The Molecular Mechanism of PRDM9 during Meiosis in the Mouse

by

Sarah Thibault-Sennett

Dissertation submitted to the Faculty of the
Molecular and Cellular Biology Graduate Program
Uniformed Services University of the Health Sciences
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy 2017

UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814

APPROVAL OF THE DOCTORAL DISSERTATION IN THE
MOLECULAR AND CELLULAR BIOLOGY GRADUATE PROGRAM

Title of Dissertation: "The Molecular Mechanism of PRDM9 during Meiosis in the Mouse"

Name of Candidate: Sarah Thibault-Sennett
Doctor of Philosophy Degree
October 11, 2017

DISSERTATION AND ABSTRACT APPROVED:

DATE:

[redacted]
10/25/17
Gabriela Dveksler, PhD
DEPARTMENT OF PATHOLOGY
Committee Chairperson

[redacted]
10/25/17
Galina Petukhova, PhD
DEPARTMENT OF BIOCHEMISTRY
Dissertation Advisor

[redacted]
10/25/17
Xin Xiang, PhD
DEPARTMENT OF BIOCHEMISTRY
Committee Member

[redacted]
10/25/17
Albert Dobi, PhD
CENTER FOR PROSTATE DISEASE RESEARCH
Committee Member

[redacted]
10/25/17
Michael Lichten. PhD
NATIONAL CANCER INSTITUTE
Committee Member

UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814

FINAL EXAMINATION/PRIVATE DEFENSE FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE MOLECULAR AND CELLULAR BIOLOGY GRADUATE PROGRAM

Name of Student: Sarah Thibault-Sennett

Date of Examination: October 11, 2017
Time: 3:30 PM
Place: A2015

DECISION OF EXAMINATION COMMITTEE MEMBERS:

	PASS	FAIL
[redacted]	✓	
Gabriela Dveksler, PhD DEPARTMENT OF PATHOLOGY Committee Chairperson		
[redacted]	✓	
Galina Petukhova, PhD DEPARTMENT OF BIOCHEMISTRY Dissertation Advisor		
[redacted]	✓	
Xin Xiang, PhD DEPARTMENT OF BIOCHEMISTRY Committee Member		
[redacted]	✓	
Albert Dobi, PhD CENTER FOR PROSTATE DISEASE RESEARCH Committee Member		
[redacted]	✓	
Michael Lichten. PhD NATIONAL CANCER INSTITUTE Committee Member		

ACKNOWLEDGMENTS

I would first like to thank my advisor, Dr. Petukhova for providing non-stop guidance and support for the last 5 or so years. You allowed me freedom in choosing my projects and allowed me to pursue many avenues that would not be possible for most graduate students. You also have spent endless hours reading drafts and watching my presentations, and I am very grateful for all the feedback. I hope that there is a noticeable difference from when I joined your lab! Thank you so much!

I would like to thank my committee members for sitting through many long committee meetings and for providing wonderful feedback and encouragement on this project. You were a fantastic sounding board and I thank you for your time and energy.

I would not have been able to pursue any of this work without the institutional support at USU: from my MCB program to GEO. Thank you!

To my parents - Once my middle school teacher informed you that "science just wasn't going to be my thing." I took that to heart, you didn't. You have been my strongest supporters from day one and encouraged me back into the field that I loved. I can never thank you enough for that and for supporting me throughout my college and graduate work. Thank you!

To Riley - you may not have performed any of these experiments or written any of this text, but this work might not have been completed without your endless support and encouragement. Thank you for the many nights you sat with me in the cell culture hood, your willingness to either celebrate small successes with pizza or drown disappointments (also in pizza), and how you have refused to let me turn into a cynic, this is for you!

DEDICATION

During my graduate career I lost two wonderful grandmothers, I would like to dedicate this work to them.

For my Mémère - who encouraged the women in our family to continue with our education though she was denied the opportunity.

For my Grandma - who had a gift to see the wonder and beauty in the world around her.

COPYRIGHT STATEMENT

The author hereby certifies that the use of any copyrighted material in the dissertation manuscript entitled: "The Molecular Mechanism of PRDM9 during Meiosis in the Mouse" is appropriately acknowledged and, beyond brief excerpts, is with the permission of the copyright owner.



Sarah Thibault-Sennett

February 6th, 2018

DISCLAIMER

The views presented here are those of the author and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences, the Department of Defense or the U.S. Government.

ABSTRACT

The Molecular Mechanism of PRDM9 during Meiosis in the Mouse

Sarah Thibault-Sennett, PhD, 2017

Thesis directed by: Galina Petukhova, PhD, Associate Professor, Biochemistry and Molecular Biology

Meiotic homologous recombination is vital for accurate segregation of the homologous chromosomes in the first meiotic division, and infertility, miscarriage and developmental disabilities can result when this process is impaired. Additionally, meiotic homologous recombination generates genetic diversity. During meiosis, recombination predominantly occurs at recombination "hotspots" determined by the DNA-binding protein PRDM9, which is essential for fertility. PRDM9 marks where homologous recombination can occur with a histone modification, H3K4me3.

We propose that (1) PRDM9 binding is regulated beyond the recognized nucleotide sequence and that (2) PRDM9 has an active role in promoting homologous recombination at the binding sites.

We investigated if DNA methylation could alter PRDM9 binding using a mouse that lacked a key component of the *de novo* methylation pathway (*Dnmt3L*$^{-/-}$). We mapped the location of PRDM9 binding through H3K4me3 ChIP-seq in the testis and mapped the location of the meiotic recombination sites with our single-stranded DNA

sequencing technique. These studies showed that DNA methylation does play a role in regulating recombination at PRDM9 binding sites and, in fact, explains the difference in hotspot usage strength between male and female mice.

PRDM9 contains three different domains which are typically involved in regulation of transcription, but the role of PRDM9 in gene expression has not been evaluated. We first confirmed that PRDM9 does not appear to be involved in gene regulatory activity through RNA-seq. We further explored the role of PRDM9 in meiosis by analyzing the effect of removing the KRAB, SSXRD and post-SET zinc finger domains followed by expression in a cell system and generating a mouse which lacked the KRAB domain. We found that although the post-SET zinc finger and the KRAB domain are not essential for the methyltransferase activity of PRDM9 *in vitro*, the KRAB domain mutant mice show only residual PRDM9 activity and undergo meiotic arrest. In aggregate, our data indicate that the PRDM9 function has likely diverged from the typical function of a transcription factor and has adapted to the new role in setting the proper chromatin environment for initiation and completion of homologous recombination.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER 1: INTRODUCTION 1

Development of the Gametes 1
Basic embryogenesis: generation of the three germ layers 1
Primordial Germ Cell Specification 3
DNA Methylation in Germ Cell Development 5
Gametogenesis 9
Differences between spermatogenesis and oogenesis 9
Male germ cell specification 10
Female Germ Cell Specification 14
Meiosis 16
Meiotic recombination overview 16
Double-strand break formation and processing 18
Homologous chromosome synapsis 20
Crossing-over resolution and homolog segregation 21
Crossing-over versus non-crossing-over pathway 22
Meiotic Recombination “Hotspots” 23
Mapping of recombination sites 23
Global mapping revealed genome-wide properties of DSB hotspots in the mouse 24
Previous evidence that PRDM9 was imperative for meiotic progression in mice 25
Global DSB mapping with single-stranded DNA sequencing (SSDS) 27
An Expanding Role for PRDM9 during Mouse Meiosis 32

CHAPTER 2: MANUSCRIPT 1 - “Extensive sex differences at the initiation of genetic recombination” 33

Abstract 34
Main text 35
Acknowledgements 44
Main Figures 45
Supplemental Tables 50
Supplementary Figures 53
Materials and Methods 74

CHAPTER 3: MANUSCRIPT 2 - “A non-repressive role for the KRAB domain and SSXRD of PRDM9” 82

Abstract 83
Introduction 84
Results 87

Inactivation of *Prdm9* does not lead to changes in gene expression. 87
The KRAB domain and the post-SET zinc finger are not essential for H3K4me3 activity of PRDM9 *in vitro*. 90
The KRAB domain is necessary for PRDM9 activity *in vivo*. 92
Discussion 94
Acknowledgements 98
Main Figures 99
Supplementary Tables 104
Supplementary Figures 110
Materials and Methods 115

CHAPTER 4: CONCLUSION 122

Specific Aims 122
Regulation of PRDM9 Binding 123
Specific DNA methylation patterns can either favor or disfavor PRDM9 binding. 123
Interplay of DNA methylation and hotspot usage between the sexes 124
Limitations of study 125
Future directions 126
Formation of DSBs at Sites Bound by PRDM9 129
The KRAB domain 129
The SSXRD 131
The lone zinc finger 132
Limitations of study 134
Future directions 136
Closing 138

REFERENCES 140

LIST OF TABLES

CHAPTER 2: MANUSCRIPT 1 - "Extensive sex differences at the initiation of genetic recombination"
Supplementary Table 1: SSDS sample details 50
Supplementary Table 2: Other sequencing sample details 51
Supplementary Table 3: Public datasets used 52
Supplementary Table 1: RNA-seq samples 104
Supplementary Table 2: Genes differentially expressed in *Prdm9*$^{-/-}$ and *Hop2*$^{-/-}$ testes at Q-value < 0.05 105
Supplementary Table 3: miRNA-seq samples 106
Supplementary Table 4: Novel miRNAs found in all four miRNA-seq samples 107
Supplementary Table 5: Sequences used in gene targeting and genotyping primers for *Prdm9*^{K} mice. 108
Supplementary Table 6: ChIP sample information 109

LIST OF FIGURES

CHAPTER 1: INTRODUCTION
Figure 1: Schematic diagram showing the derivation of mammalian embryonic germ layers. 1
Figure 2: Specification of primordial germ cells. 3
Figure 3: DNA methylation and PGC development 8
Figure 4: Sperm development. 13
Figure 5: DNA events in meiotic recombination. 18
Figure 6: Meiotic prophase progression with immunostaining. 19
Figure 7: Kinetic enrichment of ssDNA fragments for high-throughput sequencing. 29
Figure 1. Meiotic DSB hotspots can be detected in oocytes. 45
Figure 2. Extensive sex differences in DSB hotspot usage. 46
Figure 3. Large-scale influences on sex biased recombination. 47
Figure 4. Sex-dimorphic patterns of DNA methylation at sex-biased hotspots. 48
Supplementary Figure 1: Sample details and quality metrics for DSB maps 53
Supplementary Figure 2: DSB hotspots unique to O2 are likely false positive calls 54
Supplementary Figure 3: Most DSB hotspots are used in both male and female meiosis 55
Supplementary Figure 4: DSB hotspots are narrower in ovaries than in spermatocytes 56
Supplementary Figure 5: Hotspot usage differs between male and female mice 58
Supplementary Figure 6: Most meiotic DSBs occur at sex-biased hotspots 59
Supplementary Figure 7: Sex-biased hotspots are unlikely to be arising from inter-sex differences in DSB repair dynamics. 60
Supplementary Figure 8: Default hotspots are more prevalent in female meiosis 61
Supplementary Figure 9: Recombination is enriched at TSS in mouse and human females 62
Supplementary Figure 10: The physical size of hotspot clusters scales with the number of hotspots per cluster 64
Supplementary Figure 11: Similar proportions of PRDM9-defined and default hotspots occur in clusters 65
Supplementary Figure 12: Clusters of female-biased hotspots are enriched at q-arm subtelomeric regions. 66
Supplementary Figure 13: Genetic crossovers in males are elevated adjacent to the q-arm telomere, independent of the *Prdm9* allele 67
Supplementary Figure 14: Differing patterns of DNA methylation at sex-biased hotspots 68
Supplementary Figure 15: DNA methylation at PRDM9 binding sites is present across tissues and absent in the female germ line 69
Supplementary Figure 16: Schematic of how DNA methylation affects hotspot usage 70
Supplementary Figure 17: Elevated DNA methylation at female-biased hotspots outside of the PRDM9 binding site. 71

Supplementary Figure 18: DNA methylation at PRDM9 binding sites is reduced in *Dnmt3L*$^{A/A}$ mice.....72
Supplementary Figure 19: Hotspot strength variation in *Dnmt3L*$^{-/-}$ mice.....73
Figure 1. Inactivation of the *Prdm9* gene does not lead to changes in gene expression profiles. 99
Figure 2: Expression of GFP-PRDM9 in GC-1 cells leads to H3K4 trimethylation at *Dom2* recombination hotspots (HS). 100
Figure 3: GFP-PRDM9 lacking the KRAB domain retains activity in the GC-1 expression system......101
Figure 4: The KRAB domain of PRDM9 is essential for meiotic progression. 102
Figure 5: KRAB-less PRDM9 retains residual activity in vivo......103
Supplementary Figure 1: Gene expression profile in *Prdm9*$^{-/-}$ testes is most similar to 12-14 dpp wild type testes. 110
Supplementary Figure 2: Expression of GFP-PRDM9 constructs in GC-1 cells. ...111
Supplementary Figure 3: Gene targeting design for the KRAB deletion. 112
Supplementary Figure 4: *Prdm9*$^{K/K}$ mice do not produce functional gametes......113
Supplementary Figure 5: PRDM9, not DSB processing, promotes the formation of nucleosome-depleted region at hotspot centers......114
Figure 1: Model of hotspot specification by PRDM9. 138

CHAPTER 1: INTRODUCTION

DEVELOPMENT OF THE GAMETES

Basic embryogenesis: generation of the three germ layers

Following fertilization, the single-cell mouse embryo develops into the blastocyst, which forms both the embryo and the surrounding extraembryonic tissue. The early blastocyst consists of two major components: the blastocoel and the inner cell mass. The blastocoel is a primordial yolk sac contained by a layer of trophoblast cells. On top of this yolk-sac is the layer of cells termed the "inner cell mass" which goes on to form the embryo proper (Figure 1). At E3.5 (3.5 days post-conception) the cells in the inner cell mass begin expressing either Nanog or Gata6, specifying them as either epiblast cells which will go on to form the three embryonic tissues, or hypoblast cells which contribute to the yolk sac, respectively.

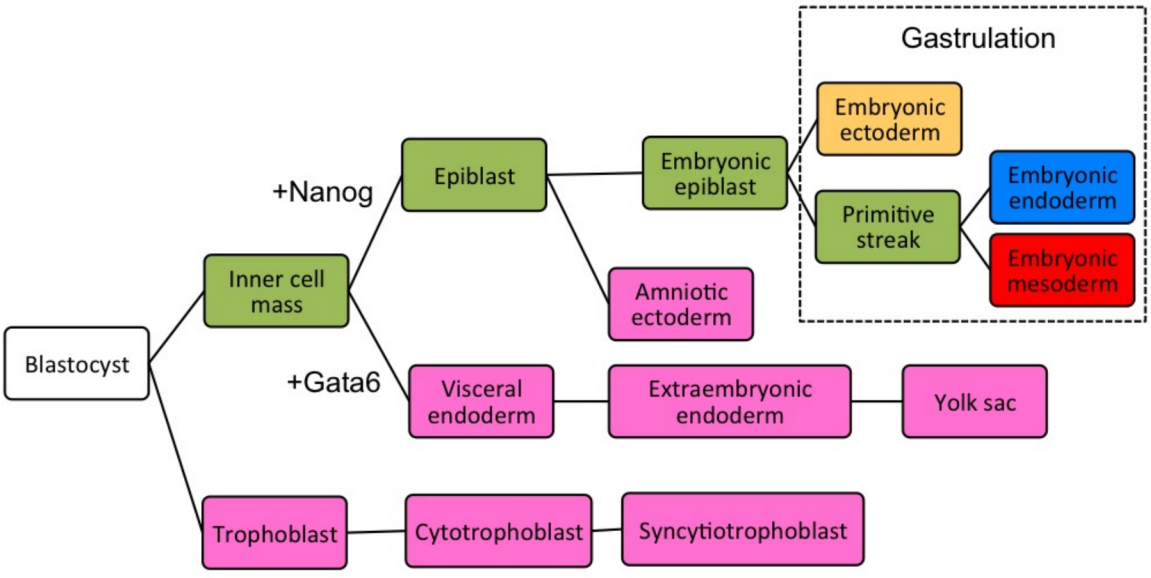


Figure 1: Schematic diagram showing the derivation of mammalian embryonic germ layers.

Figure modified from (48). Pink boxes represent extraembryonic tissues.

During embryogenesis, three tissue types are specified: endoderm, ectoderm and mesoderm. The endoderm represents the "inner-most" tissue layer and goes on to form the digestive system and the respiratory system. The ectoderm represents the "outer-most" tissue layer and forms the epidermis and interestingly, the nervous system. Finally, the mesoderm represents the "middle" layer and goes on to form all of the organ systems not mentioned above, such as the musculoskeletal system, the blood and circulatory systems and the reproductive system.

The epiblast undergoes the transition to the three germ layers through the intricate process of gastrulation. During gastrulation a physical depression -- the primitive groove -- appears in epiblast layer. Cell specification occurs as individual cells begin a migration path from the epiblast, diving into and through the primitive groove and down to form a new layer between the epiblast and the hypoblast layer. Cells first migrate through to form a hypoblast-most layer and are specified as endoderm cells, with later migrating cells forming a layer on top of the endoderm layer which make up the mesoderm.

This migration and resulting specification is coordinated by fibroblast growth factors, specifically *fgf8*, that are secreted by cells forming the primitive groove (34). In the chick embryo, *fgf8* leads to specification of the mesoderm by regulation of *snail*, *Brachyury (T)*, and *Tbx6*, and presumably specification occurs by a similar program in the mouse.

Primordial Germ Cell Specification

The formation of the germ cells involves an intricate network of gene regulation and extensive cell migration throughout the forming embryo. The specification of primordial germ cells (PGCs) occurs very early in development, surprisingly prior to gastrulation and the subsequent creation of the mesoderm. At the time the PGCs are specified the embryo consists of the hypoblast, consisting of the visceral endoderm cell layer separating the yolk sac from the epiblast cell layer which sits atop (Figure 2).

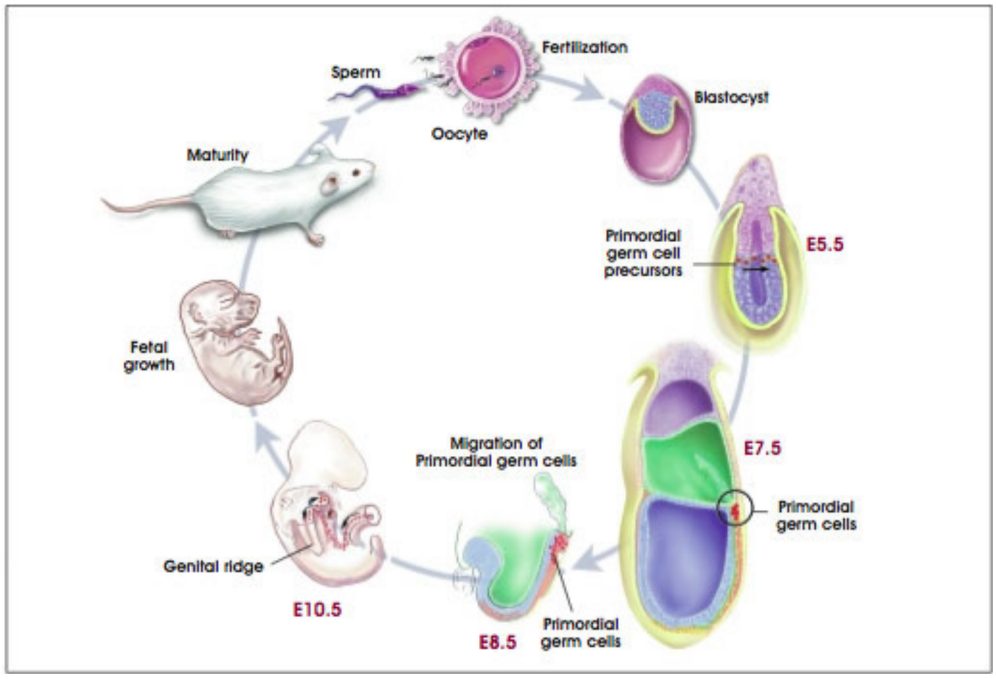


Figure 2: Specification of primordial germ cells.
Figure adapted from (126).

The visceral endoderm expresses a gradient of Wnts (a family of polarity establishing proteins in both *Drosophila* [*wingless*] and vertebrates [*integrated*]), with higher levels at what will become the posterior of the embryo. Cells exposed to high levels of Wnts are made competent to begin differentiation into PGCs by additional exposure to bone morphogenic proteins-2, -4, and -8 (BMPs, TGF-β family members,

(153). Cells specified by these BMPs begin expressing a host of genes involved in PGC identity and survival: alkaline phosphatase, Oct4, developmental pluripotency associated 3 (Dppa, (113)), and stage-specific embryonic antigen (Ssea1, (151)). PGCs also express the RNA binding proteins Nanos3 and Dnd1 (75; 176). BMP4 and the Wnts are capable of starting the expression cascade above and are sufficient for PGC specification. Embryonic cells that are exposed to BMP4 are capable of forming PGCs and produce functional sperm following implantation into a different embryo (127).

During gastrulation, the small population of specified PGCs then move into the embryonic endoderm through the primitive groove and take up residence in the hindgut between E8-10.5. The cell migration involves the crawling of the PGCs over the underlying cells through the extension of filopodia. Migrating PGCs also have the capability to penetrate cell sheets and monolayers. It is during this migration the PGCs are undergoing an extensive gene expression shift.

Blimp1 turns off the genes required for somatic development while turning on Sox2 and Nanos which both promote pluripotency (56). Blimp1 also turns on Nanos3 which is imperative for preventing apoptosis in the PGCs as they migrate through various pools of cell signaling molecules (56). Prdm14 also turns on Sox2, however it's best known role is directing the chromatin modifications to promote the specific epigenetic landscape that is required of premeiotic cells (190).

In addition to the alteration of gene expression in the PGCs themselves, the PGCs are accompanied throughout their migration with clusters of non-PGC cells that are responsible for secreting stem cell factor (SCF) throughout the migration to maintain pluripotency. And while a handful of genes involved in germ cell specification are

upregulated, the majority of the PGC genome is transcriptionally repressed during the migration.

The PGCs are proliferating as they migrate along the hindgut. When the PGCs have colonized the newly formed genital ridge they begin undergoing genome-wide demethylation and histone modifications to generate a genomic "blank slate" for the developing gametes.

DNA Methylation in Germ Cell Development

Both DNA methylation and histone modifications are involved in epigenetics, a term which describes the "mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to DNA sequence" (40). Through the course of development from a single-celled embryo through a multicellular eukaryotic being the differentiating cells obtain chemical modifications which regulate gene expression and confer cell identity. These modifications are important for a variety of purposes, including restricting gene expression to genes crucial for a cell's function, opening regions of the genome for easy access to promote gene expression and restricting expression of imprinted genes to either the maternal or the paternal chromosome.

While histone modifications and DNA methylation play similar regulatory roles, they are involved in different interactions at the chromatin level. The ubiquitous histones which populate the genome are available for a variety of chemical modifications which alter how transcription factors can bind to the DNA or affect the attraction between the DNA and the histone itself.

Unlike histone modifications, which are chemical modifications to the proteins which organize the DNA, DNA methylation is the only known covalent modification to the DNA itself (reviewed in (68)). DNA methyltransferases (DNMT) transfer a methyl group to the cytosine from S-adenosyl-L-methionine, to form 5-methylcytosine (referred to hereafter as "DNA methylation")(33; 143). The placement of this methyl group does not affect the base pairing of cytosine, but swings out into the major DNA groove. As many different DNA-binding proteins interact with the major groove this allows the modification to directly promote or inhibit protein binding. Unlike the DNA sequence which is passed on with each replication of the cell, the DNA methylation pattern is passively lost with each round of DNA replication and needs to be reestablished on the newly synthesized DNA strand.

Methylation is added to DNA using one of two mechanisms: maintenance or *de novo*. Maintenance methylation serves to counteract the passive loss of total DNA methylation in a cell which results from semi-conservative DNA replication during each cell division. DNA methylation typically occurs in CpG nucleotide sets. When one of these sets is methylated, there are actually two methyls attached: one on each cytosine in the pair on the complementary DNA strands. During semi-conservative DNA replication one DNA strand is retained while a new strand is generated. This allows each new DNA molecule to maintain partial DNA methylation which guides *Dnmt1* (DNA methyltransferase 1) to place a corresponding methyl group to the opposing newly-synthesized strand. Through this mechanism each daughter cell is ensured to have the same level of DNA methylation at the same locations as in the parent cell, maintaining

the cell identity through divisions. As expected, *Dnmt1* is crucial for viability, and knockout mice exhibit embryonic lethality (91; 94).

De novo methylation, on the other hand, leaves a new DNA methylation pattern at previously unmarked chromatin. This program is necessary for cells to differentiate and to restrict the expression of unnecessary genes. *De novo* methylation is catalyzed by *Dnmt3A* and *Dnmt3B*. Mouse knockouts for either of these proteins also results in embryonic lethality as it is incompatible with cell differentiation (72; 129). The other *de novo* pathway member *Dnmt3l*, does not have any enzymatic activity on its own, although it regulates the other *Dnmt3* proteins. It also appears that *Dnmt3l* plays slightly different roles in males and females: male knockout mice exhibit sterility while female knockouts only exhibit abnormalities in imprinting (21).

As the DNA methylation pattern is crucial for maintaining a cell's identity, there is a strong need to remove any established pattern in cells which are destined to form the gametes which will give rise to pluripotent embryonic stem cells. Correct embryogenesis is incompatible with the existing DNA methylation patterns of the differentiated precursor cells. However, there is an additional complication: once the DNA methylation pattern is completely removed the gene expression levels will be altered which will prove deleterious in actively dividing and differentiating cells.

The PGCs rid themselves of their DNA methylation patterns through passive loss. At approximately E7.5, the PGCs repress both their maintenance and *de novo* DNA methylation pathways while they continue to divide mitotically. This passive loss of DNA methylation through multiple rounds of cell division allows the global DNA

methylation to drop from 71% when the pathways are repressed to 15% at E11.5 when the PGCs begin undergoing sex-specific PGC differentiation (Figure 3, (153; 161)).

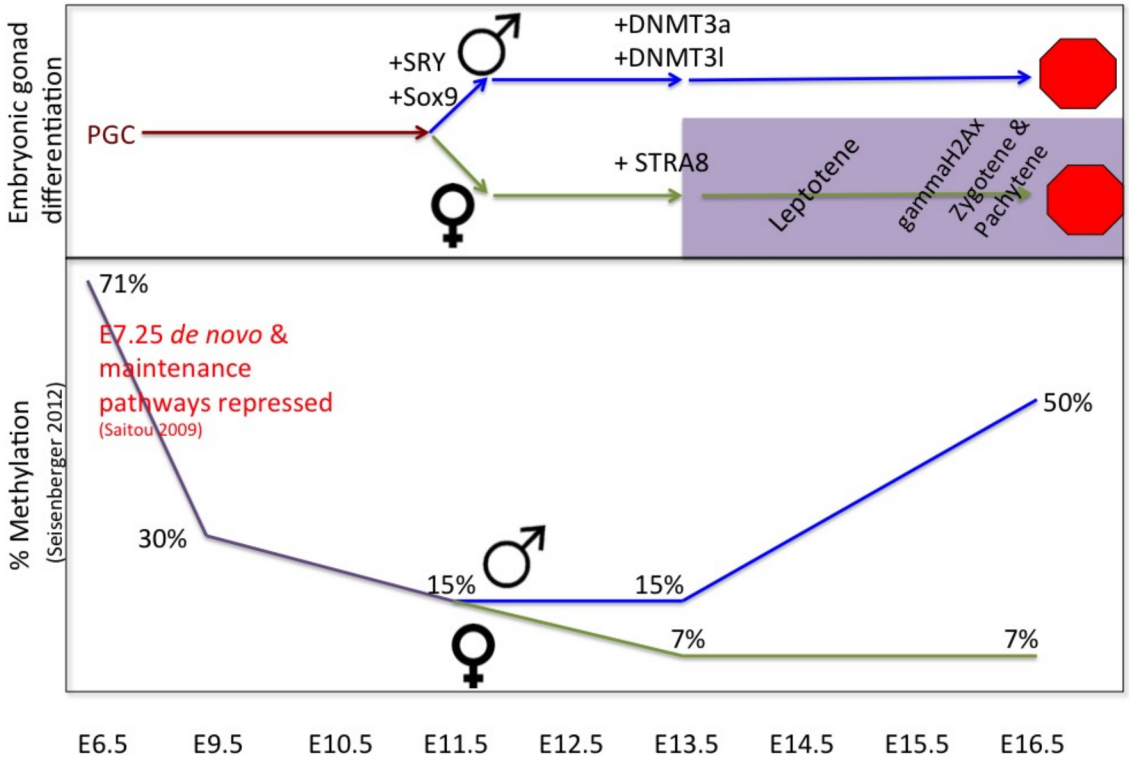


Figure 3: DNA methylation and PGC development

Timeline of embryonic development, measures in Embryonic Days. Top panel shows timing of sex specific gonad differentiation followed by arrest during embryogenesis in both sexes. The stop sign represents the meiotic arrest for females and the mitotic arrest for males. Bottom panel shows the % Methylation of PGCs throughout development compared to surrounding somatic tissue. Blue line following E11.5 represents methylation level in male PGCs, green line following E11.5 represents methylation level in female PGCs. Data for figure from (153; 161).

GAMETOGENESIS

Differences between spermatogenesis and oogenesis

Regardless of the sex of the mouse, the migration and specification of PGCs follows the same path until the cells arrive in the genital ridges. From there the further development of functional oocytes and sperm follow distinct paths which differentiate their end products. Oocytes are large, non-motile cells which have sequestered large amounts of cellular machinery, allowing an explosion of cell division cycles following fertilization. On the other hand, sperm are streamlined, motile and have lost most of their cellular machinery to allow for propulsion.

There are two additional important differences between oogenesis and spermatogenesis with regards to their timing. In females, oogenesis begins during fetal development and while the ovary continues developing into a mature organ, the gametes do not increase in number. The PGCs which are destined to become oocytes enter meiosis and freeze part-way through oogenesis in the fetal ovary. In fact, an oocyte does not complete both meiotic divisions unless fertilization occurs.

While males form testicular tissue during embryogenesis, the pre-spermatocyte cells do not begin undergoing meiosis until the animal reaches puberty. This means that the PCGs which give rise to the sperm will need to undergo many rounds of mitosis between their arrival in the genital ridges and meiosis entry. Additionally, unlike females which are born with all of their oocytes formed, males continuously produce sperm for the duration of their lifespan once they have entered puberty.

These differences in form and function require different developmental pathways, specifically in timing of meiosis and epigenetic states which allow male PGCs to continue mitotically dividing.

Male germ cell specification

PGCs begin sex-specific development when the sex-determining region of Y (*SryY*) is expressed at E10.5 (55; 84; 164). *Sry* is commonly referred to as the "male-ness" gene that resides on the Y chromosome. *Sry* is a potent gene regulator and the cascade of gene activation which follows *Sry* expression is imperative to the development of male germ cells. Interestingly, *Sry* does not exert this transformation in the PGCs themselves, but in the surrounding somatic tissue which supports the developing gametes. While the regulating mechanism is not clear, cells containing a Y chromosome in the center region of the genital ridges begin expressing *Sry* with expression turning on in a wave that proceeds towards the edges of the ridge (1; 26; 184).

Sry expression upregulates *Sox9* expression in a subset of surrounding somatic cells which specifies them as pre-Sertoli cells (27). The number of *Sox9* expressing pre-Sertoli cells is then increased through various mechanisms including increased proliferation of the specified cells (156). The expression of *Sox9* is imperative to the formation of the male gonads, and *Sox9* expression has proven to be sufficient for testis development in a genetically female embryo (179). This means that even though *Sry* provides the link between the male chromosome Y and "male-ness," it is the downstream *Sox9* which provides the signals needed for male gonad development.

Sox9 expression also plays an important role by re-activating both the maintenance and *de novo* DNA methylation pathways in the cell by promoting the transcription of both *Dnmt1*, *Dnmt3A* and *Dnmt3l* at E13.5 (161). As the original DNA methylation has been depleted, this reactivation allows the genome to be given a novel methylation pattern which is then maintained throughout the mitotic divisions of the PCGs. By turning on these pathways, the male PGCs have increased their methylation level from 15% to 50% three days later (161). It is important to note that even though these cells have reestablished their methylation levels compared to their most depleted state, they have approximately half the amount of methylation found in the surrounding somatic cells. Even though these cells have obtained methylation signatures, they are much more "plastic" than non-PGCs.

The somatic pre-Sertoli cells undergo their transition to supportive Sertoli cells by physically relocating themselves to bind to each other and, through exclusionary action, wrap around clusters of PGCs, forming a protective layer (45; 54; 112; 116). This "wrapping" action results in the formation of testis-specific tubular structures known as "testis cords," which are the precursor to the seminiferous tubules which constitute the adult testis.

While sex-determination has occurred within the embryonic gonads based on the presence or absence of a Y chromosome, the embryonic kidneys bathe the genital ridge in retinoic acid (RA) which promotes entry into meiosis of the PCGs (10). While female cells are capable of responding to this signal, the male PGCs produce Cyp261 and then Nanos2 which blocks the activity of RA either directly or indirectly, respectively (85; 172).

Upon reaching puberty, the testis cords hollow out and form the mature seminiferous tubules. This physical change is accompanied by the maturation of the pre-Sertoli cells into functional Sertoli cells. These Sertoli cells form a niche between blood vessels and testosterone-producing interstitial cells which encourages the differentiation of the PGC population into undifferentiated type A spermatogonia (192). These undifferentiated type A spermatogonia effectively act as spermatogonial germ cells moving forward.

The continuous production of sperm cells is enabled by the characteristics of these spermatogonial germ cells. When the undifferentiated type A spermatogonia divide, the daughter cells can either continue along the spermatogenic pathway to continue their differentiation or can maintain the same identity as the parental cell. Daughter cells are specified to continue on the spermatogenic path by expression of SFC (150) and exposure to low levels of glial-derived neurotrophic factor (GDNF, (115)). Alternatively, exposure to increased levels of GDNF promotes the cells to maintain the undifferentiated type A spermatogonia fate, allowing for sperm precursor cells to be replenished (115).

This switch is controlled by nearby somatic Sertoli cells which secrete GDNF. The Sertoli cells themselves are controlled by follicle-stimulating hormone (FSH) which is released by endocrine system. This response pathway allows the endocrine system to balance the current production of sperm with the maintenance of precursor cells to ensure future production.

Once the daughter cell enters the spermatogenic pathway it undergoes multiple mitotic divisions and cycles through many different cell stages (Figure 4). The last mitotic division in spermatogenesis produces the primary spermatocytes, which enter

meiosis. Due to activity from *Stra8* (the original target of RA) the primary spermatocytes undergo a round of DNA replication and then go through an extensive meiotic prophase which results in the physical pairing of the homologous chromosomes, discussed in detail later on. The homologous chromosomes are separated into the secondary spermatocytes and then a second round of division segregates the sister chromatids to the spermatids.

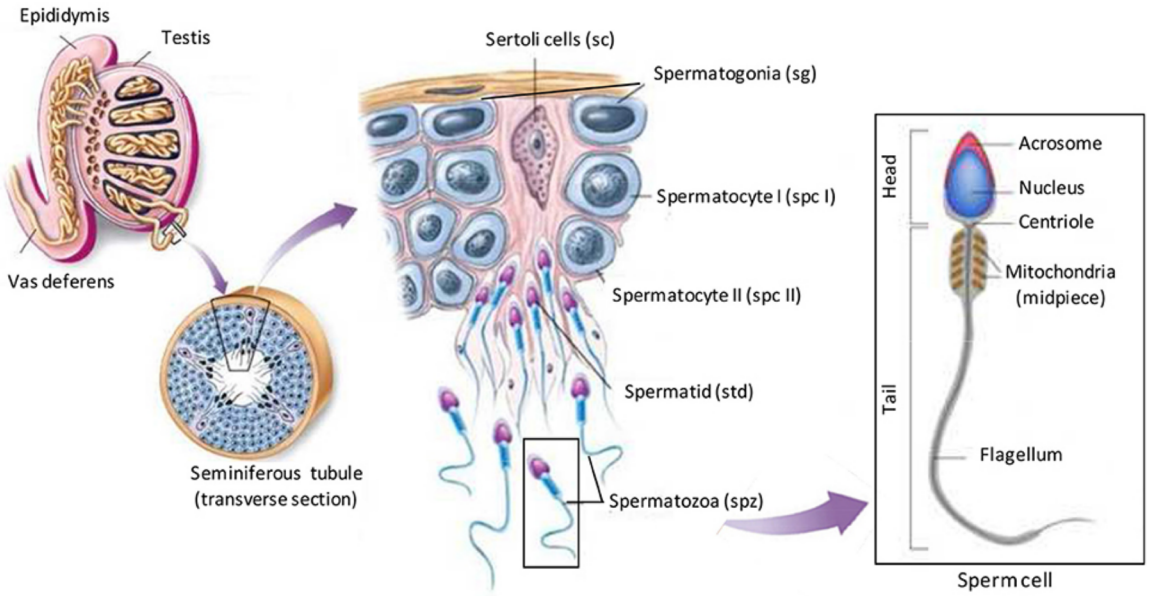


Figure 4: Sperm development.
Figure from (2).

During spermatogenesis, each of the four meiotic cell products goes on to form a fully functional, mature sperm. This maturation of the spermatid involves extensive cellular reorganization: the cytoplasm is removed, the nucleus is condensed further and a flagellum is grown while the mitochondria are relocated to power the tail. Inside the nucleus, histones are actively removed or exchanged for sperm-specific histone variants which will likely be lost upon fertilization. The removed histones are replaced with protamine, which allows condensation of the chromatin and also effectively shuts off all

transcription (49). Additionally, it is thought that the PGC pattern of DNA methylation is finally lost during this last stage of maturation, yielding nuclei containing only the male-specific imprinting (185).

Female Germ Cell Specification

If the Y chromosome is absent, as in female cells, the PGCs at the genital ridge proceed along the female gonadal developmental pathway. Retinoic acid, which is secreted from the embryonic kidneys, or mesonephros, is free to stimulate the PGC to enter meiosis through the transcription of *Stra8 (54; 85)*. As there is no further mitotic division in the PGCs, this means that each PGC at E11.5 begins oogenesis and the overall number of oocytes in the animal is determined at this stage.

The designated oocytes replicate their DNA and begin to undergo meiosis. This early timing gives the female germ cells a different nuclear environment during the early stages of meiosis: both the *de novo* and maintenance DNA methylation pathways are still inactive (Figure 3, (153; 177)). However, the meiotic progression is halted and the cells are maintained at meiotic prophase until puberty. This halt in progression is known as the dictyate state and in mice can draw out the completion of the first meiotic division for months after meiotic entry.

Similar to male mice, oocytes have a network of supportive somatic cells which provide nutrients and the signals for oogenesis to continue following puberty. In males the supportive cells form the physical structure of the testis and provide developmental niches for spermatogenesis. In females, however, the supportive cells form a tight association with each oocyte: the primordial follicle. This leads to differences in

gametogenesis entry between males and females. In male mice, the hormonal signals lead the entire population of support cells to promote the precursor germ cells to enter meiosis in a wave. In females, individual populations of support cells promote the meiotic reentry of their contained oocyte, while the majority of primordial follicles do not react to the hormonal stimulation.

