Spring 5-20-2021

HELLS and PRDM9 Form a Pioneer Complex to Open Chromatin at Meiotic Recombination Hotspots: A Technological Perspective

Catrina Spruce
catrina.spruce@maine.edu

HELLS AND PRDM9 FORM A PIONEER COMPLEX TO OPEN CHROMATIN

AT MEIOTIC RECOMBINATION HOTSPOTS:

A TECHNOLOGICAL PERSPECTIVE

By

Catrina Spruce

B.A. Dartmouth College, 2006

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Biochemistry)

The Graduate School

The University of Maine

May 2021

Advisory Committee:

Christopher Baker, Assistant Professor at The Jackson Laboratory, Advisor

Mary Ann Handel, Professor at The Jackson Laboratory

Laura Reinholdt, Professor at The Jackson Laboratory

HELLS AND PRDM9 FORM A PIONEER COMPLEX TO OPEN CHROMATIN

AT MEIOTIC RECOMBINATION HOTSPOTS:

A TECHNOLOGICAL PERSPECTIN

By Catrina Spruce

Thesis Advisor: Dr. Christopher Baker

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Biochemistry)
May 2021

Chromatin functions as a physical barrier regulating access to DNA, however pioneer factors are able to engage partial recognition motifs present within nucleosomal DNA to initiate access to specific DNA sequences. During spermatogenesis, genomic locations that become recombination hotspots are generally in regions of closed chromatin, or heterochromatin, before meiosis. However, in leptotene and zygotene stages of meiosis, PRDM9 marks nucleosomes at recombination hotspots with H3K4me3 and H3K36me3 and recruits other factors that deposit additional histone marks. Here I focus on the technological approaches by which we discovered that hotspots also transition from closed to open chromatin, dependent on the chromatin remodeler HELLS. HELLS and PRDM9 form a pioneer complex that targets hotspot sequences and provides access for the DNA double-strand break (DSB) machinery. Failure to reorganize chromatin at hotspots into an open conformation leads to DSBs at other open regions, such as promotors, causing a breakdown in recombination and cell death.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

STATEMENT OF PURPOSE

1.1. Description of thesis goal

The purpose of this document is to describe the process of discovery that took place over three years and culminated in a paper co-authored by myself and other lab members, HELLS *and* PRDM9 *form a pioneer complex to open chromatin at meiotic recombination hot spots*, Genes Dev, 2020. Though I will cover the results of our experiments and our findings, the more structured and polished scientific significance of our work is in the published paper. My thesis will focus on the technologies that I used and follow the chronological course of my work, illustrating the successes and enigmas we encountered that inspired our work, and how we built on these to elucidate the novel role for *Hells* in meiosis and fertility. In this thesis, I will use first person singular to refer to work that I conducted, and first person plural to refer to work done in collaboration with other lab members, particularly my advisor, Dr. Baker.

CHAPTER 2

INTRODUCTION

2.1. Chromatin is remodeled to permit homologous recombination

Homologous recombination is a defining event in meiosis when programmed double-stranded DNA breaks (DSBs) initiate a homology search resulting in homologous chromosome pairing and the resolution of the DSBs through exchange of genetic material. These programmed DSBs occur at hotspots, which are specialized sites that can be recognized by PRDM9 (Paigen and Petkov, 2010; Baudat et al., 2013). PRDM9 is a meiotic protein that is expressed in leptotene to zygotene stage of prophase I. The PRDM9 protein has several conserved domains; including a PR/SET domain for methylation of histones (Hohenauer and Moore, 2012, Mzoughi et al., 2016), KRAB-related (Krüppel-associated box) and an SSXRD (SSX repression domain) domain for protein interactions (Imai et al., 2017, Parvanov et al., 2017), and a DNA-binding domain composed of a tandem array of C2H2 zinc fingers (Hayashi et al., 2005). Once PRDM9 binds at a hotspot, the SET domain allows PRDM9 to trimethylate both histone H3 trimethylation on Lysine 3 (H3K4me3) and histone H3 trimethylation on Lysine 26 (H3K36me3) (Buard et al., 2009; Sun et al., 2015; Powers et al., 2016), a combination of marks that is unique to hotspots. These epigenetic modifications at hotspots by PRDM9 subsequently may play a part in allowing proteins like SPO11 to bind the hotspot and create meiotic DSBs necessary for synapsis (Romanienko and Camerini-Otero, 2000; Diagouraga et al., 2018).

2.2. Homologous recombination is necessary for the progression of meiosis

Progression of meiosis starts in leptotene stage when *Prdm9* is expressed and activates hotspots. Next is zygotene stage where DSBs occur and homology search is initiated. Then in pachytene stage synapsis is completed between homologous chromosomes and DSBs are resolved by homologous recombination. Progression from zygotene stage into pachytene stage is particularly sensitive to completion of synapsis

and repair of DSBs. Meiotic recombination plays essential roles in promoting the recognition, pairing, and accurate segregation of homologous chromosomes (Hunter, 2015). In the absence of PRDM9, DSBs are relocated to other sites marked by H3K4me3, "default sites" of PRDM9-independent H3K4me3, including a large proportion of transcription promoters and enhancers (Brick et al., 2012). For unknown reasons these DSBs are not efficiently repaired, leading to incomplete synapsis and meiotic arrest at pachytene stage resulting in cell apoptosis (Hayashi et al., 2005).

2.3. Evidence suggests hotspots are sites of open chromatin

The question of whether open chromatin plays a role in hotspot activation and recombination began with Dr. Christopher Baker's observation (Baker et al., 2014) of multiple histone marks associated with open chromatin at meiotic hotspots and evidence of nucleosome reorganization at active hotspots. Some of these histone marks are deposited by PRDM9 when it binds to hotspots, like H3K4me3 and H3K36me3. Additional active marks including H3K9ac and H3K4me1 are deposited by yet unknown factors (Spruce et al, 2020). The presence of activating histone marks led us the question whether nucleosome remodeling is also occurring to open chromatin at hotspots. Additional strong evidence of induced MNase sensitivity at hotspots upon presence of PRDM9 (Baker et al, 2014), and DNase hypersensitivity at a single specifically queried hotspot site (Shenkar et al., 1991) suggest DNA is made accessible at activated hotspots.

2.4. Description of HELLS protein and expression

HELLS, encoded by the *Hells* gene, (and also known as Lsh, PASG, or SMARCA6) is a chromatin remodeling protein in the SNF2 family (Jarvis, Geiman et al., 1996). SNF2 proteins are part of the SWI/SNF complex that remodels nucleosomes in an ATP-dependent manner to increase accessibility of DNA (Becker and Horz 2002). The HELLS protein contains seven conserved helicase domains (Jarvis, Geiman et al., 1996) and an ATPase-domain (Ren, Briones et al., 2015, Jenness, Giunta et al., 2018). *Hells* CKO mice are embryonic lethal, however not all SNF2 homologues share this phenotype (Essers,

Hendriks et al., 1997, Reyes, Barra et al., 1998, Wesoly, Agarwal et al., 2006), indicating a specific role

for *Hells* in early development. *Hells* is expressed in undifferentiated ES cells, however it is significantly

downregulated in differentiating cells (Xi, Geiman et al., 2009). *Hells* expression was originally identified

in proliferating lymphocytes (Lsh – lymphoid-specific helicase) undergoing VDJ recombination (Jarvis,

Geiman et al., 1996), and is correlated with poor T lymphocyte proliferation (Geiman and Muegge 2000).

Poor proliferation is also a phenotype in neural stem/progenitor cells, which show a 60% decrease in

replication efficiency in *Hells* CKO (Han, Ren et al., 2017), and oogenesis and spermatogenesis, which are

completely arrested in *Hells* CKO (De La Fuente, Baumann et al., 2006, Zeng, Baumann et al., 2011).

2.5. HELLS protein functional activities

Previous studies have shown different roles for HELLS depending on its binding partner. For instance,

HELLS can remodel nucleosomes in complex with a DNA binding partner CDCA7 in zebrafish (Jenness,

Giunta et al., 2018), but cannot load onto chromatin or remodel nucleosomes in absence of CDCA7.

HELLS interacts with yH2AX at DNA double-strand breaks and is required to phosphorylate yH2AX which

is critical for efficient repair of DNA damage after ionizing radiation (Burrage, Termanis et al., 2012).

HELLS participates in classical non-homologous end joining in HEK293 cells with DNA-binding protein

CDCA7 (Unoki, Funabiki et al., 2018). HELLS is required for proper nucleosome density and de novo DNA

methylation by remodeling nucleosomes through its ATPase-domain (Ren, Briones et al., 2015, Termanis

et al., 2012, Torrea et al., 2016). DNMT3A/B *de novo* DNA methylation is dependent on HELLS

interaction with DNMT3A/B, resulting in a 30-50% reduction in DNA methylation in *Hells* CKO mice,

primarily at repeat regions (Dennis, Fan et al., 2001). Reduced methylation in these normally

heterochromatic regions leads to deregulated expression of certain transposable elements and

disrupted lamin B attachment (LADs) (Yu, McIntosh et al., 2014). These data suggest that the presence

of HELLS is necessary to allow certain cell processes that are dependent on chromatin compaction.

2.6. *Hells* phenotypes

In female meiosis, *Hells* CKO oocytes fail to complete meiosis and arrest at pachytene stage due to incomplete synapsis (De La Fuente, Baumann et al., 2006). The same phenotype was shown in male germ cells during spermatogenesis where deletion of *Hells* leads to meiotic arrest at pachytene stage and infertility (Zeng, Baumann et al., 2011). In embryonic cells, *Hells* expression is necessary for the establishment of *de novo* DNMT3A/B -dependent DNA methylation to silence retroviral genes and methylate endogenous genes (Zhu, Geiman et al., 2006). Altered chromatin structure due to global demethylation in *Hells* CKO fibroblasts causes reduced proliferation through failed chromosome segregation during mitosis (Fan, Yan et al., 2003, Sun and Arceci 2005). These phenotypes all show a role for *Hells* in creating the necessary epigenetic environment for cell-specific functions. Given the fertility phenotypes associated with *Hells* and our prior knowledge that epigenetic landscape at hotspots is critical for hotspot activity, we set out to investigate the potential role of DNA accessibility at hotspots and the potential role of the nucleosome remodeling protein *Hells*.

CHAPTER 3

ATAC-SEQ ASSAY FOR OPEN CHROMATIN

3.1. ATAC-Seq is a new assay to interrogate regions of open chromatin

Increased DNA accessibility can lead to increased protein binding and many potential biological activities and is often associated with actively transcribed DNA. Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq) uses direct transposition and ligation of DNA adapters compatible with high-throughput sequencing into native chromatin through a modified transposase that can access and cut open chromatin (Buenrostro et al., 2013). The DNA adaptors allow PCR amplification to create a library of DNA fragments compatible with sequencing on the Illumina platform. This greatly increases efficiency and minimizing processing steps that lead to loss of product as compared to traditional sequencing library preparation.

3.2. Determine optimal sample prep for ATAC-Seq

We optimized ATAC-Seq in our hands by performing preliminary testing on mouse embryonic stem cells (mESCs) because they are easy to grow, tolerant to cryopreservation, and are the focus of another on-going study in our lab. The published ATAC protocol was performed on extracted nuclei, so our plan was to optimize a nuclear extraction protocol for our cells. However, the three methods we tried: a three-density sucrose gradient layered column, a single sucrose gradient, and a commercially available iodixanol gradient solution, all had very poor yields of purified nuclei and we realized it would be difficult to impossible to collect enough spermatocyte cells from a fresh prep of 12-12dpp male testis to get enough purified nuclei for the ATAC protocol. Therefore, we additionally performed the ATAC protocol on intact fresh mESCs, avoiding nuclear purification altogether, and instead briefly lysing the cells before adding Tn5. We also performed ATAC on cryopreserved mESCs (80% media, 10% FBS, 10% DMSO, cooled at -1°/min to -80°), realizing there would be a huge advantage to having all samples

6

collected and frozen before performed ATAC in one batch, instead of performed ATAC one by one on each fresh sample. In the end, we tested three cell conditions: fresh intact cells, purified nuclei, and cryopreserved cells.

3.3. Description of ATAC Tn5 mode of action

Tn5 is a hyper-active transposase loaded with forward or reverse primers. The ATAC protocol is straightforward: Tn5 transposase is added to the cells where it diffuses into the nucleus. Heterochromatinized regions of DNA are resistant to Tn5, but relaxed regions of chromatin like those at regulatory regions are vulnerable to Tn5 transposase activity. Once in contact with exposed, nucleosome-free DNA, the Tn5 makes a single-stranded cut and ligates an adaptor on the cut strand (Fig. 1A). The Tn5 must cut in two locations to liberate a DNA fragment from a nucleosome-free region. Nucleosomes are an octamer of histone proteins wrapped with ~140bp of DNA, with ~60bp of linker DNA between nucleosomes, so the DNA fragments liberated by Tn5 represent intranucleosomal, di-, and tri-nucleosome DNA lengths because the Tn5 can only access the DNA where it is NOT bound by nucleosomes. Additionally, Tn5 ligates adaptors onto the ends of DNA that are subsequently used to amplify the DNA fragment during PCR amplification.

3.4. Description of library amplification of ATAC samples

We started by using mouse embryonic stem cells (ESCs) for our initial ATAC trouble-shooting, because they are readily available and easily scalable. I used the original Greenleaf protocol (Wu et al., 2013) and the Tagmentation enzyme kit from Illumina (cat #20034198). The amount of DNA obtained after Tn5 cuts at open chromatin regions is quite small and needs to undergo PCR amplification before there is enough DNA product for sequencing. For sequencing library preparation, I used primers designed by the Greenleaf lab that include an Illumina adaptor sequence, a barcode, an Illumina primer, and a Tn5 adaptor sequence (Fig. 1B). Each part of the library primer performs a function to allow the DNA fragment to be sequenced. The Illumina adaptor sequence is complementary to the sequence on the

Illumina sequencing flowcell, it attaches the fragment of DNA to the flowcell and allows it to be amplified to make clusters. The barcode is a unique 8 base pair DNA sequence different between each primer. When individual ATAC samples are amplified with different barcodes for each ATAC-Seq library, the resulting libraries are pooled and the unique barcode sequence is used to demultiplex sequenced samples that are sequenced together on the same flowcell. The Illumina primer is complementary to the primers used in the NextSeq to initiate the sequencing reaction. Finally, the Tn5 adaptor sequence is complementary to the adaptor ligated onto the DNA by Tn5 and serves as the binding sequence to amplify the DNA fragments cut by Tn5. I used a PCR amplification program with 98°C denaturing step for 10 sec, 60°C primer annealing step for 30sec, and 72°C polymerase extension step for 1 min and repeated these 3 steps for 8-12 cycles. PCR amplification both increases the amount of starting DNA and adds the needed adaptors and barcodes to the DNA fragments.

3.4.1 Library PCR amplification considerations

The final library is sensitive to both over- and under-amplification by PCR. Too many cycles of PCR amplification can randomly over-amplify some sequences over others, undermining library diversity and resulting in duplicate fragments getting sequenced. On the other hand, too few cycles of amplification will not yield enough product for downstream sequencing. The Greenleaf protocol recommends performing quantitative PCR on a fraction of the library to determine the optimal number of amplification cycles. After 4 cycles of PCR amplification, I removed the reaction and removed a 5 µl aliquot from the total 50 µl reaction of amplified library to use in a separate qPCR reaction to determine the additional number of cycles needed. I used the 5 µl of pre-amplified library in a 15 µl qPCR reaction with SYBR green, and cycled 20X using the same amplification program. To maximize capture of the ATAC DNA fragments and avoid over-amplification of PCR duplicates, the optimal amount of library amplification is to stop the PCR at the transition to exponential amplification, as determined by the qPCR curves. For the most part, our samples required ~9-10 total cycles of library amplification (Fig. 1C),

which is exactly in the 8-12 cycle range recommended in the protocol. Samples requiring more cycles than this would indicate low starting levels of DNA and increased PCR duplication rate, both of which negatively affect the quality of sequencing results.

3.4.2. ATAC library purification

After completing library amplification, I purified the libraries using AMPure beads (Beckman Coulter #A63880). These are magnetic beads coated with carboxyl molecules that reversibly bind DNA in the presence of the "crowding agent" polyethylene glycol (PEG) and salt (20% PEG, 2.5M NaCl is optimal). PEG causes the negatively-charged DNA to bind with the carboxyl groups on the bead surface. The amount of DNA bound is dependent on the concentration of PEG and salt in the reaction, where adding more beads allows more of the smaller fragments to be bound and less beads captures only larger fragments of DNA. We experimented with using a two-step AMPure purification to exclude large fragments and small fragments. Large DNA fragments (>600bp) do not form clusters on the sequencing flowcell and therefore are not useful. Small fragments (<150bp) are mostly library primer dimers, and do not have any meaningful sample DNA sequence. To exclude these, we added a 1:1 ratio of AMPure beads to ATAC DNA library and then discarded the beads bound to large fragments. Then we added a 1.7:1 ratio of beads to library and kept the beads while discarding the supernatant containing small fragments. After purification, BioAnalyzer was used to analyze library size and quantity.

