquitination and
573 H3K27 acetylation for escape gene activation on the sex chromosomes. *PLoS Genet* 14:
574 e1007233

575 Alavattam KG, Namekawa SH (2020) Licensing meiotic progression dagger. *Biology of
576 reproduction* 103: 10-12

577 Appiah B, Fullio CL, Ossola C, Bertani I, Restelli E, Cheffer A, Polenghi M, Haffner C, Garcia-
578 Miralles M, Zeis P *et al* (2023) DOT1L activity affects neural stem cell division mode and reduces
579 differentiation and ASNS expression. *EMBO Rep* 24: e56233

580 Aslam MA, Alemdehy MF, Kwesi-Maliepaard EM, Muhaimin FI, Caganova M, Pardieck IN, van
581 den Brand T, van Welsem T, de Rink I, Song JY *et al* (2021) Histone methyltransferase DOT1L
582 controls state-specific identity during B cell differentiation. *EMBO Rep* 22: e51184

583 Bao J, Bedford MT (2016) Epigenetic regulation of the histone-to-protamine transition during
584 spermiogenesis. *Reproduction (Cambridge, England)* 151: R55-70

585 Bastos H, Lassalle B, Chicheportiche A, Riou L, Testart J, Allemand I, Fouchet P (2005) Flow
586 cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse
587 spermatogenesis. *Cytometry A* 65: 40-49

588 Bernt KM, Zhu N, Sinha AU, Vempati S, Faber J, Krivtsov AV, Feng Z, Punt N, Daigle A, Bullinger
589 L *et al* (2011) MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by
590 DOT1L. *Cancer Cell* 20: 66-78

591 Blanco M, Cocquet J (2019) Genetic Factors Affecting Sperm Chromatin Structure. *Adv Exp Med
592 Biol* 1166: 1-28

593 Blanco M, El Khattabi L, Gobe C, Crespo M, Coulee M, de la Iglesia A, Ialy-Radio C, Lapoujade
594 C, Givelet M, Delessard M *et al* (2023) DOT1L regulates chromatin reorganization and gene
595 expression during sperm differentiation. *EMBO Rep* 24: e56316

596 Borosha S, Ratri A, Ghosh S, Malcom CA, Chakravarthi VP, Vivian JL, Fields TA, Rumi MAK, Fields
597 PE (2022) DOT1L Mediated Gene Repression in Extensively Self-Renewing Erythroblasts. *Front
598 Genet* 13: 828086

- 599 Bryant JM, Donahue G, Wang X, Meyer-Ficca M, Luense LJ, Weller AH, Bartolomei MS, Blobel
600 GA, Meyer RG, Garcia BA *et al* (2015) Characterization of BRD4 during mammalian postmeiotic
601 sperm development. *Mol Cell Biol* 35: 1433-1448
602 Cao K, Ugarenko M, Ozark PA, Wang J, Marshall SA, Rendleman EJ, Liang K, Wang L, Zou L,
603 Smith ER *et al* (2020) DOT1L-controlled cell-fate determination and transcription elongation
604 are independent of H3K79 methylation. *Proceedings of the National Academy of Sciences of
605 the United States of America* 117: 27365-27373
606 Cattaneo P, Hayes MGB, Baumgarten N, Hecker D, Peruzzo S, Aslan GS, Kunderfranco P, Larcher
607 V, Zhang L, Contu R *et al* (2022) DOT1L regulates chamber-specific transcriptional networks
608 during cardiogenesis and mediates postnatal cell cycle withdrawal. *Nature communications*
609 13: 7444
610 Cattaneo P, Kunderfranco P, Greco C, Guffanti A, Stirparo GG, Rusconi F, Rizzi R, Di Pasquale E,
611 Locatelli SL, Latronico MV *et al* (2016) DOT1L-mediated H3K79me2 modification critically
612 regulates gene expression during cardiomyocyte differentiation. *Cell death and differentiation*
613 23: 555-564
614 Chen LF, Lin YT, Gallegos DA, Hazlett MF, Gomez-Schiavon M, Yang MG, Kalmeta B, Zhou AS,
615 Holtzman L, Gersbach CA *et al* (2019) Enhancer Histone Acetylation Modulates Transcriptional
616 Bursting Dynamics of Neuronal Activity-Inducible Genes. *Cell reports* 26: 1174-1188 e1175
617 Chen Y, Lyu R, Rong B, Zheng Y, Lin Z, Dai R, Zhang X, Xie N, Wang S, Tang F *et al* (2020) Refined
618 spatial temporal epigenomic profiling reveals intrinsic connection between PRDM9-mediated
619 H3K4me3 and the fate of double-stranded breaks. *Cell Res* 30: 256-268
620 Cocquet J, Ellis PJ, Yamauchi Y, Mahadevaiah SK, Affara NA, Ward MA, Burgoyne PS (2009) The
621 multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis.
622 *PLoS biology* 7: e1000244
623 Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome.
624 *Nature* 489: 57-74
625 da Cruz I, Rodriguez-Casuriaga R, Santanaque FF, Farias J, Curti G, Caprano CA, Folle GA,
626 Benavente R, Sotelo-Silveira JR, Geisinger A (2016) Transcriptome analysis of highly purified
627 mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to
628 postmeiotic-related processes at pachytene stage. *BMC genomics* 17: 294
629 Dobin A, Gingeras TR (2015) Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics*
630 51: 11 14 11-19
631 Dottermusch-Heidel C, Gartner SM, Tegeder I, Rathke C, Barckmann B, Bartkuhn M, Bhushan
632 S, Steger K, Meinhardt A, Renkawitz-Pohl R (2014a) H3K79 methylation: a new conserved mark
633 that accompanies H4 hyperacetylation prior to histone-to-protamine transition in *Drosophila*
634 and rat. *Biol Open* 3: 444-452
635 Dottermusch-Heidel C, Klaus ES, Gonzalez NH, Bhushan S, Meinhardt A, Bergmann M,
636 Renkawitz-Pohl R, Rathke C, Steger K (2014b) H3K79 methylation directly precedes the
637 histone-to-protamine transition in mammalian spermatids and is sensitive to bacterial
638 infections. *Andrology* 2: 655-665
639 Ernst C, Eling N, Martinez-Jimenez CP, Marioni JC, Odom DT (2019) Staged developmental
640 mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. *Nature
641 communications* 10: 1251
642 Ernst J, Kellis M (2017) Chromatin-state discovery and genome annotation with ChromHMM.
643 *Nat Protoc* 12: 2478-2492

- 644 Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y (2002) Methylation
645 of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* 12:
646 1052-1058
- 647 Feng Y, Borosha S, Ratri A, Lee EB, Wang H, Fields TA, Kinsey WH, Vivian JL, Rumi MAK, Fields
648 PE (2022) DOT1L Methyltransferase Regulates Calcium Influx in Erythroid Progenitor Cells in
649 Response to Erythropoietin. *Int J Mol Sci* 23
- 650 Ferrari F, Arrigoni L, Franz H, Izzo A, Butenko L, Trompouki E, Vogel T, Manke T (2020) DOT1L-
651 mediated murine neuronal differentiation associates with H3K79me2 accumulation and
652 preserves SOX2-enhancer accessibility. *Nature communications* 11: 5200
- 653 Franz H, Villarreal A, Heidrich S, Videm P, Kilpert F, Mestres I, Calegari F, Backofen R, Manke T,
654 Vogel T (2019) DOT1L promotes progenitor proliferation and primes neuronal layer identity in
655 the developing cerebral cortex. *Nucleic Acids Res* 47: 168-183
- 656 Godfrey L, Crump NT, Thorne R, Lau IJ, Repapi E, Dimou D, Smith AL, Harman JR, Telenius JM,
657 Oudelaar AM *et al* (2019) DOT1L inhibition reveals a distinct subset of enhancers dependent
658 on H3K79 methylation. *Nature communications* 10: 2803
- 659 Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ,
660 Moore BB *et al* (2018) A Comprehensive Roadmap of Murine Spermatogenesis Defined by
661 Single-Cell RNA-Seq. *Developmental cell* 46: 651-667 e610
- 662 Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and
663 transcription initiation at most promoters in human cells. *Cell* 130: 77-88
- 664 Hammoud SS, Low DH, Yi C, Carrell DT, Guccione E, Cairns BR (2014) Chromatin and
665 transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell*
666 *stem cell* 15: 239-253
- 667 Jones B, Su H, Bhat A, Lei H, Bajko J, Hevi S, Baltus GA, Kadam S, Zhai H, Valdez R *et al* (2008)
668 The histone H3K79 methyltransferase Dot1L is essential for mammalian development and
669 heterochromatin structure. *PLoS Genet* 4: e1000190
- 670 Karnuta JM, Scacheri PC (2018) Enhancers: bridging the gap between gene control and human
671 disease. *Human molecular genetics* 27: R219-R227
- 672 Kim SK, Jung I, Lee H, Kang K, Kim M, Jeong K, Kwon CS, Han YM, Kim YS, Kim D *et al* (2012)
673 Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing
674 RNA polymerase II to regulate gene expression. *The Journal of biological chemistry* 287: 39698-
675 39709
- 676 Kim W, Choi M, Kim JE (2014) The histone methyltransferase Dot1/DOT1L as a critical regulator
677 of the cell cycle. *Cell Cycle* 13: 726-738
- 678 Koster J, Rahmann S (2018) Snakemake-a scalable bioinformatics workflow engine.
679 *Bioinformatics* 34: 3600
- 680 Kwesi-Maliepaard EM, Aslam MA, Alemdehy MF, van den Brand T, McLean C, Vlaming H, van
681 Welsem T, Korthout T, Lancini C, Hendriks S *et al* (2020) The histone methyltransferase DOT1L
682 prevents antigen-independent differentiation and safeguards epigenetic identity of CD8(+) T
683 cells. *Proceedings of the National Academy of Sciences of the United States of America* 117:
684 20706-20716
- 685 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:
686 357-359
- 687 Lin H, Cheng K, Kubota H, Lan Y, Riedel SS, Kakiuchi K, Sasaki K, Bernt KM, Bartolomei MS, Luo
688 M *et al* (2022) Histone methyltransferase DOT1L is essential for self-renewal of germline stem
689 cells. *Genes & development* 36: 752-763

- 690 Lin H, Cossu IG, Leu NA, Deshpande AJ, Bernt KM, Luo M, Wang PJ (2023) The DOT1L-MLLT10
691 complex regulates male fertility and promotes histone removal during spermiogenesis.
692 *Development (Cambridge, England)* 150
- 693 Liu S, Yu H, Liu Y, Liu X, Zhang Y, Bu C, Yuan S, Chen Z, Xie G, Li W *et al* (2017) Chromodomain
694 Protein CDYL Acts as a Crotonyl-CoA Hydratase to Regulate Histone Crotonylation and
695 Spermatogenesis. *Mol Cell* 67: 853-866 e855
- 696 Liu Y, Zhang Y, Yin J, Gao Y, Li Y, Bai D, He W, Li X, Zhang P, Li R *et al* (2019) Distinct H3K9me3
697 and DNA methylation modifications during mouse spermatogenesis. *The Journal of biological*
698 *chemistry* 294: 18714-18725
- 699 Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for
700 RNA-seq data with DESeq2. *Genome biology* 15: 550
- 701 Maezawa S, Yukawa M, Alavattam KG, Barski A, Namekawa SH (2018) Dynamic reorganization
702 of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res*
703 46: 593-608
- 704 Malla AB, Rainsford SR, Smith ZD, Lesch BJ (2023a) DOT1L promotes spermatid differentiation
705 by regulating expression of genes required for histone-to-protamine replacement.
706 *Development (Cambridge, England)* 150
- 707 Malla AB, Yu H, Farris D, Kadimi S, Lam TT, Cox AL, Smith ZD, Lesch BJ (2023b) DOT1L bridges
708 transcription and heterochromatin formation at mammalian pericentromeres. *EMBO Rep* 24:
709 e56492
- 710 McCarthy DJ, Smyth GK (2009) Testing significance relative to a fold-change threshold is a
711 TREAT. *Bioinformatics* 25: 765-771
- 712 Moretti C, Serrentino ME, Ialy-Radio C, Delessard M, Soboleva TA, Tores F, Leduc M, Nitschke
713 P, Drevet JR, Tremethick DJ *et al* (2017) SLY regulates genes involved in chromatin remodeling
714 and interacts with TBL1XR1 during sperm differentiation. *Cell death and differentiation* 24:
715 1029-1044
- 716 Moretti C, Vaiman D, Tores F, Cocquet J (2016) Expression and epigenomic landscape of the
717 sex chromosomes in mouse post-meiotic male germ cells. *Epigenetics Chromatin* 9: 47
- 718 Nguyen AT, Taranova O, He J, Zhang Y (2011) DOT1L, the H3K79 methyltransferase, is required
719 for MLL-AF9-mediated leukemogenesis. *Blood* 117: 6912-6922
- 720 Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, Su L, Xu G, Zhang Y (2005) hDOT1L links histone
721 methylation to leukemogenesis. *Cell* 121: 167-178
- 722 Okuda H, Stanojevic B, Kanai A, Kawamura T, Takahashi S, Matsui H, Takaori-Kondo A,
723 Yokoyama A (2017) Cooperative gene activation by AF4 and DOT1L drives MLL-rearranged
724 leukemia. *The Journal of clinical investigation* 127: 1918-1931
- 725 Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J,
726 Gupta PB *et al* (2012) Chromatin-modifying enzymes as modulators of reprogramming. *Nature*
727 483: 598-602
- 728 Ontoso D, Kauppi L, Keeney S, San-Segundo PA (2014) Dynamics of DOT1L localization and
729 H3K79 methylation during meiotic prophase I in mouse spermatocytes. *Chromosoma* 123:
730 147-164
- 731 Park G, Gong Z, Chen J, Kim JE (2010) Characterization of the DOT1L network: implications of
732 diverse roles for DOT1L. *Protein J* 29: 213-223
- 733 Patel L, Kang R, Rosenberg SC, Qiu Y, Raviram R, Chee S, Hu R, Ren B, Cole F, Corbett KD (2019)
734 Dynamic reorganization of the genome shapes the recombination landscape in meiotic
735 prophase. *Nature structural & molecular biology* 26: 164-174

- 736 Pursani V, Bhartiya D, Tanavde V, Bashir M, Sampath P (2018) Transcriptional activator DOT1L
737 putatively regulates human embryonic stem cell differentiation into the cardiac lineage. *Stem*
738 *Cell Res Ther* 9: 97
- 739 Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2011) A unique
740 chromatin signature uncovers early developmental enhancers in humans. *Nature* 470: 279-
741 283
- 742 Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor analysis
743 of control genes or samples. *Nat Biotechnol* 32: 896-902
- 744 Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential
745 expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140
- 746 Sadate-Ngatchou PI, Payne CJ, Dearth AT, Braun RE (2008) Cre recombinase activity specific to
747 postnatal, premeiotic male germ cells in transgenic mice. *Genesis* 46: 738-742
- 748 Sin HS, Kartashov AV, Hasegawa K, Barski A, Namekawa SH (2015) Poised chromatin and
749 bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biol* 13:
750 53
- 751 Smith E, Lin C, Shilatfard A (2011) The super elongation complex (SEC) and MLL in
752 development and disease. *Genes & development* 25: 661-672
- 753 Soumillon M, Necsculea A, Weier M, Brawand D, Zhang X, Gu H, Barthes P, Kokkinaki M, Nef S,
754 Gnirke A *et al* (2013) Cellular source and mechanisms of high transcriptome complexity in the
755 mammalian testis. *Cell reports* 3: 2179-2190
- 756 Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim JE, Chen J,
757 Lazar MA *et al* (2008) DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously
758 coupled with gene transcription in mammalian cells. *Mol Cell Biol* 28: 2825-2839
- 759 Turner JM (2007) Meiotic sex chromosome inactivation. *Development (Cambridge, England)*
760 134: 1823-1831
- 761 Turner JM (2015) Meiotic Silencing in Mammals. *Annual review of genetics* 49: 395-412
- 762 van der Heijden GW, Derijck AA, Posfai E, Giele M, Pelczar P, Ramos L, Wansink DG, van der
763 Vlag J, Peters AH, de Boer P (2007) Chromosome-wide nucleosome replacement and H3.3
764 incorporation during mammalian meiotic sex chromosome inactivation. *Nature genetics* 39:
765 251-258
- 766 van der Velde A, Fan K, Tsuji J, Moore JE, Purcaro MJ, Pratt HE, Weng Z (2021) Annotation of
767 chromatin states in 66 complete mouse epigenomes during development. *Commun Biol* 4: 239
- 768 Vara C, Paytuví-Gallart A, Cuartero Y, Le Dily F, Garcia F, Salva-Castro J, Gomez HL, Julia E,
769 Moutinho C, Aiese Cigliano R *et al* (2019) Three-Dimensional Genomic Structure and Cohesin
770 Occupancy Correlate with Transcriptional Activity during Spermatogenesis. *Cell reports* 28:
771 352-367 e359
- 772 Vlaming H, van Leeuwen F (2016) The upstreams and downstreams of H3K79 methylation by
773 DOT1L. *Chromosoma*
- 774 Wang M, Chen Z, Zhang Y (2022) CBP/p300 and HDAC activities regulate H3K27 acetylation
775 dynamics and zygotic genome activation in mouse preimplantation embryos. *The EMBO*
776 *journal* 41: e112012
- 777 Wille CK, Sridharan R (2022) Connecting the DOTs on Cell Identity. *Front Cell Dev Biol* 10:
778 906713
- 779 Wu A, Zhi J, Tian T, Cihan A, Cevher MA, Liu Z, David Y, Muir TW, Roeder RG, Yu M (2021) DOT1L
780 complex regulates transcriptional initiation in human erythroleukemic cells. *Proceedings of the*
781 *National Academy of Sciences of the United States of America* 118

Résultats

- 782 Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological
783 themes among gene clusters. *OMICS* 16: 284-287
- 784 Yu G, Wang LG, He QY (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation,
785 comparison and visualization. *Bioinformatics* 31: 2382-2383
- 786 Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F *et al* (2021)
787 Fast alignment and preprocessing of chromatin profiles with Chromap. *Nature*
788 *communications* 12: 6566
- 789 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown
790 M, Li W *et al* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology* 9: R137
- 791 Zhao X, Li X, Sun H, Zhao X, Gao T, Shi P, Chen F, Liu L, Lu X (2023) Dot1l cooperates with Npm1
792 to repress endogenous retrovirus MERVL in embryonic stem cells. *Nucleic Acids Res*
- 793 Zuo X, Rong B, Li L, Lv R, Lan F, Tong MH (2018) The histone methyltransferase SETD2 is required
794 for expression of acrosin-binding protein 1 and protamines and essential for spermiogenesis
795 in mice. *The Journal of biological chemistry* 293: 9188-9197
- 796
- 797

Discussion

1 DOT1L dans la spermatogenèse

Nous avons étudié les effets d'une délétion conditionnelle de *Dot1l* (KO-*Dot1l*) dans les cellules germinales mâles murines. L'analyse phénotypique de ce modèle montre que DOT1L est nécessaire à la différenciation des spermatides. En effet, le KO de *Dot1l* entraîne une hypofertilité pouvant être expliquée par la production de spermatozoïdes non fonctionnels présentant des altérations du flagelle et des défauts au niveau de la tête du spermatozoïde, caractérisés notamment par des défauts de compaction et de composition de sa chromatine. Ces résultats démontrent que DOT1L est bien impliquée dans le remodelage de la chromatine des spermatides, comme suggéré initialement dans le cadre de l'analyse de souris déficientes en *Sly* (Moretti et al., 2017). Par analyse protéomique, nous avons observé que le KO de *Dot1l* affecte non seulement la méthylation de H3K79 mais aussi d'autres PTMs des histones, en particulier l'hyperacétylation de H4, au moment de la transition histones-protamines. Ces anomalies expliquent les résultats de l'analyse phénotypique puisque l'hyperacétylation de H4 est connue pour être indispensable à la réorganisation et à la compaction de la chromatine des spermatozoïdes.

L'analyse phénotypique montre également un défaut plus précoce, au stade spermatogonies, caractérisé par des tubes séminifères vides de cellules germinales dans environ 7 % des cas (Blanco et al., 2023). Cette observation est en lien avec les résultats obtenus par une autre équipe, en analysant un modèle de KO conditionnel plus précoce que le nôtre (utilisant la *Ddx4-Cre* dont l'expression commence au stade embryonnaire E15, à la place de la *Stra8-Cre*) (Lin et al., 2022). L'équipe de J. Wang a en effet observé un nombre de spermatozoïde très faible voire nul pour les souris KO-*Dot1l* qui semble dû à un défaut de renouvellement des cellules souches. En effet, cette équipe a également réalisé une analyse RNA-seq sur des GSC traitées avec un inhibiteur de DOT1L et a observé la dérégulation d'une trentaine de gènes dont certains nécessaires au renouvellement des cellules souches (Lin et al., 2022).

Dans notre étude RNA-seq réalisée sur les spermatides KO, nous observons que la perte de DOT1L entraîne la dérégulation de plusieurs centaines de gènes qui pourraient expliquer le phénotype flagellaire et contribuer au phénotype chromatinien. En effet, dans les spermatogonies et dans les spermatocytes primaires, l'analyse d'ontologie met en évidence des voies de régulations associées à la formation du flagelle et à partir des spermatocytes, DOT1L est associée à des voies de régulations relatives à l'organisation de la chromatine et de la transcription. De manière surprenante, l'analyse d'ontologie montre également une dérégulation de gènes impliqués dans d'autres processus, comme le métabolisme, l'apoptose et la régulation de la transcription. Nous observons également que la perte de DOT1L induit une augmentation de l'expression de plusieurs centaines de gènes (1033 gènes dérégulés dont 789 sur-exprimés et 244 sous-exprimés), à partir du stade spermatocyte primaire. Ce phénomène sera développé plus loin dans la partie 3.4 de la discussion (cf. 3.4 *DOT1L réprime la transcription de plusieurs centaines de gènes*).

Il est intéressant de noter qu'en parallèle de la publication de notre article sur le rôle de DOT1L durant la spermiogenèse, deux autres articles ont également été publiés. Dans le premier, le KO conditionnel (dans la lignée germinale, grâce à *Ddx4-Cre*) du gène codant pour le partenaire de DOT1L, MMLT10, a été généré et montre une hypofertilité des mâles associée à des anomalies de nombre et de morphologie des spermatozoïdes (Lin et al., 2023). Par *western blot*, les auteurs observent une rétention des histones dans ce KO, et également dans le KO-*Dot1l Stra8-Cre*, en accord avec nos résultats. Par immunofluorescence, les auteurs observent que MMLT10, tout comme DOT1L, co-localise avec le *sex body* dans les stades méiotiques et le PMSC dans les stades post-méiotiques (Lin et al., 2023). Cette observation concorde avec les analyses précédentes (Ontoso et al., 2014; Moretti et al., 2017) et qui seront détaillées dans la partie 3.3 de cette discussion (cf. 3.3 *DOT1L est nécessaire à la réactivation des chromosomes sexuels dans les stades post-méiotiques*). À noter qu'aucune analyse transcriptomique ou par ChIP-seq n'a été réalisée dans l'étude de Lin et al. (Lin et al., 2023).

Le second article caractérise brièvement le phénotype du KO conditionnel *Dot1l* (via *Ddx4-Cre*) et réalise une analyse transcriptomique par scRNA-seq (Malla et al., 2023a). Leurs résultats montrent, comme les nôtres, un plus grand nombre de gènes dérégulés dans les spermatides que dans les stades précédents, et plus de gènes sur-exprimés que sous-exprimés (Blanco et al., 2023; Malla et al., 2023a). Ils observent également une diminution de l'expression de gènes impliqués dans la transition histones-protamines, comme *Brdt* et *Prm2*. Malla et al., proposent que DOT1L régule le remodelage de la chromatine des spermatides, en favorisant l'élongation de la transcription au niveau du gène *Brdt*, qui code pour l'enzyme reconnaissant l'acétylation des histones dans les spermatides (Gaucher et al., 2012; Malla et al., 2023a). Il est important de noter qu'ils ont observé une diminution du nombre de spermatides en fin de spermiogenèse dans les individus KO. Ainsi, leurs résultats de RT-qPCR montrant la diminution de l'expression de gènes abondants dans les spermatides allongées dans des analyses réalisées sur testicules entiers pose question.

Pour finir, Malla et al., observent une diminution de l'expression des microsatellites majeurs, qui pourrait être responsable des défauts des remplacements des protamines observés dans les spermatides (Malla et al., 2023a). De plus, une autre étude du groupe de B. Lesch montre que DOT1L dirige l'expression des satellites permettant ainsi la stabilisation de l'hétérochromatine dans les cellules embryonnaires chez le modèle murin (Malla et al., 2023b).

2 H3K79me2 est associée à une transcription active au niveau des gènes et des enhancers

Dans notre étude, nous avons cherché à caractériser la marque H3K79me2 dans plusieurs stades de la spermatogenèse, pour cela, nous avons généré des données pour les stades GSC, SCI, RS et ES (Fig. 30). Cependant, les données provenant des spermatides allongées ont été exclues des différentes analyses, car trop peu de pics sont communs entre les deux réplicats sur ce matériel. Chaque réplicat montre cependant une bonne corrélation avec les autres stades cellulaires, en particulier avec le stade spermatide ronde, ce qui est

cohérent. Ces observations sont explicables par le fait que ce matériel est assez hétérogène, car il inclut des spermatides allongées et des spermatides en cours de condensation. La marque H3K79me2 est donc moins accessible et les résultats de ChIP-seq, moins reproductibles, à ce stade. Pour avoir des résultats plus exploitables, il serait donc nécessaire d'augmenter le nombre de réplicats dans les spermatides allongées de façon à avoir une meilleure quantification de l'enrichissement de la marque.

Notre analyse de H3K79me2 dans les différentes cellules germinales mâles montre que la marque suit un enrichissement au niveau des gènes similaire à celui qui est décrit dans la littérature pour d'autres types cellulaires (Vlaming and van Leeuwen, 2016). De plus, on observe que l'enrichissement de la marque est positivement corrélé à l'expression génique. De façon intéressante, l'analyse de l'annotation génomique de H3K79me2 montre que, dans les spermatides rondes, une grande proportion des régions enrichies en H3K79me2 est annotée comme des régions distales intergéniques. Nous avons donc cherché à caractériser l'impact de la présence de H3K79me2 dans les *enhancers* sur l'expression génique. Contrairement aux autres analyses (Godfrey et al., 2019; Cattaneo et al., 2022) qui prennent en compte uniquement la présence de H3K27ac pour déterminer la présence d'*enhancer*, nous avons catégorisé les *enhancers* en quatre catégories définies par le modèle ChromHMM : bivalents, actifs, faibles et géniques (Fig. 31). Ces analyses ainsi que la nôtre montrent que la présence de la marque au niveau des *enhancers* corrèle avec un niveau d'expression génique élevé, quel que soit le type cellulaire (Godfrey et al., 2019; Cattaneo et al., 2022). De plus, l'association gène-*enhancer*, réalisée en utilisant la méthode que nous avons développée dans le package ChromENVEE (cf. 2.3.1 Association des gènes aux *enhancers*) a été comparée avec celle utilisée dans l'article de Godfrey et al. (Godfrey et al., 2019). Ces deux méthodes montrent des résultats similaires, permettant ainsi de conclure que la présence de H3K79me2 au niveau des *enhancers* est associée à une transcription active.

L'association gène-*enhancer* mise en place dans notre étude est une approximation et ne peut pas permettre d'évaluer avec exactitude la relation gène-*enhancer*. Bien que basée sur les propriétés génomiques décrites pour les *enhancers* comme la localisation ou la distance entre le gène et l'*enhancer*, ces informations ne remplaceront pas une analyse plus fine de la structure tridimensionnelle de la chromatine par Hi-C. De ce fait, et pour confirmer les relations gènes-*enhancers* estimées dans notre analyse, il sera nécessaire de réaliser des analyses Hi-C très résolutive à différents stades de la spermatogenèse (GSC, SCI et RS) dans le but de les comparer avec les résultats déjà obtenus par ChromENVEE. Cette approche permettra non seulement de valider l'association entre H3K79me2 localisée aux *enhancers* et l'expression génique, mais également d'affiner et de valider la méthode de prédiction de ChromENVEE.

3 DOT1L régule la transcription des gènes par plusieurs mécanismes

3.1 Analyse différentielle d'expression

Pour cette analyse, plusieurs modèles ont été mis en place. Le premier modèle, réalisé à partir de données de spermatoocytes primaires, secondaires et spermatoïdes ronds, a été initialement développé dans le cadre de l'analyse de la spermiogenèse (cf. 3.2 Article 1). Le second modèle a été réalisé sur les stades spermatogonies indifférenciées et différenciées. Ce modèle a été utilisé dans le cadre d'une collaboration avec l'équipe de Pierre Fouchet (Lapoujade et al., en préparation). Un troisième modèle a été développé afin d'intégrer l'ensemble des données générées aux cinq stades de la spermatogenèse. Cependant ce modèle ne semblait pas refléter correctement la biologie de nos stades cellulaires, car la différence de niveau d'expression des gènes dans les spermatogonies en comparaison aux stades méiotiques et post-méiotiques était trop importante, entraînant ainsi un biais lors de l'analyse de la dynamique. C'est pour ces raisons que les analyses différentielles « classiques » ont été réalisées en comparant les données obtenues sur les deux premiers modèles (cf. 4.2 Article 2). Cependant, cette approche est loin d'être parfaite pour comparer tous les stades, j'ai donc utilisé la normalisation via *spike-in* pour produire un modèle contenant les 5 stades de la spermatogenèse. Les résultats obtenus par les deux méthodes sont comparables avec une majorité de gènes dérégulés en commun dans les deux analyses (cf. 4.2 Article 2).

Lors de la mise en place de ces modèles, nous avons estimé le seuil au-delà duquel un gène est considéré comme exprimé, c'est-à-dire $CPM \geq 1$ dans au moins 2 échantillons. Ce seuil est celui qui est couramment utilisé pour ce type d'analyse dans la littérature bien que ce seuil soit arbitraire. De plus, ce seuil ne permet pas de prendre en compte les gènes faiblement exprimés comme cela peut être le cas pour les gènes multi-copies de l'Y qui sont cependant essentiels à la spermiogenèse. Dans notre analyse, nous avons donc également réalisé une analyse avec comme seuil $CPM \geq 0$ ainsi, nous prenons en compte l'ensemble des gènes exprimés, permettant de garder les gènes multi-copies de l'Y. Cependant, il serait pertinent de mettre en place une méthode permettant d'estimer au mieux le seuil d'expression des gènes comme utilisé dans l'article de Moretti et al. (Moretti et al., 2016).

Autre point intéressant, il est décrit depuis 2018 que DOT1L joue un rôle dans l'épissage alternatif (Li et al., 2018). Cependant l'approche de RNA-seq que nous avons mis en place ne permet pas de déterminer avec exactitude les événements d'épissage alternatif, il serait donc pertinent de mettre en place un nouveau pipeline d'analyse à partir de nos données de RNA-seq pour savoir si DOT1L et H3K79me2 ont un impact sur l'épissage alternatif dans le cadre de la spermatogenèse. Ces analyses font, actuellement, l'objet d'un nouveau projet de thèse co-dirigé par Julie Cocquet.