Following puberty, gonadotropin-releasing hormone (GRH) is released by the hypothalamus and leads the pituitary to release follicle-stimulating hormone (FSH, reviewed in (48)). The FSH promotes small populations of primordial follicles to continue development by leading to a proliferation of support granulosa cells around the oocyte as the oocyte undergoes a 500-fold volume increase. Luteinizing hormone (LH), also from the pituitary, leads to the production of mitosis-promoting factor (MPF) through interactions in the follicle cells. Expression of MPF in the oocyte releases it from dictyate and allows for the first meiotic division to occur.

As the primordial follicles develop to become "fertilizable," the *de novo* and maintenance DNA methylation pathways are reactivated with the expression of *Dnmt1* and *Dnmt3A/B/l (177)*. Additionally, an oocyte-specific isoform of *Dnmt1* (*Dnmt1o*) is expressed in postnatal oocytes. This allows the female imprinting pattern to be applied prior to fertilization.

Unlike the meiotic divisions during spermatogenesis, which give rise to 4 meiotic products, oogenesis produces a single egg cell and this is reflected in the unequal distribution of cell materials through the meiotic divisions. The oocyte has a unique metaphase plate which borders the edge of the cell. When the division is completed one set of homologous chromosomes remains in the large egg precursor, while the other set

of homologous chromosomes is jettisoned from the cell in the form of a "polar body." The polar body consists of only chromosomes with enough cell membrane to allow separation from the egg, and generally is disintegrated without continuing through the second meiotic division. This unequal cytokinesis allows for each oogenic precursor to concentrate all of the cellular material in single cell which will go on to form the mature oocyte.

After the first meiotic division is completed, the oocyte freezes a second time during metaphase of the second meiotic division (64). At this stage the oocyte is "fertilizable" and will only continue meiotic progression if it is fertilized: the joining with a sperm leads to the second division and the jettisoning of the sister chromatids in a second polar body.

MEIOSIS

Meiotic recombination overview

In mitotic cell division (and the second meiotic division), the replicated sister chromatids are held together by the kinetochore, a proteinaceous complex which is anchored to the centromeres. This physical attachment allows for the chromosomes to be arranged on the metaphase plate and ensures they are oriented correctly. This correct orientation allows for the chromatids to be pulled towards opposite poles during anaphase and yields daughter cells with the correct assortment of chromosomes.

Similarly, during the first meiotic division, the paternal and maternal homologous chromosomes need to be correctly segregated. Unlike the replicated sister chromatids though, the homologs are not tied together by a kinetochore. In order to ensure that the

maternal and paternal homologous chromosomes are assorted correctly to the daughter cells, the homologs also form a physical attachment. Unlike the proteinaceous kinetochore attachment however, this attachment is a physical interaction of the chromosomes themselves (30; 81; 147) which is supported through a variety of protein complexes. This arrangement leads to a secondary consequence of meiotic recombination: when the homologs resolve this interaction, genetic material is exchanged between the paternal and maternal chromosomes which contributes to genetic diversity.

Meiotic homologous recombination occurs in a stepwise process following DNA replication and not only involves alterations to the individual DNA strands, but also leads to large-scale changes in chromosome structure. These large-scale structural changes can be seen via microscopy and were historically broken into stages based on the appearance of the chromosomes. The four stages are summarized below with a detailed explanation of the molecular mechanisms following:

Leptotene stage - Double-stranded DNA breaks are deliberately generated along the chromosomes, which promotes a homology search.

Zygotene stage - When homology is found the homologous chromosomes are pulled more closely together. This relocalization corresponds with the building of the proteinaceous synaptonemal complex (SC) along each condensing chromosome.

Pachytene stage - Homologous chromosomes are attached along their length through the "zipped up" SCs. During this time, homologous recombination has resulted in DNA "crossing-over" between the homologs.

Diplotene stage - The SC has disassembled and the homologs are now held together through chiasmata, the visible instance of chromosome crossing-over.

Double-strand break formation and processing

Meiotic recombination is initiated when the meiosis-specific topoisomerase SPO11 introduces a double-stranded break into the chromosome (Figure 5, reviewed in (76)). This break is then processed by exonucleases to reveal a 3' single-stranded DNA (ssDNA) overhang. The ubiquitous RAD51 (86) and meiosis-specific DMC1 recombinases (disrupted meiotic cDNA1, (19; 160)) then coat the ssDNA overhang. The ssDNA unwinds as it is wrapped in the protein filaments, allowing the base-pairs to be accessible.

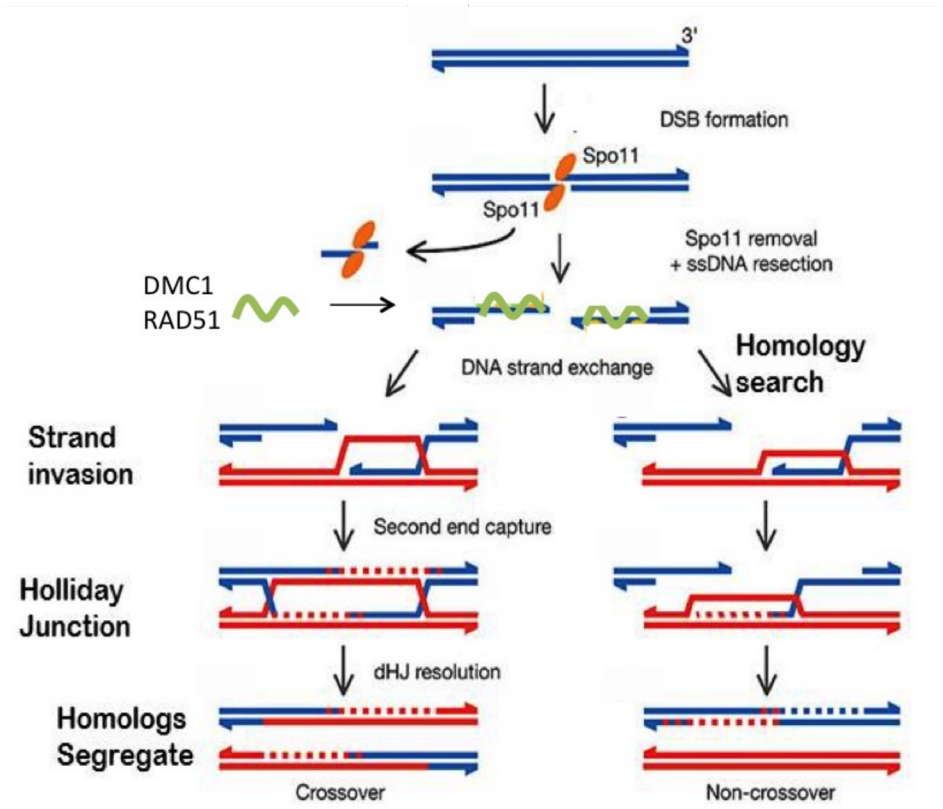


Figure 5: DNA events in meiotic recombination.
Figure adapted from (122).

The protein-coated ssDNA strands then wrap around a region of intact double-stranded DNA, forming a protein filament that bundles the three individual strands together. As the intact double-strand DNA is unwound, the recombinases promote invasion of the single-strand end. The single-strand end enters the dsDNA and displaces one strand while "sampling" the complement for homology through base-pairing (122). Once a region of homology is found, the ssDNA forms a complex through base-pairing and the original complementary strand is excluded. The locations of these breaks can be visualized through staining for γH2AX which is a common marker for DSB breaks (Figure 6).

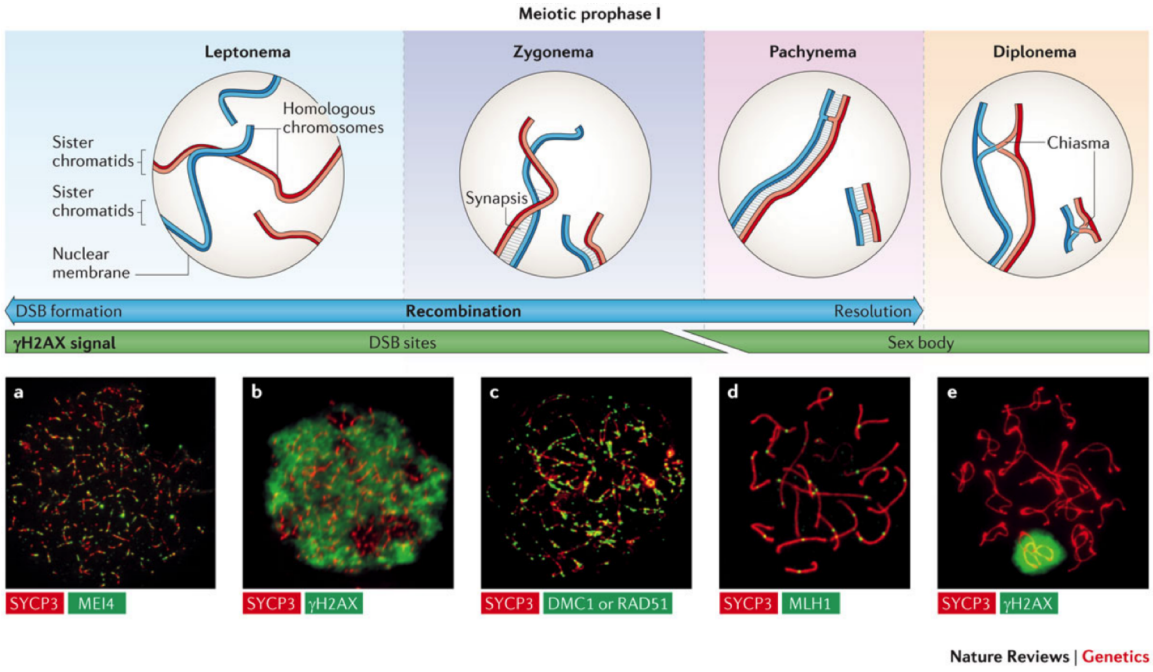


Figure 6: Meiotic prophase progression with immunostaining.
Figure from (13).

Homologous chromosome synapsis

Throughout these interactions the homologous chromosomes have been brought into close proximity. To be more energetically favorable, the displaced strand interacts with the second 3' ssDNA overhang on the homolog. These two interactions are stabilized by DNA polymerase producing a double Holliday junction ((159), see Figure 5). The protein-coated ssDNA filaments can be visualized by staining meiotic nuclei for either RAD51 or DMC1. Once the breaks have been processed and during the search for homology, multiple RAD51/DMC1 foci will be seen along each chromosome. Following strand invasion, 1-3 crossovers form along each mouse chromosome and can be visualized through staining for MLH1 (5; 108)(Figure 6).

Prior to the homology search, the homologs undergo an alignment which places them physically near their homolog. During this alignment, the chromatids are condensing by forming chromatin loops which are secured to the proteinaceous "axial element" which runs along the entire length of the chromosomes and will form the synaptonemal complex (SC). The protein complex which is responsible for initiating homologous recombination, known as the "recombinosome" (29; 31; 181), is already closely associated with the axial elements during this time as the breaks and homology search occur (reviewed in (46)). The mature SC is formed as the respective axial elements of the homologs come together and "zip" up through transverse filaments (131; 157).

Immunostaining can also be used to monitor the progress of the synaptonemal complex formation through staining for SYCP3 (one of the components of the axial element). During zygonema, SYCP3 staining shows filaments corresponding to the

length of each chromosome; as pachynema is completed, thick filaments are seen which correspond to homologs which are bound along their entire length. Another visual hallmark of pachynema is the presence of the "sex body." While the sex chromosomes, X and Y, are not homologs due to their size difference and differing gene content, they do pair during meiosis to ensure correct distribution of the sex chromosomes to the daughter cells. Since they are not fully homologous they do not pair along their lengths but rather form SC only at the pseudoautosomal region. Additionally, as they are not homologs there is no homologous region which would be used to repair the programed DSBs. These differences give the sex body a characteristic appearance. At pachynema, the X and Y chromosomes will still appear as thin SYCP3 stained filaments and the sex body will retain γH2AX throughout even though they have successfully formed the mandatory crossover. This is in contrast to the autosomes, which lose their γH2AX by the time pachynema is complete (Figure 6).

Crossing-over resolution and homolog segregation

Following the completion of a successful chiasmata for each homolog and completion of pachynema, the SC disassembles. Microscopically the thick SYCP3 staining filaments that were characteristic of pachynema revert back to their slim appearance. However, the major difference can be seen as the chromosomes (and SYCP3 filaments) form visible crosses.

The double Holliday junction continues to hold the homologs together as they are arranged on the metaphase plate. During meiotic anaphase I the junction is resolved so that the homologs can segregate faithfully. This pathway yields intact, but separate,

homologous chromosomes that have swapped a portion of chromosome arm with each other (Figure 5, (122)).

Crossing-over versus non-crossing-over pathway

As demonstrated above it is imperative that each homolog forms a crossover for correct homolog segregation. However, the dynamics of a single break finding a suitable region of homology would be time-consuming and unfavorable. Instead, ~200-300 DSB are formed simultaneously in mice, which then undergo a homology search (73). This wide search ensures that at least one break on each chromosome will form a mature crossover in a timely manner.

However, while a large number of homology searches genome-wide is favorable for ensuring at least one chiasmata per chromosome (known as "the obligate crossover rule"), mouse spermatocytes limit the number of crossovers to 1-3 per chromosome based on the chromosome length (reviewed in (12)). A regulatory process known as "crossover interference" then prevents DSBs nearby from forming into crossovers (reviewed in (50)). As the Holliday junction is formed at the sites destined to become chiasmata, the additional DSB sites are resolved through the non-crossover pathway (NCO, Figure 5, (122)). The NCO pathway involves repair of the DSB through DNA synthesis to fill in the 5' ends which were chewed back to promote the homology search followed by ligation.

MEIOTIC RECOMBINATION "HOTSPOTS"

Mapping of recombination sites

From studies looking at the histocompatibility region in mouse it became clear that recombination did not occur randomly along the chromosome, but was organized into discrete "hotspots", which consisted of a few kilobases or less (171). The locations of meiotic recombination "hotspots" were mapped through sperm typing and linkage disequilibrium/pedigree analysis.

Sperm typing (96) involves the collection of sperm from an individual, then either single sperm sorting followed by genotyping DNA markers for crossover junctions (69; 193) or sperm pool typing. While this process does allow for a large-scale investigation into recombination along the length of the chromosomes, sites will only be mapped that have a high rate of recombination. Additionally, sorting of sperm or sperm pool typing are technically demanding and costly.

Linkage disequilibrium, on the other hand, allows for the inference of "historical" hotspots based on records of single nucleotide polymorphisms (SNPs) within a population. Linkage disequilibrium refers to how frequently individual SNPs are nonrandomly associated with each other throughout multiple generations, with recombination sites inferred where there is a break in the linkage disequilibrium of two SNPs over time. Efforts to record such SNPs and subsequent recombination sites in humans have generated many different online databases, the best known being HapMap (37). This type of analysis is limited, based on diversity and limited data on SNPs across populations. In addition, this method only allows us to determine the location of historical hotspots, not to map active sites.

Another concern with these techniques is that they only locate sites of recombination that have completed the crossover pathway, a fraction of the DSBs that are generated for this process. A map of the DSBs generated through meiosis was needed in order to determine factors that regulate location of the recombination sites and to determine how discrete hotspots are formed.

A collaborative effort between our lab and the lab of Dr. Camerini-Otero at the National Institute of Health in the National Institute of Digestive, Diabetes and Kidney resulted in a novel technique to map the single-stranded DSB intermediates and generated the first genome-wide map of meiotic DSB hotspots in mice.

Global mapping revealed genome-wide properties of DSB hotspots in the mouse

Chromatin-immunoprecipitation followed by high-throughput sequencing (ChIP-seq) is commonly used to generate a map of protein binding locations within the genome and requires no initial knowledge of binding sites. The ChIP-seq technique involves the following generalized steps:

1. Tissue sample is fixed, typically using a dilute paraformaldehyde solution.
2. Nuclei are isolated and chromatin is sheared using either mechanical sonication or enzymatic digestion.
3. Antibodies for protein of choice are applied to the sheared nuclear lysate.
4. Antibody/protein/DNA complex is pulled out of solution, commonly using magnetic beads that are coated with Protein A or G which binds the heavy chain of IgGs.

5. Cross-links are removed and DNA fragments that were previously attached to protein of interest are isolated.
6. DNA fragments can then be analyzed by either qPCR (for previously documented binding sites) or used to generate libraries and undergo high-throughput sequencing (allows for discovery of novel binding sites).

However, the ability to map the binding of proteins is dependent on the efficiency of the immunoprecipitation step (134). This limit of detection had initially prevented global DSB mapping, as <2% of cells within the mouse testes have DMC1/RAD51 bound ssDNA intermediates at any given moment. This small population leads to very inefficient immunoprecipitation (14; 114). Originally, our lab bypassed this efficiency problem to generate the first mouse DSB hotspot map by using DMC1-pull down in *Hop2*$^{-/-}$ mice. *Hop2*$^{-/-}$ (also known as *Psmc3ip*) mice exhibit meiotic failure with unrepaired DSBs (142), and this impaired progression results in an abnormally high population of cells containing DSBs compared to the wild type testes.

This initial global mapping effort confirmed hypotheses that the locations that became DSB hotspots were both marked with trimethylation of Histone 3 lysine 4 (H3K4me3) and centered around a DNA-motif that was recognized by the DNA-binding protein PRDM9 (167).

Previous evidence that PRDM9 was imperative for meiotic progression in mice

PRDM9 was initially reported as an ovary and testis specific transcription factor termed Meisetz (meiosis-induced factor containing a PR/SET domain and zinc-

finger motif, (63)). The presence of the PR/SET domain immediately suggested that PRDM9 could play a role in regulation of meiosis-specific gene expression, as the knockout mouse was sterile and PR/SET domains are known for transferring methyl groups to histones to alter gene expression. However, PRDM9 was found to lack the typical H/RxxNHxC sequence that is responsible for the histone methyltransferase activity in other SET domains (146). Interestingly, even though PRDM9 contains mutations to the classic PR/SET catalytic domain, PRDM9 is capable of histone methylation both *in vitro* and *in vivo*.

Both wild type female and male mice showed PRDM9 expression in tissues through leptonema and pachynema (63). Female mice showed expression in the fetal ovary from E13.5-E16.5. While male mice exhibited PRDM9 expression in the testes from 10-18dpp which corresponds with the initial wave of meiotic entry, and with decreased testicular expression at later timepoints corresponding to the non-synchronous meiotic progression which result in lifelong sperm production. This expression pattern and histone methyltransferase activity suggested that PRDM9 was responsible for activating a meiosis-specific set of genes. In line with this hypothesis, mice with interrupted PRDM9 showed disrupted spermatogenesis and oogenesis. While *Prdm9*$^{-/-}$ and wild type mice had similar numbers of DSBs, the breaks were not repaired in *Prdm9*$^{-/-}$ mice and the spermatocytes never reached a true pachytene stage (63).

PRDM9 is capable of methylating the histone H3 at the lysine 4 position: H3K4me3. H3K4me3 is best known as a marker of active gene expression and is commonly found at promoters and enhancer elements in eukaryotes (15; 123; 155; 158). It was concluded that PRDM9 promoted the expression of genes necessary for meiotic

progression as the *Prdm9* knock out had both decreased global H3K4me3 and meiotic failure.

However, studies in yeast suggested that histone modifications and chromatin accessibility might serve a necessary function during meiosis not related to gene expression. Sites which were likely to undergo recombination were known to be "open" during meiosis (119; 128; 188). Additionally, meiotic DSBs were decreased in yeast with impaired H3K4 methyltransferase components (117; 168; 189; 191). Finally, sites likely to undergo DSBs had higher levels of H3K4me3 than expected and this level was independent of the expression of genes encoded in that location (20).

These important findings suggested that H3K4me3 might form a molecular mark which determines the location where DSBs could occur and agreed with previous findings that H3K4me3 is vital for promoting non-meiotic V(D)J recombination in mammals (101; 111).

Global DSB mapping with single-stranded DNA sequencing (SSDS)

To further investigate the role of PRDM9-specific H3K4me3 it was necessary to refine the mapping technique to no longer rely on a specific transgenic mouse. We then shifted focus from improving the efficiency of the immunoprecipitation step to selectively enriching for the single-stranded DNA fragments which had been pulled down within the population. To do this we took advantage of the unique biochemical

characteristics of single-stranded DNA compared to double-stranded DNA during the library preparation steps.

For any DNA sample headed for high-throughput sequencing, the DNA fragments must be modified so that the individual fragments can be loaded on the sequencing machine. The most common protocol involves the following three main steps:

1. Enzymatic end repair which blunts DNA fragments which were produced by random shearing or digestion.
2. Monoadenylation of the 3' end of fragments.
3. Ligation of an "adaptor" to the A-tailed fragment end. (An adaptor is a small section of DNA containing a sequence necessary to bind each fragment to the lane in the sequencing machine. In the case of multiplexed sequencing, the adaptor additionally allows different samples following sequencing to be identified by a unique "barcode" sequence.)

In solution, the single-stranded form of DNA is not favorable, leading pulled-down single-stranded DNA to spontaneously form a hairpin structure. These hairpins are formed by intramolecular interactions between micro-homologies in sequence along the fragment. During the library steps outlined above, single-stranded DNA is processed as a double-stranded fragment with a single end (Figure 7). DNA overhangs beyond the hairpin structure are either removed or filled out, to generate a blunt end for monoadenylation. Single-stranded DNA processed in this only manner makes up 4.7% of the DMC1 ChIP-seq library for wild type adult testes and can be as low as 1% in lower quality samples (79).

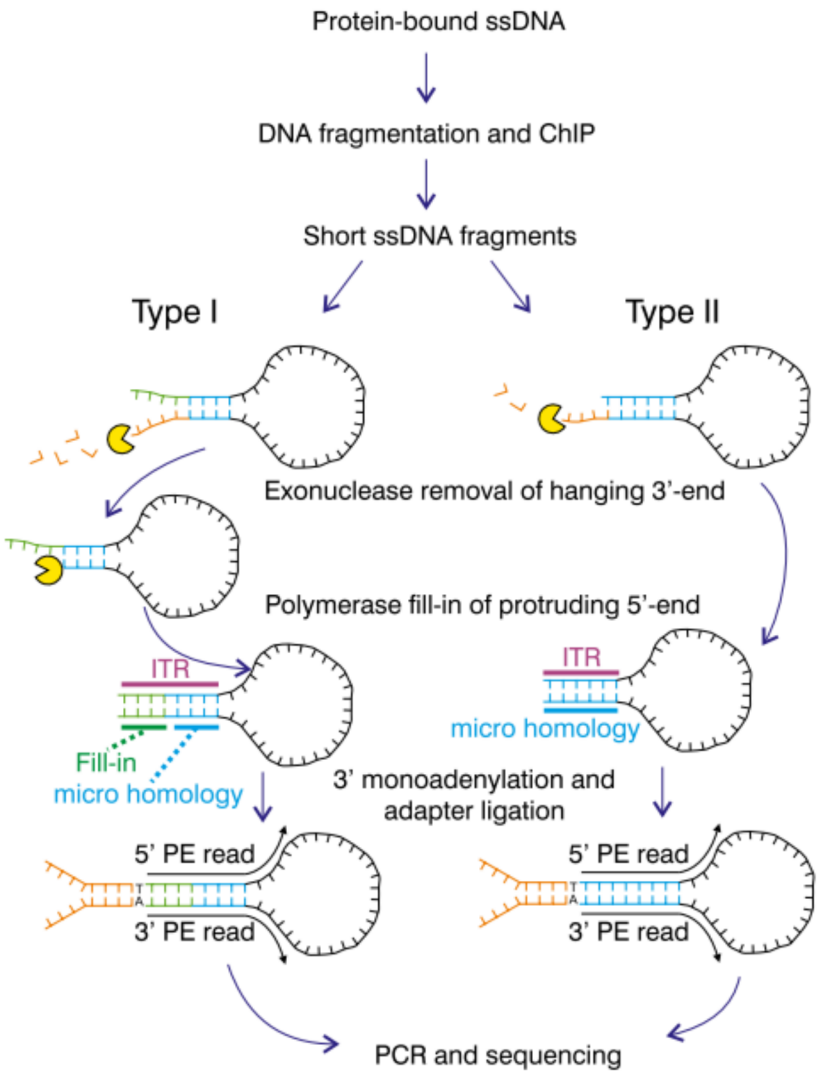


Figure 7: Kinetic enrichment of ssDNA fragments for high-throughput sequencing. Figure from (79).

However, while double-stranded DNA has a remarkably long reannealing time, single-stranded DNA in a hairpin reanneals almost instantaneously (23). The small population of blunted single-stranded fragments can then be positively selected by boiling the samples following the blunting step then letting the sample return to room temperature before ligation of the adaptors. This allows only the single-stranded

fragments to obtain an adaptor, as any double-stranded fragments will no longer be annealed and in the correct structure for addition of the adaptor. Even if these newly single-stranded pieces form hairpins with their own microhomology they will not be blunted and will also miss the adaptor addition.

This specific selection, or "kinetic enrichment," for the single-stranded DNA fragments allows us to map the DSBs in many different breeds and strains of mice following ChIP for either DMC1 or RAD51. Recently we have also been able to map the DSBs in other mammals such as humans (145) and dogs (in progress).

Through this improved method we were able to compare two C57Bl/6 mice which were genetically identical with the exception of the PRDM9 allele. As mentioned above, PRDM9 contains a DNA-binding array of zinc fingers. The DNA motif which is recognized and bound by PRDM9 depends on the contact residues of the individual zinc fingers in the array. This results in multiple alleles of PRDM9 across and within species, each with a unique PRDM9 recognized motif. If PRDM9 was specifically determining the location of the DSBs we would expect that these mice would have different DSB hotspots, in spite of being genetically identical.

As anticipated, mice containing different alleles of PRDM9 each had their own distinct set of DSB hotspots, with very little overlap to mice of a different allele (~1.1%, (22)). Further, the DSB hotspots in each mouse strain were enriched for a DNA motif which corresponded to their PRDM9 allele. These findings confirmed that PRDM9-specific H3K4me3 in the testis served as a marker for DSB formation.

However, if PRDM9 plays this crucial role in promoting DSB hotspots, why do *Prdm9* knockout mice still have similar numbers of DSBs as mentioned above? When

these DSBs were mapped in *Prdm9*$^{-/-}$ mice, it was revealed that in the absence of PRDM9 the DSBs form at locations that already were enriched for H3K4me3 (22). This led to DSBs being formed at genomic elements such as active promoters and enhancers. It was presumably these errant DSBs that led to the meiotic arrest and infertility of the *Prdm9*$^{-/-}$ mouse.

AN EXPANDING ROLE FOR PRDM9 DURING MOUSE MEIOSIS

This intriguing result opened up a new role for PRDM9: not as a promoter of DSBs, but instead as a force which *restricts* the formation of DSBs to areas of the genome where they will cause the least amount of damage. This also suggested that, while PRDM9 was originally described during a screen for meiosis-specific gene regulators, it might not be involved in gene regulation.

While PRDM9 has been established as the marker of DSB formation in mammals, the molecular mechanism of how it is regulated and promotes the formation of DSBs is unknown. In order to investigate both of these questions, we addressed the following two specific aims with this body of work:

1. How is the binding of PRDM9 regulated? While there are over a million instances of the PRDM9-recognized motif within the genome, PRDM9 binding consistently occurs at a subset of approximately 20,000 sites. **We propose that PRDM9 binding is regulated by other chromatin factors, such as DNA methylation, in Chapter 2.**

2. How is DSB formation restricted to sites marked with PRDM9-specific H3K4me3? H3K4me3 is one of the most common chromatin modifications across the genome and it is unclear how DSB formation is restricted to a subset of these sites. **We propose that PRDM9 has an active role in recruiting DSB formation machinery to the sites where it has bound through protein-protein interactions, in Chapter 3.**

CHAPTER 2: MANUSCRIPT 1 - "Extensive sex differences at the initiation of genetic recombination"

Authors:

Kevin Brick[1*], Sarah Thibault-Sennett[2*], Fatima Smagulova[2,3] Florencia Pratto[1], R. Daniel Camerini-Otero[1] and Galina V. Petukhova[2]

Affiliations:

[1] National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

[2] Department of Biochemistry and Molecular Biology, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

[3] Present address: Inserm UMR1085-Irset, 263 ave du Général Leclerc, 35042, Rennes, France

[*] Equal contribution

ABSTRACT

Homologous recombination in meiosis is initiated by programmed DNA double strand breaks (DSBs) and DSB repair as a crossover is essential to prevent chromosomal abnormalities in gametes. Sex differences in recombination have been previously observed by analyses of recombination end points. To understand when and how these differences are established, we built genome-wide maps of meiotic DSBs in both male and female mice. We find that most recombination occurs at sex-biased DSB hotspots. Local chromatin context, the choice of DSB targeting pathway and sex-specific patterns of DNA methylation give rise to these differences. Sex differences also manifest during DSB repair whereby crossovers are suppressed near chromosome ends in females but elevated in males. This uneven repair patterning may be linked to the higher aneuploidy rate in females.

MAIN TEXT

Genetic recombination brings homologous chromosomes together during meiotic prophase and facilitates their segregation at the first meiotic division. Recombination is initiated by programmed DNA double strand breaks (DSBs) that are subsequently repaired as either crossovers (COs) or as non-crossovers (NCOs). The pattern and frequency of recombination can differ between males and females of the same species: the female CO rate is higher in humans and mice (92) while in humans (24), cattle (105), sheep (70) and to some extent mice (132), COs are highly concentrated at sub-telomeric regions in males, but not in females. This pattern is not universal however and in some species, such as pigs, subtelomeric crossovers are elevated in females (174). To date, sex differences in recombination have been studied by comparing the genetic end products of recombination, primarily COs, between the sexes. However, CO-based estimates of recombination rates are limited by the number of sampled meioses and the maximum resolution of such analyses is determined by the spacing of sequence polymorphisms. Most importantly, sex-specific variation that manifests at the initiation of meiotic recombination has not been studied and it remains an open question whether sex differences originate at the time of DSB formation or whether they arise later, during DSB repair. We therefore generated quantitative, high resolution and genome-wide maps of meiotic DSBs in both male and female mice to examine when and where sex-biases in recombination are established, and to elucidate the mechanism(s) that give rise to these biases.

To map meiotic DSBs in female meiosis we exploited a method we previously developed (79) to map DSB hotspots in mouse (22) and human (145) males. This variant

of ChIP-Seq (single stranded DNA sequencing, SSDS) detects single stranded DNA (ssDNA) bound to the DMC1 protein, an early intermediate in the DSB repair process (19; 145). In female mice meiotic DSBs form in the fetal ovary. Each fetal ovary contains just 10,000 cells at the required stage (140), approximately 100 times fewer than in adult male testis. We harvested fetal ovaries at 15.5 dpc (days post coitum) to maximize the fraction of oocytes undergoing meiotic DSB repair (140). We first attempted to map DSBs in females using a pool of 90 ovaries, however the quality of the map was low (Sample O2, Fig. S1). We subsequently pooled 230 fetal ovaries to generate an ovary-derived DSB map that was of similar quality to testis-derived maps (Figure 1A, Fig. S1, S2). 13,297 DSB hotspots were detected from the pool of 230 ovaries, 88% of these were also found in males (Fig. S3A) whereas 12% were exclusively found in females. This is similar to the proportion of human crossover hotspots found only in females (15%) (83). Importantly, hotspots found only in females are weak relative to shared hotspots (Fig. S3B) and contribute just 3% of DSBs. It therefore remains possible that these hotspots are also used in males, but with a frequency below our detection threshold. Thus, we conclude that there are few, if any, hotspots that are used exclusively in female mice. Male-exclusive hotspots were not assessed, because the female samples were not of comparable quality to the best male sample. One intriguing difference between the sexes though, is that DSB hotspots in females are narrower than in males (~400 bp at the widest point; Figure 1B, Fig. S4). This may result from shorter DSB end resection length or from different repair dynamics between the sexes and adds to a growing body of evidence (36; 58) that DSB processing differs substantially between the male and female germ line.

There are striking differences in meiotic DSB repair on the sex chromosomes in males and females. Since X and Y chromosomes only share approximately 700 kb of homology (in the pseudoautosomal region, PAR) (138), an obligate meiotic crossover at the PAR is required in males, the heterogametic sex. Females have two copies of chrX, therefore crossover formation in the PAR is not essential. We found that DSBs in the PAR were highly enriched in all nine males relative to controls, but in neither female (Figure 1C,D). This is consistent with the PAR CO rate being 7-fold higher in males compared to females (169). The dynamics of DSB repair on chrX outside the PAR also differs between the sexes. On the single male copy of chrX, DSBs either remain unrepaired (88) or are continually formed (74) long after autosomal DSBs have been repaired. This prolonged presence of DSBs on chrX biases their detection, such that in wild-type male mice, hotspots on chrX appear disproportionately strong relative to autosomal hotspots. Indeed, we found that compared to autosomal hotspots, hotspots on chrX appear stronger in males than in females (Figure 1E,F). Because of these systematic differences, the sex chromosomes are excluded from subsequent analyses unless explicitly mentioned. Together, we conclude that sex differences are apparent in DSB maps of chromosome X and that the PAR is not particularly recombinogenic in female meiosis.

Although few, if any, DSB hotspots are exclusive to either sex (Fig. S3), hotspot strength differs substantially between females and males. We merged hotspots from the best male (T1) and female (O1) samples (see methods) to create a superset of 20,802 hotspots (19,600 autosomal), then calculated the strength of hotspots for each sample. Hotspot strength estimates using SSDS are extremely consistent as evidenced by the high

correlation between all male samples (Figure 2A, Fig. S5; spearman $R^2 \geq 0.90$). Hotspot strength is correlated to a lesser degree between the female-derived samples (Figure 2B; spearman $R^2 = 0.86$) which may simply reflect noise in hotspot strength estimates from samples with fewer detectable DSBs and a lower FRiP. Most strikingly, hotspot usage between males and females is far more variable than between mice of the same sex (Figure 2C, Fig. S5; spearman $R^2 \leq 0.44$). 43% of autosomal hotspots exhibit sex-biased usage ($N_{male\text{-}bias}$ = 3,978 [20%], $N_{female\text{-}bias}$ = 4,519 [23%]; $N_{unbiased}$ = 11,103 [57%] ; $P < 0.001$, MAnorm (162) - see methods). This is likely an underestimate as we have limited power to detect sex biases at weaker hotspots; at stronger hotspots, where we have the greatest power to detect sex differences, over 60% of hotspots are sex biased (Fig. S6A). DSBs at hotspots biased to their respective sex account for at least 44 ± 1 % and 45 ± 2 % (mean ± s.d.) of the DSB load in males and females, respectively (Fig. S6B). In addition, 15 - 18% of DSBs occur at hotspots biased to the other sex (female-biased hotspots in males or male-biased hotspots in females). Therefore, although most hotspots are used in both sexes, the majority of DSBs occur at sex-biased hotspots. Persistent DSB repair products may affect SSDS-based hotspot strength estimates (38; 88), however this does not appear to contribute to our sex biases (Fig. S7).

PRDM9 targets the vast majority of DSBs to hotspots (22) through DNA sequence-specific binding. In the absence of PRDM9, a default and perhaps ancient mechanism targets DSBs to hotspots at PRDM9-independent H3K4me3 modified nucleosomes (22). Although PRDM9 defines the majority of DSB hotspots in both sexes (Fig. S8) we found that the default targeting pathway is used far more frequently by females (pink & red lines; Figure 2D) than by males (blue lines; Figure 2D, Fig. S8).

Nevertheless, default hotspots are weak, contributing 0.4 - 1.0% of DSBs in females but just 0.3 - 0.4% in males (Figure 2E). Since default hotspots account for 8.5% (388/4,519) of female-biased hotspots but only 0.3% (10/3,978) of male-biased hotspots, we distinguish between hotspots defined by each pathway for subsequent analyses. Default hotspots are used despite an abundance of PRDM9 binding sites (182), which may imply that PRDM9 protein levels are limiting in oocytes. Alternatively, since default hotspots coincide with functional elements including transcription start sites (TSSs) (22), the DNA at these sites may be more accessible in female meiosis. Intriguingly, recombination (CO rate) also increases locally around TSSs in human females, but not in males(16). Although this has been proposed to be a PRDM9-dependent increase, it could be also explained by PRDM9-independent default hotspot usage (Fig. S9). It remains to be seen if the default targeting pathway is active in humans.

In addition to primary DNA sequence, the chromatin context can modulate the usage of DSB hotspots. For example, PRDM9-independent H3K4me3 modified nucleosomes and heterochromatic histone modifications disfavor hotspot usage (38; 182) whereas DSBs are more frequent in actively transcribed regions (167; 182). Since meiotic chromosome packaging differs between males and females (173) we looked for evidence of large-scale epigenetic effects that modulate sex biases. Male-biased, unbiased and female-biased hotspots all occurred in clusters more frequently than expected from random simulations (see methods; Figure 3A). This domain-scale structure was most apparent for female-biased hotspots, for which clusters of between two and nine hotspots were all significantly enriched ($P < 10^{-3}$, empirical simulations; Figure 3A). Cluster size scales with the number of hotspots per cluster and we found no evidence for a physical

limit on cluster size (Fig. S10). A similar proportion of default and PRDM9-defined hotspots were found in clusters (Fig. S11) suggesting that domain-scale regulation of hotspot usage is independent of the mechanism that targets DSBs. Female-biased hotspot domains are enriched at sub-telomeric regions (Fig. S12), and in these regions the DSB rate is high in females, but low in males (Figure 3B). Thus, DSB suppression or altered DSB repair kinetics in males likely give rise to this elevated density of female-biased hotspots. This clustering suggests that sex biases in hotspot usage are partly governed by aspects of local chromatin structure. This is consistent with the known spatial clustering of sex-biased recombination in humans (83) and lends the question whether these sex differences in humans may be established at the onset of DSB formation.

Recombination in subtelomeric regions is of particular interest in the context of female meiosis, because in human oocytes, the chromosomes that have a single telomere-adjacent crossover appear to have an elevated risk of mis-segregation (107; 130). This is likely because telomeric chiasmata are less stable through the long dictyate arrest of female oocytes (107). We found that the DSB distribution in sub-telomeric regions contrasts starkly with that of crossovers; the SSDS signal is high in females relative to males (Figure 3B), whereas, crossovers are more frequent in males (Figure 3C; Fig. S12, S13) (99). Thus, the crossover : DSB ratio increases in males close to the telomere, but decreases in females in the same region (Figure 3D). If distal crossovers indeed increase the likelihood of karyotypic defects in female gametes, these gametes would not lead to viable progeny, and the frequency of such crossovers may be underestimated in pedigree studies. Importantly however, a recent study that examined repair outcomes in fetal oocytes (where there should be negligible selection against karyotypic defects) also

implicated crossover depletion at two subtelomeric hotspots in females (39), in agreement with our results. Thus, it appears that different sex-specific mechanisms exist to modulate DSB repair outcome frequencies in subtelomeric regions. It will be intriguing to see if distal crossover suppression is more extreme in female humans or other mammals, where dictyate arrest can last ten times longer and where negative selection against distal crossovers may therefore be more pronounced.

