Revisiting the Role of Sperm Protamine Proteins in Organismal Development and Fertility

by

Lindsay Elizabeth Moritz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2022

Doctoral Committee:

        Assistant Professor Saher S. Hammoud, Chair
        Assistant Professor Jacob Kitzman
        Assistant Professor Kaushik Ragunathan
        Professor Yukiko M. Yamashita, Massachusetts Institute of Technology

Lindsay Elizabeth Moritz

moritzle@umich.edu

ORCID ID: 0000-0001-6704-8453

Dedication

To my grandmother Peggy, without whose unwavering encouragement and unconditional support I would not be where I am today. I firmly believe you gave me two of the greatest gifts: an insatiable love for learning and the empowerment to believe I'm capable of anything I set my mind to, and for that I am forever grateful.

Acknowledgements

To all of the incredible mentors I had at Miami University - Drs. Michael Crowder, Rick Page, Mahesh Aitha, and Whitney Craig – thank you for pushing me to never accept anything less than what you knew I was capable of, even when I didn't believe it myself. I am so grateful for your generous teaching and mentorship and for making research the absolute best part of my undergraduate experience.

To my family- especially my parents- thank you for giving me every opportunity possible to succeed. I would not be where I am today without your generous support.

To my unwavering group of Oxford friends who started as beach volleyball opponents and became my family- Jess, Chelsea, Tim, Lauren, and Eitan (and Lucy!)- you are my rocks. I could not have survived graduate school without all of you. I will never be able to put into words the joy you all bring to my life and the security of knowing I have such an amazing support system that has always accepted me exactly as I am, lifted me up when I needed help, and will always be there for me, no matter what. Thank you for being such incredible examples to me of what unconditional friends, diligent scientists, and dedicated spouses should look like. I feel so incredibly lucky to have such amazing people as role models and friends.

Lastly, to my incredible husband Zach, I am so happy that you failed your analytical chemistry entrance exam, forcing us to meet in class. Thank you for your unconditional support and for always stepping up in a big way to take care of me and our family. Your constant love and support truly has allowed me to get to where I am today, and I will never be able to convey how much it has meant to be. I'm constantly inspired by your dedication as a scientist and teacher. Thank you for never hesitating to drop everything to be there for me, for driving through snowstorms to visit me, and always putting in 110% to our family. It's been quite a road to get here, and I can't wait for where we're headed.

Table of Contents

List of Tables

List of Figures

Abstract

Spermatogenesis is a coordinated differentiation process that produces mature sperm throughout the reproductive lifetime of a male. The generation of sperm requires three distinct transitions: mitosis, meiosis, and spermiogenesis – the last of which encompasses extensive cytoskeletal and chromatin remodeling, resulting in compaction of the paternal genome and the near total replacement of histones for protamines. Protamines are small, arginine-rich proteins that hypercondense sperm chromatin, protecting the paternal DNA from physical and chemical damage during transit to the egg. Most mammals, including mice and humans, express two forms of protamine: protamine 1 (P1) and protamine 2 (P2), that together in a species-specific ratio wrap 90-95% of the sperm genome. Because protamines package most of the sperm genome, they are believed to be inert, structural elements that condense DNA via non-specific electrostatics. However, protamine protein sequences are highly variable across species and additionally, we and others have identified several post-translational modifications (PTMs) on protamines in sperm. Together, these observations raise the question of whether these PTMs may have a regulatory or instructive role in sperm chromatin, however given that protamine-DNA interactions are presumed to be driven by arginine residues, our understanding of the functional role for other residues or PTMs is limited.

To molecularly dissect the roles of non-arginine residues within P1, we generated a comprehensive array of tools to examine whether protamine sequence and/or PTMs may be instructive for sperm chromatin packaging or unpackaging in the zygote. Phylogenetic analysis of P1 protein sequences revealed several non-arginine residues that are conserved within, but not across species, and that bear PTMs in mouse sperm. We first investigated the functional role of K49, a rodent-specific residue that is acetylated in early elongating spermatids. *In vivo,* substitution of K49 to alanine (K49A) results in abnormal histone retention, decreased sperm motility, subfertility, and premature removal of P1 from paternal chromatin in zygotes. *In vitro,* the K49A substitution decreases P1 affinity for DNA and alters DNA compaction and decompaction rates. Together, this study illustrated the profound ability of a single residue substitution outside of the P1 arginine core to

alter developmental outcomes. Next, we extended these findings to investigate the role of phosphorylation on P1 S43/T45, a modification acquired in late-stage spermatids just prior to release from the testis. *In vivo,* substitution of S43/T45 to alanine (S43A/T45A) does not alter sperm chromatin structure or fertility, however substitution to a negatively charged glutamic acid (S43E/T45E) results in loss of sperm motility, altered sperm chromatin structure, and infertility. These results therefore raise the possibility that premature addition of negative charges caused by the glutamic acid residues perturbs P1 structure and/or protamine-DNA interactions, resulting in disrupted genome packaging/condensation.

Altogether, my studies leveraged novel molecular and genetic tools to establish the functional roles of individual residues outside the P1 arginine core, highlighting a complex role of protamine proteins outside a purely electrostatic model of sperm genome packaging. These studies support the hypothesis that evolutionary changes in protamine protein sequence across species may not be neutral, but rather may have evolved to regulate species-specific packaging and unpackaging of the paternal genome. Additionally, these studies suggest that protamine PTMs and crosstalk between PTMs may aid in conferring specific structural characteristics to protamines bound to DNA, providing a mechanism by which protamine-based packaging is fine-tuned to safeguard reproductive fitness.

Chapter 1 Introduction

1.1 Spermatogenesis allows for the continuous production of sperm

Spermatogenesis is an intricately regulated differentiation process by which diploid stem cells within the seminiferous tubules of the testis give rise to haploid spermatozoa. Continuous spermatogenesis is essential for optimal fertility throughout the reproductive lifetime of a male and ultimately, the transmission of genetic information to the next generation. The foundation of spermatogenesis relies on spermatogonial stem cells (SSCs), which reside at the basement membrane of the tubule and constantly balance between self-renewal to maintain the progenitor pool over time, and differentiation. After completing a species-specific number of amplifying divisions, differentiating spermatogonia then proceed through meiosis (spermatocytes) where cells undergo haploid reduction of the genome to produce round spermatids.[1] In the final stage of spermatogenesis, known as spermiogenesis, round spermatids then undergo dramatic morphological transformation including elongation of the spermatid nucleus and shedding of the cytoplasm, which is accompanied by global alterations of the chromatin landscape (see subsequent sections for details). Ultimately, elongated spermatids are released into the lumen of the seminiferous tubule and undergo additional nuclear condensation and maturation in the epididymis (**Figure 1-1**).

To effectively maintain continuous fertility over a male's lifetime, spermatogenesis must initiate at regular intervals to ensure constant, asynchronous production of sperm. In mice, this initiation of spermatogonia differentiation occurs every ~8.6 days, with the entire differentiation taking ~35 days to complete (known as the seminiferous epithelial cycle).[2,3] Because stem cell firing occurs more often than it takes to complete differentiation, the net result is the presence of multiple generations of germ cells present within a cross-section, where more differentiated cells are closer to the lumen of the tubule. For the last several decades, the field has adopted a consensus description of the 12 histological stages of the seminiferous epithelial cycle based on protein

markers and histological changes in germ cell populations.[4-6] Spermatid development is further broken down into 16 "steps" based on the morphology of the nucleus and the acrosome (a Golgi-derived sperm specific organelle essential for fusion with the egg). Steps 1-8 spermatids morphologically appear round and continue to undergo active transcription, steps 9-11 spermatids exhibit slightly elongated nuclear morphology and have largely ceased transcription, and steps 12-16 spermatids are condensed, with a hook-shaped head (characteristic of mouse sperm), and are transcriptionally quiescent.[2,7-9] Ultimately, this cyclic and asynchronous nature of spermatogenic differentiation allows for continuous production of sperm and optimal organismal reproductive capacity.

1.2 The histone-to-protamine exchange during spermiogenesis leads to biophysical and morphological changes in spermatid head morphology

1.2.1 *Spermatids undergo dramatic morphological transformation during spermiogenesis*
Proper restructuring of the spermatid nucleus and formation of both the acrosome and the sperm tail are essential for proper sperm function and the reproductive capacity of the organism. Numerous studies utilizing knockout mouse models have highlighted the importance and complexity of these processes, notably many of which result in globozoospermia (rounded sperm heads lacking acrosomes) and infertility.[10] Observable morphological transformation of spermatids begins in step 8 as the spermatid nucleus initiates elongation, flattens the acrosome on one end of the nucleus, and forms the sperm tail at the opposite end.[10] In addition to the microtubule-based axoneme of the sperm tail that provides the propulsive force to ensure motility, the manchette is an additional microtubule and actin filament-based structure that exists in sperm and is essential for spermatid morphological transformation.

The manchette is a transient, microtubule-based platform that first appears in step 8 spermatids and plays an essential role in both elongation of the spermatid nucleus as well as sperm head shaping.[11-13] The microtubular scaffold that comprises the manchette rapidly forms in a "skirt" type structure in the perinuclear region, connected to the nucleus through the linker of nucleoskeleton and cytoskeleton (LINC complex) and projects outward into the cytoplasm.[12,14] As the nucleus elongates, the manchette moves caudally in a process likely mediated by dynein, before

it is disassembled in step ~13.[13,15] Despite being a transient structure, the manchette serves several key functions for successful spermiogenesis- (1) facilitates cellular transport of proteins/cargo necessary for proper sperm tail formation[16,17] and (2) shaping of the spermatid nucleus, although the exact mechanism underlying the ability of the manchette to aid in shaping the nucleus is not well understood.[13,18]

Concomitant with this physical restructuring/reshaping of the spermatid nucleus is the restructuring of the chromatin landscape (discussed in detail below). While it remains unclear how these two processes are interrelated, several studies have documented that loss of certain proteins involved in chromatin packaging results in not only improper or incomplete remodeling of the chromatin, but also aberrant nuclear elongation.[19–22] These results strongly suggest that a direct relationship exists between chromatin remodeling and nuclear shaping during spermiogenesis, however the underlying mechanisms or factors involved in connecting these two processes remains unknown.

1.2.2 *The histone-to-protamine exchange is the hallmark of spermiogenesis*
The hallmark of spermiogenesis is the dramatic reorganization of chromatin in spermatids, which ultimately results in the replacement of the majority of histones by protamines and facilitates the packaging of the genome into a more highly condensed state, protecting the paternal DNA. The histone-to-protamine exchange is stepwise, beginning with the replacement of canonical histones by histone variants, histone post-translational modifications, and incorporation of transition proteins (all intermediate states discussed in detail in subsequent sections, **Figure 1-2**). It has long been appreciated that the histone-to-protamine exchange is a stepwise and highly regulated process but is incomplete, leaving about 1-5% of the genome packaged with nucleosomes in mature mouse sperm and 10-15% in human sperm.[23–25] Recent studies examining both human and mouse sperm chromatin have revealed that the low levels of histones retained in sperm are enriched at key developmental gene promoters, and additionally bear both active and repressive modifications- a unique "poised" chromatin state termed bivalency.[23,24] Subsequent studies also revealed a general overrepresentation of retained nucleosomes in CpG rich, hypomethylated regions and exons in addition to developmentally important promoters, although minor differences have been observed depending on methodology used.[26,27] Ultimately, the histone-to-protamine exchange results in a

unique packaging mechanism in sperm that is distinct from both somatic cells and female germ cells, allowing sperm to adopt a highly condensed, hydrodynamic shape that likely serves to allow for optimal sperm motility and protection against physical and chemical damage during transit to the egg.

1.2.3 *Incorporation of histone variants and histone post-translational modifications are essential for initiation of chromatin remodeling*

Prior to histone-to-protamine exchange in spermatids, male germ cells employ a canonical nucleosome-based chromatin architecture of genome packaging by histone octamers composed of H2A, H2B, H3, and H4.[28] In early elongating spermatids (in some cases, even earlier cell types, highlighted in **Figure 1-2**), these canonical histones are first replaced by histone variants, including many testis-specific histone variants, that promote a more open chromatin state accessible for remodeling. Several studies have demonstrated both biochemically and structurally that incorporation of histone variants induces destabilization of the nucleosome, thereby promoting reorganization of the chromatin.[29–31] One of the most well-characterized histone variants is testis-specific histone H2B (TH2B). Replacement of canonical H2B by TH2B begins well before spermatid elongation- in meiotic spermatocytes- with almost complete depletion of canonical H2B observed in spermatids.[32] Loss of function experiments have demonstrated that TH2B is essential for proper histone-to-protamine exchange, potentially mediated by its interaction with histone variant H2A.L.2.

The H2A.L.2 variant also plays a key role in ensuring successful histone-to-protamine exchange, as males lacking H2A.L.2 are infertile and fail to progress through subsequent intermediate states of chromatin remodeling, including the incorporation of transition proteins and protamines.[33] H2A.L.2 preferentially dimerizes with TH2B over canonical H2B and has been shown biochemically and structurally to destabilize nucleosomes individually as well as perturb higher order chromatin structure.[30] Additional variants incorporated during this remodeling process include spermatid-specific linker histone H1 (HILS)- also shown to be a poor condenser of DNA compared to its canonical counterpart[34,35] and testis-specific histone H3 (H3t).[29,36,37]

4

In addition to the incorporation of histone variants to destabilize nucleosomes and aid in promoting chromatin accessibility, certain histone post-translational modifications (PTMs) have been implicated in the histone-to-protamine exchange process. For example, hyperacetylation of histone H4 at K5 and K8 loosens nucleosome-DNA interactions, allowing for relaxation and opening of the chromatin, and has been documented in multiple species as a preceding modification for histone-to-protamine exchange.[38–43] H4 K5 and K8 acetylation marks are then read by the master chromatin remodeler BRDT which subsequently initiates chromatin reorganization and facilitates the deposition of transition proteins and protamines.[44] Deletion of the first bromodomain of BRDT – which prevents direct binding to H4K5/K8 acetyl marks – precludes nucleosome eviction and downstream chromatin intermediate states.[45]

More recent studies have discovered that methylation of H3 at K79 (predominantly the di- and trimethylated forms) is also deposited immediately prior to histone eviction concomitantly with H4 hyperacetylation in several organisms including mice and humans.[46,47] The presence of H3K79me2 and H3K79me3 have additionally been confirmed to be present in both mouse and human sperm by mass spectrometry.[48] While direct evidence for a specific role in the histone-to-protamine exchange is needed, these studies suggest that H3K79 methylation may be an additional prerequisite modification for the histone-to-protamine exchange.

In addition to acetylated H4 and methylated H3K79, ubiquitinated H2A and H2B are also present in elongating spermatids and males deficient for RNF8, the E3 ubiquitin ligase that deposits ubiquitin onto H2A/H2B, exhibit defective nucleosome removal and protamine deposition.[49] Furthermore, dysregulation of lysine crotonylation also leads to defective histone-to-protamine exchange.[50] Thus, many modifications are implicated in the histone-to-protamine exchange, however the interdependency between these modifications is unknown. Altogether, these studies highlight the necessity for both histone variant incorporation as well as histone PTMs to promote chromatin accessibility for the downstream incorporation of transition proteins and protamines.

1.2.4 Transition proteins are critical intermediates of the histone-to-protamine exchange

Following histone eviction and incorporation of histone variants, the transition proteins (TNPs) then occupy ~90% of the genome, acting as intermediates between histones and protamines.[51]

Mammals express two transition proteins- TNP1 and TNP2, both of which are relatively small and highly basic. TNP1 is a 6.5 kDa protein, rich in both arginine and lysine residues that are distributed throughout the entire length of the protein. While the exact role of transition proteins is not fully understood, early studies indicated that TNP1 exhibits DNA melting properties *in vitro*, allowing destabilization of chromatin and increased flexibility of the DNA strands, in a manner thought to be mediated by two tyrosine residues.[52] TNP2 is slightly larger (13 kDa) and its sequence is comprised of a broader spectrum of amino acids, including numerous histidine and serine residues. Mouse TNP2 contains two potential zinc finger motifs, which may play a role in condensing paternal chromatin during spermiogenesis and/or cessation of transcription.[53] Unlike TNP1, TNP2 stabilizes and condenses DNA *in vitro*.[54,55]

Knockout mice devoid of TNP1 exhibit subfertility, abnormal sperm morphology, and severely decreased progressive sperm motility. Furthermore, TNP1[-/-] males exhibit altered sperm chromatin composition, specifically a compensatory increase in the level of TNP2 in mature sperm as well as an accumulation of unprocessed P2.[56] Interestingly, TNP2[-/-] males do not exhibit an overt fertility phenotype, although they do exhibit a decrease in progressive sperm motility and some sperm morphology abnormalities. Like TNP1[-/-] males, TNP2[-/-] males also exhibit an increased level of unprocessed P2 in mature sperm. In both cases, when defects in progressive sperm motility were overcome with intracytoplasmic sperm injections (ICSI), sperm from both TNP1[-/-] males and TNP2[-/-] males were able to fertilize oocytes and form blastocysts at a normal rate.[57] TNP1 and TNP2 double knockout mice exhibit complete infertility, a near complete loss of progressive sperm motility, and an inability to incorporate protamines onto paternal chromatin. These studies illustrate that while TNPs may be able to partially compensate for one another, together they are indispensable for proper chromatin condensation. While both TNP1 and TNP2 are observed only transiently during mid to late spermiogenesis, together they allow for an accessible chromatin structure for the deposition of protamines and ultimately, proper chromatin condensation.

1.2.4 *The chromatin remodeling factors involved in the histone-to-protamine exchange are largely unknown*

Despite chromatin remodeling being essential for successful spermatid development, we lack an in depth understanding of the remodeling proteins involved in this process. It remains largely

unknown whether the histone-to-protamine exchange is mediated by known histone chaperones and chromatin remodelers or if germ cells employ specific machinery. Several candidate factors have been implicated in the remodeling process, and the most well characterized is the chromodomain helicase DNA binding protein 5 (CHD5). Sperm of males lacking CHD5 exhibit numerous morphological defects as well as decreased progressive motility and less condensed chromatin compared to wild type littermates.[58,59] Furthermore, histone-to-protamine exchange is perturbed in these mice, resulting in abnormal retention of histones, the presence of unprocessed P2, and curiously, an overall higher level of protamine protein expression, indicating a possible role for CHD5 in protamine transcriptional and/or translational control.[58]

Similar studies have additionally utilized knockout mouse models for several well-known chromatin remodelers, such as BAZ1A (also known as ATP-dependent chromatin assembly and remodeling factor 1, ACF1) and the SWI/SNF complex, however these studies are confounded by upstream functions in spermatogenesis. For example, loss of BAZ1A in mice leads to male infertility, immotile sperm, and dysregulation of global transcription cessation.[60] Moreover, deletion of BRG1 (a component of SWI/SNF) results in meiotic arrest, precluding the study of its role in chromatin remodeling during spermiogenesis.[61] Therefore, it remains difficult to address the role of chaperones/remodelers specifically in nucleosome eviction/protamine deposition and furthermore, to discern whether histone removal and protamine deposition are functionally distinct processes that require distinct proteins or if they are shared.

1.3 Protamine proteins are essential for sperm function and fertility

Protamines are small, sperm-specific, and highly arginine-rich proteins that can compact DNA >6-fold tighter than mitotic chromosomes, ultimately allowing the sperm head to confer a highly condensed, hydrodynamic shape.[62] Protamines or protamine-like proteins are found in both vertebrate and invertebrate genomes and are highly variable across species (see section 1.4).[63] Interestingly, protamines appear to be undergoing a unique form of positive selection to maintain overall arginine content, although not per se specific arginine sites (see section 1.4 for additional details on protamine evolution across species).[64–66]

Most mammals express two protamine proteins – protamine 1 (P1) and protamine 2 (P2). Almost universally across mammals, P1 contains 51 amino acids (including the first methionine), with >50% of the residues being arginine (R, humans are an exception, with 50 amino acids). P1 contains a characteristic arginine-rich core in the middle of the polypeptide, typically with several stretches of 6 or more arginine residues in a row. Given the conferred positive charge of stretches of R residues, it is postulated that the arginine-rich core of P1 constitutes the DNA binding domain, allowing for neutralization of the phosphodiester backbone of the DNA.[67] On both sides of the arginine core, P1 contains several potential phosphorylation sites (see section 1.6) that are presumed to be important for DNA binding and/or chromatin compaction. Additionally, P1 in eutherian mammals contains multiple cysteine residues that are essential for forming both inter- and intramolecular disulfide bonds.[68,69] Unlike P1 which is expressed in its mature form, P2 is initially expressed as a longer precursor (pro P2) that is first deposited onto DNA and subsequently processed by a series of proteolytic cleavages to produce its mature form (P2).[70,71] P2 has additionally been shown to be capable of binding zinc and while still high in arginine content, P2 additionally possesses multiple histidine residues, likely to coordinate zinc binding.[72,73]

Knockout and haploinsufficiency studies have demonstrated that P1 and P2 are essential for sperm chromatin packaging and fertility.[74,75] Initial experiments investigated the effects of loss of either P1 or P2 by introducing disrupted P1 or P2 sequences into embryonic stem cells from 129-strain mice and injecting those cells into blastocysts from C57BL/6 mice to generate chimeras.[76] While male chimeras (either P1[+/-] or P2[+/-]) produced the 129 strain sperm, they never produced offspring, and furthermore they observe abnormal sperm with reduced motility and incomplete processing of P2, leading the authors to conclude that both P1 and P2 are necessary for fertility and that haploinsufficiency of either gene is sufficient to drive the fertility phenotype. Subsequent studies established either P1 or P2-deficient mouse lines by CRISPR/Cas9-mediated gene editing and found that while haploinsufficiency of P1 is sufficient to cause infertility, loss of one P2 allele is tolerated and complete deletion is necessary to cause infertility.[74,75]

Furthermore, the ratio between P1 and P2 is essential for fertility and proper sperm chromatin structure. While this ratio is highly variable across species (1:1 in humans, 1:2 in mouse), maintenance of the species-specific P1:P2 ratio is essential. Numerous studies in humans have

demonstrated that deviations from this ratio correlate with decreased fertility, poor embryonic development, and poor outcomes of *in vitro* fertilization.[77–79] Studies in mice revealed that altered protamine ratios are associated with increased sperm DNA fragmentation, diminished fertilization rates, and defects in sperm morphology and motility.[80] Altogether, these studies highlight the importance of both the presence of P1 and P2 as well as their species-specific ratio for optimal sperm chromatin structure and fertility.

1.4 Protamines protein sequences are highly variable across species

In addition to a high level of variability in the P1:P2 ratio across species, protamine sequences themselves are also highly variable across species. Protamines or protamine-like proteins are found in a wide variety of both vertebrate and invertebrate genomes and have been posited to have evolved from linker histone H1.[63,81] There are two conserved structural features among vertebrate P1 protein sequences- (1) several stretches of either arginine or lysine residues (typically >3 in a row) and (2) multiple serine and threonine residues that are able to be phosphorylated.[63,64] Interestingly, there is compelling evidence that protamines are undergoing a unique form of positive selection that results in the maintenance of the overall number of arginine residues (and therefore net charge of the protein), but not per se specific sites.[65,66,82] The driving force behind this type of positive selection in protamines has been hypothesized to be as a means of increasing sperm chromatin stability, although direct supporting evidence is lacking.

Adjacent to the central arginine stretches of vertebrate P1 proteins are flanking regions typically containing multiple serine and threonine residues that are phosphorylated during spermiogenesis (see Section 1.6). The C-terminal flanking region in particular is highly variable across species, with multiple substitutions observed between closely related species (**Figure 1-3**).[64] An additional distinguishing feature of P1 in higher organisms is the presence of multiple cysteine residues, which emerged at the transition from Metatheria and Eutheria.[63,83] Interestingly, to date no species has yet been identified to have between 1 and 5 cysteine residues in P1- all known sequences contain either no cysteines at all or 6 or more. This observation suggests that these changes occurred relatively quickly over evolutionary time and that a minimum threshold number of cysteines must be present.[64]

The P2 gene is thought to be derived from a duplication of the P1 gene and at the amino acid level, P2 continues to diverge in sequence from P1 (50-70% sequence identify between P1 and P2). All Eutherian mammals contain the gene for P2, although only a subset of species (including *Mus Musculus, Homo sapiens,* and other primate species) express the protein. P2 generally is under less evolutionary constraint than P1, and interestingly the level of constraint differs considerably between primates and rodents and between the N-terminal portion of P2 (that is ultimately cleaved off) and the C-terminal portion of P2 (mature P2). Using evolutionary rate across all species, the rate of non-synonymous mutations compared to the rate of synonymous mutations (dN/dS) was <1 for the cleaved portion of P2, indicating weak purifying selection, whereas dN/dS for mature P2 was >1, indicating positive selection across all species. Additional analysis comparing selective pressure of either rodents or primates against all other species using a branch analysis revealed that cleaved P2 in both clades has a low evolutionary rate, again indicating weak purifying selection. In contrast, mature P2 in primates is undergoing positive selection in primates but is under relatively relaxed evolutionary constraint in rodents.[84,85]

1.5 Mammalian protamine-DNA interactions and dynamics are not well understood

Given the essential nature of protamine-DNA binding to sperm genome packaging, early *in vitro* studies of protamines focused on understanding the nature of their interaction with DNA, largely utilizing either salmon or bull (domestic cattle, *Bos taurus*) sperm protamine.[86–90] Raman and nuclear magnetic resonance (NMR) spectroscopy using a polyarginine (R6WGR6) peptide – a representative sequence of the central arginine-rich domain of bull P1 – suggested that protamines bind preferentially to the major groove of DNA, with one protamine molecule bound per turn of the helix.[88,91] Additionally, measurement of the total amount of phosphorous and sulfur in the nuclei of sperm from various species using Particle-induced X-ray emission estimated that bull P1 binds ~10-11 base pairs of DNA. Assuming the mode of P1 binding to DNA is conserved across species and given known P1:P2 ratios, calculated phosphorous:sulfur ratios of sperm from respective species estimated that P2 binds ~15 base pairs, although the exact footprints of P1 and P2 remain to be determined.[90]

While no crystal structure has yet been solved for either protamine alone or in combination with DNA, early *in vitro* studies using atomic force microscopy (AFM) on bull sperm protamine suggested that protamines coil DNA into toroidal structures containing up to 60 kb of DNA.[91] Similarly, light scattering and fluorescence microscopy experiments of various polyamines have investigated the energetics of toroid formation, highlighting the thermodynamically favorable nature of protamine-DNA toroids in solution.[92,93] More recent subsequent studies have attempted to reconstitute native sperm chromatin *in vitro* by combining salmon protamine with linearized plasmid DNA and have observed similar toroidal structures.[94]

Despite these extensive efforts in understanding protamine-DNA interactions and protamine-induced chromatin structures, an in depth understanding of mammalian protamine packaging is lacking. The studies described above almost exclusively used sperm from salmon or *Bos taurus*, both of which have atypical packaging mechanisms compared to most other mammalian protamines.[86,88–90] For example, the majority of mammalian genomes package chromatin using multiple protamine proteins (P1 and P2) and likely exhibit inter and intramolecular interactions between protamine forms, which cannot be captured in either salmon or *Bos taurus sperm*, which package chromatin using only P1. Additionally, protamine-mediated DNA condensation is facilitated by the formation of inter and intramolecular disulfide bonds, which cannot form in salmon protamine due to an absence of cysteine residues. Moreover, rapid evolution of protamines across species results in extensive sequence differences between *Bos taurus/*salmon protamine and protamines of more distant species (**Figure 1-3**). Therefore, additional molecular and biophysical studies are needed to gain an understanding of the complex P1 and P2 dual packaging system employed by most mammalian species, such as mice and humans.

1.6 Protamines are post-translationally modified

It is well characterized that histone proteins package the genome of somatic cells and that they acquire a plethora of PTMs that impact both transcriptional status and histone-DNA interactions.[95–97] The interactions of protamines with DNA, however, are presumed to be constitutive and largely driven by non-specific electrostatics between arginine residues and DNA, given the highly basic nature of protamines across species. Early studies have, however, suggested a role for dynamic

phosphorylation/dephosphorylation of protamines in a variety of species, which has been postulated to modulate protamine-DNA dynamics and maximize chromatin compaction.[98,99] For example, analysis of radiolabeled proteins from mouse and rat seminiferous tubules by acid urea gel electrophoresis illustrated that newly synthesized protamines are phosphorylated and subsequently dephosphorylated shortly after their deposition onto DNA.[71] Similar studies examining human sperm have also illustrated rapid phosphorylation and dephosphorylation of human protamines.[100] Subsequent studies identified the specific sites of phosphorylation in both human and mouse sperm (S9/S11 in human and S9 in mouse) and interestingly, these N-terminal sites are highly conserved as serine residues across species, suggesting a conserved function of the residue itself and phosphorylation at these sites.[101,102] Currently, it remains unknown which kinases deposit phosphorylation marks on protamines *in vivo*, however several studies have identified kinases that possess the ability to phosphorylate protamines *in vitro*, such as SR protein-specific kinase 1 (SRPK1), protein kinase A (PKA), protein kinase C (PKC), and Ca^{2+}/calmodulin-dependent protein kinase IV (CAMK4).[103–105]

Recent studies have utilized bottom-up mass spectrometry to generate a more comprehensive catalog of PTMs on both mouse and human protamines.[102,106] In mouse sperm, 7 novel modifications were identified (S9 phosphorylation, S43 phosphorylation/acetylation, T45 phosphorylation, and K50 acetylation/methylation) using a combination of bottom-up and top-down mass spectrometry with 53% of P1 peptides containing PTMs. Additionally, 4 PTMs were identified on P2 (S56 phosphorylation/acetylation, K58 acetylation, and K65 acetylation) with 16% of peptides containing modifications. It is likely, however, that additional modifications exist, especially in P1, given that the sequence coverage is severely limited due to the long arginine stretches. Interestingly, PTMs on P2 tended to co-occur- for example, acetylation at S56/K58 or S56/K65, suggesting potential crosstalk between modifications. On the other hand, modifications on P1 tend to be exclusive- peptides that were both acetylated and phosphorylated were never detected, further supporting potential crosstalk or specific functions of different PTMs.

Despite the identification of these PTMs, their function largely remains unknown, however it has been postulated that they may modulate protamine-DNA interactions, similar to histone PTMs. Because the addition of modifications alters the charge of the residue (addition of acetylation to

lysine neutralizes its positive charge, while phosphorylation itself imparts a negative charge), potentially addition of PTMs to P1 and P2 alter protein structure/DNA binding to modulate compaction during spermiogenesis or their removal from chromatin in the fertilized zygote. In line with this hypothesis, it was recently shown that dynamic phosphorylation/dephosphorylation at P2 S56 is essential for sperm maturation, however it remains unclear when S56 phosphorylation is established and what its specific function may be throughout spermiogenesis and spermatid development.[107] Additionally, Gou et. al. reported that P1 phosphorylation is absent in mature sperm and acquired during early embryogenesis and this phosphorylation is necessary for male pronuclear remodeling and protamine-to-histone exchange.[108] Therefore, these observations strongly suggest that protamine sequence itself and/or dynamics of modifications play distinct roles in chromatin packaging in sperm and unpackaging in the early embryo.