3.2 DOT1L active la transcription de plusieurs centaines de gènes via la présence de H3K79me2

Comme nous l'avons décrit précédemment et en accord avec les observations présentes dans la littérature (Vlaming and van Leeuwen, 2016), H3K79me2 est associée à

une transcription active au niveau du corps du gène et/ou des *enhancers*, et DOT1L interagit avec des protéines impliquées dans l'élongation de la transcription. Il était donc attendu que le KO de *Dot1l* entraînerait une diminution de l'expression de nombreux gènes, mais lors de l'initiation de mon projet, aucune étude ne s'était intéressée à caractériser, en détails, l'impact du KO de *Dot1l* sur l'expression génique dans un contexte de différenciation cellulaire (i.e. seules des études réalisées sur des cellules cancéreuses MLL avaient été publiées) (Chen et al., 2015a). Notre étude a montré que le KO entraîne un plus grand nombre de gènes sur-exprimés que de gènes sous-exprimés à tous les stades de la spermatogenèse que nous avons étudié (Article 2, Fig. 2). Ces résultats, bien que inédits, ont été confirmés et retrouvés depuis, dans le KO de *Dot1l* pour d'autres types cellulaires. En effet, depuis 2019, de nombreuses études de RNA-seq ont montré une dérégulation génique induite par le KO de *Dot1l* dans divers types de cellules somatiques, allant dans le même sens que ce que nous avons trouvé (Fig. 36) (Franz et al., 2019; Kwesi-Maliepaard et al., 2020; Aslam et al., 2021; Kim et al., 2021; Borosha et al., 2022; Cattaneo et al., 2022; Feng et al., 2022; Willemsen et al., 2022; Appiah et al., 2023). Dans le but de confirmer nos résultats et que cette régulation est le reflet biologique du KO de *Dot1l* et non due à un potentiel biais de normalisation, nous avons décidé de réaliser une analyse de RNA-seq *spike-in* dans les différents stades de la spermatogenèse (Fig. 30). Cette approche nous a permis de confirmer la sur-expression observée dans le KO.

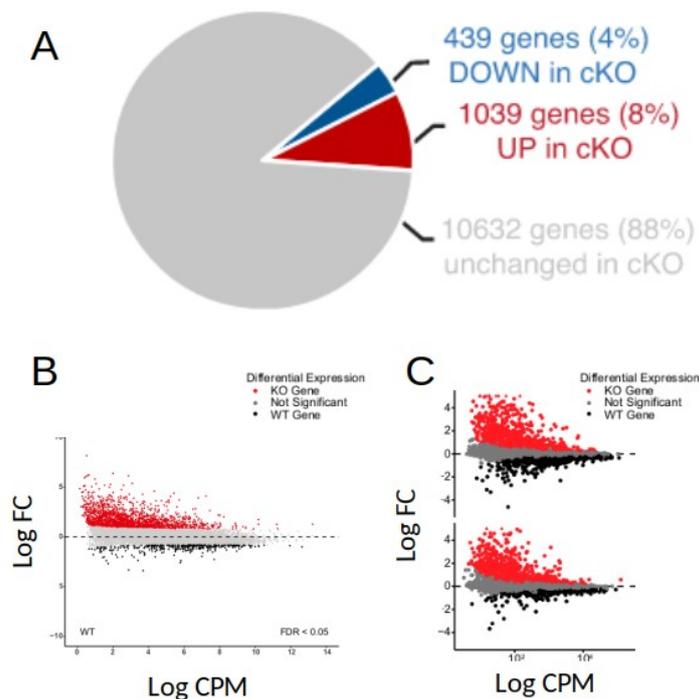


Figure 36 : Expression différentielle des gènes entre des individus contrôles et KO-Dot1l dans plusieurs types de cellules somatiques. En rouge sont représentés les gènes sur-exprimés et en noir/bleu les gènes sous-exprimés dans les KO-Dot1l par rapport aux contrôles (FDR < 5 %). A. Pie-chart provenant de Cattaneo et al, 2022, sur des cardiomyocytes. B, C. MD-plots provenant de Kwesie-Maliepaard et al, 2020 (B) et Aslam et al, 2021 (C) dans des cellules T et B respectivement.

L'analyse de la dynamique des gènes dérégulés montre un programme transcriptionnel différent avant et après méiose. En effet, la majorité des gènes dérégulés

dans les spermatogonies ne sont pas retrouvés dans les stades méiotiques et post-méiotiques et inversement (Article 2, Fig 3c). L'intégration de nos données de CHIP-seq et transcriptomiques montre que les gènes sous-exprimés dans le KO sont plus associés à la présence de la marque au niveau du corps du gène que les gènes sur-exprimés. La même observation a été faite dans d'autres types cellulaires (Franz et al., 2019; Godfrey et al., 2019; Kwesi-Maliepaard et al., 2020; Aslam et al., 2021; Cattaneo et al., 2022), confirmant le lien entre H3K79me2/3 et transcription active. En utilisant les *enhancers* définis par notre modèle (Fig. 31), nous observons que les gènes sous-exprimés sont en fait caractérisés par la co-présence de H3K79me2 au niveau du corps du gène et des *enhancers*, corrélant ainsi avec l'observation de Cattaneo et al., durant la cardiogenèse (Cattaneo et al., 2022). Ces résultats appuient le rôle d'activateur de la transcription de H3K79me2 au niveau des *enhancers*. De leur côté, Ferrari et al., supposent que l'inhibition de DOT1L pourrait affecter seulement une sous-partie des *enhancers* indépendamment de la présence de H3K79me2 (Ferrari et al., 2020). En effet, l'utilisation d'un inhibiteur de DOT1L entraîne la diminution de H3K27ac, mais celle-ci ne suit pas la diminution globale de H3K79me2.

Un article récent montre que DOT1L semble favoriser l'élongation de la transcription des microsatellites dans les régions péri-centromériques et en particulier pour les régions d'hétérochromatine (Malla et al., 2023b). L'observation faite dans cet article, initialement réalisée sur des cellules embryonnaires, a également été retrouvée par RT-PCR sur des extraits de testicule entier KO-*Dot1l* (Malla et al., 2023a). Il serait donc intéressant de confirmer dans nos échantillons KO de cellules germinales à différents stades si la même dérégulation des microsatellites est retrouvée et, si oui, à quel stade.

Nous avons également cherché à catégoriser ces gènes en fonction de leur type, de leur taille et de leur localisation chromosomique. Ainsi on observe, qu'à partir de la méiose, les gènes sur-exprimés sont significativement plus longs tandis que les gènes sous-exprimés dans les Kit+ sont plus petits. Des études précédentes montrent que les gènes plus longs sont plus sensibles aux événements d'épissage alternatif (Grishkevich and Yanai, 2014; Lopes et al., 2021). Ainsi il serait intéressant de savoir si les gènes sur-exprimés dans notre étude sont plus associés à des événements d'épissage alternatif. L'annotation fonctionnelle des gènes dérégulés est sensiblement similaire dans tous les types cellulaires, avec une majorité de gènes annotée comme des protéines codantes (« *protein coding* ») concordant avec le fait que notre analyse de RNA-seq est fait sur de l'ARNm. Nous avons cherché à vérifier la localisation chromosomique des gènes dérégulés dans notre étude ; cette analyse montre que les gènes dérégulés présentent une répartition aléatoire le long du chromosome et ne sont pas localisés autour des centromères et télomères.

3.3 DOT1L est nécessaire à la réactivation des chromosomes sexuels après la méiose

De façon intéressante, nous observons que les gènes sous-exprimés enrichis en H3K79me2 dans les spermatocytes secondaires et les spermatides rondes sont en majorité des gènes portés par les chromosomes sexuels. Ces résultats concordent avec des observations précédemment réalisées par immunofluorescence par Ontoso et al., et par

notre équipe, montrant une co-localisation de DOT1L avec la chromatine sexuelle au stade diplotène de méiose I (Ontoso et al., 2014) et dans les spermatocytes secondaires et spermatides rondes (Moretti et al., 2017) – résultats également observés par la suite par l'équipe de J. Wang (Lin et al., 2023). Ces résultats suggèrent que DOT1L pourrait permettre la réactivation des chromosomes sexuels après la prophase I de la méiose. La distribution de la marque H3K79me2 durant les différents stades de la spermatogenèse va également dans ce sens. En effet, on observe que la présence de la marque H3K79me2 dans les chromosomes sexuels est augmentée dans les stades spermatides par rapport aux stades spermatocytes et en particulier dans le chromosome X. La dynamique de la marque H3K79me2 dans les chromosomes a été comparée avec celle d'autres marques d'histones connues pour être différentiellement enrichies dans les chromosomes sexuels en comparaison des autosomes : H3K4me1, H3K4me3, H3K27ac, H3K27me3 et H3K9me3 (Article 2, Fig. S4c et S5).

Plusieurs études précédentes s'étaient déjà penchées sur la question ; cependant, aucune ne regroupait l'ensemble de ces marques sur les trois phases de la spermatogenèse : spermatogonies, spermatocytes et spermatides et le chromosome Y n'était généralement pas étudié (Sin et al., 2015; Moretti et al., 2016; Liu et al., 2019; Chen et al., 2020b). Pour pouvoir étudier cette dynamique et par souci d'homogénéité, j'ai récupéré des données de la littérature (Hammoud et al., 2014; Bryant et al., 2015; Zuo et al., 2018; Liu et al., 2019; Chen et al., 2020b) que j'ai analysé suivant le même protocole que pour l'analyse de H3K79me2. Ainsi nos résultats montrent que la marque répressive H3K9me3 est très enrichie dans les chromosomes sexuels à partir du stade spermatogonie jusqu'au stade post-méiotique, corrélant avec les observations des précédentes études (Ernst et al., 2019; Liu et al., 2019). En particulier, on observe une augmentation de l'enrichissement de la marque H3K9me3 associée à l'X dans les spermatides rondes par rapport aux spermatocytes (Article 2, Fig. S4c et S5). Ces observations corrént avec les résultats d'autres études qui montrent une augmentation de la marque H3K9me3 dans les chromosomes sexuels, en particulier de l'Y, dans les spermatides par rapport au stade précédent (Liu et al., 2019) et un fort enrichissement de H3K9me3 dans l'Y par rapport aux autres autosomes, dans les spermatides (Moretti et al., 2016). Cependant, contrairement à l'article de Moretti et al., qui utilise les données de Bryant et al., nous observons que H3K9me3 est fortement enrichie dans l'X également (Bryant et al., 2015; Moretti et al., 2016). De façon similaire à H3K9me3, on constate que la marque H3K27ac est enrichie dans les chromosomes sexuels à partir des stades méiotiques de la spermatogenèse et plus particulièrement pour l'Y dans les spermatides rondes. Cette observation corréle avec les résultats de Moretti et al., dans les spermatides (Moretti et al., 2016). Par opposition, nous observons que la présence de la marque H3K27me3 dans les chromosomes sexuels diminue au cours de la spermatogenèse, corrélant avec la perte de H3K27me3 dans l'X au moment du MSC1 (Sin et al., 2015).

Ainsi, notre analyse nous permet de confirmer et de compléter les observations de la dynamique des marques modifiées durant le MSC1 et ce, en partant des spermatogonies indifférenciées (GSC). Cette analyse est également complétée par H3K79me2 et montre que

cette marque est l'une des plus dynamiques sur les chromosomes sexuels lors de la spermatogenèse.

3.4 DOT1L réprime la transcription de plusieurs centaines de gènes

Les analyses RNA-seq montrent que le KO de *Dot1l* entraîne une sur-expression de nombreux gènes (Fig. 36) qui ne semble pas être limitée à un type cellulaire. En effet, cette sur-expression est retrouvée dans diverses cellules somatiques tout comme à différents stades de la spermatogenèse (Franz et al., 2019; Ferrari et al., 2020; Kwesi-Maliepaard et al., 2020; Aslam et al., 2021; Borosha et al., 2022; Cattaneo et al., 2022; Feng et al., 2022; Blanco et al., 2023). Il est intéressant de noter que les gènes sur-exprimés ne sont pas ou très peu enrichis en H3K79me2 au niveau du corps du gène à partir du stade méiotique, corrélant avec l'étude de Cattaneo et al., qui décrivent que les gènes sur-exprimés dans le KO de *Dot1l* sont majoritairement associés à des *enhancers* ayant H3K79me2 (sans avoir H3K79me2 au corps des gènes) (Cattaneo et al., 2022). Ce qui suggère que les gènes sur-exprimés lors du KO de *Dot1l* sont régulés par d'autres mécanismes épigénétiques.

Le mécanisme responsable de cette sur-expression n'est pas encore clairement établi bien que certaines pistes ont été proposées. Pour répondre à cette question, j'ai cherché à déterminer l'environnement chromatinien des gènes dérégulés et ainsi estimer l'état chromatinien majoritaire en utilisant la méthode développée dans ChromENVEE (cf. 2.3.2 *Caractérisation de l'environnement chromatinien*). Cette analyse a montré que les gènes sur-exprimés sont majoritairement associés à des états bivalents et répresseurs, tous les deux caractérisés par la présence de H3K27me3. Dans leur article, Ferrari et al., ont également utilisé ChromHMM pour définir des états chromatiniens et ont montré que les gènes sur-exprimés sont caractérisés par la co-présence des marques H3K27me3 et H3K4me3, caractéristiques des états chromatiniens bivalents (Ferrari et al., 2020). Ces observations corrélaient avec nos résultats et ceux d'autres études qui montrent que les gènes sur-exprimés sont plutôt associés à la présence de la marque H3K27me3 au niveau du promoteur des gènes (Ferrari et al., 2020; Kwesi-Maliepaard et al., 2020; Aslam et al., 2021; Appiah et al., 2023). Ces informations suggèrent un lien entre DOT1L et H3K27me3 qui avait déjà été observé : des analyses CHIP-seq dans les cellules MLL-AF9 avaient en effet montré que l'inhibition de DOT1L entraîne une augmentation de l'enrichissement de H3K27me3 (Chen et al., 2015a).

De leur côté, Kwesi-Maliepaard et al., montrent que, dans les cellules T CD8+, DOT1L régule positivement l'expression de gènes répresseurs, en particulier *Ezh2*, qui code pour la triméthyltransférase de H3K27 (Kwesi-Maliepaard et al., 2020). Cependant, cet effet n'est pas retrouvé dans les autres articles ni dans notre étude de la spermatogenèse, ce qui suppose que le mécanisme par lequel DOT1L réprime la transcription n'est pas le même dans tous les types cellulaires (Aslam et al., 2021; Appiah et al., 2023). Dans notre étude des gènes cibles de DOT1L, nous n'avons pas trouvé *Ezh2* ou d'autres gènes codant pour des protéines du complexe PRC2, cependant, nous avons identifié d'autres répresseurs de la transcription, comme *Hdac8*, *Zkscan17*, *Bcl6* ou *Bcor* (co-répresseur de BCL6) dans les stades méiotiques et post-méiotiques de la spermatogenèse (Article 2, Fig. 4e). *Bcl6* pourrait être

un candidat intéressant, car il intervient dans la régulation de plusieurs fonctions biologiques comme la différenciation, l'apoptose et le contrôle du cycle cellulaire. De plus, une étude de 2001 a montré que la perte de *Bcl6* chez la souris entraînerait l'apoptose des cellules germinales mâles et notamment lors de la phase de méiose, suggérant que la présence de *Bcl6* permettrait de protéger les spermatocytes lors de la méiose de l'apoptose (Kojima et al., 2001). Ces résultats sont en accord avec les nôtres, puisque le KO de *Dot1l* entraîne une augmentation de l'apoptose après la méiose. De façon intéressante, l'analyse avec HOMER (Heinz et al., 2010) des gènes sur-exprimés montre que certains de ces gènes sont enrichis pour le motif BCL6B (Article 1, Fig. S5a). *Zkscan17*, également connu sous le nom de *Nizp1*, pourrait être également un gène d'intérêt, car il est très exprimé durant la spermatogenèse et en particulier à partir des stades méiotiques.

D'un autre côté, *Hdac8* est un gène codant pour une histone désacétylase de H3K27 (Chen et al., 2019b; Wang et al., 2022) et a été décrit comme réprimant l'expression de gènes grâce à sa capacité à désacétyler les *enhancers* (Chen et al., 2019b). Cette observation suggère que la présence de DOT1L empêcherait l'expression prématurée des gènes importants pour la différenciation spermatique en empêchant l'acétylation de H3K27 et en favorisant sa triméthylation au niveau des *enhancers*. Parmi les gènes activés par DOT1L, on retrouve également *Kdm6a*, une des méthyltransférases de H3K27, dans les spermatocytes secondaires et les spermatides rondes suggérant une interaction complexe entre H3K27ac et H3K27me3. Ces résultats suggèrent que l'interaction entre H3K27ac et H3K27me3 serait nécessaire pour la régulation de l'expression des gènes régulés par les *enhancers* et les *super-enhancers* (Maezawa et al., 2020).

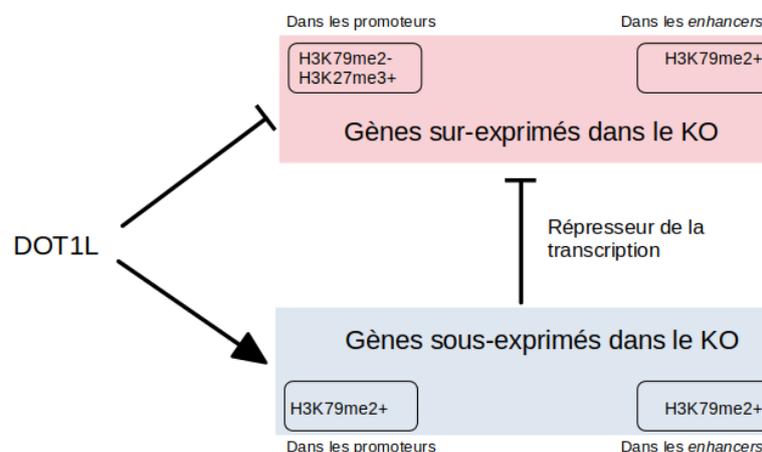


Figure 37 : Modèle expliquant les effets de DOT1L sur la régulation de l'expression génique au cours de la spermatogenèse

Prises ensemble, ces informations suggèrent que DOT1L réprime l'expression des gènes via plusieurs mécanismes indépendants de la présence de H3K79me2 au niveau du corps du gène. Cette observation corrèle avec l'hypothèse émise par Chen et al., qui supposait que DOT1L pourrait réprimer l'expression génique via plusieurs mécanismes

(Chen et al., 2015b). Dans notre étude, nous observons que la régulation des gènes par DOT1L se fait *via* l'environnement chromatinien de ces gènes, et notamment par la présence de H3K27me3 et indirectement en régulant l'expression de répresseurs de la transcription (Fig. 37).

Très récemment, un article a montré que DOT1L réprime les rétrotransposons, en particulier les rétrotransposons MERVL (*Murine Endogenous Retrovirus-L*), dans les cellules souches embryonnaires (Zhao et al., 2023). Les rétrotransposons peuvent contribuer à la régulation de l'expression de gènes « classiques » comme cela a été démontré dans les cellules germinales (Sakashita et al., 2020). Il serait donc très pertinent de regarder si les gènes que nous trouvons sur-exprimés dans les cellules KO sont associés à des MERVL. Il serait également intéressant de savoir si ces gènes sur-exprimés sont associés à des *enhancers* « *poised* » ayant H3K27ac. En effet, il a été montré que les *enhancers* « *poised* » peuvent être régulés par des rétrovirus endogènes (Sakashita et al., 2020).

4 H3K79me2 et développement embryonnaire précoce

Dans une étude précédente, l'équipe avait observé par *western blot* et immunofluorescence que H3K79me2 est une marque retenue dans les spermatozoïdes (Moretti et al., 2017). Ainsi, durant ma thèse, l'équipe a essayé de générer des données de ChIP-seq pour H3K79me2 dans les spermatozoïdes, afin d'étudier le potentiel impact intergénérationnel de la marque. Cependant de par l'extrême compaction de la chromatine, nous n'avons pas réussi à obtenir des données fiables. Pour surmonter cette difficulté nous avons étudié la possible transmission de la marque des spermatides à l'embryon, en utilisant des données récemment publiées sur l'accessibilité de la chromatine (ATAC-seq) dans des embryons issus de ROSI (issus de la fécondation par des spermatides rondes) et des embryons issus d'ICSI (issus de la fécondation par des spermatozoïdes par Injection Intracytoplasmique) (Sakamoto et al., 2022).

Pour ceci, nous avons comparé ces données d'ATAC-seq avec nos données de ChIP-seq pour H3K79me2 dans les spermatides rondes. Comme ces deux types de fécondation ont été réalisés avec des souris du même fond génétique, les deux types d'embryons partagent la même organisation génomique maternelle. Nous avons donc considéré que les régions communes entre les données provenant d'ICSI et ROSI sont d'origine maternelle, et nous sommes concentrés sur les régions différentes. Nous avons observé qu'une grande proportion des régions accessibles spécifiquement dans les embryons provenant de ROSI (au stade zygote et au stade 2 cellules) corrèlent avec la présence de la marque H3K79me2 dans les spermatides. Ces analyses préliminaires restent à poursuivre afin de déterminer si H3K79me2 transmis par les gamètes mâles pourrait jouer un rôle dans le développement embryonnaire précoce.

Pour compléter ces résultats, il serait intéressant d'avoir des informations sur H3K79me2 dans les spermatozoïdes pour comparer avec d'autres données de chromatine et transcriptomique issues d'embryons, et ainsi déterminer si H3K79me2 peut avoir un effet intergénérationnel.

Conclusion

Mon projet de thèse a permis de définir le rôle de H3K79me2 et de sa méthyltransférase DOT1L durant la spermatogenèse via des approches bioinformatiques en utilisant des données de transcriptomique et de ChIP-seq.

D'une part, ces approches nous ont permis de confirmer le lien entre transcription et H3K79me2 dans les stades de la spermatogenèse. En effet, la marque est corrélée positivement à l'expression génique lorsque celle-ci est située au niveau des gènes et des *enhancers*. Ces observations nous ont également permis de démontrer que DOT1L active l'expression de plusieurs gènes durant la spermatogenèse et en particulier à partir de la méiose caractérisée par la présence de H3K79me2 au niveau des gènes et des *enhancers*. Il est intéressant de noter que DOT1L active en particulier les gènes portés par les chromosomes sexuels après le processus de MSCI. Ces résultats montrent que DOT1L et H3K79me2 sont indispensables à la régulation des gènes sur l'ensemble de la spermatogenèse.

D'un autre côté, ce projet nous a permis de définir un nouveau rôle pour DOT1L jusqu'alors décrit comme un facteur d'élongation de la transcription. En effet, dans ce projet, nous observons que DOT1L réprime l'expression de plusieurs centaines de gènes *via* un mécanisme indépendant de la présence de H3K79me2 au niveau des gènes mais qui pourrait être lié en partie à la présence de H3K79me2 au niveau des *enhancers*. Nos résultats montrent que la répression des gènes par DOT1L est dirigée par un effet chromatinien caractérisé par la présence de H3K27me3 au niveau de l'environnement du gène mais également par la régulation de répresseur de la transcription. Notre étude suggère que la présence de DOT1L empêcherait l'acétylation de H3K27 et favoriserait sa triméthylation au niveau des *enhancers*.

Enfin, les différentes analyses réalisées dans ce projet m'ont permis le développement d'un package R, ChromENVEE permettant l'association des gènes aux *enhancers* et la caractérisation de l'environnement chromatinien d'un gène.

Bibliographie

- 1000 Genomes Project Consortium, Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. doi: 10.1038/nature09534.
- Abe, H., Yeh, Y.-H., Munakata, Y., Ishiguro, K.-I., Andreassen, P. R., and Namekawa, S. H. (2022). Active DNA damage response signaling initiates and maintains meiotic sex chromosome inactivation. *Nat Commun* 13, 7212. doi: 10.1038/s41467-022-34295-5.
- Abe, M., Tsai, S. Y., Jin, S.-G., Pfeifer, G. P., and Szabó, P. E. (2011). Sex-Specific Dynamics of Global Chromatin Changes in Fetal Mouse Germ Cells. *PLOS ONE* 6, e23848. doi: 10.1371/journal.pone.0023848.
- Al Adhami, H., Bardet, A. F., Dumas, M., Cleroux, E., Guibert, S., Fauque, P., et al. (2022). A comparative methylome analysis reveals conservation and divergence of DNA methylation patterns and functions in vertebrates. *BMC Biol* 20, 70. doi: 10.1186/s12915-022-01270-x.
- Alavattam, K. G., Maezawa, S., Sakashita, A., Khoury, H., Barski, A., Kaplan, N., et al. (2019). Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. *Nat Struct Mol Biol* 26, 175–184. doi: 10.1038/s41594-019-0189-y.
- Álvarez-González, L., Burden, F., Doddamani, D., Malinverni, R., Leach, E., Marín-García, C., et al. (2022). 3D chromatin remodelling in the germ line modulates genome evolutionary plasticity. *Nat Commun* 13, 2608. doi: 10.1038/s41467-022-30296-6.
- Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* 16, 47–57. doi: 10.1016/j.devcel.2008.11.011.
- An, J., Zhang, X., Qin, J., Wan, Y., Hu, Y., Liu, T., et al. (2014). The histone methyltransferase ESET is required for the survival of spermatogonial stem/progenitor cells in mice. *Cell Death Dis* 5, e1196–e1196. doi: 10.1038/cddis.2014.171.
- Anderson, C. J., Baird, M. R., Hsu, A., Barbour, E. H., Koyama, Y., Borgnia, M. J., et al. (2019). Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase. *Cell Rep* 26, 1681–1690.e5. doi: 10.1016/j.celrep.2019.01.058.
- Appiah, B., Fullio, C. L., Ossola, C., Bertani, I., Restelli, E., Cheffer, A., et al. (2023). DOT1L activity affects neural stem cell division mode and reduces differentiation and ASNS expression. *EMBO reports* 24, e56233. doi: 10.15252/embr.202256233.
- Aslam, M. A., Alemdehy, M. F., Kwesi-Maliepaard, E. M., Muhaimin, F. I., Caganova, M., Pardieck, I. N., et al. (2021). Histone methyltransferase DOT1L controls state-specific identity during B cell differentiation. *EMBO Rep* 22, e51184. doi: 10.15252/embr.202051184.
- Bao, J., and Bedford, M. T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction* 151, R55–R70. doi: 10.1530/REP-15-0562.
- Barbero, G., de Sousa Serro, M. G., Perez Lujan, C., Vitullo, A. D., González, C. R., and González, B. (2023). Transcriptome profiling of histone writers/erasers enzymes across spermatogenesis, mature sperm and pre-cleavage embryo: Implications in paternal epigenome transitions and inheritance mechanisms. *Front Cell Dev Biol* 11, 1086573. doi: 10.3389/fcell.2023.1086573.
- Barral, S., Morozumi, Y., Tanaka, H., Montellier, E., Govin, J., Dieuleveult, M. de, et al. (2017). Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Molecular Cell* 66, 89–101.e8. doi: 10.1016/j.molcel.2017.02.025.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., et al. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129, 823–837. doi: 10.1016/j.cell.2007.05.009.
- Bastos, H., Lassalle, B., Chicheportiche, A., Riou, L., Testart, J., Allemand, I., et al. (2005). Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry A* 65, 40–49. doi: 10.1002/cyto.a.20129.
- Bedi, Y. S., Roach, A. N., Thomas, K. N., Mehta, N. A., and Golding, M. C. (2022). Chromatin alterations during the epididymal maturation of mouse sperm refine the paternally inherited epigenome. *Epigenetics Chromatin* 15, 2. doi: 10.1186/s13072-021-00433-4.

- Bell, E. L., Nagamori, I., Williams, E. O., Del Rosario, A. M., Bryson, B. D., Watson, N., et al. (2014). SirT1 is required in the male germ cell for differentiation and fecundity in mice. *Development* 141, 3495–3504. doi: 10.1242/dev.110627.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289–300.
- Berman, B. P., Weisenberger, D. J., Aman, J. F., Hinoue, T., Ramjan, Z., Liu, Y., et al. (2011). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* 44, 40–46. doi: 10.1038/ng.969.
- Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., et al. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci U S A* 99, 8695–8700. doi: 10.1073/pnas.082249499.
- Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169–181. doi: 10.1016/j.cell.2005.01.001.
- Bernt, K. M., Zhu, N., Sinha, A. U., Vempati, S., Faber, J., Krivtsov, A. V., et al. (2011). MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. *Cancer Cell* 20, 66–78. doi: 10.1016/j.ccr.2011.06.010.
- Biswas, D., Milne, T. A., Basrur, V., Kim, J., Elenitoba-Johnson, K. S. J., Allis, C. D., et al. (2011). Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proceedings of the National Academy of Sciences* 108, 15751–15756. doi: 10.1073/pnas.1111498108.
- Bitoun, E., Oliver, P. L., and Davies, K. E. (2007). The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Human Molecular Genetics* 16, 92–106. doi: 10.1093/hmg/ddl444.
- Black, J. C., Van Rechem, C., and Whetstone, J. R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* 48, 491–507. doi: 10.1016/j.molcel.2012.11.006.
- Blanco, M., and Cocquet, J. (2019). Genetic Factors Affecting Sperm Chromatin Structure. *Adv Exp Med Biol* 1166, 1–28. doi: 10.1007/978-3-030-21664-1_1.
- Blanco, M., El Khattabi, L., Gobé, C., Crespo, M., Coulée, M., de la Iglesia, A., et al. (2023). DOT1L regulates chromatin reorganization and gene expression during sperm differentiation. *EMBO Rep*, e56316. doi: 10.15252/embr.202256316.
- Blythe, M. J., Kocer, A., Rubio-Roldan, A., Giles, T., Abakir, A., Ialy-Radio, C., et al. (2021). LINE-1 transcription in round spermatids is associated with accretion of 5-carboxylcytosine in their open reading frames. *Commun Biol* 4, 691. doi: 10.1038/s42003-021-02217-8.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170.
- Boltsis, I., Grosveld, F., Giraud, G., and Kolovos, P. (2021). Chromatin Conformation in Development and Disease. *Frontiers in Cell and Developmental Biology* 9. Available at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.723859> [Accessed July 19, 2023].
- Borosha, S., Ratri, A., Ghosh, S., Malcom, C. A., Chakravarthi, V. P., Vivian, J. L., et al. (2022). DOT1L Mediated Gene Repression in Extensively Self-Renewing Erythroblasts. *Frontiers in Genetics* 13. Available at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.828086> [Accessed December 7, 2022].
- Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., et al. (1996). Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. *Cell* 84, 843–851. doi: 10.1016/S0092-8674(00)81063-6.
- Bryant, J. M., Donahue, G., Wang, X., Meyer-Ficca, M., Luense, L. J., Weller, A. H., et al. (2015). Characterization of BRD4 during mammalian postmeiotic sperm development. *Mol. Cell Biol.* 35, 1433–1448. doi: 10.1128/MCB.01328-14.
- Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E. J., Roloff, T. C., et al. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* 17, 679–687. doi: 10.1038/nsmb.1821.
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-

- binding proteins and nucleosome position. *Nat Methods* 10, 1213–1218. doi: 10.1038/nmeth.2688.
- Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., et al. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490. doi: 10.1038/nature14590.
- Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. in.
- Caldeira-Brant, A. L., Martinelli, L. M., Marques, M. M., Reis, A. B., Martello, R., Almeida, F. R. C. L., et al. (2020). A subpopulation of human Adark spermatogonia behaves as the reserve stem cell. *Reproduction* 159, 437–451. doi: 10.1530/REP-19-0254.
- Cao, K., Ugarenko, M., Ozark, P. A., Wang, J., Marshall, S. A., Rendleman, E. J., et al. (2020). DOT1L-controlled cell-fate determination and transcription elongation are independent of H3K79 methylation. *Proc Natl Acad Sci U S A* 117, 27365–27373. doi: 10.1073/pnas.2001075117.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., et al. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* 298, 1039–1043. doi: 10.1126/science.1076997.
- Cao, Y., Liu, S., Cui, K., Tang, Q., and Zhao, K. (2023). Hi-TrAC detects active sub-TADs and reveals internal organizations of super-enhancers. *Nucleic Acids Research* 51, 6172–6189. doi: 10.1093/nar/gkad378.
- Carone, B. R., Hung, J.-H., Hainer, S. J., Chou, M.-T., Carone, D. M., Weng, Z., et al. (2014). High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell* 30, 11–22. doi: 10.1016/j.devcel.2014.05.024.
- Cattaneo, P., Hayes, M. G. B., Baumgarten, N., Hecker, D., Peruzzo, S., Aslan, G. S., et al. (2022). DOT1L regulates chamber-specific transcriptional networks during cardiogenesis and mediates postnatal cell cycle withdrawal. *Nat Commun* 13, 7444. doi: 10.1038/s41467-022-35070-2.
- Cavé, H., Acquaviva, C., Bièche, I., Brault, D., de Fraipont, F., Fina, F., et al. (2003). [RT-PCR in clinical diagnosis]. *Ann Biol Clin (Paris)* 61, 635–644.
- Champroux, A., Cocquet, J., Henry-Berger, J., Drevet, J. R., and Kocer, A. (2018). A Decade of Exploring the Mammalian Sperm Epigenome: Paternal Epigenetic and Transgenerational Inheritance. *Front. Cell Dev. Biol.* 6. doi: 10.3389/fcell.2018.00050.
- Chen, C.-W., Koche, R. P., Sinha, A. U., Deshpande, A. J., Zhu, N., Eng, R., et al. (2015a). DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia. *Nat Med* 21, 335–343. doi: 10.1038/nm.3832.
- Chen, H., Lareau, C., Andreani, T., Vinyard, M. E., Garcia, S. P., Clement, K., et al. (2019a). Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biology* 20, 241. doi: 10.1186/s13059-019-1854-5.
- Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., and Gregor, T. (2018a). Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet* 50, 1296–1303. doi: 10.1038/s41588-018-0175-z.
- Chen, L.-F., Lin, Y. T., Gallegos, D. A., Hazlett, M. F., Gómez-Schiavon, M., Yang, M. G., et al. (2019b). Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. *Cell Rep* 26, 1174–1188.e5. doi: 10.1016/j.celrep.2019.01.032.
- Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., et al. (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* 351, 397–400. doi: 10.1126/science.aad7977.
- Chen, S., Yang, Z., Wilkinson, A. W., Deshpande, A. J., Sidoli, S., Krajewski, K., et al. (2015b). The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79. *Mol Cell* 60, 319–327. doi: 10.1016/j.molcel.2015.08.019.
- Chen, X., Zheng, Y., Lei, A., Zhang, H., Niu, H., Li, X., et al. (2020a). Early cleavage of preimplantation embryos is regulated by tRNAGln-TTG-derived small RNAs present in mature spermatozoa. *J Biol Chem* 295, 10885–10900. doi: 10.1074/jbc.RA120.013003.
- Chen, Y., Lyu, R., Rong, B., Zheng, Y., Lin, Z., Dai, R., et al. (2020b). Refined spatial temporal epigenomic profiling reveals intrinsic connection between PRDM9-mediated H3K4me3 and the fate of double-stranded breaks. *Cell Res* 30, 256–268. doi: 10.1038/s41422-020-0281-1.