3.4.3. ATAC library QC

The BioAnalyzer is a chip-based capillary electrophoresis machine to analyze RNA, DNA, and protein. We use Bioanalyzer to look at fragment size and quantity in our sequencing libraries. The BioAnalyzer results also showed that our ATAC libraries had distinct banding patterns indicative of nucleosome organization (Fig. 1D). We see in our BioAnalyzer results a population of smaller-sized DNA fragments that contain intranucleosomal fragments, then another population of fragments ~140bp larger, and another population ~140bp after that. These successive peaks represent DNA fragments that

9

were cut on either side of the nucleosome they occupied, liberating either mono-, di, or tri-nucleosome fragments. The presence of this banding pattern also indicates that the chromatin was still intact when we performed ATAC, a good indicator of quality. We found through experimentation that the dual size selection was not worthwhile because we lost product during the two-step process and the large fragments do not seem to interfere with the quality or quantity of our sequencing. We kept the 1.7:1 AMPure purification, and although there is still a population of primer dimer that is not removed, we did not see evidence of these short sequences in our data. We did not want to increase the AMPure bead ratio to eliminate the primers and risk losing smaller fragments that contain valuable sample DNA sequence.
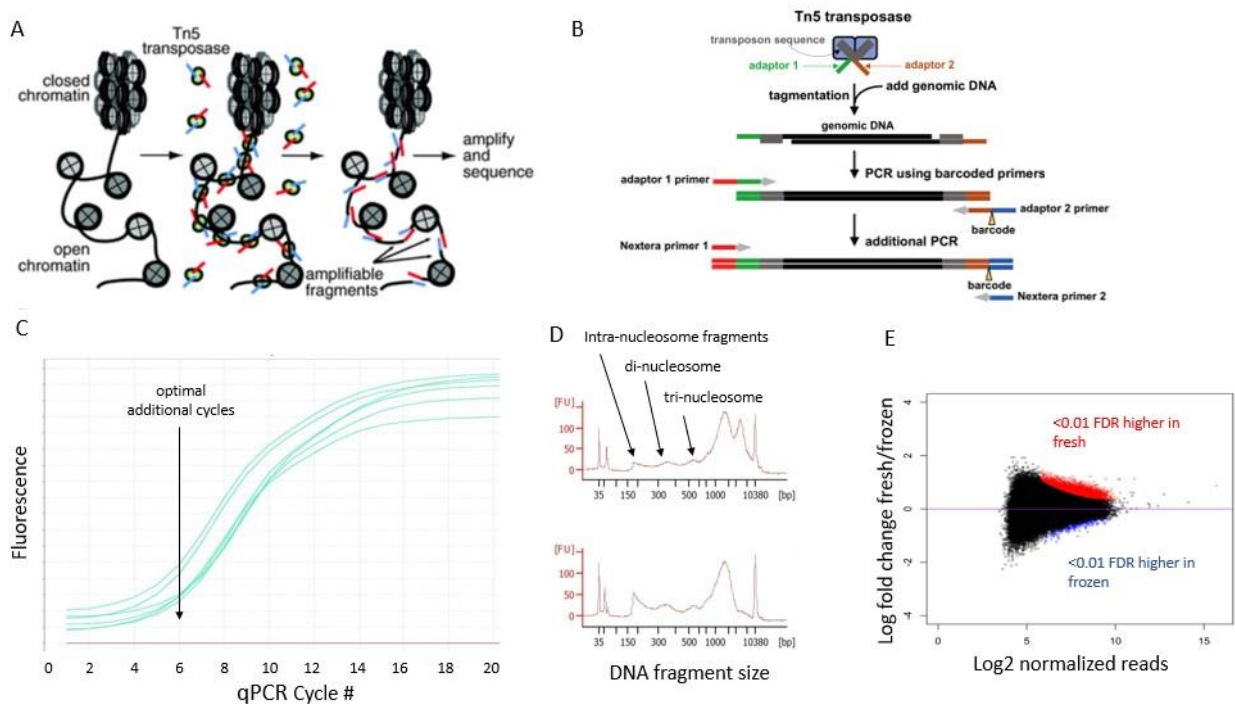
Figure 1. Description of ATAC assay. (A) Tn5 transposes cuts and ligates an adaptor on accessible DNA (Buenrostro et al., 2013). (B) Library amplification adds barcodes and adaptors for sequencing (Shashikant and Ettensohn, 2019). (C) Optimal PCR cycles for ATAC library are calculated from start of qPCR exponential amplification. (D) BioAnalyzer traces show nucleosome banding in quality ATAC libraries. (E) Comparison of ATAC peaks from fresh vs. frozen cells.

3.5. Sequence analysis of ATAC samples

We sequenced the ATAC libraries from 8 experiments to determine the quality of the experiments and enrichment for open chromatin regions. Initially we sequenced to a depth of 80-100 mln reads per sample, paired-end reads, 75bp each. The resulting sequence files (i.e. fastq files) I ran through FastQC to look for red flags in sequencing quality, over-represented sequences, and duplication rate. Overall, the sample sequences passed quality criteria. There were some flags for adapter content at the 3' end of reads and for duplication rate that stood out. Other than that, there were flags for per base sequence content at the 5' end of reads, likely part of the library amplification process and Kmer content, which commonly flags samples that are fine and is not part of the FastQC program in subsequent versions. I used the Burrows-Wheeler Alignment tool (BWA) (Li and Durbin; 2009) to efficiently align short sequencing reads against a large reference sequence, allowing mismatches and gaps. For my data, I used the 'bwa mem' command to map the 75 bp ATAC-Seq reads to the B6 mouse genome (mm10). Next I sorted the mapped reads into chromosomal order and convert the file into a compressed bam file using 'samtools sort' and samtools view' (Li et al; 2009). I checked the number of reads mapping to each chromosome and proportion of reads mapping to mitochondrial sequences using 'samtools idxstat.' To check library complexity and estimate the number of unique reads I used Picard tools 'MarkDuplicates' to identify duplicate reads. These are reads from separate clusters on the flowcell but represent the same starting DNA fragment that was amplified by PCR to make many copies. I removed all but one of each duplicated fragment, because the copies represent the same biological instance and are not new occurrences. We next used a Greenleaf lab script to shift our single-end mapped reads. If the read is on the positive strand (as determined by the sam flag) it will add 4bp to the start and subtract 5bp from the partner start. The reads are shifted because the Tn5 enzyme has a 9bp binding site, so the DNA must have been open 4.5bp on either side of the integration, therefore by shifting the reads we are more

accurately placing the open chromatin. This same script removes low quality reads and mitochondrial reads. The output of these metrics is in Table 1.

3.6. Identification of unmapped reads

Immediately we noticed a depressed rate of mapped reads (Table 1). Looking closer at the sequence identity of the unmapped reads we found sequence fragments originating from the Tn5 adaptors and Illumina primers and adaptors. This can be explained if we imagine small fragments of sample DNA being sequenced through their entire length and into the primer and adaptors sequences on the 3' side. The presence of the engineered sequence prevents alignment to the mouse genome which lacks this sequence, even though part of each sequence contains mouse DNA. This issue we resolved by using Trimmomatic ILLUMINACLIP:/opt/bin/Trimmomatic-0.39/adapters/NexteraPE-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 and specifying Nextera PE adaptors to remove the Illumina sequences and everything after, which includes the Tn5 adaptor sequence. Once the reads were trimmed, our alignment rate went back up to 97-100% (Table 1).

3.7. Call ATAC open chromatin peaks using MACS

To identify regions of open chromatin, pile-up of unique sequencing reads after alignment indicating open chromatin (i.e., peaks), was performed using MACS (Zheng et al., 2008) with the command 'macs14 -t <ATAC.bam> -f BAM -g mm -p 1e-5.' MACS identifies peaks based on significant accumulation of reads over the local background. I set the significance level at 1e-5 to filter for peaks that are clearly and strongly enriched over the background, excluding less significant peaks that could be due to noise or are barely measurable and therefore are unlikely to be biologically significant. At this threshold I recovered ~80,000-170,000 peaks per sample (Table 1). I calculated the proportion of reads that were counted in a peak compared to the total number of reads. This measurement of the fraction of reads in peaks (FRiP) usually serves as a quality test for the specificity of the assay, where more reads in peaks is

a more specific assay, vs high background and low FRiP. The ATAC-Seq I performed had the highest FRiP

and therefore sensitivity in the fresh mESC samples.

3.8. Trouble-shooting ATAC-seq protocol to reduce MT reads

One of the well-known issues with sequencing ATAC-Seq libraries is contamination with mitochondrial

DNA reads (Corces et al., 2016). In our samples, almost 40-50% of reads were mitochondrial (Table 1).

Since only chromosomal reads are informative, this represents a significant loss of sequencing capacity

and additional cost. One way to reduce mitochondrial read in ATAC-Seq data is to extract nuclei and

purify them away from the mitochondria. We had success in reducing the mitochondria by purifying

nuclei, but the cost was a huge reduction in yield, where we lost 75-95% of the nuclei during the

purification process. In order to reduce the number of reads lost to mitochondrial DNA, I tested a newly

published ATAC method called FastATAC, which was designed to reduce mitochondrial contribution. The

FastATAC protocol skips cell lysis, and instead permeabilizes the cells with digitonin, a detergent that

solubilizes lipids like those in cell membranes, making them more permeable. This allows the Tn5 to

access the chromosomal DNA while minimizing the lysis of the membrane-bound mitochondria. Using

the same number of ES cells but following the FastATAC protocol, we were able to reduce the

mitochondrial percentage from 40-50% to 5-10%. This resulted in an increase in the number of usable

chromosomal reads and allowed us to recover equal numbers of ATAC peaks from a sequencing depth

of 50-60 mln reads per sample compared to 100 mln reads per sample with the previous protocol

greatly reducing sequencing costs.

3.9. Nuclear extraction or fresh cells are not necessary for successful ATAC-Seq

Interestingly, the ATAC-Seq worked as well or better in intact cells vs nuclei and frozen vs fresh cells.

Although nuclei had fewer mitochondrial reads, the number of peaks called was reduced and the

fraction of reads in peaks was significantly lower than either fresh or frozen intact cells (Table 1). This

suggested that the process of purifying nuclei has an adverse effect on the integrity of the chromatin.

Surprisingly, freezing cells and then thawing them for ATAC did not seem to impact assay quality in

terms of # peaks and FRiP (Table 1). To confirm that the peaks I was collecting from the frozen cells were

the same locations and relative intensities, I used the DiffBind analysis package(Ross-Innes et al., 2012)

which includes functions to overlap and merge peaksets, count sequencing reads overlapping intervals

in peaksets and identify statistically significantly differentially bound sites based on evidence of binding

affinity (measured by differences in read densities). I used the 'dba.peakset' command to collate a

consensus peakset for the fresh and cryopreserved frozen samples, and then 'dba.count' compare

normalized read counts falling within those peaks. The resulting MA plot shows fresh and frozen

samples have similar read counts in each peak, shown in black, and only ~7% of peaks are significantly

different at FDR <0.01, shown in red and blue (Fig. 1E). Although the similar results for fresh and frozen

mES cells may be cell type-specific, it is reassuring to think that samples could be collected for a project

over time, and then then assayed for ATAC all at once to minimize effort and batch effects.

ATAC Prep	Raw Reads	% Mapped	% Duplication	% MT	Trimmed Reads	% Mapped	# Peaks	FRiP
fresh	61,397,131	26.0%	56.3%	31.8%	59,826,503	99.4%	122,879	26.0%
fresh	105,814,739	23.9%	61.6%	28.6%	102,370,688	99.7%	142,748	26.0%
nuclei	99,880,891	67.8%	8.9%	2.8%	96,456,979	98.6%	109,176	13.0%
nuclei	33,145,218	68.4%	6.9%	3.8%	31,959,494	98.6%	82,548	14.0%
frozen	90,958,646	35.4%	43.1%	27.1%	88,586,844	99.8%	143,866	23.0%
frozen	115,516,238	38.5%	39.0%	24.3%	111,705,333	99.8%	171,372	24.0%

Table 1. ATAC sequence analysis comparing fresh, frozen, or extracted nuclei conditions.

OPEN CHROMATIN AT HOTSPOTS IS DEPENDENT ON PRDM9 AND INDEPENDENT OF

DOUBLE -STRANDED BREAKS

4.1. ATAC-Seq on spermatocytes to look for open chromatin at hotspots

Based on the presence of histone marks that suggest open chromatin (H3K4me3, H3K36me3) and increased MNase sensitivity at hotspots, we tested whether activated hotspots are sites of open chromatin using ATAC-Seq. To investigate hotspot dynamics during male meiosis, we need to capture leptotene/zygotene (L/Z) stage spermatocytes when PRDM9 is active. Previous research has shown that the first wave of spermatogenesis in male mice is relatively synchronous and passes through L/Z stage at ~12-14 days post-partem (dpp) (Ball et al., 2016). At this time point, the mouse testes are enriched for L/Z stage spermatocytes, increasing the chances of capturing activated hotspots and minimizing background from non-meiotic cells.

4.2. Sample choice for spermatocyte ATAC-Seq

Before performing ATAC-Seq on 12dpp spermatocytes, we needed appropriate controls in order to decisively conclude the signal we might find at hotspots was truly meiosis- or PRDM9-specific. We can take advantage of a genetically-modified knock-in mouse with a different allele of PRDM9. The knock-in mouse, B6-PRDM9$^{CAST-KI}$/Kpgn (B6$^{CAST-KI}$, Baker et al., 2014), has the same genetic background as C57BL/6J with the exception of the PRDM9 zinc-finger domain, which is from the CAST/EiJ *Prdm9* allele. The zinc-finger domain for PRDM9^{Cst} has both different numbers of and different identity of zinc-fingers resulting in a protein variant that recognizes and binds different DNA sequences. Therefore, we expect that meiosis-specific open chromatin at hotspots will differentially appear at allele-specific sites for B6 PRDM9^{Dom2} and the CAST PRDM9^{Cst}, while all other open chromatin sites should stay the same. We performed ATAC-Seq on two replicates each B6$^{CAST-KI}$ and B6 12dpp spermatocytes.

4.3. Spermatocyte isolation and transposition

To extract spermatocytes from the 12dpp testes, I used the established protocol for isolating

spermatocytes for chromatin assays (Baker et al., 2014) and applied it to the ATAC-Seq protocol. Briefly,

I collected testes from 12dpp male mice into HEPES-buffered DMEM and disaggregated them using a

collagenase-based dissociation with Liberase TM (Sigma, cat# 5401119001). After two 15min

incubations at 37° with Liberase TM, the tubules become fragmented by enzymatic digestion, and I

further reduced them to single cell suspension by pipetting them vigorously 50X with a 10mL serological

pipet. After filtering for single cells using a 40um filter and washing with PBS, I put the spermatocytes

straight into the FastATAC protocol, using 100,000 spermatocyte cells per reaction.

4.4. ATAC-Seq data from spermatocytes shows good nucleosome banding and enrichment at TSS's

After sequencing and processing the data with the workflow we established for ATAC-Seq, I performed

further quality control tests looking at correlation between replicates, insert size metrics, and read

coverage over TSS's. To calculate correlation between the replicates, I used DiffBind to obtain a shared

peakset and normalized reads, then plotted a scatter plot of normalized rpms for each peak. The

correlation within replicates was excellent, Pearson's r=0.96 for the B6 replicates and r=0.99 for the

B6[CAST-KI] replicates (Fig. 2A, B6 shown). To examine DNA fragment lengths showing nucleosome

occupancy, I ran Picard 'CollectInsertSizeMetrics' which outputs a histogram of read lengths for each

sample. The insert size metrics for our ATAC libraries show a large proportion of reads with less than 100

bp, which represents the nucleosome-free regions, and then fragment size periodicity corresponding to

the approximate lengths of DNA wrapped around nucleosomes, indicating nucleosome occupancy from

quality, intact chromatin (Fig. 2B).
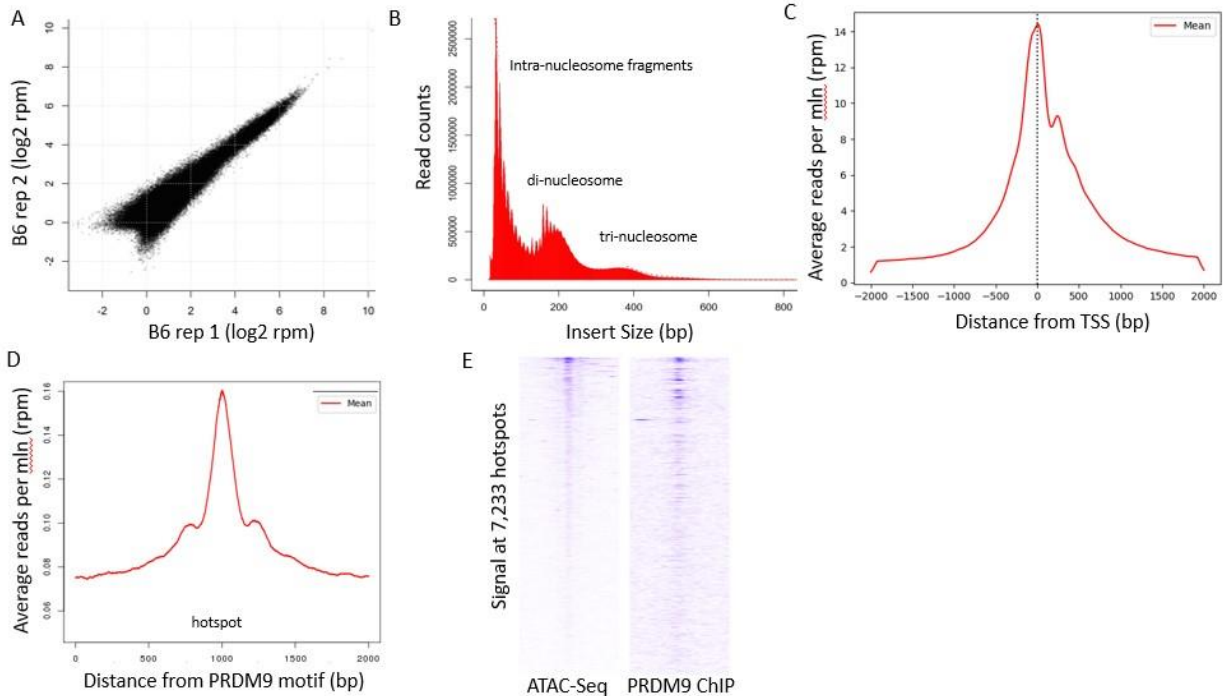
Figure 2. ATAC-Seq shows open chromatin at hotspots. (A) ATAC-Seq replicates on B6 spermatocytes show high correlation. (B) ATAC-Seq read length show evidence of nucleosome banding. (C) ATAC reads are enriched at transcription start sites (TSS's) (D) ATAC reads are enriched at hotspots. (E) ATAC and PRDM9 read depths across a 2kb region centered on hotspots for 7,233 hotspots. (F) Density plot of aggregate signal over hotspots for both ATAC and H3K4me3.

Another hallmark of quality ATAC-Seq data is enrichment of reads at transcription start sites (TSS's), where active chromatin at gene promoters are known to have open chromatin. I used CoverageView to inspect how the ATAC-Seq coverage distributed across the genome (Lowy, 2020), and specifically to plot read coverage profiles around TSS's. I used as input our ATAC bam files and a bed file of ENCODE-curated TSS's and ran CoverageView 'cov.matrix', and the program returned a matrix with read counts at each base pair location across the 2 kb region centered around each TSS. I plotted the means in a density plot to look at enrichment of ATAC reads and our libraries show a strong peak at the TSS that quickly tapers on either side to background levels, showing specificity. Our ATAC-Seq showed an enrichment of ~14-20 rpm at TSS's compared to <1 rpm in the flanking regions (Fig. 2C, B6 shown). Our enrichment at TSS's was within the range recommended by ENCODE, which suggests a TSS

enrichment of 10-15 is acceptable, and >15 is ideal (https://www.encodeproject.org/atac-seq, updated

June 2020)

4.5. ATAC peaks at hotspots show increased DNA accessibility during meiosis

I recovered 88,474 shared ATAC peaks from the B6 spermatocyte samples. I annotated the peaks based

on whether they overlapped with known TSS's, PRDM9^{Dom2}-hotspots, or other. In addition to open

chromatin at promoters, our spermatocyte samples had ATAC peaks that overlap hotspots, confirming

that hotspots are sites of open chromatin. Of 7,223 hotspots for which we know PRDM9 motif location ,

I recovered 1,527 by ATAC-Seq. Although this overlap is low, I am likely over-looking low-level ATAC

peaks at hotspots due to my stringent peak-calling (e-5), lower signal at hotpots due to heterogenous

cell types and assorted hotspot usage within meiotic cells, and the more limited dynamic range of the

ATAC assay. I tested for enrichment for ATAC reads at all hotspots by plotting normalized reads per

million from ATAC (rpm) using hotspot PRDM9 motifs as the center of a 2kb window. There is clear pile-

ups of ATAC reads at the hotspot centers compared to the flanking regions (Fig. 2D). A heatmap of the

CoverageView matrix shows ATAC enrichments at individual hotspots, where the strongest hotspots

from PRDM9 ChIP have the strongest ATAC signals (Fig. 2E). Though the largest ATAC signal is in the

middle of the hotspots, there are secondary small peaks on either side that correlate with strict

nucleosome remodeling around hotspots. Overall, these results confirm our hypothesis that hotspots

are regions of open chromatin specific to the hotspot site.

