Using TALENs to Knockout H2A.Lap1 Function in Mice

Nur Diana Anuar

A thesis submitted for the degree of

Doctor of Philosophy

of the Australian National University

February 2018

Australian National University

© Copyright by Nur Diana Anuar (2018)

All Rights Reserved

DECLARATION

The experimental work presented in this thesis constitutes original work by myself carried out at the John Curtin School of Medical Research under the supervision of Dr Tanya Soboleva and Prof David Tremethick unless otherwise stated in the methods, texts or figures.

This thesis conforms to the Australian National University guidelines and regulations. The work contained within has not been submitted for the purpose of obtaining any other degree at this or other universities.

Nur Diana Anuar

February 2018

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious Most Merciful.

"The two most important days in your life are the day you were born and the day you find out why"- Mark Twain

I am still searching the reason why I am here on this earth, but one thing for sure is to complete this thesis.

Determination, perseverance, patience and humility.

Throughout this PhD journey, I have learnt how important it is to have and understand these 4 values in pursuit of learning new knowledge.

I would not be here finishing this project and thesis without these following individuals whom I owe a special debt of gratitude for their tremendous help and kind support in every aspect of my life.

To Dr Tanya Soboleva and Prof. David Tremethick

Thank you for accepting me to be in this well-regarded team and trusting me to work on this project. Thank you for teaching and guiding me in this scientific research world. Thank you for introducing me to the exciting world of epigenetics. Thank you for all the valuable knowledge that you gave me throughout these years. Above all, thank you for your continuous encouragement academically and emotionally. Your supports have been the major driving forces through my graduate years here at the JCSMR. Thank you for opening your door to me, which I will truly treasure for the rest of my life. May Allah bless you and your beautiful family and may this beautiful friendship last forever.

To my dear Sophie (Dr Qian Shen)

Like you always said, what would I do without you. Girl, thank you for always giving me a shoulder to cry on. Thank you for all of your help inside and outside the lab. Thank you for being the understanding friend that I always need. You are indeed a wonderful person. You know I love you no matter what happens and may you and your family continue to lead a happy life.

To Dr Daniel Ryan

Thank you for all of your scientific suggestions, feedback and academic advice throughout these years. I truly admire your wisdom and I truly enjoyed working with you. All the best at your new workplace.

To the Tremethick Group and the members of the Department of Genome Sciences

Thank you for all your help in the lab and thank you for being the cheerful friends. We really had a fun journey in the lab as well as outside the lab.

To my family

Mom and Dad, thank you for all your prayers and thank you for letting me to pursue higher education. Thank you for the financial assistance throughout these years. The road was not always easy for all of us, but I promise you that one day we will look back and feel proud for what we have achieved so far. Forgive me for my absence in time of need and for all of my wrongdoings. Your continuous blessing is essential to my success in this journey. To my sisters, nephews and niece, you brought colours into my life. I love you all deeply.

To my dearest friends in Canberra

To my housemates, thank you for your continuous emotional support throughout these years. To the rest of my friends, thank you for all of your help and for bringing joy and happiness into my years in Canberra.

To my financial sponsor, Malaysian Government, Prof David Tremethick Research Grant

Thank you for providing the financial assistance. And special thanks to Prof David Tremethick, thank you for providing the stipend and visa for the final year of my study. This PhD is not possible without it.

To the JCSMR staff, The Australian Phenomics Facility (APF), the Bioinformatics and Statistical unit

Cathy, Mick, Dr Harpreet and Ann thank you for helping me with all the microscopy issues and thank you for making the MCRF an enjoyable place to be. To Barbara Burke and Phillipa Haas, thank you so much for all the mice handling training, and for all your service in handling our mice colony. I enjoyed and appreciate every single moment of it. To Dr Terry Neeman, thank you for dedicating your time for one-on-one R lesson to me. I never thought that coding can give me so much fun! I look forward to it in the future. To Dr Sebastian Kurscheid, thank you for helping me with chapter 4 and thank you for all the codes that you sent me.

To all collaborators at Sangamo, Koopman's Lab, Khochbin lab and Arkell's lab

Thank you for all of the experiments outside our expertise, the plasmids, reagents and all of the scientific advice throughout this project. Every input is valuable to this project and this project is possible with all of your help.

To Dr Kathie Brown

Thank you for editing my thesis in the requested time frame.

Finally, in exploring the world of scientific research, my experiences here have taught me that mistakes and failures do not always have to hold you back, and often take you closer to your goal. *Science,* I look forward to explore more.

ABSTRACT

The Tremethick laboratory in 2012 discovered a new mouse histone variant, which was designated H2A.Lap1 (Lack of an acidic patch), which, in the adult, is uniquely expressed in the testis and the brain. It was proposed that in the testis, H2A.Lap1 is involved in a new mechanism of gene regulation whereby H2A.Lap1 co-ordinately activates the expression of many genes by being targeted to and directly opening the chromatin region encompassing their transcription start site (TSS). More recently, this laboratory also identified H2A.Lap1 as a component of the intron-exon boundary of active genes expressed in the testis and the brain. These observations suggested a role in transcription and splicing but all the experiments to date have been correlative in nature. Therefore, the major aim of this thesis was to generate an *in vivo* mouse knockout model in order to directly test the importance of H2A.Lap1.

However, the generation of a H2A.Lap1 knockout (KO) mouse was potentially a technically challenging feat because the H2A.Lap1 protein is expressed from three different genes (all located on the X chromosome). In this thesis, we report a simplified TALEN approach that achieved this goal whereby we used one pair of TALENs to simultaneously disrupt the three gene copies of H2A.Lap1. To our knowledge, this has not been done before. Importantly, bioinformatics analyses and wet-lab validation on exome sequencing data of our KO mice did not reveal any TALEN-induced off-target mutations, proving the reliability of this genome editing approach in generating a H2A.Lap1-specific KO model.

Since H2A.Lap1 is predominantly expressed in the testis, we first explored whether are H2A.Lap1 KO mice are fertile. Our results showed that the loss of H2A.Lap1 caused a mild subfertility phenotype that manifested itself in smaller litter sizes when the hemizygous KO males (H2A.Lap1 $^{-/Y}$) were mated with wild type females (H2A.Lap1 $^{+/+}$). Given that these hemizygous KO mice are still able to reproduce despite the absence of H2A.Lap1 suggested that H2A.Lap1 is important but not essential for fertility. Alternatively though, H2A.Lap1 may be critical but compensatory epigenetic mechanisms may come into play in its absence.

To test this hypothesis, I examined the immunofluorescence distribution of several active promoter-associated histone post translational modifications (PTMs) in germ

cells, as well as investigating another H2A.Lap1 like histone variant, H2A.L.2 (expressed at latter spermatogenesis stages compared to H2A.Lap1) in H2A.Lap1 $^{-/Y}$ mice and their wt siblings. The immunostaining analyses showed that, in comparison to the wt mice, H3K4me3 become more enriched in the euchromatin of round spermatids of KO mice, implying a possible involvement of this PTM in a compensation mechanism. In addition, and most interestingly, H2A.L.2 changed its timing of expression being expressed earlier in the round spermatid stage thus also possibly compensating for the loss of H2A.Lap1.

Another striking observation in this thesis is that, in the absence of H2A.Lap1, the nuclei of round spermatids become significantly compacted consistent with its *in vitro* ability to decompact chromatin. Further, the nuclei of round spermatids in KO males showed RNA Polymerase II (RNA Pol II) no longer occupies microscopically distinct transcription hubs. Significantly, RNA-Seq data revealed an increase in intron retention in the absence of H2A.Lap1. Therefore, despite possible compensatory mechanisms, clear changes to the functional organisation of the nucleus and splicing are observed.

During meiosis, sex chromosome becomes highly condensed and silenced to ensure that there is no recombination between the X and Y chromosomes. This inactivation continues post-meiotically, which is referred to post meiotic sex chromosome (PMSC). However, as early round spermatids (ERS) develop into late round spermatids (LRS), approximately 20 % of X-chromosome genes required for spermiogenesis become activated. In the absence of H2A.Lap1, the expression of several X-linked genes (as well as examined autosomal genes) is significantly reduced.

Another interesting observation arising from the immunostaining of active and repressive marks are the dynamic epigenetic changes between ERS and LRS, which has not been observed before. The repressive mark H3K9me3 becomes more enriched in LRS on PMSC which coincides with a depletion in the active marks H4K8Ac and Kcr. At euchromatin, we also observed other active marks (H3K36me3 and H3K4me3) become depleted in LRS in wt mice. This reduction of active marks as round spermatids differentiate suggests that these cells are preparing for a global transcription shut down, which occurs in subsequent developmental stages.

Finally, we examined a possibly phagocytic role for H2A.Lap1 in Sertoli cells. Microscopic analysis of seminiferous tubules of the testes showed that the KO mice have

an increased number of clogged tubules indicating an inefficient clearance of residual bodies (RBs) by Sertoli cells, which could also explain why H2A.Lap1 KO mice are subfertile.

ABBREVIATIONS

Abbreviation	Full name
Arc	Animal Resource Centre
BLAST	Basic Local Alignment Search Tool
CC	Chromocentre
CS	Condensing spermatids
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic acid
DLSSA	Dual Luciferase Single Strand Annealing Assay
DSB	Double strand break
EC/EU	Euchromatin
ERS	Early round spermatid
FVB	Friend Virus B-type
G1	1st generation
G2	2nd generation
G3	3rd generation
G4	4th generation
HC	Heterochromatin
ICI	Intracytoplasmic injection
Indel	Insertion/deletion
JAX	Jackson laboratory
Kcr	Lysine crotonylation
KO	Knock out
LRS	Late round spermatid
NDR	Nucleosome depleted region
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
PCR	Polymerase chain reaction
PMSC	Post meiotic sex chromosome
PNI	Pronucleus injection

PTMs	Post translational modifications
RGN	RNA guide nuclease
RNA	Ribonucleic acid
RNA Pol II	RNA Polymerase II
RS	Round spermatid
RVD	Repeat variable di-residue
SNP	Small nucleotide polymorphism
SNV	Small nucleotide variation
SV	Structural variant
TALEN	Transcription activator-like effector nuclease
T7EI	T7 endonuclease I
TSS	Transcription start site
VCF	Variant call format
WT/wt	Wild type
ZFN	Zinc finger nuclease

Conference and paper arising from this research

Nur D Anuar, David J Tremethick, Philip D Gregory, Lei Zhang, Josephine Bowles, Tara-Lynne Davidson, Peter Koopman, Ruth M Arkell and Tatiana A Soboleva (2013). Targeting a multicopy histone variant gene for knockout using TALEN technology in mice. Epigenetics 2013, Australian Scientific Conference, NSW, Australia.
Poster presentation

Nur Diana Anuar, Matt Field, Sebastian Kurscheid, Lei Zhang, Edward Rebar, Philip Gregory, Josephine Bowles, Peter Koopman, David J. Tremethick and Tatiana Soboleva (2017). Gene editing of the multi-copy H2A.B gene family by a single pair of TALENS. Submitted as preprint;
https://www.biorxiv.org/content/early/2017/12/13/233379.article

Table of contents

List of figures

List of tables

CHAPTER 1
INTRODUCTION

1.1 Epigenetics and chromatin

In 1942, C.H. Waddington introduced the term ‘epigenetics’, which he defined as changes in phenotype without changes in genotype. Today, epigenetics is defined as a number of mechanisms that ensure the inheritance of patterns gene expression without changes occurring in the corresponding DNA sequence. Epigenetic-based mechanisms influence the structure and compaction of chromatin, a complex of DNA, histones and many other associated proteins. Chromatin serves as a reversible structure for complex gene regulation and governs diverse processes including gene transcription, DNA replication, and DNA repair.

The fundamental unit of chromatin is the nucleosome, consisting on average, of 147 bp of DNA wound approximately 1.6 turns around a histone octamer, which contains one H3–H4 tetramer and two H2A–H2B dimers (Figure 1-1 A) (Luger et al. 1997). The linker histone consists of a winged helix motif flanked by N- and C-terminal tails. It binds to linker DNA that connects adjacent nucleosomes and is responsible for the higher order folding into a 30-nm chromatin fibre. Significantly, the structure of a histone H1-bound nucleosome was recently solved (Bednar et al. 2017) (Figure 1-1). Each of the core histones is composed of a structured globular domain and unstructured N- and C-terminal tails. A striking feature of histones is that their unstructured tails can be post-translationally modified. Different types of modifications such as acetylation, phosphorylation methylation, ubiquitylation, and sumoylation among others influence the charge and the repertoire of chromatin binding proteins, thus changing the structure and function of a nucleosome (Grant 2001).

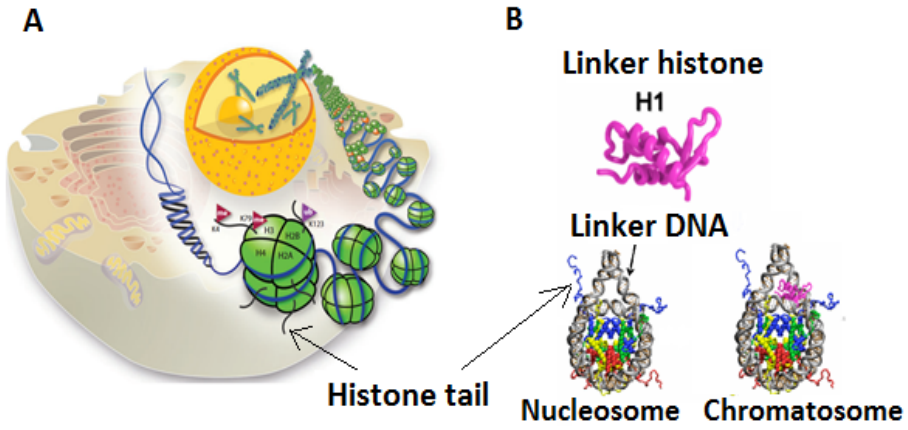


Figure 1-1: Schematic diagram of chromatin. A) The breakdown unit of chromatin, which is a nucleosome. **B)** Crystal structure of nucleosome.

The states of chromatin compaction can be simplistically categorized as active euchromatin (EC) (‘open’ state) and inactive heterochromatin (HC) (‘closed’ state) with each of these states marked by a distinct epigenetic modification (Kouzarides 2007; Woodcock & Ghosh 2010). Euchromatin, or ‘active’ state, consists primarily of coding and regulatory sequences (e.g., enhancers and promoters), which only account for a small fraction (less than 4%) of the genome, in mammals (Allis et al. 2007). Heterochromatin tends to be located at the nuclear periphery, where specific interactions with the envelope may occur. It often forms blocks surrounding the nucleolus and is usually associated with repression of transcription. Heterochromatin may be further classified as constitutive or facultative. Constitutive heterochromatin typically remains condensed throughout the cell cycle and tends to be enriched in repetitive, gene-poor and late replicating DNA sequences. Examples of constitutive heterochromatin are telomeric, centromeric and pericentromeric regions. In contrast, facultative heterochromatin can undergo reversible transition from a more open transcriptionally active state to a compact, transcriptionally inactive state, as occurs during differentiation (Bell et al. 2011; Woodcock & Ghosh 2010).

Berger et al. (2009) stated that there are three categories of signals that culminate in the establishment of a given heritable epigenetic state of chromatin. Firstly, the "Epigenator", which originates from changes in the environment of the cell (e.g., temperature) and triggers an intracellular pathway (either protein-protein or modification-based). The signal is short-lived, but sufficient for the epigenetic phenotype to occur. Secondly, the Epigenetic Initiator signal which responds to the Epigenator and is necessary to define the precise location of the epigenetic chromatin environment (e.g., a DNA-binding protein, non-coding RNA (ncRNA) or any molecule(s) that can define the coordinates of chromatin to be assembled). It should be able to self-reintroduce and self-renew through positive feedback mechanisms. Thirdly, the Epigenetic Maintainer signal which sustains the chromatin environment throughout the generations by a variety of different epigenetic processes as outlined below (Berger et al. 2009).

The complexity of higher organisms is associated with the evolution of epigenetics to maintain stable gene expression throughout cell division. To date, a number of epigenetic machineries that maintain this stable and heritable state of genome expression have been identified. They include DNA modifications (e.g., methylation), post-translational modifications of histones, non-coding RNA, ATP-

dependent remodelers, nucleosome positioning and histone variants (Berger et al. 2009), which will be further discussed in the following section.

1.2 DNA methylation and hydroxymethylation

DNA methylation is an epigenetic mechanism that is widely studied in animals, plants and fungi (Feng et al. 2010). It occurs by the addition of a methyl (CH_3) group at the carbon 5 of cytosine, resulting in 5-methylcytosine (5mC), and informally known as the "fifth base" of DNA. In mammals, methylation predominantly occurs at CG dinucleotides. CG dinucleotides are overall depleted in the mammalian genome except at short DNA sequences (on average 1 kb), termed ‘CpG' islands (which in mammals is globally depleted in methylation). CpG islands typically contain around 5-10 CpGs per 100 bp and often co-localise with gene promoters and regulatory regions (Auclair & Weber 2012; Jones 2012). The methyl groups of 5mC project into the major groove of DNA and thus can either inhibit transcription directly or indirectly by the recruitment of methyl DNA binding complexes (Guibert & Weber 2013; Rose & Klose 2014; Clouaire & Stancheva 2008). Gene silencing by DNA methylation is important in developmental processes such as cell differentiation, tumour suppression and reprogramming in plants and mammals (Feng et al. 2010).

According to Song et al. (2016), 5-methylcytosine is generated by DNA methyltransferases (DNMTs) of which there are three conserved in mammals that regulate methylation. DNMT1 plays a role in the maintenance of DNA methylation and restoration of 5mC, after replication (Goll & Bestor 2005). DNMT3A and DNMT3B are de novo DNA methyltransferases and are essential during early development in mammals (Okano et al. 1999). DNMT1 is ubiquitously expressed in proliferating cells where it localises at replication foci and interacts with the proliferating cell nuclear antigen (PCNA) (Chuang et al. 1997; Leonhardt et al. 1992). Studies by Song et al. (2011) revealed that DNMT1 methylates hemi methylated cytosines at an initial rate 5- to 30-fold greater than unmethylated residues.

Both DNMT3A and DNMT3B are responsible for de novo methylation in mammalian cells (Goll & Bestor 2005). Structural analyses revealed that DNMT1 and DNMT3 families have highly conserved C-5 methyltransferase motifs at their C termini but showed no similarity in N-terminal regions. In contrast to DNMT1, DNMT3A and DNMT3B are highly expressed in embryonic stem cells and during early embryo development but down-regulated in somatic tissues of postnatal animals

(Borgel et al. 2010; Okano et al. 1999). It has been shown by Chen et al. (2004) that the conserved PWWP domain of DNMT3A and DNMT3B, characterised by the presence of the highly conserved proline-tryptophan-tryptophan-proline motif, is involved in the functional specialisation of these enzymes. The disruption of this domain prevents their association with pericentric heterochromatin and abolishes their ability to methylate major satellite repeats (Chen et al. 2004). Double knockouts of DNMT3A and DNMT3B in ES cells have revealed a complete lack of de novo methylation activity which supports our understanding of their functional importance and redundancy in the process (Okano et al. 1999).

Hydroxymethylation (5hmC) results from the conversion of 5mC to 5hmC by the ten-eleven translocation (TET) family of dioxygenases and is reported to be mostly enriched in ES cells (Stroud et al. 2011) and in the brain (Tahiliani et al. 2009; Kriaucionis & Heintz 2009; Khare et al. 2012). Analysis of chromosomal distribution in ES cells has shown that 5hmC is enriched at gene-rich genomic regions and in enhancers (Stroud et al. 2011). Further, 5hmC also has the tendency to localise at other protein-DNA binding sites, such as transcription factor binding sites (TFBs), and with the insulator binding protein, CTCF. 5hmC is also reported to be associated with GC-skew, (where Gs residues are enriched over Cs residues at 5' ends and vice versa at 3′ ends) leading to the suggestion that sequence composition serves as a signal for the deposition of this epigenetic mark (Stroud et al. 2011).

The most commonly used methods to study DNA methylation, such as methylation-sensitive restriction enzymes and sodium bisulphite treatment, cannot distinguish between 5mC and 5hmC (Gao & Das 2014). However, several chemical/enzymatic modifications to these methods in combination with bisulphite sequencing have made possible the base-resolution mapping of 5hmC and provided a more accurate picture of its global distribution (Kamdar et al. 2016; Song et al. 2016; Lister et al. 2013). For example, genome-wide sequencing following methyl-binding protein capture and hydroxymethyl-selective chemical labeling (hMeSeal-seq), conducted by Kamdar et al. (2016) on xenograft-derived prostate cancer cell lines, 22Rv1, (in comparison with the normal prostate cell line, RWPE-1), has shown that 5mC and 5hmC have different locus-specific distribution. The overall distribution of 5mC and 5hmC marks, within each cell line, showed a peak increase in methylation and peak decrease in hydoxymethylation at promoter, exonic and DNAse I hypersensitive sites (DHS) in cancer.

The lack of significant intronic 5hmC differences between cell lines, coupled with the observation that in prostate cancer, intronic regions are the most likely to gain 5hmC (Kamdar et al. 2016) suggests that intronic hydroxymethylation is both tightly regulated and potentially critical for basic cellular function (Kamdar et al. 2016). Similarly, the proportional gain of intergenic 5hmC and its status as the most likely feature to retain 5hmC in cancer may indicate either a lack of importance for intergenic 5hmC marks or, conversely, a key regulatory function in oncogenic formation (Kamdar et al. 2016). 5hmC alterations are generally reported as being associated with various neurological disorders. This includes autism (5hmC enrichments are detected on autism related genes) (Wang et al. 2012; Khare et al. 2012; Zhubi et al. 2014); Rett syndrome (5hmC is globally decreased in the genome) (Szulwach et al. 2011); Angelman syndrome (5hmC is globally increased in the genome) (Szulwach et al. 2011); Fragile X syndrome (5hmC is enriched in FXS related genes) (Wang et al. 2012), Alzheimer's disease (5hmC is increased or decreased in the genome) (Villar-Menéndez et al. 2013; Chouliaras et al. 2013; Condliffe et al. 2014; Coppieters et al. 2014; Bradley-Whitman & Lovell 2013); and Huntington's disease (5hmC is globally decreased in the genome) (Wang et al. 2013).

1.3 Post-translational modifications of histones (PTMs)

A striking feature of histones, especially of their unstructured tails, is that they are extensively post-translationally modified (Grant 2001). The widely-reported modifications include acetylation (Sterner & Berger 2000), methylation (Zhang & Reinberg 2001), phosphorylation (Nowak & Corces 2004), ubiquitination (Shilatifard 2006), sumoylation (Nathan et al. 2006), ADP-ribosylation (Hassa et al. 2006), deimination (Cuthbert et al. 2004; Wang et al. 2004)and proline isomerization (Nelson et al. 2006). Such chemical changes can be enzymatically added or removed. They regulate many aspects of an organism's development including determination of a cell's fate, differentiation, and, in many cases, disease onset (Pedersen & Helin 2010), by altering the chromatin architecture. Lysine is reportedly the most frequently modified amino acid residue and also contains the widest range of PTMs (Azevedo & Saiardi 2016). Lysine is a positively charged amino acid and has a similar structure to arginine. It contains a long flexible side-chain with three methyl groups and a primary positively charged ε-amino group. Arginine, on the other hand, has positively charged guanidinium groups. The primary reactive groups of the ε-amino group in lysine and

the guanidinium groups of arginine contribute to protein stability by forming ionic interactions and hydrogen bonds with proteins and DNA, in addition to acting as a general base in catalysis (Azevedo & Saiardi 2016; Patel et al. 2011). Although both lysine and arginine have side chains of similar reactivity, lysine is more prone to PTMs. The guanidinium group of arginine undergoes three-dimensional ionic interactions that allow arginine to drive protein folding and stability. Lysine polarity contributes to protein structure by way of its amine group that can form a single ionic interaction. Lysine residues are thus more flexible and more easily modified. This unusual chemical plasticity within the lysine residue eliminates steric hindrance and allows histone-modifying enzymes, central to transcriptional regulation, to perform acetylation and methylation as well as subsequent deacetylation and demethylation (Patel et al. 2011).

Over 100 different modification sites have been described to date on histones, and depending on the type of modification, they can function as docking sites for trans-acting factors or influence the structural organization of chromatin and thus affect the accessibility of DNA (Rando & Chang 2009; Tan et al. 2011). PTMs of histones are essential for changing the dynamic state of chromatin in order to allow access of DNA binding complexes to the binding site. These dynamic changes include open/active states (associated with transcriptional activation) and closed/inactive states (associated with transcriptional repression). These modifications correlate with cellular processes such as transcription (Rando & Chang 2009). The elimination of histone-modifying enzymes can affect transcription rates, chromosome stability, and other chromosomal processes. Thus, mapping histone modifications has become an important way to characterize genome function and regulation (Rando & Chang 2009).

PTMs are written by ‘writer’ enzymes, which serve as docking modules for proteins (usually as part of a complex) (‘readers’) that bind to the modification and are removed by ‘eraser’ enzymes (Pande 2016). One such modification that has been well characterised is histone methylation which is catalysed by lysine methyltransferase (the ‘writer') and methylates lysine and arginine residues (Bannister & Kouzarides 2011). Examples of this include H3K36me3 and H3K4me3, both of which accumulate at the gene body and the TSS of actively transcribed genes and are commonly associated with transcriptional activation (Wozniak & Strahl 2014; Barski et al. 2007; Martin & Zhang 2005). The ‘writer’ enzymes responsible for H3K36me3 and H4K4me3 are lysine N-methyltransferase (SETD2) (Martin & Zhang 2005) and lysine methyltransferase 2 (KMT2) (Eissenberg & Shilatifard 2010), respectively. Conversely, the demethylation

of these two sites is executed by lysine-specific demethylase 4A (KDM4A) (for H3K36me3) and lysine demethylase five families (KDM5) (for H3K4me3) (Kooistra & Helin 2012). The ‘reader' recognizes methylated histones, including PHD, chromo, WD40, Tudor, double/tandem Tudor, MBT, Ankyrin Repeats, zf-CW and PWWP (Yun et al. 2011). Li (2013) demonstrated the mechanism whereby H3K36me3 interacts with the PWWP domain, *in vivo* and *in vitro,* by examining the regulation of human DNA mismatch repair (MMR) with MutSα. It was shown that the epigenetic mark H3K36me3 is required (*in vivo)* to recruit the mismatch recognition protein hMutSα onto chromatin, through direct interaction with the PWWP domain of hMutSα. The recruitment of hMutSα is cell cycle-dependent, such that H3k36me3 and the histone methyltransferase SETD2 are required for MMR, *in vivo.* The abundance of H3K36me3, in G1 and early S phase, ensures that hMutSα is enriched in chromatin before mis-pairs are introduced during DNA replication. They showed that cells with depleted SETD2 displayed microsatellite instability and elevated spontaneous mutation frequency, characteristics of MMR-deficient cells. Further, cells with depleted H3K36me3 fail to recruit hMutSα, and the restoration of H3K36me3 in SETD-depleted cells restored the localization of hMutSα to the chromatin, via its PWWP domain. Therefore, H3K36me3 and the PWWP domain regulate human MMR *in vivo*.

Tan et al. (2011) recently identified histone crotonylation as another example of histone modification and a mechanism for epigenetic regulation. Lysine crotonylation (Kcr) marked active promoters and potentially enhancers in both mouse somatic and male germ cell genome (Tan et al. 2011). A report by (Li et al. 2016) showed that Kcr has high-affinity binding to AF9 YEATS domain via an extended aromatic sandwich pocket. Also, AF9 co-localises with crotonylation of H3 and positively regulates gene expression in YEATS domain-dependent manner. They proposed the YEATS domain as a family of a crotonyllysine readers and AF9 YEATS domain directly links Kcr to activate transcription.

In addition to active transcription, PTMs have been widely linked to a silenced chromatin state. A typical repressive histone mark is the trimethylation of lysine 9 and 27 on histone H3 (H3K9me3 and H3K27me3) (Saksouk et al. 2015; Sienski et al. 2012). H3K9me3 is usually enriched at constitutive heterochromatin, while H3K27me3 is enriched at facultative heterochromatin (Saksouk et al. 2015). The methylation of H3K9 relies on the action of H3K9-specific lysine methyltransferase, KMT1A/B (SUV39H1/2), which di- and tri-methylation occur specifically at

pericentric heterochromatin through direct interaction with chromodomain-containing protein (HP1) (Bannister et al. 2001). Meanwhile, the Polycomb repressive complex 2 (PRC2) ensures trimethylation of H3K27 (Young et al. 2011). In ES cells, H3K27me3 is enriched at the promoter region of the Polycomb group (PcG) protein associated genes, which includes the *Hox* genes as well as the inactive X chromosome (Boyer et al. 2006; Escamilla-Del-Arenal et al. 2013; Zhang et al. 2015). There is evidence that shows that in some instances a crosstalk through PRC2 is necessary for the coexistence of H3K9me3 and H3K27me3 in order for gene repression to occur (Boros et al. 2014; Cooper et al. 2014). For example, Boros et al. (2014) showed that binding of HP1α to H3K9me3 is significantly increased in the presence of H3K27me3 and that this is dependent on PRC2. In addition, they have shown that PRC2 is an indirect binder of H3K9me3. The knockdown of two PRC2 subunits, EZH2 and SUZ12, leads to the depletion of H3K27me3, a decrease in the total cellular level of HP1α and a slight reduction of H3K9me3. This is most likely due to a reduced ability to recruit SUV39H1/2 H3K9me methyltransferase to chromatin in response to HP1α loss.

1.4 Non-coding RNA

The developments in next-generation sequencing have increased our ability to sequence cDNA to an unprecedented scale. Transcriptome sequencing, such as RNA-sequencing, coupled with computational analyses allowed researchers to obtain in-depth sequence information that led to the identification and mapping of novel alternatively spliced transcripts and of non-coding transcripts. Protein coding genes account for approximately 1.5% of the genome, while more than 80% is transcribed into non-coding RNA (ncRNA) (Greco & Condorelli 2015). Non-coding RNA is classified according to its sequence length. There are two main categories that differentiate ncRNAs, small (< 200 bp) and long ncRNA (>200 bp). Small ncRNAs include microRNA (miRNA), piwi-interacting RNA (piRNA), and endogenous short-interfering RNA (siRNA), while long ncRNAs (lncRNA) form a highly diverse class with less protein-coding potential (Wang & Chang 2011).

Small ncRNAs impact gene regulation by acting as sequence-specific silencers of gene expression. They are cis-acting molecules that bind to mRNA, thus either interfering with its translational or inducing its degradation (reviewed by Collins et al. 2011). For instance, miRNA is perhaps the most well-known of the small regulatory ncRNA molecules. miRNAs are initially single-stranded RNA (ssRNAs) produced via

transcription or through splicing, which fold into stem-loop structures to form imperfect double-stranded RNA (dsRNA) molecules. These are then processed by RNase III endoribonuclease (generally Dicer) before being denatured. One of the RNA strands binds to the RNA-induced silencing complex (RISC), which later binds to a specific target of mRNA, containing a sequence complementary to the miRNA. The association induces either cleavage, degradation or blocks translation to protein (Collins et al. 2011). siRNAs, on the other hand, are produced as dsRNA and can enter the post-transcriptional gene-silencing (PTGS) pathway, which leads to mRNA degradation in the cytoplasm. Alternatively, siRNAs enter the transcriptional gene-silencing (TGS) pathway, which involves chromatin modification. piRNAs and ssRNAs are produced in clusters and cleaved to individual mRNA by means of a yet unidentified processing mechanism(s) (Collins et al. 2011).

Long non-coding RNAs (lncRNAs) are heterogeneous transcripts produced predominantly by RNA polymerase II. They are defined as RNA molecules of more than 200 bases in length with no protein-coding capacity (Aliperti & Donizetti 2016). lncRNA have been further categorized according to the anatomical properties of their gene loci. For instance, antisense lncRNA overlap with known protein-coding genes; intronic lncRNA are encoded within intronic regions of protein-coding genes; overlapping transcripts are those that overlap protein-coding genes; and lincRNA are encoded completely within the intergenic genomic space between protein-coding loci (Rinn & Chang 2012). In humans, estimates for the number of different types of lncRNA have ranged from 5,400 to 53,000, with only a small fraction being found at levels high enough to suggest evidence of functionality (Rinn & Chang 2012). The association of ncRNA with cellular development has been reported across many mammalian cell types and tissues, including human neuronal cells (Aliperti & Donizetti 2016), mouse cardiomyocytes (Wang et al. 2014), human B-cells (Petri et al. 2015) and human keratinocytes (Kretz et al. 2012).

One widely-cited function of lncRNA, specifically Xist and antisense transcript Tsix, is in X-chromosome inactivation (XCI) (Collins et al. 2011; Lee 2009). Xist is a 17 kb RNA produced exclusively from the Xi (inactive X) and is thought to initiate chromosome-wide silencing as it accumulates and blankets the selected X chromosome, triggering extensive histone methylation as well as other chromatin changes (Collins et al. 2011; Lee 2009). Txis appears to restrict Xist activity on the active X chromosome, Xa, by two mechanisms: firstly Txis prevents the activation of

Xist allele in cis (even as cell differentiation signals trigger XCI on the selected X chromosome) by directing euchromatic modifications to the Xist locus; and secondly, Txis associates with Dnmt3a at the Xist promoter and facilitates de novo CpG methylation and stable silencing of active X chromosome Xist allele (Collins et al. 2011; Lee 2009).

In addition to gene silencing, lncRNA has been associated with regulation of the mitochondrial network (Wang et al. 2014). Specifically, a cardiac apoptosis-related lncRNA (CARL) was shown to suppress mitochondrial fission and apoptosis, by targeting miR-539 and PHB2. PHB2 is an oestrogen receptor (ER)-binding protein, which represses transcriptional activation and is negatively regulated by miR-539. CARL binds to miR-539 directly and participates in the regulation of mitochondrial network and apoptosis, through the miR-539/PHB2 pathway (Wang et al. 2014). One approach that has been used to predict lncRNA function was to co-express it with well-characterized protein-coding genes (Petri et al. 2015). Co-expression alone may not be sufficient to assign a function to lncRNA, but the information obtained from embedding lncRNA in a transcriptional network associated with cellular development, provides a starting point for functional studies. To further determine the function of lncRNA, it will be necessary to perform knock out experiments (Palazzo & Lee 2015).

1.5 Nucleosome positioning

Before the era of genome-wide, next-generation sequencing, mapping of nucleosome occupancy through the whole genome was not feasible, and the question of whether nucleosomes are positioned randomly or at specific DNA sequences, remained an open question. In recent years, technologies like MNase-ChIP-seq, Hydroxyl-radical-seq and MNase exoII-seq have allowed us to map nucleosome positions to a base-pair resolution first in *S. cerevisiae* and then in *D. melanogaster* and to determine their influence(s) in gene regulation (Jiang & Pugh 2009). Given the large size of the mammalian genome, such base pair resolution has been challenging (Soboleva et al. 2014)

According to Cui et al. (2014), nucleosome positioning is usually characterized by two parameters, "rotational positioning," which refers to the orientation of the DNA helix facing the nucleosome, and "translational positioning" of the nucleosome core particle along the DNA sequence i.e., which bp is at the central dyad (or midpoint) position of the core particle. Nucleosome positioning is a dynamic process, but

sequencing-based mapping approaches have been able to identify the positions of individual nucleosomes as a cell population average (Barski et al. 2007) and more recently in single cells at a specific time (Struhl & Segal 2013).

Genome-wide studies have provided a view of the nucleosome positions in a typical gene (Radman-Livaja & Rando 2010). Promoters and other functional regions, such as enhancers and replication origins, are depleted of nucleosomes relative to transcribed genes (Lieleg et al. 2015). These regions are known as "nucleosome-depleted regions" (NDR) and are localized upstream of transcription start sites (TSS) (Radman-Livaja & Rando 2010). However, as shown by this laboratory, rather than being nucleosome depleted, NDRs may be occupied by unstable or fragile nucleosomes (Soboleva et al. 2014)

Genome-wide mapping of the nucleosome has also provided insights into the organization of nucleosomes around protein-coding genes. The organization of nucleosomes is well documented in *S.cerevisiae*, which provides the clearest example of a consensus pattern of organization. The first predominant nucleosome located upstream of the TSS (designated -1) covers a region from -300 to -150 (relative to the TSS) and can regulate the accessibility of the promoter in that region (Jiang & Pugh 2009). During a transcription cycle, the -1 nucleosome experiences many changes that affect its stability, including histone replacement, acetylation, and methylation, as well as translational positioning and ultimately eviction after pre-initiation complex (PIC) formation to RNA polymerase II such as TFIID and SAGA. Further, in mammals, it has been reported that nucleosomes occupy exons more frequently than introns, providing that the exon has a higher guanine-cytosine (GC) content than its neighbouring introns. The nucleosomes positioned over exon sequences produce so-called “speed bumps” that are believed to slow the rate of RNA pol II, promoting exon inclusion during pre-mRNA splicing. In addition to the -1 nucleosome, there exists a highly positioned nucleosome downstream of the TSS (the +1 nucleosome), which may also act as a barrier to the initial movement of RNA Pol II (Soboleva et al. 2014)

Nucleosomes prevent gene expression from taking place by inhibiting the access of transcriptions factors that are recruited to bind to their DNA binding sites. However, some transcription factors are able to alter the nucleosome at a specific DNA location, i.e., near entry/exit point of a nucleosome, by transiently dissociating DNA from the octamer and exposing any underlying binding sites (Anderson & Widom 2000). In fact, it was shown that there are five nucleosome positioning patterns found

in mouse C2C12 myoblast cells that appear around transcription factor binding sites which can be used to predict gene expression (Maehara & Ohkawa 2016). This study also showed that the 5' nucleosome positioning patterns act as functional units on chromatin and describe the transcriptional state of the transcription factors. This indicates that transcription factors are a major determinant of nucleosome positioning.

Nucleosome positioning is strongly affected by DNA sequence (Struhl & Segal 2013). There are two major sequences that determine nucleosome positioning. First, the intrinsic sequence property that refers to the variability of dinucleotide bending properties. Optimal nucleosome formation occurs when bendable nucleotides (AT and TA) occur on the face of the helix that directly interacts with histones (Struhl & Segal 2013). Other than DNA sequence, nucleosome positioning is also determined by ATP-dependent nucleosome remodelling enzymes. For instance, the combined action of nucleosome remodelling enzymes, ISW1 and CHD1 in yeast, are required to maintain the positioning of nucleosomes within the coding regions of eukaryotic genes and in alignment with transcriptional start sites (TSS) (Gkikopoulos et al. 2011). In addition, a recent study also showed these two nucleosome spacing enzymes compete to set the spacing on most genes, such that CHD1 dominates genes with shorter spacing, and ISW1 dominates genes with longer spacing. In contrast, ISW2 (another yeast nucleosome enzyme) plays a minor role, limited to transcriptionally inactive genes (Ocampo et al. 2016).

1.6 ATP remodelers

ATP-dependent chromatin remodelling enzymes comprise a specialized class of enzymes involved in the epigenetic mechanisms that direct the compaction and decompaction of DNA in the nucleus (Runge et al. 2016; Kadoch & Crabtree 2015; Clapier & Cairns 2009). Remodelers are large multi-subunit complexes that contain an ATPase subunit that hydrolyses ATP to produce energy which is used to facilitate nucleosome sliding, to disrupt nucleosome conformation, for histone variant exchange (Latrick et al. 2016), and for promoter clearance on a genome-wide scale (Runge et al. 2016). Remodelers bind to chromatin at enhancers, promoters and replication origins, to expose certain cis DNA elements in a regulated manner, thereby facilitating processes like transcription, DNA replication, DNA repair and DNA recombination (Clapier & Cairns 2009). Although some redundancy occurs, most remodelers are essential for cellular growth, development, or differentiation (Lans et al. 2012).

There are four major classes of remodelers, namely, i) Swi/Snf (switching defective/sucrose nonfermenting); ii) INO80 (inositol requiring 80); iii) CHD (chromodomain, helicase, DNA binding); and iv) ISW1 (imitation switch) (Clapier & Cairns 2009). They are considered derivatives of the SNF2 ATPase family which is evolutionarily conserved across eukaryotic organisms (all have a conserved ATPase domain), reflecting their conserved function and importance in gene regulation (Runge et al. 2016). Swi/snf are well known remodelling complexes, composed of 8 to 14 subunits (Clapier & Cairns 2009), and were initially identified in yeast (Winston & Carlson 1992). They are identified by their ability to activate transcription of the HO and SUC2 genes. Their mechanism of action, similar to many others that are regulated by this complex in yeast, involves binding in proximity to the promoter in order to assist the subsequent binding of specific transcription factors (Masliah-Planchon et al. 2015). In mammals there are 15 subunits that make up the Swi/Snf complex. Among these subunits, only two ATPases, Brg1 or Brm, are sufficient to remodel nucleosomes *in vitro* (Phelan et al. 1999), while the others are likely providing specificity *in vivo*. In addition, there are reports suggesting that mutations in Swi/Snf subunits may contribute to various human malignancies, as evidenced by the high frequency of mutations, within genes encoding BAF subunit (mammalian Swi/Snf complex), occurring in >20% of human cancers (Kadoch & Crabtree 2015).

1.7 Histone variants: Structure and versatility of function

The DNA is packaged by an octamer of core histones, consisting of two H2A-H2B dimers and an H3-H4 tetramer (Figure 1.2) to form a 10-nm nucleosome as the primary component of chromatin. Histones are proteins comprising 100-140 amino acid long and structurally consist of 3 motifs: i) the histone fold, ii) the extra fold structured elements unique to the different histones, and iii) the labile termini, which vary from 13 to 42 amino acids in length (Arents & Moudrianakis 1995; Luger et al. 1997). The 10 nm polynucleosome arrays are folded and stabilized by linker H1 histones. As discussed above, histones are constantly modified, shifted, evicted and re-deposited as chromatin continues to remodel, in order to maintain proper organization and gene regulation (Marzluff et al. 2008). Major chromatin disruption requiring new histone synthesis also occurs during DNA/chromatin replication at S-phase. These continuous disassembly-reassembly of chromatin requires rapid transcription and translation of histone genes.

The core histone proteins are encoded by a family of replication-dependent genes. In eukaryotes, the mRNA of these genes is the only known cellular non-polyadenylated mRNA that has a unique 3' stem-loop (26 bp), crucial in modulating mRNA stability, transport, and translation. These mRNAs are rapidly expressed at the beginning of S phase and are maintained at high levels throughout the S phase to coincide with DNA replication (Marzluff et al. 2008).

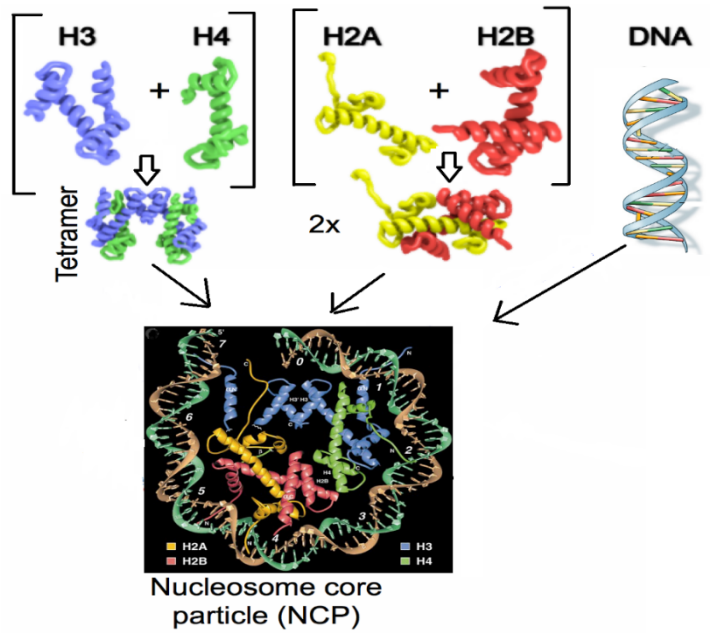


Figure 1-2: Nucleosome core particle schematic diagram: half ribbon traces (73 bp) of the 146-bp DNA phosphodiester backbones (brown and turquoise) and eight histone protein main chains (blue: H3; green: H4; yellow: H2A; red: H2B. Adapted from Luger et al. (1997).

Histone variants, unlike the core histones, are expressed and incorporated into chromatin throughout the cell cycle. The replication-independent gene and mRNA structure of histone variants also differ from core histones. The histone variant genes often contain introns, and the transcripts are often polyadenylated; features that are thought to be important in the post-transcriptional regulation (Kamakaka & Biggins 2005). The most prominent sequence divergence is found in H2A and H2B, with H2A having the most variants. In contrast, H4 has no reported variants, and H3 variants have only minimal sequence variations (Mattiroli et al. 2015).

Histone variants have evolved to contribute to the complexity of chromatin and have specialized functions in regulating chromatin dynamics. Their sequence divergence can alter chromatin states and the fate of cell decisions (Weber & Henikoff 2014). The role of histone variants in regulating transcription has been widely reported. It was initially understood that in order for transcription to take place, a nucleosome-depleted region (NDR) must form at the TSS, to allow binding of transcription machinery. Based on experimental evidence, the accessibility of micrococcal nuclease supported this notion, however, this has been disputed upon finding that nucleosomes containing labile H2A variants flanked the NDR. Jin et al. (2009) used chromatin immunoprecipitation (ChIP) and re-ChIP, followed by high throughput Solexa sequencing on HeLa cells expressing Flag-tagged H3.3 histone to reveal that hybrids of H3.3-H2A.Z nucleosome core particle (NCP) localized to TSS of active genes. Subsequently, this laboratory has shown that heterotypic H2A.Z-H2A nucleosomes occupy the TSS in stem cells thus raising the possibility that H3.3-H2A.Z nucleosomes are in fact heterotypic with respect to H2A.Z (Nekrasov et al. 2012; Soboleva et al. 2014). As mentioned above, we also found that H2A.Lap1-containing nucleosomes are found at the TSS of active genes in the testis (Soboleva et al. 2012; Soboleva et al. 2014). These unstable nucleosomes may serve as a ‘place holder’ to prevent the region from being covered by adjacent stable NCP and/or nonspecific factors, when the region is nucleosome-free. Moreover, these fragile nucleosomes could be more easily displaced by transcription factors because of its instability (Jin et al. 2009).

Histone variants have also been reported to participate in their own epigenetic inheritance, maintaining correct localization on newly synthesized daughter strands following DNA replication. For instance, during the cell cycle of mouse trophoblast stem cells, a decrease in the level of H2A.Z, at promoters, during S phase occurred. This coincided with the H2A.Z-H2A.Z nucleosome, flanking the TSS, becoming

heterotypic. This nucleosome persisted as heterotypic until M phase (Nekrasov et al. 2012). The depleted expression of H2A.Z at promoter, after DNA replication, is possibly due to its redistribution to constitutive heterochromatin and/or at the centromere. This cell cycle-dependent possible dynamic movement of H2A.Z suggests its functional importance in providing transcriptional memory, that is, active marks established at the G1 phase and remaining at M phase may facilitate the reestablishment of gene transcription after cell division (Nekrasov et al. 2012).

Histone variants have also been linked to the response to DNA damage, following exposure to gamma-radiation (Volle & Dalal 2014). H2A variants, γH2A.X. H2A.Z and macroH2A, specifically, have been implicated in this role. The various functions of histone variants in eukaryotes are briefly presented in Table 1.

Table 1-1: List of reported functions of histone variants in eukaryotic organisms (Yuan & Zhu 2012)

Histone variants	Function
CENP-A or cenH3	Epigenetic marker of the centromere
H3.3	Involved in transcription
	Enriched at transcriptionally active regions, telomerase, and pericentric regions
H3.Z	Regulation of cellular response to outside stimuli
H3.Y	Regulation of cellular response to outside stimuli
H2A.Z	Transcriptional control rheostat
	Heterochromatin formation and maintenance
H2A.X	Double strand break repair
	Meiotic remodelling of sex chromosome
macroH2A	Gene silencing
	X-chromosome inactivation
H2A.Bbd/H2A.B or H2A.Lap1	Epigenetic mark of active chromatin. Found at the TSS and intron-exon boundaries.

1.8 Histone H2A.Z variants: The structure and function

H2A.Z is the most widely studied of the histone variants, with respect to structure and function (Chakravarthy et al. 2005; Sarma & Reinberg 2005; Yuan & Zhu 2012; Suto et al. 2000). It is an essential histone found in all higher eukaryotes and there is approximately 90 % sequence homology between species (Suto et al. 2000; Chakravarthy et al. 2004; Yuan & Zhu 2012). The H2A.Z sequence differs by approximately 60 % from canonical H2A and the other H2A variants, mainly in its "docking" domain in the C-terminus and in the L1 loop where two H2A molecules contact each other at the back face of the nucleosome (Suto et al. 2000). The docking domain of the H2A.Z histone variant contains several negatively charged amino acid residues that come together on the surface of the nucleosome to create an 'acidic patch' (Luger et al. 2012). The larger acidic patch of H2A.Z-H2B displays higher affinity with H4 N-terminal tail than that of the H2A-H2B dimer, resulting in the compaction of chromatin by facilitating intra-nucleosome-nucleosome interactions. Specifically, the acidic patch of a H2A.Z-containing nucleosome interacts better with the basic histone H4 tail originating from a neighbouring nucleosome thus driving intra-nucleosome interactions (Fan et al. 2004; Luger et al. 2012). Hence, DNA accessibility is suppressed and transcription is inhibited in vitro from a chromatin template that is enriched with H2A.Z to a greater extent than a H2A-containing chromatin template (Zhou et al. 2007)

H2A.Z has been shown to be crucial in metazoan development, as the absence of this variant has been reported to be lethal to *Tetrahymena thermophile* (Liu et al. 1996), Drosophila (Clarkson et al. 1999) and in blastocysts of the mouse embryo (Faast et al. 2001). Immunofluorescence staining of differentiating inner cell mass (ICM) in mice showed that H2A.Z is enriched at the pericentric heterochromatin, but absent from totipotent ICM. Interestingly, initially H2A.Z enrichment occurred at pericentric heterochromatin before shifting to other regions of the nucleus. This suggested a novel role for H2A.Z in providing a signal for the assembly of an architecturally distinct structure at pericentric heterochromatin (Rangasamy et al. 2003). Indeed, findings by Greaves et al. (2007) showed that H2A.Z has a major structural role in maintaining the integrity of pericentric heterochromatin in a variety of mammalian cell lines, as down regulation of H2A.Z impairs mitotic division and leads to chromosomal segregation, defects in cytokinesis and the loss of HP1α from pericentric heterochromatin (Greaves et al. 2007). In addition, H2A.Z in combination with either H3K9me3 or H3K4me2

contribute to the specialised 3-D organization of a centromere at metaphase (Greaves et al. 2007). H2A.Z also has a unique and specific role in assembling the X chromosome into a silenced heterochromatic state, following meiosis (Greaves et al. 2006)

In both yeast and mammals, H2A.Z has been shown to associate with transcription activation. Genome-wide chromatin immunoprecipitation (ChIP) showed that H2A.Z localizes primarily to the promoter region (Zhang et al. 2005; Barski et al. 2007) Formaldehyde crosslinking followed by ChIP experiments on budding yeast *S. cerevisiae* showed that H2A.Z is enriched at intergenic regions upstream of *PHO5* and *GAL1,* even under repressed conditions (Santisteban et al. 2000). Subsequent genome-wide studies showed that H2A.Z is enriched at the promoter-proximal +1 and -1 nucleosome (Albert et al. 2007). The presence of H2A.Z nucleosome surrounding most yeast promoters, in the absence of transcription, led to the proposal that H2A.Z-containing nucleosomes helped poise genes for transcription (reviewed by Luk et al. 2010).

Evidence is accumulating to associate H2A.Z with cancers, such as melanoma (Vardabasso et al. 2015), prostate (Baptista et al. 2013), bladder (Kim et al. 2013) and breast (Hua et al. 2008; Svotelis et al. 2010). It was observed by Svotelis et al. (2010) that H2A.Z mRNA is elevated in breast cancer biopsies. This correlates with increased levels of oestrogen receptor gene α (ERα) in low-grade breast tumours, and elevated levels of H2A.Z in high-grade tumours. ERα was shown to regulate H2A.Z in breast cancer by way of a positive feedback loop that appears important for ERα signalling under normal and pathological conditions. This study correlates elevation of H2A.Z expression with cellular proliferation, when E2 (ligand for oestrogen receptor) levels are low and during tamoxifen (drug used widely to treat breast cancer due to its capacity bind to ERα) treatment. H2A.Z therefore could be an important factor in the transformation of breast cancer cells and in the establishment of endocrine resistance, which is of major clinical importance in breast cancer treatment.

1.8.1 H2A.Bbd (human) and the recently discovered mouse histone H2A.Lap1

The importance of the acidic patch is exemplified by the fact that in addition to H2A.Z having an extended acid patch, the human histone variant H2A.Bbd and the mouse orthologue H2A.Lap1 (recently re-named H2A.B) have a reduced acidic patch (missing three acid amino acid residues). Not surprisingly then, H2A.Bbd nucleosome

arrays cannot compact and de-represses chromatin mediated repression *in vitro* (Zhou et al. 2007). Further, restoration of the acidic residues in H2A.Bbd was shown to restore the intranucleosomal interactions, needed for the folding of a nucleosomal array, and repression of transcription (Zhou et al. 2007). In addition, H2A.Bbd has a truncated C-terminus/docking domain making this nucleosome highly unstable leading to unwrapping of the DNA from the octamer surface (Bao et al. 2004; Doyen et al. 2006). Specifically, this domain is responsible for the interaction with the N-terminus H3, which guides the DNA at the entry and exit points. Therefore, micrococcal nuclease digestion of H2A.Bbd-NCP (nucleosome core particle) revealed that only 118 ± 2 bp of DNA, compared with 145 bp for the wild type nucleotide, was protected (Bao et al. 2004).

H2A.Bbd was first discovered in humans and shown to be excluded from the female X-chromosome (Chadwick & Willard 2001), a feature that gave this variant its name, Barr body deficient (Bbd). H2A.Bbd protein consists of 115 amino acids with a molecular weight of 12.7 kDa and is only 48 % identical to the core H2A histone, making it the most evolutionary divergent histone variant known to date (Chadwick & Willard 2001; González-Romero et al. 2008). It has been proposed that the major hallmarks of Bbd, compared to core H2A (Bao et al. 2004; Doyen et al. 2006), are: i) the presence of a continuous stretch six arginine residues and the conspicuous absence of lysine in the N-terminal tail; ii) the absence of a C-terminus including the last segment of the docking domain that is responsible for the interaction with H3 in major NCP; iii) major sequence differences in the docking domain of H2A (residues 81-119); iv) the presence of only one lysine in H2A.Bbd (compared to 14 in major H2A); and v) the absence of 'acidic patch as discussed above.

To establish the function of histone H2A.Bbd *in vivo*, our laboratory turned to the mouse model. An extensive analysis of the mouse genome database revealed, however, that several H2A.Bbd 'like' genes exist but the most similar to H2A.Bbd was a new histone variant, which we designated, H2A.Lap1 (**L**ack of **a**cidic **p**atch) (see below). H2A.Lap1 is a multi-copy gene, coded by at least three genes, H2Afb3 (henceforth referred to as mH2ALap1a) which is a major form, Gm14920 (mH2ALap1b) and EG624153 (mH2ALap1c) (Soboleva et al. 2012). Although recently, H2A.Lap1 has been re-designated as H2A.B.3, in this thesis, I will refer to the histone as H2A.Lap1 for ease of distinction from the human H2A.Bbd. H2A.Lap1 encoded genes are over 90% identical in DNA sequences (Figure 1.3). We found that

mice have a family of four H2A.Lap1-like variants (which we designated H2A.Lap 1-4), all multi-copy genes, related to a human variant, H2A.Bbd. The H2A.Lap1 is the closest variant, being 68% similar to the human H2A.Bbd and the focus of this study.

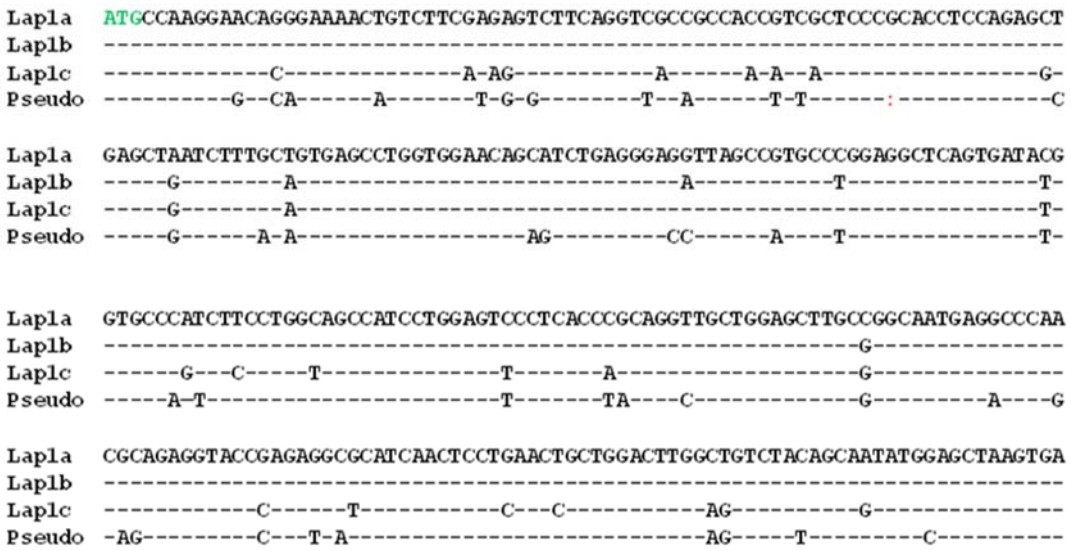


Figure 1-3: The alignment of partial sequence of H2A.Lap1 genes.

1.8.2 H2A.Lap1 is a tissue-specific histone

1.8.2.1 H2A.Lap1 in spermatogenesis

In mammals, germ cells move from the periphery to the lumen of a seminiferous tubule and are subsequently transported to the epididymis, a long tubule that stores mature sperm. Mammalian spermatogenesis takes place in the seminiferous tubules of the testes.

Spermatogenesis consists of three stages (Figure 1-4), i) somatic diploid cell, or spermatogonia (2n): establishment of primordial germ cell (PGC) in the embryo, and mitosis of spermatogonial cells; ii) tetraploid cells (4n) or spermatocytes: meiosis of diploid cells which involves reductional divisions in primary and secondary spermatocytes to produce haploid cells. The primary spermatocytes undergo differentiation through preleptotene (in Prophase I)→leptotene→zygotene→early pachytene→ mid-pachytene→late pachytene. Secondary spermatocytes undergo further meiosis divisions to produce haploid spermatids. In the final stage, iii) haploid (n) spermatids that are represented by round, elongating and condensing stage of differentiation resulting in final stage, mature haploid spermatozoa (Clermont 1972).

A characteristic feature of spermatogenesis is that transcription stops at a defined point during the differentiation of germ cells, at the beginning of elongating spermatid stage. Thus, prior to this stage (the round spermatid stage), high levels of transcription occur to produce a store of mRNA, e.g. mRNA encoding protamines and other key proteins essential for the completion of spermatid development. The translation of those mRNA species is uncoupled from transcription and happens in elongating-condensing spermatids. Mammalian spermiogenesis involves a dramatic reorganization of chromatin from a histone-based nucleosome structure to a structure primarily based on protamines, which condense the DNA for packaging within the sperm head. Gradual replacement of histones with protamines requires first, the utilization of transitional proteins (TP; in mammals contain two major types, TP1 and TP2), followed by incorporation of protamine dimers, Prt1/2.

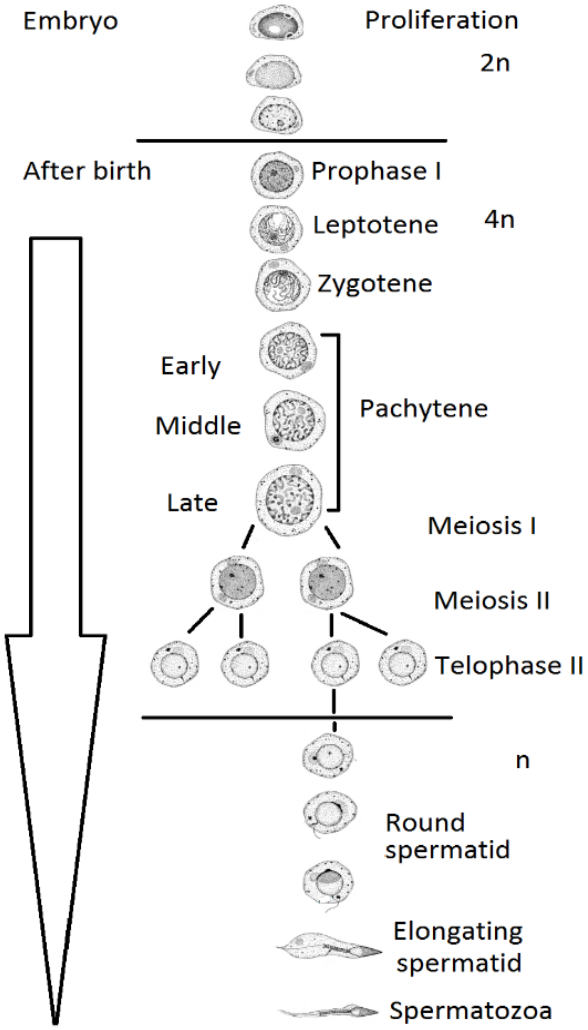


Figure 1-4: Stages of spermatogenesis in mammals. Adapted from Clermont (1972)

Many histone variants of H2A and H2B are specifically expressed in the testis. For instance, TH2A and TH2B have been detected in post-meiotic spermatids. During the condensation of spermatid nuclei, TH2A and TH2B gradually decrease, before being degraded during DNA compaction (Moss et al. 1989). An increase in histone acetylation in elongating spermatids serves as a signal for the recruitment of specific machinery acting on acetylated histones needed for the histone-protamine exchange (Govin et al. 2004). Investigation of nucleoproteins on condensing spermatids has led to the identification of five more H2A and H2B histone variants that are expressed mainly in testes, namely H2A.L1 (H2A.Lap2), H2A.L2 (H2A.Lap3), H2A.L3 (H2A.Lap4), H2B.L1 (of which mRNA is strongly enriched at round and elongating spermatids) and H2B.L2 (which has very low mRNA present at meiotic and post-meiotic stages). Immunostaining using antibodies that recognize H2A.L1 and H2A.L2 (H2A.L1/H2A.L2) or H2B.L1 showed an accumulation of H2A.L1/L2 and H2B.L1 in condensing spermatids, similar to that of transition proteins and protamines (Govin et al. 2007).

Further investigation to identify testis-specific histones has led to the in-depth characterization of the function of H2A.Lap1 variant *in vivo,* by Soboleva and colleagues (2012) in our laboratory. Mechanistically, and similar to H2A.Bbd, H2A.Lap1 inhibited chromatin compaction *in vitro* (Soboleva et al. 2012). *In vivo*, H2A.Lap1 expression begins at the pachytene stage and peaks in the round spermatid stage coincident with the highest levels of transcription. In elongating spermatids, H2A.Lap1 is exported out of the nucleus (Soboleva et al. 2012). ChIP-Seq and RNA-Seq analysis revealed that H2A.Lap1 is targeted to the TSS of active genes including those few genes on the X chromosome that escape the X chromosome inactivation process (Soboleva et al. 2012).

Most recently, H2A.Lap1 has also been shown to be incorporated into the intron-exon boundary of active genes suggesting that, in addition to a role in transcriptional initiation, it also has a specific function in pre-mRNA splicing (Soboleva et al. 2017). To explore this further, CLIP (Cross-Linking and Immunoprecipitation) experiments were conducted and was it demonstrated that H2A.Lap1 associated with mRNA. Finally, a mass spec analysis revealed that H2A.Lap1 directly interacts with splicing factors (Soboleva et al. 2017). Taken together, this indeed provides support for the notion that H2A.Lap1 is a novel epigenetic regulator linking transcriptional initiation with RNA splicing.

1.8.2.2 H2A.Lap1 in the brain.

The role of H2A.Lap1 in the mouse brain remains unknown, but similar to the testis, H2A.Lap1 ChIP-Seq and RNA-Seq in the hippocampus revealed that H2A.Lap1 is also found at the TSS and intron-exon boundaries of active genes (Soboleva et al. 2017). An immunohistochemical staining of mouse brain tissues showed that H2A.Lap1 is expressed in neurons throughout the adult mouse brain, including the hippocampal areas of dentate gyrus (DG), CA1 and CA3 (unpublished data). In cerebellum, H2A.Lap1 localized specifically in the Purkinje neurons, while H2A.Z is expressed in the granular layer (unpublished data). Further, H2A.Lap1 expression was found to coincide with the maturation of the Purkinje neurons, implicating H2A.Lap1 in Purkinje neuron development (unpublished data).

In spite of these successes in characterizing many aspects of H2A.Lap1 function in mouse brain and testes, the ultimate approach to deciphering a function of a gene is to create and characterize a biological system where the function of the gene of interest is disrupted. Therefore, the overall aim of this thesis was to generate a mouse model in which all three H2A.Lap1-coding genes are knocked-out.

1.9 Targeting H2A.Lap1 encoded genes

The purpose of gene targeting is to discern a particular gene's biological function. There are two ways to achieve this, either by overexpression of the gene of interest, or by down regulation of its expression. Overexpressing a gene can be employed by transgenic technology, whereas, homologous recombination is typically employed to create a mutation that leads to 'loss-of-function' (Hall et al. 2009). Where overexpression can violate balanced gene dosage, affecting protein folding, complex assembly and downstream regulation (Gibson et al. 2013), most researchers opt for a knockdown or knockout targeting strategy to disrupt a gene's open reading frame and block its expression, as a direct means to delineate its function (Hall et al. 2009).

Since H2A.Lap1 is a multicopy gene (three genes and one pseudogene), the conventional knockout approach that only able to target one gene at a time is not a suitable tool to target H2A.Lap1 for knockout. However, new methodologies in genome editing, which rely on recognition of specific DNA sequences by either proteins or complementary RNA, have allowed researchers to induce a custom modification, in cells and model organisms.

1.9.1 Genome editing approach to disrupt genes

1.9.1.1 ZFN, TALE or CRISPR

Zinc finger nuclease (ZFN)

Specificity of targeting remains the focus in disrupting a gene of interest. A recent approach called ‘genome-editing' has enabled scientists to target virtually any sequence. This approach uses engineered nucleases which consist of a specific sequence of DNA binding domain fused to a non-specific DNA cleavage motif (Gaj et al. 2013; Boettcher & McManus 2015; Cermak et al. 2011). The first genome editing introduced was zinc finger nucleases (ZFN). A zinc finger nuclease contains a DNA-binding domain composed of at least three tandem arrays of Cys2His2 zinc fingers, commonly found in eukaryotes and representing the second most frequently encoded protein domain in the human genome. This motif is fused with a nonspecific DNA cleavage domain derived from the endonuclease FokI which requires dimerization for enzymatic activity (Gutschner et al. 2011; Gupta et al. 2011) (Figure 1-5). The specificity of targeting lies in the DNA binding domain that can be customized by the researcher. One zinc finger recognizes three base-pairs on one DNA strand. For example, a protein comprising of three zinc fingers recognizes a 9- to 10-base-pair segment. But the zinc-finger triplet also means you have less flexibility in how the recognition sequence must be read i.e. although there are hundreds of zinc-finger modules, not every possible sequence is represented (DeFrancesco 2011).

Zinc finger nucleases generate sequence-specific, double-strand breaks that are repaired mainly by error-prone nonhomologous end joining (NHEJ) or by high-fidelity homologous recombination (Cui et al. 2011) (Figure 1-6). Zinc finger nucleases have reportedly been used to create knockouts in rats (Geurts et al. 2009), mice (Carbery et al. 2010) and zebrafish (Gupta et al. 2011) with remarkable germ line transmission rates and efficiency (Cui et al. 2011). However, mutations are unpredictable owing to the variable nature of DNA repair by NHEJ and are limited to knockouts (Cui et al. 2011). Although ZFN has demonstrated flexibility in the specificity of the targeted sequence, this modular assembly requires the use of a preselected library of zinc-finger motifs generated by selection of large combinatorial libraries or by rational design (Cui et al. 2011). Further, off-target effects resulting from the application of ZFN are widely reported across all eukaryotic organisms. Also, Pattanayak et al. (2011) identified 37 sites and 2,652 sites in the human genome that can be cleaved by ZFN CCR5-224 and VF2468 that were designed to target only the *CCR5A* and *VEGF* genes. Further, Sander

et al. (2013) revealed that transient transfection and using a strong constitutive CMV promoter may induce an even greater number of undesired cleavage sites by ZFN. Numerous efforts have been made to improve the design and specificity of ZFN. Prior to the introduction of the TALEN system, another limitation of the ZFN approach is the complexity involved in designing ZFN to recognise specific DNA sequences. TALEN and CRISPR were subsequently developed as an alternative to ZFN. These systems are characterized by their simplicity of design and high functional efficiency at enabling various genomic manipulations.

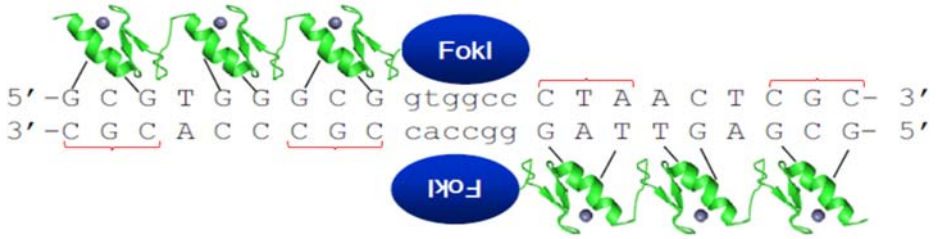


Figure 1-5: ZFN recognize nucleotide in triplet manner. The engineered zinc finger nuclease which consists of DNA binding domain (tandem repeat of Cys2His2). Since FokI works as a dimer, two constructs are required, one to target the sense strand and the other one for the antisense. Adapted from Durai et al. (2005)

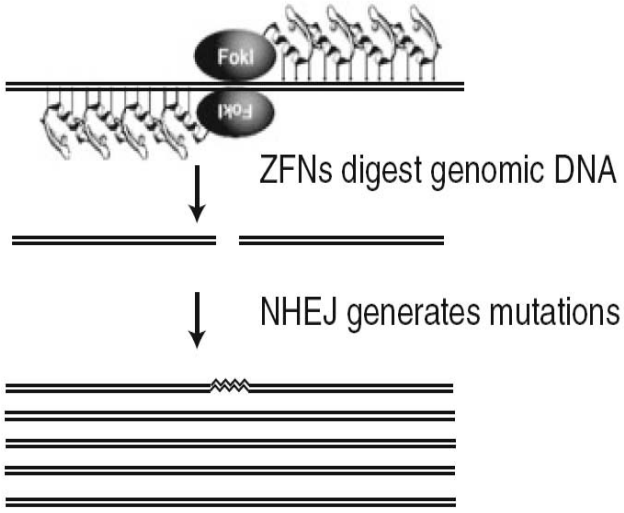


Figure 1-6: FokI will only start the enzymatic cleavage activity when ZFNs bind to both sense and antisense strand. This creates double-strand breaks (DSB) in the genomic strand which will be naturally repaired by NHEJ or HR.

Transcription activation-like effector nuclease (TALEN)

In 2011, *Nature Methods* named the methods of precise genome editing, including the TALEN system, as the method of the year. TAL effector is a newly described class of DNA binding proteins, a family of virulence factors produced by plant pathogen *Xanthomonas spp*, which when injected into a host plant, bind to a specific host promoter sequence that regulates genes affecting disease processes, both negatively and positively (Cermak et al. 2011; DeFrancesco 2011). TAL effectors contain tandem, 33-35 polymorphic amino acid repeats, which independently specify single, contiguous nucleotides in the DNA target. The targeting specificity is determined by the amino acids 12-13, referred to as repeat variant di-residues (RVDs), which are HD, NG, NI and NN corresponding to the nucleotide C, T, A and G/A respectively (Cermak et al. 2011; Bogdanove & Voytas 2011) (Figure 1-7).