A striking epigenetic difference between the sexes at the time of DSB formation is that the genome is globally demethylated in females (161) but not in males (154). Since DNA methylation can alter the site preferences of DNA binding proteins (32; 183), it may also affect PRDM9 binding and drive sex dimorphism of hotspot usage. We found that in spermatocytes, the PRDM9 binding site is differentially methylated at male-biased and female-biased hotspots (Figure 4A, Fig. S14; methylation data from (59)). The pattern at unbiased hotspots (Figure 4A) likely stems from methylation at hotspots where we lack power to confidently call biases (Fig. S15). The methylation signal is most pronounced at the hotspots showing the strongest sex bias (Fig. S15) and is absent in females at the time of DSB formation (primordial germ cells at 16.5 dpc (161) (see methods)) (Figure 4B). Thus, DNA methylation at putative PRDM9 binding sites indeed differs between the sexes at the time of meiosis and these sites do not escape demethylation in the germ line (57). Qualitatively similar DNA methylation patterns at biased hotspots are also seen in DNA from somatic tissue of both sexes (mammary gland (42) and male liver (125)) and in spermatocyte precursor cells (spermatogonia (59)) (Fig. S15) indicating that meiosis-specific processes do not give rise to these patterns. The differing patterns of DNA methylation at female-biased and male-biased hotspots

suggest a dual role of DNA methylation in driving sex biases. At female-biased hotspots, two 5' end CpGs incorporating highly conserved guanines are prominently methylated (Figure 4C; green boxes). Methylation of these residues in males may suppress PRDM9 binding and disfavor DSB formation. At male biased hotspots, methylated cytosines are distributed further from the 5' end of the binding site (Figure 4C; yellow box). Many of these nucleotides coincide with a region of the putative binding site that lacks specific sequence preferences but that is nonetheless, important for DSB targeting (165). Such methylation appears to promote DSB formation in males, perhaps by favoring PRDM9 binding (Fig. S16). This may imply that in the absence of DNA methylation at these cytosines in females, PRDM9 has less affinity to its DNA targets, and therefore, the default hotspot targeting pathway is used more frequently. More broadly, DNA methylation affects DNA accessibility by contributing to DNA compaction into heterochromatin. Intriguingly, outside the PRDM9 binding site, the DNA at female-biased hotspots in males is slightly, but significantly more methylated than the DNA at male-biased hotspots (Fig. S17). This may reflect a more compact chromatin organization that inhibits DSB formation at these loci in males.

Bisulfite sequencing, the base-pair resolution mapping of methylated DNA, does not distinguish between methylated (5-mC) and hydroxymethylated (5-HmC) cytosines. Therefore, putatively "methylated" nucleotides may be a compound signature of 5-mC and 5-HmC. We find that 5-HmC in spermatocytes (59) is highly enriched at male-biased hotspots, with lesser enrichment at unbiased and female-biased hotspots (Figure 4D). The 5-HmC signal in sperm is greatly diminished and in contrast to 5-mC, the 5-HmC signal is absent in somatic cells (Figure 4D). Thus, 5-HmC is transiently enriched

at hotspots around the time of recombination, making it unlikely that 5-HmC is contributing to the methylation signal at PRDM9 binding sites (Figure 4A). Intriguingly, 5-HmC has been implicated in the DNA damage response in mitotic cells (71). The transient enrichment of 5-HmC at hotspots in spermatocytes strongly suggests that 5-HmC also plays a role in meiotic DSB repair.

To directly assess if DNA methylation affects DSB frequency, we examined hotspots in mice that lack *Dnmt3L*, a DNA methyltransferase critical to meiosis (21). In *Dnmt3L* mutant mice DNA methylation is reduced genome-wide (180) and at PRDM9 binding sites (Fig. S18). Intriguingly, female-biased hotspots are stronger in *Dnmt3L*$^{-/-}$ mice compared to *wild type*, whereas male-biased hotspots are weaker (Figure 4E, Fig. S19). Thus, DNA methylation indeed appears to mediate sex differences in hotspot usage by favoring DSB formation at male-biased hotspots and by disfavoring DSB formation at female-biased hotspots. Although significant, it is difficult to quantify the effect of demethylation, because DNA methylation is only partially depleted in *Dnmt3L*$^{-/-}$ mice (21). It will be interesting to see if sex differences in recombination patterning are less pronounced in species such as ferrets, where meiosis in both sexes occurs post-partum, possibly in the context of a methylated genome.

Altogether, these data represent the first comprehensive map of meiotic DSBs in a female organism and reveal that extensive and heretofore unappreciated sex biases occur at the onset of meiotic recombination. These biases do not linearly translate into differences in the crossover landscape, revealing the multi-layered complexity of sex-biased control of recombination. We have established that several mechanisms modulate sex biases in hotspot formation including local chromatin structure, choice of targeting

pathway and DNA methylation. Future examination of these mechanisms will yield additional insights into how females and males differentially contribute to shaping evolution of the genome.

ACKNOWLEDGEMENTS

We thank P. Hsieh for critical feedback on the manuscript and members of the Camerini-Otero and Petukhova labs for helpful discussions and suggestions. We also thank the NIDDK genomics core for assistance with sequencing. This study used the high performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov). This research was supported by NIH grant R01GM084104 from National Institute of General Medical Sciences (G.V.P.), March of Dimes Foundation grant 1-FY13-506 (G.V.P.) and by the NIDDK Intramural Research Program (R.D.C.O.). The sequencing data reported in this paper are archived at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) as accession no. GSE99921.

MAIN FIGURES

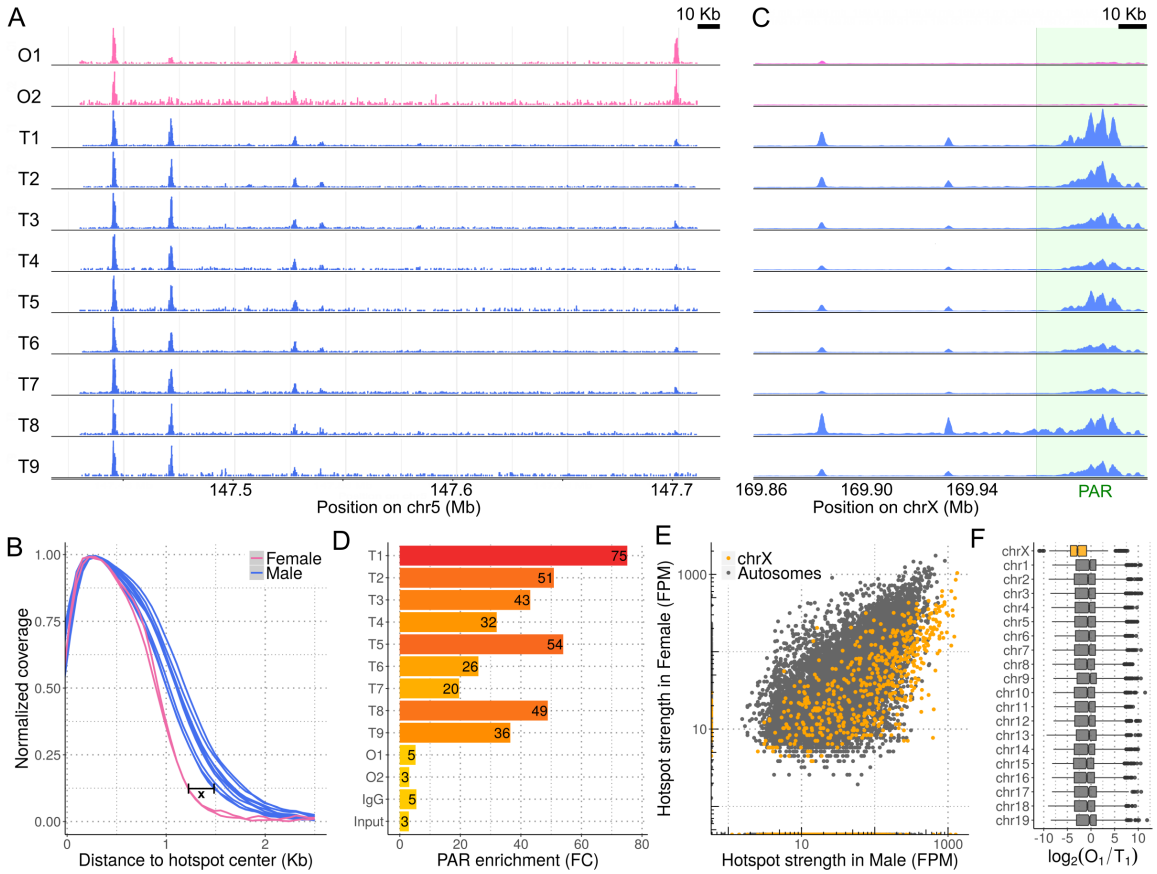


Figure 1. Meiotic DSB hotspots can be detected in oocytes.
(A) DSB maps from ovaries (O1,O2) and from testis (T1-9). (B) SSDS coverage at DSB hotspots is wider in male than in female mice. Coverage is plotted as the 3' distance along ssDNA to the hotspot center. Male samples are shown as blue lines, female as pink. (C,D) PAR DSBs are enriched in male, but not in female meiosis. (C) Coverage is normalized by ssDNA fragments on chrX. The ~50 kb of the mouse PAR adjacent to the pseudo-autosomal boundary (green box) is the only assembled PAR sequence in the genome. (D) Enrichment is the 50 Kb PAR SSDS signal divided by the mean coverage for all chrX 50 Kb intervals; fold change (FC). All male libraries show enrichment relative to controls, whereas female samples do not. The PAR is also enriched relative to other regions for input and IgG samples, likely because the mouse PAR is highly repetitive and poorly assembled. (E,F) DSB hotspots on chromosome X appear stronger in male meiosis. (E) Strength is shown as SSDS fragments per million (FPM). (F) Hotspots in females appear weak relative to males on chromosome X compared to all autosomes.

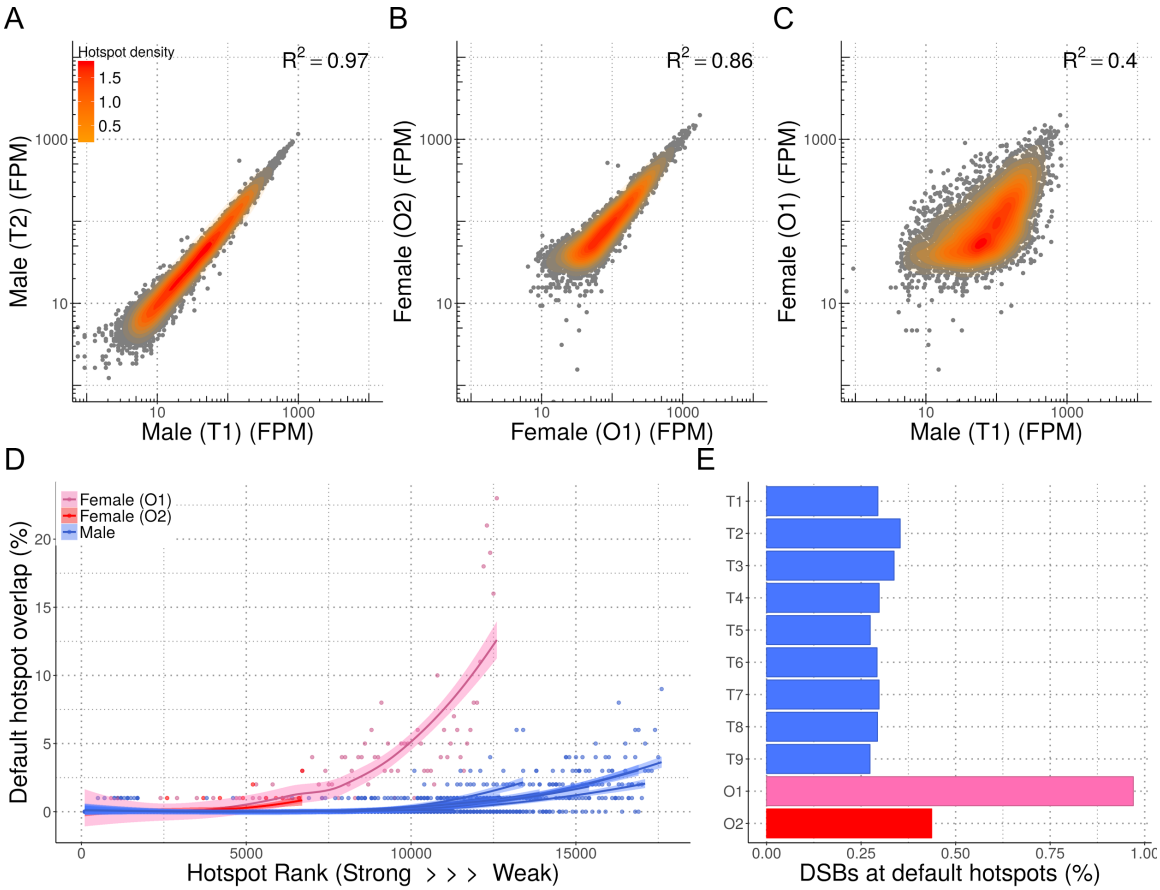


Figure 2. Extensive sex differences in DSB hotspot usage.
(A) Hotspot strength is highly correlated between male samples. In (A,B,C) only hotspots that were called in both samples are considered. (B) Hotspot strength is also highly correlated between female-derived DSB maps. (C) Hotspot strength substantially differs between males and females. (D) Default hotspots are used more frequently in female meiosis. Hotspots identified in each library were ranked by strength and split into 250 hotspot bins. Hotspots that overlap hotspots in a *Prdm9*$^{-/-}$ mouse but that lack a putative PRDM9 binding site were designated as "default" hotspots for each strength-ranked bin. In the female samples, default hotspots were relatively strong. (E) More DSBs occur at default hotspots in female meiosis than in male meiosis.

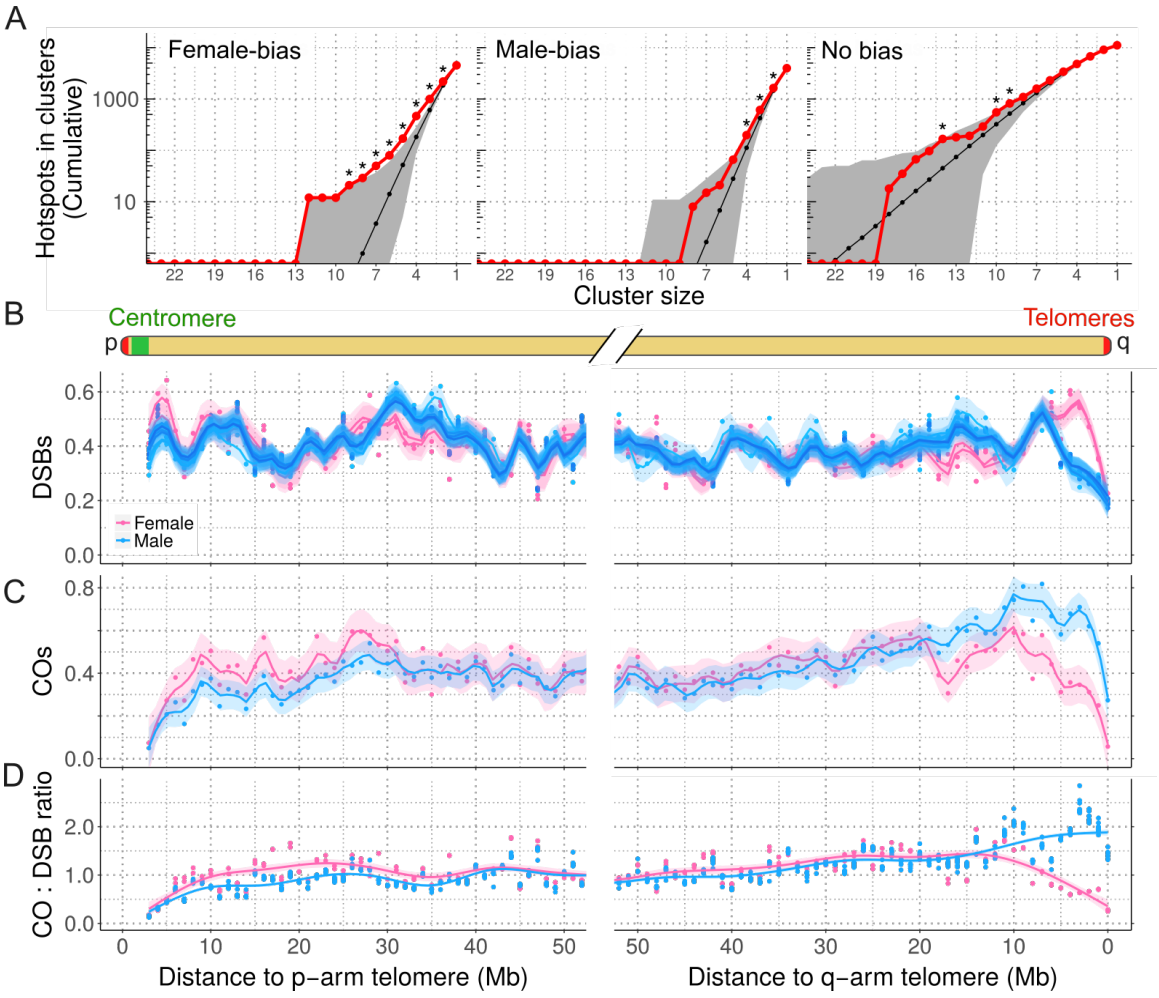


Figure 3. Large-scale influences on sex biased recombination.
(A) Female-biased DSB hotspots are frequently found in clusters. We define a cluster as a group of consecutive hotspots with the same sex bias and cluster size is simply the number of hotspots per cluster. Red line depicts the observed number of hotspots in clusters ≤ a given size. The expected hotspot counts were calculated by randomly shuffling the sex biases of hotspots and generating clusters in the same way (grey area: limits of 10,000 bootstraps; black lines: median bootstrapped count). Asterisks designate cluster sizes that were found more often than expected (empirical $P < 0.001$). (B) Male and female DSB maps deviate adjacent to the q-arm telomere. DSBs decrease in males relative to females in this region (consistent across all male (blue) and both female samples (pink)). We primarily focus on the sub-telomeric region of the q-arm because unassembled centromeric DNA (38) abuts the p-arm telomere in mice. (C) Genetic crossovers (99) also exhibit sex biases adjacent to the non-centromeric telomere. Unlike DSBs, crossovers in this region are elevated in males as compared to in females. (D) Sex dimorphism in the CO:DSB ratio implies that sub-telomeric DSBs are repaired

differently in male and female mice. All profiles are generated from 1Mb non-overlapping windows.

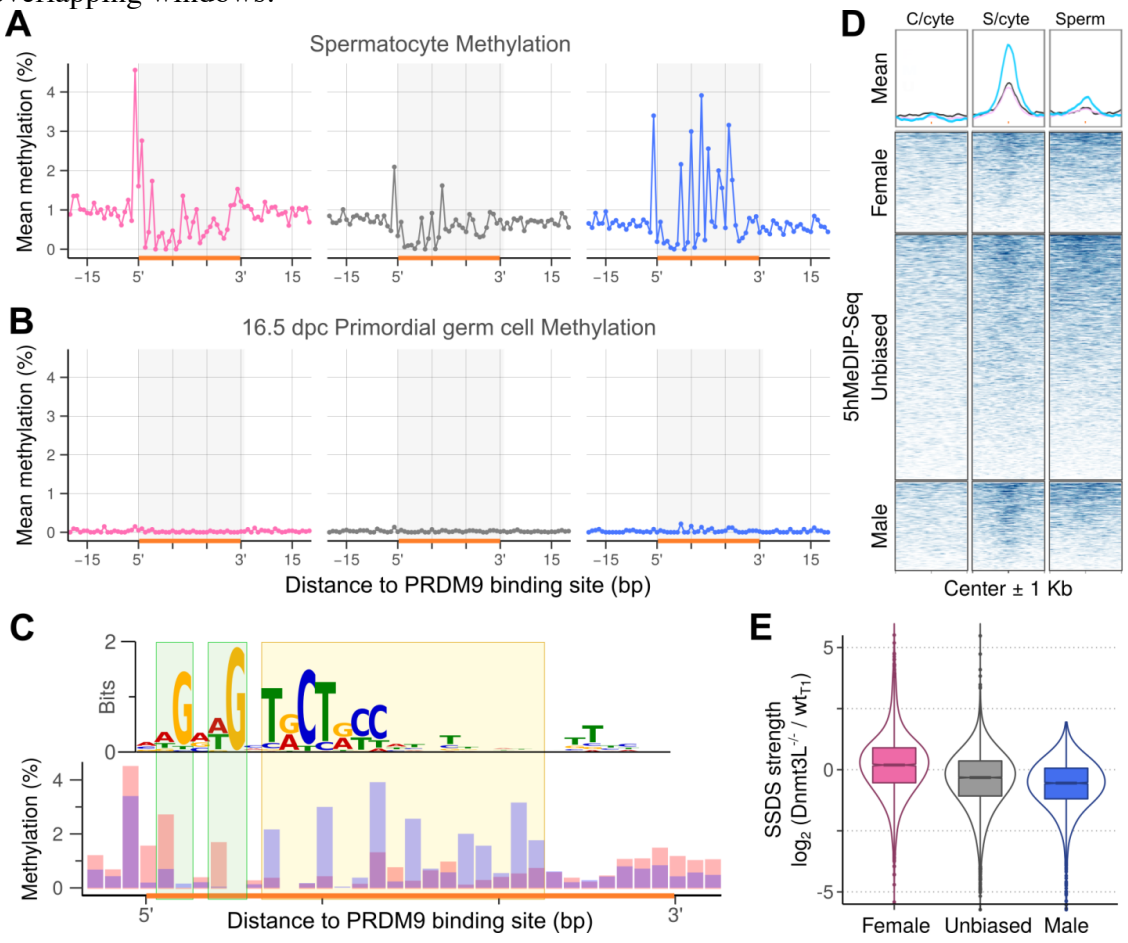


Figure 4. Sex-dimorphic patterns of DNA methylation at sex-biased hotspots. (A) Dimorphic DNA methylation patterns at sex-biased hotspots. The left, center and right panels show average per base methylation profiles for female-biased, unbiased and male biased hotspots, respectively. Sex-biased DSB hotspots exhibit different patterns of DNA methylation. A-D are plotted relative to the position of the putative PRDM9 binding site (orange bar). (B) PRDM9 binding sites are not methylated in fetal ovary. (C) Specific residues in the PRDM9 binding site that are methylated at sex-biased hotspots. The consensus binding site for PRDM9 from C57Bl/6J is shown. Bars represent average methylation in males (blue) and females (red). (D) 5-hydroxymethylcytosine (5-HmC) is enriched at DSB hotspots. Heatmaps depict coverage of hMeDIP-Seq coverage at hotspots in cardiomyocytes (c/cyte) (51), spermatocytes (s/cyte) and sperm. Deeper colors represent higher coverage. 5-HmC is particularly enriched at male-biased hotspots in spermatocytes, although female-biased and unbiased hotspots also show enrichment. (E) Sex-biased hotspot usage is altered in male mice with reduced DNA methylation (*Dnmt3L*$^{-/-}$). In these males, female-biased hotspot strength is increased ($P = 10^{-123}$,

Wilcoxon test) and male-biased hotspot strength is decreased ($P = 10^{-28}$, Wilcoxon test) relative to unbiased hotspots.

SUPPLEMENTAL TABLES

Supplementary Table 1: SSDS sample details

Sample	Source	Antibody	Genotype	Accession #	Start material
N1	Testis (adult)	SC 8973 (c20)	B6 *Hop2*$^{-/-}$	n/a	2 x 10^5 cells
T1	Testis (adult)	SC 8973 (c20)	B6 *wt*	GSM869781	Whole testis
T2	Testis (adult)	SC 8973 (c20)	B6 *wt*	n/a	Whole testis
T3	Testis (adult)	SC 8973 (c20)	B6 *wt*	n/a	Whole testis
T4	Testis (adult)	SC 8973 (c20)	B6 *wt*	n/a	Whole testis
T5	Testis (adult)	SC 8973 (c20)	B6 *wt*	n/a	Whole testis
T6	Testis (adult)	SC 8973 (c20)	B6 *wt*	n/a	Whole testis
T7	Testis (adult)	αDMC1-GP1	B6 *wt*	n/a	Whole testis
T8	Testis (adult)	SC 8973 (c20)	B6 *wt*	n/a	Whole testis
T9	Testis (adult)	SC 8973 (c20)	B6 *wt*	GSM869782	Whole testis
O1	Ovary (15.5 dpc)	αDMC1-GP1	B6 *wt*	n/a	230 ovaries
O2	Ovary (15.5 dpc)	αDMC1-GP1	B6 *wt*	n/a	90 ovaries
TP1	Testis (adult)	SC 8973 (c20)	*Prdm9*$^{-/-}$	GSM869796	Whole testis
DN1	Testis (30 dpp)	SC 8973 (c20)	*Dnmt3L*$^{-/-}$	n/a	Whole testis
iO1	Ovary (15.5 dpc)	n/a : input	B6 *wt*	n/a	230 ovaries

Testis samples were named T1-9. Ovary samples were named O1,O2. Sample N1 was generated from *Hop2*$^{-/-}$ mice using 2 x 10^5 cells. This sample was run in single-end mode, therefore it was not processed through the ssDNA pipeline. Samples that have been previously published are referenced by their GEO accession number.

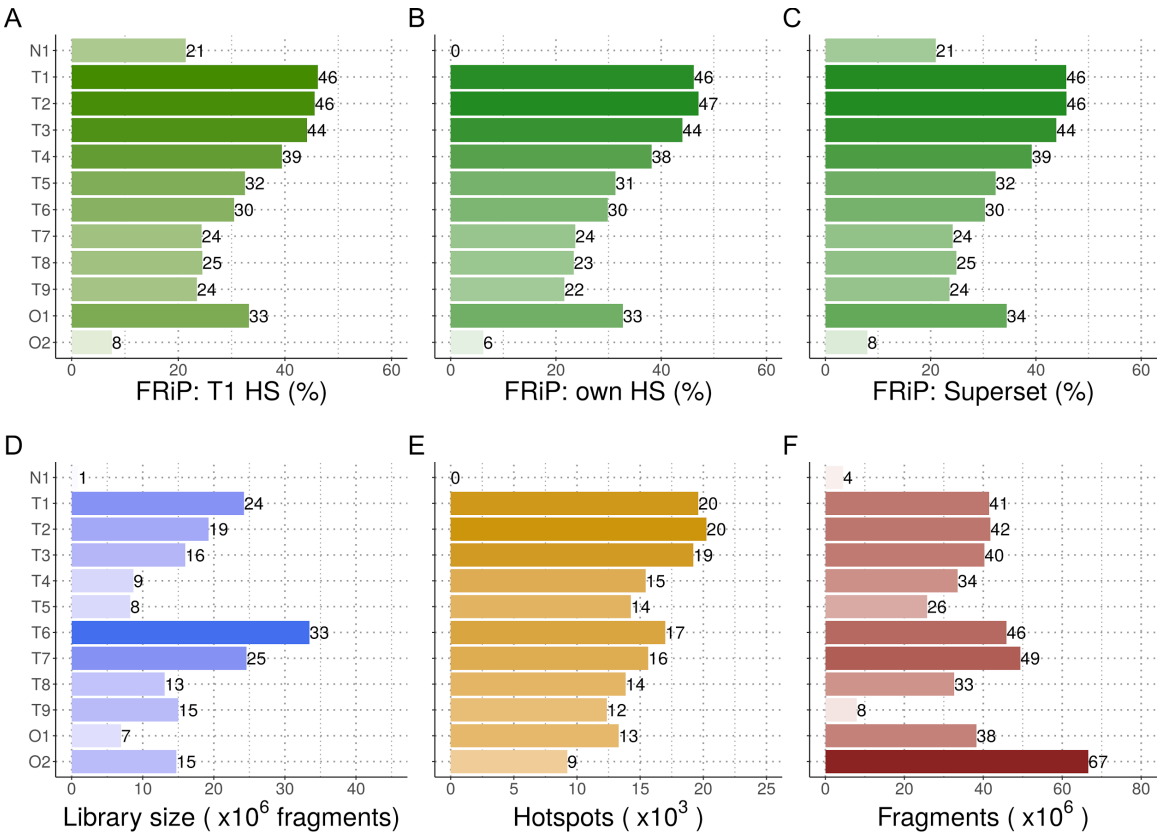
Supplementary Table 2: Other sequencing sample details

ID	**Sample**	**Source**	**Antibody**	**Genotype**
$H3_{W30}$	H3K4me3 ChIP-Seq	Testis (30 dpp)	Millipore (138-473)	B6 *wt*
$H3_{D30}$	H3K4me3 ChIP-Seq	Testis (30 dpp)	Millipore (138-473)	*Dnmt3L*$^{-/-}$

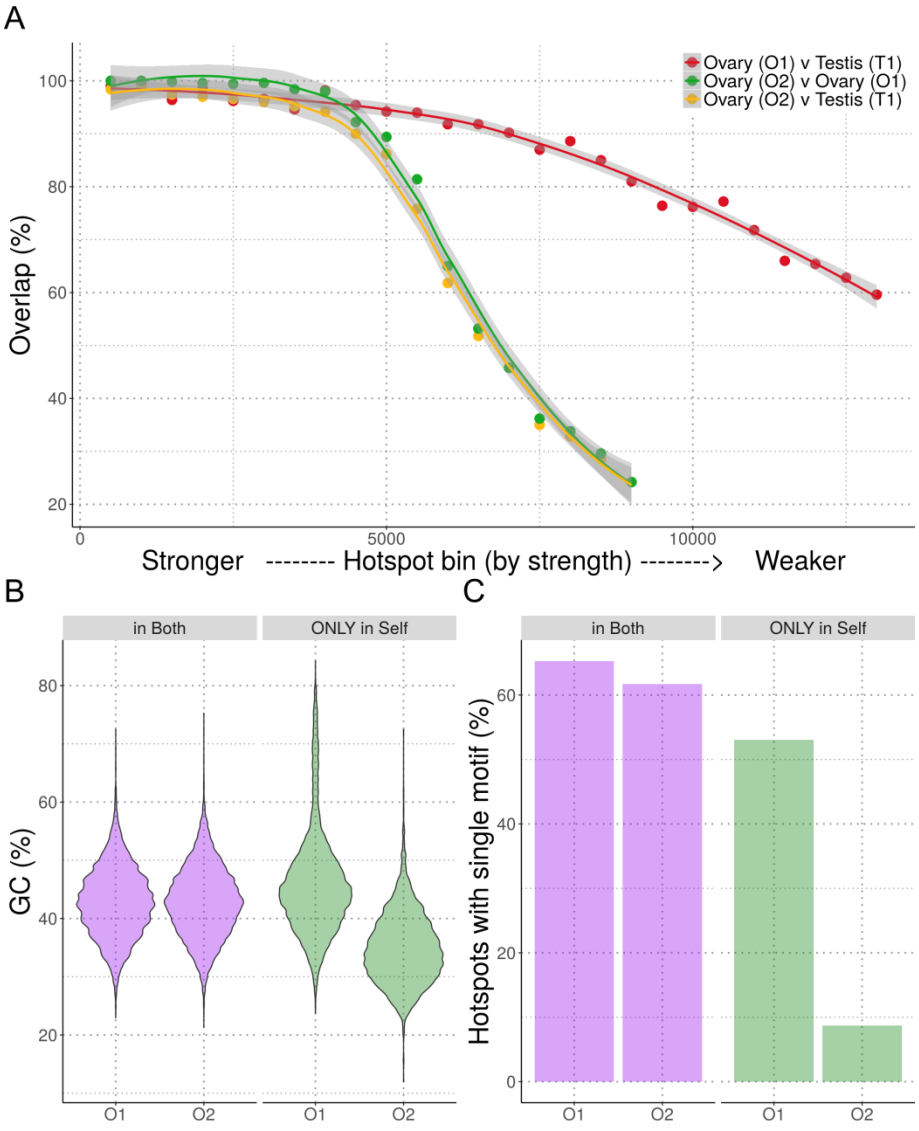
Supplementary Table 3: Public datasets used

ID	Sample	Data accession number	Citation
T1	C57Bl/6 SSDS from testis	GSM869781	(22)
T9	C57Bl/6 SSDS from testis	GSM869782	(22)
PR	*Prdm9*$^{-/-}$ SSDS from testis	GSM869790	(22)
$H3_{H2}$	C57Bl/6 Hop2$^{-/-}$ H3K4me3 ChIP-Seq	GSM869791	(22)
SPO	Spo11-associated oligo mapping from C57Bl/6 mice	GSM2247727	(88)
BS_{SC}	C57Bl/6 whole genome bisulfite sequencing from sorted spermatocytes	GSM1202742 GSM1202743	(59)
BS_{SG}	C57Bl/6 whole genome bisulfite sequencing from sorted spermatogonia	GSM1202738 GSM1202739	(59)
BS_{MG}	Balb/C whole genome bisulfite sequencing from parous basal differentiated cells in female mammary gland	GSM1646791	(42)
BS_{LIV}	C57Bl/6 whole genome bisulfite sequencing from male liver	GSM1857046 : SRR2173867 to SRR2173871	(125)
BS_{PGC}	C57Bl/6 whole genome bisulfite sequencing from female E16.5 primordial germ cells	ERX167033	(161)
BS_{DAA}	*Dnmt3L*$^{A/A}$ whole genome bisulfite sequencing from prospermatogonia	GSM1400327	(180)
BS_{DWT}	*wt* whole genome bisulfite sequencing from prospermatogonia	GSM1400326	(180)
hM_{SC}	C57Bl/6 hMeDIP-Seq from sorted spermatocytes	GSM1202717	(59)
hM_{SP}	C57Bl/6 hMeDIP-Seq from sorted sperm	GSM1202719	(59)
hM_{CM}	C57Bl/6 hMeDIP-Seq from cardiomyocytes	GSM2107809	(51)

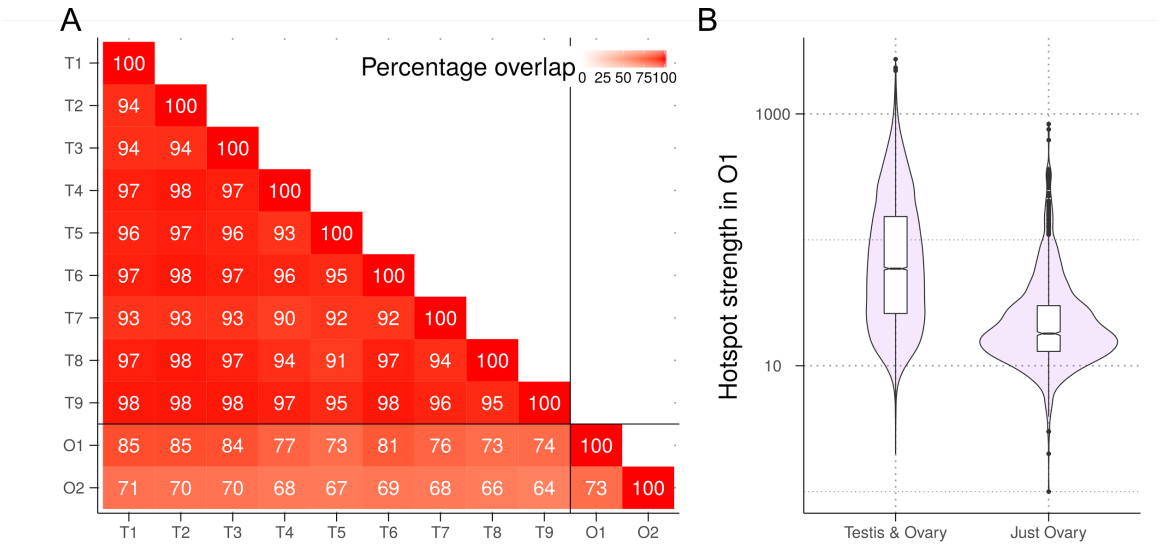
SUPPLEMENTARY FIGURES



Supplementary Figure 1: Sample details and quality metrics for DSB maps
The fraction of reads in peaks is calculated for all samples at (A) hotspots identified in the T1 sample. For sample identifiers, see Table S1. (B) hotspots identified in each respective sample or (C) hotspots in the combined T1/O1 superset. Peak calling was not performed for N1. (D) The estimated library size (x) was inferred using bisection root finding for $f(x) = (1-N_{NR}/x)\text{-}\exp(N_{TOT}/x)$, $10^4 <= x <= 10^{12}$; N_{NR} = # unique fragments; N_{TOT} = # total fragments. (E) The number of hotspots identified in each sample. (F) The number of ssDNA fragments sequenced for each sample.

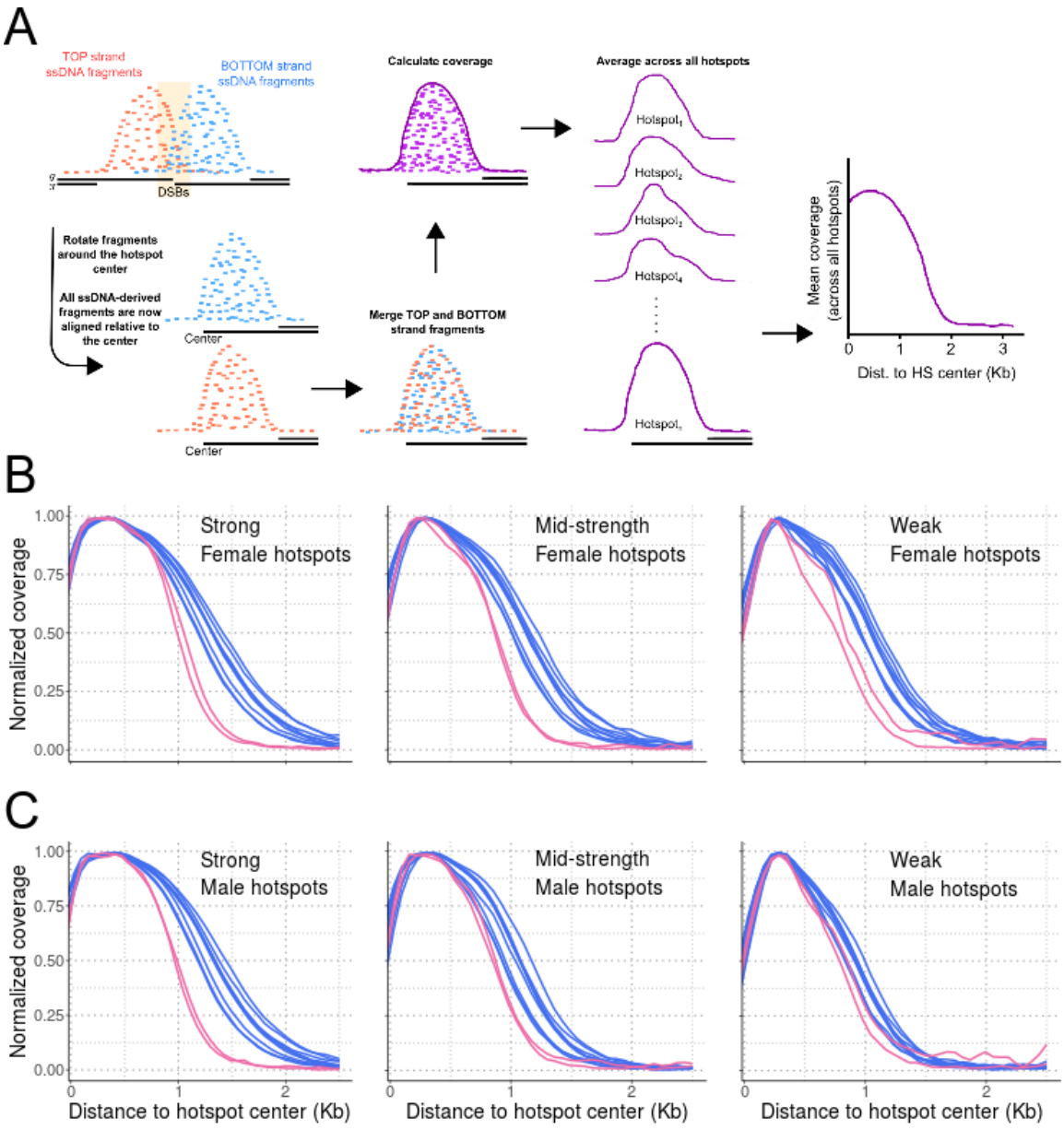


Supplementary Figure 2: DSB hotspots unique to O2 are likely false positive calls (A) Weak DSB hotspots in Sample O2 do not coincide with DSB hotspots in the better testis or ovary-derived samples. Conversely, for the better ovary map (O1), a large proportion of hotspots overlap hotspots identified in testis. DSB hotspots were ranked by strength in the first sample (see key) and for each bin, the percentage of hotspots that are found in the second sample (see key) is shown. (B) Unlike shared hotspots, or hotspots exclusively found in O1, hotspots exclusive to O2 have low GC-content. This indicates that a large proportion of these hotspots likely arise from AT-rich sequencing artifacts. (C) Very few hotspots unique to O2 contain a match to the PRDM9 consensus binding site in the central ± 500bp. These data together suggest that the majority of peaks found in O2 but not in O1 are false positive calls that result from peak calling in a low quality library.