- Chen, Y., Zheng, Y., Gao, Y., Lin, Z., Yang, S., Wang, T., et al. (2018b). Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res* 28, 879–896. doi: 10.1038/s41422-018-0074-y.
- Cheng, E.-C., Hsieh, C.-L., Liu, N., Wang, J., Zhong, M., Chen, T., et al. (2021). The Essential Function of SETDB1 in Homologous Chromosome Pairing and Synapsis during Meiosis. *Cell Reports* 34. doi: 10.1016/j.celrep.2020.108575.
- Clermont, Y. (1966). Renewal of spermatogonia in man. *Am J Anat* 118, 509–524. doi: 10.1002/aja.1001180211.
- Cobb, J., and Handel, M. A. (1998). Dynamics of meiotic prophase I during spermatogenesis: from pairing to division. *Seminars in Cell & Developmental Biology* 9, 445–450. doi: 10.1006/scdb.1998.0202.
- Cocquet, J., Ellis, P., Yamauchi, Y., Mahadevaiah, S., Affara, N., Ward, M., et al. (2009). The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. *PLoS biology* 7, e1000244. doi: 10.1371/journal.pbio.1000244.
- Crespo, M., Damont, A., Blanco, M., Lastrucci, E., Kennani, S. E., Ialy-Radio, C., et al. (2020). Multi-omic analysis of gametogenesis reveals a novel signature at the promoters and distal enhancers of active genes. *Nucleic Acids Res* 48, 4115–4138. doi: 10.1093/nar/gkaa163.
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 107, 21931–21936. doi: 10.1073/pnas.1016071107.
- da Cruz, I., Rodríguez-Casuriaga, R., Santiñaque, F. F., Fariás, J., Curti, G., Capoano, C. A., et al. (2016). Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genomics* 17, 294. doi: 10.1186/s12864-016-2618-1.
- Darde, T. A., Lecluze, E., Lardenois, A., Stévant, I., Alary, N., Tüttelmann, F., et al. (2019). The ReproGenomics Viewer: a multi-omics and cross-species resource compatible with single-cell studies for the reproductive science community. *Bioinformatics* 35, 3133–3139. doi: 10.1093/bioinformatics/btz047.
- Darde, T. A., Sallou, O., Becker, E., Evrard, B., Monjeaud, C., Le Bras, Y., et al. (2015). The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Res* 43, W109–116. doi: 10.1093/nar/gkv345.
- Davie, J. R., and Murphy, L. C. (1990). Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. *Biochemistry* 29, 4752–4757. doi: 10.1021/bi00472a002.
- de Mateo, S., and Sassone-Corsi, P. (2014). Regulation of spermatogenesis by small non-coding RNAs: Role of the germ granule. *Seminars in Cell & Developmental Biology* 29, 84–92. doi: 10.1016/j.semcd.2014.04.021.
- de Rooij, D. G. (2017). The nature and dynamics of spermatogonial stem cells. *Development* 144, 3022–3030. doi: 10.1242/dev.146571.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311. doi: 10.1126/science.1067799.
- Deshpande, A. J., Chen, L., Fazio, M., Sinha, A. U., Bernt, K. M., Banka, D., et al. (2013). Leukemic transformation by the MLL-AF6 fusion oncogene requires the H3K79 methyltransferase Dot1l. *Blood* 121, 2533–2541. doi: 10.1182/blood-2012-11-465120.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635.
- Dottermusch-Heidel, C., Gärtner, S. M. K., Tegeder, I., Rathke, C., Barckmann, B., Bartkuhn, M., et al. (2014a). H3K79 methylation: a new conserved mark that accompanies H4 hyperacetylation prior to histone-to-protamine transition in *Drosophila* and rat. *Biol Open* 3, 444–452. doi: 10.1242/bio.20147302.
- Dottermusch-Heidel, C., Klaus, E. S., Gonzalez, N. H., Bhushan, S., Meinhardt, A., Bergmann, M., et al. (2014b). H3K79 methylation directly precedes the histone-to-protamine transition in mammalian spermatids and is sensitive to bacterial infections. *Andrology* 2, 655–665. doi: 10.1111/j.2047-2927.2014.00248.x.

- Dunn, O. J. (1961). Multiple Comparisons among Means. *Journal of the American Statistical Association* 56, 52–64. doi: 10.1080/01621459.1961.10482090.
- Eberwine, J., Sul, J.-Y., Bartfai, T., and Kim, J. (2014). The promise of single-cell sequencing. *Nat Methods* 11, 25–27. doi: 10.1038/nmeth.2769.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247.
- Engström, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., Räscher, G., et al. (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods* 10, 1185–1191. doi: 10.1038/nmeth.2722.
- Erkek, S., Hisano, M., Liang, C.-Y., Gill, M., Murr, R., Dieker, J., et al. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* 20, 868–875. doi: 10.1038/nsmb.2599.
- Ernst, C., Eling, N., Martinez-Jimenez, C. P., Marioni, J. C., and Odom, D. T. (2019). Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. *Nat Commun* 10, 1251. doi: 10.1038/s41467-019-09182-1.
- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* 9, 215–216. doi: 10.1038/nmeth.1906.
- Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nature Protocols* 12, 2478–2492. doi: 10.1038/nprot.2017.124.
- Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Research* 48, D941–D947. doi: 10.1093/nar/gkz836.
- Fang, L., Zhou, Y., Liu, S., Jiang, J., Bickhart, D. M., Null, D. J., et al. (2019). Comparative analyses of sperm DNA methylomes among human, mouse and cattle provide insights into epigenomic evolution and complex traits. *Epigenetics* 14, 260–276. doi: 10.1080/15592294.2019.1582217.
- Farré, M., Robinson, T. J., and Ruiz-Herrera, A. (2015). An Integrative Breakage Model of genome architecture, reshuffling and evolution: The Integrative Breakage Model of genome evolution, a novel multidisciplinary hypothesis for the study of genome plasticity. *Bioessays* 37, 479–488. doi: 10.1002/bies.201400174.
- Feil, R., Charlton, J., Bird, A. P., Walter, J., and Reik, W. (1994). Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res* 22, 695–696. doi: 10.1093/nar/22.4.695.
- Feng, Y., Borosha, S., Ratri, A., Lee, E. B., Wang, H., Fields, T. A., et al. (2022). DOT1L Methyltransferase Regulates Calcium Influx in Erythroid Progenitor Cells in Response to Erythropoietin. *Int J Mol Sci* 23, 5137. doi: 10.3390/ijms23095137.
- Ferrari, F., Arrigoni, L., Franz, H., Izzo, A., Butenko, L., Trompouki, E., et al. (2020). DOT1L-mediated murine neuronal differentiation associates with H3K79me2 accumulation and preserves SOX2-enhancer accessibility. *Nat Commun* 11, 5200. doi: 10.1038/s41467-020-19001-7.
- Flavahan, W. A., Drier, Y., Liao, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., et al. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114. doi: 10.1038/nature16490.
- Franz, H., Villarreal, A., Heidrich, S., Videm, P., Kilpert, F., Mestres, I., et al. (2019). DOT1L promotes progenitor proliferation and primes neuronal layer identity in the developing cerebral cortex. *Nucleic Acids Research* 47, 168–183. doi: 10.1093/nar/gky953.
- Frederiks, F., Tzouros, M., Oudgenoeg, G., van Welsem, T., Fornerod, M., Krijgsveld, J., et al. (2008). Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. *Nat Struct Mol Biol* 15, 550–557. doi: 10.1038/nsmb.1432.
- Fuchs, G., Hollander, D., Voickek, Y., Ast, G., and Oren, M. (2014). Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. *Genome Res* 24, 1572–1583. doi: 10.1101/gr.176487.114.
- Gan, H., Wen, L., Liao, S., Lin, X., Ma, T., Liu, J., et al. (2013). Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat Commun* 4, 1995. doi: 10.1038/ncomms2995.
- Gaucher, J., Boussouar, F., Montellier, E., Curtet, S., Buchou, T., Bertrand, S., et al. (2012). Bromodomain-dependent stage-specific male genome programming by Brdt. *EMBO J* 31, 3809–3820. doi: 10.1038/emboj.2012.233.

- Genome Reference Consortium (n.d.). Available at: <https://www.ncbi.nlm.nih.gov/grc> [Accessed July 26, 2023].
- Gilan, O., Lam, E. Y. N., Becher, I., Lugo, D., Cannizzaro, E., Joberty, G., et al. (2016). Functional interdependence of BRD4 and DOT1L in MLL leukemia. *Nat Struct Mol Biol* 23, 673–681. doi: 10.1038/nsmb.3249.
- Godfrey, L., Crump, N. T., Thorne, R., Lau, I.-J., Repapi, E., Dimou, D., et al. (2019). DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat Commun* 10, 1–15. doi: 10.1038/s41467-019-10844-3.
- Godmann, M., Auger, V., Ferraroni-Aguiar, V., Di Sauro, A., Sette, C., Behr, R., et al. (2007). Dynamic regulation of histone H3 methylation at lysine 4 in mammalian spermatogenesis. *Biol Reprod* 77, 754–764. doi: 10.1095/biolreprod.107.062265.
- Govin, J., Escoffier, E., Rousseaux, S., Kuhn, L., Ferro, M., Thévenon, J., et al. (2007). Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol* 176, 283–294. doi: 10.1083/jcb.200604141.
- Green, C. D., Ma, Q., Manske, G. L., Shami, A. N., Zheng, X., Marini, S., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev Cell* 46, 651–667.e10. doi: 10.1016/j.devcel.2018.07.025.
- 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. doi: 10.1038/s41580-019-0159-6.
- Grigsby, S. M., Friedman, A., Chase, J., Waas, B., Ropa, J., Serio, J., et al. (2021). Elucidating the Importance of DOT1L Recruitment in MLL-AF9 Leukemia and Hematopoiesis. *Cancers (Basel)* 13, 642. doi: 10.3390/cancers13040642.
- Grishkevich, V., and Yanai, I. (2014). Gene length and expression level shape genomic novelties. *Genome Res* 24, 1497–1503. doi: 10.1101/gr.169722.113.
- Guo, J., Grow, E. J., Mlcochova, H., Maher, G. J., Lindskog, C., Nie, X., et al. (2018). The adult human testis transcriptional cell atlas. *Cell Res* 28, 1141–1157. doi: 10.1038/s41422-018-0099-2.
- Guo, J., Grow, E. J., Yi, C., Mlcochova, H., Maher, G. J., Lindskog, C., 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. doi: 10.1016/j.stem.2017.09.003.
- Guzman, C., and D'Orso, I. (2017). CIPHER: A flexible and extensive workflow platform for integrative next-generation sequencing data analysis and genomic regulatory element prediction. *BMC Bioinformatics* 18. doi: 10.1186/s12859-017-1770-1.
- Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J., and Fazio, T. G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* 177, 1319–1329.e11. doi: 10.1016/j.cell.2019.03.014.
- Hammoud, S. S., Low, D. H. P., Yi, C., Carrell, D. T., Guccione, E., and Cairns, B. R. (2014). Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis. *Cell Stem Cell* 15, 239–253. doi: 10.1016/j.stem.2014.04.006.
- Hammoud, S. S., Nix, D. A., Zhang, H., Purwar, J., Carrell, D. T., and Cairns, B. R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 460, 473–478. doi: 10.1038/nature08162.
- Handel, M. A. (2004). The XY body: a specialized meiotic chromatin domain. *Experimental Cell Research* 296, 57–63. doi: 10.1016/j.yexcr.2004.03.008.
- Hasegawa, K., Sin, H.-S., Maezawa, S., Broering, T. J., Kartashov, A. V., Alavattam, K. G., et al. (2015). SCML2 establishes the male germline epigenome through regulation of histone H2A ubiquitination. *Dev Cell* 32, 574–588. doi: 10.1016/j.devcel.2015.01.014.
- Hazzouri, M., Pivot-Pajot, C., Faure, A. K., Usson, Y., Pelletier, R., Sèle, B., et al. (2000). Regulated hyperacetylation of core histones during mouse spermatogenesis: involvement of histone deacetylases. *Eur J Cell Biol* 79, 950–960. doi: 10.1078/0171-9335-00123.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38, 576–589. doi: 10.1016/j.molcel.2010.05.004.
- Hermann, B. P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K. N., Chen, I.-C., et al. (2018). The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Rep* 25, 1650–1667.e8. doi: 10.1016/j.celrep.2018.10.026.

- Hess, R. A., and Renato de Franca, L. (2008). Spermatogenesis and cycle of the seminiferous epithelium. *Adv Exp Med Biol* 636, 1–15. doi: 10.1007/978-0-387-09597-4_1.
- Hildebrand, E., and Dekker, J. (2020). Mechanisms and functions of chromosome compartmentalization. *Trends Biochem Sci* 45, 385–396. doi: 10.1016/j.tibs.2020.01.002.
- Hirota, T., Blakeley, P., Sangrithi, M. N., Mahadevaiah, S. K., Encheva, V., Snijders, A. P., et al. (2018). SETDB1 Links the Meiotic DNA Damage Response to Sex Chromosome Silencing in Mice. *Developmental Cell* 47, 645–659.e6. doi: 10.1016/j.devcel.2018.10.004.
- Holliday, R. (1994). Epigenetics: an overview. *Dev Genet* 15, 453–457. doi: 10.1002/dvg.1020150602.
- Hu, K., Ting, A. H., and Li, J. (2015). BSPAT: a fast online tool for DNA methylation co-occurrence pattern analysis based on high-throughput bisulfite sequencing data. *BMC Bioinformatics* 16, 220. doi: 10.1186/s12859-015-0649-2.
- Hua, M., Liu, W., Chen, Y., Zhang, F., Xu, B., Liu, S., et al. (2019). Identification of small non-coding RNAs as sperm quality biomarkers for in vitro fertilization. *Cell Discov* 5, 20. doi: 10.1038/s41421-019-0087-9.
- Huff, J. T., Plocik, A. M., Guthrie, C., and Yamamoto, K. R. (2010). Reciprocal intronic and exonic histone modification regions in humans. *Nat Struct Mol Biol* 17, 1495–1499. doi: 10.1038/nsmb.1924.
- Hutcheon, K., McLaughlin, E. A., Stanger, S. J., Bernstein, I. R., Dun, M. D., Eamens, A. L., et al. (2017). Analysis of the small non-protein-coding RNA profile of mouse spermatozoa reveals specific enrichment of piRNAs within mature spermatozoa. *RNA Biol* 14, 1776–1790. doi: 10.1080/15476286.2017.1356569.
- Jang, C.-W., Shibata, Y., Starmer, J., Yee, D., and Magnuson, T. (2015). Histone H3.3 maintains genome integrity during mammalian development. *Genes Dev* 29, 1377–1392. doi: 10.1101/gad.264150.115.
- Jang, S., Kang, C., Yang, H.-S., Jung, T., Hebert, H., Chung, K. Y., et al. (2019). Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase. *Genes Dev* 33, 620–625. doi: 10.1101/gad.323790.118.
- Johnson, G. D., Lalancette, C., Linnemann, A. K., Leduc, F., Boissonneault, G., and Krawetz, S. A. (2011). The sperm nucleus: chromatin, RNA, and the nuclear matrix. *Reproduction* 141, 21–36. doi: 10.1530/REP-10-0322.
- Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., et al. (2008). The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. *PLoS Genet* 4. doi: 10.1371/journal.pgen.1000190.
- Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13, 484–492. doi: 10.1038/nrg3230.
- Jonkers, I., Kwak, H., and Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 3, e02407. doi: 10.7554/eLife.02407.
- Jung, Y. H., Sauria, M. E. G., Lyu, X., Cheema, M. S., Ausio, J., Taylor, J., et al. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* 18, 1366–1382. doi: 10.1016/j.celrep.2017.01.034.
- Karnuta, J. M., and Scacheri, P. C. (2018). Enhancers: bridging the gap between gene control and human disease. *Hum Mol Genet* 27, R219–R227. doi: 10.1093/hmg/ddy167.
- Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K., and Henikoff, S. (2020). Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* 15, 3264–3283. doi: 10.1038/s41596-020-0373-x.
- Kerry, J., Godfrey, L., Repapi, E., Tapia, M., Blackledge, N. P., Ma, H., et al. (2017). MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia. *Cell Rep* 18, 482–495. doi: 10.1016/j.celrep.2016.12.054.
- Kessler, S., Minoux, M., Joshi, O., Ben Zouari, Y., Ducret, S., Ross, F., et al. (2023). A multiple super-enhancer region establishes inter-TAD interactions and controls Hoxa function in cranial neural crest. *Nat Commun* 14, 3242. doi: 10.1038/s41467-023-38953-0.
- Kevil, C. G., Walsh, L., Laroux, F. S., Kalogeris, T., Grisham, M. B., and Alexander, J. S. (1997). An Improved, Rapid Northern Protocol. *Biochemical and Biophysical Research Communications* 238, 277–279. doi: 10.1006/bbrc.1997.7284.

- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 14, R36. doi: 10.1186/gb-2013-14-4-r36.
- Kim, K.-P., Choi, J., Yoon, J., Bruder, J. M., Shin, B., Kim, J., et al. (2021). Permissive epigenomes endow reprogramming competence to transcriptional regulators. *Nat Chem Biol* 17, 47–56. doi: 10.1038/s41589-020-0618-6.
- Kim, S.-K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., et al. (2012). Human Histone H3K79 Methyltransferase DOT1L Methyltransferase Binds Actively Transcribing RNA Polymerase II to Regulate Gene Expression *. *Journal of Biological Chemistry* 287, 39698–39709. doi: 10.1074/jbc.M112.384057.
- Kim, W., Choi, M., and Kim, J.-E. (2014). The histone methyltransferase Dot1/DOT1L as a critical regulator of the cell cycle. *Cell Cycle* 13, 726–738. doi: 10.4161/cc.28104.
- Kimmins, S., and Sassone-Corsi, P. (2005). Chromatin remodelling and epigenetic features of germ cells. *Nature* 434, 583–589. doi: 10.1038/nature03368.
- Kojima, S., Hatano, M., Okada, S., Fukuda, T., Toyama, Y., Yuasa, S., et al. (2001). Testicular germ cell apoptosis in Bcl6-deficient mice. *Development* 128, 57–65. doi: 10.1242/dev.128.1.57.
- Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* 184, 868–871. doi: 10.1126/science.184.4139.868.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005.
- Krivtsov, A. V., Feng, Z., Lemieux, M. E., Faber, J., Vempati, S., Sinha, A. U., et al. (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell* 14, 355–368. doi: 10.1016/j.ccr.2008.10.001.
- Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., et al. (2003). The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation. *Molecular Cell* 11, 721–729. doi: 10.1016/S1097-2765(03)00091-1.
- Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. doi: 10.1093/bioinformatics/btr167.
- Kwesi-Maliepaard, E. M., Aslam, M. A., Alemdehy, M. F., van den Brand, T., McLean, C., Vlaming, H., et al. (2020). The histone methyltransferase DOT1L prevents antigen-independent differentiation and safeguards epigenetic identity of CD8+ T cells. *Proc Natl Acad Sci U S A* 117, 20706–20716. doi: 10.1073/pnas.1920372117.
- La Spina, F. A., Romanato, M., Brugo-Olmedo, S., De Vincentiis, S., Julianelli, V., Rivera, R. M., et al. (2014). Heterogeneous distribution of histone methylation in mature human sperm. *J Assist Reprod Genet* 31, 45–49. doi: 10.1007/s10815-013-0137-4.
- Lambrot, R., Chan, D., Shao, X., Aarabi, M., Kwan, T., Bourque, G., et al. (2021). Whole-genome sequencing of H3K4me3 and DNA methylation in human sperm reveals regions of overlap linked to fertility and development. *Cell Rep* 36, 109418. doi: 10.1016/j.celrep.2021.109418.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359. doi: 10.1038/nmeth.1923.
- Lee, S., Oh, S., Jeong, K., Jo, H., Choi, Y., Seo, H. D., et al. (2018). Dot1 regulates nucleosome dynamics by its inherent histone chaperone activity in yeast. *Nat Commun* 9, 240. doi: 10.1038/s41467-017-02759-8.
- Lesch, B. J., Dokshin, G. A., Young, R. A., McCarrey, J. R., and Page, D. C. (2013). A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proc Natl Acad Sci U S A* 110, 16061–16066. doi: 10.1073/pnas.1315204110.
- Lesch, B. J., Silber, S. J., McCarrey, J. R., and Page, D. C. (2016). Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat Genet* 48, 888–894. doi: 10.1038/ng.3591.
- Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., et al. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12, 1725–1735. doi: 10.1093/hmg/ddg180.

- Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr Biol* 20, R754-763. doi: 10.1016/j.cub.2010.06.070.
- Li, E., and Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 6, a019133. doi: 10.1101/cshperspect.a019133.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324.
- Li, T., Liu, Q., Garza, N., Kornblau, S., and Jin, V. X. (2018). Integrative analysis reveals functional and regulatory roles of H3K79me2 in mediating alternative splicing. *Genome Med* 10, 30. doi: 10.1186/s13073-018-0538-1.
- Li, W., Wu, J., Kim, S.-Y., Zhao, M., Hearn, S. A., Zhang, M. Q., et al. (2014). Chd5 orchestrates chromatin remodelling during sperm development. *Nat Commun* 5, 3812. doi: 10.1038/ncomms4812.
- Li, X., and Wang, C.-Y. (2021). From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci* 13, 1–6. doi: 10.1038/s41368-021-00146-0.
- Liao, J., and Szabó, P. E. (2020). Maternal DOT1L is dispensable for mouse development. *Sci Rep* 10, 20636. doi: 10.1038/s41598-020-77545-6.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi: 10.1126/science.1181369.
- Lin, C., Garrett, A. S., De Kumar, B., Smith, E. R., Gogol, M., Seidel, C., et al. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev* 25, 1486–1498. doi: 10.1101/gad.2059211.
- Lin, H., Cheng, K., Kubota, H., Lan, Y., Riedel, S. S., Kakiuchi, K., et al. (2022). Histone methyltransferase DOT1L is essential for self-renewal of germline stem cells. *Genes Dev*. doi: 10.1101/gad.349550.122.
- Lin, H., Cossu, I. G., Leu, N. A., Deshpande, A. J., Bernt, K. M., Luo, M., et al. (2023). The DOT1L-MLLT10 complex regulates male fertility and promotes histone removal during spermiogenesis. *Development* 150, dev201501. doi: 10.1242/dev.201501.
- Lin, Z., Hsu, P. J., Xing, X., Fang, J., Lu, Z., Zou, Q., et al. (2017). Mettl3-/Mettl14-mediated mRNA N6-methyladenosine modulates murine spermatogenesis. *Cell Res* 27, 1216–1230. doi: 10.1038/cr.2017.117.
- Linn, E., Ghanem, L., Bhakta, H., Greer, C., and Avella, M. (2021). Genes Regulating Spermatogenesis and Sperm Function Associated With Rare Disorders. *Frontiers in Cell and Developmental Biology* 9. Available at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.634536> [Accessed December 21, 2022].
- Lismer, A., Dumeaux, V., Lafleur, C., Lambrot, R., Brind'Amour, J., Lorincz, M. C., et al. (2021). Histone H3 lysine 4 trimethylation in sperm is transmitted to the embryo and associated with diet-induced phenotypes in the offspring. *Dev Cell* 56, 671-686.e6. doi: 10.1016/j.devcel.2021.01.014.
- Lismer, A., and Kimmins, S. (2023). Emerging evidence that the mammalian sperm epigenome serves as a template for embryo development. *Nat Commun* 14, 2142. doi: 10.1038/s41467-023-37820-2.
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322. doi: 10.1038/nature08514.
- Liu, Y., Zhang, Y., Yin, J., Gao, Y., Li, Y., Bai, D., et al. (2019). Distinct H3K9me3 and DNA methylation modifications during mouse spermatogenesis. *J Biol Chem* 294, 18714–18725. doi: 10.1074/jbc.RA119.010496.
- Long, H. K., Prescott, S. L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167, 1170–1187. doi: 10.1016/j.cell.2016.09.018.
- Lopes, I., Altab, G., Raina, P., and de Magalhães, J. P. (2021). Gene Size Matters: An Analysis of Gene Length in the Human Genome. *Front Genet* 12, 559998. doi: 10.3389/fgene.2021.559998.
- Lord, T., and Oatley, J. M. (2017). A revised A single model to explain stem cell dynamics in the mouse male germline. *Reproduction* 154, R55–R64. doi: 10.1530/REP-17-0034.
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550. doi: 10.1186/s13059-014-0550-8.

- Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., et al. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334. doi: 10.1016/j.cell.2013.03.036.
- Luense, L. J., Donahue, G., Lin-Shiao, E., Rangel, R., Weller, A. H., Bartolomei, M. S., et al. (2019). Gcn5-Mediated Histone Acetylation Governs Nucleosome Dynamics in Spermiogenesis. *Developmental Cell* 51, 745–758.e6. doi: 10.1016/j.devcel.2019.10.024.
- Luense, L. J., Wang, X., Schon, S. B., Weller, A. H., Lin Shiao, E., Bryant, J. M., et al. (2016). Comprehensive analysis of histone post-translational modifications in mouse and human male germ cells. *Epigenetics & Chromatin* 9, 24. doi: 10.1186/s13072-016-0072-6.
- Lukassen, S., Bosch, E., Ekici, A. B., and Winterpacht, A. (2018). Single-cell RNA sequencing of adult mouse testes. *Sci Data* 5, 180192. doi: 10.1038/sdata.2018.192.
- Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. *Nat Rev Mol Cell Biol* 13, 543–547. doi: 10.1038/nrm3417.
- Luo, Z., Wang, X., Jiang, H., Wang, R., Chen, J., Chen, Y., et al. (2020). Reorganized 3D Genome Structures Support Transcriptional Regulation in Mouse Spermatogenesis. *iScience* 23, 101034. doi: 10.1016/j.isci.2020.101034.
- Maezawa, S., Hasegawa, K., Yukawa, M., Kubo, N., Sakashita, A., Alavattam, K. G., et al. (2018a). Polycomb protein SCML2 facilitates H3K27me3 to establish bivalent domains in the male germline. *Proceedings of the National Academy of Sciences* 115, 4957–4962. doi: 10.1073/pnas.1804512115.
- Maezawa, S., Sakashita, A., Yukawa, M., Chen, X., Takahashi, K., Alavattam, K. G., et al. (2020). Super-enhancer switching drives a burst in gene expression at the mitosis-to-meiosis transition. *Nat Struct Mol Biol* 27, 978–988. doi: 10.1038/s41594-020-0488-3.
- Maezawa, S., Yukawa, M., Alavattam, K. G., Barski, A., and Namekawa, S. H. (2018b). Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res* 46, 593–608. doi: 10.1093/nar/gkx1052.
- Malla, A. B., Rainsford, S. R., Smith, Z. D., and Lesch, B. J. (2023a). DOT1L promotes spermatid differentiation by regulating expression of genes required for histone-to-protamine replacement. *Development* 150, dev201497. doi: 10.1242/dev.201497.
- Malla, A. B., Yu, H., Farris, D., Kadimi, S., Lam, T. T., Cox, A. L., et al. (2023b). DOT1L bridges transcription and heterochromatin formation at mammalian pericentromeres. *EMBO Rep*, e56492. doi: 10.15252/embr.202256492.
- Mao, Y., and Zhang, G. (2022). A complete, telomere-to-telomere human genome sequence presents new opportunities for evolutionary genomics. *Nat Methods* 19, 635–638. doi: 10.1038/s41592-022-01512-4.
- Marson, A., Levine, S. S., Cole, M. F., Frampton, G. M., Brambrink, T., Johnstone, S., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134, 521–533. doi: 10.1016/j.cell.2008.07.020.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200.
- McKittrick, E., Gafken, P. R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proceedings of the National Academy of Sciences* 101, 1525–1530. doi: 10.1073/pnas.0308092100.
- Meistrich, M. L. (1977). “Chapter 2 Separation of Spermatogenic Cells and Nuclei from Rodent Testes,” in *Methods in Cell Biology*, ed. D. M. Prescott (Academic Press), 15–54. doi: 10.1016/S0091-679X(08)60207-1.
- Mereu, E., Lafzi, A., Moutinho, C., Ziegenhain, C., McCarthy, D. J., Álvarez-Varela, A., et al. (2020). Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat Biotechnol* 38, 747–755. doi: 10.1038/s41587-020-0469-4.
- Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323, 373–375. doi: 10.1126/science.1163601.
- Milekic, M. H., Xin, Y., O'Donnell, A., Kumar, K. K., Bradley-Moore, M., Malaspina, D., et al. (2015). Age-related sperm DNA methylation changes are transmitted to offspring and associated with abnormal behavior and dysregulated gene expression. *Mol Psychiatry* 20, 995–1001. doi: 10.1038/mp.2014.84.