TALEN has been shown to act similarly to the zinc finger nuclease. However, in contrast to the zinc finger domain which predominantly recognizes nucleotide triplets, the TAL effector domain consists of monomers with the ability to recognize a single nucleotide, which has dramatically improved DNA targeting (DeFrancesco 2011). The targeting domain of TAL can be fused with the catalytic domain of the FokI nuclease generating a fusion protein that creates site-specific DNA double-strand breaks (DSB). This protein functions as a dimer, hence, TALENs are designed in pairs for both strands of DNA separated by a spacer, allowing the FokI monomer to come together and cleave. The creation of DSB in turn triggers an *in vivo* DNA repair pathway, which can be harnessed to create specific DNA sequence modifications at the break site. In non-homologous end joining (NHEJ), the broken chromosome may be re-joined imprecisely, resulting in small insertions or deletions at the break site. In homologous recombination, the DNA surrounding the break site is replaced with a repair template of a similar sequence, which can be modified or amended to introduce specific mutations, or different DNA sequences. Unlike ZFN, TALEN has been reported also to have low cellular toxicity and higher target specificity (Mussolino et al. 2014). Reports regarding off-target sites by TALEN are not as common as they are for ZFN (and CRISPR, see below). The only limitation of the TALEN design is in the selection of the TALEN nuclease site that is needed for nucleotide thymine -before the 5'- end of the target sequence (Nemudryi et al. 2014). Based on crystal structure analysis of TALE effector by Mak et al. (2012), the W32 residue in the N-terminal region of the DNA-binding domain was demonstrated to interact with 5'-T, affecting

the efficiency of TALEN binding to the target site (Nemudryi et al. 2014). However, this limitation was fixed by selecting TALEN N-terminal mutant variants with the capability to bind A, G, or C (Nemudryi et al. 2014). Despite the specificity shown by TALENs design, the TALEN approach has been super ceded by the CRISPR-Cas9 system.

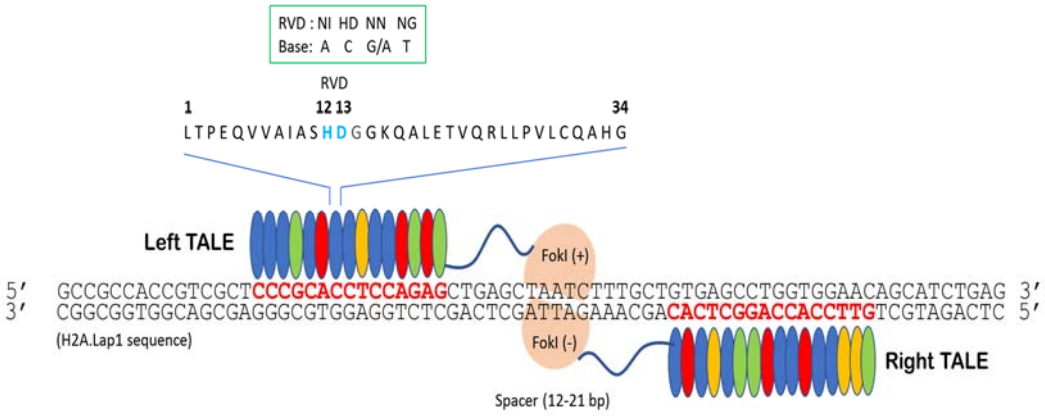


Figure 1-7: The design of TALEN is much more flexible than ZFN. The DNA binding domain consists of repeat di-residue of amino acid that able to recognize one nucleotide, providing more specificity and flexibility for targeting in comparison to ZFN. Adapted from Kim & Kim (2014)

Clustered regularly interspaced short palindromic repeats -Cas9 (CRISPR-Cas9)

About two years after the discovery of chimeric TALEN proteins, a new genome-editing tool called CRISPR-Cas was introduced. Unlike ZFN and TALEN that used protein motifs to recognize DNA sequences, CRISPR uses antisense RNA to bind to the targeted RNA-guided nuclease, (RGN). Since it was first described in 1987, CRISPR was not further developed for genome editing purposes until 2010 (Doudna & Charpentier 2014). Briefly, the Cas9 protein, found in bacteria and algae, serves as an adaptive immune response against invading viruses and plasmid DNA by degrading complementary DNA sequences (Chandrasegaran & Carroll 2016). Targeting of Cas9 to specific genomic sites is mediated by a 20-nucleotide guide sequence, within an associated CRISPR RNA (crRNA), and requires *trans*-activating crRNA (trcrRNA) that recruits the crRNA into the Cas9 complex. Recognition of cleavage sites is determined by crRNA-DNA base pairing and protospacer-adjacent motif (PAM), a three-nucleotide sequence (NGG) juxtaposed to the complementary DNA region (Doudna & Charpentier 2014). It is notable that a single guide RNA (gRNA) that mimics the trcrRNA-crRNA complex recruits Cas9 to targeted genomic sites and generate DSBs in DNA (Doudna & Charpentier 2014).

Despite its simplicity for targeting, and with improvements continually being made (Doench et al. 2016; Kleinstiver et al. 2016; Shin et al. 2017), one major drawback of the CRISPR-Cas9 system that has been widely discussed is its high probability of producing off-target mutations, which has been clearly illustrated when a direct comparison between CRISPR and TALENs were made (Duan et al. 2014; Kim et al. 2016; Schaefer et al. 2017; Xu et al. 2015). As noted above, one attractive feature of TALENs, which enables a higher level of specificity, is that the FokI nuclease domain will only cleave DNA when dimerized, which occurs when the two TALE-domains bind to DNA (on opposite strands) in close proximity to each other. Therefore, we used the TALEN approach to knockout H2A.Lap1 (and when I started this project, CRISPR was still in its infancy).

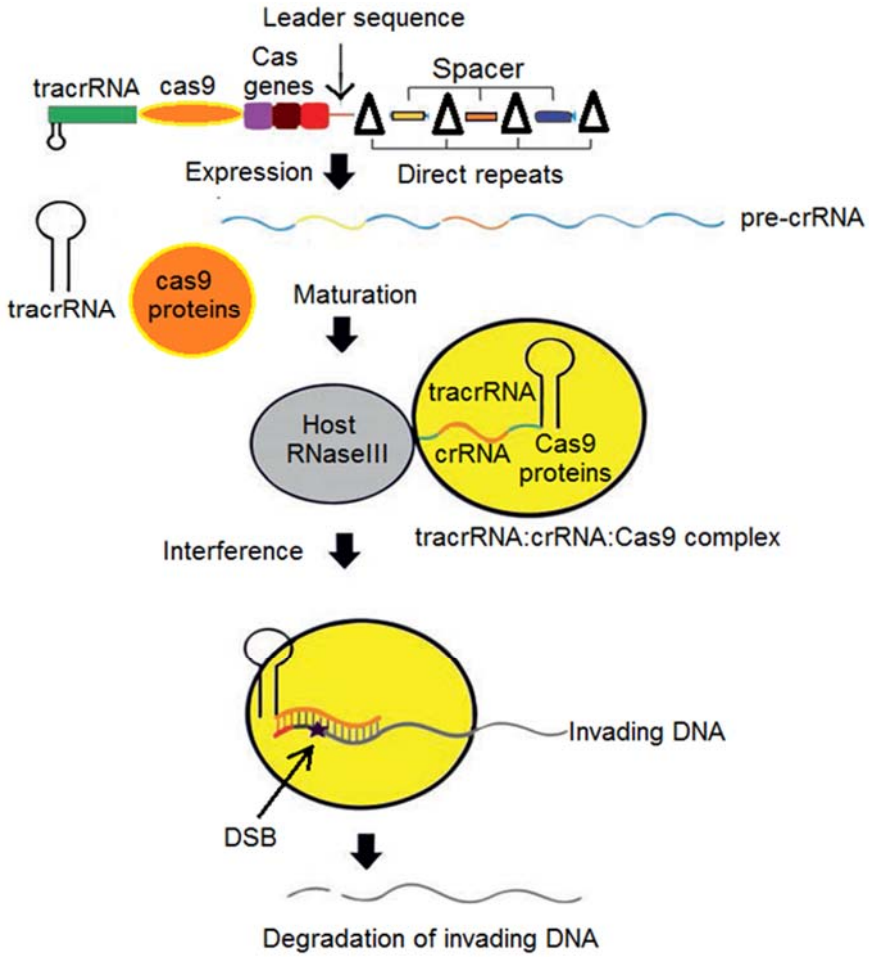


Figure 1-8: A mechanism of CRISPR-Cas9 action in bacterial cells. Adapted from Nemudryi et al. (2014)

1.10 Project objectives and chapters summary

The availability of a H2A.Lap1 knockout would be an invaluable tool for investigating the importance of this histone variant, however, as the H2A.Lap1 protein is expressed from three different genes, this is a technically challenging feat. Our aim was to devise the most efficient and specific strategy to inhibit the function of all three genes. In this thesis, we report a simplified TALEN approach that achieves this goal. Specifically, we used only one pair of TALENS to do this. To our knowledge, this has not been done before.

The aims of the project were to:

1- Generate an H2A.Lap1 knockout (KO) mouse colony using the TALEN system.
2- To determine the specificity of the TALEN approach by investigating, whether there were any off-target effects.
3- Perform an initial characterisation of the genotype and phenotype of H2A.Lap1 knockout mice.

In chapter 3, in collaboration with Phillip Gregory and Lei Zhang of Sangamo Biosciences, we present the design of the TALEN plasmids that specifically target the H2A.Lap1 encoding genes. We screen their activity efficiencies in transiently transfected cells: this includes a list of plasmids designed, a description of the Dual-Luciferase Single Strand Annealing Assay (DLSSA) to test plasmids' activities in cells, and the surveyor nuclease assay to detect mutations in cells. Next, we show how the plasmids' cleavage activities were validated in our cell lines, and how we developed our screening methods for mutations using a surveyor nuclease assay, T7 endonuclease I assay, and the use of the PCR cloning sequencing assay (which we simplified) for efficient high-throughput genotype screening in mice. Next, the microinjection of H2A.Lap1TALEN mRNA into single cell embryos of FVB/NJArc strain mice, which were implanted it into pseudopregnant females to generate H2A.Lap1KO founder mice is described (in collaboration with Prof Peter Koopman, the University of Queensland). Nineteen founders were born and transported to JCSMR so we could determine their genotype. In all 19 founders that we have screened, we have shown that H2A.Lap1 encoded genes were successfully disrupted by frame shift mutations which led to the loss of H2A.Lap1 protein. We also demonstrate how we continued to successfully breed the founders to produce KO pups. In this chapter, I also describe how we further

characterized the genotype of the pups that stably carried their parents' mutation throughout several generations. Our findings also include the discovery of chimera mice resulting from genomic rearrangement.

In Chapter 4, I asked the question: “Do TALENs cause any off-target mutations?” To answer this, we conducted exome sequencing on H2A.Lap1KO mice from 3 generations (G0, G1 and G2). By using bioinformatics analysis on the exome sequencing data, we hypothesized that if there were any off-target effects, the number of variants picked up by bioinformatics analysis would be diluted by half, from one generation to the next, when heterozygous female (+/-) bred with wild-type (wt) (+/Y) partner (Mendelian law of segregation). We were able to show, in collaboration with our Bioinformatics collaborators (Dr. Matthew Field and Dr. Sebastian Kurscheid, JCSMR) that the variations observed showed no dilution effects in all three generations of mice, providing evidence of TALEN's specificity. However, their analysis identified 19 putative indels that were consistently present in all 3 generations. We validated these variants experimentally, using PCR and Sanger sequencing analysis, and proved that the indels were mouse strain variation specific to our inbred colony and not the result of TALEN off-target activity. This proved the remarkable specificity of a single pair of TALENs as genome editing tools on multicopy genes.

In Chapter 5, I then characterized the phenotype of our H2A.Lap1KO mice colony. Since H2A.Lap1 is highly expressed in testes, we explored the ability of the KO mice to reproduce. Here we show that male H2A.Lap1 KO mice are subfertile and that their subfertility is likely due to aberrant nuclear morphology epigenetic landscape in the round spermatids of the KO mice. Since, the differentiation of the germ cells underwent significant epigenetic and chromosome organisation changes, we examined whether the absence of H2A.Lap1 affected the localisation and abundance of important active and repressive chromatin marks. Finally, we observed a connection between loss of H2A.Lap1 and aberrant function of somatic Sertoli cells that do not express H2ALap1.

References

Albert, I. et al., 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature*, 446(7135), pp.572–576. Available at: http://www.nature.com/articles/nature05632 [Accessed January 9, 2018].

Aliperti, V. & Donizetti, A., 2016. Long Non-coding RNA in Neurons: New Players in Early Response to BDNF Stimulation. *Frontiers in molecular neuroscience*, 9, p.15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26973456 [Accessed March 23, 2017].

Allis, D.C., Jenuwein, T. & Reinberg, D., 2007. Chapter 3: Overview and Concepts. *Epigenetics*, pp.23–40.

Anderson, J.. & Widom, J., 2000. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *Journal of Molecular Biology*, 296(4), pp.979–987. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10686097 [Accessed March 26, 2017].

Arents, G. & Moudrianakis, E.N., 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, 92(24), pp.11170–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7479959%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC40593.

Auclair, G. & Weber, M., 2012. Mechanisms of DNA methylation and demethylation in mammals. *Biochimie*, 94(11), pp.2202–2211. Available at: http://dx.doi.org/10.1016/j.biochi.2012.05.016.

Azevedo, C. & Saiardi, A., 2016. Why always lysine? The ongoing tale of one of the most modified amino acids. *Advances in Biological Regulation*, 60, pp.144–150. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26482291 [Accessed March 23, 2017].

Bannister, A.J. et al., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824), pp.120–124. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11242054 [Accessed March 23, 2017].

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21321607 [Accessed March 23, 2017].

Bao, Y. et al., 2004. Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *The EMBO Journal*, 23(16), pp.3314–3324. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15257289 [Accessed March 24, 2017].

Baptista, T. et al., 2013. Regulation of histone H2A.Z expression is mediated by sirtuin 1 in prostate cancer. *Oncotarget*, 4(10), pp.1673–1685. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24127549 [Accessed March 24, 2017].

Barski, A. et al., 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4), pp.823–837. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867407006009 [Accessed March 23, 2017].

Bednar, J. et al., 2017. Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Molecular Cell*, 66(3), p.384–397.e8. Available at: http://www.sciencedirect.com/science/article/pii/S109727651730268X?via%3Dihub [Accessed January 7, 2018].

Bell, O. et al., 2011. Determinants and dynamics of genome accessibility. *Nature reviews. Genetics*, 12(8), pp.554–64. Available at:

http://www.nature.com/doifinder/10.1038/nrg3017%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/21747402.

Berger, S.L. et al., 2009. An operational definition of epigenetics An operational definition of epigenetics. , pp.781–783.

Boettcher, M. & McManus, M.T., 2015. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular Cell*, 58(4), pp.575–585. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26000843 [Accessed March 24, 2017].

Bogdanove, A.J. & Voytas, D.F., 2011. TAL Effectors: Customizable Proteins for DNA Targeting. *Science*, 333(6051), pp.1843–1846. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21960622 [Accessed March 24, 2017].

Borgel, J. et al., 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nature genetics*, 42(12), pp.1093–100. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21057502.

Boros, J. et al., 2014. Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9 methylation to maintain heterochromatin protein 1α at chromatin. *Molecular and cellular biology*, 34(19), pp.3662–74. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4187721&tool=pmcentrez&rendertype=abstract.

Boyer, L.A. et al., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441(7091), pp.349–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16625203.

Bradley-Whitman, M.A. & Lovell, M.A., 2013. Epigenetic changes in the progression of Alzheimer's disease. *Mechanisms of Ageing and Development*, 134(10), pp.486–495. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24012631 [Accessed March 23, 2017].

Carbery, I.D. et al., 2010. Targeted Genome Modification in Mice Using Zinc-Finger Nucleases. *Genetics*, 186, pp.451–459. Available at: http://ko.cwru.edu/publications/Carbery.pdf [Accessed March 26, 2017].

Cermak, T. et al., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, 39(12), p.e82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21493687 [Accessed March 24, 2017].

Chadwick, B.P. & Willard, H.F., 2001. 2001 - Chadwick, Willard - A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome.pdf. , 152(2), pp.375–384.

Chakravarthy, S. et al., 2004. Structural Characterization of Histone H2A Variants. *Cold Spring Harbor Symposia on Quantitative Biology*, 69(0), pp.227–234. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16117653 [Accessed March 24, 2017].

Chakravarthy, S. et al., 2005. Structural characterization of the histone variant macroH2A. *Molecular and cellular biology*, 25(17), pp.7616–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16107708 [Accessed March 24, 2017].

Chandrasegaran, S. & Carroll, D., 2016. Origins of Programmable Nucleases for Genome Engineering. *Journal of Molecular Biology*, 428(5), pp.963–989. Available at: http://dx.doi.org/10.1016/j.jmb.2015.10.014.

Chen, T., Tsujimoto, N. & Li, E., 2004. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol Cell Biol*, 24(20), pp.9048–9058. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15456878.

Chouliaras, L. et al., 2013. Consistent decrease in global DNA methylation and

hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiology of Aging*, 34(9), pp.2091–2099. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23582657 [Accessed March 23, 2017].

Chuang, L.S.H. et al., 1997. Human DNA (cytosine-5) methyltransferase PCNA complex as a target for p21(WAF1). *Science*, 277(1992), p.1996–2000 ST–Human DNA (cytosine–5) methyltrans.

Clapier, C.R. & Cairns, B.R., 2009. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, 78(1), pp.273–304. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19355820 [Accessed March 23, 2017].

Clarkson, M.J. et al., 1999. Regions of variant histone His2AvD required for Drosophila development. *Nature*, 399(6737), pp.694–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10385122.

Clermont, Y., 1972. Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiological reviews*, 52(1), pp.198–236.

Clouaire, T. & Stancheva, I., 2008. Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cellular and Molecular Life Sciences*, 65(10), pp.1509–1522. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18322651 [Accessed January 8, 2018].

Collins, L.J., Schönfeld, B. & Chen, X.S., 2011. The Epigenetics of Non-coding RNA. *Handbook of Epigenetics: The New Molecular and Medical Genetics. Elsevier Inc.*, pp.49–61. Available at: http://pubman.mpdl.mpg.de/pubman/item/escidoc:1509110:5/component/escidoc:1509109/Epigenetics_n [Accessed March 26, 2017].

Condliffe, D. et al., 2014. Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiol Aging*, 35(8), pp.1850–1854. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4066184/pdf/nihms568398.pdf [Accessed March 23, 2017].

Cooper, S. et al., 2014. Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Reports*, 7(5), pp.1456–1470. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24857660 [Accessed March 23, 2017].

Coppieters, N. et al., 2014. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiology of Aging*, 35(6), pp.1334–1344. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24387984 [Accessed March 23, 2017].

Cui, F. et al., 2014. Prediction of nucleosome rotational positioning in yeast and human genomes based on sequence-dependent DNA anisotropy. *BMC bioinformatics*, 15(1), p.313. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25244936 [Accessed March 23, 2017].

Cui, X. et al., 2011. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nature biotechnology*, 29(1), pp.64–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21151125.

Cuthbert, G.L. et al., 2004. Histone deimination antagonizes arginine methylation. *Cell*, 118(5), pp.545–553.

DeFrancesco, L., 2011. Move over ZFNs. *Nature biotechnology*, 29(8), pp.681–4.

Doench, J.G. et al., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), pp.184–191. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26780180 [Accessed January 10, 2018].

Doudna, J.A. & Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), pp.1258096–1258096. Available at:

http://www.sciencemag.org/content/346/6213/1258096.long.
Doyen, C.-M. et al., 2006. Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *The EMBO Journal*, 25(18), pp.4234–4244. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16957777 [Accessed January 9, 2018].
Duan, J. et al., 2014. Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Research*, 24(8), pp.1009–1012. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24980957 [Accessed January 10, 2018].
Durai, S. et al., 2005. Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Research*, 33(18), pp.5978–5990.
Eissenberg, J.C. & Shilatifard, A., 2010. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental biology*, 339(2), pp.240–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19703438 [Accessed March 23, 2017].
Escamilla-Del-Arenal, M. et al., 2013. Cdyl, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. *Molecular and cellular biology*, 33(24), pp.5005–20. Available at: http://mcb.asm.org/content/33/24/5005.long.
Faast, R. et al., 2001. Histone variant H2A.Z is required for early mammalian development. *Current Biology*, 11(15), pp.1183–1187.
Fan, J.Y. et al., 2004. H2A.Z Alters the Nucleosome Surface to Promote HP1α-Mediated Chromatin Fiber Folding. *Molecular Cell*, 16(4), pp.655–661. Available at: https://www.sciencedirect.com/science/article/pii/S1097276504006495 [Accessed January 9, 2018].
Feng, S., Jacobsen, S.E. & Reik, W., 2010. Epigenetic Reprogramming in Plant and Animal Development. , 330(6004), pp.622–627.
Gaj, T., Gersbach, C.A. & Barbas, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, 31(7), pp.397–405. Available at: http://dx.doi.org/10.1016/j.tibtech.2013.04.004.
Gao, F. & Das, S.K., 2014. Epigenetic regulations through DNA methylation and hydroxymethylation: clues for early pregnancy in decidualization. *Biomolecular concepts*, 5(2), pp.95–107. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25372745 [Accessed March 22, 2017].
Geurts, A.M. et al., 2009. Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. *Science*, 325(5939), pp.433–433. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19628861 [Accessed March 24, 2017].
Gibson, T.J., Seiler, M. & Veitia, R. a, 2013. The transience of transient overexpression. *Nature methods*, 10(8), pp.715–21. Available at: http://dx.doi.org/10.1038/nmeth.2534.
Gkikopoulos, T. et al., 2011. A Role for Snf2-Related Nucleosome-Spacing Enzymes in Genome-Wide Nucleosome Organization. *Science*, 333(6050), pp.1758–1760. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21940898 [Accessed March 23, 2017].
Goll, M.G. & Bestor, T.H., 2005. Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry*, 74(1), pp.481–514. Available at: http://www.annualreviews.org/doi/10.1146/annurev.biochem.74.010904.153721.
González-Romero, R. et al., 2008. Quickly evolving histones, nucleosome stability and chromatin folding: All about histone H2A.Bbd. *Gene*, 413(1–2), pp.1–7.
Govin, J. et al., 2007. Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *The Journal of Cell Biology*, 176(3),

pp.283–294. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17261847 [Accessed March 24, 2017].
Govin, J. et al., 2004. The role of histones in chromatin remodelling during mammalian spermiogenesis. *European Journal of Biochemistry*, 271(17), pp.3459–3469. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15317581 [Accessed March 24, 2017].
Grant, P. a, 2001. A tale of histone modifications. *Genome biology*, 2(4), p.REVIEWS0003.
Greaves, I.K. et al., 2007. H2A.Z contributes to the unique 3D structure of the centromere. *Proceedings of the National Academy of Sciences of the United States of America*, 104(2), pp.525–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17194760 [Accessed March 24, 2017].
Greaves, I.K. et al., 2006. The X and Y chromosomes assemble into H2A.Z-containing [corrected] facultative heterochromatin [corrected] following meiosis. *Molecular and cellular biology*, 26(14), pp.5394–405. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16809775 [Accessed March 24, 2017].
Greco, C.M. & Condorelli, G., 2015. Epigenetic modifications and noncoding RNAs in cardiac hypertrophy and failure. *Nature Reviews Cardiology*, 12(8), pp.488–497. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25962978 [Accessed March 23, 2017].
Guibert, S. & Weber, M., 2013. *Functions of DNA Methylation and Hydroxymethylation in Mammalian Development* 1st ed., Copyright © 2013 Elsevier Inc. All rights reserved. Available at: http://dx.doi.org/10.1016/B978-0-12-416027-9.00002-4.
Gupta, A. et al., 2011. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Research*, 39(1), pp.381–392.
Gutschner, T., Baas, M. & Diederichs, S., 2011. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Research*, 21(11), pp.1944–1954.
Hall, B., Limaye, A. & Kulkarni, A.B., 2009. Overview: Generation of Gene Knockout Mice. In *Current Protocols in Cell Biology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. Unit 19.12 19.12.1-17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19731224 [Accessed March 24, 2017].
Hassa, P.O. et al., 2006. Nuclear ADP-Ribosylation Reactions in Mammalian Cells: Where Are We Today and Where Are We Going? *Microbiology and Molecular Biology Reviews*, 70(3), pp.789–829. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16959969 [Accessed March 23, 2017].
Hua, S. et al., 2008. Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression. *Molecular systems biology*, 4, p.188. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18414489 [Accessed March 24, 2017].
Jiang, C. & Pugh, B.F., 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics*, 10(3), pp.161–172. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19204718 [Accessed March 23, 2017].
Jin, C. et al., 2009. H3.3/H2A.Z double variant–containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nature Genetics*, 41(8), pp.941–945. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19633671 [Accessed March 24, 2017].
Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews. Genetics*, 13(7), pp.484–92. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/22641018.
Kadoch, C. & Crabtree, G.R., 2015. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Science Advances*, 1(5), pp.e1500447–e1500447. Available at: http://advances.sciencemag.org/content/1/5/e1500447.abstract.
Kamakaka, R.T. & Biggins, S., 2005. Histone variants: deviants? *Genes & Development*, 19(3), pp.295–316. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15687254 [Accessed March 23, 2017].
Kamdar, S.N. et al., 2016. Dynamic interplay between locus-specific DNA methylation and hydroxymethylation regulates distinct biological pathways in prostate carcinogenesis. *Clinical epigenetics*, 8(1), p.32. Available at: http://www.clinicalepigeneticsjournal.com/content/8/1/32%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4791926&tool=pmcentrez&rendertype=abstract.
Khare, T. et al., 2012. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature Structural & Molecular Biology*, 19. Available at: http://www.nature.com/nsmb/journal/v19/n10/pdf/nsmb.2372.pdf [Accessed March 22, 2017].
Kim, D. et al., 2016. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Research*, 26(3), pp.406–415. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26786045 [Accessed January 10, 2018].
Kim, H. & Kim, J.-S., 2014. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*, 15(5), pp.321–334. Available at: http://www.nature.com/doifinder/10.1038/nrg3686 [Accessed July 19, 2017].
Kim, K. et al., 2013. Gene dysregulation by histone variant H2A.Z in bladder cancer. *Epigenetics & chromatin*, 6(1), p.34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24279307 [Accessed March 26, 2017].
Kleinstiver, B.P. et al., 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), pp.490–495. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26735016 [Accessed January 10, 2018].
Kooistra, S.M. & Helin, K., 2012. Molecular mechanisms and potential functions of histone demethylases. *Nature Reviews Molecular Cell Biology*, 13(5), pp.297–311. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22473470 [Accessed March 23, 2017].
Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705.
Kretz, M. et al., 2012. Suppression of progenitor differentiation requires the long noncoding RNA ANCR Email alerting service Suppression of progenitor differentiation requires the long noncoding RNA ANCR. , pp.338–343.
Kriaucionis, S. & Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)*, 324(5929), pp.929–30. Available at: http://www.sciencemag.org/content/324/5929/929.long.
Lans, H., Marteijn, J.A. & Vermeulen, W., 2012. ATP-dependent chromatin remodeling in the DNA-damage response. *Epigenetics & Chromatin*, 5(1), p.4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22289628 [Accessed March 23, 2017].
Latrick, C.M. et al., 2016. Molecular basis and specificity of H2A.Z–H2B recognition and deposition by the histone chaperone YL1. *Nature Structural & Molecular Biology*, 23(4), pp.309–316. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/26974126 [Accessed January 8, 2018].
Lee, J.T., 2009. Lessons from X-chromosome inactivation: Long ncRNA as guides and tethers to the epigenome. *Genes and Development*, 23(16), pp.1831–1842.
Leonhardt, H. et al., 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, 71(5), pp.865–873.
Li, G.-M., 2013. Decoding the Histone Code: Role of H3K36me3 in Mismatch Repair and Implications for Cancer Susceptibility and Therapy. *Cancer Research*, 73(21), pp.6379–6383. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24145353 [Accessed January 10, 2018].
Li, Y. et al., 2016. Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Molecular Cell*, 62(2), pp.181–193. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27105114 [Accessed March 23, 2017].
Lieleg, C. et al., 2015. Nucleosome spacing generated by ISWI and CHD1 remodelers is constant regardless of nucleosome density. *Molecular and cellular biology*, 35(9), pp.1588–605. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25733687 [Accessed March 23, 2017].
Lister, R. et al., 2013. Global epigenomic reconfiguration during mammalian brain development. *Science (New York, N.Y.)*, 341(mC), pp.629–643.
Liu, X., Li, B. & Gorovsky, M.A., 1996. Essential and Nonessential Histone H2A Variants in Tetrahymena thermophila. *MOLECULAR AND CELLULAR BIOLOGY*, 16(8), pp.4305–4311. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC231429/pdf/164305.pdf [Accessed March 24, 2017].
Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648), pp.251–260. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9305837 [Accessed March 26, 2017].
Luger, K., Dechassa, M.L. & Tremethick, D.J., 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology*, 13(7), pp.436–447. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22722606 [Accessed January 9, 2018].
Luk, E. et al., 2010. Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. *Cell*, 143(5), pp.725–736. Available at: http://dx.doi.org/10.1016/j.cell.2010.10.019.
Maehara, K. & Ohkawa, Y., 2016. Exploration of nucleosome positioning patterns in transcription factor function. *Scientific Reports*, 6(1), p.19620. Available at: http://www.nature.com/articles/srep19620 [Accessed March 23, 2017].
Mak, A.N.-S. et al., 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (New York, N.Y.)*, 335(6069), pp.716–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22223736 [Accessed March 24, 2017].
Martin, C. & Zhang, Y., 2005. The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology*, 6(11), pp.838–849. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16261189 [Accessed March 23, 2017].
Marzluff, W.F., Wagner, E.J. & Duronio, R.J., 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature Reviews Genetics*, 9(11), pp.843–854. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18927579 [Accessed March 23, 2017].
Masliah-Planchon, J. et al., 2015. SWI/SNF Chromatin Remodeling and Human Malignancies. *Annual Review of Pathology: Mechanisms of Disease*, 10(1), pp.145–171. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25387058 [Accessed March 23, 2017].

Mattiroli, F., D'Arcy, S. & Luger, K., 2015. The right place at the right time: chaperoning core histone variants. *EMBO reports*, 16(11), pp.1454–1466. Available at: http://embor.embopress.org/cgi/doi/10.15252/embr.201540840 [Accessed March 23, 2017].

Moss, S.B., Challoner, P.B. & Groudine, M., 1989. Expression of a novel histone 2B during mouse spermiogenesis. *Developmental Biology*, 133(1), pp.83–92.

Mussolino, C. et al., 2014. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Research*, 42(10), pp.6762–6773. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24792154 [Accessed March 24, 2017].

Nathan, D. et al., 2006. Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes & Development*, 20(8), pp.966–976. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16598039 [Accessed March 23, 2017].

Nekrasov, M. et al., 2012. Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. *Nature Structural & Molecular Biology*, 19(11), pp.1076–1083. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23085713 [Accessed March 24, 2017].

Nelson, C.J., Santos-Rosa, H. & Kouzarides, T., 2006. Proline Isomerization of Histone H3 Regulates Lysine Methylation and Gene Expression. *Cell*, 126(5), pp.905–916. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16959570 [Accessed March 23, 2017].

Nemudryi, A.A. et al., 2014. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta naturae*, 6(3), pp.19–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25349712 [Accessed March 24, 2017].

Nowak, S.J. & Corces, V.G., 2004. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends in Genetics*, 20(4), pp.214–220. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15041176 [Accessed March 23, 2017].

Ocampo, J. et al., 2016. The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing *in vivo*. *Nucleic Acids Research*, 44(10), pp.4625–4635. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26861626 [Accessed March 23, 2017].

Okano, M. et al., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3), pp.247–257.

Palazzo, A.F. & Lee, E.S., 2015. Non-coding RNA: what is functional and what is junk? *Frontiers in Genetics*, 6, p.2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25674102 [Accessed March 23, 2017].

Pande, V., 2016. Understanding the Complexity of Epigenetic Target Space. *Journal of Medicinal Chemistry*, 59(4), pp.1299–1307.

Patel, J., Pathak, R.R. & Mujtaba, S., 2011. The biology of lysine acetylation integrates transcriptional programming and metabolism. *Nutrition & metabolism*, 8, p.12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21371315 [Accessed March 23, 2017].

Pattanayak, V. et al., 2011. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature Methods*, 8(9), pp.765–770. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21822273 [Accessed March 24, 2017].

Pedersen, M.T. & Helin, K., 2010. Histone demethylases in development and disease. *Trends in Cell Biology*, 20(11), pp.662–671. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20863703 [Accessed March 23, 2017].

Petri, A. et al., 2015. Long noncoding RNA expression during human B-cell

development. *PLoS ONE*, 10(9), pp.1–19.
Phelan, M.L. et al., 1999. Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. *Molecular cell*, 3(2), pp.247–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10078207 [Accessed March 23, 2017].
Radman-Livaja, M. & Rando, O.J., 2010. Nucleosome positioning: How is it established, and why does it matter? *Developmental Biology*, 339(2), pp.258–266. Available at: http://dx.doi.org/10.1016/j.ydbio.2009.06.012.
Rando, O.J. & Chang, H.Y., 2009. Genome-Wide Views of Chromatin Structure. *Annual Review of Biochemistry*, 78(1), pp.245–271. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19317649 [Accessed March 23, 2017].
Rangasamy, D. et al., 2003. Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *The EMBO journal*, 22(7), pp.1599–607. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12660166 [Accessed March 24, 2017].
Rinn, J.L. & Chang, H.Y., 2012. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, 81(1), pp.145–166. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22663078 [Accessed March 23, 2017].
Rose, N.R. & Klose, R.J., 2014. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(12), pp.1362–1372. Available at: http://dx.doi.org/10.1016/j.bbagrm.2014.02.007.
Runge, J.S., Raab, J.R. & Magnuson, T., 2016. Epigenetic Regulation by ATP-Dependent Chromatin-Remodeling Enzymes. In *Current topics in developmental biology*. pp. 1–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26969969 [Accessed March 23, 2017].
Saksouk, N. et al., 2015. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin*, 8(1), p.3. Available at: http://www.epigeneticsandchromatin.com/content/8/1/3 [Accessed March 23, 2017].
Sander, J.D. et al., 2013. In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. *Nucleic Acids Research*, 41(19), pp.1–7.
Santisteban, M.S., Kalashnikova, T. & Smith, M.M., 2000. Histone H2A.Z Regulates Transcription and Is Partially Redundant with Nucleosome Remodeling Complexes. *Cell*, 103(3), pp.411–422. Available at: http://www.sciencedirect.com/science/article/pii/S0092867400001331 [Accessed January 9, 2018].
Sarma, K. & Reinberg, D., 2005. Histone variants meet their match. *Nature reviews. Molecular cell biology*, 6(2), pp.139–149.
Schaefer, K.A. et al., 2017. Unexpected mutations after CRISPR–Cas9 editing in vivo. *Nature Methods*, 14(6), pp.547–548. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28557981 [Accessed January 10, 2018].
Shilatifard, A., 2006. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. *Annual Review of Biochemistry*, 75(1), pp.243–269. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16756492 [Accessed March 23, 2017].
Shin, J. et al., 2017. Disabling Cas9 by an anti-CRISPR DNA mimic. , (July), pp.1–10. Available at: http://advances.sciencemag.org/content/advances/3/7/e1701620.full.pdf.
Sienski, G., Dönertas, D. & Brennecke, J., 2012. Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene

Expression. *Cell*, 151(5), pp.964–980. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23159368 [Accessed March 23, 2017].
Soboleva, T.A. et al., 2017. A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. Available at: http://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.1006633&type=printable [Accessed March 24, 2017].
Soboleva, T.A. et al., 2012. A unique H2A histone variant occupies the transcriptional start site of active genes. *Nature Structural & Molecular Biology*, 19(1), pp.25–30. Available at: http://www.nature.com/doifinder/10.1038/nsmb.2161 [Accessed March 24, 2017].
Soboleva, T.A. et al., 2014. Histone variants at the transcription start-site. *Trends in Genetics*, 30(5), pp.199–209. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24768041 [Accessed January 8, 2018].
Song, C.-X. et al., 2016. Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation. *Proceedings of the National Academy of Sciences* , 113(16), pp.4338–4343. Available at: http://www.pnas.org/content/113/16/4338.abstract.
Song, J. et al., 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science (New York, N.Y.)*, 331(6020), pp.1036–40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4689315&tool=pmcentrez&rendertype=abstract.
Sterner, D.E. & Berger, S.L., 2000. Acetylation of histones and transcription-related factors. *Microbiology and molecular biology reviews : MMBR*, 64(2), pp.435–459.
Stroud, H. et al., 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome biology*, 12(6), p.R54. Available at: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-6-r54.
Struhl, K. & Segal, E., 2013. Determinants of nucleosome positioning. *Nature Structural & Molecular Biology*, 20(3), pp.267–273. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23463311 [Accessed March 23, 2017].
Suto, R.K. et al., 2000. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nature structural biology*, 7(12), pp.1121–1124.
Svotelis, A. et al., 2010. H2A.Z overexpression promotes cellular proliferation of breast cancer cells. *Cell Cycle*, 9(2), pp.364–370. Available at: http://www.tandfonline.com/doi/abs/10.4161/cc.9.2.10465 [Accessed March 24, 2017].
Szulwach, K.E. et al., 2011. 5-hmC–mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature Neuroscience*, 14(12), pp.1607–1616. Available at: http://www.nature.com/doifinder/10.1038/nn.2959 [Accessed March 22, 2017].
Tahiliani, M. et al., 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. , 324(5929), pp.930–935.
Tan, M. et al., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, 146(6), pp.1016–1028.
Vardabasso, C., Hake, S.B. & Bernstein, E., 2015. Histone variant H2A.Z.2: A novel driver of melanoma progression. *Http://Dx.Doi.Org/10.1080/23723556.2015.1073417*, 3556(July), pp.9–11. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/27308593%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4905390.

Villar-Menéndez, I. et al., 2013. Increased 5-Methylcytosine and Decreased 5-Hydroxymethylcytosine Levels are Associated with Reduced Striatal A2AR Levels in Huntington's Disease. *NeuroMolecular Medicine*, 15(2), pp.295–309. Available at: http://link.springer.com/10.1007/s12017-013-8219-0 [Accessed March 22, 2017].

Volle, C. & Dalal, Y., 2014. Histone variants: The tricksters of the chromatin world. *Current Opinion in Genetics and Development*, 25(1), pp.8–14. Available at: http://dx.doi.org/10.1016/j.gde.2013.11.006.

Wang, F. et al., 2013. Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease. *Human Molecular Genetics*, 22(18), pp.3641–3653. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23669348 [Accessed March 23, 2017].

Wang, K. et al., 2014. CARL lncRNA inhibits anoxia-induced mitochondrial fission and apoptosis in cardiomyocytes by impairing miR-539-dependent PHB2 downregulation. *Nature Communications*, 5, p.3596. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24710105 [Accessed March 23, 2017].

Wang, K.C. & Chang, H.Y., 2011. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell*, 43(6), pp.904–914. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21925379 [Accessed March 23, 2017].

Wang, T. et al., 2012. Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. *Human Molecular Genetics*, 21(26), pp.5500–5510. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23042784 [Accessed March 22, 2017].

Wang, Y. et al., 2004. Human PAD4 Regulates Histone Arginine Methylation Levels via Demethylimination. *Science*, 306(5694), pp.279–283. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15345777 [Accessed March 23, 2017].

Weber, C.M. & Henikoff, S., 2014. Histone variants: dynamic punctuation in transcription. *Genes & development*, 28(7), pp.672–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24696452 [Accessed March 24, 2017].

Winston, F. & Carlson, M., 1992. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends in genetics : TIG*, 8(11), pp.387–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1332230 [Accessed March 23, 2017].

Woodcock, C.L. & Ghosh, R.P., 2010. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology*, 2(5), p.a000596. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20452954 [Accessed March 24, 2017].

Wozniak, G.G. & Strahl, B.D., 2014. Catalysis-dependent stabilization of Bre1 fine-tunes histone H2B ubiquitylation to regulate gene transcription. *Genes & development*, 28(15), pp.1647–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25085417 [Accessed March 23, 2017].

Xu, P. et al., 2015. Both TALENs and CRISPR/Cas9 directly target the HBB IVS2-654 (C > T) mutation in β-thalassemia-derived iPSCs. *Scientific reports*, 5, p.12065. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26156589 [Accessed January 10, 2018].

Young, M.D. et al., 2011. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Research*, 39(17), pp.7415–7427. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21652639 [Accessed March 23, 2017].

Yuan, G. & Zhu, B., 2012. Histone variants and epigenetic inheritance. *Biochimica et*

Biophysica Acta - Gene Regulatory Mechanisms, 1819(3–4), pp.222–229. Available at: http://dx.doi.org/10.1016/j.bbagrm.2011.06.007.

Yun, M. et al., 2011. Readers of histone modifications. *Cell research*, 21(4), pp.564–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21423274 [Accessed March 23, 2017].

Zhang, H., Roberts, D.N. & Cairns, B.R., 2005. Genome-Wide Dynamics of Htz1, a Histone H2A Variant that Poises Repressed/Basal Promoters for Activation through Histone Loss. *Cell*, 123(2), pp.219–231. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16239141 [Accessed March 26, 2017].

Zhang, T., Cooper, S. & Brockdorff, N., 2015. The interplay of histone modifications - writers that read. *EMBO Rep*, 16(11), pp.1467–1481. Available at: http://embor.embopress.org/content/embor/16/11/1467.full.pdf.

Zhang, Y. & Reinberg, D., 2001. Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes and Development*, 15(18), pp.2343–2360.

Zhou, J. et al., 2007. The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat Struct Mol Biol*, 14(11), pp.1070–1076. Available at: http://www.nature.com/nsmb/journal/v14/n11/pdf/nsmb1323.pdf.

Zhubi, A. et al., 2014. Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Translational Psychiatry*, 4(1), p.e349. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24448211 [Accessed March 22, 2017].

CHAPTER 2

MATERIALS AND METHODS

2.1 Reagents

Name	Supplier
1,4-Dithiothreitol (DTT)	Life Technologies
2-mercaptoethanol	Sigma
30 % Acrylamide:Bis (29:1)	Biorad
AEBSF	Enzo life science
Agarose Powder	Amresco
Agencourt AMpure XP beads	Beckman Coulter
Ammonium chloride	Sigma
Ampicillin	Sigma
BigDye system	ABI
Bolt® LDS sample buffer	Life Technologies
Bolt® Bis-Tris Polyacrylamide gel (4-12 %)	Life Technologies
Bolt® buffer system for protein gel (20x)	Life Technologies
BSA (Bovine serum albumin)	Sigma, Biomatik
Calcium chloride	AnalaR, Merck BDH
Complete protease inhibitor	Roche
Cel1 Surveyor ` Mutation Detection Kit	Transgenomic
DAPI	Roche
Dynabeads® Protein G	Life Technologies
RPMI	Sigma
di-Sodium hydrogen phosphate	Merck
DMSO	Sigma
dNTP mix	Life Technologies
EDTA	Life Technologies
EGTA	Sigma
Ethidium bromide	Amresco
FBS (fetal bovine serum)	Life Technologies
Flourescent mounting medium (Vectashield)	Vector laboratories
Formaldehyde	Chem-supply
Gel extraction kit	Qiagen
Glycerol	Sigma
HEPES	Sigma

Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma
Lipofectamine 2000	Invitrogen
L-glutamine	Life Technologies
Magna ChIP ™ Protein A magnetic beads	Milipore
Magnesium sulphate	Merck
N,N-dimethylformamide	Sigma
Nonidet P-40	Fluka
Nuclease-free water	Ambion
Paraformaldehyde	Life Technologies
pGEMT-easy vector	Promega
Plasmid mini purification kit	Promega
Potassium chloride	Merck
Potassium hydrogen carbonate	Merck
Power Syb Green kit	ABI
Propidium iodide	Sigma
Proteinase K	Sigma
PureYield ™ plasmid midiprep	Promega
Qubit® fluorometer protein kit	Life Technologies
Quick extract DNA solution	Epicenter
Recombinant *Taq* DNA polymerase	Life Technologies
RNase A	Qiagen
RPMI 1640	Life Technologies
SDS	Sigma
Sodium di-hydrogen phosphate	Merck
Sodium DL-lactate	Sigma, Santa Cruz
Sodium pyruvate	Life Technologies
SureSelect XT2 Mouse All Exon Kit	Agilent
Sucrose	Merck
SuperScript III first strand synthesis system	Life Technologies
T4 DNA ligase and buffer	Promega
Tris	Sigma
Triton® X-100	Sigma
Trizol reagent	Ambion
Trypan blue stain (0.4 %)	Invitrogen
Trypsin-EDTA 0.25 %	Life Technologies
Tween® 20	Sigma

2.2 Antibodies

Specific antigen	Species antibodies were raised	Isotype	Conjugated to	Clone & source/ manufacturer	Catalogue number
Mouse H2A.Z	Sheep	IgG	-	Polyclonal, South Australia Health & Medical Research Institute (SAHMRI) Antibody Production Services, SA, Australia.	-
Mouse H2A.Lap1	Sheep	IgG	-	Polyclonal, SAHMRI, Antibody Production Services, SA, Australia.	-
Mouse H2A.Lap2	Rabbit	IgG	-	Polyclonal, SAHMRI, Antibody Production Services, SA, Australia.	-
Mouse H3K9Me2	Rabbit	IgG	-	Abcam	ab7220
Mouse H3K9me3	Rabbit	IgG	-	Abcam	ab8898
LectinPNA Alexa Fluor® 488 conjugate	-	-	FITC	Invitrogen	L21409
Sheep IgG	Donkey		Cyanine Cy™3	Polyconal, Jackson ImmunoResearch	713-165-003
Rabbit IgG	Donkey	-	Cyanine Cy™3	Polyclonal, Jackson ImmunoResearch	713-165-152
Goat/Sheep IgG	Donkey	-	HRP	Polyclonal, Millipore	AB324P

2.3 Methods

2.3.1 Purification of antibody

2.3.1.1 Thrombin digest of 6-His tagged protein

To remove a Histidine tag from a 6xHis-tagged H2A.Lap1 and 6xHis-H2A.Lap2 proteins, a 0.1 u/µl of Thrombin was prepared by diluting with Thrombin cleavage buffer (TCB) which contains 200 mm of Tris, pH 8.4, 1.5 M NaCl, 25 mM of $CaCl_2$. In 200 µl reaction, 0.2-0.5 mg of protein was cleaved with 3 U of 0.1 u/µl of Thrombin. The reaction consists of 75 µl of 8 M urea in H_2O, 0.2-0.5 mg protein, 30 µl of 0.1u/µl Thrombin, 20 µl of 10x TCB and H_2O. The reagents were mixed well and dispensed into equal amount of 8 PCR tubes. Then the reaction was incubated at 25 °C for 16 hrs. Following incubation, 6 µl of 4x Bolt sample buffer was added to the tube and boiled for 3 minutes. Then, after a brief spin, the sample was loaded into 10-well 4-12% gradient acrylamide gel, with 1 well loaded with uncut sample and one well for protein marker. The proteins were separated on the gel at 200V for 28 minutes. Alternatively, the cleaved sample were stored at -20 °C until use.

2.3.1.2 Affinity purification of antibody

Separated proteins were transferred into PVDF membrane as mentioned in section 2.3.5.3. The transfer was done at 30 V overnight. Following that, the PVDF membrane was stained with 0.2 % Ponceau S containing 1 % acetic acid. Once the protein is visible, the membrane was rinsed several times with mili Q H_2O. Then, the strip containing the desired protein was cut off using a scalpel blade. After that, the cut strip containing protein of interest was blocked using 3 % BSA at room temperature for 1 hr. Next, the serum containing the appropriate IgGs was diluted 1:1 with the blocking solution and incubated with the pre-blocked strip at room temperature for 5 hrs with rotation.Following that, the strip was washed with 10 ml of 1x PBS twice, 10 minutes each, at room temperature with rotation. To elute the antibody, a small boat was made out of paraffin film and stuck onto a petri dish. Inside the petri dish, wet tissues were placed around the paraffin boat to create humidity. Then, the strip was placed gently into the paraffin boat and 200 µl of elution buffer containing 0.2 M glycine pH 2.8 and 1mM EGTA was added. The Petri dish was agitated at 280 rev/min for 20 minutes. Then, immediately, the eluted antibody was transferred into microcentrifuge tube and 0.1 volume of Tris (unpHed) was added to

the tube. The pH of the eluted antibody was measured by dropping 5 µl onto pH paper. Then, 0.1 volume of 10x PBS was added. The antibody was stored at 4 °C.

2.3.2 Preparation of gDNA

2.3.2.1 Preparation of gDNA for exome capture and for Next Generation Sequencing (NGS)

Three to five ear punches or 5 mm of tail were collected from each selected mouse. The tissues were snapped frozen in dry ice immediately after punches. Then, 500 µl of tissue lysis buffer (50mM Tris pH 8.0, 100 mM EDTA, 0.5 % SDS) and 2.5 µl of 20 mg/ml of Proteinase K were added to the tissues. The sample was incubated at 55 °C for 1 hour until fully digested. In every 10 minutes interval during the incubation, the tissues were passed through 19 G and 23 G needles 10 times to further aid the lysis. After that, 5 µl of RNase A (100 mg/ml) were added to the reaction tube and mixed several times by inversion. The tissues were further incubated at 37 °C for 30 minutes. Following the incubation, the digested sample was centrifuged at 16,000 g for 10 minutes at room temperature. Then the supernatant was transferred to a new tube and 2.5 M of NaCl were added to the tube. The sample was mixed by inversion and left at room temperature for at least 5 minutes. After that, the sample was centrifuged at 16,000 g for 10 minutes at 4 °C, and after that the supernatant was transferred to a new tube. This step was repeated another time to remove the white precipitate in the supernatant. Following that, an equal amount of ice-cold isopropanol was added to the supernatant and the tube was mixed by inversion, followed by incubation at -20 °C overnight to precipitate the DNA. The following day, the tube was centrifuged at 16,000 g for 20 minutes at 4 °C. The supernatant was carefully removed without disturbing the pellet. Then, the pellet was washed with 1 ml of freshly prepared ice-cold 70% ethanol followed by brief gentle vortexing to lift the pellet of the wall. After that, the tube was centrifuged at 16,000 g for 10 minutes. The washing step was repeated with 500 µl of cold 70% ethanol and spun for another 5 minutes. The supernatant was carefully removed without disturbing the pellet. Then the pellet was air dried to remove any remaining ethanol residue. The pellet was dissolved in pre-warmed dH_2O and incubated further at 55 °C for 10 minutes to completely dissolve the DNA. The concentration was measured using Nanodrop and the visualisation of DNA was done using 1 % agarose gel, post-stained with EtBr.

One µg of was used for the preparation of indexed gDNA libraries using SureSelect XT2 Mouse All Exon kit as per the manufacturer protocol. Prior to exome enrichment, the DNA was iniatially sheared and the quality of DNA was assessed using 2100 Bioanalyzer DNA Assay to ensure the average DNA fragments is around 150 to 200 bp. To capture the exome using the SureSelect XT2 Mouse All Exon Kit, the DNA ends were initially repaired and then purified using the AMPure XP beads. Then the 3′ ends of the DNA fragments were adenylated. Following that, the A-tailed DNA fragments were ligated to the indexing adaptor and subsequently purified using the AMPure XP beads. Then, the indexed libraries were purified via PCR program as described in the kit. After that, the amplified libraries were purified using AMPure XP beads followed by quality assessment using the 2100 Bioanalyzer DNA Assay. Then the sample was ready to be sent for sequencing using Illumina HiSeq 2000 as 100 bp paired-end reads. The output result was sent for bioinformatics analysis.

2.3.2.2 Preparation of gDNA for standard PCR.

One to three ear punches from each selected mouse were placed in a 1.5 ml tube containing 50 µl of Epicentre Quick Extract Solution 2.0. For gDNA from cell lines, about 1 x10^6 cells were washed with 1x PBS and harvested using 0.25 % trypsin-EDTA at 37°C for 5 minutes and spun at 1,200 rpm for 5 minutes in 15 ml tube. After that that the cell pellet was resuspended with 1 ml the culture media and transferred into 1.5 ml tube and lysed with 100 µl of Epicentre Quick Extract Solution. Then, the tube was spun at 10,000g for 1.5 minutes. Following that the tube was incubated at 65 °C for 15 minutes with agitation and centrifuged briefly and further incubated at 98 °C for 8 minutes. Subsequently, the tube was vortex for 10 seconds and centrifuged at 10,000 g for 3 minutes to pellet the digested tissues. The supernatant containing DNA can be used directly in a standard PCR setup.

2.3.3 PCR

2.3.3.1 Standard PCR for genotyping purpose

For each mouse, all three genes of H2A.Lap1 were amplified to determine the genotype (wild type (WT), homozygous, heterozygous or hemizygous). For each primer, the reagents in Table 1 were prepared in master mix and aliquoted in a 96-well plate. The PCR condition was set to 1 cycle of initial denaturation at 95 °C for 10 minutes, a 28 cycle of denaturation at 95 °C for 30 secs, annealing at 60 °C (for H2A.Lap1a), 65°C

(for H2A.Lap1b), 56 °C (for H2A.Lap1c) for 30 secs, extension at 72 °C for 1 minutes and 1 cycle of final extension at 72°C for 4 minutes. For gDNA PCR reaction on other than genotyping purpose, the annealing temperature was adjusted and optimised according to the primer's melting temperature.

Table 2-1: List of PCR ingredients.

Ingredients	1x Volume (µl)
10x Normal Platinum Buffer	2.5
50 mM $MgCl_2$	1.0
10 mM dNTP	0.5
10 µM Forward Primer	0.5
10 µM Reverse Primer	0.5
Platinum Taq polymerase 5 u/µl	0.1
gDNA	1.0
H_2O	18.9
TOTAL	25.0

2.3.3.2 Amplicons visualisation.

To visualise the amplicons, we used native polyacrylamide gel electrophoresis in order to detect small deletions such as 5 bp. The gel was prepared by adding the reagents in Table 2 in a 50 ml polypropylene tube. The mixture was homogenised by inverting the tube 4 times. Then immediately casted into 1 mm cassette. Twelve-well comb was placed on top of the cassette slowly from left to right to prevent bubble trap. The gel was left at room temperature for 15 minutes to polymerise. Higher gel concentration will solidify faster, hence, the mixing needs to be done quickly. Once the gel solidified, it was placed in Bolt® electrophoresis tank filled with 0.5 % TBE buffer (diluted with dH_2O from 5x TBE which consists of 54 g Tris base, 27.5 g boric acid and 20 ml of 0.5 M EDTA). While waiting for the gel to solidify, the PCR product was diluted to 20 times with 1 x TBE sample buffer (5x TBE buffer, 20 % glycerol, 0.004 % bromophenol blue and 0.004 % xylene cyanol, and H_2O). Then 5 µl of the diluted amplicons were loaded in to the well and run at 200 V for 60 minutes (for H2A.Lap1a, c) and 90 minutes (for H2A.Lap1b). Then, the visualisation was done by post-stain the gel in 0.5 µg/ml of EtBr for 10 minutes and viewed under UV.

For amplicons documentation other than genotyping, the presence of PCR products was visualised using 2 % agarose gel, pre-stained with 0.5 µg/ml of EtBr.

Table 2-2: List of ingredients for native PAGE gels

Ingredients	10 %	12 %	14 %
30% Acrylamide: Bis (29:1)	4.0 ml	4.8 ml	5.6 ml
sdH_2O	5.6 ml	4.8 ml	4.0 ml
5x TBE	2.4 ml	2.4 ml	2.4 ml
10 % APS (fresh)	200 µl	200 µl	200 µl
10 µl Temed	10 µl	10 µl	10 µl

2.3.3.3 Purification of PCR products using Agencourt AMPure XP kit

10 µl of PCR product were transferred from plate into 0.2 ml tube. Then, 20 µl of H_2O was added to the PCR product. After that, 1.4x (40 µl) of AMPure XP solution was added into the reaction tube. The mixture was homogenised thoroughly by pipetting 20 times. The sample was incubated at RT for 5 minutes. Then the tube was placed onto a magnet plate for 2 minutes. Beads were separated from solution to the wall of the tube. Whilst tube still on magnet plate, the cleared solution from tube was aspirated and discarded (Carefully not to disturb beads). Whilst tube still on magnet plate, 200 µl of 70 % EtOH was dispensed into the reaction tube and incubated for 30 seconds. Then the EtOH was aspirated and discarded. The washing step was repeated 2 times. Off the magnet plate, 30 µl of elution buffer was eluted into reaction tube and mixed for 10 times until homogenise. The tube was placed back onto the magnet plate and incubated at RT for 1 minute to separate beads. The eluent was transferred to a new tube. The purity of the amplicons was measured using nanodrop and then visualised using 2 % agarose gel that was pre-stained with EtBr.

2.3.3.4 Ligation of amplicons into pGEMT vector

Cloning and sequencing was done to confirm the sequence of the amplicons. To distinguish heterozygous mutations, individual PCR products were cloned into and at least 8-10 clones for each gene were sequenced by Sanger method.

The ligation was done according to the manufacturer protocol of pGEMT-easy vector. Briefly, in a 1.5 ml tube, the following reagents were added: 5 µl of 2x rapid ligation buffer, 1 µl of pGEMT-easy vector plasmid, 1 µl of T4 DNA ligase, 2 µl of purified PCR product and H_2O to 10 µl of total reaction volume. The mixture was homogenised by pipetting. Then it was left at room temperature for at least 2 hrs. Alternatively, the reaction can take place at 4 °C overnight.

2.3.3.5 Transformation

Following ligation, the recombinant vector was transformed into *E.coli* TOP10 competent cells. The tube with ligation reaction was centrifuged briefly. Then, 1 tube of TOP10 competent *E. Coli* cells was thawed on ice. For each ligation,100 µl of competent cells was added into 5 µl ligation followed by gentle flick to mix. And then the mixture was left on ice for 20 minutes. Following that, a heat shock treatment was done by

placing the tube in a water bath at 42 °C for 45-50 seconds followed by immediate cooling of the tube on ice for 2 minutes. Then the 900 µl of pre-warmed LB was added into reaction tube and incubated at 37 °C with shaking for 1.5 hours. After that, the reaction tube was centrifuged at 8,000 rpm to pellet the bacterial cells. Then 800 µl of medium was aspirated and discarded. Then the bacterial pellet was resuspended in the remaining 200 µl medium. A100 µl onto of the bacterial suspension was spread on to LB plate which contains 100 µg/ml of ampicillin, 80 µg/ml of X-gal and 0.5 mM of IPTG. The plate was incubated at 37 °C for 16-20 hrs.

2.3.3.6 DNA amplification from *E. Coli* colonies by PCR

The following day, only white colonies (which contain the insert) were picked from the plate and placed into 96-well plate containing 20 µl of H_2O. The suspension was subjected to PCR reaction to screen for the inserts. Briefly, in 96-well plate reaction, each well contains 5 µl of the bacterial suspension, 20 mM of universal T7 primer, 20 mM of SP6 primer, 10 µl of Immomix and H_2O to 20 µl of total reaction volume. Stand The standard PCR condition was set to 1 cycle of initial denaturation at 95 °C for 10 minutes, a 28 cycle of denaturation at 94 °C for 30 secs, annealing at 50 °C for 30 secs, extension at 72 °C for 1 minutes and 1 cycle of final extension at 72°C for 5 minutes. Following that, the PCR product was purified using Agencourt AMPure XP beads as described in section 2.3.3.3.

2.3.3.7 Sanger sequencing

The plasmids containing inserts were subjected to Sanger sequencing. Briefly, in 96-well plate, 5 µl of BigDye buffer was added with 3 pmole/µl of universal T7primer, 200-500 ng of plasmid, 1 µl of BigDye fluorophore and H_2O to 20 µl of total reaction. The mixture was subjected to PCR using the following parameters. Single cycle of initial denaturation at 94 °C for 5 minutes, 30 cycles of denaturation at 96 °C for 10 secs, annealing at 50 °C for 5 secs and extension at 60 °C for 4 minutes. Then the PCR product was subjected to purification: 5 µl of 125 mM EDTA was added at the bottom of the well, followed by the addition of 60 µl of 100 % EtOH. The plate was covered with aluminium seal and inverted 4 times to homogenise the mixture. Then the plate was subjected to centrifugation at 2,500 *g* for 30 minutes. Then, the supernatant was removed by inverting and centrifuged for up to 185 *g* for 1 minute. The washing step was repeated

with 100 µl of EtOH. The pellet was left to dry for 1 minutes before it was sent to the Biomolecular Research Facility, JCSMR for sequencing.

2.3.4 Preparation of cDNA

2.3.4.1 Extraction of RNA

RNA extraction from tissues and cell lines was performed using the TRIzol® reagent. Firstly, 1 ml of TRIzol® reagent was added to the frozen tissue sample. The tissues were passed through 19 G and 23 G needles attached to a 1mL syringe 10 times to homogenize. Then, the homogenized sample was incubated for 5 minutes at room temperature to permit complete dissociation of the nucleoprotein complex. Following the incubation, 2 ml of chloroform per 1 ml of TRIzol® reagent was added to the tube. The tube was securely capped and shook vigorously for 15 seconds. Then the tube was incubated at room temperature for 2-3 minutes. After that, the tube was centrifuged at 12,000 x*g* for 30 minutes at 4°C. Then the aqueous phase was transferred to a new tube. 10 µg of glycogen (used as an RNA carrier) was added to the aqueous phase and mixed by pipetting. Then, RNA was precipitated by addition of 0.5 ml of 100% isopropanol and mixed by inverting 6 times. Following that, the RNA was left to precipitate overnight at -20 °C. The next day, the tube was centrifuged at 12,000 x*g* for 30 minutes at 4°C. Then, the supernatant was removed by pipetting. The pellet was washed with freshly made 75 % ethanol. After that the tube was centrifuged at 10, 000 x*g* for 10 minutes at 4 °C. After that, the supernatant was removed by pipetting leaving about 5-10 µl of ethanol. Then, the tube was briefly centrifuged using tabletop microcentrifuge to bring down all solution residues. Following that, using 10 µl tips, the remaining ethanol was carefully aspirated and the pellet was left to dry for 1-2 minutes. The pellet was dissolved and mixed in pre-warmed nucleus-free water and incubated at 55 °C for 10 minutes. The concentration was measured by Nanodrop.

2.3.4.2 DNAse treatment of purified RNA

Following RNA extraction, 8-10 µg of RNA was subjected to DNAse treatment to eliminate remaining DNA. In 1.5 ml tube, 200 ng/µl of RNA was treated with TURBO 2 units of DNAse Turbo in 1x Turbo buffer in a 50 ul reaction at 37 °C for 30 minutes. Then, extra 2 units of Turbo DNAse were added and mixed by pipetting. The tube was further incubated another 30 minutes. The reaction was inactivated by mixing with 1/10 volume of DNAse inactivation reagent and left at room temperature for 5 minutes with

occasional mixing. Then the tube was centrifuged at 10,000 xg for 1.5 minutes. The supernatant containing treated RNA was transferred to a new tube and stored at -80 °C.

2.3.4.3 First strand cDNA synthesis using SuperScript III reverse transcriptase

The first strand cDNA synthesis was performed in a 200 µl PCR tube using 1-2 µg of Turbo DNase treated RNA (making sure that the volume does not not exceed 40% of total volume), 3 µl of 5x first strand buffer, 1 µl of 10 mM dNTPs mix, 0.5 µl of 50 µM oligo (dT) primer, 0.5 µl of random primer and nuclease-free H_2O to 15 µl. The mixture was mixed by pipetting followed by a quick spin and incubated at 65 °C for 5 minutes. Following the incubation, the tube was immediately coolled on ice for 2 minutes then spun briefly. Then, the rest of the reagents were added: 2 µl of 0.1 M DTT, 1 µl of first strand buffer, 0.4 µl (per µg of RNA) of 200u/µl SuperScript III reverse transcriptase and nuclease-free H_2O to 20 µl. The mixture were mixed and centrifuged briefly. The reaction tube was further incubated at 50 °C for 50 minutes followed by inactivation at 70 °C for 15 minutes. The synthesised cDNA was stored at -20 °C until used.

2.3.5 Real time PCR

The expression of genes was assessed using relative quantification and comparative ΔΔCt method. In a total reaction of 10 µL, 2.5 µM of forward and reverse primer (mixed), 5 µl of Power Syber green, 5-10 ng of cDNA and nuclease-free water were mixed. The PCR reaction took place in a 384-well plate in an ABI SDS 7900 real-time PCR machine using the following parameters: initial denaturation at 95 °C for 10 minutes, a total of 40 cycle of denaturation at 95 °C for 15s and annealing/extension at 60 °C for 1 min.

2.3.6 Preparation of protein

2.3.6.1 Preparation of total tissue lysate for Western blot

To prepare protein samples for PAGE, 2 different lysis buffers was used based on the type of tissues. The buffers are as listed in Table 2-3. Following extraction, the concentration of the protein was determined using nanodrop at absorbance 280 nM or Qubit Fluorometer® Protein Kits

2.3.6.2 Polyacrylamide gel electrophoresis (PAGE)

For comparison of expression, total protein concentration was standardised and lysates were mixed with 4x Bolt® sample loading buffer in reduced conditions by adding of 2.5 µl of 14.3M β-mercaptoethanol for every 20 µl of sample buffer. The sample was boiled for 5 minutes and instantly place on ice before loading into 4-12 % Bis-Tris polyacrylamide gel. Electrophoresis was performed in 1X Bolt® MOPS SDS Running Buffer at 200 V for 28 minutes.

2.3.6.3 Western blot transfer

In all western blot experiments, wet transfer was applied. All sponges and papers were pre-wet in 1x western transfer buffer (WTB, 25 mM Tris-HCL, 192 mM glycine, 10 % v/v methanol, pH8.3). PVDF membrane was pre-wet in 100 % methanol for 15 secs, rinsed in mili Q water for 2 minutes and in ice-cold 1x WTB for 5 minutes. Briefly, the gel and membrane are sandwiched between sponge and paper (sponge/paper/gel/membrane/paper/sponge) and all are clamped tightly together after ensuring that there is no air bubble between layers. The sandwich was submerged in 1x WTB and placed in an electrophoresis tank. The transfer process was done at 30 V for overnight in a cold room. Alternatively, it was carried out at 30v for 2 hours in a cold room.

Table 2-3: List of lysis buffer used for protein extraction

Sample	Lysis Buffer	Components
Nero2a, Neuroblastoma cell line	RIPA	50 mM Tris, pH 8.0, 1 % NP-40, 0.5 % sodium deoxycholate, 0.1 % SDS, 150 mM NaCl, 2 mM EDTA
Mouse germ cells (for IP)	iCLIP	50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate
Mouse germ cells for western blot	1x LSB	60 mM Tris-HCL pH 6.8, 2 % SDS, 10 % glycerol, 2.5 % β-mercaptoethanol

2.3.6.4 Incubation with primary and secondary antibody

All incubation and washing steps were done with agitation. Following transfer process, the membrane was instantly blocked with 3 % BSA (in 1x PBS and 0.05 % Tween 20) for 1 hour at room temperature. The remaining gel from the transfer was stained in Instant blue stain (Expedeon, cat. no. ISB1L) for 1 hour to assess the efficiency of the transfer process. Then, the membrane was incubated with primary antibody (H2A.Lap1- 1:5000) diluted in 1 % BSA/ PBS/0.05 % Tween at room temperature for 1 hour with agitation. Alternatively, the incubation was also done at 4 °C overnight. Following primary antibody incubation, the membrane was washed with PBST (1x PBS, 0.5 % Tween, 0.5 M NaCl) for 50 minutes at room temperature. After that, the membrane was washed with PBS/ 0.5 % Tween for 10 minutes at room temperature. Then, the membrane was incubated with secondary antibody (anti-sheep HRP- 1: 15,000) which was diluted in 1 % BSA/PBS,/0.05 % Tween) for 1 hour at room temperature. Following that, the membrane was washed 3 times with PBS/ 0.05 % Tween for 5 minutes at room temperature.

2.3.6.5 Image development

For HRP-conjugated antibodies, Immobilon Western Chemiluminescent HRP substrate was used. Working HRP substrate was prepared by combining equal volumes of Luminol Reagent and Peroxide Solution. The substrate was spread onto the membrane by pipetting and left for 5 minutes. Then the membrane was exposed to chemiluminescence in LAS 400 machine with various time of exposure (from 10 s to 3 minutes) depending on the strength of the signal. Alternatively, for a more sensitive signal detection, X-ray film was used.

2.3.7 Preparation of protein for immunoprecipitation (IP)

2.3.7.1 Beads preparation

To assess H2A.Lap1 protein from testes, 1 whole testicle was used in preparing the total cell lysate. Firstly, the procedure was begun by preparing the protein A and protein G beads for coupling to H2A.Lap1 antibody. Fifty µl of Magna ChIP™ Protein A Magnetic Beads (Milipore, cat. no. 16-661) and 50 µl of Dynabeads® Protein G (Life Technologies, cat. no. 10004D) were mixed in a 1.5 ml tube. The beads were washed by adding 900 µl of iCLIP buffer (50 mM Tris-HCl, pH 7.4; 100 mM NaCl; 1% NP-40; 0.1% SDS; 0.5% sodium deoxycholate) followed by inverting the tubes 10 times. To

remove the buffer, the tube was place in a magnetic holder and the supernatant was removed by pipetting. The washing step was repeated 3 times. Then the beads were resuspended with 200 µl of iCLIP buffer and 20 µl of sheep anti-H2A.Lap1 antibody. The mixture was mixed at RT for 1 hour on a rotator to allow conjugation of the protein to the antibody. Following the incubation, the beads were washed for 5 minutes with 900 µl of iCLIP buffer on a rotator. The washing step was repeated 3 times. Then, the beads were resuspended with 200 µl of iCLIP elution buffer (iCLIP buffer added with 1x Protease Inhibitor cocktail (Roche, cat. no. 05892791001) and 1 mM of AEBSF (Enzo Life Science, cat. no. ALX-270-022-G001)). The beads were kept at 4 °C until use.

2.3.7.2 Total cell lysate preparation

A cell pellet from 1 whole testicle was resuspended in 900 µl of iCLIP elution buffer in a 1.5 ml tube by pipetting on ice for 20 minutes. To facilitate the lyse, the cell pellet was passed through pre-chilled 1 ml syringe with 25G and 23G needle for 10 times. Then, to lyse the DNA from the cell, 4 µl of 2U/ µl TURBO DNase (Ambion, cat. no. AM1907) was added to the tube and mixed well prior to incubation at 37 °C for 10 minutes. Following the incubation, the reaction tube was centrifuged at 16,000 xg at 4 °C for 30 minutes. The supernatant was transferred to a new 1.5 ml tube and kept on ice. The pellet containing unbroken nuclei was resuspended with 200 µl of iCLIP elution buffer. The pellet was sonicated on ice for 5 minutes with 15 seconds pulse. Then the cleared supernatant was combined and centrifuged at 16,000 rpm for 15 minutes at 4 °C to pellet any remaining debris. The supernatant containing the total cell lysate were transferred to a new tube and the OD is measured at 280 nM. The sample was kept at -20 °C until use.

2.3.7.3 Immunoprecipitation assay

The total cell lysate (about 950 µl) were combined with 100 µl of antibody-conjugated beads (from section 3.1.1). The sample was mixed for 4 hours on a rotator at 4 °C. Following the incubation, the tube was placed on a magnetic holder and the flow through was transferred to a fresh 1.5 ml tube and kept at -20°C until use. The beads were washed for 5 minutes with high salt wash buffer (50mM Tris-HCl, pH7.4; 1M NaCl; 1mM EDTA; 1% NP-40; 0.1% SDS; 0.5% sodium deoxycholate) 2 times on a rotator at 4 °C. Following the high salt buffer washing step, the beads were washed twice with wash

buffer (20 mM Tris-HCl, pH 7.4; 10 mM MgCl2; 0.2% Tween-20) for 5 minutes on a rotator at 4 °C. Then, the buffer was removed. The beads were resuspended with 20 µl of reduced 1x Bolt® LDS Sample Buffer (which included 12.5 mM β-mercaptoethanol). The beads were heated at 80 °C with shaking at 800 rpm for 10 minutes. Following the heating, the tube was immediately placed on a magnetic holder and the eluate was collected and transferred to a new 1.5 ml tube and instantly placed on ice. The sample was kept at -20 °C until use. 097714898

2.4 Extraction of mouse tissues

2.4.1 Extraction of germ cells from testes for hypotonic spreads and fixation

All procedures that involved extraction of tissues were done under the approval of Biomedical Animal Ethics Protocol number A2014/33. All procedures of extraction were conducted in Containment Suite #3, JCSMR building.

A day before the tissues extraction, the following buffers were prepared and stored at 4 °C. For testes collection, the PBSGL buffer was used which consisted of 1x PBS, 5.6 mM glucose, 5.4 mM sodium DL-lactate, EDTA-free 1x protease inhibitor cocktail (was freshly added to the 1x PBSGL just prior to extraction).