Supplementary Figure 3: Most DSB hotspots are used in both male and female meiosis

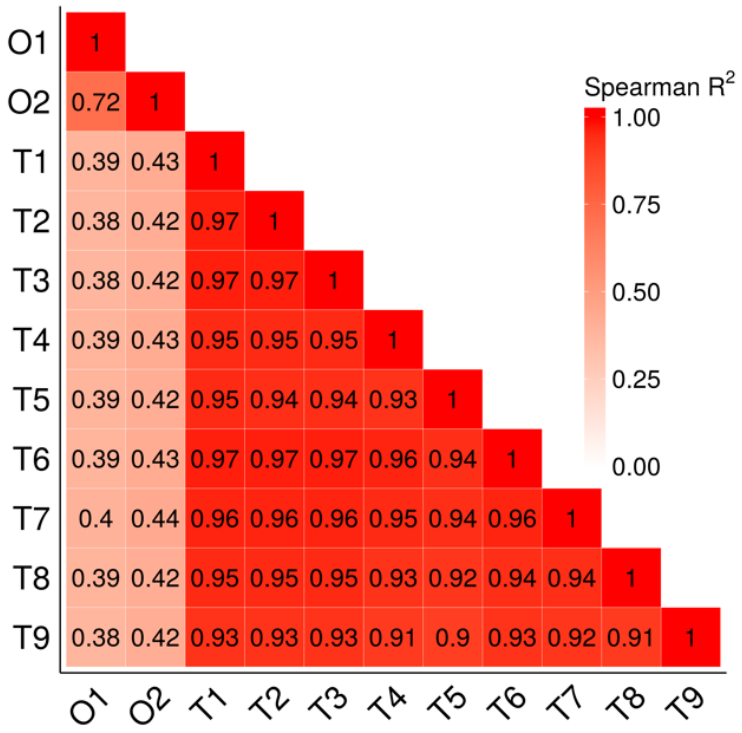
(A) The overlap between hotspots in each sample was calculated using only the central ± 200 bp of hotspots. The maximum reciprocal overlap between samples is shown. As shown in Fig. S1, the overlap with O2 is low, partly because of spurious peak calls that were not removed for this analysis. (B) DSB hotspots found only in the O1 sample and not in any testis sample are weak.



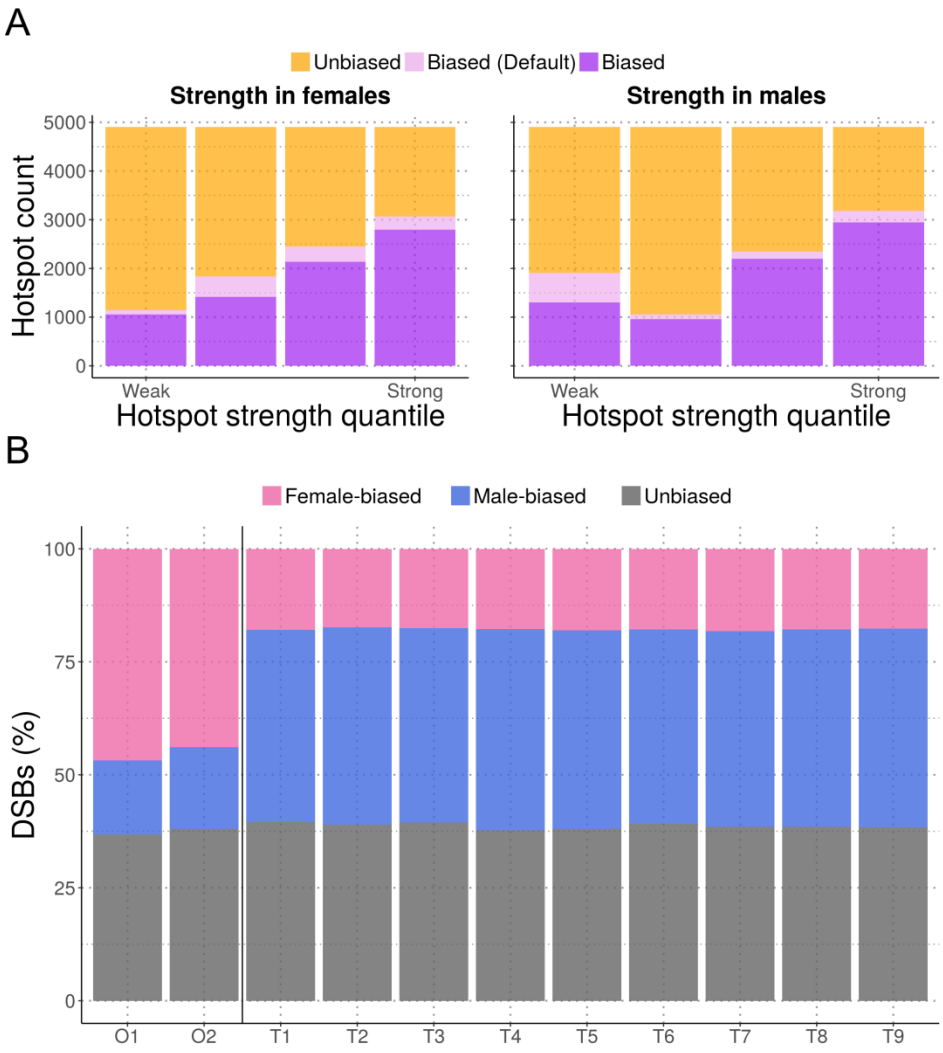
Supplementary Figure 4: DSB hotspots are narrower in ovaries than in spermatocytes

(A) The SSDS coverage is a measure of DMC1-bound ssDNA either side of each meiotic DSB. In a population of meiocytes, DSBs will occur in a several hundred nucleotide window around the hotspot center (orange rectangle). To assess coverage, we first convert the position of each SSDS fragment into the distance along ssDNA from the hotspot center. Merging the top and bottom strand fragments in this way increases coverage two-fold and minimizes the influence of asymmetric gaps and fluctuations in coverage. Coverage at each hotspot was normalized by the maximum value at the hotspot to prevent strong hotspots from dominating the average profile. The average normalized coverage across all hotspots was then calculated. DSB hotspots identified in (B) females (ovary sample O1) and (C) males (testis sample T1) were each split into three bins by

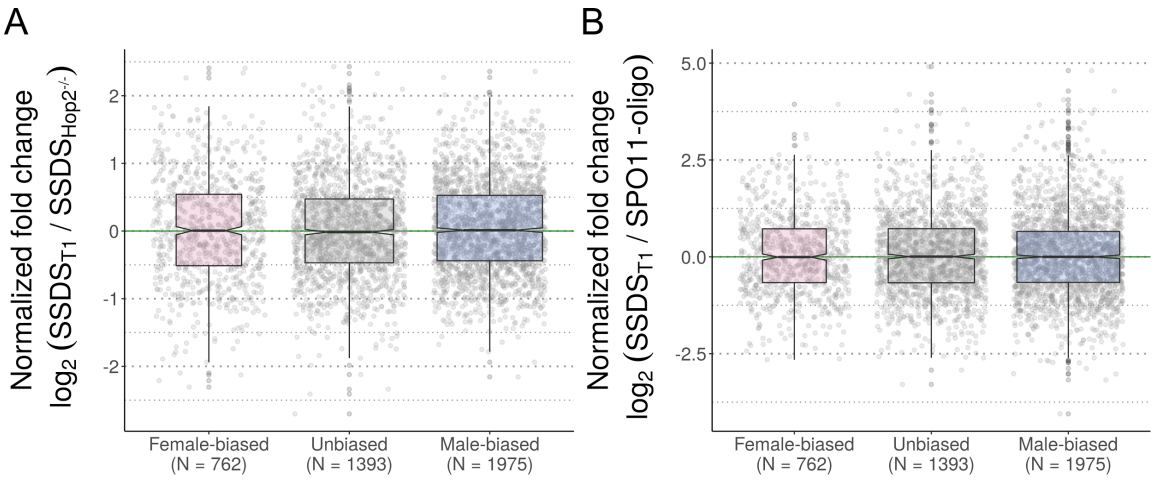
strength. Coverage was calculated for all nine male and two female samples for each set. The SSDS signal is narrower for all female samples compared to male samples. The difference is particularly pronounced at stronger hotspots, where coverage estimates are most accurate. At the widest point, the mean male and female profiles diverge by ~0.4 Kb.



Supplementary Figure 5: Hotspot usage differs between male and female mice
The correlation of hotspot strength was calculated between all testis and ovary-derived samples. Hotspots on sex chromosomes were excluded. The R^2 values differ from those in Figure 2 because here we examine all 19,600 hotspots in the male & female hotspot superset (all hotspots in O1 or T1) for each comparison; in Figure 2, only hotspots identified in the pair of samples being compared were used.

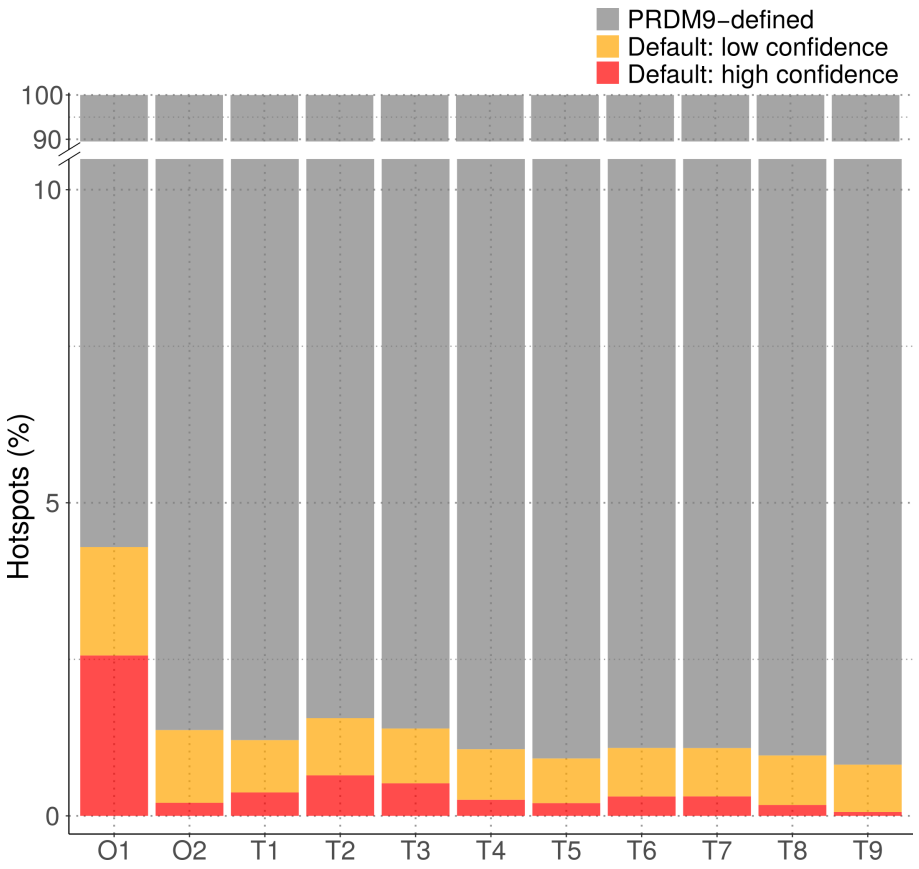


Supplementary Figure 6: Most meiotic DSBs occur at sex-biased hotspots
(A) The hotspots in the testis/ovary super set were split into quintiles by strength in either females (left panel) or males (right panel). In both sexes, over 60% of the strongest hotspot subset exhibit sex-biased DSB formation. This is a proxy for the true amount of sex-biased DSB formation. In progressively weaker hotspot sets, fewer biased hotspots are detected. One outlier is the set of weak male hotspots. This set contains many female-biased default hotspots that form independently of PRDM9. (B) We quantified the total in-hotspot SSDS signal at female-biased, unbiased and male-biased hotspots in the two ovary-derived samples and in the nine testis samples. In all cases, over half of the in-hotspot sequencing tags (referred to as total DSBs) occur at sex-biased hotspots. Hotspots biased towards usage in females are enriched in ovary samples, while those biased towards male usage are enriched in testis-derived samples.

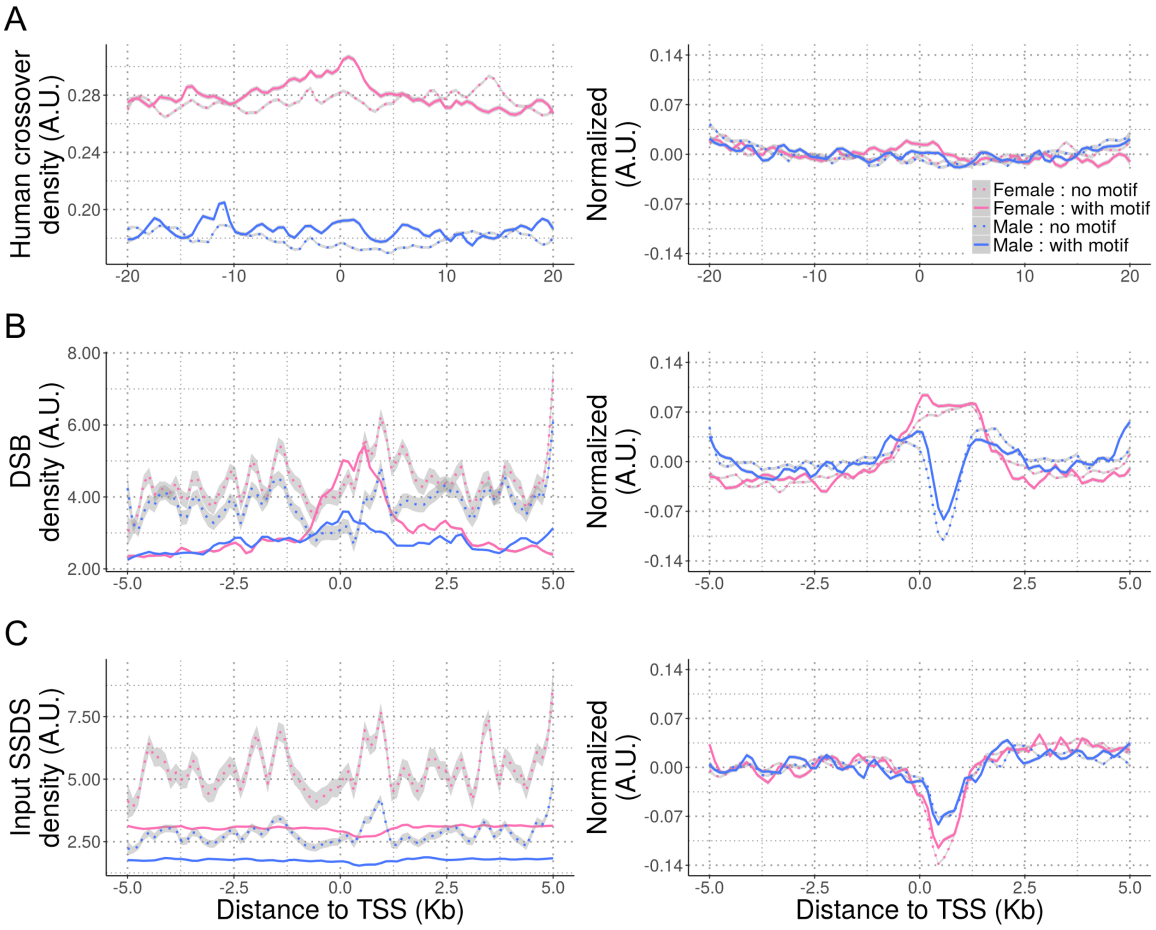


Supplementary Figure 7: Sex-biased hotspots are unlikely to be arising from inter-sex differences in DSB repair dynamics.

Persistent or long-lived DSB repair intermediates can affect SSDS strength estimates. To assess if persistent hotspots disproportionately affect sex-biased hotspots, we compared hotspot strength determined by SSDS signal in *wild-type* males (T1) to two other estimates of hotspot strength that should be unaffected by persistence: (A) SSDS signal in mice that cannot repair meiotic DSBs (*Hop2*$^{-/-}$) and in which all DSBs will persist equally, and (B) SPO11-oligonucleotide mapping. (A) The *wt* : *Hop2*$^{-/-}$ SSDS ratio at unbiased hotspots was no different to that at male-biased (P = 0.25; wilcoxon test) or female-biased hotspots (P = 0.67; wilcoxon test). (B) The wt SSDS : SPO11-oligo ratio at unbiased hotspots was no different to that at male-biased (P = 0.09; wilcoxon test) or female-biased hotspots (P = 0.20; wilcoxon test). For these analyses, to ensure that we used biased and unbiased hotspots of similar strength, the strongest 33% of hotspots were used (strength from the T1 SSDS sample). Log ratios were normalized by the mean log ratio for all hotspots. *Hop2*$^{-/-}$ SSDS and SPO11-oligo data are from male mice. Together, these data suggest that persistence does not influence detection of sex-biased hotspots.



Supplementary Figure 8: Default hotspots are more prevalent in female meiosis
Default hotspots are enriched in the O1 female sample. Because default hotspots are weak and the O2 sample is of poor quality, we lack the sensitivity to determine if default hotspot targeting occurs in the O2 sample. DSB hotspots that overlapped hotspots in *Prdm9*$^{-/-}$ mice were identified as "default" hotspots. Some "default" hotspots contain a match to the PRDM9 binding site for C57Bl/6J (FIMO p-value > 10^{-4}) in the central ± 250 bp. These are classified as low confidence (orange) because they may also be targeted by PRDM9. High confidence default hotspots did not contain a motif match (red). All other hotspots were classified as PRDM9 defined (grey).

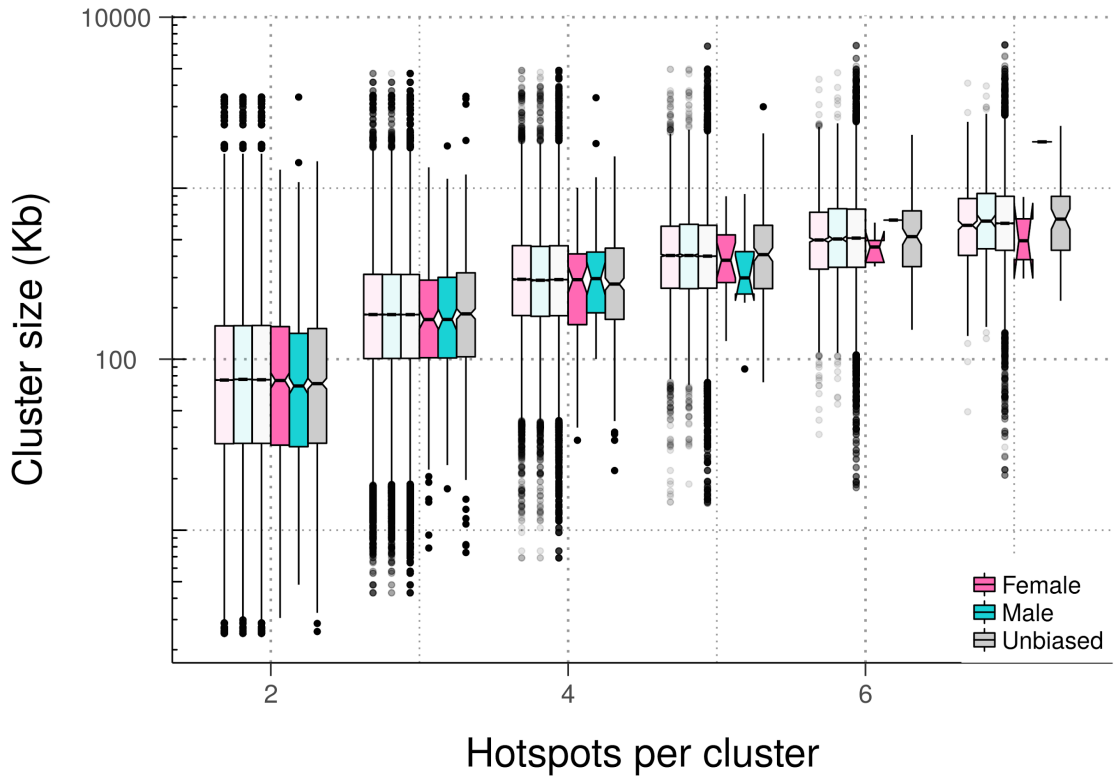


Supplementary Figure 9: Recombination is enriched at TSS in mouse and human females

(A) In human females, crossovers are elevated near TSS that contain a putative PRDM9 binding site. No enrichment is seen at TSS in males or at TSS that lack a putative binding site. This implies that PRDM9-dependent crossover formation may be elevated in human females near TSS(24). (B) In mice, we see a similar pattern: DSB density is highest in females around TSS that contain a putative PRDM9 binding site. We also see a lesser increase in the DSB frequency at TSS containing a motif in males; this is not seen in humans, perhaps because the resolution and the quantitative estimates of recombination are far less precise using crossover mapping. Importantly, if we normalize the DSB density per TSS, we see that the DSB density at TSSs in females increases irrespective of the presence or absence of a PRDM9-binding motif. Furthermore, the center of the TSS-associated peak is downstream of the TSS, consistent with previous data showing that PRDM9-independent hotspots coincide with the +1 nucleosome, just inside the TSS(8). Thus, the DSB enrichment at TSS in female mice is mostly a result of PRDM9-independent DSB targeting. Importantly, comparing the absolute (left) and normalized (121) figures suggests that in mice, the non-normalized DSB density profiles are biased by a few strong outliers; this affects motif-containing TSSs more than those without a motif. It therefore remains possible that PRDM9-independent DSB targeting in humans also contributes to enrichment at TSSs in females. Notably, the male DSB frequency

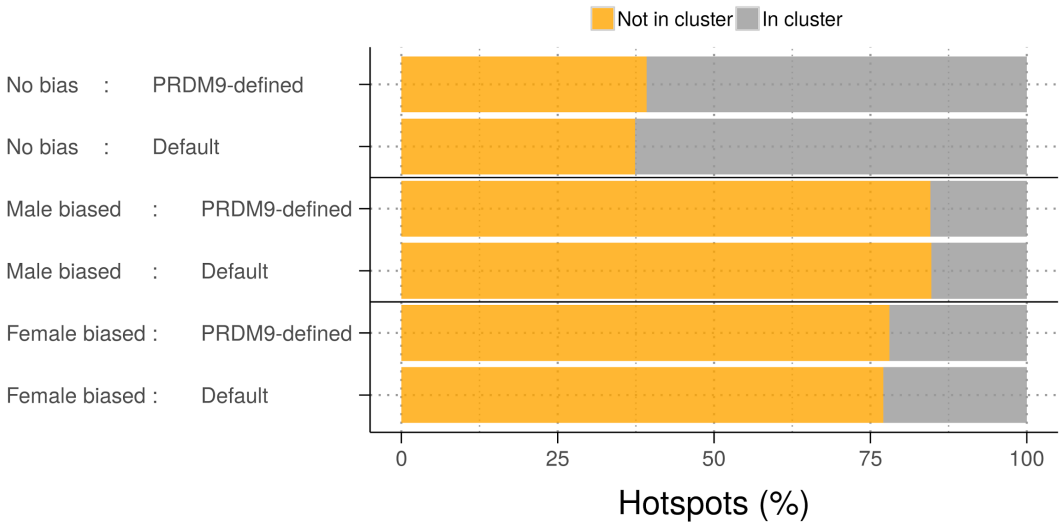
exhibits a pronounced dip downstream of the TSS. (C) A similar dip is seen in input SSDS coverage for both males and females and likely represents a difficult to sequence locus.

Crossovers from male and female humans were obtained from (*43*). To directly compare crossovers with DSB mapping, we generated 10 pseudoreads of 50bp for each crossover. Pseudoreads were randomly distributed across the crossover interval. Pseudoread density was then plotted relative to GenCODE TSSs. For both human (A) and mouse (B,C), we split GenCODE TSSs into two groups; those that contained a putative PRDM9 binding site (solid lines) and those that did not (dashed lines). In humans, we looked for a match to the 13-mer motif for human $PRDM9_A$ (CCNCCNTNNCCNC)(*44*) within 5 Kb of the TSS. For mice, we used TSS with a C57Bl/6J PRDM9 binding site within the central 500bp. (A,B,C) Left panels show the average density around all TSS. Right panels show the average normalized density around TSS. Density at each TSS is normalized by the mean and divided by the standard deviation. The normalized density is better to analyze shape and is not influenced by strong outliers.



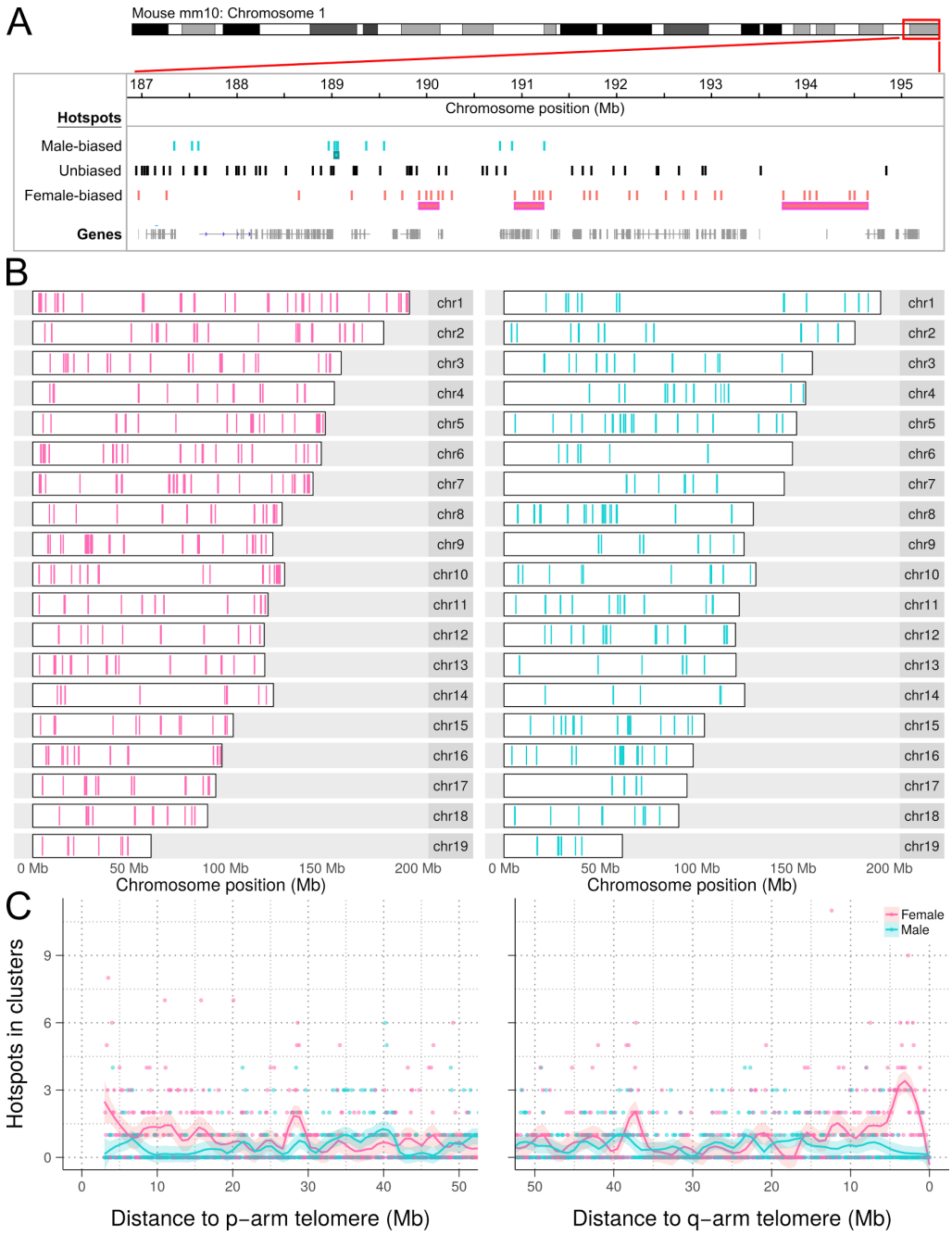
Supplementary Figure 10: The physical size of hotspot clusters scales with the number of hotspots per cluster

It therefore seems unlikely that clustering results from a physical size constraint imposed by sex-specific chromatin structure. Importantly however, the presence of such a size constraint may be masked by the presence of a large number of clusters that occur by chance. Semi-transparent boxplots show the expected size distribution for randomly distributed clusters (n = 1,000 bootstraps). Clusters of three male or female biased hotspots are marginally smaller than expected. There are no significant differences for clusters of other sizes.



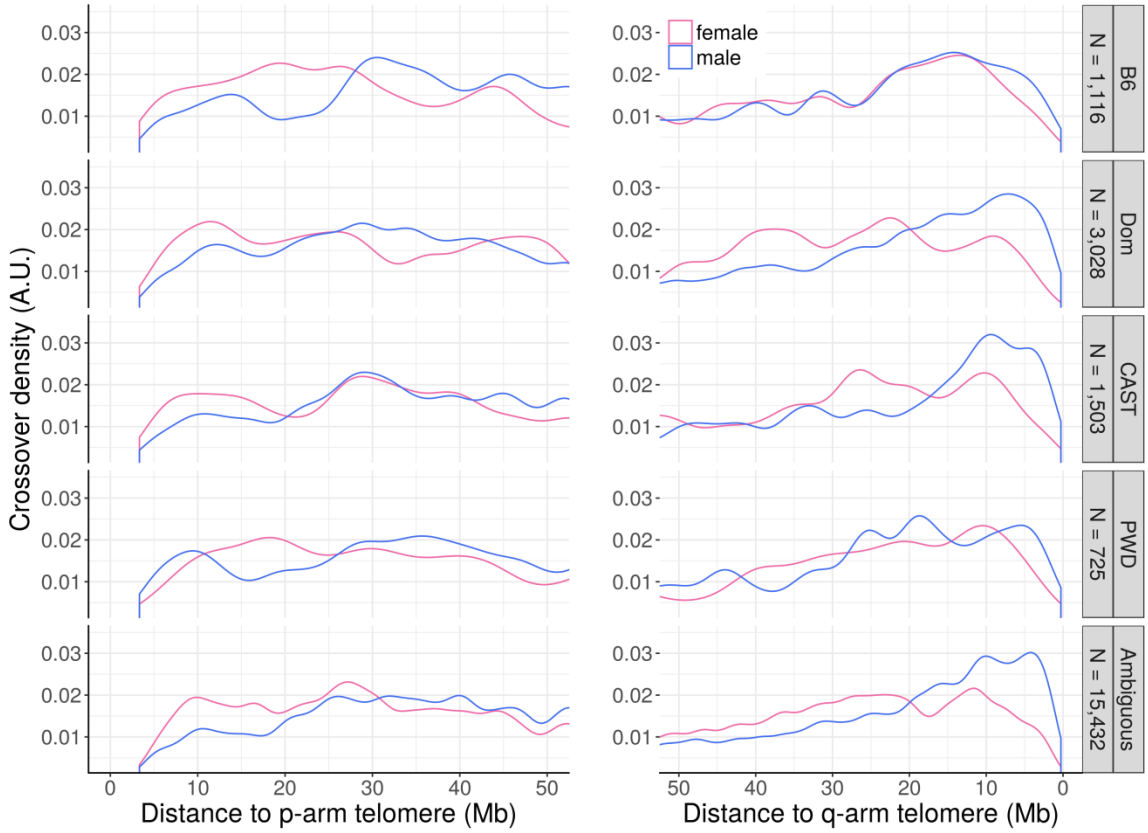
Supplementary Figure 11: Similar proportions of PRDM9-defined and default hotspots occur in clusters

Hotspots in clusters of ≥ 3 consecutive hotspots of the same type were counted.



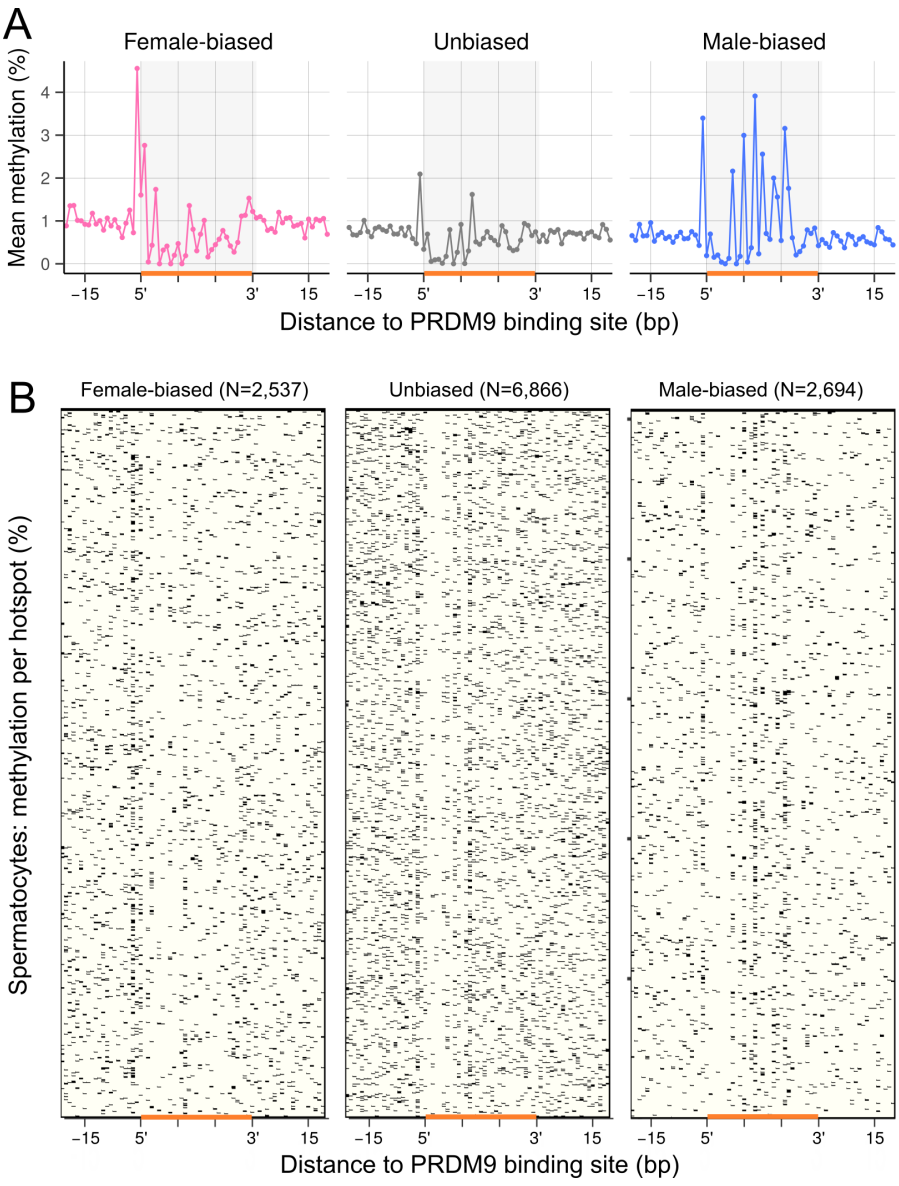
Supplementary Figure 12: Clusters of female-biased hotspots are enriched at q-arm subtelomeric regions.

(A) 17 female-biased hotspots but no male-biased hotspots occur in the distal 4 Mb of chromosome 1. This includes a cluster of 7 consecutive female-biased hotspots at the distal end. Clusters shown as larger boxes underneath the respective hotspot track. (B) Female-biased (left) and male-biased (121) hotspot clusters on all autosomes. (C) Male-biased hotspots do not exhibit particular spatial patterning, female-biased hotspots are found frequently in distal regions, adjacent to the q-arm (non-centromeric) telomere. Enrichment adjacent to the p-arm is also seen, however the centromere may influence this end. The trend in the data is visualized using LOESS smoothing.