- Misteli, T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* 183, 28–45. doi: 10.1016/j.cell.2020.09.014.
- Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K. C., Zhang, Y., et al. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev.* 24, 574–589. doi: 10.1101/gad.1898410.
- Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W. R., Hannon, G. J., et al. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146, 1029–1041. doi: 10.1016/j.cell.2011.08.016.
- Monesi, V. (1964). RIBONUCLEIC ACID SYNTHESIS DURING MITOSIS AND MEIOSIS IN THE MOUSE TESTIS. *J Cell Biol* 22, 521–532. doi: 10.1083/jcb.22.3.521.
- Moore, L. D., Le, T., and Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacol* 38, 23–38. doi: 10.1038/npp.2012.112.
- Moretti, C., Blanco, M., Ialy-Radio, C., Serrentino, M.-E., Gobé, C., Friedman, R., et al. (2020). Battle of the Sex Chromosomes: Competition between X and Y Chromosome-Encoded Proteins for Partner Interaction and Chromatin Occupancy Drives Multicopy Gene Expression and Evolution in Muroid Rodents. *Molecular Biology and Evolution* 37, 3453–3468. doi: 10.1093/molbev/msaa175.
- Moretti, C., Serrentino, M.-E., Ialy-Radio, C., Delessard, M., Soboleva, T. A., Tores, F., et al. (2017). SLY regulates genes involved in chromatin remodeling and interacts with TBL1XR1 during sperm differentiation. *Cell Death Differ.* 24, 1029–1044. doi: 10.1038/cdd.2017.32.
- Moretti, C., Vaiman, D., Tores, F., and Cocquet, J. (2016). Expression and epigenomic landscape of the sex chromosomes in mouse post-meiotic male germ cells. *Epigenetics & Chromatin* 9, 47. doi: 10.1186/s13072-016-0099-8.
- Moritz, L., and Hammoud, S. S. (2022). The Art of Packaging the Sperm Genome: Molecular and Structural Basis of the Histone-To-Protamine Exchange. *Frontiers in Endocrinology* 13. Available at: <https://www.frontiersin.org/articles/10.3389/fendo.2022.895502> [Accessed July 25, 2023].
- Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.-P., Monroe, S., Sreekumar, A., et al. (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood* 110, 4445–4454. doi: 10.1182/blood-2007-05-090514.
- Mueller, J. L., Mahadevaiah, S. K., Park, P. J., Warburton, P. E., Page, D. C., and Turner, J. M. A. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet* 40, 794–799. doi: 10.1038/ng.126.
- Murat, F., Mbengue, N., Winge, S. B., Trefzer, T., Leushkin, E., Sepp, M., et al. (2022). The molecular evolution of spermatogenesis across mammals. *Nature*. doi: 10.1038/s41586-022-05547-7.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349. doi: 10.1126/science.1158441.
- Namekawa, S. H., Park, P. J., Zhang, L.-F., Shima, J. E., McCarrey, J. R., Griswold, M. D., et al. (2006). Postmeiotic Sex Chromatin in the Male Germline of Mice. *Current Biology* 16, 660–667. doi: 10.1016/j.cub.2006.01.066.
- Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., et al. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347, 1017–1021. doi: 10.1126/science.1262088.
- Nguyen, A. T., Taranova, O., He, J., and Zhang, Y. (2011). DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis. *Blood* 117, 6912–6922. doi: 10.1182/blood-2011-02-334359.
- Novo, C. L., Javierre, B.-M., Cairns, J., Segonds-Pichon, A., Wingett, S. W., Freire-Pritchett, P., et al. (2018). Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. *Cell Rep* 22, 2615–2627. doi: 10.1016/j.celrep.2018.02.040.
- Okuda, H., Stanojevic, B., Kanai, A., Kawamura, T., Takahashi, S., Matsui, H., et al. (2017). Cooperative gene activation by AF4 and DOT1L drives MLL-rearranged leukemia. *J Clin Invest* 127, 1918–1931. doi: 10.1172/JCI91406.

- Ontoso, D., Kauppi, L., Keeney, S., and San-Segundo, P. A. (2014). Dynamics of DOT1L localization and H3K79 methylation during meiotic prophase I in mouse spermatocytes. *Chromosoma* 123, 147–164. doi: 10.1007/s00412-013-0438-5.
- Ooga, M., Inoue, A., Kageyama, S., Akiyama, T., Nagata, M., and Aoki, F. (2008). Changes in H3K79 Methylation During Preimplantation Development in Mice. *Biology of Reproduction* 78, 413–424. doi: 10.1095/biolreprod.107.063453.
- Ooga, M., Suzuki, M. G., and Aoki, F. (2013). Involvement of DOT1L in the remodeling of heterochromatin configuration during early preimplantation development in mice. *Biol Reprod* 89, 145. doi: 10.1095/biolreprod.113.113258.
- Pang, B., and Snyder, M. P. (2020). Systematic identification of silencers in human cells. *Nat Genet* 52, 254–263. doi: 10.1038/s41588-020-0578-5.
- Pantano, L., Jodar, M., Bak, M., Ballescà, J. L., Tommerup, N., Oliva, R., et al. (2015). The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *RNA* 21, 1085–1095. doi: 10.1261/rna.046482.114.
- Patel, L., Kang, R., Rosenberg, S. C., Qiu, Y., Raviram, R., Chee, S., et al. (2019). Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. *Nat Struct Mol Biol* 26, 164–174. doi: 10.1038/s41594-019-0187-0.
- Peters, A. H. F. M., Kubicek, S., Mechtler, K., O’Sullivan, R. J., Derijck, A. A. H. A., Perez-Burgos, L., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* 12, 1577–1589. doi: 10.1016/s1097-2765(03)00477-5.
- Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., and Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes* 12, 106. doi: 10.1186/s13104-019-4137-z.
- Plank, J. L., and Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. *Mol Cell* 55, 5–14. doi: 10.1016/j.molcel.2014.06.015.
- Qu, J., Hodges, E., Molaro, A., Gagneux, P., Dean, M. D., Hannon, G. J., et al. (2018). Evolutionary expansion of DNA hypomethylation in the mammalian germline genome. *Genome Res.* 28, 145–158. doi: 10.1101/gr.225896.117.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Bruggmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283. doi: 10.1038/nature09692.
- Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., et al. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* 9, 189. doi: 10.1038/s41467-017-02525-w.
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021.
- Rathke, C., Baarends, W. M., Awe, S., and Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839, 155–168. doi: 10.1016/j.bbagr.2013.08.004.
- Risso, D., Ngai, J., Speed, T. P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 32, 896–902. doi: 10.1038/nbt.2931.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative Genomics Viewer. *Nat Biotechnol* 29, 24–26. doi: 10.1038/nbt.1754.
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616.
- Roosen-Runge, E. C. (1962). The Process of Spermatogenesis in Mammals*. *Biological Reviews* 37, 343–376. doi: 10.1111/j.1469-185X.1962.tb01616.x.
- Sakamoto, M., Ito, D., Inoue, R., Wakayama, S., Kikuchi, Y., Yang, L., et al. (2022). Paternally inherited H3K27me3 affects chromatin accessibility in mouse embryos produced by round spermatid injection. *Development*, dev.200696. doi: 10.1242/dev.200696.
- Sakashita, A., Maezawa, S., Takahashi, K., Alavattam, K. G., Yukawa, M., Hu, Y.-C., et al. (2020). Endogenous retroviruses drive species-specific germline transcriptomes in mammals. *Nat Struct Mol Biol* 27, 967–977. doi: 10.1038/s41594-020-0487-4.

- Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., et al. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265, 687–695. doi: 10.1038/265687a0.
- Santiago, J., Silva, J. V., Howl, J., Santos, M. A. S., and Fardilha, M. (2021). All you need to know about sperm RNAs. *Hum Reprod Update* 28, 67–91. doi: 10.1093/humupd/dmab034.
- Sanyal, A., Lajoie, B., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109–113. doi: 10.1038/nature11279.
- Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* 20, 437–455. doi: 10.1038/s41576-019-0128-0.
- Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* 25, 582–597. doi: 10.1101/gr.185272.114.
- Schuster-Böckler, B., and Bateman, A. (2007). An introduction to hidden Markov models. *Curr Protoc Bioinformatics* Appendix 3, Appendix 3A. doi: 10.1002/0471250953.bia03as18.
- Shanower, G. A., Muller, M., Blanton, J. L., Honti, V., Gyurkovics, H., and Schedl, P. (2005). Characterization of the grappa gene, the Drosophila histone H3 lysine 79 methyltransferase. *Genetics* 169, 173–184. doi: 10.1534/genetics.104.033191.
- Sharma, P., Sattarifard, H., Fatemiyani, N., Lakowski, T. M., and Davie, J. R. (2022). Bioinformatic Analyses of Broad H3K79me2 Domains in Different Leukemia Cell Line Data Sets. *Cells* 11, 2830. doi: 10.3390/cells11182830.
- Sharma, U. (2019). Paternal Contributions to Offspring Health: Role of Sperm Small RNAs in Intergenerational Transmission of Epigenetic Information. *Front Cell Dev Biol* 7, 215. doi: 10.3389/fcell.2019.00215.
- Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., Bing, X. Y., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* 351, 391–396. doi: 10.1126/science.aad6780.
- Shaytan, A. K., Landsman, D., and Panchenko, A. R. (2015). Nucleosome adaptability conferred by sequence and structural variations in histone H2A–H2B dimers. *Current Opinion in Structural Biology* 32, 48–57. doi: 10.1016/j.sbi.2015.02.004.
- Shea, J. M., Serra, R. W., Carone, B. R., Shulha, H. P., Kucukural, A., Ziller, M. J., et al. (2015). Genetic and Epigenetic Variation, but Not Diet, Shape the Sperm Methylome. *Dev Cell* 35, 750–758. doi: 10.1016/j.devcel.2015.11.024.
- Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120. doi: 10.1038/nature11243.
- Shiota, H., Barral, S., Buchou, T., Tan, M., Couté, Y., Charbonnier, G., et al. (2018). Nut Directs p300-Dependent, Genome-Wide H4 Hyperacetylation in Male Germ Cells. *Cell Rep* 24, 3477–3487.e6. doi: 10.1016/j.celrep.2018.08.069.
- Sin, H.-S., Barski, A., Zhang, F., Kartashov, A. V., Nussenzweig, A., Chen, J., et al. (2012). RNF8 regulates active epigenetic modifications and escape gene activation from inactive sex chromosomes in post-meiotic spermatids. *Genes Dev* 26, 2737–2748. doi: 10.1101/gad.202713.112.
- Sin, H.-S., Kartashov, A. V., Hasegawa, K., Barski, A., and Namekawa, S. H. (2015). Poised chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biology* 13, 53. doi: 10.1186/s12915-015-0159-8.
- Singer, M. S., Kahana, A., Wolf, A. J., Meisinger, L. L., Peterson, S. E., Goggin, C., et al. (1998). Identification of high-copy disruptors of telomeric silencing in *Saccharomyces cerevisiae*. *Genetics* 150, 613–632.
- Skene, P. J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* 6, e21856. doi: 10.7554/eLife.21856.
- Soldi, M., Mari, T., Nicosia, L., Musiani, D., Sigismondo, G., Cuomo, A., et al. (2017). Chromatin proteomics reveals novel combinatorial histone modification signatures that mark distinct subpopulations of macrophage enhancers. *Nucleic Acids Res* 45, 12195–12213. doi: 10.1093/nar/gkx821.
- Song, N., Liu, J., An, S., Nishino, T., Hishikawa, Y., and Koji, T. (2011). Immunohistochemical Analysis of Histone H3 Modifications in Germ Cells during Mouse Spermatogenesis. *Acta Histochem Cytochem* 44, 183–190. doi: 10.1267/ahc.11027.

- Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., et al. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep* 3, 2179–2190. doi: 10.1016/j.celrep.2013.05.031.
- Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98, 503–517. doi: 10.1016/S0022-2836(75)80083-0.
- Squazzo, S. L., O’Geen, H., Komashko, V. M., Krig, S. R., Jin, V. X., Jang, S., et al. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res* 16, 890–900. doi: 10.1101/gr.5306606.
- Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495. doi: 10.1038/nature10716.
- Steger, D. J., Lefterova, M. I., Ying, L., Stonestrom, A. J., Schupp, M., Zhuo, D., et al. (2008). DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells. *Molecular and Cellular Biology* 28, 2825–2839. doi: 10.1128/MCB.02076-07.
- Steilmann, C., Paradowska, A., Bartkuhn, M., Vieweg, M., Schuppe, H.-C., Bergmann, M., et al. (2011). Presence of histone H3 acetylated at lysine 9 in male germ cells and its distribution pattern in the genome of human spermatozoa. *Reprod Fertil Dev* 23, 997–1011. doi: 10.1071/RD10197.
- Stewart, K. R., Veselovska, L., and Kelsey, G. (2016). Establishment and functions of DNA methylation in the germline. *Epigenomics* 8, 1399–1413. doi: 10.2217/epi-2016-0056.
- Stovner, E. B., and Sætrom, P. (2019). epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* 35, 4392–4393. doi: 10.1093/bioinformatics/btz232.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102.
- Sun, S., Jiang, Y., Zhang, Q., Pan, H., Li, X., Yang, L., et al. (2022). Znhit1 controls meiotic initiation in male germ cells by coordinating with Stra8 to activate meiotic gene expression. *Developmental Cell* 57, 901–913.e4. doi: 10.1016/j.devcel.2022.03.006.
- Sun, Z.-W., and Allis, C. D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 418, 104–108. doi: 10.1038/nature00883.
- Suzuki, H., Takatsuka, S., Akashi, H., Yamamoto, E., Nojima, M., Maruyama, R., et al. (2011). Genome-wide Profiling of Chromatin Signatures Reveals Epigenetic Regulation of MicroRNA Genes in Colorectal Cancer. *Cancer Research* 71, 5646–5658. doi: 10.1158/0008-5472.CAN-11-1076.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., et al. (2011). Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* 146, 1016–1028. doi: 10.1016/j.cell.2011.08.008.
- Temple-Smith, P. D., Ravichandran, A., and Nunez, F. E. H. (2018). Sperm: Comparative vertebrate. *Encyclopedia of Reproduction: Volume 6: Comparative Reproduction*, 210–220. doi: 10.1016/B978-0-12-809633-8.20558-X.
- Tomizawa, S.-I., Kobayashi, Y., Shirakawa, T., Watanabe, K., Mizoguchi, K., Hoshi, I., et al. (2018). Kmt2b conveys monovalent and bivalent H3K4me3 in mouse spermatogonial stem cells at germline and embryonic promoters. *Development* 145, dev169102. doi: 10.1242/dev.169102.
- Turner, J. M. A. (2007). Meiotic sex chromosome inactivation. *Development* 134, 1823–1831. doi: 10.1242/dev.000018.
- Turner, J. M. A. (2015). Meiotic Silencing in Mammals. *Annu Rev Genet* 49, 395–412. doi: 10.1146/annurev-genet-112414-055145.
- Valencia-Sánchez, M. I., De Ioannes, P., Wang, M., Truong, D. M., Lee, R., Armache, J.-P., et al. (2021). Regulation of the Dot1 histone H3K79 methyltransferase by histone H4K16 acetylation. *Science* 371, eabc6663. doi: 10.1126/science.abc6663.
- van Leeuwen, F., Gafken, P., and De, G. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* 109. doi: 10.1016/s0092-8674(02)00759-6.

- van Welsem, T., Korthout, T., Ekkebus, R., Morais, D., Molenaar, T. M., van Harten, K., et al. (2018). Dot1 promotes H2B ubiquitination by a methyltransferase-independent mechanism. *Nucleic Acids Res* 46, 11251–11261. doi: 10.1093/nar/gky801.
- Vara, C., Paytuví-Gallart, A., Cuartero, Y., Le Dily, F., Garcia, F., Salvà-Castro, J., et al. (2019). Three-Dimensional Genomic Structure and Cohesin Occupancy Correlate with Transcriptional Activity during Spermatogenesis. *Cell Rep* 28, 352–367.e9. doi: 10.1016/j.celrep.2019.06.037.
- Veloso, A., Kirkconnell, K. S., Magnuson, B., Biewen, B., Paulsen, M. T., Wilson, T. E., et al. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* 24, 896–905. doi: 10.1101/gr.171405.113.
- Vignieri, S. (2023). Zoonomia. *Science* 380, 356–357. doi: 10.1126/science.adi1599.
- Vlaming, H., and van Leeuwen, F. (2016). The upstreams and downstreams of H3K79 methylation by DOT1L. *Chromosoma* 125, 593–605. doi: 10.1007/s00412-015-0570-5.
- Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., et al. (2003). mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol Cell* 12, 475–487. doi: 10.1016/j.molcel.2003.08.007.
- Wang, J., Lin, B., Zhang, Y., Ni, L., Hu, L., Yang, J., et al. (2020). The Regulatory Role of Histone Modification on Gene Expression in the Early Stage of Myocardial Infarction. *Front. Cardiovasc. Med.* 7. doi: 10.3389/fcvm.2020.594325.
- Wang, M., Chen, Z., and Zhang, Y. (2022). CBP/p300 and HDAC activities regulate H3K27 acetylation dynamics and zygotic genome activation in mouse preimplantation embryos. *EMBO J* 41, e112012. doi: 10.15252/embj.2022112012.
- Wang, M., Liu, X., Chang, G., Chen, Y., An, G., Yan, L., et al. (2018). Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. *Cell Stem Cell* 23, 599–614.e4. doi: 10.1016/j.stem.2018.08.007.
- Wang, T., Gao, H., Li, W., and Liu, C. (2019a). Essential Role of Histone Replacement and Modifications in Male Fertility. *Front Genet* 10, 962. doi: 10.3389/fgene.2019.00962.
- Wang, Y., Wang, H., Zhang, Y., Du, Z., Si, W., Fan, S., et al. (2019b). Reprogramming of Meiotic Chromatin Architecture during Spermatogenesis. *Mol Cell* 73, 547–561.e6. doi: 10.1016/j.molcel.2018.11.019.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63. doi: 10.1038/nrg2484.
- Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., et al. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39, 457–466. doi: 10.1038/ng1990.
- Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., et al. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319. doi: 10.1016/j.cell.2013.03.035.
- Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239–1243. doi: 10.1038/nature07002.
- Wille, C. K., and Sridharan, R. (2022). Connecting the DOTs on Cell Identity. *Front Cell Dev Biol* 10, 906713. doi: 10.3389/fcell.2022.906713.
- Willemsen, L., Prange, K. H. M., Neele, A. E., van Roomen, C. P. A. A., Gijbels, M., Griffith, G. R., et al. (2022). DOT1L regulates lipid biosynthesis and inflammatory responses in macrophages and promotes atherosclerotic plaque stability. *Cell Rep* 41, 111703. doi: 10.1016/j.celrep.2022.111703.
- Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., et al. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* 46, W11–W16. doi: 10.1093/nar/gky504.
- Wood, K., Tellier, M., and Murphy, S. (2018). DOT1L and H3K79 Methylation in Transcription and Genomic Stability. *Biomolecules* 8. doi: 10.3390/biom8010011.
- Worden, E. J., Hoffmann, N. A., Hicks, C. W., and Wolberger, C. (2019). Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* 176, 1490–1501.e12. doi: 10.1016/j.cell.2019.02.002.

- Xu, S., Grullon, S., Ge, K., and Peng, W. (2014). Spatial Clustering for Identification of ChIP-Enriched Regions (SICER) to Map Regions of Histone Methylation Patterns in Embryonic Stem Cells. *Methods Mol Biol* 1150, 97–111. doi: 10.1007/978-1-4939-0512-6_5.
- Yamaguchi, K., Hada, M., Fukuda, Y., Inoue, E., Makino, Y., Katou, Y., et al. (2018). Re-evaluating the Localization of Sperm-Retained Histones Revealed the Modification-Dependent Accumulation in Specific Genome Regions. *Cell Rep* 23, 3920–3932. doi: 10.1016/j.celrep.2018.05.094.
- Yan, F., Powell, D. R., Curtis, D. J., and Wong, N. C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biology* 21, 22. doi: 10.1186/s13059-020-1929-3.
- Yao, T., Jing, W., Hu, Z., Tan, M., Cao, M., Wang, Q., et al. (2019). Structural basis of the crosstalk between histone H2B monoubiquitination and H3 lysine 79 methylation on nucleosome. *Cell Res* 29, 330–333. doi: 10.1038/s41422-019-0146-7.
- Yao, X., Tang, Z., Fu, X., Yin, J., Liang, Y., Li, C., et al. (2015). The Mediator subunit MED23 couples H2B mono-ubiquitination to transcriptional control and cell fate determination. *The EMBO Journal* 34, 2885–2902. doi: 10.15252/emj.201591279.
- Yin, Q., Yang, C.-H., Strelkova, O. S., Wu, J., Sun, Y., Gopalan, S., et al. (2023). Revisiting chromatin packaging in mouse sperm. 2022.12.26.521943. doi: 10.1101/2022.12.26.521943.
- Yoh, S. M., Lucas, J. S., and Jones, K. A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* 22, 3422–3434. doi: 10.1101/gad.1720008.
- Yoshida, K., Muratani, M., Araki, H., Miura, F., Suzuki, T., Dohmae, N., et al. (2018). Mapping of histone-binding sites in histone replacement-completed spermatozoa. *Nat Commun* 9, 3885. doi: 10.1038/s41467-018-06243-9.
- Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383. doi: 10.1093/bioinformatics/btv145.
- Yu, T., Tang, C., Zhang, Y., Zhang, R., and Yan, W. (2017). Microfluidics-based digital quantitative PCR for single-cell small RNA quantification. *Biol Reprod* 97, 490–496. doi: 10.1093/biolre/iox102.
- Zentner, G. E., Tesar, P. J., and Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* 21, 1273–1283. doi: 10.1101/gr.122382.111.
- Zhang, K., Sridhar, V. V., Zhu, J., Kapoor, A., and Zhu, J.-K. (2007). Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. *PLoS One* 2, e1210. doi: 10.1371/journal.pone.0001210.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* 9, R137. doi: 10.1186/gb-2008-9-9-r137.
- Zhang, Y., Shi, J., Rassoulzadegan, M., Tuorto, F., and Chen, Q. (2019). Sperm RNA code programmes the metabolic health of offspring. *Nat Rev Endocrinol* 15, 489–498. doi: 10.1038/s41574-019-0226-2.
- Zhao, X., Li, X., Sun, H., Zhao, X., Gao, T., Shi, P., et al. (2023). Dot1l cooperates with Npm1 to repress endogenous retrovirus MERVL in embryonic stem cells. *Nucleic Acids Research*, gkad640. doi: 10.1093/nar/gkad640.
- Zhao, Z., and Zhang, F. (2006). Sequence context analysis in the mouse genome: Single nucleotide polymorphisms and CpG island sequences. *Genomics* 87, 68–74. doi: 10.1016/j.ygeno.2005.09.012.
- Zhou, C., Halstead, M. M., Bonnet-Garnier, A., Schultz, R. M., and Ross, P. J. (2023). Histone remodeling reflects conserved mechanisms of bovine and human preimplantation development. *EMBO reports* 24, e55726. doi: 10.15252/embr.202255726.
- Zhu, B., Zheng, Y., Pham, A.-D., Mandal, S. S., Erdjument-Bromage, H., Tempst, P., et al. (2005). Monoubiquitination of Human Histone H2B: The Factors Involved and Their Roles in HOX Gene Regulation. *Molecular Cell* 20, 601–611. doi: 10.1016/j.molcel.2005.09.025.
- Zuo, X., Rong, B., Li, L., Lv, R., Lan, F., and Tong, M.-H. (2018). The histone methyltransferase SETD2 is required for expression of acrosin-binding protein 1 and protamines and essential for spermiogenesis in mice. *J Biol Chem* 293, 9188–9197. doi: 10.1074/jbc.RA118.002851.

Annexes

Vignette ChromENVEE

ChromENVEE: Chromatin ENVironment and Enhancer-dependent Expression

Manon Coulée

2023-01-02

Contents

Abstract	1
Citation	2
Introduction	2
Data initialization	2
Distribution of the chromatin states in the genome	4
Annotation of enhancers	5
Association of enhancers to genes	5
Gene expression information	6
Visualization of enhancer annotation	8
Enhancer annotation comparison	9
Characterization of chromatin states in the gene environment	12
Chromatin states at gene promoters	12
Predominant chromatin state at gene promoters	12
Session Information	14
References	15

Abstract

Standard analyses on ChIPseq data provide information (annotation, enrichment level) at the gene body level but do not necessarily investigate other genomic regions. ChromHMM R package allows to go further by predicting chromatin states using ChIPSeq datasets for several histone marks. The present R package ChromENVEE uses the chromatin states obtained by ChromHMM and compare them with transcriptomic data (RNAseq) and other ChIP-Seq data.

Specifically, ChromENVEE implements functions to associate all the neighbouring genes to a list of enhancers and to define the chromatin environment of genes using chromatin states informations. Several visualization functions are available to summarize the distribution of chromatin states, characterize genes associated with enhancers and also assign chromatin environment to genes.

Citation

If you use ChromENVEE in published research, please cite:

- Manon Coulee, Guillaume Meurice, Julie Cocquet* and Laila El Khattabi* (2022). ChromENVEE: Chromatin Environment and Enhancer-dependent Expression. R package version 1.1.8. *co-authorship

Introduction

ChromENVEE (Chromatin **ENV**ironment and **ENH**ancer-dependent **EXP**ression) is a package that was developed to define chromatin dynamics in a specific cell type and to characterize a histone mark at the enhancer level and its chromatin environment.

ChromHMM (*Ersnt et al, 2012*) is a tool using the Hidden Markov Model (HMM) method to predict the most likely chromatin state of each genomic segment. The tool uses ChIPseq data from multiple epigenetic marks to predict chromatin states, each characterized by at least one epigenetic mark. In the case of this present study, six epigenetic marks from 15 different cell types were used to build a model of 18 chromatin states.

The package contains several applications all using the results obtained with ChromHMM tools.

- It characterizes the distribution of the chromatin states in a given cell type.
- The package can associate chromatin states defined as enhancers with genes located nearby.
- Using transcriptomic (RNAseq) data it can also analyze the expression of those nearby genes and produce graphs to visualize the results. ChromENVEE can also determine the chromatin environment of a gene and estimate the predominant chromatin state.

The package was developed to in depth characterize a chromatin mark and correlate it with gene expression and chromatin environment in given cell types. In the present study, we focused on the chromatin mark H3K79me2 because two recent studies had shown that the presence of H3K79me2 at a subset of active enhancers can regulate gene expression (*Ferrari et al, 2020; Godfrey et al, 2019*).

```
# Loading package
library(ChromENVEE)
```

Data initialization

`colorTable` is a dataframe that gives the following information: chromatin state numbers (`stateNumber`), chromatin state names (`stateName`) and chromatin state colors (`colorValue`). This table is necessary for plot generation. `colorValue` accepts as value hex code and/or color name code.

```
data(colorTable)
```

	stateNumber	stateName	colorValue
1	U1	TSSA	#B71C1C
2	U2	TSSFlnk	#E65100
3	U3	TSSFlnkD	#E65100
4	U4	Tx	#43A047
5	U5	TxWk	#1B5E20
6	U6	EnhG	#99FF66
7	U7	EnhG	#99FF66
8	U8	EnhA	#F5B041
9	U9	EnhWk	#FFEB3B
10	U10	ZNFRpts	#48C9B0
11	U11	Het	#B39DDB
12	U12	TssBiv	#880E4F
13	U13	EnhBiv	#666633
14	U14	ReprPC	#424949
15	U15	ReprPCWk	#7B7D7D
16	U16	Quies	#D0D3D4
17	U17	Quies	#D0D3D4
18	U18	Quies	#D0D3D4

genomeFile is a GRanges object generated from an annotation bed file. In the case of this present study, we used the mouse Ensembl annotation file.

genomeFile should contain the following information: chromosome (chr), gene position (start and end), strand information (strand) and gene name (gene_ENS). The score information is suggested but not mandatory.

```
data(genomeFile)
```

```
#> GRanges object with 6 ranges and 2 metadata columns:
#>      seqnames      ranges strand |      score      gene_ENS
#>      <Rle>        <IRanges> <Rle> | <character> <character>
#> 24099 chr8 3056294-3056445 + | . ENSMUSG00000108612.2
#> 24102 chr8 3104279-3105802 + | . ENSMUSG00000108971.1
#> 24105 chr8 3259704-3260641 + | . ENSMUSG00000086065.1
#> 24108 chr8 3457105-3467680 - | . ENSMUSG00000069633.12
#> 24111 chr8 3493138-3497208 + | . ENSMUSG00000047264.8
#> 24114 chr8 3567990-3584939 + | . ENSMUSG00000065952.13
#> -----
#> seqinfo: 22 sequences from mm10 genome; no seqlengths
```

chromatinState is a GRanges object that contains chromatin states information. It is generated with the output of the ChromHMM tool.

chromatinState should contain the following information: chromosome (chr), genomic regions (start and end), chromatin states (state and state_name) and sample name (name).

```
data(chromatinState)
```

```
#> GRanges object with 6 ranges and 3 metadata columns:
#>      seqnames      ranges strand |      state      name state_name
```

```

#>      <Rle>      <IRanges> <Rle> | <character> <character> <factor>
#> [1] chr10      0-3100000    * |      U16      RS      Quies
#> [2] chr10 3100000-3109200    * |      U11      RS      Het
#> [3] chr10 3109200-3110600    * |      U12      RS      TssBiv
#> [4] chr10 3110600-3111000    * |      U14      RS      ReprPC
#> [5] chr10 3111000-3111200    * |      U13      RS      EnhBiv
#> [6] chr10 3111200-3117200    * |      U12      RS      TssBiv
#> -----
#> seqinfo: 22 sequences from mm10 genome; no seqlengths

```

Distribution of the chromatin states in the genome

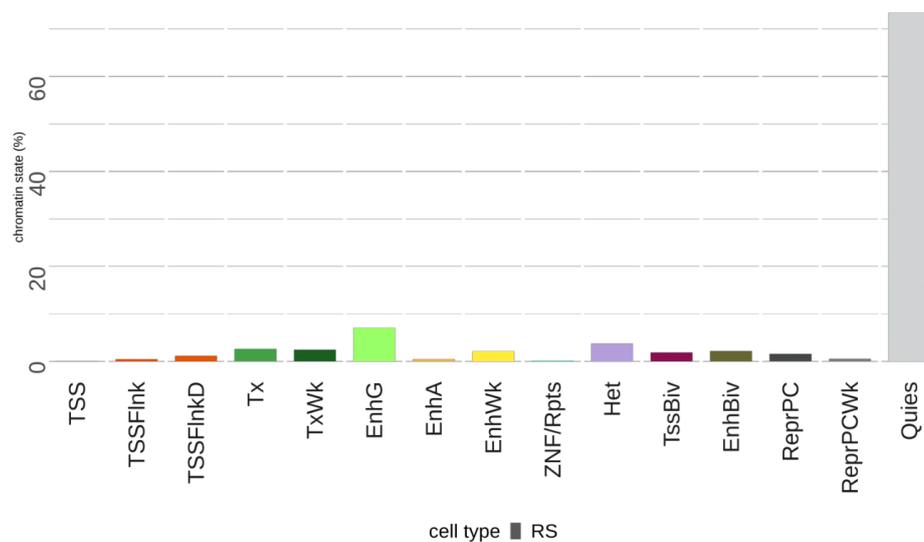
`plotChromatinState()` calculates the percentage of each chromatin state at a given genomic region. The output consists of a dataframe with the percentage of coverage for each chromatin state relative to the length of the genomic region. It is possible to plot the results in PNG file with the argument `plot = TRUE`. If you have a list of dataframe, it is possible to merge all the dataframe in a unique merged dataframe and in a unique plot with the argument `merge = TRUE`.

```
summary_chromatin_state = plotChromatinState(chromatinState, merge = TRUE, plot = FALSE,
colorTable = colorTable, filename = "")
```

```

#>      state coverage sample_name
#> TSSA      TSSA 0.08519426      RS
#> TSSFlnk  TSSFlnk 0.45530134      RS
#> TSSFlnkD TSSFlnkD 1.18900667      RS
#> Tx        Tx 2.60257103      RS
#> TxWk      TxWk 2.44911129      RS
#> EnhG      EnhG 7.10081351      RS

```



Annotation of enhancers

Enhancers are cis-regulatory regions that (locate more or less) near or even within their regulated gene. We assume that an enhancer, may regulate all its neighbouring genes within a given distance (in this present study, the distance is 500kb). We focus on enhancer chromatin states (in this study, we merged them into four types: bivalent enhancers (EnhBiv), genic enhancers (EnhG), active enhancers (EnhA) and weak enhancers (EnhWk)).