The method is based on the well-known protocol (Peters et al. 1997). Briefly, mouse was ethically culled using cervical dislocation method. Once the mouse was confirmed dead, it was placed on its back on a paper towel. The abdominal part was cleaned with 70 % EtOH and an V section was made at about 1 cm above genitals. The skin was torn open and a cut was made in the muscle tissue. A long cut was made to expose the fat pads and the testicles. Then, the testes were pulled out and placed in a well of 12-well plate containing 2 ml PBSGL. Then, a small incision was made in the tunica albuginea and the seminiferous tubule was squeezed and transferred into a petri dish on ice containing 1 ml of PBSGL. The tubules were rinsed to remove blood and interstitial cells and cut for 1 minutes (on ice) with scissors to release the germ cells. Then, when the media became cloudy, all of the mixture was mixed gently by pipetting 10 times before transferring it into 1.5 ml tube. Then, the mixture was centrifuged at 1200 rpm for 5 minutes at 4 °C. The supernatant was aspirated and discarded, leaving about 20 µl of media in the tube. The cell pellet was resuspended gently by flicking the tube and 80-160 µl (depending on the size of the pellet) of hypotonic buffer (100mM sucrose, 3mM Tris-HCL pH 8.) was added. Then, the cells were left on ice for 5 minutes to swell up. Whilst waiting for the cells to swell up, hydrophobic circles were drawn on the slide.

Then, a thin layer of 2 % PFA in water, containing 0.05 % Triton, pH at 9.2, was spread inside the circle. Then, about 5-10 µl of the hypotonic cells was drop onto the slide and spread by rotating the slide. The cells were fixed in a humidity chamber for 2 hrs at room temperature. After that, the lid of the chamber was lifted off to let the slides to dry completely. The slides were stored in -80 °C.

For whole testicle fixation for IHC analysis, 10 % Neutral-Buffers Formalin was used (NBF), which consists of 9.2 M NaH_2PO_4, 33.3 mM Na_2HPO_4, 3.65% formaldehyde. A whole testicle was placed in a 15 ml tube containing 10 ml of 10% NBF. The tissue was left in the fixation solution and incubated at room temperature on a wheel with rotation. After 1 hour, the solution was carefully removed and a fresh 10 % NBF was added. The tissue was left to fix with rotation for 4-24 hours. Then, tissue was washed 3 times with fresh 70 % ethanol, 10 minutes for each wash. Finally, the fixed tissue was sent to the Histology lab of JCSMR Imaging and Cytometry Resource Facility for processing and paraffin embedding.

2.4.2 Sperm extraction from mouse cauda epididymis for sperm swim-up

Before the extraction, 4 ml of Donners medium (135 mM NaCl, 5 mM KCl 1 mM $MgSO_4$, 2 mM $CaCl_2$, 30 mM HEPES, pH 7.4, 25 mM $NaHCO_3$, 20 mg/ml BSA, 1 mM sodium pyruvate and 0.53 % (v/v) sodium DL-lactate) was placed in a 5 ml round bottom tube. The caudal epididymis was dissected using fine scissors and forceps. Then, 2-3 cuts were made in each epididymis. After that, the epididymis was gently placed at the bottom of round-bottom tube. Then, another 3 ml of Donners medium was added along the wall of the tube. Following that, the tube was incubated at 37 °C for 1 hr to allow sperm to swim up to the upper phase of the medium. After incubation, 2.5 ml of the upper phase was removed into 15 ml tube. Then, 10 µl of the suspension, was diluted 1:1 with 0.4 % trypan blue stain and viewed under microscope for morphological observation and cell count.

2.5 Immunofluorescence staining

Slides that were stored at -80 °C was equilibrated in a cold room for 10-15 minutes. Then, slides were placed in glass chamber containing ice cold PBS and placed at room temperature for 15-20 minutes to warm up. Alternatively, for wet-prep slide, after 2 hr fixation with 2 % PFA, the excess tissues were washed with water in the upright glass chamber gently. Then, the antigen blocking step was conducted by pipetting onto the

slide 3 % BSA/PBS for 2 hrs at room temperature in humidity chamber. In the meantime, primary antibody mix were prepared in 1 % BSA/PBS containing 0.1 % Tween®20. Then, the blocking mixture was removed by tapping the slides on a tissue. Immediately, primary antibody was pipetted onto the slide. The incubation was done at 4 °C overnight in humidity chamber. Alternatively, the slide was also incubated at 37 °C for 1 hr. After primary antibody incubation, the secondary antibody was prepared the same way as the primary. Then, after the removal of primary antibody, the slide was washed 3 times with 1x PBS 5 minutes for each wash. Then, the slide was incubated with secondary antibody which was prepared in the same buffer as the primary antibody and incubated for 1 hr at room temperature in dark. Following secondary antibody incubation, all steps were conducted in dark. The slide was washed 2 times with 1x PBS. Then, in the upright glass chamber, the side was stained with 0.2 mg/ml DAPI in water for 2-5 minutes. Then, the slide was washed with water for 1 minute. To protect the fluorescence signal, 10-20 µl of mounting medium (Vectashield) was dropped onto a coverslip, placed on the slide to cover the cell-containing area and sealed with nail polish. The slide was stored in the dark at 4 °C. Visualisation and documentation was done using Leica SP5 confocal microscope.

2.6 Mammalian cell culture and transfection

2.6.1 Mouse neuroblastoma 2a (Neuro2a) cell culture

Mouse Neuro 2a cells were cultured at 37 °C in a humidified atmosphere with 5 % CO_2 in RPMI medium supplemented with 10 % FBS and 2 nM L-glutamine supplemented with penicillin (100 U/ml) and streptomycin (100µg/ml).

2.6.2 Transfection with Lipofectamine 2000

A day before transfection, 5 x 10^5 of cells per well were seeded in a 6-well dish. Then, transfection was done according to manufacturer's protocol. The plasmid DNA concentration was adjusted to 0.5 µg/µl and a total of 250 ng of DNA were transfected into the cells in a serum-antibiotic-free condition. After 5 hrs, the medium was changed to normal growth medium. The cells were allowed to grow for another 24-48 hrs, prior to collection for further experiments.

2.6.3 Cel1 assay

Forty-eight hours post-transfection, the cells were harvested and subjected to gDNA extraction as described in section 2.3.2.2. Then, the PCR amplification was conducted to amplify all 3 H2A.Lap1 genes as described in section 2.3.3.1. After confirming the presence of amplicons in 2 % agarose gel electrophoresis, 10 µl of the PCR product was transferred into 200 µl PCR tube and subjected to annealing program as follow: 95 °C 10 mins, slow-gradient decrease from 95 °C to 85 °C at 2°C per second, rapid-gradient decrease from 85 °C to 25 °C at 0.1 °C per second and 4 °C as holding temperature. Then, 1 µl of Cel1 enzyme was added and mixed to the tube and incubated at 42 °C for 30 minutes. After that, the tube was placed on ice to stop the reaction. Then, digested sample were prepared for native PAGE for visualisation and documentation as described in section 2.3.3.2.

Reference

Peters, A.H. et al., 1997. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 5(1), pp.66–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9088645 [Accessed January 10, 2018].

CHAPTER 3

TALEN PLASMID DESIGN AND THE ESTABLISHMENT OF H2A.LAP1 KNOCKOUT (H2A.LAP1KO) MOUSE COLONY

3.1 TALEN plasmid design and preliminary testing by Sangamo

Transcription activator-like (TAL) effector nucleases (TALENs) are highly effective in inducing mutations at specific genomic loci and are a potential alternative to conventional gene targeting in mice. TAL proteins are transcriptional activators from *Xanthomonas* bacteria, that can bind to specific host DNA sequences and regulate plant genes during pathogenesis. Within the TAL structure, a central repeat domain mediates DNA recognition, with each repeat unit of 33–35 amino acids specifying one target base. DNA binding specificity is determined by the amino acids at positions 12 and 13 within each repeat, referred to as the repeat variant di-residue (RVD) (Figure 3-1). These motifs determine the base preference (Boch et al. 2009; Christian et al. 2010; Miller et al. 2011).

The design of TALENs to target the H2A.Lap1 encoding gene was carried out by our collaborators from US-based company, Sangamo Ltd. At the time of TALEN design, in 2011, the Ensembl release 57 (March 2010) database listed four H2A.Lap1-coding genes in mice: H2Afb3 (mH2A.Lap1a), Gm14920 (mH2A.Lap1b), H2Afb2 (mH2A.Lap1c), and a pseudogene Gm14904, which was subsequently removed from all assemblies in later releases. The H2A.Lap1 genes are over 92% identical in DNA sequence (with H2A.Lap1a and H2A.Lap1b being 98% identical), while the pseudogene is only 87% identical. All four genes are located on the X-chromosome (Figure 3-2, Table 3-1). The alignments of the four H2A.Lap1 genes are shown in Figure 3-3.

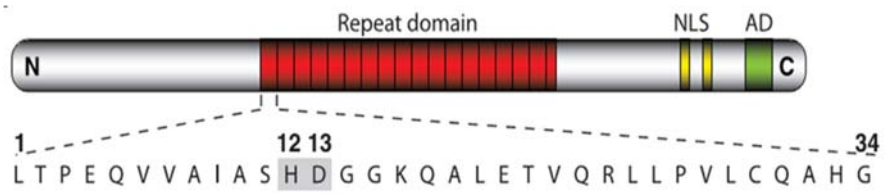


Figure 3-1: Schematic diagram of TALEN construct.

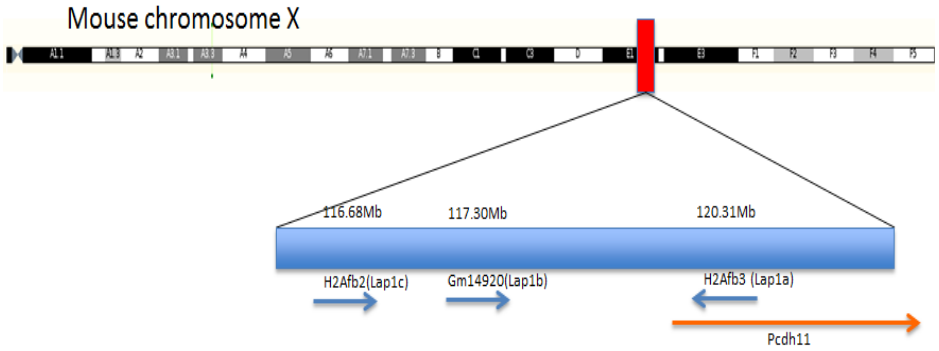


Figure 3-2: H2A.Lap1 encoded genes on the X-chromosome

Table 3-1: Genomic coordinates for H2A.Lap1 encoding genes.

Gene Names	Origin strand	Chromosome Name	Locus start	Locus end	% identical	Length
H2afb3 ENSMUSG00000083616 (H2A.Lap1a)	-	X	117426357	117426704	100	348
Gm14920 ENSMUSG00000067441 (H2A.Lap1b)	+	X	114128366	114128713	98.28	348
H2afb2 EG624153 (H2A.Lap1c)	+	X	113794787	113795134	92.55	349
Gm14904 ENSMUSG00000084053 pseudo	+	X	113311798	113312139	87.83	345

Figure 3-3: The alignment of H2A.Lap1 encoded genes.

Although the aim of this project was to target all H2A.Lap1 genes with one pair of TALEN constructs, two groups of H2A.Lap1-targeting TALENs were designed to test the specificity of TALENs given the high similarity between the three H2A.Lap1 genes. One group targeted sequences present in all four H2A.Lap1 genes, while the second group targeted H2A.Lap1a gene only, subsequently shown to be the main gene expressed in the testes (see Table 3-2). The second group was designed to test how specific TALEN targeting can be in the case of highly homologous genes. The alignment of these plasmids in the H2A.Lap1 genes is shown in Figure 3-4 (A). The nucleotide specificity of these plasmids lies in the 12^{th} and 13^{th} position of the repeat variable di-residue (RVD) in TAL effector repeats (shown in Figure 3-4 (B)).

Zhang and colleague of Sangamo Bioscience, Inc, designed 19 plasmids (Table 3-2) to target various regions of the genes on both DNA strands. The plasmids were paired to 22 pairs to find the pair that produces the highest activity, ensuring that a construct targeting a plus strand is paired with a construct targeting a minus strand, in its vicinity. In order to test the nuclear activity of FokI cleavage, each combination of two TALEN plasmids was transfected into mouse neuroblastoma 2a (Neuro 2a) cells using Lipofectamine 2000, as described in Chapter 2, section 2.7.2. The cells were harvested 24 h post-transfection for gDNA extraction, as described in Chapter 2, section 2.3.2.2. The H2A.Lap1 targeted sequence for each plasmid is shown in Figure 3-4 (A). The ability of TALEN to cleave all H2A.Lap1 genes was tested by Sangamo, using two methods, Dual-Luciferase Single Strand Annealing Assay (DLSSA) and CelI cleavage assay in mammalian neuroblastoma 2a (Neuro 2a) cells.

A

```
                                                    101408 CGCCGCCACCGTCGCT
                                                  101406 CAGGTCGCCGCCACCGTCG
                                                                      101421 CCCGCACCTCCAGAG
Lap1a     ATGCCAAGGAACAGGGAAAACTGTCTTCGAGAGTCTTCAGGTCGCCGCCACCGTCGCTCCCGCACCTCCAGAGCT
Lap1b     ---------------------------------------------------------------------------
Lap1c     ------------C--------------A-AG-----------A------A-A--A-----------------G-
Pseudo    ----------G--CA------A-------T-G-G--------T--A------T-T------:------------C

101409     TCGATTAGAAACGACAC
Lap1a     GAGCTAATCTTTGCTGTGAGCCTGGTGGAACAGCATCTGAGGGAGGTTAGCCGTGCCCGGAGGCTCAGTGATACG
Lap1b     -----G--------A-----------------------------A-----------T--------------T-
Lap1c     -----G--------A-------------------------------------------------------T-
Pseudo    -----G------A-A-----------------AG---------CC------A----T--------------T-
                      CACTCGGACCACCTTG 101422
                         TCGGACCACCTTGTCG 101423

Lap1a     GTGCCCATCTTCCTGGCAGCCATCCTGGAGTCCCTCACCCGCAGGTTGCTGGAGCTTGCCGGCAATGAGGCCCAA
Lap1b     ------------------------------------------------------------G--------------
Lap1c     ------G---C-----T-------------T-------A------------------G--------------
Pseudo    -----A—T----------------------T-------TA----C------------G---------A----G

Lap1a     CGCAGAGGTACCGAGAGGCGCATCAACTCCTGAACTGCTGGACTTGGCTGTCTACAGCAATATGGAGCTAAGTGA
Lap1b     ---------------------------------------------------------------------------
Lap1c     ------------C------T-----------C---C----------AG----------G---------------
Pseudo    -AG---------C---T-A---------------------------AG-----T---------C----------
```

B

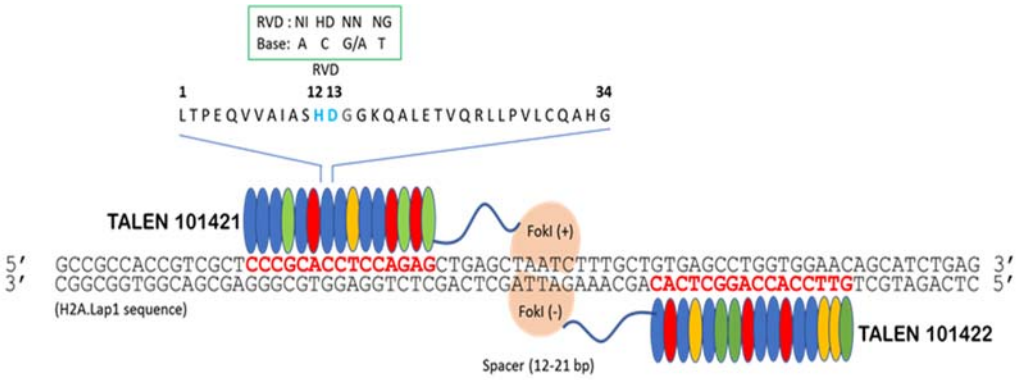


Figure 3- 4: TALENs design are specific to target H2A.Lap1 genes. A) The alignment of TALEN's targeting sequence to H2A.Lap1 genes**.** Blue coded is top strand, red coded is bottom strand. H2A.Lap1 gene targeted sequences that are bound by TALEN RVDs were designed by Zhang et al (2011) as listed in Table 3-2. Not all RVDs sequences were shown on H2A.Lap1 sequence. Only RVDs sequence that produced the highest activity in DLSS assay were chosen for transfection into Neuro 2a cells. **B)** Each TALEN consists of a DNA binding domain (obtained from TAL effector protein) fused to Fok1 endonuclease. The nucleotide specificity lies in the 12^{th} and 13^{th} position (of total 30-34 amino acids) of the RVD in TAL effector repeats. This RVD recognition is one nucleotide base.

Table 3-2: List of TALEN plasmids designed by Zhang et. al.

Gene specificity	Design #	Name	Binding site
Specific to H2A.Lap1a	101406	mH2A.Lap1--Tf1054a1	ttCAGGTCGCCGCCACCGTcg
	101407	mH2A.Lap1--Tf1058a1	gtCGCCGCCACCGTCGCTcc
	101408	mH2A.Lap1--Tf1059a1	gtCGCCGCCACCGTCGCTCcc
	101409	mH2A.Lap1--Tr1077a1	ctCACAGCAAAGATTAGCTca
	101410	mH2A.Lap1--Tr1078a1	ctCACAGCAAAGATTAGCtc
	101412	mH2a.Lap1--Tr1081a1	ctCACAGCAAAGATTag
	101413	mH2A.Lap1--Tf1090a1	ctGAGCTAATCTTTGCTgt
	101415	mH2A.Lap1--Tf1092b1	ctAATCTTTGCTGTga
	101416	mH2A.Lap--Tf1094a1	ctAATCTTTGCTGTGAgc
	101417	mH2A.Lap1--Tr1107a1	ctAACCTCCCTCAGATGCTgt
	101418	mH2A.Lap1--Tr1110a1	ctAACCTCCCTCAGATgc
	101419	mH2A.Lap1--Tr1119a1	ctCCGGGCACGGCTAACCTcc
Conserved among H2A.Lap1a,b and c	101421	mH2ALap1--Tf1073a1	ctCCCGCACCTCCAGAGct
	101422	mH2ALap1--Tr1091a1	ctGTTCCACCAGGCTCACag
	101423	mH2A.Lap1--Tr1094a1	ctGCTGTTCCACCAGGCTca
	101424	mH2ALap1--Tf1113a1	ctGGTGGAACAGCATCTga
	101426	mH2ALap1--Tr1134a1	gtATCACTGAGCCTCCgg
	101427	mH2ALap1--Tf1207a1	gtTGCTGGAGCTTgc
	101429	mH2ALap1--Tr1224a1	gtACCTCTGCGTTcc

DLSSA is a novel system used to optimize and quantify ZFN or TALEN activities in transiently transfected cells. DLSSA is based on the Dual-Luciferase Reporter Assay System from Promega. Based on Gregory et al. (2014), the system allows for sequential measurement of two individual reporter enzymes, firefly (*Photinus pyralis*) and sea pansy (*Renilla reniformis*) luciferases, within a single tube. Both the firefly and the *Renilla* luciferase reporters have been re-engineered so that the firefly luciferase reporter construct contains two incomplete copies of the firefly coding regions, separated by DNA binding sites for either ZFNs or TALENs (Gregory et al. 2014). The 5' copy is derived from approximately two thirds of the N-terminal part of the firefly gene, and the 3' copy is derived from approximately two thirds of the C-terminal part of the firefly gene. The two incomplete copies contain 600-bp homology arms. The separated firefly fragments have no luciferase activity. ZFN and TALEN create a DNA double strand break which will trigger recombination that the flanking repeat, using the single strand annealing pathway, and subsequently restore the firefly luciferase function. The co-transfected promoterless *Renilla* luciferase plasmid provides an internal control. The luminescent activity of each reporter is read on a luminometer. Normalizing the activity of the experimental reporter (firefly) to the activity of the internal control (*Renilla*) minimizes experimental variability caused by differences in cell viability and/or transfection efficiency. The normalized value is used to determine the activity of a given ZFN or TALEN pair (Gregory et al. 2014).

3.1.1 Pair 101421:101422 showed the highest activity in all four H2A.Lap1 genes

The plasmid pair 101421:101422 was transfected into Neuro 2a cells and the Fok1 activity assessed using DLSSA assay, as shown in Figure 3-5. This assay was performed by Ranier Amora (2011) of Sangamo Biosciences Inc.

The DLSSA were conducted on pairs that were intended to target the major form of H2A.Lap1, that is H2A.Lap1a gene, and also pairs that target all four H2A.Lap1 encoded genes. Remarkably, all TALEN pairs that were designed to target H2A.Lap1a gene only, had a very high activity for H2A.Lap1a gene, with 10 out of 14 pairs displaying preferential specificity to that gene. The pair that showed the highest specificity to the H2A.Lap1a gene only was 101418:101412. Meanwhile, the pair that showed the highest activity in all four genes was 101421:101422.

Based on this observation, the CelI assay was used next to determine the TALEN-induced gene modification activity in transiently transfected mouse Neuro 2a cells.

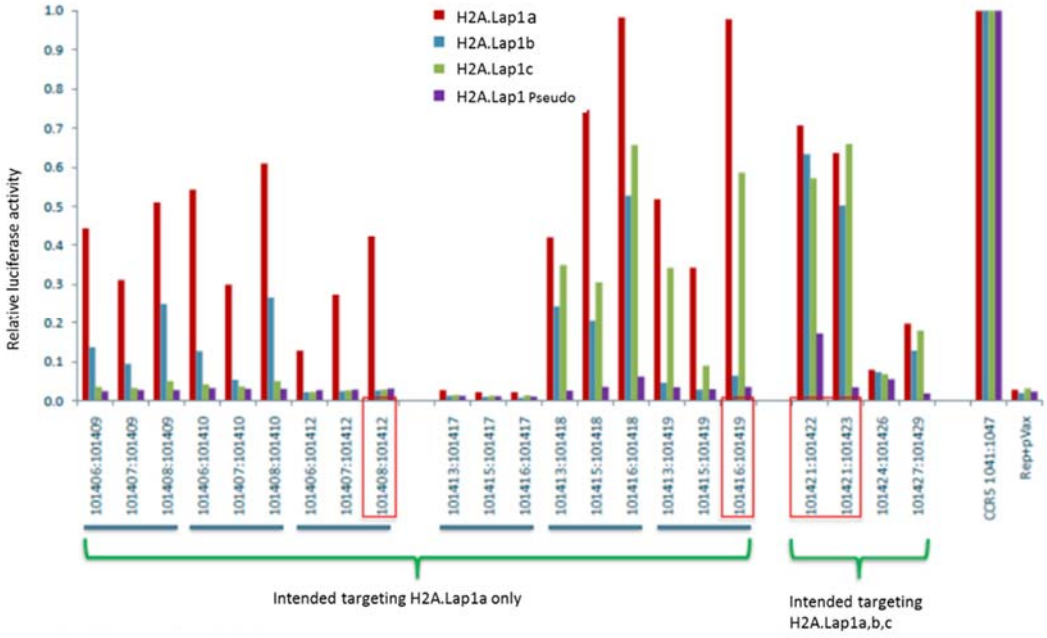


Figure 3-5: TALEN plasmid pair 101421:101422 showed a high enzyme activity on all H2A.Lap1 genes. The graph showed DLSS assay on 22 pairs of TALEN plasmids. The histogram showed that TALEN pairs 101421:101422 and 101421:101423 have the highest activity of all four H2A.Lap1 genes compared to other TALEN plasmid combinations, while 101418:101412 pair was the most specific against the main H2A.Lap1 gene, H2A.Lap1a.

3.1.2 Surveyor (CelI) assay showed pair 101421:101422 to have the highest activity

The principle behind the CelI assay is that, once a double stranded break (DSB) is introduced by Fok1, the affected DNA will be naturally repaired by the error-prone non-homologous end joining (NHEJ) or by homologous recombination. The NHEJ process will create mismatch in some of the base pairs. When these strands are amplified via PCR, the mutated and wild type (wt) loci copy number is increased. The heterogenous PCR products are then subjected to a denaturation step, followed by a slow gradual re-annealing process. Because of the slow re-annealing process, some mutated strands will bind to wt complementary strands creating DNA heteroduplexes (Figure 3-6), producing a "bubble" in the mispaired region.

CelI surveyor nuclease has a high specificity for mismatched DNA heteroduplexes, and when added to the reaction, it only cleaves heteroduplexes, producing shorter DNA fragments that can be visualised on agarose or polyacrylamide gel. The results of the CelI digestion are shown in Figure 3-7.

Only TALEN pairs designed to target all four H2A.Lap1 genes were tested by CelI assay. The gel in Figure 3-7 showed that not all pairs of plasmids were able to produce an efficient digestion. Based on the percentage shown at the bottom of the gel, plasmid pair 101421:101422 showed the highest percentage of digestion in all three genes. The quantification of the bands is represented in the graph below (Figure 3-8).

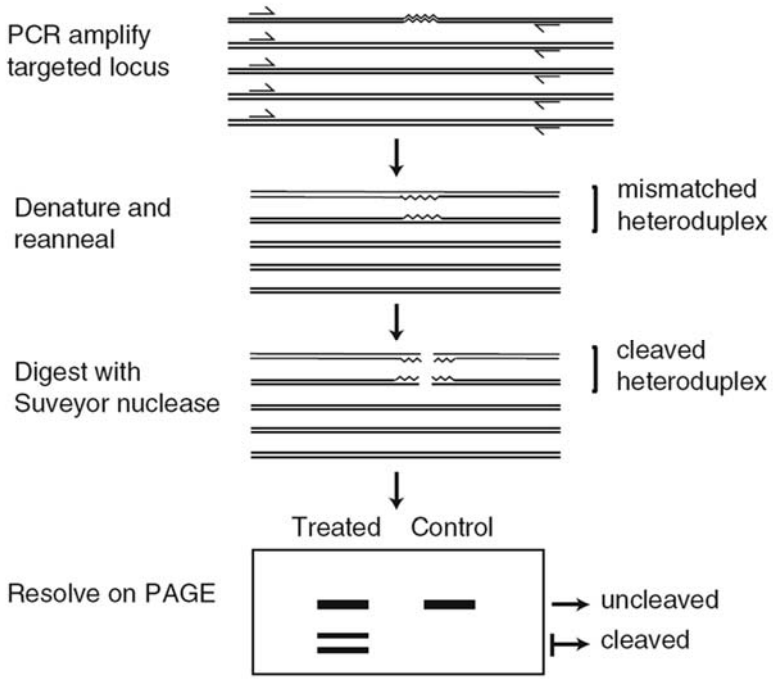


Figure 3-6: Schematic diagram of CelI assay principle. This assay involved heteroduplex formation by denaturing the amplicons at 95°C for 10 mins, followed by slow and gradual annealing, by decreasing the temperature to 25 °C. During the slow temperature decrement, some wt strands bind to mutated strands forming a "bubble" that results from mismatched DNA. The CelI enzyme recognizes this mismatch and digests it. Hence, different sizes of DNA are visibly distinguishable via polyacrylamide gel electrophoresis.

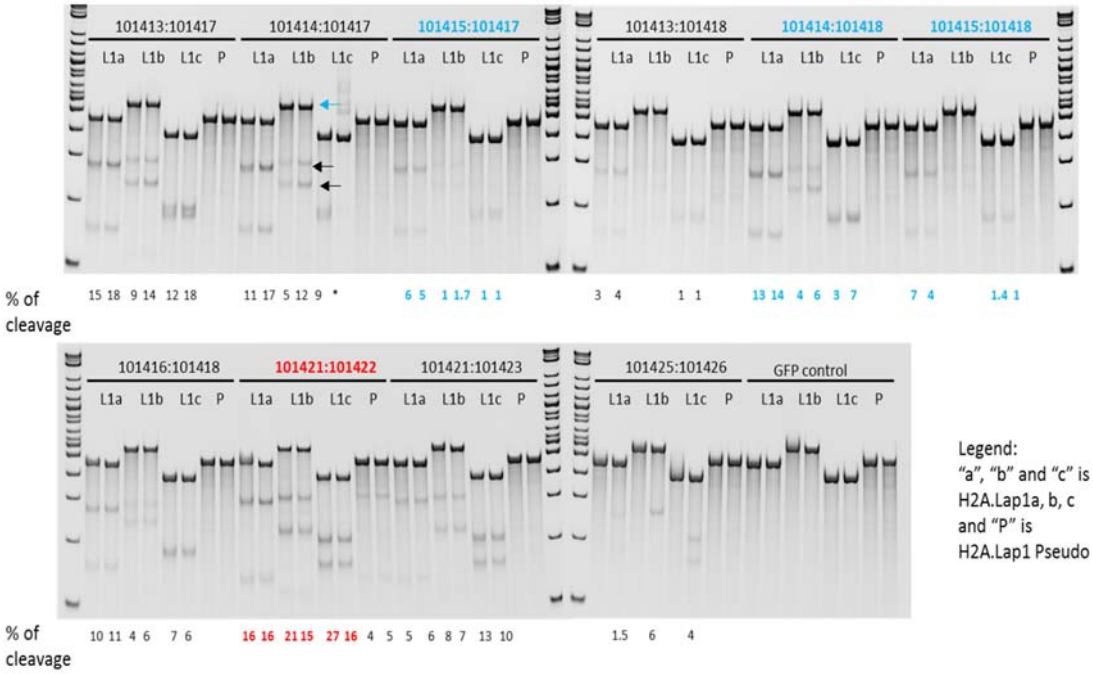


Figure 3-7: Three TALEN plasmid pairs showed a visible CelI digestion product on all H2ALap genes. The gel image showed CelI assay on all H2A.Lap1 amplicons amplified from the gDNA of Neuro 2a cells transfected with TALEN plasmids. The gel showed the digested amplicons of four H2A.Lap1 genes from CelI assay using Neuro 2a cells. The wt uncleaved H2A.Lap1 band is indicated by the blue arrow and gave the most intense band on the gel. The cleaved amplicons (heteroduplex strand) are indicated by the black arrow. The plasmid pairs highlighted in red and blue were sent to us for confirmation of this observation.

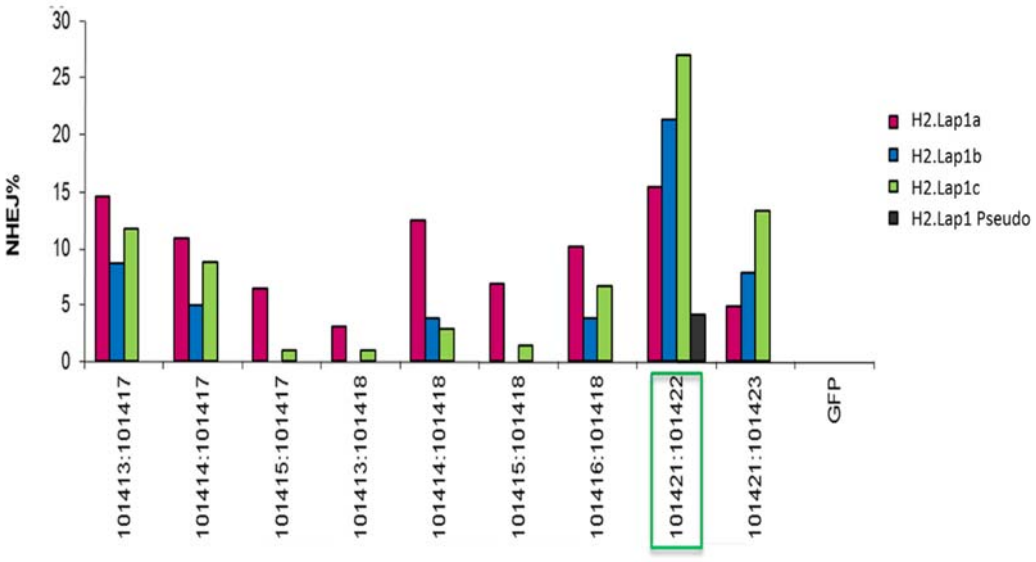


Figure 3-8: TALEN plasmid pair 101421:101422 has the highest CelI enzyme cleavage in H2A.Lap1 genes. This graph represents quantification of non-homologous end joining, expressed as percentage produced by Cel1 assay depicted in Figure 3-7. The pair in green box are the plasmid pair which have the highest cleavage percentage present in all four genes.

3.2 Validation of TALEN activity assay in cultured Neuro 2a cells

3.2.1 Optimisation of Neuro 2a transient transfection using CMV-driven GFP plasmid

To replicate observations by Sangamo, we cultured Neuro 2a cells in our laboratory and transfected them with two plasmid pairs, 101421:101422 and 101414:101418. The transfection efficiency test was done prior to transfecting the cells with TALENs plasmids. To examine the efficiency of transfection conditions, we tested Neuro 2a cells using pEGFP-N1 because it contains the same CMV promoter as the TALEN plasmids. Briefly, in each well of a 24-well dish, the cells were transfected with 0.5 µg of plasmid DNA using 2.5 µl of Lipofectamine 2000 in a serum-antibiotic free medium for 24h. Figure 3-9 shows the three biological replicates of the transfected cells. Table 3-3 indicates the intensity of GFP and the percentage of cells transfected. The average number of cells that were successfully transfected by pEGFP-N1 is 67.9 % which was comparable to Neuro 2a transfection efficiency by Sangamo.

The next step was to confirm the results obtained by colleagues from Sangamo and to choose the best TALEN pair for injection into mouse embryos, to create a H2A.Lap1 knock out (H2A.Lap1KO) mouse model. We cultured our Neuro 2a cells and transfected them with two TALEN plasmid pairs (101421:101422 and 101414:101418) to obtain transient TALEN expression. Twenty-four hours post-transfection, we harvested the cells and subjected them to genomic DNA (gDNA) extraction using Quick Extract Solution 2.0, as described in Chapter 2, section 2.3.2.2. The gDNA was then subjected to PCR amplification, as described in Chapter 2, section 2.3.3.1

The list of primers used for amplification is shown in Table 3-4. To optimise the mutation detection method, we re-designed the primers for H2A.Lap1b gene to obtain a shorter amplicon (398 bp), for a better gel resolution. The amplicons were subjected to CelI assay as described in Chapter 2, section 2.7.3. The products of CelI enzyme digest were loaded onto 10% TBE gel and post stained with 0.5 µg/ml of EtBr. Figure 3-10 shows a gel picture of all 4 H2A.Lap1 genes (in duplicate) isolated from Neuro 2a cells, transiently transfected with TALEN plasmid pair 101421:101422 and 101414:101418. CelI assay shows that cells transfected with plasmid pair 101421:101422 have cleaved bands present in all four genes and that they digest more efficiently than plasmid pair 101414:101418. This observation confirmed the findings

of Sangamo and gave us confidence to use this particular TALEN plasmid pair for microinjection into mouse embryos.

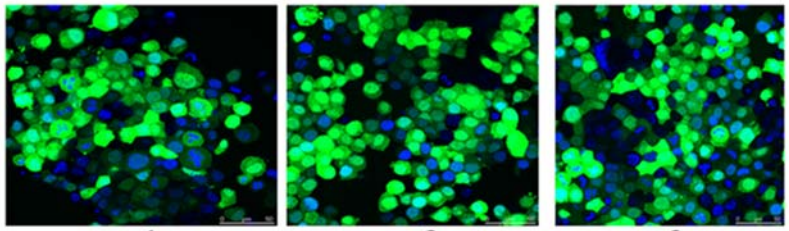


Figure 3-9: Optimised parameters with high transfection efficiency was used in the transfection of TALEN pair 101421:101422 into Neuro 2a cells. Images of Neuro 2a cells transfected with pEGFP-N1. The green florescent cells indicate cells that were transfected with the vector. Blue florescence represents DAPI stained nuclei.

Table 3-3: The percentage of transfected Neuro 2a cells by pEGFP-N1.

Replicate	Total cells transfected (GFP/DAPI) (%)	Percentage of transfection (Green channel only)			Total cells (DAPI only)
		Weak	Medium	High	
1	96 (70%)	44	30	22	137
2	137 (75.3 %)	43	68	26	182
3	116 (58.6 %)	38	41	37	198

Table 3-4: List of standard PCR primers for H2A.Lap1 encoding genes.

H2A.Lap1 gene family	Sequence (5'-3')	Size of amplicon
H2A.Lap1a (by Sangamo)	Fwd- CAGCAGAAAGCAGCCAAGTGG Rev- GCAGGTCAGCCAAGAAGCA	440 bp
H2A.Lap1b (by Sangamo)	Fwd- GTACGGTACAAAGGGAGATG Rev- GAGCAGGTCAGCCAAGCAGAG	536 bp
H2A.Lap1b	Fwd- GTTGGGCATTGGACTTGGAC Rev- CAGCCAAGTCCAGCAGTTC	398 bp
H2A.Lap1c (by Sangamo)	Fwd- CAGGTCAGCAGAGAGCAATT Rev- CTCCATACTGCTGTAGACCT	374 bp
H2A.Lap1Pseudo (by Sangamo)	Fwd- GTCAGCAGAATGCAGCCAAATAT Rev- CAAGCCAGTAGCCAACATCAAG	459 bp

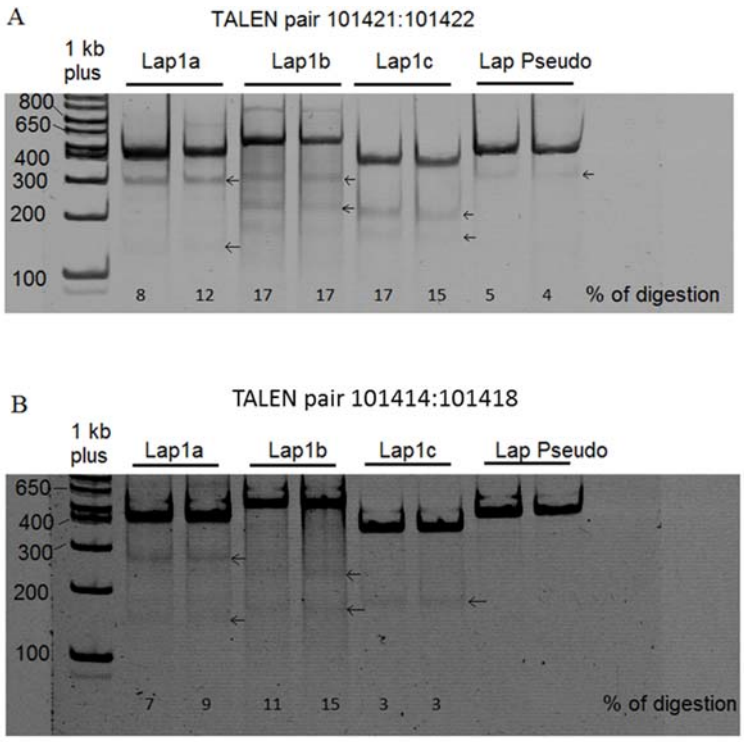


Figure 3-10: TALEN 101421:101422 works efficiently in our mouse Neuro 2a cell culture. CelI assay on all H2A.Lap1 genes amplified from the gDNA of Neuro 2a cells transfected with TALEN plasmids. A) CelI assay of H2A.Lap1 amplicons (in duplicate for technical control) from Neuro 2a cells transfected with plasmid pair 101421:101422 and B) CelI assay of H2A.Lap1 amplicons from Neuro 2a cells transfected with plasmid pair 101414:101418.

3.2.2 T7 endonuclease I (T7EI) digests more efficiently than CelI surveyor nuclease in Neuro 2a cells

Just like CelI, T7EI recognizes and cleaves imperfectly matched DNA. It also cleaves cruciform DNA structures, Holliday structures or junctions, heteroduplex DNA and more slowly, nicked double-stranded DNA. The cleavage site is at the first, second or third phosphodiester bond that is 5′ to the mismatch. In addition to being a cheaper alternative than the CelI surveyor nuclease, T7E1 has been shown to outperform CelI surveyor nuclease, in terms of sensitivity, with deletion in heteroduplex strands (Vouillot et al. 2015).

Initially, we employed CelI assay to confirm TALEN activity in our Neuro 2a cells. However, we found that incubation with T7EI enzyme gave a better digestion efficiency (Figure 3-11).

We tested the same Neuro 2a cells transfected with TALEN plasmid pair 101421:101422, and amplified H2A.Lap1a, via the same PCR conditions. The cleaved band was quantified using ImageJ (Figure 3-11). We set up various digestion conditions, that include various enzyme concentrations (1 U-10 U) and duration of incubation (20 mins to 1 h). The results showed higher digestion efficiencies and more abundant cleaved bands in T7EI enzyme digestions in Neuro 2a cells. However, as we further employed T7EI enzyme to detect mismatch in our mice colony, we found that gDNA extracted from mouse ear punch showed better digestion efficiency with CelI enzyme. Therefore, we used the CelI enzyme for assaying our mice colony.

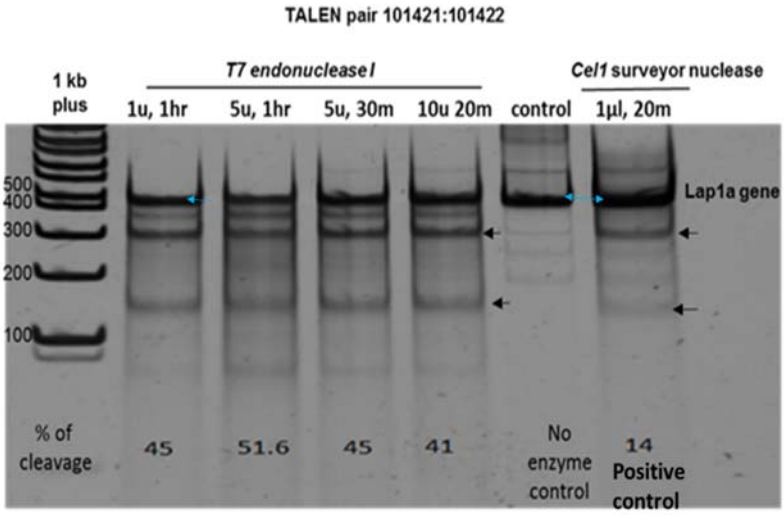


Figure 3-11: T7 endonuclease I digests more efficiently than CelI. The 10% TBE gel showed digested H2A.Lap1a amplicons from Neuro 2a cells (transfected with 101421:101422) under various conditions. As a positive control, we used the digestion conditions suggested by Zhang et. al. Black arrows show digested heteroduplex strands, whilst blue arrows indicate homoduplex strands. Numbers in each lane represent the percentage of cleavage/digestion.

3.2.3 H2A.Lap1 pseudogene Gm14904 is likely an assembly error

As mentioned earlier, the design of H2A.Lap1-specific TALENs was done in 2010, when Ensembl assembly (Ensembl Release 57, March2010) listed four H2A.Lap1 genes in mice genomes. The TALEN pair 101421:101422 successfully targeted three H2A.Lap1 genes but didn't produce detectable alterations in the H2A.Lap1 pseudogene, Gm14904. Later mouse genome assemblies removed this gene altogether. Although Gm14904 is a pseudogene and, in theory, shouldn't be expressed, many novel genes are often listed as pseudogenes until their expression is shown experimentally. Our qPCR analysis showed that three H2A.Lap1 genes, H2afb3 (H2A.Lap1a), Gm14920 (H2a.Lap1b) and H2afb2 (H2A.Lap1c) were expressed in the testis of wt mice, while the expression of a pseudogene Gm14904 was not detected (Figure 3-12). Hence, we concluded from our experimental analysis and Ensembl's latest Assemblies that only three H2A.Lap1 genes are expressed, and TALEN 101421:101422 pair's inactivity on the pseudogene, should not affect the knockout of H2A.Lap1.

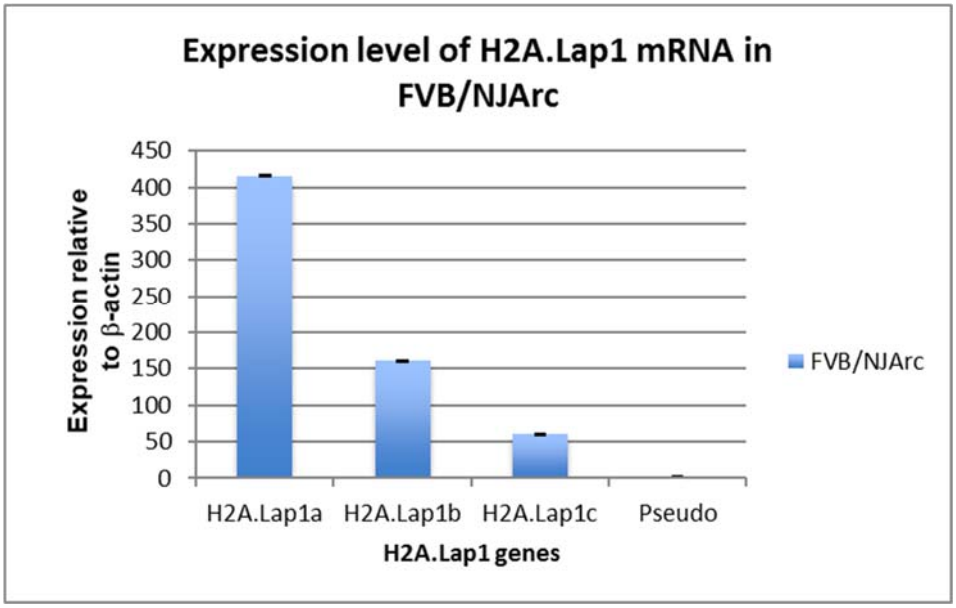


Figure 3-12: The expression level of H2A.Lap1 mRNA in FVB/NJArc mouse. qPCR results using gene-specific primers. Expression relative to mouse β-actin.

3.2.4 Screening potential founder mice for single nucleotide polymorphisms within three H2A.Lap1 genes

One of the last steps we performed before attempting to produce a H2A.Lap1 KO in FVB/NJArc mice, was to investigate the presence of single nucleotide polymorphisms (SNPs) within all H2A.Lap1 genes, in this mouse strain. Although most laboratories use inbred strains, one cannot rule out that mice are not completely inbred or, due to handling errors, may have been bred with a different strain, that is phenotypically similar, and thus may contain an appreciable number of SNPs. Since SNPs in the vicinity of H2A.Lap1 genes will create a signal in the surveyor nuclease assay, their presence would produce a false-positive result. Therefore, prior to microinjection of the TALEN plasmid, we performed the CelI assay on potential founder mice (Figure 3-13).

FVB/NJArc are commonly used for production of transgenic mice because their fertilized eggs contain large and prominent pronuclei which facilitate the microinjection of DNA. This strain also produces a large number of pups (12-15 per litter), which is crucial for creating enough P_o founders. (JAX® Mice Database). As shown in Figure 3-13, we have amplified three H2A.Lap1 genes from gDNA extracted from WT male and female FVB/NJArc mice ear notches. The amplicons were subjected to CelI assay and no detectable cleaved bands were identified. Hence, no SNPs were detected in the amplified regions, in this particular mouse strain. We therefore concluded that the FVB/NJArc mice from this particular facility can be safely used for production of H2A.Lap1 KO mice.

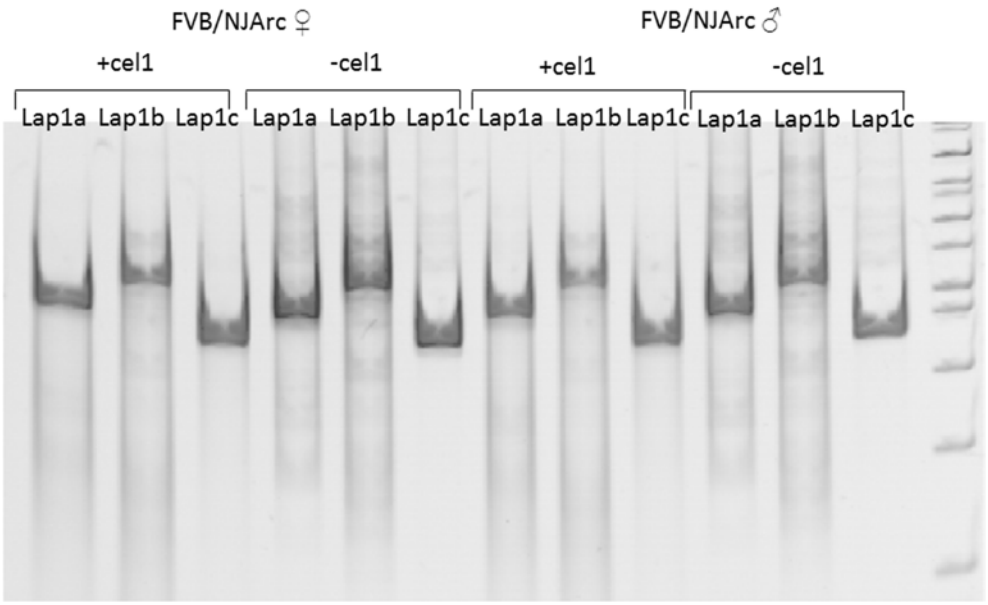


Figure 3-13: CelI assay on H2A.Lap1 genes from FVB/NJArc strain of female and male mouse gDNA. Symbol: ♂, male; ♀, female; +, added; -, without.

3.3 Production of H2A.Lap1KO founder pups by the delivery of TALENs to one-cell embryos

TALENs can be delivered into single cell embryos either by pronuclear injection (PNI) of plasmid DNA or mRNA, or via intracytoplasmic injection (mRNA only). The experiments involving the delivery of ZFNs into rat embryos showed that injection of mRNA produces higher rates of mutants compared to plasmid DNA, with intracytoplasmic injection (ICI) of mRNA producing the highest rate of mutant founders (up to 75%) (Geurts et al. 2009). Thus, we chose to deliver the mRNA by ICI or PNI for the production of H2A.Lap1 KO founders.

In vitro production of stable and ready-to translate TALEN mRNA involves several steps. First, the capped RNA was produced by *in vitro* transcription using T7 RNA polymerase. The incorporation of specially modified cap, anti-reverse cap analogue (ARCA), results in the synthesis of capped RNA that is more efficiently translated and is more stable *in vivo* than the standard cap. Immediately following *ín vitro* transcription, polyadenosine (polyA) tail was added to protect the mRNA molecule from enzymatic degradation in the cytoplasm and to enhance the efficiency of translation initiation *in vivo*. The summary of TALEN mice production is presented in Figure 3-14.

Capped, polyadenylated and purified *in vitro* transcribed mRNA was adjusted to 10 ng/µl for TALEN 101421 and 101422 and injected into fertilised one-cell embryos. We chose to use high concentrations of mRNA (10µg/µl), as it produced the highest percentage of mutant founder pups, although it resulted in the lowest birth rate (Geurts et al. 2009)**.** The 1-2 µl injections were performed as either PNI or ICI, as indicated in Table 3-5. The single cell embryo was allowed to culture overnight to the 2-cell stage in a microdrop of embryo culture at 37 °C, in 5% CO_2. All microinjection procedures were performed by Dr Jo Bowles and Ms Tara-Lynn Davidson (P. Koopman Laboratory, University of Queensland).

The surviving two-cell embryos, displaying normal morphology, were implanted into pseudo pregnant female mice of the same strain. In cages where there was only one transfer mouse, another pseudo pregnant mouse was placed to act as a "nanny" (as indicated in Table 3-5). The "nanny" mouse helped to lessen stress on the transfer mother by acting as a partner in caring for the litter (as it would have had the same pregnancy schedule as the mother, eg: lactate/feed, natural instinct to care for the pups) and it also acted as a cage mate. After the pups were born and weaned, they were exported to us here at JCSMR, to be screened for mutation.

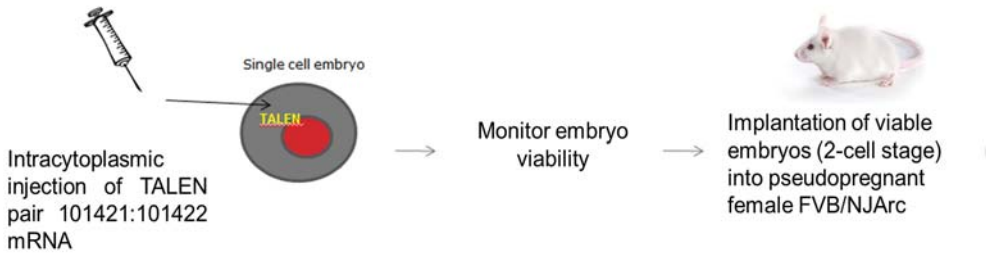


Figure 3-14: The summary of TALEN mice production

Table 3-5: List of microinjections performed by Koopman's lab to produce viable founder.

Injection Date	Implant #	Construct Name	Background Strain	F Plug Score	Harvest Data			Cell Division				# Implanted	Due Date	DOB	Total Pups/Alive
					#of F harvested	# Oocytes Collected	# Oocytes Injected	2 cell	1 cell	mutated	lysed				
7/11/2012	L68	TALENS (10 ng/µl) ICI	FVB	10 of 15	15	106	101	44	0	0	57	22	Nov 26/12	not pregnant	0
												22	Nov 26/12	not preg	0
14/11/2012	L69	TALENS (10 ng/µl) ICI	FVB	14 of 15	15	107	99	23	0	0	76	23	Dec 3/12	not pregnant	0
												n/a	nanny	n/a	n/a
28/11/2012	L74	TALENS (10 ng/µl) in ICI	FVB	10 of 15	15	123	123	22	5	0	96	22	Dec 17/12	Dec 18/12	5
												n/a	nanny	n/a	
6/12/2012	L79	TALENS (10 ng/µl) i PNI	F2	3 of 10	10	60	41	26	2	0	13	26	Dec 25/12	Dec 31/12	1
												n/a	nanny	n/a	
7/02/2013	L89	TALENS (10 ng/µl) ICI	FVB	6 of 15	15	159	58	10	5	2	41	10	Feb 26/13	Feb 27/13	3
												n/a	nanny	n/a	
11/02/2013	L90	TALENS (10 ng/µl) ICI	FVB	7 of 15	15	114	110	28	1	1	80	28	Mar 2/13	Mar 3/13	10
												n/a	nanny	n/a	
TOTAL							**532**	**153**				**153**			**19**

*Symbols: ICI, intracytoplasmic injection; PNI, pronucleus injection.

As indicated in Table 3-5, a total of 532 oocytes were successfully injected and incubated for the next 24 h. Of those, 153 oocytes survived to 2-cell stage and were implanted into five pseudopregnant females FVB/NJArc. We also implanted the embryos into one F2 strain mice (PNI), which is a hybrid of B6D2/F1 strains, to see whether this strain can produce surviving litter. Table 3-5 also shows that the highest number of litters was produced by FVBN/JARc mice, when injections of mRNA into the cytoplasm (L90, 10 pups and L74, 5 pups) were used, suggesting that delivery of RNA via cytoplasm yields better results. From these six female mice, 19 healthy pups were born and when they were two weeks post weaning, they were transported to the JCSMR, to be genotyped. The efficiency of TALEN-mediated microinjections into mouse embryo is presented in Table 3-6.

Table 3-6: The efficiency of TALEN-mediated microinjections into mouse oocytes.

TALEN mRNA concentration	Mouse strain	Total oocytes injected	Total oocytes surviving to 2-cell stage/ total oocytes (%)	Total oocytes implanted	Pups born/implanted oocytes (%)	Number of KO mice (%) (mutation rate)
TALEN mRNA (10 ng)	FVB (ICI)	491	127 (25.8 %)	127	18/127 (14.1 %)	8/18 (44 %)
TALEN mRNA (10 ng)	F2 (PNI)	41	26 (63 %)	26	1/26 (3.8 %)	1/1 (100 %)
TOTAL		**532**	**153 (28.7%)**	**153**	**19/153 (12.4%)**	**9/19 (47.3 %)**

3.3.1 A modification to CelI surveyor nuclease assay to detect homozygous mutant progeny in mice.

In screening for mutations with Surveyor and T7 endonucleases, one relies on the heterogeneity of the sample pool, where WT sequences present alongside of mutated sequences. Thus, detection of mutations in Neuro 2a cells, where transient transfection efficiency is never 100%, is a perfect system for CelI and T7EI. However, when one attempts to produce a gene knock-out in a mouse where mutations are introduced into single cell embryos by injection with TALEN RNA, the outcome is anticipated to be i) wild-type (none of the alleles underwent NHEJ), ii) homozygous mutation (both alleles underwent NHEJ) or iii) heterozygous (only one allele was mutated via NHEJ). Surveyor and T7 endonuclease will only cleave mismatched DNA in the case of (iii), while (i) and (ii) will not produce short digestion products in the absence of heterogeneity. Moreover, all H2A.Lap1 genes are located on the X chromosome, of which male mice have only one copy. Hence, for male mice, they can either be hemizygous or wild type, neither of which can be distinguished by Surveyour nuclease. We therefore created the heterogeneity in male mice by adding equimolar concentrations of amplicons from wild type (or the background strain) into the amplicons of mutant mice, in a single tube prior to the T7EI or CelI enzyme digestion assay, as shown in Figure 3-15.

Upon receiving the pups, we first genotyped them using CelI assay. In Figure 3-16, an example of the CelI assay is shown in three of the founders. In male founder L74-5, amplicons for each gene were mixed with WT amplicons to detect hemizygous mutations.

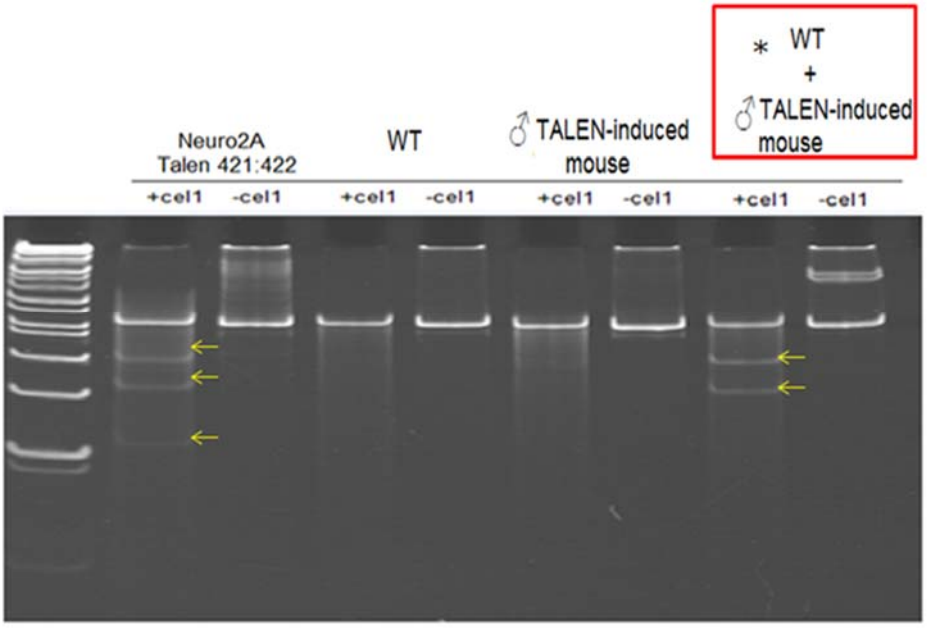


Figure 3-15: CelI assay on H2A.Lap1 amplicons from male founder TALEN-induced mouse requires improvisation, by adding wt amplicons to create heterogeneity. The gel image showed H2A.Lap1a amplicons resolved in 2% agarose gel. As a control, amplicons from Neuro 2a transfected cells were included. The red box (*) indicates the reaction that contained equimolar amplicons of wt and TALEN-induced mouse added in one tube for CelI assay, and produced digested bands showed by yellow arrows. Symbol: ♂, male; ♀, female; +, added; -, without; arrows, digested amplicons**.**

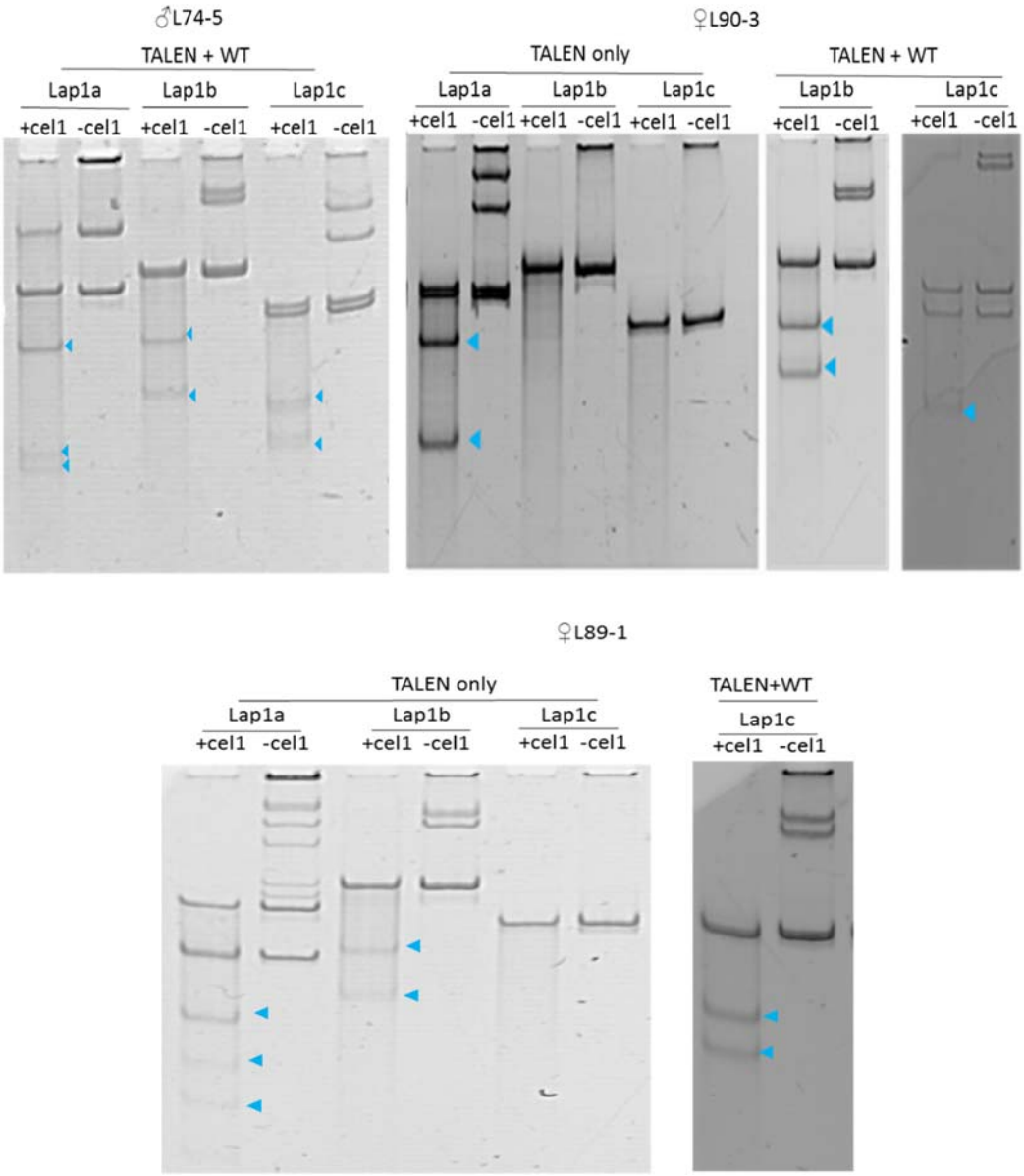


Figure 3-16: CelI assay on H2A.Lap1a, b and c genes showed a positive digestion of gDNA from male founder L74-5 and female founders L90-3 and L89-1. Blue arrows indicate the digested band from CelI cleavage activity. Symbol: ♂, male; ♀, female; +, added; -, without; arrows, digested amplicons.

The results showed that H2A.Lap1a, b and c genes were all mutated, which suggest the genotype of this male mouse is hemizygous mutant. Meanwhile, for female L90-3, we did not mix the sample with wt amplicons at first. This is because female mice have two copies of the gene, hence, heterogeneity is possible. However, the gel image for CelI assay on TALEN-only sample indicated that the H2A.Lap1a gene produced cleaved bands (suggesting the gene is a heterozygous mutant), while the other two genes did not. We then repeated the assay by mixing H2A.Lap1b and c gene with WT amplicons and the result showed that CelI did detect mismatches in these samples, suggesting H2A.Lap1b and c genes of female mouse L90-3 carry homozygous mutations on both alleles as well. The same approach was employed in determining the genotype of L89-1 female mouse. The gel image of CelI assay on L89-1 suggests that H2A.Lap1a and H2A.Lap1b are heterozygous mutants, whilst H2A.Lap1c is a homozygous mutant. As a result, CelI assay identified that 9 out of 19 pups contained TALEN-induced mutations.

3.3.2 Improvement in the screening method to characterise the mutation pattern method in TALEN-induced mice

Next, in order to charcterize the exact mutation pattern in mice where TALENs induced them, the PCR products, obtained by amplification of all H2A.Lap1 genes, were cloned and 8-10 individual clones for each gene were sequenced by the Sanger method.

Analysis of a single gene by cloning and sequencing individual clones is a laborious procedure. It becomes even more laborious and time consuming when a multiple-gene family needs to be screened in that manner. To characterise mutations introduced in all H2A.Lap1 genes, we had to amplify all three H2A.Lap1 encoding genes, ligate the amplification product, transform bacteria and subsequently purify and screen plasmids for inserts, before subjecting them to Sanger sequencing. For each gene, at least eight to ten clones had to be sequenced.

To simplify the screening method, particularly the cloning and colony screening, I introduced several modifications to the work flow. First, I took advantage of a TA-cloning technology, that allowed me to clone PCR products directly into a vector without restriction digests. Secondly, I set up a quick and robust colony PCR method using universal promoter T7 and SP6 primer, and introduced subsequent purification of the PCR products using magnetic beads Agencourt® AMpure® XP kit by Beckman Coulter. This approach provided a rapid screening of bacterial clones in 96 and 384-

well format. The Agencourt® AMPure® XP PCR Purification systems utilize Agencourt's solid-phase paramagnetic bead technology for high-throughput purification of PCR amplicons. Agencourt® AMPure® XP utilizes an optimized buffer to selectively bind PCR amplicons, 100bp and larger, to paramagnetic beads. Excess primers, nucleotides, salts, and enzymes can be removed using a simple washing procedure with 70 % EtOH. The resulting purified PCR product is essentially free of contaminants. Hence, magnetic separation based purification gave us rapid, high throughput screening and good quality purified DNA. Furthermore, unlike column-based purification, the magnetic separation method does not involve centrifugation and contains no salt carry-over that may interfere with the sequencing reaction. The workflow that we employed to genotype our mice is shown in Figure 3-17.

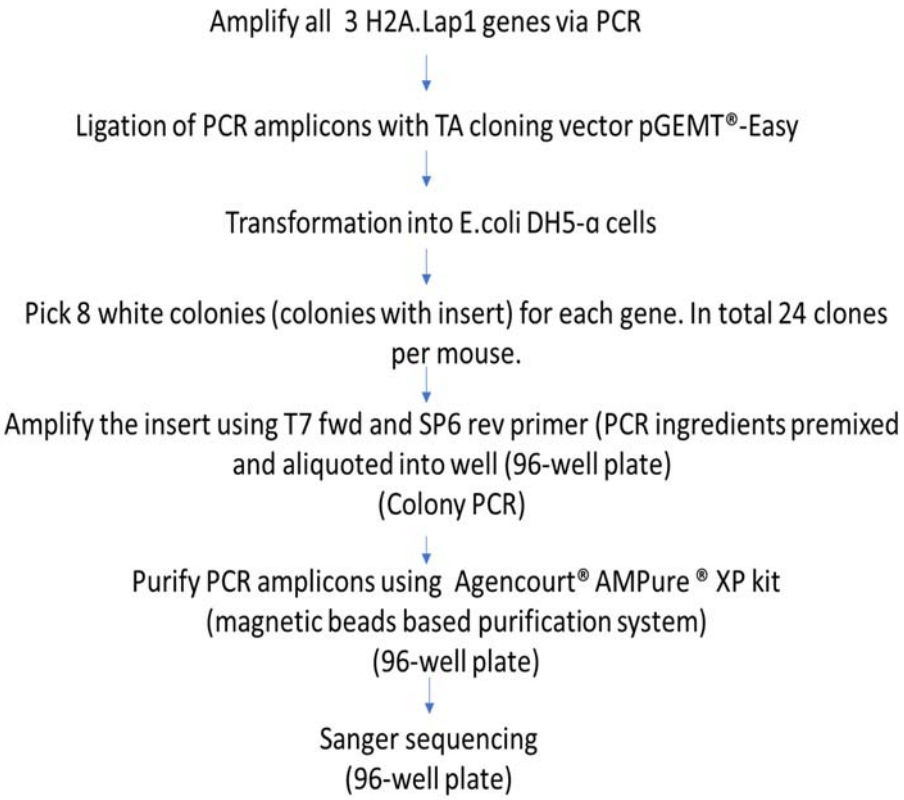


Figure 3-17: The workflow employed to characterise the genotype of TALEN mouse.

Table 3-7 shows the genotype and the exact mutations of all 19 founders. Our sequence analysis showed 57 % (4 of 7) of male founders and 41.6 % (5 of 12) of female founders were mutants. Most of the mutants carry nucleotide deletions in H2A.Lap1 genes, while L89-2 had a two nucleotides substitution in H2A.Lap1b gene and L74-5 contained a 286-bp insertion in the H2A.Lap1a gene, in addition to a 5-bp deletion. The longest deletion (160 bp) was found in H2A.Lap1a gene of L89-1 male mouse. Interestingly, in two founders, L90-4 and L90-8, at least one of the H2A.Lap1 genes were not amplifiable, using gene specific primers. This observation often occurred in H2A.Lap1b and H2A.Lap1c genes. However, when we mixed primer pairs (eg: Lap1**b** FWD and Lap1**c** REV) we were able to detect a template, suggesting that TALEN activity induced the formation of a chimera between H2A.Lap1b and H2A.Lap1c genes, in these founder mice. We further sequenced these templates, and indeed, the sequences were hybrids of H2A.Lap1b and H2A.Lap1c genes. These particular chimeras were observed in five out of nine mutant founders (L79-1, L89-2, L89-3, L90-6 and L90-8). The alignment of H2A.Lap1 genes from the founders to WT is shown in Figure 3-18 below.