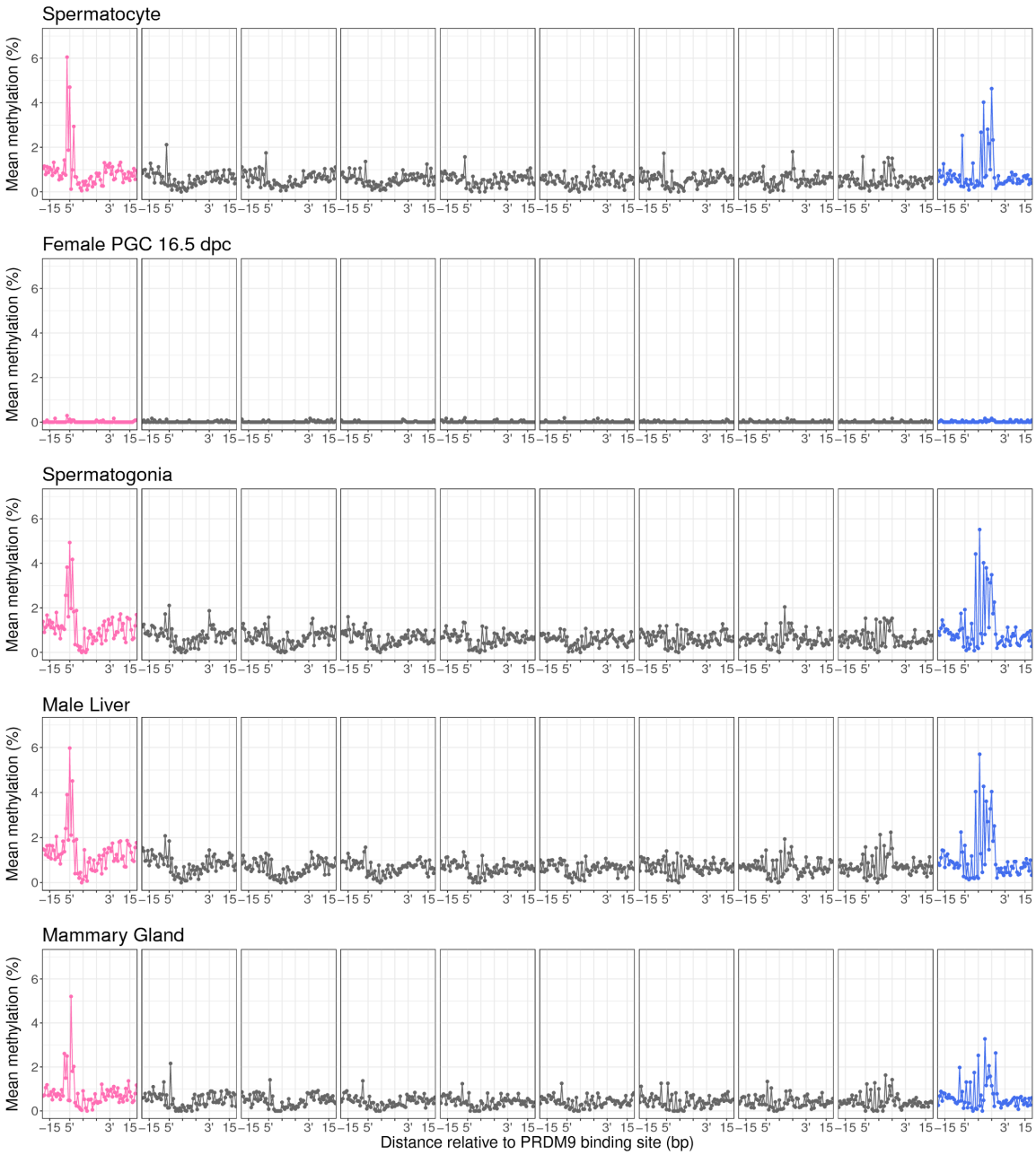
Supplementary Figure 13: Genetic crossovers in males are elevated adjacent to the q-arm telomere, independent of the *Prdm9* allele

To assess if the elevated male crossover rate in the q-arm subtelomeric region is *Prdm9* dependent, we inferred genetic crossovers that were likely to have been formed by the B6, CAST or PWD alleles of *Prdm9*. We also inferred COs that were likely of *Mus musculus domesticus* origin. Briefly, for each grand-parental derived CO, we could infer the two *Prdm9* alleles that may have given rise to that CO. We then compared these CO positions with maps of DSB hotspots for each allele of PRDM9 (if available) (9). If a CO coincided with a DSB in the respective DSB map, it was designated as having been formed by that allele. COs that did not overlap any DSB hotspot from either parental genotype and COs from crosses for which DSB hotspot maps were not available were designated as "Ambiguous". Graphs above show the LOESS smoothed crossover density for each sex.



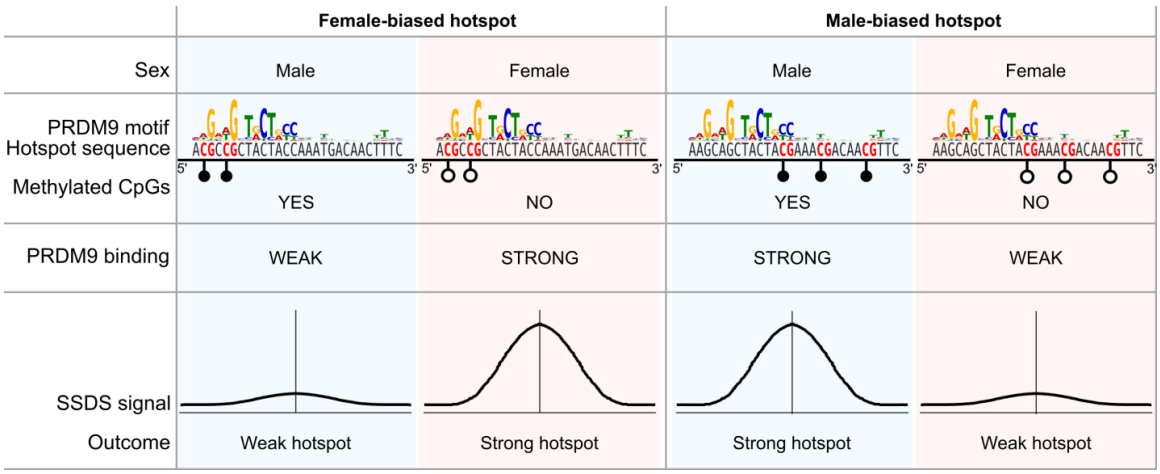
Supplementary Figure 14: Differing patterns of DNA methylation at sex-biased hotspots

(A) Mean DNA methylation at the putative PRDM9 binding site (orange bar) of female-biased (pink), unbiased (grey) and male-biased (blue) hotspots. Note that this panel is also shown in Figure 4A. (B) Heatmap rows depict methylation at individual hotspots. Note that the density of methylation appears higher at unbiased hotspots because rows are more densely spaced.

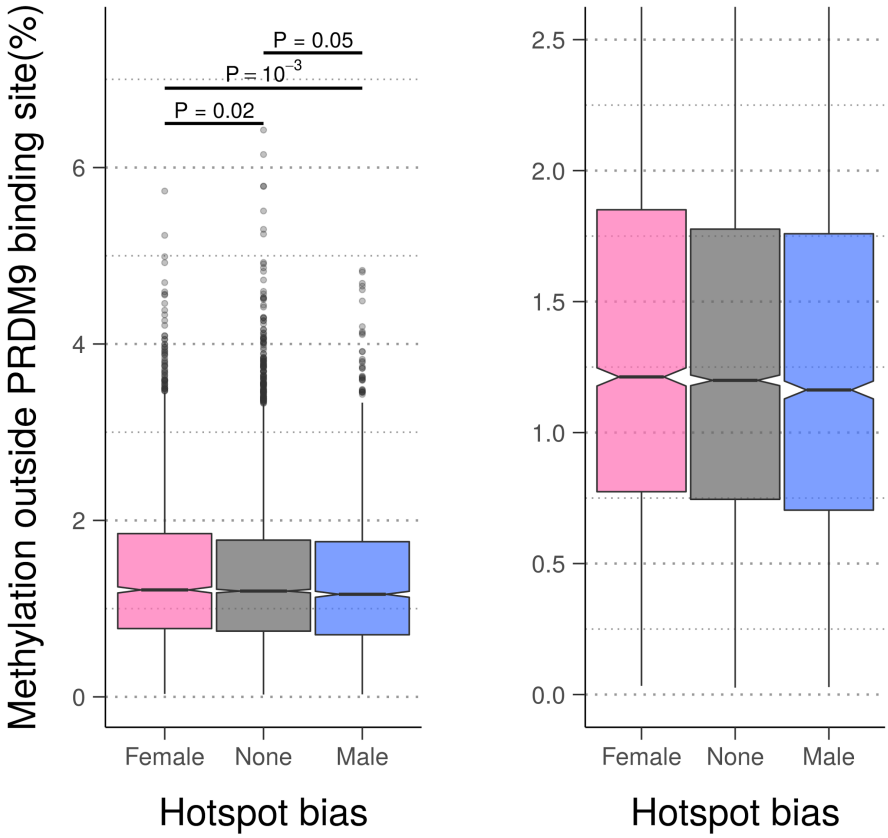


Supplementary Figure 15: DNA methylation at PRDM9 binding sites is present across tissues and absent in the female germ line

Mean DNA methylation across all hotspots. Variations in the magnitude of the signal may be expected for technical reasons. Hotspots are split in deciles by strength fold change and ordered from left (most female biased; pink) to right (most male biased; blue). All plots span -18 bp upstream of the PRDM9 binding site to +18 bp downstream.

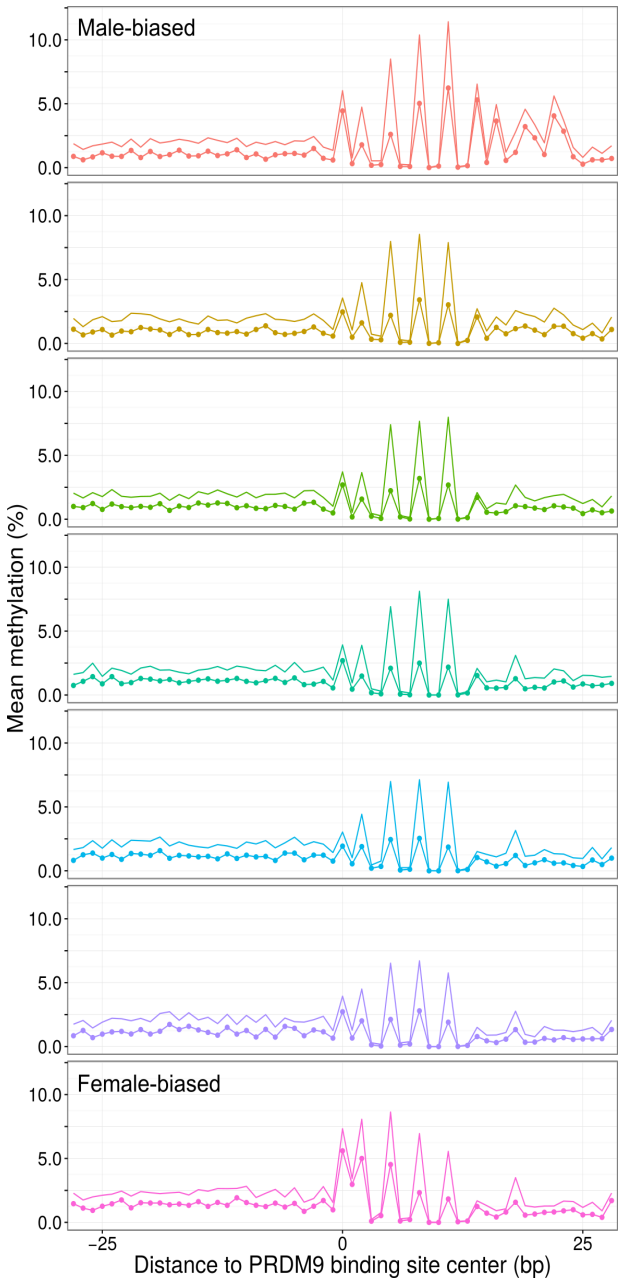


Supplementary Figure 16: Schematic of how DNA methylation affects hotspot usage At female-biased hotspots, CpG dinucleotides primarily occur near the 5' end of the PRDM9 binding site. In males, these CpGs are methylated (black circles), whereas in females they are not (white circles). Methylation at this part of the PRDM9 binding site appears to suppress PRDM9 binding and / or DSB formation. This results in a relatively weak DSB hotspot in males, but a relatively strong hotspot in females, where DNA methylation is absent. At male-biased hotspots, CpG dinucleotides occur primarily near the 3' end of the PRDM9 binding site. In males, these CpGs are methylated and this methylation appears to favor PRDM9 binding and subsequent DSB formation. This results in a relatively strong DSB hotspot in males, but a relatively weak hotspot in females, where DNA methylation at these sites is absent.



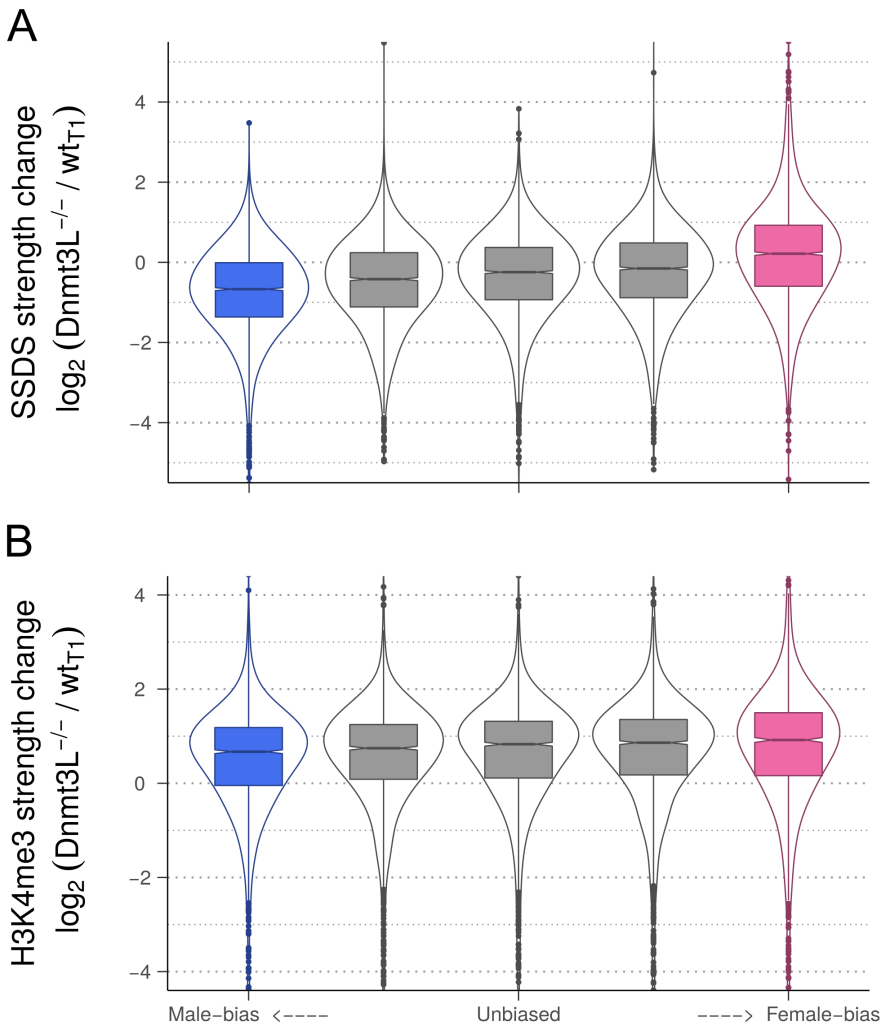
Supplementary Figure 17: Elevated DNA methylation at female-biased hotspots outside of the PRDM9 binding site.

Spermatocyte DNA methylation was calculated in the ± 1 Kb regions flanking each DSB hotspot. To remove the effects of methylation at the PRDM9 binding site, hotspots were centered at the 5' end of the PRDM9 binding site and only DSB hotspots with a single PRDM9 binding site were considered (see methods). Subsequently, the central ± 50 bp was excluded from analysis. Female-biased hotspots had significantly more methylated DNA than male-biased hotspots (Wilcoxon test; $P = 10^{-3}$) although the magnitude of this difference is small with extensive overlap between the female-biased and male-biased hotspot populations. The left and right panels show the same data however the y-axis of the left panel has been limited to show the difference between the median values.



Supplementary Figure 18: DNA methylation at PRDM9 binding sites is reduced in *Dnmt3L*$^{A/A}$ mice

Methylation in *wt* mice is shown as a solid line and in *Dnmt3L*$^{A/A}$ mice as a solid line with dots. Hotspots are split in seven equal sized sets by sex bias. Methylation is universally depleted in *Dnmt3L*$^{A/A}$ mice relative to *wild type*.



Supplementary Figure 19: Hotspot strength variation in *Dnmt3L*$^{-/-}$ mice
The usage of sex-biased hotspots is altered in mice where DNA methylation is reduced (*Dnmt3L*$^{-/-}$). Hotspots are split into quintiles by the fold change between O1 and T1 SSDS samples. The log2 fold change between the tags per million normalized signal at hotspots in *Dnmt3*$^{-/-}$ and *wt* (T1) male mice is shown. Relative to unbiased hotspots, female-biased hotspot strength is increased and male-biased hotspot strength is decreased in *Dnmt3L*$^{-/-}$ mice. (A) SSDS signal. The most male (blue) and female (pink) biased hotspots differ significantly from each other ($P = 10^{-200}$, wilcoxon test) and from all unbiased hotspot quintiles (grey; $P < 10^{-21}$, wilcoxon test). (B) H3K4me3 signal at each hotspot was calculated as the sum of overlapping H3K4me3 peak strengths. This is a proxy for DSB hotspot strength because PRDM9 trimethylates histone H4 lysine 3 before DSB formation. The most male (blue) and female (pink) biased hotspots differ significantly from each other ($P = 10^{-17}$, wilcoxon test) and from all unbiased hotspot quintiles (grey; $P < 0.004$; wilcoxon test). The magnitude of the differences is less pronounced in the H3K4me3 experiment, likely because SSDS is more sensitive for hotspot detection.

MATERIALS AND METHODS

Animal procedures

All animal procedures have been approved by the USUHS Animal Care and Use or were performed according to NIH Guide for the Care and Use of Laboratory Animals.

Sample preparation and sequencing

Embryos were collected at E15.5, ovaries were dissected in cold PBS and stored at -80ºC until use. 90 or 230 ovaries were fixed in 1 ml PBS with 1% paraformaldehyde for 3 min, quenched and homogenized with Dounce homogenizer. Cells were collected by 10 minute centrifugation at 900g using a bucket rotor. The pellet was washed in 1 ml of the following buffers: 1) PBS 2) 0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH8. Cells were lysed in 0.5 ml of the lysis buffer (1% SDS, 10 mM EDTA, 50 mM TrisCl pH8 with complete protein inhibitor cocktail (Roche)) and the chromatin was sheared with Misonix sonicator with the following parameters: efficiency 1, 10 s on, 20 s off, total sonication time 4 min. Chromatin was cleared by 10 min centrifugation at 12,000g at 4C. The supernatant was diluted 2-fold by ChiP buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM TrisHCl , 167 mM NaCl) and dialyzed against the same buffer for 5 hour at 4ºC.

Chromatin was incubated with 6μg of custom made anti-DMC1 antibody and 20μl Dynabeads (10002D, Invitrogen) at 4ºC overnight followed by washing with 500μl of the following buffers: 1) 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM TrisHCl, 150 mM NaCl; 2) 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM TrisCl pH8, 500 mM NaCl; 3) 0.25 M LiCl, 1% Igepal, 1 mM EDTA, 10 mM TrisCl, pH8, 1%

Deoxycholic acid. DNA protein complexes were eluted by two consecutive 15 min incubations at 65^0C using elution buffer (0.1M NaHCO3, 1% SDS, 5mM DTT). The eluates were combined and crosslinking was reversed at 65^0C for 5 hours. The samples were deproteinized and cleaned up with MinElute PCR purification kit (QIAGEN).

The sequencing library was prepared as previously described (22) with small modifications. Briefly: the end repair step was done in 1X T4 DNA ligase buffer with 10mM ATP in the presence of 0.25mM dNTPs, 0.6U T4 DNA polymerase, 0.5U Klenow Enzyme, 2 Units T4 Polynucleotide kinase for 30 min at 20^0C. DNA was purified with MinElute kit. The second step was done in the presence of 1mM dATP and 1U of Klenow Exo minus. Reaction was incubated at 37^0C for 30 min and DNA was purified with MinElute kit. To enrich for single stranded DNA the sample was denatured for 2 min at 95^0C, then cooled to room temperature. The sequencing adaptor mix (Illumina) was diluted 1:200 and added for the ligation step. DNA was purified by MinElute kit and amplified using Phusion Polymerase (0.5 μl per reaction), in the presence of 1 μl of each Illumina PE primers. The following parameters were used: initial denaturation at 98ºC for 30 sec; 21 cycles: 98ºC 10 s, 65ºC 30s, 72ºC 30s; final extension for 5 min at 72ºC. Size selection was done in 2% agarose gel, 180-250bp slice was excised and purified using MinElute Gel Extraction kit.

Testis sample preparation was performed as described previously (22).The following antibodies were used: anti-DMC1: Santa Cruz (C-20, sc 8973), anti-DMC1: (custom), anti-H3K4me3: Millipore (138-473) . All sequencing was performed on an Illumina HiSeq 2500 at the NIDDK genomics core.

Alignment of sequencing reads

For SSDS, reads were aligned to the genome and ssDNA derived reads were identified using the single stranded DNA sequencing (SSDS) processing pipeline (79). Briefly, the first read of each mate pair is mapped to the genome with bwa (v0.7.12) (95). The second read is then mapped to the genome using a modified bwa algorithm that finds the longest mapping suffix for each read. ssDNA is determined from the structure of inverted terminal repeats on the first and second end reads.

Data from Spo11-oligonucleotide mapping were aligned to the reference genome using bwa mem (0.7.12). All other sequencing data were aligned to the reference genome using bwa aln (0.7.12).

Evaluation of SSDS sensitivity for DSB detection in ovary

To estimate the lower hotspot detection limit using SSDS, we generated a DSB map using 2 x 10^5 testis cells from *Hop2*$^{-/-}$ (*Psmc3ip*$^{-/-}$) mice. HOP2 is required for DSB repair (141; 142), and ~37% of cells in testis of *Hop2*$^{-/-}$ mice (167) harbor unrepaired DSBs (N≅74,000; data not shown). The fraction of reads in peaks (FRiP) for this sample was 21% (Sample N1; Fig. S1A); this is slightly lower than the FRiP for DSB maps from whole testis (23-46%; Samples T1-9; Fig. S1A), but far above the 2% FRiP expected by chance. The estimated library size (see methods) of N1 was about 10-times smaller than for the smallest whole testis sample (Fig. S1D), likely because of the limited starting amount of DNA. Small library size complicates hotspot detection, therefore we first attempted to map DSBs in females using approximately 10 times more target cells (a pool of 90 ovaries; approximately 9 x 10^5 target cells). This sample (O2) had a library size

close to that of wt whole testis samples (Fig. S1D), however a low FRiP (7%; Fig. S1A) suggested that hotspot DNA recovery from oocytes was less efficient than from testis. Subsequently, we pooled 230 fetal ovaries to generate a second ovary-derived DSB map (O1). This map was of similar quality to DSB maps derived from testis (FRiP = 33%; library size = 3.8 x 10^7 fragments; Sample O1; Figure 1, Fig. S1A,D).

DSB hotspot identification

Uniquely mapping fragments unambiguously derived from ssDNA (ssDNA type 1) and having both reads with a mapping quality score ≥ 30 were used for identifying hotspot locations (peak calling). NCIS was used to estimate the background fraction for each library. Peak calling was performed using MACS (v.2.1.0.20150420) (194) with the following parameters : --ratio [output from NCIS] -g mm --bw 1000 --keep-dup all --slocal 5000. DSB hotspots within regions previously blacklisted (165) were removed and hotspot strength was calculated as described previously (165).

9,236 hotspots were detected from the O2 ovary-derived sample, however just 6,765 coincided with hotspots from the 230 ovary pool. Most of the 2,471 hotspots unique to this sample appear to be peak calling artifacts (Fig. S2; see methods) and were discarded.

H3K4me3 peak calling

Uniquely mapping reads with a mapping quality score ≥ 30 were used for peak calling. NCIS was used to estimate the background fraction relative to an input DNA

library. Peak calling was performed using MACS (v.2.1.0.20150420) with the following parameters : --ratio [output from NCIS] -g mm --bw 1000 --keep-dup all --slocal 5000. Peak strength was subsequently calculated by subtracting the NCIS normalized input read count from the ChIP-Seq read count.

Hotspot overlaps and merging hotspot sets

Unless otherwise stated, when assessing if hotspots occur at the same location, we restrict the overlap to the ± 200 bp region of DSB hotspots. Previously, we have shown that using a ± 200 bp regions is sufficient for detecting true overlaps and limits the number of spurious overlaps (22). We merged DSB hotspots from the best testis and ovary samples (T1 and O1 respectively). The center of overlapping hotspots was defined as the mean center point of the T1 and O1 hotspot and the flanks were defined as the maximum distance from this center to the original T1 and O1 hotspot edges. Non-overlapping hotspots from each sample were retained as originally defined.

Strength metrics at hotspots

Hotspot strength measured by SSDS was calculated as described previously (22). Briefly, the center-point of the Watson- and Crick-strand ssDNA fragment distributions was used to define the hotspot center. ssDNA fragments on the Watson (top) strand to the left and Crick (bottom) strand to the right of this center were considered signal. Background was extrapolated from the count of ssDNA fragments of opposite polarity around the center (excluding the very center of the hotspot). Hotspot strength is then calculated as the signal - background fragment count. This strength is used as a proxy for

DSB frequency (22; 165). Hotspot strength measured from Spo11-oligo mapping was calculated as the sum of SPO11 oligo-associated reads overlapping each hotspot. H3K4me3 strength at hotspots was calculated as the sum of the strength of overlapping H3K4me3 peaks.

Default hotspots

Hotspots that overlap a hotspot in mice that lack functional *Prdm9* were categorized as putative "default" hotspots. Despite this overlap, these sites may still be targeted by PRDM9. We used FIMO to identify matches to the PRDM9 binding motif for C57Bl/6J in the central ± 250 bp of hotspots (P-value threshold = 10^{-4}). Putative "default" hotspots that contained a PRDM9 binding site were considered low confidence default hotspots. Those that did not contain a motif match were considered high confidence default hotspots.

Sex bias determination

We used MAnorm (163) to infer differential usage of DSB hotspots between the T1 and O1 samples. Hotspots on chromosomes X, Y and M were excluded from this analysis. Hotspots with a Bonferroni corrected MAnorm p-value < 10^{-3} were considered differential.

Cluster analysis

We first identified clusters of hotspots that shared the same sex-bias (female-biased, male-biased, unbiased). Only uninterrupted runs of hotspots with the same bias

were considered. To estimate the expected numbers of clusters, hotspot bias designations were shuffled and the aforementioned process was repeated. 10,000 iterations of this randomization process were performed. P-values are calculated from the empirical distribution of expected values for clusters of each size.

Generation of randomized maps of DSB hotspots

DSB hotspot locations were randomized as described in (22). Briefly, hotspots were uniformly distributed per chromosome, but prohibited from being placed at unmappable regions. Specifically, a mapability score for 40bp sequencing reads at each mm10 base was calculated using the GEM library (20100419-003425 (41)). Hotspot excluded regions were defined as annotated assembly gaps, in addition to 1 kb genomic intervals with <50% uniquely mappable bases plus 1 kb either side. Hotspot width and strength were preserved at the randomized location for each hotspot.

Genetic crossovers

The locations of mouse crossovers in were obtained from the Collaborative Cross (99). Crossovers likely defined by the B6, CAST and PWD/PWK alleles of PRDM9 were determined by first identifying crossovers that occurred in hybrids involving any of these pure strains (MGP,MGM,PGP,PGM). Crossovers that coincided with a single DSB hotspot from only one of the parental PRDM9 alleles were then designated as having originated from that allele. All hotspots between B6, 129 and WSB mice that coincided with a B6, C3H or 13R-defined DSB hotspot were designated as having originated from Mus musculus domesticus (Dom).

DNA methylation

For BS_{SC}, BS_{SG}, BS_{DAA}, BS_{DWT}(Table S1), bismark (87) was used to align whole genome bisulfite sequencing (WGBS-Seq) data to the reference mm10 genome. For BS_{PGC}, BS_{LIV}, BS_{MG} (Table S1), pre-processed nucleotide-resolution methylation data were available, therefore UCSC liftover was used to convert these data from mouse mm9 to mouse mm10 genomic coordinates. hMeDIP-Seq reads were mapped to the genome using bwa mem (0.7.12) (95) and default parameters. Most of the WGBS-Seq datasets did not distinguish between methylation on the two DNA strands, therefore, we only considered methylation at top-strand cytosines.

To examine methylation patterns, we first inferred a high confidence set of PRDM9 binding sites at DSB hotspots. We used FIMO (38) to identify matches to the B6 PRDM9 binding site (22) PWM within hotspots. The number of matches to the position weight matrix (PWM) of the PRDM9 binding site depends on the alignment score threshold used, therefore we identified the PWM p-value alignment threshold that yielded the maximal number of DSB hotspots with a single match to the PRDM9 PWM within the central ± 250 bp. We tested the following threshold values: $P \leq 5x10^{-3}$, $1x10^{-4}$, $2x10^{-4}$, $4x10^{-4}$, $6x10^{-4}$, $8x10^{-4}$, $1x10^{-5}$, $2x10^{-5}$, $4x10^{-5}$, $4x10^{-5}$, $6x10^{-5}$, $8x10^{-5}$, $8x10^{-5}$,$1x10^{-6}$, $1x10^{-7}$, $1x10^{-8}$. 12,097 / 19,600 hotspots (62%) contained a single PRDM9 binding site at the optimal threshold ($P \leq 4x10^{-4}$).

CHAPTER 3: MANUSCRIPT 2 - "A non-repressive role for the KRAB domain and SSXRD of PRDM9"

Authors:

Sarah Thibault-Sennett*[1], Qi Yu*[2], Fatima Smagulova[1,3], Jeff Cloutier[2], Kevin Brick[2], R. Dan Camerini-Otero[2] and Galina V. Petukhova[1].

Affiliations:

* These authors contributed equally to the manuscript.

[1] Department of Biochemistry and Molecular Biology, Uniformed Services University of Health Sciences, Bethesda, Maryland 20814, USA

[2] Genetics and Biochemistry Branch, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH), Bethesda, Maryland 20892, USA

[3] Present address: Inserm UMR1085-Irset, 263 ave du Général Leclerc, 35042, Rennes, France

ABSTRACT

Homologous recombination is required for proper segregation of homologous chromosomes during meiosis. It predominantly occurs at recombination hotspots that are defined by the DNA binding specificity of the PRDM9 protein. PRDM9 belongs to three different families of proteins involved in regulation of transcription, yet, the role of PRDM9 in gene expression control has not been evaluated. Here we analyze the germline transcriptome of the *Prdm9*$^{-/-}$ male mice in comparison to *Prdm9*$^{+/+}$ males and find no apparent differences in the mRNA and miRNA profiles. We further explore the role of PRDM9 in meiosis by analyzing the effect of the KRAB, SSXRD and post-SET zinc finger domain deletion mutants in a cell expression system and the KRAB domain deletion in mice. We found that although the post-SET zinc finger and the KRAB domain are not essential for the methyltransferase activity of PRDM9 *in vitro*, the KRAB domain mutant mice show only residual PRDM9 activity and undergo meiotic arrest. In aggregate, our data indicate that the PRDM9 function has likely diverged from the typical function of a transcription factor and adapted the new role in setting the proper chromatin environment for initiation and completion of homologous recombination.

INTRODUCTION

Homologous recombination is required for faithful segregation of homologous chromosomes during meiosis. It assures that homologous chromosomes find each other and stay connected until the first meiotic division, when they segregate to different daughter cells. Failure to segregate properly results in aneuploidy in gametes and offspring which accounts for 35% of miscarriages and developmental disabilities in 0.3% of live births (61). The most severe recombination defects lead to meiotic arrest and consequent infertility (60).

During meiotic recombination the chromosomes undergo programmed DNA double-stranded breaks (DSBs) (reviewed in (78)). This triggers a genome-wide search for the intact homologous chromosome to be used for repair. RAD51 and DMC1 recombinases bind to single-stranded tails of the resected DSB ends and initiate recombinational repair of DSBs that results in either crossover or non-crossover products (78). Meiotic DSBs preferentially occur in discrete sites of the genome called recombination hotspots (reviewed in (6; 25; 35; 97; 133)). Hotspot locations are defined by the DNA binding specificity of the PR/SET domain containing 9 (PRDM9) protein that trimethylates Lysine 4 of the histone H3 (H3K4me3) at the sites where recombination may subsequently occur (11; 22; 52; 120; 135). The DNA binding domain of PRDM9 consists of an array of zinc finger domains (ZFs) that vary only at three contact amino acids that define the sequence specificity of the DNA binding (reviewed in (139; 186)). Remarkably, the ZF array is extremely polymorphic and may consist of different ZFs in different individuals, leading to a completely different distribution of

recombination hotspots. When PRDM9 is absent, DSBs are formed at so-called default hotspots, the other available H3K4me3 sites within the genome, mostly at gene promoters and enhancer elements (22). *Prdm9*$^{-/-}$ mice undergo meiotic arrest due to recombination failure, and both sexes are sterile. Interestingly, fertility seems to be preserved in *Prdm9* mutants at least in some species (121).

The PR/SET domain of PRDM9 is responsible for the histone methyltransferase activity including mono-, di- and trimethylation of the H3K4 (63; 187), as well as H3K36 trimethylation (44; 144). This domain is distantly related to the classic SET domain found in the histone lysine methyltransferases, the protein family involved in regulation of gene expression. The PR/SET crystal structures with and without substrate suggest that PRDM9 methyltransferase activity is autoregulated through interaction between pre-SET and post-SET regions that block the substrate binding site. The lone ZF located in the post-SET region is involved in this interaction (187).

PRDM9 belongs to the PRDM protein family, the group that is defined by the presence of a ZF array and the PR/SET domain (66). PRDM family members either directly methylate their histone targets or recruit other histone modifiers to the chromatin. Through these activities the PRDM family is involved in a diverse array of developmental processes through regulation of gene expression. PRDM9, however, is unique in that it also has the Kruppel associated box (101) domain. This puts PRDM9 into the second family of transcription factors, the KRAB-ZFPs (KRAB Zinc Finger Proteins). Members of this family also contain the ZF DNA-binding array, but have the

KRAB domain instead of PR/SET (104). While none of these domains is unique in isolation, it appears that no other protein besides PRDM9 contains all three (18).

The KRAB domain is typically involved in gene silencing through interaction with the TRIM28 protein (also known as KAP-1). TRIM28 interacts with the Heterochromatin Protein 1 (HP-1) and serves as a scaffold protein for recruiting chromatin modifying enzymes and remodeling complexes to promoters of the target genes bound by the KRAB-ZFPs. Interestingly, the KRAB domain of PRDM9 misses the amino acids required for interaction with TRIM28 (100) and recent studies found no direct interaction between the two proteins (67; 137). Indeed, the KRAB domain of PRDM9 has higher similarity to the atypical KRAB domain found in the Synovial Sarcoma X chromosome breakpoint (SSX) protein family (98). PRDM9 and the SSX family members have another similarity: they both contain the SSX repressive domain (SSXRD) following the atypical KRAB domain. Although the SSXRD was shown to be a potent transcription repressor molecular mechanism behind the repressive activity of SSXRD has not been described.

Considering that PRDM9 has three domains implicated in regulation of transcription it was originally thought of as a transcription factor (28; 63; 110; 118; 124). This role was particularly attractive, because *Prdm9* is the only known speciation gene in mammals, and the conflict between regulatory mechanisms is perceived as the most likely cause of hybrid incompatibilities (106; 124). Furthermore, recent studies found that during meiosis PRDM9 is directly bound to a set of active promoters in a manner that is

not dependent on PRDM9 binding specificity (53). In this study we evaluate the potential role of PRDM9 in regulation of transcription and the contribution of the PRDM9 domains implicated in transcription control to PRDM9 function.

We first used RNA sequencing (RNA-seq) to assess whether deletion of *Prdm9* leads to changes in gene expression profiles genome-wide. We found very little, if any, difference in the transcript levels between $Prdm9^{-/-}$ and $Prdm9^{+/+}$ testes. We then showed that PRDM9 is unlikely to be involved in regulation of microRNA expression, as no difference was found between $Prdm9^{-/-}$ and $Prdm9^{+/+}$ small RNA-seq profiles. We next assessed the effect of KRAB domain, SSXRD and post-SET ZF domain deletions on methyltransferase activity of PRDM9 expressed in cell culture. Although SSXRD deletion resulted in completely inactive protein, the KRAB and post-SET ZF mutants retained significant activity. To assess if this level of activity is sufficient to maintain meiotic program we generated the KRAB domain deletion mouse strain. We found only trace amount of H3K4me3 at the expected hotspots and very few DSB hotspots at the wild type locations. In aggregate, we found that PRDM9 is not involved in regulation of gene expression, but the KRAB domain of PRDM9 is essential for the role of PRDM9 in formation of recombination hotspots.

RESULTS

Inactivation of *Prdm9* does not lead to changes in gene expression.

Since both the KRAB domain and the SSXRD are required for regulation of gene expression we assessed whether PRDM9 is indeed a transcription factor. We carried out

high throughput sequencing of mRNA (RNA-seq) from testes of *Prdm9*$^{-/-}$ mice. As *Prdm9*$^{-/-}$ spermatocytes undergo meiotic arrest (62) the cell type composition in *Prdm9*$^{-/-}$ testis is vastly different from that in wild type. Consequently, comparing the transcriptome between wild type and *Prdm9* knockout mice would not be meaningful. To get around this problem we used the *Hop2*$^{-/-}$ mice as the *Prdm9*$^{+/+}$ control. In *Hop2*$^{-/-}$ mice meiotic DSBs are formed, but not repaired, which leads to meiotic arrest at the stage similar to *Prdm9*$^{-/-}$ with a very similar cell type composition (62; 142).

Two biological replicates of RNA-seq data for each *Prdm9*$^{-/-}$ and *Hop2*$^{-/-}$ samples were used to evaluate gene expression levels (Supplementary Table S1). To count the number of expressed transcripts within each sample, we defined expressed transcripts as genes with the value of fragments per kilobase per million (FPKM) larger than 0.1. With this threshold, 15,368 genes were expressed in both *Prdm9*$^{-/-}$ replicates while 15,349 genes were expressed in both *Hop2*$^{-/-}$ replicates, representing 65.3% and 65.2% of annotated Refseq mouse genes, respectively (Supplementary Table S1). The correlation between *Prdm9*$^{-/-}$ and *Hop2*$^{-/-}$ expression is very strong across 18,969 annotated genes that are expressed in at least one of four samples (Figure 1A, Spearman's R=0.98-0.99). This indicates that the absence of *Prdm9* does not perturb the global gene expression pattern.

As expected, when RNA-seq was performed using testes from adult wild type mice, 17,948 genes were expressed over our threshold level, representing 76.2% of annotated mouse genes (Supplementary Table S1). This increase in approximately 2,000 genes reflects the presence of the late meiotic and post-meiotic cells in the wild type

testes that are absent in the *Hop2* and *Prdm9* mutants. Consequently, the global gene expression pattern between wild type and the mutants shows low similarity (Figure 1A, Spearman's R=0.56-0.58). At the same time, the gene expression profile in *Prdm9*$^{-/-}$ testes is highly correlated to that in wild type pups aged 12-14 days *post partum* (Supplementary Figure S1, Spearman's R=0.88). At this age male mice undergo the first wave of meiosis, reaching the pachytene stage, when meiosis is arrested in *Prdm9* mutants. This confirms that the difference in gene expression profile between *Prdm9* knockout and adult wild type mice reflects the composition of the cell population within testes, rather than misregulation of gene expression.