4.6. Hotspots show differential DNA accessibility dependent on PRDM9 allele

To compare peaks between different samples, I concatenated the peak sets from all samples and then

merged overlapping peaks with BEDTools (Quinlan and Hall, 2010) using 'bedtools merge' to make a

shared "peakome." This allowed me to standardize my analysis by using the same peaks for every

sample. Using bedtools 'multicov,' I counted reads from bam files for each sample that map to the

peakome. I then used edgeR (Robinson et al., 2010) to investigate differential expression of count data

from different samples. edgeR uses an overdispersed Poisson model to account for both biological and

technical variability and correctly identify true differences and minimize background (Robinson et al.,

2010). I used 'calcNormFactors' to produce a matrix of normalized read counts per peak for each

sample. I filtered for peaks with less than 3 reads to minimize low-level and sample-specific peaks that

skew statistical analysis. I used this normalized peak matrix to test for differential ATAC peaks between

B6$^{CAST-KI}$ and B6.

I plotted an MA plot to determine the change in ATAC peak accessibility between the two

strains. The MA plot showed that most ATAC peaks did not change between the two strains of mice,

which would be expected since the mice have the same genetic background (Fig. 3A). However, there

was another set of ATAC peaks that showed significantly different DNA accessibility, and these locations

correspond to hotspots that are determined by either PRDM9^{Dom2} (blue) or PRDM9^{Cst} (red). I uploaded

bedgraph tracks for each of the ATAC-Seq samples to the UCSC genome browser to visualize ATAC reads

along the genome. I looked at allele-specific hotspots and confirmed that the B6$^{CAST-KI}$ showed no DNA

accessibility at PRDM9^{Dom2} hotspots but did have novel DNA accessibility at PRDM9^{Cst} hotspots compared

to B6 (Fig. 3B). For the ATAC peaks that were significantly higher (FDR<0.01) in the B6$^{CAST-KI}$, 92% (n =

5,305) overlap with reported PRDM9^{Cst}-dependent hotspots (Baker et al., 2014). Reciprocally, for the

ATAC peaks that are higher in B6, 94% (n = 1,643) overlap PRDM9^{Dom2}-dependent hotspots (Baker et al.,

2014). The difference in the number of ATAC-enriched hotspots detected in the B6$^{CAST-KI}$ vs the B6 is due

to PRDM9^{Cst} having stronger binding affinity in the B6 genetic background than the endogenous

PRDM9^{Dom2}, which has been previously published (Baker et al., 2014; Baker et al., 2015a; Baker et al.,

2015b; Grey et al., 2017). Clearly, hotspots are sites of PRDM9-specific open chromatin

4.7. Open chromatin at hotspots is independent of double-stranded breaks

An alternative explanation for open chromatin at hotspots is the presence of double stranded DNA

breaks (DSBs). It is known that DSB repair machinery can remodel nucleosomes and create DNA

accessibility to facilitate DNA repair (Price and D'Andrea, 2013). In meiosis, hotspots become sites of

programmed DSBs after PRDM9 binds and the DNA is cut by meiosis-specific SPO11 (Baudat et al., 2000;

Romanienko and Camerini-Otero, 2000). Therefore, we investigated whether the appearance of

increased DNA accessibility at hotspots is due to meiotic DSBs or happens in the absence of DSBs.
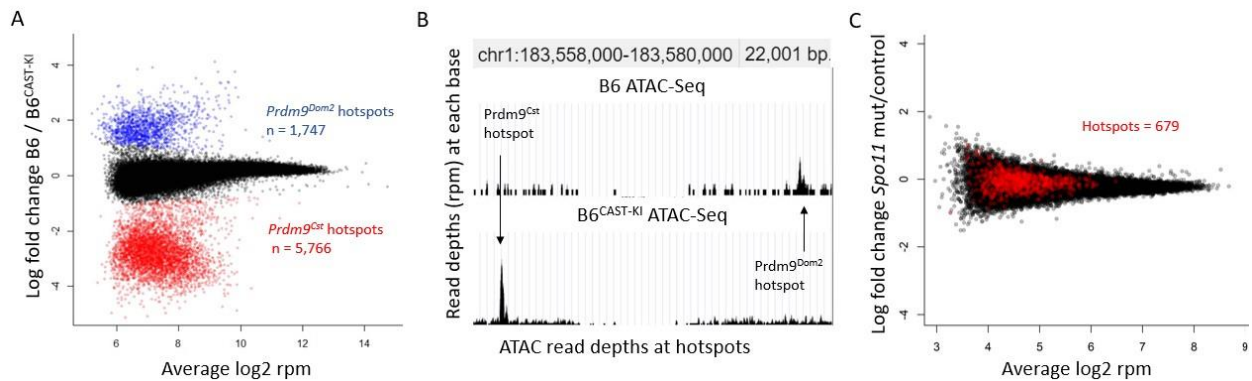
Figure 3. Open chromatin at hotspots is dependent on PRDM9 binding and independent of DSBs. (A) MA plot of differentially open chromatin peaks between B6 and B6$^{CAST-KI}$ with known hotspots for each allele in color. (B) Bedgraph pile-ups of ATAC reads from B6 and B6$^{CAST-KI}$ ATAC-Seq over a Prdm9^{Cst} and a Prdm9^{Dom2} hotspot. (C) MA plot of differentially open chromatin peaks between Spo11 KO and WT. Hotspots are highlighted in red.

4.8. ATAC-Seq on *Spo11* mutant mice shows open chromatin at hotspots

We used a knock-out mouse with an inactive allele of *Spo11* (B6.129X1-Spo11^{tm1Mjn}/J) (Baudat et al.,

2000) and performed ATAC-Seq to look for open chromatin hotspots before DSBs by SPO11. Since

SPO11 is catalytically dead and does not make DSBs, any open chromatin at hotspots in the *Spo11*

mutant will be independent of DSB repair machinery suggesting it is dependent on PRDM9. Indeed,

when I compared ATAC-Seq peaks from 12dpp *Spo11* mutant to wildtype littermates, there were no

differences in ATAC peaks, including at hotspots (Fig. 3C). Therefore, open chromatin at hotspots exists

prior to and independent of SPO11 double-stranded breaks.

CHAPTER 5

HELLS AND PRDM9 ARE CO-EXPRESSED IN L/Z OF MEIOSIS AND FORM A PROTEIN COMPLEX

5.1. HELLS is a good candidate for meiotic chromatin remodeling

Above I showed that hotspots are sites of open chromatin and that open chromatin at hotspots is dependent on PRDM9. Although PRDM9 can bind DNA and modify histone marks at hotspots, it lacks the domains for nucleosome remodeling. If there is another protein creating open chromatin at hotspots, like we hypothesize, then loss of that protein should show an infertility phenotype similar to loss of PRDM9. We examined mouse phenotypes in mice with mutations in known chromatin remodeling proteins and found three candidates with a fertility phenotype: *Brg1*, *Ino80*, and *Hells*. We further narrowed the candidates by looking for a fertility phenotype that mimics PRDM9 knockout phenotype (Hayashi et al., 2005), specifically a block in early pachytene due to incomplete synapsis and persistent DSBs. Loss of *Ino80* (Serber et al., 2016) and *Hells* (Zeng et al., 2011) mutations exhibited meiotic infertility phenotypes similar to loss of *Prdm9*, however *Ino80* is involved in H2A.Z and H2A variant exchange (Papamichos-Chronakis et al., 2011), and previous work did not find evidence of H2A.Z at recombination hotspots (Spruce et al., 2020). Given that HELLS is capable of remodeling chromatin and creating DNA accessibility, and has a fertility phenotype, we sought to determine whether HELLS can associate with PRDM9, particularly in spermatocytes in L/Z stage meiosis.

5.2. *Hells* RNA is expressed in meiotic spermatocytes

On indication that Hells would play a role in meiotic recombination would be it is expressed during leptotene/zygotene stage spermatocytes, just like *Prdm9*. It would add strength to our hypothesis is *Hells* expression is also specific to this stage over others. We searched for *Hells* RNA expression using data from Jung et al., 2019 that performed single cell RNA-Seq on testes to define stages of spermatogenesis. We saw strong expression of *Hells* in L/Z spermatocytes, which correlated with

expression of *Prdm9* (Fig. 4A and B). However, unlike PRDM9, *Hells* is additionally expressed in spermatogonia at an earlier stage in spermatogenesis (Fig. 4B). As a side note, *Ino80* is expressed during meiosis, but more broadly and not correlated with Prdm9. This data shows that HELLS is present when PRDM9 is acting to establish recombination hotspots and during the acquisition of meiotic DSBs.

5.3. Confirm HELLS protein expression in testis

Now that we know *Hells* RNA is expressed in L/Z stage cells, we wanted to confirm that HELLS protein was indeed present, since RNA expression does not necessarily mean that the protein is being translated. Commercial anti-HELLS antibody are available through several vendors. The rabbit polyclonal anti-HELLS antibody from Millipore (cat# ADB41), was reported to work for Western blot and recognize both mouse and human HELLS. To test whether the anti-HELLS antibody worked, I extracted protein from mouse tissues that should be either positive (testis) or negative (heart and liver)for HELLS protein. I took a ~10mg piece of tissue for each sample and used the previously published RIPA extraction protocol to lyse the tissues and extract protein (Baker et al., 2015b). Samples were heated to 95°C for 5 min with 1X Laemmli's buffer, which contains 1% SDS that helps denature and unfold proteins, allowing them to run on the gel at their proper size and expose latent epitopes. I loaded the Invitrogen 4-12% gel heavily with 60ug protein/lane to maximize my chances of seeing a signal I followed the manufacturer's recommended concentration of 1:5000 anti-HELLS antibody in our primary incubation conditions of TBS + 0.3% Tween, overnight, at 4°. The next day, I washed the blot and incubated it for an hour in 1:10,000 anti-rabbit secondary conjugated with horseradish peroxidase (HRP) and washed it again. I developed the blot using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (ThermoFisher Scientific, 34579), which provides the substrate to the HRP enzyme to produce luminescence. I exposed the blot on film using at 1min, 30sec, 15sec, and 5sec intervals and then developed the film to look for signal at 97kD, the expected weight for HELLS.

5.4. Testis shows strong, specific signal for HELLS

The testis sample showed strong, specific signal for HELLS around 110kD (Fig. 4D). In contrast, none of the other tissues were positive for HELLS. The protein band for HELLS is higher than the calculated weight, but both the manufacturer's example Western and Westerns for HELLS from other papers both show bands the same size as mine. Post-translational modifications, such as phosphorylation or glycosylation, could cause the protein to run slower on a Western gel. I concluded from these preliminary results that we could detect HELLS with our antibody, and I confirmed that HELLS is specifically expressed in testis.

5.5. Immunohistochemistry on testis cross-sections shows co-localization of PRDM9 and HELLS in the same cells

In addition to the positive results from the RNA-Seq and the western, we wanted proof that HELLS and PRDM9 are co-expressed in the same cells. Immunohistochemistry is a technique to detect proteins of interest in fixed tissue sections by fluorescently labeling them with modified antibodies.

5.5.1 Fixation of adult testis and mounting cross-section onto slides

Samples for immunohistochemistry are fixed to preserve the structure and cellular context of the tissue. I fixed 6-week old adult C57BL/6J testis in 4% PFA overnight at 4°. The next day, I moved the fixed tissue into 70% ethanol to wash away the fixative before submitting my sample to the Histology core where the fixed testis was embedding in paraffin and cross-sectioned onto slides. I requested 5µm thick sections spaced ~1 cm apart along the slide so I could perform multiple IHC assays on the same slide without touching between sections.

5.5.2. Deparaffinization, rehydration, and antigen unmasking

Embedding the tissue in paraffin supports the tissue structure for fine sectioning but interferes with IHC analysis. In addition, the sections need to be dehydrated to remove excess fixation solution and tissue fluids that will interfere with the IHC reaction. Both of these are accomplished by washing twice for 8

min in xylene, followed by 2 min washes in 100%, 95%, 80%, and 70% EtOH. I rinsed the slides for 5 min in H2O and rehydrated them with three 5 min washes in PBS.

The process of sample fixation can lead to protein cross-linking, which masks antigens and can restrict antigen-antibody binding. Antigen unmasking permits antibodies to access the target protein within the tissue. To unmask antigens, I boiled the slides submersed in 10mM sodium citrate in the microwave for 10 min to break protein cross-links and unmask antigens. Afterwards, I let them cool to room temperature undisturbed before rinsing them three time for 5 min in PBS.

5.5.3. Permeabilization of cells and blocking of non-specific binding

IHC signals can be improved by increasing antibody access to cells by permeabilizing the cell membranes with a detergent like Triton X-100. I incubated my slides in 0.05% Triton-X100 in PBS for 30 min at room temperature to permeabilize, and then washed three times for 5 min in PBS. IHC signals can also be improved by reducing background noise from antibody binding to non-specific targets. Normal donkey serum (NDS)contains antibodies that do not recognize any mouse proteins. Therefore, any interactions it forms to tissues are non-specific, and pre-emptively displace potential non-specific interactions with our target antibody. I blocked my slides with 10% normal donkey serum in PBS for 30 min. I used a humid chamber to inhibit evaporation and allow me to reduce the amount of reagent I needed for each section. Instead of submerging the slide in 10% NDS, I circled each of my tissue sections with a waterproof marking pen to restrict run-off, and then deposited a 30-50 µl droplet of blocking solution into the circle and onto the section. After 30 min of blocking, I tilted the slide and tapped it gently to knock off the droplets from the sections, and then added the primary antibody.

5.5.4. Titrating anti-HELLS IHC staining

First, I needed to ensure our anti-HELLS antibody worked in our IHC assay and determine the optimum concentration of antibody. I based my anti-HELLS dilutions on the typical range of dilutions for IHC, which was ~1:200 to 1:500 for most antibodies. I have 5 testis sections per slide, so I diluted the anti-

HELLS antibody to 1:1000, 1:500, 1:250, and 1:100 in PBS + 1% BSA and kept one section as a secondary-only control. I deposited 20-30 µl of primary antibody solution to each of the sections, and PBS + 1% BSA only to the secondary-only control and incubated overnight in a humid chamber at 4°.

5.5.5. Addition of fluorescently-conjugated secondary antibody and imaging

The next day, I tapped off the primary antibody solution and submersed my slide in three 5min washes in PBS. I re-blocked with 10% NDS for 15 min before adding the secondary antibody. Secondary antibodies can be conjugated with several different fluorophores, and it is important to have a range of fluorophore options to perform co-staining with multiple antibodies and not have the secondary fluorophores overlap. I used goat anti-rabbit IgG, Alexa Fluor 488 conjugate (ab150077) to detect HELLS at 1:300 in PBS + 1% BSA for 2 hrs. It is crucial to perform the secondary incubation in the dark to protect the fluorophores from light. After secondary antibody incubation, I washed the slides and mounted them with antifade reagent with DAPI. I imaged the slides with the Zeiss AxioImager and saw strong, specific staining for HELLS in all four antibody dilutions, and nothing in the secondary-only control. I chose 1:250 as the concentration, because it appeared to be the best balance between strong staining and not overloading the signal.

5.5.6. Immunostaining with antibodies to HELLS, PRDM9 and marker protein yH2AX

The goal of my IHC staining was to confirm colocalization of PRDM9 and HELLS in the same cell. I performed IHC on adult testis sections and co-stained for HELLS, PRDM9, and yH2AX. I included yH2AX as a marker to help distinguish stages in spermatocytes. The yH2AX protein localizes to double-stranded DNA breaks, therefore it stains strongly and diffusely in L/Z stage when DSBs are present due to SPO11 activity. In pachytene stage, yH2AX staining becomes restricted to the sex body. I used the optimized concentration for anti-HELLS (1:250 Millipore/Sigma, ABD41) along with previously optimized concentrations for anti-yH2AX (1:2000, Millipore 05636I, lot#2888552) and anti-PRDM9 (1:100, Parvanov et al. 2017) and was careful to use non-overlapping fluorophores for each of my secondary

antibodies (goat anti-rabbit IgG Alexa Fluor 594 conjugate (A-11037) 1:300, goat anti-mouse IgG Alexa

Fluor 647 conjugate (A-21236) 1:300, goat anti-guinea pig IgG Alexa Fluor 488 conjugate (A-11073)

1:300). I imaged the sections and saw clear, specific staining in each of my channels. The PRDM9 staining

was not as strong or specific as the others, I needed to increase the exposure time and accept some

non-specific signal in between the tubules. I tried trouble-shooting the protocol to increase PRDM9

signal but was not able to improve it. Still, within tubules it was clear which cells were positive for

PRDM9 and which were not, even if the staining was not as crisp as the other two antibodies.
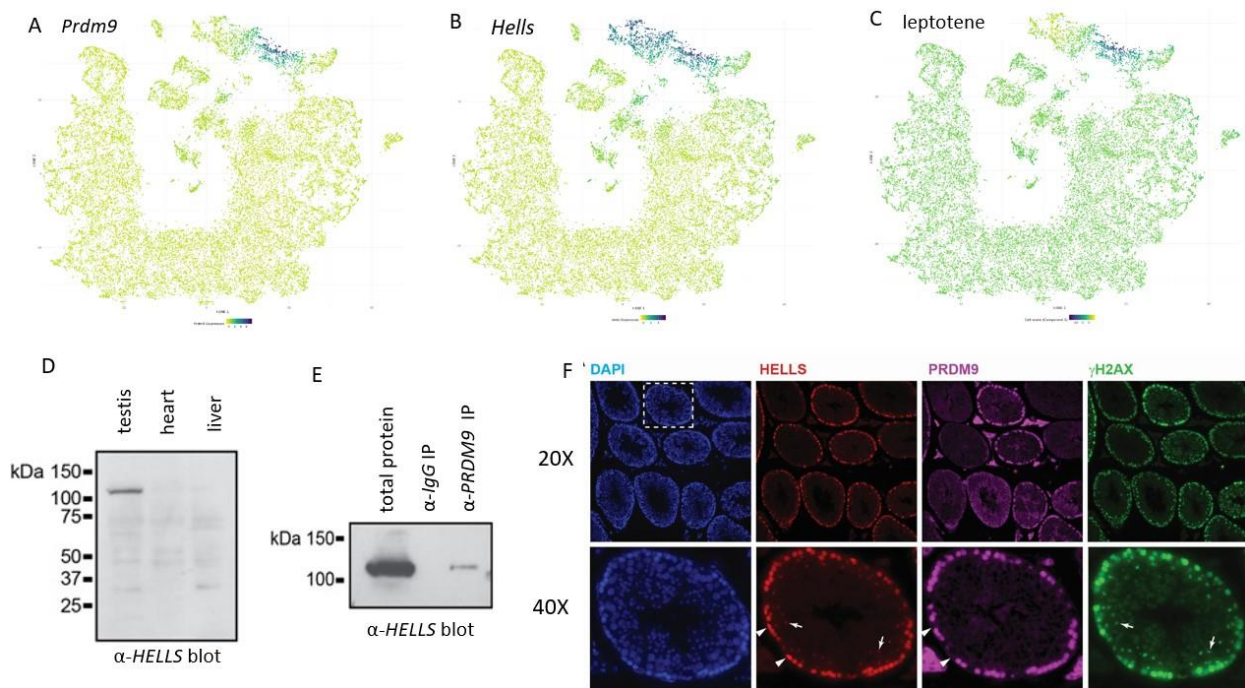
Figure 4. *Hells* is co-expressed in testis with *Prdm9*. Single-cell RNA-Seq tSNE plot of (A) *Prdm9* (B) *Hells* (C) leptotene stage expression in spermatogenesis. (D) Western for *HELLS* of testis and various other tissues. (E) Western for *HELLS* of immunoprecipitation (IP) using *IgG* or *PRDM9* antibodies on testis protein. (F) Immunohistochemistry (IHC) on adult testis sections staining for DAPI, *HELLS*, *PRDM9*, and *yH2AX*.