Figure 3-18: Sequence alignment of H2A.Lap1 genes from mutant founders relative to H2A.Lap1 genes of wt FVB mice showed that successful mutations are created in coding region of H2A.Lap1 genes. In L79-1, H2A.Lap1c sequence is a hybrid of Lap1b and Lap1c genes. Symbols: -, Δ deletion; /, substitute by. Blue coded font, deleted nucleotide; red coded font, nucleotide substitution**.**

Table 3-7: Table showed the genotype and mutation pattern (by sequencing) of all 19 founder mice

#	Founders ID	Genotype by sequencing/ Protein sequence implication		
		H2A.Lap1a	**H2A.Lap1b**	**H2A.Lap1c**
1	♀L74-1	XX (wt)	XX (wt)	XX (wt)
2	♀L74-2	XX (wt)	XX (wt)	XX (wt)
3	♀L74-3	XX (wt)	XX (wt)	XX (wt)
4	♂L74-4	XY (wt)	XY (wt)	XY (wt)
5	♂L74-5*	$\mathbf{X^{\Delta5\ +286}Y}$ (Hemizygous) Stop codon after 33 aa	$\mathbf{X^{\Delta15}Y}$ (Hemizygous) Deletion of aa *ELIFA*	$\mathbf{X^{\Delta17}Y}$ (Hemizygous) Stop codon after 35 aa
6	♀L79-1	$\mathbf{X^{\Delta10}X^{\Delta10}}$ (Homozygous) Stop codon after 35 aa	$\mathbf{X^{\Delta12}X^{\Delta12}}$(Homozygous) Chimera of b/Δ/c Deletion of aa *ELIF*	$\mathbf{X^{\Delta15}X^{\Delta15}}$ (Homozygous) Chimera of c/Δ/b Deletion of aa *ELIFA*
7	♀L89-1*	$\mathbf{X^{\Delta22}X^{\Delta160}}$ (Homozygous) Stop codon after 19, 14 and 35 aa	$\mathbf{X^{\Delta5}X^{\Delta17}}$ (Homozygous) Chimera c/Δ/b Stop codon after 39 and 35 aa	$\mathbf{X^{\Delta5}X^{\Delta5}}$ (Homozygous) Stop codon after 39 aa
8	♂L89-2	$\mathbf{X^{\Delta65}Y}$ (Hemizygous) Stop codon after 18 aa	$\mathbf{X^{C/T,A/G}Y}$ (Hemizygous) Substitution of 1 aa	$\mathbf{X^{\Delta42}Y}$ (Chimera b/Δ/c) Deletion of 14 aa and absence of stop codon
9	♂L89-3	$\mathbf{X^{\Delta26}Y}$ (Hemizygous) Chimera c/Δ/a Stop codon after 32 aa	$\mathbf{X^{\Delta10}Y}$ (Hemizygous) Chimera b/Δ/c Stop codon after 27 aa	$\mathbf{X^{\Delta10}Y}$ (Hemizygous) Chimera b/Δ/c Stop codon after 27 aa
10	♀L90-1	XX (wt)	XX (wt)	XX (wt)
11	♀L90-2	XX (wt)	XX (wt)	XX (wt)
12	♀L90-3*	$\mathbf{X^{\Delta28}X^{\Delta5}}$ (Homozygous) Deletion of aa *ELIFA* Stop codon after 28 and 39 aa	$\mathbf{X^{\Delta10}X^{\Delta10}}$ (Homozygous) Stop codon after 27 aa	$\mathbf{X^{\Delta64}X^{\Delta64}}$ (Homozygous) Stop codon after 34 aa
13	♀L90-4	$\mathbf{X^{\Delta15}X^{\Delta10}}$ (Homozygous) Deletion of aa *ELIFA* Stop codon after 27 aa	No successful amplification. Locus had major genomic rearrangement	No successful amplification. Locus had major genomic rearrangement.
14	♀L90-5	XX (wt)	XX (wt)	XX (wt)
15	♀L90-6	$\mathbf{X^{\Delta17}X^{\Delta17}}$ (Homozygous) Chimera c/Δ/a Deletion of aa *ELIFA* Stop codon after 35 aa	$\mathbf{X^{\Delta5}X^{\Delta5}}$ (Homozygous) Chimera b/Δ/c Stop codon after 39 aa	$\mathbf{X^{\Delta5}X^{\Delta5}}$ (Homozygous) Chimera c/b/Δ/c Stop codon after 39 aa
16	♀L90-7	XX (wt)	XX (wt)	XX (wt)
17	♂L90-8	$\mathbf{X^{\Delta16}Y}$ (Hemizygous) Stop codon after 27 aa	$\mathbf{X^{\Delta17}Y}$ (Hemizygous) Chimera c/Δ/c/b	No successful amplification. Locus had major genomic rearrangement.
18	♂L90-9	XY (wt)	XY (wt)	XY (wt)
19	♂L90-10	XY (wt)	XY (wt)	XY (wt)

Symbols: ♂**,** male; ♀, female; Δ, deletion; +, insertion; /, substitute by; wt, wild type; aa, amino acids; *, mouse that was chosen as parent breeder to produce H2A.Lap1KO colony.

3.3.3 In-frame mutation in founder L74-5 may be as detrimental as a frameshift mutation

The male founder (L74-5) that we used for establishment of two strains of the H2A.Lap1 KO colony, carried an in-frame deletion mutation (Δ15) in the H2A.Lap1b gene (Figure 3-19). This particular 15 bp deletion results in the removal of five amino acids (^{26}*ELIFA*^{30}) after the 25th amino acid in the protein sequence (Figure 3-20). In order to predict whether the deletion of these five amino acids may be deleterious to H2A.Lap1 function, we performed some *in silico* structural analysis. There are no existing high-resolution crystal structures of H2A.Lap1-containing nucleosome thus, we used an existing nucleosome structure containing canonical histone H2A protein as a proxy for H2A.Lap1 (Figure 3-20).

The amino acids in canonical H2A, equivalent to H2A.Lap1 residues 26-30, (*ELIFA)* correspond to a short loop between two helices towards the N-terminus of H2A (Figure 3-20). This loop interacts with other structural elements of H2A as well as histone H2B. Thus, removal of this loop is likely to cause significant structural disruption, removing key protein-protein contacts as well as potentially disrupting histone-DNA contacts that occur in close proximity to this region. This could lead to destabilisation of the nucleosome and consequently create a dominant negative phenotype. However, this conclusion is drawn from computer-based analysis and was not confirmed experimentally. That is why, most of our analyses of the H2A.Lap1 KO phenotype were done on the progeny of L90-3 and L89-1 founders that contained premature stop codons in all three H2A.Lap1 genes.

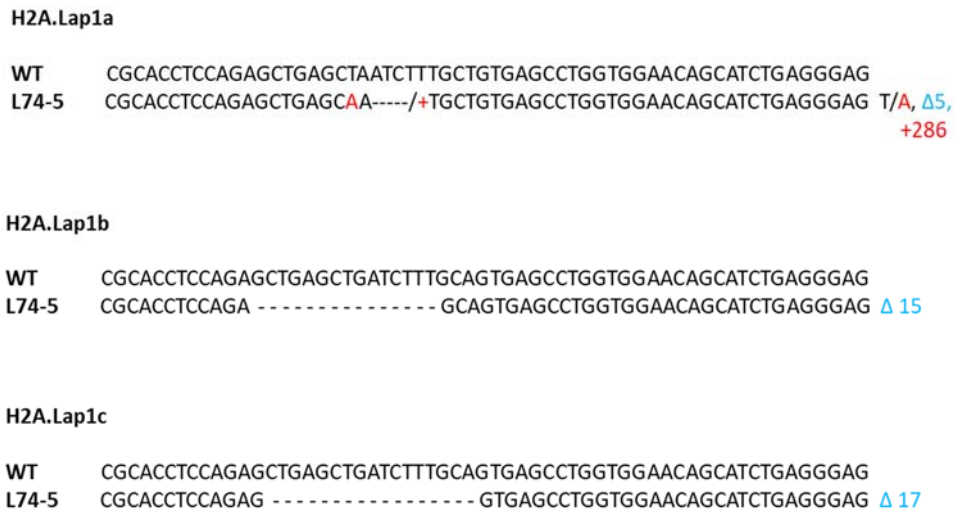


Figure 3-19: TALEN induced mutations in all H2A.Lap1 genes of male founder L74-5. The image showed the sequence alignment of H2A.Lap1 genes from mutant male L74-5, relative to WT FVB mouse. Symbols: -, Δ deletion; +, insertion. Blue coded font, deleted nucleotide; red coded font, nucleotide substitution.

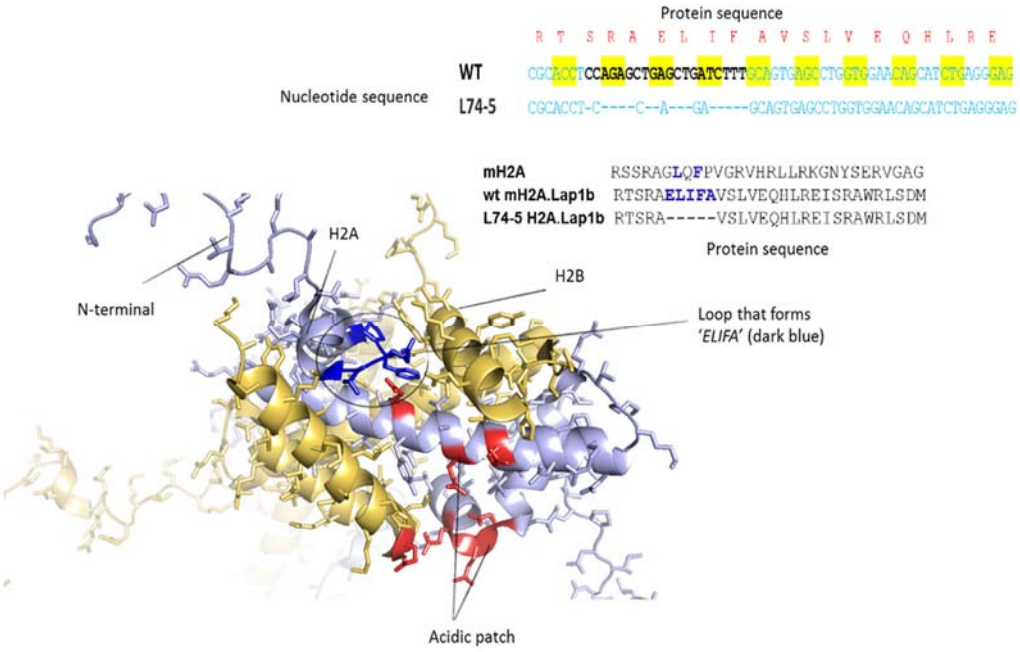


Figure 3-20: Structure of a canonical H2A-H2B dimer as observed in the context of a nucleosome (PDB ID: 1KX5). H2A is shown in purple and H2B in yellow. The region equivalent to the deleted amino acids in H2A.Lap1b of L74-5 (sequence alignment inset) is shown in dark blue. Amino acids in red form the 'acidic patch', a key interaction interface on the nucleosome surface.

3.3.4 Choice of breeders to establish a H2A.Lap1 knockout colony

To produce the next generation of mice, we chose founders i) that carried mutations in all H2A.Lap1 genes; ii) whose mutations produced a frame shift and a premature stop codon or an in-frame deletion that is deleterious to the function of H2A.Lap1; iii) in which chimera templates were not detected.

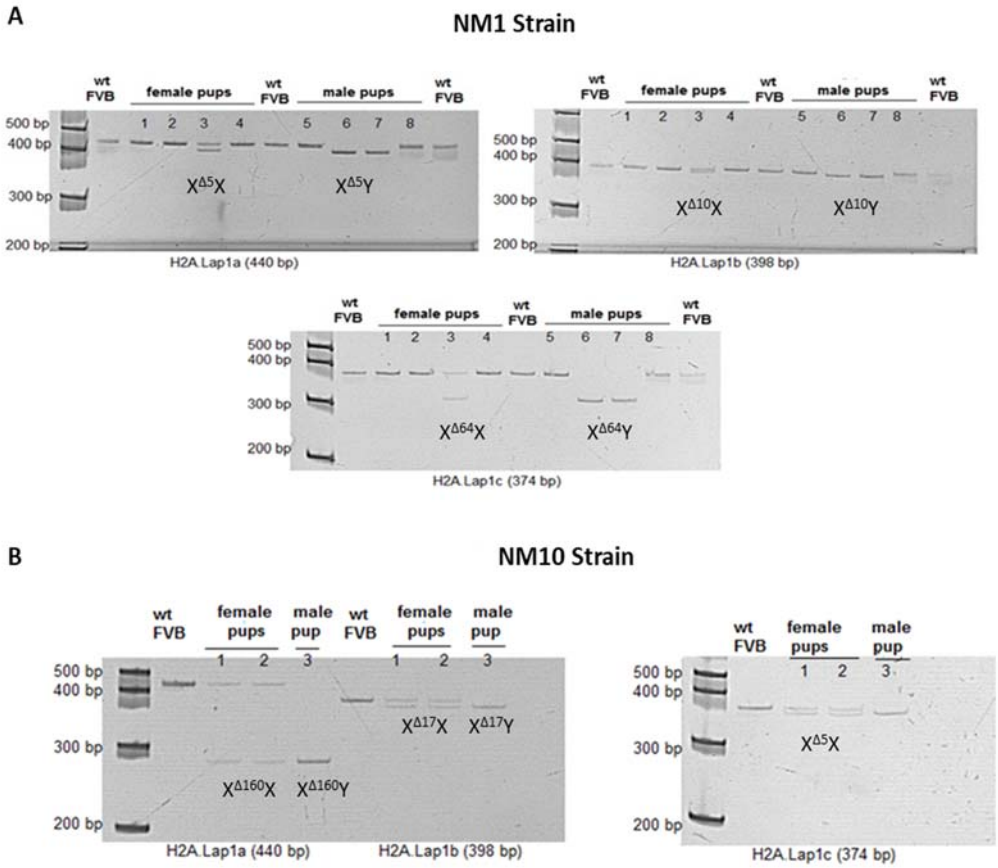
Only two founders, male L74-5 and female L90-3, passed all three criteria. However, we also included a female founder L89-1, in the breeding program, that had a chimera between H2A.Lap1b and H2A.Lap1c on one of the alleles. We predicted that successive breeding would filter away that chimera in future generations, which has indeed happened by generation G3.

Every founder was mated with a wt partner to produce hemizygous and heterozygous G1 offspring (Table 3-8). Mating with a wt partner is an important strategy to remove any nonspecific alterations that might have been introduced into the genome by TALENs. In each new generation, we filtered out 50 % of the original genome, to produce a specific H2A.Lap1KO mouse colony. Thus, progeny of heterozygous female L90-3 with her wt partner were designated NM1; progeny of homozygous female L89-1 and wt male, designated NM10; and progeny of hemizygous male L74-5 and his wt female partner were named NM5. We also mated hemizygous male L74-5 with the homozygous female L90-3 to generate hemizygous and homozygous mutant G1 offspring (strain NM4), in order to produce more mutant pups (refer to Table 3-8). For genotyping the pups that were born to the above couples, we eliminated the CelI assay step and ran the PCR products directly on polyacrylamide gel. This is because their predicted mutations, based on their parent's mutations, were more than or equal to five nucleotide indels, able to be resolved on a polyacrylamide gel, provided the sample was run side by side with wt sample. Further, to confirm genotype, Sanger sequencing was undertaken. Genotyping results confirmed that most mutant pups inherited their parents' mutations (see Table 3-8 and Figure 3-21). However, 28% of pups from all strains combined had unexpected aberrations that are described in the subsequent section.

Table 3-8: Breeding strategy of TALEN founders in order to produce H2A.Lap1KO colony

Genotype of parents	Progeny G1 mouse ID	Predicted progeny genotypes		
		H2A.Lap1a	H2A.Lap1b	H2A.Lap1c
♀**L90-3** + ♂**WT** L1a: $X^{\Delta28}X^{\Delta5}$ L1b: $X^{\Delta10}X^{\Delta10}$ L1c: $X^{\Delta61}X^{\Delta61}$	NM1-G1 ♀/♂	$X^{\Delta5}X, X^{\Delta5}Y$ $X^{\Delta28}X, X^{\Delta28}Y$	$X^{\Delta10}X, X^{\Delta10}Y$	$X^{\Delta64}X, X^{\Delta64}Y$
		(heterozygous female, hemizygous male)		
♀**L89-1** + ♂**WT** L1a: $X^{\Delta160}X^{\Delta22}$ L1b: $X^{\Delta17}X^{\Delta17}$ L1c: $X^{\Delta5}X^{\Delta5}$	NM10-G1 ♀/♂	$X^{\Delta160}X, X^{\Delta160}Y$ $X^{\Delta22}X, X^{\Delta22}Y$	$X^{\Delta17}X, X^{\Delta17}Y$	$X^{\Delta5}X, X^{\Delta5}Y$
		(heterozygous female, hemizygous male)		
♀**L90-3** + ♂**L74-5** Lap1a: $X^{\Delta28}X^{\Delta5}$ + $X^{\Delta5,+286}Y$ L1b: $X^{\Delta10}X^{\Delta10}$ + $X^{\Delta15}Y$ L1c: $X^{\Delta64}X^{\Delta64}$ +$X^{\Delta17}Y$	NM4-G1 ♀/♂	$X^{\Delta5}X^{\Delta5,+286}, X^{\Delta5}Y$ $X^{\Delta28}X^{\Delta5,+286}, X^{\Delta28}Y$	$X^{\Delta10}X^{\Delta15}$ $X^{\Delta10}Y$	$X^{\Delta64}X^{\Delta17}$ $X^{\Delta64}Y$
		(homozygous female, hemizygous male)		
♂**L74-5** + **WT** L1a: $X^{\Delta5,+286}Y$ L1b: $X^{\Delta15}Y$ L1c: $X^{\Delta17}Y$ *only produce all WT male and all heterozygous female.	NM5-G1 ♀/♂	$X^{\Delta5,+286}X$	$X^{\Delta15}X$	$X^{\Delta17}X$
		(heterozygous female, hemizygous male)		

Symbols: ♂, male; ♀, female; Δ, deletion; +, insertion; wt, wild type; G1, 1st generation.


A
NM1 Strain
wt FVB
female pups
wt FVB
male pups
wt FVB
1 2 3 4 5 6 7 8
500 bp
400 bp
300 bp
200 bp
X^Δ5X
X^Δ5Y
H2A.Lap1a (440 bp)
X^Δ10X
X^Δ10Y
H2A.Lap1b (398 bp)
X^Δ64X
X^Δ64Y
H2A.Lap1c (374 bp)
B
NM10 Strain
wt FVB
female pups
male pup
1 2 3
X^Δ17X
X^Δ17Y
X^Δ160X
X^Δ160Y
H2A.Lap1a (440 bp)
H2A.Lap1b (398 bp)
X^Δ5X
H2A.Lap1c (374 bp)

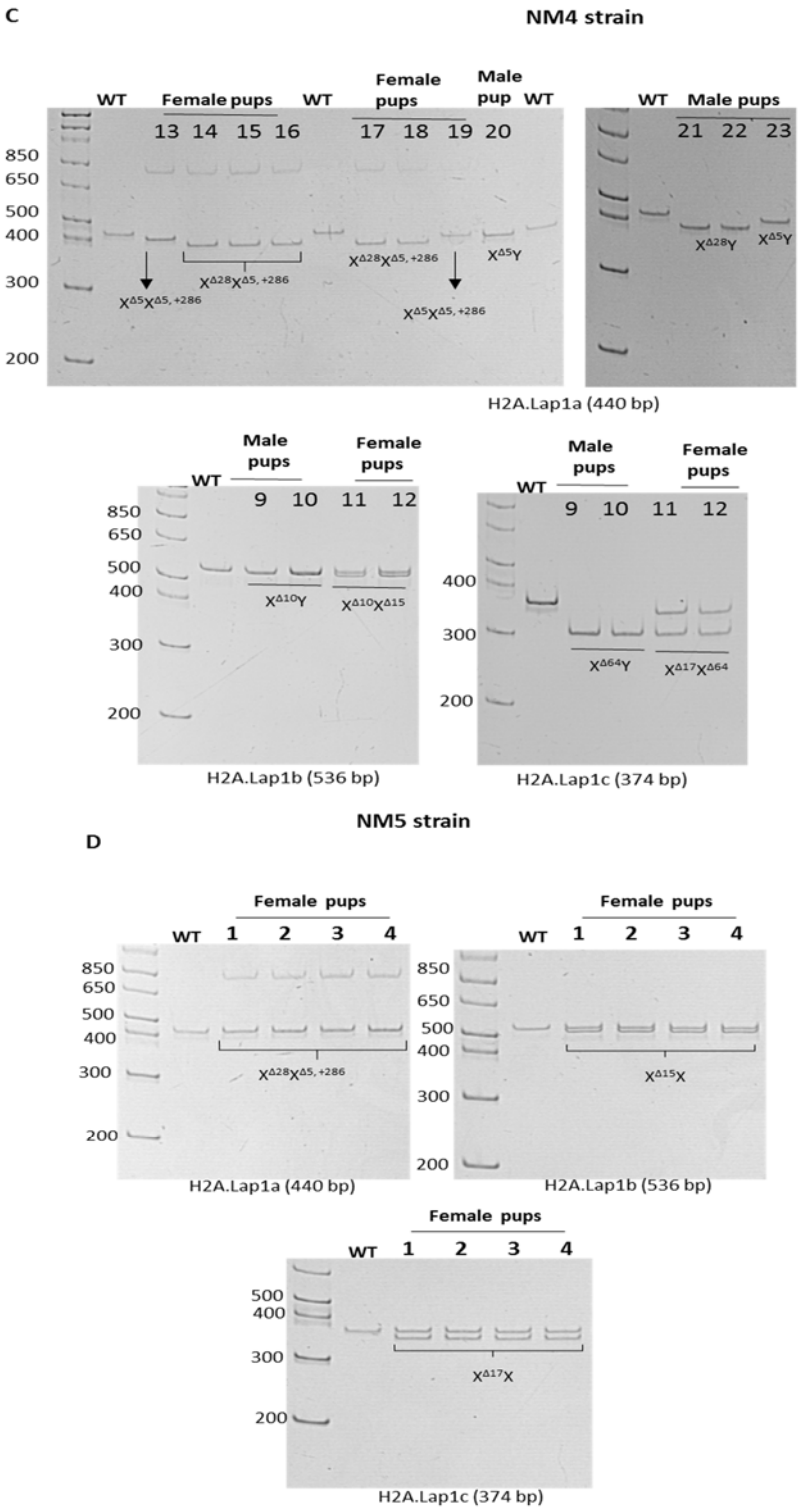


Figure 3-21: Mutations are stably inherited by H2ALap1 TALEN progeny. The gel images indicate the amplicons of each H2A.Lap1 gene from the progeny of TALEN founders. A, pups from NM1 strain; B, pups from NM10 strain; C, pups from NM4 strain; and D, pups from NM5 strain. H2A.Lap1 genes from wt FVB mice were also amplified and loaded in each gel as a control. Symbols: ♂, male; ♀, female; Δ, deletion; +, insertion; wt, wild type; numbers, ID of pups.

3.3.5 TALENs induce stable and inheritable mutations in multi-copy genes but TALEN activity is retained for several embryonic divisions

Two of the founders used to establish H2A.Lap1 KO colonies, L74-5 and L90-3 (referred to as strains NM1, NM4 and NM5) did not present with any chimeric H2A.Lap1 genes or other detectable aberrations, except for the described mutations in the H2A.Lap1 genes. It was surprising to us to see that some of their G1 progeny suddenly presented with novel aberrations within H2A.Lap1 genes that were not detected in their parents. Even the strain NM10, which was produced by crossing wt male with a female mouse L89-1 (harbouring a chimeric H2A.Lap1b/c gene in one of the alleles) produced G1 pups with a novel, previously undetected aberration in H2A.Lap1 genes. Figure 3-22 shows examples of NM10 G1 progeny, in which H2A.Lap1b and H2A.Lap1c genes were lost and a chimera H2A.Lap1b/c was formed instead. Altogether, in G1 generation of all four H2A.Lap1 strains, 22 (28.2%) out of 78 pups contained novel, previously unobserved aberrations in the H2A.Lap1 genes (refer to Table 3-9). Further, the novel aberration spans 3.0 kb (or perhaps even more) in the chromosome (Figure 3-23). For this analysis, gene specific primers were designed to amplify over 2 kb chromosomal regions, including the region containing H2A.Lap1a, b and c genes. The primers were specific to each region. However, the same observation occurred with NM1, where H2A.Lap1b and c are lost and H2A.Lap1b/c chimera is detected instead.

The detection of novel aberrations may be attributed to two possibilities: First, if TALENs induce mutations at multiple nonspecific loci, the genome may become unstable and undergo spontaneous mutations, possibly during mitosis and meiosis. Second, while TALEN activity is specific, it is retained after several early embryo divisions and novel mutations may arise from TALEN's action in embryos of two, four or even eight cells, where extra mutations are introduced *de novo* in newly formed cells.

To test both hypotheses, we continued breeding all four strains, following our breeding strategy in which a heterozygous female was mated with a wt male, so that in each consecutive generation, the original genome exposed to active TALENs was reduced by 50%. In that manner, we produced G2, G3 and G4 generations of pups. If the genome was altered to become unstable (hypothesis 1), we would expect that previously detected mutations may disappear, novel mutations may appear, and most importantly, the frequency of those mutations would not be reduced in each

consecutive generation. However, our analysis of those four generations by PCR and Sanger sequencing revealed that all previously detected mutations were maintained, no more novel aberrations were detected from generation G2 onwards and the frequency of the mutations decreased rapidly. Only 2.3% of G2 pups displayed aberrations seen in G1 and none of the pups of generations G3 and G4 had novel aberrations, but they did retain the original mutations in H2A.Lap1 genes (Table 3-9).

Therefore, we conclude that TALEN technology is capable of generating multi-copy gene knockout mice, where several genes are modified simultaneously by only one pair of TALENs. TALEN mRNA is able to induce stable and inheritable mutations but the activity of the TALENs is retained even after several divisions (in our case two divisions, to a four-cell stage) creating a mosaic mouse. The mosaicism may not be detectable in the founders and mice need to be bred until generation G3, to produce a pure non-mosaic colony. This phenomenon was also observed by others (Tesson et al. 2011)

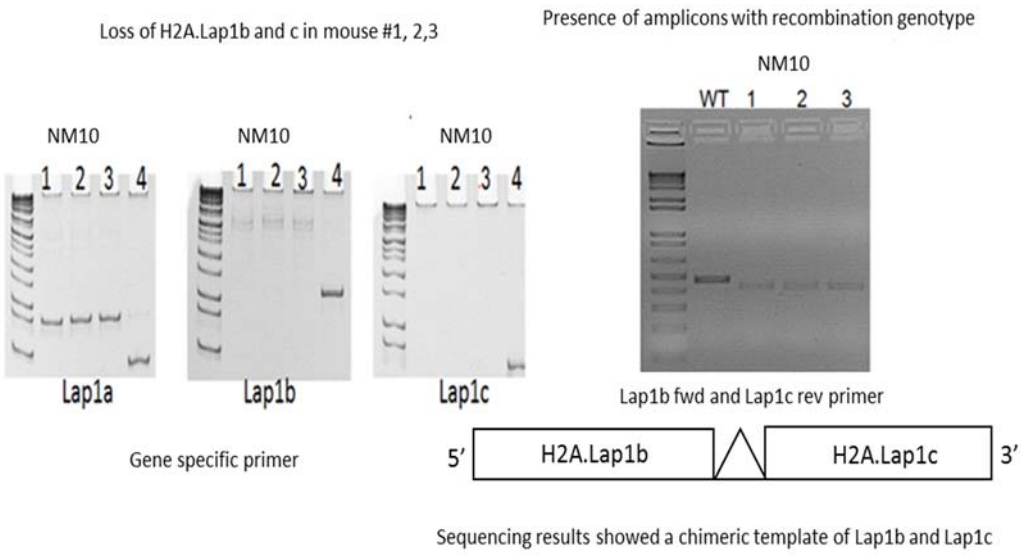


Figure 3-22: The gel image of three H2A.Lap1 genes amplified from pups with recombination genotype. The left panel indicates H2A.Lap1 genes amplified using each gene specific primers. Right panel is the gel image indicating the presence of amplicons when amplified using combination of H2A.Lap1b and H2A.Lap1c primers. Symbol: Δ, deletion; numbers, ID of pups.

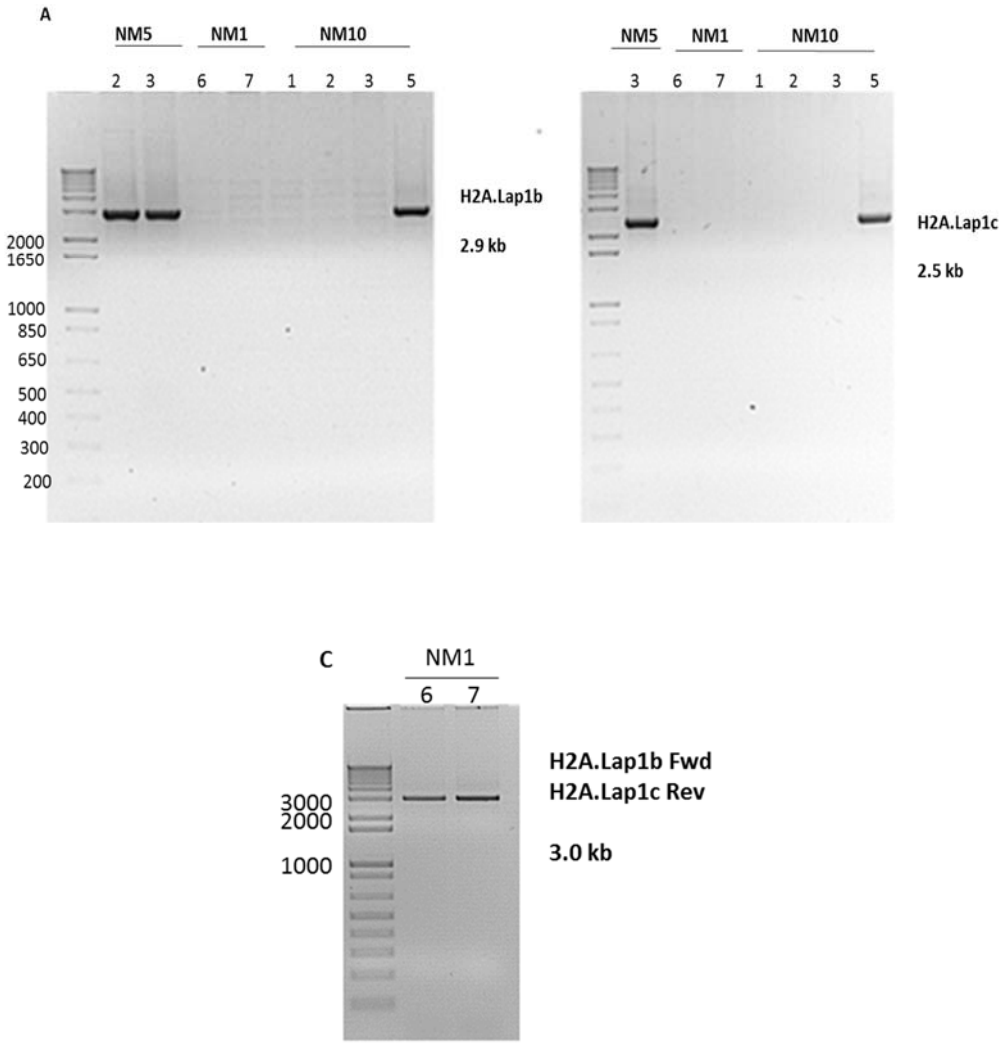


Figure 3-23: The TALEN pups have recombination genotype in H2A.Lap1b and H2A.Lap1c gene. The gel image of H2A.Lap1b and c amplicons from gDNA of pups with recombination genotype. The top panel shows the absence of H2A.Lap1b and c genes in NM1 (#6 and &) and NM10 (#1, 2, and 3). Lower panel indicates the formation of chimeric genes in NM1 which is likely a result of NHEJ between homologous genes, H2A.Lap1b and H2A.Lap1c. No specific band was detected for NM10 using recombinant primers. Agarose gel image of approximately ± 1.5kb region flanking H2A.Lap1b and H2A.Lap1c genes amplified from the mice with recombination genotype. Symbols: Fwd, forward primer; Rev, reverse primer; numbers, ID of pups.

Table 3-9: The percentage of offspring that inherited the TALEN founder mutation and offspring with recombination genotype

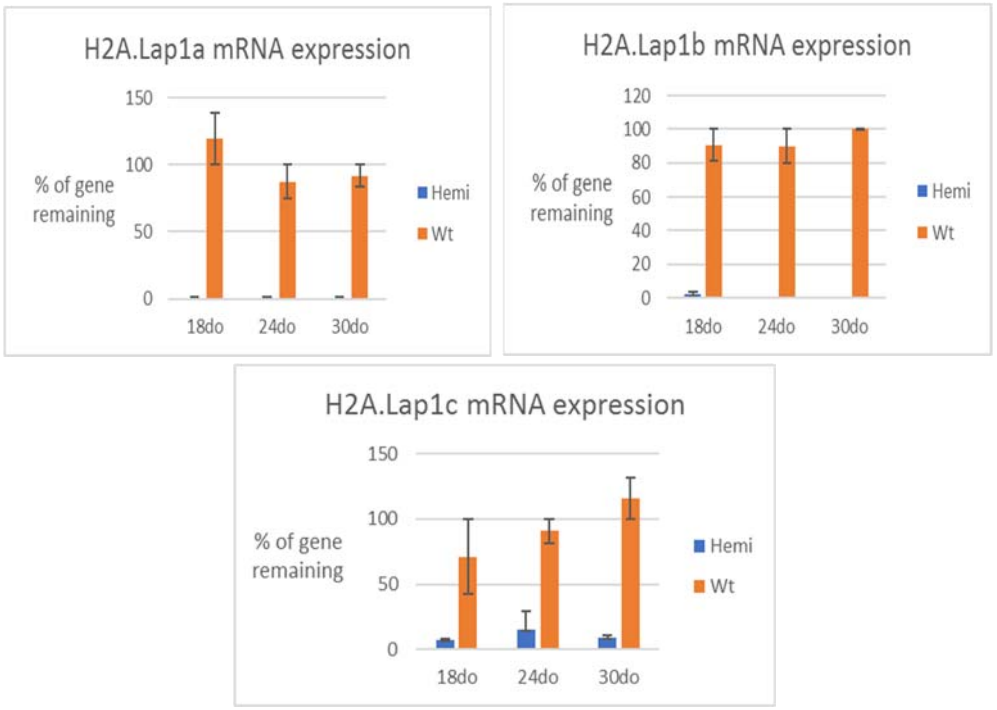
Mice lineage	Total mice produced	% of mice that inherit parents' mutation	% of mice with novel aberration
1st generation (G1)	78	37.0	28.2 (22/78)
2nd generation (G2)	171	39.4	2.3 (4/171)
3rd generation (G3)	198	49.0	0
4th generation (G4)	241	48.0	0

3.3.6 H2A.Lap1 protein loss in KO mice is stably inherited by the progeny

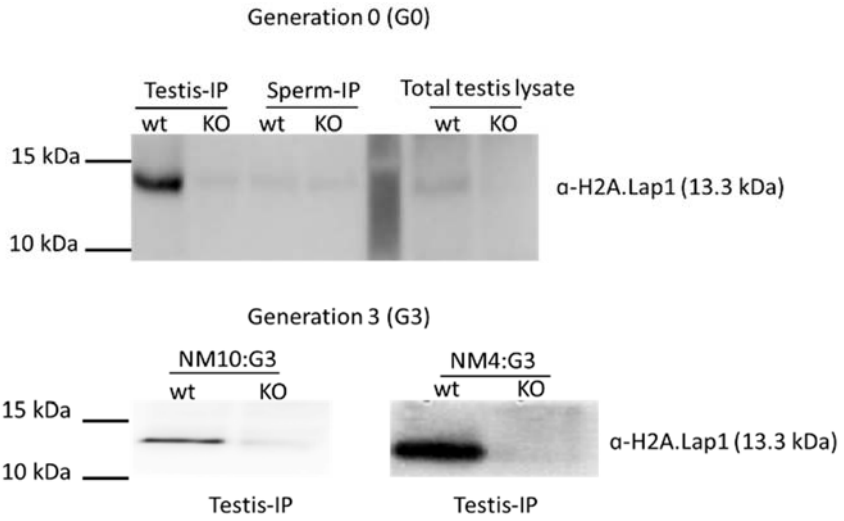
To confirm that any biochemical and phenotypic changes can be attributed to the depletion of H2A.Lap1 protein, we must first confirm the loss of H2A.Lap1 protein in the KO mice, as well as the absence of the wt form of H2A.Lap1 mRNA. We conducted qPCR experiments on the testes of wt and KO mice to check whether the wt form of H2A.Lap1 genes is still present. We designed primers (based on wt sequence) for all H2A.Lap1 genes of which either the forward and/or reverse were located within the known mutated (deleted) region. Figure 3-25 (A) shows that we could not detect the presence of wt H2A.Lap1 genes in any KO testes, suggesting that disruption of the H2A.Lap1 gene successfully produced truncated and non-functional mRNA for all H2A.Lap1 genes.

We then proceeded to determine whether the full-length protein was lost in the KO mice, by conducting immunoprecipitation (IP) of H2A.Lap1 protein, from the testes of KO mice, followed by immunoblot with H2A.Lap1 antibody. We also performed immunofluorescent staining of hypotonic spreads of male germ cells from wt and KO mice using anti-H2A.Lap1 antibody. Indeed, as indicated in Figure 3-24 (B), the immunoblot showed that H2A.Lap1 protein was lost in all samples from the testes of hemizygous founder mice (G0) and their progeny in the third generation (G3). Here the sperm samples were used as a negative control. In addition, immunofluorescent staining (Figure 3-24 (C)) showed undetectable signal with anti-H2A.Lap1 antibody in the KO male germ cells. Hence, this confirmed that the protein of H2A.Lap1 is lost in the KO mice, and this loss is stably inherited by the progeny.

A


H2A.Lap1a mRNA expression
% of gene remaining
150
100
50
0
18do
24do
30do
Hemi
Wt
H2A.Lap1b mRNA expression
% of gene remaining
120
100
80
60
40
20
0
18do
24do
30do
Hemi
Wt
H2A.Lap1c mRNA expression
% of gene remaining
150
100
50
0
18do
24do
30do
Hemi
Wt

B


Generation 0 (G0)
Testis-IP
Sperm-IP
Total testis lysate
wt
KO
wt
KO
wt
KO
15 kDa
10 kDa
α-H2A.Lap1 (13.3 kDa)
Generation 3 (G3)
NM10:G3
wt
KO
NM4:G3
wt
KO
15 kDa
10 kDa
α-H2A.Lap1 (13.3 kDa)
Testis-IP
Testis-IP

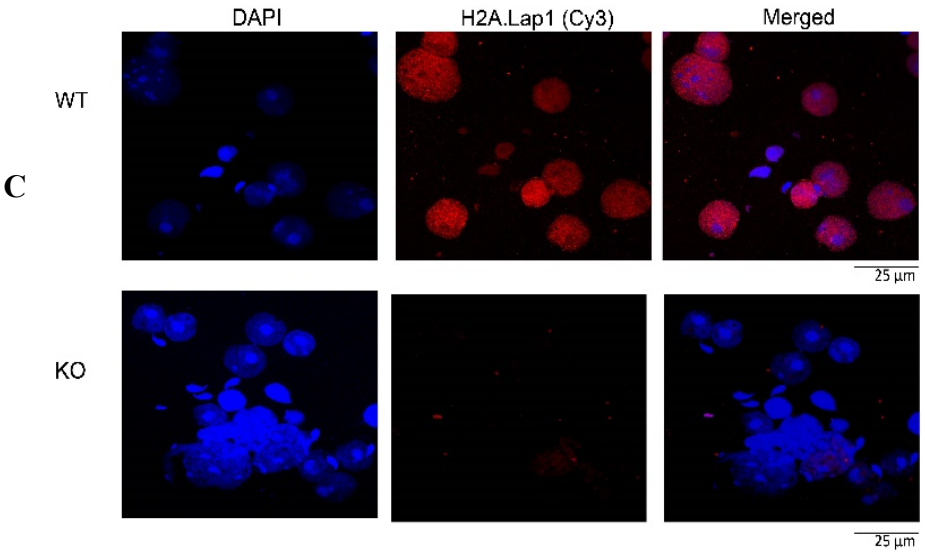


Figure 3-24: The loss of WT mRNA H2A.Lap1 genes and protein. A) the qPCR of three H2A.Lap1 genes showed that the mRNA of all three genes was undetected, showing that the genes have been successfully disrupted by TALEN. B) Western blot showed the immunoprecipitation (IP) of H2A.Lap1 protein from both testis and sperm cells. C) Immunofluorescence staining with H2A.Lap1 antibody on the hypotonic spread of male germ cells from KO mice and their WT siblings.

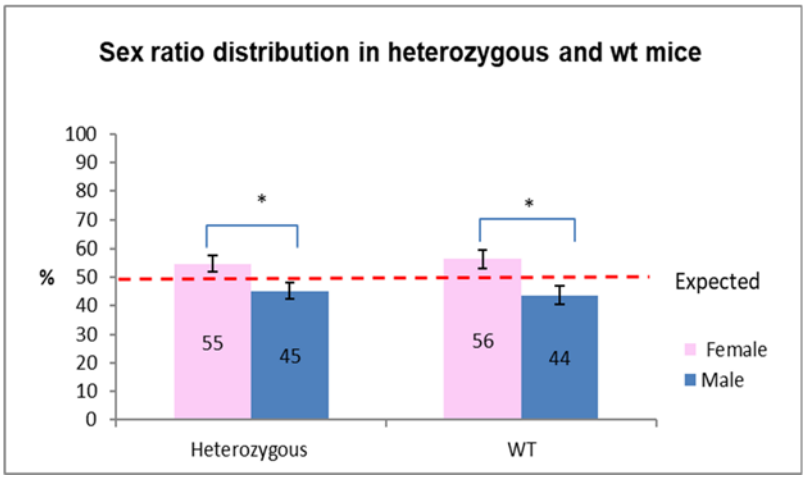
3.3.7 No defect in sex ratio and genotype ratio distribution between KO mice and wt siblings

Next, we examined the distribution of the sex and genotype ratio. The ratio of sexually producing species tends to be 1:1 (50 % male, 50% female). We hypothesised that in the absence of H2A.Lap1, the reproductivity of KO mice might be affected. To test this hypothesis, we conducted several strategies of breeding and examined the distribution of males and females amongst the offspring. We bred i) heterozygous female (H2A.Lap1 $^{+/-}$) with wt male (H2A.Lap1$^{+/+}$), ii) wt female with hemizygous male (H2A.Lap1$^{-/Y}$) and iii) homozygous female (H2A.Lap1$^{-/-}$) with hemizygous male (H2A.Lap1$^{-/Y}$). As illustrated in Figure 3-25 (A), there was a skewness in the distribution of male and female pups born to the heterozygous female (H2A.Lap1 $^{+/-}$) breed with wt male (H2A.Lap1$^{+/+}$) (less male pups were born) (i). Curiously, although the difference was significant, the pattern of the sex distribution shows the same skewness when wt females are bred with wt males indicating that this is a strain specific effect. With regard to the genotype distribution, we observed no skewness in breeding pair (i) i.e. the ratio of wt versus H2A.Lap1 $^{+/-}$ females, and wt versus H2A.Lap1$^{-/Y}$ males both did not change (Figure 3.25 (B)). Figure 3-25 (C) and (D) shows no significant difference in the distribution of male and female progeny from breeding pairs (ii) and (iii) as well. We could not compare the genotype distribution for breeding pairs (ii) and (iii) because breeding pair (ii) only produced all mutant females and all wt male, while breeding pair (iii) produced all mutant male and female offspring.

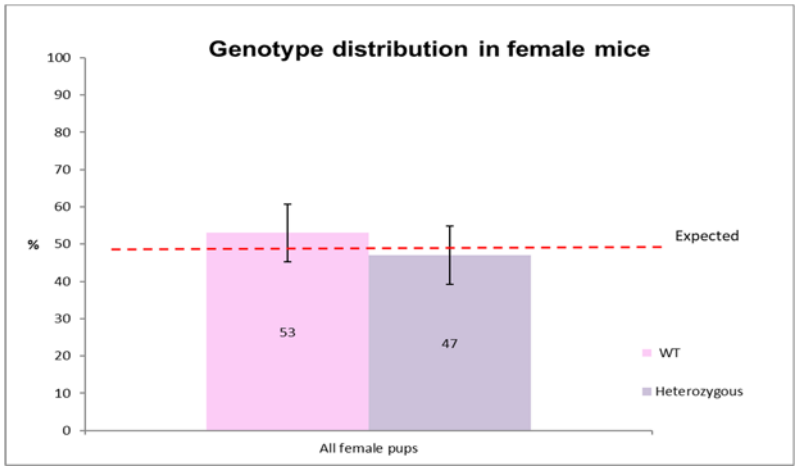
Hence, the disruption of H2A.Lap1 did not cause any abnormal effects in terms of the ratio of males to females for all types of KO mice. However, this was not the case when we crossed the hemizygous (H2A.Lap1$^{-/Y}$) males with the wt females (H2A.Lap1$^{+/+}$) as we observed less pups per litter being born when hemizygous males were bred with wt females, the result of which will be discussed in Chapter 5, section 5.2.

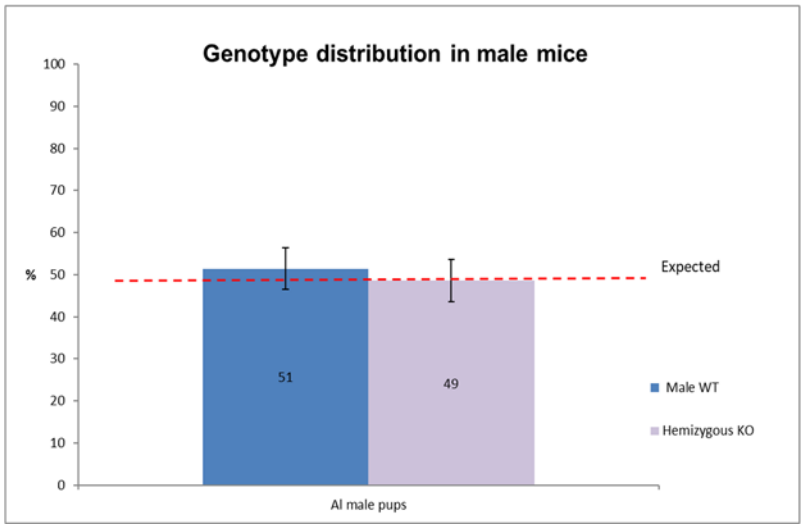
i- Breeding pair: heterozygous female ($H2A.Lap1^{+/-}$) + WT male ($H2A.Lap1^{+/Y}$)

A



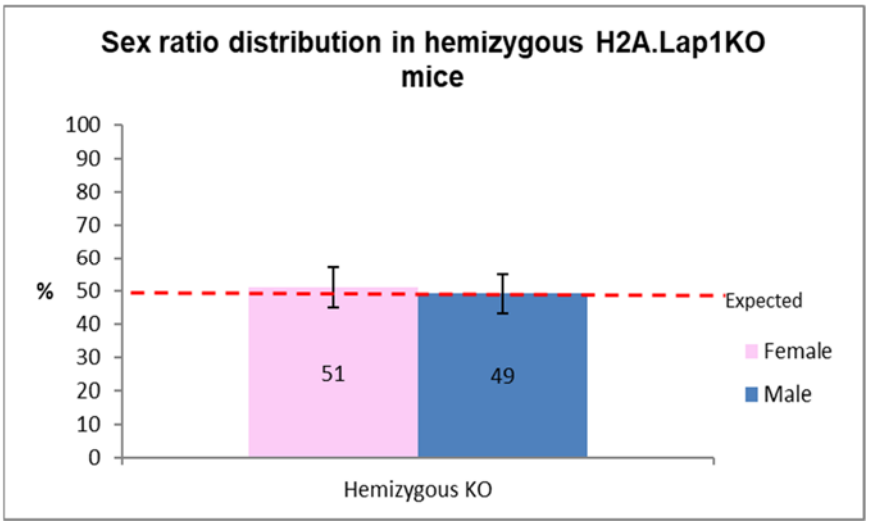
B





ii- Breeding pair: WT female (H2A.Lap1$^{+/+}$) + hemizygous KO male (H2A.Lap1$^{-/Y}$)

C



iii- Breeding pair: Homozygous KO female (H2A.Lap1 $^{-/-}$) + hemizygous KO male (H2A.Lap1 $^{-/Y}$)

D

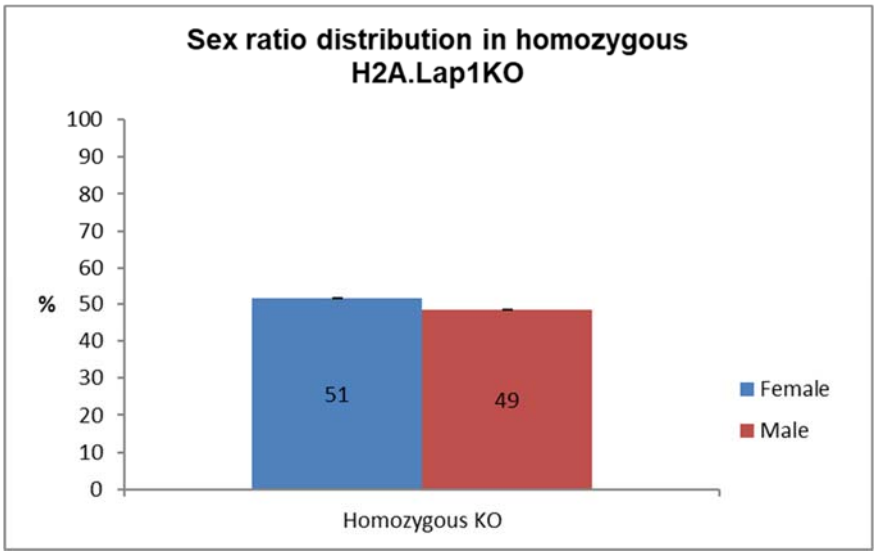


Figure 3-25- The sex ratio and genotype distribution of the KO mice progeny showed no difference compared to the wt mice progeny. (A) Graph indicates skewness in terms of sex ratio distribution in breeding pair (i), although significant, the pattern is the same between WT and KO mice. (B) Graph indicates that there is no significant difference in terms of genotype distribution between both male and female pups produced by breeding pair (i). (C) Graph indicates that there is no significant difference in the sex ratio distribution of offspring produced by breeding pair (ii). (D) Graph indicates that there is no signifcant difference in terms of sex ratio distribution in breeding pair (iii).

3.4 Discussion

In this collaborative study, we have successfully applied TALEN technology to generate the first H2A.Lap1 knockout mouse. This kind of simultaneous knock-out in mice, where several genes were disrupted with only one pair of TALENs, has not been done before to our knowledge.

We tested engineered TALEN pairs for activity in two *in vitro* assays, Dual Luciferase Single Strand Annealing Assay and CelI/T7 endonuclease assay, using cultured mouse Neuro 2a cells. We have found T7EI to be more affordable and more efficient than the CelI enzyme, in cultured cells. Perhaps a more notable improvement in screening for TALEN-induced mutations was the development of a simplified workflow and the introduction of a robust high throughput DNA purification step, using magnetic beads.

Having established that TALEN pair 101421:101422 was the most active against all H2A.Lap1 genes, we used it for microinjections into 532 oocytes, of which 153 (28.8 %) survived to two-cell stage and were implanted into pseudopregnant females. Nineteen pups (12.4 % of initial embryos implanted) were born, of which nine had all three H2A.Lap1 genes disrupted (47.3%). By breeding founder mice, we have shown that TALEN technology induces stable and heritable genomic alterations at multiple specified locations of the genome, thus allowing simultaneous disruption of several multicopy genes with only one pair of TALENs. We have also observed that TALEN mRNA can stay active through several early embryonic divisions, a phenomenon also observed by others (Tesson et al. 2011; Li et al. 2014) and similarly in the CRISPR-Cas9 system (Sung et al. 2014). We have discovered that disruption of highly homologous genes can lead to the formation of gene chimeras and genomic rearrangements which is a concern for TALEN-mediated gene editing and other related technologies, like CRISPR. However, this unwanted side-effect can be eliminated simply by breeding-out unwanted alterations. In our case, all genomic rearrangements were eliminated by generation three.

We have made several important observations about TALEN technology during the establishment of the H2A.Lap1 KO. Firstly, by targeting only one of the H2A.Lap1 genes, H2A.Lap1a of a multicopy gene family, we had a chance to test the specificity and selectivity of TALEN technology. The high specificity was demonstrated when we used the Dual Luciferase Single Strand Annealing assay, on cultured cells (Figure 3-5). TALEN pairs designed against the H2A.Lap1a gene only,

showed no or minimal activity on the other two H2A.Lap1 genes (H2A.Lap1b and H2A.Lap1c, that are 98% and 92% identical to H2A.Lap1a). This is a remarkable result, given the high degree of homology between the genes, and given that the targeted sequences differed by only 2 (in most cases) to 3 nucleotides in the 15 nucleotide target sequences. On the other hand, 13 TALEN pairs were designed to target all three H2A.Lap1 genes, and 10 of these pairs were successful, although with varying activity, so the best pair was selected for an *in vivo* disruption in mouse embryos.

Secondly, we have confirmed the findings of other researchers (Tesson et al. 2011) that TALEN mRNA retains its activity through several embryonic divisions, so that several unique modifications of each H2A.Lap1 gene were detected in some founder mice and their progeny, creating a mosaic genotype. This finding calls for fine-tuning of the stability of *in vitro* produced mRNA, possibly by incorporation of less stable 5'-cap structures.

Thirdly, we observed a novel phenomenon that seems to be specific to simultaneous knock-outs of highly homologous genes – formation of chimeras between H2A.Lap1 genes. Every chimera contained a small deletion of one of the genes, suggesting that the NHEJ mechanism was acting. It is also possible that a subtype of the NHEJ mechanism, called terminal microhomology (Lieber 2010), was employed here, in which distal DNA ends that display at least four nucleotide homology can be joined together, even in the absence of all NHEJ components, at the sites of DSBs. As H2A.Lap1b and H2A.Lap1c genes are 92% identical, such a mechanism is highly plausible. The reason that H2A.Lap1a gene was not a frequent part of a chimera (except in 2 out of 5 chimera cases, L89-3 and L90-6) is that it is positioned on the opposite strand from H2A.Lap1b and c and is 3.631-3.298 Mb away, while H2A.Lap1b and H2A.Lap1c are clustered together (refer to Table 3-1).

This important observation suggests that TALEN knockout technology may cause chromosomal translocations and genomic rearrangements when genes belonging to a multi-copy family are disrupted simultaneously. It is important to point out that other technologies that rely on NHEJ for inducing mutations, like CRISPR/Cas9 technology, are also likely to have a similar effect on homologous genes. However, several rounds of backcrossing of heterozygous mutants to WT mice can successfully eliminate these undesired genomic modifications. In our case, the third generation of pups was free of rearrangement and mosaic mutations.

Finally, TALEN technology showed an interesting trend: it seems that TALEN produced all or nothing activity in our 19 founders. Either all three genes (six alleles in females) were disrupted (nine founders were mutants for all H2A.Lap1 genes) or no disruption occurred at all (ten founders were wt). The TALEN technology is comparable to the widely used alternatives, CRISPR-Cas9, Cpf1, a putative class 2 CRISPR effector (Zetsche et al. 2015) in terms of producing genetically altered offspring.

Injections of Cas9 mRNA and sgRNA at the same concentration as we used (10 ng each) via intracytoplasmic into FVB/N mouse single cell embryo resulted in 87 % of embryos being implanted. Following implantation, 54 % of the transferred embryos produced viable pups, of which 65 % were mutant (Sung et al. 2014). Another study showed a cytoplasmic injection of hCas9 mRNA and sgRNA at 10 ng produced 29.4 % viable pups and of that 45 % were mutants (Mashiko et al. 2013). Meanwhile, injections of "low" concentrations (10 ng) of AsCpf1 RNA, via the pronucleus, yielded only 3 % surviving pups, of which 3.4 % were mutant (Watkins-Chow et al. 2017). In our case, we experienced a rather lower survival rate of embryos, but we believe that it might have been due to adjusting and fine-tuning the injection technique and embryo handling at the beginning of the experiment. The first two rounds of injections resulted in no pups born, while the last round of injections resulted in the highest number of healthy pups born (n=10). Nevertheless, in spite of unsuccessful first rounds, our overall statistics are well within the range of CRISPR and TALEN experiments performed by others, with the mutation rate of 47.3%. The mutation rate observed by Qiu et al. (2013)in their TALEN-modified mice was between 13-67 %. which suggests that the mutation rate occurring in our hands was reasonably high.

Although we did not see a significant difference in terms of distribution of the sexes, this does not rule out that the disruption of H2A.Lap1 might have an effect on a different aspect of reproduction, including testis morphology and a number of pups produced by hemizygous male mice, as mentioned in the earlier part of section 3.3.7, and this will be discussed in Chapter 5. However, first, it was important to show that our TALEN approach did not produce any off-target mutations.

References

Boch, J. et al., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(5959), pp.1509–12. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/19933107.
Christian, M. et al., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186(2), pp.756–761.
Geurts, A.M. et al., 2009. Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. *Science*, 325(5939), pp.433–433. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19628861 [Accessed March 24, 2017].
Gregory, P.D. et al., 2014. Novel DNA-binding proteins and uses thereof. Available at: https://www.google.com/patents/US20140134740.
Li, C. et al., 2014. Simultaneous Gene Editing by Injection of mRNAs Encoding Transcription Activator-Like Effector Nucleases into Mouse Zygotes. *Molecular and cellular biology*, 34(February).
Lieber, M.R., 2010. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry*, 79(1), pp.181–211. Available at: http://www.annualreviews.org/doi/10.1146/annurev.biochem.052308.093131 [Accessed July 25, 2017].
Mashiko, D. et al., 2013. Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Scientific reports*, 3, p.3355. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24284873%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3842082.
Miller, J.C. et al., 2011. A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*, 29(2), pp.143–148. Available at: http://dx.doi.org/10.1038/nbt.1755%5Cnhttp://www.nature.com/doifinder/10.1038/nbt.1755.
Qiu, Z. et al., 2013. High-efficiency and heritable gene targeting in mouse by transcription activator-like effector nucleases. *Nucleic Acids Research*, 41(11).
Sung, Y. et al., 2014. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Research*, 24, pp.125–131. Available at: http://genome.cshlp.org/content/24/1/125.short.
Tesson, L. et al., 2011. Knockout rats generated by embryo microinjection of TALENs. *Nature Biotechnology*, 29(8), pp.695–696. Available at: http://www.nature.com/doifinder/10.1038/nbt.1940.
Vouillot, L., Thélie, A. & Pollet, N., 2015. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda, Md.)*, 5(3), pp.407–15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25566793%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4349094.
Watkins-Chow, D.E. et al., 2017. Highly Efficient Cpf1-Mediated Gene Targeting in Mice Following High Concentration Pronuclear Injection. *G3: Genes|Genomes|Genetics*, 7(2), pp.719–722. Available at: http://g3journal.org/lookup/doi/10.1534/g3.116.038091.
Zetsche, B. et al., 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System Cpf1 is a RNA-guided DNA nuclease that provides immunity in bacteria and can be adapted for genome editing in mammalian cells. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163, pp.759–771. Available at: http://dx.doi.org/10.1016/j.cell.2015.09.038 [Accessed July 26, 2017].

CHAPTER 4

EXOME SEQUENCING TO DETERMINE WHETHER TALENS GENERATE OFF-TARGET MUTATION(S)

4.1 Introduction

In the previous chapter, we described the generation of H2A.Lap1KO mice using TALEN technology. The mutations generated in all three H2A.Lap1 genes in the KO mice were shown to be stably inherited by the progeny up to the current G7 generation. We have shown that TALENs, designed against only one of the H2A.Lap1 genes, have a very high activity for the desired gene and low activity against the other two H2A.Lap1 genes, which provides strong evidence in support of the high specificity and selectivity of TALEN technology. However, we wanted to investigate the effect of TALEN activity on the rest of the genome and to understand whether TALENs, designed against H2A.Lap1, generated any off-target genomic alterations. This kind of validation was important for two reasons: firstly, we wanted to ensure that any physical and biochemical manifestations in the KO mice were a direct consequence of H2A.Lap1 depletion. Secondly, we wished to see whether TALENs were prone to generations of multiple off-target genomic alterations, as has been predicted by several research groups but never validated experimentally (Boel et al. 2016; Tsai et al. 2015; Cho et al. 2014). This is obviously important if genome editing tools are going to be broadly used in research, medicine and agriculture. However, the detection of off-target mutations is more challenging than on-target mutations due to the number of potential genome-wide locations (Kanchiswamy et al. 2016)

There are several methods used to identify off-target sites, caused by site specific nucleases (SSN) including TALEN (Zischewski et al., 2016). The methods include amplification and sequencing of pre-selected off-target sites, exome sequencing, whole genome sequencing (WGS), BLESS (direct *in situ* breaks labelling, enrichment on streptavidin, and next generation sequencing), GUIDE-seq (genome-wide, unbiased identification of DSBs enabled by sequencing), LAM-HTGTS (linear amplification-mediated high-throughput genome-wide translocation sequencing) and Digenome-seq (*in vitro* Cas9-digested whole genome sequencing). Of all these

methods, the targeted sequencing of exons allows the identification of relevant variants in the exomes, but costs only a fraction of whole genome sequencing. Therefore, I decided to identify any potential TALEN- induced off-target sites and use the exome sequencing approach to validate any observed phenotypic changes in the KO mice as resulting from an H2A.Lap1 depletion.

The strategy for identification of off-target mutations in H2A.Lap1KO mice was based on the sequencing of exomes from three mice of the same strain, in three consecutive generations. The breeding scheme was set up in such a way that every new generation would contain only 50% of the TALEN modified genome of the H2A.Lap1 altered parent. The logic was that if there were off-target mutations, the number of genomic alterations detected in KO mice would be reduced by 50% in each subsequent generation of progeny, due to our breeding strategy of heterozygous female (H2A.Lap1$^{+/-}$) mated with wt male (H2A.Lap1$^{+/Y}$). Thus, if for example, the number of off-target alterations in G1 mice was 100, then in G2 that number would reduce to 50, and by G3 the number would be 25. If we see this trend, it would indicate the presence of initial off-target events, and would allow us to evaluate the effectiveness of the TALEN approach.

4.2 Results

4.2.1 Preparation of samples for exome sequencing

Successful next generation sequencing (NGS) depends on the high quality of gDNA input. We extracted gDNA (using a method described in Chapter 2 section 2.3.2.1) from three hemizygous KO (H2A.Lap1$^{-/Y}$) male mice (NM4 strain) originating from three consecutive generations (NM4-G1-#28, NM4-G2-#18 and NM4-G3-#32). One hemizygous male mouse from the NM1 strain, NM1-G2-#23, was also used to determine if any off-target events were common between different TALEN injection rounds (Figure 4-1). Our DNA quantification indicated that the concentration of DNA was between 200 and 320 ng/µl (except for NM1-G2-#23, which was 95.5 ng/µl): the 260/280 ratio was between 1.8-1.9 and 260/230 ratio was between 2.0-2.2, which met the requirement for NGS.

The mice that we chose were genotyped prior to sequencing. The genotype of H2A.Lap1 genes determined for NM4-G1, G2, and G3; and NM1-G2 are H2A.Lap1a ($X^{\Delta 5}Y$), H2A.Lap1b ($X^{\Delta 10}Y$) and H2A.Lap1c ($X^{\Delta 64}Y$). The purified gDNA (Figure 4-1) was sent to the Australian Phenomics Facility (APF) for exome enrichment using the Agilent SureSelect XT2 All Exon Kit, and each exome sample sequenced using paired-end reads at the Biomolecular Research Facility (BRF), on the Illumina HiSeq 2000 sequencer. As discussed above, the strategy of sequencing the genome of mice from three consecutive generations, originating from the same strain, was crucial for identifying and tracking the dilution pattern.

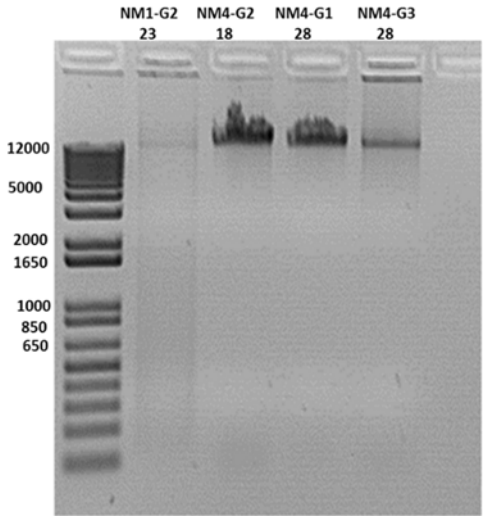


Figure 4-1: Purified gDNA from KO mice strain NM1 and NM4 that was used for exome enrichment and sequencing. Symbols: numbers, ID of pups**.**

4.2.2 Reference sequence used in bioinformatics analysis is of FVB/NJ, not FVB/NJArc

It is important, for reasons outlined below, to describe the background strain of mice that we used in generating the H2A.Lap1KO colony. We used FVB/NJArc mice to generate our KO strain. The FVB mouse line originated from an outbred colony of Swiss mice, established at the National Institute of Health (NIH) in the 1930s for general purposes (N:GP colony). From the N:GP colony, a second colony (N:NIH) was established in the 1940s and selected based on their sensitivity to histamine diphosphate (HSFS/N=sensitive and HSFR/N=resistance) (Taketo et al. 1991; Wong et al. 2012). The HSFS/N colony was found to carry the *Fv-l*^{b} gene which is susceptible to the B strain of Friend leukaemia virus (Taketo et al. 1991). The mice were then inbred to homozygosity and designated FVB/N (for Friend Virus B-type susceptibility), where N refers to NIH, the origin of the mice (Taketo et al. 1991). The strain was exported to the Jackson laboratory (JAX) in 1988 at F38 and designated FVB/NJ. The mice were then bred by the Animal Resources Centre (Arc), Western Australia in 2003 and designated as FVB/NJArc. To date, the genome of FVB/NJArc has not been sequenced. The only genome that was sequenced is of FVB/NJ strain, but the genome has not been assembled (Wong et al. 2012). Therefore, despite being the same strain, there may exist genetic variation between FVB/NJ and FVB/NJArc. As described below, the first step in looking for TALEN's off target effect was to use the fully assembled and annotated mm10 genome which is primarily based on the wild type C57BL/6 mouse strain.

4.2.3 TALENs do not recognise off-target locations that have close sequence homology to targeted H2A.Lap1 gene regions

We collaborated with bioinformaticians from the Department of Immunology, JCSMR, Dr Field and Dr Andrews who have developed a workflow to use massively parallel sequencing reads as a sole data source to identify exonic mutations from ENU-mutated C57/BL6J strain of mice (Andrews et al. 2012). To detect the off-target sites or variants from TALEN-induced mice, they used the mouse exome analysis pipeline, shown in Figure 4-2.

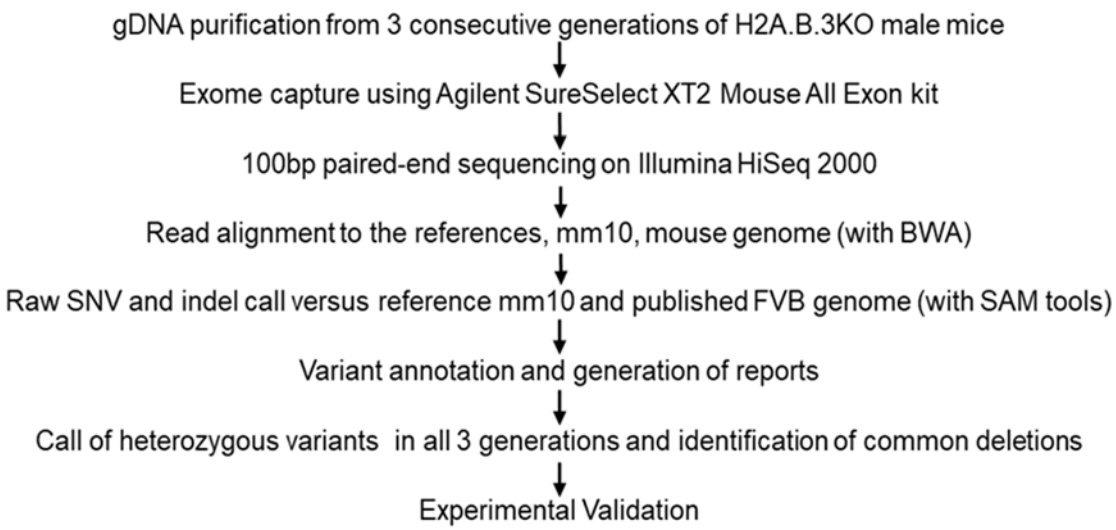


Figure 4-2: Workflow and filtering strategy used to identify off-target sites

Since no reference genome specific to FVB/NJArc strain was available, the published mm10 genome was used as reference sequence in the first step of our analysis. Our analysis of the TALEN DNA sequences (labelled 101421 and 101422, see Chapter 3 Figure 3.4) designed to target H2A.Lap1 genes started firstly with a BLAST (allowing 3-5 mismatches in the sequence) search of either TALEN sequence 101421 or 101422 (which is a target sequence for all H2A.Lap1 genes) against the mm10 reference genome. The search revealed 583 possible mutations, of which 54 overlapped with exons or splice sites; however, no match was detected when the 54 hits were aligned with the exome sequencing data from our KO mice. Secondly, when the TALEN sequences, 101421 and 101422, were analysed in BLAST as a pair, no hits were found. Therefore, based on these two observations, we concluded that the design of the TALEN pair is highly specific to H2A.Lap1 genes and since the double strand break on the targeted sequence will only take place in the presence of a TALEN pair, it is highly unlikely that the off-target mutation will be introduced even if one TALEN plasmid still persists.

4.2.4 Higher coverage was obtained with our sample and a higher number of variants detected relative to mm10 mouse reference genome

The exome of our KO mice was sequenced using 100-bp paired end reads generated by the Illumina HiSeq 2000. The exomes of three mice (NM4-G1-28, NM4-G2-18, and NM4-G3-32) were sequenced to a depth of approximately 350 to over 600x coverage (Table 4-1), providing more data for high confidence variant and indel (insertion-or-deletion) calling. The resultant reads were aligned to the C57BL/6 mouse reference genome (mm10). The lower number of reads resulting from the NM1-G2-23 sample are probably due to degradation of the sample, as seen in Figure 4-1. The exome capture efficiency was uniformly high with approximately 45-50 % of all DNA sequenced being exonic.

Raw sequence reads can be transformed into meaningful information by employing various software applications to detect variants in the genome. In our analysis, the software SAMtools, Pindel and Janda were used to detect raw and structural variants in the KO mice. The initial filtering was done by detecting raw variants using the Sequence Alignment/Map (SAM) format and tools. This software allows the analysis and manipulation of alignments (such as sort, merge and remove duplicates) in SAM/BAM format (Li et al. 2009). It can also be used to detect single

nucleotide polymorphisms (SNPs) and short indels (up to 60 bp), and to present it in the alignment of text-based viewer (Li et al. 2009).

Differences in the number of raw variants, (comprising single nucleotide variations, SNVs), indels and structural variants (SVs), were detected by comparing our exon sequencing data and the mm10 reference sequence (Table 4-2). The numbers resulted from variants with a score ≥ 40 anywhere in the genome. The typical SNVs observed in C57BL/6J at exome covered 30x depth is 7000 SNVs, 1950 deletions, and 1880 insertions. Our colony has shown SNVs at least 22 times higher, small deletions at least 3.3 times higher, and small insertions at least 3.8 times higher than the typical observations (Table 4-3). However, knowing that we align our sequences with a non-related C57BL/6J strain used for mm10 reference genome, we predicted that those differences due to the strain variability. Indeed, the number of SNVs and indels was dramatically reduced when our data was compared to FVB/NJ genome (see below) Furthermore, a much larger size of the sequencing library at 600x depth may also contribute to the increment number of variants.

Table 4-1: The total reads sequenced and number of reads aligned to exonic target regions per sample. The read length is based on Illumina HiSeq 2000 and the approximate size of mouse exome was obtained from Andrews et al. (2012).

Samples	Total reads	Read length (bp)	Total sequence bases (read count x read length) (A)	Approximate size of mouse genome (Mb) (B)	Approximate coverage (A/B)
NM4-G1-28	226905460	100	$22.69x10^9$	37	613
NM4-G2-18	228920578	100	$22.89x10^9$	37	619
NM4-G3-32	129857884	100	$12.98x10^9$	37	351

*Symbols: G1, 1st generation, G2, 2nd generation, G3, 3rd generation.

Table 4-2: The raw variants detected from KO mice sequencing data

Samples	SNVs	Small deletions	Small insertions
NM4-G1-28	1745670	82379	90762
NM4-G2-18	1768957	81956	90005
NM4-G3-32	693686	33823	38321

*Symbols: G1, 1st generation, G2, 2nd generation, G3, 3rd generation.

Table 4-3: Variants output after strain FVB/NJ filter was applied

Samples	SNVs	Small deletions	Small insertions
NM4-G1-28	21256	3889	4378
NM4-G2-18	26646	4064	4445
NM4-G3-32	13417	1589	2032

*Symbols: G1, 1st generation, G2, 2nd generation, G3, 3rd generation.

Given the differences in mouse strains, we applied a variant call format (VCF) filter specific to the FVB/NJ strain to remove the strain variant differences from the list. VCF is a text file that stores sequence data about the position of variants in the genome, based on the reference sequence. In addition, VCF also contains genotype information which allows the user to estimate allele frequencies (Danecek et al. 2011). Indeed, application of the FVB/NJ filter drastically reduced number of SNVs and indels (Table 4-3) but the numbers were still high and required further investigation

Importantly, however, after the variant filter was applied to our exome sequencing data, there was no dilution effect observed from one generation to the next, rather, the variant numbers correlated with the original number of reads. This provides strong evidence that the discrepancies between the reference genome, filtered with FVB/NJ filter, and our exome data are not due to off-target activity of TALENs.