1.7 Summary and Aims of Dissertation

Protamine-based compaction of paternal DNA has been appreciated since the initial discovery of protamines in the 1800s. Despite the well-established relationship between proper protamine-based chromatin condensation and male fertility, we lack a comprehensive understanding of the mechanisms governing protamine deposition, protamine-based sperm genome packaging, and paternal genome unpackaging in the fertilized zygote. This is largely due to the lack of genetic and molecular reagents needed to identify regulatory factors involved, as well as a lack of *in vitro* experimental systems. Therefore, the generation of novel tools and a detailed molecular investigation of the regulation of mammalian protamine-induced DNA condensation and subsequent decondensation in the zygote are needed to elucidate how protamine-based DNA compaction is fine-tuned to safeguard reproductive fitness.

Given the highly basic nature of protamines and evolutionary selection for high arginine content, their interactions with DNA and resulting higher order genome condensation in sperm has been presumed to be dominated by non-specific electrostatic interactions between these arginine residues and DNA. Therefore, most early studies focused on arginine residues, which has severely hindered our understanding of the potential functional roles of other residues or the modifications these residues bear.[64–66] More recent studies have investigated the role of a few non-arginine

residues and their corresponding phosphorylation marks. Itoh et. al. reported that dephosphorylation of P2 at S56 is necessary for proper sperm chromatin condensation and sperm morphology, potentially for removing the repulsive interactions between the phosphorylation mark and the DNA.[107] Additionally, Gou et. al. reported that P1 phosphorylation acquired in the zygote immediately after fertilization is essential to weaken protamine-DNA interactions, allowing for remodeling of the male pronucleus.[108] While both studies provided essential insight into functions of non-arginine residues and protamine PTMs, both support a model in which electrostatics primarily drive regulation of chromatin organization.

To examine whether protamine-DNA interactions truly are governed exclusively by non-specific electrostatics or if the protamine amino acid sequence itself and/or PTMs are instructive for sperm chromatin packaging and/or unpackaging in the zygote, we generated a host of molecular tools and reagents to investigate this open area of protamine and sperm chromatin biology. First, we investigated the functional role of P1 lysine 49, a residue outside the arginine-rich core of P1 that is highly conserved across the rodent lineage and is acetylated in mature mouse sperm. Extensive biochemical, molecular, and genetic investigation revealed an indispensable role for this residue in sperm genome packaging and embryonic development, establishing generally a role for P1 amino acids outside of an exclusively electrostatically based model (Chapter 2). To extend these findings to other residues in P1 outside the arginine core, we investigated the role of P1 S43/T45, two residues present in the C-terminal portion of P1 that are phosphorylated in mature sperm. Our findings revealed that while substitution of S43/T45 to a neutral, but non-modifiable alanine residue does not perturb overall sperm chromatin structure, substitution to a negatively charged glutamic acid dramatically alters sperm chromatin structure and fertility (Chapter 3).

Altogether, this work generated novel genetic, biochemical, and biophysical tools and highlights the potential that P1 residues outside of the arginine core can exhibit functional consequences. Our observations that single amino acid changes in mouse P1 can dramatically alter the functional properties of the P1 protein strongly supports the potential that evolutionary changes in protamine amino acid sequences across species are highly unlikely to be neutral. Instead, it is conceivable that these non-arginine residues (which diverge greatly across species) have evolved to modulate species-specific sperm genome packaging and unpackaging in the fertilized zygote, possibly to

ensure compatibility between species upon fertilization and optimal reproductive capacity. Furthermore, these results suggest the potential for a structural role for protamine PTMs in sperm chromatin packaging, possibly by conferring structural changes to protamine proteins to promote favorable packaging in a manner that allows for fine-tuning of the overall chromatin landscape before condensation patterns are fixed across the genome. Moreover, it is likely that there is crosstalk occurring between protamine PTMs during spermiogenesis and/or in mature sperm and this interplay may additionally aid in promoting proper genome condensation.
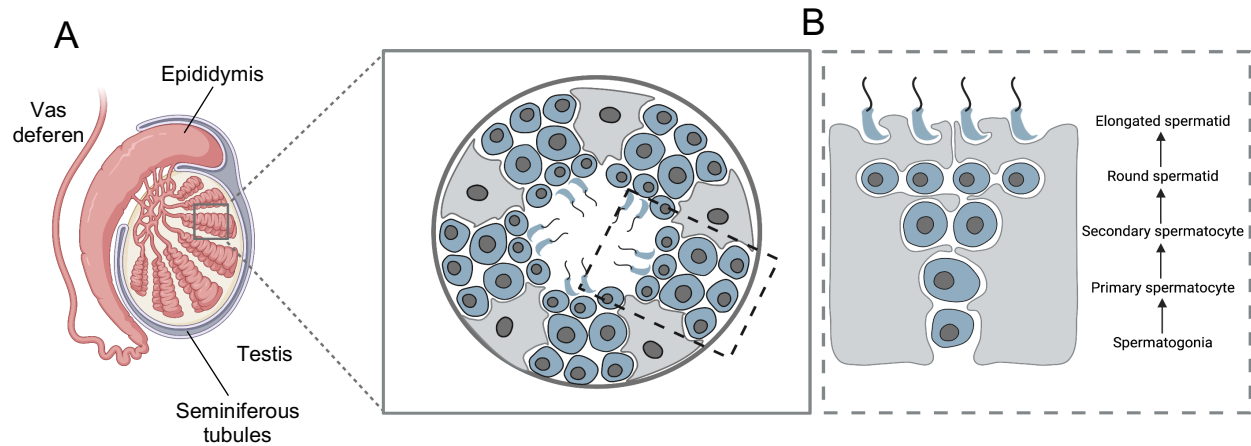
Figure 1-1 An overview of spermatogenesis (A) General testis architecture highlighting the major structures necessary for sperm production (left) and a cartoon representation of seminiferous tubule organization in cross section (right). **(B)** Cartoon representation of the progression of spermatogenesis within the seminiferous tubule, highlighting the major germ cell types observed in the testis. This figure was created using BioRender.com and cartoon representations are adapted from [109].
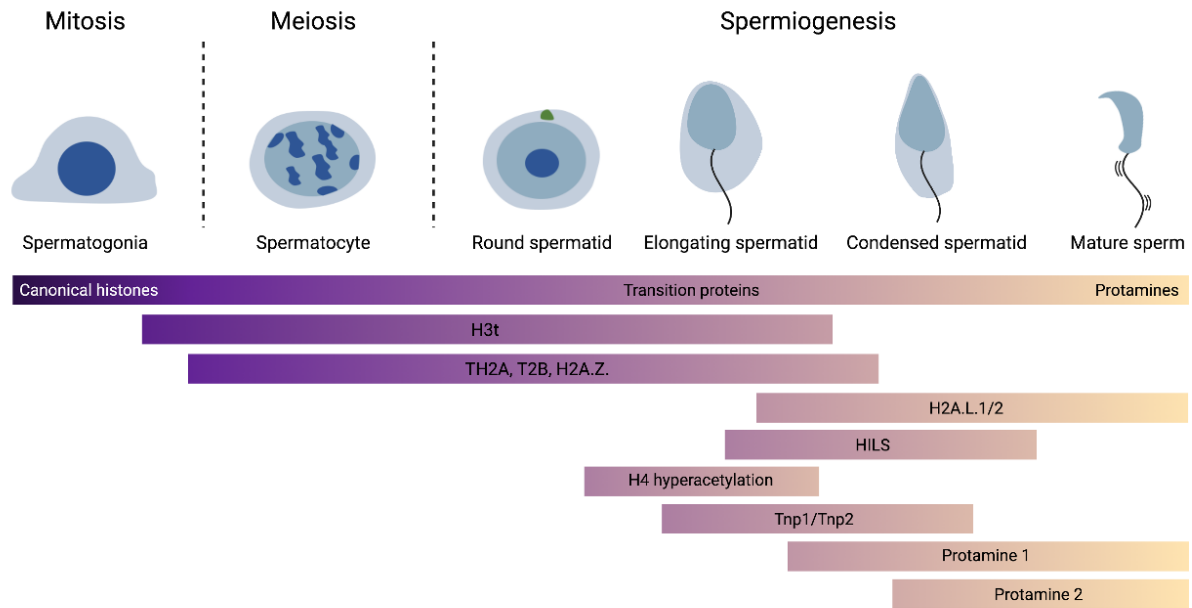
Figure 1-2 Major remodeling events of the histone-to-protamine exchange
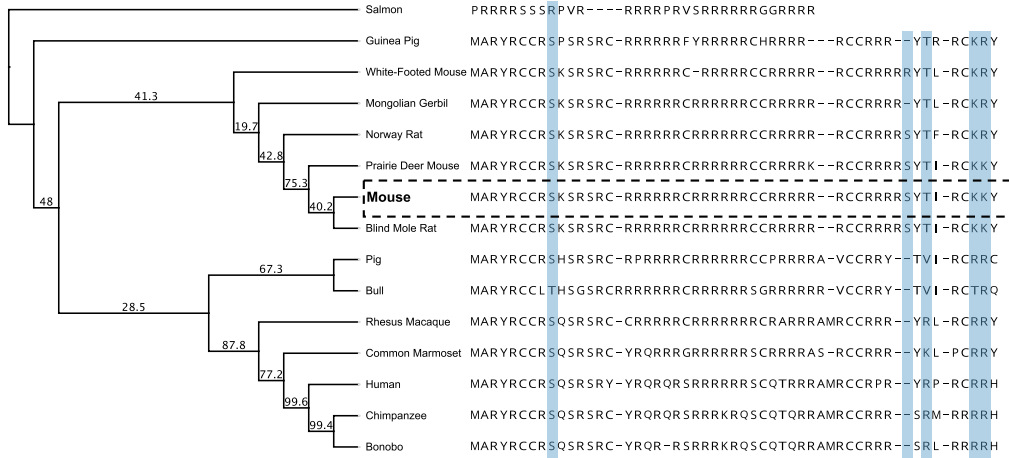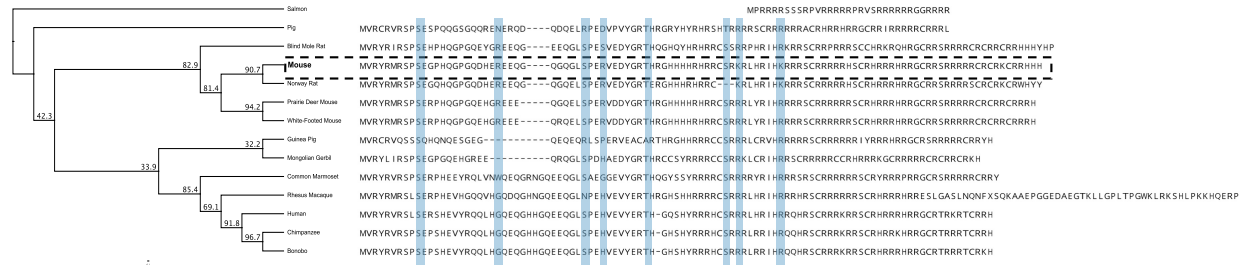
Figure 1-3 Sequence alignments of *P1* and *P2* protein sequences across numerous species. Blue shaded bars indicate residues that are post-translationally modified in mouse sperm.

1.8 References

1. Fayomi, A. P. & Orwig, K. E. Spermatogonial stem cells and spermatogenesis in mice, monkeys and men. *Stem Cell Res* **29**, 207–214 (2018).
2. Oakberg, E. F. A description of spermiogenesis in the mouse and its use in analysis of the cycle of the seminiferous epithelium and germ cell renewal. *Am J Anat* **99**, 391–413 (1956).
3. Clermont, Y. Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiol Rev* **52**, 198–236 (1972).
4. Meistrich, M. L. & Hess, R. A. Assessment of spermatogenesis through staging of seminiferous tubules. *Methods Mol Biol* **927**, 299–307 (2013).
5. Green, C. D. *et al.* A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental Cell* **46**, 651-667.e10 (2018).
6. Hasegawa, K. & Saga, Y. Retinoic acid signaling in Sertoli cells regulates organization of the blood-testis barrier through cyclical changes in gene expression. *Development (Cambridge, England)* **139**, 4347–55 (2012).
7. Zheng, J. *et al.* Erasure of the paternal transcription program during spermiogenesis: the first step in the reprogramming of sperm chromatin for zygotic development. *Dev Dyn* **237**, 1463–1476 (2008).
8. Monesi, V. Ribonucleic Acid Synthesis during Mitosis and Meiosis in the Mouse Testis. *J Cell Biol* **22**, 521–532 (1964).
9. Monesi, V., Geremia, R., D'Agostino, A. & Boitani, C. Biochemistry of male germ cell differentiation in mammals: RNA synthesis in meiotic and postmeiotic cells. *Curr Top Dev Biol* **12**, 11–36 (1978).
10. Khawar, M. B., Gao, H. & Li, W. Mechanism of Acrosome Biogenesis in Mammals. *Front Cell Dev Biol* **7**, 195 (2019).
11. Kierszenbaum, A. L. & Tres, L. L. The acrosome-acroplaxome-manchette complex and the shaping of the spermatid head. *Arch Histol Cytol* **67**, 271–284 (2004).
12. Toshimori, K. & Ito, C. Formation and organization of the mammalian sperm head. *Arch Histol Cytol* **66**, 383–396 (2003).
13. Russell, L. D., Russell, J. A., MacGregor, G. R. & Meistrich, M. L. Linkage of manchette microtubules to the nuclear envelope and observations of the role of the manchette in nuclear shaping during spermiogenesis in rodents. *Am J Anat* **192**, 97–120 (1991).
14. Gob, E., Schmitt, J., Benavente, R. & Alsheimer, M. Mammalian sperm head formation involves different polarization of two novel LINC complexes. *PLoS One* **5**, e12072 (2010).
15. Yoshida, T., Ioshii, S. O., Imanaka-Yoshida, K. & Izutsu, K. Association of cytoplasmic dynein with manchette microtubules and spermatid nuclear envelope during spermiogenesis in rats. *J Cell Sci* **107 (Pt 3)**, 625–633 (1994).
16. Kierszenbaum, A. L., Rivkin, E. & Tres, L. L. Cytoskeletal track selection during cargo transport in spermatids is relevant to male fertility. *Spermatogenesis* **1**, 221–230 (2011).
17. Kierszenbaum, A. L. *et al.* GMAP210 and IFT88 are present in the spermatid golgi apparatus and participate in the development of the acrosome-acroplaxome complex, head-tail coupling apparatus and tail. *Dev Dyn* **240**, 723–736 (2011).
18. Lehti, M. S. & Sironen, A. Formation and function of the manchette and flagellum during spermatogenesis. *Reproduction* **151**, R43-54 (2016).

19. Martianov, I. *et al.* Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. *Proceedings of the National Academy of Sciences* **102**, (2005).

20. Adham, I. M. *et al.* Teratozoospermia in mice lacking the transition protein 2 (Tnp2). *Molecular human reproduction* **7**, (2001).

21. Zhuang, T. *et al.* CHD5 is required for spermiogenesis and chromatin condensation. *Mechanisms of development* **131**, (2014).

22. Li, W. *et al.* Chd5 orchestrates chromatin remodelling during sperm development. *Nature communications* **5**, (2014).

23. Brykczynska, U. *et al.* Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* **17**, 679–687 (2010).

24. Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473–478 (2009).

25. Hammoud, S. S. *et al.* Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. *Hum Reprod* **26**, 2558–2569 (2011).

26. Erkek, S. *et al.* Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat Struct Mol Biol* **20**, 868–875 (2013).

27. Yamaguchi, K. *et al.* Re-evaluating the Localization of Sperm-Retained Histones Revealed the Modification-Dependent Accumulation in Specific Genome Regions. *Cell Rep* **23**, 3920–3932 (2018).

28. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997).

29. Tachiwana, H. *et al.* Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc Natl Acad Sci U S A* **107**, 10454–10459 (2010).

30. Syed, S. H. *et al.* The incorporation of the novel histone variant H2AL2 confers unusual structural and functional properties of the nucleosome. *Nucleic Acids Res* **37**, 4684–4695 (2009).

31. Li, A. *et al.* Characterization of nucleosomes consisting of the human testis/sperm-specific histone H2B variant (hTSH2B). *Biochemistry* **44**, 2529–2535 (2005).

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

33. Barral, S. *et al.* Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Mol Cell* **66**, 89-101 e8 (2017).

34. Yan, W., Ma, L., Burns, K. H. & Matzuk, M. M. HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. *Proc Natl Acad Sci U S A* **100**, 10546–10551 (2003).

35. Mishra, L. N. *et al.* Spermatid-specific linker histone HILS1 is a poor condenser of DNA and chromatin and preferentially associates with LINE-1 elements. *Epigenetics Chromatin* **11**, 43 (2018).

36. Tachiwana, H., Osakabe, A., Kimura, H. & Kurumizaka, H. Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins in vitro. *Nucleic Acids Res* **36**, 2208–2218 (2008).

37. Ueda, J. *et al.* Testis-Specific Histone Variant H3t Gene Is Essential for Entry into Spermatogenesis. *Cell Rep* **18**, 593–600 (2017).

38. Shiota, H. *et al.* Nut Directs p300-Dependent, Genome-Wide H4 Hyperacetylation in Male Germ Cells. *Cell Rep* **24**, 3477-3487 e6 (2018).

39. Ketchum, C. C., Larsen, C. D., McNeil, A., Meyer-Ficca, M. L. & Meyer, R. G. Early histone H4 acetylation during chromatin remodeling in equine spermatogenesis. *Biol Reprod* **98**, 115–129 (2018).

40. Meistrich, M. L., Trostle-Weige, P. K., Lin, R., Allis, C. D. & Bhatnagar, Y. M. Highly acetylated H4 is associated with histone displacement in rat spermatids. *Molecular Reproduction and Development* **31**, (1992).

41. Grimes, S. R. & Henderson, N. Hyperacetylation of histone H4 in rat testis spermatids. *Experimental Cell Research* **152**, (1984).

42. Hong, L., Schroth, G. P., Matthews, H. R., Yau, P. & Bradbury, E. M. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *Journal of Biological Chemistry* **268**, (1993).

43. Candido, E. P. M. & Dixon, G. H. Trout Testis Cells. *Journal of Biological Chemistry* **247**, (1972).

44. Gaucher, J. *et al.* Bromodomain-dependent stage-specific male genome programming by Brdt. *EMBO J* **31**, 3809–3820 (2012).

45. Gaucher, J. *et al.* Bromodomain-dependent stage-specific male genome programming by Brdt. *The EMBO Journal* **31**, (2012).

46. Dottermusch-Heidel, C. *et al.* H3K79 methylation directly precedes the histone-to-protamine transition in mammalian spermatids and is sensitive to bacterial infections. *Andrology* **2**, (2014).

47. Dottermusch-Heidel, C. *et al.* H3K79 methylation: a new conserved mark that accompanies H4 hyperacetylation prior to histone-to-protamine transition in *Drosophila* and rat. *Biology Open* **3**, (2014).

48. Luense, L. J. *et al.* Comprehensive analysis of histone post-translational modifications in mouse and human male germ cells. *Epigenetics & chromatin* **9**, (2016).

49. Lu, L. Y. *et al.* RNF8-dependent histone modifications regulate nucleosome removal during spermatogenesis. *Dev Cell* **18**, 371–384 (2010).

50. Liu, S. *et al.* Chromodomain Protein CDYL Acts as a Crotonyl-CoA Hydratase to Regulate Histone Crotonylation and Spermatogenesis. *Mol Cell* **67**, 853-866 e5 (2017).

51. Meistrich, M. L., Mohapatra, B., Shirley, C. R. & Zhao, M. Roles of transition nuclear proteins in spermiogenesis. *Chromosoma* **111**, (2003).

52. Singh, J. & Rao, M. R. Interaction of rat testis protein, TP, with nucleic acids in vitro. Fluorescence quenching, UV absorption, and thermal denaturation studies. *Journal of Biological Chemistry* **262**, (1987).

53. Baskaran, R. & Rao, M. R. S. Mammalian spermatid specific protein, TP2, is a zinc metalloprotein with two finger motifs. *Biochemical and Biophysical Research Communications* **179**, (1991).

54. Kundu, T. K. & Rao, M. R. S. DNA Condensation by the Rat Spermatidal Protein TP2 Shows GC-Rich Sequence Preference and Is Zinc Dependent. *Biochemistry* **34**, (1995).

55. Baskaran, R. & Rao, M. R. Interaction of spermatid-specific protein TP2 with nucleic acids, in vitro. A comparative study with TP1. *Journal of Biological Chemistry* **265**, (1990).

56.   Yu, Y. E. *et al.* Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proceedings of the National Academy of Sciences* **97**, (2000).

57.   Shirley, C. R., Hayashi, S., Mounsey, S., Yanagimachi, R. & Meistrich, M. L. Abnormalities and Reduced Reproductive Potential of Sperm from Tnp1- and Tnp2-Null Double Mutant Mice1. *Biology of Reproduction* **71**, (2004).

58.   Li, W. *et al.* Chd5 orchestrates chromatin remodelling during sperm development. *Nature Communications* **5**, (2014).

59.   Zhuang, T. *et al.* CHD5 is required for spermiogenesis and chromatin condensation. *Mechanisms of Development* **131**, (2014).

60.   Dowdle, J. A. *et al.* Mouse BAZ1A (ACF1) is dispensable for double-strand break repair but is essential for averting improper gene expression during spermatogenesis. *PLoS genetics* **9**, (2013).

61.   Wang, J., Gu, H., Lin, H. & Chi, T. Essential Roles of the Chromatin Remodeling Factor Brg1 in Spermatogenesis in Mice1. *Biology of Reproduction* **86**, (2012).

62.   Pienta, K. J. & Coffey, D. S. A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome. *J Cell Sci Suppl* **1**, 123–135 (1984).

63.   Lewis, J. D., Song, Y., de Jong, M. E., Bagha, S. M. & Ausió, J. A walk though vertebrate and invertebrate protamines. *Chromosoma* **111**, (2003).

64.   Queralt, R. *et al.* Evolution of protamine P1 genes in mammals. *Journal of Molecular Evolution* **40**, (1995).

65.   Rooney, A. P., Zhang, J. & Nei, M. An Unusual Form of Purifying Selection in a Sperm Protein. *Molecular Biology and Evolution* **17**, (2000).

66.   Torgerson, D. G., Kulathinal, R. J. & Singh, R. S. Mammalian Sperm Proteins Are Rapidly Evolving: Evidence of Positive Selection in Functionally Diverse Genes. *Molecular Biology and Evolution* **19**, (2002).

67.   Balhorn, R. The protamine family of sperm nuclear proteins. *Genome Biology* **8**, (2007).

68.   Hutchison, J. M., Rau, D. C. & DeRouchey, J. E. Role of Disulfide Bonds on DNA Packaging Forces in Bull Sperm Chromatin. *Biophysical Journal* **113**, (2017).

69.   Vilfan, I. D., Conwell, C. C. & Hud, N. v. Formation of Native-like Mammalian Sperm Cell Chromatin with Folded Bull Protamine. *Journal of Biological Chemistry* **279**, (2004).

70.   Yelick, P. C. *et al.* Mouse protamine 2 is synthesized as a precursor whereas mouse protamine 1 is not. *Mol Cell Biol* **7**, 2173–2179 (1987).

71.   Green, G. R., Balhorn, R., Poccia, D. L. & Hecht, N. B. Synthesis and processing of mammalian protamines and transition proteins. *Mol Reprod Dev* **37**, 255–263 (1994).

72.   Bench, G., Corzett, M. H., Kramer, C. E., Grant, P. G. & Balhorn, R. Zinc is sufficiently abundant within mammalian sperm nuclei to bind stoichiometrically with protamine 2. *Molecular Reproduction and Development* **56**, (2000).

73.   Gatewood, J. M., Schroth, G. P., Schmid, C. W. & Bradbury, E. M. Zinc-induced secondary structure transitions in human sperm protamines. *Journal of Biological Chemistry* **265**, (1990).

74.   Takeda, N. *et al.* Viable offspring obtained from Prm1-deficient sperm in mice. *Scientific Reports* **6**, (2016).

75.   Schneider, S. *et al.* Re-visiting the Protamine-2 locus: deletion, but not haploinsufficiency, renders male mice infertile. *Sci Rep* **6**, 36764 (2016).

76. Cho, C. *et al.* Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat Genet* **28**, 82–86 (2001).

77. Aoki, V. W. *et al.* Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. *Fertil Steril* **86**, 1408–1415 (2006).

78. de Mateo, S. *et al.* Protamine 2 precursors (Pre-P2), protamine 1 to protamine 2 ratio (P1/P2), and assisted reproduction outcome. *Fertil Steril* **91**, 715–722 (2009).

79. AOKI, V. *et al.* Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. *Fertility and Sterility* **86**, (2006).

80. Cho, C. *et al.* Protamine 2 deficiency leads to sperm DNA damage and embryo death in mice. *Biol Reprod* **69**, 211–217 (2003).

81. Lewis, J. D. *et al.* Histone H1 and the origin of protamines. *Proceedings of the National Academy of Sciences of the United States of America* **101**, (2004).

82. Wyckoff, G. J., Wang, W. & Wu, C.-I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, (2000).

83. Ausio, J. & van Holde, K. E. The histones of the sperm of Spisula solidissima include a novel, cysteine-containing H-1 histone. *Cell differentiation* **23**, (1988).

84. Lüke, L., Tourmente, M., Dopazo, H., Serra, F. & Roldan, E. R. S. Selective constraints on protamine 2 in primates and rodents. *BMC Evolutionary Biology* **16**, (2016).

85. Lüke, L., Vicens, A., Tourmente, M. & Roldan, E. R. S. Evolution of Protamine Genes and Changes in Sperm Head Phenotype in Rodents1. *Biology of Reproduction* **90**, (2014).

86. Brewer, L. R. Protamine-Induced Condensation and Decondensation of the Same DNA Molecule. *Science* **286**, (1999).

87. Balhorn, R., Brewer, L. & Corzett, M. DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. *Mol Reprod Dev* **56**, 230–234 (2000).

88. Prieto, M. C., Maki, A. H. & Balhorn, R. Analysis of DNA-protamine interactions by optical detection of magnetic resonance. *Biochemistry* **36**, 11944–11951 (1997).

89. Brewer, L., Corzett, M., Lau, E. Y. & Balhorn, R. Dynamics of Protamine 1 Binding to Single DNA Molecules. *Journal of Biological Chemistry* **278**, (2003).

90. Bench, G. S., Friz, A. M., Corzett, M. H., Morse, D. H. & Balhorn, R. DNA and total protamine masses in individual sperm from fertile mammalian subjects. *Cytometry* **23**, 263–271 (1996).

91. Hud, N. V., Allen, M. J., Downing, K. H., Lee, J. & Balhorn, R. Identification of the Elemental Packing Unit of DNA in Mammalian Sperm Cells by Atomic Force Microscopy. *Biochemical and Biophysical Research Communications* **193**, (1993).

92. Widom, J. & Baldwin, R. L. Cation-induced toroidal condensation of DNA. *Journal of Molecular Biology* **144**, (1980).

93. Watanabe, F. & Schwarz, G. Thermodynamics and kinetics of co-operative protein-nucleic acid binding. *Journal of Molecular Biology* **163**, (1983).

94. Allen, M. AFM analysis of DNA-protamine complexes bound to mica. *Nucleic Acids Research* **25**, (1997).

95. Brehove, M. *et al.* Histone Core Phosphorylation Regulates DNA Accessibility. *Journal of Biological Chemistry* **290**, (2015).

96. Kiefer, C. M., Hou, C., Little, J. A. & Dean, A. Epigenetics of β-globin gene regulation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **647**, (2008).

97. Shogren-Knaak, M. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* **311**, (2006).
98. Ingles, C. J. & Dixon, G. H. Phosphorylation of protamine during spermatogenesis in trout testis. *Proc Natl Acad Sci U S A* **58**, 1011–1018 (1967).
99. Seligman, J., Zipser, Y. & Kosower, N. S. Tyrosine phosphorylation, thiol status, and protein tyrosine phosphatase in rat epididymal spermatozoa. *Biol Reprod* **71**, 1009–1015 (2004).
100. Pruslin, F. H., Imesch, E., Winston, R. & Rodman, T. C. Phosphorylation state of protamines 1 and 2 in human spermatids and spermatozoa. *Gamete Res* **18**, 179–190 (1987).
101. Chira, F. *et al.* Phosphorylation of human sperm protamines HP1 and HP2: identification of phosphorylation sites. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1203**, (1993).
102. Brunner, A. M., Nanni, P. & Mansuy, I. M. Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics Chromatin* **7**, 2 (2014).
103. Papoutsopoulou, S. SR protein-specific kinase 1 is highly expressed in testis and phosphorylates protamine 1. *Nucleic Acids Research* **27**, (1999).
104. Wu, J. Y. *et al.* Spermiogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. *Nature Genetics* **25**, (2000).
105. Pirhonen, A., Linnala-Kankkunen, A. & Menpaa, P. H. P2 protamines are phosphorylated in vitro by protein kinase C, whereas P1 protamines prefer cAMP-dependent protein kinase. A comparative study of five mammalian species. *Eur J Biochem* **223**, 165–169 (1994).
106. Soler-Ventura, A. *et al.* Characterization of Human Sperm Protamine Proteoforms through a Combination of Top-Down and Bottom-Up Mass Spectrometry Approaches. *J Proteome Res* **19**, 221–237 (2020).
107. Itoh, K. *et al.* Dephosphorylation of protamine 2 at serine 56 is crucial for murine sperm maturation in vivo. *Sci Signal* **12**, (2019).
108. Gou, L. T. *et al.* Initiation of Parental Genome Reprogramming in Fertilized Oocyte by Splicing Kinase SRPK1-Catalyzed Protamine Phosphorylation. *Cell* **180**, 1212-1227 e14 (2020).
109. Larose, H. *et al.* Gametogenesis: A journey from inception to conception. *Current topics in developmental biology* **132**, (2019).