The results clearly show co-staining of both HELLS and PRDM9 (Fig. 4F), showing that HELLS and

PRDM9 are active and expressed at the same time in the same cell. The co-stained cells appear to be L/Z

stage judging by the diffuse yH2AX signal. In addition, I could see HELLS signal in type B spermatogonia,

which are pre-meiotic cells showing no yH2AX signal, and also in pachytene-stage cells, where both HELLS and yH2AX are localized to the sex body. The immunohistochemistry results on testis sections shows co-localization of HELLS and PRDM9 and support our hypothesis that the two proteins may interact to activate hotspots in L/Z stage meiosis.

5.6. HELLS protein forms a complex with PRDM9 protein

HELLS has no DNA binding domain of its own and so likely needs to be tethered to a binding partner with DNA-binding specificity to be recruited to DNA (Zhu et al., 2006; Jenness et al., 2018) . I performed coIP for HELLS and PRDM9 to determine whether HELLS protein can form a complex with PRDM9 protein, which would provide a mechanism for HELLS recruitment to hotspots. Protein co-immunoprecipitation is a well-established technique to look for protein interactions. The technique works by using a bead-bound antibody to bind and purify the primary protein and whatever protein complex that is co-bound to it. Then a Western blot for the second protein of interest is run on the purified protein material, to interrogate whether the two proteins form a complex that can be co-immunoprecipitated. Two working antibodies, one for each protein, are required for co-immunoprecipitation. Luckily, we were gifted a working PRDM9 antibody from a collaborating lab (Parvanov et al., 2017).

5.6.1. Protein extraction to preserve protein-protein interactions for co-IP

For the co-immunoprecipitation, I extracted testes from four 12dpp mice and pooled them into one sample to get enough protein for a negative control co-IP, HELLS experimental co-IP, and untreated control protein. My previous protein extractions used RIPA buffer, which contains harsh detergents good for lysing cells and denaturing proteins. However, for the co-IP I used HEPES extraction buffer (50 mM Hepes-KOH pH 7.4, 137 mM NaCl, 10% Glycerol, 0.4% NP-40) to minimize disruption of potential protein complexes and keep them intact for pull down. I added benzonuclease to degrade DNA and RNA. This is good for two reasons: to prevent clumping due to sticky DNA released from dying cells, and

to avoid spurious interactions from co-binding at the same chromatin site, rather than true protein-protein binding.

5.6.2. Co-IP experiment and controls

The goal of the co-IP experiment is to test whether HELLS and PRDM9 form a complex in L/Z spermatocytes. I performed co-IP with anti-PRDM9 antibody to purify PRDM9 protein and any other proteins associated with it. In addition, I also performed co-IP with anti-IgG antibody which has no specificity to any protein and serves as a negative control to show non-specific background signal. First, I incubated 4 µl of either anti-PRDM9 or anti-IgG with 20 µl of Protein G magnetic beads for >20 min to bind the antibody to the beads. After washing to remove excess antibody, I added 40% of the testis protein to each co-IP with HEPES buffer with protease inhibitors. The leftover 20% of protein I saved as untreated input control. I incubated the co-IP overnight at 4° to allow the antibody to bind with its target protein. The next day I placed the co-IP solution against a magnet to draw the antibody-beads out of solution, and along with it the protein of interest and any co-bound proteins. I washed the away the unbound protein fraction from the beads and then eluted the bound protein from the beads by adding Laemmli's buffer and heating to 95° for 10 min, allowing the SDS to simultaneously release the proteins and linearize them.

5.6.3. co-IP for PRDM9 co-precipitates HELLS

To determine whether HELLS is concurrently pulled down in an IP for PRDM9, I ran the resultant co-IP protein on a Western and blotted with anti-HELLS. The protein pulled down in the anti-IgG co-IP showed no enrichment for HELLS, a confirmation of the assay's specificity, however, the anti-PRDM9 co-IP protein had a positive signal for HELLS, showing that the two proteins are in a complex together (Fig. 4E). The strength of the HELLS signal in the anti-PRDM9 co-IP was weaker than the total signal from the untreated input protein, which could be due to either a weak bond between the two proteins that could not hold up to processing, or not every molecule of the PRDM9 protein is bound to a molecule of HELLS

protein. Still, the result of a co-IP is a qualitative rather than quantitative result and the positive

interaction is clear. This evidence of PRDM9 interacting with HELLS encouraged us to believe that if

PRDM9 binds to HELLS, it could be recruiting HELLS to hotspots.

CHAPTER 6

CREATION OF A CONDITIONAL KNOCK-OUT OF HELLS IN MICE

6.1. Creation of a conditional knock-out mouse model to investigate loss of HELLS *in vivo*

In order to prove that *Hells* is the operative protein inducing DNA accessibility at hotspots, we needed to

inactive or remove it during meiosis and determine effect on chromatin. Unfortunately, homozygous

Hells null mice are not viable and die within a few hours of birth, probably due to liver problems

(Geiman et al., 2001). Previous labs had attempted to circumvent this by surgically grafting newborn

Hells KO testis cells into adult wildtype hosts (Zheng et al., 2011). Doing so they found a meiotic arrest at

pachytene in the *Hells* KO spermatocytes. To improve on the previous study and determine the

mechanism of the meiosis-specific functions of *Hells in vivo* we created a *Hells* conditional knock-out

(CKO) mouse.

6.1.1. Creation of mice from EUCOMM conditional KO mESCs

An international consortium generated conditional knock-outs alleles for of most mouse genes,

including *Hells*, so we did not need to design and construct a *Hells* KO allele *de novo*. We ordered

conditional *Hells* KO embryos from EUCOMM, C57BL/6NTac- *Hells*^tm1a(EUCOMM)Wtsi^/Ieg, EM04583. In the

case of the *Hells* allele, critical exon 12 is flanked by loxP sites (Fig. 5A). All conditional alleles from

EUCOMM follow the same structure: a 5' intronic integration upstream of a critical exon containing an

FRT-flanked cassette that contains lacZ and neo genes followed by one loxP site, and another intronic

integration containing just a loxP site on the 3' side of a critical exon (Fig. 5B).

6.1.2. Genotyping pups from rederivation

The EUCOMM repository shipped frozen 2-cell embryos from a heterozygous C57BL/6NTac-

Hells^tm1a(EUCOMM)Wtsi^/Ieg male by a wildtype female. We employed the Reproductive Services core at The

Jackson Laboratory to transfer the frozen embryos into pseudo-pregnant female B6 mice. The transfer

was successful, and we received 16 pups that must be either heterozygous *Hells*-tm1a or wildtype. To

genotype the pups, I used the PCR assay provided by EUCOMM. The forward primer, *Hells*-5'arm-F, for

both the wildtype and mutant alleles is upstream of the 5' integration cassette in wildtype sequence.

The reverse primer for the wildtype assay, *Hells*-3'arm-R, is 3' of the integration site, and amplifies the

native sequence in the wildtype allele. However, in a mutant allele, the presence of the lacZ and neo

cassette will put the wildtype reverse primer too far away from the *Hells*-5'arm-F to amplify a PCR

product. The reverse primer for the mutant allele, LAR3-R, is located within the integrated cassette and

amplifies a product only when the conditional allele is present (Fig. 5B and C).

6.1.3. Removal of lacZ and neo cassette by mating to FLP

The *Hells*-tm1a allele that we currently had is a targeted trap allele with a splice acceptor upstream of

the lacZ gene. Left intact, the tm1a will interrupt proper splicing and generate a null allele. To remove

the cassette, I mated *Hells*-tm1a heterozygous mice to homozygous transgenic ACTFlpe mice that

constitutively express flippase recombinase protein (FLP) that will excise the DNA flanked by Frt sites,

leaving behind a single inactive Frt site. I genotyped the offspring from the ACTFlpe/ACTFlpe x *Hells*-

tm1a/+ mating using the common forward primer that sits just outside of the integration cassette, *Hells*-

3'arm-F, and the reverse primer I designed that sits 3' of the lacZ neo cassette integration, *Hells* 3'

integration-R, and saw the smaller PCR product expected if the cassette was removed, indicative of FLP-

excision. I also double-checked to be sure there was no *Hells*-tm1a allele remaining by genotyping with

the *Hells* 5'arm-F and the LAR3-R and saw no evidence of the tm1a allele. After crossing to FLP, the

tma1a allele was converted into tm1c, containing just the loxP-flanked *Hells* exon 12. I genotyped for the

Hells-tm1c allele using a primer that overlaps the 5' loxP site, giving it specificity to the conditional allele,

with *Hells* 3'arm-R (Fig. 5C). This *Hells*-tm1c ("c" for conditional) allele is what I used in subsequent

matings.
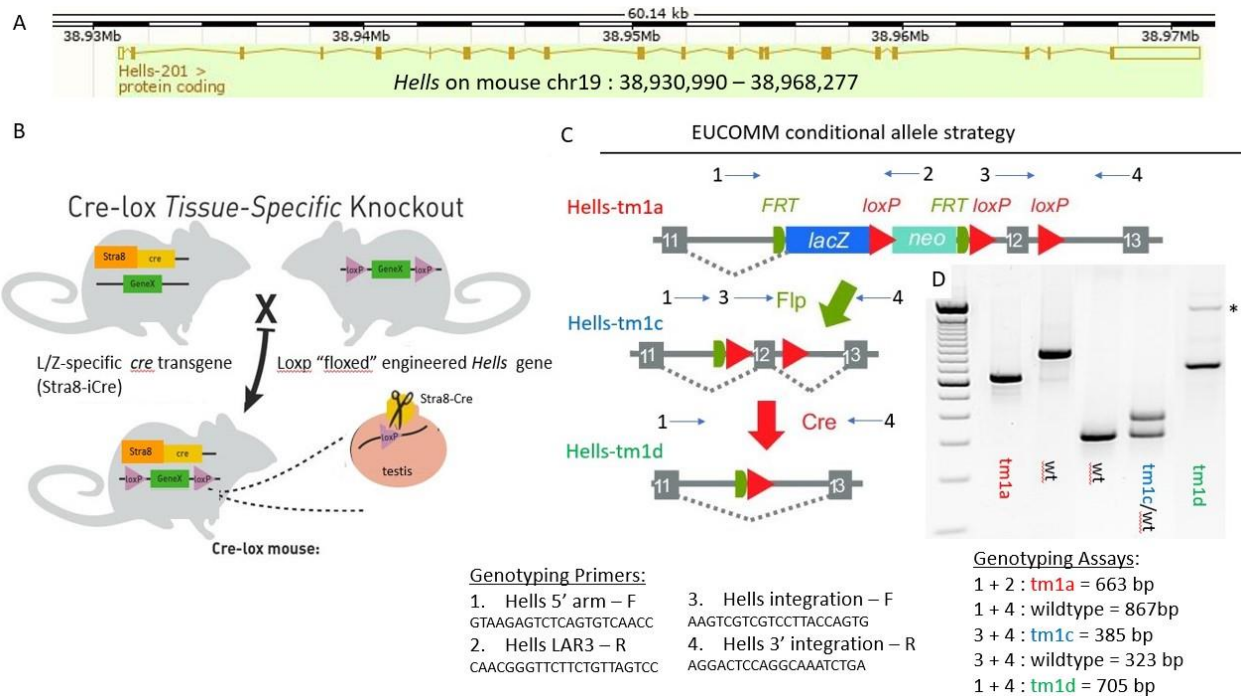
Figure 5. Visual description of EUCOMM conditional allele and *Hells* CKO genotyping assays. (A) Overview of *Hells* gene and integration of conditional knock out allele. (B) Cre mating scheme to generate conditional knock out in testis. (C) Visual representation of elements in tm1a, tm1b, and tm1c alleles. (D) Example gel showing bands from different genotyping assays.

6.1.4. Mating of *Hells*-tm1c to *Stra8*-iCre to generate heterozygous null and on-board *Stra8*-iCre allele

To improve my chances of a complete null, my strategy was to first ensure complete deletion of one conditional allele, and then pursue cell type-specific deletion of the second conditional allele in only spermatocytes. First, I made a constitutive heterozygous knock out of *Hells* by crossing *Hells*-tm1c to *Stra8*-iCre. Heterozygous *Hells* mice are viable and fertile, with no overt phenotypes. My goal is to completely delete *Hells* in L/Z stage spermatocytes, however Cre excision is not always completely efficient. By completely excising one of the floxed exons, I will increase my chances of getting complete excision of the one remaining floxed allele.

The *Stra8*-iCre mouse expresses iCre under control of the *Stra8* promoter, which is expressed specifically at the onset of meiosis in spermatocytes. Using *Stra8*-iCre is advantageous for two reasons:

first, the *Stra8*-iCre is expressed in male germ cell development, so offspring from a *Stra8*-iCre; *Hells*-tm1c sire will be constitutive *Hells* null for the conditional allele; second, the *Stra8*-iCre is specifically-expressed in L/Z stage meiosis, which is critical for the second step when I will generate conditional *Hells* null spermatocytes. By generating mice through these two matings, I can create a *Hells* conditional knock out in testis while the rest of the animal remains heterozygous for *Hells*, as depicted in Fig. 5B.

I mated *Hells*-tm1c/+; ACTFlpe/+ mice to *Stra8* iCre/+ to both breed away from the ACTFlpe allele and gain the *Stra8* iCre allele. I genotyped the offspring and looked for male pups that were positive for *Hells*-tm1c and *Stra8*-iCre, and negative for ACTFlpe, which given the 50% probabilities of each of these alleles make the combined probability of about $1/16^{th}$ of all pups. The resulting male *Hells*-tm1c; *Stra8*-iCre offspring will produce heterozygous null germ cells. The second generation indeed genotyped positive for the deletion, becoming *Hells*-tm1d ("d" for deletion) (Fig. 5D).

6.1.5. Mating of *Hells*-tm1c/*Hells*-tm1c females to *Hells*-tm1d; *Stra8*-iCre males to produce the genotype of interest

Meanwhile, I mated *Hells*-tm1c/+; ACTFlpe/+ mice to each other to produce homozygous *Hells*-tm1c/*Hells*-tm1c females that were negative for ACTFlpe ($1/32^{th}$ probability). To obtain males that were *Hells* null in L/Z spermatocytes, I mated *Hells*-tm1c/*Hells*-tm1c females to *Hells*-tm1c/+; *Stra8*-iCre/+ males (will be *Hells*-tm1d in germ cells) to produce offspring that are *Hells*-tm1c/*Hells*-tm1d; *Stra8*-iCre males, the final genotype of interest. I maximized my chances of getting *Hells*-tm1c/*Hells*-tm1d; *Stra8*-iCre males through this final mating, but still the probability is only $1/8^{th}$. The only way I could increase my odds is by making the *Stra8*-iCre allele homozygous in the sire, which I was able to do in subsequent matings by identifying *Hells*-tm1d; *Stra8*-iCre/*Stra8*-iCre males through progeny testing, given that the generic iCre primers cannot differentiate between heterozygous and homozygous *Stra8*-iCre.

I genotyped the male offspring from my final cross for *Stra8*-iCre, *Hells*-tm1c, and *Hells*-tm1d. The *Hells* tm1d allele I genotyped using the *Hells* 5'arm-F primer with *Hells* 3' integration-R, which span

the floxed exon. In the tm1c conditional allele, the primers are 1200 bp apart, but in the *Hells*-tm1d

excised allele, they are only 705 bp apart (Fig. 5C). I also genotyped for the maternal *Hells*-tm1c allele

and the paternal *Stra8*-iCre. In my genotype of interest, the mouse tail DNA will genotype as *Hells*-

tm1d/*Hells*-tm1c; however, in the spermatocytes the *Stra8*-iCre allele will be active and will excise the

floxed *Hells*-tm1c allele, making them null for *Hells* in spermatocytes. I used these *Hells*-tm1d/*Hells*-

tm1c; *Stra8*-iCre males to study the effects of loss of *Hells* in meiosis.

In some progeny from the *Hells*-tm1c; *Stra8*-iCre sires there was evidence of incomplete

conversion of the *Hells*-tm1c allele to *Hells*-tm1d (Fig. 5D, star). To mitigate this effect, I used the *Hells*-

tm1d/+; *Stra8*-iCre males from my final cross as sires in future matings, so the germ cells of these sires

would have another round of exposure of Cre, which should eliminate any residual intact *Hells*-tm1c.

The additional round of Cre exposure worked, where I sometimes saw residual *Hells*-tm1c in offspring

from a *Hells*-tm1c/+; *Stra8*-iCre/+ sire, but not in offspring from a *Hells*-tm1d/+; *Stra8*-iCre sire.

CHAPTER 7

HELLS CONDITIONAL KNOCK DOWN PHENOTYPE

7.1. *Hells* conditional knock down males are infertile and have reduced testis size

The *Hells*-tm1d/*Hells*-tm1c; *Stra8*-iCre mice were phenotypically normal, grew well, and looked similar

to their littermates. I performed fertility testing on four *Hells*-tm1d/*Hells*-tm1c; *Stra8*-iCre adult males by

mating them to proven fertile female B6 mice for 12 weeks, and none of the sires produced pregnant

dams or any litters. I sacrificed the sires and compared their testis morphology to wildtype controls. All

of the *Hells* CKO testis were much smaller in size than the wildtype, similar to the PRDM9 CKO

phenotype (Fig. 6A). To quantify this comparison, I plotted the ratio of the total body weight of the

males to the total weight of its two testes and could see that the testis:body weight ratio for the *Hells*

CKO was significantly lower than the wildtype (Fig. 6B).