PRDM9 was implicated in the regulation of the *Morc2b* gene because of altered expression level in *Prdm9*$^{-/-}$ testes (63), as well as variable expression in certain mouse strains and hybrids carrying different *Prdm9* alleles (118). We found that there is a 300-fold change in the *Morc2b* expression over the course of spermatogenesis with a sharp increase starting at the onset of pachytene stage (Figure 1B). Therefore, even very slight changes in the testes cell population due to meiotic arrest (*Prdm9*$^{-/-}$ and sterile hybrids) or delay (subfertile hybrids) would lead to substantial changes in the abundance of the *Morc2b* transcript, explaining prior observations. Indeed, we observe no significant difference in the *Morc2b* expression level between *Prdm9*$^{-/-}$ (FPKM=1.23, 0.82, mean=1.1) and *Prdm9*$^{+/+}$*Hop2*$^{-/-}$ mutants (FPKM=0.21,0.34, mean=0.28) (Supplementary Table S2), indicating that *Morc2b* transcription is unlikely to be regulated by PRDM9. Overall, we did not detect any differentially expressed genes between the *Prdm9*$^{-/-}$ and *Hop2*$^{-/-}$ samples (q-value < 0.01). Less stringent criteria yielded 29 differentially

expressed genes that were not further evaluated (q-value < 0.05, Supplementary Table S2).

We next explored whether PRDM9 could play a role in regulating miRNAs, as they have been shown to play crucial roles in meiosis (152). To compare miRNA levels we considered miRNAs that were detected at ≥10 transcripts per million reads (TPM) in miRNA sequencing libraries prepared from two *Hop2*$^{-/-}$ and two *Prdm9*$^{-/-}$ samples (Supplementary Table S3). This resulted in 207 mature miRNAs expressed in all four libraries, which represent 20% of the total mouse miRNAs in miRbase. We found that none of these 207 miRNAs showed expression difference between the *Hop2*$^{-/-}$ and *Prdm9*$^{-/-}$ samples (q value > 0.01) and there is a strong overall correlation of miRNA profiles between *Prdm9*$^{-/-}$ and *Hop2*$^{-/-}$ samples ($R = 0.97\sim0.98$, Fig. 1C). In addition to annotated miRNAs we identified mature miRNAs that were not described in the miRBase (release 21), 11 of which were present in all four samples (Supplementary table S4). With one exception, these novel miRNAs had very low expression level (1000 times lower than annotated miRNAs) and none of them showed differences between *Prdm9*$^{-/-}$ and *Hop2*$^{-/-}$ samples. In aggregate, our data indicate that inactivation of the *Prdm9* gene does not lead to significant changes in transcription, either of protein coding genes or of miRNAs.

The KRAB domain and the post-SET zinc finger are not essential for H3K4me3 activity of PRDM9 *in vitro*.

Since we found no evidence of PRDM9 being involved in regulation of transcription, we decided to assess whether the KRAB domain and the SSXRD were required for other PRDM9 activities. We first expressed the wild type GFP-tagged

PRDM9 (*Dom2* allele, (135)) in the GC-1 mouse spermatogonial cell line (65) and evaluated the methyltransferase activity of PRDM9 by H3K4me3 Chromatin Immunoprecipitation followed by High-Throughput Sequencing (ChIP-seq). We found a marked enrichment of H3K4me3 reads at the known *Dom2* hotspots (22), but not at hotspots defined by a different *Prdm9* allele, *13R* (22) (Figure 2). This enrichment was not due to expression of the endogenous *Prdm9*, because no H3K4me3 signal was detected in GC-1 cells before transfection (GC-1 cells carry *Dom2 Prdm9* allele). Next, we generated a point mutant with a single Y341F amino acid change in the active site of the PR/SET domain (SET, Figure 2). This mutation was previously shown to abolish methyltransferase activity of PRDM9 (187) and serves in this study as a negative control.

Following confirmation that the expressed GFP-PRDM9 was active, we used targeted mutagenesis to generate two alternative GFP-PRDM9 constructs that lacked the KRAB domain (KRAB-1 and KRAB-2, Figure 2) and assessed their activity by H3K4me3 ChIP-seq. H3K4me3 is a relatively common epigenetic modification that is found mostly at gene promoters and enhancer elements. As H3K4 trimethylation at these sites is independent of PRDM9 it can be used as the internal control for the H3K4me3 ChIP efficiency. Therefore, we normalized all H3K4me3 ChIP-seq samples by the number of reads at promoters (see methods). We found that both KRAB mutants retained the methyltransferase activity, although efficiency was lower and the number of hotspots with H3K4me3 mark was reduced compared to wild type PRDM9 (Figure 3). Since our second KRAB mutant lacks a significant portion of the SSXRD we assessed whether SSXRD is required for PRDM9 function. Unexpectedly, removing the SSXRD

(SSXRD, Figure 2) resulted in complete inactivation of PRDM9 with no trace of H3K4me3 enrichment at the *Dom2* hotspots (Figure 3). We confirmed that the protein product was expressed at comparable to wild type level and properly localized to the nucleus (Supplementary Figure S2), therefore, lack of activity could not be explained by inability of PRDM9 to reach its target.

We next tested whether the post-SET zinc finger had an effect on the H3K4 methyltransferase activity of PRDM9. This lone zinc finger has been proposed to regulate the PR/SET catalytic activity by blocking access to the active site in the absence of histones (187). We found that deletion of this zinc finger (ZincF, Figure 2) did not abolish PRDM9 activity and H3K4 trimethylation marks were still detected at expected hotspots (Figure 3). Nevertheless, just as in the case of KRAB deletion there was a noticeable decrease in the trimethylation efficiency.

The KRAB domain is necessary for PRDM9 activity *in vivo*.

After we found that deletion of the KRAB domain does not abolish methyltransferase activity of PRDM9 we wanted to investigate the effect of the KRAB domain deletion in meiosis. Since the KRAB domain was proposed to tether the hotspot DNA to chromosome cores to enable DSB formation (67; 136) we expected that DSBs would not be targeted exclusively to hotspot regions despite of the presence of the H3K4me3 marks. To test this we used CRISPR-based gene targeting to generate a knock-in mouse strain that lacks the KRAB domain of PRDM9, $Prdm9^{K/K}$ (Supplementary Figure S3). The mutant *Prdm9* allele matched the KRAB1 deletion that was tested in the

GC-1 expression system and lacks the majority of the KRAB domain while retaining three N- and three C-terminal residues of the domain to preserve protein folding.

We found that both male and female *Prdm9*$^{K/K}$ mice were infertile. *Prdm9*$^{K/K}$ ovaries were greatly reduced in size and had no follicles (Supplementary Figure S4). Testes size was also reduced and seminiferous tubules lacked spermatids, although primary spermatocytes were easily detectable. To evaluate more precisely the time of arrest we analyzed meiotic progression by immunostaining chromosome spreads of the *Prdm9*$^{K/K}$ spermatocytes. We found that spermatocytes arrested at the stage resembling pachynema, but DSBs were only partially repaired and synapsis of the homologous chromosomes was incomplete (Figure 4). Overall, the phenotype of the *Prdm9*$^{K/K}$ mice was indistinguishable from that of the *Prdm9*$^{-/-}$ mice, indicating that the KRAB domain of PRDM9 is indispensable for meiosis. We confirmed that the expression level of the PRDM9 KRAB deletion mutant was not different from wild type PRDM9 (Supplementary Figure S2), making this unlikely that the null phenotype we observe was due to lack of the protein product.

Since the KRAB mutants retained methyltransferase activity in cell expression system (Figure 3) we evaluated whether H3K4me3 was deposited at hotspots in the *Prdm9*$^{K/K}$ mice. We detected only trace levels of H3K4 trimethylation activity at *Dom2* hotspots (Figure 5) suggesting that the activity observed in the cell culture expression system was due to much higher expression level of *Prdm9* (Supplementary Figure S2). We next mapped the DSB hotspots in the *Prdm9*$^{K/K}$ mice by ChIP-seq with DMC1

antibodies followed by single stranded DNA sequencing (SSDS, (79)). Not surprisingly, this analysis revealed that the vast majority of DSBs in the *Prdm9*$^{K/K}$ spermatocytes form at promoters and enhancers, the default hotspots where DSBs form in the absence of PRDM9 (22) (Figure 5). Only a few hotspots in *Prdm9*$^{K/K}$ mice were formed at hotspots active in wild type mice and, with a few exceptions, they were much weaker. This indicates that the PRDM9 mutant without the KRAB domain is largely incapable of methylating H3K4 at its binding sites or of targeting the DSBs formation.

DISCUSSION

PRDM9 compiles three different domains that have been implicated in regulation of transcription in other protein families: the KRAB domain, SSXRD and the PR/SET domain. Based on this, PRDM9 was assumed to be a transcription factor and early reports supported this assumption (28; 63; 110; 118; 124).

The KRAB domain of PRDM9 has a noncanonical structure and shows ~25 fold weaker repression activity than the classic KRAB domains of the KRAB-ZFP family in *in vitro* assays (98). Therefore, involvement of the KRAB domain of PRDM9 in gene expression was somewhat unlikely. Indeed, two recent papers reported that the role of the KRAB domain is to anchor the PRDM9-bound regions to the chromosomal core of meiotic chromosomes where the factors required for DSB formation are located. PRDM9 exerts this function through interaction with the CXXC1 protein, the mammalian homolog of yeast Spp1 (67; 136). Spp1 has a PHD domain that binds H3K4me3, but it also interacts with Mer2, the axis-associated protein required for DSB formation.

Therefore, Spp1 provides a link allowing H3K4 trimethylated chromatin to be physically anchored to the chromosome axis where DSB formation machinery resides. Yeast lack the PRDM9 protein, and recombination hotspots occur at H3K4me3 sites, usually promoters, regardless of underlying DNA sequence. A similar situation occurs in *Prdm9* knockout mice, where DSBs are formed in promoter regions. In the presence of PRDM9, CXXC1 would interact with both PRDM9 (67; 136) and H3K4me3 (43) and then be tethered to the chromosome axis through interaction with IHO1 (67), the mammalian homolog of Mer2.

Although the role of PRDM9 in bringing hotspot regions to the axis-associated DSB promoting factors is plausible, it does not explain why removal of the KRAB domain leads to elimination of the methyltransferase activity of PRDM9. The purified PR/SET domain is capable to carry out methyltransferase activity *in vitro* using synthetic peptides as targets (187). The fact that removal of the KRAB domain reduces H3K4me3 activity in our cell expression system may suggest that the KRAB domain is necessary to support the PR/SET function within chromatin, for example, by assisting in accessing DNA or reaching the histone target. Chromatin remodeling activity was previously proposed for PRDM9, when H3K4me3 profiling at recombination hotspots revealed nucleosome depleted regions at PRDM9 binding sites (8). To determine whether such remodeling was performed by PRDM9 itself rather than the downstream DSB formation and processing we carried out H3K4me3 profiling in *Spo11*$^{-/-}$ mice, which are deficient in DSB formation (148). We found a clear nucleosome depleted region at hotspot locations, confirming that chromatin remodeling at these sites occurs before the formation of DSBs

(Supplemental Figure 5). Yeast Spp1 and CXXC1 are members of the multisubunit complex COMPASS that is involved in regulation of transcription. In both cases, the SET domain containing subunit (Set1 in yeast, SET1 or MLL1 through MLL4 in human COMPASS complexes) catalyzes the histone methylation. It is conceivable that PRDM9 acts as the methyltransferase subunit in a meiosis-specific COMPASS-like complex that both remodels chromatin to facilitate PRDM9 binding and delivers PRDM9-bound chromatin to chromosome axes. This would explain the reduction of the H3K4me3 activity in both KRAB mutants in the cell expression system.

The deletion in the KRAB2 mutant included the entire KRAB domain and 24 out of 32 amino acids of the SSXRD. Since the resulting protein showed the same level of the methyltransferase activity in cell culture as the KRAB1 mutant, it appeared that the SSXRD did not have additional role in PRDM9 function in non-meiotic environment. It was therefore surprising that deletion of the entire SSXRD resulted in a completely inactive protein. Although it is conceivable that the last 8 amino acids of the SSXRD are critical for its function, it is more likely that this deletion resulted in misfolding of the protein and inhibition of the PR/SET domain by means not related to biological function of the SSXRD.

The post-SET ZnF was implicated in the autoregulation of the methyltransferase activity of PRDM9 (188). However, studies of the DNA binding affinity of the ZF array of PRDM9 *in vitro* revealed that extending the DNA binding construct to include the post-SET ZnF lead to drastic reduction of the DNA binding (182). Protein misfolding

was proposed to explain this effect, consistent with our data showing substantial H3K4me3 activity in the post-SET deletion mutant. The proposal that the post-SET ZF is involved in regulation of PRDM9 (188) to prevent automethylation (82) is consistent with decrease of the H3K4me3 activity of the post-SET ZF mutant in our cell culture assays. Removal of the post-SET ZF would allow PRDM9 to be automethylated, which is potentially deleterious to either methyltransferase activity of the PR/SET domain or/and PRDM9 interactions with other proteins or DNA.

A recent study described the development of the *Prdm9* mutant that lacked the N-terminal part of the protein including the KRAB domain (68). Although it was not clear whether the mutant phenotype was due to removal of the KRAB domain or because of substantially reduced expression level of the product, it matched the phenotype observed in our $Prdm9^{K/K}$ strain. We could not assess the level of PRDM9 protein in our mutant due to lack of antibodies of sufficient quality. However, the mRNA levels of $Prdm9^{K/K}$ and wild type *Prdm9* were similar (Supplemental Figure 3) and in cell culture, wild type *Prdm9* and the KRAB mutant produced comparable amount of protein (Supplemental Figure 2). Furthermore, we were able to detect residual H3K4me3 activity at the sites of expected hotspots (Figure 5). Surprisingly, we also detected the evidence of DSB formation at such hotspots. If the KRAB domain of PRDM9 is required to communicate with the DSB formation machinery than deletion of this domain should eliminate the preference of PRDM9-dependent H3K4me3 sites over H3K4me3 at promoters by the DSB machinery. Although the vast majority of DSBs in the $Prdm9^{K/K}$ mice form at default hotspots, the fraction of DSBs that form at wild type hotspot sites is unexpectedly

high. Although the vast majority of DSBs in the *Prdm9*$^{K/K}$ mice form at default hotspots, there is evidence that DSBs also form at PRDM9-dependent locations. Out of 359 H3K4me3 peaks detected in *Prdm9*$^{K/K}$ mice at wild type locations 35% resulted in the formation of DSB hotspots (data not shown). Although it is possible that a large number of PRDM9-dependent H3K4me3 peaks remained undetected due to their low intensity in the *Prdm9*$^{K/K}$ mice, it is also possible that interaction of PRDM9 with DSB machinery involves additional players that bind additional regions of PRDM9.

ACKNOWLEDGEMENTS

We thank members of the Camerini-Otero and Petukhova labs for helpful discussions and suggestions. We also thank both the NIDDK genomics core and CHIRP at the American Genome Center at USU, specifically G.S. and C.L.D., for assistance with sequencing. This study used the high performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov). This research was supported by NIH grant R01GM084104 from National Institute of General Medical Sciences (G.V.P.), March of Dimes Foundation grant 1-FY13-506 (G.V.P.) and by the NIDDK Intramural Research Program (R.D.C.O.). The sequencing data reported in this paper will be archived at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo).

MAIN FIGURES

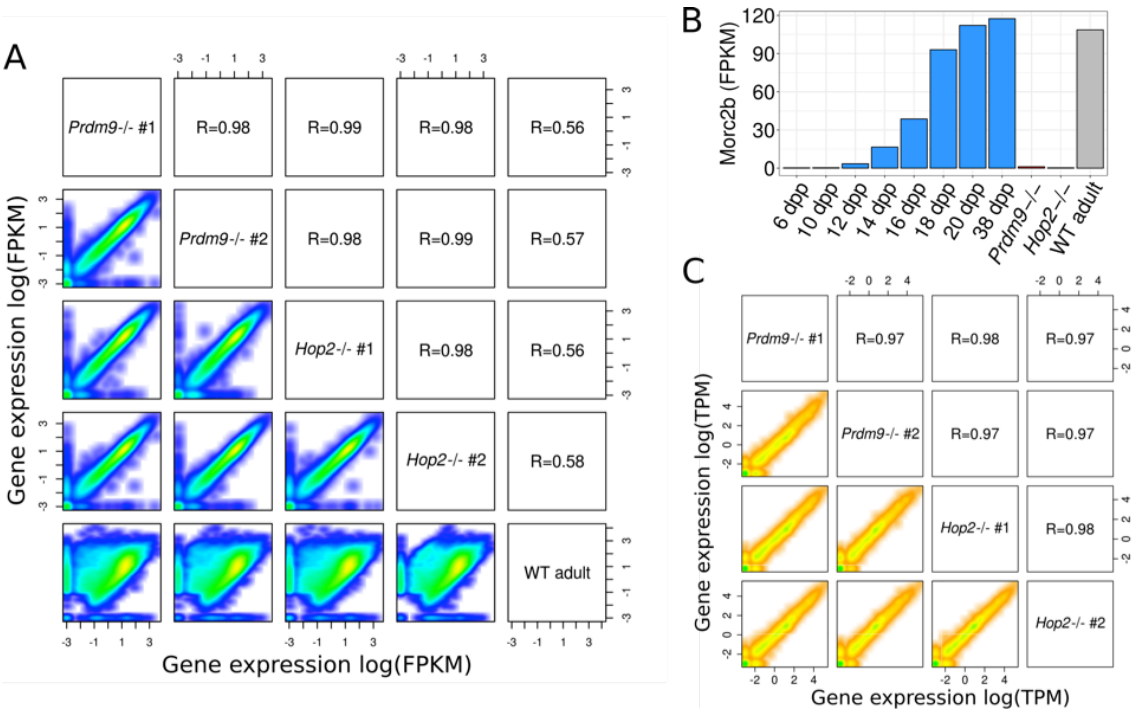


Figure 1. Inactivation of the *Prdm9* gene does not lead to changes in gene expression profiles.

(A) Scatter plot of gene expression level measured by log(FPKM). The numbers represent Spearman's correlation coefficient. B) Expression of the *Morc2b* gene in *Prdm9*$^{-/-}$(average from two library), *Hop2*$^{-/-}$ (average from two library) and juvenile wild type mice. (C) Scatter plot of miRNA expression level measured by log (TPM). The numbers represent Spearman's correlation coefficient. 2228 annotated miRNAs are shown.

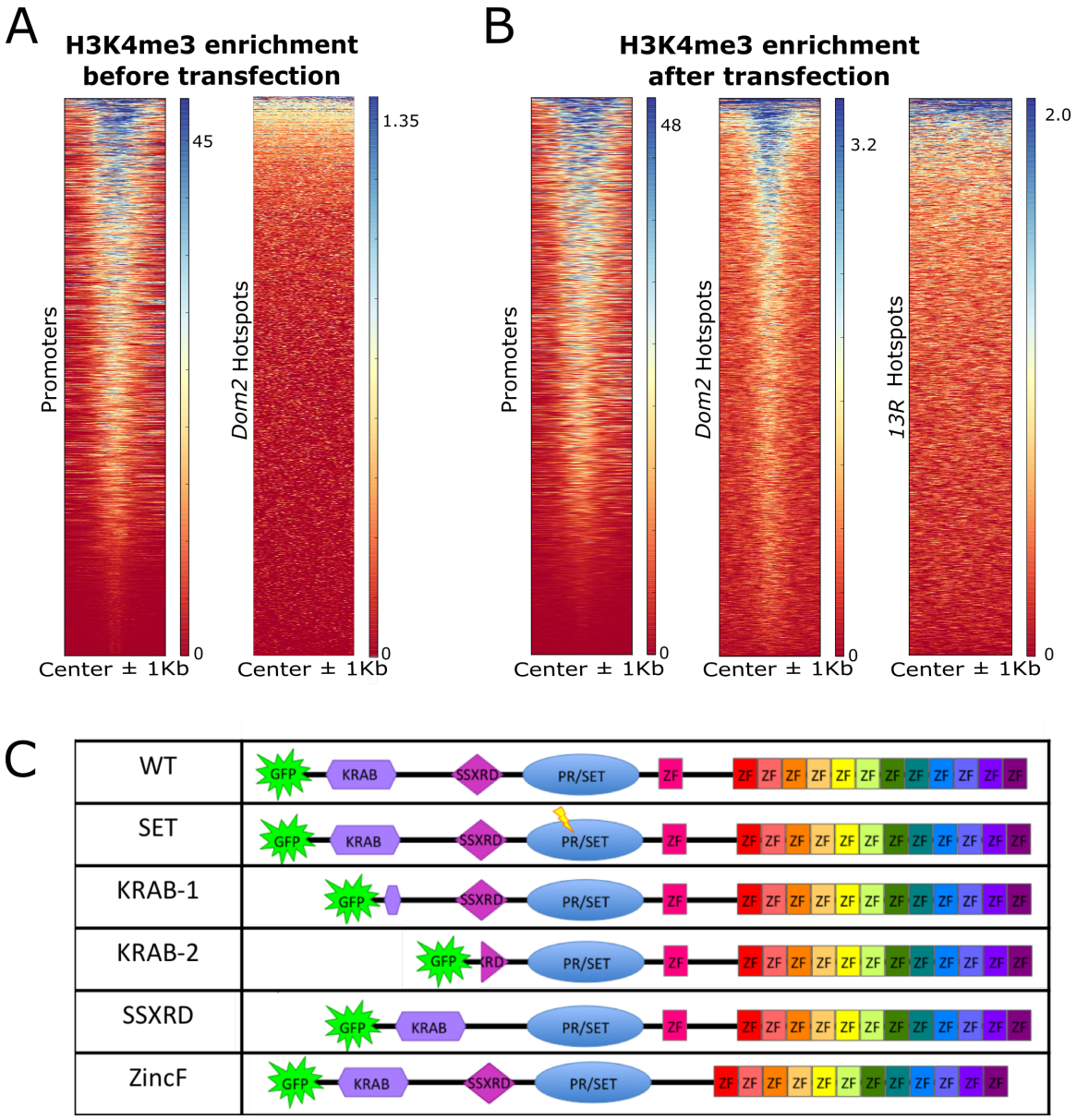


Figure 2: Expression of GFP-PRDM9 in GC-1 cells leads to H3K4 trimethylation at *Dom2* recombination hotspots (HS).

H3K4me3 ChIP-seq was performed using cells (A) prior to and (B) post transfection of the GFP-PRDM9 construct. Heatmaps represent strength of H3K4me3 signal (as reads per genomic content, RPGC) across 12,840 promoters, 19,528 *Dom2* hotspots or 9,834 13R hotspots. Blue represents strong H3K4me3 signal while red represents lack of H3K4me3. Enrichment of the H3K4me3 at promoters serves as an internal control for ChIP efficiency. (C) Constructs used for expression in GC-1 cell system.

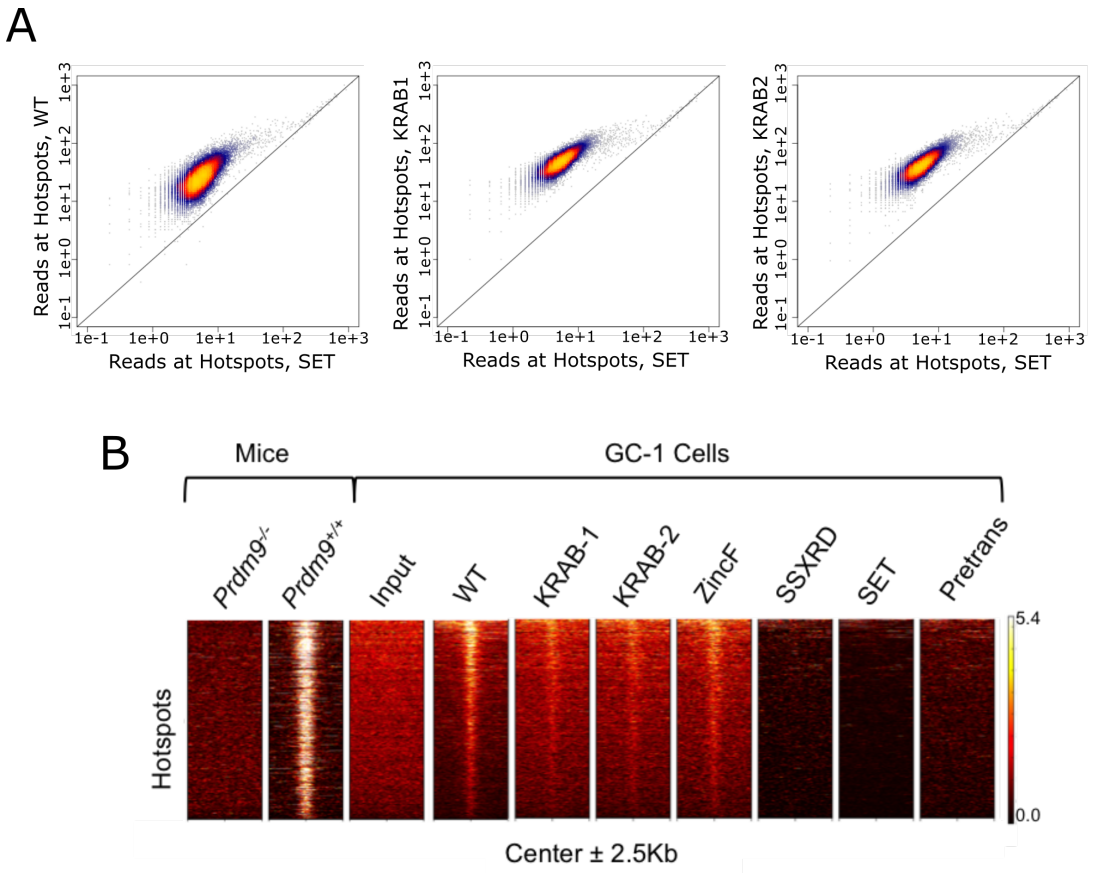


Figure 3: GFP-PRDM9 lacking the KRAB domain retains activity in the GC-1 expression system.

Various GFP-PRDM9 constructs from Figure 1C were expressed in GC-1 cells followed by H3K4me3 ChIP-seq to assess PRDM9 activity at known binding sites. (A) Normalized sequencing read count at 11,642 *Dom2* hotspots for WT, KRAB1 and KRAB2 versus the catalytically-dead SET mutant. As ChIP pull-down efficiency varied between samples, read count per sample was normalized by read count at 150 promoters (see methods). Colors represent density of overlapping points: yellow, the highest and blue, the lowest. (B) Heatmaps show enrichment of sequencing tags across 11,642 *Dom2* hotspots. Experiments were conducted in triplicates for WT and KRAB1 and duplicates for SET, SSXRD, ZincF and KRAB2.

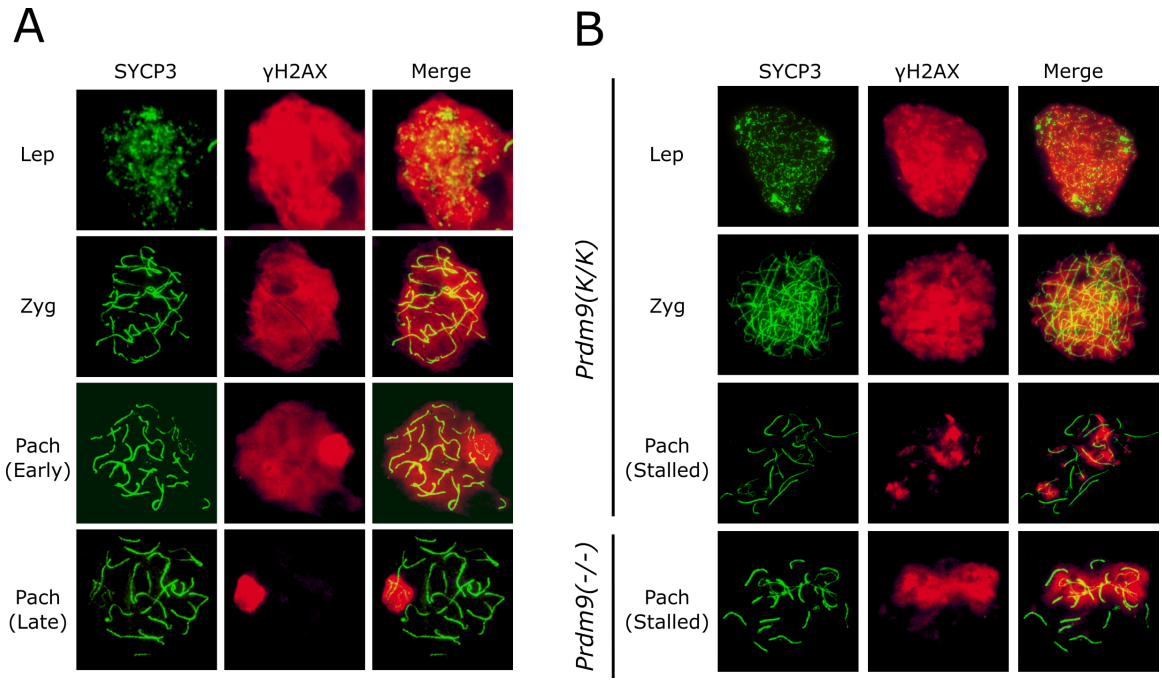


Figure 4: The KRAB domain of PRDM9 is essential for meiotic progression. Chromosome spreads from (A) *Prdm9*$^{+/+}$ or (B) *Prdm9*$^{K/K}$ and *Prdm9*$^{-/-}$ spermatocytes immunostained with SYCP3 and γH2AX antibodies as indicated. *Prdm9*$^{K/K}$ and *Prdm9*$^{-/-}$ arrest at pachytene-like stage with incomplete synapsing and DSB repair. L, leptonema; Z, zygonema; EP, early pachynema; P, pachynema; P(A), spermatocytes arrested in pachytene-like stage.

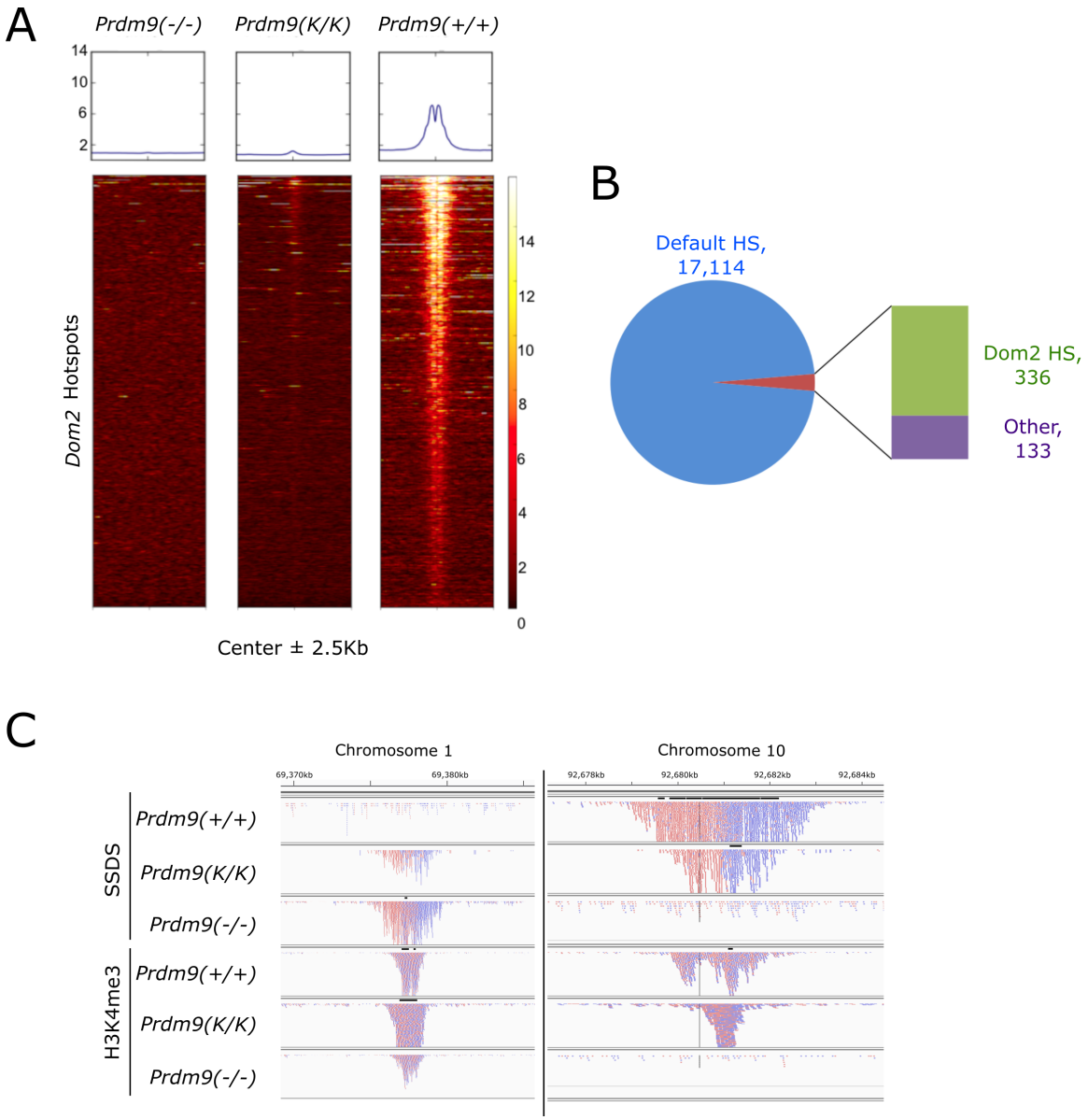


Figure 5: KRAB-less PRDM9 retains residual activity in vivo.
(A) H3K4me3 ChIP-seq heat maps for *Prdm9*$^{-/-}$, *Prdm9*$^{K/K}$ and *Prdm9*$^{+/+}$ spermatocytes across known 11,642 *Dom2* hotspots. (B) DSB hotspots were mapped with SSDS in *Prdm9*$^{K/K}$ spermatocytes and compared to known hotspot locations. While the vast majority of DSBs occur at non-PRDM9-directed default hotspots (hotspots used in *Prdm9*$^{-/-}$ mice), there is a small number of seemingly PRDM9-directed DSBs which are not present in *Prdm9*$^{-/-}$ mice. (C) Integrated Genome Viewer images of representative default hotspot (left) and PRDM9-directed hotspot (121) aligned to the genome in *Prdm9*$^{+/+}$, *Prdm9*$^{K/K}$, and *Prdm9*$^{-/-}$ mice. "SSDS" tracks show sequencing reads from DMC-1 SSDS in *Prdm9*$^{+/+}$, *Prdm9*$^{K/K}$, and *Prdm9*$^{-/-}$ mice. "H3K4me3" tracks show sequencing reads from H3K4me3 ChIP-seq in *Prdm9*$^{+/+}$, *Prdm9*$^{K/K}$, and *Prdm9*$^{-/-}$ mice.

SUPPLEMENTARY TABLES

Supplementary Table 1: RNA-seq samples

Samples	Mapped reads	Annotated Genes (FPKM > 0.1)	Fraction of total genes
***Prdm9*-/- #1**	323,291,400	15626	66.44%
***Prdm9*-/- #2**	360,742,655	15790	67.14%
***Hop2*-/- #1**	321,854,537	15600	66.33%
***Hop2*-/- #2**	311,040,911	15857	67.42%
WT adult	331,114,512	17948	76.31%

Supplementary Table 2: Genes differentially expressed in *Prdm9*$^{-/-}$ and *Hop2*$^{-/-}$ testes at Q-value < 0.05

Transcript name	Location	RPKM *Hop2-/-*	RPKM *Prdm9-/-*	Log2 (Fold Change)	P-value	Q-value
Aass	chr6:23072172-23132986	6.66	13.48	1.02	0.00005	0.03
Aim1l	chr4:134068451-134092504	0.98	2.88	1.56	0.00005	0.03
Aldh3a1	chr11:61208741-61218416	40.68	19.09	-1.09	0.00005	0.03
Atp10d	chr5:72203328-72298771	1.11	4.84	2.13	0.00005	0.03
C4b	chr17:34728380-34743897	12.49	28.37	1.18	0.00005	0.03
Ccdc80	chr16:45094052-45127924	1.83	14.01	2.94	0.00005	0.03
Cntnap1	chr11:101176116-101190720	0.63	2.02	1.67	0.00005	0.03
Col11a1	chr3:114030539-114220326	0.66	1.95	1.57	0.00005	0.03
Cyp21a1	chr17:34801347-34804426	9.84	38.22	1.96	0.00005	0.03
Dnahc8	chr17:30626935-30875264	0.55	2.83	2.38	0.00005	0.03
Dynlt1b	chr17:6430111-6436295	13.70	39.31	1.52	0.00005	0.03
E330017L17Rik	chr4:129906225-129908996	5.29	1.17	-2.18	0.00005	0.03
Filip1	chr9:79815561-79977882	0.46	1.99	2.10	0.00005	0.03
Gm5506	chr4:150237196-150248873	9.30	20.82	1.16	0.00005	0.03
Hddc3	chr7:80343136-80346097	4.04	12.36	1.61	0.00005	0.03
Hsd3b3	chr3:98741475-98763128	1.79	0.05	-5.09	0.00005	0.03
Morc2b	chr17:33135589-33139683	0.31	1.21	1.96	0.00005	0.03
Mt3	chr8:94152606-94154148	105.50	35.85	-1.56	0.00005	0.03
Myo7b	chr18:31959233-32036931	0.03	0.24	3.21	0.00005	0.03
Myom3	chr4:135759714-135815564	0.65	2.74	2.06	0.00005	0.03
Prdm9	chr19:42725857-42752775	4.35	0.27	-4.00	0.00005	0.03
Pyroxd2	chr17:8128590-8147832	1.50	4.63	1.63	0.00005	0.03
Rnaset2b	chr4:133600555-133602168	15.97	3.20	-2.32	0.00005	0.03
Sfn	chr17:7925999-7934897	23.14	6.97	-1.73	0.00005	0.03
Tagap	chr7:97412956-97417510	0.37	2.06	2.47	0.00005	0.03
Thrsp	chr3:88328652-88334433	7.57	19.54	1.37	0.00005	0.03
Tmem79	chr8:43183121-43193884	3.57	7.92	1.15	0.00005	0.03
Triml2	chr10:23894687-23905343	0.14	1.01	2.87	0.00005	0.03
Vnn1	chr6:23072172-23132986	3.49	9.85	1.49	0.00005	0.03

Supplementary Table 3: miRNA-seq samples

Sample	Read count	After Filter by length	Mapped to the genome	Annotated miRNAs	Annotated miRNA, >10 TPM	Novel miRNA, > 10 TPM
***Prdm9-/-*, #1**	177,780,539	162,905,112 (91.63%)	144,448,922 (88.7%)	510	232	28
***Prdm9-/-*, #2**	140,133,738	134,060,783 (95.67%)	110,819,813 (82.7%)	490	232	34
***Hop2-/-*, #1**	171,444,695	162,971,979 (95.06%)	146,742,487 (90%)	477	231	34
***Hop2-/-*, #2**	131,711,523	119,420,490 (90.67%)	93,370,945 (78.2%)	457	209	40

Supplementary Table 4: Novel miRNAs found in all four miRNA-seq samples

Novel miRNA	TPM_ *Prdm9*-/- #1	TPM_ *Prdm9*-/- #2	TPM_ *Hop2*-/- #1	TPM_ *Hop2*-/- _#2	log2 (Fold Change)	P-value
chr4:132310004..132310067:+	433.50	529.80	416.03	382.35	-0.03	0.93
chr3:86138789..86138857:-	34.63	17.48	26.79	23.20	0.18	0.86
chr6:51529803..51529870:+	87.88	62.22	48.97	50.18	-0.36	0.54
chr2:168311969..168312013:+	31.77	94.99	74.30	44.74	0.14	0.85
chr4:58983817..58983875:+	83.74	76.76	61.62	64.93	-0.10	0.86
chr4:132309972..132310026:+	433.30	529.67	415.84	382.18	-0.03	0.92
chr6:59705791..59705844:-	18.59	16.09	13.63	13.47	-0.14	0.93
chr7:121040347..121040416:-	81.70	145.51	50.95	87.46	-0.46	0.44
chr5:138881971..138882048:+	131.01	43.89	115.74	60.77	0.23	0.77
chr13:84058941..84058983:+	2768.83	2164.81	1534.95	2176.41	-0.16	0.62
chr7:127527940..127528000:+	61.37	83.79	68.48	47.16	-0.09	0.90

Supplementary Table 5: Sequences used in gene targeting and genotyping primers for *Prdm9*^{K} mice.