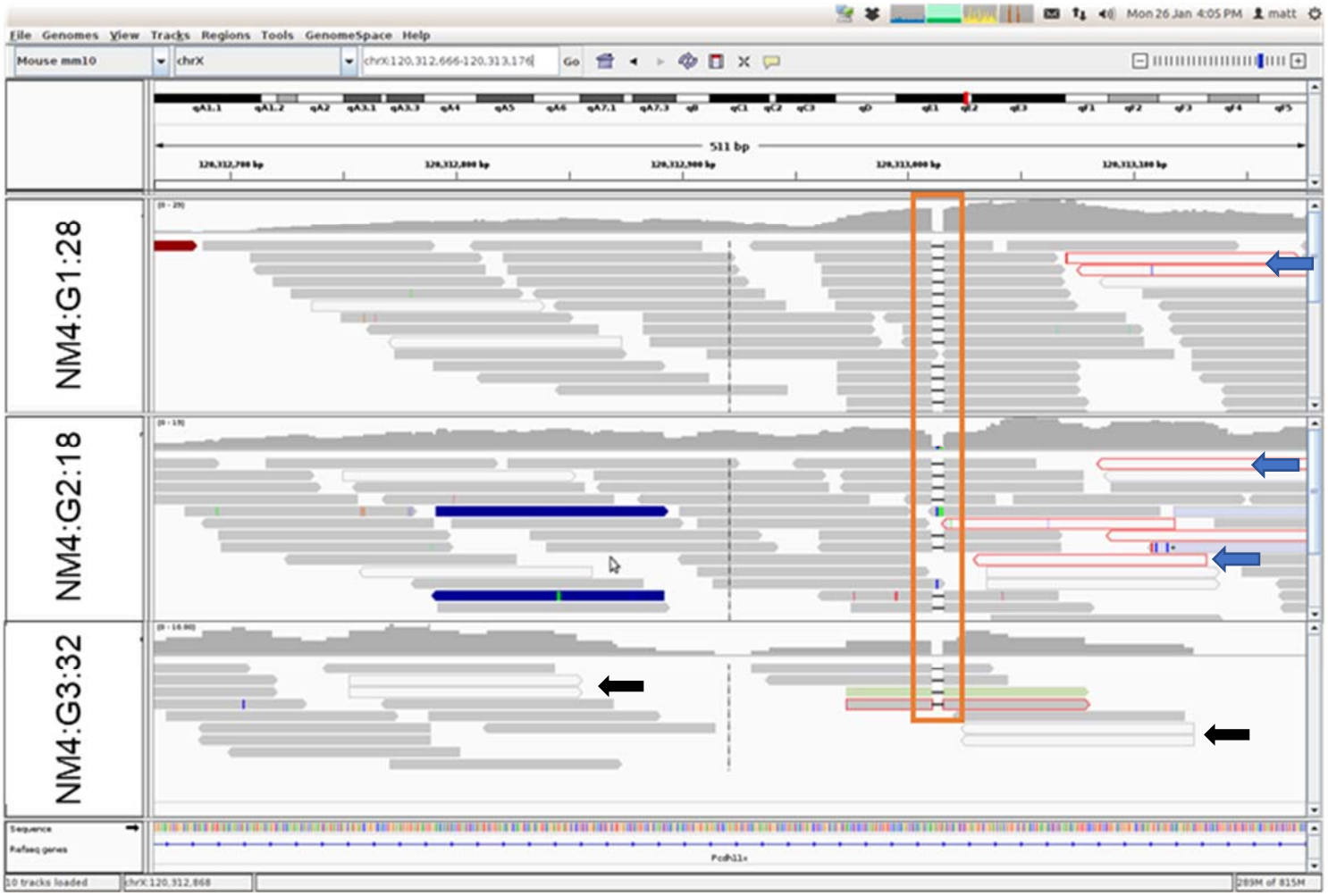
4.2.5 Structural variant (SV) and indel analysis detected expected mutation in all three samples

Further analysis was done to detect SVs because any off-target effects would cause deletions or insertions (indels). Structural variants are genomic rearrangements of 50 bp or more; they comprise balanced as well as unbalanced events. Deletions, insertions, and duplications alter the copy number of the genome and are thus known as unbalanced SVs. Inversions and translocations do not change the copy number and are known as balanced SVs (Tattini et al. 2015; Guan & Sung 2016). Detecting structural variants is a challenging task, hence multiple tools were used to ensure all SVs in the sample were detected. Our samples were analysed using two separate SV algorithms; Janda (unpublished) and Pindel.

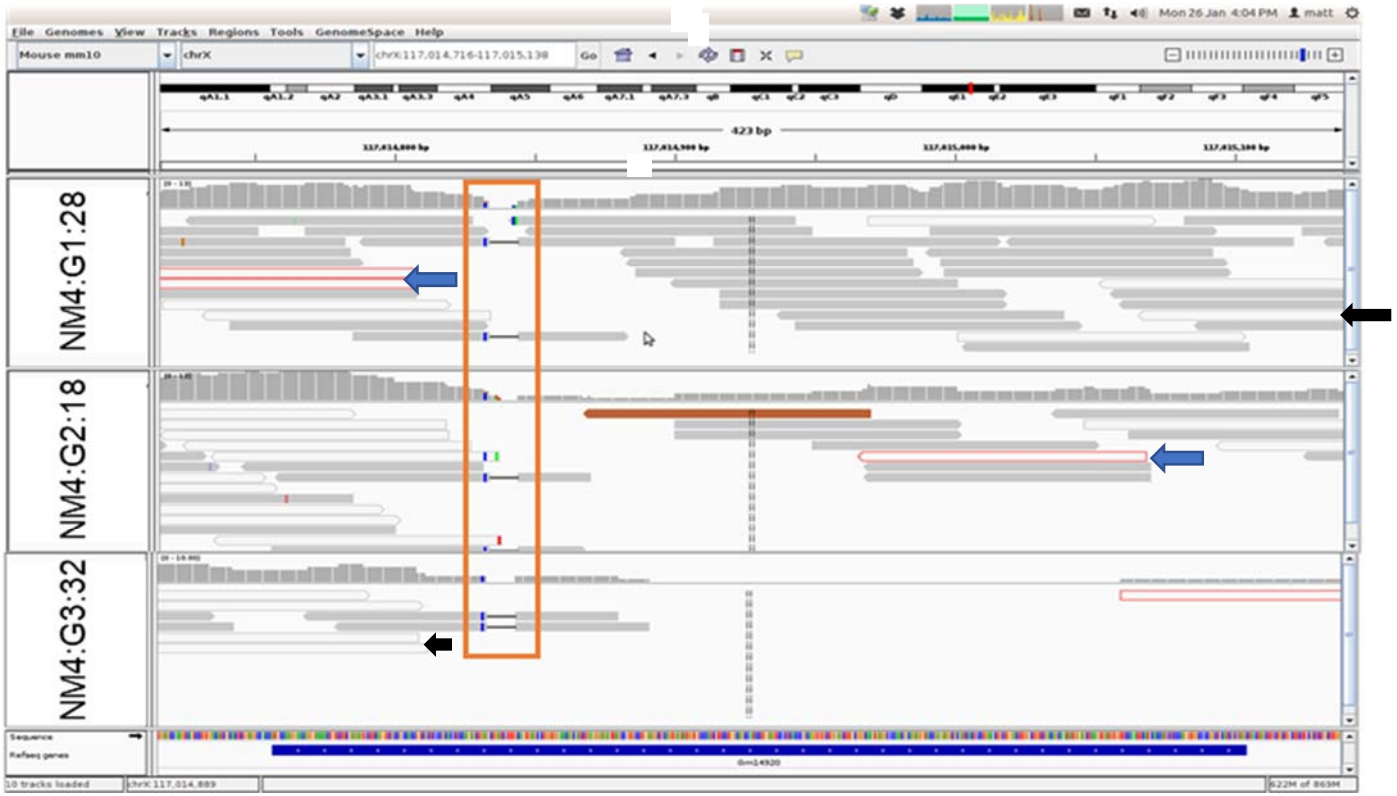
Janda is used to detect and identify SVs (deletions, duplications, translocations and inversions) from aberrant mapped pair reads and for realigning the junction from large SVs in whole genome sequencing data. Meanwhile Pindel uses pattern growth algorithm to identify the breakpoints of large deletions (1 bp -10 kb) and medium sized insertions (1-20 bp) (Ye et al. 2009). While Janda did not detect any structural variants in the region of interest, Pindel detected only three expected deletions in H2A.Lap1 genes: H2A.Lap1a (5 bp) (Figure 4-3 (A)), H2A.Lap1b (10 bp) (Figure 4-3 (B)), as well as a large deletion of 64 bp in H2A.Lap1c (Figure 4-3 (C)). The

coverage for H2A.Lap1b and H2A.Lap1c was low (0-12 and 0-7 times, respectively); nonetheless, the deletion was still detectable. These variants represent the TALEN-specific generated mutations and support our genotype characterization, as described earlier in section 4.2.1. Based on Figure 4-3, we have summarized the indels detected in all three genes, as listed in Table 4-4. Our structural variant analysis did not detect any evidence of unexpected insertion or genomic rearrangement.

A) The alignment of exome sequencing data to H2A.Lap1a



B) **The alignment of exome sequencing data to H2A.Lap1b**



C) The alignment of exome sequencing data to H2A.Lap1c

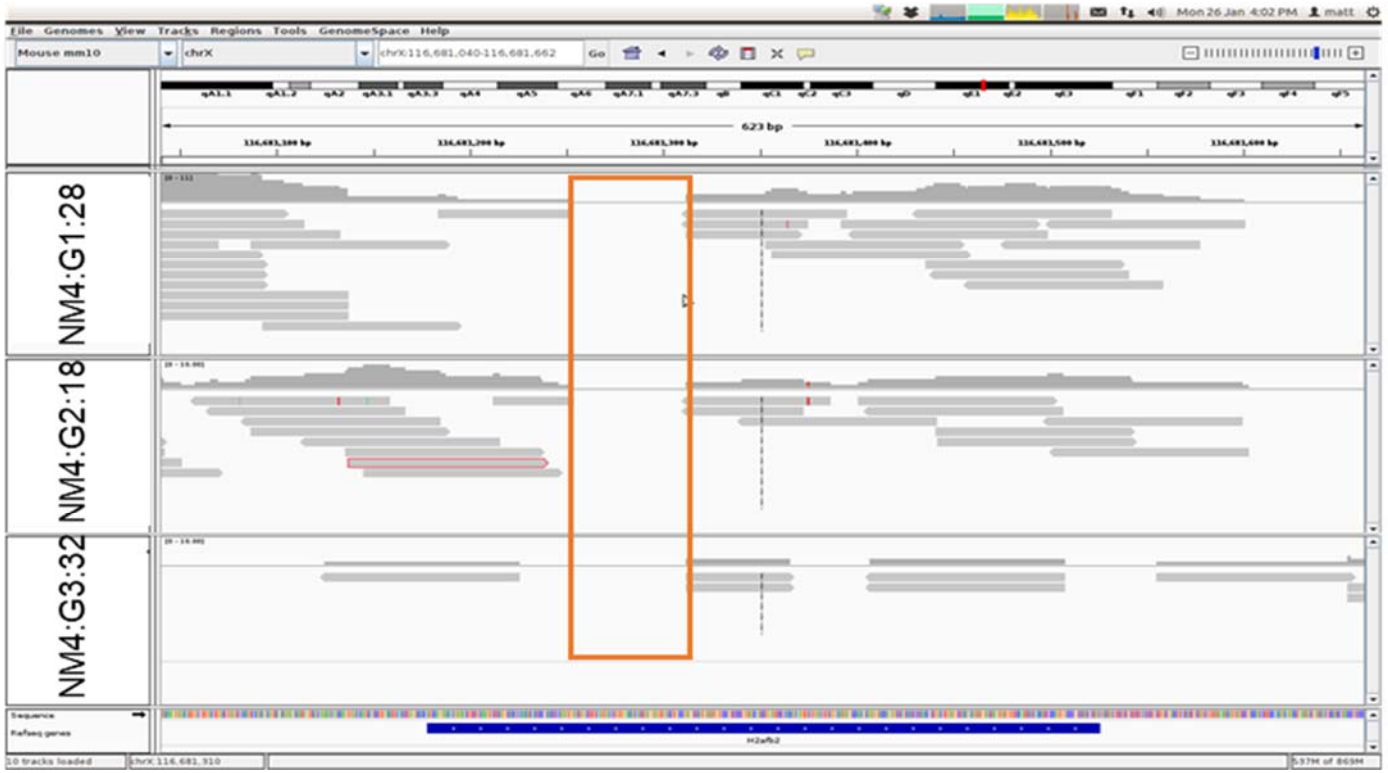


Figure 4-3: Exome sequencing allows to detect all 3 expected mutations in H2A.Lap1 genes. The images showed the output from SV calls by Pindel in all H2A.Lap1 genes. The highlighted region (yellow box) indicates the deleted region detected in all samples by SAMtools. The black arrows pointing at the white box indicate mapping quality of 0; occurs with ambiguously mapped reads. Blue arrows pointed at red box indicate unmapped mate. **A**) The alignment of H2A.Lap1a gene with mm10 reference genome. The yellow box highlighted 5 bp gap; **B**) The alignment of H2A.Lap1b gene with mm10 reference genome. The yellow box highlighted 10 bp gap; **C**) The alignment of H2A.Lap1c with mm10 reference genome. The yellow box highlighted 64 bp gap. Symbols: Δ, deletion

Table 4-4: Expected mutations detected from sequencing data

Gene	Deletion	Insertion
H2A.Lap1c chrX:116681178-116681525	3 samples show 64 bp deletion (116681247-116681312)	0
H2A.Lap1b chrX:117014757-117015104	4 samples show 10 bp deletion (~1171014833-117101842)	0
H2A.Lap1a chrX:120312748-120313095	4 samples show 5 bp deletion (120313011-120313015)	0

Table 4-5: Variants with heterozygous and homozygous deletions detected from sequencing data of KO mice

Samples	No Call	Genotype (deletions, Q>100, whole genome, FVB/N strain variants removed)						Total
		Homozygous			Heterozygous			
		0/0	1/1	2/2	0/1	0/2	1/2	
NM4-G1	111	131	732	2	388	2	27	1393
NM4-G2	116	127	727	3	392	3	25	1393
NM4-G3	314	219	527	8	306	4	15	1393

*Symbols: G1, 1st generation, G2, 2nd generation, G3, 3rd generation.

4.2.6 Further analysis to detect heterozygous and homozygous deletions

Further analysis of the VCF output was conducted to find heterozygous and homozygous deletions within the exome (in collaboration with Dr Sebastian Kurscheid, JCSMR). Indels common to the wild-type FVB/NJ were filtered by applying a cut off quality score of >100. Table 4-5 indicates the detection of all variants with deletions. For a diploid organism, the genotype indicated by two alleles carried by the sample, encoded by a 0 for reference allele, 1 for the first alternative allele and 2 for the second alternative allele. Whenever we found a 0/0, 1/1, and 2/2 it is a homozygous call. The 0/0 means the sample is homozygous with regard to the reference, 1/1 is homozygous with alternative allele 1, and 2/2 is homozygous with alternative allele 2. The heterozygous calls 0/1, 0/2 and 1/2 indicate the heterozygous variants with one reference allele and one alternative allele (0/1, 0/2) and even heterozygous variants with 2 alternative alleles (1/2) (Van der Auwera et al. 2013). In other words, homozygous indels are the ones where both homologous chromosomes of the FVB/NJ genome have the same allele but different from the reference mm10 genome and therefore these are strain specific. On the other hand, if TALENS did produce off-targets, this would produce heterozygous indels.

As shown in Table 4-5, after filtering against the FVB/NJ genome, 417 NM4-G1, 420 NM4-G2 and 325 NM4-G3 heterozygous (0/1+0/2+1/2) indels remained. These numbers again did not display a 50% dilution pattern from G1-G3. To provide further evidence that these remaining indels are due to strain differences between FVB/NJArc and FVB/NJ mice, we interrogated the 19 indels common to all three NM4 G1-G3 mice (Figure 4-4, the location of these putative indels are shown in Table 4.6).

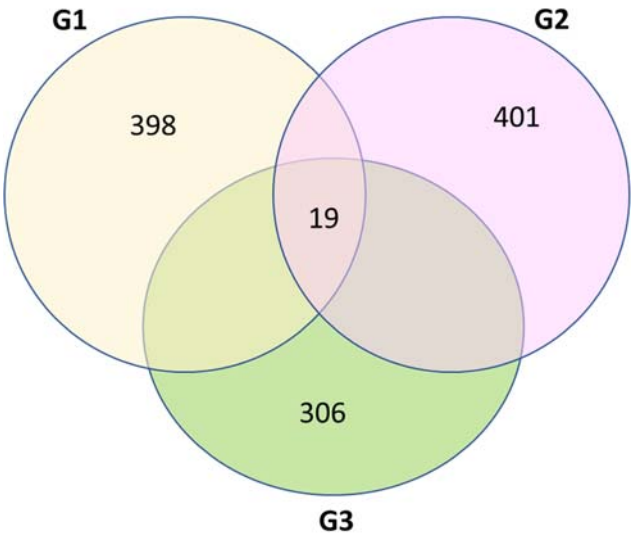


Figure 4-4: Computational analysis predicted 19 putative heterozygous deletions. The number of heterozygous deletion detected in all three generations. The 19 deletions (overlapped) are commonly detected in all 3 mice**.**

Table 4-6: Common variants with deletions detected in all three KO mice that were sequenced

#	Chromosome	Locus start	Locus end	Deletion size (bp)	Gene ID
1	1	171573961	171573965	4	ENSMUSG00000004709
2	1	173270636	173270648	12	ENSMUSG00000049605
3	1	173760152	173760189	38	ENSMUSG00000073489
4	12	115158545	115158549	5	ENSMUSG00000093894
5	12	115335195	115335199	5	ENSMUSG00000095197
6	12	115808522	115808526	21	ENSMUSG00000091087
7	12	115833994	115834007	14	ENSMUSG00000096020
8	17	23882468	23882474	7	ENSMUSG00000096445
9	17	35266213	35266218	6	ENSMUSG00000073411
10	17	36189406	36189415	10	ENSMUSG00000054128
11	17	48145527	48145531	5	ENSMUSG00000073386
12	2	119618224	119618250	27	ENSMUSG00000072980
13	3	7604075	7604079	5	ENSMUSG00000040329
14	4	41195246	41195333	88	ENSMUSG00000028433
15	4	42871823	42871856	34	ENSMUSG00000050141
16	5	113819659	113819689	31	ENSMUSG00000048163
17	7	8245023	8245027	5	ENSMUSG00000053720
18	7	47981089	47981093	5	ENSMUSG00000067173
19	X	95940663	95940678	16	ENSMUSG00000057421

4.2.7 Validation of 19 indels from sequencing data using PCR amplification to determine whether they are true indels or strain specific variations

In order to prove that the 19 indels are indeed strain specific and not off-target effects, I amplified ~300-450 bp regions spanning all 19 indels and compared them to the wt FVB/NJArc mice. To do this, I designed a primer pair for each indel and conducted PCR amplification from gDNA from our hemizygous KO mice (NM4-G3) and the WT FVB/NJArc, as a positive control. The primers are listed in Table 4-7. The amplicons from both hemizygous KO and WT mice were visualised by PAGE. In Figure 4-5 D, we incorporated the amplicon from the first generation of knockout mice (G1) to show that the genotypes of the indels were heritable. Table 4-7 shows that the deletion size of indels varied from 5 to 88 bp. These are the indels with heterozygous deletions, based on the analysis in the previous section. If the indels are specific to our strain, when compared to the WT, they should show the same amplicon size as our hemizygous KO mice sample. As we demonstrated in Chapter 3, a 5 bp deletion in the genotype of our hemizygous KO mice progeny can be detected, therefore a heterozygous sample should be visible on PAGE. We did not detect any differences in the size of the amplicons between hemizygous KO and WT mice (Figure 4-5), nor did we detect any deletion (heterozygosity) in the samples. Furthermore, we sequenced 5 of these putative indels, which confirmed that the hemizygous KO and wt mice are identical in the nucleotide sequence where the indels were predicted to be (Figure 4-5 C).

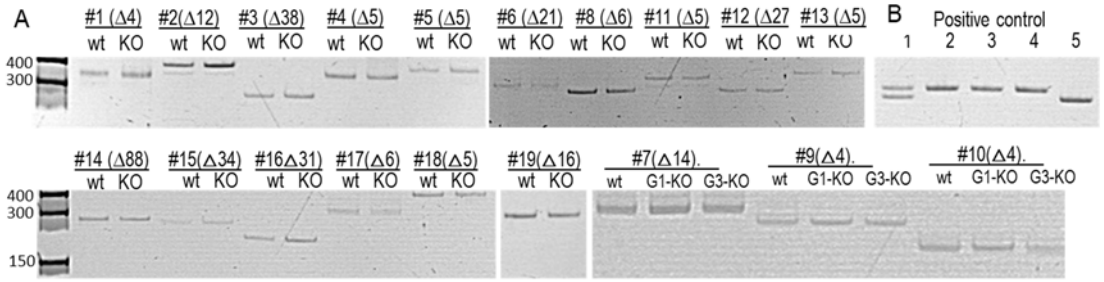
Not all of the actual lengths of the indels determined experimentally corresponded to the predicted size of the indels (Figure 4-5 D), but the results unequivocally show that these 19 indels are specific to our inbred strain of mouse and not a consequence of TALEN activity. The discrepancy between the bioinformatics analysis and the PCR validation experiments (Table 4-8), might be due to an assembly error in the database, used in the variant calling.

In conclusion, based on exome sequencing, we could not detect off-target sites in the exonic region of our KO mice, which confirms the specificity of TALEN targeting to their recognition sequences. Further, the mutations generated by TALEN are stably inherited by the KO mice progeny.

Table 4-7: List of primers used to amplify 19 indels from TALEN mutant mice

Indel #	Gene ID	Sequence (5'-3')	Expected product length (bp)
1	ENSMUSG00000004709 (Δ4)	F- AAGAAGACAGAACAGGGCTCAC	419
		R- TAGCTCACATTGTCATCCTTGG	
2	ENSMUSG00000049605 (Δ12)	F- GACAGTGTACACATTGGTCATT	458
		R- TGGAGATGATGAGAATGTAAGA	
3	ENSMUSG00000073489(Δ38)	F- TTAATGCCACATGATCAGTTAT	426
		R- CCTAGACCATCTATTTTGGTTT	
4	ENSMUSG00000093894 (Δ5)	F-AAGTCTACAAGCAAACCTTCTC	452
		R-CATATACATGGGTGACAATGAC	
5	ENSMUSG00000095197 (Δ5)	F-GCTGTATCTGAGCAAGTCTACA	464
		R-ATACACTTGGGTGACAATGAC	
6	ENSMUSG00000091087 (Δ21)	F-TAGATGGACATTTACAAAGCTG	412
		R-TATCCTCACCTGTTCCAAAC	
7	ENSMUSG00000096020 (Δ14)	F-TAGATGGACATTTACAAAGCTG	402
		R-GCTACACCTTCACAAGCTACTA	
8	ENSMUSG00000096445 (Δ6)	F-GTTGTGATGTCTTCTGTGCTC	405
		R-GTACTCAGAACTGCCCACATA	
9	ENSMUSG00000073411 (Δ5)	F-GGAAGGAGCAGAATTACACAT	468
		R-GTTCAAGGAAGATCTTGACACT	
10	ENSMUSG00000054128 (Δ10)	F-GTGCACTCCTGTGTCTTTTT	451
		R-GCTGGACTACTACAACCTGAGT	
11	ENSMUSG00000073386 (Δ5)	F-TTCTTGAGTCCTCTTCAACTCT	413
		R-CTCTCTGACAGGAAGCATTTAT	
12	ENSMUSG00000072980 (Δ27)	F-ACTCGGGAAAACAAAACAAAC	418
		R-AAAGAAATATGGCGACTCTCT	
13	ENSMUSG00000040329 (Δ5)	F-CAATGAAAGGGGTATAGTTCAT	445
		R-AATTGTGAGAAGTTTCATTGTG	
14	ENSMUSG00000028433 (Δ88)	F-CAGGGTCTAGTTTGTCCAGTAT	401
		R-TTGCCATCTTTATCCTACAGAC	
15	ENSMUSG00000050141 (Δ34)	F-AGGAGCTGACAGAGTCAATTT	401
		R-GACGACTCCTGTGCTCAGAT	
16	ENSMUSG00000048163 (Δ31)	F-CAGACGTTGTAAACAGAGTTTT	409
		R-ACACAGTGGAGTCTAACCTCAG	
17	ENSMUSG00000053720 (Δ5)	F-CAGACACTGAAGAACACTAGCA	409
		R-TCACATTTGGAATTGTATTCAC	
18	ENSMUSG00000067173 (Δ5)	F-CAGGTAGAGGCTCTTCATCA	418
		R-ACCCGATGTTTTTGTTTATAGT	
19	ENSMUSG00000057421 (Δ16)	F-TCTACATCATCACTTTCCTCAA	411
		R-GCGAGTCTGTTCCATTTATACT	

Symbols: Δ, deletion



C

```
Deletion #1
mm10 CTGAAGGCCCAGTGGAAGCCCTGGACTAATGGGACTTGTCAACTGTTTTTGTCCTGCTT
wt   CTGAAGGCCCAGTGGAGAC----------------TTGTCAACAGTTTTTGTCCTGCTT
KO   CTGAAGGCCCAGTGGAGAC----------------TTGTCAACAGTTTTTGTCCTGCTT
     ****************** *               ******** ****************

Deletion #3
mm10 AGTCCCCGCCTGAGCTGTGGAAGTCCTCGCCTGAGCTGTGGAAGTCCCCGCCTGAGCTGTG
wt   AGTCCCCGCC------------------------------------------TGAGCTGTG
KO   AGTCCCCGCC------------------------------------------TGAGCTGGG
     **********                                          ******* *

Deletion #16
mm10 TGGCTTTGAGGTGTCTGCCTCCATGGGTGCTGGCTGCGAGGTGTCTGCCTCCATGGGTGCTGGCTGCGAGGTCTCTGCCTCCATG
wt   TGGCTATGAGGTG-------------------------------------------------------------TCTGCCTCCATG
KO   TGGCTTTGAGGTG-------------------------------------------------------------TCTGCCTCCATG
     ***** *******                                                              ************

Deletion #14
mm10 CGCTCACCGGGGCATCCTGCTGTAGGTGGTGGTGCAACAGCTGGGAGTGGGGCTGCTG
wt   CGCTCACCGGGGCATCCTGCTGTAGGTGGTGGTGCAACAGCTGGGAGTGGGGCTGCTG
KO   CGCTCACCGGGGCATCCTGCTGTAGGTGGTGGTGCAACAGCTGGGAGTGGGGCTGCTG
     ************************************************************

Deletion #15
mm10 TTTGGAGAGCTCTGAACTCTGGCCATTCAACACTTTGGAGAGCTCTGAACTCTGGCC
wt   TTTGGAGAG--------------------------------CTCTGAACTCTGGCC
KO   TTTGGAGAG--------------------------------CTCTGAACTCTGGCC
     *********                                 ***************
```

D

Deletion	#1	#3	#14	#15	#16
Predicted	△4	△38	△88	△34	△31
Observed in FVB/NJArc and H2A.Lap$^{-/y}$ KO	△16	△42	Nil	△33	△60

Figure 4-5: Interrogation of predicted off-target heterozygous deletions in KO mice. A: 300-400bp regions surrounding 19 common predicted heterozygous deletions were amplified from gDNA of KO and FVB/NJArc in-house wt mice. PCR products size of a wt and KO was compared by resolving on 10 % polyacrylamide gel, side by side. The predicted deletion is indicated with △, followed by a number of nucleotides. The image represents the template amplified from G3 KO mouse (NM4:G3-32) and our inbred wt FVB/NJArc. The templates showed the putative indels constantly detected in SNVs analysis. **B)** Positive controls showing that a deletion as small as 5-10 nt can be resolved on 10 % polyacrylamide gel. Lane 1, a heterozygous △5nt, lanes 2, 3, 4 wt.; lane 5, homozygous △10nt **C)** 5 out of the 19 PCR products were sequenced by Sanger sequencing to show that FVB/NJArc and KO mice have identical sequences but they are different from the mm10 reference genome. **D)** Table showing that the prediction tools are often inaccurate at predicting the presence or the size of a putative deletion. Symbols: Δ, deletion; #, indel's number as indicated in table 4-6; G1, 1st generation, G2, 2nd generation, G3, 3rd generation.).

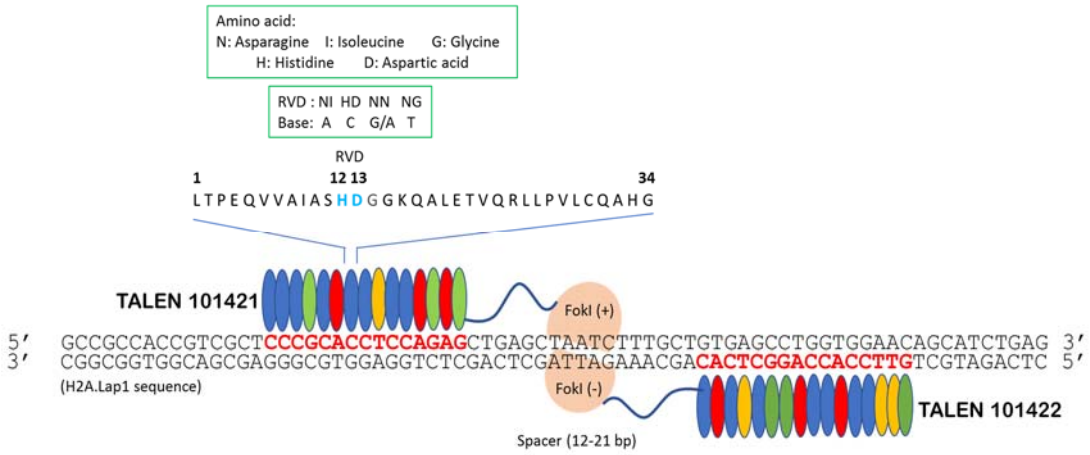


Figure 4-6: FokI will only cleave when TALENs bind to both sense and antisense strand. Image showed a schematic representation of a TALEN pair (adapted and modified from Kim & Kim (2014)). Each TALEN is composed of transcription activator-like effectors (TALEs) at the 5' terminus and the Fok1 nuclease domain at the 3' terminus. Each TALE (~34 amino acids) repeat independently recognizes its corresponding nucleotide base with 2 repeat variable di-residue (RVD) located at 12th and 13th position**.**

4.3 Discussion

In this chapter, we have described how we employed bioinformatic filtering following exome sequencing, as a strategy to determine the location of possible indels to detect off-target sites as well as to validate on-target mutation. This strategy was important in determining the specificity of the TALEN genome editing tool. The key element in genome editing is the specific recognition of the targeted DNA sequence, within a large genome. In TALEN, the specificity of sequence recognition comes from protein-DNA interactions. TALENs are artificial endonucleases in which a DNA binding domain (obtained from TAL effector (TALE) protein) is fused to the cleavage domain of Fok1 endonuclease which is dimeric in nature. The nucleotide specificity is determined by amino acids 12 and 13 of the 30 to 34 amino acids in the repeat domains of TAL (Figure 4-6).

TALEN was shown to have specific cleaving activity for all three functional H2A.Lap1 genes (Chapter 3). This is also supported by our BLAST search on the nucleotide sequence designed to target H2A.Lap1 genes from the 101421 and 101422 TALEN plasmids. We used mm10 as a reference genome with either one of the 15 nt sequences, corresponding to H2A.Lap1 genes, that were targeted by TALENs, and our BLAST search resulted in 54 possible sites on the exonic region. However, none of the indels matched to the indel list generated from exome sequencing data of FVB/NJARc KO mice. The same was the case even with 3-4 mismatched nucleotide allowed. Thus, this validates the specificity of Sangamo's design of TALEN's for all functional variants of H2A.Lap1.

Next, the analysis of detected SNVs and indels was performed. In order to do that our approach was to monitor whether there is any dilution in the variant output from each generation of mouse (G1-G3). We applied several variant callers to filter our sequencing data, and for every variant caller (GATK. Pindel, Janda) applied, we did not see any pattern of dilution that could suggest the presence of unexpected indels in our mice. Although we still observed a high number of variants in the output, the number did not correspond to the dilution pattern but to the sequencing depth. Importantly, TALEN activity usually does not generate single nucleotide variants (SNVs), so the presence of a high number of SNVs was also indicative of the differences between strains rather than the being a product of TALEN activity. In addition, we detected the expected mutations in the sequencing data for all H2A.Lap1 genes, in all three mice. Hence, this validates our genotype analysis

presented in Chapter 3. Further, not only have we proved that there is no dilution pattern from one generation to the next, we have also shown that the high number of variants detected was specific to our inbred FVB/NJArc mouse strain.

Since we detected a high number of variants that are not present in the reference mouse genome database (primarily based on C57BL/6 mouse), we conducted further variant filtering by calling the genotype of the indels to see whether there was any dilution pattern. Again, we did not observe a dilution pattern in heterozygous indels. Interestingly, we did detect 19 heterozygous indels that were shared between G1 to G3 mice. Our careful experimental validation that included amplification of each genomic region and analysis of its sequence, revealed that those putative indels are a feature of in-house FVB/NJArc strain and have not resulted from TALEN activity.

According to the published literature, most studies conduct initial computational filtering to predict off-target sites, prior to establishing their KO model. Following this, various sequencing methods are utilized to further identify any potential off-target sites. However, not many have taken the sequencing output and tested it experimentally (Boel et al. 2016; Liu et al. 2015; Cho et al. 2014). We have extended the analysis in order to prove, experimentally, that the 19 heterozygous indels are true variants specific to our inbred strain. Our validation of the 19 indels via PCR reaction, and 5 by Sanger sequencing, showed that there is no difference between the indels in the WT and KO mice, proving that these 19 heterozygous indels are indeed variation specific to our inbred FVB/NJArc strain. We also propose that this experimental validation step should become standard practice when determining the specificity of indels from NGS output. This is especially important when the KO model established, does not originate from the same background strain as the reference genome used in filtering variants bioinformatically.

Our analysis has proven that TALENs display high specificity in gene targeting. However, during the course of this study, the CRISPR/Cas9 system has overtaken TALENs as the tool of choice for genome editing. These two editing tools are often compared for their ease of customization. The Cas9 nuclease can be easily retargeted to new DNA sequences by simply purchasing a pair of oligos encoding the 20-nt guide sequence. In contrast, retargeting of TALENs to a new DNA sequence requires the construction of two new TALEN genes. Although a variety of protocols exist for TALEN construction, it takes substantially more time to construct a new

pair of TALENs. However, it appears that TALEN specificity is higher than CRISPR. Studies have shown that mutagenesis of sequences with a high similarity to the on-target site is 77 % for CRISPR and only 1 % for TALENs (Veres et al. 2014). In addition, it has been shown that RNA guided nuclease (RGN) can introduce unintended indels at off-target sites with as many as five mismatches (Tsai et al. 2015). Although the efficiency of CRISPR RNA guide nucleases (RGNs) in targeting has improved in recent years (Lee et al. 2016), the report on off-target sites, until now, has been unavailable (Nakajima et al. 2016). Further, despite the development of various computational programs to predict off-target sites, they still miss a fair number of undetected off-target regions. For instance, the MIT CRISPR Design Tool (http://crispr.mit.edu) and the E-CRISP (http://www.e-crisp.org/E-CRISP), which were used to predict off-target cleavage generated by CRISPR, have been utilized widely by researchers (Singh et al. 2015; Puente et al. 2015; Sánchez-Rivera et al. 2014; Heigwer et al. 2014; Doench et al. 2014; Long et al. 2014)to predict their off-target sites based on their degree of similarity to the intended targeted site. However, they are still unable to pick up a number of off-target sites because these two programs simply do not consider off-target sites bearing more than 3 or 4 mismatches (Tsai et al. 2015). One concern of the off-target effect is that they might introduce alterations to a tumour suppressor gene, such as a chromosomal translocation, which could lead to unfavourable clinical outcomes (Tsai et al. 2015). Many steps have been taken with CRISPR to reduce the problem of off-target. One such approach is the double-nicking strategy (Tsai et al. 2015) which has added more complexity to the design for the CRISPR/Cas9 system. Even with the improved double nicking strategy, CRISPR still has off-target sites, as demonstrated in (Tsai et al. 2015).

In an approach similar to that of CRISPR, we generated several sequences to target the intended sites, however, we tested them in the luciferase assay and measured their activity efficiency in cell lines (as described in Chapter 3 section 3.1). This approach of combining the screening methods in order to ensure TALEN cleaves specifically is similar to that suggested by Ran et al. (2013) to be conducted on CRISPR-Cas9 system. TALEN, on the other hand, cleaves specifically with a higher degree of flexibility in design. Researchers can virtually target any sequence simply by manipulating the four-main repeat variable di-residues (RVDs) in the DNA binding domain of the TAL construct. This simple one to one recognition of

TALEN, together with the requirement of flanking 5' thymine, not only have given TALEN the advantage over ZFN, but also over CRISPR.

The main objective of this chapter was to determine whether TALENs could cause any off-target cleavage and we have successfully shown, to the best of our ability, that TALEN did not induce any off-target effects in the exonic region of our KO mice. Although we did not conduct whole genome sequencing, our sequencing result from the exome region clearly indicates that no disease-causing genes were affected by TALEN, which is a significant concern when using gene targeting in biomedical and clinical applications. Therefore, this technology is suitable for generating a specific KO model, with low probability of unwanted cleavage occurring in the genome. Another very important conclusion from our investigation presented here, is that computer generated predictions need to be tailored carefully to the experimental set up, and even then, predicted aberrations must be checked in the wet lab in order to make sound conclusions about specificity of a method used for targeting. Once the specificity of the mutation is confirmed, the phenotypic and biochemical changes that we observed in our KO mice can only be attributed to the loss of H2A.Lap1. Our assessment is confined to the biological function of H2A.Lap1 expression, predominantly in the testes, at low but detectable levels in the brain and in white blood cells, and will be presented in the following chapter.

References

Andrews, T.D. et al., 2012. Massively parallel sequencing of the mouse exome to accurately identify rare, induced mutations: an immediate source for thousands of new mouse models. *Open biology*, 2(5), p.120061. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3376740&tool=pmcentrez&rendertype=abstract.

Van der Auwera, G.A. et al., 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. 11.10.1-11.10.33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25431634 [Accessed July 19, 2017].

Boel, A. et al., 2016. BATCH-GE: Batch analysis of Next-Generation Sequencing data for genome editing assessment. *Scientific Reports*, 6(1), p.30330. Available at: http://www.nature.com/articles/srep30330 [Accessed April 7, 2017].

Cho, S.W. et al., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, 24(1), pp.132–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24253446 [Accessed December 20, 2017].

Danecek, P. et al., 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15),

pp.2156–2158.
Doench, J.G. et al., 2014. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature Biotechnology*. Available at: http://www.nature.com/nbt/journal/v32/n12/pdf/nbt.3026.pdf [Accessed April 7, 2017].
Guan, P. & Sung, W.-K., 2016. Structural variation detection using next-generation sequencing data: A comparative technical review. *Methods*, 102, p. Available at: http://www.sciencedirect.com/science/article/pii/S1046202316300184.
Heigwer, F., Kerr, G. & Boutros, M., 2014. E-CRISP: fast CRISPR target site identification. *Nature Publishing Group*, 11(2). Available at: http://www.nature.com/nmeth/journal/v11/n2/pdf/nmeth.2812.pdf [Accessed April 7, 2017].
Kanchiswamy, C.N. et al., 2016. Fine-Tuning Next-Generation Genome Editing Tools. *Trends in Biotechnology*, 34(7), pp.562–574. Available at: http://dx.doi.org/10.1016/j.tibtech.2016.03.007.
Kim, H. & Kim, J.-S., 2014. A guide to genome engineering with programmable nucleases. *Nature Publishing Group*, 15. Available at: http://www.nature.com/nrg/journal/v15/n5/pdf/nrg3686.pdf [Accessed April 7, 2017].
Lee, C.M. et al., 2016. Nuclease Target Site Selection for Maximizing On-target Activity and Minimizing Off-target Effects in Genome Editing. *Molecular Therapy*, 24(3), pp.475–487. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26750397%5Cnhttp://www.nature.com/mt/journal/v24/n3/pdf/mt20161a.pdf.
Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.
Liu, M.-J. et al., 2015. Determinants of nucleosome positioning and their influence on plant gene expression. *Genome Research*, 25(8), pp.1182–1195. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26063739 [Accessed March 23, 2017].
Long, C. et al., 2014. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science (New York, N.Y.)*, 345(6201), pp.1184–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4398027&tool=pmcentrez&rendertype=abstract.
Nakajima, K. et al., 2016. Exome sequencing in the knockin mice generated using the CRISPR/Cas system. *Scientific Reports*, 6(October 2015), p.34703. Available at: http://www.nature.com/articles/srep34703.
Puente, X.S. et al., 2015. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*, 526(7574), pp.519–524. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26200345%5Cnhttp://www.nature.com/doifinder/10.1038/nature14666.
Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 8(11), pp.2281–2308.
Sánchez-Rivera, F.J. et al., 2014. Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature*, 516. Available at: http://www.nature.com/nature/journal/v516/n7531/pdf/nature13906.pdf [Accessed April 7, 2017].
Singh, R. et al., 2015. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Research*, 43(18). Available at: https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/nar/43/18/10.1093_nar_gkv575/2/

gkv575.pdf [Accessed April 7, 2017].
Taketo, M. et al., 1991. FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6), pp.2065–2069.
Tattini, L., D'Aurizio, R. & Magi, A., 2015. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Frontiers in bioengineering and biotechnology*.
Tsai, S.Q. et al., 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, 33(2), pp.187–197. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25513782%5Cnhttp://www.nature.com/nbt/journal/v33/n2/pdf/nbt.3117.pdf.
Veres, A. et al., 2014. Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Cell Stem Cell*, 15(1), pp.27–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24996167 [Accessed March 27, 2017].
Wong, K. et al., 2012. Sequencing and characterization of the FVB/NJ mouse genome. *Genome Biology*, 13(8), p.R72. Available at: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-8-r72.
Ye, K. et al., 2009. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, 25(21), pp.2865–2871.
Zischewski, J., Fischer, R. & Bortesi, L., 2016. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology Advances*, 35(1), pp.95–104. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0734975016301586.

CHAPTER 5
CHARACTERISATION OF H2A.LAP1KO MICE

5.1 Introduction

In chapter 4, it was shown that TALEN approach was successful at editing a multi-copy gene family of H2A.Lap1 by a single pair of TALENs. It was also shown that this approach offers a high level of specificity, as the exome sequencing analysis and wet lab validation did not detect any off-target sites of TALEN activity. Chapter 5 is dedicated to the investigation on the effect of the loss of H2A.Lap1 in spermatogenesis of H2A.Lap1$^{-/Y}$ males. Firstly, the fertility of H2A.Lap1$^{-/Y}$ males was assessed. Secondly, the epigenetic landscape of round spermatids, the main germ cell subpopulation that expresses H2A.Lap1, was examined. As H2A.Lap1 is enriched in euchromatin of autosomal chromosomes, as well as within the inactive X chromosome (at genes that escape X chromosome inactivation), we examined these two nuclear locations separately. Thirdly, the overall anatomy of seminiferous tubules of wild type versus H2A.Lap1$^{-/Y}$ mice was compared.

5.2 Results and discussion

5.3 The litter size produced by hemizygous H2A.Lap1^{-Y} males is significantly reduced in comparison to their wild type litter mates.

In order to test the fertility of hemizygous H2A.Lap1$^{-/Y}$ males, 5 hemizygous male mice were paired with 10 wt females aged 3-6 months. As a control, 5 age-matched wt male siblings were also mated at the same time with the wt females aged between 3 and 6 months. Each male was paired continuously with two females for a period of 2-3 months until the statistically sound number of pups for each trio could be obtained. Each litter size was recorded on the morning after the birth. In most cases, each litter was assigned to a particular mother, as most females birthed on different days. Altogether, 39 litters and 259 pups were produced by 5 hemizygous males and 23 litters and 209 pups were produced by the wt male siblings. All males tested were from generation 3 (G3) thus were free from mosaicism/recombination genotype. Our analysis showed that the hemizygous H2A.Lap1$^{-/Y}$ males produced a significantly low number of pups per litter compared to their wt siblings, with hemizygous H2A.Lap1^{-Y} males producing an average litter size of 6.48 pups/litter

and their wt siblings producing 8.46 pups/litter on average (figure 5-1). This reduced litter size indicates a subfertility phenotype for H2A.Lap1^{-Y} males and implied that H2A.Lap1 is required for fertility.

In order to uncover the reasons for H2A.Lap1$^{-/Y}$ subfertility, the epigenetic landscape of haploid round spermatids (RS), the subpopulation of germ cells that displays the highest expression of H2A.Lap1 in wt mice, was examined using immunofluorescent microscopy analysis and the drying-down meiotic spread technique (Peters et al. 1997).

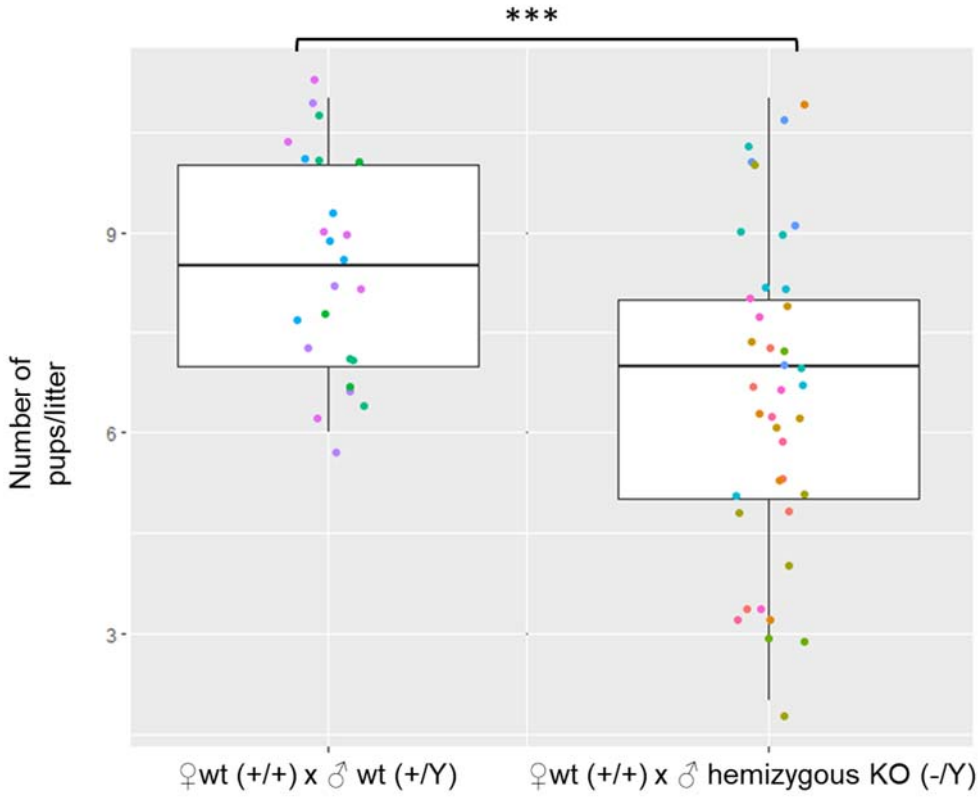


Figure 5-1: The litter size produced by hemizygous H2A.Lap1^{-Y} males is significantly reduced in comparison to wild type litter mates. The scatter plot showed the size of litter produced by breeding pairs of wt female (H2A.Lap1 +/+) with wt male (H2A.Lap1 +/Y) and wt female (H2A.Lap1 +/+) with hemizygous KO male (H2A.Lap1 -/Y). The analysis was conducted on 10 age matched females (3-6 mo) which were bred with age matched wt (n=5) and hemizygous KO (n=5) males. Statistical test, ANOVA, showed that there is a significant reduction in the number of litter produced from the hemizygous KO pair. ***p value=0.0006547 ($p \leq 0.001$).

5.4 The nuclei of round spermatids in H2A.Lap1 KO mice are significantly smaller in hypotonic conditions.

Historically, spermatocyte and spermatid spreading techniques were used to study chromatin state of those cells. The cells are incubated in hypotonic solution containing sucrose, followed by a solution containing detergent and formalin, as a fixative, at pH 9.2. These conditions allow cells to swell up and burst their membrane but also result in enlarged nuclei and overall slight decondensation of chromatin, improving its accessibility for immunolabeling.

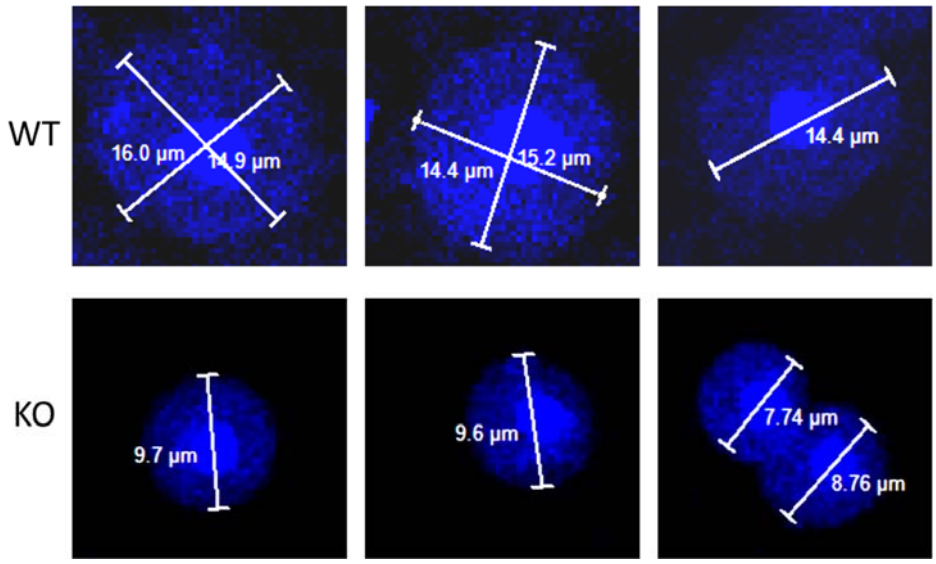
In our experiments, when H2A.Lap1$^{-/Y}$ germ cells were spread on the glass slides for immunofluorescence staining, we consistently noticed that the size of the round spermatid nuclei in the absence of H2A.Lap1 was smaller than the size of wt RS. In order to rule out any discrepancies in the preparation of spreads, wt and H2A.Lap1 $^{-/Y}$ germ cells were spread at the same time, using the same type slides, reagents and incubation conditions. The prepared spreads were stained with DAPI and the nuclei size of round spermatids was measured and compared between wt and KO mice. The analysis of 77 wt RS and 81 KO RS showed that KO RS nuclei display a significant reduction in size (Figure 5-2). Thus, the average size of nuclei in wt mice in hypotonic conditions was 16.01 μm whilst in KO mice it was only 9.5 μm. Microscopy analysis also showed a more uniform shape of nucleus in KO mice, suggesting that the hypotonic conditions did not induce decondensation of chromatin of H2A.Lap1 KO RS to the same extend as in wt mice.

As H2A.Lap1 is known to be associated with decompacted, transcriptionally active euchromatin (EU) and is absent from the highly compacted heterochromatic chromocenter (CC) of RS (which stains strongly with DAP1) (Soboleva et al. 2012), the CC would not be affected by the absence of H2A.Lap1 but the surrounding euchromatin would. Therefore, we measured the fluorescence intensity of the DAPI signal in chromocenter and in the surrounding euchromatin and calculated their ratio (CC-B)/(EU-B), where B is a background (or autofluorescence). We hypothesised that the ratio would decrease in H2A.Lap1 KO RS versus wt mice due to a more compacted nature of euchromatin in the absence of H2A.Lap1 (because the DAP1 staining of euchromatin would be stronger). We also scored early round spermatids (ERS) separately from late round spermatids (LRS), to reflect the observation that H2A.Lap1 has a higher expression in LRS and it also acquires an additional function of re-activating X-chromosome-coded genes in LRS (Soboleva et al. 2012). To

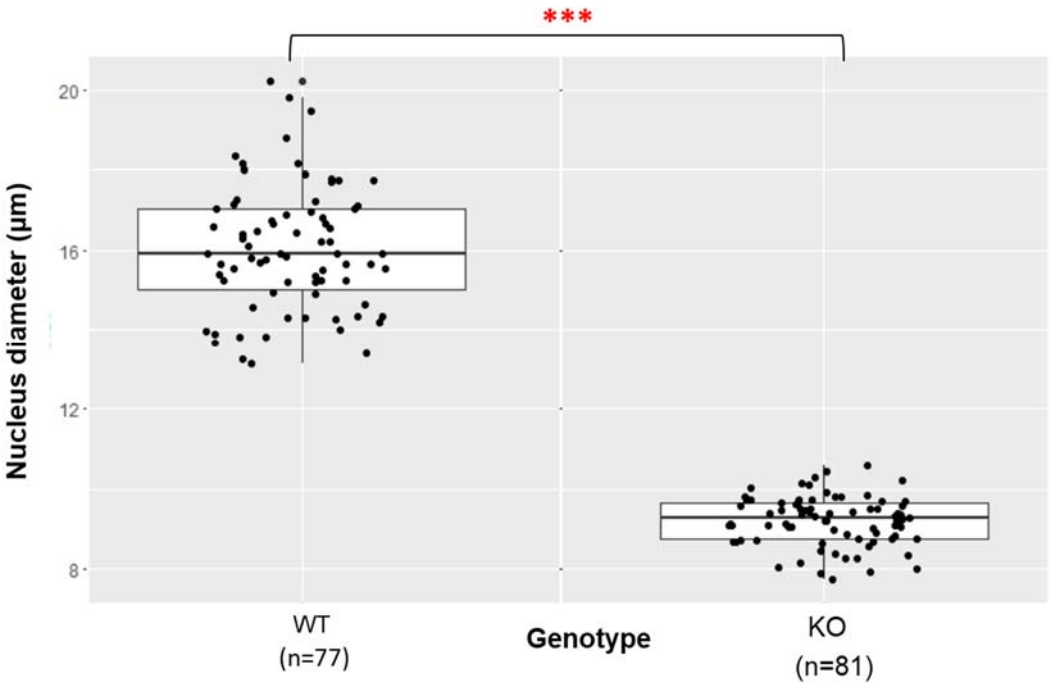
distinguish early and late RS, we used lectin-PNA staining that visualised the growing acrosome in RS.

Indeed, we observed that the average (CC-B)/(EU-B) ratio in wt ERS and LRS was 3.77 and 3.97, respectively, while in the H2A.Lap1$^{-/Y}$ the ratios were significantly smaller, 2.98 and 2.68, for ERS and LRS, respectively (Figure 5-2, C). This represents a 1.4-fold difference indicating that the euchromatin of H2A.Lap1 KO RS is 1.4 times more condensed in hypotonic conditions compared to wt. (Table 5.1). These *in vivo* findings confirm our previous *in vitro* results, that showed that H2A.Lap1 decondenses the chromatin fibre upon its incorporation (Soboleva et al. 2012). In conclusion, we show here, for the first time, that the lack of H2A.Lap1 causes chromatin condensation *in vivo*.

A


WT
16.0 µm
14.9 µm
14.4 µm
15.2 µm
14.4 µm
KO
9.7 µm
9.6 µm
7.74 µm
8.76 µm

B


***
Nucleus diameter (µm)
20
16
12
8
WT
(n=77)
Genotype
KO
(n=81)

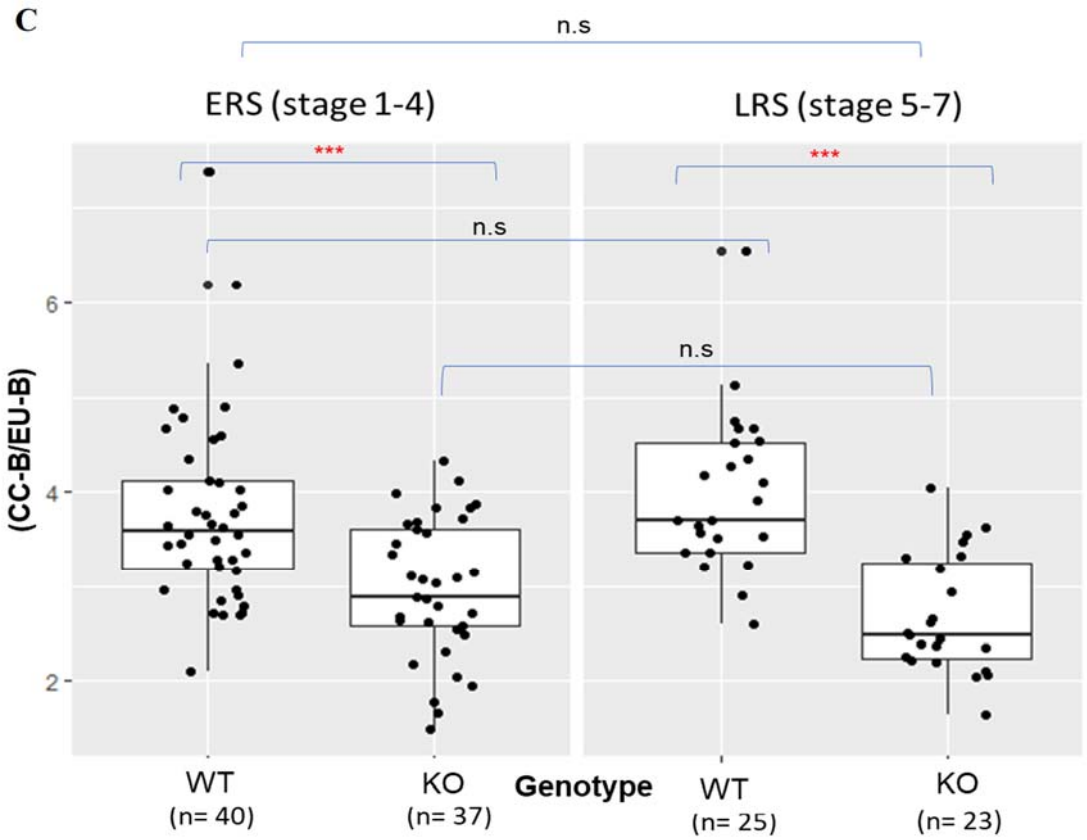


Figure 5-2: The nuclei of round spermatids in H2A.Lap1 KO mice are significantly smaller in hypotonic conditions A) Immunofluorescence images of Round spermatid nuclei stained with 4',6-diamidino-2-phenylindole (DAPI) in wt and hemizygous KO mice. The diameter shown is a direct measurement by LAS-AF software, size is in µm. **B)** Detailed analysis of RS in wt (n=77) and hemizygous KO (n=81) showed a significant difference in the diameter of nucleus between wt and hemizygous KO mice in hypotonic conditions (observation is consistent in 4 independent experiments, wt n=3 and KO n=4, aged between 3 mo and 10 mo). **C)** The ratio of DAPI fluorescence between the chromocentre and euchromatin. This quantification acts as internal control for the subsequent immunofluorescence quantification and controls for the fact that euchromatin in H2A.Lap1 -/Y round spermatids is more condensed. ***, $p \leq 0.001$,

5.5 The active promoter mark, H3K4me3, accumulates in euchromatin of H2A.Lap1 KO round spermatids.

Although we have shown that euchromatin of RSs that lacks H2A.Lap1 does not decompact in hypotonic conditions to the same level as in wt RS, and that the H2A.Lap1$^{-/Y}$ males are subfertile and produce smaller litters, the spermiogenesis and haploid spermatid differentiation still appears to proceed in the absence of H2A.Lap1, which implies that there may be a compensatory mechanism(s) in place. Spermatogenesis is a vital process that ensures the propagation of a specie and thus has evolved to be extremely robust and error-proofed. In addition, the epigenetic machinery is known to be highly dynamic and adaptable to changing conditions. One of the most striking examples of this adaptability is when the histone TH2B variant is knocked-out in mice. In spite of a complete lack of TH2B, the mice were fertile and displayed no deficiencies in spermatogenesis. It appeared that a compensational mechanism evolved in mice lacking TH2B, where nucleosomes containing somatic H2B were decorated with H3K122 and H4K77 crotonylation rendering those nucleosomes unstable mirroring TH2B-containing chromatin (Montellier et al. 2013).

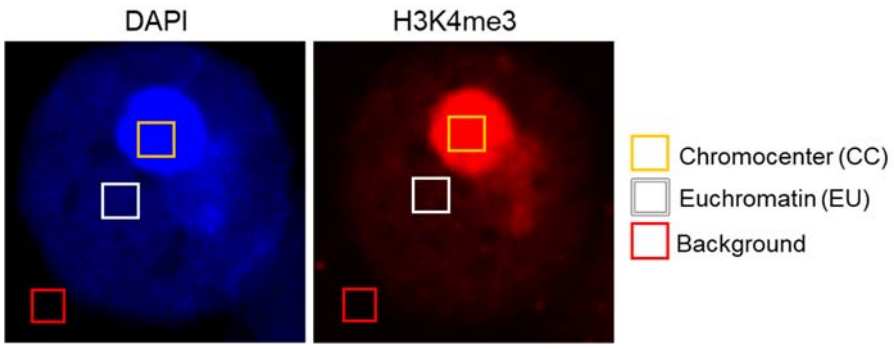
Thus, we wondered whether a similar epigenetic compensatory mechanism may take place in H2A.Lap1 KO mice. In order to address this hypothesis, we stained wt and H2A.Lap1 KO round spermatids with H3K4me3, an epigenetic mark that is associated with active gene promoters similar to H2A.Lap1 and thus may be a good candidate for such a compensatory mechanism.

The staining results were quite striking, showing that in the absence of H2A.Lap1, H3K4me3 accumulates in the euchromatin of RS to a much higher level than in the wt RS, and that this accumulation phenomenon is not an artefact of the lack of decondensation in hypotonic conditions of H2A.Lap1-less euchromatin but a true enrichment. The accumulation of H3K4me3 in euchromatin was compared to the chromocenter staining intensity, and the ratio of (CC-B)/(EU-B), where B is a fluorescence background, was calculated as described in Figure 5-2. The (CC-B)/(EU-B) ratio in ERS was 9.18 and 2.62 in wt and H2A.Lap1KO respectively; and for LRS, 11.69 and 3.24 in wt and H2A.Lap1 KO respectively. Even after taking in account the 1.4-fold more condensed euchromatin in KO round spermatids, this result shows that there was an ~2.6 fold more H3K4me3 in the euchromatin of H2A.Lap1 KO RS than in wt RS. This indicates that the H3K4me3 mark accumulates in the euchromatin in the absence of H2A.Lap1 and may be one of the ways of

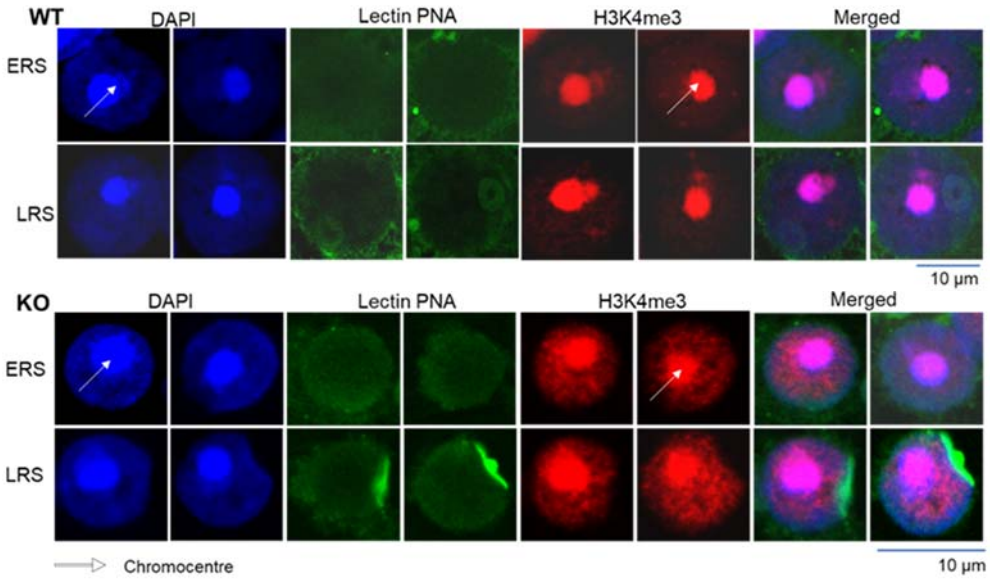
compensating for the lack of H2A.Lap1 at the TSS of active genes (Figure 5-3, Table 5-1). Unexpectedly, unlike H2ALap1, H3K4me3 is also found at the heterochromatic chromocentre. We have set up cross reactivity controls and found that H3K4me3 antibodies were also cross-reacting with H3K9me3. However, staining with H3K9me3 did not show any difference between KO and wt round spermatids (data not shown). We concluded that euchromatin staining produced by anti-H3K4me3 antibodies reflected true H3K4me3 status in euchromatin.

Since H3K4me3 is preceded by H3K4me2, we also assessed the (CC-B)/(EU-B) ratio in order to investigate if the H3K4me2 mark also accumulates in the euchromatin regions of RS in a similar fashion. We found that in ERS, H3K4me2 accumulation is higher in KO spermatids than in wt but this accumulation is less pronounced, which can be attributed simply to the lesser level of decondensation of KO RS in hypotonic preparations (Figure 5-4). Interestingly, this small difference disappears in LRS, probably due to the fact that wt LRS also acquire more H3K4me2 in euchromatin (Figure 5-4). Therefore, we can conclude that H3K4me3 but not H3K4me2 may be involved in a compensatory mechanism to replace H2A.Lap1.

A


DAPI
H3K4me3
Chromocenter (CC)
Euchromatin (EU)
Background

B


WT
DAPI
Lectin PNA
H3K4me3
Merged
ERS
LRS
10 µm
KO
DAPI
Lectin PNA
H3K4me3
Merged
ERS
LRS
Chromocentre
10 µm

C

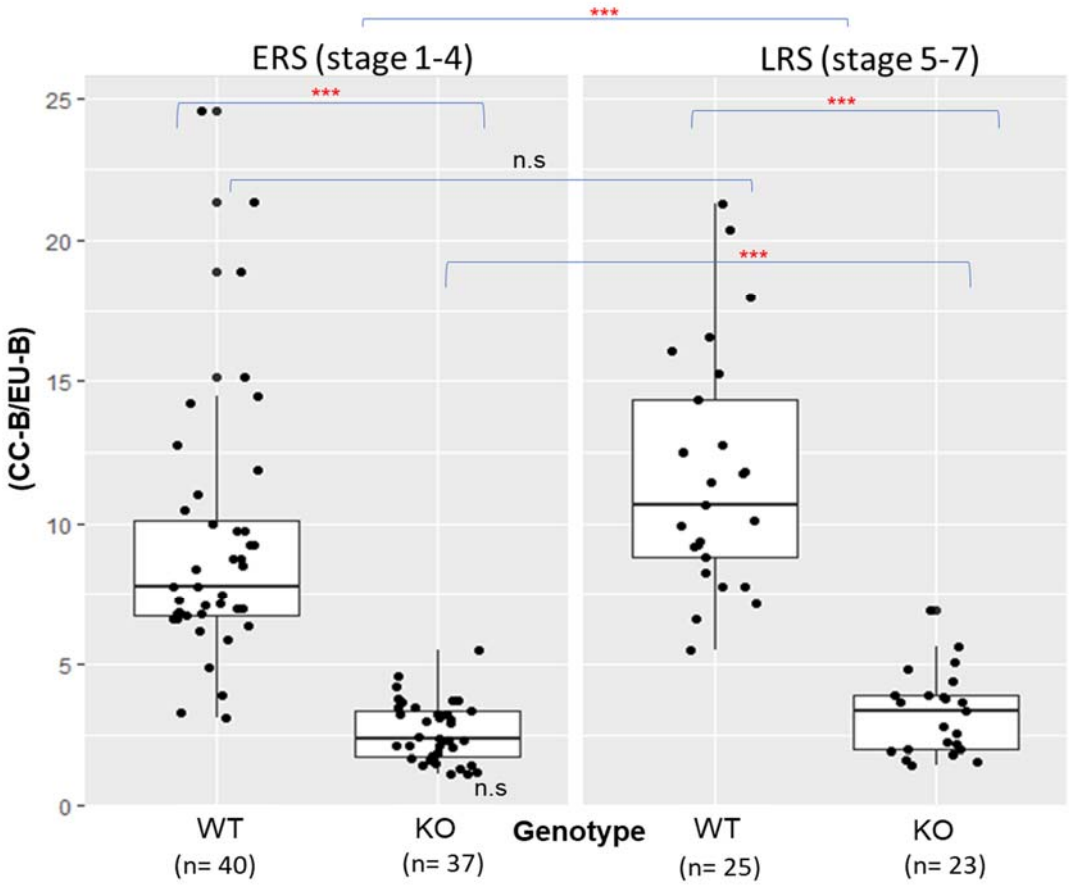
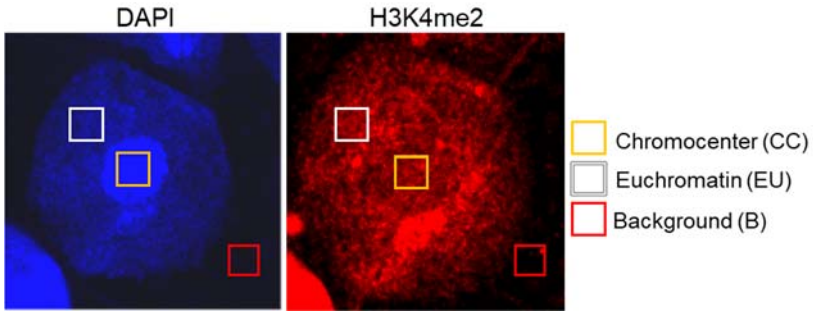


Figure 5-3: **The active promoter mark, H3K4me3, accumulates in euchromatin of H2A.Lap1 KO round spermatids. A)** The quantification strategy used to examine the enrichment of H3K4me3 in euchromatin. In order to compare different slides with each other, fluorescent ratio of (CC-B)/(EU-B) was measured to assess enrichment of H3K4me3 in the euchromatin. **B)** Images of early round spermatids (ERS) and late round spermatids (LRS) from the hypotonic spread of male germ cells. The sample was indirectly stained with active promoter mark, H3K4me3 (red) and Lectin PNA (conjugated with Alexa Fluor 488, green), a marker for acrosome to determine the stages of RS cells. DNA was co-stained with DAPI (blue). Scale bar is 10 µm. **C)** The quantification of H3K4me3 signal based on the ratio of (CC-B)/(EU-B). *** $p \leq 0.001$, ANOVA test. Number of mice used in this analysis wt n=1, KO n=1.