`listTableEnhancer` is a `GRanges` object or a list of `GRanges` objects (produced by `GenomicRanges` package). Similar to `chromatinState` dataframe, `listTableEnhancer` should contain genes and chromatin states informations. Sample name (`sample_name`) is mandatory to compare enhancer annotation (see Enhancer annotation comparison).

```
data(listTableEnhancer)

#> GRanges object with 119 ranges and 2 metadata columns:
#>      seqnames      ranges strand | chromatin_state sample_name
#>      <Rle>        <IRanges> <Rle> | <character> <character>
#> 2 chr10 9342200-9344000 * | U13 EnhBiv
#> 5 chr10 20952400-20952600 * | U13 EnhBiv
#> 8 chr10 22944400-22944600 * | U13 EnhBiv
#> 11 chr10 41044400-41044800 * | U13 EnhBiv
#> 14 chr10 43482600-43482800 * | U13 EnhBiv
#> ... ..
#> 1851 chr8 120539600-120543400 * | U13 EnhBiv
#> 1854 chr8 121854400-121855000 * | U13 EnhBiv
#> 1857 chr8 123851800-123852800 * | U13 EnhBiv
#> 1860 chr8 124769400-124769600 * | U13 EnhBiv
#> 1863 chr8 125388200-125389400 * | U13 EnhBiv
#> -----
#> seqinfo: 22 sequences from mm10 genome; no seqlengths
```

Association of enhancers to genes

To determine which genes are associated to which enhancers, we assign to each enhancer all the genes located within an interval. To do that, `enhancerAnnotation()` uses a `GRanges` object.

The function takes few minutes to process depending on the size of your enhancer table. It is possible to preformed multithreading using the `nCore` parameter. To each enhancer position, we obtain the list of associated genes and their distance from the enhancer (in bp).

```
table_enhancer_gene = enhancerAnnotation(listTableEnhancer[[1]], genome = genomeFile,
interval = 500000, nCore = 1)

#> GRanges object with 6 ranges and 7 metadata columns:
#>      seqnames      ranges strand | chromatin_state sample_name
#>      <Rle>        <IRanges> <Rle> | <character> <character>
#> 1.2 chr10 9342200-9344000 * | U13 EnhBiv
#> 1.5 chr10 20952400-20952600 * | U13 EnhBiv
#> 1.8 chr10 22944400-22944600 * | U13 EnhBiv
#> 1.11 chr10 41044400-41044800 * | U13 EnhBiv
#> 1.14 chr10 43482600-43482800 * | U13 EnhBiv
#> 1.17 chr10 53752800-53753000 * | U13 EnhBiv
#>      start_500kb end_500kb gene_association      distance
#>      <numeric> <numeric>      <integer>      <character>
```

```

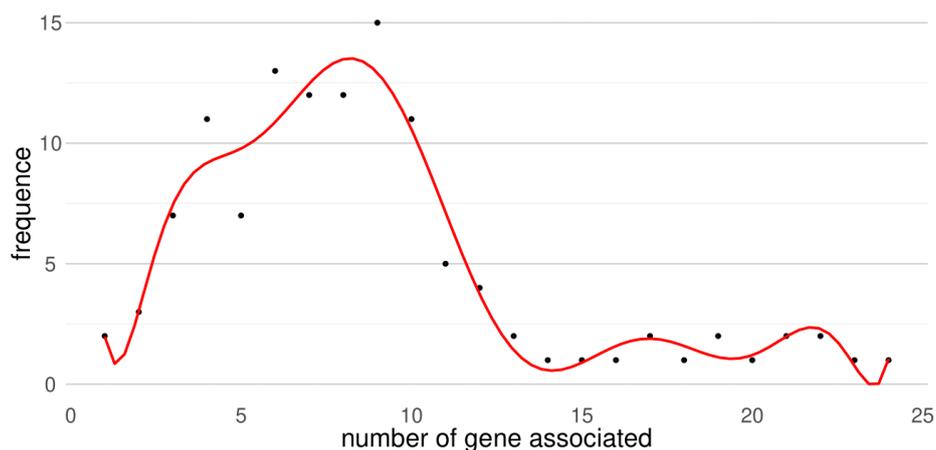
#> 1.2      8842200  9844000                6 397063;118229;12532;..
#> 1.5      20452400 21452600             7 121097;77890;208384;..
#> 1.8      22444400 23444600             3  212462;153081;11984
#> 1.11     40544400 41544800             9 477728;361118;105830..
#> 1.14     42982600 43982800             9 328384;17718;109112;..
#> 1.17     53252800 54253000             8 469114;415047;277331..
#>
#>          gene_list
#>          <character>
#> 1.2 ENSMUSG00000111731.1..
#> 1.5 ENSMUSG00000111091.1..
#> 1.8 ENSMUSG00000071359.1..
#> 1.11 ENSMUSG00000111108.1..
#> 1.14 ENSMUSG00000099893.1..
#> 1.17 ENSMUSG00000112709.1..
#> -----
#> seqinfo: 22 sequences from mm10 genome; no seqlengths

```

Number of genes associated with an enhancer

With the `enhancerAnnotation()` function, each enhancer region can be associated at least one genes. The function `plotGeneAssociation()` allows to represent the distribution of the number of genes associated with the enhancers. The function uses polynomial linear regression for the graph representation.

```
plotGeneAssociation(table_enhancer_gene, all = FALSE)
```



Gene expression information

`geneExpression` is a dataframe that contains information on the gene expression level.

It is generated with the results from RNAseq gene expression analysis. `geneExpression` should contain the following information: chromosome (`chr`), gene position (start and end), gene name (`gene_ENS`), strand information (`strand`), level of gene expression (`gene_expression`). The score is not necessary for the analysis. For the gene name, the same name than the one used to generate the `genomeFile` dataframe should be used.

```
data(geneExpression)

#>      gene_ENS seqnames      start      end  width strand
#> 1 ENSMUSG00000000290.13 chr10 77530252 77565708 35457 +
#> 2 ENSMUSG00000000295.13 chr10 31313383 31328204 14822 +
#> 3 ENSMUSG00000000296.8 chr10 31332376 31445958 113583 -
#> 4 ENSMUSG00000000303.12 chr8 106603351 106670246 66896 +
#> 5 ENSMUSG00000000374.8 chr10 78186725 78244641 57917 -
#> 6 ENSMUSG00000000711.3 chr10 128677175 128696264 19090 -
#> gene_expression
#> 1 1.0547488
#> 2 0.3407228
#> 3 0.4008302
#> 4 0.2292526
#> 5 29.3456974
#> 6 41.4675071
```

enhancerExpression() is able to associate the level of gene expression to each gene-enhancer pair that was determined by the enhancerAnnotation function. When a gene-enhancer pair is not associated to an expression level, the function indicates NA.

```
table_enhancer_gene_expression = enhancerExpression(table_enhancer_gene,
geneExpressionTable = geneExpression)
```

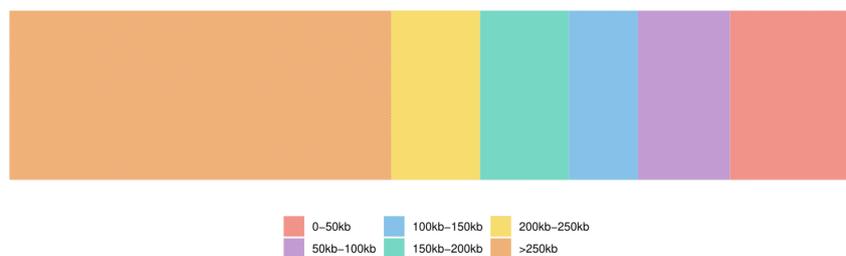
```
#> GRanges object with 6 ranges and 8 metadata columns:
#>      seqnames      ranges strand | chromatin_state sample_name
#>      <Rle>      <IRanges> <Rle> | <character> <character>
#> 1.2 chr10 9342200-9344000 * | U13 EnhBiv
#> 1.5 chr10 20952400-20952600 * | U13 EnhBiv
#> 1.8 chr10 22944400-22944600 * | U13 EnhBiv
#> 1.11 chr10 41044400-41044800 * | U13 EnhBiv
#> 1.14 chr10 43482600-43482800 * | U13 EnhBiv
#> 1.17 chr10 53752800-53753000 * | U13 EnhBiv
#>      start_500kb end_500kb gene_association distance
#>      <numeric> <numeric> <integer> <character>
#> 1.2 8842200 9844000 6 397063;118229;12532;..
#> 1.5 20452400 21452600 7 121097;77890;208384;..
#> 1.8 22444400 23444600 3 212462;153081;11984
#> 1.11 40544400 41544800 9 477728;361118;105830..
#> 1.14 42982600 43982800 9 328384;17718;109112;..
#> 1.17 53252800 54253000 8 469114;415047;277331..
#>      gene_list      gene_expression
#>      <character> <character>
#> 1.2 ENSMUSG00000111731.1.. NA;NA;2.918723039239..
#> 1.5 ENSMUSG00000111091.1.. NA;NA;0.025907853489..
#> 1.8 ENSMUSG00000071359.1.. 606.043748379438;NA;NA
#> 1.11 ENSMUSG00000111108.1.. NA;NA;14.11411792553..
#> 1.14 ENSMUSG00000099893.1.. NA;1.37378184530106;..
#> 1.17 ENSMUSG00000112709.1.. 11.2589040047601;1.0..
#> -----
#> seqinfo: 22 sequences from mm10 genome; no seqlengths
```

Visualization of enhancer annotation

Distribution of genes according to their distance from the enhancer

`plotGeneDistance()` enables the generation of a plot showing gene distribution according to their distance from the associated enhancer. The distance is calculated using the `limit` argument and clustered into six groups as shown in the plot below.

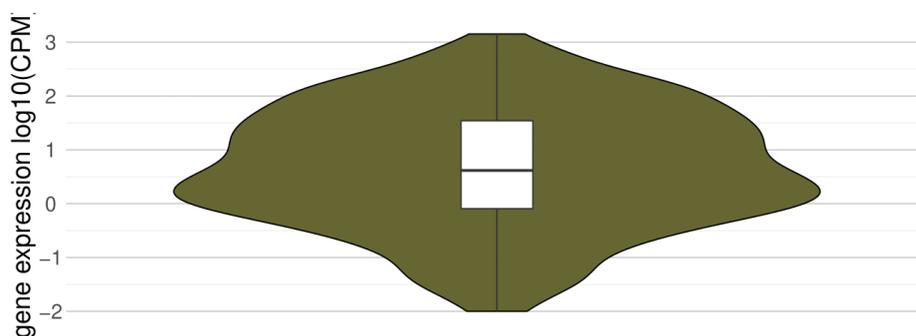
```
plotGeneDistance(table_enhancer_gene_expression, limit = 500000, xlab = "",
ylab = "distance enhancer-gene (bp)")
```



Expression of a gene associated with a given enhancer

`plotEnhancerExpression()` allows to generate a plot of gene expression distribution according to the type of enhancer. It is possible to rescale the plot using the `scale` argument ('none', 'log10' and 'log2' are accepted).

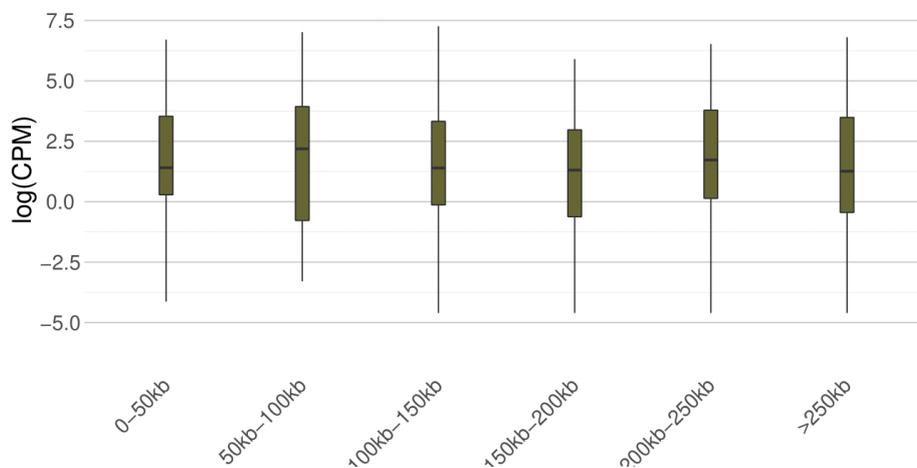
```
plotEnhancerExpression(table_enhancer_gene_expression, scale = "log10",
colorTable = colorTable, ylab = "gene expression log10(CPM)")
```



Gene expression according to gene-enhancer distance

`plotDistanceExpression()` enables the generation of a plot of the level of gene expression according to the gene-enhancer distance. The distance is calculated using `limit` argument and clustered into six groups as illustrated in the plot below.

```
plotDistanceExpression(table_enhancer_gene_expression, colorTable = colorTable,
limit = 500000)
```



Enhancer annotation comparison

It is possible to compare different categories of enhancers by means of a list of GRanges objects, each containing input information similar to the one in `listTableEnhancer`. Unlike the individual analysis, each GRanges object in the list requires sample information (`sample_name`).

The first step is to assign to each enhancer all the genes located within an interval using `enhancerAnnotation()`. After gene association, we associate the gene expression at enhancer using `enhancerExpression()`.

```
list_table_enhancer_gene = lapply(listTableEnhancer, enhancerAnnotation,
genome = genomeFile, interval = 500000, nCore = 1)
listTableEnhancerGeneExpression = lapply(list_table_enhancer_gene, enhancerExpression,
geneExpressionTable = geneExpression)
```

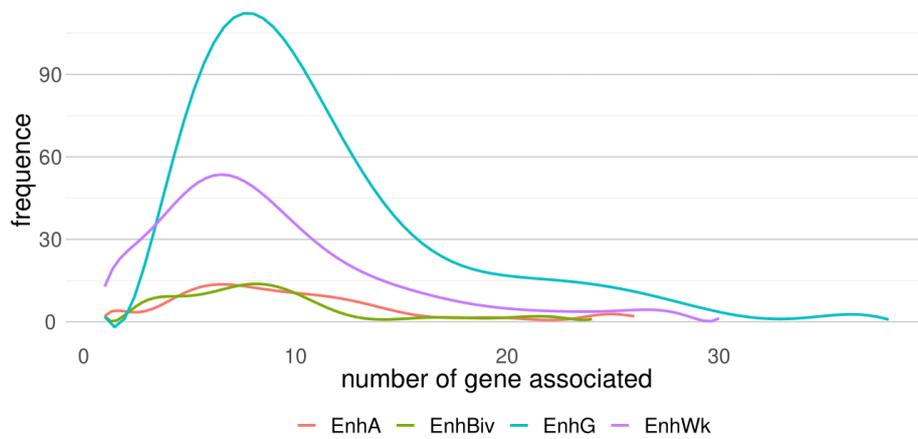
This process takes a few minutes. To reduce time, you can load the `listTableEnhancerGeneExpression` data to process the following analyses.

```
data(listTableEnhancerGeneExpression)
```

Number of genes associated with the enhancer

With the `enhancerAnnotation()` function, each enhancer region can be associated with at least one gene. The function `plotGeneAssociation()` allows to represent the distribution of the number of genes associated with the enhancers. The function uses polynomial linear regression for the graph representation. `all = TRUE` parameter is used to compile all enhancer tables in same 'png' file.

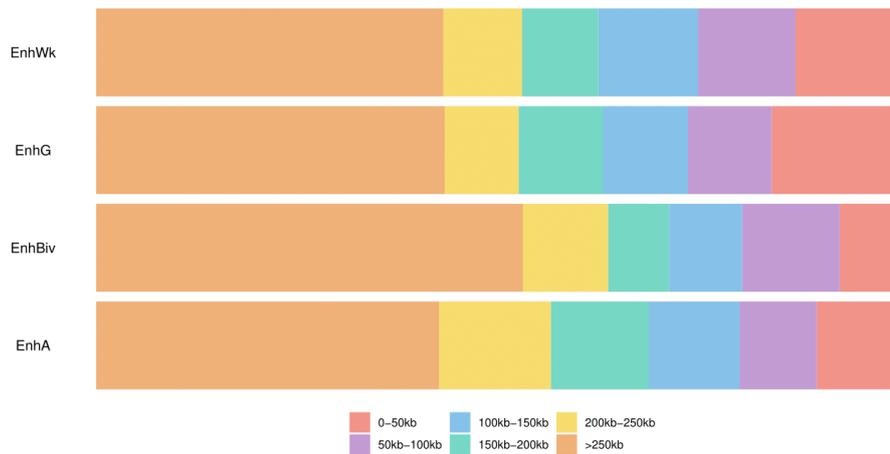
```
plotGeneAssociation(listTableEnhancerGeneExpression, all = TRUE)
```



Distribution of genes according to the gene-enhancer distance

plotGeneDistance() allows to generate a plot of gene distribution according to gene-enhancer distance. The distance is calculated with the limit argument and clustered into six groups as illustrated in the plot below.

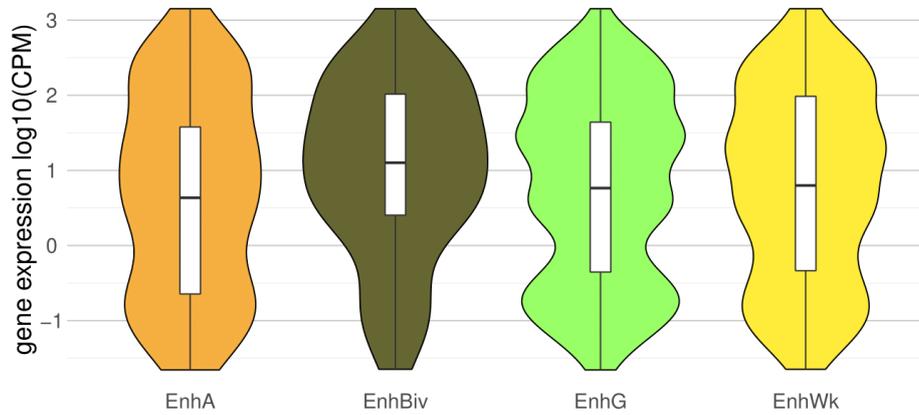
```
plotGeneDistance(listTableEnhancerGeneExpression, limit = 500000,
xlab = "", ylab = "distance enhancer-gene (bp)")
```



Expression of a gene associated with enhancers

plotEnhancerExpression() allows to generate a plot of gene expression distribution according to the type of enhancer. It is possible to rescale the plot using the scale argument ('none', 'log10' and 'log2' are accepted).

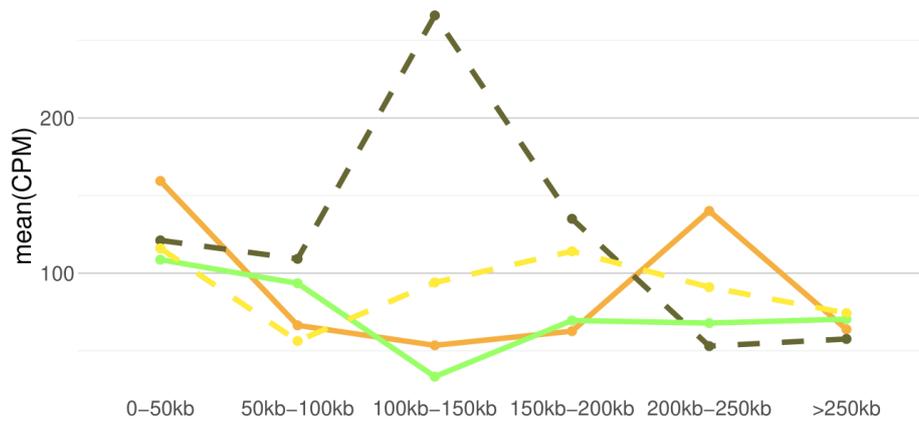
```
plotEnhancerExpression(listTableEnhancerGeneExpression, scale = "log10",
colorTable = colorTable, ylab = "gene expression log10(CPM)")
```



Expression of genes according to their distance from their associated enhancers

This function generates a plot to visualize the level of gene expression according to the distance between a gene and its associated enhancer, using `plotDistanceExpression`. The distance is calculated using the `limit` argument and clustered into six groups as shown in the plot below. In case of list of enhancer, the function shows the average expression of all genes associated with each enhancer.

```
plotDistanceExpression(listTableEnhancerGeneExpression, colorTable = colorTable,
limit = 500000)
```



Characterization of chromatin states in the gene environment

This aims at analyzing the chromatin landscape within genes. To perform this analysis, gene expression data from RNAseq analysis (`geneExpression`) as well as chromatin state data from ChromHMM analysis (`chromatinState`) are needed.

```
data(geneExpression)
data(chromatinState)
```

Chromatin states at gene promoters

The `geneEnvironment()` function calculates the percentage of overlap of each chromatin state with each genes promoters using the `interval` parameter.

`geneEnvironment()` may take a few minutes depending on the number of genes analyzed.

```
table_overlapping = geneEnvironment(geneExpression, chromatinState,
stateOrder = unique(colorTable$stateName), interval = 3000)

#>      gene_ENS seqnames      start      end  width strand
#> 1 ENSMUSG00000000290.13 chr10 77530252 77565708 35457 +
#> 2 ENSMUSG00000000295.13 chr10 31313383 31328204 14822 +
#> 3 ENSMUSG00000000296.8 chr10 31332376 31445958 113583 -
#> 4 ENSMUSG00000000303.12 chr8 106603351 106670246 66896 +
#> 5 ENSMUSG00000000374.8 chr10 78186725 78244641 57917 -
#> 6 ENSMUSG00000000711.3 chr10 128677175 128696264 19090 -
#>  gene_expression      TSS TSS_moins_3kb TSS_plus_3kb TSSA  TSSFlnk TSSFlnkD
#> 1 1.0547488 77530252 77527252 77533252 0 0.0000000 0
#> 2 0.3407228 31313383 31310383 31316383 0 0.0000000 0
#> 3 0.4008302 31445958 31442958 31448958 0 0.0000000 0
#> 4 0.2292526 106603351 106600351 106606351 0 0.0000000 0
#> 5 29.3456974 78244641 78241641 78247641 0 0.2333333 0
#> 6 41.4675071 128696264 128693264 128699264 0 0.0000000 0
#>      Tx TxWk      EnhG EnhA      EnhWk ZNF.Rpts Het      TssBiv      EnhBiv
#> 1 0.0000000 0 0.0000000 0 0.0000000 0 0 0 0.0000000 0.5753333
#> 2 0.0000000 0 0.0000000 0 0.0000000 0 0 0 0.1333333 0.4000000
#> 3 0.0000000 0 0.0000000 0 0.0000000 0 0 0 0.1666667 0.3666667
#> 4 0.0000000 0 0.0000000 0 0.0000000 0 0 0 0.3000000 0.1000000
#> 5 0.3068333 0 0.2265000 0 0.0333333 0 0 0 0.2000000 0.0000000
#> 6 0.0000000 0 0.7333333 0 0.0000000 0 0 0 0.2666667 0.0000000
#>      ReprPC ReprPCWk      Quies
#> 1 0.3246667 0 0.1000000
#> 2 0.0000000 0 0.4666667
#> 3 0.0000000 0 0.4666667
#> 4 0.3585000 0 0.2415000
#> 5 0.0000000 0 0.0000000
#> 6 0.0000000 0 0.0000000
```

Predominant chromatin state at gene promoters

`predominantState()` estimates the predominant chromatin state at gene promoter, which corresponds to the state with the largest overlap with the gene promoter environment. Genes are then clustered according to their chromatin state using `umap` package. The output contains information on the predominant chromatin state and the corresponding UMAP dimension.

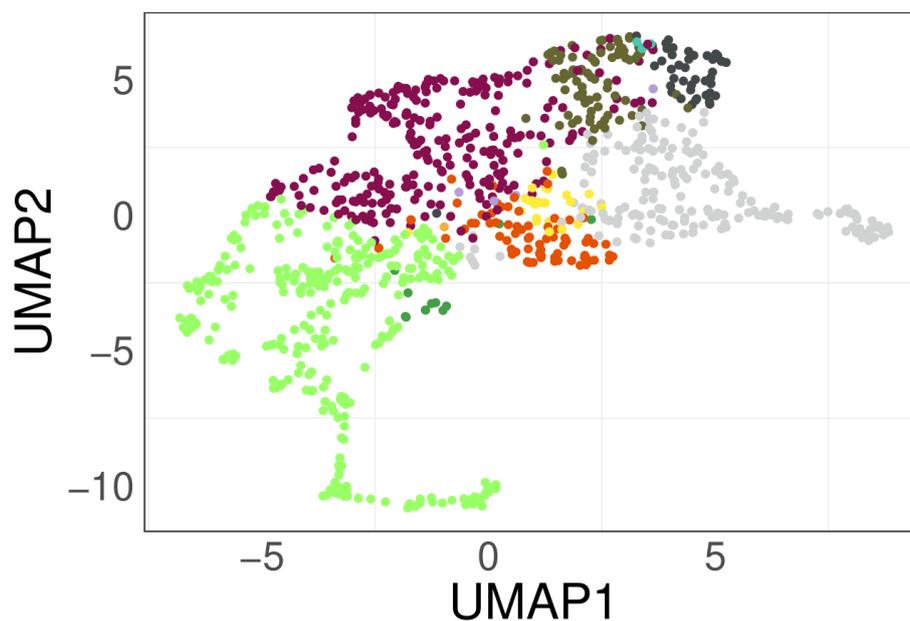
```
result_umap = predominantState(table_overlapping, state = unique(colorTable$stateName),
header = unique(colorTable$stateName), neighbors = 32, metric = "euclidean", dist = 0.5)
#> ==> It will be take few minutes to process
```

```
#>  TSSA  TSSFlnk TSSFlnkD      Tx TxWk      EnhG EnhA      EnhWk ZNF.Rpts Het
#> 1  0 0.0000000      0 0.0000000  0 0.0000000  0 0.0000000  0 0.0000000  0 0
#> 2  0 0.0000000      0 0.0000000  0 0.0000000  0 0.0000000  0 0.0000000  0 0
#> 3  0 0.0000000      0 0.0000000  0 0.0000000  0 0.0000000  0 0.0000000  0 0
#> 4  0 0.0000000      0 0.0000000  0 0.0000000  0 0.0000000  0 0.0000000  0 0
#> 5  0 0.2333333      0 0.3068333  0 0.2265000  0 0.0333333  0 0.0333333  0 0
#> 6  0 0.0000000      0 0.0000000  0 0.7333333  0 0.0000000  0 0.0000000  0 0
#>  TssBiv  EnhBiv  ReprPC ReprPCWk  Quies  UMAP1  UMAP2  state
#> 1 0.0000000 0.5753333 0.3246667      0 0.1000000 3.023172 5.811569 EnhBiv
#> 2 0.1333333 0.4000000 0.0000000      0 0.4666667 3.876110 2.703313 Quies
#> 3 0.1666667 0.3666667 0.0000000      0 0.4666667 3.824936 2.450091 Quies
#> 4 0.3000000 0.1000000 0.3585000      0 0.2415000 4.043915 4.546622 ReprPC
#> 5 0.2000000 0.0000000 0.0000000      0 0.0000000 -2.067456 -2.039908 Tx
#> 6 0.2666667 0.0000000 0.0000000      0 0.0000000 -6.453034 -4.011411 EnhG
```

Below is an example of UMAP representation to visualize the predominant chromatin state in each gene. Each dot corresponds to a gene and is colored according to its predominant chromatin state. The resulting figure may not be exactly the same than the one presented in this thumbnail since the order of display?? of dimension axes may differ; however, the clusters remain the same.