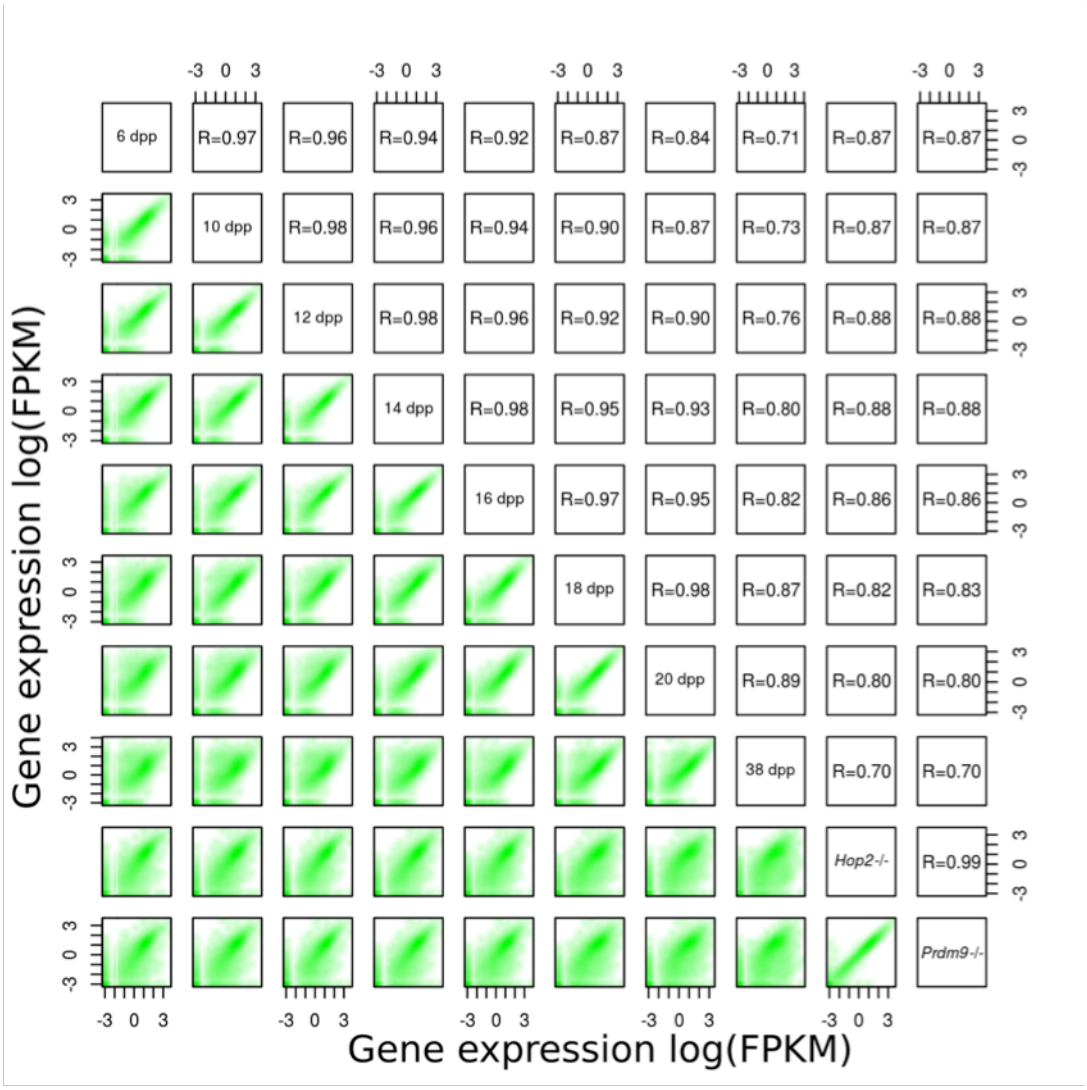
CRISPR-Cas9 Targeting:		
GuideRNAs:	5'-GGTCAAAGATGAATTCAAAG-3'	5'-CAAATAAATGACAGTGAGGA-3'
ssODN:	5'-AACCTTGGTCACTGAGACCACATTCCAGGGTCCCAAGCCCGGCCCAGCATGTTCCGTTTTTTGTTTTTATTTTTTTTCCCCCAGGTCAAACAAATAAATGACAGTGAGGATTCTGATGAAGAGTGGACACCTAAGCAACAAGGTGAGGAG-3'	
Genotyping Primers:	5'-GCATGAACATCCAGTGTTTGCTTAAAGAAGG-3'	5'-CTAAATCTGGCTTGGTTTCC-3'
qPCR Primers:		
Prdm9: (Exons 3-5)	5'-GATTCTGATGAAGAGTGGAC-3'	5'-GTTCTGAGCCACTTGTATTC-3'
Prdm9: (Exons 5-7)	5'-CCTCTGGACAGCACTCTGG-3'	5'-CTGAGTGGTTGGGATGCCC-3'
Prdm9: (Exons 7-9)	5'-GATGAAGAGGCAGCCAACAG-3'	5'-CATCCCGGGCACAGTTCA-3'
Gapdh:*	5'-AGGTCGGTGTGAACGGATTTG-3'	5'-GGGGTCGTTGATGGCAACA-3'

***** PrimerBank ID: 126012538c1 (170).

Supplementary Table 6: ChIP sample information

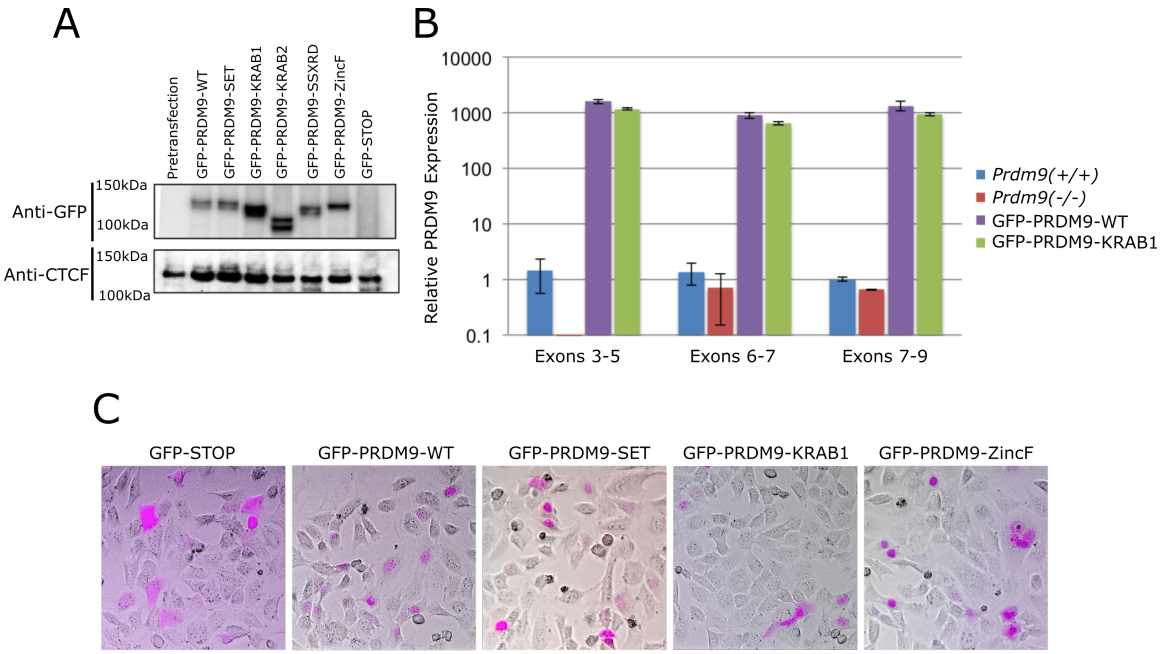
Sample	Source	Antibody	Genotype or Construct	Starting Material	Accession Number
GC-1 Cell Studies					
Pretransfection	SPG GC-1	Millipore (138-473)	n/a	~20x10^6 cells	
WT	SPG GC-1	Millipore (138-473)	GFP-PRDM9-WT	~20x10^6 cells	
SET	SPG GC-1	Millipore (138-473)	GFP-PRDM9-SET	~20x10^6 cells	
KRAB1	SPG GC-1	Millipore (138-473)	GFP-PRDM9-KRAB1	~20x10^6 cells	
KRAB2	SPG GC-1	Millipore (138-473)	GFP-PRDM9-KRAB2	~20x10^6 cells	
SSXRD	SPG GC-1	Millipore (138-473)	GFP-PRDM9-SSXRD	~20x10^6 cells	
ZincF	SPG GC-1	Millipore (138-473)	GFP-PRDM9-ZincF	~20x10^6 cells	
GFP-STOP	SPG GC-1	Millipore (138-473)	GFP-STOP	~20x10^6 cells	
Prdm9^{krab} Mouse Studies					
H3K4me3	Testis (60dpp)	Millipore (138-473)	*Prdm9$^{K/K}$*	Whole testis	
SSDS	Testis (60dpp)	SC 8973(C-20)	*Prdm9$^{K/K}$*	Whole testis	
Micrococcal Nuclease Studies					
WT	Testis (60dpp)	Millipore (138-473)	C57Bl/6	Whole testis	
Spo11$^{-/-}$	Testis (60dpp)	Millipore (138-473)	*Spo11$^{-/-}$*	Whole testis	

SUPPLEMENTARY FIGURES

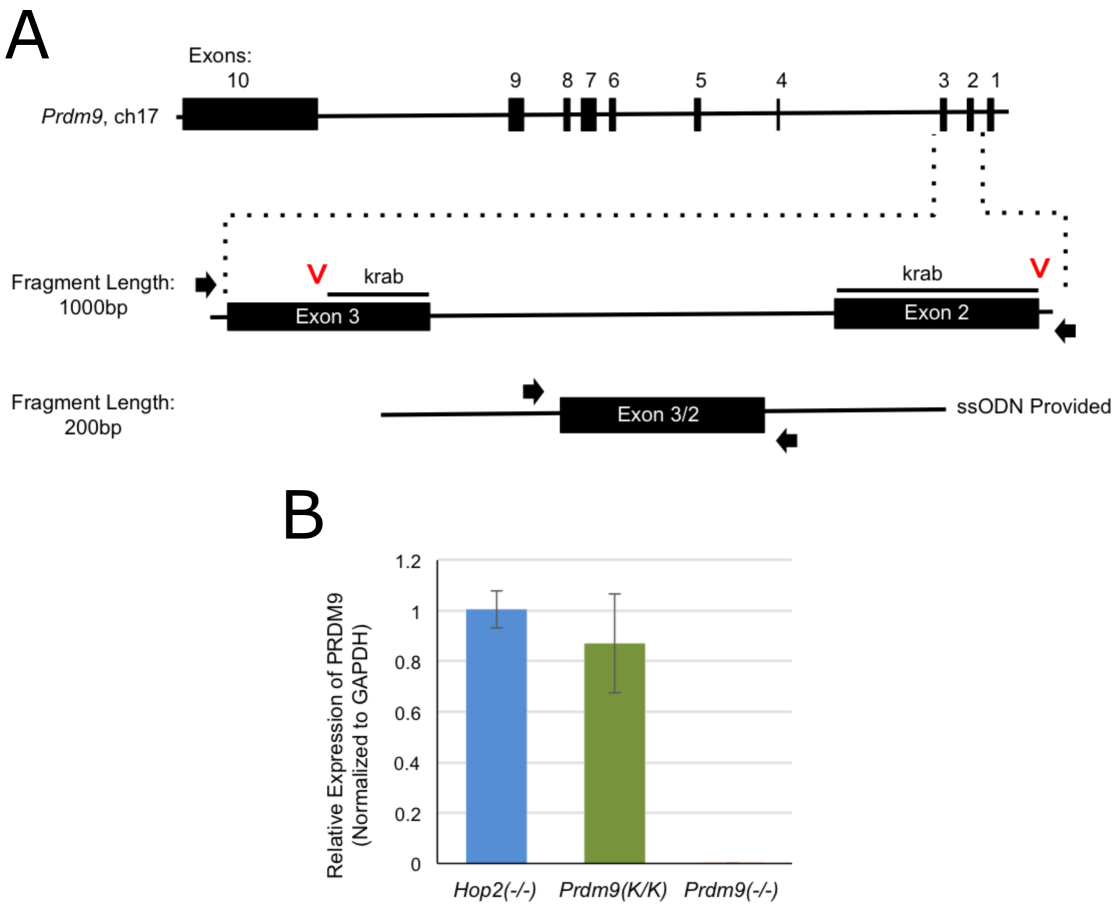


Supplementary Figure 1: Gene expression profile in *Prdm9*$^{-/-}$ testes is most similar to 12-14 dpp wild type testes.

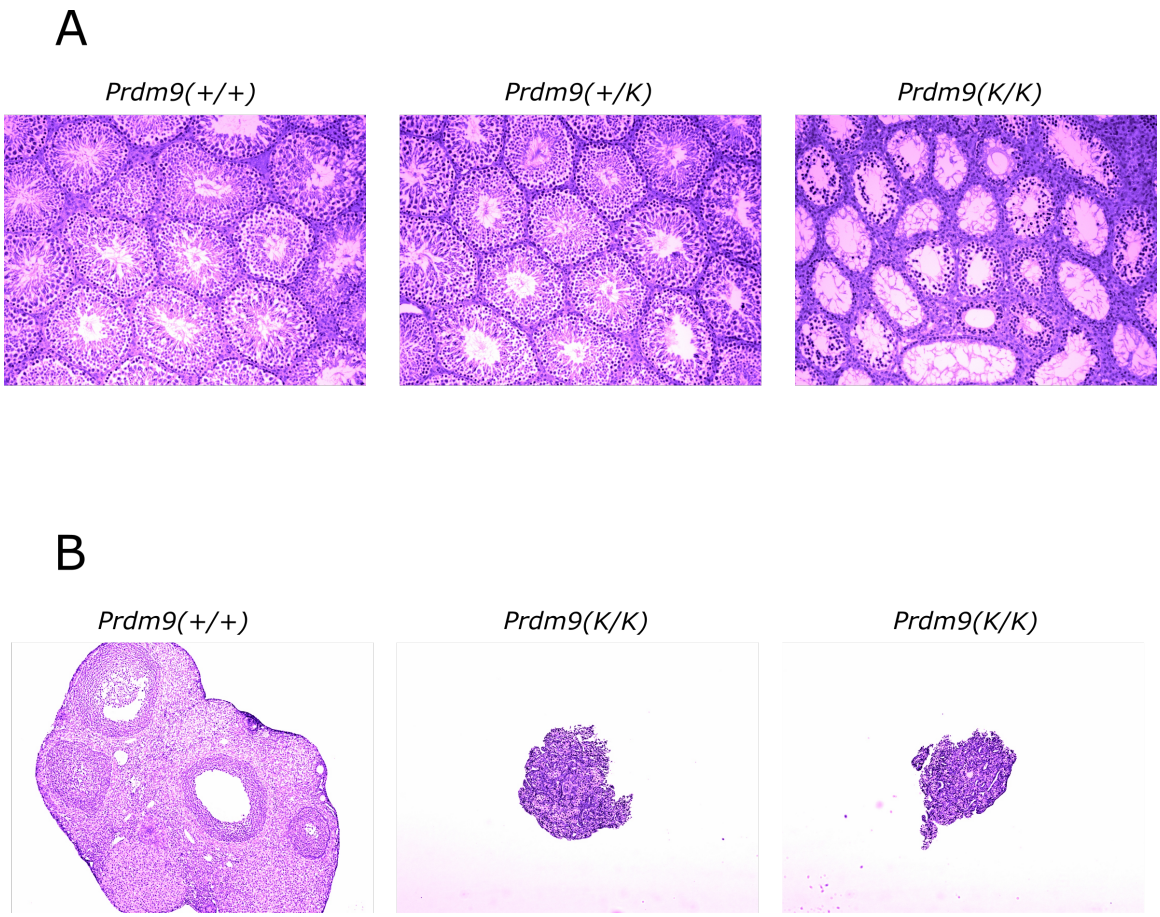
Scatter plots of RNA expression levels for *Prdm9*$^{-/-}$, *Hop2*$^{-/-}$ and juvenile 6 dpp - 38 dpp wild type mice measured by log (FPKM). Time course data from (109). Each point is the Fragment per million (FPKM) value for an individual gene in one sample against the FPKM value of the same gene in another sample. The number represents Spearman's correlation coefficient R.



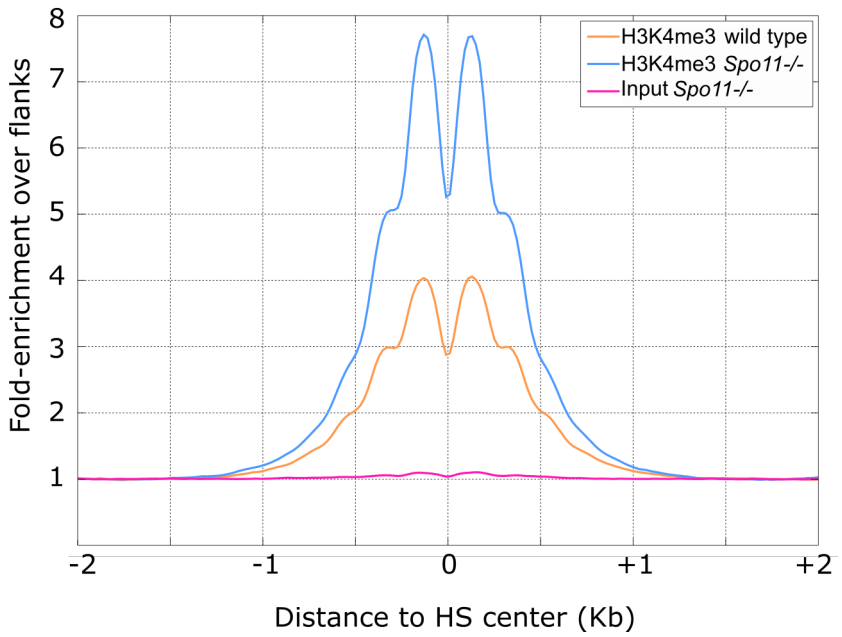
Supplementary Figure 2: Expression of GFP-PRDM9 constructs in GC-1 cells. GC-1 cells were transfected with indicated GFP-PRDM9 constructs described in Figure 2. (A) Western blot analysis of PRDM9 expression across all mutants. CTCF serves as a loading control. (B) qPCR analysis of *Prdm9* mRNA expression level across three exon-spanning junctions for *Prdm9*$^{+/+}$ mouse testes (blue), *Prdm9*$^{-/-}$ mouse testes (red) and GC-1 cells expressing GFP-PRDM9-WT (purple) or GFP-PRDM9-KRAB1 (38). *Prdm9*$^{-/-}$ mouse shows alternative transcript starting at Exon 5. (C) Live-cell imaging of GC-1 cells 13 hours post transfection, images shown at 100X magnification. GFP is shown in pink against phase contrast. All constructs show nuclear localization with limited presence in the cytoplasm, except GFP-STOP which localizes to both nucleus and cytoplasm.



Supplementary Figure 3: Gene targeting design for the KRAB deletion.
(A) DNA breaks were guided to form in exons 2 and 3 of *Prdm9* using guide RNAs and Cas9. A single-stranded oligonucleotide (ssODN) with desired deletion of the KRAB domain was provided (sequences listed in Supplementary Table S5). Red arrows indicate break sites. Genotyping primers are indicated as black arrows. (B) Expression of $Prdm9^{K/K}$ transcript was confirmed by qPCR for *Prdm9* exons 3-5, normalized to GAPDH.



Supplementary Figure 4: *Prdm9*$^{K/K}$ mice do not produce functional gametes.
(A) Testes cross-sections from *Prdm9*$^{+/+}$, *Prdm9*$^{+/K}$ or *Prdm9*$^{K/K}$ mice stained with hematoxylin and eosin (H&E), images were taken at 100X. (B) Ovary cross-sections from either *Prdm9*$^{+/+}$ or *Prdm9*$^{K/K}$ mice stained with H&E, images were taken at 100X. No follicles are visible in *Prdm9*$^{K/K}$ ovaries and no spermatids are present in in *Prdm9*$^{K/K}$ testes.



Supplementary Figure 5: PRDM9, not DSB processing, promotes the formation of nucleosome-depleted region at hotspot centers.

MNase-treated chromatin from wild type and *Spo11*$^{-/-}$ mouse testes was used for H3K4me3 ChIP-seq, allowing nucleosome-resolution H3K4me3 mapping around HS. Since DSBs do not form in *Spo11*$^{-/-}$ mice (77; 149) the nucleosome reorganization is due to PRDM9 action and is not the consequence of downstream DSB formation and processing. Meiosis is arrested in *Spo11*$^{-/-}$ mice, leading to exaggerated PRDM9-specific H3K4me3 enrichment above the wild type levels.

MATERIALS AND METHODS

Mice and cell lines

Gene expression studies were performed using the GC-1 spg immortalized BALB/c cell line derived from spermatogonial cells (ATCC CRL-2053, (65)). All animal procedures have been approved by the USUHS Animal Care and Use Committee. *Hop2* knockout mice have been previously described (142). *Prdm9* knockout mice have been previously described (62) and were obtained from the Jackson Laboratory (stock number 010719). *Spo11* knockout mice (strain *deltaSpo.BC/B6*) have been described in (166). C57Bl/6 was used as a wild type strain and all mutant strains were on a C57Bl/6 background. Adult (2-6 month old) mice were used for the analyses unless specifically mentioned in the text.

RNA sequencing

Total RNA was isolated using Qiagen RNeasy mini kit (74104). Polyadenylated RNA was purified using Oligotex beads (Qiagen, 70022). Quality of the RNAs was verified on Bioanalyzer and RNA was fragmented to ~200 bp. Sequencing libraries were prepared using mRNA-Seq Sample Preparation Kit (Illumina, RS-930-10) and sequencing was performed using Illumina X2000 platform with 50 bp paired-end reads. Reads were mapped to the mouse genome mm10 using bowtie 2.1.0.0 (89) which is built-in by Tophat v2.0.8b (80). Cufflink 2.1.0 (175) was used to assemble the transcripts, and calculate FPKM. Cuffdiff 2.1.0 (175) was used to perform the differential expression tests of genes and differential splicing tests of primary transcripts from genes with two or more isoforms. Transcripts were identified as present if FPKM > 0.1 in two replicates (103).

miRNA analysis

Small RNAs were isolated from whole testes using the mirVana miRNA Isolation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The resulting RNA (~ 50 μg) was heat-denatured, loaded and resolved on the 15% denaturing acrylamide gel. Two radiolabeled RNA oligos of 18 and 28 bases were used as the ladder. Gel was run for 3h at 45W. The gel slice containing RNA from 19 to 27 bases was excised and RNA was eluted in 800 μl of 0.4 M NaCl overnight. RNA was precipitated with ethanol and resuspended in 6 μl of H_2O (~ 200 ng of RNA). Library was prepared with the NEBNext Small RNA Library Prep Set for Illumina (New England BioLabs E7330S) following manufacturer's instructions. The microRNA samples were sequenced using Illumina HiSeq 2500 using single-end reads. After trimming the adapter reads were filtered using the length threshold of 17-29 bp. Reads were mapped to the mm10 genome using bowtie 0.12.9 (90) with parameters "bowtie -f -n 0 -e 80 -l 18 -a -m 5 –best –strata", which are set as default in MiRDeep2 v2.0.0.5 (47). Significance testing for the differential analysis was performed using DESeq (4) (Bioconductor version 2.11). The novel miRNAs were discovered using miRDeep2 (Version 21). We define common miRNAs between the four samples as miRNAs that have at least 50% sequence overlap between the samples. These miRNAs are considered expressed if their TPM is larger than 10. We used human miRNAs as templates for predicting novel miRNAs.

GFP-PRDM9 constructs

Full length B6 mouse PRDM9 following GFP was cloned into pcDNA4-Myc-His-B (ThermoFisher). A linker region of the following sequence separated GFP from PRDM9 "GGAGGCTCTGGCGGATCTGGAGGGTCTGGAGGT." PRDM9 start codon was removed but the stop codon was maintained. Mutant constructs were generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). The specific mutations for constructs in Figure 3 are as follows:

SET: Y345F; KRAB1: del(D30-P87), KRAB2: del(K27-V198V), SSXRD: del(L175-Y206); ZincF: del(H392-W413). GFP-STOP was generated by the following mutation in linker region above: G1T, G2A.

Cell culture and transfection

GC-1 cells were grown in DMEM supplemented with 10% FBS and 5% Penn-Strep at 37°C in 5% CO_2. Cells were plated and allowed to grow for 15 hours before transfection using Lipofectamine 3000 (Thermo Fisher Scientific). Cells were allowed to grow in transfection solution for 15 hours before collection.

ChIP-seq

GC-1 cells were fixed directly on the plate as described in (7). ChIP for testis and GC-1 cell samples was carried out as previously described (22).

For single nucleosome resolution H3K4me3 ChIP-seq instead of sonication the fixed nuclei were resuspended in MNase buffer (50 mmol TrisHCl, 1 mmol $CaCl_2$, 4 mmol $MgCl_2$, 4% NP-40) at 26 µl/10^6 cells, supplemented with 1 mmol PMSF and 1X

cOmplete, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich). The Micrococcal nuclease digestion was then carried out as described in (17). Following digestion, the sample was diluted 3:1 in ChIP buffer (0.1% SDS, 1% NP-40, 2mM EDTA, 20 mM Tris-HCl pH8, 150mM NaCl) and H3K4me3 ChIP was performed as in (22). The following antibodies were used: anti-DMC1, Santa Cruz sc8973; anti-H3K4me3, Millipore #07-473. H3K4me3 sequencing was performed with 75 bp SE reads on an Illumina NextSeq500 and SSDS sequencing was performed with 50bp PE reads on an Illumina HiSeq 2500. H3K4me3 ChIP-sequencing reads were aligned to the mouse genome mm10 using bwa aln. Overlap of reads with either promoters or known hotspots was performed using bedtools 2.26.0 Intersect. Bigwig files were generated with deeptools 1.5.11 bamCoverage with the following parameters: --binSize 10 --centerReads --normalizeUsingRPKM. Bigwig files were then used to generate heatmaps across promoters or hotspots using deeptools computeMatrix reference point followed by plotHeatmap. SSDS reads were aligned to the genome using bwa aln (0.7.12) and ssDNA derived reads were identified using the single stranded DNA sequencing (SSDS) processing pipeline (79).

Normalization of H3K4me3 ChIP-seq samples

As the H3K4me3 ChIP-seq experiments produced samples of varying quality we normalized reads across promoters to more accurately assess the activity of PRDM9. As all the experiments were performed under similar conditions in the same cell type we can expect that biologically they should have similar H3K4me3 enrichment at the promoters. We then use this built-in positive control to normalize for the quality of the sample by

averaging the sequencing tags for each sample across 150 random promoters. We then calculate a normalization factor which allows us to "scale down" the sequencing tag enrichment to the lowest quality sample. Tag counts across 12,840 additional promoters and 11,642 B6 hotspots were then calculated and normalized in this fashion. We confirm that data processed in this way gives a reasonable result by comparing the normalized tag enrichment between different samples (Figure 4 A-C). All samples cluster around x=y line, showing successful normalization between samples and allowing us to estimate the activity of various PRDM9 mutant activity compared to WT.

Identification of DSB hotspots

Uniquely mapped fragments unambiguously derived from ssDNA (ssDNA type 1) and having both reads with a mapping quality score ≥ 30 were used for identifying hotspot locations (peak calling). NCIS was used to estimate the background fraction for each library. Peak calling was performed using MACS (v.2.1.0.20150420) (194) . with the following parameters : --ratio [output from NCIS] -g mm --bw 1000 --keep-dup all --slocal 5000. DSB hotspots within regions previously blacklisted (165) were removed and hotspot strength was calculated as described previously (165).

Assessment of PRDM9 expression in GC-1 cells

Live cell images were taken 12 hours post transfection to confirm nuclear localization and estimate expression level of GFP-PRDM9 constructs. During ChIP procedure, 50 µl aliquot of sheared chromatin lysate was removed and frozen at -80°C. The sample was incubated at 95°C for 10 minutes in SDS loading buffer with DTT and resolved in Bolt

Bis-Tris Plus acrylamide Gel followed by transfer onto PVDF membrane. The following antibodies were used: rabbit anti-GFP (Abcam, AB290) and rabbit anti-CTCF (Abcam, AB70303). The blots were developed using SuperSignal West Pico PLUS Chemiluminescent Substrate.

Assessment of PRDM9 expression in mouse testis

Testes were removed and stored at -80°C in RNALater (Thermo Fisher Scientific). Total RNA was extracted using RNeasy plus mini kit (Qiagen) and the cDNA was prepared with the Transcriptor First Strand cDNA Synthesis Kit (168). Quantitative PCR was performed using the SYBR Green PCR Master Mix (Thermo Fisher Scientific) using a 7500 AB Real-Time PCR system with ABI SDS Software. The primers used are listed in Supplemental Table 5.

Generation of the mutant *Prdm9* mouse strain without the KRAB domain

Prdm9^{krab} mice were generated on a C57Bl/6 background as outlined in Supplementary Figure 3 and Supplementary Table 5. Targeting design and pronuclear injections were done by University of California Irvine transgenic mouse facility.

Histology, Chromosome Spreads and Immunostaining.

Testes or ovaries were dissected and stored in buffered Formalin (1:10 dilution). Tissue sections with hematoxylin/eosin staining was done by American HistoLabs Inc (Gaithersburg, MD). Meiotic chromosomes spreads were made and immunostained as described in (166). The following primary antibodies were used: anti-SCP3 (142); anti-

γH2AX (Trevigen, 4411-PC-100). Samples were imaged using Leica fluorescent DM5500 microscope with Retiga Camera and Volocity Acquisition System.

CHAPTER 4: CONCLUSION

SPECIFIC AIMS

During this project we investigated the following two specific aims:

1. How is the binding of PRDM9 regulated? While there are over a million instances of the PRDM9-recognized motif within the genome, PRDM9 binding consistently occurs at a subset of approximately 20,000 sites. We show that distinct DNA methylation patterns regulate how frequently PRDM9 is able to bind to certain motifs. Excitingly, this regulation was correlated with a biological difference in DSB hotspot strength between males and females.

2. How is DSB formation restricted to sites marked with PRDM9-specific H3K4me3? H3K4me3 is one of the most common chromatin modifications across the genome and it is unclear how DSB formation is restricted to a subset of these sites. PRDM9 has two domains which are typically involved in gene repressive functions, the KRAB and SSXR domains. We show that PRDM9 has diverged from this typical function and has adapted a new role in setting the proper chromatin environment for initiation and completion of homologous recombination. However, we propose that the mechanism of DSB machinery recruitment is more complex than a simple tethering mechanism. Additionally we propose that these protein-protein interacting domains are important for histone trimethylation activity at the PRDM9 binding site.

REGULATION OF PRDM9 BINDING

Specific DNA methylation patterns can either favor or disfavor PRDM9 binding

The C57Bl/6 mouse genome has over 1 million instances of the PRDM9 binding motif, however only a subset representing approximately 20,000 of these motifs appear to be used consistently. This suggests that PRDM9 DNA-binding is regulated beyond recognition of the primary sequence alone. Originally we suspected that DNA methylation might be used to enhance a primary sequence for PRDM9 attraction or repulsion. DNA methylation results in a physical alteration to the DNA bases along the major groove surface, the same region which is recognized by the DNA binding domain. Additionally, there are instances where DNA methylation is known to regulate DNA binding of other proteins (32; 102; 183).

When we began to look into the role that DNA methylation could play in regulating PRDM9 binding we anticipated large-scale alterations in binding. However, the overall usage of PRDM9 binding sites in the hypomethylated germ cells of the *Dnmt3l*$^{-/-}$ mice did not display a dramatic alteration: we did not see any hotspots formed at previously unused PRDM9 motifs compared to the wild type mouse. Unexpectedly, it became clear that although the same locations happened to be used by PRDM9 binding in both the hypomethylated and the wild type mouse, the frequency that each individual site was used differed between the two. For example, even though a particular site might be frequently bound by PRDM9 in the wild type mouse (yielding strong H3K4me3 and SSDS enrichment), the same site in the hypomethylated mouse might yield weaker enrichment, although it was still used.

Interplay of DNA methylation and hotspot usage between the sexes

As mentioned in the Introductory Chapter, male and female germ cells have dramatically different epigenetic landscapes during the period when PRDM9 is active in gametogenesis. Specifically, the male germ cells have reestablished some level of DNA methylation while the female germ cells are globally depleted for methylation when meiotic entry occurs. The discovery that sex biased hotspot locations could be explained by the DNA methylation pattern in the respective germ cells was very exciting. These results were underscored by our discovery that the hypomethylated *Dnmt3l*$^{-/-}$ mouse had both a higher usage of female-biased hotspots and lower usage of male-biased hotspots. As we were able to alter the hotspot usage simply by altering the DNA methylation pattern, we confirmed that the methylation specifically helped lead to sex-biased hotspot usage.

While the altered usage of these sites is notable on its own, what was truly striking was the fine-scale effect which DNA methylation played in either encouraging or discouraging PRDM9 binding at a particular motif. We had originally suspected that the presence or absence of DNA methylation at a motif would influence binding. However, we did not expect that methylated bases at one region of the motif could be inhibitory while methylated bases at a different region within the motif could promote binding.

It is interesting to note that beyond the different PRDM9-specified DSBs, there appear to be other sex-specific pathways at work which continue to modify meiotic progression. An interesting example we report is the inverse relationship between DSB formation and crossover formation at subtelomeric regions in females. Even though female germ cells are more likely to form DSBs in the subtelomeric region, there appears

to be a mechanism at work which disfavors these sites developing into crossovers. This result is in contrast to males, which exhibit the reverse trend: while fewer DSBs are formed in the subtelomeric region, they are more likely to develop into crossovers than other sites.

This is an intriguing sex-specific adaptation, as chromosomes with subtelomeric crossovers are more likely to mis-segregate in female meiosis (107; 130). It has been suggested that these subtelomeric crossovers are physically less stable than non-subtelomeric regions over the long meiotic arrest experienced in females. It follows that females would have developed a pathway to prevent this from happening, while males would not need a similar mechanism to prevent against breakdown over time. Additionally, it must be noted that subtelomeric crossovers serve a vital purpose in male meiosis: the X and Y chromosome can only crossover at the subtelomeric PAR. As males are more likely to form crossovers at subtelomeric sites than across the rest of the chromosome length it suggests an opposite mechanism which promotes crossover formation at the subtelomeric regions.

However this pathway(s) functions, this represents clear evidence that DSBs are processed differently in males versus females. Further, it suggests that the mechanism that directs DSBs to become crossovers interacts with other genomic features to determine the intrachromosomal location.

Limitations of study

The main limitation of this study is the use of the *Dnmt3l*$^{-/-}$ mouse. While this mouse is globally hypomethylated compared to the wild type mouse, it is not a major

depletion. Perhaps these results would have been exaggerated if there was further DNA methylation depletion. Additionally, as DNMT3l is important for spermatogenesis, these mice are sterile. To assess if early meiotic failure could lead to the altered PRDM9-binding phenotype, we performed meiotic spreads and confirmed that 30dpp mice contained spermatocytes which had progressed beyond the stage when PRDM9 becomes active (data not shown). Even though we do not think this meiotic arrest is a compounding factor, we cannot definitively rule it out.

In order to get around this problem we considered two additional experiments: using chemotherapeutic agents to decrease the testis DNA methylation and using a chemotherapeutic agent to decrease the global DNA methylation in the GC-1 cells followed by PRDM9 expression. Ultimately pilot experiments for both of these procedures did not appear to decrease the DNA methylation as we had hoped. For this reason we focused only on the *Dnmt3l*$^{-/-}$ mouse as the hypomethylation is well documented.

Future directions

These studies could be continued by in depth analysis into how DNA methylation affects PRDM9 binding and/or attempting to elucidate the sex-specific pathways during meiosis.

Biochemical assays and the additional experiments attempted above could help assay the impact of methylation at individual motif bases on PRDM9 binding. Additionally, while this study was performed in mice, it would be interesting to investigate how DNA methylation could play a role in hotspot formation in humans.

Similar to mice, humans also have PRDM9, and human PRDM9 binding is likely affected in a similar way. This would be very interesting to study in the context of the wide genetic diversity of PRDM9 recognized motifs in humans. While there have been multiple clinical studies which link DNA methyltransferase mutations to infertility (reviewed in (178)), it is likely that this infertility is due to meiotic failure from incorrect genome silencing (as seen in the *Dnmt3l*$^{-/-}$) mouse and is not a consequence of altered hotspot formation.

Additionally, while all of this work was done looking at the C57Bl/6 *Prdm9* allele (*Dom2*) in the B6 mouse it would be very interesting to assess the methylation patterns at motifs of different PRDM9 alleles in various strains of mice. We, along with others, have demonstrated that PRDM9 alleles appear to have a pseudo-dominant effect. For example, in a cross between genetically similar mice with two different PRDM9 alleles, one allele will determine the location of the majority of the hotspots which are generated (22; 165). When we originally reported this dominance effect we performed an analysis to see if hotspot usage varied within a subset of regions known to have imprinting marks, including DNA methylation, and found no association (165). In light of our recent findings that trace amounts of DNA methylation at specific bases in the motif impacts PRDM9 binding, it would be interesting to revisit this analysis.

Finally, as there is a drastic difference in nondisjunction rates of homologs between the sexes, any biological differences which would impact homologous recombination are of great interest. Not only does this work make a direct link between the epigenomic environment on meiotic progression between the sexes, it also shows evidence that there are sex-specific mechanisms in how sites of recombination are

utilized. It will be of great interest in the field to continue with this line of reasoning and to expand the scope of studies to other organisms, such as humans, which exhibit a longer meiotic arrest.

Additionally, in recent years there has been both public and scientific interest in the effect of inadvertently altering the epigenetic markers (for example through chemical exposure) on fertility and fecundity. As our results show evidence that epigenetic markers do play a regulatory role in determining locations of homologous recombination this is another consequence we should keep in mind while forming future studies.