7.2. Western blot analysis of testis protein from *Hells* CKO shows absence of HELLS

To confirm that the knock down of *Hells* in spermatocytes was successful, I performed western blot on

12dpp testis protein from *Hells* CKO, heterozygous, and wildtype littermates. I blotted for HELLS and saw

absence of HELLS in the CKO and reduced level of HELLS in the heterozygote, compared to normal levels

in the wildtype (Fig. 6C). This confirmed that our mating strategy and the Cre excision were effective at

generating a tissue-specific *Hells* null.

I also blotted for PRDM9 on the same protein samples to check that PRMD9 expression

remained unaffected by knock out of HELLS. Surprisingly, PRDM9 levels were slightly reduced in the

Hells CKO compared to wildtype (Fig. 6C). I collected three additional replicates of *Hells* CKO and control

littermates and performed HELLS and PRDM9 blots in triplicate with b-TUBULIN as a loading control and

confirmed that indeed there was an ~20% reduction in PRDM9 protein in *Hells* CKO compared to

matched controls (Fig. 6D). However, we ruled out that this slight reduction in PRDM9 protein could

cause the infertility phenotype in the *Hells* CKO since heterozygous PRDM9 mice have ~50% reduction in

PRDM9 protein but still activate hotspots, albeit at a slightly reduced efficiency, and complete successful

meiosis (Baker et al., 2015). We can only speculate that the loss of binding between HELLS and PRDM9

in the *Hells* CKO somehow affects PRDM9 protein half-life or that presence of HELLS somehow regulates

PRDM9 transcription.
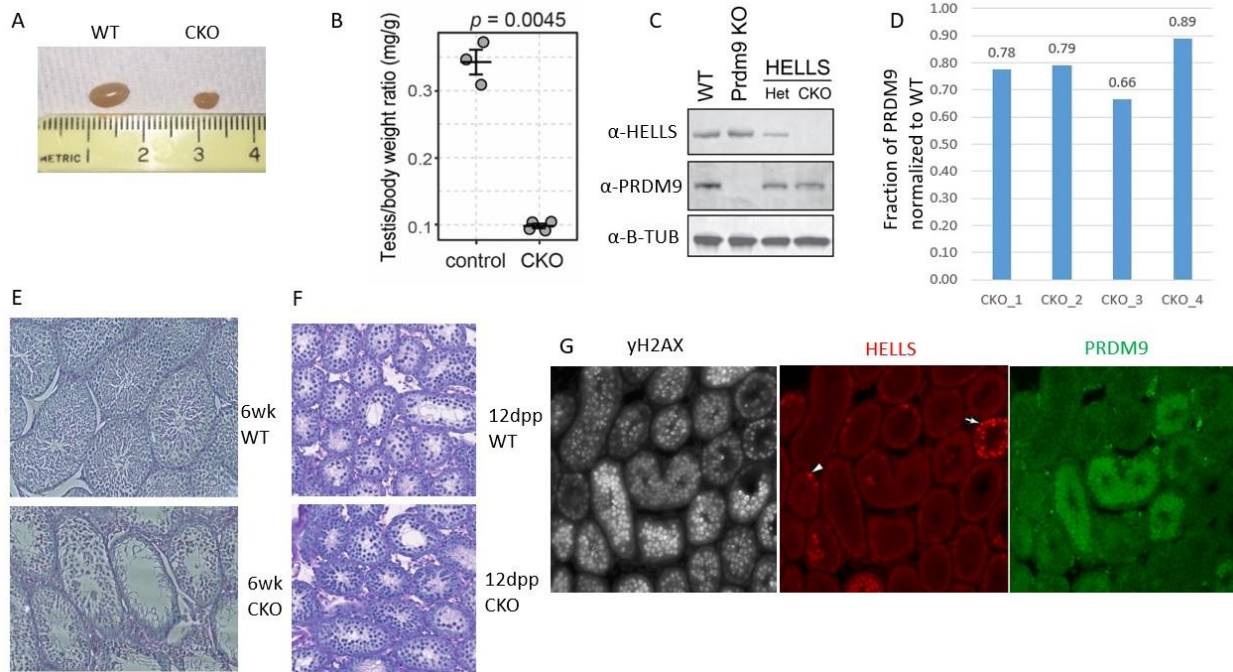
Figure 6. *Hells* CKO testis show infertility phenotype. (A) and (B) Comparison of testis size between *Hells CKO* and wildtype. (C) Western of testis protein showing levels of HELLS and PRDM9. (D) Normalized PRDM9 expression from 4 replicate *Hells* CKO testis as a fraction of wildtype. PAS staining of testis sections from (E) 6-week adult and (F) 12 dpp juvenile from wildtype and *Hells* CKO. (G) IHC staining for yH2AX, HELLS, and PRDM9 in adult *Hells* CKO testis.

7.3. Histological analysis of *Hells* CKO seminiferous tubules show arrest at stage 10 of

spermatogenesis

I investigated the effects of *Hells* CKO in testis by examining their morphology in histological sections. I

harvested testis from *Hells* CKO and *Hells* het littermates, as well as C57BL/6J wildtype control. I

collected testis from both juvenile 12dpp males and adult 6-week old males and fixed the testis

overnight at room temperature in Bouin's solution, which is a good fixative for fixing soft tissue and gives crisper and better staining for nuclei. The next day, I washed the fixed testis in three 1 hour washes of 70%, 50%, and 25% EtOH and then stored them in H_2O to remove excess pigment from the fixed tissue. I brought the fixed tissue to the Histology core to be embedded in paraffin, cross-sectioned, and stained with PAS to look at cellular morphology and organization. I took the 5 μm PAS-stained slides and submitted them to our Microscopy for Nanozoomer imaging at 40X.

I evaluated the images and it was immediately evident that the *Hells* CKO tubules were devoid of mature sperm and had empty tubules lacking cells (Fig. 6E). Because in adult testes spermatogenesis is asynchronous, each cross-section of tubules will capture individual tubules at different stages in spermatogenesis. Thus an experienced eye can identify the whole spectrum of stages in a sufficiently representative testis cross-section. Because I do not have the requisite experience, I consulted another scientist, Hui Tian, for her expert opinion. She analyzed the histological images and concluded that the *Hells* CKO were arresting at stage 10 of spermatogenesis, where she could find evidence of cells types leading up to that stage but none of the more mature cell types from subsequent stages. Stage 10 in spermatogenesis includes spermatocytes progressing through the pachytene stage of meiosis. Arrest at stage 10 means arrest at pachytene stage in meiosis, the same as seen in PRDM9 knock out and many other fertility phenotypes involving meiotic arrest.

In contrast, PAS-staining on 12 dpp testis for both *Hells* CKO and wildtype showed no difference in morphology. The tubules in both *Hells* CKO and wildtype 12 dpp mice have not yet progressed into stage 10 and there appear to be no effects from loss of *Hells* (Fig. 6F). Both testes show equal proportions of tubules in each stage, so the meiotic arrest we see later is not due to pre-meiotic retardation in spermatogenesis in the Hell CKO but is only initiated when *Hells* CKO cells progress into pachytene stage.

7.4. IHC on *Hells* CKO testis shows loss of HELLS in spermatocytes

I performed IHC for HELLS, PRDM9, and yH2AX on adult *Hells* CKO and wildtype testis sections and saw

loss of HELLS in meiotic cells in the *Hells* CKO. The L/Z tubules still stained well and normally for PRDM9,

but there was no concurrent HELLS signal in the *Hells* CKO (Fig. 6G). I could see evidence of specificity of

the *Stra8*-iCre by the presence of HELLS in pre-meiotic spermatogonia, before *Stra8* is expressed, in cells

lacking PRDM9 signal. This is good for the integrity of our experiment, because it means the phenotype

we see in the *Hells* CKO is specific to meiosis and is not due to some unknown function of *Hells* in

spermatogonia. The staining for PRDM9 looked normal and equal intensity in the *Hells* CKO testis

compared to the wildtype, indicating that meiotic arrest is not due to loss or misexpression of PRDM9.

As a side note, I did see a small number of *Hells* CKO tubules that were able to progress through

meiosis and produce sperm. When I performed IHC for HELLS in the *Hells* CKO sections I saw that some

tubules still expressed HELLS despite carrying *Stra8*-iCre, and these were the tubules with sperm. This is

due to incomplete excision of the conditional allele by Cre, which has also been described by other

investigators using the *Stra8*-iCre allele (Tian et al., 2018).

7.5. TUNEL staining shows increased apoptosis in *Hells* CKO tubules

We predicted that absence of mature sperm and empty tubules meant that cells were undergoing

apoptosis after arrest at stage 10. To confirm this, I stained Bouin's fixed sections from *Hells* CKO and

Hells heterozygous testis with the *In Situ* Cell Death Detection Kit (Roche). The kit labels dying cells by

detecting DNA degradation that is a hallmark of apoptosis. The DNA double-stranded breaks can be

detected by polymerization of fluorescently labelled nucleotides that cleave to the free 3' OH of the

breaks. This method has also been termed TUNEL (TdT-mediated dUTP-X nick end labeling). To use the

kit, I first needed to prepare my sections by dehydrating them to remove all fixation solution and tissue

fluid and then rehydrate them in PBS. I performed two washes in xylene followed by 100%, 95%, 80%,

and 70% ethanol and then two washes each in H_2O and PBS. Next, I incubated the tissue with 20 µg/µl

proteinase K for 5 min at RT to permeabilize the cells. I rinsed twice with PBS to remove proteinase K

and labelled according to manufacturer's recommendation. Once the sections were labeled, I co-stained with DAPI to identify nuclei and imaged the slides for fluorescence from the TUNEL staining (488nm, green) and fluorescence from the DAPI (405nm, blue). The difference in number of cells stained for apoptosis for the *Hells* CKO was immediately evident. *Hells* CKO tubules in stage 10 of spermatogenesis had lots of dying cells, compares to almost none in the wildtype control (Fig. 7A). I counted numbers of TUNEL-positive cells in six tubules from *Hells* CKO compared to six tubules from *Hells* wildtype and saw a large increase in apoptosis due to the loss of later stage germ cells (Fig. 7B).

7.6. Meiotic chromosome spreads show lack of pachytene and diplotene cells in *Hells* CKO

My results from IHC on testis sections showed arrest at stage 10 of spermatogenesis, but to look closer at individual cells undergoing meiosis I made chromosome spreads of spermatocytes to identify the cause of arrest.

7.6.1. Preparing chromosome spreads

I harvested testis from 12 dpp *Hells* CKO and wildtype littermates and placed them in hypotonic extraction buffer (30 mM Tris pH8.2, 50 mM sucrose, 17 mM sodium citrate, 5 mM EDTA, 2.5 mM DTT, 1 mM PMSF). I used Dumont forceps to gently pull apart the tubules into a fluffy cloud to allow access of the lysis buffer. After 30 min lysis, I took small amount (half a grain of rice-sized) of testis tissue and placed it in 0.1 M sucrose. I used forceps to cut tissue in 'steak-like' manner, lifting and dropping tubules to release cells into the sucrose until the mixture became cloudy. Next, I coated a clean slide in 100 µl fresh 1% PFA, then dropped 7µl of testis/sucrose solution in one upper corner and tilted slowly back and forth like a making a "Z" all the way down the length of the slide. I put the slides with chromosome spreads in a humid chamber overnight at room temperature, then the next day took them out and let them air dry until all liquid evaporates, about 30min. After drying, I rinsed the slides in 0.04% Kodak Photo-Flo for 60min before storing them at -20°C.

7.6.2 Immunofluorescence on chromosome spreads

For immunofluorescence, I first blocked slides and permeabilized them by covering them completely in 3 mg/mL BSA, 1% normal donkey serum, and .005% Triton-X 100 in 0.05 M TBS, laying a strip of Parafilm on top of the slide to further reduce evaporation, and incubating at room temperature for ~6 hrs in a humid chamber. After blocking, I co-stained the chromosome spreads with anti-HELLS and anti-SYCP3 (1:500 Santa Cruz SC74569, lot#J1314) or anti-HELLS and anti-yH2AX, using the same dilutions as I did for testis sections and incubated them overnight at 4°, covered with coverslips and sealed with rubber cement to prevent evaporation. The next day I performed washes and added fluorophore-labeled secondaries, and finally added antifade with DAPI before sealing the slides with coverslips and sewing them shut with clear nail polish.

I chose SYCP3 as a marker protein to show the progress of synapsis of chromosomes. SYCP3 is a synaptonemal complex protein that coats paired homologous chromosomes. The signal is diffuse in leptotene phase, but then starts to condense along paired chromosome in zygotene phase, where the signal looks thin and "tangled" as the chromosomes start to pair along their length. In pachytene phase the chromosomes become completely paired and compacted, and the signal looks fat and stubby. I also used yH2AX as a marker protein to show unresolved double-stranded breaks, which we would expect to see if PRDM9 does not activate hotspots and SPO11 instead cuts at promoters or other H3K4me3-marked regions of the genome, locations that will not be resolved in the course of meiosis (Brick et al., 2012).

7.6.3. Chromosome spreads from *Hells* CKO are devoid of pachytene cells and carry unrepaired DSBs

The composition of chromosome spreads from Hell CKO testis was very different from wildtype spreads. In the wildtype, most cells were either pachytene or diplotene phase. In contrast, most of the cells in the *Hells* CKO were in zygotene phase and had very few cells in pachytene phase. I counted 200 cells from each *Hells* CKO and *Hells* wildtype spreads and scored them for meiotic stage to quantify the arrest of meiotic progression (Fig. 7C). The *Hells* CKO showed 56% of cells in zygotene and only 20% in pachytene,

as opposed to 6% of cells in zygotene for the wildtype and 58% in pachytene. Most of the *Hells* CKO

pachytene phase cells showed evidence of unresolved DSBs, which appeared as yH2AX staining on parts

of autosomes outside of the sex body, or SYCP3 staining that looked uncompacted in parts of the

chromosomes (Fig. 7D). The *Hells* CKO appears enriched for zygotene-stage cells because the

spermatocytes cannot progress into or past pachytene stage.
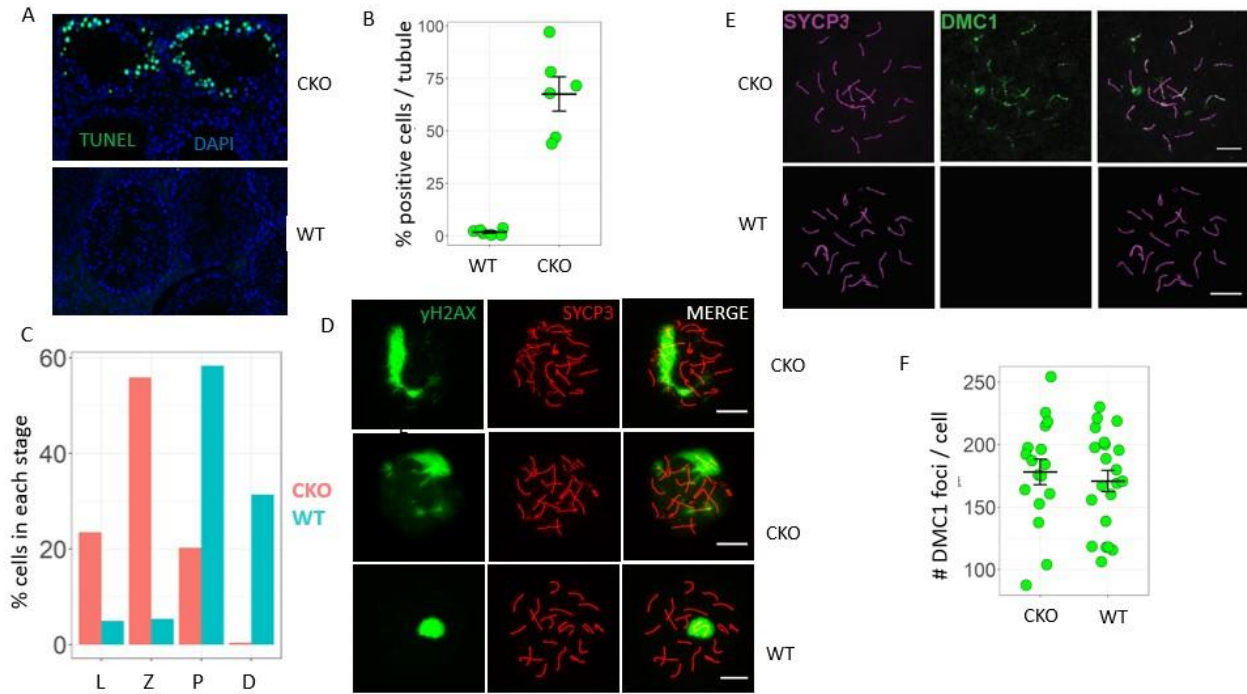
Figure 7. *Hells* CKO testis show arrest at pachytene stage and apoptosis compared to wildtype. (A) TUNEL staining on *Hells* CKO and wildtype adult testis sections. (B) Number of TUNEL-positive cells per tubule. (C) Percentages of cells in each stage of meiosis counted from chromosome spreads. (D) Example of incomplete synapsis and persistence of yH2AX staining in *Hells* CKO "pachytene-like" cells compared to wildtype pachytene cells. (E) Staining for DMC1 foci in zygotene stage cells and (F) number of DMC1 foci per cell for *Hells* CKO and wildtype.

7.7. Staining of chromosome spreads for DMC1 shows unresolved meiotic DSBs

DMC1 is a meiosis-specific protein that binds to single-stranded DNA caused by DSBs at PRDM9-bound

hotspots. Because we saw persistent yH2AX staining we wondered whether these were meiotically-

programmed DSBs and would label with DMC1. I stained a slide of chromosome spreads from 12 dpp

Hells CKO and wildtype with DMC1 and SYCP3. In the wildtype spreads I saw punctate DMC1 staining at DSB foci along the length of zygotene-stage chromosomes and no DMC1 staining in fully synapsed pachytene-stage chromosomes. Interestingly, the *Hells* CKO also showed DMC1 foci along zygotene-stage chromosomes but the foci persist in pachytene-like chromosomes (Fig. 7E). This evidence shows that meiotic double-stranded breaks are persisting after zygotene and are preventing the chromosomes from fully synapsing. The phenotype is similar to the *Prdm9* knock out, where misdirected meiotic DSBs are unable to participate in homologous recombination and remain unrepaired, causing a block in pachytene. I checked to be sure that the persistent DSBs in the *Hells* CKO were not due to an increase in the number of DSBs compared to wildtype by counting the numbers of DMC1 foci per cell during zygotene. The numbers of DSBs in *Hells* CKO and wildtype zygotene-stage cells are not significantly different (Fig. 7F). Therefore, the meiotic block is due to unrepaired DSBs that impede chromosomal synapsis.

CHAPTER 8

HOTSPOTS FAIL TO ACQUIRE ACTIVE HISTONE MARKS AND OPEN CHROMATIN IN *HELLS* CKO

Given our phenotype of meiotic arrest at pachytene phase and evidence of unresolved meiotic DSBS, we next investigated how the loss of HELLS affects chromatin configuration at hotspots.