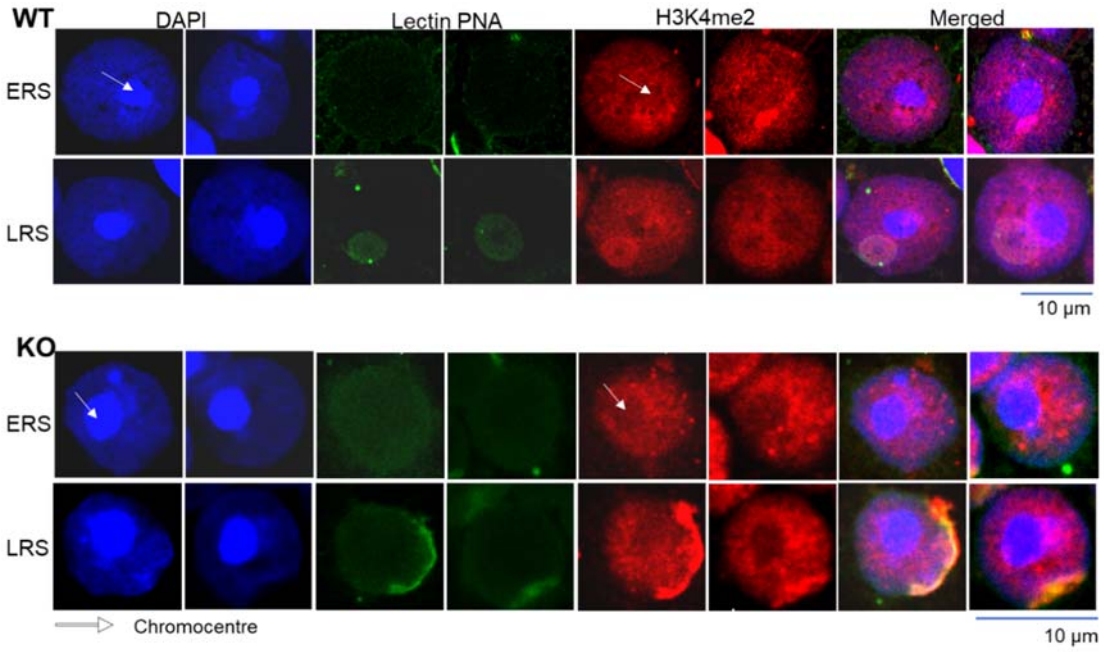
Table 5-1: H3K4me3 enrichment in chromocenter (CC) and euchromatin (EU) in the differentiating round spermatids (RS) of KO mice. Numbers indicate ration intensity CC/EU

Genotype	WT		KO	
Stages of RS	Early round spermatids (ERS) (Stage 1-4)	Late round spermatids (LRS) (Stage 5-7)	Early round spermatids (ERS) (Stage 1-4)	Late round spermatids (LRS) (Stage 5-7)
DAPI	3.77	3.97	2.98	2.68
H3K4me3	9.18	11.69	2.62	3.24

A


DAPI
H3K4me2
Chromocenter (CC)
Euchromatin (EU)
Background (B)

B


WT
DAPI
Lectin PNA
H3K4me2
Merged
ERS
LRS
10 µm
KO
ERS
LRS
Chromocentre
10 µm

C

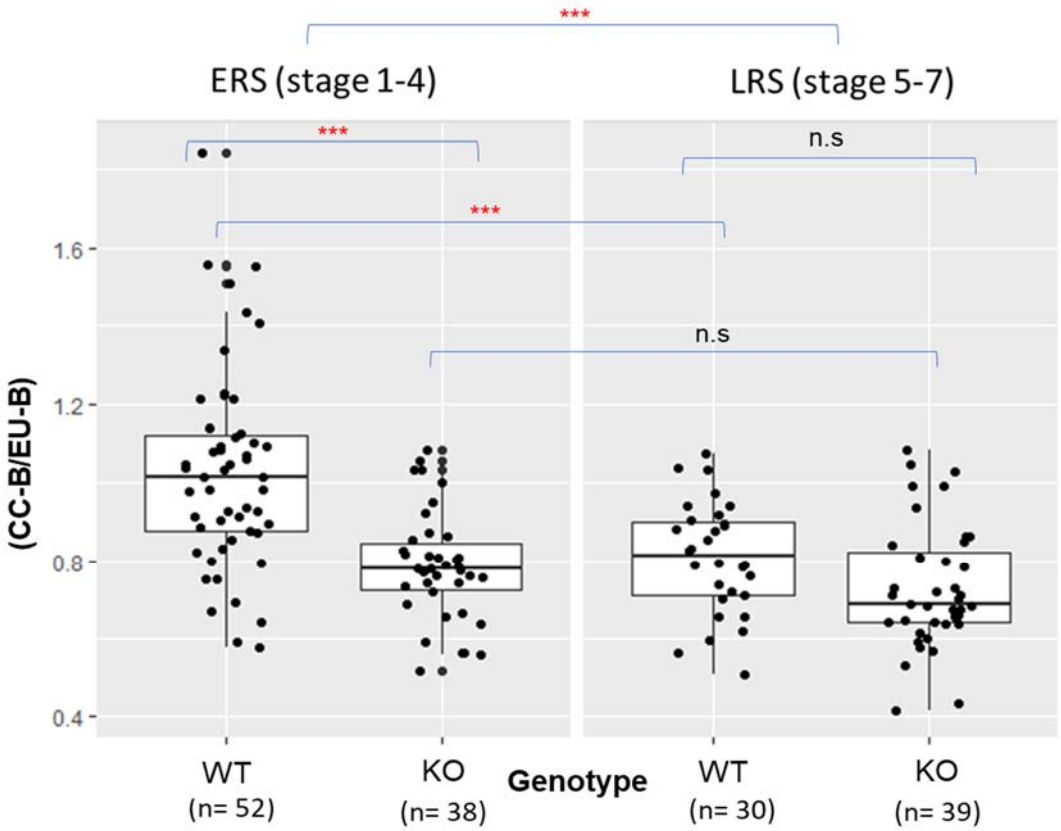


Figure 5-4: H3K4me2 does not accumulate in euchromatin of round spermatids in KO mice. A) The quantification strategy used in examining the enrichment of H3K4me2 in euchromatin. **B)** Images of early round spermatids (ERS) and late round spermatids (LRS) from the hypotonic spread of male germ cells. The sample was indirectly stained with active promoter mark, H3K4me2 (red) and Lectin PNA (conjugated with Alexa Fluor 488, green), a marker for acrosome to determine the stages of RS cells. DNA was co-stained with DAPI (blue). Scale bar is 10 µm**. C)** The quantification of H3K4me2 intensity in chromocenter and euchromatin determined as described in previous figures. ANOVA test ***$p<0.001$. Number of mice used in this analysis wt n=1, KO n=1.

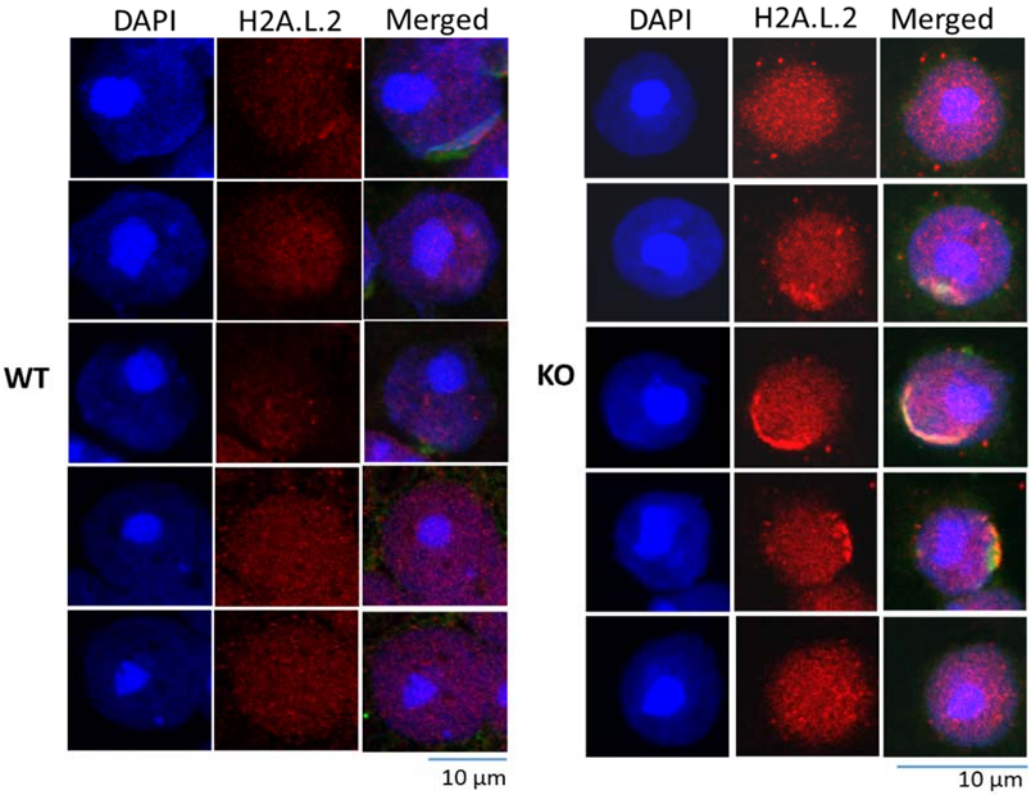
5.6 H2A.L.2 histone variant displays an unusual behaviour in round and condensing spermatids in the absence of H2A.Lap1.

Another interesting possibility is that histone variants, that belong to the same family as H2A.Lap1, may be involved in the compensation mechanism when H2A.Lap1 is absent. As have been reported previously and described in detail in the introduction chapter, H2A.Lap1 belongs to the family of testis-specific histone variants that comprise of H2A.L.1, H2A.L.2, H2A.L.3 and H2A.Lap1 (Govin et al. 2007; Soboleva et al. 2012). All these histone variants contain a short acidic patch and upon incorporation into the chromatin, relax chromatin fibre and destabilise the nucleosomes *in vitro* (Soboleva et al 2012). Extensive work has been done on deciphering the function of H2A.L.2 variant by the Khochbin laboratory (Barral et al. 2017; Govin et al. 2007) in which it was shown that H2A.L.2 histone variant is expressed by condensing spermatids (at the time when H2A.Lap1 has already ceased from the nucleus) and is associated with pericentric heterochromatin. H2A.L.2 is actively involved in histone-protamine exchange by destabilising the nucleosomes allowing for transitional proteins to associate with the DNA (Barral et al. 2017).

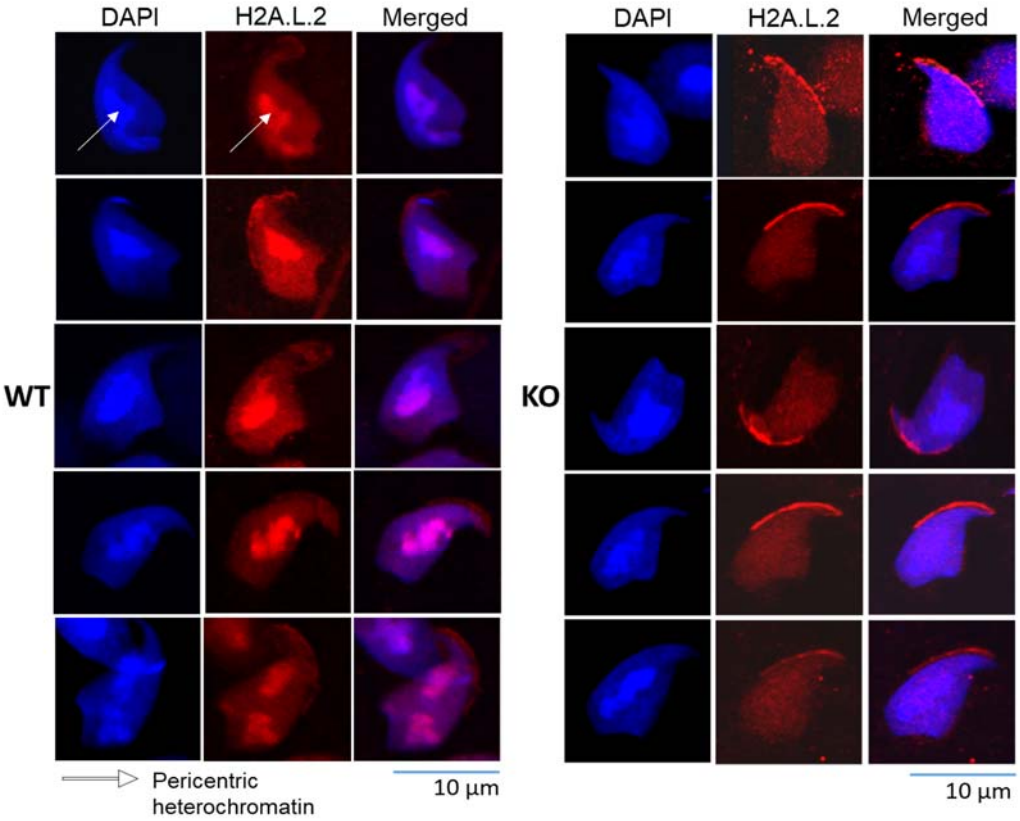
We entertained a possibility, that in the absence of H2A.Lap1, H2A.L.2 may change its expression timing and be expressed earlier, in order to compensate for the lack of H2A.Lap1. Our immunofluorescent staining with anti-H2A.L.2 serum (a gift from S. Khochbin) have shown that in the wt round spermatids, the signal is very weak and the laser power and the gain needs to be raised in order to detect any signal (Laser at 30% (normally 20%), gain at 880nm (normally 400-600nm). However, in the KO round spermatids, when the identical settings were used, the signal seemed to be stronger (Figure 5-5 A). Although the signal was quite weak in both cases, it was definitely more obvious and looked more specific in KO RS. In addition, we found a very interesting change in H2A.L.2 localisation in the condensing spermatid nuclei in the absence of H2A.Lap1. Thus, the wt condensing spermatid staining was identical to the previously reported, where H2A.L.2 was enriched in pericentric heterochromatin (Govin et al. 2007), while in the KO condensing spermatids, we could not see H2A.L.2 enrichment in the pericentric heterochromatin in the majority of condensing spermatids. Instead, the H2A.L.2 signal was uniform and weak throughout the condensing spermatid nuclei (Figure 5-5 B). This finding implied that in the absence of H2A.Lap1, H2A.L.2 is mislocalised and that the histone-protamine exchange may

not be as efficient in the H2A.Lap1–deficient mice, which may contribute to the subfertility phenotype observed in the H2A.Lap1$^{-/Y}$ male mice.

A Round spermatids



B Condensing spermatid



C

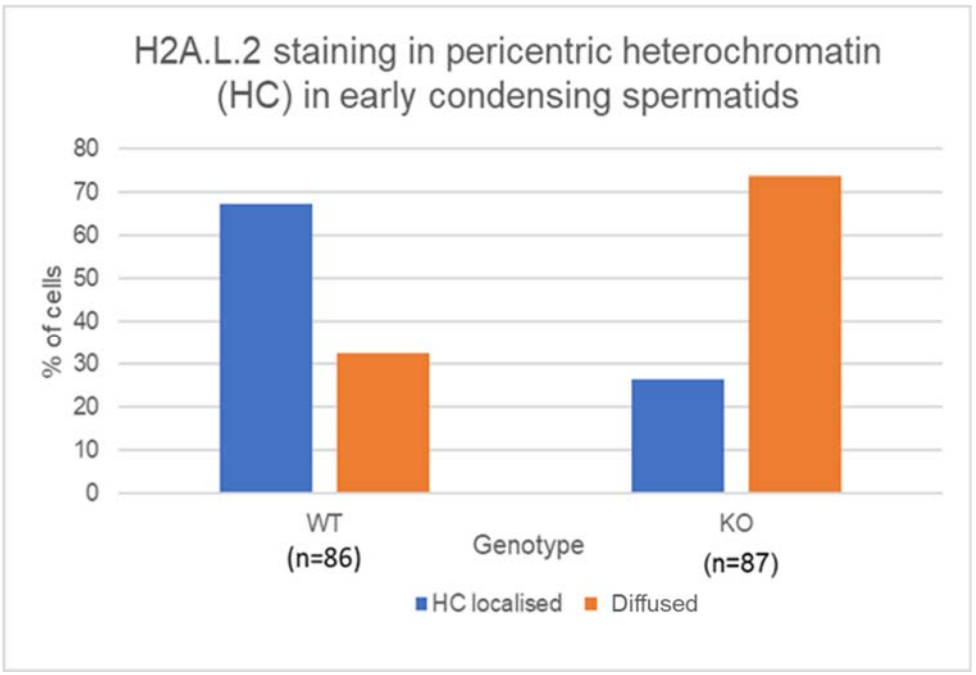


Figure 5-5: H2A.L.2 localisation in the nuclei of round and condensing spermatid in H2A.Lap1$^{-/Y}$ mice is different from the wild type. A) Images of the round spermatid cells from the hypotonic spread of germ cells were indirectly immunostained with serum raised against H2A.L.2. Under the same laser/gain settings, H2A.L.2 signal is stronger in the hemizygous KO sample than in wt. **B) HA.L.2 localisation in condensing spermatid cells.** White arrows indicate pericentric heterochromatin. **C)** The histogram depicting the number of cells with H2A.L.2 enriched in pericentric heterochromatin region and the number of cells with diffused signal of H2A.L.2. D) qPCR on other testis- specific histone variant on 18 do, 24 do and 30 do testes in wt and KO mice. Number of mice used in this analysis wt n=1, KO n=1.

5.7 Localisation of RNA polymerase II within splicing speckles is affected in H2A.Lap1$^{-/Y}$ round spermatids.

Previously, it was shown in our laboratory, that H2A.Lap1 directly interacts with active form of RNA Polymerase II (both initiating and elongating forms of RNA Pol II with CTD S2P and S5P) and that RNA Pol II and H2A.Lap1 reside within splicing speckles, the transcriptional and splicing hubs of round spermatid nuclei (Soboleva et al 2017). We have performed immunofluorescent staining of round spermatids from wt and H2A.Lap1 KO mice using RNA Pol II S5P antibodies. As it was expected, the staining in the wt round spermatids produced a signal in the euchromatin with very bright foci corresponding to the splicing speckles that surround the chromocenter (Figure 5-6, wt panel). Surprisingly, RNA Pol II S5P distribution in more than 70% of KO round spermatids was quite different: the pattern was a lot more diffused and the signal in the splicing speckles was no longer shaped into well-defined foci (Figure 5-6, KO panel).

Intrigued by this observation, we investigated whether the splicing speckle morphology is affected by the absence of H2A.Lap1, so we used splicing speckle marker, Y12, to identify splicing speckles in the KO and wt rounds spermatids. The staining, shown in Figure 5-7, did not reveal any morphological changes in the splicing speckles appearance. The Y12 signal in KO and wt round spermatids was observed as bright well-defined foci and was perfectly colocalised with the DAPI depleted regions of the nuclei, as expected (white arrows, Figure 5-7). This finding implied that H2A.Lap1 plays an important part in the anchoring of the RNA Pol II at the splicing speckle region and in the absence of H2A.Lap1, even in the presence of upregulated H3K4me3, RNA Pol II is not efficiently recruited or retained at these important transcriptional and splicing hubs of the round spermatids. In other words, this finding uncovers a novel function for H2A.Lap1 in the splicing speckles of round spermatids: it seems that direct interaction between H2A.Lap1 and active form of Pol II may serve as an anchor for Pol II, or a detailed postal address, as without H2ALap1, RNA Pol II can only find the approximate region within and around the splicing speckles but not the exact location.

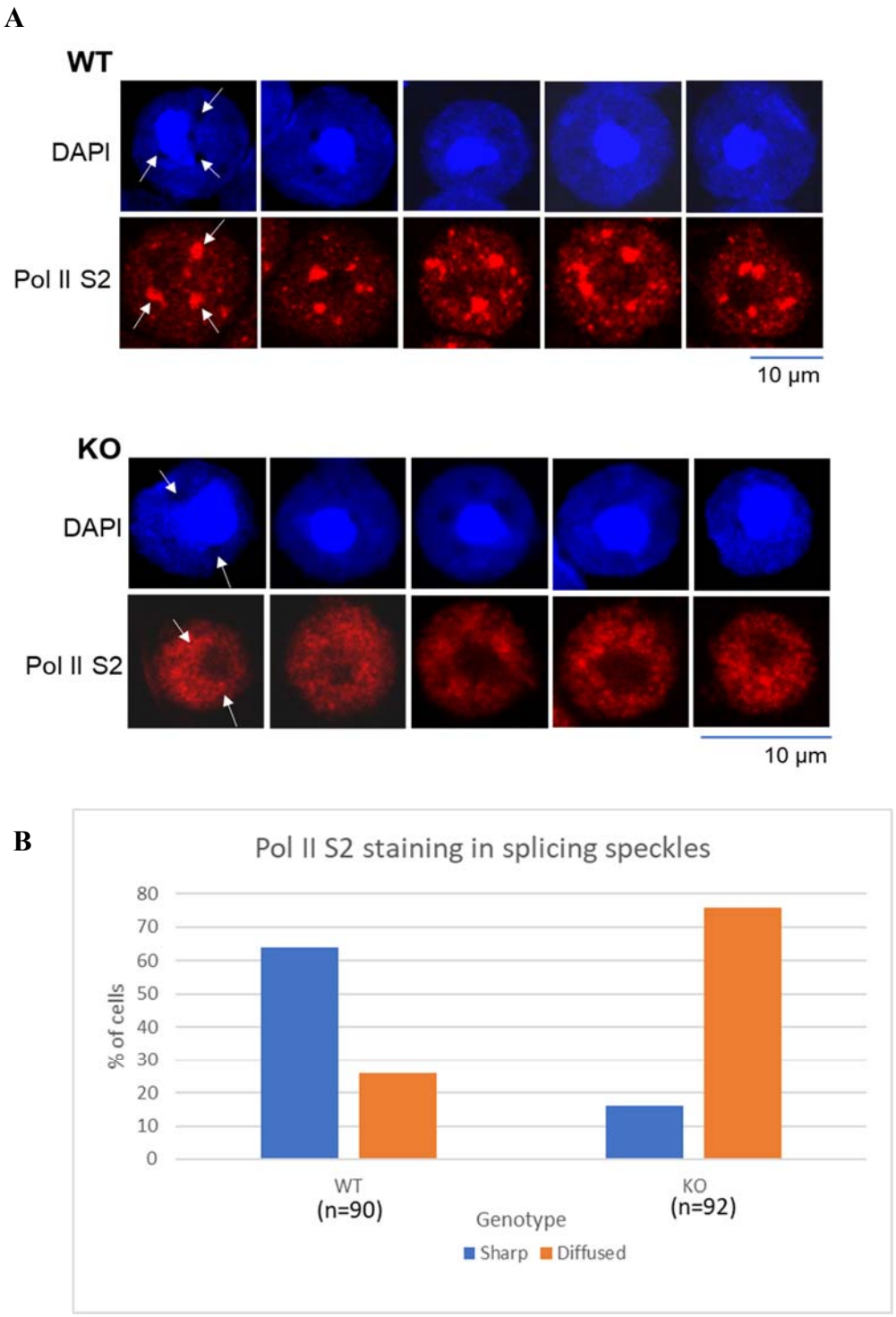


Figure 5-6: Localisation of the active form of RNA polymerase II within splicing speckles is affected in H2A.Lap1$^{-/Y}$ round spermatids. A) Immunofluorescent staining of hypotonic spreads of male germ cells from wt and hemizygous KO mice testis showing round spermatids. Cells were indirectly labelled with anti- Pol II S2P antibodies and co-stained with DAPI to visualise DNA. Scale bar is 10 µm. White arrows show accumulation of Pol II S2P signal in splicing speckles of WT round spermatid cells. The signal is diffused in hemizygous KO sample. **B)** The histogram plot showing that majority of RS cells in hemizygous KO mice have diffused Pol II S2 localisation in the euchromatin compartment of round spermatids. Number of mice used in this analysis wt n=1, KO n=1.

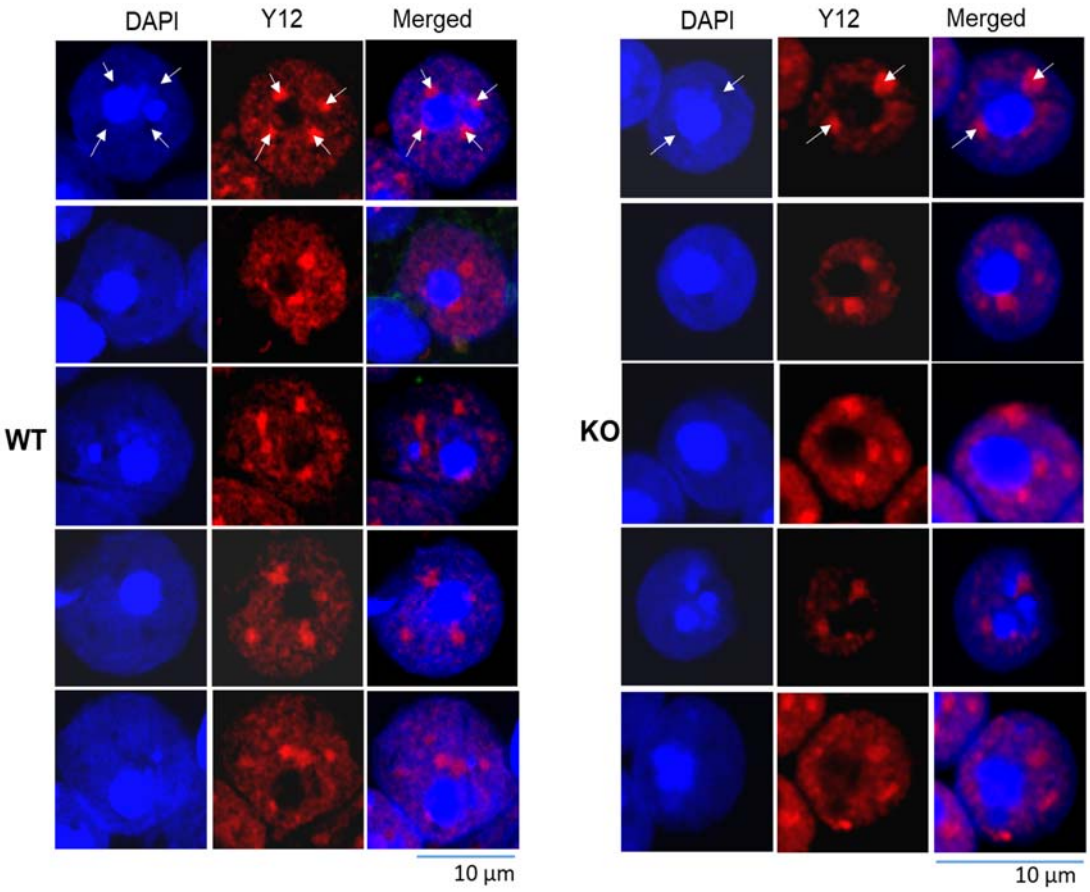
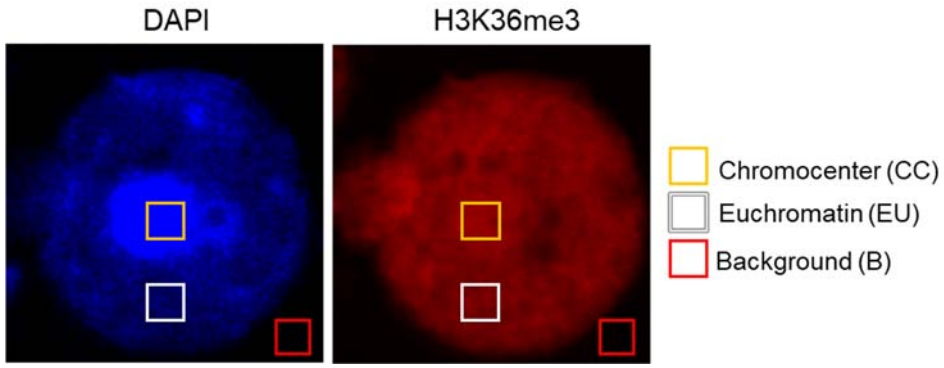


Figure 5-7: Splicing speckle morphology is not affected by the absence of H2A.Lap1. The hypotonic spreads of germ cells from wt and hemizygous KO adult mice testis are shown. Cells were indirectly immunostained with the antibody against splicing speckles marker, Y12 and counterstained with DNA-binding dye, DAPI. Scale bar is 10 µm. White arrows show accumulation of Y12 signal in DAPI depleted regions. The staining showed no difference in Y12 localisation in hemizygous H2A.Lap1 $^{-/Y}$ samples. Number of mice used in this analysis wt n=1, KO n=1.

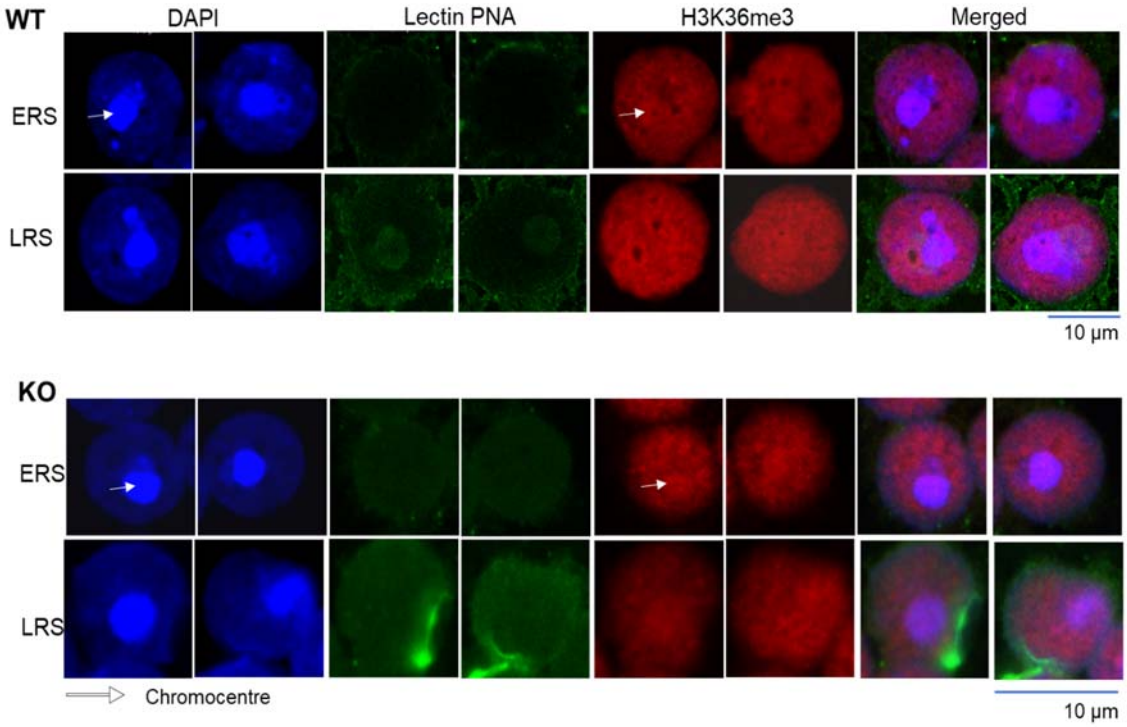
5.8 H3K36me3 enrichment in autosomal euchromatin of round spermatids is affected by the absence of H2A.Lap1.

Actively transcribed genes have well-defined epigenetic landscape that decorates the different regions of the gene. Thus, promoter regions and TSSs are decorated by H3K4me3 (and H2A.Lap1), while the body of the active gene is decorated with H3K36me3. We assessed early and late round spermatids for the H3K36me3 to see if we can uncover any differences in KO vs wt RS. Interestingly, we found that H3K36me3 is significantly depleted in the euchromatin of ERS of KO mice (Figure 5-8). This trend disappears in the LRS, as both wt and KO mice display less H3K36me3 in the euchromatin, probably reflecting the overall lower transcriptional activity of LRS in comparison to ERS. However, the fact that H3K36me3 intensity is lower in KO ERS may indicate that the overall transcription efficiency may be lower in KO round spermatid and represents an important finding that requires further investigation.

A


DAPI
H3K36me3
Chromocenter (CC)
Euchromatin (EU)
Background (B)

B


WT
DAPI
Lectin PNA
H3K36me3
Merged
ERS
LRS
10 µm
KO
ERS
LRS
Chromocentre
10 µm

C

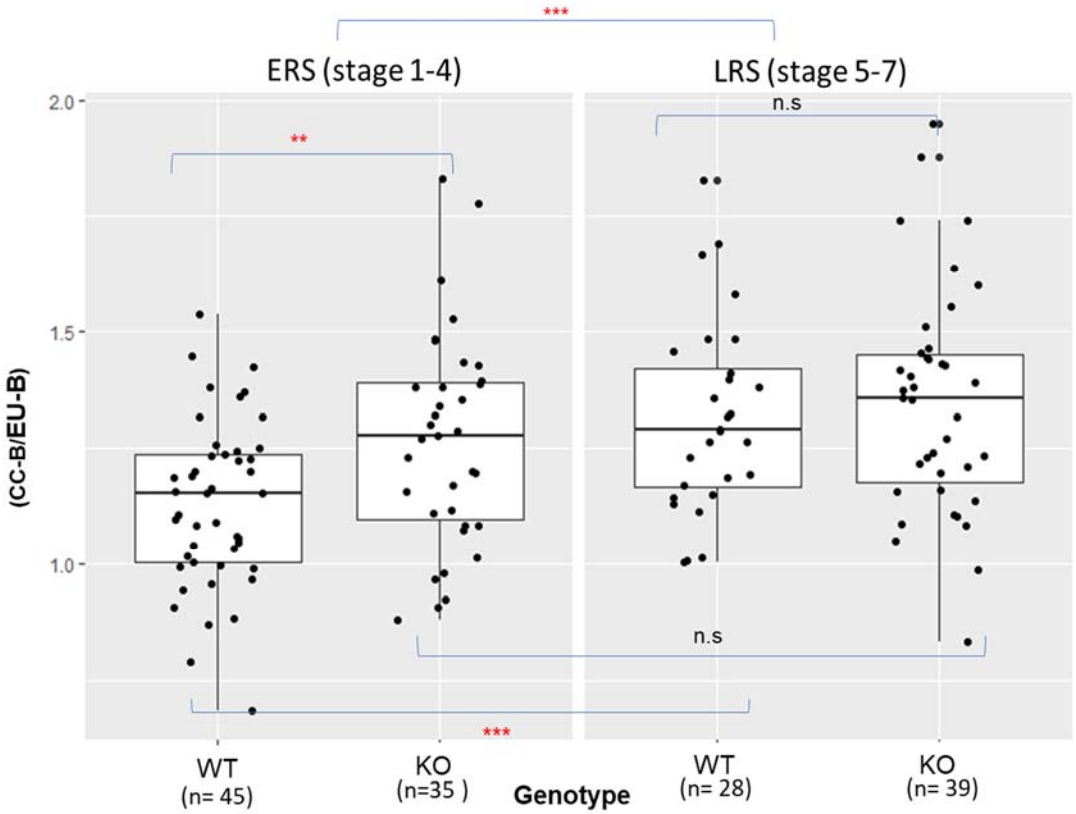


Figure 5-8: H3K36me3 accumulation in euchromatin of hemizygous KO mice reduced in early round spermatid. A) The quantification strategy used in examining the enrichment of H3K36me3 in euchromatin. **B)** Images of early round spermatids (ERS) and late round spermatids (LRS) from the hypotonic spread of male germ cells. The sample was indirectly stained with active promoter mark, H3K36me3 (red) and Lectin PNA (conjugated with Alexa Fluor 488, green), a marker for acrosome to determine the stages of RS cells. DNA was co-stained with DAPI (blue). Scale bar is 10 µm. **C)** The quantification of H3K36me3 intensity in chromocenter and euchromatin was determined as a ratio of (CC-B)/(EU-B). ANOVA test, ** $p<0.01$, ***$p<0.001$. Number of mice used in this analysis wt n=1, KO n=1.

5.9 Effect of H2A.Lap1 absence on overall gene transcription and splicing efficiency.

Inability of active form of RNA Pol II to anchor within slicing speckles and the observation that H3K36me3 mark may be depleted within euchromatin of H2A.Lap1 KO round spermatids raised a possibility that the transcription of highly expressed genes, that usually contain H2A.Lap1 within their TSS, may be affected in the absence of H2A.Lap1. To test that, 17 genes, whose expression is restricted to RS stage and whose TSS contains H2A.Lap1-associated nucleosome were chosen for qPCR. The level of expression of those genes was compared between wt and H2A.Lap1 of 30-day old testis. The 30-day old testis were chosen because we have shown previously (Soboleva et al. 2012). So, that at this stage of seminiferous tubule development, round spermatids represent the majority of cell population (up to 80% of tubules contain round spermatids, stage 1-8), with no elongating-condensing spermatids. The results showed that the expression of 17 genes was significantly affected in the KO sample (Figure 5-9).

We also subjected 30-day old testis from a different group of 3 wt and 3 H2A.Lap1$^{-/Y}$ mice to Next generation RNA-sequencing experiment to assess whether overall gene expression is affected by the lack of H2A.Lap1. Unfortunately, we didn't see any difference between the wt and KO mice (apart from down-regulation of H2A.Lap1 coding genes and one other gene, Abr), partly due to the fact that we did not use spike-in control, which is necessary to compare global changes in gene expression, partly due to the possibility that in this particular KO trio, all samples may have had a very high degree of compensation, so that RNA-seq experiment needs to be repeated using spike-in control and larger number of biological replicates per group.

However, the same RNA-seq data was used to analyse whether RNA splicing was affected in the H2A.Lap1 KO testis. It was shown previously (Soboleva et al. 2017), that H2A. Lap1-containing nucleosomes are present at the intron/exon boundaries where H2A.Lap1 directly interacts with the spliceosome subunits and with the transcribed RNA, implying a function for H2A.Lap1 in regulation of preRNA splicing. Thus, we analysed the level of intron retention (IR) in the wt and KO samples. The program IRFinder was used (Middleton et al. 2017) with default settings to search for retained introns in each individual samples. The results were pooled for each group (wt n=3, KO n=3) and differential IR was analysed as described by the authors, using

scripts provided with IRFinder. After the initial analysis, a total of 1289 IR events were found, which included many events of low significance (p-value > 0.1), low or non-uninform coverage of introns (A/B-IntronDepth, A/B-IntronCover), or events which were associated with annotated/known minor isoforms of transcripts. These events of IR were systematically excluded, thus resulting in a list of 131 retained introns.

We observed that differential IR upon KO of H2A.Lap1 is characterised by an even split between increased and decreased IR levels (66 increased/65 decreased). However, we also observed that there is a shift in the levels of intron-retention ratios to such an extent that mean IR ratios of introns with increased retention in the KO mice (IR ratio 0.13) are higher than the mean IR ratios with increased retention in WT mice (IR ratio 0.9) (Figure 5-10). The results support the conclusion that there is higher IR in the KO mice. Interestingly, the data also shows that there is less variability of IR in the absence of H2A.Lap1 but further experiments are required to confirm the observed trend.

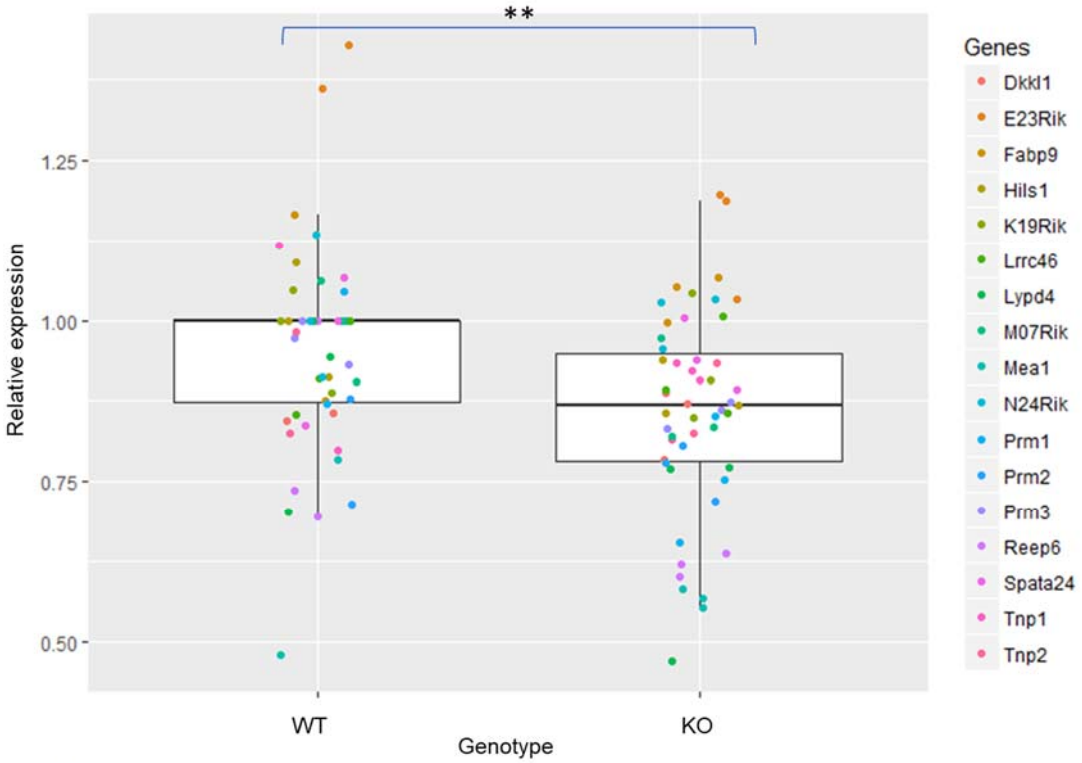


Figure 5-9: Expression level of autosomal round spermatid-specific genes (17 genes) that normally contain H2A.Lap1 nucleosome in their TSS in wt and H2A.Lap1 KO mice testis. The expression level was determined using qPCR, first normalised to β-actin and then to one of the WT mice to obtain the expression fold difference. WT (n=3) and H2A.Lap1 KO (n=3), ANOVA test, **p<0.01.

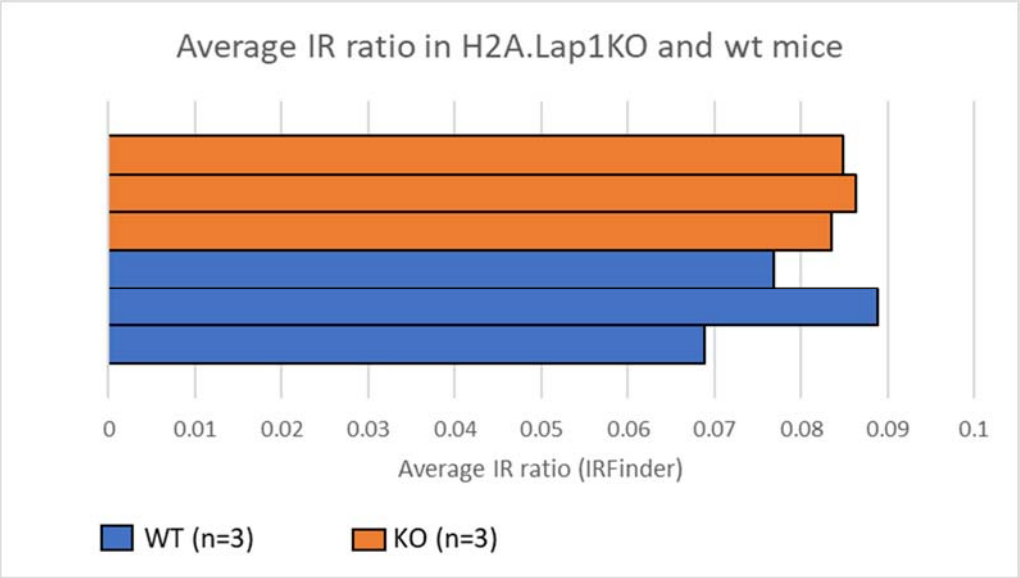


Figure 5-10: H2A.Lap1 KO mice showed higher intron retention. The IR ratio was determined by using IRfinder program (Middleton et al. 2017). The average IR ratio for each mouse is from an average ratio from 131 retained introns.

5.10 H2A.Lap1 depletion affects the epigenetic landscape of post meiotic sex chromatin (PMSC) and expression of X-chromosome coded genes that escape inactivation.

One of the interesting features of H2A.Lap1 was its incorporation into the chromatin of inactivated sex chromosomes in the late round spermatids (stages 5-8), and its involvement in reactivation of post-meiotically silent sex-chromatin (PMSC) associated genes (Soboleva et al. 2012). Therefore, it was very interesting to assess epigenetic landscape of the PMSC in round spermatids that lack H2A.Lap1 and the level of expression of the X-chromosome-coded genes that normally contain H2A.Lap1 nucleosomes at their TSS.

There is a number of epigenetic marks that are known to be associated with the PMSC and appear to be enriched at that site, when microscopic immunofluorescent analysis is used. It has been shown by our laboratory and others that, apart from H2A.Lap1, histone variant H2A.Z, as well as multiple post-translational histone modifications, such as H3K9me2, H4K8ac, H4K12ac, H3K4me2 and 3 and Lysine crotonylation (Kcr) decorate silenced sex chromosomes (Greaves et al. 2006; Namekawa et al. 2006; van der Heijden et al. 2007; Montellier et al. 2012).

As in immunofluorescent staining above, enrichment at PMSC was compared to round spermatid chromocenter (CC) and expressed as a ratio of PMCS/CC. Also, early and late round spermatids were examined separately to reveal a possible dynamic of those marks during rounds spermatid differentiation, which was not done by previously published works.

Thus, staining with anti-Lysine crotonylation antibody (Kcr) revealed that in the early round spermatids, Kcr of PMSC is significantly depleted in the absence of H2A.Lap1. Interestingly, it was also observed that the PMSC decoration by Kcr becomes generally depleted in LRS in both, wt and KO round spermatids. That trend has not been observed before (Figure 5-11). This finding indicates that Krc is not upregulated at the PMSC to compensate for H2A.Lap1 absence, even though it was shown to be present at the TSS of X-chromosome coded genes that escape inactivation during PMSC silencing (Montellier et al. 2012). Moreover, it seems to be affected by the absence of H2A.Lap1 in early round spermatids (at the time when H2A.Lap1 is not present on PMSC) for reasons that are unclear.

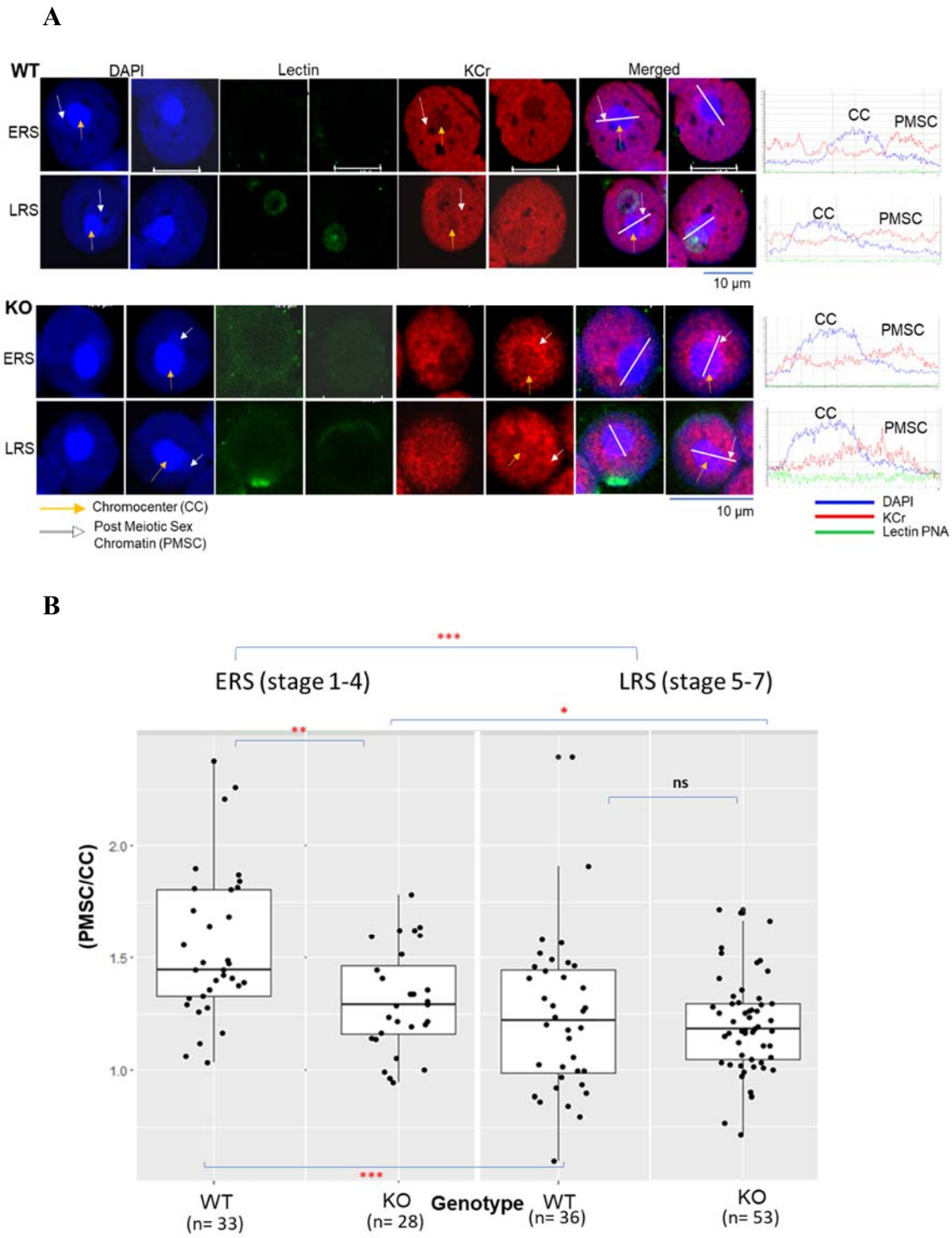


Figure 5-11: Kcr intensity is reduced in PMSC of hemizygous KO round spermatids. A) Images of hypotonic spreads of early round spermatids (ERS) and late round spermatids (LRS). The sample was indirectly stained with pan anti-Kcr antibody and Lectin PNA (conjugated with Alexa Fluor 488), to determine the stages of RS cells. DAPI was used to counterstain nuclei. White lines in merged images indicate the path used to determine fluorescence intensity across the PMSC and chromocenter. Scale bar is 10 µm. **B)** The quantification of KCr intensity as a ratio of PMSC to chromocenter fluorescence. ANOVA test, * $p \leq 0.05$, ** $p \leq 0.01$, ***$p \leq 0.001$. Number of mice used in this analysis wt n=1, KO n=1.

Another interesting finding was that a chromatin mark, H3K4me2, was depleted at the PMSC in the LRS in the absence of H2A.Lap1 (Figure 5-12) suggesting that it is also not involved in the compensation for the loss of H2A.Lap1 at the TSS of X-chromosome coded genes that become reactivated in late round spermatids.

Other active mark, H4K8ac, and a repressive mark, H3K9me2 were also assessed. H3K9me2 was of interest because it was hypothesised that in the absence of decondensing function of H2A.Lap1, PMSC may display a more condensed behaviour. However, no difference between wt and KO round spermatids were observed for both marks (Figures 5-13 and 5-14).

The lack of compensation in PMSC, at least at the level of 4 marks examined, prompted us to check transcription levels of the X-chromosome coded genes that are known to become re-activated at the time, when H2A.Lap1-nucleosome is loaded at their TSS in the late RS (Soboleva et al. 2012). Six genes were accessed in 30 do testis by quantitative real time PCR and their relative expression was compared between the wt and H2ALap1 KO. Interestingly, all 6 genes were downregulated with 5 out of 6 genes showing consistent downregulation by approximately 20% in all H2A.Lap1 $^{-/Y}$ mice that were examined (Figure 5-15). This important finding implies that H2A.Lap1 is important factor that ensures high level of expression of X-linked genes and that in the absence of H2A.Lap1, expression of these genes becomes less efficient.

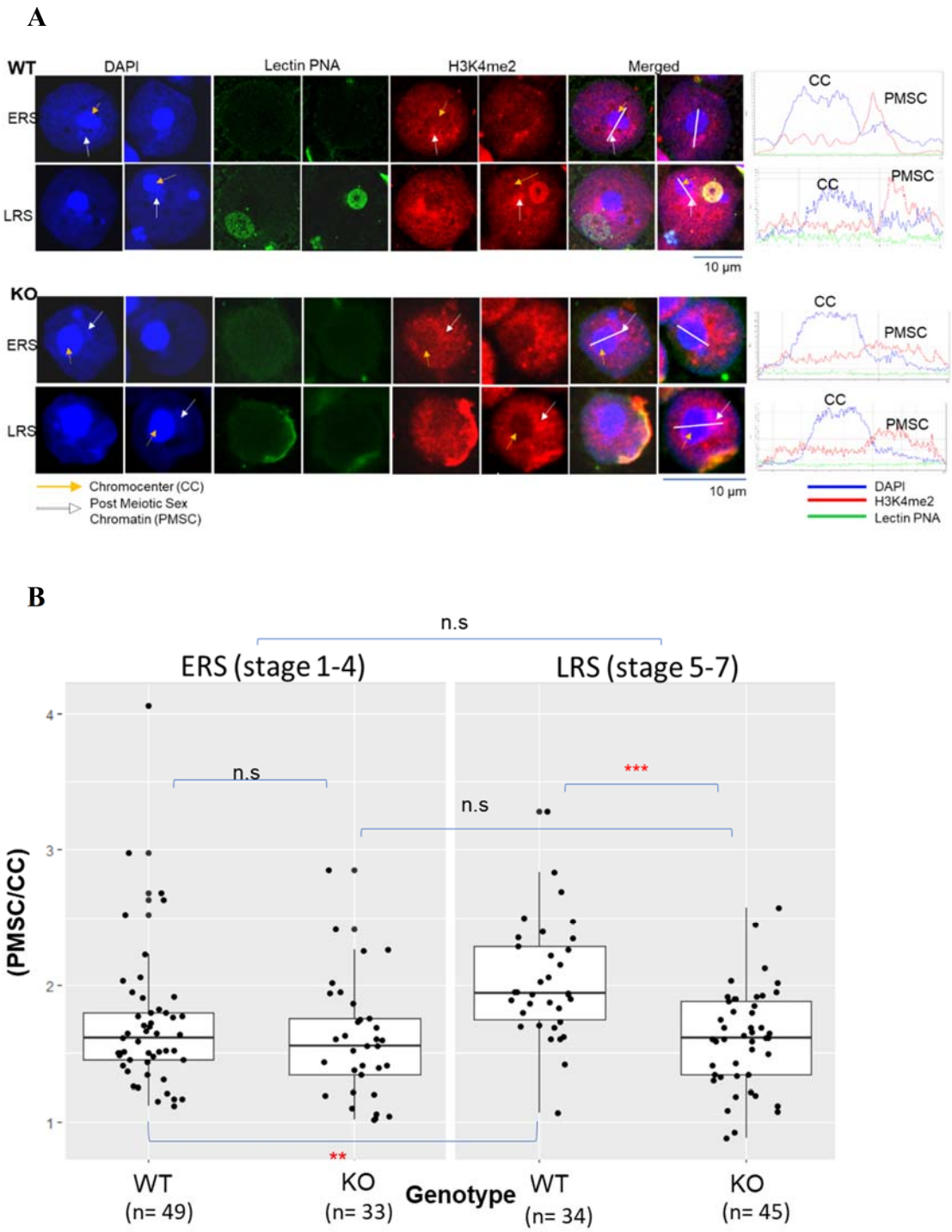


Figure 5-12: H3K4me2 was depleted at the PMSC in the late RS in the absence of H2A.Lap1 A) Hypotonic spreads of early round spermatids (ERS) and late round spermatids (LRS) from wt and H2A.Lap1 KO testis. The sample was indirectly stained with anti-H3K4me2 (red signal), Lectin PNA (conjugated with Alexa Fluor 488) and DAPI (blue), to visualise nuclei. White lines in merged image indicate the path used to determine fluorescence intensity across the post meiotic sex chromatin and chromocenter. Scale bar is 10 µm**. B)** The quantification of H3K4me2 intensity in post meiotic sex chromatin shown as a ratio of PMSC to CC. ANOVA test, **, $p<0.01$, *** $p<0.001$. Number of mice used in this analysis wt n=1, KO n=1.

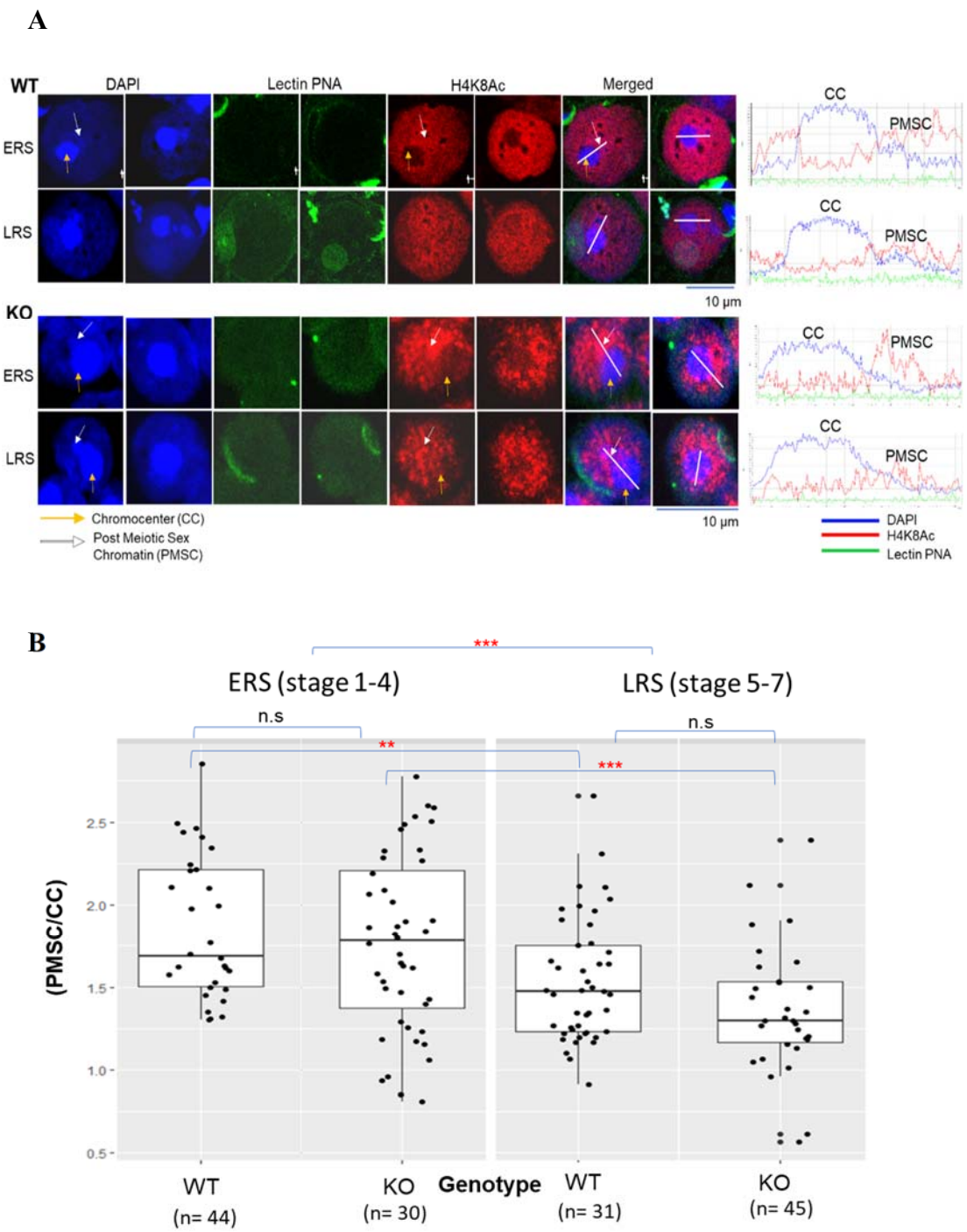


Figure 5-13: H4K8Ac displays no difference between wt and KO round spermatids but is reduced in PMSC of late round spermatids in both wt and hemizygous KO mice. A) Images of early round spermatids (ERS) and late round spermatids (LRS) of hypotonic spreads of male germ cells. The samples were indirectly stained with anti- H4K8Ac antibody (red), and Lectin PNA (conjugated with Alexa Fluor 488), a marker for acrosome to determine the stages of RS cells. Nuclei were counterstained with DAPI. White lines in merged image indicate the path used to determine fluorescence intensity across the post meiotic sex chromatin and chromocenter. Scale bar is 10 µm. B) The quantification of H4K8Ac intensity in post meiotic sex chromatin. ANOVA test, ** $p \leq 0.01$, ***$p \leq 0.001$. Number of mice used in this analysis wt n=1, KO n=1.

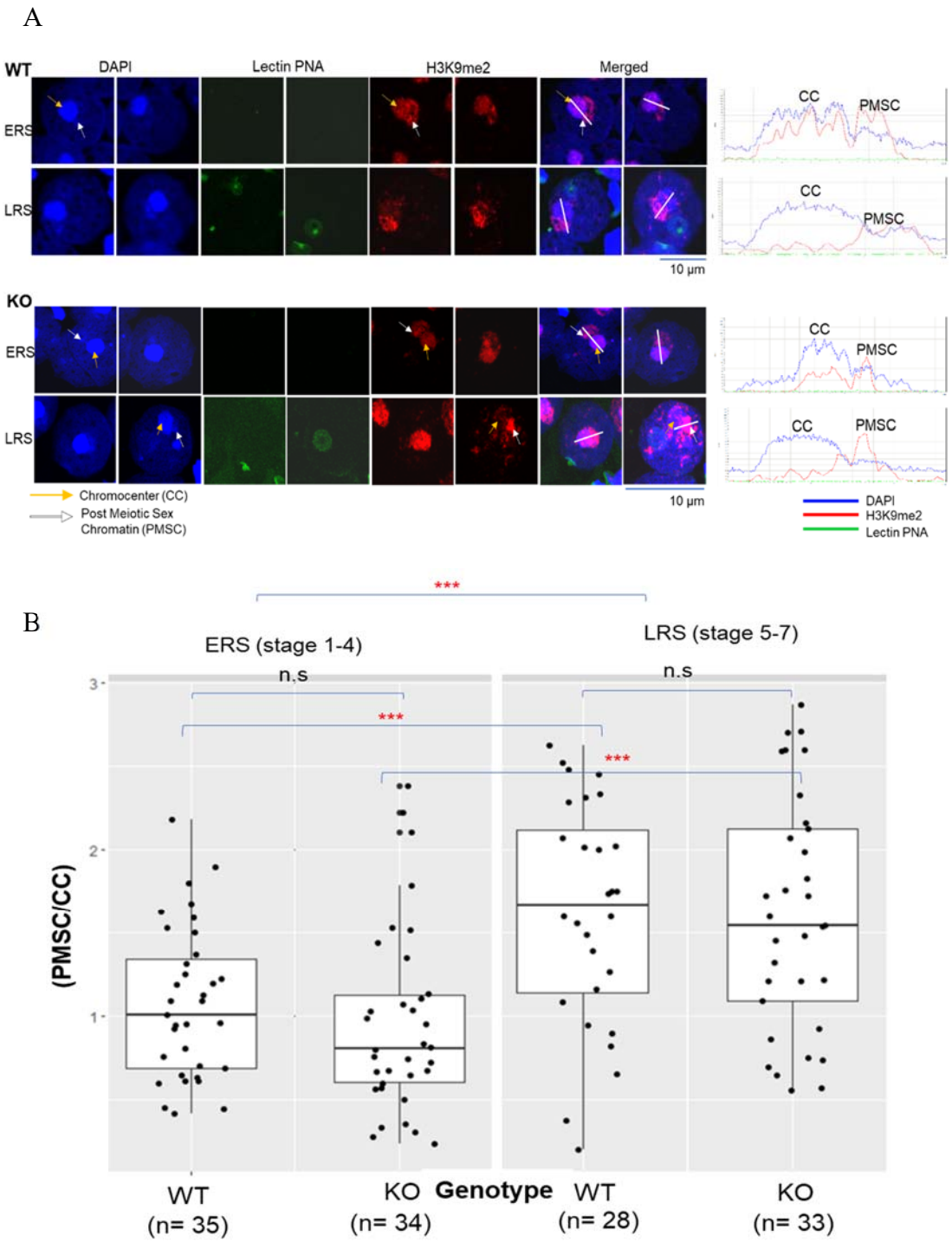


Figure 5-14: The intensity of repressive mark, H3K9me2 displays no difference between wt and KO round spermatids but is significantly increased in PMSC of late round spermatids (LRS) in both wt and hemizygous KO mice. A) Images of early round spermatids (ERS) and late round spermatids (LRS) from the hypotonic spread of male germ cells. The sample was indirectly stained with anti- H3K9me2 antibody (red), Lectin PNA (conjugated with Alexa Fluor 488), a marker for acrosome to determine the stages of RS cells and counterstained with DAPI to visualise nuclei. White lines in merged image indicate the path used to determine fluorescence intensity across the post meiotic sex chromatin and chromocenter. Scale bar is 10 µm**. B)** The quantification of H3K9me2 intensity in post meiotic sex chromatin shown as a ratio of PMSC/CC. ANOVA test, *** $p<0.001$. Number of mice used in this analysis wt n=1, KO

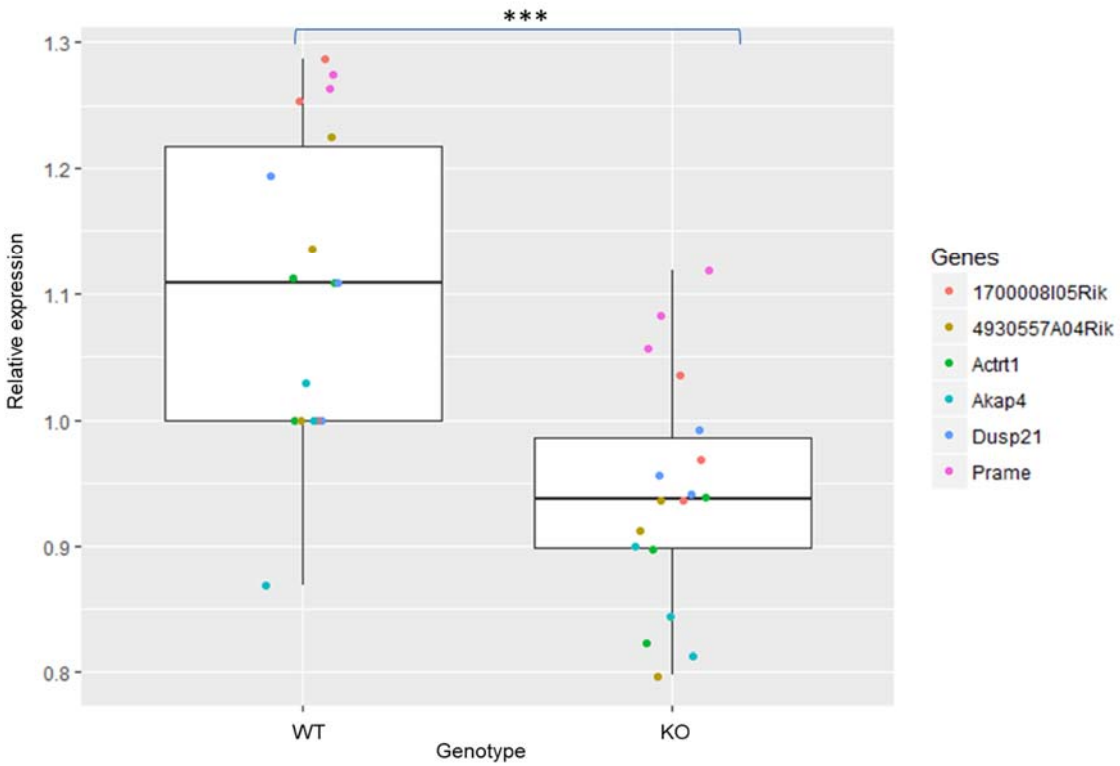


Figure 5-15: Round spermatid-specific X-linked genes that contain H2A.Lap1 in their TSS, showing reduced level of expression in H2A.Lap1 KO mice. The expression level was determined using qPCR, first normalised to β-actin and then to one of the WT sample to obtain the expression fold difference. WT (n=3) and H2A.Lap1 KO (n=3), ANOVA test, *** $p<0.001$.

5.11 H2A.Lap1 KO testis contain more clogged seminiferous tubules

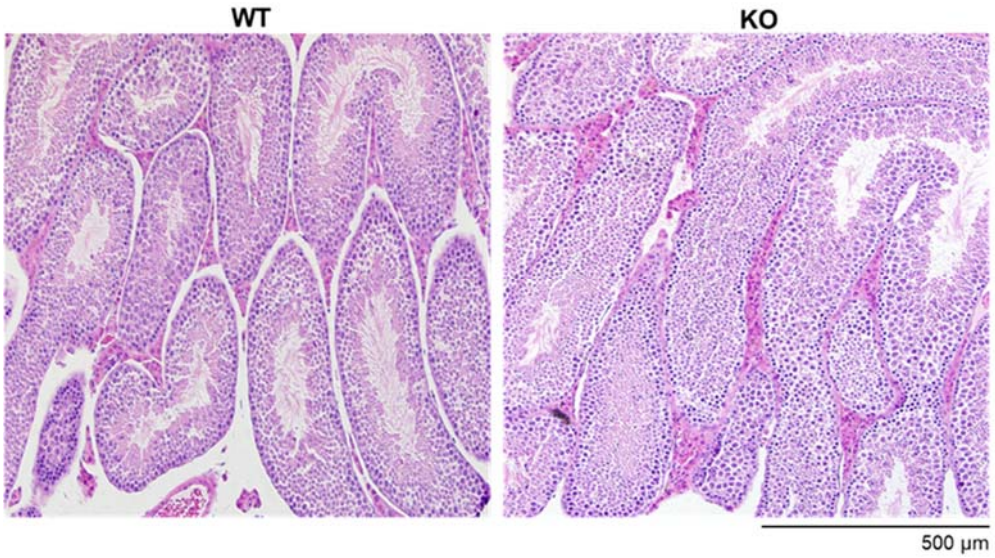
Finally, the overall testis morphology was assessed of formalin fixed, paraffin embedded testis sections, stained with Hematoxylin and Eosin (H&E). Hematoxylin binds to DNA and RNA, thus stains nuclei blue/violet, while eosin binds to cytoplasmic proteins, rich in Arginines and Lysines, and stains cytoplasm pink/red (Baker 1962). The testis size and seminiferous tubule cellular make up appeared similar in wt and hemizygous KO testis (data not shown). All cell types were represented and cellular layers corresponded to what was expected for each seminiferous tubule stage. As hypotonic conditions were not used for this preparation, round spermatid size of hemizygous KO spermatids did not differ from the wt cells. However, it was noticed that in some H2A.Lap1$^{-/Y}$ mice, seminiferous tubules were more often clogged so that the lumen was blocked by cytoplasmic residues (also known as residual bodies, RB), as indicated by pink staining of lumen in H&E-stained sections. In order to access this phenomena in more detail, sections from 20 mice (9 wt and 11 hemizygous KO siblings) were analysed and it was found that significant percentage of tubules were clogged in the H2A.Lap1 KO tubules (Figure 5-16).

The presence of clogged lumen in H2A.Lap1 KO testis indicates that the clearance of residual bodies by somatic Sertoli cells is inadequate in the absence of H2A.Lap1. Residual bodies are formed by the shedding cytoplasm of maturing elongating spermatids, as during maturation, spermatids must acquire motility which is achieved via an almost complete loss of cytoplasm, as well as nucleus shrinkage and a development of a flagellum. The residual bodies are cleared via phagocytosis by Sertoli cells in a stage-dependent manner (Garza et al. 1991).

Interestingly, H2A.Lap1 was shown by us previously to be a constituent of RBs (Soboleva et al. 2012), (Figure 5-17, D, E). Staining of paraffin-embedded seminiferous tubule sections with anti-H2A.Lap1 antibody also showed H2A.Lap1 signal in a form of multiple small foci in the lumen of tubules during stages III-VI, that co-insides with the formation of the residual bodies (Figure 5-17, A-C). Most surprisingly, we detected a signal within the nuclei of the Sertoli cells at the stages of active phagocytosis of RB by Sertoli cells. Intrigued by this novel observation, that H2A.Lap1 signal is detected not in the cytoplasm, which would be expected from the phagocytosis of H2A.Lap1-containing RBs, but in the nucleus of Sertoli cells, the mRNA expression of H2A.Lap1 in Sertoli cell line, TM4, was measured and compared

to H2A.Lap1 expression in total testis as well as in brain regions that are known to express small but detectable level of H2A.Lap1 mRNA. To our surprise, mRNA expression of H2A.Lap1 was undetectable in Sertoli cells (Figure 5-17, F). This finding prompted us to hypothesise that H2A.Lap1 is not expressed by SC but instead becomes a part of the nuclear makeup of the Sertoli cells in a protein form, through being engulfed by Sertoli cells. Further, upon incorporation into Sertoli cell chromatin, H2A.Lap1 may be involved in activation of a particular gene expression program, so that in the absence of H2A.Lap1, genes that may control function of Sertoli cells in clearing Residual bodies are not efficiently expressed, resulting in inefficient mechanism and clogged tubules. Alternatively, it is plausible that H2A.Lap1 may also bring important factors to the Sertoli cells that are required for a proper function of SC. Although these suggestions are highly speculative and require further investigation, if proven to be correct, this finding will represent a novel mechanism, where epigenetic information may be transferred extra-cellular, from germ cells to somatic Sertoli cells.