Here is an example of code to generate the figure below:

```
library(ggplot2)
ggplot(result_umap,aes(UMAP1,UMAP2, color = factor(state,
levels = unique(colorTable$stateName)))) +
  geom_point() +
  scale_color_manual(values = getStateColor(colorTable)$stateName) +
  theme_bw() + theme(strip.background = element_blank(),
text = element_text(size=25, angle = 0),
panel.grid.major = element_blank(),
axis.ticks = element_blank(),
strip.text.x = element_text(size = 25, angle = 0, hjust = 1),
legend.position = "none")
```



Session Information

Here is the output of `sessionInfo()` on the system on which this document was compiled:

```
#> R version 4.1.3 (2022-03-10)
#> Platform: x86_64-conda-linux-gnu (64-bit)
#> Running under: Ubuntu 18.04.6 LTS
#>
#> Matrix products: default
#> BLAS/LAPACK: /home/mcoulee/anaconda3/envs/R_package_3/lib/libopenblas-r0.3.20.so
#>
#> locale:
#>  [1] LC_CTYPE=fr_FR.UTF-8      LC_NUMERIC=C
#>  [3] LC_TIME=fr_FR.UTF-8      LC_COLLATE=fr_FR.UTF-8
#>  [5] LC_MONETARY=fr_FR.UTF-8  LC_MESSAGES=fr_FR.UTF-8
#>  [7] LC_PAPER=fr_FR.UTF-8     LC_NAME=C
#>  [9] LC_ADDRESS=C             LC_TELEPHONE=C
#> [11] LC_MEASUREMENT=fr_FR.UTF-8 LC_IDENTIFICATION=C
#>
#> attached base packages:
#> [1] stats      graphics  grDevices  utils      datasets  methods   base
#>
#> other attached packages:
#> [1] ggplot2_3.3.6      ChromENVEE_0.99.8
#>
```

```

#> loaded via a namespace (and not attached):
#> [1] Rcpp_1.0.9          lattice_0.20-45      png_0.1-7
#> [4] prettyunits_1.1.1    ps_1.7.1            assertthat_0.2.1
#> [7] rprojroot_2.0.3      digest_0.6.29       utf8_1.2.2
#> [10] RSpectra_0.16-1     R6_2.5.1            GenomeInfoDb_1.30.1
#> [13] stats4_4.1.3        evaluate_0.15       highr_0.9
#> [16] pillar_1.8.1         zlibbioc_1.40.0     rlang_1.0.5
#> [19] callr_3.7.1         S4Vectors_0.32.4   Matrix_1.4-1
#> [22] reticulate_1.26     rmarkdown_2.14     labeling_0.4.2
#> [25] splines_4.1.3       devtools_2.4.3     stringr_1.4.1
#> [28] RCurl_1.98-1.8      munsell_0.5.0      umap_0.2.9.0
#> [31] compiler_4.1.3      xfun_0.31          askpass_1.1
#> [34] pkgconfig_2.0.3     BiocGenerics_0.40.0 pkgbuild_1.3.1
#> [37] mgcv_1.8-40         htmltools_0.5.3    openssl_2.0.3
#> [40] tidyselect_1.1.2    tibble_3.1.8       GenomeInfoDbData_1.2.7
#> [43] IRanges_2.28.0      fansi_1.0.3        crayon_1.5.1
#> [46] dplyr_1.0.9         withr_2.5.0        bitops_1.0-7
#> [49] grid_4.1.3          nlme_3.1-158       jsonlite_1.8.0
#> [52] gtable_0.3.1        lifecycle_1.0.2    DBI_1.1.3
#> [55] magrittr_2.0.3      scales_1.2.1       cli_3.4.0
#> [58] stringi_1.7.8       cachem_1.0.6       farver_2.1.1
#> [61] XVector_0.34.0      fs_1.5.2           remotes_2.4.2
#> [64] ellipsis_0.3.2     vctrs_0.4.1        generics_0.1.3
#> [67] tools_4.1.3         glue_1.6.2         purrr_0.3.4
#> [70] processx_3.7.0      pkgload_1.3.0     parallel_4.1.3
#> [73] fastmap_1.1.0       yaml_2.3.5         colorspace_2.0-3
#> [76] GenomicRanges_1.46.1 sessioninfo_1.2.2  memoise_2.0.1
#> [79] knitr_1.39          usethis_2.1.6

```

References

- Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*, 9:215-216, 2012
- Ferrari, F. et al. DOT1L-mediated murine neuronal differentiation associates with H3K79me2 accumulation and preserves SOX2-enhancer accessibility. *Nat. Commun.* 11, 5200 (2020)
- Godfrey, L. et al. DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat. Commun.* 10, 1–15 (2019).
- Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). “Software for Computing and Annotating Genomic Ranges.” *PLoS Computational Biology*, 9. doi: 10.1371/journal.pcbi.1003118, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- McInnes, Leland, and John Healy. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.” arXiv:1802.03426.

Figures supplémentaires – Article 1

Mélina Blanco et al

EMBO reports

Expanded View Figures

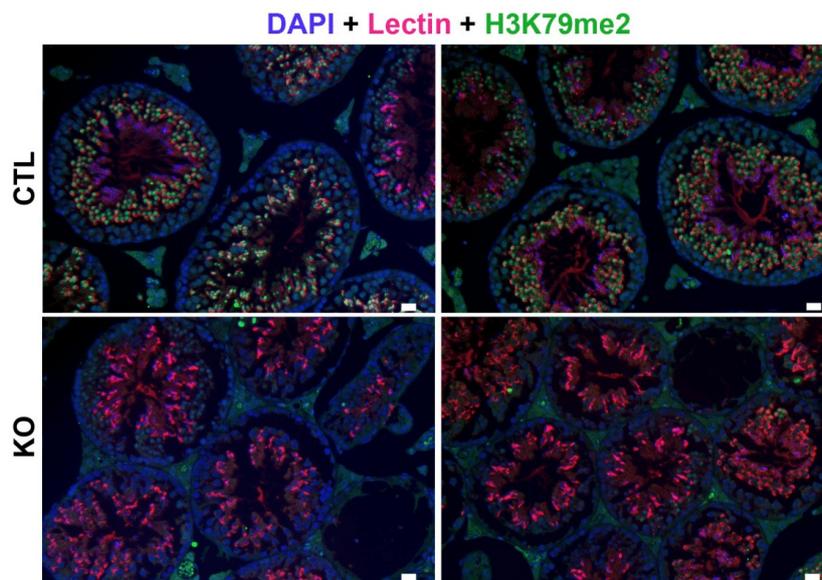


Figure EV1. Immunofluorescence detection of H3K79me2 on testicular sections from adult CTL or *Dot1l*-KO mice.

H3K79me2 signal is shown in green. DAPI (blue) was used to stain nuclei. Lectin (pink) was used to stain acrosome. Scale bar indicates 20 μ m.

Figure EV2. Detailed analyses of CTL and *Dot1l*-KO spermatozoa.

- A Scatter plots showing the quantification of sperm abnormalities (mean \pm SEM; $N = 4-5$ and four biological replicates for CTL and KO, respectively). Stars indicate significant differences obtained using t-tests after angular transformation of the percentages (* P -value < 0.05 , ** P -value < 0.005 , and *** P -value < 0.0005).
- B Motility parameters from *Dot1l*-KO and CTL spermatozoa following CASA (computer-assisted sperm analyses) (mean \pm SEM; $N = 12$ and 10 biological replicates for CTL and KO, respectively). Specifically, *Dot1l*-KO spermatozoa are slower and more static. They have a more linear movement (LIN) and a decreased straightness (STR), decreased lateral displacements of their heads (ALH) with a lower curvilinear velocity (VCL) resulting in a straight-line velocity (VSL) similar to CTL spermatozoa but a higher Average Path Velocity (VAP). Stars indicate significant differences obtained using Mann-Whitney t -tests after angular transformation of the percentages (** P -value < 0.01 , *** P -value < 0.001 , **** P -value < 0.0001 , ns' indicates non-significance). The Beat Cross Frequency (BCF) of CTL and *Dot1l*-KO spermatozoa is similar.

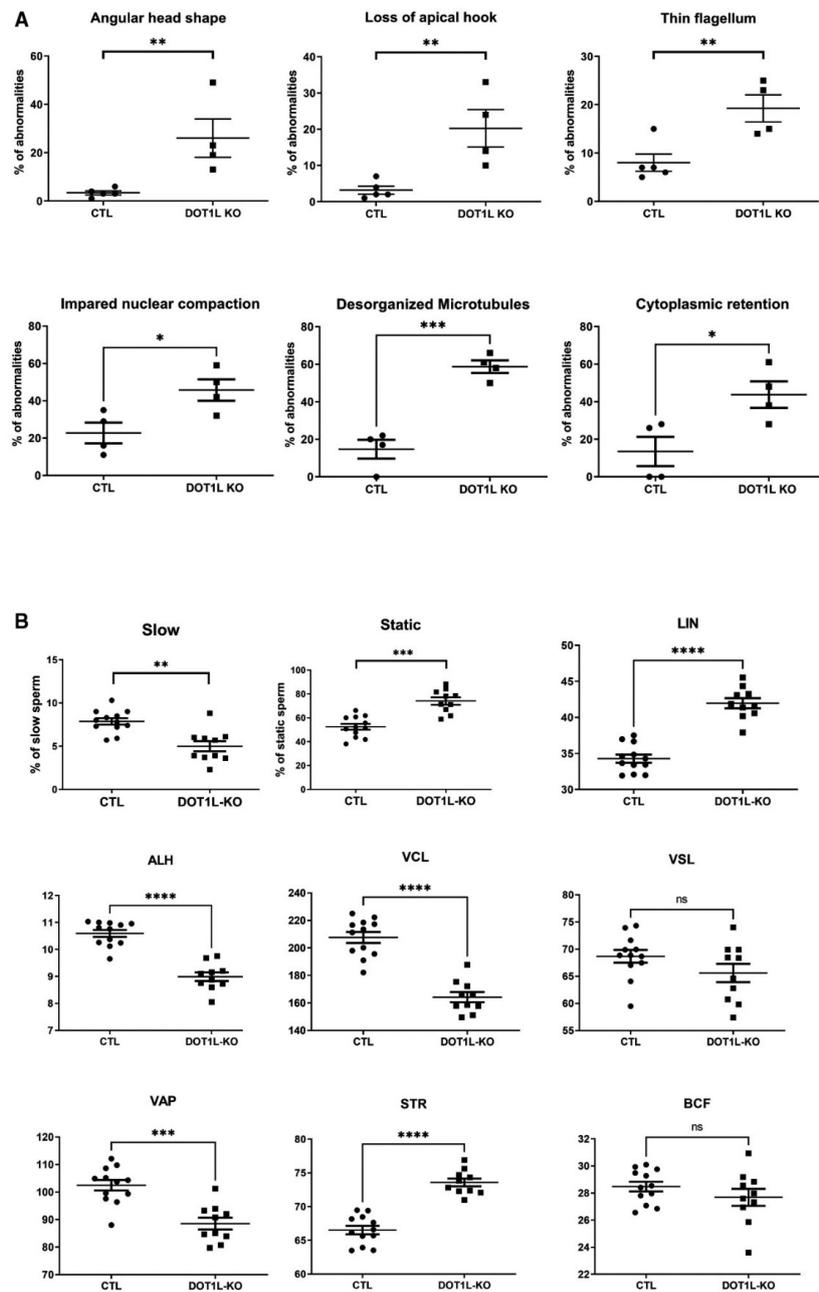


Figure EV2.

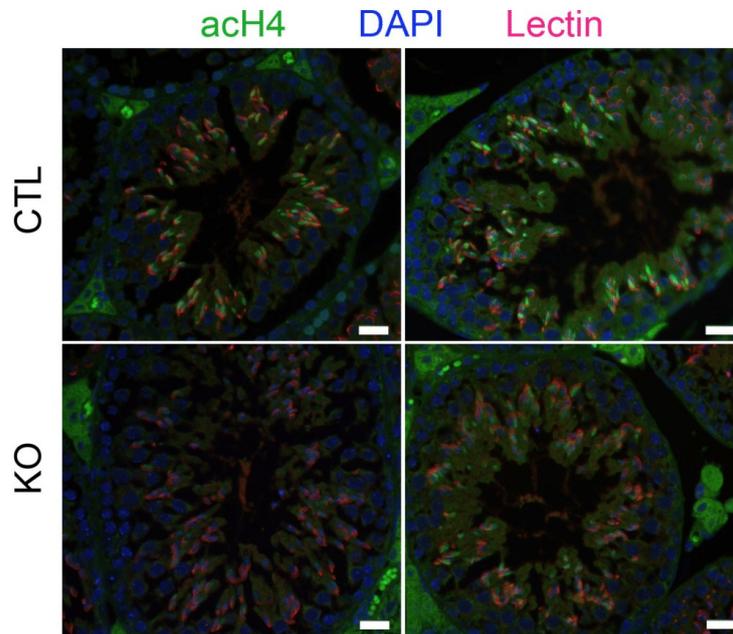


Figure EV3. Confirmation of a decrease in H4 acetylation in elongating spermatids (ES) by immunofluorescence.

Representative pictures of immunofluorescence detection of acH4 (anti-poly-H4ac, green) on stage XI testicular sections. At this stage, acH4 level is strong in CTL elongating spermatids but appears weaker in *Dot1L*-KO. DAPI (blue) was used to stain nuclei and lectin (red) to stage the acrosome and facilitates tubule staging. Scale bar indicates 20 μ m.

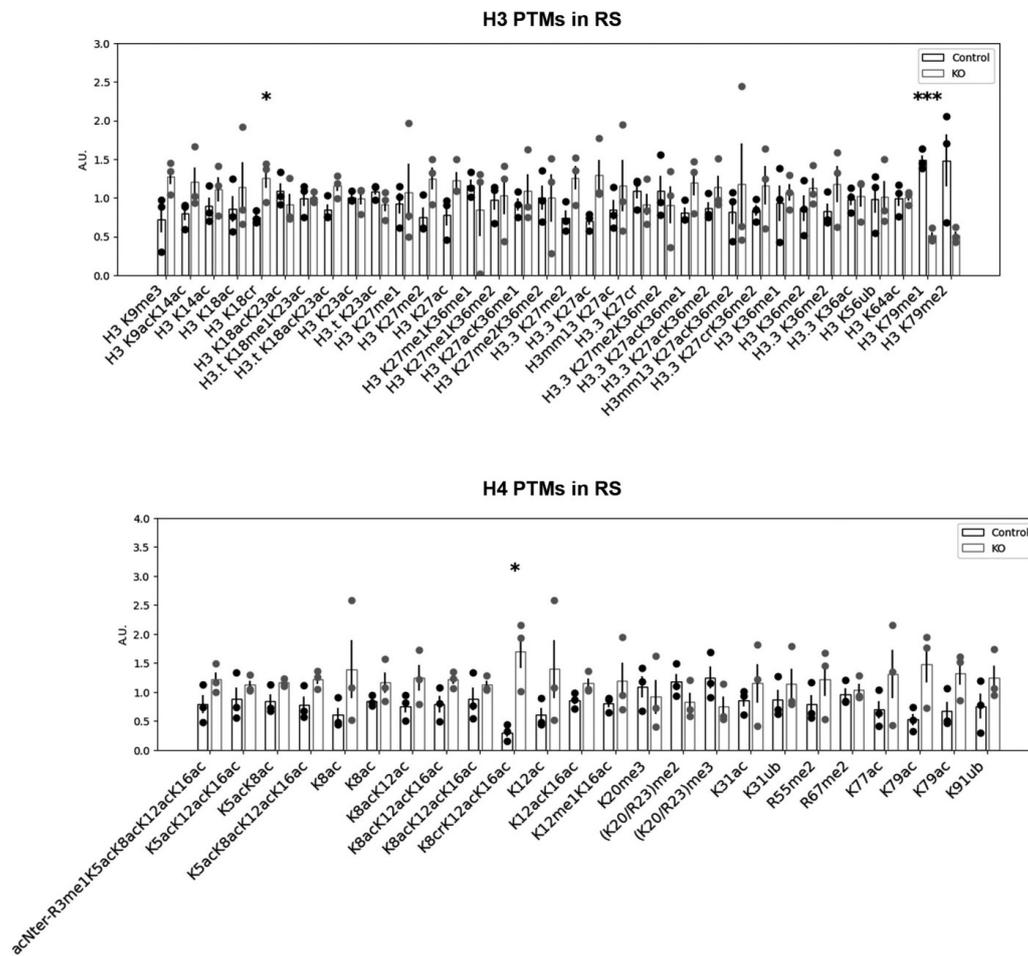


Figure EV4. Quantification of histone H3 and H4 PTMs in *Dot1L*-KO and CTL round spermatids.
 LC-MS/MS data acquired on H3 and H4 were processed as described in Fig 4. Quantitative data obtained on biological replicates ($N = 3$ for CTL and 3 for KO) were plotted as dots, and the height of the bars indicates the average value (mean \pm SD). A.U. = arbitrary units. * P -value < 0.05 , ** P -value < 0.005 , and *** P -value < 0.0005 , obtained with t -tests.

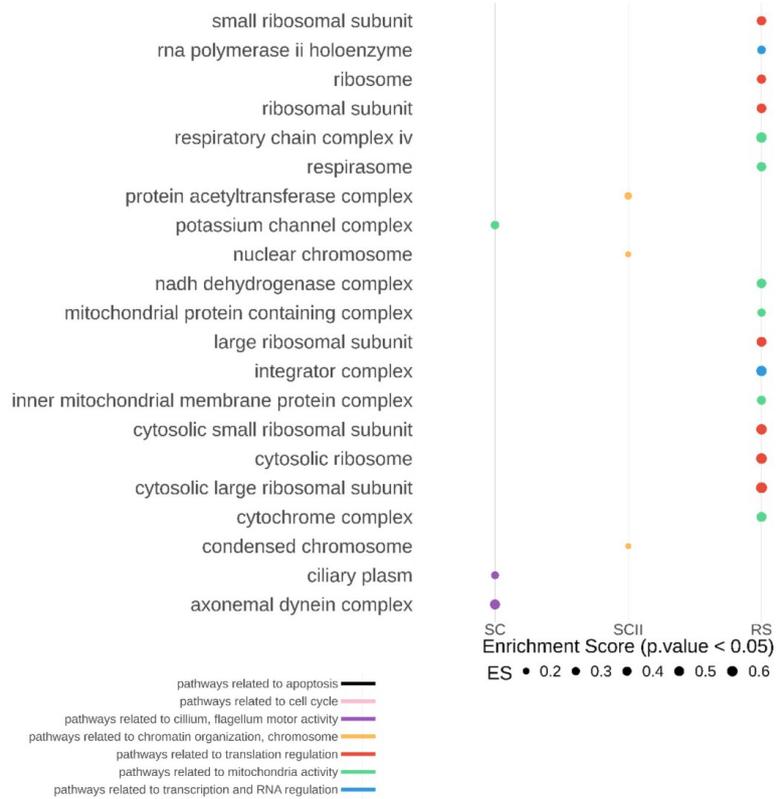


Figure EV5. GSEA analysis of RNA-seq data from *Dot1l*-KO versus CTL SC, SCII, and RS.
The figure shows all the “cellular components” found significantly downregulated in *Dot1l*-KO primary spermatocytes (SC), secondary spermatocytes (SCII) and round spermatids ($P < 0.05$), ranked by their enrichment score (ES).

Blanco et al. Appendix

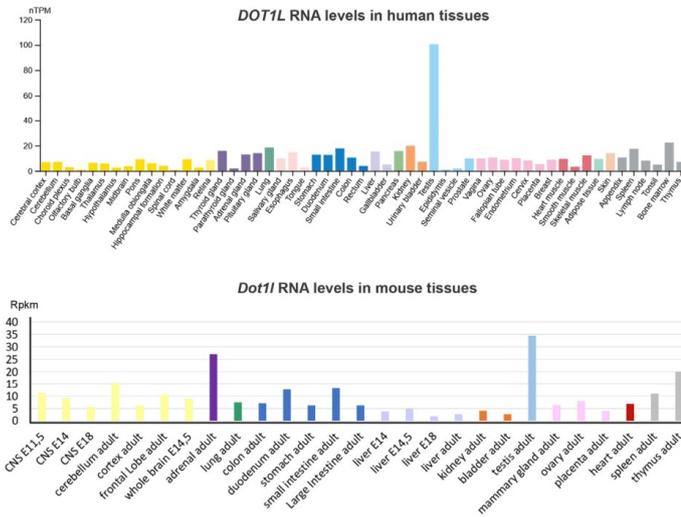
Table of contents

Appendix Figure S1.....	2
Appendix Figure S2.....	8
Appendix Figure S3.....	10
Appendix Figure S4.....	13
Appendix Figure S5.....	18
Appendix Table S1.....	21
Appendix Table S2.....	22
References.....	23

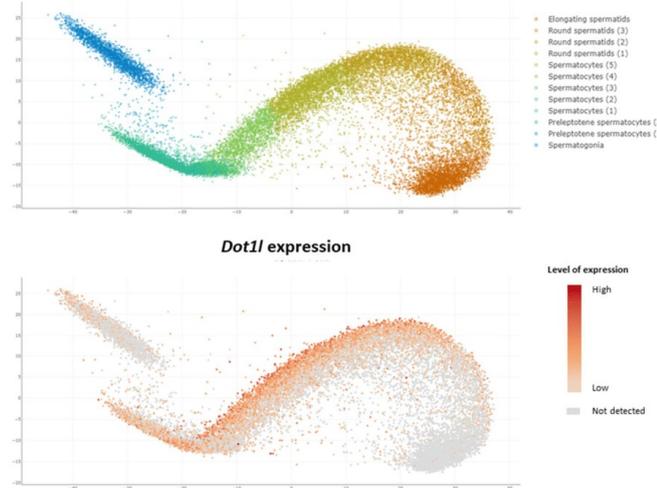
Blanco et al. Appendix

Appendix Figure S1

1A



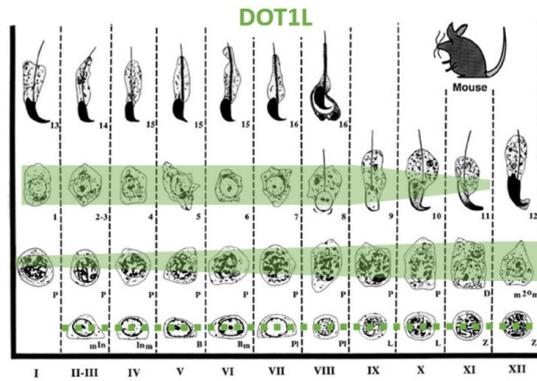
1B



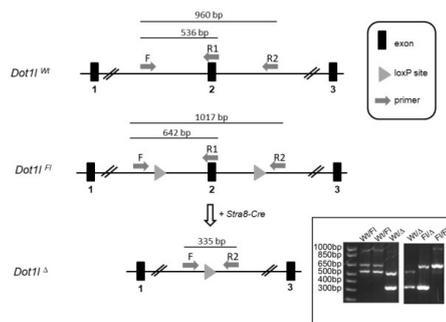
1469178, 2023, 6, Downloaded from https://www.onlinelibrary.wiley.com/doi/10.1111/1365-3113.12088 by Institut National de la Santé et de la Recherche Médicale, Wiley Online Library on [22/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Blanco et al. Appendix

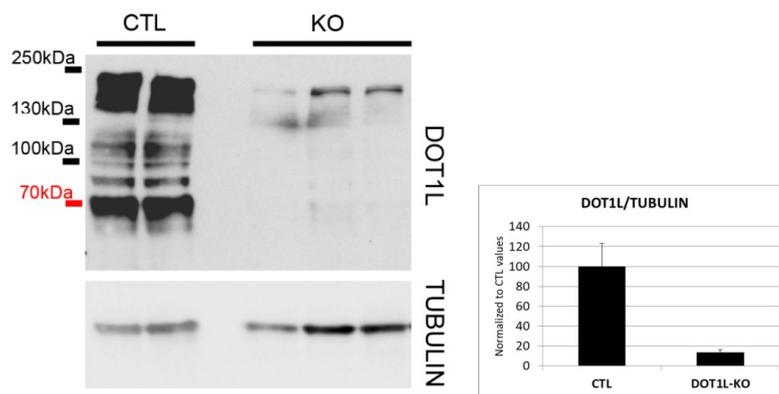
1C



1D



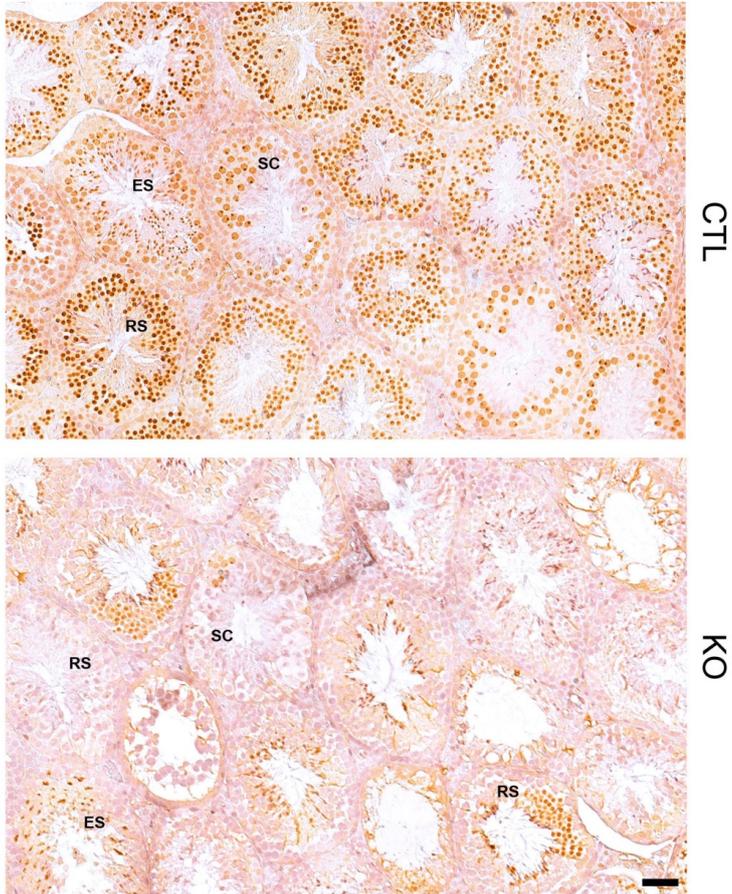
1E



3

Blanco et al. Appendix

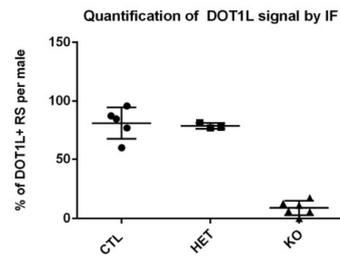
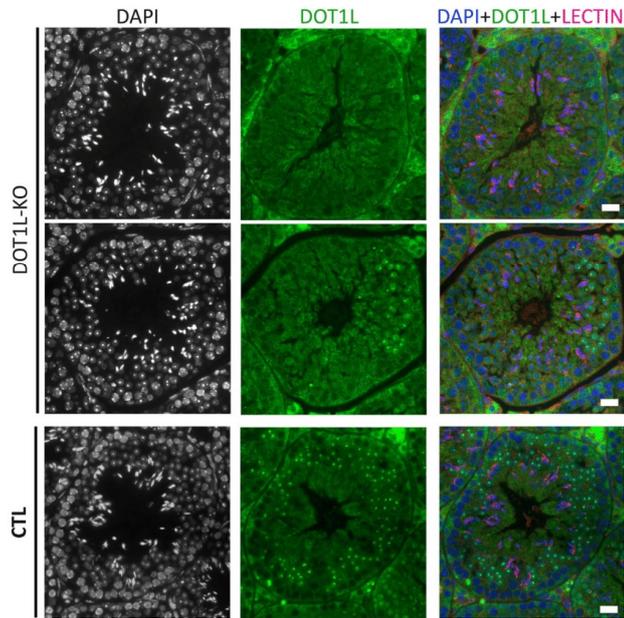
1F



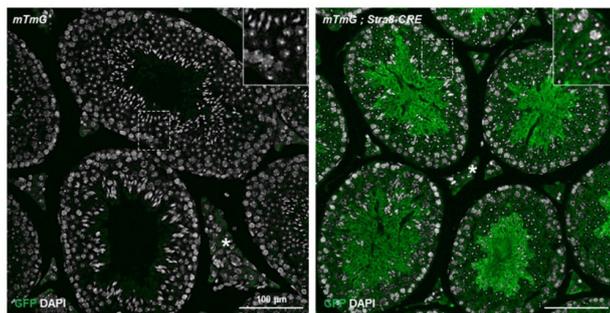
1469178, 2023, 6, Downloaded from https://www.onlinelibrary.wiley.com/doi/10.1111/j.1365-2214.2022.56161 by Insem Dsc, Inc, Wiley Online Library on [22/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Blanco et al. Appendix

1G

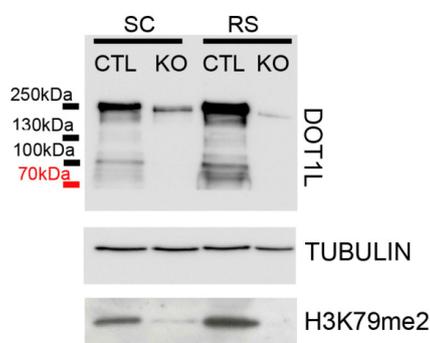


1H



Blanco et al. Appendix

11



Appendix Figure S1. Complement to Figure 1. A) RNA expression overview of *DOT1L* RNA in human and mouse tissues. Data were obtained from The Human Protein Atlas [<https://www.proteinatlas.org/>; consensus dataset, consisting of normalized expression (nTPM) levels for 55 tissue types, created by combining the HPA and GTEx transcriptomics datasets] or from mouse tissues from mouse ENCODE transcriptome data (Data ref: Mouse ENCODE transcriptome data, NCBI Sequence Read Archive PRJNA66167, 2011). **B) Expression of *Dot1l* mRNA based on single cell RNA-seq analyses from adult testicular cells** (Data ref: (Green *et al*, 2018) analyzed using <https://rgv.genouest.org/> (Darde *et al*, 2019; Darde *et al*, 2015)). **C) Summary of DOT1L protein expression in mouse germ cells based on immunofluorescence detection on testicular sections.** Scheme modified after Russell *et al*. (Russell *et al*, 1990). DOT1L protein is detected in all germ cells but is particularly abundant in late spermatocytes and in round spermatids. **D) Generation of *Dot1l* KO allele.** Drawing representing the beginning of mouse *Dot1l* gene (from exon 1 to exon 3), the location of loxP sites and genotyping primers. Bottom right insert shows typical pictures of products from PCR using *Dot1l* genotyping primers (F+R1+R2) following run on a 2% agarose gel. **E) Estimation of *Dot1l* knockout efficiency by western blot (uncropped western blot image of Figure 1B).** Detection of DOT1L and TUBULIN in whole testicular protein extracts from CTL and *Dot1l*-KO mice. The most abundant and high molecular weight bands are DOT1L testis-specific (~185kDa) and canonical proteoform. Additional bands most likely corresponding to smaller, less abundant proteoforms (such as Q679P5 at ~122kDa or Q6XZL7 at ~68kDa) are also visible in case of long exposure. All DOT1L proteoforms are markedly reduced in *Dot1l*-KO testicular extracts. On the right is shown a graphic representation of the quantification of DOT1L signal (all bands) normalized to TUBULIN signal. **F) Immunohistochemistry detection of DOT1L in testicular sections from CTL and *Dot1l*-KO (KO) adult mice counterstained with hematoxylin (complement panel to Fig 1C).** SC = Spermatocytes, RS = Round spermatids, ES = Elongating spermatids. Pictures were taken using the same parameters. Scale bars indicate 50 μ m. **G) Estimation of *Dot1l* knockout efficiency by immunofluorescence.** Representative pictures of immunofluorescence detection of DOT1L (Abcam 64077 antibody, in green) on testicular sections from adult CTL or KO mice. DAPI (grey or blue) was used to stain nuclei; Lectin (pink) was used to stain acrosome. Scale bar indicates 20 μ m. In CTL testicular sections, tubules containing round spermatids are uniformly stained (bottom panel). In *Dot1l*-KO sections, the majority of tubules do not show any DOT1L protein signal (top panel) but in some tubules, DOT1L signal can still be visible (middle panel). Below the immunofluorescence pictures is shown a graphic representation of the percentage of round spermatids with a visible DOT1L signal. N=5 CTL, 3 HET and 6 KO males between 2.5-4 month old (6 to 10 tubules per individual). Quantification was performed on stages I to V tubules. **H) Verification of *Stra8-Cre* recombinase activity and specificity in adult mouse testis.** The Cre reporter mouse line *mTmG* carries a *GFP* transgene expressed upon Cre induction. In this transgene, GFP is targeted to the cell membrane (Muzumdar *et al*, 2007). In testicular sections from *mTmG* mice (left panel), using an anti-GFP antibody, no signal is detected, excepted for faint auto-fluorescence from Leydig cells (asterisk). In *mTmG* ; *Stra8-Cre* mice (right panel) the Cre recombinase is specifically expressed in germ cells and GFP (in green) is detected in most germ

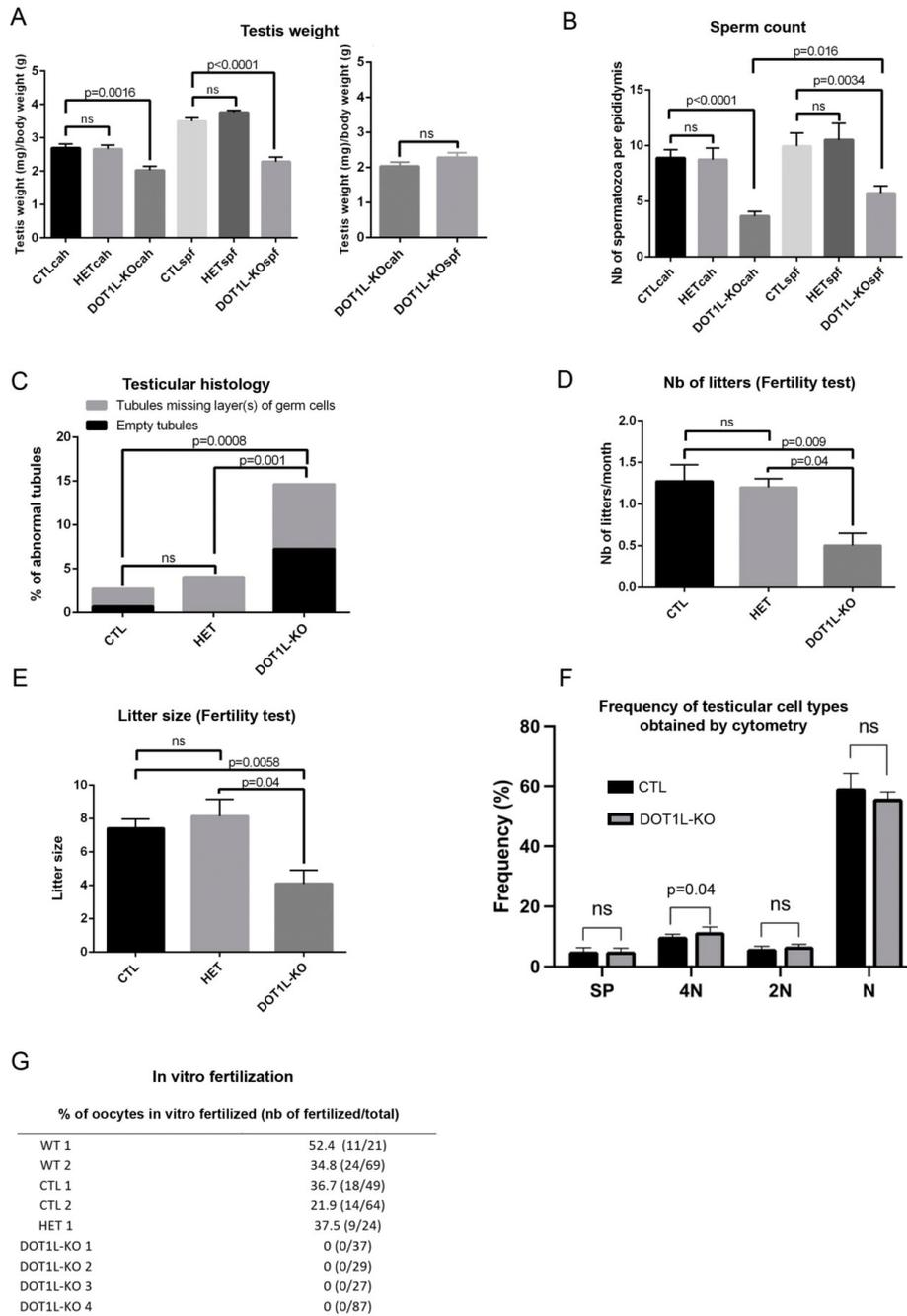
6

Blanco et al. Appendix

cells. DAPI (grey) has been used to stain cell nuclei. Top right panel represents a 2X magnification. **1) Uncropped western blot image of Fig 1D.** Note that the membrane was cut at ~70kDa and ~35kDa to generate three pieces which were incubated with anti-DOT1L antibody (top part), with anti-TUBULIN antibody (mid part) or with anti-H3K79me2 antibody (bottom part).