Chapter 2 Sperm Chromatin Structure and Reproductive Fitness are Altered by Substitution of a Single Amino Acid in Mouse Protamine 1[1]

2.1 Abstract

Conventional dogma presumes that protamine-mediated DNA compaction in sperm is achieved by passive electrostatics between DNA and the arginine-rich core of protamines. However, phylogenetic analysis reveals several non-arginine residues that are conserved within, but not across, species. The functional significance of these residues or post-translational modifications are poorly understood. Here, we investigated the functional role of K49, a rodent-specific lysine residue in mouse protamine 1 (P1) that is acetylated early in spermiogenesis and retained in sperm. *In vivo*, an alanine substitution (P1 K49A) results in ectopic histone retention, decreased sperm motility, decreased male fertility, and in zygotes, premature P1 removal from paternal chromatin. *In vitro*, the P1 K49A substitution decreases protamine-DNA binding and alters DNA compaction/decompaction kinetics. Hence, a single amino acid substitution outside the P1 arginine core is sufficient to profoundly alter protein function and developmental outcomes, suggesting that protamine non-arginine residues are essential to ensure reproductive fitness.

2.2 Introduction

Spermatogenesis is a highly regulated differentiation process by which spermatogonial stem cells give rise to mature haploid spermatozoa throughout life. In its final stage, known as spermiogenesis, haploid round spermatids elongate to form sperm. During this morphological transformation, the global chromatin composition is completely altered, resulting in the near total replacement of histones by protamines. This transition from nucleosome-based to protamine-based chromatin in spermatids is a stepwise process. It begins with the exchange of canonical histones for testis-specific histone variants, such as spermatid-specific linker histone H1-like protein (HILS),[1] H2AL1/2[2,3], testis-specific histone H2B (TH2B)[4,5], and histone H3T.[6–8] Histone variants (both canonical and testis-specific) then acquire post-translational modifications (PTMs), notably hyperacetylation of H4[9,10] and ubiquitination of H2A/H2B[11], which initiates loosening of the chromatin structure to facilitate the incorporation of transition proteins 1 (TNP1) and 2 (TNP2), and subsequent replacement by protamines.[2,12,13]

Protamines are small, arginine-rich sperm-specific structural proteins that condense sperm chromatin.[14,15] Most mammals, including mice and humans, express two forms of protamine: protamine 1 (P1) and protamine 2 (P2). P1 is directly expressed in its mature form, while P2 is expressed as a longer precursor (pro P2) that is initially deposited onto DNA and subsequently processed by a series of selective proteolytic cleavages to produce its mature form (P2).[16,17] Together, P1 and P2 wrap 90-95% of the mammalian sperm genome.[15,18] Numerous studies in both mice and humans have demonstrated that maintenance of a species-specific ratio of P1:P2 (1:1 in humans, 1:2 in mouse) is necessary for fertility, and that alterations in this ratio correlate with decreased fertility and poor embryonic development.[19–21] Furthermore, from knockout and haploinsufficiency studies, we know that P1 and P2 are essential for sperm chromatin packaging and fertility. [22–25]

Although protamines' role in packaging the majority of the sperm genome and requirement for fertility are widely known, the regulation of mammalian protamine-induced DNA condensation and subsequent decondensation in the zygote is not well understood. Much of our understanding of protamine-DNA dynamics arises from early *in vitro* studies utilizing either salmon or bull

(domestic cattle, *Bos taurus*) sperm protamine, both of which express only one form of protamine. [26–29] However, most mammalian genomes encode multiple protamine proteins (P1, P2, and/or P3) that may engage in complex inter and intramolecular interactions between different protamine forms, which cannot be captured or monitored in species which encode for a single protamine protein[26,30,31] or species that lack cysteine residues (like Salmon). Hence, our understanding of a complex, multi-protamine packaging system is limited and based on the assumption that the biochemical and biophysical properties of protamines are conserved, despite striking differences in amino acid sequence and composition across animal species.[32–34] Consequently, the framework based on our current knowledge is unlikely to accurately describe functional differences and/or kinetics of mammalian protamines or systems that employ a dual protamine (P1 and P2) packaging system.

In somatic cells, histone proteins package the DNA, and these proteins acquire various PTMs which impact histone-DNA interaction strength and chromatin and transcriptional states.[35–37] In sperm, a similar series of modifications have been reported for mouse and human protamines, but given the basic nature of protamine proteins and apparent evolutionary selection for high arginine content, most studies of protamine-DNA interactions centered on arginine residues, limiting our understanding of functional roles for other residues or protamine PTMs. [38–40]However, a few studies that examined protamine phosphorylation and dephosphorylation have suggested that protamine phosphorylation during spermiogenesis is important for modulating protamine-DNA dynamics and maximizing chromatin compaction.[16,41–43] More recently, Gou. et. al. reported that several serine residues in P1 acquired phosphorylation during early embryogenesis and these modifications were required to weaken protamine-DNA interactions to allow male pronuclear remodeling and protamine-to-histone exchange—further supporting a model where electrostatic interactions are the primary mode of regulation of sperm chromatin organization.[44]

To examine whether protamine-DNA interactions can solely be explained by simple electrostatics or if the protamine sequence itself and/or the dynamics of its modifications are instructive for proper chromatin packaging in sperm and unpackaging in the early embryo, we performed mass spectrometry and phylogenetic analysis of P1 sequences. We find that the sites of protamine PTMs are conserved within species, but not across, suggesting the possibility of lineage specific function.

To dissect this phenomenon genetically, we chose to focus on P1 lysine 49—a residue that is conserved across the rodent lineage and we find to be acetylated in mature mouse sperm. Specifically, we report that P1 K49 acetylation is acquired in early elongating spermatids (stage IX-XI) and persists in mature sperm. The substitution of K49 for alanine (A) results in severe male subfertility in mice. Biochemical analysis of sperm nuclei reveals alterations in sperm chromatin composition and histone eviction. K49A mutant embryos prematurely dismiss P1 from paternal chromatin and many of these embryos arrest at the 1-cell and blastocyst stages. *In vitro*, bulk and single molecule assays reveal that the K49A mutant protein has significantly lower affinity for DNA, slower rates of DNA condensation, and accelerated de-condensation—consistent with premature dismissal of P1 in embryos. All together, these findings establish an indispensable role for P1 amino acids outside of a general electrostatic model, highlighting a more complex role for protamine protein sequence in governing protamine-DNA genome packaging and embryonic development.

2.3 Results

2.3.1 *Post-translational modifications on P1 are lineage specific*

Previous top-down and bottom-up mass spectrometry studies identified several P1 and P2 PTMs in both human and mouse mature sperm, however their function (aside from observations of P1 phosphorylation in the early embryo and P2 phosphorylation/dephosphorylation during spermatogenesis) remains unknown. [41,44–46] We were intrigued by these data, but because sperm are transcriptionally quiescent, modifications cannot be involved in germ cell transcriptional regulation. Therefore, we set out to investigate possible alternative functions of protamine PTMs. First, we used mass spectrometry to independently validate that these modifications are present in mature mouse sperm. Through this analysis, we confirmed previously identified P1 modifications, such as phosphorylation at serine (S) 9, S43, and threonine (T) 45 and acetylation at lysine (K) 50, but we also identified additional modifications, such as acetylation of P1 at K49 (P1 K49ac) (**Table 2-1**, summarized in **Figure 2-1A**). To gain a deeper understanding of sequence conservation of PTM-bearing sites, we constructed a phylogenetic tree for species across the orders Rodentia, Primate and Artiodactyla (hoofed animals, including *Bos taurus*) using maximum likelihood inferred from P1 protein sequences (**Figure 2-1B**). We found that the P1 S9 position is highly

conserved, and its phosphorylation is well-established in both mouse and human, likely reflecting that both the amino acid position and modification serve a necessary function across species.[45,47] In contrast, several C-terminal modified residues (S43, T45, K49, K50) are all highly conserved within the mouse lineage but are all largely occupied by alternative residues in more distant species (**Figure 2-1B**). Given that the K49 residue is conserved across rodents but not higher species, we developed a host of molecular and genetic reagents to begin to dissect the role of non-arginine residues, and specifically explore the functional role of K49 in the mouse germline.

2.3.2 *P1 K49 acetylation is acquired in the testis in a stage-specific manner and persists in mature sperm*

To detect the presence of K49 acetylation (K49ac) in sperm, we generated a polyclonal antibody against P1 K49ac. In immunoblots of acid extracted protein from mature sperm, we found that our antibody detected a distinct band. This band was lost when outcompeted by an acetylated P1 K49 peptide, but not when we used a nonspecific peptide from an unrelated protein, or a non-acetylated P1 peptide, thus confirming both the presence of K49ac *in vivo* and the specificity of our antibody (**Figure 2-2A**). To precisely define at which stage or stages of the seminiferous tubule cycle P1 K49ac is established, we co-stained testes cross-sections using our custom antibody combined with the acrosomal marker PNA-Lectin. We found that specific signal was initiated in stage IX (containing early elongating spermatids) and peaked at stages X-XI (100% of tubules) but then diminished in stages XII-VIII (late-stage spermatids, **Figure 2-1C,D, 2-2B**). Although the fluorescent signal for P1 K49ac is diminished in later stages of spermatid maturation, the modification remained detectable in elongating spermatid-enriched lysates from the testis, as well as in epididymal and vas deferens sperm by immunoblotting (**Figure 2-1E**). These data point to two possible interpretations: loss of signal may be the result of high compaction of spermatids or low abundance below the immunofluorescence detection limit for our antibody.

2.3.3 *Substitution of P1 K49 for alanine results in sperm motility defects, abnormal sperm morphology, and subfertility*

To investigate the functional significance of P1 K49ac *in vivo,* we used CRISPR/Cas9 to generate a lysine to alanine mutant mouse (K49A). We then used Sanger sequencing to confirm the presence of the target mutation (**Figure 2-3A**) and the absence of potential off-target genetic modifications

(**Figure 2-4A**). Overall, P1$^{K49A/+}$ or P1$^{K49A/K49A}$ mice appeared phenotypically normal; we observed no significant differences in testes/body weight ratio and all germ cell populations were detected (**Figure 2-4B,C**). However, while overall sperm counts were normal in P1$^{K49A/K49A}$ mice, progressive sperm motility (the ability of sperm to swim forward) was severely impaired (**Figure 2-3B,C**) and various sperm structural abnormalities were noted including coiled midpieces, bent back heads, and abnormal head morphology (**Figure 2D,E**). Furthermore, P1$^{K49A/K49A}$ males were severely subfertile (**Figure 2-3F**). To ensure that this phenotype is not caused by loss of the P1 protein itself, we stained both P1$^{+/+}$ and P1$^{K49A/K49A}$ testis cross-sections using a custom P1 antibody (**Figure 2-4D, specificity test in Figure 2-4E**). We found that P1 is detectable in both cases, and that P1 protein levels are comparable between genotypes (**Figure 2-4F, 2-5A**), suggesting that the phenotype is not simply due to loss of P1 expression.

2.3.4   P1 K49A mutants progress through key chromatin intermediate stages of the histone-to-protamine exchange, yet have abnormal histone retention

Given the abnormal sperm motility and morphology, we next analyzed the effects of the K49A substitution on the histone-to-protamine exchange and mature sperm chromatin composition. When we compared protamine levels and ratios in a fixed number of P1$^{+/+}$, P1$^{K49A/+}$, and P1$^{K49A/K49A}$ sperm, we found that ratios in P1$^{K49A/K49A}$ sperm were significantly shifted from the expected 1:2 P1:P2 ratio, to a ratio closer to ~1:1. This decrease in P1:P2 ratio is not caused by a change in P1 levels, but is instead the result of accumulation of unprocessed P2 (pro P2, **Figure 2-5A**). Despite the defects in P2 processing, the total level of P2 (processed and unprocessed) remained unchanged, and the corrected ratio using P1: total P2 was ~1:2.4. These results collectively suggest that the P1 K49A substitution does not affect P1 or P2 expression or overall protamine levels in sperm, but does affect the amount of processed P2 in sperm. However, we found that P1$^{K49A/K49A}$ sperm had ~3.5 fold higher levels of histones retained than P1$^{+/+}$ (**Figure 2-5B,C**), suggesting that the P1 K49A protein disrupts overall histone eviction.

As H4 hyperacetylation is indispensable for histone-to-protamine exchange[48–52] we next asked whether such initial triggering events occurred normally in P1$^{K49A/K49A}$ mice. To this end, we analyzed P1$^{+/+}$ and P1$^{K49A/K49A}$ testes using an anti-H4 tetra-acetyl (referred to as ac-H4) antibody but observed no significant difference in ac-H4 levels by immunostaining (**Figure 2-5D, left**

panels) or immunoblotting of testes lysates (**Figure 2-6A**). Next, we investigated transition proteins (TNP1 and TNP2), well known chromatin intermediate components of the histone-to-protamine exchange. Specifically, loss of TNP1 and TNP2 perturbs sperm morphology, chromatin composition, and final chromatin packaging—similar to our observations in P1$^{K49A/K49A}$ mice.[12,13] When we analyzed TNP1 and TNP2 proteins in both testis cross-sections by immunostaining and testes lysates by immunoblotting, we did not observe any overt differences when comparing P1$^{+/+}$ and P1$^{K49A/K49A}$ mice (**Figure 2-5D, 2-6A,C**). Taken together, the P1$^{K49A/K49A}$ mutants appear to progress normally through several key intermediate processes, yet the ultimate chromatin packaging is strikingly abnormal. These observations raise the question of whether acetylation of K49 itself may be required in the remodeling process, or whether other intermediate histone variants may be improperly loaded.

We next asked whether the retained histones in P1$^{K49A/K49A}$ sperm are selectively enriched for specific histone PTMs which would indicate possible regional or programmatic alterations in histone eviction. To answer this question, we performed histone PTM immunoblots of protein extracts from increasing numbers of P1$^{+/+}$, P1$^{K49A/+}$ and P1$^{K49A/K49A}$ sperm, and probed for a series of both activating and repressive modifications including ac-H4, H3K27ac, H3K9me3, and H3K27me3, and H4K20me3 (**Figure 2-6B**). When comparing across genotypes, it was evident that all PTMs analyzed appeared to be globally increased in the P1$^{K49A/K49A}$ mutant, except for ac-H4, which was consistently lower in multiple biological replicates (**Figure 2-6B, data not shown**). In conclusion, the P1 K49A substitution does not compromise P1 protein stability, but rather the mutation causes functional changes to sperm chromatin, likely compromising the ability of protamines to compete against stably bound nucleosomes.

2.3.5 *P1 K49A substitution results in decreased blastocyst formation and accelerated P1 dismissal from paternal chromatin*

Given the severely reduced motility observed in the sperm of P1$^{K49A/K49A}$ males and severe subfertility, we next performed intracytoplasmic sperm injections (ICSI) to examine early developmental consequences of the P1 K49A substitution. In agreement with our natural mating data, blastocyst formation was significantly impaired when we injected P1$^{K49A/K49A}$ sperm (13.5% vs. 48.0% using P1$^{+/+}$ sperm, **Figure 2-7A**). When assessing embryo development and survival

every ~24 hours, we noticed two significant blocks in development in embryos derived from P1$^{K49A/K49A}$ sperm: the first at the 1-cell to 2-cell transition, and, surprisingly, the other from morula to blastocyst stages (**Figure 2-7A**). This later block is striking and suggestive of abnormal transcriptional or epigenetic landscapes caused by the P1 K49A substitution.

Since proper decompaction of the paternal genome and the replacement of protamines by histones is critical for embryonic development following fertilization, we first evaluated male pronuclear remodeling and protamine dismissal from paternal chromatin by immunostaining zygotes derived from either P1$^{+/+}$ or P1$^{K49A/K49A}$ sperm (collected 4 hours post-fertilization [hpf], outlined in **Figure 2-7B**). By DAPI staining alone, we found that the average relative pronuclear size (male/female) in zygotes derived from P1$^{K49A/K49A}$ sperm was significantly larger than in those derived from P1$^{+/+}$ sperm, suggesting potentially an accelerated decompaction of paternal chromatin in the mutant (**Figure 2-7C**).

If paternal chromatin decompaction is affected by the K49A substitution, we would then expect to observe differences in P1 removal in the zygote. When we stained both P1$^{+/+}$ and P1$^{K49A/K49A}$ zygotes at 4hpf with our custom P1 antibody, we observed differences in P1 distribution patterns. In P1$^{+/+}$ zygotes at 4hpf, P1 is more broadly and densely distributed throughout the male pronucleus and appears more often to be associated with DNA. In contrast, in P1$^{K49A/K49A}$ zygotes, P1 adopts a speckle-like distribution in the male pronucleus and the intensity of a concentrated (possibly phase separated) ring-like pattern immediately inside the prenucleolar body is higher (**Figure 2-7D**). When we measured the pixel distance from the edge of the DAPI signal to the most intense P1 signal as a proxy of P1 eviction, we found a significantly higher distance for P1$^{K49A/K49A}$ zygotes, further indicating an accelerated P1 dismissal (**Figure 2-7E, 2-8A**). Moreover, while ~24% of P1$^{+/+}$ measurements indicate complete overlap between P1 and DNA (a pixel distance of 0, scale corresponds to ~6 pixels/µm), only ~14% of P1$^{K49A/K49A}$ measurements exhibited complete overlap, with 71% of P1$^{K49A/K49A}$ measurements having a pixel distance >3 (compared to 50% of P1$^{+/+}$ zygotes, **Figure 2-8A**). The increased number of P1$^{K49A/K49A}$ zygotes with premature P1 dismissal is consistent with the higher percentage of arrested 1-cell zygotes in the mutant.

Curiously, while tracking P1 in both WT and mutant zygotes, we noticed that P1 localized to both female and male pronuclei. Moreover, P1 in P1$^{K49A/K49A}$ zygotes displays similar localization patterns in the female pronucleus as the male pronucleus (**Figure 2-7D, 2-8C**). Localization of protamines in the female pronucleus was previously reported but was assumed to be an antibody artifact.[53] To rule out this alternative explanation, we knocked in a V5 tag to the endogenous P1 locus to create an N-terminal V5-tagged P1 protein (V5-P1) and confirmed that the addition of the V5 tag on P1 does not affect sperm parameters or fertility (**Figure 2-8B**). We detected V5-P1 in both pronuclei using an anti-V5 antibody, consistent with our anti-P1 antibody staining and supporting that localization to the female pronucleus is not an antibody artifact (**Figure 2-8C**). The protamines detected in the female pronuclei are male derived since a reciprocal IVF (using P1$^{+/+}$ sperm and V5-P1 oocytes) revealed no V5 expression but P1 localization in both pronuclei (data not shown). Although localization to the female pronucleus is an intriguing observation, future studies are required to better understand its potential functional implications.

Altogether, our observations suggest that a single amino acid substitution in P1 results in an increased number of embryos arresting at the 1-cell stage and accelerated P1 removal in P1$^{K49A/K49A}$ zygotes. Thus, these results confirm the functional significance of the K49 residue *in vivo* and at the organismal level.

2.3.6 *The substitution of P1 K49 to alanine decreases P1 DNA binding ability*

Given our *in vivo* observations of the P1 K49A substitution and the difficultly of assessing how protamine-DNA binding or dynamics are regulated *in vivo*, we turned to bulk biochemical and single molecule assays *in vitro* to examine protamine-DNA interactions. Because of the high arginine content in protamines, generation of recombinant P1 and P2 proteins in bacteria in sufficient quantity and purity has been extremely challenging. To overcome these challenges, we developed a method to successfully purify P1 and P2 proteins (amino acid sequences shown in **Figure 2-9A**) from both P1$^{+/+}$ (referred to as WT P1 or WT P2) and P1$^{K49A/K49A}$ mature sperm (referred to as P1 K49A or pro P2) using a combination of acid extraction of basic proteins and size exclusion chromatography (**Figure 2-9B**). The combination of these two methods enabled purification and efficient separation of P1 and P2 not only from each other but also from other basic proteins such as histones (**Figure 2-10A,B**).

To test the binding affinity of WT P1 or P1 K49A to DNA, we performed electrophoretic mobility shift assays (EMSAs). Using a ~300 bp linear DNA we found that WT P1 robustly bound DNA in a concentration-dependent manner after 1 hour with a $K_{d,app}$ of 0.68 μM. In contrast, P1 K49A exhibited a marked decrease in DNA binding affinity, $K_{d,app}$ = 0.95 μM (**Figure 2-9C,D**). Since we noted that the protamine-DNA complex never entered the gel, regardless of experimental conditions, type of gel, or DNA fragment size (data not shown), we inferred that protamine-DNA complexes are forming large, higher order structures that preclude migration into the gel. To confirm this, we repeated the EMSA experiments, in the presence or absence of proteinase K. As expected, the addition of proteinase K dissolved the complex and restored movement of the DNA into the gel, suggesting that the well shift is representative of protamine-DNA interactions and not a technical artifact (**Figure 2-10H**).

Next, to ensure that the reaction had reached binding equilibrium, we performed EMSAs after incubating DNA and proteins for 10 minutes, 1 hour, or 4 hours, but as we did not observe any difference in binding for either WT P1 or P1 K49A, we concluded that the reaction reaches equilibrium within 10 minutes (**Figure 2-10D,E**). Interestingly, both P1 and P1 K49A appeared capable of cooperative binding behavior, as evidenced by Hill coefficients >1 (4.2 for WT P1 and 3.0 for P1 K49A). Interestingly, the lower values in Hill coefficients in the P1 K49A mutant might reflect a lower DNA affinity or defect in the ability of the mutant to initiate binding and/or polymerize on DNA. To explore whether this apparent cooperative behavior is a general property of protamine proteins or restricted to P1, we also assessed the binding of WT P2 and pro P2. Similarly, WT P2 and pro P2 also displayed cooperative-like behavior *in vitro* (Hill coefficients of 4.2 and 3.7, respectively). Moreover, WT P2 had a higher binding affinity, $K_{d,app}$ = 0.67 μM (similar to that of WT P1), than pro P2, $K_{d,app}$ = 0.84 μM (**Figure 2-9E, 2-10C**) and similarly neither binding affinity was affected by incubation time (**Figure 2-10F,G**).

Since mouse sperm (and most mammalian sperm) use both P1 and P2 to package chromatin, we next aimed to understand how the presence of both proteins influences their affinity to DNA, and specifically whether the P1 K49A substitution may alter this affinity. To this end, we repeated the EMSAs using WT P1 or P1 K49A in combination with either WT P2 or pro P2 in a 1:2 ratio

(P1:P2), the expected ratio in mouse. As expected, the combination of WT P1 and WT P2 bound more efficiently to DNA than either protein alone (**Figure 2-9F, 2-9G top left panel**). However, upon mixing WT P1 with pro P2, we observed a significant decrease in affinity and an overall shift in the binding curve (**Figure 2-9F, 2-9G bottom left panel**). When comparing the binding properties of P1 K49A with either WT P2 or pro P2, the binding appeared nearly identical, in both cases requiring a higher protein concentration (~1.2 uM of total protamine) to reach a fully bound state (**Figure 2-9F, 2-9G right panels**).

Taken together, we show that although P1 K49A maintains a cooperative binding mode, the mutant protein has a marked decrease in DNA binding affinity. Furthermore, we show that protamine-DNA binding affinity and cooperative behavior is enhanced in the presence of both P1 and P2 together, but the mutant P1 protein loses its preferred selectivity for mature P2 and instead can interact equally well with either P2 or pro P2.

2.3.7   *P1 K49A substitution causes altered DNA compaction and decompaction kinetics*

Since EMSAs ultimately measure an equilibrium between both the protein on rate (k_{on}) and off rate (k_{off}) and therefore cannot provide kinetic information, we turned to a DNA curtain assay. Here, we used DNA from bacteriophage λ (λ-DNA) to investigate the real time compaction and decompaction kinetics of wild type and mutant protamines at single molecule resolution (**Figure 2-11A**). Protamines are expected to bind to 10-15 bp sites [54] and notably the 50 kb of λ-DNA contains a large diversity of 10-15 bp sites which are all represented thousands of times within the mouse genome, therefore this DNA source allows us to probe representative and relevant general interactions between protamines and DNA. As protamines cause the dissociation of DNA-intercalating dyes such as YOYO-1 (data not shown), we instead labeled each λ-DNA molecule with a fluorescent dCas9 at the untethered end and monitored changes in DNA length to calculate compaction and decompaction (**Figure 2-11B**).

We found that WT P1 largely failed to initiate compaction at 100 nM, but induced robust and complete DNA compaction at 200 nM, with an average velocity of 1.57 μm/s (~6kbp/s) (**Figure 2-11C,E,G**). In contrast, P1 K49A failed to initiate compaction at 200 nM but achieved robust compaction at 275 nM (**Figure 2-11D,G,F**). In addition, the average velocity for P1 K49A at the

maximum rate of compaction was slower than WT P1 (1.09 μm/s at 275 nM compared to P1, 1.57 μm/s at 200 nM). Curiously, at the low/intermediate protein concentrations, both WT P1 and P1 K49A displayed a unique pattern: a few molecules condensed >10 kilobases of DNA (**Figure 6G**), but the majority of DNA molecules remained uncompacted, with a few strands only initiating compaction stochastically. At intermediate protein concentrations (125 nM – 175 nM) the extent of DNA compaction is non-uniform (**Figure 2-11G**), meaning that not all DNA strands within a single experiment compact to the same level and compaction exhibits a start-and-stop behavior (**Figure 2-11G**). Given that even directly adjacent DNA molecules separated by only a couple microns do not exhibit the same behavior (i.e. one molecule could compact completely, whereas the neighboring DNA molecule does not compact at all, **Figure 2-11H**), this highlights the need for a better understanding of mechanisms underlying cooperativity and competition in the context of protamine binding, and the local signals that license regions of DNA to compact. Therefore, consistent with the initial cooperative-like behavior we observed in the EMSA experiments, the DNA curtain experiments support that protamines, when present in a limited pool, preferentially bind to a small number of DNA molecules to generate a high level of compaction, as opposed to distributing evenly among all available DNA molecules to produce a uniform but low level of compaction.

Next, we assessed the kinetics of WT P2 and pro P2. Pro P2 initially binds to DNA in elongating spermatids and then undergoes proteolytic cleavage to generate mature P2 (WT P2). To this end, we found that WT P2 required 275 nM protein to achieve robust compaction, similar to P1 K49A, but compacted DNA at twice the rate of P1 K49A at this concentration (2.01 μm/s at 275 nM, **Figure 2-12A,C**). Pro P2 compacts DNA across a similar concentration range but at slightly slower rates (pro P2: 1.26 μm/s vs. WT P2: 1.97 μm/s at 250 nM, **Figure 2-12B,D,E**). At intermediate concentrations, and similar to WT and mutant P1, both WT P2 and pro P2 compact a fraction of DNA molecules, leaving many uncompacted (**Figure 2-12E**). However, the average extent of compaction generated by WT P2 is much greater (e.g. ~9 μm for WT P2 and ~4 μm for pro P2 after 5 seconds at 275 nM), consistent with increased genome compaction occurring once P2 has undergone processing (**Figure 2-12C,D**). Unlike for the other tested protamines, we found that WT P2 at intermediate concentrations initiated DNA compaction but then rapidly de-compacted (**Figure 2-12A,C,E**). This result suggests that the stability of WT P2-DNA complexes

is more sensitive to local concentration changes and initial WT P2-DNA complexes are stabilized by further protein binding. We surmise that these opposing characteristics of WT P2; robust DNA compaction and more sensitive concentration-dependent decompaction, may be central to the opposing roles of protamines in sperm versus zygotes. Overall, the compaction behavior we observe across protamines suggests that tight regulation of local protamine concentration provides a general mechanism for controlling chromatin condensation during spermiogenesis *in vivo.*

The DNA curtain experiments also allowed us to monitor DNA decompaction, the rate at which protamines passively dissociate from DNA. Here, we find that DNA condensed by P1 K49A decompacted significantly faster than DNA decompacted by WT P1 (0.97 μm/min, ~3.6 kbp/min for P1 K49A vs. 0.45 μm/min, ~1.7 kbp/min for WT P1) (**Figure 2-11I**). Likewise, pro P2-compacted DNA decompacted at a faster rate than DNA stably compacted by WT P2 (0.45 μm/min, ~1.7 kbp/min vs. 0.37 μm/min, ~1.4 kbp/min) (**Figure 2-12F**). In short, these data demonstrate that mutant P1, even at higher concentrations relative to WT P1, compacts DNA slower and dissociates from DNA faster, consistent with our bulk measurements of $K_{d,app}$. Furthermore, compared to pro P2, WT P2 requires more protein to initiate compaction, but compacts DNA at a faster rate and to a greater extent, and dissociates from DNA more slowly, again consistent with our bulk results. Moreover, our *in vitro* studies also show that P1 K49A-compacted DNA decompacts significantly faster than DNA compacted by WT P1, explaining our observation in embryos. Therefore, although electrostatics may be a prominent driver of sperm DNA condensation, other regulatory factors, like PTMs or individual residues, fine-tune the DNA compaction and decompaction to ensure correct packaging and developmental sequence of events.

2.4 Discussion

Efficient eviction of histones and subsequent addition of protamines to optimally package paternal chromatin during spermiogenesis is essential to safeguard fertility throughout life. Given the arginine-rich composition of protamines, previous work assumed a non-specific protamine-DNA binding mechanism, leaving the contribution of individual P1 or P2 residues to chromatin condensation in spermatids unresolved. Here, we pioneered a series of complementary molecular, genetic, biochemical, and biophysical assays to explore how the single amino acid substitution of P1 K49 to an alanine—a residue outside the central arginine core—systematically perturbs sperm

37

genome packaging. Our systematic *in vitro* and *in vivo* analysis of efficacy, development, and biophysical properties of the P1 protein support a possible regulatory role for the K49 residue. Based on our findings, we propose a reevaluation of the conventional view of protamines as purely electrostatic structural components to instead consider that protamine protein sequence variants outside of the arginine core may have evolved to execute species-specific and regulated packaging and unpackaging processes.

The conservation of K49 in P1 across the rodent lineage led us to hypothesize that it plays an essential and species-specific role in spermiogenesis and/or embryonic development, either through the K49 residue itself or through its acetylation. By using a modification-specific antibody against K49ac, we find that acetylation is acquired in early elongating spermatids, and that the K49A substitution leads to a ~3.5-fold increase in canonical histone retention and accumulation of pro P2. Given that the K49A substitution affects P1-DNA binding affinity, more P1 K49A protein is likely needed *in vivo* to overcome this decrease in affinity. Hence, it is possible that the efficiency of histone eviction is secondarily hindered, resulting in increased histone retention. Earlier studies have elegantly shown that histone acetylation is a prerequisite for spermatid maturation and histone-to-protamine exchange, and furthermore, disruption of the testes-specific dual bromodomain containing protein, BRDT—an acetyl-lysine reader—also precludes nucleosome eviction in round spermatids.[48,50,55,56] Here, we show that the P1 K49A mutation alters the sites of and possible mechanism underlying nucleosome eviction or retention. Our findings suggest that P1 K49ac may be important in the histone-to-protamine exchange process, thereby expanding the pool of factors implicated in this process. Furthermore, given that both bromodomains of BRDT are required to induce a large-scale acetylation-dependent chromatin reorganization in sperm, our future studies will explore the possibility that BRDT may interact with both acetylated histone H4 and acetylated P1 to modulate this process.