A



B

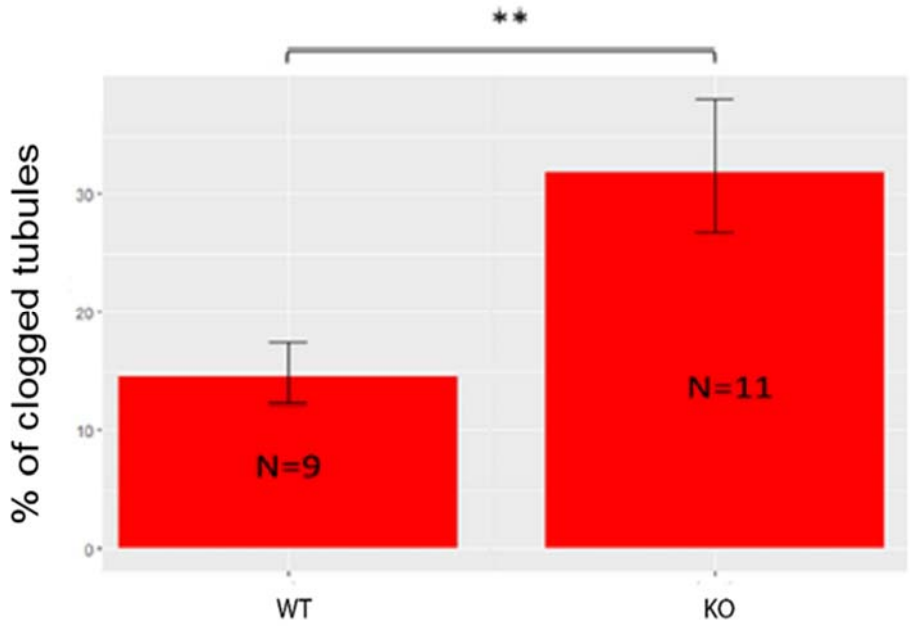


Figure 5-16: Hemizygous KO mice has more clogged tubules. A) image of hematoxylin and eosin (H&E) stained seminiferous tubule from wt and hemizygous KO mice. **B)** Histogram plot of the percentage of seminiferous tubule with clogged lumen in hemizygous KO mice (n=9) and their wt siblings (n=11). ANOVA test, **p=0.006547 (≤0.01)

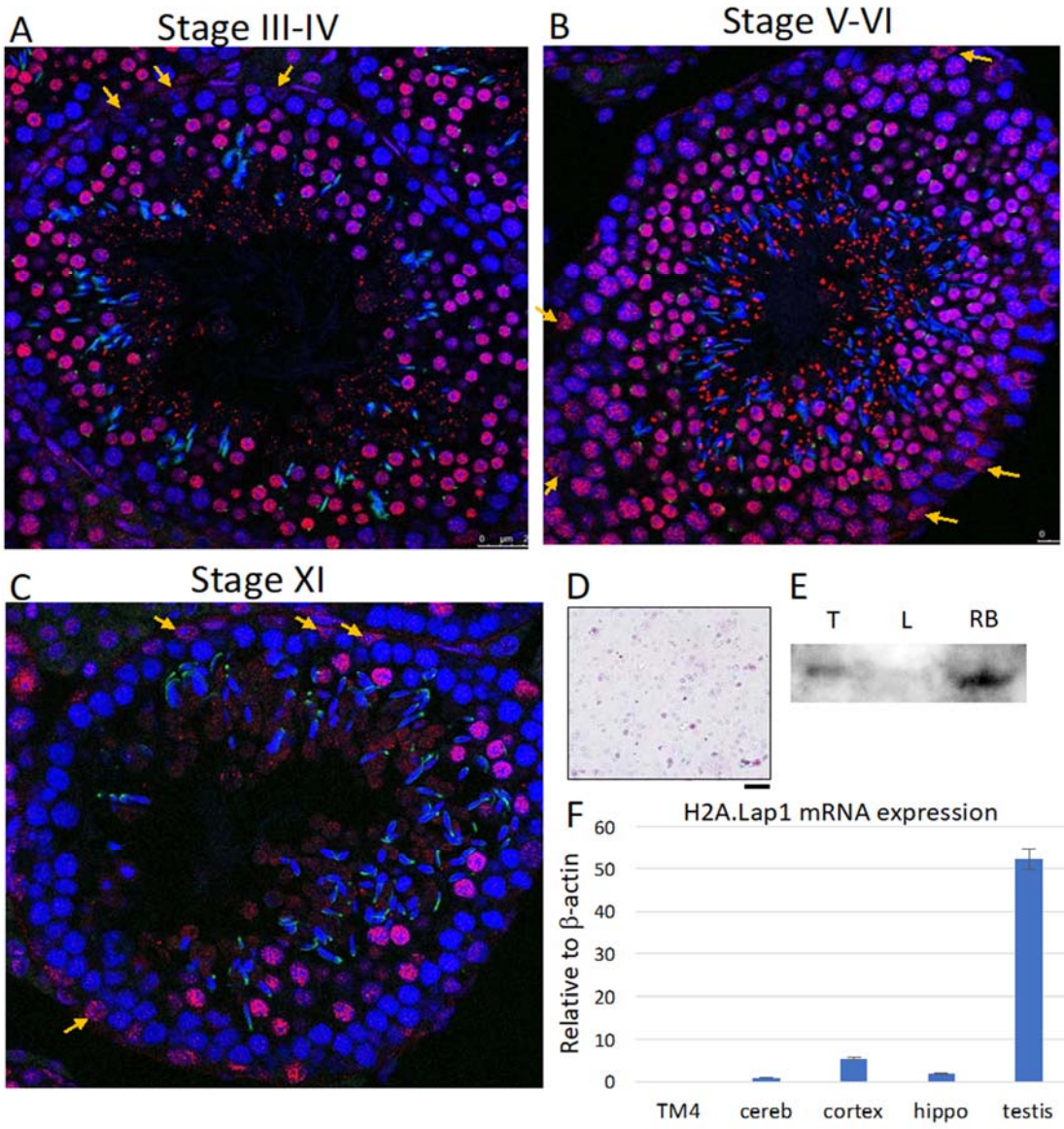


Figure 5-17: H2A.Lap1 is a component of residual bodies and Sertoli cell nuclei. A-C) adult mouse testis paraffin embedded sections were stained with anti-H2A.Lap1 antibodies and co-stained with DAPI to visualize nuclei. **A**: stage III-IV tubule shows that H2A.Lap1 appears in RB but not yet enriched in Sertoli cells. B: in stage V-VI tubules H2A.Lap1 is enriched in RB and SC nuclei; **C:** during stage XI RB disappear but H2A.Lap1 is still present in SC nuclei. Yellow arrows indicate SCs. **D:** Purified RBs stained with Pyronin Y, bar 20µm. **E:** Western blot. Lane T: 130µg of whole testis extract; lane L:110µg of whole liver extract; lane RB: 13 µg of RB extract. **F**: Sertoli cells do not express H2A.Lap1 mRNA. Level of expression of H2A.Lap1, measured by qPCR in Sertoli cell line, TM4, cerebellum, cortex, hippocampus and testis of adult mice. N=3, Error bars represent standard deviations.

REFERENCES

Baker, JR. 1962. 'Experiments on the action of mordants. 2. Aluminium-haematein.', *Quart. J. Microsc. Sci.*, 103: 493-517.

Barral, S., Y. Morozumi, H. Tanaka, E. Montellier, J. Govin, M. de Dieuleveult, G. Charbonnier, Y. Coute, D. Puthier, T. Buchou, F. Boussouar, T. Urahama, F. Fenaille, S. Curtet, P. Hery, N. Fernandez-Nunez, H. Shiota, M. Gerard, S. Rousseaux, H. Kurumizaka, and S. Khochbin. 2017. 'Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells', *Mol Cell*, 66: 89-101 e8.

de França, L.R. et al., 1993. Surface and surface-to-volume relationships of the Sertoli cell during the cycle of the seminiferous epithelium in the rat. Biology of reproduction, 49(6), pp.1215–28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8286604 [Accessed December 19, 2017].

Deng, T., Chen, Q. & Han, D., 2016. The roles of TAM receptor tyrosine kinases in the mammalian testis and immunoprivileged sites. Frontiers in bioscience (Landmark edition), 21, pp.316–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26709775 [Accessed December 20, 2017].

Di Liegro, C.M., Schiera, G. & Di Liegro, I., 2017. Extracellular Vesicle-Associated RNA as a Carrier of Epigenetic Information. Genes, 8(10), p.240. Available at: http://www.mdpi.com/2073-4425/8/10/240 [Accessed December 19, 2017].

Garza, M. M., L. K. Schwarz, J. M. Bonner, and F. R. Boockfor. 1991. 'Sertoli cell function varies along the seminiferous tubule: the proportion and response of transferrin secretors differ between stage-associated tubule segments', *Endocrinology*, 128: 1869-74.

Govin, J., E. Escoffier, S. Rousseaux, L. Kuhn, M. Ferro, J. Thevenon, R. Catena, I. Davidson, J. Garin, S. Khochbin, and C. Caron. 2007. 'Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis', *J Cell Biol*, 176: 283-94.

Greaves, I. K., D. Rangasamy, M. Devoy, J. A. Marshall Graves, and D. J. Tremethick. 2006. 'The X and Y chromosomes assemble into H2A.Z-containing [corrected] facultative heterochromatin [corrected] following meiosis', *Mol Cell Biol*, 26: 5394-405.

Le Borgne, M. et al., 2014. The RNA-binding protein Mex3b regulates the spatial organization of the Rap1 pathway. Development, 141(10), pp.2096–2107. Available at: http://dev.biologists.org/cgi/doi/10.1242/dev.108514.

Montellier, E., F. Boussouar, S. Rousseaux, K. Zhang, T. Buchou, F. Fenaille, H. Shiota, A. Debernardi, P. Hery, S. Curtet, M. Jamshidikia, S. Barral, H. Holota, A. Bergon, F. Lopez, P. Guardiola, K. Pernet, J. Imbert, C. Petosa, M. Tan, Y. Zhao, M. Gerard, and S. Khochbin. 2013. 'Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B', *Genes Dev*, 27: 1680-92.

Montellier, E., S. Rousseaux, Y. Zhao, and S. Khochbin. 2012. 'Histone crotonylation specifically marks the haploid male germ cell gene expression program: post-meiotic male-specific gene expression', *Bioessays*, 34: 187-93.

Namekawa, S. H., P. J. Park, L. F. Zhang, J. E. Shima, J. R. McCarrey, M. D. Griswold, and J. T. Lee. 2006. 'Postmeiotic sex chromatin in the male germline of mice', *Curr Biol*, 16: 660-7.

Peters, A. H., A. W. Plug, M. J. van Vugt, and P. de Boer. 1997. 'A drying-down technique for the spreading of mammalian meiocytes from the male and female germline', *Chromosome Res*, 5: 66-8.

Soboleva, T. A., M. Nekrasov, A. Pahwa, R. Williams, G. A. Huttley, and D. J. Tremethick. 2012. 'A unique H2A histone variant occupies the transcriptional start site of active genes', *Nat Struct Mol Biol*, 19: 25-30.

Soboleva, T.A. et al., 2017. A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. Available at: http://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.1006633&type=printable [Accessed March 24, 2017].

van der Heijden, G. W., A. A. Derijck, E. Posfai, M. Giele, P. Pelczar, L. Ramos, D. G. Wansink, J. van der Vlag, A. H. Peters, and P. de Boer. 2007. 'Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation', *Nat Genet*, 39: 251-8.

CHAPTER 6
GENERAL DISCUSSION

Previous studies by Soboleva et al. (2012, 2017) identified H2A.Lap1 as a component of the TSS and the intron-exon boundary of active genes expressed in the testis and the brain. In the testis, H2A.Lap1 is expressed meiotically and post-meiotically, from late pachytene to the late round spermatid stage coincident with the highest levels of expression that occurs during spermatogenesis. These observations suggested a role in transcription and splicing but all the experiments to date where correlative in nature. The availability of a H2A.Lap1 knockout would be an invaluable tool for investigating the importance of this histone variant. However, as the H2A.Lap1 protein is expressed from three different genes, this was a technically challenging feat. Our aim was to devise the most efficient and specific strategy to inhibit the function of all three genes. In this thesis, we report a simplified TALEN approach that achieved this goal whereby we used one pair of TALENs to specifically and simultaneously disrupt three gene copies of H2A.Lap1. To our knowledge this has not been done before (the TALEN methodology for producing the H2A.Lap1 KO mouse was recently submitted as a preprint; https://www.biorxiv.org/content/ early/2017/12/13/233379.article).

The major findings of this study are that H2A.Lap1 KO mice: 1) are subfertile, 2) have nuclei in round spermatids that are more condensed in hypotonic salt conditions, 3) display an apparent epigenetic compensatory mechanism with an increase in H3K4me3 and possibly H2A.L.2, 4) show a disruption of the hubs of transcription that are enriched with RNA Pol II, 4) have altered patterns of intron retention and reduced expression of genes, especially on the X chromosome, 5) an altered PMSC epigenetic landscape and most interestingly 6) exhibit a clogged seminiferous tubule phenotype.

6.1 TALENs versus CRISPR.

A major ongoing question is whether a RNA (e.g. CRISPR)- or protein (e.g. TALEN)-based mechanism will ultimately prove to be more specific for genome editing and this will be critical for future medical therapies. In this study, we have successfully applied the TALEN technology, whereby we used one pair of TALENs to specifically and simultaneously disrupt three gene copies of H2A.Lap1. Many human diseases are

caused by gene copy number variations (de Smith et al. 2008) and therefore the use of a limited number of genome editing enzymes might help in the future treatment of such diseases. The use of only one pair of TALENs, rather than multiple pairs, reduces the complexity of the TALEN approach (and hence the cost) and would reduce the likelihood of mosaicism and most importantly, reduce the possibility of off-target mutations. Indeed, our rigorous exome analyses did not reveal any TALEN-induced off-target mutations. We also found that the computational tools used to predict putative SNVs and indels have limitations, and therefore wet-lab validation is required before any conclusions can be made about off-target mutations.

The TALEN approach did expose several interesting phenomena. Firstly, the simultaneous knockout of highly homologous H2A.Lap1 genes leads to the formation of chimeras between these genes. This can be explained by the fact that TALEN activity can be retained for several cell divisions following their injection into the 2-cell stage, thus producing a mosaic genotype, as previously observed (Tesson et al. 2011; Li et al. 2014). As every chimera contained a small deletion, it is likely that terminal microhomology, a mechanism of NHEJ, was involved (Lieber 2010). It is important to point out that other technologies that rely on NHEJ for inducing mutations, like CRISPR/Cas9 technology, will also likely produce chimeras between highly homologous genes. Several rounds of backcrossing can successfully eliminate such small-scale genomic rearrangements. Secondly, we observed an all-or-nothing effect whereby the TALENs either modified all 3 H2A.Lap1 genes (6 and 3 alleles from female and male pups received its own unique mutation, respectively) or modified none, which reduced the need for a more complex breeding and screening strategy to establish a H2A.Lap1 KO colony.

6.2 H2A.Lap1 KO mice are viable but subfertile suggesting a compensatory mechanism may exist.

One of the main findings of this work was that the loss of H2A.Lap1 caused mild subfertility that manifested itself in smaller litter sizes when hemizygous (H2A.Lap1$^{-/Y}$) males were mated with wt females. This suggests that while H2A.Lap1 is important for fertility, it is not essential. However, there are a number other factors and/or compensatory mechanisms that could explain why only a mild subfertile phenotype is observed.

The fertility issues of hemizygous KO males may be somewhat obscured by the fact that the FVB/NJ strain of mouse is known to be exceptionally fertile characterised by >90% of successful mating's enabling the production of very large litters (9.5 pups per litter on average). In comparison, another widely used inbred strain, Balb/c, is characterised by less than 50% of successful mating's and only 5.2 pups per litter on average (Silver, Lee 1995). In the future, it would be important to re-analyse the fertility of hemizygous KO on a less fertile mouse genetic background i.e. backcross our hemizygous KO males with Balb/c mice.

An alternative view to one that H2A.Lap1 may not be essential for fertility is that it does play crucial role and hence compensatory epigenetic mechanisms come into play. Several active promoter-associated histone PTMs as well as a related histone variant, H2A.L.2, were examined by immunofluorescence. A major finding was that H3K4me3 became more enriched in the euchromatin of autosomal genes in the absence of H2A.Lap1, implying that this PTM may indeed play a compensatory role. Similarly, it was also observed that H2A.L.2 changed its expression timing, being expressed earlier in round spermatids, suggesting that H2A.L.2 may also play a compensatory role. In the future, ChIP-Seq experiments need to be performed to show that H3K4me3 and/or H2A.L.2 replaces H2A.Lap1 on promoters in H2A.Lap1 KO mice that were previously occupied by H2A.Lap1.

Ultimately, to prove that H2A.Lap1 is crucial for spermatogenesis, a conditional transgenic mouse needs to be produced whereby H2A.Lap1 is specifically knocked out only in the adult testis.

6.3 H2A.Lap1 deficient mice display major changes to nuclear compaction and to the organisation of transcriptional hubs

Perhaps the most remarkable discovery in this thesis is that in the absence of H2A.Lap1, the nuclei of round spermatids become significantly more compacted and therefore if compensatory mechanisms do exist, they are unable to decompact chromatin to the same extent as H2A.Lap1. Previously, we demonstrated that H2A.Lap1 could inhibit the folding of a chromatin fibre *in vitro* (Soboleva et al. 2012). The hypotonic spread technique employed here allowed us to show for the first time that H2A.Lap1 regulates compaction of the chromatin fibre *in vivo*. In the future, it will be of great interest to investigate how this H2A.Lap1 induced loss of compaction alters high-order genome organisation e.g. by performing Hi-C experiments on isolated round spermatids.

Another exciting finding is that in the absence of H2A.Lap1, RNA Pol II no longer occupies microscopically distinct transcription hubs or factories. Previously (Soboleva et al. 2017), this laboratory demonstrated that H2A.Lap1 co-localises with splicing speckles, which were also sites of active transcription because both the initiation and elongation form of RNA Pol II were enriched at these distinct nuclear bodies. We show here that in the absence of H2A.Lap1, RNA Pol II no longer occupies the same territory as splicing speckles becoming more diffuse within round spermatid nuclei. This suggests that H2A.Lap1, perhaps given its location at the TSS of active genes, has an important role in the recruitment and/or anchoring of RNA Pol II to active territories of transcription and splicing. This is consistent with previous mass spec data (Soboleva et al. 2012) where it was shown that H2A.Lap1 directly interacts with RNA Pol II. Alternatively, perhaps the loss of nuclear organisation caused by the increased compaction of H2A.Lap1 deficient nuclei may indirectly cause the mis-localisation of RNA Pol II.

6.4 RNA-Seq analysis of 30-day mouse testes reveals no changes in gene expression but alterations to intron retention.

Given the major changes in nuclear condensation, RNA Pol II nuclear localisation and the fact that H2A.Lap1 KO male mice are subfertile, it was a surprise to find that our RNA-Seq experiments revealed no genes whose expression were altered significantly in the absence of H2A.Lap1 (data not shown). One interpretation of this observation is that H2A.Lap1 may have no role in transcription initiation but only in splicing. However, we do not believe that this is the case (especially given that H2A.Lap1 is uniquely located at the TSS of highly expressed genes). There are several non-mutually exclusive reasons why no apparent change in global gene expression was observed.

First, as discussed in section 5-9, each individual wt and KO mouse testicle (three mice per each condition) displayed large individual gene transcriptome variability, and current RNA-Seq analysis pipelines cannot take into account such expression variability. In other words, such variability would mask any significant changes in gene expression. Indeed, we have come to the realisation that different ways of analysing RNA-Seq data needs to be developed that actually examines this expression variability/noise because it is possible that such patterns of variability may in fact be the distinct expression signature for a particular phenotype (normal or impaired). Second, as discussed above,

the replacement of H2A.Lap1 with H3K4me3 or H2A.L.2, or any other mark not yet investigated e.g. H2A.Z may result in restored levels of gene expression. However, our qPCR results produced using a different set of mice and a small number of autosomal genes showed an overall small but significant reduction of gene expression in the absence on H2A.Lap1, supporting our hypothesis that H2A.Lap1 is a crucial factor involved in control of high level of gene expression.

A third possibility is that H2A.Lap1 may have a more important role in inducing gene activation in response to intrinsic or external signals rather than maintaining steady state levels of gene expression. There is experimental support for this notion. Preliminary experiments (performed by Tanya Soboleva) revealed that when mice testes were subjected to physiological stress (by elevating testicular temperature to 42ºC for 30 minutes), this had a devastating effect on the seminiferous tubules of H2A.Lap1 KO mice compared to wt mice (Figure 6-1). In KO mice, the seminiferous tubules contain almost no meiotic and post-meiotic cells plus a large number of typical apoptotic cells are also visible. On the other hand, wt tubules appear largely normal. This example illustrates that under physiological stress, H2A.Lap1 becomes essential.

Recently, we established a collaboration with Timothy Bredy (The Brain Institute, University of Queensland) who has established the mouse model to investigate the memory of fear (Lin et al. 2011). In this paradigm, a conditioned stimulus (CS, in this case, a sound) is presented with an unconditioned stimulus (US, such as a foot shock). After several CS and US pairings, the CS alone induces the state of fear, evidenced by 'freezing' and an increased heart rate, among other reactions, which are monitored. The regions of the brain critical for this conditioned fear response are the hippocampus and the medial prefrontal cortex (mPFC) (regions where H2A.Lap1 is expressed). Most excitingly, their laboratory performed fear-conditioning experiments comparing hemizygous KO versus wt mice and found that H2A.Lap1 KO mice exhibited a slower rate of fear acquisition and impaired long-term memory of this fear (Figure 6-2). Significantly, RNA-Seq experiments (by T. Soboleva) revealed that before fear training, ~ 50 genes had an altered expression when wt mPFC was compared with KO mPFC. However, following fear training, the expression of > 300 fear-related genes were impaired (data not shown). In other words, the major impact of the loss of H2A.Lap1 was the induction of fear gene expression but not the steady state expression of constitutive genes in the mPFC.

Finally, while our RNA-Seq testis data did not reveal any changes in gene expression when wt versus H2A.Lap1 KO mice were compared, the loss of H2A.Lap1 increased the extent of intron retention providing support for our proposal that H2A.Lap1 has a role in the regulation of pre-mRNA splicing. This is the first demonstration that H2A.Lap1 has a role in splicing in its proper physiological context, the testes.

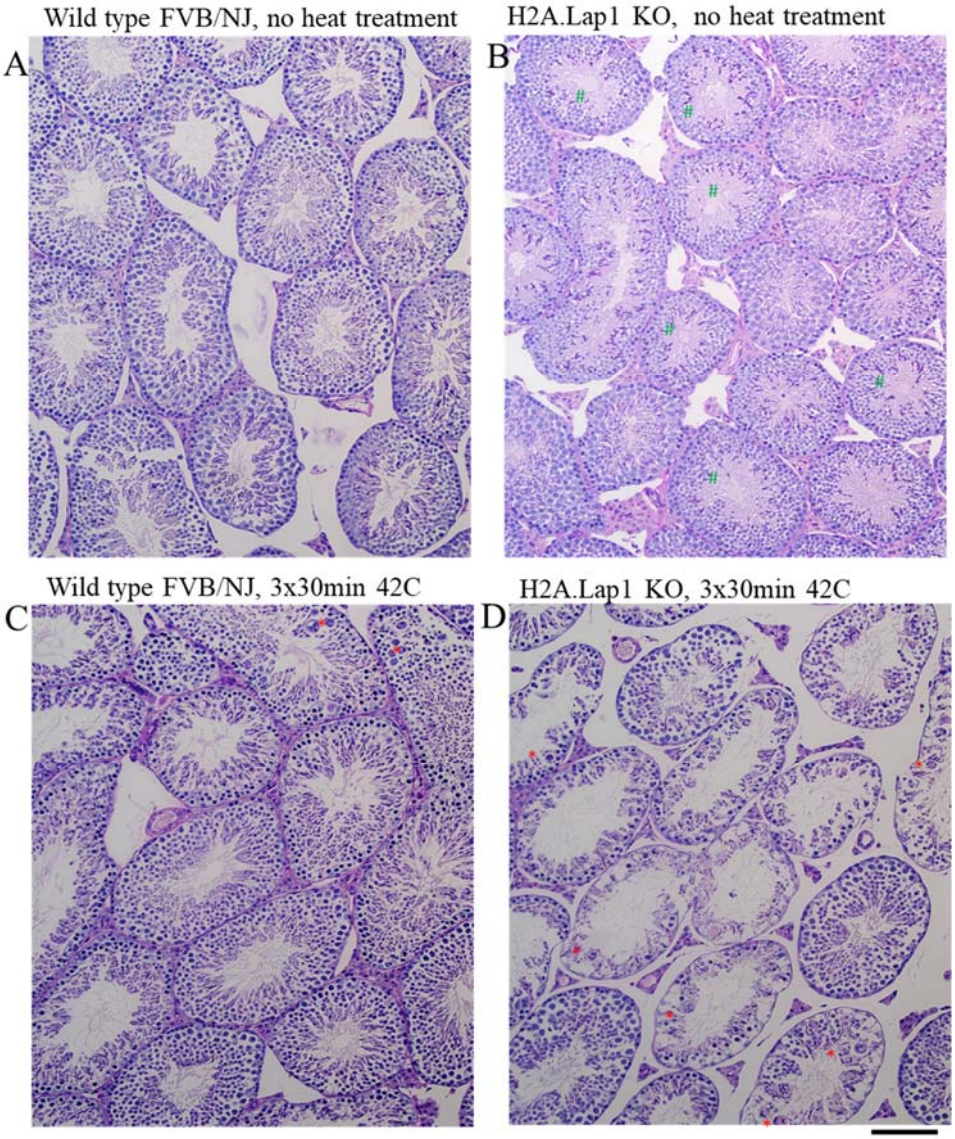


Figure 6-1: Heat treatment of mice testis. Panels A and C – wt FVB/NJArc, B and D- H2A.Lap1 KO. A and B – non-treated controls, C and D – mice testes were submerged into 42°C water bath for 30 minutes three times on three consecutive days and sacrificed on day 4. Note that control H2A.Lap1 KO mouse (panel B) has clogged seminiferous tubules, marked with #. Upon heat treatment, H2A.Lap1 KO mouse (panel D) lost majority of spermatocytes and round spermatids and contains numerous apoptotic cells marked by *, while the wt FVB/NJArc tubules (panel C) display almost normal phenotype with all cell types present and with a small number of apoptotic cells in the tubules (*). Scale bar is 250µm.

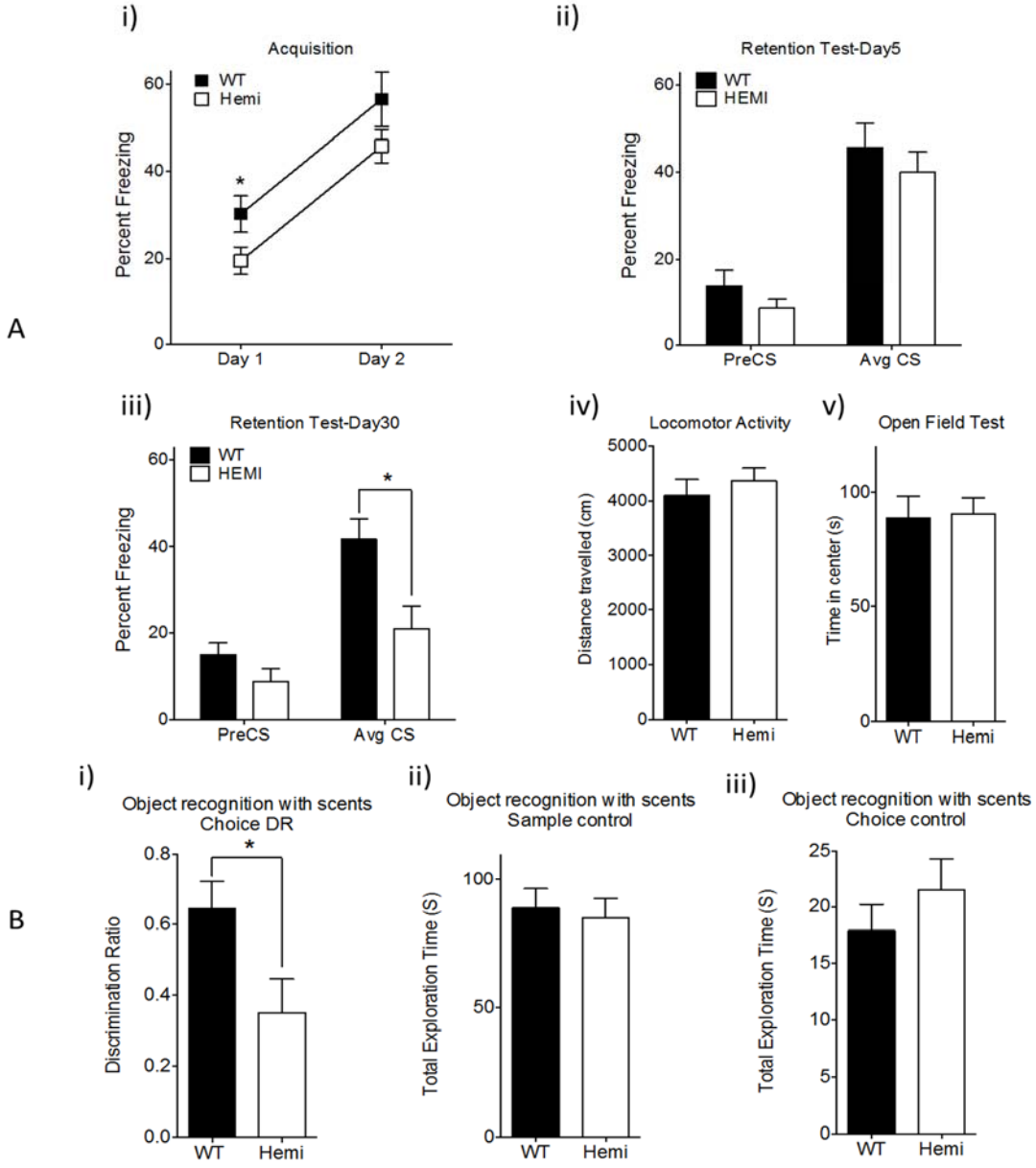


Figure 6-2: H2A.Lap1 depletion associated with impaired long-term memory. A i) Hemizygous KO mice exhibit a slower rate of acquisition of fear compared with wild type mice following 1 day of behavioural training (*p<0.05, t test, t=2.068 df=30). Although this difference goes away after a second day of training, there remains a trend, suggesting that the hemizygous KO mice have impaired or less stable fear memory. **A ii)** A retention test for short-term memory (Day 5) and long-term memory (Day 30, **A (iii)**), revealed a reduction in freezing at day 30. These initial observations strongly suggest that the hemizygous KO mice have less robust fear memory compared with wild type mice. This effect on fear memory is not due to any group differences in locomotor activity (**A (iv))** or generalized anxiety (**A (v)),** suggesting a fundamental role of H2A.Lap1 in the formation and stability of long-term memory**. B i)** showed that both hemizygous KO and wt mice have memory compare with chance, and hemizygous KO mice show less preference to the novel object, because their memory about the familiar object is not as good as wt mice. Figure **B ii)** and **iii)** are all control data: **B ii)** shows that wt and hemi spent same amount of time to explore objects during learning period (sample control); **B iii)** shows they spent similar amount of time during test (choice phase).

6.5 Loss of H2A.Lap1 alters the post-meiotic sex chromatin epigenetic landscape

During meiosis the sex chromosomes become highly condensed and silenced in order to avoid illegitimate recombination between the X and Y chromosome. This inactivation continues post-meiotically and this inactivation in round spermatids is referred to as post-meiotic sex chromatin (PMSC). Despite this silencing however, ~20 % of X-chromosome coded genes become activated in order to produce proteins that are vital for later stages of spermiogenesis. Others and we have shown that PMSC are decorated by a number of specific histone variants (H2A.Lap1 and H2A.Z) and PTMs necessary to ensure either gene silencing or activation of the required genes (Greaves et al. 2007; Moretti et al. 2016; Montellier et al. 2013). H2A.Lap1 is targeted to the inactive X chromosome in late round spermatids and this is coincident with X-linked gene activation. In addition to H2A.Lap1, histone crotonylation (Kcr) also decorates the TSS of reactivated X chromosome genes (Montellier et al. 2013). Other active marks, including H3K4me2 and H4K8Ac are also associated with PMSC (Khalil et al. 2004).

Our results have shown that neither Kcr, H3K4me2 nor H4K8ac increased at PMSC to compensate for the loss of H2A.Lap1. In fact, Kcr became depleted at PMSC in early round spermatids (ERS) when H2A.Lap1 was absent. On the other hand, H3K4me2 was depleted in late round spermatids (LRS) of H2A.Lap1 KO mice for reasons that are unclear. Given that there was no apparent epigenetic compensation at PMSC (at least for the active marks investigated), it was perhaps not surprising that the overall expression efficiency of X-linked genes was reduced by ~ 20% in H2A.Lap1 KO mice. However, to confirm that no epigenetic compensation process occurs on the X chromosome (unlike autosomes), it will be necessary to perform ChIP-Seq experiments examining these as well as other active marks.

6.6 Dynamic epigenetic changes between early and late round spermatids

This thesis has yielded other interesting observations concerning the changing epigenetic landscape between ERS and LRS that has not been observed before. In spite of lack of difference between wt and KO round spermatids, we have observed that the repressive mark H3K9me2 becomes more enriched in LRS on PMSC which coincides with a depletion in the active marks H4K8Ac and Kcr. At euchromatin, we also observed other active marks (H3K36me3 and H3K4me3) become depleted in LRS in wt mice. These reductions in active marks at the latter stages of round spermatid differentiation

might be in preparation for the global shut down of transcription that occurs when the round spermatids begin to elongate and condense.

Another intriguing observation is that in the absence of H2A.Lap1, H2A.L.2 loses its heterochromatin enrichment in elongating-condensing spermatids. This could be because it begins to be expressed earlier in KO round spermatids being incorporated into euchromatin and therefore by the time the condensing spermatid stage is reached, there are limiting amounts of H2A.L.2 protein to be incorporated into heterochromatin. Alternatively, in ways that are not understood, the proper expression H2A.Lap1 is subsequently required for the proper expression and incorporation of H2A.L.2 into heterochromatin in condensing spermatids e.g. if H2A.Lap1 regulates the expression of a H2A.L.2 chaperone. It is worth noting that H2A.L.2 KO mice are infertile (Barral et al. 2017) and therefore mis-incorporation of H2A.L.2 could provide one possible explanation as to why H2A.Lap1 KO mice are subfertile. Clearly, future experiments are required to test these speculative ideas.

6.7 A possible role for H2A.Lap1 in somatic Sertoli cell function.

A final intriguing and unexpected finding of this thesis was that H2A.Lap1 was detected in the nucleus of Sertoli cells even though, based on a Sertoli cell line, is not expressed in these cells. Sertoli cells are crucial for the proper differentiation of male germ cells. They send intracellular signals to control differentiation of spermatogonial stem cells, they clear multiple apoptotic events by phagocytosis (up to 50% of germ cells undergo apoptosis (Nakanishi & Shiratsuchi 2004), they are solely responsible for clearing out residual bodies from the lumen, and they create a blood-testis barrier that allows germ cells to proliferate and differentiate in an immuno-privileged environment (reviewed in Sharpe et al. (2003).

Previously, this laboratory demonstrated that H2A.Lap1 is a constituent of residual bodies (RB) (Soboleva et al. 2012). Here, we observed that RB are not cleared efficiently by Sertoli cells in H2A.Lap1 KO testes thus producing tubules that are clogged. Given that Sertoli cells phagocytose RB, it would not be surprising to find H2A.Lap1 in the cytoplasm. However, what is fascinating is that H2A.Lap1 is found in the nucleus indicating that H2A.Lap1 is imported into the nucleus presumably to perform an important function i.e to regulate gene expression and/or splicing. Alternatively, given that H2A.Lap1 is a RNA binding protein, and that RB is enriched with RNA, H2A.Lap1 may transport regulatory RNA molecules into the nuclei of Sertoli

cells. Future experiments would include purifying Sertoli cells (Anway et al. 2003) and performing H2A.Lap1 ChIP-Seq experiments to confirm its incorporation into chromatin and determining its genomic locations. In addition, it would be necessary to purify RB and perform mass spec as well as iCLIP experiments to determine the protein and RNA partners of H2A.Lap1. If proven to be true, this would represent a novel mechanism of epigenetic regulation whereby an epigenetic regulator (H2A.Lap1) produced by one cell type can be transferred to a different cell type to regulate the expression and function of its genome.

REFERENCES

Anway, M.D. et al., 2003. Isolation of sertoli cells from adult rat testes: an approach to ex vivo studies of Sertoli cell function. *Biology of reproduction*, 68(3), pp.996–1002. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12604653 [Accessed January 15, 2018].

Barral, S. et al., 2017. Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Molecular Cell*, 66(1), p.89–101.e8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28366643 [Accessed January 15, 2018].

Greaves, I.K. et al., 2007. H2A.Z contributes to the unique 3D structure of the centromere. *Proceedings of the National Academy of Sciences of the United States of America*, 104(2), pp.525–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17194760 [Accessed March 24, 2017].

Khalil, A.M., Boyar, F.Z. & Driscoll, D.J., 2004. Dynamic histone modifications mark sex chromosome inactivation and reactivation during mammalian spermatogenesis. *Proc Natl Acad Sci U S A*, 101(47), pp.16583–16587. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15536132.

Li, C. et al., 2014. Simultaneous Gene Editing by Injection of mRNAs Encoding Transcription Activator-Like Effector Nucleases into Mouse Zygotes. *Molecular and cellular biology*, 34(February).

Lieber, M.R., 2010. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry*, 79(1), pp.181–211. Available at: http://www.annualreviews.org/doi/10.1146/annurev.biochem.052308.093131 [Accessed July 25, 2017].

Lin, Q. et al., 2011. The brain-specific microRNA miR-128b regulates the formation of fear-extinction memory. *Nature Neuroscience*, 14(9), pp.1115–1117. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21841775 [Accessed January 15, 2018].

Montellier, E. et al., 2013. Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes and Development*, 27(15), pp.1680–1692.

Moretti, C. et al., 2016. Expression and epigenomic landscape of the sex chromosomes in mouse post-meiotic male germ cells. *Epigenetics and Chromatin*, 9(1), pp.1–18.

Nakanishi, Y. & Shiratsuchi, A., 2004. Phagocytic removal of apoptotic spermatogenic cells by Sertoli cells: mechanisms and consequences. *Biological &*

pharmaceutical bulletin, 27(1), pp.13–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14709891 [Accessed January 15, 2018].
Sharpe, R.M. et al., 2003. Proliferation and functional maturation of Sertoli cells, and their relevance to disorders of testis function in adulthood. *Reproduction (Cambridge, England)*, 125(6), pp.769–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12773099 [Accessed January 15, 2018].
Silver, Lee, M., 1995. *Mouse Genetics Concepts and Applications*, Oxford University Press.
de Smith, A.J. et al., 2008. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenetic and genome research*, 123(1–4), pp.17–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19287135 [Accessed January 15, 2018].
Soboleva, T.A. et al., 2017. A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. Available at: http://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.1006633&type=printable [Accessed March 24, 2017].
Soboleva, T.A. et al., 2012. A unique H2A histone variant occupies the transcriptional start site of active genes. *Nature Structural & Molecular Biology*, 19(1), pp.25–30. Available at: http://www.nature.com/doifinder/10.1038/nsmb.2161 [Accessed March 24, 2017].
Tesson, L. et al., 2011. Knockout rats generated by embryo microinjection of TALENs. *Nature Biotechnology*, 29(8), pp.695–696. Available at: http://www.nature.com/doifinder/10.1038/nbt.1940.

APPENDIX

Supplementary Table S1: List of high confidence differential Intron Retention events as estimated by IRFinder in H2A.Lap1KO and wt mice

												WT					KO		
	Chr	Start	End	Intron-GeneName/ GeneID	Direction	Excluded Bases	p-diff	p-increased	p-decreased	IR ratio	Intron Cover	Intron Depth	Splices Max	Splices Exact	IR ratio	Intron Cover	Intron Depth	Splices Max	Splices Exact
1	17	24594595	24594671	Pkd1/ ENSMUSG00000032855	+	10	0.00086152	0.999138	0.00086152	0.156426	1	32.8214	174	174	0.061088	1	14.9643	229	229
2	15	58430098	58430178	D15Ertd621e/ ENSMUSG00000037119	+	10	0.00162941	0.00162941	0.998371	0	0	0	306	306	0.027825	1	8.21429	286	286
3	11	58042642	58042743	Larp1/ ENSMUSG00000037331	+	10	0.0042067	0.0042067	0.995793	0.024978	1	33.8919	1206	1206	0.043358	1	61.5946	1259	1259
4	2	30030370	30030451	Sptan1/ ENSMUSG00000057738	+	10	0.00455201	0.995448	0.00455201	0.011764	1	21.6897	1822	1821	0.004246	1	8.65517	2029	2029
5	5	108655364	108655439	Dgkq/ ENSMUSG00000004815	-	10	0.00534799	0.994652	0.00534799	0.205882	1	4.66667	18	18	0	0	0	32	32
6	8	3636421	3636493	Stxbp2/ ENSMUSG00000004626	+	10	0.00673497	0.00673497	0.993265	0.015367	1	11.3462	727	726	0.035253	1	27.0769	741	735
7	4	155587784	155587860	Nadk/ ENSMUSG00000029063	+	10	0.00727491	0.00727491	0.992725	0.034411	1	16.5357	464	445	0.072941	1	31	394	379
8	8	84003002	84003115	1700067K01Rik/ ENSMUSG00000046408	+	10	0.00872941	0.991271	0.00872941	0.029904	1	13.1628	427	427	0.008197	1	4	484	484
9	9	108814351	108814429	Nckipsd/ ENSMUSG00000032598	+	10	0.0107419	0.989258	0.0107419	0.187713	1	3.92857	17	17	0.014032	0.441	0.357143	31	31
10	11	120562545	120562654	P4hb/ ENSMUSG00000025130	-	10	0.0123026	0.987697	0.0123026	0.013173	1	11.8537	888	888	0.003845	1	3.36585	872	872
11	5	145043556	145044096	1700018F24Rik/ ENSMUSG00000029620	+	10	0.0126676	0.0126676	0.987332	0.05673	1	132.132	2197	2132	0.073283	1	168.594	2132	2070
12	7	122220739	122220849	Chp2/ ENSMUSG00000030865	+	10	0.0128993	0.0128993	0.987101	0.010649	0.98	1.55	144	144	0.065476	1	11	157	157
13	15	76596698	76596776	Cpsf1/ ENSMUSG00000034022	-	10	0.0130521	0.986948	0.0130521	0.038286	1	19.7857	497	497	0.015302	1	8.5	546	546
14	7	25303497	25303581	Prr19/ ENSMUSG00000058741	+	10	0.013681	0.013681	0.986319	0.026437	1	25.9333	955	947	0.045531	1	48.4667	1016	1012
15	10	127720020	127720152	Myo1a/ ENSMUSG00000025401	+	10	0.0157843	0.984216	0.0157843	0.038462	1	12.08	301	301	0.011028	1	3.78	339	339
16	15	89282004	89282090	Ppp6r2/ ENSMUSG00000036561	+	10	0.0166393	0.983361	0.0166393	0.03639	1	12.5	329	329	0.01369	1	5.34375	383	383
17	8	56532514	56532775	Cep44/ ENSMUSG00000038215	-	10	0.0168562	0.0168562	0.983144	0.097925	1	33.2178	280	280	0.157553	1	47.1287	234	234
18	17	24678052	24678595	Zfp598/ ENSMUSG00000041130	+	10	0.0170308	0.0170308	0.982969	0.10352	1	47.3442	405	405	0.154513	1	65.2419	351	351
19	17	34027522	34027622	H2-Ke6/ ENSMUSG00000073422	-	10	0.0171102	0.0171102	0.98289	0.075614	1	14.8056	181	181	0.148354	1	28.9167	166	166
20	7	44500438	44500967	Fam71e1/ ENSMUSG00000051113/ known-exon	+	68	0.0189671	0.0189671	0.981033	0.04649	1	168.211	3082	3057	0.057507	1	208.795	3054	3018
21	14	55719142	55719239	Rabggta/ ENSMUSG00000040472	-	10	0.0193371	0.0193371	0.980663	0.066407	1	20.4857	288	288	0.116162	1	34.1714	260	260
22	5	110342404	110342499	P2rx2/ ENSMUSG00000029503	-	10	0.0208594	0.979141	0.0208594	0.110528	1	39.1429	315	315	0.06645	1	24.4857	344	344

23	15	76248900	76248977	Grina/ ENSMUSG00000022564	+	10	0.0218147	0.978185	0.0218147	0.013051	1	55.963	4224	4224	0.00855	1	40.5556	4688	4688
24	9	21833173	21833363	Dock6/ ENSMUSG00000032198	-	10	0.0225117	0.977488	0.0225117	0.033329	1	3.93056	114	113	0.000332	0.0444444	0	133	133
25	15	89185464	89185537	Dennd6b/ ENSMUSG00000015377/ known-exon	-	10	0.0233786	0.976621	0.0233786	0.143243	1	106	634	620	0.107609	1	83.4444	692	681
26	2	24926194	24926271	Arrdc1/ ENSMUSG00000026972/ known-exon	-	10	0.0234023	0.976598	0.0234023	0.446062	1	24.963	26	26	0.23057	1	13.1852	38	38
27	14	54468781	54469035	Lrp10/ ENSMUSG00000022175	+	10	0.0241515	0.0241515	0.975848	0.023757	1	7.39796	304	304	0.053316	1	14.6429	260	260
28	17	25235111	25235181	Gnptg/ ENSMUSG00000035521	-	10	0.0244091	0.0244091	0.975591	0.06821	1	18.6667	246	246	0.12535	1	26.0833	180	179
29	18	84947939	84948022	Timm21/ ENSMUSG00000024645	-	10	0.0263483	0.0263483	0.973652	0.111015	1	32.9677	264	248	0.173333	1	47.3871	226	214
30	17	25829163	25829239	Jmjd8/ ENSMUSG00000025736	+	10	0.0267616	0.973238	0.0267616	0.030678	1	3.89286	120	120	0.003021	0.409091	0.285714	132	131
31	7	44860911	44860988	Pnkp/ ENSMUSG00000002963	+	10	0.0267624	0.0267624	0.973238	0.008621	0.925373	2	227	227	0.036394	1	11.4815	301	301
32	7	25704278	25704783	Tgfb1/ ENSMUSG00000002603	+	10	0.0290665	0.970933	0.0290665	0.024472	0.785859	3.13568	125	125	0.002105	0.383838	0.211055	182	182
33	1	171222153	171222374	Tomm40l/ ENSMUSG00000005674	-	41	0.0295543	0.0295543	0.970446	0.089864	1	5.43056	51	50	0.219411	1	11.8056	36	35
34	3	145578358	145578444	Znhit6/ ENSMUSG00000074182	+	10	0.0306191	0.969381	0.0306191	0.010899	1	4	363	362	0	0	0	370	369
35	7	126862720	126862817	Hirip3/ ENSMUSG00000042606	+	10	0.0307826	0.969217	0.0307826	0.106735	1	73.4857	615	604	0.077013	1	63.9143	766	749
36	X	74276556	74276645	Dnase1l1/ ENSMUSG00000019088	-	10	0.03125	0.03125	0.96875	0.000602	0.0126582	0	19	19	0.183406	1	3.81818	17	17
37	2	25616312	25616531	Ccdc183/ ENSMUSG00000026940/ known-exon	-	10	0.0320405	0.0320405	0.967959	0.138908	1	191.482	1182	1182	0.166628	1	212.941	1064	1063
38	1	36520227	36520518	Cnnm3/ ENSMUSG00000001138	+	10	0.033357	0.966643	0.033357	0.38109	1	11.6991	19	18	0.170681	1	8.02655	39	36
39	Y	821360	821449	Uba1y/ ENSMUSG00000069053	+	10	0.0336092	0.0336092	0.966391	0	0	0	168	168	0.023121	1	4	169	169
40	1	182442297	182442632	Trp53bp2/ ENSMUSG00000026510	+	10	0.0337359	0.966264	0.0337359	0.056584	1	2.51908	42	42	0	0	0	60	60
41	17	25149101	25149210	Clcn7/ ENSMUSG00000036636	+	10	0.0340887	0.965911	0.0340887	0.115744	1	11.7805	90	90	0.049459	1	5.46341	105	105
42	13	24831929	24832009	Tdp2/ ENSMUSG00000035958	+	10	0.0342681	0.0342681	0.965732	0.01	1	3	297	294	0.031064	1	9.71429	303	299
43	10	81261715	81261794	Matk/ ENSMUSG00000004933/ known-exon	+	10	0.034556	0.034556	0.965444	0.152308	1	3.41379	19	19	0.423547	1	9.55172	13	13
44	17	34590260	34590460	Gpsm3/ ENSMUSG00000034786	+	10	0.0364772	0.0364772	0.963523	0.088235	1	3	31	28	0.294664	1	8.35526	20	19
45	3	94430651	94430817	Tdrkh/ ENSMUSG00000041912/ known-exon	+	10	0.0371743	0.0371743	0.962826	0.47262	1	126.359	141	141	0.583856	1	155.734	106	106

				Ube2d2a/														
46	18	35800029	35800109	ENSMUSG00000091896	+	10	0.0377401	0.96226	0.0377401	0.01332	1	48.0714	3561	3557	0.00881	1	33.5	3769	3767
				Mast3/															
47	8	70783809	70784142	ENSMUSG00000031833	-	10	0.0394907	0.960509	0.0394907	0.035466	1	2.61069	70	70	0.000874	0.0804954	0	92	92
				4930519G04Rik/															
48	5	114870265	114870718	ENSMUSG00000029564	+	10	0.0401014	0.0401014	0.959899	0.155203	1	160.017	871	871	0.187688	1	175.832	761	761
				Cad/															
49	5	31075853	31075954	ENSMUSG00000013629	+	10	0.0417683	0.958232	0.0417683	0.020905	1	4.27027	200	200	0.001516	0.274725	0	181	181
				Mocs1/															
50	17	49449161	49449434	ENSMUSG00000064120	+	10	0.0422586	0.0422586	0.957741	0	0	0	58	58	0.061971	1	1.91589	29	29
				Idh3g/															
51	X	73779526	73779613	ENSMUSG00000002010	-	10	0.0424464	0.0424464	0.957554	0.002801	0.844156	1	356	356	0.016129	1	5	305	305
				Lck/															
52	4	129551870	129551947	ENSMUSG00000000409	-	10	0.0426883	0.0426883	0.957312	0	0	0	28	28	0.12	1	3	22	22
				Idh3g/															
53	X	73779171	73779465	ENSMUSG00000002010	-	10	0.0433323	0.0433323	0.956668	0.001415	0.369718	0.175439	260	260	0.013778	1	3.01754	216	216
				Mrpl28/															
54	17	26125391	26125477	ENSMUSG00000024181	+	10	0.0435337	0.0435337	0.956466	0.001437	0.631579	0.8125	439	435	0.012058	1	5.6875	466	459
				Arhgap21/															
55	2	20859796	20859880	ENSMUSG00000036591	-	10	0.0442231	0.0442231	0.955777	0.02381	0.945946	1	41	40	0.131222	1	4.83333	32	32
				Rapgef3/															
56	15	97749741	97750005	ENSMUSG00000022469	-	10	0.0445239	0.955476	0.0445239	0.054546	0.858268	1.5	26	26	0.00568	0.291339	0	51	51
				Tstd3/															
57	4	21759529	21759622	ENSMUSG00000028251	-	10	0.0451364	0.954864	0.0451364	0.013348	1	3.57143	263	263	5.24E-05	0.0120482	0	226	226
				Uhrf1/															
58	17	56318150	56318246	ENSMUSG00000001228	+	10	0.0461142	0.953886	0.0461142	0.054154	1	45.8611	801	801	0.036967	1	35.5833	927	926
				Pkd1/															
59	17	24590937	24591014	ENSMUSG00000032855	+	10	0.0474164	0.952584	0.0474164	0.128535	1	16.6667	105	103	0.069431	1	10.2963	131	128
				Fgfr3/															
60	5	33730059	33730143	ENSMUSG00000054252	+	10	0.0490416	0.950958	0.0490416	0.089005	1	2.83333	29	29	0.002247	0.0810811	0	36	36
				Cdc42bpg/															
61	19	6315096	6315171	ENSMUSG00000024769	+	10	0.0495059	0.0495059	0.950494	0	0	0	84	84	0.04	0.938462	3	72	72
				Fanca/															
62	8	123268549	123268744	ENSMUSG00000032815	-	10	0.051309	0.948691	0.051309	0.036882	0.956757	3.29333	86	86	0.001542	0.151351	0	98	98
				Nras/															
63	3	103064157	103064233	ENSMUSG00000027852	+	10	0.0521193	0.947881	0.0521193	0.012228	1	13.0357	1017	1016	0.005505	1	6	1042	1042
				Tmem131/															
64	1	36825346	36825428	ENSMUSG00000026116	-	10	0.052319	0.947681	0.052319	0.015152	1	3	195	195	0.001284	0.277778	0	216	216
				Adad2/															
65	8	119615507	119615610	ENSMUSG00000024266	+	10	0.0531216	0.946878	0.0531216	0.086279	1	59.2051	620	620	0.062333	1	44.8718	674	673
				Dock6/															
66	9	21832945	21833034	ENSMUSG00000032198	-	10	0.0546569	0.0546569	0.945343	0.002552	0.240506	0	94	94	0.030083	1	2.63636	85	85
				Gramd4/															
67	15	86130190	86130256	ENSMUSG00000035900	+	10	0.0552617	0.0552617	0.944738	0.005225	0.0892857	0	17	16	0.1875	1	3	13	13
				Kdm5d/															
68	Y	927891	927969	ENSMUSG00000056673	+	10	0.0566219	0.0566219	0.943378	0.004184	0.0882353	0	21	21	0.143885	1	2.85714	17	17
				Rpusd4/															
69	9	35272627	35272701	ENSMUSG00000032044	+	10	0.0572766	0.0572766	0.942723	0.034163	1	3.96154	112	112	0.085121	1	9.76923	105	105
				Tle6/															
70	10	81595245	81595322	ENSMUSG00000034758	-	10	0.0585013	0.0585013	0.941499	0.00622	0.38806	0.222222	62	62	0.05	1	3	57	57
				Ckb/															
71	12	111671347	111671424	ENSMUSG00000001270	-	10	0.0592346	0.940765	0.0592346	0.089783	1	9.66667	98	98	0.033058	1	3.55556	104	103

				Il17re/															
72	6	113466154	113466252	ENSMUSG00000043088	+	10	0.0597225	0.940277	0.0597225	0.146919	1	7.75	45	44	0.063291	1	5	74	74
				Prickle4/															
73	17	47690261	47690573	ENSMUSG00000096549	-	10	0.0604604	0.93954	0.0604604	0.200524	0.966887	2.5082	10	10	0.047889	0.910596	1.10656	22	22
				Ddx51/															
74	5	110656641	110656815	ENSMUSG00000029504	+	10	0.0606755	0.939325	0.0606755	0.029664	1	13.8485	453	446	0.014894	1	7.10606	470	470
				Dda1/															
75	8	71472117	71472194	ENSMUSG00000074247	+	10	0.0614111	0.938589	0.0614111	0.01532	1	24.037	1545	1525	0.009349	1	15.2222	1613	1581
				Fads3/															
76	19	10054121	10054226	ENSMUSG00000024664	+	10	0.0615068	0.0615068	0.938493	0.006818	0.315789	0.0512821	46	46	0.081581	1	2.48718	26	26
				Matk/															
77	10	81260182	81260907	ENSMUSG00000004933	+	10	0.0625	0.0625	0.9375	0.013198	0.427972	0.320557	30	30	0.09296	0.854545	2.97213	29	29
				Esyt3/															
78	9	99324240	99324896	ENSMUSG00000037681	-	10	0.0626425	0.0626425	0.937358	0.054951	1	8.37308	144	143	0.104674	1	15.3154	131	130
				Cd209g/															
79	8	4135020	4135131	ENSMUSG00000079168	+	10	0.062814	0.062814	0.937186	0.189723	1	4.68293	20	20	0.465991	1	7.85366	9	8
				Gpkow/															
80	X	7708021	7708110	ENSMUSG00000031148	+	10	0.0642543	0.0642543	0.935746	0.002825	0.405063	0.30303	143	143	0.02069	1	3	142	142
				AI481877/															
81	4	59071752	59072284	ENSMUSG00000038598	-	10	0.0649458	0.935054	0.0649458	0.042665	0.831418	1.7381	38	38	0.00363	0.222222	0	61	61
				Nrn1l/															
82	8	105894480	105894622	ENSMUSG00000044287	+	27	0.0649679	0.935032	0.0649679	0.214286	1	3	11	11	0.025065	0.591304	0.723404	23	23
				Nckipsd/															
83	9	108814747	108814927	ENSMUSG00000032598	+	10	0.0649928	0.935007	0.0649928	0.268275	1	10.6324	29	29	0.121951	1	5	36	36
				St5/															
84	7	109525349	109525538	ENSMUSG00000031024	-	10	0.0662378	0.933762	0.0662378	0.040992	1	20.6027	482	482	0.024953	1	13.2055	516	516
				Ppp2r5b/															
85	19	6232400	6232527	ENSMUSG00000024777	-	10	0.0694062	0.930594	0.0694062	0.052767	0.786325	3.06383	55	55	0	0	0	55	55
				Pkd1l3/															
86	8	109667185	109667526	ENSMUSG00000048827	+	10	0.0701059	0.0701059	0.929894	0.00193	0.386707	0.218045	200	200	0.013381	0.963746	2.83459	207	207
				Asah2/															
87	19	32016566	32016683	ENSMUSG00000024887	-	10	0.0705808	0.929419	0.0705808	0.132161	1	4.72093	26	26	0.002796	0.0560748	0	16	16
				Itgad/															
88	7	128203033	128203157	ENSMUSG00000070369	+	10	0.0719873	0.0719873	0.928013	0.211203	1	16.0652	54	54	0.336425	1	28.3913	50	50
				Mnt/															
89	11	74837472	74837611	ENSMUSG00000000282	+	10	0.0723232	0.927677	0.0723232	0.057143	1	4	63	63	0.007692	0.643411	0.849057	75	75
				Npm2/															
90	14	70647549	70647636	ENSMUSG00000047911	-	10	0.0727771	0.0727771	0.927223	0	0	0	62	62	0.049215	1	4.45161	85	85
				Ttc19/															
91	11	62283707	62284130	ENSMUSG00000042298	+	10	0.0730313	0.0730313	0.926969	0.006374	0.622276	0.958084	97	92	0.053205	1	4.38323	78	77
				Fam193b/															
92	13	55542657	55542783	ENSMUSG00000021495	-	10	0.0733485	0.926651	0.0733485	0.226463	1	44.5	151	151	0.163443	1	32.0417	163	163
				Lck/															
93	4	129556005	129556287	ENSMUSG00000000409	-	10	0.0759715	0.924028	0.0759715	0.043478	0.720588	1.54545	34	32	0.008333	0.411765	0.281818	49	48
				Slc28a1/															
94	7	81169062	81169166	ENSMUSG00000025726	+	10	0.07776	0.07776	0.92224	0	0	0	23	23	0.120752	1	4.39474	31	31
				Dph5/															
				ENSMUSG00000033554/															
95	3	115888277	115888439	known-exon	+	10	0.0778693	0.922131	0.0778693	0.02132	0.743421	1.67742	66	66	0	0	0	91	91
				Cdc42bpg/															
96	19	6315787	6315863	ENSMUSG00000024769	+	10	0.0781546	0.921845	0.0781546	0.069552	1	3.21429	42	42	0	0	0	40	40

97	9	26766072	26767076	Glb1l2/ ENSMUSG00000036395	-	10	0.0787172	0.0787172	0.921283	0.005277	0.466801	0.439698	88	88	0.026449	0.779678	1.73869	63	63
98	10	78182722	78182945	Pwp2/ ENSMUSG00000032834	-	10	0.0788994	0.921101	0.0788994	0.185647	1	5.47126	24	24	0.043956	0.591549	1.6092	35	35
99	7	127933114	127933370	Prss36/ ENSMUSG00000070371	-	10	0.0793608	0.920639	0.0793608	0.092794	1	3.58	35	34	0.03169	0.97561	1.44	44	44
100	3	67355626	67356220	Rsrc1/ ENSMUSG00000034544	+	10	0.0794349	0.0794349	0.920565	0.043336	1	30.1239	642	642	0.061195	1	41.4573	617	617
101	14	55719345	55719429	Rabggta/ ENSMUSG00000040472	-	10	0.0798407	0.0798407	0.920159	0.021828	1	8.03333	360	360	0.03986	1	15.9	383	383
102	3	94484402	94484491	Celf3/ ENSMUSG00000028137	+	10	0.0799223	0.0799223	0.920078	0.436606	1	105.394	133	131	0.522864	1	134.788	122	120
103	5	147356373	147356843	Flt3/ ENSMUSG00000042817	-	10	0.0805198	0.0805198	0.91948	0.017094	0.434783	0.336957	25	23	0.101666	0.836957	1.92391	17	17
104	17	78655659	78655748	Strn/ ENSMUSG00000024077	-	10	0.0807355	0.919265	0.0807355	0.022512	1	2.30303	100	100	0	0	0	134	134
105	8	117453220	117453822	Cmip/ ENSMUSG00000034390	+	10	0.0830894	0.916911	0.0830894	0.012996	0.858108	1.86975	142	142	0.001288	0.239865	0	186	186
106	19	47127109	47127250	Pdcd11/ ENSMUSG00000025047	+	10	0.0833279	0.916672	0.0833279	0.041629	1	6.03774	139	139	0.013423	1	2	147	147
107	4	156016273	156016361	Tnfrsf4/ ENSMUSG00000029075	+	10	0.0843714	0.0843714	0.915629	0.000801	0.025641	0	32	32	0.084577	1	2.125	23	21
108	19	3898638	3898717	Tcirg1/ ENSMUSG00000001750	-	10	0.0849191	0.0849191	0.915081	0.008566	0.449275	0.37931	52	52	0.04918	1	3	58	58
109	14	54638191	54638326	Cdh24/ ENSMUSG00000059674	-	10	0.0860852	0.0860852	0.913915	0.041812	0.48	0.45098	11	11	0.204368	0.984	2.56863	10	10
110	19	45050725	45050875	Sfxn3/ ENSMUSG00000025212	+	10	0.0868991	0.913101	0.0868991	0.061114	1	10.0893	155	155	0.030064	1	5.98214	193	193
111	10	127185172	127185286	Arhgef25/ ENSMUSG00000019467	-	10	0.0876231	0.0876231	0.912377	0.217311	1	33.5952	121	121	0.297581	1	51.2619	121	121
112	2	25288552	25288764	Tmem210/ ENSMUSG00000026963	+	10	0.0886251	0.0886251	0.911375	0.029981	1	58.4146	1732	1732	0.0376	1	70.5976	1673	1673
113	4	12054832	12055030	Tmem67/ ENSMUSG00000049488	-	10	0.0890738	0.910926	0.0890738	0.010997	1	3.94737	355	351	0.002538	0.81383	1	393	388
114	8	105330009	105330094	Fhod1/ ENSMUSG00000014778	-	10	0.089207	0.910793	0.089207	0.098144	1	18.9355	161	161	0.059941	1	13.6452	196	196
115	5	124114645	124114714	Ogfod2/ ENSMUSG00000023707	+	10	0.0903724	0.0903724	0.909628	0.038603	1	20.96	522	518	0.056573	1	34.24	571	562
116	5	137642013	137642095	Sap25/ ENSMUSG00000079165	+	10	0.0907339	0.909266	0.0907339	0.259477	1	29.4333	84	84	0.174392	1	20.0667	95	95
117	8	83316461	83316800	Elmod2/ ENSMUSG00000035151	-	10	0.092368	0.092368	0.907632	0.027322	0.917933	1.48872	53	53	0.079169	1	3.95489	46	46
118	10	79866585	79866667	Plppr3/ ENSMUSG00000035835	-	10	0.0926591	0.0926591	0.907341	0	0	0	13	13	0.185804	1	2.96667	13	12
119	7	126373836	126373921	Spns1/ ENSMUSG00000030741	-	10	0.0934002	0.9066	0.0934002	0.28497	1	24.7097	62	59	0.185938	1	14.1613	62	59
120	7	4551135	4552627	Ptprh/ ENSMUSG00000035429	-	140	0.0945848	0.0945848	0.905415	0.140981	0.987426	6.89299	42	38	0.29037	1	7.36531	18	17
121	11	120284555	120284719	Bahcc1/ ENSMUSG00000039741	+	10	0.0948123	0.905188	0.0948123	0.12	1	3	21	21	0.016064	0.571429	0.677419	35	35
122	7	27212383	27212959	Itpkc/ ENSMUSG00000003752	-	10	0.0957519	0.904248	0.0957519	0.038009	0.980565	4.70175	118	118	0.015485	0.90636	2.2807	145	145

123	11	78190390	78190477	Rab34/ ENSMUSG00000002059	+	10	0.0959955	0.904005	0.0959955	0.040755	1	6.96774	164	164	0.01517	1	2.74194	178	178
124	1	88256549	88256650	Mroh2a/ ENSMUSG00000079429	+	10	0.0964254	0.903575	0.0964254	0.047971	1	52.7568	1014	1013	0.036458	1	40.6757	1047	1046
125	3	89217941	89218027	Thbs3/ ENSMUSG00000028047	+	10	0.0964467	0.903553	0.0964467	0.275951	1	36.9688	97	97	0.204741	1	33.4688	130	130
126	2	25278505	25279171	Anapc2/ ENSMUSG00000026965	+	189	0.0969528	0.0969528	0.903047	0.006907	0.995807	3.34555	479	479	0.014723	1	7.78534	515	515
127	17	49479965	49480429	Daam2/ ENSMUSG00000040260	-	10	0.0973731	0.902627	0.0973731	0.052672	1	9.78571	176	152	0.028375	1	6.95055	238	223
128	9	108978162	108978349	Col7a1/ ENSMUSG00000025650	+	10	0.0978127	0.0978127	0.902187	0.03223	0.632768	0.84507	19	19	0.218635	1	4.19718	14	14
129	19	6954299	6954380	Plcb3/ ENSMUSG00000024960	-	10	0.0979069	0.0979069	0.902093	0.018313	1	2.27586	122	122	0.047132	1	6.03448	122	122
130	13	21319079	21319480	Gpx6/ ENSMUSG00000004341/ known-exon	+	10	0.0989707	0.0989707	0.901029	0.293095	1	332.522	802	795	0.324192	1	331	690	686
131	4	43495621	43495696	Ccdc107/ ENSMUSG00000028461/ known-exon	+	10	0.0994974	0.0994974	0.900503	0.085963	1	26.3333	279	279	0.118811	1	39.3704	291	290

Supplementary Table S2: The IR ratio of individual mouse.