Blanco et al. Appendix

Appendix Figure S2



14691778, 2023, 6, Downloaded from https://www.embopress.org/doi/10.15252/emmm.20225616 by Inserm Dsc, Inc. Wiley Online Library on [22/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

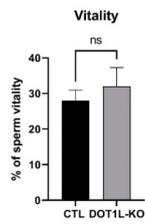
Blanco et al. Appendix

Appendix Figure S2. Complement to Figure 2. Detailed analyses of reproductive parameters from *Dot1l*-KO males and CTL. ‘cah’ indicates animals produced in a conventional animal house. ‘spf’ indicates animals produced in the specific pathogen-free animal house. HET are heterozygous animals of the following genotype *Dot1l*^{F/ Δ} . Obtained p-values (following unpaired t-tests) are indicated above; “ns” indicates non-significant difference. **A)** Average testis weight (normalized to body weight, mean value \pm SEM) in different groups of males (N=9 for CTLcah, N=6 for HETcah, N= 9 for DOT1Lcah, N=12 for CTLspf, N=2 for HETspf and N= 17 for DOT1Lspf). **B)** Average number of epididymal spermatozoa (mean value \pm SEM) in different groups of males (N=12 for CTLcah, N=6 for HETcah, N=12 for DOT1Lcah, N=9 for CTLspf, N=2 for HETspf and N= 13 for DOT1Lspf). **C)** Graphic representation of the percentage of abnormal tubules in *Dot1l*-KO (N=5) CTL (N=6) and HET (N=4) testes (from both cah and spf). Obtained p-values (following unpaired t-tests on empty tubules values) are indicated above. **D)** and **E)** Results from the tests of fertility (natural mating in spf with wild type females for \sim 3 months) of *Dot1l*-KO, HET and CTL males (mean \pm SEM; N=7 for KO, N= 2 for HET and N=6 for CTL). **F)** Schematic diagram representing the frequency of each cell type in CTL and *Dot1l*-KO testes, as calculated by FACS (mean \pm SEM; N= for CTL and 3 for KO). 4N = Primary spermatocytes, 2N = Secondary spermatocytes, N = spermatids, SP = side population representing premeiotic germ cells (i.e. spermatogonia). **G)** Results from *in vitro* fertilization assays. The percentage of fertilized oocytes is indicated for each male. Values obtained for *Dot1l*-KO spermatozoa are significantly different from WT or CTL spermatozoa (chi-square, $p < 0.01$).

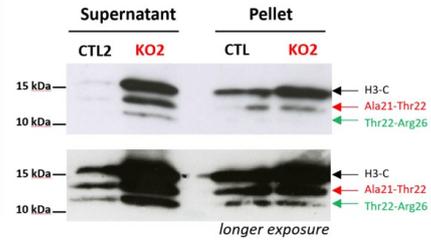
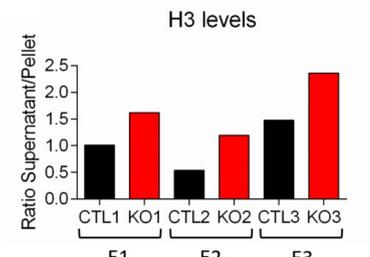
Blanco et al. Appendix

Appendix Figure S3

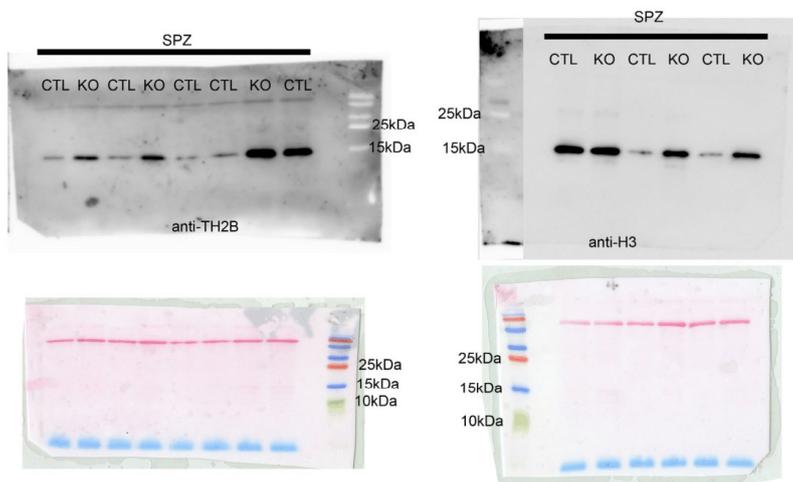
3A



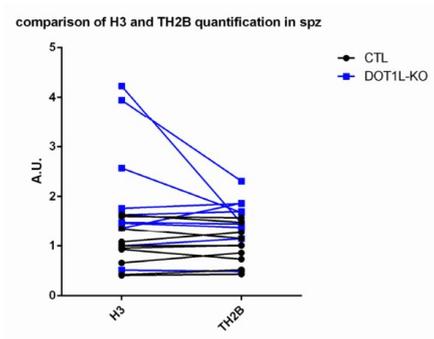
3B



3C

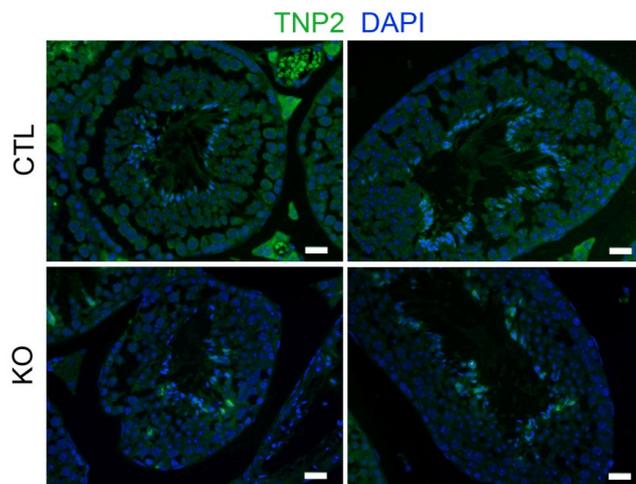
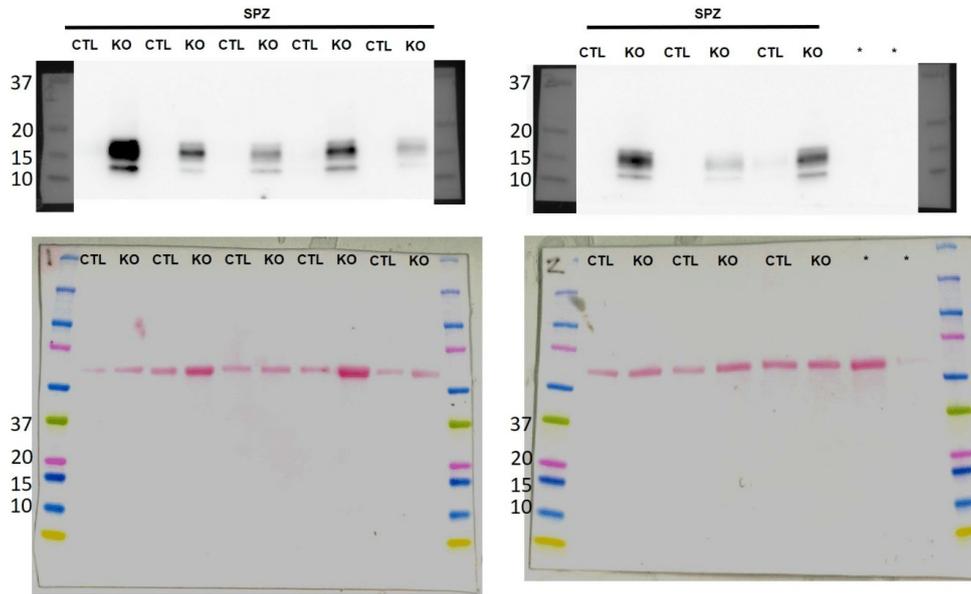


3D



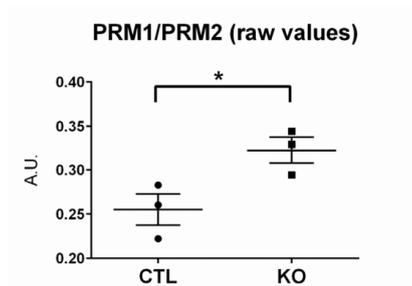
Blanco et al. Appendix

3E



Blanco et al. Appendix

3F

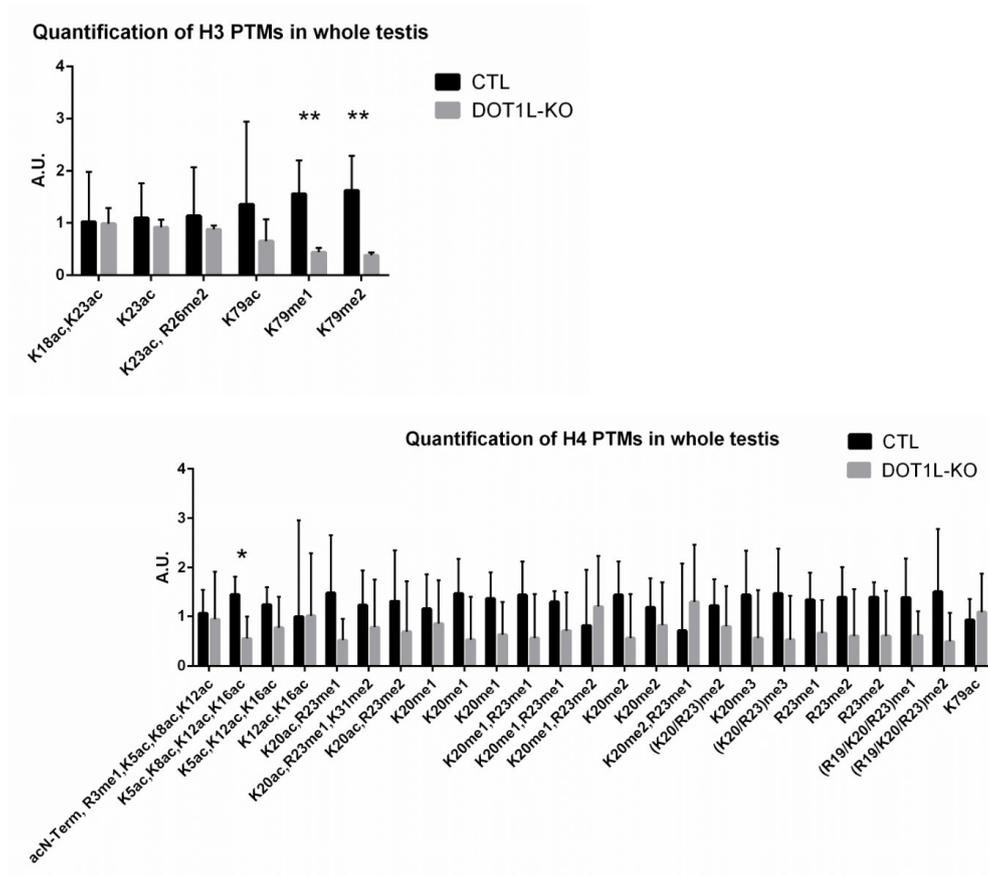


Appendix Figure S3. Complement to Figure 3. A) Sperm vitality. Bar graph showing sperm vitality measured by eosin-nigrosine staining in which dead spermatozoa appear violet (mean \pm SEM). Sperm samples from 4 CTL and 4 *Dot1l*-KO males were analyzed. NS indicates non-significant difference between the two genotypes (t-test). **B) *Dot1l*-KO sperm chromatin is sensitive to nucleoplasmin-induced decompaction.** Quantification of histone H3 levels following NPM (nucleoplasmin) partial decondensation of CTL and *Dot1l*-KO (KO) epididymal spermatozoa. Quantification was performed on the two fractions obtained after centrifugation: (i) the supernatant which contains histones from the NPM-de-compacted chromatin, and ii) the pellet which contains compact chromatin, i.e. “resistant” to the NPM treatment we used. Left panel: Bar graph showing quantification results of supernatant/pellet levels in 3 independent experiments (E1, E2 and E3, which are biological replicates). Right panel: Images of western blot using anti-histone H3 antibody (PanH3), corresponding to CTL2 and KO2 (the bottom image was used for quantification). PanH3 is directed against the C-terminal part of histone H3 and detects full length histone H3 at 17 kDa (black arrow, H3-C) and its two cleaved forms previously described in (Yamaguchi et al., 2018): one occurring between Ala21-Thr22 (red arrow) and the other one between Thr22-Arg26 (the smallest form at 13 kDa, green arrow). **C)** Uncropped western blot images from TH2B and H3 detections in spermatozoa (shown in Fig 3E), and corresponding Ponceau staining of the membranes. A non-specific band is visible in all lanes with Ponceau; it indicates that similar amounts of material were loaded in all lanes. **D)** Comparison of TH2B and H3 levels in the same CTL (black) and *Dot1l*-KO (blue) samples. The graph shows a good correlation between TH2B and H3 quantifications. **E)** Top panel: uncropped western blot images from TNP2 detection in spermatozoa (shown in Fig 3F), and corresponding Ponceau staining of the membranes. A non-specific band is visible in all lanes with Ponceau; it indicates that similar quantity of material was loaded in all lanes. Bottom panel: immunofluorescence detection of TNP2 (green) on stage VI-VIII testicular sections. At this stage, the chromatin of condensed spermatids is packaged with protamines, and TNP2 signal is not (or weak) in CTL. In *Dot1l*-KO, some condensed spermatids (inside of the tubule) have a stronger signal. DAPI (blue) was used to stain nuclei. **F)** Box plot of the raw PRM1/PRM2 ratio following acid urea gel electrophoresis (mean \pm SEM) complement to Fig 3G. A.U. = arbitrary units. *p-value < 0.05 (unpaired t-test).

Blanco et al. Appendix

Appendix Figure S4

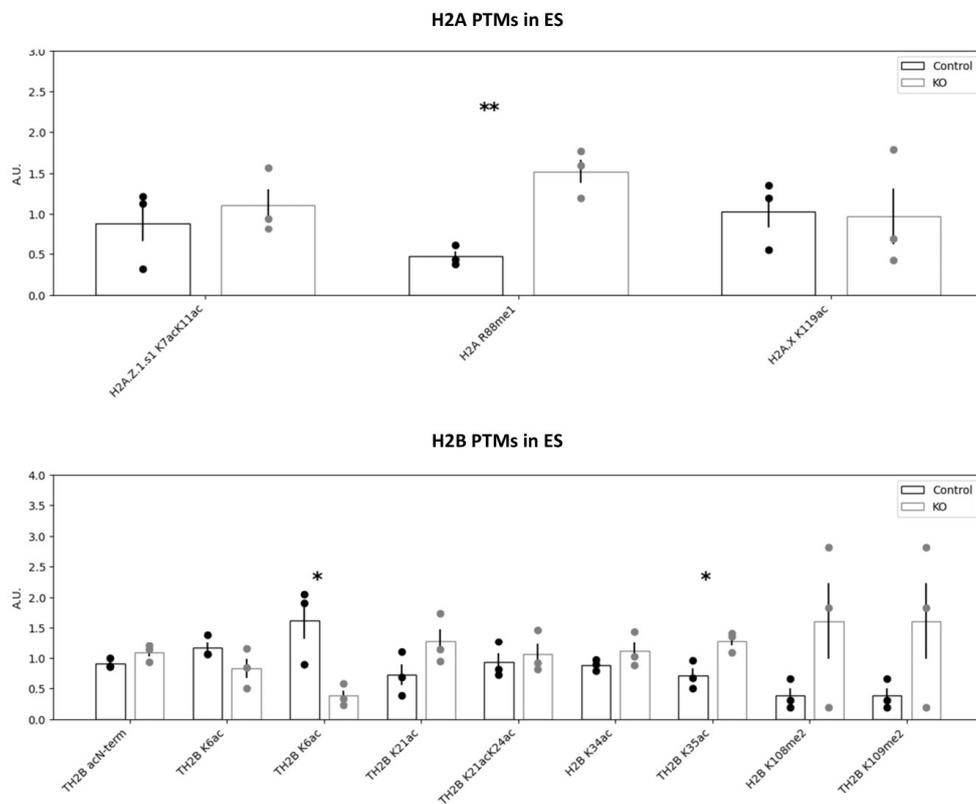
4A



1469178, 2023, 6, Downloaded from https://www.embopress.org/doi/10.15252/emmm.202256161 by Inserm Dsc, Inc. Wiley Online Library on [22/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Blanco et al. Appendix

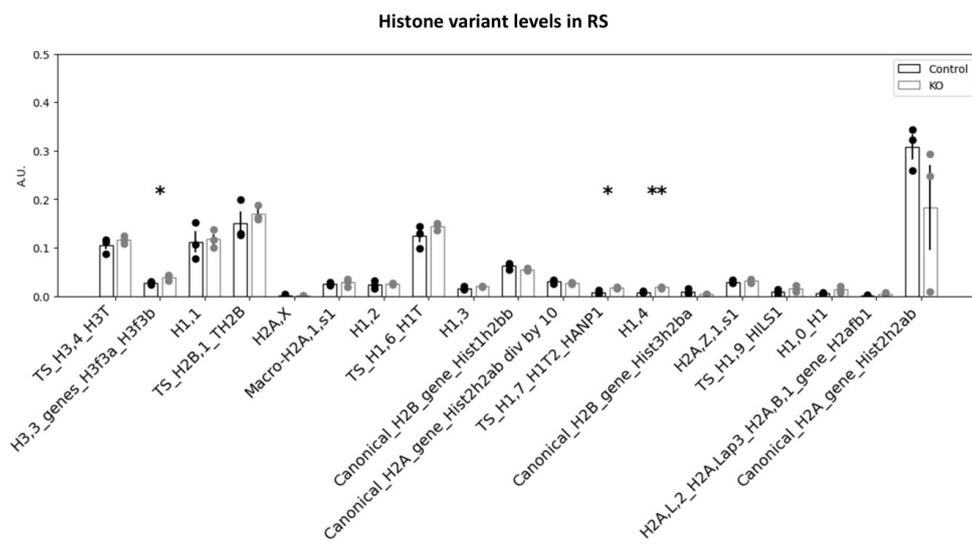
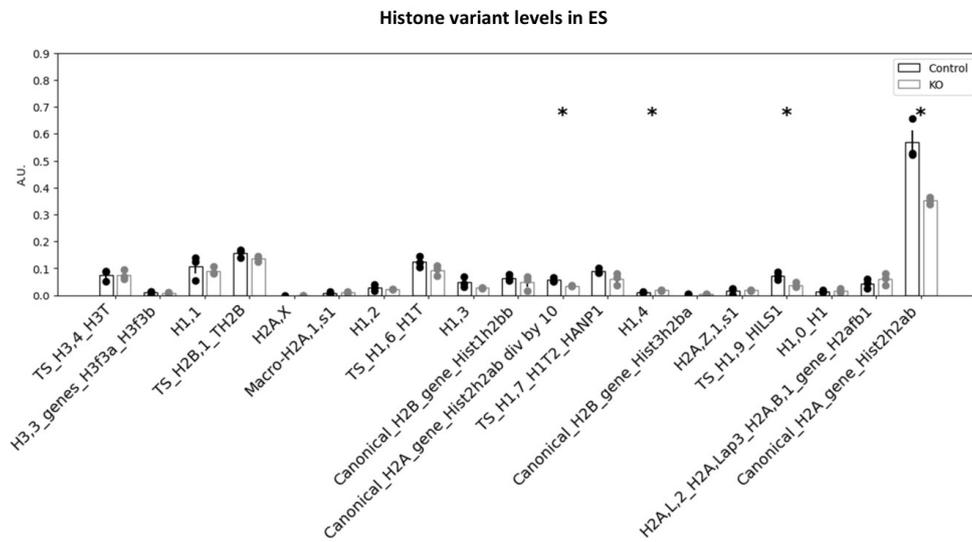
4B



1469178, 2023, 6, Downloaded from https://www.onlinelibrary.wiley.com/doi/10.1111/j.1365-2214.2023.03616 by Institut Dac, Inc. Wiley Online Library on [22/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

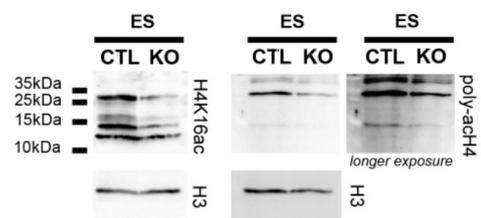
Blanco et al. Appendix

4C

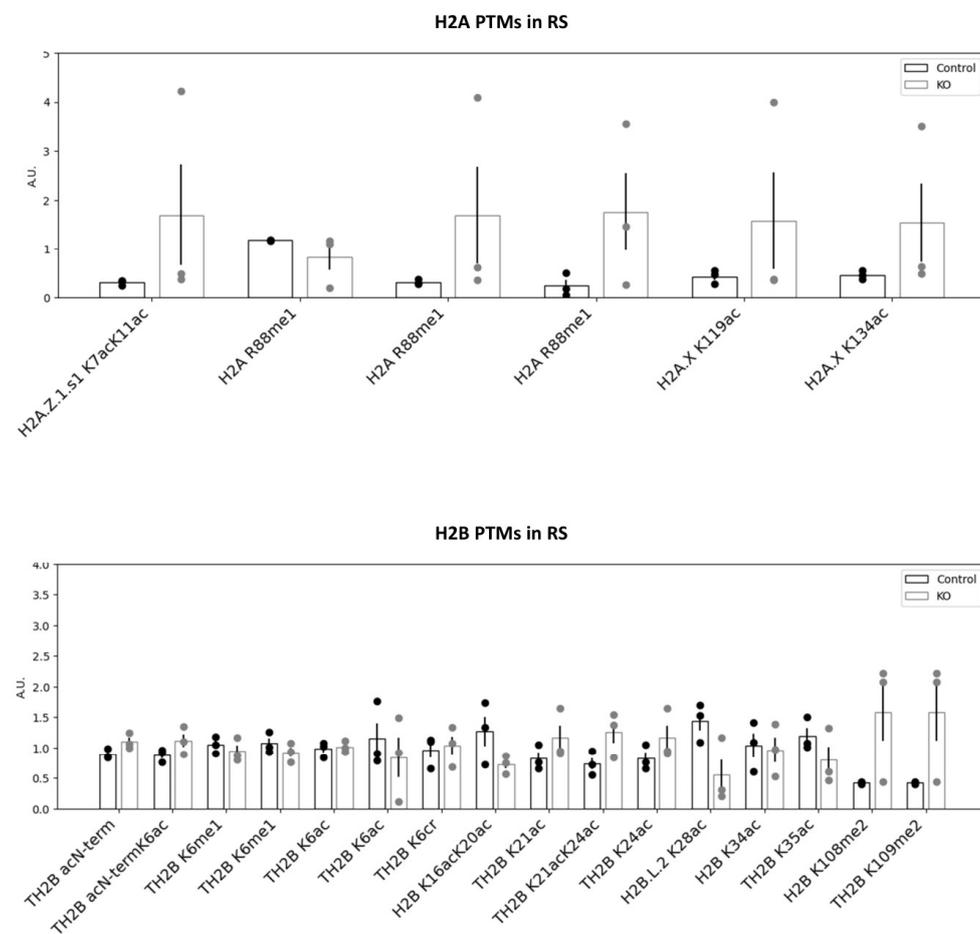


Blanco et al. Appendix

4D

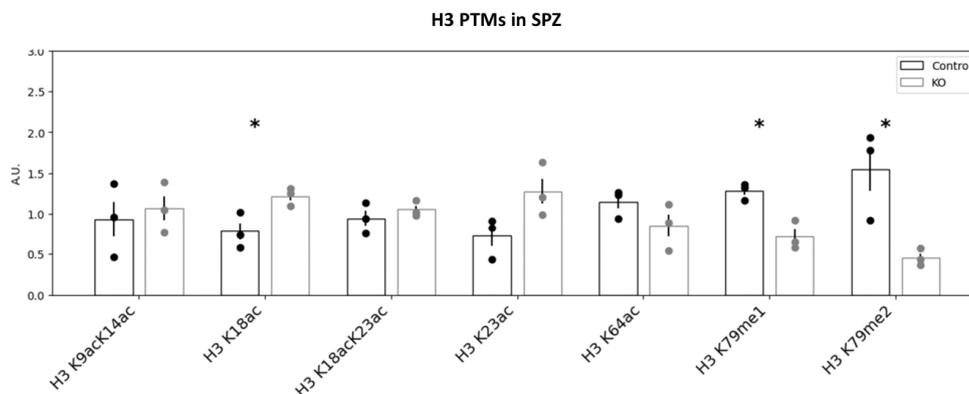


4E



Blanco et al. Appendix

4F



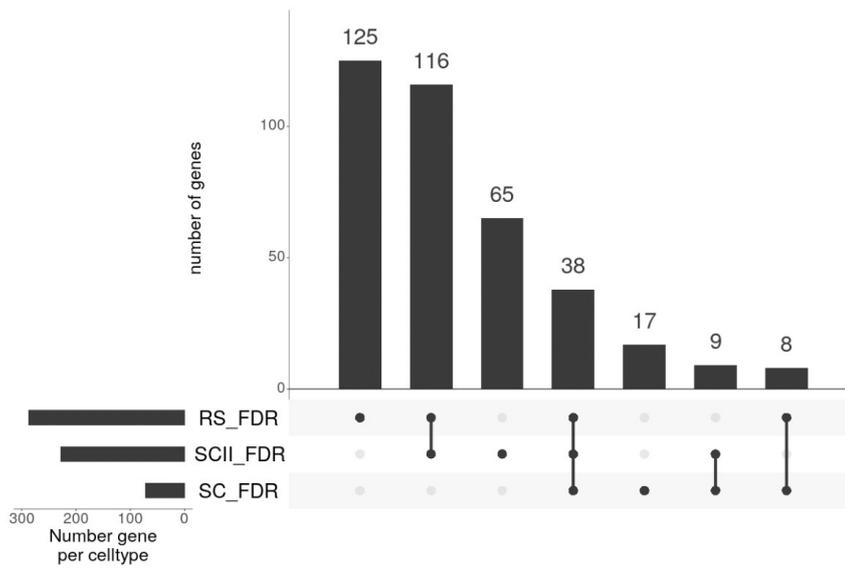
Appendix Figure S4. Complement to Figure 4. Quantification of histone PTMs and variants in *Dot1l*-KO and CTL samples. One star indicates a p-value <0.05, two stars a p-value <0.005 and three stars a p-value <0.0005, obtained with t-tests. **A) Mass spectrometry quantification of H3 and H4 PTMs in whole testes (N= 4 CTL and 4 KO).** After normalization to be at constant amount of histone H3 or H4 in each analyzed sample (see Material and methods), mass spectrometry signals were divided by the average signal in both conditions (CTL and *Dot1l*-KO), so as to be able to represent all peptides in the same figure, whatever their MS intensity. Error bars represent standard deviations calculated on the measurements made on biological replicates. Each modified peptide identified and quantified by MS analysis is defined as a list of PTMs separated by commas. Some modified sites were detected and quantified in several tryptic peptides due to possible missed cleavages (e.g. H4K20me1 was quantified in K20VLRDNIQGITK, K20VLRDNIQGITKPAIR and K20VLRDNIQGITKPAIRR); we indicated the measurements made on such peptides with the same label (K20me1 in the aforementioned example), which allows verifying the coherence of quantitative measurements. In a few instances, a precise localization of the modification site was not possible from the MS/MS spectra. This was typically the case between H4R19, K20 and R23 on which a methyl (me1) and dimethyl (me2) were detected. We indicated such uncertainties by (R19/K20/R23)me1 and (R19/K20/R23)me2, respectively. **B) Mass spectrometry quantification of H2A and H2B in elongating/condensing spermatids (ES) (N=3 CTL and 3 KO).** LC-MS/MS data acquired on H2A and H2B were processed as described above for H3 and H4 in whole testes, except that MS signals detected on modified peptides were normalized to be at constant amount of H2A or H2B in each analyzed sample. **C) Mass spectrometry quantification of histone variants in elongating/condensing spermatids (ES) and in round spermatids (RS).** The abundance of each histone variant was estimated from the sum of intensities of its identified specific proteolytic peptides. The abundance of each protein was then normalized to be at constant amount of histone H4 between samples. For some variants, MS values were divided by 10, so as to be able to visualize all the proteins on the same figure, whatever their MS intensity. Error bars represent standard deviations calculated on the measurements made on biological replicates. **D) Confirmation of a decrease in H4 acetylation in elongating spermatids (ES).** Western blot detection of H4K16ac, poly-H4ac and H3. Higher molecular weight bands likely represent non-denatured forms of histone H4. Western blot normalization was performed by detecting the same membranes with anti-H3 antibody. Anti-H3 rather than anti-H4 antibody was chosen to avoid interference with acetylated H4 signal. Histones H3 and H4 are organized as obligate tetramers (or occasionally heterodimers) in nucleosomes; therefore, they are expected to be present in the chromatin in the same quantity. **E) Mass spectrometry quantification of histone H2A and H2B PTMs in round spermatids (N= CTL and 3 KO).** LC-MS/MS data acquired on H2A and H2B were processed as described above for ES. **F) Mass spectrometry quantification of histone H3 PTMs in spermatozoa (N= 4 CTL and 4 KO).** Same legend as above. H4 PTMs levels were too weak to be quantified.