In addition to altered sperm chromatin composition, we find that fertility in P1[K49A/K49A] males is significantly decreased due to a near-total loss of progressive sperm motility. The morphological defects we observed in P1[K49A/K49A] sperm overlap largely with those observed in mice that are haploinsufficient for P1 (P1[+/-]), lack P2, or lack TNP1 and TNP2.[12,13,24,25] Interestingly, many of the morphological abnormalities localize to the midpiece (sperm head/neck connection), raising

the possibility that protamine incorporation is linked to cytoskeleton remodeling or manchette formation in elongating spermatids. This conclusion is also supported by the previously reported interaction between phosphorylated P1 and the inner nuclear membrane protein Lamin B receptor.[57] Hence, active cytoskeletal remodeling in the setting of nuclear remodeling in sperm could be a cellular mechanism that overrides the repulsive forces of positively charged proteins, such as protamines. Conversely, this remodeling process may increase local protamine concentration and enhance protamine-DNA cooperativity to ensure regulated initiation and polymerization of protamines on genomic segments—potentially ensuring that sperm DNA compaction is able to reach an energetically favorable structure while also maintaining instructive information for programmatic unfolding during development.

The large net positive charge of arginine-rich protamines remains a challenge to a more complex model of protamine behavior and regulation, as their interaction with DNA is undoubtedly electrostatic. Our data suggest that electrostatic interactions are important but are not the only determinants of protamine-DNA interactions. The differences in P1 DNA binding affinity we observed in protamines across species suggest a possible role for protamine sequence, protein-protein oligomerization, or possibly differences in protamine PTMs (subjects of future investigations). For example, *in vitro*, we observe that P1 K49A has a much lower binding affinity for DNA than WT P1. This impact from the loss of a single lysine residue is unexpected, bearing in mind that this protein contains more than 30 positively charged residues (**Figure 5C,D, 6C-F**).[58,59] Similarly, in the DNA curtain assay, we found that each protein displayed a sharp concentration dependence on the level of DNA compaction measured on DNA curtains. Moreover, compaction was not uniform across the curtain as was observed for DNA compaction by HP1.[60] Remarkably, even for DNA molecules side-by-side within our experiments, we observed drastically different levels of compaction when incubated with protamines, indicative of distinct levels of protamine binding on a molecule-to-molecule basis. Furthermore, small differences in protein concentration resulted in unexpectedly large changes in compaction velocity. For example, increasing WT P1 from 175 nM to 200 nM resulted in an increase in average velocity from 0.31 μm/s to 1.57 μm/s. Taken together, we hypothesize that protamines engage in cooperative binding modes with DNA and that tight regulation of their local concentration is a mechanism for achieving precise control over chromatin condensation during spermiogenesis *in vivo*. In addition to

differences in compaction rates, the P1 K49A protein dissociated from DNA significantly faster than DNA compacted by WT P1. Strikingly, this result is in agreement with the accelerated dismissal of P1 from the paternal genome in zygotes (**Figure 4D, S4D**) and suggests that modifications apart from phosphorylation may help to regulate the protamine-to-histone transition following fertilization.

Altogether, our findings highlight an indispensable role for P1 K49 in protamine biology, as illustrated by the significant perturbations in protamine-DNA interactions, sperm chromatin packaging, and embryonic development that occur when substituting K49 for alanine. Furthermore, these results highlight the potential that other amino acid residues in P1 outside the central arginine-rich DNA binding core may have functional consequences and perturb biological processes. Our observation that a single amino acid substitution can cause such dramatic alterations in the biophysical and functional properties of mouse P1 lends strong support to the conclusion that evolutionary changes in protamine protein sequences across species are unlikely to be neutral. Future studies are needed to test whether additional residues or PTMs have an impact on fertility. Curiously, sites of modification in both mouse and human protamines are enriched in the N- and C-terminal sequences flanking the arginine core and tend to be highly conserved within a species, but not across species. Therefore, it is conceivable that such non-arginine residues evolved to regulate the species-specific sperm genome packaging and subsequent unpackaging in the zygote to ensure both species compatibility upon fertilization and optimal organismal reproductive fitness.

2.5 Methods

Mice

All experiments using animals were carried out with prior approval of the University of Michigan Institutional Committee on Use and Care of Animals (Protocols: PRO00006047, PRO00008135, PRO00010000) and in accordance with the guidelines established by the National Research Council Guide for the Care and Use of Laboratory Animals. Mice were housed in the University of Michigan animal facility, in an environment controlled for light (12 hours on/off) and

temperature (21 to 23°C) with ad libitum access to water and food (Lab Diet #5008 for breeding mice, #5LOD for non-breeding animals).

P1[K49A/K49A] knock-in mice were generated on the C57BL/6N background using CRISPR/Cas9-mediated genome editing by the Cincinnati Children's Hospital Transgenic Animal and Genome Editing Core Facility. The sgRNA and donor oligo were designed as previously described.[61,62] The guide RNA target sequence was selected according to the on- and off-target scores provided by the web tool CRISPOR[61] (http://crispor.tefor.net) and proximity to the target site. Ribonucleoprotein (RNP) complexes were formed by mixing the sgRNA (80 ng/uL) with Cas9 protein (IDT, 120 ng/uL) in Opti-MEM (ThermoFisher) and incubating at 37 °C for 10 minutes, at which time the donor oligo (IDT, 500 ng/uL) containing the intended mutation was added. Zygotes from super-ovulated C57BL/6N females were electroporated with 7 uL of the RNP/donor oligo mix on ice using a Genome Editor electroporator (BEX; 30V, 1 ms width, 5 pulses with 1 s interval). Two minutes after electroporation, zygotes were moved to 500 ul cold M2 medium (Sigma), warmed to room temperature, and transferred to oviductal ampullas of pseudopregnant CD-1 females. All animal procedures were carried out in accordance with the Institutional Animal Care and Use Committee and approved protocol of Cincinnati Children's Hospital Medical Center. Offspring were genotyped for the P1 K49A mutation by extraction of genomic DNA from a small ear biopsy. Mutant males and control mice were used for all experiments between 8-16 weeks of age for all studies.

Antibodies

Rabbit polyclonal antibodies against total P1 and acetylation at P1 K49 were generated at GeneMed Synthesis Inc. via immunization of rabbits with the following synthesized peptides: P1-CRRRRSYTIRSKKY, P1 K49ac- CRRRRSYTIRCK(ac)KY. All other antibodies used are provided in Table 2-1.

Acid extraction of sperm basic proteins

Extraction of basic proteins from mature sperm was performed as previously described.[63] Briefly, sperm pellets were subjected to hypotonic lysis in 1 mM PMSF and subsequently spun down at 8,000xg for 8 min. Sperm pellets were then resuspended in 100 uL of 100 mM Tris pH 8.0, 20

mM EDTA, and 1 mM PMSF followed by denaturation of proteins with 100 uL of 6 M Guanidine-HCl, 575 mM DTT and alkylation with 200 uL of 522 mM sodium iodoacetate for 30 min in the dark. Protein pellets were then washed twice with 1 ml ice cold ethanol and extracted with 800 uL of 0.5 M HCl, 50 mM DTT at 37°C for 10 min. Supernatants were precipitated overnight at -20°C with TCA to a final concentration of 20%. The following day, precipitates were spun down at 12,000xg for 8 minutes and protein pellets were washed twice in 1 ml of 1% 2-mercaptoethanol in cold acetone. Final protein pellets were then resuspended in water.

Peptide competition assay to assess antibody specificity

Protamines were first acid-extracted using the method described above. An increasing amount of protein (0.5 ug and 1 ug for non-specific peptide, 1 ug and 3 ug for non-acetylated P1 peptide) was loaded on each immunoblot. Prior to adding to immunoblots, antibodies were incubated at room temperature for 30 minutes, with either 10-fold excess of specific or non-specific peptide, or alone with no peptide. After blocking, blots were then incubated for 1.5 hours in either antibody only, antibody with specific peptide, or antibody with non-specific peptide. The non-specific peptide used in these assays was N-DSNKEFGTSNESTE-C and the non-acetylated P1 peptide used was N-CRRRRSYTIRSKKY-C.

Mass spectrometry analysis of mouse protamines

Mass spectrometry of mouse sperm was performed at MS BioWorks in Ann Arbor, MI. Briefly, Protamines were first acid-extracted using the method described above. Approximately 20 ug of acid-extracted protein was run in triplicate on a 4-20% SDS-PAGE gel (BioRad) and a single band corresponding to P1 and P2 was cut out for processing. Gel bands were washed once with 25 mM ammonium bicarbonate followed by three washes in 100% acetonitrile. Bands were then reduced with 10 mM dithiothreitol at 60°C followed by alkylation with 50 mM light iodoacetamide at room temperature. Bands were then digested with either trypsin (Promega) at 37°C for 4 hours, Chymotrypsin (Promega) at 37°C for 12 hours, or Lys-C (Promega) at 37°C for 12 hours. For all enzymes used, digests were quenched with formic acid and the supernatants were analyzed. Digests were analyzed by nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher Q Exactive. Peptides were loaded on a trapping column and eluted over a 75 um analytical column at 350 nL/min. Both columns were packed with Luna C18 resin

(Phenomenex). The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 70,000 FWHM and 17,500 FWHM resolution, respectively. The fifteen most abundant ions were selected for MS/MS. Data were searched using a local copy of Byonic with the following parameters: Enzyme: Semi-Trypsin or None (for Chymotrypsin and Lys-C), Database: Swissprot Mouse (forward and reverse appended with common), fixed modification: carbamidomethyl (C), variable modifications: oxidation (O), acetyl (protein N-term), deamidation (NQ), phosphor (STY), methyl (KR), dimethyl (KR), trimethyl (K), mass values: monoisotopic, peptide mass tolerance (10 ppm), fragment mass tolerance (0.02 Da), max missed cleavages: 2. Mascot DAT files were parsed into the Scaffold software for validation, filtering and to create a non-redundant list per sample. Data were filtered using a minimum protein value of 95%, a minimum peptide value of 50% (Prophet scores) and requiring at least two unique peptides per protein. Site localization probabilities were assigned using A-Score.[64]

Evolutionary analysis of P1 sequence conservation

The phylogenetic tree was constructed using maximum likelihood (PhyML) inferred from P1 protein sequences for species across the orders Rodentia, Primate, and Artiodactyla, using the Whelan and Goldman matrix (WAG) substitution strategy. Sequences were downloaded from NCBI and aligned using standard parameters of MUSCLE.[65] Bootstrap support with 1,000 replicates is shown for each node, with values >95 indicating strong support.

Immunofluorescence and quantification of seminiferous tubule staging

Adult testes were fixed overnight in 4% PFA at 4°C before submerging in ethanol and processing for formalin fixed paraffin embedding (FFPE). Five-micron thick tissue sections were first deparaffinized followed by permeabilization and subsequent antigen retrieval via boiling in 10 mM sodium citrate pH 6.0 for 10 minutes. Following blocking in 1X PBS, 3% BSA, 500 mM glycine, sections were incubated with primary antibodies overnight at 4°C. PNA-Lectin (GeneTex) was used to stain acrosomes and DAPI was used as a nuclear counterstain. All AlexaFlour-conjugated secondary antibodies (Life Technologies/Molecular Probes) were used at 1:1000. For assessment of staging, seminiferous tubules were split into categories (I-III, IV-VI, VII-VIII, IX, X-XI, XII) according to their Lectin staining pattern and cell types present as previously described.[66]

Phenotypic assessment of P1$^{+/+}$, P1$^{K49A/+}$, and P1$^{K49A/K49A}$ males

All phenotyping was carried out in males between 64 and 71 days of age (9-10 weeks). Sperm were counted using a Makler chamber and performed as n=3 independent technical replicates per mouse (n=4 mice per genotype). For progressive sperm motility assessment, a minimum of 100 total sperm were counted and forward (progressive) movement was assessed in comparison to the total number of sperm counted in a total of n=4 mice per genotype. For quantification of fecundity, 8-week-old males (n=3 per genotype) were individually housed for 3 days before 8-week-old C57BL/6J females were added. Females (n=3 females per male, for a total of n=9 females per genotype of male) were checked daily for the presence of copulatory plugs and once plugs were noted, females were removed and placed in a new cage. The percent of females that were successfully impregnated was recorded (fecundity).

Acid urea gel electrophoresis for the separation of sperm basic proteins

Protamines were first acid-extracted from a fixed number of sperm cells per genotype as described above, with slight modification.[67] Following hypotonic lysis in 1 mM PMSF, sperm pellets were resuspended in 1 ml of 6 M Guanidine-HCl, 500 mM Hepes pH 7.5, 10 mM DTT for 1 hour at room temperature. Cysteines were then alkylated using vinylpyridine to a final concentration of 250 mM and incubated for 1.5 hours at room temperature. Proteins were then extracted with 0.9 M HCl and dialyzed overnight at 4°C against 0.2 M HCl. The following day, insoluble proteins were removed by centrifugation at 12,000xg for 5 minutes. Soluble proteins were then precipitated with TCA to a final concentration of 20% for 4 hours at -20°C. Precipitated proteins were then washed twice with acetone before being resuspended in 0.9 M acetic acid, 8 M urea, 100 mM betamercaptoethanol. Acid urea gels were prepared as previously described [63]. As an identical number of input sperm was used for extraction, an identical volume of protein was loaded for each genotype. P1:P2 ratios were calculated using ImageJ.

Sperm protein extraction for the assessment of histone retention

Histone levels in P1$^{+/+}$, P1$^{K49A/+}$, and P1$^{K49A/K49A}$ sperm (sperm pooled from 2-3 animals per genotype) were assessed as previously described.[50] Briefly, sperm pellets were resuspended in lysis buffer (20 mM Tris pH 7.5, 1 mM MgCl$_2$, 1 mM CaCl$_2$, 137 mM NaCl, 10% glycerol, 1%

NP-40, 12.5 U/ml Benzonase, 1X protease inhibitors), sonicated briefly, and rotated for 1 hour at 4°C. For immunoblotting, lanes were loaded by input number of sperm cells. Due to variability between antibodies, the following sperm numbers were loaded for each corresponding antibody: histone H3, histone H2B, ac-H4, H4K20me3, and H3K9me2: 25000, 50000,10000; histone H4: 100000, 250000, 400000; H3K27ac and H3K27me3: 250000, 500000, 750000. Each blot was probed for alpha tubulin as a loading control to ensure comparative loading.

Protamine purification and *in vitro* electrophoretic mobility shift assays
Acid-extraction of sperm basic proteins was first performed as described above. Following TCA precipitation, the protein pellet was resuspended in 50 ul of water and brought up to 500 ul in gel filtration buffer (25 mM Hepes pH 7.5, 150 mM NaCl, 5 mM TCEP (TCEP was not pH neutralized). The solution was then subsequently subjected to size exclusion chromatography using a Superdex S75 column. Peak fractions were identified by absorbance at 214 nm and confirmed by immunoblotting. For *in vitro* electrophoretic mobility shift assays, varying concentrations of purified protamines were incubated with 40 nM DNA (280 bp) after briefly incubating the proteins at 37°C for 10 minutes in reaction buffer. DNA was prepared by PCR amplification of mouse genomic DNA using the primers specified in Table 2-4. After 1 hour of incubation, EMSA reactions were then run on a non-denaturing 0.5X TBE 6% polyacrylamide gel and stained with ethidium bromide (Sigma). Band intensities were quantified using ImageJ.

DNA Curtains
Microfluidic devices were constructed, and DNA curtain assays were performed as previously described.[60,68] Briefly, a lipid bilayer was coated on the surface of the sample chamber and biotinylated phage DNA was anchored to biotinylated lipids within the bilayer via streptavidin. DNA were then aligned at microfabricated barriers using buffer flow. In all experiments, care was taken to ensure that DNA molecules were separated by at least 2µm to prevent protamine interactions across DNA molecules.

DNA labeling with Cas9
Recombinant dCas9 protein was purchased from IDT (Alt-R S.p. dCas9 protein V3). The dCas9 protein was loaded with a dual guide RNA according to IDT's "Alt-R CRISPR-Cas9 system - in

vitro cleavage of target DNA with RNP complex" protocol with slight modifications for DNA curtains. dCas9 protein was diluted to 200nM and incubated with gRNA targeting position 47,752 (AUCUGCUGAUGAUCCCUCCG) at a 1:10 ratio in imaging buffer (1mg/mL BSA, 40mM Tris-HCl, pH=7.5, 50 mM NaCl, 5mM MgCl2, and 1mM DTT) and incubated on ice for at least 15min. Then, anti-HisAlexa555 (Invitrogen) was added to the reaction at a 1:2 ratio to dCas9 and incubated in the dark for 15min at room temperature. The fluorescent RNP complex was then diluted to 4nM in imaging buffer and incubated in the flow cell with DNA for 10 minutes. Finally, to remove nonspecifically bound proteins, the flow cell was washed with 500uL of imaging buffer containing 100ug/mL Heparin.

DNA compaction and decompaction experiments

DNA were maintained in flow at a rate of 0.6mL/min (average extension to 90% of contour length) for the duration of compaction and decompaction experiments. Prior to introduction into the flowcell, protamines were incubated at 37°C for 15 min in imaging buffer. Then, protamine was injected into the flowcell, images were collected at 10Hz, and compaction was monitored by tracking the motion of dCas9 molecules. Immediately following compaction, collection was shifted to 0.2Hz and decompaction was monitored by tracking the position of dCas9 molecules.

Intracytoplasmic sperm injections and embryo immunofluorescence

Oocyte collection, sperm collection, and piezo-actuated intracytoplasmic sperm injections were performed as previously described.[69] Embryos were cultured for 4 hours in KSOM (Millipore-Sigma) prior to immunofluorescence analysis.

For immunofluorescence, embryos were collected at the indicated timepoints after washing in KSOM, treated briefly (30 seconds-1 minute) with Acidic Tyrodes solution (EMD Millipore) to remove the zona pellucida, and fixed in 4% PFA for 10 minutes. Embryos were washed and gently permeabilized in PBS supplemented with 0.1% Triton and 3% BSA overnight at 4 °C before a 1-hour permeabilization in PBS supplemented with 0.5% Triton and 3% BSA. Embryos were then blocked in PBS with 0.1% Triton, 3% BSA, and 10% fetal bovine serum for 1 hour and stained with primary antibodies overnight in blocking buffer at 4°C. The following day, embryos were washed in PBS with 0.1% Triton and 3% BSA five times for 15 minutes each, followed by

incubation in secondary antibodies (Life Technologies/Molecular Probes) and DAPI (Sigma) for 2 hours at room temperature. All images were taken on a Nikon A1R-HD25 confocal microscope and processed with ImageJ.

2.6 Acknowledgements

Table 2-1 List of reagents and resources used in Chapter 2

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Protamine 1	This study	N/A
Protamine 1 K49ac	This study	N/A
Protamine 1	Briarpatch Biosciences	Cat# Hup1N RRID: AB_2651186
Protamine 2	Briarpatch Biosciences	Cat#: Hup2B RRID: AB_2687949
Histone H2B	Abcam	Cat#:ab52484 RRID: AB_1139809
Histone H3	Abcam	Cat#: ab1791 RRID: AB_302613
Histone H4	Protein Tech	Cat#: 16047-1-AP RRID: AB_2118625
H3K9me3	Abcam	Cat#: ab8898 RRID: AB_306848
Tetra-acetyl H4	EMD Millipore	Cat#: 06-866 RRID: AB_310270
H3K27ac	Abcam	Cat#: ab4729 RRID: AB_3118291
H3K27me3	Cell Signaling	Cat#: 9733s RRID: AB_2616029
H4K20me3	Abcam	Cat#: ab9053 RRID: AB_306969
Alpha tubulin	Protein Tech	Cat#: 66031-1-Ig RRID:AB_11042766
H3K9me2	Abcam	Cat#: ab1220 RRID: AB_449854
Tnp1	Protein Tech	Cat#17178-1-AP RRID: AB_2206757
Tnp2	Santa Cruz	Cat#: sc- 393843 RRID: N/A
V5	BioRad	Cat#: MCA1360 RRID: AB_322378
PNA-Lectin	Genetex	Cat#GTX01508 RRID: N/A
His Tag (clone 4E3D10H2/E3)-AlexaFluor555	Invitrogen	Cat# MA1-135-A555 RRID: AB_2610636
Chemicals, peptides, and recombinant proteins		
Trichloroacetic acid	Sigma	Cat#**T0699**
4-vinylpyridine	Sigma	Cat#V3204
Micrococcal Nuclease	Affymetrix/USB	Cat#AAJ70196ZCR
Pregnant Mare Serum Gonadotropin	Prospec Protein Specialist	Cat#Hor-272

Human Chorionic Gonadotropin	Sigma	Cat#CG5-1VL
KSOM	Millipore-Sigma	Cat#MR-106
FspI	New England Biolabs	Cat#R0135
dCas9	Integrated DNA technologies	Cat# 1081067
Benzonase	EMD Millipore	Cat#71206
Acidic Tyrodes Solution	EMD Millipore	Cat# MR-004-D
P1 peptide	This study	See Table 2-3
P1 K49ac peptide	This study	See Table 2-3
Tcf21 (non-specific protein) peptide	This study	See Table 2-3
Critical commercial assays		
BLOXALL® Endogenous Peroxidase and Alkaline Phosphatase Blocking Solution	Vector Labs	Cat#SP-6000-100
Accel-NGS 2S Plus DNA Library Kit	Swift Biosciences	Cat#21024
Swift Biosciences Indexing Kit	Swift Biosciences	Cat#26148
AMPure beads	Beckman Coulter	Cat#A63880
Experimental models: Organisms/strains		
C57BL/6J	Jackson Labs	Stock #000664
P1$^{K49A/K49A}$	This study	N/A
P1$^{V5/+}$	This study	N/A
Oligonucleotides		
P1 K49A genotyping primer_Forward, see Table 2-4	This study	
P1 K49A genotyping primer_Reverse, see Table 2-4	This study	
sgRNA for P1$^{K49A/K49A}$ generation, see Table 2-4	This study	
Donor DNA for P1$^{K49A/K49A}$ generation, see Table S3	This study	
280 bp dsDNA oligo, see Table 2-4	This study	
Bacteriophage λ (λ-DNA)	New England Biolabs	Cat#N3011
Software and algorithms		
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
GraphPad Prism 9	Graphpad Software	N/A
Scaffold PTM	Scaffold PTM software	N/A

Table 2-2 Mass spectrometry analysis of mouse sperm protamines

Peptide sequence	Variable Modifications	Localization probability	Ascore	Peptide score	Scaffold:Peptide Probability	NTT	Actual Mass	Observed Mass	Charge	Delta AMU	Delta PPM	Start	Stop
RsYTIRcK	S2 Phospho, C7 Carbamidomethyl	100%, 100%	55.92, 1,000.00	106.35	0.979	2	1162.529848	582.2722	2	-0.00202	-1.74	42	49
RSYtIRcK	T4 Phospho, C7 Carbamidomethyl	50%, 100%	0.00, 1,000.00	37.61	0.7833	2	1162.533448	582.274	2	0.00158	1.36	42	49
RSYtIRcK	T4 Phospho, C7 Carbamidomethyl	0%, 100%	0.00, 1,000.00	77.59	0.99	2	1162.529772	388.5172	3	-0.00209	-1.8	42	49
RSYTIRcK	C7 Carbamidomethyl	100%	1,000.00	233.32	0.99	2	1082.564172	361.862	3	-0.00136	-1.26	42	49
RSYTIRcK	C7 Carbamidomethyl	100%	1,000.00	226.74	0.99	2	1082.564472	361.8621	3	-0.00106	-0.981	42	49
RSYTIRcK	C7 Carbamidomethyl	100%	1,000.00	100.14	0.9525	2	1082.564648	542.2896	2	-0.000887	-0.819	42	49
RSYTIRcK	C7 Carbamidomethyl	100%	1,000.00	139.47	0.9812	2	1082.564848	542.2897	2	-0.000687	-0.634	42	49
SYtIRcK	T3 Phospho, C6 Carbamidomethyl	33%, 100%	0.00, 1,000.00	48.03	0.9	2	1006.429848	504.2222	2	-0.000919	-0.912	43	49
SYtIRcK	T3 Phospho, C6 Carbamidomethyl	33%, 100%	0.00, 1,000.00	26.49	0.7087	2	1006.436248	504.2254	2	0.00548	5.44	43	49
SYtIRcK	T3 Phospho, C6 Carbamidomethyl	98%, 100%	20.98, 1,000.00	48.34	0.9	2	1006.437072	336.4863	3	0.00631	6.26	43	49
SYtIRcK	T3 Phospho, C6 Carbamidomethyl	100%, 100%	27.96, 1,000.00	68.42	0.9	2	1006.430248	504.2224	2	-0.000519	-0.515	43	49
SYtIRcK	T3 Phospho, C6 Carbamidomethyl	97%, 100%	17.98, 1,000.00	46.33	0.7746	2	1006.429572	336.4838	3	-0.00119	-1.19	43	49
SYtIRcK	T3 Phospho, C6 Carbamidomethyl	100%, 100%	26.20, 1,000.00	64.85	0.9	2	1006.431648	504.2231	2	0.000881	0.875	43	49
SYTIRck	C6 Carbamidomethyl, K7 Methyl	100%, 0%	1,000.00, 0.00	167.24	0.9	2	940.479248	471.2469	2	-0.000837	-0.889	43	49
SYTIRck	C6 Carbamidomethyl, K7 Methyl	100%, 50%	1,000.00, 0.00	120.55	0.8889	2	940.477872	314.4999	3	-0.00221	-2.35	43	49
SYTIRcK	C6 Carbamidomethyl	100%	1,000.00	191.63	0.9	2	926.463048	464.2388	2	-0.00139	-1.5	43	49
SYTIRcK	C6 Carbamidomethyl	100%	1,000.00	125.89	0.9	2	926.467572	309.8298	3	0.00314	3.38	43	49
SYTIRcK	C6 Carbamidomethyl	100%	1,000.00	170.73	0.9	2	926.467048	464.2408	2	0.00261	2.82	43	49
SYTIRcK	C6 Carbamidomethyl	100%	1,000.00	136.23	0.9	2	926.463372	309.8284	3	-0.00106	-1.15	43	49
SYTIRckK	C6 Carbamidomethyl, K7 Acetyl	100%, 50%	1,000.00, 0.00	65.21	0.7447	2	1096.567272	366.5297	3	-0.00273	-2.49	43	50
SYtIRckKY	T3 Phospho, C6 Carbamidomethyl, K7 Acetyl	96%, 100%, 33%	18.44, 1,000.00, 0.00	27.44	0.5609	2	1339.605972	447.5426	3	0.00634	4.73	43	51
SYtIRcKKY	T3 Phospho, C6 Carbamidomethyl	98%, 100%	21.95, 1,000.00	49.25	0.5292	2	1297.594872	433.5389	3	0.00581	4.47	43	51
SYtIRcKKY	T3 Phospho, C6 Carbamidomethyl	100%, 100%	27.96, 1,000.00	71.06	0.871	2	1297.597448	649.806	2	0.00838	6.45	43	51
SYTIRckKY	C6 Carbamidomethyl, K7 Acetyl	100%, 1%	1,000.00, 0.00	112.44	0.8213	2	1259.638872	420.8869	3	0.00557	4.42	43	51
SYTIRcKKY	C6 Carbamidomethyl	100%	1,000.00	112.26	0.9499	2	1217.628848	609.8217	2	0.00611	5.02	43	51
YTIRcK	C5 Carbamidomethyl	100%	1,000.00	82.38	0.5	2	839.431248	420.7229	2	-0.00119	-1.41	44	49
TIRckKY	C4 Carbamidomethyl, K5 Acetyl	100%, 50%	1,000.00, 0.00	80.74	0.8927	2	1009.537572	337.5198	3	-0.00043	-0.426	45	51
TIRckKY	C4 Carbamidomethyl, K5 Acetyl	100%, 100%	1,000.00, 49.90	140.68	0.9	2	1009.542248	505.7784	2	0.00425	4.2	45	51
TIRcKkY	C4 Carbamidomethyl, K6 Acetyl	100%, 100%	1,000.00, 24.95	86.57	0.9	2	1009.539672	337.5205	3	0.00167	1.65	45	51
TIRcKKY	C4 Carbamidomethyl	100%	1,000.00	102.9	0.9	2	967.526172	323.516	3	-0.00126	-1.3	45	51
TIRcKKY	C4 Carbamidomethyl	100%	1,000.00	83.44	0.8741	2	967.526172	323.516	3	-0.00126	-1.3	45	51
TIRcKKY	C4 Carbamidomethyl	100%	1,000.00	159.99	0.9	2	967.526248	484.7704	2	-0.00119	-1.23	45	51

Table 2-3 Peptides used in Chapter 2

Peptide	Sequence
Protamine 1	CRRRRSYTIRSKKY
Protamine 1 K49ac	CRRRRSYTIRSK(ac)KY
Tcf21 (non-specific protein)	DSNKEFGTSNESTE

Table 2-4 Oligonucleotides used in Chapter 2

Oligonucleotide	Sequence
P1 K49A genotyping primer, Forward	5'-GAAGATGTCGCAGACGGAGGAG-3'
P1 K49A genotyping primer, Reverse	5'-ACGGTGGCATTTTTCAAGATGTG-3'
sgRNA for P1$^{K49A/K49A}$ generation	5'-GCCGCCGCTCATACACCATAAGG-3'
Donor DNA for P1$^{K49A/K49A}$ generation	5'-TTCTAGGATGCTGCCGTCGCCGCCGCTCATACACAATTAGATGCGCAAAATACTAGATGCACAGAATAGCAAGTCCATCAAAACTCCTGCGTGAGAATTTTACCAGACTTCAAGAGCATCTCGCC-3'
280 bp dsDNA oligo used for EMSAs	5'CTGGACAAGACCATGAACGCGAGGAGCAGGGGCAGGGGCAAGGGCTGAGCCCAGAGCGCGTAGAGGACTATGGGAGGACACACAGGGGCCACCACCACCACAGACACAGGCGCTGCTCTCGTAAGAGGCTACATAGGATCCACAAGAGGCGTCGGTCATGCAGAAGGCGGAGGAGACACTCCTGCCGCCACAGGAGGCGGCATCGCAGAGGTAAGCACCCCACAGCCGACCCCCTGGCCACCTGTGCTGCTGCTGCCCATCTAAACCCTGCTGCCTTCCA-3'

A

Protamine 1

MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCCRRRRRRRCCRRRRSYTIRCKKY
9   43 45   49 50