	Intron retention (IR) ratio					
	WT			KO		
Intron-GeneName/GeneID	1	2	3	1	2	3
Pkd1/ENSMUSG00000032855	0.17732	0.154874	0.136886	0.0789474	0.0501157	0.0526316
D15Ertd621e/ENSMUSG00000037119	0	0	0	0.0113636	0.0510204	0.0215545
Larp1/ENSMUSG00000037331	0.0220175	0.0262009	0.0279472	0.0307566	0.0541576	0.0441512
Sptan1/ENSMUSG00000057738	0.0104325	0.0120385	0.0129808	0.00450817	0.00592527	0.00231653
Dgkq/ENSMUSG00000004815	0.125	0.25	0.285714	0	0	0
Stxbp2/ENSMUSG00000004626	0.0170564	0.0231776	0.00979672	0.0508001	0.0313725	0.0236194
Nadk/ENSMUSG00000029063	0.0202703	0.049157	0.0310657	0.0652504	0.0830409	0.0634146
1700067K01Rik/ENSMUSG00000046408	0.0169446	0.055566	0.0247854	0.012987	0.0107527	0
Nckipsd/ENSMUSG00000032598	0.125	1	0.166667	0	0.0385604	0
P4hb/ENSMUSG00000025130	0.013289	0.0120928	0.0151246	0.00650575	0.00310559	0.00347222
1700018F24Rik/ENSMUSG00000029620	0.0459699	0.0672308	0.0569627	0.0756882	0.0780602	0.0673047
Chp2/ENSMUSG00000030865	0.0178341	0.010101	0.00384481	0.058566	0.11593	0.0188983
Cpsf1/ENSMUSG00000034022	0.0307461	0.0302302	0.0517349	0.0060241	0.0171649	0.0218341
Prr19/ENSMUSG00000058741	0.0339031	0.0192893	0.0242816	0.0435914	0.0418167	0.0509045
Myo1a/ENSMUSG00000025401	0.0201842	0.036571	0.0507484	0	0.00990099	0.0171411
Ppp6r2/ENSMUSG00000036561	0.0409836	0.0293542	0.0427689	0.00787402	0.02	0.015655
Cep44/ENSMUSG00000038215	0.0920193	0.115437	0.0919081	0.158045	0.156193	0.151261

Zfp598/ENSMUSG00000041130	0.0923586	0.114346	0.102972	0.135899	0.172741	0.148544
H2-Ke6/ENSMUSG00000073422	0.0381139	0.0839695	0.1	0.120069	0.181818	0.128205
Fam71e1/ENSMUSG00000051113/known-exon	0.0448524	0.0501135	0.0436462	0.0501286	0.061797	0.0611411
Rabggta/ENSMUSG00000040472	0.0711253	0.0717982	0.0566997	0.115032	0.129682	0.110923
P2rx2/ENSMUSG00000029503	0.0969439	0.108816	0.125413	0.058296	0.0666667	0.0758024
Grina/ENSMUSG00000022564	0.0149297	0.0118431	0.0119326	0.00726054	0.0112781	0.00722429
Dock6/ENSMUSG00000032198	0.0427596	0.0416667	0.00374933	0	0.000837872	0
Dennd6b/ENSMUSG00000015377/known-exon	0.133954	0.151293	0.143892	0.0913905	0.113424	0.11441
Arrdc1/ENSMUSG00000026972/known-exon	0.454545	0.374659	0.546218	0.15917	0.312102	0.268859
Lrp10/ENSMUSG00000022175	0.0280833	0.0250696	0.0212353	0.037224	0.0550651	0.0705718
Gnptg/ENSMUSG00000035521	0.032967	0.074518	0.09342	0.095082	0.109198	0.18429
Timm21/ENSMUSG00000024645	0.13149	0.117055	0.0932203	0.136791	0.189542	0.186813
Jmjd8/ENSMUSG00000025736	0.025	0.0474485	0.0212766	0	0.00811542	0
Pnkp/ENSMUSG00000002963	0	0.0117647	0.0121951	0.0294118	0.047619	0.0317623
Tgfb1/ENSMUSG00000002603	0.0301199	0.00857642	0.0153064	0	0.00605578	0
Tomm40l/ENSMUSG00000005674	0.0342052	0.13253	0.105127	0.136986	0.191011	0.250975
Znhit6/ENSMUSG00000074182	0.00511811	0.0234279	0.00519584	0	0	0
Hirip3/ENSMUSG00000042606	0.110133	0.103527	0.0983414	0.0927176	0.08168	0.058611
Dnase1l1/ENSMUSG00000019088	0	0.00180505	0	0.134247	0.166667	0.185185

Ccdc183/ENSMUSG00000026940/known-exon	0.133525	0.145798	0.137327	0.148757	0.188476	0.163934
Cnnm3/ENSMUSG00000001138	0.22451	0.389189	0.489458	0.203205	0.163615	0.1805
Uba1y/ENSMUSG00000069053	0	0	0	0.0408163	0.0188679	0.0140845
Trp53bp2/ENSMUSG00000026510	0.026113	0.0769231	0.112867	0	0	0
Clcn7/ENSMUSG00000036636	0.113194	0.153846	0.086217	0.0144181	0.0338369	0.0828295
Tdp2/ENSMUSG00000035958	0.0159744	0	0.0153846	0.0231362	0.017094	0.0551524
Matk/ENSMUSG00000004933/known-exon	0.2	0	0.194444	0.387324	0.608108	0.416107
Gpsm3/ENSMUSG00000034786	0.126437	0.0322581	0.0739236	0.3	0.198946	0.380098
Tdrkh/ENSMUSG00000041912/known-exon	0.515052	0.479487	0.430983	0.543049	0.602599	0.603426
Ube2d2a/ENSMUSG00000091896	0.0137551	0.0166199	0.0102389	0.00847547	0.00985076	0.00821603
Mast3/ENSMUSG00000031833	0.0111438	0.0421176	0.0454545	0	0.00308642	0
4930519G04Rik/ENSMUSG00000029564	0.158253	0.160534	0.156773	0.177395	0.181514	0.213752
Cad/ENSMUSG00000013629	0.0165639	0.0473961	0.0113636	0	0	0.00506175
Mocs1/ENSMUSG00000064120	0	0	0	0.0798834	0.0666667	0.0311093
Idh3g/ENSMUSG00000002010	0	0	0.00769231	0.0173445	0.00869565	0.0229885
Lck/ENSMUSG00000000409	0	0	0	0.0833333	0.333333	0.0428571
Idh3g/ENSMUSG00000002010	0.00454368	0.00053198	0.000786016	0.00599379	0.0232558	0.0188679
Mrpl28/ENSMUSG00000024181	0	0.00422089	0	0.0138675	0.0178752	0.00433683
Arhgap21/ENSMUSG00000036591	0	0	0.0526316	0.333333	0.0666667	0.125
Rapgef3/ENSMUSG00000022469	0.1	0.0553231	0.0321761	0	0	0.0237028

Tstd3/ENSMUSG00000028251	0.0125	0.0217391	0.0063313	0.000160617	0	0
Uhrf1/ENSMUSG00000001228	0.0631598	0.0523438	0.0464548	0.0338576	0.041069	0.0368486
Pkd1/ENSMUSG00000032855	0.109039	0.185798	0.11556	0.0666667	0.0370121	0.0939597
Fgfr3/ENSMUSG00000054252	0.1	0.0769231	0.111111	0	0	0.00504202
Cdc42bpg/ENSMUSG00000024769	0	0	0	0.027326	0.08	0.0134267
Fanca/ENSMUSG00000032815	0.0277778	0.010266	0.108659	0	0.00407391	0
Nras/ENSMUSG00000027852	0.00849858	0.0156631	0.0147783	0.00527704	0.00543478	0.0058309
Tmem131/ENSMUSG00000026116	0.0195122	0.0094583	0.0153846	0.000243605	0	0.00325733
Adad2/ENSMUSG00000024266	0.101586	0.088451	0.0730838	0.0498584	0.0645679	0.0715222
Dock6/ENSMUSG00000032198	0.000683761	0.00667976	0	0.0357143	0.0294118	0.0384615
Gramd4/ENSMUSG00000035900	0	0.0146628	0	0.5	0.0372493	0.166667
Kdm5d/ENSMUSG00000056673	0	0	0.0109091	0.305785	0.0285714	0.142857
Rpusd4/ENSMUSG00000032044	0.0285714	0.027027	0.0301954	0.110375	0.0490857	0.103448
Tle6/ENSMUSG00000034758	0	0.0211039	0	0.0259166	0.1	0.038565
Ckb/ENSMUSG00000001270	0.0971787	0.121212	0.075	0.0427529	0.0263158	0.0304679
Il17re/ENSMUSG00000043088	0.114754	0.0833333	0.217391	0.0423645	0.08	0.0759358
Prickle4/ENSMUSG00000096549	0.159555	0.266533	0.0944528	0.0637733	0.0895522	0
Ddx51/ENSMUSG00000029504	0.0242424	0.0397862	0.0268456	0.0091257	0.0143369	0.0193078
Dda1/ENSMUSG00000074247	0.0202575	0.0137378	0.011415	0.00824728	0.00956938	0.0109605
Fads3/ENSMUSG00000024664	0	0	0.0163488	0.0577096	0.0428212	0.141809

Matk/ENSMUSG00000004933	0.0174072	0	0.0360634	0.0769089	0.0381166	0.065534
Esyt3/ENSMUSG00000037681	0.0593743	0.0585395	0.037037	0.080939	0.121082	0.113321
Cd209g/ENSMUSG00000079168	0.0225806	0.226415	0.23125	0.705036	0.528736	0.349206
Gpkow/ENSMUSG00000031148	0.0083682	0	0.00521059	0.0243902	0.0212766	0.0175439
Al481877/ENSMUSG00000038598	0.0210617	0.0311251	0.0362164	0	0.012966	0.00735294
Nrn1l/ENSMUSG00000044287	0.2	0.25	0.2	0	0.0778923	0
Nckipsd/ENSMUSG00000032598	0.235294	0.246377	0.43956	0.11194	0.151515	0.111111
St5/ENSMUSG000000310244	0.0437484	0.0397121	0.0393223	0.0208032	0.0371219	0.0201342
Ppp2r5b/ENSMUSG00000024777	0.0415074	0.0454545	0.0688217	0	0	0
Pkd1l3/ENSMUSG00000048827	0	0	0.00711032	0.0161127	0.00781635	0.00809631
Asah2/ENSMUSG00000024887	0.022335	0.171084	0.163889	0	0.0110906	0
Itgad/ENSMUSG00000070369	0.263158	0.2	0.185185	0.380054	0.34944	0.279926
Mnt/ENSMUSG000000002 82	0.054339	0.0914286	0.0221574	0	0	0.0197103
Npm2/ENSMUSG00000047911	0	0	0	0.00668896	0.0381862	0.0716878
Ttc19/ENSMUSG00000042298	0	0.0149506	0.00770264	0.0151033	0.0607927	0.0579165
Fam193b/ENSMUSG00000021495	0.21194	0.231043	0.227981	0.170868	0.158137	0.17414
Lck/ENSMUSG00000000409	0.0465699	0.0442727	0.0374688	0	0.0285714	0.0110725
Slc28a1/ENSMUSG00000025726	0	0	0	0.0309278	0.142857	0.208333
Dph5/ENSMUSG00000033554/known-exon	0.00605473	0.047619	0.0199871	0	0	0
Cdc42bpg/ENSMUSG00000024769	0.0728477	0.133333	0.0104948	0	0	0

Glb1l2/ENSMUSG00000036395	0.00996016	0.00422974	0.0106344	0.0211435	0.0155042	0.0286645
Pwp2/ENSMUSG00000032834	0.232353	0.25	0.1	0.065277	0.0308742	0
Prss36/ENSMUSG00000070371	0.0381955	0.231508	0.071002	0.0324484	0.0327654	0.0249703
Rsrc1/ENSMUSG00000034544	0.0418033	0.0392442	0.0470043	0.0508355	0.0572944	0.0723051
Rabggta/ENSMUSG00000040472	0.0263405	0.0157729	0.0174854	0.0404387	0.0293454	0.05
Celf3/ENSMUSG00000028137	0.418744	0.464239	0.424419	0.563095	0.511563	0.498054
Flt3/ENSMUSG00000042817	0	0	0.0323625	0.071511	0.0935961	0.103896
Strn/ENSMUSG00000024077	0.0285714	0.0156639	0.03125	0	0	0
Cmip/ENSMUSG00000034390	0.0106194	0.00392597	0.0140835	0.00351162	0	0.00171264
Pdcd11/ENSMUSG00000025047	0.0432493	0.0466699	0.0402414	0	0.000520201	0.039707
Tnfrsf4/ENSMUSG00000029075	0	0.00319489	0	0.0468432	0.0346535	0.164179
Tcirg1/ENSMUSG00000001750	0	0	0.0183758	0.0588235	0.0416667	0.05
Cdh24/ENSMUSG00000059674	0	0.0740741	0	0.34555	0.25	0.142857
Sfxn3/ENSMUSG00000025212	0.0401097	0.0685598	0.0705394	0.0299698	0.0298507	0.0298507
Arhgef25/ENSMUSG00000019467	0.201983	0.228864	0.232	0.335897	0.286372	0.28
Tmem210/ENSMUSG00000026963	0.0255347	0.032928	0.0312161	0.0351705	0.0420348	0.0372953
Tmem67/ENSMUSG00000049488	0.00854701	0.0243551	0.00477241	0	0.0044335	0.00366898
Fhod1/ENSMUSG00000014778	0.0937355	0.113311	0.0891434	0.0748609	0.0446841	0.0707434
Ogfod2/ENSMUSG00000023707	0.0461611	0.0443686	0.0281433	0.0491284	0.0657022	0.0566858
Sap25/ENSMUSG00000079165	0.324645	0.256198	0.222222	0.178571	0.160156	0.182331

Elmod2/ENSMUSG00000035151	0.0203474	0.0588235	0.00580201	0.0714286	0.0916031	0.0723627
Plppr3/ENSMUSG00000035835	0	0	0	0.0501319	0.2	0.333333
Spns1/ENSMUSG00000030741	0.23945	0.400455	0.212198	0.2	0.208885	0.133824
Ptprh/ENSMUSG00000035429	0.170549	0.100347	0.177428	0.206829	0.290045	0.288153
Bahcc1/ENSMUSG00000039741	0.0525727	0.246201	0.125	0	0.0366972	0
Itpkc/ENSMUSG00000003752	0.0280843	0.0422683	0.027027	0.0114066	0.0204082	0.0087204
Rab34/ENSMUSG00000002059	0.055176	0.0535957	0.03125	0.0020903	0.0178571	0.022739
Mroh2a/ENSMUSG00000079429	0.039667	0.0610807	0.0416727	0.0343487	0.0395847	0.0346739
Thbs3/ENSMUSG00000028047	0.25641	0.304348	0.251637	0.209581	0.208217	0.208989
Anapc2/ENSMUSG00000026965	0.0074641	0.00292386	0.00975907	0.0108634	0.0201217	0.0108467
Daam2/ENSMUSG00000040260	0.0415029	0.0475908	0.0674353	0.0223711	0.0321516	0.0305867
Col7a1/ENSMUSG00000025650	0.0634921	0.0622517	0	0.209354	0.204036	0.249471
Plcb3/ENSMUSG00000024960	0.0208333	0.00918993	0.0285714	0.0588235	0.0339084	0.0662879
Gpx6/ENSMUSG00000004341/known-exon	0.281335	0.293448	0.300869	0.331383	0.303319	0.348662
Ccdc107/ENSMUSG00000028461/known-exon	0.0849282	0.102095	0.0686275	0.102407	0.123635	0.134031
AVERAGE RATIO	0.068862605	0.088836566	0.076887997	0.083511173	0.08626762	0.084904982

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

Gene editing of the multi-copy H2A.B gene family by a single pair of TALENS

Nur Diana Anuar[1], Matt Field [1, 2, 6], Sebastian Kurscheid[1, 6], Lei Zhang[3], Edward Rebar[3], Philip Gregory[3,4], Josephine Bowles[5], Peter Koopman[5], David J. Tremethick[#1] and Tatiana Soboleva[#1]

[1]The John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia.

[2]Current address: James Cook University, PO Box 6811, Cairns, QLD 4870, Australia

[3]Sangamo Therapeutics, 501 Canal Blvd, Richmond, CA 94804, USA.

[4]Current address: bluebird bio, 60 Binney St, Cambridge, MA 02142, USA.

[5]Institute for Molecular Bioscience, The University of Queensland, Brisbane QLD 4072, Australia

[6]Contributed equally

Corresponding authors:

Tatiana Soboleva, Tel: +61 2 6125 4391; Fax: +61 2 6125 2499; Email: Tanya.Soboleva@anu.edu.au

David Tremethick, Tel: +61 2 6125 2326; Fax: +61 2 6125 2499; Email: David.Tremethick@anu.edu.au

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

In view of the controversy related to the generation of off-target mutations by gene editing approaches, we tested the specificity of TALENs by disrupting a multi-copy gene family using only one pair of TALENS. We show here that TALENS do display a high level of specificity by simultaneously knocking out the function of the three genes that encode for H2A.B.3. This represents the first described knockout of this histone variant.

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

Despite being discovered over a decade ago [1], the *in vivo* importance of the mammalian histone H2A variant H2A.B remains unknown. H2A.B appeared late in evolution in mice (H2A.B.3) and humans (H2A.B), and it is predominantly expressed in the testis with low expression levels in the brain [2,3]. The availability of a H2A.B.3 knockout would be an invaluable tool for investigating the importance of H2A.B, however, as the H2A.B.3 protein is expressed from three different genes, this is a technically challenging feat. Our aim was to devise the most efficient and specific strategy to inhibit the function of all three genes. Here, we report a simplified TALEN approach that achieves this goal.

Previously, transcription activator-like effector nucleases (TALENs) were the choice to perform gene and genome editing [4,5] but more recently this technology has been superseded by the CRISPR-Cas9 system [6]. However, while improvements are continually being made [7-9], one issue has been the extent of off-targets mutants generated with the use of CRISPR-Cas9, which appears to be greater when compared with TALENs [10-13]. One attractive feature of TALENs, which enables a higher level of specificity, is that the FOK1 nuclease domain will only cleave DNA when dimerized, which occurs when the two TALE-domains bind to DNA (on opposite strands) in close proximity to each other. Here, we used TALENs to genetically impair the function of all three H2A.B.3 encoding genes and moreover, we used only one pair of TALENS to do this. To our knowledge, this has not been done before.

There are three H2A.B.3-encoding genes (H2Afb3, Gm14920, H2Afb2, which are >92% identical) plus a pseudogene (Gm14904) all located on the X chromosome in the mouse (this pseudogene was present in the Ensemble release 57 but subsequently removed) (**Supplementary Table 1**). We confirmed the expression of the three H2A.B.3-encoding genes in the testis and as expected, the pseudogene was not expressed (**Supplementary Fig 1**). In order to test the specificity of TALENs against the H2A.B.3 gene family, two groups of H2A.B.3-targeting TALENs were designed. The first group of 12 TALENs were designed to target only one gene, H2Afb3. The second group of 7 TALENs was design to target all three H2A.B.3 genes (**Supplementary Table 2**).

The activity and specificity of the TALENs were tested by employing a Dual-Luciferase Single Strand Annealing Assay (DLSSA) and by the Cel1 cleavage assay. The TALEN plasmids were co-transfected into Neuro2a cells in pairs in various combinations (**Supplementary Fig. 2**). The results for the first group showed that all TALEN pairs, in different combinations, had

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

the highest activity for the H2Afb3 gene, with some TALEN pairs indeed only cleaving the H2Afb3 gene (e.g. 101408:101412). The second group of TALENs showed nuclease activity for all genes, with 2 pairs (101421:101422 and 101421:101423) having the highest activity. Cel1 assays confirmed these findings (**Supplementary Fig. 3**). These *in vitro* results demonstrate the high specificity of TALENs; some of the targeted DNA sequences of the first group, designed against the H2Afb3 gene, differed by only 2-3 base pairs compared to the second group, and yet specificity for only the H2Afb3 gene was achieved (**Supplementary Fig. 4**). The TALEN pair 101421:101422 was used to knock out (KO) the function of all three H2A.B.3 genes in mice (**Supplementary Fig. 4**).

Prior to creating the H2A.B.33 KO, we determined whether any off-target genomic locations, which have limited sequence homology to the TALEN pair 101421:101422, can be identified by using NCBI BLAST against the mouse reference genome mm10 built on the C57BL/6J strain. Even relaxing this BLAST search to only 8 bp out of the 15 bp DNA recognition sequence returned no matches (data not shown).

In order to produce H2A.B.3 KO mice, FVB/NJArc mice (derived in Jackson Laboratories and sourced from Australian Animal Resource Centre (ARC), hence an extra suffix, JArc) were used. Capped, poly-adenylated and purified *in vitro* transcribed mRNA was adjusted to 10 ng/µl for each TALEN pair (101421:101422) and injected into 532 fertilised one-cell embryos and cultured overnight to the 2-cell stage. 153 surviving 2-cell embryos that displayed normal morphology were implanted in pseudo pregnant females, resulting in the birth of 19 pups. Of these 19 pups, 9 contained TALEN-induced H2A.B.3 mutations. Most interestingly, TALENs activity produced an all-or-nothing effect, i.e., for each of the 9 founder pups, all three H2A.B.3 genes were mutated while for the 10 unmodified pups, no gene copies of H2A.B.3 were modified (Table 1). Moreover, in female pups, both alleles of H2Afb3, Gm14920 and H2Afb2 were modified, and each H2A.B.3 gene contained a unique indel as the result of TALEN activity (identified by Sanger sequencing). Therefore, one pair of TALENS can edit 6 homologous alleles simultaneously.

Most of the mutants carried small deletions. Interestingly, in 2 founders, L90-4 and L90-8, at least one of the H2A.B.3 genes (usually gm14920 or H2Afb2) was not amplifiable by PCR. However, when we used mixed primer pairs for PCR amplification (e.g. gm14920-forward and H2Afb2-reverse) we were able to detect an amplification product and following Sanger DNA sequencing, demonstrated that a fusion between gm14920 and H2Afb2 had

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

occurred (data not shown). Also unexpectedly, when the founder L74-5 mouse, which only displayed the deletions and insertions described in Table 1, was bred with a wild-type (wt) female, the resulting G1 progeny displayed a chimera between the gm14920 and H2Afb2 genes (**Supplementary Fig. 5**). Continued breeding from these chimeric G1 H2A.B.3$^{-/x}$ females with wt males (up to 4 generations), revealed that these additional mutations were lost after G2, but maintained original founder mutations (H2Afb3 X$^{\Delta+286}$, gm14290 X$^{\Delta15}$, H2Afb2 X$^{\Delta17}$). This can be explained by the fact that TALEN activity can be retained for several cell divisions following their injection into the 2-cell stage, thus producing a mosaic genotype, as previously observed [14,15]. Collectively, the G1 generation of the 9 founder H2A.B.3 genetically modified mice produced 22/78 mice displaying mosaicism, but by G3 a pure non-mosaic mouse colony was derived.

The strategy for the identification of possible TALEN-induced off-target mutations in H2A.B.3$^{-/y}$ mice was based on the sequencing of exomes of three related non-mosaic H2A.B.3$^{-/y}$ mice from three consecutive generations (G1-G3). This strategy avoids the problem observed in previous studies where unrelated mice were compared to assess off-target effects [12]. Each exome sample was sequenced using 100bp paired-end reads on the Illumina HiSeq 2000 sequencer to a depth of 350 - 600x coverage, yielding between 129x10^6 and 228x10^6 reads per sample (**Supplementary Table 3)**. The breeding began by crossing ♀L90-3 with ♂L74-5 to produce phenotype NM4-G1 H2Afb3 (X$^{\Delta5}$Y), gm14920 (X$^{\Delta10}$Y) and H2Afb2 (X$^{\Delta64}$Y) mice. Subsequent crosses between genetically modified H2A.B.3 female mice and wt males produced NM4-G2 and G3 H2Afb3 (X$^{\Delta5}$Y), gm14920 (X$^{\Delta10}$Y) and H2Afb2 (X$^{\Delta64}$Y) mice for exon sequencing. By crossing mutant females with wt. males any true off-target mutations would be identified because they would be diluted by 50% after each generation. On the other hand, natural sequence variations due to mouse strain differences would not be diluted by breeding (see below). Sequenced mouse exomes were run through our in-house variant detection pipeline (see Online Methods) to detect SNVs, small indels and larger structural variants. Importantly, TALENS do not usually produce SNVs, thus this type of variant serves as an internal measure for mouse strain differences.

As expected, alignment with the mm10 genome identified a large number of SNVs (~1.7x10^6 for G1 and G2, and ~0.69x10x^6 for G3) and indels (~1.7x10^5 for G1 and G2, ~0.6x10^5 for G3) (**Supplementary Table 4**), however neither followed a dilution pattern but rather correlated with the original number of sequencing reads. To remove the effect of

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

mouse strain differences, a filter for FVB/NJ was included into the alignment, which indeed greatly reduced the number of SNVs and indels (**Supplementary Table 5**). Again, the remaining SNVs and putative indels did not follow a dilution pattern following successive generations.

An additional analysis was conducted to separate heterozygous from homozygous indels [16]. As shown (**Supplementary Table 6**), after filtering against the FVB/NJ genome, 417 NM4-G1, 420 NM4-G2 and 325 NM4-G3 heterozygous (0/1+0/2+1/2) indels remained after alignment with the mm10 genome, but again did not display a 50% dilution pattern from G1-G3. To provide further evidence that these remaining indels are due to strain differences between FVB/NJArc and FVB/NJ mice, we interrogated the 19 indels common to all three NM4 G1-G3 mice (**Supplementary Fig. 6**, the location of these putative indels are shown in **Supplementary Table 7**).

We amplified ~300-450 bp regions spanning all 19 indels and compared them to the wt FVB/NJArc mice. The results for all 19 regions (**Fig. 1a**) clearly show that the amplified fragments were identical in size for the H2A.B.3$^{-/y}$ and wt FVB/NJArc mice. Furthermore, we sequenced 5 of these putative indels, which confirmed that the H2A.B.3$^{-/y}$ and wt mice are identical in the nucleotide sequence where the indels were predicted to be (**Fig. 1b and c**). Further, the computational prediction tools are often inaccurate, not only in predicting the size of a putative deletion but also in falsely predicting that a deletion exists (**Fig. 1c** and **d**). Overall, our results suggest that the predicted indels in our H2A.B.3 $^{-/y}$ are due to DNA sequence differences in our in-house FVB/NJArc mouse strain, and computational errors in predicting indels. Finally, to detect structural variants in NM4-G1, NM4-G2 and NM4-G3, analyses using Pindel and Janda were performed. In addition to the three expected deletions in H2A.B.3 genes: H2Afb3 (5 bp), gm14920 (10 bp), a larger deletion of 64 bp in H2Afb2 was also detected **(Supplementary Fig. 7)**.

Despite the fact that H2A.B.3-deficient mice produce neither wt H2A.B.3 mRNA nor protein (**Supplementary Fig. 8a** and **b**), they can obviously reproduce. To investigate whether the loss of H2A.B.3 does affect fertility, we set up a breeding regime whereby age-matched male wt H2A.B.3 and H2A.B.3$^{-/y}$ were mated with 3-6 month female mice and the litter sizes were recorded (**Supplementary Fig. 8c**). The results showed that H2A.B.3$^{-/y}$ mice were subfertile, producing significantly smaller litter sizes and thus demonstrating that H2A.B.3 is required for fertility.

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

In conclusion, a major ongoing question is whether a RNA (e.g. CRISPR)- or protein (e.g. TALEN)-based mechanism will ultimately prove to be more specific for genome editing and this will be critical for future medical therapies. In this study, we have successfully applied the TALEN technology, whereby we used one pair of TALENs to specifically and simultaneously disrupt three gene copies of a gene family. Many human diseases are caused by gene copy number variations [17] and therefore the use of a limited number of genome editing enzymes might help in the future treatment of such diseases. The use of only one pair of TALENs, rather than multiple pairs, reduces the complexity of the TALEN approach (and hence the cost) and would reduce the likelihood of mosaicism and most importantly, reduce the possibility of off-target mutations. Indeed, our rigorous exome analyses did not reveal any TALEN-induced off-target mutations. We also found that the computational tools used to predict putative SNVs and indels have limitations, and therefore wet-lab validation is required before any conclusions can be made about off-target mutations.

Finally, the TALEN approach did expose several interesting phenomena. Firstly, the simultaneous knockout of highly homologous H2A.B.3 genes leads to the formation of chimeras between these genes. As every chimera contained a small deletion, it is likely that terminal microhomology, a mechanism of NHEJ, was involved [18]. It is important to point out that other technologies that rely on NHEJ for inducing mutations, like CRISPR/Cas9 technology, will also likely produce chimeras between highly homologous genes. Several rounds of backcrossing can successfully eliminate such small-scale genomic rearrangements. Secondly, we observed an all-or-nothing effect whereby the TALENs either modified all 3 H2A.B.3 genes or modified none, which reduced the need for a more complex breeding and screening strategy to establish a H2A.B.3 KO colony.

METHODS

Methods are available in the online version of the paper.

Note: Any Supplementary Information is available on the online version of the paper.

ACKNOWLEDGMENTS

We thank Tara Davidson for assistance with microinjections of the TALENs and Dan Andrews for bioinformatics support in analyzing the exome sequencing data. We acknowledge the

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

excellent high-throughput DNA sequencing service provided by our in-house Bimolecular Research Service headed by Stephanie Palmer, and the mouse technicians responsible for the breeding and care of the mice. This work was supported by a NHMRC project grant to D.T. (1058941).

AUTHOR CONTRIBUTIONS

T.S. and D.T conceived the study and wrote the paper. T.S. devised the analyses. M.F. and S.K. performed exome analyses. L.Z., E.R., and P.G. designed the TALENs and performed the Dual-Luciferase Single Strand Annealing Assay (DLSSA) and the Cel1 cleavage assays. J.B. and P.K. generated the TALEN founder mice. N.D.A. performed the screening, genotyping, Cel1 assays, prepared exome-seq samples and produced the KO mice.

COMPETING FINANCIAL INTERESTS

The authors declare no financial interests.

REFERENCES

1 Chadwick, B. P. & Willard, H. F. A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *The Journal of cell biology* **152**, 375-384 (2001).

2 Soboleva, T. A. *et al.* A unique H2A histone variant occupies the transcriptional start site of active genes. *Nature structural & molecular biology* **19**, 25-30, doi:10.1038/nsmb.2161 (2012).

3 Soboleva, T. A. *et al.* A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLoS genetics* **13**, e1006633, doi:10.1371/journal.pgen.1006633 (2017).

4 Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology* **29**, 731-734, doi:10.1038/nbt.1927 (2011).

5 Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* **29**, 143-148, doi:10.1038/nbt.1755 (2011).

6 Kim, J. S. Genome editing comes of age. *Nature protocols* **11**, 1573-1578, doi:10.1038/nprot.2016.104 (2016).

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

7 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology* **34**, 184-191, doi:10.1038/nbt.3437 (2016).

8 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).

9 Shin, J. *et al.* Disabling Cas9 by an anti-CRISPR DNA mimic. *Science advances* **3**, e1701620, doi:10.1126/sciadv.1701620 (2017).

10 Duan, J. *et al.* Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell research* **24**, 1009-1012, doi:10.1038/cr.2014.87 (2014).

11 Kim, D., Kim, S., Kim, S., Park, J. & Kim, J. S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome research* **26**, 406-415, doi:10.1101/gr.199588.115 (2016).

12 Schaefer, K. A. *et al.* Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nature methods* **14**, 547-548, doi:10.1038/nmeth.4293 (2017).

13 Xu, P. *et al.* Both TALENs and CRISPR/Cas9 directly target the HBB IVS2-654 (C > T) mutation in beta-thalassemia-derived iPSCs. *Scientific reports* **5**, 12065, doi:10.1038/srep12065 (2015).

14 Li, C. *et al.* Simultaneous gene editing by injection of mRNAs encoding transcription activator-like effector nucleases into mouse zygotes. *Molecular and cellular biology* **34**, 1649-1658, doi:10.1128/MCB.00023-14 (2014).

15 Sung, Y. H. *et al.* Knockout mice created by TALEN-mediated gene targeting. *Nature biotechnology* **31**, 23-24, doi:10.1038/nbt.2477 (2013).

16 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).

17 de Smith, A. J., Walters, R. G., Froguel, P. & Blakemore, A. I. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenetic and genome research* **123**, 17-26, doi:10.1159/000184688 (2008).

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

18 Lieber, M. R., Gu, J., Lu, H., Shimazaki, N. & Tsai, A. G. Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. *Sub-cellular biochemistry* **50**, 279-296, doi:10.1007/978-90-481-3471-7_14 (2010).

ONLINE METHODS

TALEN Design.

The TALENs were assembled using the method previously described [5,19]. The TALEN target sites were searched using a computer script developed in Sangamo. The four most common repeat variable diresidue (RVD), NI, HD, NN and NG, were used to recognize bases A, C, G and T, respectively. The NK RVD was sometimes used to specify the 3' half repeat T base. TALENs were generated by linking several pre-made TALE repeat blocks in the forms of monomer, dimer, trimer, or tetramer. The desired repeat blocks were PCR amplified using position-specific primers. The PCR products were pooled, digested with BsaI and ligated into the TALEN expression vector, pVAXt-3Flag-NLS-TALE-Bsa3-C63-FokI, which had been linearized with the BsaI restriction enzyme. Ligations were transformed into *E. coli* competent cells, and single clones were picked and sequence verified.

Dual-Luciferase Single Stranded Annealing Assay.

The Dual-Luciferase Single Stranded Annealing Assay (DLSSA) is based on the Dual-Luciferase Reporter Assay System from Promega (Madison, WI). In the DLSSA, the Firefly luciferase reporter gene is split into two inactive fragments with overlapping repeated sequences separated by stop codons and one of the three H2A.B.3 genes. Introduction of a double strand break in the H2A.B.3 gene by the TALEN pair initiates recombination between the flanking repeats by the single stranded annealing pathway and produces an active luciferase gene. Neuro2a cells (ATCC #CCL-131) were cultured in Dulbecco's Modified Eagle Medium (DMEM: Cellgro Cat#: 10-013-CV) plus 10% FBS and 2mM L-Glu (PSG optional antibiotic). One day before transfection, 20,000 cells per well were seeded into a 96-well plate. Cells were transiently transfected using Lipofectamine 2000 (Thermo Fisher Scientific) with four plasmids including a pair of TALENs, the SSA Firefly luciferase reporter,

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

and the internal control Renilla Luciferase reporter (6.25 ng DNA for each component). The luciferase assay was performed 24 hours post-transfection according to manufacturer's instructions. TALEN activity was measured as the ratio of the Firefly and the Renilla luminescence units.

Surveyor Nuclease Assay.

The assay has been described in detail previously [5]. Briefly, mouse Neuro2A cells (2 x 10^5) were seeded into a 96-well plate the day before transfection. The cells were then transfected with a pair of TALEN DNAs (400ng of each) using Lipofectamine 2000 (Thermo Fisher Scientific). The genomic DNA was purified 48 hours post-transfection. Gene specific primer pairs were used to amplify each of the H2A gene variants. The primer sequences are: H2Aafb3 forward (5'-CAGCAGAAAGCAGCCAAGTGG) and reverse (5'-GCAGGTCAGCCAAGAAGCA); Gm14920 forward (5' GTACGGTACAAAGGGAG ATG) and reverse (5'-GAGCAGGTCAGCCAAGCAGAG); H2afb2 forward (5'-CAGGTCAGCAGAGAGCAATT) and reverse (5'-CTCCATACTGCTGTAGACCT); and pseudo Lap forward (5'-GTCAGCAGAATGCAGCCAAATAT) and reverse (5'-CAAGCCAGTAGCCAACATCAAG). The PCR products were treated with Surveyor Nuclease (Cel-1) and resolved on PAGE. The nuclease activity of TALENs was measured by quantifying the proportion of the cleaved DNA fragments.

Talen mRNA synthesis.

Plasmids (in Sangamo backbone) were linearized with *XbaI* and the products were purified (PCR Cleanup kit, Qiagen). Using linearized plasmid as template, capped and poly(A)-tailed mRNA transcripts were produced using mMESSAGE mMACHINE T7 Ultra Transcription Kit (Life Technologies) following manufacturer's instructions. mRNA was purified using a MEGAclear kit (Life Technologies) and was eluted in RNase-free water and single-use aliquots were frozen.

Production of Talen-mediated genetically modified mice.

All mouse work was conducted according to protocols approved and authorised by the University of Queensland Animal and the Australian National University Ethics Committee (Protocol A2014/33). Prior to injection, mRNA was diluted to 10 ng/µl in filtered RNase-free

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

microinjection buffer (10 mM Tris-HCl, pH 7.4, 0.25 mM EDTA). TALEN mRNA was delivered into the pronucleus of single-cell embryos (FVB), using standard techniques. Injected embryos were cultured overnight to the two-cell stage, then surgically transferred into oviducts pseudopregnant CD1 mice, using standard techniques.

Analysis of potential founders and genotyping of successive generations.

Transgenic founders with mutated H2A.B.3 genes were identified among live born mice by PCR (**Supplementary Table 8**) genotyping of ear notch tissue. Genomic DNA was extracted from ear punches using Quick Extract Solution (Epicentre) following manufacturer's instructions. The H2afb3, gm14920 and H2afb2 genes were amplified in 25µl volume using the same gene-specific primers as for the Cel1 assay. PCR mix (1x Platinum Taq buffer, 1.5mM Mg2Cl; 0.2mM dNTP; 0.2µM of forward and reverse primer; 2 units of Platinum Taq polymerase (Thermo Fisher)) was amplified using 28 cycles of: 95°C 30s denaturation, followed by 56-65°C, 30s annealing and 72°C, 4min extension.

Exome Sequencing and bioinformatic analyses.

The male mice that were chosen for exome sequencing were genotyped prior to sequencing to confirm that all three generations had the same H2A.B.3 genotype (H2Afb3 ($X^{\Delta 5}Y$), gm14920 ($X^{\Delta 10}Y$) and H2Afb2 ($X^{\Delta 64}Y$)). The purified gDNA was subjected to exome enrichment using the Agilent SureSelect^{XT2} All Exon Kit. The exome capture efficiency was uniformly high with approximately 45-50 % of all DNA sequenced being exonic. Sequenced mouse exomes were analyzed using our in-house variant detection pipeline described in detail previously [20,21]. In short, reads were aligned to the mouse reference genome mm10 using BWA [22] and sorted BAM alignment files were generated using SAMTools [23]. Candidate PCR duplicate reads were annotated using Picard software (http://broadinstitute.github.io/picard/), specifically the MarkDuplicate command. SNVs and small indels were called using SAMTools Mpileup [24] and variants were annotated using the Variant Effect Predictor [25]. To detect larger structural variants, both Pindel [26] and Janda (unpublished; https://sourceforge.net/projects/janda/) were utilised. To reduce the number of total variants called, variants specific to the FVBN mouse strain were downloaded from the Sanger Institute ftp site (ftp://ftp-mouse.sanger.ac.uk/REL-1206-FVBNJ/). As these

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

variants were originally reported relative to reference genome mm9, the UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver) was used to map the FVBN variants to mm10. These FVBN specific variants were annotated and not considered for further off-target analysis.

For the detection of INDELs in the TALEN mouse exome sequencing data, we followed the GATK recommendations "Best Practices for Germline SNP & Indel Discovery in Whole Genome and Exome Sequence" [16]. Briefly, the alignment files in BAM format were realigned and base re-calibration was performed. The cleaned alignments of all sequenced mouse exomes were then used as input for the "HaplotypeCaller" program resulting in raw variants in VCF format, containing both SNVs and INDELs. We then performed hard filtering of the called variants to only include INDELS and applied a quality cut-off score of 100 to include only high-confidence INDEL calls. In order to identify potential TALEN off-targets, we inspected the frequency of INDELs in each generation of TALEN mice, expecting a "dilution" effect due to back-crosses.

In order to search the mm10 reference genome for potential off-target sites, we used the *in silico* PCR function of the UCSC Genome Browser at http://genome.ucsc.edu/cgi-bin/hgPcr [27]. We used the TALEN target sequences as forward and reverse primers, and relaxed the minimum perfect match parameter to 8 nucleotides (nt) and constrained the maximum product size to 100bp.

Experimentally validation of putative indels.

To interrogate whether the 19 putative indels are present in NM4 G1-G3 mice, the gDNA of wild type FVB/NJArc in-house mice and NM4 G1-G3 KO mice were amplified with primers surrounding the predicted indels, listed in Supplementary Table 9. The PCR products were analyzed by electrophoresis (7% acrylamide gel) and Sanger sequencing.

19 Miller, J. C. *et al.* Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nature methods* **12**, 465-471, doi:10.1038/nmeth.3330 (2015).

20 Andrews, T. D. *et al.* Massively parallel sequencing of the mouse exome to accurately identify rare, induced mutations: an immediate source for thousands of new mouse models. *Open biology* **2**, 120061, doi:10.1098/rsob.120061 (2012).

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

21 Field, M. A., Cho, V., Andrews, T. D. & Goodnow, C. C. Reliably Detecting Clinically Important Variants Requires Both Combined Variant Calls and Optimized Filtering Strategies. *PloS one* **10**, e0143199, doi:10.1371/journal.pone.0143199 (2015).

22 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

23 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

24 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).

25 McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-2070, doi:10.1093/bioinformatics/btq330 (2010).

26 Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871, doi:10.1093/bioinformatics/btp394 (2009).

27 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome research* **12**, 656-664, doi:10.1101/gr.229202. Article published online before March 2002 (2002).

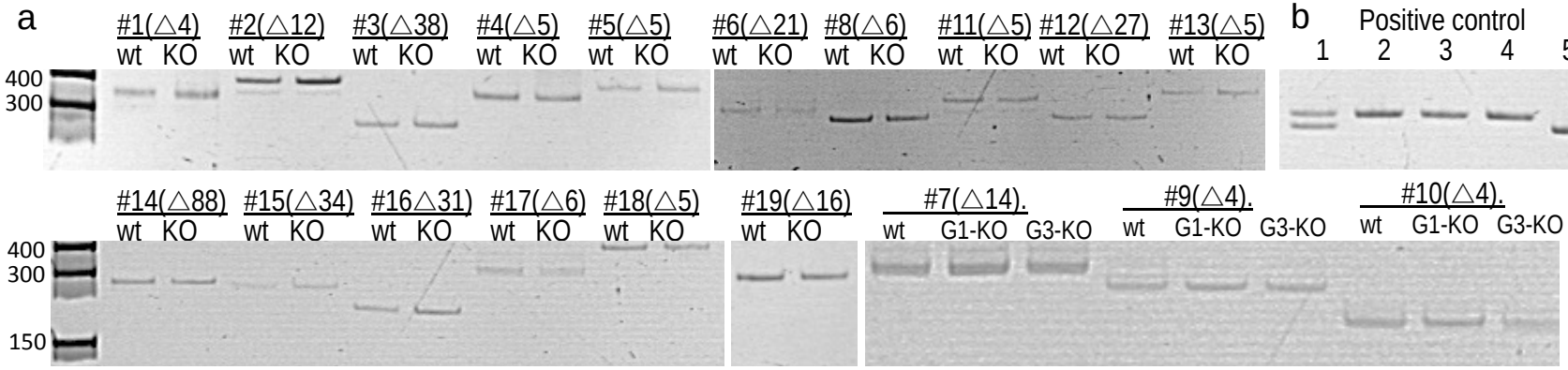
FIGURES

Figure 1. Interrogation of predicted off-target heterozygous deletions in H2A.B.3.3$^{-/y}$ mice. (**a**) 300-400bp regions surrounding 19 common predicted heterozygous deletions were amplified from gDNA of H2A.B.3.3$^{-/y}$ KO and FVB/NJArc in-house wt mice. PCR product sizes for wt and H2A.B.3.3$^{-/y}$ were compared by resolving the products on a 7% polyacrylamide gel, side by side. The predicted deletion is indicated with Δ, followed by the number of nucleotides (nt) predicted to be deleted. (**b**) Positive controls showing that a deletion as small as 5-10 nt can be resolved on 7% polyacrylamide gel. Lane 1, a heterozygous Δ5nt,

bioRxiv preprint first posted online Dec. 13, 2017; doi: http://dx.doi.org/10.1101/233379. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

lanes 2, 3, 4 wt.; lane 5, homozygous △10nt. (**c**) 5 out of the 19 PCR products were sequenced by Sanger sequencing to show that FVB/NJArc and H2A.B.3.3$^{-/y}$ mice have identical sequences but they are different from the mm10 reference genome. (**d**) Table showing that the prediction tools are often inaccurate at predicting the presence or the size of a putative deletion.

a



b

c

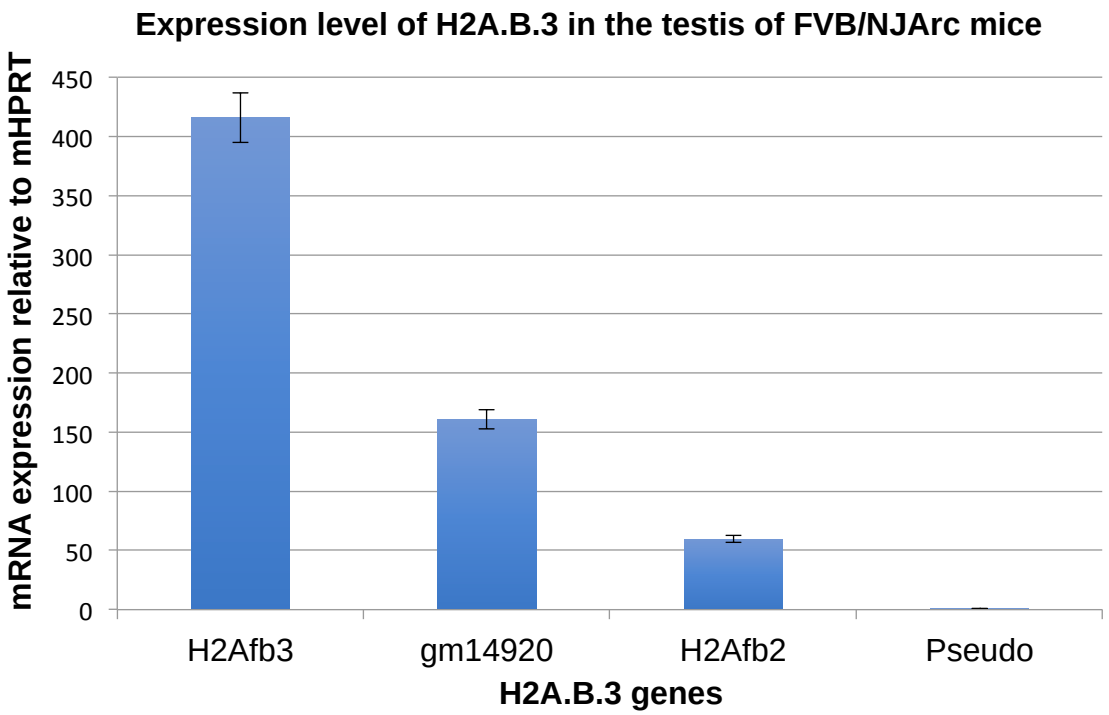
```
Deletion #1
mm10 CTGAAGGCCCAGTGGAAGCCCTGGACTAATGGGACTTGTCAACTGTTTTTGTCCTGCTT
wt   CTGAAGGCCCAGTGGAGAC----------------TTGTCAACAGTTTTTGTCCTGCTT
KO   CTGAAGGCCCAGTGGAGAC----------------TTGTCAACAGTTTTTGTCCTGCTT
     ***************** *                ******** ***************

Deletion #3
mm10 AGTCCCCGCCTGAGCTGTGGAAGTCCTCGCCTGAGCTGTGGAAGTCCCCGCCTGAGCTGTG
wt   AGTCCCCGCC-----------------------------------------TGAGCTGTG
KO   AGTCCCCGCC-----------------------------------------TGAGCTGGG
     **********                                         ******* * *********

Deletion #16
mm10 TGGCTTTGAGGTGTCTGCCTCCATGGGTGCTGGCTGCGAGGTGTCTGCCTCCATGGGTGCTGGCTGCGAGGTCTCTGCCTCCATG
wt   TGGCTATGAGGTG-------------------------------------------------------------TCTGCCTCCATG
KO   TGGCTTTGAGGTG-------------------------------------------------------------TCTGCCTCCATG
     ***** *******                                                             ************

Deletion #14
CGCTCACCGGGGCATCCTGCTGTAGGTGGTGGTGCAACAGCTGGGAGTGGGGCTGCTG
CGCTCACCGGGGCATCCTGCTGTAGGTGGTGGTGCAACAGCTGGGAGTGGGGCTGCTG
CGCTCACCGGGGCATCCTGCTGTAGGTGGTGGTGCAACAGCTGGGAGTGGGGCTGCTG
************************************************************

Deletion #15
TTTGGAGAGCTCTGAACTCTGGCCATTCAACACTTTGGAGAGCTCTGAACTCTGGCC
TTTGGAGAG--------------------------------CTCTGAACTCTGGCC
TTTGGAGAG--------------------------------CTCTGAACTCTGGCC
                                           ***************
```

d

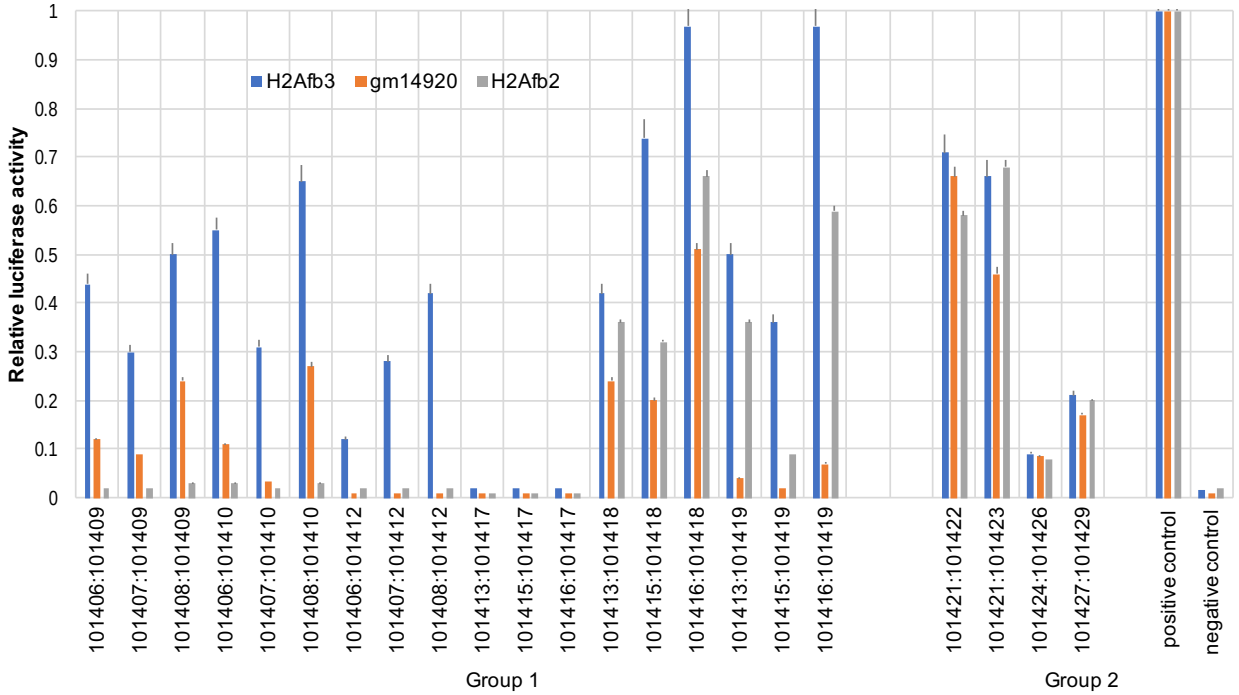
Deletion	**#1**	**#3**	**#14**	**#15**	**#16**
Predicted	△4	△38	△88	△34	△31
Observed in FVB/NJArc and H2A.B.3$^{-/y}$ KO	△16	△42	Nil	△33	△60

#	Pup ID	Genotype by Sanger sequencing		
		H2Afb3	**Gm14920**	**H2Afb2**
1	♀ L74-1	XX (wt)	XX (wt)	XX (wt)
2	♀ L74-2	XX (wt)	XX (wt)	XX (wt)
3	♀ L74-3	XX (wt)	XX (wt)	XX (wt)
4	♂ L74-4	XY (wt)	XY (wt)	XY (wt)
5	♂ L74-5*	**$X^{\Delta 5\ +286}Y$**	**$X^{\Delta 15}Y$**	**$X^{\Delta 17}Y$**
6	♀ L79-1	**$X^{\Delta 10}X^{\Delta 10}$**	**$X^{\Delta 12}X^{\Delta 12}$** Chimera of b/Δ/c	**$X^{\Delta 15}X^{\Delta 15}$** Chimera of c/Δ/b
7	♀ L89-1*	**$X^{\Delta 22}X^{\Delta 160}$**	**$X^{\Delta 5}X^{\Delta 17}$** Chimera c/Δ/b)	**$X^{\Delta 5}X^{\Delta 5}$**
8	♂ L89-2	**$X^{\Delta 65}Y$**	**$X^{C/T,\ A/G}Y$**	$X^{\Delta 42}Y$ Chimera b/Δ/c)
9	♂ L89-3	**$X^{\Delta 26}Y$** Chimera c/Δ/a)	**$X^{\Delta 10}Y$** Chimera b/Δ/c)	**$X^{\Delta 10}Y$** Chimera b/Δ/c)
10	♀ L90-1	XX (wt)	XX (wt)	XX (wt)
11	♀ L90-2	XX (wt)	XX (wt)	XX (wt)
12	♀ L90-3*	**$X^{\Delta 28}X^{\Delta 5}$**	**$X^{\Delta 10}X^{\Delta 10}$**	**$X^{\Delta 64}X^{\Delta 64}$**
13	♀ L90-4	**$X^{\Delta 15}X^{\Delta 10}$**	No successful PCR amplification	No successful PCR amplification.
14	♀ L90-5	XX (wt)	XX (wt)	XX (wt)
15	♀ L90-6	**$X^{\Delta 17}X^{\Delta 17}$** Chimera c/Δ/a)	**$X^{\Delta 5}X^{\Delta 5}$** Chimera b/Δ/c)	**$X^{\Delta 5}X^{\Delta 5}$** Chimera c/b/Δ/c)
16	♀ L90-7	XX (wt)	XX (wt)	XX (wt)
17	♂ L90-8	**$X^{\Delta 16}Y$**	**$X^{\Delta 17}Y$** Chimera c/Δ/c/b)	No successful PCR amplification.
18	♂ L90-9	XY (wt)	XY (wt)	XY (wt)
19	♂ L90-10	XY (wt)	XY (wt)	XY (wt)

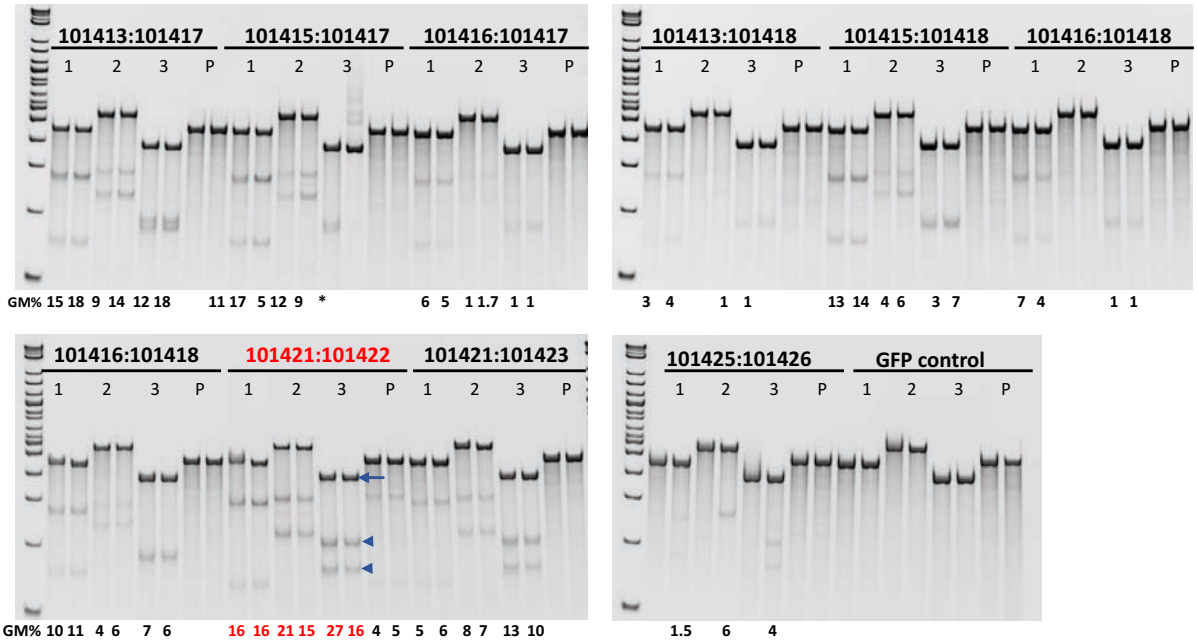
Table 1. Genotype of all 19 pups born following TALEN injections. Wild type (Wt.) offspring. Δn,deletion of n nucleotides; +n, insertion of n nucleotides. * denotes mice that were used as founders to establish H2A.B.3KO colonies.



Supplementary Figure 1: The relative expression of H2A.B.3 gene family members in the testis of FVB/NJArc mice. Gene expression was assessed relative to mHPRT using qPCR.

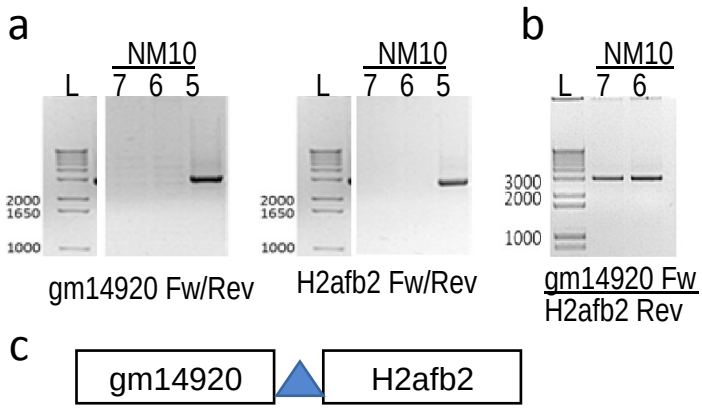


Supplementary Figure 2. TALEN pair specificity tested by the DLLS assay. TALEN pairs designed to target the H2Afb3 gene only (Group 1) or all three H2A.B.3 genes (Group 2). One day before transfection, 20,000 cells per well were seeded into a 96-well plate. Cells were transiently transfected with four plasmids including a pair of TALENs, the SSA Firefly luciferase reporter, and the internal control Renilla Luciferase reporter. The luciferase assay was performed 24 hr post-transfection. The pair 101421:101422 was used to create the H2A.B.3 KO mouse.

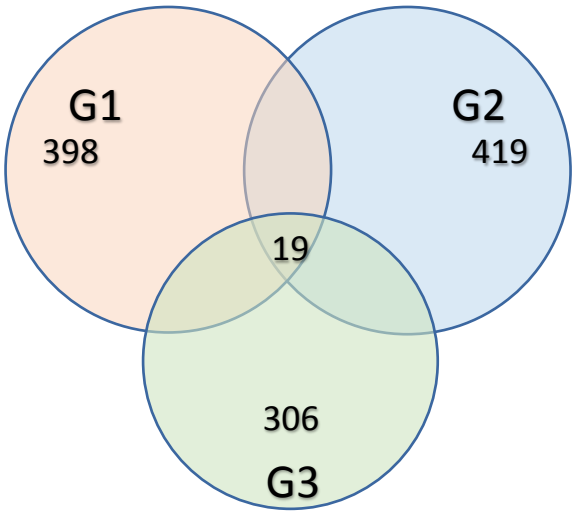


Supplementary Figure 3. Cel 1 assays**.** Neuro2a cells were transiently transfected with various TALEN pairs and extracted genomic DNA was amplified by PCR with gene-specific primers and subjected to Cel1 digestion. TALEN activity is represented by the % cleaved (% genomic mutation, GM) shown below each lane. Cleaved DNA products (arrow heads), uncut DNA (arrow). TALEN pair 101421:101422, highlighted in red, showed the highest activity for all three H2A.B.3 genes. 1,2,3 and P denotes H2Afb3, gm14920, H2Afb2 and pseudo H2A.B.3 genes, respectively. Two biological replicates are shown. Empty GFP-vector was used as a negative control.

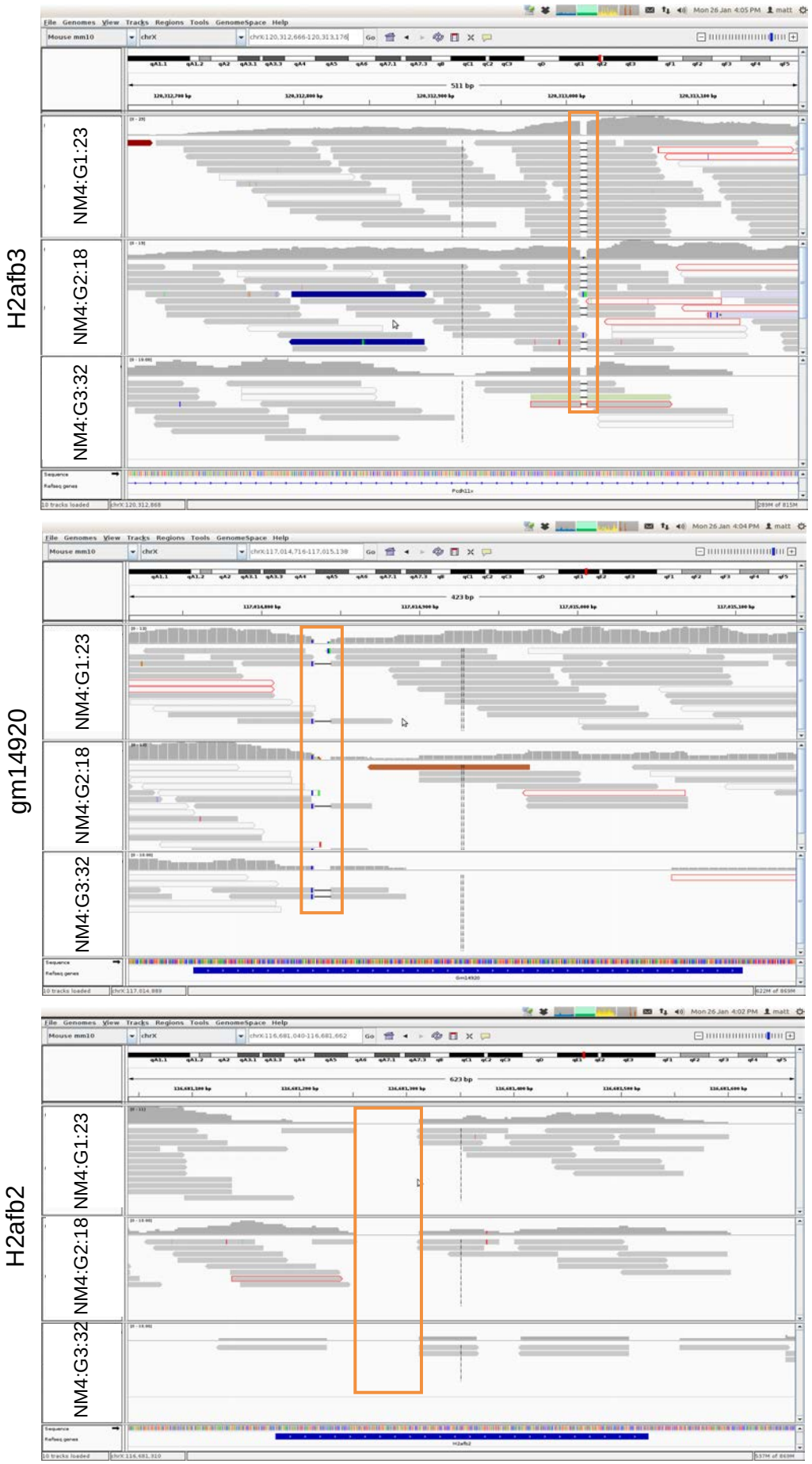
Supplementary Figure 4. TALEN-targeted DNA sequences. The alignment of H2A.B.3 genes and the pseudogene. TALENs specific to H2afb3 are in blue and those common for all H2A.B3 genes are in red.



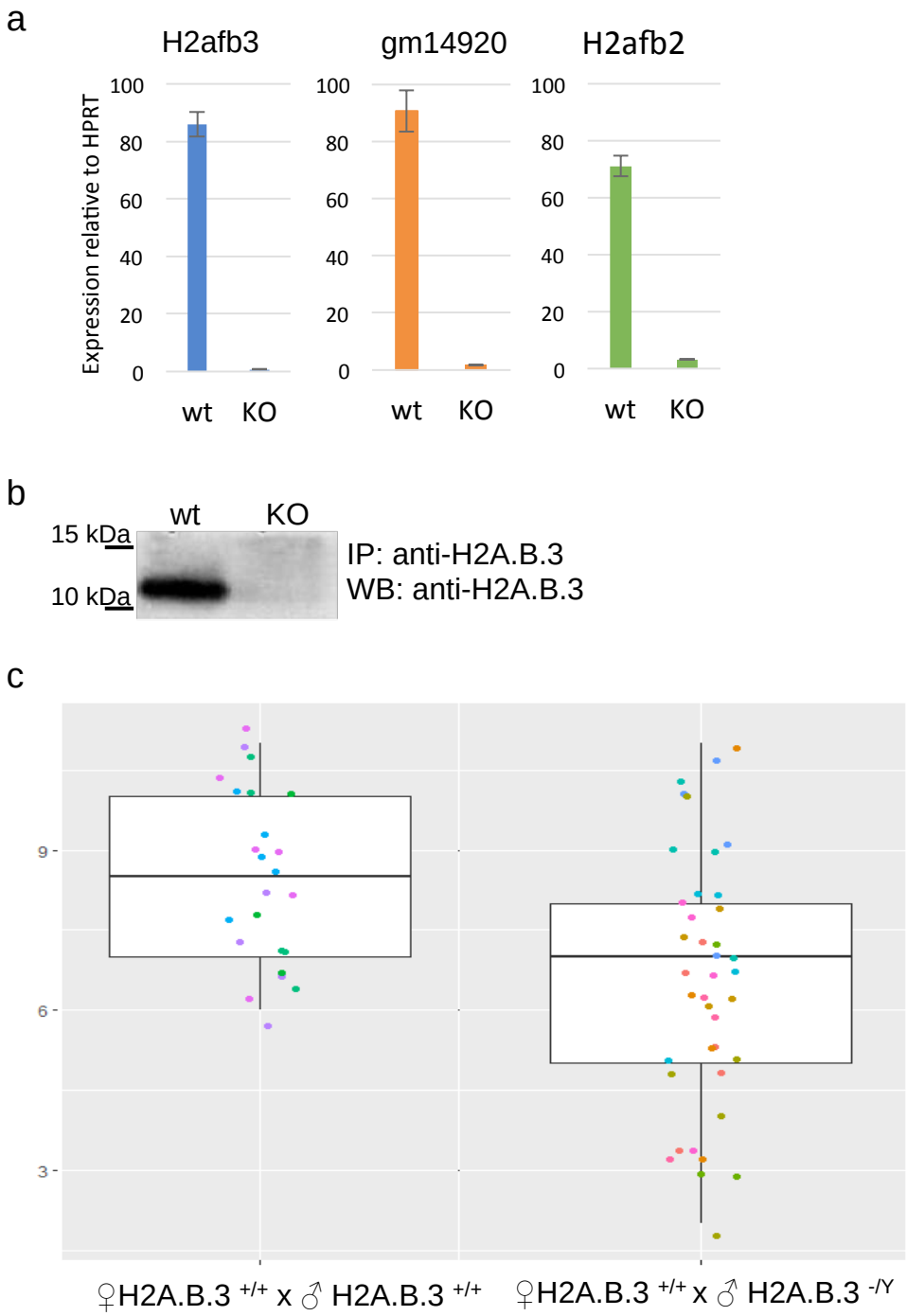
Supplementary Figure 5. An example of chimeras formed between gm14920 and H2afb2 in some G1 mice produced by crossing founder L74-5 with a wt. female. (**a**) genes gm14920 and H2afb2 are not amplified with gene-specific primers in pups #6 and #7, respectively but is amplified in a pup #5. (**b**) amplification with gm14920-Fw and H2afb2-Rev primers shows the chimeric product in pups #6 and 7. (**c**) A schematic diagram showing that a chimera was formed between gm14920 and H2afb2 genes with a small deletion between these fused genes suggesting that NHEJ mechanisms were involved. Fw, forward primer; Rev, reverse primer; number; L, DNA ladder.



Supplementary Figure 6: Computational analysis predicted 19 putative heterozygous deletions shared between all 3 generations of H2A.B.3$^{-/y}$ KO mice. If TALENs introduced off-target mutations in G0 founder mice, then those mutations would be inherited by their progeny as heterozygous.



Supplementary Figure 7. Exome sequencing reveals all the three expected mutations in H2A.B.3 genes. The screen shots are the output from SVs call by Pindel for all H2A.B.3 genes. The orange border box indicates the deleted region detected for each gene by SAMtools.



Supplementary figure 8. H2A.B.3$^{-/Y}$ KO males have impaired fertility. (**a**). The mRNA level of each H2A.B.3 gene was determined by real-time quantitative PCR, relative to HPRT. Primers were designed to anneal within the TALEN induced deletions. (**b**). Confirmation that H2A.B.3 protein is not produced in in H2A.B.3$^{-/Y}$ mice. H2A.B was immunoprecipitated from total testis lysates followed by Western Blotting using anti-H2A.B.3 polyclonal antibodies. (**c**) The breeding was conducted using age matched wt females (3-6 months) and age matched wt (n=5) and H2A.B.3$^{-/Y}$ (n=5) males. Total number of pups produced was 201 and 256 for wt and KO males, ***p value=≤0.001(ANOVA test).

Gene names/length	Strand	Chromosome	Start	End	% Identity
H2afb3, 348bp ENSMUSG00000083616	-	X	117426357	117426704	100
Gm14920, 348bp ENSMUSG00000067441	+	X	114128366	114128713	98.28
H2afb2, 349bp EG624153	+	X	113794787	113795134	92.55
Gm14904(Pseudo), 345bp ENSMUSG00000084053	+	X	113311798	113312139	87.83

Supplementary Table 1. The genomic coordinates of H2A.B.3 coding genes. The degree of homology between the genes is indicated as % Identity.

Target genes	Design #	Targeted sequence
H2Afb3 (Group 1)	101406	ttCAGGTCGCCGCCACCGTcg
	101407	gtCGCCGCCACCGTCGCTcc
	101408	gtCGCCGCCACCGTCGCTCcc
	101409	ctCACAGCAAAGATTAGCTca
	101410	ctCACAGCAAAGATTAGCtc
	101412	ctCACAGCAAAGATTag
	101413	ctGAGCTAATCTTTGCTgt
	101415	ctAATCTTTGCTGTga
	101416	ctAATCTTTGCTGTGAgc
	101417	ctAACCTCCCTCAGATGCTgt
	101418	ctAACCTCCCTCAGATgc
	101419	ctCCGGGCACGGCTAACCTcc
H2Afb3, Gm14920, H2afb2 (Group 2)	101421	ctCCCGCACCTCCAGAGct
	101422	ctGTTCCACCAGGCTCACag
	101423	ctGCTGTTCCACCAGGCTca
	101424	ctGGTGGAACAGCATCTga
	101426	gtATCACTGAGCCTCCgg
	101427	gtTGCTGGAGCTTgc
	101429	gtACCTCTGCGTTcc

Supplementary Table 2. List of TALENs designed to target H2A.B.3 genes.

Sample ID	Total reads	Read length (bp)	Approximate coverage (x)
NM4-G1-28	226,900,000	100	613
NM4-G2-18	228,900,000	100	619
NM4-G3-32	129,800,000	100	351

Supplementary Table 3. The exome sequencing coverage. 3 consecutive generations of mice from the NM4 H2A.B.3 KO colony were sequenced using paired-end sequencing with a 100bp read length.

Sample ID	SNVs	Small deletions	Small insertions
NM4-G1-28	1,745,670	82,379	90,762
NM4-G2-18	1,768,957	81,956	90,005
NM4-G3-32	693,686	33,823	38,321

Supplementary Table 4. Putative SNVs and Indels identified in H2A.B.3 KO mice using the mm10 mouse genome as a reference.

Sample ID	SNVs	Small deletions	Small insertions
NM4-G1-28	21,256	3,889	4,378
NM4-G2-18	26,646	4,064	4,445
NM4-G3-32	13,417	1,589	2,032

Supplementary Table 5. Putative SNVs and Indels identified in H2A.B.3 KO mice after applying the FVB/NJ strain filter.

Sample ID	Genotype							
	Homozygous			Heterozygous				Total
	0/0	1/1	2/2	0/1	0/2	1/2	No Call	
NM4-G1-28	131	732	2	388	2	27	111	1393
NM4-G2-18	127	727	3	392	3	25	116	1393
NM4-G3-32	219	527	8	306	4	15	314	1393

Supplementary Table 6. Predicted putative homozygous and heterozygous polymorphic variants (off-target deletions) in three consecutive generations of H2A.B.3$^{-/y}$ mice. Exome data was referenced to mm10 genome, followed by filtering of FVB/N-specific variants and analysed for genome polymorphism using Pindel tool. Homozygous polymorphisms: no polymorphic alleles (0/0), both alleles differ from reference (1/1), and both alleles differ from reference and from 1/1 (2/2). Heterozygous polymorphisms: one allele differs from the reference genome (0/1), one allele defers from the reference genome and from 0/1 (0/2), both alleles differ from the reference genome and from each other (1/2). No call, not assigned to any type.

#	Chromosome	Locus start	Locus end	Deletion size (bp)	Gene ID
1	1	171573961	171573965	4	ENSMUSG00000004709
2	1	173270636	173270648	12	ENSMUSG00000049605
3	1	173760152	173760189	38	ENSMUSG00000073489
4	12	115158545	115158549	5	ENSMUSG00000093894
5	12	115335195	115335199	5	ENSMUSG00000095197
6	12	115808522	115808526	21	ENSMUSG00000091087
7	12	115833994	115834007	14	ENSMUSG00000096020
8	17	23882468	23882474	7	ENSMUSG00000096445
9	17	35266213	35266218	6	ENSMUSG00000073411
10	17	36189406	36189415	10	ENSMUSG00000054128
11	17	48145527	48145531	5	ENSMUSG00000073386
12	2	119618224	119618250	27	ENSMUSG00000072980
13	3	7604075	7604079	5	ENSMUSG00000040329
14	4	41195246	41195333	88	ENSMUSG00000028433
15	4	42871823	42871856	34	ENSMUSG00000050141
16	5	113819659	113819689	31	ENSMUSG00000048163
17	7	8245023	8245027	5	ENSMUSG00000053720
18	7	47981089	47981093	5	ENSMUSG00000067173
19	X	95940663	95940678	16	ENSMUSG00000057421

Supplementary Table 7. 19 common putative heterozygous polymorphisms identified in all 3 generations of H2A.B.3$^{-/y}$ mice.

H2A.B gene family	Sequence (5'-3')	Size of amplicon
H2A.Afb3	Fwd: CAGCAGAAAGCAGCCAAGTGG Rev: GCAGGTCAGCCAAGAAGCA	440 bp
Gm14920	Fwd: GTTGGGCATTGGACTTGGAC Rev: CAGCCAAGTCCAGCAGTTC	398 bp
H2Afb2	Fwd- CAGGTCAGCAGAGAGCAATT Rev- CTCCATACTGCTGTAGACCT	374 bp

Supplementary Table 8. PCR primers for genotyping of H2A.B.3 encoding genes in founders and the H2A.B3KO colonies.

# Indel	Gene ID	Sequence (5'-3')	Expected product length (bp)
1	ENSMUSG00000004709 (Δ4)	F- AAGAAGACAGAACAGGGCTCAC R- TAGCTCACATTGTCATCCTTGG	419
2	ENSMUSG00000049605 (Δ12)	F- GACAGTGTACACATTGGTCATT R- TGGAGATGATGAGAATGTAAGA	458
3	ENSMUSG00000073489(Δ38)	F- TTAATGCCACATGATCAGTTAT R- CCTAGACCATCTATTTTGGTTT	426
4	ENSMUSG00000093894 (Δ5)	F-AAGTCTACAAGCAAACCTTCTC R-CATATACATGGGTGACAATGAC	452
5	ENSMUSG00000095197 (Δ5)	F-GCTGTATCTGAGCAAGTCTACA R-ATACACTTGGGTGACAATGAC	464
6	ENSMUSG00000091087 (Δ21)	F-TAGATGGACATTTACAAAGCTG R-TATCCTCACCTGTTCCAAAC	412
7	ENSMUSG00000096020 (Δ14)	F-TAGATGGACATTTACAAAGCTG R-GCTACACCTTCACAAGCTACTA	402
8	ENSMUSG00000096445 (Δ6)	F-GTTGTGATGTCTTCTGTGCTC R-GTACTCAGAACTGCCCACATA	405
9	ENSMUSG00000073411 (Δ5)	F-GGAAGGAGCAGAATTACACAT R-GTTCAAGGAAGATCTTGACACT	468
10	ENSMUSG00000054128 (Δ10)	F-GTGCACTCCTGTGTCTTTTT R-GCTGGACTACTACAACCTGAGT	451
11	ENSMUSG00000073386 (Δ5)	F-TTCTTGAGTCCTCTTCAACTCT R-CTCTCTGACAGGAAGCATTTAT	413
12	ENSMUSG00000072980 (Δ27)	F-ACTCGGGAAAACAAAACAAAC R-AAAGAAATATGGCGACTCTCT	418
13	ENSMUSG00000040329 (Δ5)	F-CAATGAAAGGGGTATAGTTCAT R-AATTGTGAGAAGTTTCATTGTG	445
14	ENSMUSG00000028433 (Δ88)	F-CAGGGTCTAGTTTGTCCAGTAT R-TTGCCATCTTTATCCTACAGAC	401
15	ENSMUSG00000050141 (Δ34)	F-AGGAGCTGACAGAGTCAATTT R-GACGACTCCTGTGCTCAGAT	401
16	ENSMUSG00000048163 (Δ31)	F-CAGACGTTGTAAACAGAGTTTT R-ACACAGTGGAGTCTAACCTCAG	409
17	ENSMUSG00000053720 (Δ5)	F-CAGACACTGAAGAACACTAGCA R-TCACATTTGGAATTGTATTCAC	409
18	ENSMUSG00000067173 (Δ5)	F-CAGGTAGAGGCTCTTCATCA R-ACCCGATGTTTTTGTTTATAGT	418
19	ENSMUSG00000057421 (Δ16)	F-TCTACATCATCACTTTCCTCAA R-GCGAGTCTGTTCCATTTATACT	411

Supplementary Table 9: List of primers to amplify 19 putative Indels from TALEN mutant mice. Symbols: Δ, deletion.