FORMATION OF DSBS AT SITES BOUND BY PRDM9

The KRAB domain

Over the course of this project, multiple other labs undertook independent investigations into the function of the PRDM9 KRAB domain. Thus, we had the benefit of comparing and contrasting our results with two other papers which were published in 2017. Both papers utilized yeast 2-hybrid assays and concluded that the KRAB domain was involved in protein-protein interactions with the loop-axis machinery. To confirm these results one group used a CoIP for PRDM9 in mouse testes (136) and the other made a transgenic mouse with an N-terminal truncated version of PRDM9 (67). Interestingly, the mouse with a truncated KRAB domain appeared to have lost all the histone methylation activity and had a similar phenotype as the PRDM9 knockout mouse.

While this loss of methylation activity was not addressed extensively in the paper, it did contrast with our results showing that PRDM9 lacking the KRAB domain appeared to be active in cell culture and showed low levels of activity *in vivo*. As we report, this difference in expression level does not alter the overall phenotype of the mouse, as the majority of the hotspots still occur at default sites and the mouse is sterile. This suggests that though the KRAB domain does not appear to be imperative for the enzymatic function of PRDM9, it does prove necessary for PRDM9 function during meiosis. Underscoring this result, a large-scale survey of organisms showed that animals which had lost the KRAB domain/SSXRD of PRDM9 appeared to no longer have PRDM9 directed hotspot formation (9).

Regardless of how the KRAB domain of PRDM9 influences the activity, the frequent loss of PRDM9 and specifically the KRAB region of PRDM9, throughout

evolution is striking. This begs the question: is PRDM9-dependent hotspot formation an improvement on non-PRDM9 dependent hotspot formation?

PRDM9-dependent hotspot formation does prevent hotspots from forming at deleterious regions while also promoting genetic diversity through the wide range of alleles. However, it comes with a host of problems, including the overarching "hotspot paradox." When PRDM9 binds strongly to a motif, it triggers the formation of DSB at this location which is then repaired using the homolog with a presumably weaker PRDM9 motif (as PRDM9 did not bind there). Over time this will erode the best PRDM9 motifs from the genome, resulting in decreased PRDM9 fitness in a genome with only weaker motifs remaining. In fact, this mechanism has been shown many times over, most strikingly it was predicted that the strongest PRDM9 motifs used in present humans will no longer exist in 3 million years (93).

Additionally, as PRDM9 divergence is a speciation event it renders individuals from two similar populations incapable of interbreeding (118). This could have negative consequences for very small populations and might be deleterious in some instances. Further, the evidence that so many organisms have reverted "back" to non-PRDM9-dependent hotspot determination suggests that PRDM9 is not imperative for evolutionary success. And unlike in the knockout mouse, a genome-wide population study uncovered at least one record of a natural "PRDM9 human knockout" who was capable of reproducing (121). It will be very interesting to study animals which do not have a functional PRDM9 and/or a PRDM9 which has naturally lost the KRAB domain to compare differences in regulatory pathways which determine the formations of DSBs and COs.

The SSXRD

While not the main focus of this work, we also used our cell system to compare the function of PRDM9 lacking either the SSXRD or the post-SET lone zinc finger. These domains, like the KRAB domain, have yet to be explored in detail for PRDM9. To our knowledge this is the first study to look at the SSXRD of PRDM9, and possibly only the second publication to assess the function of the SSXRD in any protein. While the SSXRD is a small portion of PRDM9, the potential tie in as a gene regulator with an unknown molecular mechanism made it an interesting target for our studies.

Unexpectedly, removing this domain led to complete inactivity of the protein. While we confirmed that the protein was present in similar levels to the WT and retained the nuclear localization, we can not rule out protein misfolding as a cause of the phenotype. Additionally, this complete loss of function was unexpected as a similar mutant, KRAB2, lacked approximately two thirds of the SSXRD in addition to removing the KRAB domain. However, unlike the full SSXRD removal, this partially removed construct retained the same activity as the KRAB1 construct. If the SSXRD does play an imperative role in PRDM9 function or structure, we would expect that more than the 8 C-terminal-most residues would need to be retained for proper protein function. Alternatively, it is possible that these last 8 amino acids are crucial for SSXRD activity and prove capable of modulating the negative effects of removing the entire KRAB domain on PRDM9 structure.

Assessing the role of the PRDM9 domains is complicated as the nomenclature and amino acid coordinates for the domains has varied throughout the publication history. An interesting example involves the partial crystal structure of PRDM9, spanning the last

amino acids of the SSXRD through the post-SET zinc finger (187). This led to the proposal that a pre- and post-SET domain "zinc finger" interacted to promote or restrict access of substrates to the active site as necessary. However, the proposed "pre-SET zinc finger" appears to be located somewhere between the end of the SSXRD and the SET domain, a region that classically has not been annotated as a potential domain (although it was recently referred to in a publication as the "zinc knuckle" (3)). It is possible, though, that there is an undescribed zinc finger domain between the KRAB and SET domains which has been missed by annotation. If this is indeed the case, it is possible that our inactive phenotype is actually due to a downstream effect, where the structure of this undescribed zinc finger is disrupted by alteration of the SSXRD.

This study shows the need for further characterization of the PRDM9 structure. It is very difficult to assess how individual domains are involved in the function of PRDM9 when their location and orientation in the folded protein is a mystery. While an interaction between the pre- and post-SET domain regions was proposed to have a regulatory role, it could also be the case that this interaction forms a "backbone" which allows for protein stability. In any case, it will be very interesting to continue the studies of the SSXRD and the region immediately surrounding it. As extremely little is known about the mechanism of this domain, any investigations can further our understanding of the evolutionary history of the SSXRD.

The lone zinc finger

Unlike the KRAB domain and the SSXRD, the lone zinc finger following the SET domain has been involved in previous studies. While this zinc finger is located outside of

the DNA-binding array of zinc fingers, we were curious if it still interacted with the DNA through non-specific interactions. Alternatively, it was proposed that this finger is involved in intra-molecular interactions which regulate the SET domain (187).

In either of these cases, our finding that removing this zinc finger does not appear to greatly impact PRDM9 activity is unexpected. We had originally suspected that the zinc finger played a role in strengthening the DNA binding, and expected that we would see a strong decrease in H3K4me3 at the hotspots which correlated to weak DNA binding. Alternatively, if this lone zinc finger was involved in regulating PRDM9 as mentioned above, it was possible we would see either no sign of activity or higher activity than in the wild type due to misregulation.

Unlike the other domains, an earlier study suggested that the lone zinc finger of PRDM might play a role in the stability of PRDM9 (182). The PRDM9 DNA-binding array of zinc fingers was expressed with and without the additional lone post-SET domain zinc finger. This protein segment was then incubated with DNA fragments and the DNA-binding ability was assessed. It was determined that when the lone zinc finger was included with the zinc finger array there was no DNA-binding activity. Alternatively when the zinc finger binding domain was expressed without the lone zinc finger it retained full DNA-binding activity. This would initially suggest that the lone zinc finger has an adverse effect on the DNA binding of PRDM9. However, this loss of function could be rescued in the construct containing the lone zinc finger if the protein product was incubated in testes extract with a high protein concentration. This led to the conclusion that the lack of activity was a protein folding problem which was mitigated by

a high protein concentration acting as a chaperone, leading the PRDM9 fragment to fold correctly.

This is an interesting idea regarding the potential of high protein concentration to encourage correct folding of these protein domains. Beyond that it highlights the potential differences between studying the DNA-binding ability of PRDM9 in a binding assay versus assessing the binding *in vivo*. While it is likely that the results will be mostly similar between the two, potential differences in substrate and the meiotic environment should be kept in mind.

Limitations of study

The main weakness when conducting any studies into PRDM9 activity is the lack of reliable PRDM9 antibody. We attempted to get around this problem by both generating our own antibody and side-stepping the antibody production to generate a GFP-tagged-PRDM9 knock-in mouse line. While we isolated an antibody which worked nicely in our cell culture assays when PRDM9 was overexpressed, we never saw similar results with testes extracts. Additionally, we did not have success integrating our GFP-tag while using CRISPR/Cas9 single cell embryo targeting (results not shown). However, multiple labs have attempted similar tagging schemes which proved unsuccessful due to protein instability (personal correspondence). Without the antibody or tagged transgenic mouse we are unable to separate the DNA-binding ability of PRDM9 with the histone methyltransferase activity. It is also possible that the methyltransfer phenotype we see both with the *Prdm9*$^{K/K}$ mouse and the cell system could be due to altered DNA-binding

affinities. For example, PRDM9 that is largely incapable of binding DNA will show the same overall decreased histone methyltransferase phenotype.

Further, we are unable to discern between the contribution of the KRAB domain to PRDM9 enzymatic function versus the contribution of the KRAB domain to the structural integrity of PRDM9. To get around this problem we first confirmed that PRDM9 was still active in the cell line following removal of the KRAB domain. However, as the KRAB-less PRDM9 appeared to have a much lower activity in the mouse, this finding was complicated. There are two main possibilities: that the KRAB domain is structurally vital to PRDM9 or that the KRAB domain is needed specifically for full PRDM9 function in the mouse but not in the cells. In either case, if the KRAB domain is necessary for correct folding of PRDM9 or its enzymatic function, it is possible the cells were overexpressing the protein to such a degree that the impaired protein still reached a threshold of activity which appeared similar to the wild type. Ideally the absolute methyltransferase activity could be compared between the WT and KRAB-less PRDM9 construct through stable transfection of the PRDM9 construct under an inducible promoter using the GC-1 cells. If the same amount of PRDM9 can be expressed in all cells in the experiment, and at near biological levels, then qPCR can be used to reliably compare the activities. However, this comparison can not be made with our current system, which has variable expression levels between individual cells in the experiment that far exceed biological levels.

Another limitation is separating the enzymatic activity of PRDM9 from the DSB formation activity. Again, we attempted to discern this difference by using a cell expression system which was isolated from the downstream DSB formation. Our

assumption was that the protein-protein interactions which promoted the bound sites to form DSBs would be separate from the trimethylation activity. However, if PRDM9 instead uses protein-protein interactions to form a proteinaceous complex which then couples both the trimethylation activity and the formation of DSBs this will compound the result. And further, if PRDM9 is capable of DNA-binding and histone methylation either in isolation or as part of a larger protein complex, these differences could not be elucidated.

Future directions

It would be very difficult to generate an environment to truly test the function of the KRAB domain of PRDM9. Biochemical DNA binding and histone methyltransferase assays would show if the KRAB domain was necessary for specific PRDM9 activity, however with PRDM9 there has been very little success in producing large amounts of the full-length isolated protein. However, for this study we successfully generated a transient cell line which expressed relatively large amounts of GFP-PRDM9. It is possible that this system could be altered to allow protein isolation and then various biochemical assays could be performed.

While up to this point the focus has been on PRDM9, perhaps it would be best to elucidate the function by studying the proteins we suspect it interacts with. It would be very interesting to see what PRDM9 activity would look like in a model which lacked CXXC1, the proposed binding partner of PRDM9. It was proposed that PRDM9 and CXXC1 binding helps facilitate the formation of DSBs at PRDM9-bound sites (67). Alternatively, we propose that the interactions between PRDM9 and CXXC1 occur

within a larger complex which is necessary for full histone trimethylation. If this is indeed the case we could expect that depleting CXXC1 might abolish PRDM9 activity as well. Realistically, this would be difficult as CXXC1 is present in virtually every cell type. It would be fascinating to generate a transgenic mouse with a spermatocyte-specific knock-out of CXXC1. If CXXC1 is only imperative for linking PRDM9 marked chromatin to the chromosome core we would expect that the mouse would have a very similar phenotype: PRDM9-specific H3K4me3 but DSBs formed at both the default and the PRDM9-specific hotspot locations.

Also, while there is no reliable antibody for PRDM9, there are many commercial antibodies available for CXXC1. If ChIP can be performed for CXXC1 we might be able to determine if it is truly forming a link between the hotspot locations and PRDM9-binding sites. It would be very interesting to perform this same IP in the *Prdm9*$^{K/K}$ mouse to determine if this proposed interaction is truly abolished.

CLOSING

While it is widely accepted within our field that PRDM9 promotes the formation of DSBs through the placement of H3K4me3 in mice, the various mechanisms which regulate this process are unknown. It was unclear how PRDM9 binding is restricted to a relatively small number of available recognized motifs. Additionally, it was unclear how DSB machinery was restricted to act at the locations of PRDM9-specific H3K4me3, as opposed to endogenous H3K4me3 (Figure 1).

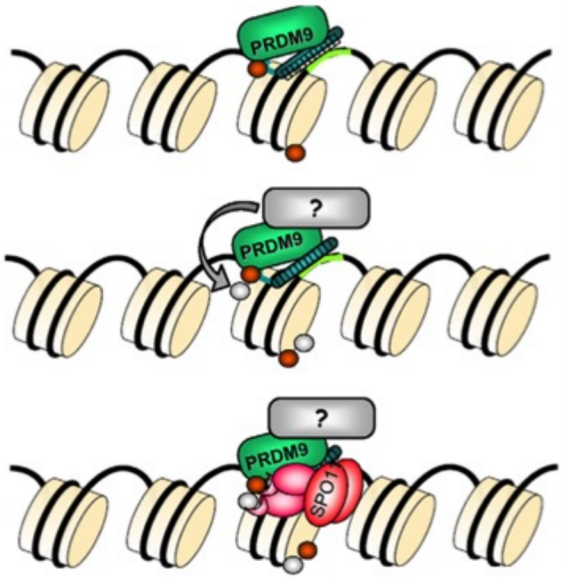


Figure 1: Model of hotspot specification by PRDM9.
Figure adapted from (52).

This work has advanced the understanding of this regulation with two proposed mechanisms. First, we show evidence that existing epigenetic factors can influence the strength of PRDM9 binding to the motif. Second, we propose that PRDM9 acts within a meiosis-specific complex which restricts DSBs from forming at sites enriched with non-PRDM9 H3K4me3. Additionally, our work shows conclusively that PRDM9 is not

involved in gene regulation, a role which has been discussed since the initial identification of PRDM9.

REFERENCES

1. Albrecht KH, Eicher EM. 2001. Evidence that Sry is expressed in pre-Sertoli cells and Sertoli and granulosa cells have a common precursor. *Dev. Biol.* 240:92-107
2. Allais-Bonnet A, Pailhoux E. 2014. Role of the prion protein family in the gonads. *Front Cell Dev Biol* 2:56
3. Altemose N, Noor N, Bitoun E, Tumian A, Imbeault M, et al. 2017. Human PRDM9 Can Bind And Activate Promoters, And Other Zinc-Finger Proteins Associate With Reduced Recombination In cis. *bioRxiv*
4. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, et al. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* 8:1765-86
5. Anderson LK, Reeves A, Webb LM, Ashley T. 1999. Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. *Genetics* 151:1569-79
6. Arnheim N, Calabrese P, Tiemann-Boege I. 2007. Mammalian meiotic recombination hot spots. *Annu. Rev. Genet.* 41:369-99
7. Baker CL, Petkova P, Walker M, Flachs P, Mihola O, et al. 2015. Multimer Formation Explains Allelic Suppression of PRDM9 Recombination Hotspots. *PLoS Genet.* 11:e1005512
8. Baker CL, Walker M, Kajita S, Petkov PM, Paigen K. 2014. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res.* 24:724-32
9. Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, et al. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *Elife* 6
10. Baltus AE, Menke DB, Hu Y-C, Goodheart ML, Carpenter AE, et al. 2006. In germ cells of mouse embryonic ovaries, the decision to enter meiosis precedes premeiotic DNA replication. *Nat. Genet.* 38:1430-4
11. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, et al. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327:836-40
12. Baudat F, de Massy B. 2007. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Res.* 15:565-77
13. Baudat F, Imai Y, de Massy B. 2013. Meiotic recombination in mammals: localization and regulation. *Nat. Rev. Genet.* 14:794-806
14. Bellvé AR, Cavicchia JC, Millette CF, O'Brien DA, Bhatnagar YM, Dym M. 1977. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J. Cell Biol.* 74:68-85
15. Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, et al. 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. U. S. A.* 99:8695-700

16. Bhérer C, Campbell CL, Auton A. 2017. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nat. Commun.* 8:14994
17. Billings T, Parvanov ED, Baker CL, Walker M, Paigen K, Petkov PM. 2013. DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biol.* 14:R35
18. Birtle Z, Ponting CP. 2006. Meisetz and the birth of the KRAB motif. *Bioinformatics* 22:2841-5
19. Bishop DK, Park D, Xu L, Kleckner N. 1992. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* 69:439-56
20. Borde V, Robine N, Lin W, Bonfils S, Géli V, Nicolas A. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* 28:99-111
21. Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431:96-9
22. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485:642-5
23. Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161:529-40
24. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* 63:861-9
25. Buard J, de Massy B. 2007. Playing hide and seek with mammalian meiotic crossover hotspots. *Trends Genet.* 23:301-9
26. Bullejos M, Koopman P. 2001. Spatially dynamic expression of Sry in mouse genital ridges. *Dev. Dyn.* 221:201-5
27. Bullejos M, Koopman P. 2005. Delayed Sry and Sox9 expression in developing mouse gonads underlies B6-Y(DOM) sex reversal. *Dev. Biol.* 278:473-81
28. Capilla L, Sánchez-Guillén RA, Farré M, Paytuví-Gallart A, Malinverni R, et al. 2016. Mammalian Comparative Genomics Reveals Genetic and Epigenetic Features Associated with Genome Reshuffling in Rodentia. *Genome Biol. Evol.* 8:3703-17
29. Carpenter AT. 1987. Gene conversion, recombination nodules, and the initiation of meiotic synapsis. *Bioessays* 6:232-6
30. Carpenter AT. 1994. Chiasma function. *Cell* 77:957-62
31. Carpenter ATC. 1975. Electron microscopy of meiosis in Drosophila melanogaster females: II: The recombination nodule—a recombination-associated structure at pachytene? *Proceedings of the National Academy of Sciences* 72:3186-9
32. Chen P-Y, Feng S, Joo JWJ, Jacobsen SE, Pellegrini M. 2011. A comparative analysis of DNA methylation across human embryonic stem cell lines. *Genome Biol.* 12:R62

33. Chen T, Li E. 2004. Structure and function of eukaryotic DNA methyltransferases. *Curr. Top. Dev. Biol.* 60:55-89
34. Ciruna B, Rossant J. 2001. FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev. Cell* 1:37-49
35. Clark AG, Wang X, Matise T. 2010. Contrasting methods of quantifying fine structure of human recombination. *Annu. Rev. Genomics Hum. Genet.* 11:45-64
36. Cole F, Baudat F, Grey C, Keeney S, de Massy B, Jasin M. 2014. Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. *Nat. Genet.* 46:1072-80
37. Consortium IH, Others. 2005. A haplotype map of the human genome. *Nature* 437:1299
38. Davies B, Hatton E, Altemose N, Hussin JG, Pratto F, et al. 2016. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* 530:171-6
39. de Boer E, Jasin M, Keeney S. 2015. Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hot spots in mice. *Genes Dev.* 29:1721-33
40. Deans C, Maggert KA. 2015. What do you mean, "epigenetic"? *Genetics* 199:887-96
41. Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, et al. 2012. Fast computation and applications of genome mappability. *PLoS One* 7:e30377
42. Dos Santos CO, Dolzhenko E, Hodges E, Smith AD, Hannon GJ. 2015. An epigenetic memory of pregnancy in the mouse mammary gland. *Cell Rep.* 11:1102-9
43. Eberl HC, Spruijt CG, Kelstrup CD, Vermeulen M, Mann M. 2013. A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Mol. Cell* 49:368-78
44. Eram MS, Bustos SP, Lima-Fernandes E, Siarheyeva A, Senisterra G, et al. 2014. Trimethylation of histone H3 lysine 36 by human methyltransferase PRDM9 protein. *J. Biol. Chem.* 289:12177-88
45. Escalante-Alcalde D, Merchant-Larios H. 1992. Somatic and germ cell interactions during histogenetic aggregation of mouse fetal testes. *Exp. Cell Res.* 198:150-8
46. Fraune J, Schramm S, Alsheimer M, Benavente R. 2012. The mammalian synaptonemal complex: protein components, assembly and role in meiotic recombination. *Exp. Cell Res.* 318:1340-6
47. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40:37-52
48. Gilbert SF. 2014. *Developmental Biology*. Sinauer Associates. 719 pp.
49. Govin J, Caron C, Lestrat C, Rousseaux S, Khochbin S. 2004. The role of histones in chromatin remodelling during mammalian spermiogenesis. *Eur. J. Biochem.* 271:3459-69
50. Gray S, Cohen PE. 2016. Control of Meiotic Crossovers: From Double-Strand Break Formation to Designation. *Annu. Rev. Genet.* 50:175-210

51. Greco CM, Kunderfranco P, Rubino M, Larcher V, Carullo P, et al. 2016. DNA hydroxymethylation controls cardiomyocyte gene expression in development and hypertrophy. *Nat. Commun.* 7:12418
52. Grey C, Barthès P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B. 2011. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.* 9:e1001176
53. Grey C, Clément JAJ, Buard J, Leblanc B, Gut I, et al. 2017. In vivo binding of PRDM9 reveals interactions with noncanonical genomic sites. *Genome Res.* 27:580-90
54. Griswold MD, Hogarth CA, Bowles J, Koopman P. 2012. Initiating meiosis: the case for retinoic acid. *Biol. Reprod.* 86:35
55. Gubbay J, Koopman P, Collignon J, Burgoyne P, Lovell-Badge R. 1990. Normal structure and expression of Zfy genes in XY female mice mutant in Tdy. *Development* 109:647-53
56. Günesdogan U, Magnúsdóttir E, Surani MA. 2014. Primordial germ cell specification: a context-dependent cellular differentiation event [corrected]. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369
57. Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, et al. 2012. Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science* 339:448-52
58. Halldorsson BV, Hardarson MT, Kehr B, Styrkarsdottir U, Gylfason A, et al. 2016. The rate of meiotic gene conversion varies by sex and age. *Nat. Genet.* 48:1377-84
59. Hammoud SS, Low DHP, Yi C, Carrell DT, Guccione E, Cairns BR. 2014. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* 15:239-53
60. Handel MA, Schimenti JC. 2010. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nat. Rev. Genet.* 11:124-36
61. Hassold T, Hall H, Hunt P. 2007. The origin of human aneuploidy: where we have been, where we are going. *Hum. Mol. Genet.* 16 Spec No. 2:R203-8
62. Hayashi K, Matsui Y. 2006. Meisetz, a novel histone tri-methyltransferase, regulates meiosis-specific epigenesis. *Cell Cycle* 5:615-20
63. Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* 438:374-8
64. Hillier SG, Smitz J, Eichenlaub-Ritter U. 2010. Folliculogenesis and oogenesis: from basic science to the clinic. *Mol. Hum. Reprod.* 16:617-20
65. Hofmann MC, Narisawa S, Hess RA, Millán JL. 1992. Immortalization of germ cells and somatic testicular cells using the SV40 large T antigen. *Exp. Cell Res.* 201:417-35
66. Hohenauer T, Moore AW. 2012. The Prdm family: expanding roles in stem cells and development. *Development* 139:2267-82
67. Imai Y, Baudat F, Taillepierre M, Stanzione M, Toth A, de Massy B. 2017. The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. *Chromosoma*
68. Jeltsch A. 2002. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem* 3:274-93

69. Jiang Z, Zhang X, Deka R, Jin L. 2005. Genome amplification of single sperm using multiple displacement amplification. *Nucleic Acids Res*. 33:e91
70. Johnston SE, Berenos C, Slate J, Pemberton JM. 2016. Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries). *Genetics* 203:583-98
71. Kafer GR, Li X, Horii T, Suetake I, Tajima S, et al. 2016. 5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability. *Cell Rep*. 14:1283-92
72. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, et al. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429:900-3
73. Kauppi L, Barchi M, Baudat F, Romanienko PJ, Keeney S, Jasin M. 2011. Distinct properties of the XY pseudoautosomal region crucial for male meiosis. *Science* 331:916-20
74. Kauppi L, Barchi M, Lange J, Baudat F, Jasin M, Keeney S. 2013. Numerical constraints and feedback control of double-strand breaks in mouse meiosis. *Genes Dev*. 27:873-86
75. Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JAF, Slanchev K, et al. 2007. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* 131:1273-86
76. Keeney S. 2001. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol*. 52:1-53
77. Keeney S, Baudat F, Angeles M, Zhou ZH, Copeland NG, et al. 1999. A mouse homolog of the Saccharomyces cerevisiae meiotic recombination DNA transesterase Spo11p. *Genomics* 61:170-82
78. Keeney S, Neale MJ. 2006. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem. Soc. Trans*. 34:523-5
79. Khil PP, Smagulova F, Brick KM, Camerini-Otero RD, Petukhova GV. 2012. Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Res*. 22:957-65
80. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14:R36
81. Kleckner N. 1996. Meiosis: how could it work? *Proceedings of the National Academy of Sciences* 93:8167-74
82. Koh-Stenta X, Poulsen A, Li R, Wee JLK, Kwek PZ, et al. 2017. Discovery and characterisation of the automethylation properties of PRDM9. *Biochem. J* 474:971-82
83. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467:1099-103
84. Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. 1991. Male development of chromosomally female mice transgenic for Sry. *Nature* 351:117-21

85. Koubova J, Menke DB, Zhou Q, Capel B, Griswold MD, Page DC. 2006. Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. U. S. A.* 103:2474-9
86. Krogh BO, Symington LS. 2004. Recombination proteins in yeast. *Annu. Rev. Genet.* 38:233-71
87. Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571-2
88. Lange J, Yamada S, Tischfield SE, Pan J, Kim S, et al. 2016. The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. *Cell* 167:695-708.e16
89. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-9
90. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25
91. Lei H, Oh SP, Okano M, Jüttermann R, Goss KA, et al. 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122:3195-205
92. Lenormand T, Dutheil J. 2005. Recombination difference between sexes: a role for haploid selection. *PLoS Biol.* 3:e63
93. Lesecque Y, Glémin S, Lartillot N, Mouchiroud D, Duret L. 2014. The red queen model of recombination hotspots evolution in the light of archaic and modern human genomes. *PLoS Genet.* 10:e1004790
94. Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915-26
95. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754-60
96. Li HH, Gyllensten UB, Cui XF, Saiki RK, Erlich HA, Arnheim N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 335:414-7
97. Lichten M. 2008. Meiotic Chromatin: The Substrate for Recombination Initiation. In *Recombination and Meiosis*:165-93: Springer, Berlin, Heidelberg. Number of 165-93 pp.
98. Lim FL, Soulez M, Koczan D, Thiesen HJ, Knight JC. 1998. A KRAB-related domain and a novel transcription repression domain in proteins encoded by SSX genes that are disrupted in human sarcomas. *Oncogene* 17:2013-8
99. Liu EY, Morgan AP, Chesler EJ, Wang W, Churchill GA, Pardo-Manuel de Villena F. 2014. High-resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male germline. *Genetics* 197:91-106
100. Liu H, Chang L-H, Sun Y, Lu X, Stubbs L. 2014. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol. Evol.* 6:510-25
101. Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S. 2007. A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* 27:561-71

102. Liu Y, Zhang X, Blumenthal RM, Cheng X. 2013. A common mode of recognition for methylated CpG. *Trends Biochem. Sci.* 38:177-83
103. Lundberg E, Fagerberg L, Klevebring D, Matic I, Geiger T, et al. 2010. Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* 6:450
104. Lupo A, Cesaro E, Montano G, Zurlo D, Izzo P, Costanzo P. 2013. KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Curr. Genomics* 14:268-78
105. Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, et al. 2015. Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. *PLoS Genet.* 11:e1005387
106. Mack KL, Nachman MW. 2017. Gene Regulation and Speciation. *Trends Genet.* 33:68-80
107. MacLennan M, Crichton JH, Playfoot CJ, Adams IR. 2015. Oocyte development, meiosis and aneuploidy. *Semin. Cell Dev. Biol.* 45:68-76
108. Marcon E, Moens P. 2003. MLH1p and MLH3p localize to precociously induced chiasmata of okadaic-acid-treated mouse spermatocytes. *Genetics* 165:2283-7
109. Margolin G, Khil PP, Kim J, Bellani MA, Camerini-Otero RD. 2014. Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genomics* 15:39
110. Matsui Y, Hayashi K. 2007. Epigenetic regulation for the induction of meiosis. *Cell. Mol. Life Sci.* 64:257-62
111. Matthews AGW, Kuo AJ, Ramón-Maiques S, Han S, Champagne KS, et al. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* 450:1106-10
112. McCoshen JA. 1982. In vivo sex differentiation of congeneic germinal cell aplastic gonads. *Am. J. Obstet. Gynecol.* 142:83-8
113. McLaren A. 2003. Primordial germ cells in the mouse. *Dev. Biol.* 262:1-15
114. Meistrich ML. 1977. Separation of spermatogenic cells and nuclei from rodent testes. *Methods Cell Biol.* 15:15-54
115. Meng X, Lindahl M, Hyvönen ME, Parvinen M, de Rooij DG, et al. 2000. Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* 287:1489-93
116. Merchant H. 1975. Rat gonadal and ovarioan organogenesis with and without germ cells. An ultrastructural study. *Dev. Biol.* 44:1-21
117. Mieczkowski PA, Dominska M, Buck MJ, Lieb JD, Petes TD. 2007. Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* 104:3955-60
118. Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323:373-5
119. Murakami H, Borde V, Shibata T, Lichten M, Ohta K. 2003. Correlation between premeiotic DNA replication and chromatin transition at yeast recombination initiation sites. *Nucleic Acids Res.* 31:4085-90

120. Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, et al. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* 327:876-9
121. Narasimhan VM, Hunt KA, Mason D, Baker CL, Karczewski KJ, et al. 2016. Health and population effects of rare gene knockouts in adult humans with related parents. *Science* 352:474-7
122. Neale MJ, Keeney S. 2006. Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* 442:153-8
123. Noma K, Allis CD, Grewal SI. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* 293:1150-5
124. Nowick K, Carneiro M, Faria R. 2013. A prominent role of KRAB-ZNF transcription factors in mammalian speciation? *Trends Genet*. 29:130-9
125. Oey H, Isbel L, Hickey P, Ebaid B, Whitelaw E. 2015. Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions. *Epigenetics Chromatin* 8:54
126. of Health NI, Others. 2001. Stem cells: scientific progress and future research directions (Department of Health and Human Services). *Stem Cell Information, Bethesda*
127. Ohinata Y, Ohta H, Shigeta M, Yamanaka K, Wakayama T, Saitou M. 2009. A signaling principle for the specification of the germ cell lineage in mice. *Cell* 137:571-84
128. Ohta K, Shibata T, Nicolas A. 1994. Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J*. 13:5754-63
129. Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247-57
130. Oliver TR, Feingold E, Yu K, Cheung V, Tinker S, et al. 2008. New insights into human nondisjunction of chromosome 21 in oocytes. *PLoS Genet*. 4:e1000033
131. Page SL, Hawley RS. 2004. The genetics and molecular biology of the synaptonemal complex. *Annu. Rev. Cell Dev. Biol*. 20:525-58
132. Paigen K, Kenneth P, Szatkiewicz JP, Kathryn S, Nicole L, et al. 2008. The Recombinational Anatomy of a Mouse Chromosome. *PLoS Genet*. 4:e1000119
133. Paigen K, Petkov P. 2010. Mammalian recombination hot spots: properties, control and evolution. *Nat. Rev. Genet*. 11:221-33
134. Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet*. 10:669-80
135. Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science* 327:835
136. Parvanov ED, Tian H, Billings T, Saxl RL, Spruce C, et al. 2017. PRDM9 interactions with other proteins provide a link between recombination hotspots and the chromosomal axis in meiosis. *Mol. Biol. Cell* 28:488-99
137. Patel A, Horton JR, Wilson GG, Zhang X, Cheng X. 2016. Structural basis for human PRDM9 action at recombination hot spots. *Genes Dev*. 30:257-65
138. Perry J, Palmer S, Gabriel A, Ashworth A. 2001. A short pseudoautosomal region in laboratory mice. *Genome Res*. 11:1826-32

139. Persikov AV, Osada R, Singh M. 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics* 25:22-9
140. Peters H. 1970. Migration of gonocytes into the mammalian gonad and their differentiation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 259:91-101
141. Petukhova GV, Pezza RJ, Vanevski F, Ploquin M, Masson J-Y, Camerini-Otero RD. 2005. The Hop2 and Mnd1 proteins act in concert with Rad51 and Dmc1 in meiotic recombination. *Nat. Struct. Mol. Biol.* 12:449-53
142. Petukhova GV, Romanienko PJ, Camerini-Otero RD. 2003. The Hop2 protein has a direct role in promoting interhomolog interactions during mouse meiosis. *Dev. Cell* 5:927-36
143. Portela A, Esteller M. 2010. Epigenetic modifications and human disease. *Nat. Biotechnol.* 28:1057-68
144. Powers NR, Parvanov ED, Baker CL, Walker M, Petkov PM, Paigen K. 2016. The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS Genet.* 12:e1006146
145. Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. 2014. DNA recombination. Recombination initiation maps of individual human genomes. *Science* 346:1256442
146. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406:593-9
147. Roeder GS. 1997. Meiotic chromosomes: it takes two to tango. *Genes Dev.* 11:2600-21
148. Romanienko PJ, Camerini-Otero RD. 2000. The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol. Cell* 6:975-87
149. Romanienko PJ, Camerini-Otero RD. 2000. The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol Cell* 6:975-87
150. Rossi P, Sette C, Dolci S, Geremia R. 2000. Role of c-kit in mammalian spermatogenesis. *J. Endocrinol. Invest.* 23:609-15
151. Rossitto M, Philibert P, Poulat F, Boizet-Bonhoure B. 2015. Molecular events and signalling pathways of male germ cell differentiation in mouse. *Semin. Cell Dev. Biol.* 45:84-93
152. Royo H, Seitz H, ElInati E, Peters AHFM, Stadler MB, Turner JMA. 2015. Silencing of X-Linked MicroRNAs by Meiotic Sex Chromosome Inactivation. *PLoS Genet.* 11:e1005461
153. Saitou M. 2009. Germ cell specification in mice. *Curr. Opin. Genet. Dev.* 19:386-95
154. Sanford JP, Clark HJ, Chapman VM, Rossant J. 1987. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes Dev.* 1:1039-46
155. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, et al. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* 419:407-11
156. Schmahl J, Eicher EM, Washburn LL, Capel B. 2000. Sry induces cell proliferation in the mouse gonad. *Development* 127:65-73
157. Schmekel K, Daneholt B. 1995. The central region of the synaptonemal complex revealed in three dimensions. *Trends Cell Biol.* 5:239-42

158. Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell Biol.* 6:73-7
159. Schwacha A, Kleckner N. 1995. Identification of double Holliday junctions as intermediates in meiotic recombination. *Cell* 83:783-91
160. Schwacha A, Kleckner N. 1997. Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. *Cell* 90:1123-35
161. Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, et al. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* 48:849-62
162. Shao Z, Zhang Y, Yuan G-C, Orkin SH, Waxman DJ. 2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol.* 13:R16
163. Shao Z, Zhen S, Yijing Z, Guo-Cheng Y, Orkin SH, Waxman DJ. 2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol.* 13:R16
164. Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, et al. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240-4
165. Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. 2016. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes Dev.* 30:266-80
166. Smagulova F, Brick K, Pu Y, Sengupta U, Camerini-Otero RD, Petukhova GV. 2013. Suppression of genetic recombination in the pseudoautosomal region and at subtelomeres in mice with a hypomorphic Spo11 allele. *BMC Genomics* 14:493
167. Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. 2011. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* 472:375-8
168. Sollier J, Lin W, Soustelle C, Suhre K, Nicolas A, et al. 2004. Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. *EMBO J.* 23:1957-67
169. Soriano P, Keitges EA, Schorderet DF, Harbers K, Gartler SM, Jaenisch R. 1987. High rate of recombination and double crossovers in the mouse pseudoautosomal region during male meiosis. *Proc. Natl. Acad. Sci. U. S. A.* 84:7218-20
170. Spandidos A, Wang X, Wang H, Seed B. 2010. PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Res.* 38:D792-9
171. Steinmetz M, Minard K, Horvath S, McNicholas J, Srelinger J, et al. 1982. A molecular map of the immune response region from the major histocompatibility complex of the mouse. *Nature* 300:35-42
172. Suzuki A, Saga Y. 2008. Nanos2 suppresses meiosis and promotes male germ cell differentiation. *Genes Dev.* 22:430-5
173. Tease C, Hultén MA. 2004. Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. *Cytogenet. Genome Res.* 107:208-15

174. Tortereau F, Servin B, Frantz L, Megens H-J, Milan D, et al. 2012. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* 13:586
175. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31:46-53
176. Tsuda M, Sasaoka Y, Kiso M, Abe K, Haraguchi S, et al. 2003. Conserved role of nanos proteins in germ cell development. *Science* 301:1239-41
177. Uysal F, Akkoyunlu G, Ozturk S. 2015. Dynamic expression of DNA methyltransferases (DNMTs) in oocytes and early embryos. *Biochimie* 116:103-13
178. Uysal F, Akkoyunlu G, Ozturk S. 2016. DNA methyltransferases exhibit dynamic expression during spermatogenesis. *Reprod. Biomed. Online* 33:690-702
179. Vidal VP, Chaboissier MC, de Rooij DG, Schedl A. 2001. Sox9 induces testis development in XX transgenic mice. *Nat. Genet.* 28:216-7
180. Vlachogiannis G, Niederhuth CE, Tuna S, Stathopoulou A, Viiri K, et al. 2015. The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis. *Cell Rep.* 11:990
181. von Wettstein D. 1984. The Synaptonemal Complex in Genetic Segregation. *Annu. Rev. Genet.* 18:331-413
182. Walker M, Billings T, Baker CL, Powers N, Tian H, et al. 2015. Affinity-seq detects genome-wide PRDM9 binding sites and reveals the impact of prior chromatin modifications on mammalian recombination hotspot usage. *Epigenetics Chromatin* 8:31
183. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, et al. 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 22:1680-8
184. Wilhelm D, Martinson F, Bradford S, Wilson MJ, Combes AN, et al. 2005. Sertoli cell differentiation is induced both cell-autonomously and through prostaglandin signaling during mammalian sex determination. *Dev. Biol.* 287:111-24
185. Wilkins JF. 2005. Genomic imprinting and methylation: epigenetic canalization and conflict. *Trends Genet.* 21:356-65
186. Wolfe SA, Nekludova L, Pabo CO. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29:183-212
187. Wu H, Mathioudakis N, Diagouraga B, Dong A, Dombrovski L, et al. 2013. Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. *Cell Rep.* 5:13-20
188. Wu TC, Lichten M. 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* 263:515-8
189. Yamada T, Mizuno K-I, Hirota K, Kon N, Wahls WP, et al. 2004. Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. *EMBO J.* 23:1792-803
190. Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, et al. 2008. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat. Genet.* 40:1016-22

191. Yamashita K, Shinohara M, Shinohara A. 2004. Rad6-Bre1-mediated histone H2B ubiquitylation modulates the formation of double-strand breaks during meiosis. *Proc. Natl. Acad. Sci. U. S. A.* 101:11380-5
192. Yoshida S, Sukeno M, Nabeshima Y-I. 2007. A vasculature-associated niche for undifferentiated spermatogonia in the mouse testis. *Science* 317:1722-6
193. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. 1992. Whole genome amplification from a single cell: implications for genetic analysis. *Proc. Natl. Acad. Sci. U. S. A.* 89:5847-51
194. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137