● Phosphorylation   ■ Methylation   ▲ Acetylation

B

Salmon	PRRRRSSSRPVR----RRRRPRVSRRRRRRGGRRRR
Guinea Pig	MARYRCCRSPSRSRC-RRRRRFYRRRRRCHRRRR---RCCRRR--YTR-RCKRY
White-Footed Mouse	MARYRCCRSKSRSRC-RRRRRRRCRRRRRRRCRRRRR--RCCRRRR-YTL-RCKRY
Mongolian Gerbil	MARYRCCRSKSRSRC-RRRRRRRCRRRRRRRCRRRRR--RCCRRRR-YTL-RCKRY
Norway Rat	MARYRCCRSKSRSRC-RRRRRRRCRRRRRRRCRRRRR--RCCRRRRSYTF-RCKRY
Prairie Deer Mouse	MARYRCCRSKSRSRC-RRRRRRRCRRRRRRRCRRRRR--RCRRRRRSYTI-RCKRY
Mouse	MARYRCCRSKSRSRC-RRRRRRRCRRRRRRRCRRRRR--RCRRRRRSYTI-RCKKY
Blind Mole Rat	MARYRCCRSKSRSRC-RRRRRRRCRRRRRRRCRRRRR--RCRRRRRSYTI-RCKRY
Pig	MARYRCCRSPSRSRC-RPRRRRCRRRRRRRCCPRRRRA-VCCRRY--TVI-RCRRC
Bull	MARYRCCLTHSGSRCRRRRRRRCRRRRRRRRCRRRRR---TVI-RCTRQ
Rhesus Macaque	MARYRCCRSRSRSRC-CRRRRRCRRRRRRRCRARRRAMRCCRRR--YRL-RCRRY
Common Marmoset	MARYRCCRSQSRSRC-YRQRRGRRRRRRSCRRRRAS-RCCRRR--YKL-PCRRY
Human	MARYRCCRSRSRSRY-YRQRQRSRRRRRRSCQTRRRAMRCCRPR--YRP-RCRRH
Chimpanzee	MARYRCCRSRSRSRC-YRQRQRSRRRRRRQSCQTQRRAMRCCRRR--SRM-RRRRH
Bonobo	MARYRCCRSQSRSRC-YRQR-RSRRRKRQSCQTQRRAMRCCRRR--SRL-RRRRH

Mus musculus (house mouse)	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
WSB/EiJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
PWK/PhJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
NZO/HlLtJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
ND/ShiLtJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
LP/J	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
FVB/NJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
DBA/2J	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
CBA/J	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
CAST/EiJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
C57BL/6NJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
C57BL6	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
C3H/HeJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
BALB/cJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
AKR/J	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY
129S1/SvImJ	MARYRCCRSKSRSRCRRRRRRRCRRRRRRRCRRRRRRRCCRRRRSYTIRCKKY

C

DAPI   Lectin   P1 K49ac   DAPI Lectin / P1 K49ac

IX
X-XI
XII
I-III
IV-VI
VII-VIII

Spermatid maturation

D

Stage Specificity

Frequency

100

50

0

IX X-XI XII I-III IV-VII VII-VIII

Seminiferous Tubule Stage

E

	Testis		Epididymal sperm		Vas sperm	
	5 ug	10 ug	0.25 ug	0.75 ug	0.25 ug	0.75 ug

20 — P1 K49ac

20 — P2

52
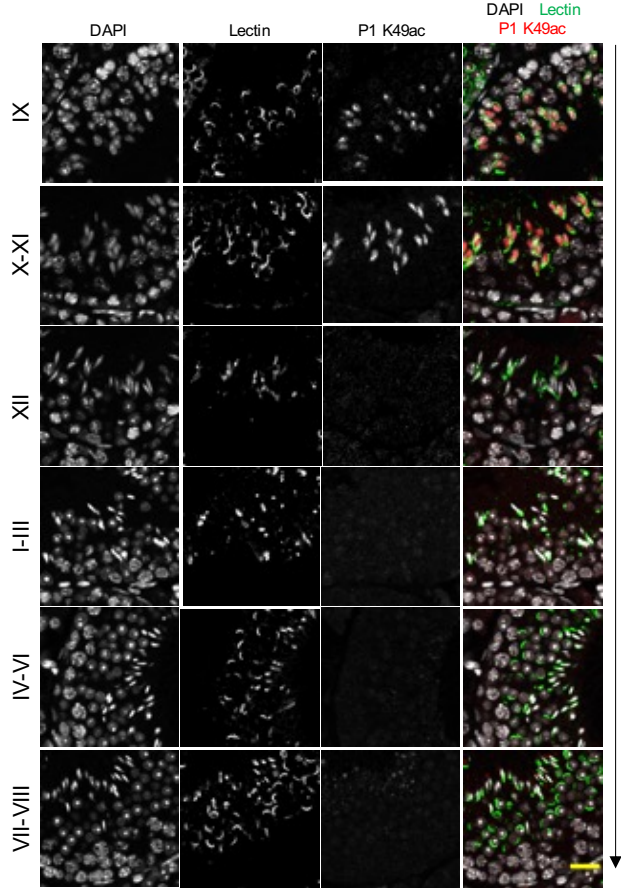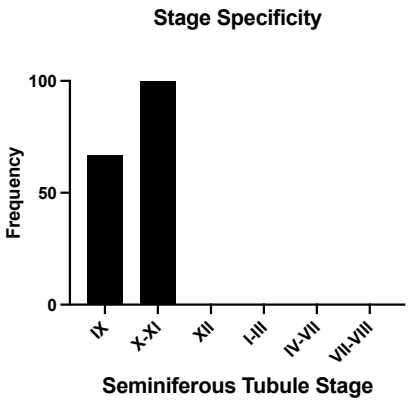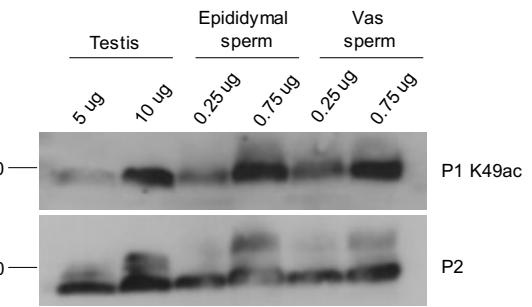
Figure 2-1: P1 lysine 49 acetylation is acquired in the testis in a stage-specific manner and is present in mouse sperm. (A) Schematic representation of modifications identified on mouse P1 using bottom-up mass spectrometry. **(B)** Phylogenetic tree constructed using maximum likelihood inferred from P1 protein sequences for species across the orders Rodentia, Primate, and Artiodactyla, using the WAG substitution strategy. Bootstrap support with 1,000 replicates is shown for each node, with values >95 indicating strong support. S9 is highlighted across species in blue and K49 is highlighted across rodents in green and in gray across more distant species that occupy alternative residues at this site. **(C)** Immunofluorescence staining of P1 K49ac in adult testes cross sections at various seminiferous tubule stages using PNA-Lectin as the acrosomal marker. Representative images from n=4 mice. Scale bar: 20 μm. **(D)** Quantification of P1 K49ac stage specificity across all stages, highlighting specificity to stage IX-XI tubules. A total of n=438 tubules were counted across all stages from a total of n=4 mice. **(E)** Western blot analysis of P1 K49ac from elongating spermatid-enriched testes lysate, mature sperm from the epididymis, and mature sperm from the vas deferens highlights the persistence of the acetylation mark into mature sperm.
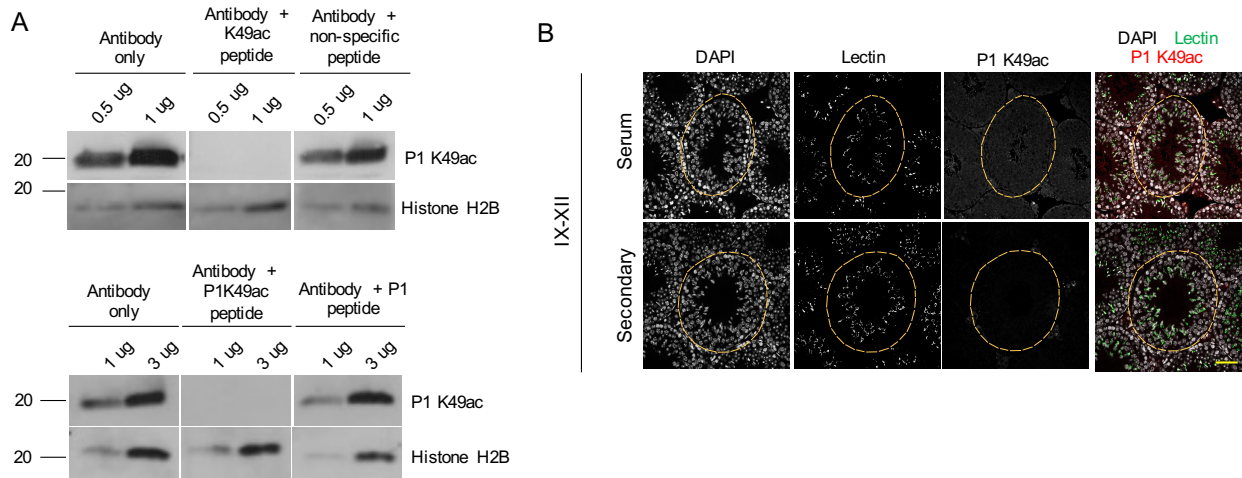
Figure 2-2: P1 K49ac antibody is highly specific. (A) Immunoblot of acid extracted protein lysates from mature sperm illustrates a clear band for P1 K49ac that is competed off only in the presence of a specific peptide containing acetylated P1 at K49. Top western blot was performed using a non-specific peptide from an unrelated protein and bottom western blot was performed using a P1 non-acetylated peptide. **(B)** Immunofluorescence of adult testes cross sections using pre-immune serum from rabbits used for antibody generation (top panels) or secondary antibody only (bottom panels) further illustrates specificity of the antibody. Dotted yellow line indicates perimeter of the seminiferous tubule. Scale bar: 50 μm.
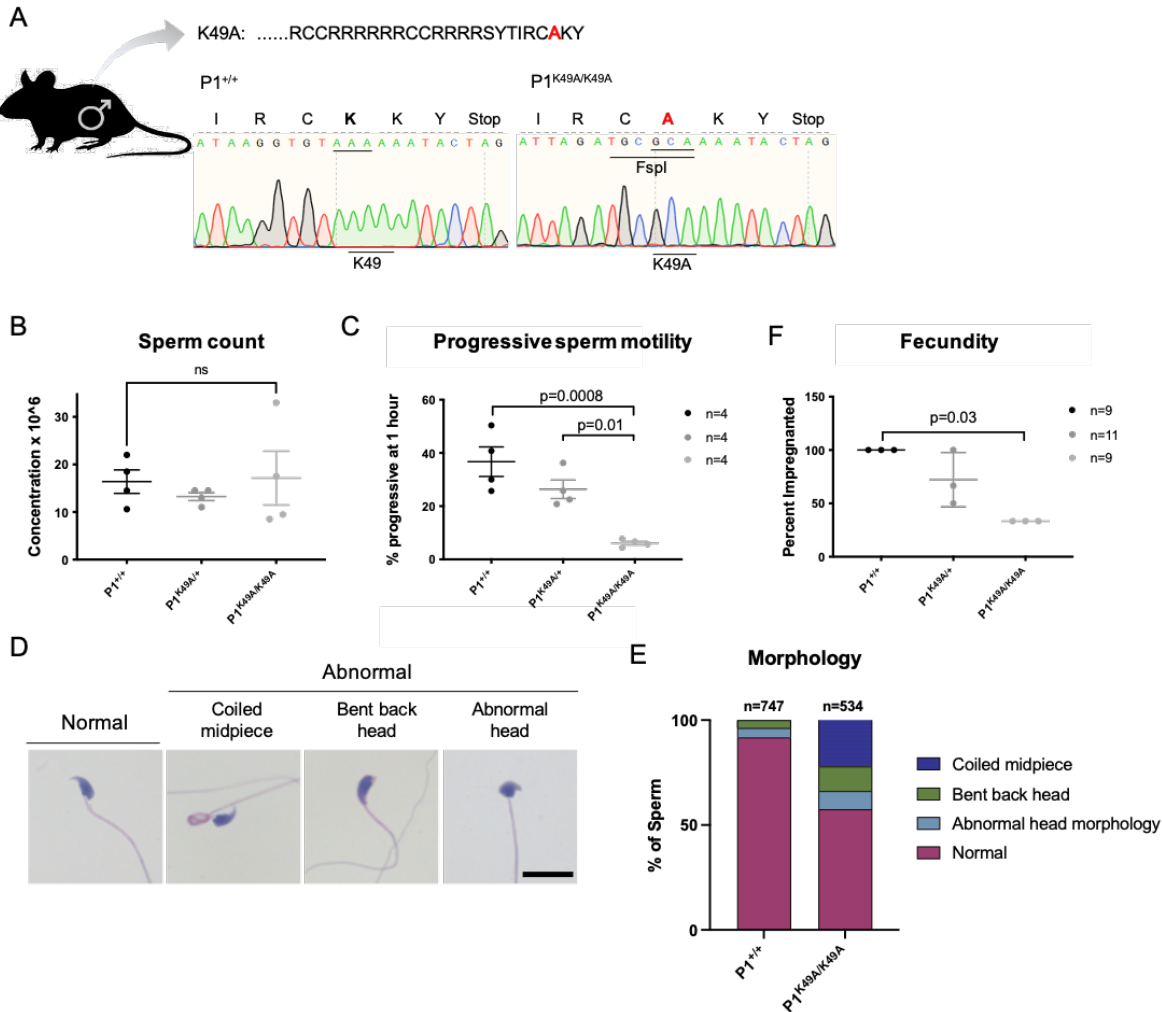
Figure 2-3: P1 K49A substitution results in sperm motility defects and subfertility. (A) Schematic of modification made to the mouse P1 sequence and corresponding Sanger sequencing traces illustrating successful mutation of K49 to alanine. Note that a few synonymous mutations were incorporated into the donor DNA to introduce an FspI site for genotyping purposes. **(B-C)** Total epididymal sperm count (**B**, n=4 for each genotype) and epididymal sperm progressive motility after 1 hour of incubation at 37 °C (**C**, n=4 for each genotype.) Each dot represents measurement from a single animal. Statistical tests were performed using a one-way ANOVA and adjusted for multiple comparisons. Center line represents the mean and error bars represent standard deviation. **(D)** Representative Hematoxylin and eosin-stained mature sperm from a P1$^{K49A/K49A}$ adult male, highlighting observed major abnormalities observed. Scale bar: 20 μm. **(E)** Quantification of major abnormalities observed in P1$^{+/+}$ and P1$^{K49A/K49A}$ mature sperm. Sperm was assessed from n=3 P1$^{+/+}$males and n=3 P1$^{K49A/K49A}$ males. **(F)** Fertility assessment of 3 adult males per genotype as measured by percent of females impregnated (fecundity). Each dot represents measurement from a single animal. Statistical test was performed using a Kruskal-Wallis test and was adjusted for multiple comparisons, p=0.03. Center line represents the mean and error bars represent standard deviation.

Genomic location	Number of mismatches	Sequence (including mismatches)	Genomic location
Chr6: 129535854	0	GCAGTGGCTCATACACCATAGGG	Intergenic
Chr12: 27066405	0	GCTACCACTCTTACACCATAGGG	Intergenic
Chr12: 5373801	0	GCCGCCTCGCAAACACCATAGGG	Intron: Klhl29
Chr10: 129362223	0	GCTGTCGATAATACACCATAGAG	Intergenic
Chr14: 102981772	0	ACCGCGGCTCCTACACCATCGGG	Exon: Kctd12
Chr6: 72119256	0	GCAGCCTCTCCTACACAATAAGG	Intron: St3gal5
Chr1: 167267189	0	GCCCCCTCCCATACACCACAGGG	Intron: Uck2
Chr2: 168467316	0	GCCACCACTCAGACACCAGATGG	Intergenic
Chr12: 4868024	0	GTCGCCCCTCATCCACAATAAGG	Intron: Mfsd2b

B
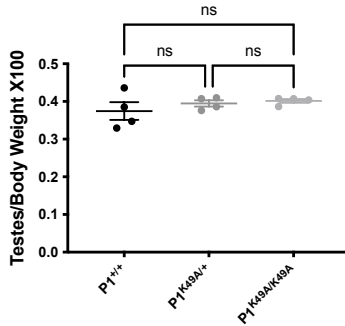
C
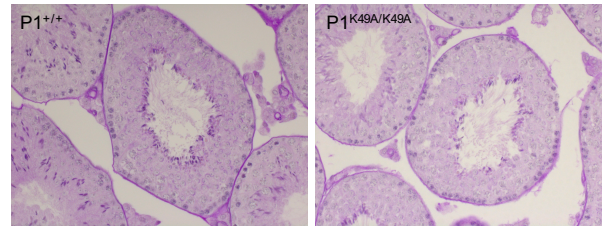
D
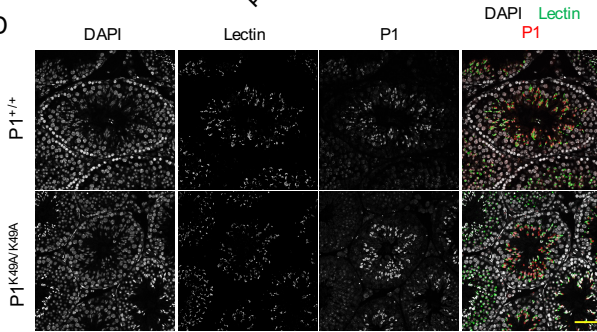
E
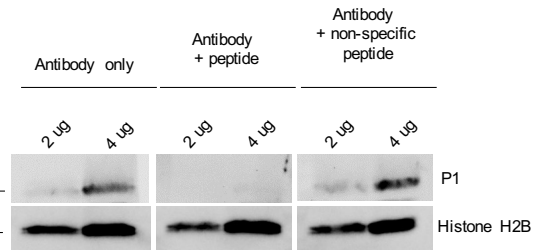
F
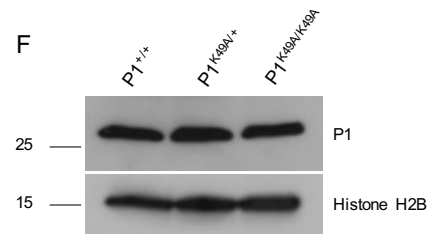
Figure 2-4: P1 K49A substitution does not affect testis parameters or P1 expression. (A) List of potential off-targets and corresponding sequencing results verify no off-target modifications generated by CRISPR/Cas9 editing. **(B)** Testes/body weight ratio of $P1^{+/+}$, $P1^{K49A/+}$, and $P1^{K49A/K49A}$ males (n=4 per genotype) suggests no loss of germ cell populations due to P1 K49A substitution. Statistical test was performed using a one-way ANOVA and adjusted for multiple comparisons, p=0.4391. Center line represents the mean and error bars represent standard deviation. **(C)** Periodic acid Schiff (PAS)-stained adult testes cross sections highlights normal testis morphology in $P1^{K49A/K49A}$ males. **(D)** Immunofluorescence of adult testes cross sections from $P1^{+/+}$ and $P1^{K49A/K49A}$ testes stained for P1 shows no loss of P1 expression upon substitution of K49 for alanine. Scale bars: 50 μm. **(E)** Peptide competition immunoblot for custom P1 antibody highlights antibody specificity. **(F)** Immunoblot of RIPA-extracted testes from $P1^{+/+}$, $P1^{K49A/+}$, and $P1^{K49A/K49A}$ males shows comparable expression of P1 across all genotypes.
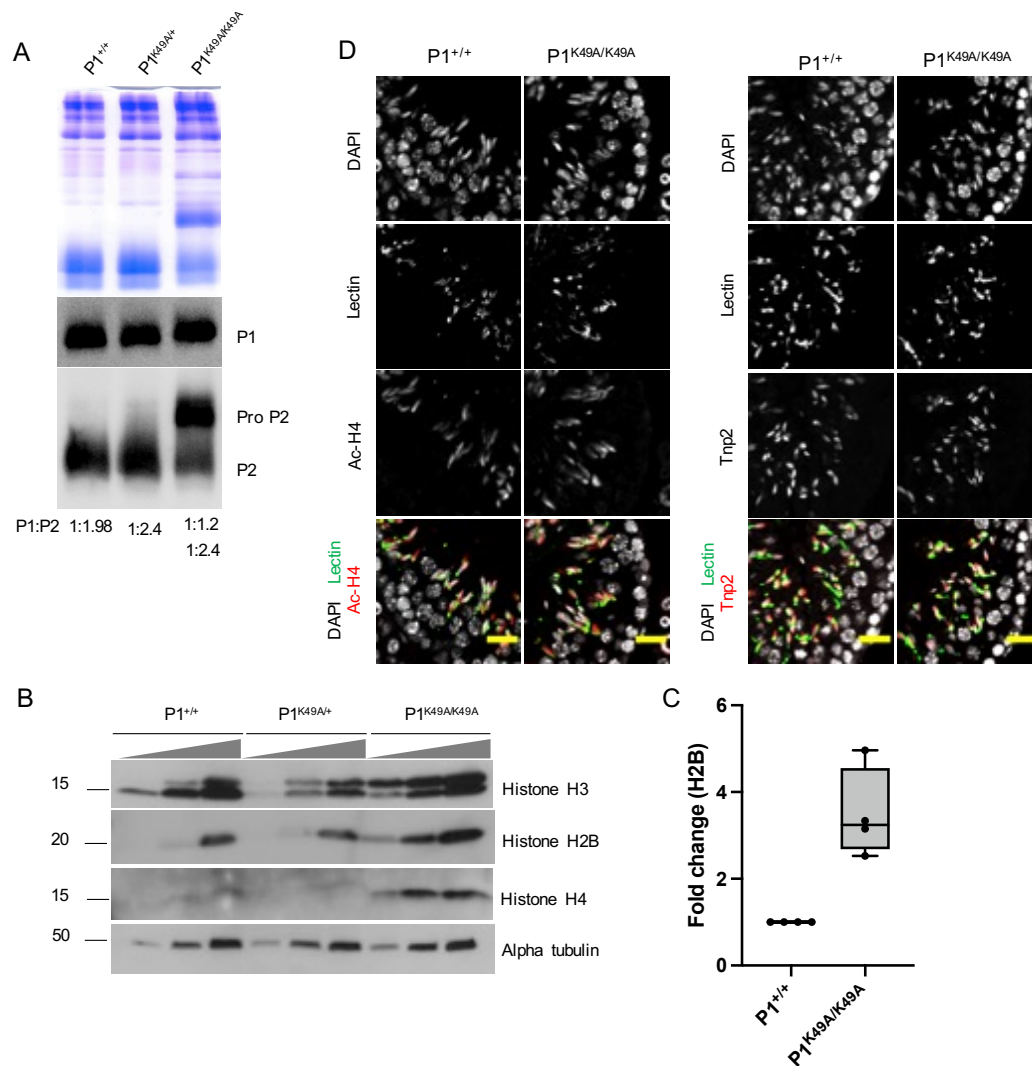
Figure 2-5: P1 K49A substitution alters sperm chromatin composition. (A) Acid urea gel electrophoresis of sperm basic proteins reveals a dramatic shift in P1:P2 ratio in P1$^{K49A/K49A}$ males by Coomassie blue staining (top). Immunoblotting of the gel reveals no difference in P1 level but an accumulation of pro P2 (bottom panels). P1:P2 ratios as quantified in ImageJ are displayed below the immunoblot. **(B)** Immunoblotting of sperm protein extracts reveals an abnormal retention of histones in P1$^{K49A/K49A}$ sperm. Blots were loaded by total input sperm number. Exact sperm numbers for the various antibodies provided in Methods section. **(C)** Quantification of immunoblots showing fold change of histone retention in P1$^{K49A/K49A}$ males. Histone H2B was used as a representative protein for quantification and quantification was performed from n=4 technical replicates; sperm was pooled from n=2 mice. Statistical test was performed using an unpaired, two-sided t test, p=0.0286. Error bars represent standard deviation. **(D)** Immunofluorescence staining of adult P1$^{+/+}$ or P1$^{K49A/K49A}$ testes cross sections for ac-H4 (left panels) or Tnp2 (right panels). Scale bar: 20 μm.
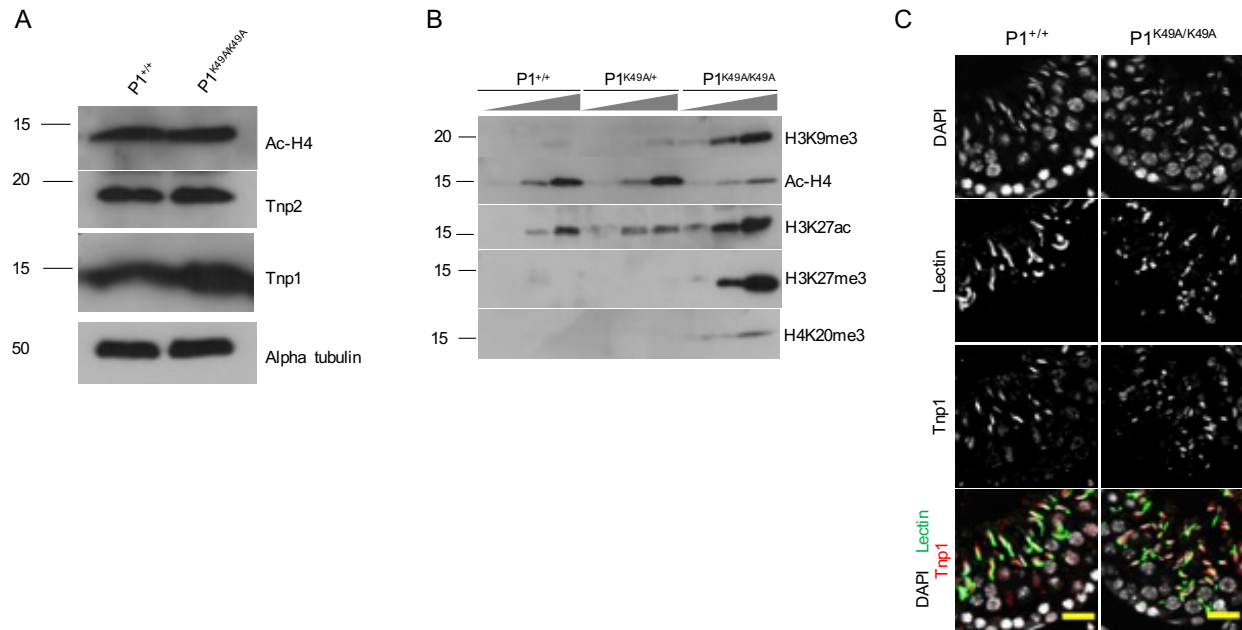
Figure 2-6: P1 K49A substitution does not alter acetylation of histone H4 or expression of Tnp1/Tnp2. (A) Immunoblots of protein lysates from P1$^{+/+}$ and P1$^{K49A/K49A}$ elongating spermatid-enriched testes lysate illustrates no difference in ac-H4, TNP2, or TNP1 levels. **(B)** Immunoblotting of sperm protein extracts reveals an abnormal retention of modified histones in P1$^{K49A/K49A}$ sperm. Blots were loaded by total input sperm number. Exact sperm numbers for each antibody are provided in the Methods section. **(C)** Immunofluorescence staining of adult P1$^{+/+}$ or P1$^{K49A/K49A}$ testes cross sections stained for TNP1. Scale bars: 20 μm.