8.1. Loss of HELLS leads to loss of ATAC peaks at hotspots

I performed ATAC-Seq on 12dpp *Hells* CKO and *Hells* wildtype, two replicates of each genotype. I processed the sequencing data with our pipeline for mapping, calling peaks, and normalizing read counts. For both genotypes, the ATAC-Seq passed QC and showed good enrichment at TSS's. However, there was a serious batch effect in the ATAC-Seq data between the first set of ATAC performed on *Hells* CKO and het ATAC littermates and the second set, where batch explains 67.9% of the variance (Fig. 7A). I used edgeR 'removeBatchEffect' to correct for known batch effect in the samples. After batch correction, genotype now explains 62.7% of the variance (Fig. 8B) and the replicates are highly correlated (Pearson's r=0.98 for CKO and Pearson's r=0.98 for het).

After annotating the peaks as hotspots, promoters, or other, I plotted the change in open chromatin at peaks between *Hells* het and *Hells* CKO. There was a clear population of sites that were drastically reduced in the *Hells* CKO that appear to be exclusively hotspots (Fig. 8C). Of the 1,391 hotspots found in the ATAC peakset, 91% of them are reduced in the *Hells* CKO. In contrast, promoters show very little variation between *Hells* CKO and het (Fig. 8D). I looked at ATAC-Seq read pile-ups on the UCSC browser and confirmed the *Hells* CKO lost enrichment of reads at known hotspot locations compared to the het (Fig. 8E). Using CoverageView to count ATAC reads across the 7,233 hotspots shows loss of ATAC enrichment in the *Hells* CKO, except for some low strength hotspots that appear to have generalized open chromatin around them in both the CKO and the het (Fig. 8F). These results indicate that open chromatin at hotspots is obviated with loss of *Hells*.

These ATAC-Seq results from *Hells* CKO are in keeping with our hypothesis that *Hells* is the causative protein that opens chromatin at hotspots in meiosis. To validate this conclusion, we still needed to prove that PRDM9 was present at hotspots in the *Hells* CKO. Otherwise, the loss in chromatin remodeling could be due to loss pf PRDM9 binding, and not lack of *Hells*.
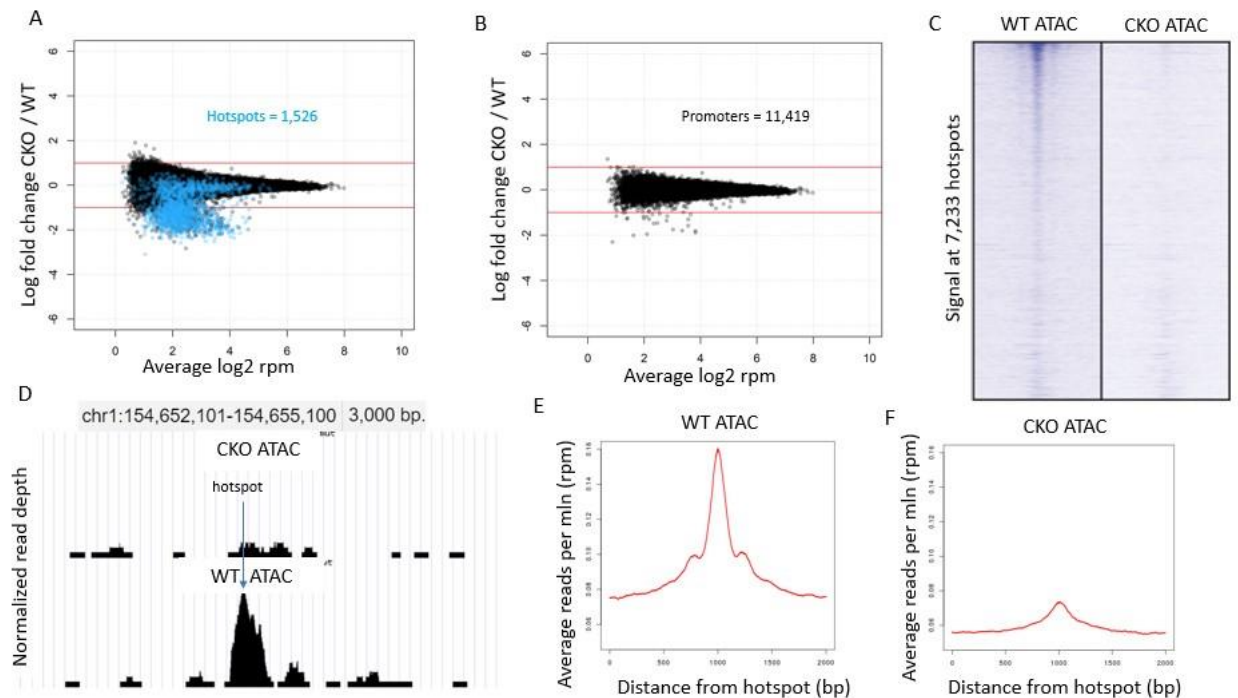
Figure 8. ATAC-Seq on *Hells* CKO shows loss of open chromatin at hotspots. (A) MA plot of differentially accessible ATAC peaks between *Hells* CKO and wildtype. Hotspots are colored in blue. (B) MA plot of differentially accessible peaks at promoters. (C) Heatmap of ATAC read counts across a 2 kb region centered over 7,233 individual hotspots. (D) Bedgraph of read pile-ups at a single hotspot. (E) Density plot of wildtype ATAC reads across a 2 kb region centered on all hotspots. (F) Density plot of *Hells* CKO ATAC reads across a 2 kb region centered on all hotspots.

8.2. ChIP for H3K4me3 shows failure of hotspots to acquire activated marks in *Hells* CKO

Chromatin immunoprecipitation (ChIP) is a technique to look at interactions between protein and chromatin. Like co-IP, the method uses magnetic beads conjugated with an assay-specific antibody to pull down a target protein along with the chromatin to which it was bound. Using ChIP, we can investigate what DNA sequences are bound by the protein of interest or locations of histone

modifications. For our first ChIP on *Hells* CKO, I used an antibody directed to H3K4me3. Activated

hotspots are known to acquire H3K4me3 only after binding by PRDM9 (Baker et al., 2014; Buard et al.,

2009; Eram et al., 2014; Sun et al., 2015; Wu et al., 2013). In fact, our lab typically uses H3K4me3 ChIP as

a proxy for PRDM9 binding because the H3K4me3 histone mark is so strong in response to PRDM9

binding.

8.2.1. Preparation of cross-linked chromatin for H3K4me3 ChIP

For each ChIP, I harvested testes from 4-5 12 dpp male pups and pooled them to make one replicate. I

collected 3 replicates of each *Hells* CKO and *Hells* het littermates and extracted spermatocytes as

described for H3K4me3 ChIP (Baker et al., 2014). After washing with PBS, I cross-linked the cells with 1%

formaldehyde for 10 min to preserve the protein-DNA complexes before stopping the reaction by

adding 125 mM glycine to quench unreacted formaldehyde. Cross-linking cells forms covalent bonds

between protein and DNA or protein and other proteins, stabilizing interactions subsequent

experiments.

8.2.2. Lysis of spermatocytes and digestion of cross-linked chromatin with MNase to fragment DNA

After cross-linking, the cells are hypotonically lysed in 10mM Tris-HCL pH 8.0, 1mM KCl, 1.5 mM MgCl2

with 1 mM PMSF and 1 X protease inhibitor cocktail for 30 min to release nuclei containing the protein-

bound DNA. Hypotonic lysis conditions challenge the cell membrane with low-salt conditions, causing

the cell to swell and rupture to release nuclei, then high-speed pelleting recovers the nuclei from

solution. The intact chromatin from the recovered nuclei is much too large to be efficiently pulled down

by ChIP, and even if it was, the resulting DNA would be too long to determine precisely where the

protein of interest in bound. For these two reasons, chromatin is fragmented before being used in ChIP.

For ChIP for histone modifications, like H3K4me3, we take advantage of the fact that DNA wrapped

around nucleosomes is relatively compacted, making the histone-bound DNA of interest protected from

micrococcal nuclease (MNase), an enzyme that cleaves protein-unbound regions of DNA. Digesting the

chromatin with MNase results in soluble mononucleosome-length fragments ~150bp, plus any surviving

di- and tri-nucleosome fragments. Two high-speed spins help purify the digested chromatin that remains

in solution away from other cellular debris.

8.2.3. Performing IP for H3K4me3 and collection of DNA fragments

The protocol for H3K4me3 ChIP is described in Baker et al., 2014. The IP is performed in RIPA buffer

augmented with 50 mg/mL BSA to reduce non-specific protein interactions with the beads and

0.5mg/mL Herring sperm DNA to reduce non-specific DNA interactions with the beads. First, I

conjugated 7.5 µl of anti-H3K4me3 antibody (Millipore, 07-473) to 60 µl Protein G Dynabeads for

>20min in 300 µl buffer, and afterward washed to remove unbound antibody. I added 3 mln cells' worth

of chromatin to the antibody-bound beads in 300 µl total buffer and incubated at 4° for 2 hrs. After

washing to remove unbound chromatin, I harvested the DNA fragments using an elution buffer

containing high SDS to elute the complexes from the beads, high NaCl to reverse cross-links, and

proteinase K to digest the protein away. The remaining DNA I purified using a Zymo DNA Clean and

Concentrator kit.

8.2.4. Analytical ChIP-qPCR shows reduced H3K4me3 at Pbx1 hotspot in *Hells* CKO

To test whether our H3K4me3 ChIP was successful, I tested our purified H3K4me3 ChIP DNA by qPCR to

look for enrichment of sequences pulled down from hotspots. I would have tested my ATAC libraries by

qPCR and I tried to develop an assay, but the variable fragment length as locations probably contributed

to me being unable to amplify a product from ATAC. However, qPCR is an excellent measure of histone

ChIP enrichment for specific sequences. For PRDM9^{Dom2} on a B6 genetic background, I chose two

hotspots that are particularly active and show strong H3K4me3 binding at the flanking nucleosomes next

to the PRDM9 binding site and used primer sets that amplify the nucleosome-occupied region next to

the strong B6 hotspot. I also used primers sets that amplify next to a strong hotspot for PRDM9^{Cst} as a

negative control and the *Actin* promoter region as a positive control. Contrary to our expectations, I saw

strongly reduced levels of enrichment at Pbx1 in the *Hells* CKO, compared to strong enrichment in the

Hells heterozygous control (Fig. 6A). Both samples showed low enrichment for the PRDM9^{Cst} hotspot

negative control and the high enrichment for the housekeeping gene promoter.

8.2.5. H3K4me3 ChIP-Seq shows reduced H3K4me3 at all hotspots in *Hells* CKO

We sequenced the three replicates of H3K4me3 ChIP for both *Hells* het and *Hells* CKO, ~50mln reads per

sample, single-end. I mapped the reads with bwa, called peaks with MACS, and normalized reads in a

shared peakome. I looked at QC metrics for the H3K4me3 ChIPs to confirm the assay was efficient and

specific. The fraction of H3K4me3 reads that were found in peaks (FRiP) was high in all samples (30-50%)

and replicates showed very good agreement with each other (Pearson's r > 0.95). Most of the 71,877

peaks annotated as hotspots that showed significant reduction in the *Hells* CKO (Fig. 9B). Of the 6,821
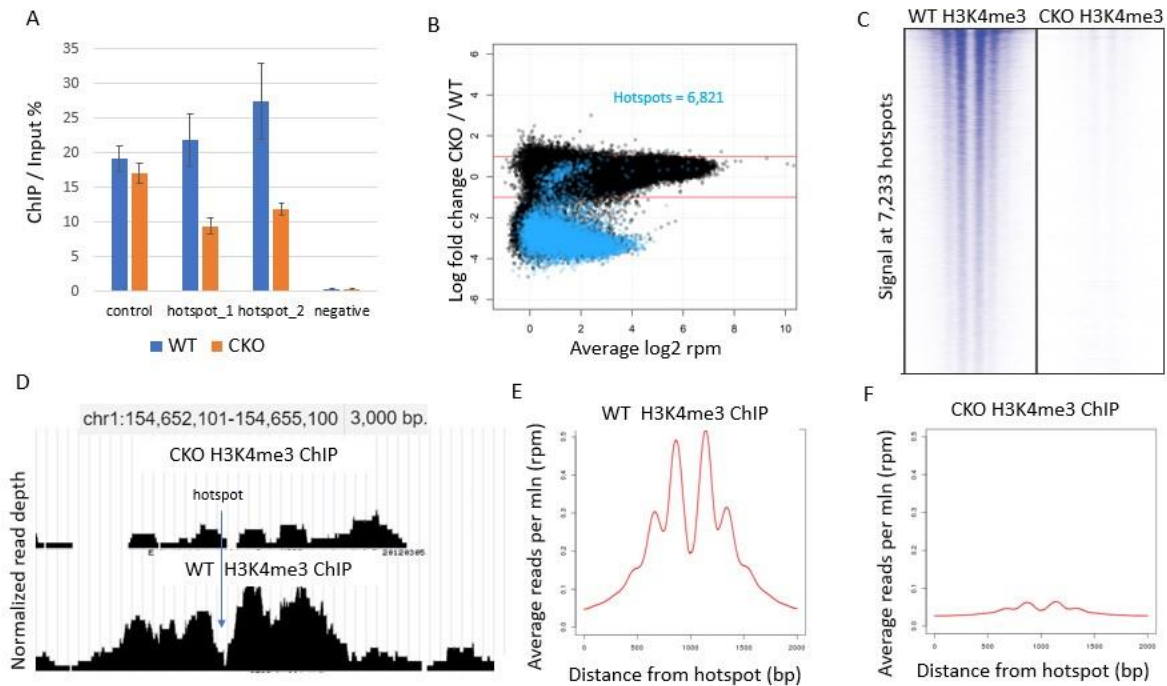
Figure 9. H3K4me3 ChIP-Seq on *Hells* CKO shows loss of signal at hotspots. (A) H3K4me3 ChIP qPCR showing loss of enrichment at hotspots in the *Hells* CKO. (B) MA plot of differential H3K4me3 peaks between *Hells* CKO and wildtype, with PRDM9 ChIP-defined hotspots colored in blue (C) Heatmap of H3K4me3 read counts across a 2 kb region centered over 7,233 PRDM9 ChIP-defined hotspots. (D) Read depths across a 3kb region at a single hotspot. (E) Density plot of wildtype H3K4me3 reads across all hotspots. (F) Density plot of *Hells* CKO H3K4me3 reads across all hotspots.

H3K4me3 peaks showed similar trimethylation patterns in both the *Hells* CKO and *Hells* het. However, H3K4me3 peaks annotated as hotspots, 98% of them are reduced in *Hells* CKO compared to wildtype. A heatmap of read depths across all hotspots shows comprehensive depletion in the *Hells* CKO (Fig. 9C).

I visualized the read coverage profile for *Hells* CKO and wildtype H3K4me3 ChIP using the UCSC browser and found significant decrease, loss, or disorganization (nucleosomes not remodeled around PRDM9 motif) of the H3K4me3 signal at hotspots in the *Hells* CKO (Fig. 9D). Likewise, aggregate signal from all hotspot regions show almost complete loss of H3K4me3 organization at hotspots in the *Hells* CKO (Fig. 9F) compared to the control.

There were 11,070 peaks that were greater than 2 logFC reduced in the *Hells* CKO, which I hypothesized were hotspots. However, for most of my analyses I used a smaller hotspot peak set of 7,233 hotspots identified by PRDM9 ChIP binding that have PRDM9 motifs identified within them, a conservative measure of a true hotspot. I found 6,821 of these hotspots in my H3K4me3 ChIP data, of which 6,677 (98%) are reduced in the *Hells* CKO, but that leaves many H3K4me3 peaks that are significantly reduced in the *Hells* CKO unannotated. I wanted to annotate more of the 11,070 H3K4me3 peaks that are reduced in *Hells* CKO as hotspots by using a larger list of hotspots from Spruce et al, 2020. Here we defined hotspots based on their unique chromatin signature using ChromHMM analysis (Ernst and Kellis, 2017) to assign chromatin-state signatures at genomic regions by using a multivariate hidden Markov model (HMM) that explicitly models the combinatorial presence or absence of each mark. Using the larger pool of 17,843 hotspots identified by ChromHMM I annotated 11,125 H3K4me3 peaks as hotspots, of which 11,121 were reduced in the *Hells* CKO. These results strongly support my hypothesis that the peaks reduced in *Hells* CKO are hotspots. There were still ~919 peaks that are reduced greater than 2 logFC in the *Hells* CKO but are not annotated. These unannotated peaks represent 1.2% of the total 71,877 peaks called in the H3K4me3 ChIP data, compared to 9.5% - 15.5% that are annotated as

hotspots. It is unclear if the remaining peaks are unannotated hotspots or something else, but the

overwhelming majority of the peaks that decrease or disappear in the *Hells* CKO are hotspots.

CHAPTER 9

HELLS LOCALIZES TO HOTSPOTS DURING MEIOSIS

9.1. ChIP for HELLS shows binding at hotspots

The molecular and morphological phenotype of the *Hells* CKO strongly indicate that HELLS in necessary

for initiation of meiotic recombination. To provide a functional link between HELLS and hotspots I

wanted to test whether HELLS bound on chromatin at hotspots by performing HELLS ChIP. Before I could

do this, we needed to develop a protocol for ChIP on a protein that may not directly bind DNA (HELLS)

and interacts through another protein partner with DNA-binding capabilities (PRDM9).

9.1.1. Dual cross-linking spermatocytes to preserve HELLS interactions with chromatin

For the HELLS ChIP, I modified our cross-linking to include a two-step cross-linking process to target both

protein-DNA interaction and protein-protein interactions. Since HELLS has no DNA binding domain of its

own, we speculated that it could only be recruited to chromatin through another DNA-binding protein,

like PRDM9. Therefore, in order to recover chromatin from the HELLS ChIP, we would need to perform

dual cross-linking to first cross-link protein-protein interactions (e.g. HELLS to PRDM9), then crosslink

protein to DNA (e.g. PRDM9 to chromatin). Protein-protein interactions were cross-linked starting from

12 dpp spermatocytes using fresh disuccinimidyl glutarate (DSG) at 2 mM in PBS with 1 mM $MgCl_2$ for 30

min at room temperature with rotation. Next, I added 1% fresh paraformaldehyde to cross-link protein-

DNA interactions and rotated another 5 min before stopping the reaction by adding 125 mM glycine to

quench unreacted formaldehyde.

9.1.2. Sonication of chromatin to fragment the DNA before ChIP

In order to make DNA compatible with high-throughput sequencing and improve enrichment over

background, longer pieces of chromatin need to be fragmented. Since HELLS is expected to create

nucleosome-free DNA, I could not use MNase-digestion without potentially digesting away the DNA of

interest. Instead, I used sonication to break the DNA through ultrasound waves that can cleave

hydrogen bonds and shear DNA while leaving proteins unharmed. There are four considerations for

successful sonication: time, SDS concentration, heat, and intensity. Sonication sheared DNA by vibration,

which also produces heat that can damage proteins. To control the heat, the Covaris sonication

equipment maintains the sample in a temperature-controlled water bath. The Covaris has adjustable

setting for intensity of sonication and length of time of sonication bursts. We followed the

manufacturer's recommendation for intensity: Duty 2%, Peak Incident Power 105 watts, Cycles per burst

200.