Blanco et al. Appendix

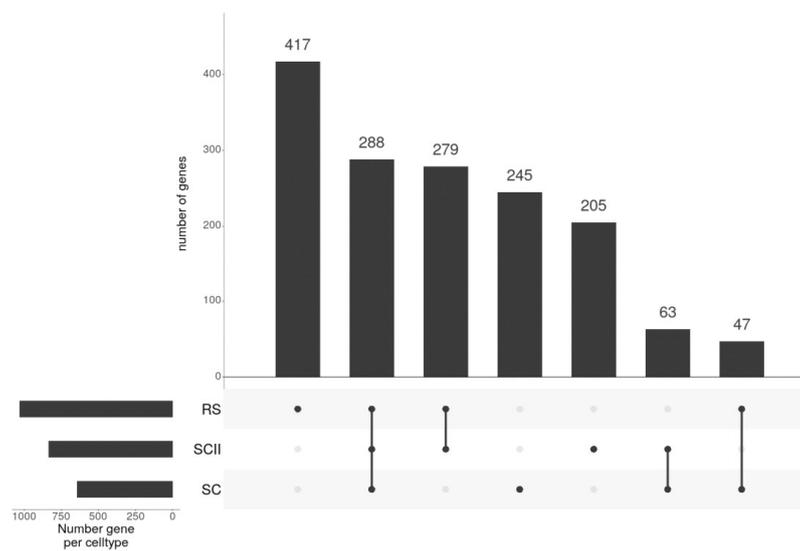
Appendix Figure S5

5A

Deregulated genes identified with DEseq2, with a Fold change > 1.5 and FDR < 0.05

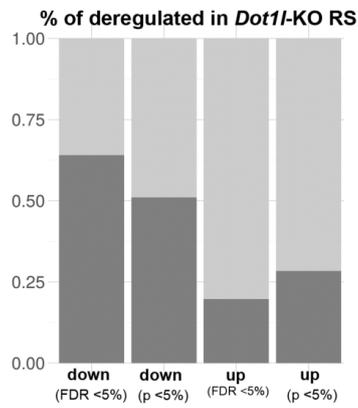


Deregulated genes identified with DEseq2, with a Fold change > 1.5 and p.value < 0.05



Blanco et al. Appendix

5B



5C

gene_name	logFC	logCPM	PValue	FDR	gl	H3K79me2 enrichment
<i>Slc10a3</i>	-1.26745740995669	0.999004290396908	0.00390006777314618	0.1545955595244	Down-regulated	yes
<i>Slc6a8</i>	-2.46727862572467	0.52781470462503	7.72654921275588e-07	0.000364372904035715	Down-regulated	yes
<i>Slc7a2</i>	-1.0665290941501	2.27338480123705	0.0150834110914719	0.387940102031817	Down-regulated	no
<i>Slc12a1</i>	1.3955625700193	0.681122253169346	0.000169908367523562	0.0174799371397939	Up-regulated	no
<i>Slc13a4</i>	1.54000399067509	2.76366740898566	4.27870310972635e-05	0.00657135254678289	Up-regulated	no
<i>Slc18a2</i>	1.25758541410194	-0.793930612861824	0.00237996137428981	0.110886263192696	Up-regulated	no
<i>Slc1a4</i>	1.10997163431429	0.860924966152978	0.00103762614100171	0.0601552770604185	Up-regulated	no
<i>Slc1a5</i>	1.06878712737505	2.2418481205103	0.00176357549579316	0.0883504245162868	Up-regulated	yes
<i>Slc25a23</i>	0.926874320430452	2.4842629146059	0.00744458216430107	0.240033828551142	Up-regulated	yes
<i>Slc25a37</i>	0.83494237600095	2.8363663126306	0.0494170665013017	0.856402093077086	Up-regulated	no
<i>Slc25a48</i>	0.844791121775937	0.601566750878029	0.035227161221668	0.691260215461133	Up-regulated	yes
<i>Slc26a11</i>	1.65452313717142	1.74006812886016	0.000661650590507997	0.044270891766396	Up-regulated	yes
<i>Slc28a2</i>	0.882471439124064	0.348580484195426	0.0475801028844169	0.835942882356513	Up-regulated	no
<i>Slc29a3</i>	1.00130602592956	3.99897275785481	0.024747552656584	0.549198182271674	Up-regulated	yes
<i>Slc37a3</i>	0.864250867428252	1.06283489334379	0.0157601580162183	0.398622878559951	Up-regulated	no
<i>Slc38a2</i>	1.19621193912572	2.2217849521669	0.000373937024375424	0.0301146206327321	Up-regulated	yes
<i>Slc38a8</i>	1.03918406075334	-0.0156846590541735	0.0201914826533337	0.472852642452675	Up-regulated	no
<i>Slc45a1</i>	1.7551977185066	-0.803109163412898	0.000635433125120995	0.0430016682924086	Up-regulated	no

Blanco et al. Appendix

5D

Homer de novo Motif Results (/home/mcoulee/Documents/Mouse/RNaseq/Motif_analysis/RS_UP_2kb//)

[Known Motif Enrichment Results](#)
[Gene Ontology Enrichment Results](#)
 If Homer is having trouble matching a motif to a known motif, try copying/pasting the matrix file into STAMP
 More information on motif finding results: [HOMER](#) | [Description of Results](#) | [Tips](#)
 Total target sequences = 656
 Total background sequences = 29647
 * - possible false positive

Rank	Motif	P-value	log P-value	% of Target	% of Background	STD (Bp STD)	Best Match Details	Motif File
1	GTCCATCTAGCA	1e-12	-2.788e+01	2.44%	0.17%	558.9bp (550.1bp)	PB0006_1_Bcl6_1/Jaspar(0.613) More Information Similar Motif Found	motif file (matrix)
2*	TCCACCCAG	1e-11	-2.703e+01	31.25%	19.61%	581.8bp (656.2bp)	PB0029_1_Hic1_1/Jaspar(0.758) More Information Similar Motif Found	motif file (matrix)
3*	TGATAGATG	1e-11	-2.643e+01	26.37%	15.69%	569.7bp (590.7bp)	H0XA1(Homeobox)inES-Hoxa1-ChIP-Seq(SRP064292)/Homer(0.742) More Information Similar Motif Found	motif file (matrix)
4*	CACICCGACCA	1e-11	-2.584e+01	5.34%	1.26%	574.9bp (629.9bp)	ZBTB14/MA1650_1/Jaspar(0.591) More Information Similar Motif Found	motif file (matrix)
5*	GCAGATCCATT	1e-11	-2.538e+01	4.27%	0.82%	510.8bp (658.3bp)	HNF4A/MA0114_4/Jaspar(0.749) More Information Similar Motif Found	motif file (matrix)
6*	TGTCGAGTTCAC	1e-10	-2.500e+01	2.90%	0.34%	620.5bp (655.8bp)	VDR/MA0693_2/Jaspar(0.682) More Information Similar Motif Found	motif file (matrix)
7*	CTACCCATAAAA	1e-10	-2.480e+01	2.59%	0.26%	476.8bp (615.7bp)	Cdx2(Homeobox)inES-Cdx2-ChIP-Seq(GSE14586)/Homer(0.736) More Information Similar Motif Found	motif file (matrix)
8*	AAAGTAAATTT	1e-10	-2.446e+01	3.96%	0.73%	508.5bp (549.2bp)	PB0165_1_Sox11_2/Jaspar(0.626) More Information Similar Motif Found	motif file (matrix)
9*	GCASCAGTGA	1e-10	-2.374e+01	4.88%	1.15%	645.9bp (600.2bp)	BCL6(BHLH)/HPC7-Sci-CAP-Seq(GSE13511)/Homer(0.651) More Information Similar Motif Found	motif file (matrix)
10*	GTCTCACGCC	1e-10	-2.339e+01	12.35%	5.59%	597.3bp (637.1bp)	Npas4(BHLH)/Neuron-Npas4-ChIP-Seq(GSE12793)/Homer(0.612) More Information Similar Motif Found	motif file (matrix)

gene_name	logFC	logCPM	PValue	FDR	gl	H3K79me2 enrichment
Cacna1e	1.381540782	2.204690766	0.002726556	0.121318141	Up-regulated	Yes
Gpsm1	1.115505708	1.018472463	0.002648857	0.119051423	Up-regulated	No
Tmem74b	1.021769193	0.247820597	0.013177832	0.354170753	Up-regulated	No
Gata5	1.673646912	-1.21958131	0.000102450	0.012941114	Up-regulated	No
Padi1	1.006558376	-0.525140701	0.027404915	0.589073299	Up-regulated	No
Ptpn13	0.896515079	3.191552902	0.026370274	0.572363590	Up-regulated	No
Megf8	0.875273628	5.337855272	0.023642554	0.532645802	Up-regulated	No
Kctd21	0.837477696	0.230864728	0.024428532	0.543473782	Up-regulated	No
Irf1	1.521852615	0.397338881	0.006670929	0.224722281	Up-regulated	No
Lhfp12	1.068600482	-0.608875694	0.028270799	0.599286273	Up-regulated	No
Grid1	1.387320345	2.179807865	0.000243895	0.022484676	Up-regulated	Yes
Dscc1	1.087546432	3.096511586	0.000497737	0.036606338	Up-regulated	No
Baiap2l2	0.874498322	0.165478650	0.034369221	0.684231999	Up-regulated	No
Zfp52	1.351603089	-0.975868434	0.001088020	0.062633064	Up-regulated	No
Plin3	0.983536252	2.207539245	0.007122447	0.233024479	Up-regulated	Yes

Appendix Figure S5. Complement to Fig 5. A) Upset plot showing the number of deregulated genes in *Dot1l*-KO vs CTL primary spermatocytes (SC), secondary spermatocytes (SCII) and round spermatids (RS) obtained using DEseq2 differential gene expression analyses. Top panel: with Fold Change > 1.5 and FDR < 0.05. Bottom panel: with Fold Change > 1.5 and p value < 0.05 (FDR was calculated *a posteriori* on the selected gene lists and found to be < 0.05). **B)** Bar graph showing the percentage of deregulated genes in *Dot1l*-KO RS found with the two types of differential analyses and which are enriched in H3K79me2 at their gene body in WT RS. **C)** List of *Sic* genes found deregulated in *Dot1l*-KO round spermatids. **D)** Top: results from HOMER motif analysis on the promoter (2kb) of upregulated genes (p-value < 0.05) in KO RS (top 10). BCL6/BCL6B binding DNA motif is the most significant of the identified motifs (and the only one which does not appear as “possible false positive”). Bottom: list of genes identified with HOMER which have BCL6 DNA binding site in their promoter.

Blanco et al. Appendix

Appendix Table S1

Table S1		
genotyping primers	sequence	PCR condition
Dot1l-F	GCAAGCCTACAGCCTTCATC	Annealing temperature = 58°C
Dot1l-R1	CACCGGATAGTCTCAATAATCTCA	
Dot1l-R2	GAACCACAGGATGCTTCAG	
STRA8CRE_oIMR9266F	AGATGCCAGGACATCAGGAACCTG	Annealing temperature = 62°C
STRA8CRE_oIMR9267R	ATCAGCCACACCAGACACAGAGATC	
Ymtxfp10_ internal control	TCACACAGATAAGAGGGTATTG	
Ymtxrp13_ internal control	GTTTTCCTATCAGGCCATCCA	

Blanco et al. Appendix

Appendix Table S2

Table S2			
Western Blot - Primary antibodies			
Target	Manufacturer	Reference	Usage
DOT1L	CST	D402T	1/500 - 1/1000
Tubulin	Upstate	05-661	1/3000
H3K79me2	Abcam	ab3594	1/500
TH2B	Upstate	07-680	1/5000
H3	Abcam	ab 1791	1/3000 - 1/10000
H4K16ac	Upstate	07-329	1/750
Poly ac-H4	Millipore	06-866	1/250
TNP2	Santa Cruz Biotechnology	sc21106	1/250
PRM2	Briar Patch Biosciences	MAB-Hup2B	1/500
Western Blot - Primary antibodies			
Target	Manufacturer	Reference	Usage
Donkey anti-goat HRP	Santa Cruz Biotechnology	sc2020	1/5000
Goat anti-mouse HRP	ThermoFisher	31430	1/5000
Goat anti-rabbit HRP	ThermoFisher	31460	1/5000
Immunofluorescence-Immunohistochemistry			
Target	Brand	Reference	Usage
DOT1L	CST	D402T	1/100
DOT1L	Abcam	64077	1/100
H3K79me2	Diagenode	C15410051	1/100
GFP	ThermoFisher	CAB4211	1/100
Poly ac-H4	Millipore	06-866	1/200
TNP2	Santa Cruz Biotechnology	sc21106	1/50
Chromatin Immunoprecipitation			
Target	Brand	Reference	
H3K79me2	Diagenode	C15410051	

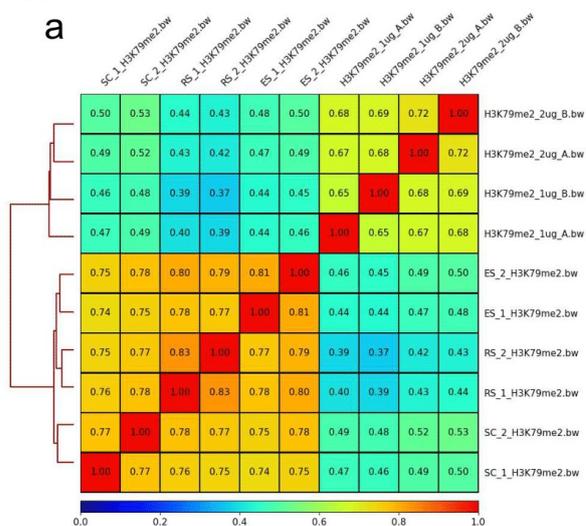
Blanco et al. Appendix

References

- Darde TA, Lecluze E, Lardenois A, Stevant I, Alary N, Tuttelmann F, Collin O, Nef S, Jegou B, Rolland AD et al (2019) The ReproGenomics Viewer: a multi-omics and cross-species resource compatible with single-cell studies for the reproductive science community. *Bioinformatics* 35: 3133-3139
- Darde TA, Sallou O, Becker E, Evrard B, Monjeaud C, Le Bras Y, Jegou B, Collin O, Rolland AD, Chalmel F (2015) The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Res* 43: W109-116
- Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB et al (2018) A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental cell* 46: 651-667 e610
- Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB et al (2018) Gene Expression Omnibus GSE112393 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112393>) [DATASET].
- Mouse ENCODE transcriptome data (2011) NCBI Sequence Read Archive PRJNA66167 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA66167>) [DATASET].
- Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L (2007) A global double-fluorescent Cre reporter mouse. *Genesis* 45: 593-605
- Russell LD, Hikim APS, Ettlin RA, Clegg ED (1990) *Histological and Histopathological Evaluation of the Testis*. Cache River Press

Figures supplémentaires – Article 2

Fig. S1



b

	GSC rep1	GSC rep2	GSC rep3	GSC rep4	SCI rep1	SCI rep2	RS rep1	RS rep2	ES rep1	ES rep2
Number of reads	21 260 743,00	23 051 274,00	24 537 676,00	25 195 852,00	45 210 180,00	54 045 728,00	56 859 200,00	53 051 834,00	43 379 619,00	55 919 709,00
MAC2 peak calling (broad parameter)										
Number of peaks	26 856,00	26 463,00	25 571,00	28 820,00	53 603,00	53 895,00	70 458,00	75 418,00	7 678,00	1 276,00
Number of annotated peaks	26 601,00	26 189,00	25 325,00	28 505,00	52 977,00	53 292,00	69 485,00	74 382,00	7 607,00	1 225,00
common analysis										
Number of peaks in common				15 896,00		46 034,00		67 918,00		385,00
ChIPseeker										
Number of peaks annotated as "promoter"				10 089,00		19 480,00		25 555,00		4 040,00
Number of peaks annotated as "intragenic"				5 621,00		13 880,00		21 348,00		4 267,00
Number of peaks annotated as "distal intergenic"				170,00		2 122,00		8 788,00		159,00
Number of peaks annotated as "downstream"				15,00		219,00		535,00		7,00

c

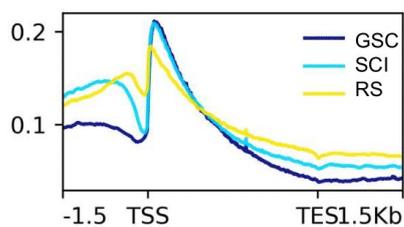


Figure S1. H3K79me2 ChIP-Seq.

a. Spearman correlation of H3K79me2 ChIP-seq data. **b.** Overall results from H3K79me2 ChIP-seq data. **c.** H3K79me2 enrichment profile for GSC, SCI and RS.

Figure S2. Chromatin environment.

a. Mean gene expression level (log(cpm)) of genes associated with enhancers enriched (+) or not (-) in H3K79me2 in GSC, SCI and RS. Top panels: method from ChromENVEE, and bottom panels: method from (Rada-Iglesias et al., 2011). Bivalent enhancers correspond to only 23 H3K79me2+ regions in GSC. The statistical analysis is therefore not powerful enough to confidently determine a p-value, which may explain the discrepancy between the two methods.

Fig. S3

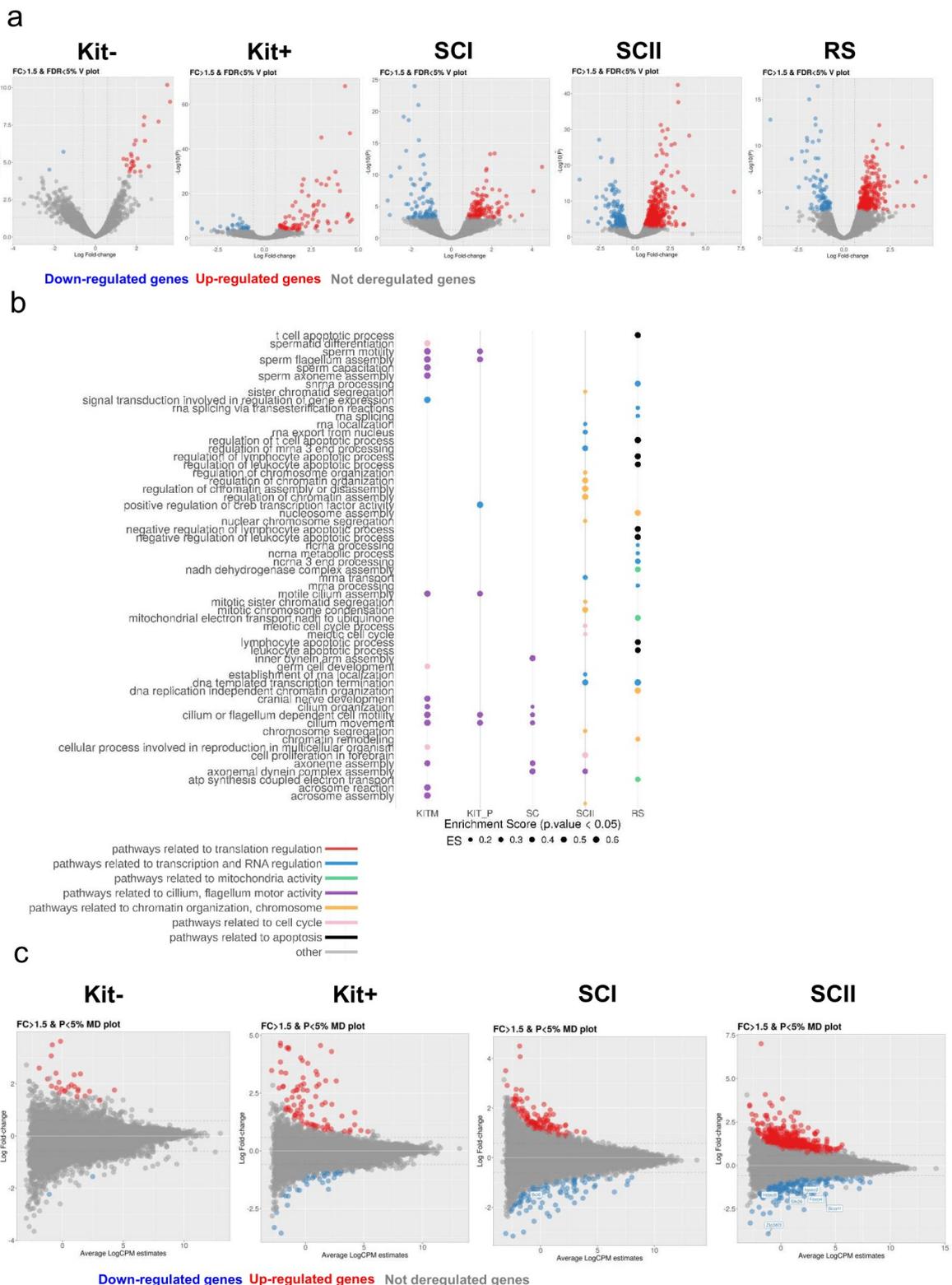


Figure S3. RNA-Seq.

a. Volcano plots showing differentially expressed genes between Dot1l-KO vs CTL cells. **b.** GSEA for Kit⁻ and Kit⁺ cells compared to results obtained for SCI, SCII and RS, which were previously shown in (Blanco et al 2023). **c.** MD plot showing deregulated genes in Kit⁻, Kit⁺, SCI and SCII (spike-in RNA-seq analysis). Significantly downregulated genes that are enriched in H3K79me2 and known transcriptional repressors are indicated in the figure.

Fig. S4

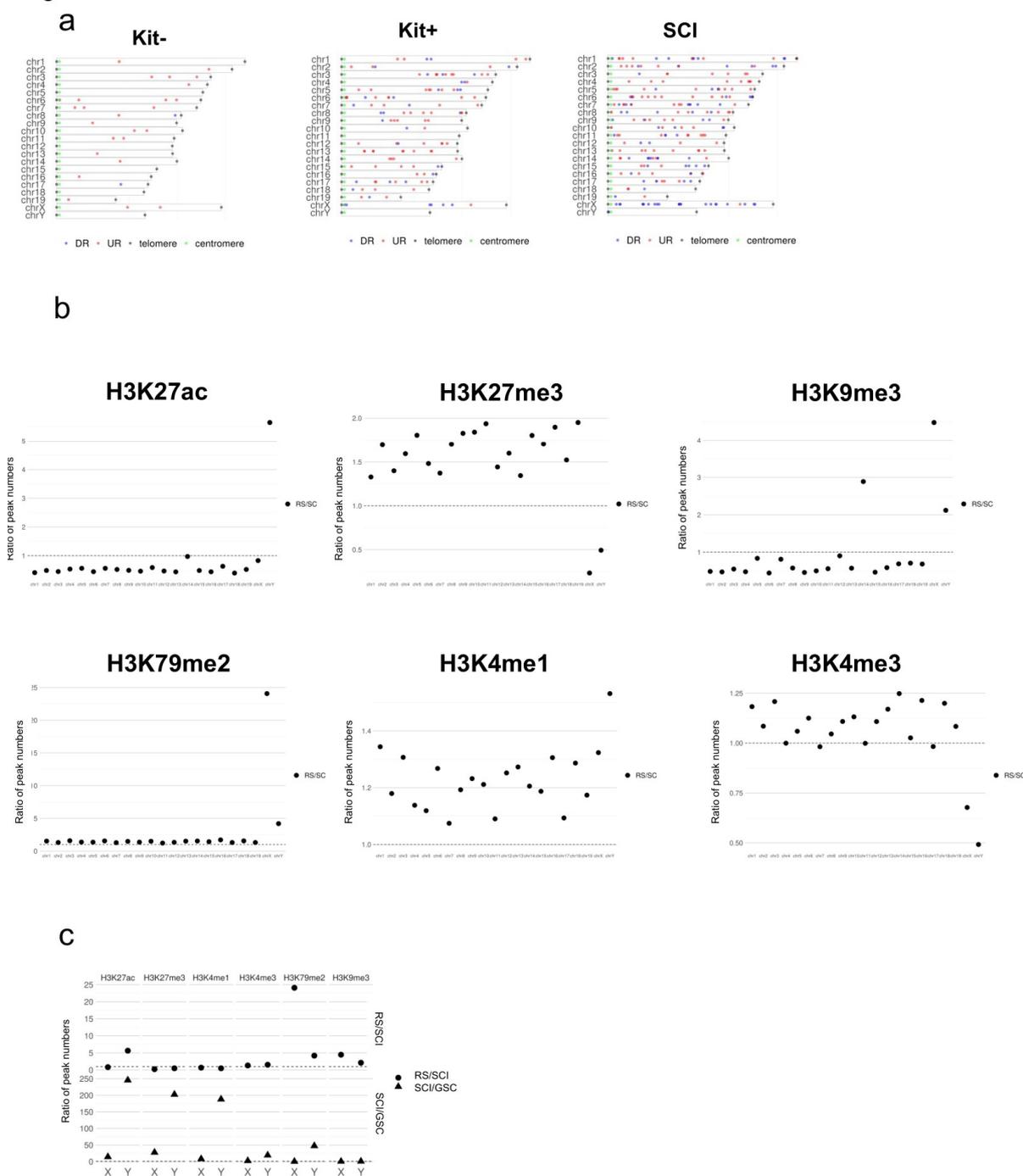


Figure S4. XY gene regulation.

a. Chromosome localization of differentially expressed genes (“spike-in” analysis) in Kit-, Kit+ and SCI. Blue: downregulated genes (DR), red: upregulated genes (UR), the green dot indicates the centromere, and the grey dot, the telomere. **b.** Ratio of peak numbers at RS stage compared to SCI stage, per chromosome, for several histone marks. The dashed line indicates a ratio = 1 corresponding to a stable number of peaks. **c.** Ratio of peak numbers on the X or Y chromosome in RS vs SCI (circle) and in SCI vs GSC (triangle), for several histone marks. The dashed line indicates a ratio = 1

Fig. S5

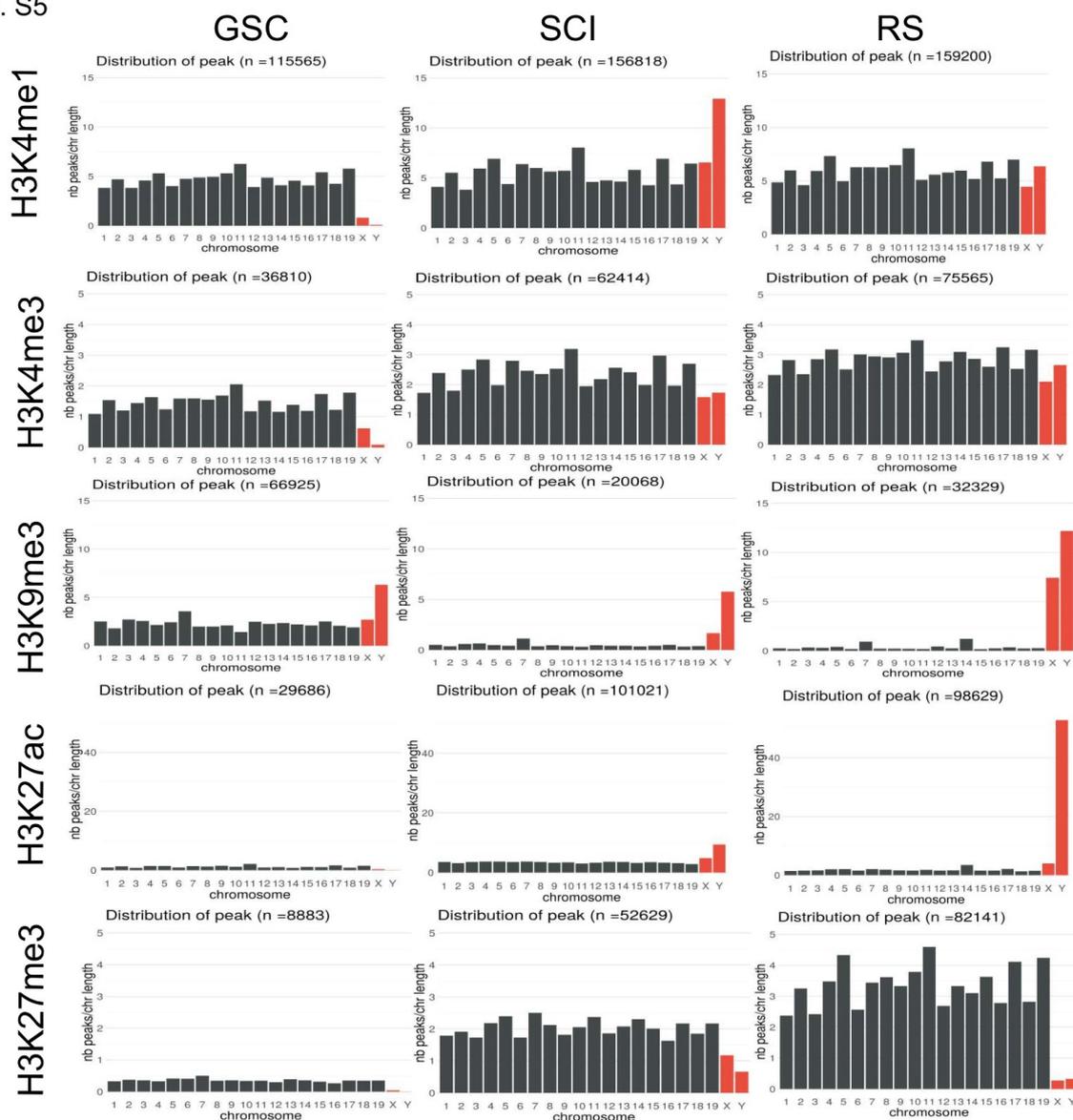


Figure S5. Peaks distribution of different histone marks along the chromosomes. The number of peaks is normalized to the chromosome length. Analysis is performed for GSC, SCI and RS.