A

Genotype	Day 1	Day 2	Day 3	Day 4	Day 5
P1+/+	25	1 cell: 6	1 cell: 5	1 cell: 5	1 cell: 5
		2 cell: 19	2 cell: 1	2 cell: 1	2 cell: 1
			4-8 cell: 19	4-8 cell: 3	4-8 cell: 3
				Morula: 16	Morula: 4
					Blast: 12
		76%	100%	84%	75%

Genotype	Day 1	Day 2	Day 3	Day 4	Day 5
P1K49A/K49A	52	1 cell: 26	1 cell: 17	1 cell: 17	1 cell: 17
		2 cell: 26	2 cell: 5	2 cell: 5	2 cell: 3
			4-8 cell: 30	4-8 cell: 8	4-8 cell: 6
				Morula: 23	Morula: 19
					Blast: 7
		50%	100%	77%	30%

B

Pronuclear size measurement

Pixel distance

P1

Paternal P1 dismissal measurement

MII Oocyte   4 hr

C

Relative Pronuclear size (Male/female)

**

P1+/+   P1K49A/K49A

D

DAPI   H3K9me2   P1   DAPI P1 H3K9me2   DAPI P1

P1+/+

P1K49A/K49A

E

P1+/+   P1K49A/K49A

Signal intensity   Distance (um)

DAPI intensity
P1 intensity

60
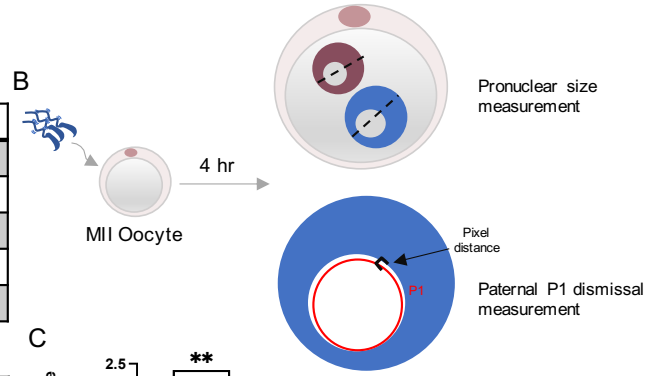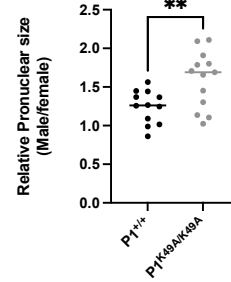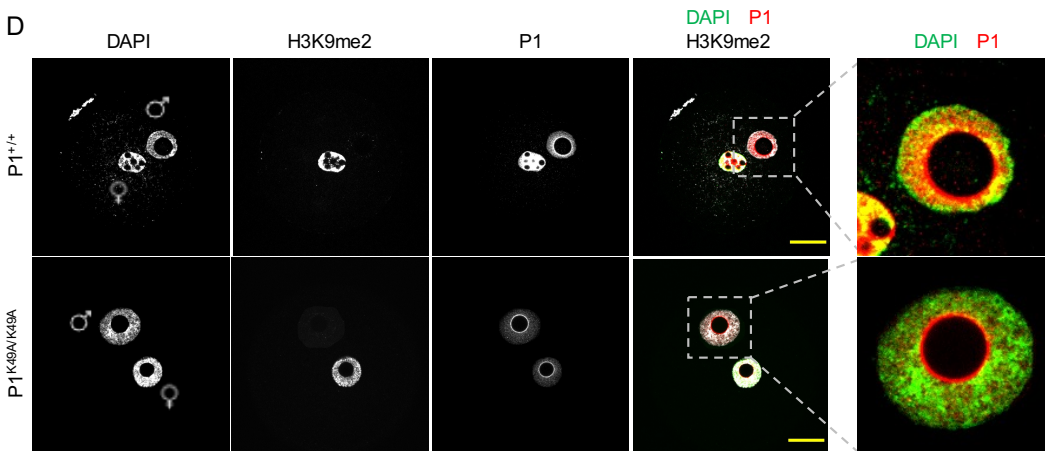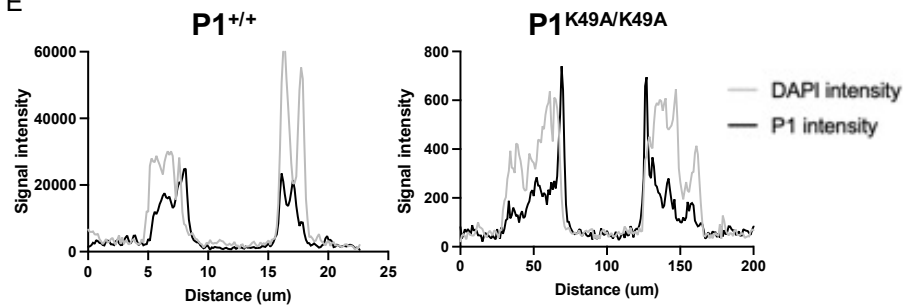
Figure 2-7: P1 K49A substitution results in decreased blastocyst formation and accelerated P1 dismissal from paternal chromatin. (A) Developmental outcomes of intracytoplasmic sperm injections using either $P1^{+/+}$ or $P1^{K49A/K49A}$ sperm reveals a marked decrease in blastocyst formation rate using $P1^{K49A/K49A}$ sperm. Statistical test was performed using a two-sided Fisher's exact test, p=0.0009. **(B)** Experimental scheme for assessing pronuclear size and P1 removal from paternal chromatin. **(C)** Quantification of relative pronuclear size (male/female) in zygotes derived from either $P1^{+/+}$ or $P1^{K49A/K49A}$ sperm. Statistical test was performed using an unpaired, two-sided t-test. Quantification was performed from a total of n=12 $P1^{+/+}$ zygotes and n=13 $P1^{K49A/K49A}$ zygotes, p=0.0075. Center line represents the mean and error bars represent standard deviation. **(D)** Representative immunofluorescence images of zygotes derived from $P1^{+/+}$ or $P1^{K49A/K49A}$ sperm collected at 4hpf and stained for P1. Scale bar = 20 μm. **(E)** Representative histograms of signal intensity for a single line drawn through the pronuclei shown in Figure 4D. A total of n=4 measurements were taken from each zygote and a total of n=12 $P1^{+/+}$ zygotes and n=13 $P1^{K49A/K49A}$ zygotes were used for measurements.
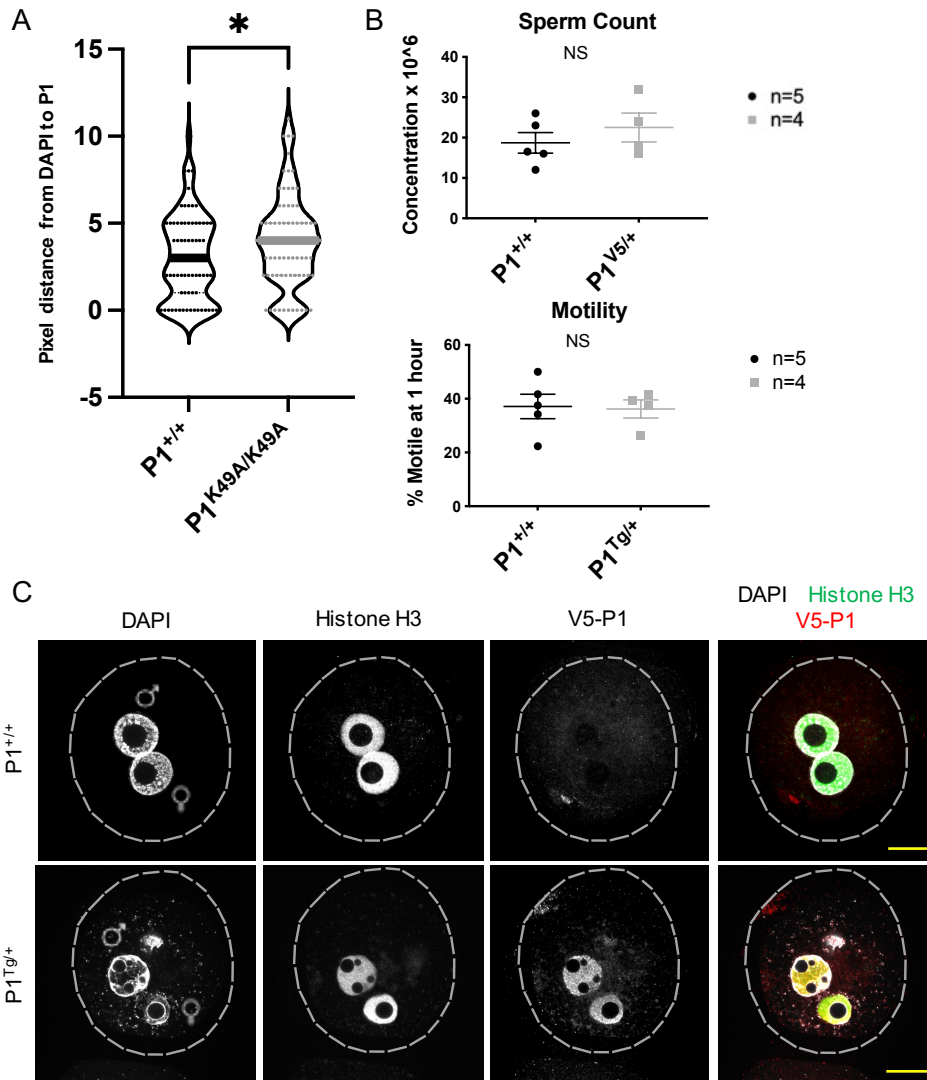
Figure 2-8: P1 localizes to the female pronucleus using a genetic approach. (A) Pixel distance between edge of DAPI signal and highest P1 signal for P1$^{+/+}$ and P1$^{K49A/K49A}$ zygotes. A total of n=8 measurements were made for each embryo and a total of n=10 P1$^{+/+}$ zygotes and a total of n=12 P1$^{K49A/K49A}$ zygotes were used for quantification. Statistical test was performed using an unpaired, two-sided t test, p=0.0123. Center line indicates the mean. **(B)** Sperm count and sperm motility for P1$^{+/+}$ and V5-P1$^{tg/+}$ males illustrate no difference upon addition of V5 tag. A total of n=5 P1$^{+/+}$ and n=4 V5-P1$^{tg/+}$ males were used. Statistical analysis was performed using an unpaired, two-sided t test. **(C)** Representative immunofluorescence images of zygotes derived from P1$^{+/+}$ or V5-P1$^{tg/+}$ sperm, collected at 4hpf, and stained for V5 highlights localization of P1 to the maternal pronucleus. White dotted line indicates perimeter of each zygote. Scale bars: 20 μm.
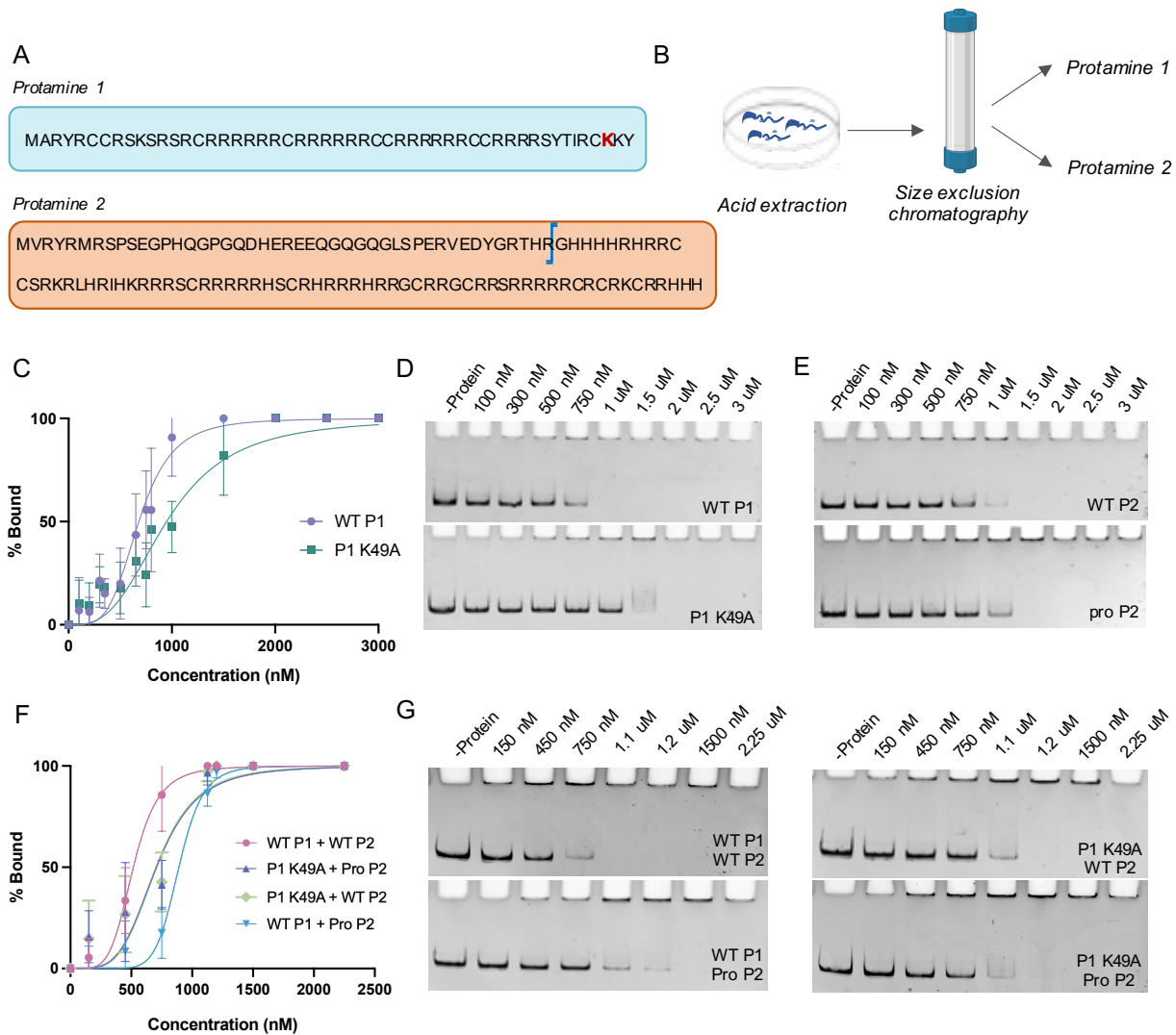
Figure 2-9: P1 K49A substitution negatively affects DNA binding. (A) Schematic of mouse P1 and P2 sequences. Blue bar in P2 indicates cleavage site. **(B)** Purification scheme used for purifying P1 and P2 from mature mouse sperm. **(C)** Quantification of the binding affinities of WT P1 and P1 K49A to a linear ~300 bp DNA fragment. $K_{d,app}$ values were calculated using the Hill equation and were taken from at least 3 technical replicates per protein. Error bars represent standard deviation. **(D)** Representative EMSAs of a titration of increasing amounts of WT P1 (top) or P1 K49A (bottom) illustrating their interaction with a ~300 bp DNA fragment. **(E)** Representative EMSAs of a titration of increasing amounts of WT P2 (top) or pro P2 (bottom) illustrating their interaction with a ~300 bp DNA fragment. **(F)** Quantification of the binding affinities of P1 (either WT or K49A) and P2 (either WT or pro P2) mixed at a 1:2 ratio to a linear ~300 bp DNA fragment. Error bars represent standard deviation. **(G)** Representative EMSAs of titrations of increasing amounts of indicated P1 and P2 mixed at a 1:2 ratio.
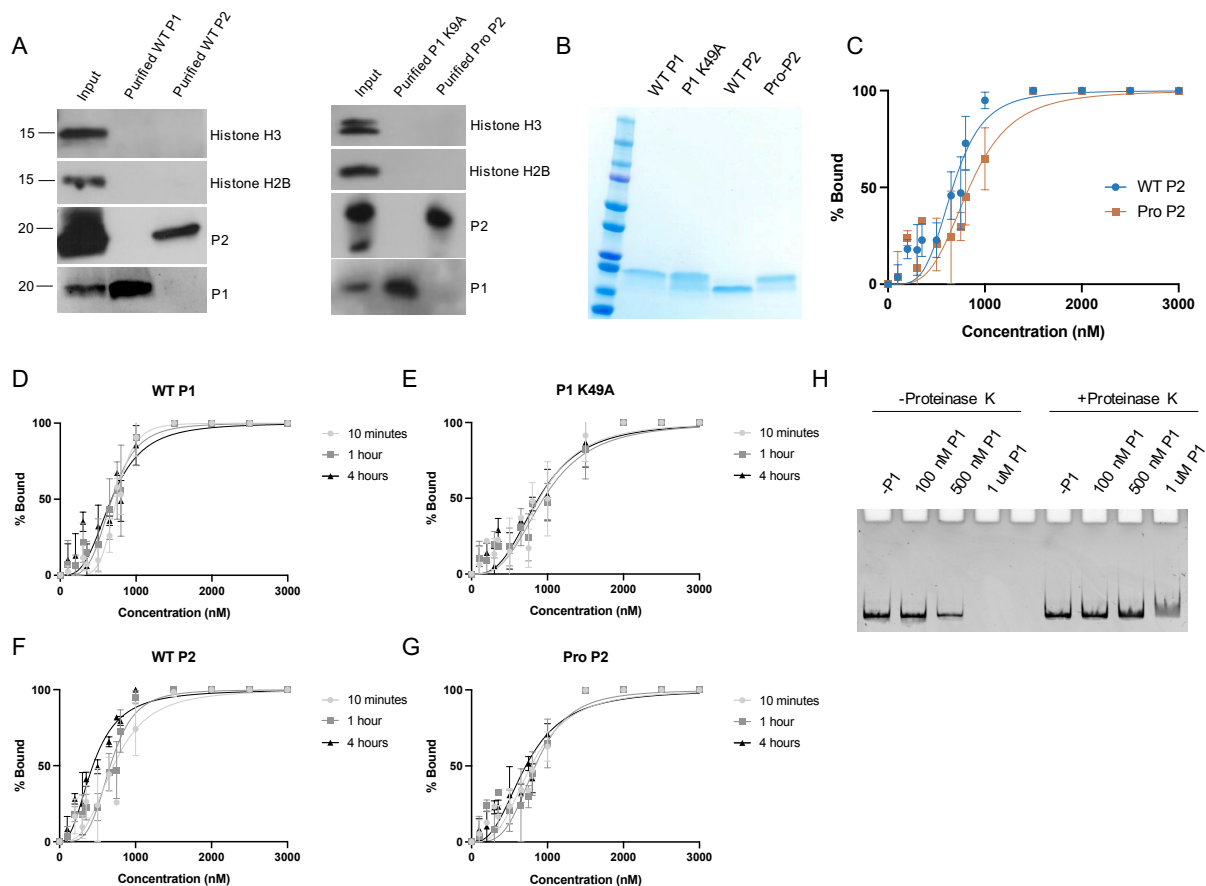
Figure 2-10: Binding equilibrium of P1 and P2 is reached within 10 minutes. (A) Immunoblot of input (prior to gel filtration chromatography) protein, purified P1 (WT or K49A), and purified P2 (WT or pro P2) illustrating efficient separation of P1 and P2 from each other, as well as the absence of histones in the final purified protein. **(B)** Coomassie-stained SDS-PAGE gel of purified proteins illustrating high purity and equivalent concentrations. **(C)** Quantification of the binding affinities of WT P2 and pro P2 to a linear ~300 bp DNA fragment. $K_{d,app}$ values were calculated using the Hill equation and were taken from at least 3 technical replicates per protein. Error bars represent standard deviation. **(D-G)** Quantification of binding affinities of WT P1 **(D),** P1 K49A **(E),** WT P2 **(F),** and pro P2 **(G)** after 10 minutes, 1 hour, or 4 hours of equilibration with DNA. Data were fit to a Hill curve and were taken from at least 3 technical replicates per protein. Error bars represent standard deviation. (H) Proteinase K treatment of EMSA reactions after 1 hour of equilibration with DNA.

A

Barriers DNA

Lipid bilayer

B

dCas9

C WT P1

100nM

125nM

150nM

175nM

200nM

10s 5µm

Protamine Concentration

Time

D P1 K49A

200nM

225nM

250nM

275nM

10s 5µm

Protamine Concentration

Time

E

Compacted DNA (kbps)

Time (s)

WT P1 200nM
WT P1 175nM
WT P1 150nM
WT P1 125nM
WT P1 100nM

F

Compacted DNA (kbps)

Time (s)

P1 K49A 275 nM
P1 K49A 250 nM
P1 K49A 225 nM
P1 K49A 200 nM

G

WT P1, 100 nM

Relative end position (µm)

Time (s)

P1 K49A, 225 nM

Relative end position (µm)

Time (s)

WT P1, 150 nM

Relative end position (µm)

Time (s)

P1 K49A, 250 nM

Relative end position (µm)

Time (s)

WT P1, 200 nM

Relative end position (µm)

Time (s)

P1 K49A, 275 nM

Relative end position (µm)

Time (s)

H

0s 1.5s 3.0s 7.0s

I

Relative position (µm)

Time (s)

WT P1
P1 K49A

0.45 µm/min

0.97 µm/min

Figure 2-11: P1 K49A substitution alters DNA compaction and decompaction kinetics *in vitro*. **(A)** Schematic of DNA curtains. DNA molecules are labeled at the 3' end by dCas9 (shown in pink). **(B)** Cartoon representation shown side-by-side with actual images of protamine-driven DNA compaction. **(C)** Representative kymographs of WT P1 induced DNA compaction at increasing protein concentrations. **(D)** Representative kymographs of P1 K49A induced DNA compaction at increasing protein concentrations. **(E)** Average DNA compaction by WT P1 at increasing concentrations. Error bars represent standard deviations (n=78 traces for 100 nM, n=67 for 125 nM, n=78 for 150 nM, n=63 for 175 nM, and n=66 for 200 nM). **(F)** Average DNA compaction by P1 K49A at increasing concentrations. Error bars represent standard deviations (n=48 traces for 200 nM, n=74 for 225 nM, n=68 for 250 nM, and n=81 for 275 nM). **(G)** Traces of individually tracked DNA molecules over time at low, intermediate, or high concentration of either WT P1 (left panels) or P1 K49A (right panels) illustrating cooperative behavior. **(H)** Representative image of WT P1-driven compaction of adjacent DNA molecules within the curtain highlighting differences in compaction even between DNA molecules that are side-by-side. **(I)** Decompaction of DNA initially compacted by WT P1 and P1 K49A over time illustrates differences in decompaction rates. Error bars represent SEM.
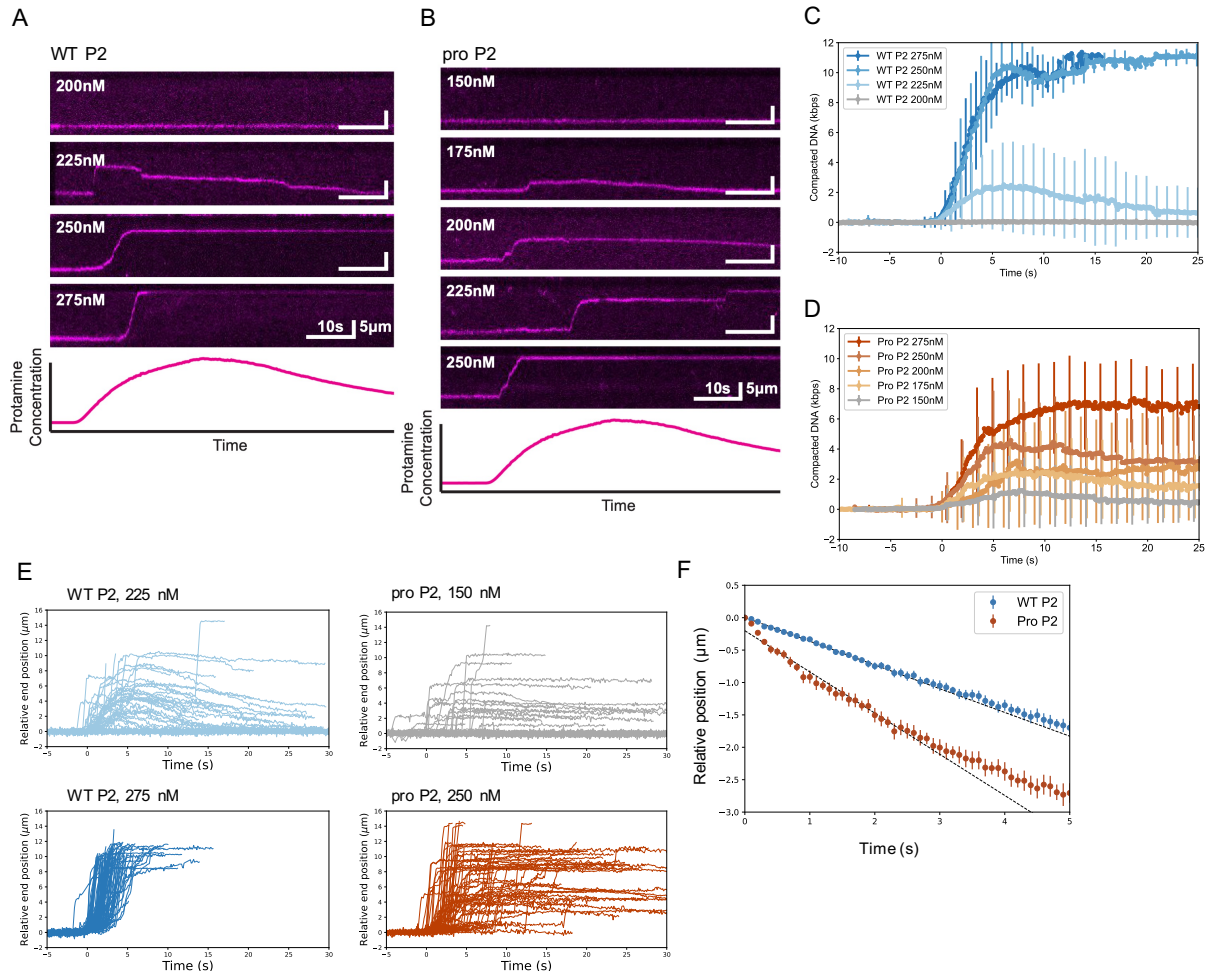
Figure 2-12: WT P2 and pro P2 exhibit different compaction and decompaction kinetics *in vitro*. (A) Representative kymographs of WT P2 induced DNA compaction at increasing protein concentrations. **(B)** Representative kymographs of pro P2 induced DNA compaction at increasing protein concentrations. **(C)** Average DNA compaction by WT P2 at increasing concentrations. Error bars represent standard deviations (n=71 traces for 200 nM, n=63 for 225 nM, n=95 for 250 nM, and n=108 for 275 nM). **(D)** Average DNA compaction by Pro P2 at increasing concentrations. Error bars represent standard deviations (n=74 traces for 150 nM, n=54 for 175 nM, n=62 for 200 nM, n=64 for 225 nM, and n=65 for 250 nM). **(E)** Traces of individually tracked DNA molecules over time at low or high concentration of either WT P2 (left panels) or pro P2 (right panels) illustrating cooperative behavior. **(F)** Decompaction of DNA initially compacted by WT P2 and pro P2 over time illustrates differences in decompaction rates. Error bars represent SEM.
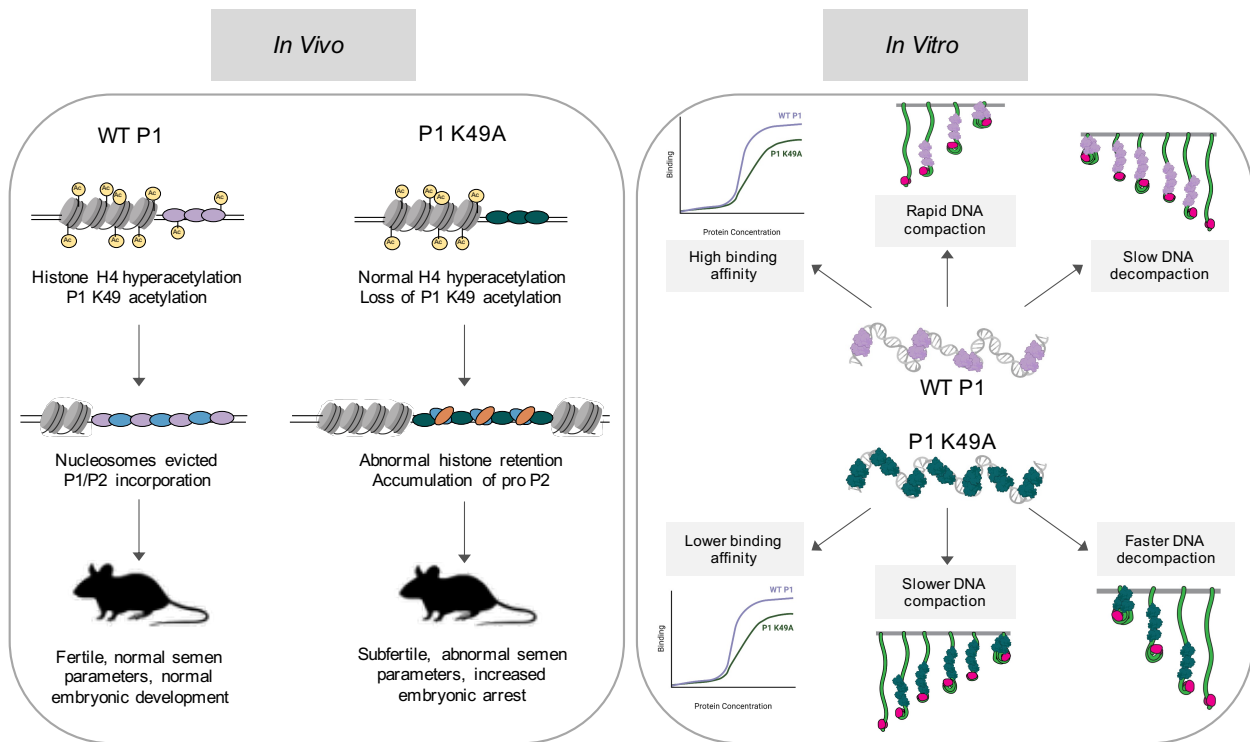
Figure 2-13: Model of alterations to both *in vivo* and *in vitro* protamine functions caused by P1 K49A substitution.

2.7 References

1. Yan, W., Ma, L., Burns, K. H. & Matzuk, M. M. HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. *Proc Natl Acad Sci U S A* **100**, 10546–10551 (2003).
2. Barral, S. *et al.* Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Mol Cell* **66**, 89-101 e8 (2017).
3. Govin, J. *et al.* Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol* **176**, 283–294 (2007).
4. Montellier, E. *et al.* Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev* **27**, 1680–1692 (2013).
5. Shinagawa, T. *et al.* Disruption of Th2a and Th2b genes causes defects in spermatogenesis. *Development* **142**, 1287–1292 (2015).
6. Tachiwana, H., Osakabe, A., Kimura, H. & Kurumizaka, H. Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins in vitro. *Nucleic Acids Res* **36**, 2208–2218 (2008).
7. Tachiwana, H. *et al.* Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc Natl Acad Sci U S A* **107**, 10454–10459 (2010).
8. Ueda, J. *et al.* Testis-Specific Histone Variant H3t Gene Is Essential for Entry into Spermatogenesis. *Cell Rep* **18**, 593–600 (2017).
9. Shirakata, Y., Hiradate, Y., Inoue, H., Sato, E. & Tanemura, K. Histone h4 modification during mouse spermatogenesis. *J Reprod Dev* **60**, 383–387 (2014).
10. Meistrich, M. L., Trostle-Weige, P. K., Lin, R., Bhatnagar, Y. M. & Allis, C. D. Highly acetylated H4 is associated with histone displacement in rat spermatids. *Mol Reprod Dev* **31**, 170–181 (1992).
11. Lu, L. Y. *et al.* RNF8-dependent histone modifications regulate nucleosome removal during spermatogenesis. *Dev Cell* **18**, 371–384 (2010).
12. Shirley, C. R., Hayashi, S., Mounsey, S., Yanagimachi, R. & Meistrich, M. L. Abnormalities and Reduced Reproductive Potential of Sperm from Tnp1- and Tnp2-Null Double Mutant Mice1. *Biology of Reproduction* **71**, (2004).
13. Yu, Y. E. *et al.* Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proceedings of the National Academy of Sciences* **97**, (2000).
14. Pienta, K. J. & Coffey, D. S. A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome. *J Cell Sci Suppl* **1**, 123–135 (1984).
15. Ward, W. S. & Coffey, D. S. DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol Reprod* **44**, 569–574 (1991).
16. Green, G. R., Balhorn, R., Poccia, D. L. & Hecht, N. B. Synthesis and processing of mammalian protamines and transition proteins. *Mol Reprod Dev* **37**, 255–263 (1994).
17. Yelick, P. C. *et al.* Mouse protamine 2 is synthesized as a precursor whereas mouse protamine 1 is not. *Mol Cell Biol* **7**, 2173–2179 (1987).
18. Wykes, S. M. & Krawetz, S. A. The structural organization of sperm chromatin. *J Biol Chem* **278**, 29471–29477 (2003).
19. AOKI, V. *et al.* Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. *Fertility and Sterility* **86**, (2006).
20. de Mateo, S. *et al.* Protamine 2 precursors (Pre-P2), protamine 1 to protamine 2 ratio (P1/P2), and assisted reproduction outcome. *Fertil Steril* **91**, 715–722 (2009).