      For length of time I tested three conditions, 5min, 10min, and 15min. I also tested two

concentrations of SDS. SDS increases sonication efficiency and chromatin yield, but downstream is

strongly interferes with ChIP binding efficiency. Therefore, I tested the recommended concentration for

sonication, which was 1% SDS, and the recommended maximum concentration for ChIP, which was

0.1%. For my sonication test I hypotonically lysed dual cross-linked cells and aliquoted them into six 1

million cell aliquots in 130 µl for sonication, the maximum cell concentration recommended for

sonication. I tested three different time conditions using two different SDS conditions. I took the

resulting sonicated DNA, reversed cross-links and digested protein, and then ran the DNA fragments on

a gel to analyze fragment size (Fig. 11A). The low SDS condition lost most chromatin, and what was left

appeared unsonicated. The high SDS concentration showed good sonication in all three times. My aim

was to recover DNA fragments around 300-600bp, which appeared to be in between 10 and 15 min.

Based on these results I decided to sonicate my samples for 12min.

9.1.3. Dialysis of chromatin to reduce SDS

The optimum concentration of SDS in a ChIP assay is 0.1%, and my sonicated chromatin was at 1%. If I

diluted my chromatin to 0.1% in ChIP buffer, the chromatin would be too dilute to bind efficiently in the

ChIP. To reduce SDS concentration without increasing volume, I dialyzed the sonicated chromatin

overnight in low-SDS dialysis buffer (0.01% SDS, 1.1% triton X100, 1.2 mM EDTA, 16.7 mM Tris-Cl pH 8.0,

167 mM NaCl) using Slide-A-Lyzer Mini Dialysis Devices, 10K MWCO, that will allow small molecules like

SDS to pass through and equilibrate but keep the chromatin inside. Previous work in our lab (Baker et

al., 2015a) determined that optimal dialysis requires an initial 1:1 dilution of sonicated chromatin to

dialysis buffer and then two >8hr incubations in successive 1L dialysis buffer baths at 4°.

9.1.4. Performing IP for HELLS and recover co-bound DNA

The HELLS ChIP required more gentle buffer conditions than the H3K4me3 ChIP. Instead of using RIPA

buffer, which is relatively harsh buffer with high detergents (50 mM Tris pH 8.0, 150 mM NaCl, 1.0% NP-

40, 0.5% Na deoxycholate, 0.1% SDS), I switched to using dialysis buffer (16.7 mM Tris-Cl pH 8.0, 167

mM NaCl, 1.1% triton X100, 1.2 mM EDTA, 0.01% SDS) that is less likely to disrupt weak antibody binding

but is also more prone to non-specific background signal. As blocking agents, I used 50mg/mL BSA and

2mg/mL polyglutamate. Polyglutamate is water-soluble synthetic polymer that mimics DNA in its ability

to neutrally coat but not bind the components in the ChIP assay. We made this change after finding

evidence of non-mouse DNA sequence in our ChIPs using the gentle buffer, which we concluded came

from Herring sperm DNA that was able to bind to beads in the less stringent buffer conditions.

For the HELLS ChIP, I performed three triplicate 500 µl ChIP reactions containing 25 µl Protein G

beads, 4 µl anti-HELLS antibody (Millipore, ADB41), and chromatin from 3 million dual cross-linked cells.

I also performed a negative control ChIP using anti-IgG antibody that does not bind protein and

represents non-specific binding. The next day I performed gentle washing with low salt, high salt, and

LiCl washes described in the PRDM9 ChIP protocol in Baker et al., 2015a. To purify the DNA from the

HELLS ChIP, I used a three-step process to elute from the beads, reverse cross-links, and digest protein,

also described in the PRDM9 ChIP protocol in Baker et al., 2015a. The yield from the HELLS ChIP after

purifying the ChIP DNA with the Zymo DNA Clean & Concentrator-5 (Zymo # D4013) kit was significantly

lower than the yield from H3K4me3 ChIP, about 3.2 ng total ChIP DNA for HELLS compared to 10 ng ChIP

DNA for H3K4me3.

9.1.5. qPCR on HELLS ChIP DNA shows enrichment at hotspots

I chose to use B6$^{CAST-KI}$ mice for the HELLS ChIP because my previous HELLS ChIPs on B6 mice were very

low yield, and though the qPCR showed the results we expected (HELLS enrichment at hotspots), it was

low efficiency. We have seen poor outcomes from B6 in the PRDM9 ChIP previously, and the resolution

was to use B6$^{CAST-KI}$ with the stronger allele of PRDM9 to capture more fragments. I compared HELLS ChIP

on replicates of B6 vs B6$^{CAST-KI}$ spermatocytes and saw that I could double my yield (3.2ng vs 1.7ng) and

relative enrichments at hotspots (3.1-fold vs 1.6-fold over control) by using B6$^{CAST-KI}$ vs B6. I performed

HELLS ChIP on three replicates of B6$^{CAST-KI}$ spermatocytes followed by qPCR to measure enrichment at

hotspots and found HELLS was preferentially bound at PRDM9^{Cst} hotspots compared to PRDM9^{Dom2}

hotspots (Fig. 11B). Compared to the replicates of HELLS ChIP on B6 spermatocytes, the enrichment at

hotspots in HELLS ChIPs was specific to binding of PRDM9 and allele-specific hotspots.

9.1.6. Sequence analysis of HELLS ChIP shows enrichment at hotspots

We sequenced our two best HELLS ChIP replicates based on yield and qPCR enrichments and one IgG

ChIP control to a depth of ~50mln reads per sample, single-end reads and I performed mapping, peak-

calling, and normalization on the data as before. I checked the correlation between HELLS ChIP

replicates as a quality control and they showed good agreement (Pearson's r=0.91). Despite >94% reads

mapping, there were only 2,675 HELLS ChIP peaks called at e-5 by MACS, which is low and only about

1,000 more than the number of peaks called on IgG ChIP negative control

I uploaded bedgraph files containing normalized reads from the HELLS ChIP to the UCSC browser

to look at read pile-ups at HELLS ChIP peaks. There was enrichment HELLS reads at hotspots, though not

every hotspot showed enrichment, perhaps as a consequence of poor ChIP efficiency. The hotspots with

HELLS signal showed read pile ups centered over the PRDM9 binding motif, similar to ATAC peaks and in

between H3K4me3 nucleosome peaks (Fig. 11C). It is interesting that although HELLS is a nucleosome

remodeler, we did not see reads at nucleosomes flanking hotspots, like we see in H3K4me3 ChIP.
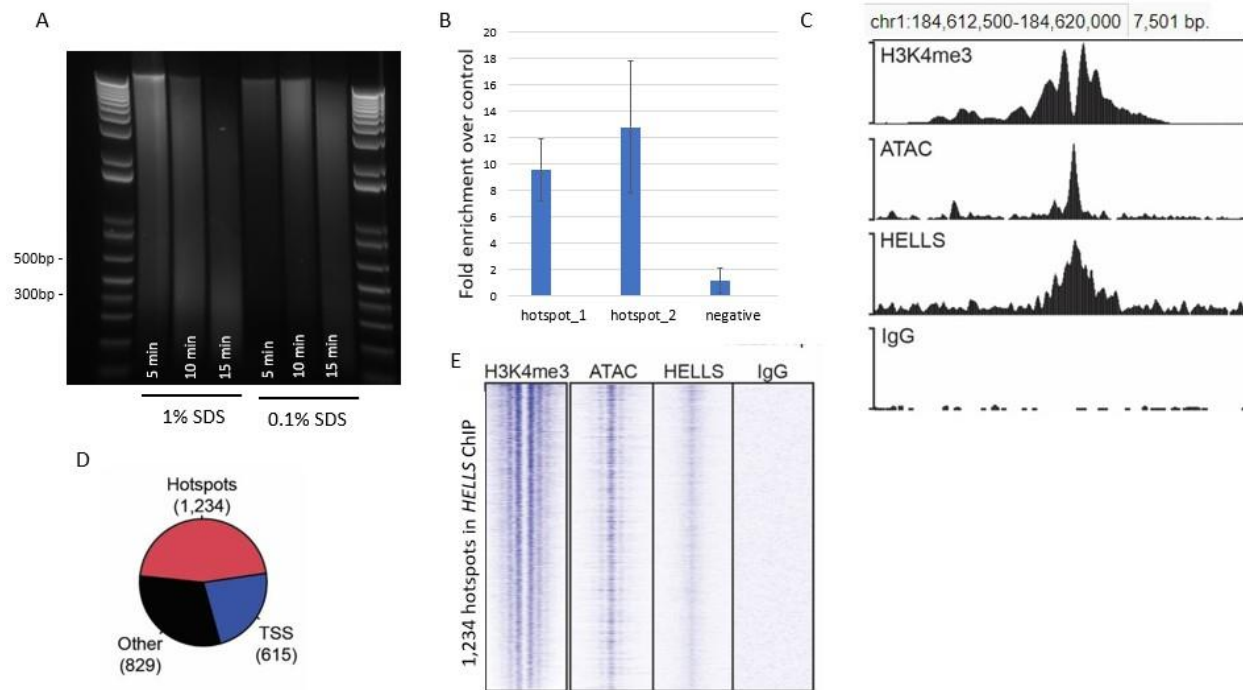
Figure 10. HELLS ChIP shows binding at hotspots. (A) Gel of sonication fragments from two SDS concentrations (1% and 0.1%) and three time lengths (5, 10, and 15min). (B) *HELLS* ChIP qPCR for hotspots (C) Bedgraphs of read pile-ups for H3K4me3 ChIP, ATAC, HELLS ChIP, and IgG ChIP at one hotspot. (D) Pie chart showing annotations of peaks recovered in *HELLS* ChIP. (E) Heatmap of *HELLS* ChIP read counts across a 2 kb region centered over 1,234 individual hotspots.

I annotated the peaks as hotspots, TSS's, or other and found that 46% of the 2,675 HELLS ChIP peaks overlapped recombination hotspots, compared to 23% of the peaks at TSS's (promoters) and 31% that were "other" (Fig. 11D). This result finding hotspots were the largest proportion of peaks supports the hypothesis that HELLS plays a role at hotspots. We also saw HELLS peaks at promoters, consistent with other findings that HELLS can act as an epigenetic regulator at chromatin regulatory regions like promoters and enhancers (Ren et al., 2019; Han et al., 2017; von Eyss et al., 2011). In contrast, the IgG negative control ChIP had 1,777 peaks, similar to one of our HELLS ChIP replicates, however only 2% of those peaks overlapped with HELLS peaks and there was no enrichment for IgG at hotspots, proving that the HELLS ChIP was truly enriched for at hotspots despite having low sensitivity.

To examine enrichment across all hotspots, I plotted read counts across a 2kb region centered over the 1,234 hotspots recovered from the HELLS ChIP and compared it to ATAC, H3K4me3, and IgG heatmaps. The read depths from HELLS ChIP at hotspots was correlated with read depths from H3K4me3 ChIP and ATAC, showing that the strongest hotspots had the strongest signal in all three marks (Fig. 11D). In contrast, the IgG ChIP I performed on the same spermatocytes at the same time showed no enrichment of any hotspots. These results show that HELLS is bound at hotspots.

CHAPTER 10

LOSS OF HOTSPOTS IS NOT DUE TO PRDM9 BINDING AT OTHER SITES IN THE *HELLS* CKO

10.1. No evidence of novel H3K4me3 peaks in *Hells* CKO

The loss of H3K4me3 was unexpected, since presumably if PRDM9 is still present in the Hell CKO it

should be able to bind and trimethylated hotspots. We questioned whether in the *Hells* CKO the

hotspots are relocated to other genomic sites by alternative binding by PRDM9 to other sites in the

absence of HELLS. To investigate potential non-canonical binding of PRDM9 in the *Hells* CKO, I looked

more closely at our H3K4me3 ChIP for emergence of increased trimethylated at sites in the *Hells* CKO

that could indicate binding by PRDM9. There are very few peaks that are more trimethylated in the *Hells*

CKO. The changes in H3K4me3 peaks that are differentially trimethylated in the *Hells* CKO are very

uneven, only 7 peaks increase greater than 2 logFC in the *Hells* CKO compared to 5,513 hotspot peaks

that decrease greater than -2 logFC reduced in the *Hells* CKO (Fig. 9B). PRDM9 can deposit H3K4me3 *in*

vitro on its own without binding partners (Chen et al., 2020), so we would expect to see a new cohort of

H3K4me3 peaks emerge if PRDM9 was binding elsewhere, therefore the H3K4me3 ChIP data shows

strong evidence against relocation of hotspots.

10.2. PRDM9 is not sequestered to TEs in absence of HELLS

There is a potential role of *Hells* before meiosis occurs in *de novo* methylation of the germline.

Transposable elements (TEs) become demethylated in fetal germ cells and must be silenced by *de novo*

methyltransferases or piRNAs. In spermatogenesis, *de novo* methylation occurs in fetal

prospermatogonia and is completed by E16.5. Interestingly, defects in *de novo* methylation also cause

defects in meiotic progression and pachytene arrest, a phenotype similar to *Hells* CKO and Prdm9 KO. De

La Fuente et al., 2006, found evidence of demethylation of transposable elements in *Hells* KO oocytes in

addition to meiotic arrest at pachytene stage due to unresolved DSBs (De La Fuente et al., 2006). In our

conditional mouse model, *Stra8*-iCre is first observed in 3 dpp testes (Sadate-Ngatchou et al., 2008), so

Hells is expressed normally when de novo methylation occurs and the phenotype we see is not due to

demethylation of TEs.

However, it did make us wonder if TEs have the potential to provide binding sites for the zinc-

finger in PRDM9, given that zinc-finger proteins evolved primarily to recognize and silence TEs. In the

Hells CKO we observe no evidence of PRDM9 binding at hotspots in the *Hells* CKO but we know PRDM9

protein is present, and we also see no novel genomic locations showing PRDM9 binding based on

increased H3K4me3, therefore we wondered if PRDM9 could be sequestered to unmapped, repetitive

TEs in the absence of HELLS.

To investigate potential PRDM9 binding at TEs, I used SalmonTE to map TE reads from the

H3K4me3 ChIP data. Most mapping algorithms exclude TE sequences because of their repetitive nature

and the accompanying difficulty of assigning reads to an individual TE or location. SalmonTE uses an

expectation-maximization (EM) algorithm and light-weight mapping approach to assign reads to

consensus sequences of TE families (Jeong et al., 2018). I ran SalmonTE to search for reads in the

H3K4me3 ChIPs that show sequence similarity to 275 consensus sequences of known mouse TEs. As

control datasets, I included a *Prdm9* KO H3K4me3 ChIP as a negative control, reasoning if we do see

increase in H3K4me3 at TEs in the *Hells* CKO, that effect should be ablated in the *Prdm9* KO. I also also

analyzed previously published H3K4me3 ChIP-seq data from a *Dnmt3L* (Brick et al., 2018). Since *de nov*o

methylation TEs is dependent on *Dnmt3L*, we should see an increase in H3K4me3 at TEs in *Dnmt3L*

knock out if the two are linked.

After normalization and differential analysis, 0 TEs had an increase in *Hells* CKO greater than 2

logFC of the 275 TE families queried. There were 23 (8%) TEs that showed a logFC > 1, which seems too

low of a difference to be a true change in PRDM9 binding, considering that there are many individual TEs

and therefore many potential binding sites represented by each SalmonTE consensus entry. I normalized

TE read counts to data from the *Hells* heterozygous animals and plotted a heatmap to show the change

in H3K4me3 in the *Hells* CKO, *Prdm9* knock out, and *Dnmt3L* knock out (Fig. 11A). After filtering for TEs

that had a >1 logFC in the *Hells* CKO, I saw the TEs that appeared to be differentially trimethylated in the

Hells CKO are similarly trimethylated in the *Prdm9* knock out, suggesting the increase in H3K4me3 at

these TEs is not specific to PRDM9 binding but may be due to differences in the meiotic stage in the

infertile phenotypes. The *Dnmt3L* knock out showed stronger differences as expected, but perhaps

unexpectedly for this subset of TEs there was less H3K4me3 at TEs in the *Dnmt3L* knock out compared

to the *Hells* het, so clearly our *Hells* CKO phenotype is not reminiscent of failure of *de novo* methylation.

In conclusion, the changes in H3K4me3 at TEs in the *Hells* CKO are not indicative of PRDM9

sequestration to TEs.
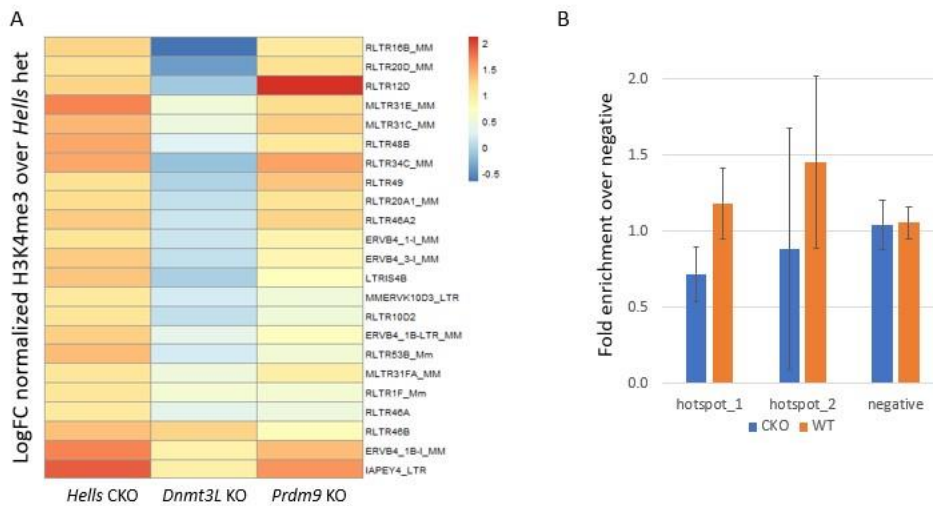
Figure 11. PRDM9 is not binding at non-hotspot locations. (A) Heatmap of log2 fold change
compared to *Hells* het in H3K4me3 read counts mapping to TEs. (B) PRDM9 ChIP qPCR for hotspots

10.3. PRDM9-ChIP on *Hells* CKO does not show binding of PRDM9

Although there was no evidence of alternative PRDM9 binding in the H3K4me3 data, I queried PRDM9

directly by performing PRDM9 ChIP on three replicates of *Hells* CKO 12dpp spermatocytes compared to

CKO het controls. I followed the ChIP protocol described for HELLS ChIP, but with single cross-linking

instead of dual cross-linking (though I tested PRDM9 ChIP on dual cross-linked cells and saw no

difference in outcome). As mentioned above, the PRDM9 ChIP on B6 with the native $Prdm9^{Dom2}$ allele is

low efficiency, and the total recovery from all $Hells$ PRDM9 ChIP samples was low. I performed qPCR for

two $Prdm9^{Dom2}$ hotspots (positives) and one $PRDM9^{Cst}$ hotspot (negative control), along with a genomic

location with no PRDM9 binding that should be equal in all samples to use for normalization. I saw

reduced binding at B6 hotspots in the $Hells$ CKO, supporting previous observations that PRDM9 does not

bind hotspots in the absence of HELLS (Fig. 11B). It is hard to say conclusively that PRDM9 is not binding

at all in the $Hells$ CKO, since enrichment is so low for this ChIP and we see background ChIP signal at

supposedly negative locations. Overall, the PRDM9 ChIP suggests that PRDM9 binding at hotspots is

impaired in $Hells$ CKO, but more optimization would be needed to be absolutely certain.