21. Zatecka, E. *et al.* The effect of tetrabromobisphenol A on protamine content and DNA integrity in mouse spermatozoa. *Andrology* **2**, 910–917 (2014).
22. Cho, C. *et al.* Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat Genet* **28**, 82–86 (2001).
23. Cho, C. *et al.* Protamine 2 Deficiency Leads to Sperm DNA Damage and Embryo Death in Mice1. *Biology of Reproduction* **69**, (2003).
24. Schneider, S. *et al.* Re-visiting the Protamine-2 locus: deletion, but not haploinsufficiency, renders male mice infertile. *Sci Rep* **6**, 36764 (2016).
25. Takeda, N. *et al.* Viable offspring obtained from Prm1-deficient sperm in mice. *Sci Rep* **6**, 27409 (2016).
26. Balhorn, R., Brewer, L. & Corzett, M. DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. *Mol Reprod Dev* **56**, 230–234 (2000).
27. Bench, G. S., Friz, A. M., Corzett, M. H., Morse, D. H. & Balhorn, R. DNA and total protamine masses in individual sperm from fertile mammalian subjects. *Cytometry* **23**, 263–271 (1996).
28. Brewer, L. R. Protamine-Induced Condensation and Decondensation of the Same DNA Molecule. *Science* **286**, (1999).
29. Prieto, M. C., Maki, A. H. & Balhorn, R. Analysis of DNA-protamine interactions by optical detection of magnetic resonance. *Biochemistry* **36**, 11944–11951 (1997).
30. Prieto, M. C., Maki, A. H. & Balhorn, R. Analysis of DNA-protamine interactions by optical detection of magnetic resonance. *Biochemistry* **36**, 11944–11951 (1997).
31. Brewer, L. R., Corzett, M. & Balhorn, R. Protamine-induced condensation and decondensation of the same DNA molecule. *Science* **286**, 120–123 (1999).
32. Krawetz, S. A. & Dixon, G. H. Sequence similarities of the protamine genes: Implications for regulation and evolution. *Journal of Molecular Evolution* **27**, (1988).
33. Lewis, J. D., Song, Y., de Jong, M. E., Bagha, S. M. & Ausió, J. A walk though vertebrate and invertebrate protamines. *Chromosoma* **111**, (2003).
34. Wyckoff, G. J., Wang, W. & Wu, C.-I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, (2000).
35. Brehove, M. *et al.* Histone Core Phosphorylation Regulates DNA Accessibility. *Journal of Biological Chemistry* **290**, (2015).
36. Kiefer, C. M., Hou, C., Little, J. A. & Dean, A. Epigenetics of β-globin gene regulation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **647**, (2008).
37. Shogren-Knaak, M. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* **311**, (2006).
38. Queralt, R. *et al.* Evolution of protamine P1 genes in mammals. *Journal of Molecular Evolution* **40**, (1995).
39. Rooney, A. P., Zhang, J. & Nei, M. An Unusual Form of Purifying Selection in a Sperm Protein. *Molecular Biology and Evolution* **17**, (2000).
40. Torgerson, D. G., Kulathinal, R. J. & Singh, R. S. Mammalian Sperm Proteins Are Rapidly Evolving: Evidence of Positive Selection in Functionally Diverse Genes. *Molecular Biology and Evolution* **19**, (2002).
41. Itoh, K. *et al.* Dephosphorylation of protamine 2 at serine 56 is crucial for murine sperm maturation in vivo. *Sci Signal* **12**, (2019).

42.   Pirhonen, A., Linnala-Kankkunen, A. & Menpaa, P. H. P2 protamines are phosphorylated in vitro by protein kinase C, whereas P1 protamines prefer cAMP-dependent protein kinase. A comparative study of five mammalian species. *Eur J Biochem* **223**, 165–169 (1994).

43.   Seligman, J., Zipser, Y. & Kosower, N. S. Tyrosine phosphorylation, thiol status, and protein tyrosine phosphatase in rat epididymal spermatozoa. *Biol Reprod* **71**, 1009–1015 (2004).

44.   Gou, L. T. *et al.* Initiation of Parental Genome Reprogramming in Fertilized Oocyte by Splicing Kinase SRPK1-Catalyzed Protamine Phosphorylation. *Cell* **180**, 1212-1227 e14 (2020).

45.   Brunner, A. M., Nanni, P. & Mansuy, I. M. Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics Chromatin* **7**, 2 (2014).

46.   Soler-Ventura, A. *et al.* Characterization of Human Sperm Protamine Proteoforms through a Combination of Top-Down and Bottom-Up Mass Spectrometry Approaches. *J Proteome Res* **19**, 221–237 (2020).

47.   Chira, F. *et al.* Phosphorylation of human sperm protamines HP1 and HP2: identification of phosphorylation sites. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1203**, (1993).

48.   Dong, Y. *et al.* EPC1/TIP60-Mediated Histone Acetylation Facilitates Spermiogenesis in Mice. *Molecular and Cellular Biology* **37**, (2017).

49.   Ketchum, C. C., Larsen, C. D., McNeil, A., Meyer-Ficca, M. L. & Meyer, R. G. Early histone H4 acetylation during chromatin remodeling in equine spermatogenesis. *Biol Reprod* **98**, 115–129 (2018).

50.   Luense, L. J. *et al.* Gcn5-Mediated Histone Acetylation Governs Nucleosome Dynamics in Spermiogenesis. *Developmental Cell* **51**, (2019).

51.   Meistrich, M. L., Trostle-Weige, P. K., Lin, R., Allis, C. D. & Bhatnagar, Y. M. Highly acetylated H4 is associated with histone displacement in rat spermatids. *Molecular Reproduction and Development* **31**, (1992).

52.   Shiota, H. *et al.* Nut Directs p300-Dependent, Genome-Wide H4 Hyperacetylation in Male Germ Cells. *Cell Rep* **24**, 3477-3487 e6 (2018).

53.   McLay, D. W. & Clarke, H. J. Remodelling the paternal chromatin at fertilization in mammals. *Reproduction* **125**, 625–633 (2003).

54.   Balhorn, R. The protamine family of sperm nuclear proteins. *Genome Biology* **8**, (2007).

55.   Boussouar, F. *et al.* A specific CBP/p300-dependent gene expression programme drives the metabolic remodelling in late stages of spermatogenesis. *Andrology* **2**, (2014).

56.   Gaucher, J. *et al.* Bromodomain-dependent stage-specific male genome programming by Brdt. *EMBO J* **31**, 3809–3820 (2012).

57.   Mylonis, I. *et al.* Temporal Association of Protamine 1 with the Inner Nuclear Membrane Protein Lamin B Receptor during Spermiogenesis. *Journal of Biological Chemistry* **279**, (2004).

58.   Berg, O. G. & von Hippel, P. H. Selection of DNA binding sites by regulatory proteins. *Journal of Molecular Biology* **193**, (1987).

59.   Shultzaberger, R. K. *et al.* Correlation between binding rate constants and individual information of E. coli Fis binding sites. *Nucleic Acids Research* **35**, (2007).

60.   Larson, A. G. *et al.* Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

61. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148 (2016).

62. Yuan, C. L. & Hu, Y. C. A Transgenic Core Facility's Experience in Genome Editing Revolution. *Adv Exp Med Biol* **1016**, 75–90 (2017).

63. de Yebra, L. & Oliva, R. Rapid analysis of mammalian sperm nuclear proteins. *Anal Biochem* **209**, 201–203 (1993).

64. Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24**, 1285–1292 (2006).

65. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, (2004).

66. Nakata, H., Wakayama, T., Takai, Y. & Iseki, S. Quantitative analysis of the cellular composition in seminiferous tubules in normal and genetically modified infertile mice. *J Histochem Cytochem* **63**, 99–113 (2015).

67. Giorgini, F., Davies, H. G. & Braun, R. E. Translational repression by MSY4 inhibits spermatid differentiation in mice. *Development* **129**, (2002).

68. Gallardo, I. F. *et al.* High-Throughput Universal DNA Curtain Arrays for Single-Molecule Fluorescence Imaging. *Langmuir* **31**, 10310–10317 (2015).

69. Yoshida, N. & Perry, A. C. Piezo-actuated mouse intracytoplasmic sperm injection (ICSI). *Nature Protocols* **2**, (2007).

Chapter 3 Investigating the Role of Protamine 1 C-terminal Phosphorylation on Sperm Chromatin Structure and Fertility

3.1 Abstract

Protamine-mediated condensation of the paternal genome during spermiogenesis is an essential transition to ensure proper sperm function and reproductive success. Phosphorylation of protamines has long been hypothesized to play a role in this process via modulation of protamine-DNA binding, however genetic evidence for a functional role of protamine phosphorylation or generally of the residues that bear phosphorylation is lacking. Here, we investigated the functional role of two residues in P1- S43 and T45, both of which are phosphorylated in late-stage spermatids just prior to release from the testis, and these modifications persist into mature sperm. *In vivo*, substitution of S43/T45 to glutamic acid (P1$^{S43E/T45E}$), but not to alanine (P1$^{S43A/T45A}$) results in sperm DNA damage, loss of sperm motility, and infertility in male mice. *In vitro,* S43E/T45E substitutions result in a slightly decreased binding affinity of P1 for DNA, both alone and when combined with P2. Altogether, our observations demonstrate that substitution of S43 and T45 to a non-modifiable alanine does not disrupt histone-to-protamine exchange, sperm chromatin composition/morphology, or fertility, suggesting that the phenotypes observed in P1^{S43ET45E} males are likely the result of premature addition of a negatively charged moiety imposed by glutamic acid. Furthermore, these results suggest that temporally controlled phosphorylation at these sites may modulate structural characteristics of the P1 protein, thereby promoting favorable, yet fluid packaging as the genome condenses.

3.2 Introduction

Proper sperm function and reproductive success are reliant on three distinct transitions during spermatogenesis: mitosis, meiosis, and spermiogenesis. During spermiogenesis, extensive cytoskeletal remodeling results in elongation of the spermatid nucleus and shedding of the cytoplasm. Additionally, global alteration in the chromatin landscape results in the exchange of histones for protamines. This transition in chromatin architecture involves the incorporation of histone variants, histone post-translational modifications (PTMs), transition proteins, and finally protamines, which ultimately hypercondense the sperm genome.[1–12]

Protamines are small, highly basic proteins essential for condensation of the sperm genome and fertility.[13–15] In mice and humans (and most mammals), packaging is carried out by two forms of protamine- protamine 1 (P1) and protamine 2 (P2).[16,17] While P1 is expressed in its mature form of an ~7 kDa protein, P2 is initially expressed as a precursor (pro P2, ~13.6 kDa) that is then proteolytically processed once deposited onto DNA to its mature form (P2, ~9 kDa).[18,19] Together, protamines package 90-95% of the sperm genome and numerous studies have demonstrated that P1 and P2 are present in a species specific ratio (P1:P2 ratio) that is highly variable across species (1:2 in mouse, 1:1 in humans), but must be maintained for optimal fertility.[20–24]

We previously performed mass spectrometry on mature mouse sperm and identified several PTMs on P1 that we hypothesized may play a role in sperm chromatin packaging including several phosphorylated residues and acetylation of lysine (K) 49. Our previous analysis (Chapter 2) showed that genetic substitution of K49 for alanine (A) results in altered sperm chromatin composition, decreased sperm motility, decreased male fertility, and premature removal of P1 from paternal chromatin in the zygote. It remains unknown, however, whether there is crosstalk between modifications on P1 or whether other modifications, such as phosphorylation, additionally play a role in sperm chromatin structure.

Phosphorylation of P1 and P2 has long been hypothesized to control protamine deposition/chromatin condensation in spermatids, as several early studies illustrated that protamines are phosphorylated shortly after their synthesis and dephosphorylated after deposition onto DNA.[18,25–29] Subsequent studies have confirmed both the presence and sites of

phosphorylation in mouse and human protamines by mass spectrometry.[30,31] Additionally, loss of CAMK4- a serine/threonine kinase capable of phosphorylating P2 *in vitro*- disrupts protamine deposition and displacement of TNPs, providing genetic evidence for a functional role of P2 phosphorylation in spermiogenesis.[32] More recently, Gou. et. al. reported that P1 phosphorylation at several serine residues is absent in the testis and in sperm but is acquired in the zygote to initiate remodeling of the male pronucleus.[33] However, given that protamine phosphorylation in sperm is well established, the role of P1 phosphorylation in spermiogenesis remains unclear.

Phylogenetic analysis of P1 sequences reveals that potential phosphorylation sites are conserved within the mouse lineage but are absent or substituted to alternative residues in more distant species (S43 is absent, whereas T45 is occupied by either K or R in primates), suggesting possible lineage specific function, as we previously hypothesized for P1 K49 acetylation. To investigate the functional role of these residues/modifications genetically, we substituted S43/T45 for either alanine (A) to introduce non-modifiable sites or glutamic acid (E) as a phosphomimetic. The genetic substitution of S43/T45 for glutamic acid but not alanine results in complete male infertility in mice, loss of sperm motility, abnormal sperm morphology, and increased sperm DNA damage. Sperm chromatin analysis of P1^{S43ET45E} sperm revealed decreased P1 incorporation onto chromatin and a slight accumulation of pro P2. *In vitro,* the P1 S43E/T45E protein has a slightly decreased binding affinity to DNA both alone and in combination with P2. Altogether, these findings illustrate that while the S43A/T45A substitutions do not significantly alter sperm chromatin compaction, premature introduction and/or persistence of the negative charges imposed by phosphorylation (S43E/T45E) severely perturb sperm chromatin packaging, ultimately resulting in severe sperm defects and infertility. These results therefore suggest that temporally controlled phosphorylation at these sites may aid in conferring structural changes to either P1 or the P1-DNA complex, thereby promoting favorable packaging that remains fluid until complete compaction is achieved.

3.3 Results

3.3.1 P1 phosphorylation sites are conserved in rodents, but not higher species

Transient phosphorylation on both P1 and P2 has been reported in multiple species, and these modifications are thought to have a role in protamine-DNA binding/dynamics, however their

function in mammalian spermatogenesis remains largely unknown.[25–27,30] We and others recently identified several sites of phosphorylation on P1 and interestingly, several of these sites reside in the C-terminal portion of P1, which is known to be highly divergent across species (**Table 2-1,** phosphorylation events summarized in **Figure 3-1A**).[31,34] A close examination of P1 sequences across the orders Rodentia, Primate, and Artiodactyla reveals a high degree of conservation of the S9 position in P1, with the exceptions being salmon and *Bos taurus,* likely reflecting a critical function of the residue position and PTM across species (**Figure 3-2A**).[30,31] In contrast, the two C-terminal phosphorylation sites (S43 and T45) are highly conserved within the mouse lineage but are either absent or occupied by alternative residues in more distant species. Specifically, we found that S43 is conserved within a subset of the rodent lineage but is completely absent in all Primate and Artiodactyla species analyzed. Additionally, T45 is highly conserved across the rodent lineage, but is occupied by a valine (V) in Artiodactyla and either R or K in Primate species (**Figure 3-2A**). Therefore, these substitutions in higher organisms introduce non-modifiable sites but also introduce different chemical properties at this position from non-charged, polar residues (S and T) to either non-polar (V) or polar but positively charged (K and R). This observation prompted us to investigate the function of the S43 and T45 residues along with phosphorylation at these sites, and additionally to determine whether any crosstalk exists between S43/T45 phosphorylation and other PTMs.

3.3.2 P1 S43/T45 phosphorylation is acquired in late elongating spermatids and persists in mature sperm

To verify the presence of S43 and T45 phosphorylation (S43/T45ph) in sperm, we generated a polyclonal antibody against S43/T45ph (a single antibody that recognizes the simultaneous phosphorylation at these two sites). Immunoblotting of acid extracted protein from mature sperm identified a distinct band that was abolished when competed by a phosphorylated P1 S43/T45 peptide, but not when using a nonspecific peptide of an unrelated protein, or a non-phosphorylated P1 peptide, confirming both the presence of P1 S43/T45ph *in vivo* and the specificity of our antibody (**Figure 3-2B**). To further confirm the specificity of our antibody for the phosphorylated form of P1, we treated lysates with calf intestinal phosphatase (CIP) to remove protein phosphorylation and found significantly decreased signal using our P1 S43/T45ph antibody on immunoblots, further indicating antibody specificity (**Figure 3-2C**). To define the stage or stages

of the seminiferous epithelial cycle in which P1 S43/T45ph is established, we co-stained adult testes cross-sections using our custom antibody with the acrosomal marker PNA-Lectin. We found that specific signal was initiated in stages IV-VI (mid-stage spermatids) and peaked at stages VII-VIII (late-stage spermatids, 100% of tubules) (**Figure 3-1B,C, 3-2D**). We additionally confirmed that P1 S43/T45ph is acquired in spermatids and persists in epididymal and vas deferens sperm by immunoblotting, further confirming our observed patterns by immunostaining (**Figure 3-1D**).

3.3.3 Substitution of P1 S43/T45 for glutamic acid, but not alanine, results in sperm motility defects and infertility

To investigate the functional role of P1 S43/T45ph *in vivo*, we used CRISPR/Cas9 to generate a serine/threonine to alanine mutant mouse (S43A/T45A) to introduce non-modifiable sites (homozygotes referred to as P1^{S43AT45A}). Additionally, to distinguish between phosphorylation itself and charge state, we also generated a phosphomimetic mouse (S43E/T45E, homozygotes referred to as P1^{S43ET45E}). We used Sanger sequencing to confirm the presence of the target mutations (**Figure 3-3A**) as well as the absence of any potential off-target genetic mutations (**Figure 3-4A**). Overall, both P1^{S43AT45A} and P1^{S43ET45E} males exhibit normal phenotypic parameters; we observed no difference in testes/body weight ratio for either line and we detected all germ cell populations in both cases (**Figure 3-4B,C**). Additionally, overall sperm counts were normal in both P1^{S43AT45A} and P1^{S43ET45E} males (**Figure 3-3B**). However, while progressive sperm motility was normal in P1^{S43AT45A} males, it was severely impaired in P1^{S43ET45E} males (**Figure 3-3C**) and several sperm structural abnormalities were observed in P1^{S43ET45E}, but not P1^{S43AT45A} males (**Figure 3-3D,E**). Notably, over ~70% of P1^{S43ET45E} sperm were abnormal, with the majority of abnormalities being related to head morphology (specifically, elongated heads that are reduced in size), although ~20% of sperm exhibited tail abnormalities (**Figure 3-3E**).

Given the dramatic reduction in sperm motility observed in P1^{S43ET45E} males, we next investigated potential molecular causes for this defect, including DNA damage and sperm tail mitochondrial content, as it is well established that alterations in both affect sperm motility.[35–37] To assess levels of DNA damage in P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} sperm, we performed a Comet assay.[38] Strikingly, we found that ~25% of P1^{S43ET45E} sperm formed "comets" compared to ~3-4% of P1$^{+/+}$ and P1^{S43AT45A} sperm, indicating high levels of DNA damage caused by the introduction of two

glutamic acid residues in P1 (**Figure 3-4D**). Additionally, we also stained sperm using Mitotracker to assess mitochondrial content and distribution in the midpiece and found that P1^{S43ET45E} sperm exhibit a decrease in overall mitochondrial content as well as an abnormal distribution of mitochondria (**Figure 3-3F**), which may in part explain the loss of sperm motility in these mice. Moreover, P1^{S43ET45E} males were completely infertile, whereas P1^{S43AT45A} males exhibited fecundity rates comparable to P1$^{+/+}$ littermates (**Figure 3-3G**). To ensure that introduction of two negatively charged residues did not compromise P1 protein levels or P1 localization, we stained P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} adult testes cross-sections using a custom P1 antibody and found that nuclear P1 is detectable in all cases, suggesting that these substitutions do not destabilize the P1 protein (**Figure 3-4E, specificity test shown in Figure 2-4E**). Additionally, by immunoblotting we found that P1 protein levels in the testis are comparable across all genotypes (**Figure 3-4F**). Therefore, despite normal P1 expression and localization, P1^{S43ET45E} males exhibit high levels of DNA damage in sperm.

3.3.4 P1^{S43AT45A} and P1^{S43ET45E} mutants undergo proper histone-to-protamine exchange

Sperm DNA damage has been postulated to be a result of increased histone retention given that histones package DNA less tightly than protamines. Additionally, several human studies have illustrated an inverse correlation between protamine/DNA ratio and DNA damage.[39–41] Therefore, we investigated sperm histone levels and the overall histone-to-protamine exchange process to determine whether it is perturbed in P1^{S43ET45E} males. By immunoblotting of sperm protein lysates, we found no difference in overall levels of histone H3 or histone H2B in either P1^{S43AT45A} or P1^{S43ET45E} sperm (**Figure 3-5A**), suggesting that histone eviction occurs normally. To further confirm proper initiation of histone-to-protamine exchange, we analyzed P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} adult testes using an anti-tetra-acetyl H4 antibody (referred to as ac-H4, an essential triggering event to initiate histone eviction) and found no differences in abundance or pattern of ac-H4 by immunostaining (**Figure 3-5B, left panels**). Additionally, we investigated transition proteins 1 and 2 (TNP1 and TNP2)- intermediate proteins of the histone-to-protamine exchange that are necessary for proper sperm chromatin composition and packaging.[11,12] By immunostaining of adult testes cross-sections and immunoblotting of testes lysates, we observed no differences in TNP1 or TNP2 when comparing P1$^{+/+}$, P1^{S43AT45A} and P1^{S43ET45E} males (**Figure 3-5B, right panels, Figure 3-6A,B**). Taken together, these observations suggest that the overall histone-to-

protamine exchange process initiates and progresses normally in both P1^{S43AT45A} and P1^{S43ET45E} males, resulting in proper histone eviction. Therefore, it is possible that the high levels of DNA damage/loss of sperm motility observed in P1^{S43ET45E} sperm may stem from decreased stability of the P1 protein or alterations in sperm chromatin composition.

3.3.5 P1^{S43ET45E}, but not P1^{S43AT45A} mutants exhibit altered sperm chromatin composition

To investigate the effects of the S43A/T45A and S43E/T45E substitutions on the final composition of mature sperm chromatin, we compared overall protamine levels and P1:P2 ratios in a fixed number of P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} sperm. We found that while P1$^{S43AT45A/+}$, P1^{S43AT45A}, and P1$^{S43ET45E/+}$ sperm exhibited normal P1 and P2 levels as well as the expected P1:P2 ratio, P1^{S43ET45E} sperm exhibited a subtle shift in P1:P2 ratio from the expected 1:2 to a ratio of 1:1.3 (**Figure 3-5C**). This alteration was caused by both a modest accumulation in unprocessed P2 (pro P2) as well as a decrease in P1 (~30% reduction).

Given that the alteration in P1:P2 ratio in P1^{S43ET45E} sperm is modest, we wondered whether other phosphorylation events in P1 are affected by the S43E/T45E substitutions. Multi-site phosphorylation of numerous proteins has been documented and serves several distinct functions including increasing binding affinity for other proteins, inducing conformational changes, and in some cases allows for downstream phosphorylation at additional sites. Given that our mass spectrometry analysis identified phosphorylation at S9 (S9ph) and that a recent study by Gou et. al. reported that P1$^{S9A/S43A}$ males are infertile, we investigated whether S9ph levels in both P1^{S43AT45A} and P1^{S43ET45E} sperm. We found that while levels of this modification are unaffected by the S43A/T45A substitutions, the S43E/T45E substitutions result in near complete loss of this modification (**Figure 3-5C**). Together, these results suggest that the S43E/T45E substitutions may hinder P1 protein stability leading to decreased incorporation onto chromatin, however it cannot be ruled out that the observed phenotype in P1^{S43ET45E} males is due to loss of S9 phosphorylation.

3.3.6 The substitution of P1 S43/T45 for glutamic acid, but not alanine, decreases DNA binding ability of P1

Given normal histone-to-protamine exchange, yet decreased P1 incorporation onto chromatin in P1^{S43ET45E} males, we next sought to investigate protamine-DNA interactions directly to determine whether the S43E/T45E substitutions alter binding of the P1 protein to DNA. Previous attempts to

recombinantly generate P1 and P2 proteins in *E. coli* were unsuccessful due to the high arginine content of protamines. Therefore, to overcome this challenge, we developed a purification scheme to purify P1 and P2 from P1$^{+/+}$ (WT P1 or WT P2), P1^{S43AT45A} (P1 S43A/T45A and WT P2), and P1^{S43ET45E} (P1 S43E/T45E and pro P2) mature sperm (note that based on acid urea gel electrophoresis and immunoblotting, P2 in P1^{S43ET45E} sperm consists of ~50% unprocessed P2 and 50% processed P2, but is referred to here as pro P2 for simplicity, **Figure 3-7A**). Acid extraction of basic proteins from mature sperm followed by size exclusion chromatography allowed for acquisition of WT and mutant proteins at high purity (**Figure 3-8A**).

To measure the binding affinity of WT P1, P1 S43A/T45A, and P1 S43E/T45E proteins to DNA, we performed electrophoretic mobility shift assays (EMSAs) using a ~300 bp linear DNA fragment, with protein concentration in excess of DNA. WT P1 exhibited robust concentration-dependent DNA binding after 1 hour with an apparent K_d ($K_{d,app}$) of 0.54 μM that was unaffected by incubation time with DNA (**Figure 3-7B,C, Figure 3-8B**). Similarly, P1 S43A/T45A also robustly bound DNA with a $K_{d,app}$ of 0.57 μM, indicating that the S43A and T45A substitutions do not significantly alter the DNA binding ability of the protein (**Figure 3-7B,C, Figure 3-8B**). In contrast, P1 S43E/T45E exhibited a decrease in DNA binding affinity with a $K_{d,app}$ of 0.79 μM, indicating that the S43E and T45E substitutions do slightly perturb P1 binding to DNA (**Figure 3-7B,C, Figure 3-8B**).

Given that the mouse sperm genome is packaged with both P1 and P2, we next sought to investigate whether the S43A/T45A or S43E/T45E substitutions alter the affinity of P1 and P2 together for DNA. To this end, we performed EMSAs using either WT or mutant P1 in combination with either WT P2 or pro P2 (~50% pro P2, 50% processed P2) in a 1:2 ratio. As we have observed previously, WT P1 and WT P2 together bind DNA more efficiently than either protein alone (**Figure 3-7D,E**). Similarly, P1 S43A/T45A bound just as efficiently with WT P2 to DNA as did WT P1, and both binding efficiencies were slightly lowered when using pro P2, although the P1 S43A/T45A was more greatly affected (**Figure 3-7D,E**). Consistent with our EMSA results for each protein alone, the P1 S43E/T45E exhibited less efficient binding in combination with either WT P2 or pro P2 compared to WT P1 or P1 S43A/T45A (**Figure 3-7D,E**)

suggesting that the overall packaging of the genome is likely affected in P1^{S43ET45E} males due to this biochemical difference of the P1 S43E/T45E protein.

3.4 Discussion

The exchange of histones for protamines is essential during spermiogenesis to allow for optimal sperm genome packaging and ultimately, reproductive fitness throughout the lifetime of a male. Here, we set out to investigate the functional role of P1 phosphorylation at two C-terminal residues–S43 and T45, both of which are phosphorylated in late-stage spermatids just prior to release from the testis. Genetic substitution of S43/T45 to a non-modifiable alanine did not disrupt histone-to-protamine exchange, overall sperm chromatin composition, or sperm motility/morphology, therefore resulting in no observable phenotype *in vivo.* In contrast, genetic substitution of S43/T45 to glutamic acid, which constitutively mimics the negative charges imposed by phosphorylation, results in broad and striking alterations to sperm morphology/motility and ultimately results in loss of fertility. Together, these results therefore demonstrate that phosphorylation at S43/T45 likely is not essential during spermiogenesis, however glutamic acid residues at these positions are not tolerated, suggesting that the premature addition of negative charges at these sites is destructive in chromatin remodeling. Given these striking differences between P1^{S43AT45A} and P1^{S43ET45E} mice, we propose several potential mechanisms to explain these observations.

Premature addition of negative charges at S43 and T45 may alter protein structure, protein-protein interactions, or other PTMs on P1 necessary for chromatin remodeling

While no crystal structure for either P1 or P2 alone or bound to DNA has been solved, several models have been suggested to describe how P1 interacts with both other P1 molecules and with DNA.[42,43] The model proposed by Balhorn posits that the central arginine core of P1 must adopt an extended, non-helical conformation in order to completely bind the entire length of genomic DNA within the sperm cell and further, that only the central arginine segment inserts into the DNA grooves. The N-terminal and C-terminal "tails" containing S43 and T45 instead form hairpin-like structures that can interact with other protamine molecules and participate in intramolecular disulfide bonds (**Figure 3-9A, P1$^{+/+}$**). Therefore, phosphorylation at these sites may function to prevent interaction with DNA and instead, promote hairpin formation within the P1 molecule to

aid in intramolecular disulfide bond formation and interaction with other P1 molecules. It is possible that premature addition of negative charges to the 43 and 45 positions, because of their proximity to arginine residues, may introduce attractive electrostatics that alter the hairpin structure, which may result in either loss of intramolecular disulfide bond formation in the C-terminus or may affect P1 oligomerization.

Our *in vitro* analysis illustrated that protamines likely employ a cooperative binding mechanism to DNA, therefore necessitating a minimum threshold of protein for any level of binding/compaction. This then suggests that oligomerization of protamines may be necessary for genome compaction. Given that phosphorylation imparts a net negative charge, it is possible then that the negatively charged phosphates interact with the positively charged arginine residues of an adjacent P1 molecule, allowing for oligomerization of P1 molecules which is then stabilized by the formation of intermolecular disulfide bonds. Once stable P1 oligomerization is achieved, likely a portion of phosphorylation marks need to be removed to prevent electrostatic clashes with the DNA. Therefore, in this model the premature addition of negative charges via S43E/T45E substitutions may cause premature P1 oligomerization and/or electrostatic clashes with the DNA, resulting in the decrease in P1 incorporation into sperm chromatin that we observe in P1^{S43ET45E} males. The observation that P1^{S43AT45A} males do not exhibit an overt phenotype then may be explained by the significant increase in local protamine concentration during spermiogenesis, which may allow the protamine molecules to overcome the loss of the advantage phosphorylation provides in bringing P1 molecules together.

Lastly, a recent study reported that P1$^{S9A/S43A}$ males are infertile[33] and given that phosphorylation at P1 S9 is present at normal levels in P1^{S43AT45A} sperm but is almost completely absent in P1^{S43ET45E} sperm, it cannot be ruled out that the phenotype in P1^{S43ET45E} males is due to loss of S9 phosphorylation. Possibly, S9 phosphorylation can perform the function of S43/T45 phosphorylation in P1^{S43AT45A} sperm. Interestingly, levels of P1 K49ac are comparable between P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} males, suggesting there is crosstalk specifically between phosphorylation events. Multi-site phosphorylation has been documented on numerous proteins and in many cases serves to induce a conformational change, promoting either protein-protein interactions, downstream phosphorylation events, or disulfide bond formation.[44–46] Therefore, it is

possible that loss of the complete phosphorylation cascade in P1^{S43ET45E} males prevents necessary conformational changes for either protein structure or protein-protein interactions. Alternatively, premature negative charge addition may induce this conformational change too early, leading to chaotic protein oligomerization and chromatin reorganization.