CHAPTER 11

CONCLUSIONS

11.1. The chromatin remodeler HELLS creates open chromatin at hotspots

In total my thesis work demonstrated 1. activated hotspots are regions of open chromatin, 2. open

chromatin at hotspots is dependent on PRDM9 and independent of SPO11, 3. HELLS forms a complex

with PRDM9 in testis, 4. Loss of *Hells* prevents chromatin reorganization at hotspots leading to failure of

homologous recombination. These results advance the field of meiosis by showing a role for open

chromatin at hotspots to allow homologous recombination, and by describing HELLS as a binding

partner of PRDM9 that is necessary for PRDM9 function.

In my first set of experiments, I showed that activated hotspots are locations of open chromatin

by using ATAC-Seq to query nucleosome-depleted regions in the genome. Open chromatin at hotspots is

dependent on PRDM9 binding as shown by allelic differences in open chromatin sites and is

independent of DSBs as shown by presence open chromatin at hotspots in the *Spo11* KO where no DSBs

have occurred. In my next set of experiments, I introduced the chromatin remodeler HELLS. I used coIP

to show that HELLS and PRDM9 form a protein complex and performed IHC on testis sections to show

protein co-expression of HELLS and PRDM9 in L/Z stage cells.

To investigate HELLS in male meiosis, I imported a mouse with a conditional knock out allele of

Hells and mated it to *Stra8*-iCre to knock out *Hells* expression in spermatocytes starting at L/Z stage in

meiosis. The *Hells* CKO spermatocytes showed meiotic arrest and apoptosis at pachytene stage due to

unresolved DSBs. The *Hells* CKO spermatocytes had no open chromatin at hotspots but normal open

chromatin everywhere else, indicating that in the absence of HELLS, nucleosomes are not remodeled at

hotspots. Concurrently in the *Hells* CKO there was a lack of H3K4me3 histone modifications at hotspots

suggesting loss of PRDM9 binding, however there is no evidence of PRDM9 binding at novel locations in

the absence of HELLS by H3K4me3 ChIP. Finally, I confirmed that HELLS is preferentially bound at

hotspots by performing a HELLS ChIP in normal spermatocytes, verifying that HELLS is part of the protein

complex recruited by PRDM9 to remodel chromatin at hotspots and allow homologous recombination in

meiosis.

11.2. The role of HELLS in PRDM9 function is unclear

One question that came up repeatedly during this project is what is happening to PRDM9 in the *Hells*

CKO. I assumed in the *Hells* CKO that PRDM9 would still bind hotspots and trimethylate H3K4 however it

appears that PRDM9 does not bind at all in the absence of HELLS. In fact, we performed an shRNA knock

down of *Hells* in cell culture and confirmed loss of PRDM9 binding without HELLS as a co-factor. The lack

of binding is not an issue with protein quantity: we see plenty of PRDM9 protein in the *Hells* CKO, even if

protein levels are slightly reduced compared to wildtype. However, impaired binding could be explained

if HELLS is a critical binding partner for PRDM9 to be active. There are many protein complexes that

require all protein components to be present before the complex is competent and active. However,

previous papers have shown PRDM9 has methylase activity in purified in vitro assays (Powers et al.,

2016; Chen et al., 2020), so it appears that HELLS protein is not necessary for PRDM9 to methylate

purified histones. HELLS could potentially play a role in allowing PRDM9 to bind DNA, although again

other studies have used PRDM9 *in vitro* and showed that PRDM9 can bind naked DNA without additional

protein partners (Billings et al.; 2013; Grey et al., 2011).

Another possibility I explored was the potential that PRDM9 was binding elsewhere. I tried

several assays (H3K4me3 and PRDM9 ChIP) and looked at possible binding to TEs, but there was no

evidence to support non-canonical binding. I could not even recover enough DNA from PRDM9 ChIP in

Hells CKO to sequence, suggesting that PRDM9 is not binding anywhere in the genome. We also

performed a DMC1 ChIP on Hells CKO with the help of a collaborator (Hui Tian) from the Paigen Lab to

look for non-canonical DSB sites, but the data returned promoter regions and other H3K4me3-enriched

loci exactly like *Prdm9* KO with no sign of alternative sites. Together these observations suggest that co-binding with HELLS is necessary for PRDM9 to be competent to interact with intact chromatin.

In total, this work describes PRDM9 as a pioneer factor that can recognize binding motifs in condensed chromatin and deposit open chromatin histone marks to encourage other factors to bind. Most described pioneer factors behave this way, where the pioneer factor binds first and then open chromatin and histone marks appear before DNA accessibility (Iwafuchi-Doi and Zaret; 2014; King and Klose, 2017). In our case the role seems to be reversed: seemingly PRDM9 cannot trimethylate histones without DNA being accessible. The previous studies showing PRDM9 binding to DNA in vitro used naked DNA, i.e., DNA that is already accessible and is parsimonious with this explanation. In fact, previous studies have already shown that chromatin compaction negatively impacts PRDM9 binding efficiency (Walker et al., 2015). PRDM9 shows transient binding and a short residency time at hotspots (Spruce et al., 2020), suggesting a weak association that could be easily lost without stabilization by other factors. Future work will be required to test whether HELLS may act to stabilize PRDM9 binding before activation of hotspots can occur.

REFERENCES

Baker CL, Kajita S, Walker M, Petkov PM, Paigen K. 2014. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res* **24**: 724-732.

Baker CL, Kajita S, Walker M, Saxl RL, Raghupathy N, Choi K, Petkov PM, Paigen K. 2015a. PRDM9 Drives Evolutionary Erosion of Hotspots in Mus musculus through Haplotype-Specific Initiation of Meiotic Recombination. *PLoS Genet* **11**: e1004916.

Baker CL, Petkova P, Walker M, Flachs P, Mihola O, Trachtulec Z, Petkov PM, Paigen K. 2015b. Multimer Formation Explains Allelic Suppression of PRDM9 Recombination Hotspots. *PLoS Genet* **11**: e1005512.

Baker CL, Walker M, Arat S, Ananda G, Petkova P, Powers NR, Tian H, Spruce C, Ji B, Rausch D et al. 2019. Tissue-Specific Trans Regulation of the Mouse Epigenome. *Genetics* **211**: 831-845.

Ball RL, Fujiwara Y, Sun F, Hu J, Hibbs MA, Handel MA, Carter GW. 2016. Regulatory complexity revealed by integrated cytological and RNA-seq analyses of meiotic substages in mouse spermatocytes. *BMC Genomics* **17**: 628.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**: 836-840.

Baudat F, Imai Y, de Massy B. 2013. Meiotic recombination in mammals: localization and regulation. *Nat Rev Genet* **14**: 794-806.

Baudat F, Manova K, Yuen JP, Jasin M, Keeney S. 2000. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol Cell* **6**: 989-998.

Becker PB, Hörz W. ATP-dependent nucleosome remodeling. Annu Rev Biochem. 2002;71:247-73. doi: 10.1146/annurev.biochem.71.110601.135400. Epub 2001 Nov 9. PMID: 12045097.

Billings T, Parvanov ED, Baker CL, Walker M, Paigen K, Petkov PM. DNA binding specificities of the long zinc-finger recombination protein PRDM9. Genome Biol. 2013 Apr 24;14(4):R35.

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**: 642-645.

Buard J, Barthes P, Grey C, de Massy B. 2009. Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *EMBO J* **28**: 2616-2624.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-1218.

Burrage J, Termanis A, Geissner A, Myant K, Gordon K, Stancheva I. The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. J Cell Sci. 2012 Nov 15;125(Pt 22):5524-34.

Chen Y, Lyu R, Rong B, Zheng Y, Lin Z, Dai R, Zhang X, Xie N, Wang S, Tang F, Lan F, Tong MH. Refined spatial temporal epigenomic profiling reveals intrinsic connection between PRDM9-mediated H3K4me3 and the fate of double-stranded breaks. Cell Res. 2020 Mar;30(3):256-268.

Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, Majeti R, Chang HY. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. 2016 Oct;48(10):1193-203. doi: 10.1038/ng.3646. Epub 2016 Aug 15. PMID: 27526324; PMCID: PMC5042844.

De La Fuente R, Baumann C, Fan T, Schmidtmann A, Dobrinski I, Muegge K. Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. Nat Cell Biol. 2006 Dec;8(12):1448-54.

Diagouraga B, Clément JAJ, Duret L, Kadlec J, de Massy B, Baudat F. PRDM9 Methyltransferase Activity Is Essential for Meiotic DNA Double-Strand Break Formation at Its Binding Sites. Mol Cell. 2018 Mar 1;69(5):853-865.e6.

Eram MS, Bustos SP, Lima-Fernandes E, Siarheyeva A, Senisterra G, Hajian T, Chau I, Duan S, Wu H, Dombrovski L et al. 2014. Trimethylation of histone H3 lysine 36 by human methyltransferase PRDM9 protein. *J Biol Chem* **289**: 12177-12188.

Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc. 2017 Dec;12(12):2478-2492.

Essers J, Hendriks RW, Swagemakers SM, Troelstra C, de Wit J, Bootsma D, Hoeijmakers JH, Kanaar R. Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination. Cell. 1997 Apr 18;89(2):195-204.

Fan T, Yan Q, Huang J, Austin S, Cho E, Ferris D, Muegge K. Lsh-deficient murine embryonal fibroblasts show reduced proliferation with signs of abnormal mitosis. Cancer Res. 2003 Aug 1;63(15):4677-83. PMID: 12907649.

Geiman TM, Muegge K. Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes. Proc Natl Acad Sci U S A. 2000 Apr 25;97(9):4772-7.

Geiman TM, Tessarollo L, Anver MR, Kopp JB, Ward JM, Muegge K. Lsh, a SNF2 family member, is required for normal murine development. Biochim Biophys Acta. 2001 May 3;1526(2):211-20.

Grey C, Barthès P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLoS Biol. 2011 Oct;9(10):e1001176.

Grey C, Clement JA, Buard J, Leblanc B, Gut I, Gut M, Duret L, de Massy B. 2017. In vivo binding of PRDM9 reveals interactions with noncanonical genomic sites. *Genome Res* **27**: 580-590.

Iwafuchi-Doi M, Zaret KS. 2014. Pioneer transcription factors in cell reprogramming. Genes Dev 28: 2679-2692.

Jarvis CD, Geiman T, Vila-Storm MP, Osipovich O, Akella U, Candeias S, Nathan I, Durum SK, Muegge K. A novel putative helicase produced in early murine lymphocytes. Gene. 1996 Mar 9;169(2):203-7.

Jeong HH, Yalamanchili HK, Guo C, Shulman JM, Liu Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. Pac Symp Biocomput. 2018;23:168-179. PMID: 29218879.

Jung M, Wells D, Rusch J, Ahmad S, Marchini J, Myers SR, Conrad DF. Unified single-cell analysis of testis gene regulation and pathology in five mouse strains. Elife. 2019 Jun 25;8:e43966.

Han Y, Ren J, Lee E, Xu X, Yu W, Muegge K. Lsh/HELLS regulates self-renewal/proliferation of neural stem/progenitor cells. Sci Rep. 2017 Apr 25;7(1):1136.

Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438**: 374-378.

Hohenauer T, Moore AW. The Prdm family: expanding roles in stem cells and development. Development. 2012 Jul;139(13):2267-82.

Hunter N. Meiotic Recombination: The Essence of Heredity. Cold Spring Harb Perspect Biol. 2015 Oct 28;7(12):a016618.

Imai Y, Baudat F, Taillepierre M, Stanzione M, Toth A, de Massy B. The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. Chromosoma. 2017 Dec;126(6):681-695.

Jenness C, Giunta S, Müller MM, Kimura H, Muir TW, Funabiki H. HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proc Natl Acad Sci U S A. 2018 Jan 30;115(5):E876-E885.

King HW, Klose RJ. 2017. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. eLife 6

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.

Lowy E (2020). CoverageView: Coverage visualization package for R. R package version 1.28.0.

Mzoughi S, Tan YX, Low D, Guccione E. The role of PRDMs in cancer: one family, two sides. Curr Opin Genet Dev. 2016 Feb;36:83-91. doi: 10.1016/j.gde.2016.03.009. Epub 2016 May 7. PMID: 27153352.

Papamichos-Chronakis M, Watanabe S, Rando OJ, Peterson CL. 2011. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**: 200-213.

Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**: 835.

Parvanov ED, Tian H, Billings T, Saxl RL, Spruce C, Aithal R, Krejci L, Paigen K, Petkov PM. 2017. PRDM9 interactions with other proteins provide a link between recombination hotspots and the chromosomal axis in meiosis. *Mol Biol Cell* **28**: 488-499.

Powers NR, Parvanov ED, Baker CL, Walker M, Petkov PM, Paigen K. The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. PLoS Genet. 2016 Jun 30;12(6):e1006146.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2.

Ren J, Briones V, Barbour S, Yu W, Han Y, Terashima M, Muegge K. The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. Nucleic Acids Res. 2015 Feb 18;43(3):1444-55.

Reyes JC, Barra J, Muchardt C, Camus A, Babinet C, Yaniv M. Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha). EMBO J. 1998 Dec 1;17(23):6979-91.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139-40.

Romanienko PJ, Camerini-Otero RD. The mouse Spo11 gene is required for meiotic chromosome synapsis. Mol Cell. 2000 Nov;6(5):975-87.

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, Ali S, Chin SF, Palmieri C, Caldas C, Carroll JS. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012 Jan 4;481(7381):389-93.

Sadate-Ngatchou PI, Payne CJ, Dearth AT, Braun RE. Cre recombinase activity specific to postnatal, premeiotic male germ cells in transgenic mice. Genesis. 2008 Dec;46(12):738-42.

Serber DW, Runge JS, Menon DU, Magnuson T. 2016. The Mouse INO80 Chromatin-Remodeling Complex Is an Essential Meiotic Factor for Spermatogenesis. *Biol Reprod* **94**: 8.

Shenkar R, Shen MH, Arnheim N. 1991. DNase I-hypersensitive sites and transcription factor-binding motifs within the mouse E beta meiotic recombination hot spot. *Mol Cell Biol* **11**: 1813-1819.

Spruce C, Dlamini S, Ananda G, Bronkema N, Tian H, Paigen K, Carter GW, Baker CL. HELLS and PRDM9 form a pioneer complex to open chromatin at meiotic recombination hot spots. Genes Dev. 2020 Mar 1;34(5-6):398-412.

Sun F, Fujiwara Y, Reinholdt LG, Hu J, Saxl RL, Baker CL, Petkov PM, Paigen K, Handel MA. Nuclear localization of PRDM9 and its role in meiotic chromatin modifications and homologous synapsis. Chromosoma. 2015 Sep;124(3):397-415.

Sun LQ, Arceci RJ. Altered epigenetic patterning leading to replicative senescence and reduced longevity. A role of a novel SNF2 factor, PASG. Cell Cycle. 2005 Jan;4(1):3-5.

Termanis A, Torrea N, Culley J, Kerr A, Ramsahoye B, Stancheva I. 2016. The SNF2 family ATPase LSH promotes cell-autonomous de novo DNA methylation in somatic cells. *Nucleic Acids Res* **44**: 7592-7604.

Tian H, Billings T, Petkov PM. CXXC1 is not essential for normal DNA double-strand break formation and meiotic recombination in mouse. PLoS Genet. 2018 Oct 26;14(10):e1007657.

Unoki M, Funabiki H, Velasco G, Francastel C, Sasaki H. CDCA7 and HELLS mutations undermine nonhomologous end joining in centromeric instability syndrome. J Clin Invest. 2019 Jan 2;129(1):78-92.

von Eyss B, Maaskola J, Memczak S, Mollmann K, Schuetz A, Loddenkemper C, Tanh MD, Otto A, Muegge K, Heinemann U et al. 2012. The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation. *EMBO J* **31**: 972-985.

Walker M, Billings T, Baker CL, Powers N, Tian H, Saxl RL, Choi K, Hibbs MA, Carter GW, Handel MA, Paigen K, Petkov PM. Affinity-seq detects genome-wide PRDM9 binding sites and reveals the impact of prior chromatin modifications on mammalian recombination hotspot usage. Epigenetics Chromatin. 2015 Sep 7;8:31.

Wesoly J, Agarwal S, Sigurdsson S, Bussen W, Van Komen S, Qin J, van Steeg H, van Benthem J, Wassenaar E, Baarends WM, Ghazvini M, Tafel AA, Heath H, Galjart N, Essers J, Grootegoed JA, Arnheim N, Bezzubova O, Buerstedde JM, Sung P, Kanaar R. Differential contributions of mammalian Rad54 paralogs to recombination, DNA damage repair, and meiosis. Mol Cell Biol. 2006 Feb;26(3):976-89.

Wu ZK, Getun IV, Bois PR. 2010. Anatomy of mouse recombination hot spots. *Nucleic Acids Res* **38**: 2346-2354.

Xi S, Geiman TM, Briones V, Guang Tao Y, Xu H, Muegge K. Lsh participates in DNA methylation and silencing of stem cell genes. Stem Cells. 2009 Nov;27(11):2691-702.

Yu W, McIntosh C, Lister R, Zhu I, Han Y, Ren J, Landsman D, Lee E, Briones V, Terashima M, Leighty R, Ecker JR, Muegge K. Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Res. 2014 Oct;24(10):1613-23.

Zeng W, Baumann C, Schmidtmann A, Honaramooz A, Tang L, Bondareva A, Dores C, Fan T, Xi S, Geiman T, Rathi R, de Rooij D, De La Fuente R, Muegge K, Dobrinski I. Lymphoid-specific helicase (HELLS) is essential for meiotic progression in mouse spermatocytes. Biol Reprod. 2011 Jun;84(6):1235-41.

Zhang, Y., Liu, T., Meyer, C.A. et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008).

BIOGRAPHY OF THE AUTHOR

Catrina Spruce was born December 30, 1981. She attended Dartmouth College and graduated in 2006 with Bachelor's degrees in Biology and Environmental Science. After being hired at The Jackson Laboratory in 2008, she started taking classes part-time at The University of Maine, eventually enrolling in the Biochemistry Master's program in 2015. Catrina is a candidate for the Master of Science degree in Biochemistry from the University of Maine in May 2021.