Premature addition of negative charges at S43 and T45 may disrupt spermatid polarization and morphological transformation

In addition to a decrease in P1 incorporation into sperm chromatin, we also find that P1^{S43ET45E} males exhibit a complete loss of progressive sperm motility, which may in part be explained by the significant decrease and abnormal distribution of mitochondria in sperm from these males, as mitochondrial ATP generation is essential for sperm motility.[47] The morphological defects observed in P1^{S43ET45E} sperm are slightly overlapping, but largely distinct from those observed in P1$^{K49A/K49A}$ sperm, which we previously noted to have significant similarity to defects observed in P1$^{+/-}$, P2$^{-/-}$, and TNP1/TNP2$^{-/-}$ sperm. Interestingly, the majority of abnormalities in P1^{S43ET45E} sperm are related to head morphology, specifically with many of the sperm exhibiting an elongated shape and an overall smaller size, a phenotype that is likely reflective of the high level of DNA damage and altered chromatin composition/condensation. Future studies are needed to discern whether DNA breaks are occurring downstream of genome compaction as a consequence of less condensed chromatin or if the programmed breaks made earlier in spermiogenesis are not being properly repaired.

Interestingly, we also observe a significant number of sperm with tail defects, specifically broken tails. Together with our observation of decreased mitochondrial content in the midpieces of P1^{S43ET45E} sperm, it is possible that structural disruptions to the P1 protein caused by the premature addition of negative charges alters cytoskeletal remodeling/manchette formation during spermiogenesis, both of which are essential for proper formation of the sperm tail and reshaping of the sperm head. In fact, phosphorylated P1 has previously been reported to interact with the inner nuclear membrane protein Lamin B receptor and it was hypothesized that the nuclear lamina may act as a scaffold where P1 is initially sequestered, and regions of TNP-bound chromatin are brought to allow for controlled exchange.[48] Additionally, dephosphorylation then may initiate detachment of P1 from the nuclear lamina and subsequent deposition onto chromatin. Therefore,

in the absence of phosphorylation (the case in P1^{S43AT45A} males), P1 is not sequestered to the nuclear lamina, however given the extremely high levels of P1 that are present in the spermatid nucleus, likely this sheer overabundance in protein concentration can again overcome the loss of the advantage sequestration of P1 at the lamina confers. However, in the case of constitutive negative charge (the case in P1^{S43ET45E} males), it is possible that P1 is properly sequestered at the nuclear lamina, but never detaches, causing chaotic sperm cell reorganization (as the nuclear lamina plays an extensive role in cytoskeletal remodeling/spermatid polarization) and genome condensation. Future studies will aim to investigate the relationship between protamines and cytoskeletal remodeling/spermatid polarization, as it is possible that the tethering of P1 to the nuclear lamina may be involved in this process. Interestingly, unpublished immunoprecipitation followed by mass spectrometry (IP-MS) from our lab identified numerous P1 and P2-interacting proteins that are associated with the nuclear lamina, therefore supporting a direct relationship between protamine-based chromatin reorganization and spermatid remodeling.

Altogether, our *in vitro* and *in vivo* analysis of P1 S43/T45 phosphorylation and the genetic substitution of these residues to both alanine and glutamic acid supports a model in which phosphorylation at S43/T45 likely is not essential for chromatin reorganization during spermiogenesis. It is possible that temporally controlled phosphorylation at these sites aids in conferring structural characteristics to either the P1 protein itself or to P1 bound to DNA to promote favorable, yet fluid packaging as the genome undergoes extensive remodeling and condensation. Lack of an observable phenotype in P1^{S43AT45A} males may be explained by an overabundance of P1 protein in the spermatid nucleus, allowing this increase in local protein concentration to enable P1 to overcome the loss of the advantage imposed by phosphorylation. The loss of temporal control of phosphorylation at these sites by introducing constitutive negative charges, however (S43E/T45E) may abrogate precise control of sperm cell reorganization and genome condensation, resulting in altered chromatin condensation, DNA damage, and infertility.

3.5 Methods

Mice

All experiments utilizing rodents were carried out with prior approval of the University of Michigan Institutional Committee on Use and Care of Animals (Protocols: PRO00006047,

PRO00008135, PRO00010000) and in compliance with all guidelines established by the National Research Council Guide for the Care and Use of Laboratory Animals. Mice were housed in the University of Michigan animal facility with ad libitum access to food (Lab Diet #5008 for breeding mice and #5LOD for non-breeding animals) and water. The environment was controlled for light (12 hours on/off) and temperature (21 to 23°C).

P1$^{S43A/T45A}$ and P1$^{S43E/T45E}$ knock-in mice were generated by the Cincinnati Children's Hospital Transgenic Animal and Genome Editing Core Facility (see section 2.5 for complete details). Briefly, the guide RNA target sequence was selected based on proximity to the target site and according to the on- and off-target scores provided by CRISPOR.[49] Ribonucleoprotein (RNP) complexes were first formed via mixing of sgRNA with Cas9 protein (IDT) and incubation at 37 °C for 10 minutes followed by addition of the donor oligo (IDT) containing the intended mutation (either S43A/T45A or S43E/T45E). Zygotes from super-ovulated C57BL/6N females were electroporated with the RNP/donor oligo mix and subsequently transferred to pseudopregnant CD-1 females. All animal procedures were carried out in accordance with the Institutional Animal Care and Use Committee and approved protocol of Cincinnati Children's Hospital Medical Center. Offspring were genotyped for the intended mutations by extraction of genomic DNA from a small ear biopsy. Mutant males and control mice were used for all experiments between 8-16 weeks of age for all studies.

Antibodies

Rabbit polyclonal antibodies against total P1 and phosphorylation at P1 S43/T45 were generated at GeneMed Synthesis Inc. via immunization of rabbits with the following synthesized peptides: P1- CRRRRSYTIRSKKY, P1 S43/T45- RRCCRRRRS(ph)YT(ph)IRCKKY. All other antibodies used are provided in Table 3-1.

Acid extraction of sperm basic proteins

Acid extraction of basic proteins from mature mouse sperm was performed as previously described (see also Section 2.5).[50] Sperm pellets were resuspended in 1 mM PMSF (in water) and spun down at 8,000xg for 8 minutes. Following removal of the supernatant, sperm pellets were then resuspended in 100 uL of 100 mM Tris pH 8.0, 20 mM EDTA, and 1 mM PMSF. Proteins were

then denatured by addition of 100 uL of 6 M Guanidine-HCl, 575 mM DTT and alkylated with 200 uL of 522 mM sodium acetate for 30 minutes in the dark. Following two washes of the pellets in 100% ethanol, basic proteins were extracted with 800 uL of 0.5 M HCl, 50 mM DTT at 37°C for 10 minutes. After centrifugation at 12,000xg for 5 minutes, supernatants containing soluble basic proteins were precipitated overnight at -20°C with TCA to a final concentration of 20%. The following day, precipitated proteins were spun down at 12,000xg for 8 minutes, washed twice in 1 ml of 1% 2-mercaptoethanol in cold acetone, and finally resuspended in water after drying.

Peptide competition assay to assess antibody specificity

Basic proteins were first extracted directly from sperm as described above. Varying protein quantities were loaded onto each immunoblot (exact quantities specified in figure). Prior to addition to immunoblots, antibodies and peptides were incubated together at room temperature for 30 minutes, with either 10-fold excess of specific or non-specific peptide, or alone. Following blocking in 5% non-fat milk in TBST, blots were incubated for 1.5 hours with antibody, antibody with specific peptide, or antibody with non-specific peptide. The peptide used against a non-specific protein was N-DSNKEFGTSNESTE-C and the non-phosphorylated P1 peptide used was N-CRRRRSYTIRSKKY-C.

Mass spectrometry analysis of mouse protamines

Mass spectrometry of mouse sperm was performed at MS BioWorks in Ann Arbor, MI. Prior to MS analysis, sperm basic proteins were extracted using the method described above. 20 ug of extracted protein was then run in triplicate on a 4-20% SDS-PAGE gel (BioRad) and a single band from each lane was cut out for analysis. Gel bands were then washed once with 25 mM ammonium bicarbonate, washed three times in 100% acetonitrile, reduced with 10 mM dithiothreitol at 60°C, and proteins were alkylated with 50 mM light iodoacetamide at room temperature. Bands were then digested with either trypsin (Promega) at 37°C for 4 hours, Chymotrypsin (Promega) at 37°C for 12 hours, or Lys-C (Promega) at 37°C for 12 hours. For all enzymes used, digests were quenched with formic acid and the supernatants were analyzed. Digests were analyzed by nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher Q Exactive. Peptides were loaded on a trapping column and eluted over a 75 um analytical column at 350 nL/min. Both columns were packed with Luna C18 resin (Phenomenex). The mass spectrometer

was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 70,000 FWHM and 17,500 FWHM resolution, respectively. The fifteen most abundant ions were selected for MS/MS. Data were searched using a local copy of Byonic with the following parameters: Enzyme: Semi-Trypsin or None (for Chymotrypsin and Lys-C), Database: Swissprot Mouse (forward and reverse appended with common), fixed modification: carbamidomethyl (C), variable modifications: oxidation (O), acetyl (protein N-term), deamidation (NQ), phosphor (STY), methyl (KR), dimethyl (KR), trimethyl (K), mass values: monoisotopic, peptide mass tolerance (10 ppm), fragment mass tolerance (0.02 Da), max missed cleavages: 2. Mascot DAT files were parsed into the Scaffold software for validation, filtering and to create a non-redundant list per sample. Data were filtered using a minimum protein value of 95%, a minimum peptide value of 50% (Prophet scores) and requiring at least two unique peptides per protein. Site localization probabilities were assigned using A-Score.[51]

Phylogenetic analysis of P1 sequence conservation across species

Construction of the phylogenetic tree was performed using maximum likelihood (PhyML) inferred from protamine 1 sequences across the orders Rodentia, Primate, and Artiodactyla using the Whelan and Goldman matrix (WAG) substitution strategy. Protein sequences were acquired from NCBI and aligned using MUSCLE (standard parameters).[52] Bootstrap support with 1,000 replicates is designated for each node of the tree and values above 95 indicate strong support.

Immunofluorescence and quantification of seminiferous tubule staging

Adult testes were fixed in 4% PFA at 4°C overnight followed by 18 hours in PBS at 4°C then submersion in 70% ethanol and subsequent processing for formalin fixed paraffin embedding (FFPE). Deparaffinization of five-micron thick testes sections was followed by permeabilization and antigen retrieval via boiling in 10 mM sodium citrate pH 6.0 for 10 minutes. Primary antibodies were incubated with tissue sections overnight at 4°C. PNA-Lectin (GeneTex) was used as a marker of the acrosome and DAPI was used as the nuclear counterstain in all images. All AlexaFlour-conjugated secondary antibodies (Life Technologies/Molecular Probes) were used at a dilution of 1:1000. For assessment of seminiferous tubule staging, Lectin was used to stage each individual tubule and tubules were split into four categories for quantification (I-III, IV-VI, VII-VIII, IX-XII) as previously described.[53]

Phenotyping of P1$^{S43A/T45A}$ and P1^{S43ET45E} mice

Males were phenotyped between 9-10 weeks of age. Sperm counting was performed using a Makler chamber, with 3 technical replicates per mouse (n values for males per genotype are indicated in each figure panel). A minimum of 100 sperm were counted for assessment of progressive sperm motility where forward movement was quantified in comparison to the total sperm count. For fecundity quantifications, 8-9 week old males (n=3 per genotype) were placed with 8-week-old C57BL/6J females after a 3 day period of being individually housed. A total of 3 females per male (for a total of 9 females per genotype of male) were checked daily for the presence of copulatory plugs to confirm mating. Once mating occurred, females were separated into a new cage and the percent of females that were successfully impregnated was recorded as fecundity.

Mitochondrial content assessment by Mitotracker

Sperm were collected from both the epididymis and vas deferens of adult P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} males and incubated at 37°C for 1 hour. A small aliquot (~5 uL) of sperm was then plated on a slide and allowed to air dry (sperm were not pelleted prior to plating to prevent removal of the sperm tails). Sperm were then fixed in 4% paraformaldehyde for 15 minutes at room temperature and washed three times in PBS for 10 minutes each. Sperm were then permeabilized and blocked for 40 minutes in PBS with 0.5% Triton and 3% BSA at room temperature, followed by incubation with DAPI, PNA-Lectin (as an acrosomal stain), and Mitotracker for 1 hour at room temperature. Mitotracker was used at a final concentration of 1 uM. Slides were then washed three times with PBS before mounting in Vectashield.

Comet assay for assessing sperm DNA damage

A neutral comet assay was performed using a Comet Assay Kit (Trevigen), with modifications for sperm as previously described.[14] Briefly, sperm were collected from the vas and epididymis as per above and immediately resuspended in PBS (-/-) at a concentration of 1 x 10^6. Cells were combined with molten agarose at a ratio of 1:10 and 50µL immediately pipetted onto the provided CometSlide. The slides were immersed for 1 hour at 4°C in pre-chilled lysis buffer and then incubated for 18 hours in lysis buffer containing 500µg/mL proteinase K followed by

electrophoresis at 21 V for 45 minutes. Slides were immersed in DNA precipitation and 70% ethanol as per the kit and then stained with SYBR gold. Over 500 sperm were assessed for each group to determine the presence or absence of a comet tail. The proportion of sperm with comets was then calculated and compared using the Chi-Square test. P<0.05 was considered statistically significant.

Acid urea gel electrophoresis for separation of sperm basic proteins

Sperm were harvested from both the epididymis and vas deferens of adult males and pelleted at 2500xg (an identical number of sperm was used for each genotype). Sperm were washed once in 1 mM PMSF, pelleted at 8,000xg for 8 minutes, then resuspended in 1 ml of 6 M Guanidine-HCl, 500 mM Hepes pH 7.5, 10 mM DTT for 1 hour at room temperature. Alkylation of cysteine residues was then carried out by addition of vinylpyridine to a final concentration of 250 mM for 1.5 hours at room temperature. Concentrated HCl was then added to a final concentration of 0.9 M and the samples were then dialyzed overnight against 0.2 M HCl at 4°C with gentle stirring of the buffer. Samples were then centrifuged at 12,000xg for 5 minutes to remove insoluble proteins followed by precipitation of basic proteins with TCA to a final concentration of 20% for 4 hours at -20°C. Proteins pellets were washed twice with cold acetone and finally resuspended 0.9 M acetic acid, 8 M urea, 100 mM betamercaptoethanol. For acid urea gel electrophoresis, an identical volume of protein was loaded for each sample (afforded by an identical number of input sperm to the extraction). Acid urea gels were made as previously described and P1:P2 ratios were calculated using band intensities from Coomassie stained gels in ImageJ.

Assessment of histone levels in mature sperm

Sperm pellets from P1$^{+/+}$, P1$^{S43AT45A/+}$, P1^{S43AT45A}, P1$^{S43ET45E+/}$, and P1^{S43ET45E} males (sperm was pooled from 2-3 males per genotype) were resuspended in 20 mM Tris pH 7.5, 1 mM MgCl$_2$, 1 mM CaCl$_2$, 137 mM NaCl, 10% glycerol, 1% NP-40, 12.5 U/ml Benzonase, 1X protease inhibitors, sonicated briefly, and rotated for 1 hour at 4°C. For immunoblots, each lane was loaded by input sperm number, which varied for each antibody given differences in antibody efficiency. For histone H3 and histone H2B, 25000, 50000, or 100000 sperm were loaded. Alpha tubulin was used as a loading control in all cases.

Purification of protamines from sperm and *in vitro* electrophoretic mobility shift assays

Proteins were initially extracted directly from sperm as described above. After precipitation of proteins with TCA and subsequent washes of the pellets with acetone, final pellets were resuspended in 50 uL of water and then brought up to 500 uL with 25 mM Hepes pH 7.5, 150 mM NaCl, 5 mM TCEP (TCEP was not pH neutralized on its own). Acid extracted proteins were then subjected to size exclusion chromatography using a Superdex S75 column. Absorbance at 214 nm was used to identify fractions containing P1 and P2 and fractions were confirmed by immunoblotting. *In vitro* electrophoretic mobility shift assays were performed using varying concentrations of purified protamines and 40 nM dsDNA of 280 bp. DNA was prepared by PCR amplification of mouse genomic DNA (primers listed in Table 3-3). Proteins were first incubated in reaction buffer for 10 minutes at 37°C at which point DNA was added. After incubating for 1 hour at room temperature, reactions were loaded onto a non-denaturing 0.5X TBE 6% polyacrylamide gel and stained with ethidium bromide (Sigma). Quantification of unbound DNA was performed using ImageJ.

3.6 Acknowledgements

Table 3-1 List of reagents and resources used in Chapter 3

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Protamine 1	This study	N/A
Protamine 1 S43/T45ph	This study	N/A
Protamine 1	Briarpatch Biosciences	Cat# Hup1N RRID: AB_2651186
Protamine 2	Briarpatch Biosciences	Cat#: Hup2B RRID: AB_2687949
Histone H2B	Abcam	Cat#:ab52484 RRID: AB_1139809
Histone H3	Abcam	Cat#: ab1791 RRID: AB_302613
Tetra-acetyl H4	EMD Millipore	Cat#: 06-866 RRID: AB_310270
Alpha tubulin	Protein Tech	Cat#: 66031-1-Ig RRID:AB_11042766
Tnp1	Protein Tech	Cat#17178-1-AP RRID: AB_2206757
Tnp2	Santa Cruz	Cat#: sc- 393843 RRID: N/A
PNA-Lectin	Genetex	Cat#GTX01508 RRID: N/A
Chemicals, peptides, and recombinant proteins		
Trichloroacetic acid	Sigma	Cat#**T0699**
4-vinylpyridine	Sigma	Cat#V3204
NdeI	New England Biolabs	Cat#R0111
NruI	New England Biolabs	Cat#R3192
Benzonase	EMD Millipore	Cat#71206
P1 peptide	This study	See Table 3-3
P1 S43/T45ph peptide	This study	See Table 3-3
Tcf21 (non-specific protein) peptide	This study	See Table 3-3
Mitotracker		Cat#M7512
Critical commercial assays		
BLOXALL® Endogenous Peroxidase and Alkaline Phosphatase Blocking Solution	Vector Labs	Cat#SP-6000-100
Comet Assay Kit	Trevigen	Cat#4250-050-K
Experimental models: Organisms/strains		
C57BL/6J	Jackson Labs	Stock #000664
P1^{S43AT45A}	This study	N/A
P1^{S43ET45E}	This study	N/A
Oligonucleotides		

P1 K49A genotyping primer_Forward, see Table S3	This study	
P1 K49A genotyping primer_Reverse, see Table S3	This study	
sgRNA for P1$^{K49A/K49A}$ generation, see Table S3	This study	
Donor DNA for P1$^{K49A/K49A}$ generation, see Table S3	This study	
280 bp dsDNA oligo, see Table S3	This study	
Software and algorithms		
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
GraphPad Prism 9	Graphpad Software	N/A
Scaffold PTM	Scaffold PTM software	N/A

Table 3-2 Peptides used in Chapter 2

Peptide	Sequence
Protamine 1	CRRRRSYTIRSKKY
Protamine 1 S43/T45ph	RRCCRRRRS(ph)YT(ph)IRCKKY
Tcf21 (non-specific protein)	DSNKEFGTSNESTE

Table 3-3 Oligonucleotides used in Chapter 3

Oligonucleotide	Sequence
Genotyping primer, Forward	5'-GAAGATGTCGCAGACGGAGGAG-3'
Genotyping primer, Reverse	5'-ACGGTGGCATTTTTCAAGATGTG-3'
sgRNA for P1^{S43AT45A} and P1^{S43ET45E} generation	5'- ACCTTATGGTGTATGAGCGGCGG-3'
Donor DNA for P1^{S43AT45A} generation	5'-TTTCTTACCTTTCTAGGATGCTGCCGTCGCCGCCGCgCATAtgCCATAAGGTGTAAAAAATACTAGATGCACAGAATAGCAAGTCCATCAAAACTCCTGCGTGAGAATTTTACCAGACTTCAAGAGCATCTCG-3'
Donor DNA for P1^{S43ET45E} generation	5'-CACCACTTTTCTTACCTTTCTAGGATGCTGCCGTCGCCGTCGCGAATACGAGATAAGGTGTAAAAAATACTAGATGCACAGAATAGCAAGTCCATCAAAACTCCTGCGTGAGAATTTTACCAGACTTCAAGAGCATCTCGCCA-3'
280 bp dsDNA oligo used for EMSAs	5'CTGGACAAGACCATGAACGCGAGGAGCAGGGGCAGGGGCAAGGGCTGAGCCCAGAGCGCGTAGAGGACTATGGGAGGACACACAGGGGCCACCACCACCACAGACACAGGCGCTGCTCTCGTAAGAGGCTACATAGGATCCACAAGAGGCGTCGGTCATGCAGAAGGCGGAGGAGACACTCCTGCCGCCACAGGAGGCGGCATCGCAGAGGTAAGCACCCCACAGCCGACCCCCTGGCCACCTGTGCTGCTGCTGCCCATCTAAACCCTGCTGCCTTCCA-3'
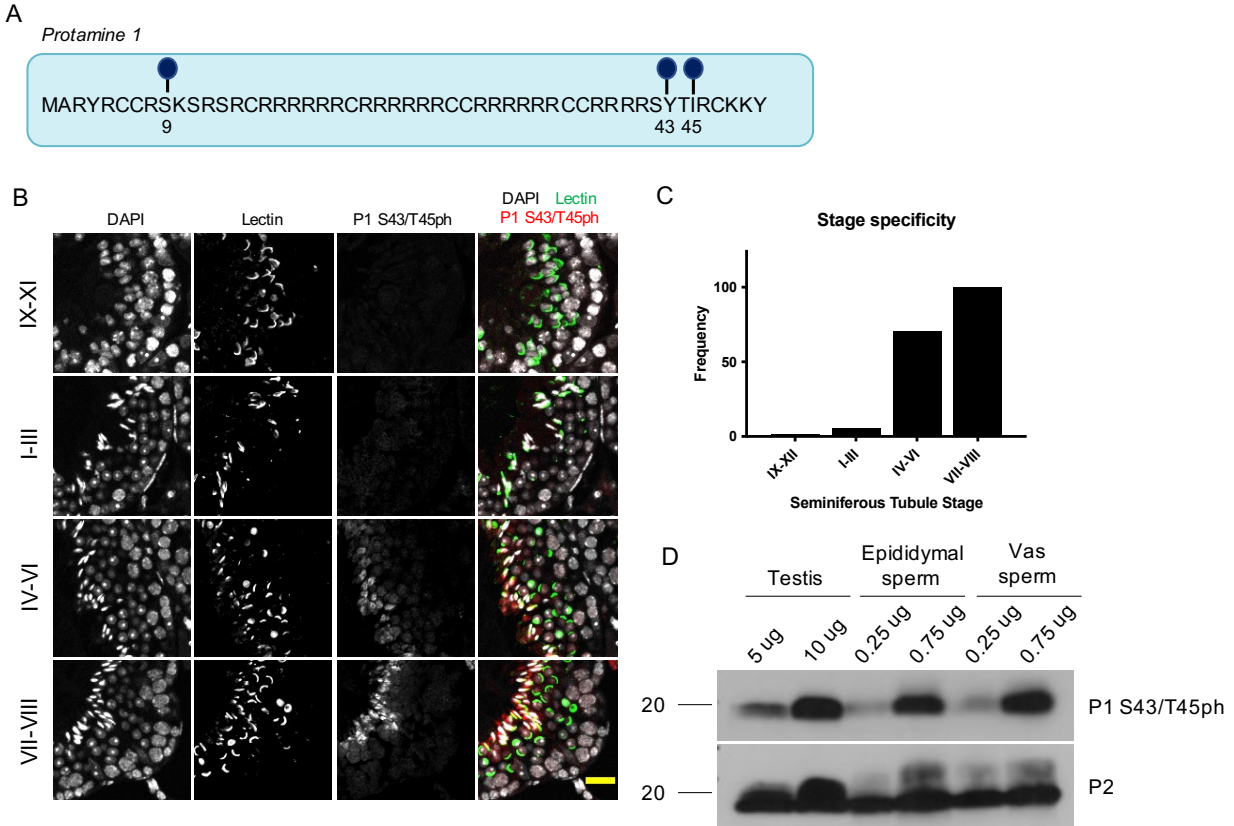
Figure 3-1 P1 S43/T45 phosphorylation is acquired in late-stage spermatids and persists in mature mouse sperm. (A) Schematic of phosphorylation marks identified on mouse P1 using bottom-up mass spectrometry. **(B)** Immunofluorescence staining of P1 S43/T45ph in adult testes cross sections across the seminiferous epithelial cycle. PNA-Lectin was used as the acrosomal marker. Representative images from n=3 mice. Scale bar: 20 μm. **(C)** Quantification of P1 S43/T45ph stage specificity across seminiferous tubule stages, illustrating a gradual increase in P1 S43/T45ph across spermatid maturation**. (D)** Immunoblotting of P1 S43/T45ph from testes lysates, sperm from the epididymis, and sperm from the vas deferens, highlighting persistence of S43/T45 phosphorylation from the testis to mature sperm.
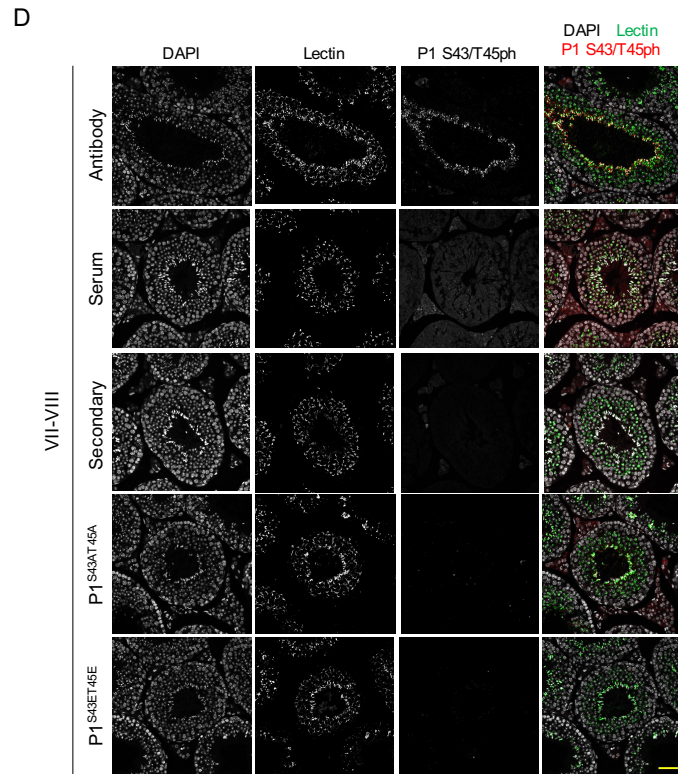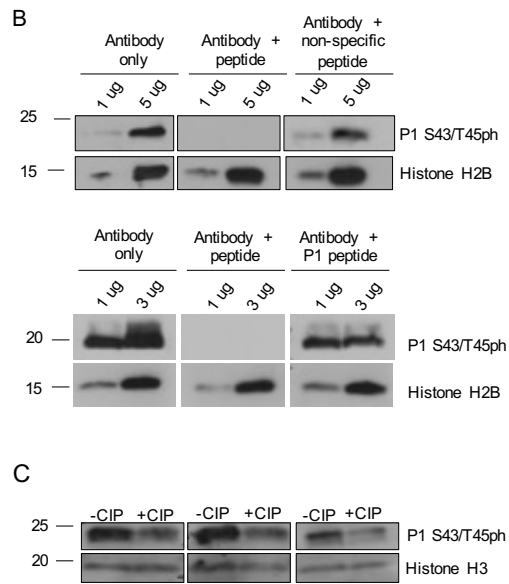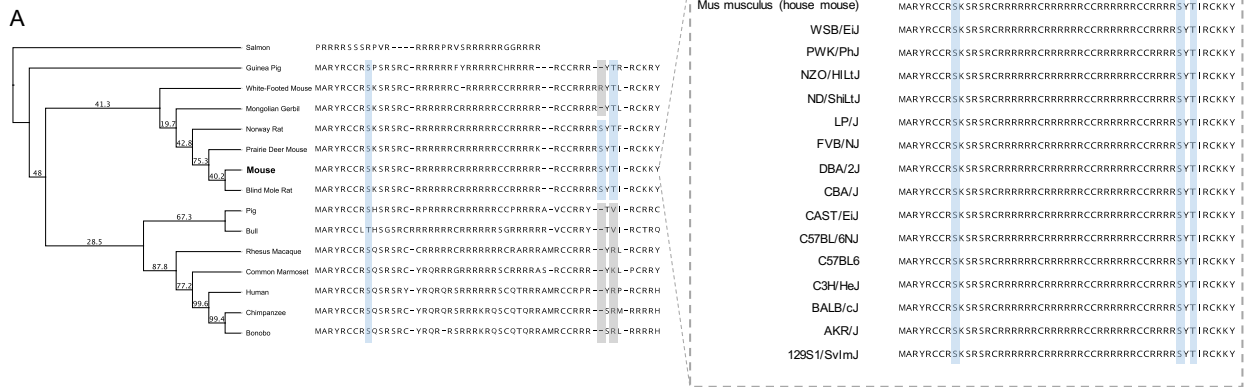
Figure 3-2 P1 modifications are present on lineage specific residues. (A) Phylogenetic tree constructed using maximum likelihood of P1 protein sequences across Rodentia, Primate, and Artiodactyla. The tree was constructed using the WAG substitution strategy, with bootstrap support with 1,000 replicates shown per node. S43 and T45 are highlighted in blue where conserved and in gray when occupied by alternative residues or absent. S9 is highlighted in blue across all species. **(B)** Peptide competition immunoblots of acid extracted protein from mature sperm illustrates the presence of a clear band for P1 S43/T45 phosphorylation that is abolished only when competed by a phosphorylated P1 peptide. **(C)** Immunoblots of calf intestinal phosphatase (CIP)-treated sperm protein lysates shows a clear reduction in signal intensity for P1 S43/T45ph only when phosphatase treated. **(D)** Immunofluorescence of adult testes cross sections using P1 S43/T45ph antibody, pre-immune serum from rabbits used for antibody generation (top panels), secondary antibody only, or cross sections from P1^{S43AT45A} or P1^{S43ET45E} testes illustrates specificity of the antibody.

A

S43A/T45A: ......RCCRRRRRRCCRRRR**AYA**IRCAKY
S43E/T45E: ......RCCRRRRRRCCRRRR**EYE**IRCKKY

P1+/+

R R S Y T I R

P1^S43A/T45A

R R A Y A I R

NdeI

P1^S43E/T45E

R R A Y A I R

NruI

B Sperm Count

C Progressive sperm motility

G Fecundity

D

Normal | Abnormal head | Bent back head | Broken tail

P1^S43AT45A

P1^S43ET45E

E Morphology

n=454 | n=534

■ Broken/abnormal tail
■ Bent back head
■ Abnormal head morphology
■ Normal

F

P1+/+ | P1^S43AT45A | P1^S43ET45E

DAPI Lectin Mitotracker

Figure 3-3 Substitution of P1 S43 and T45 to glutamic acid, but not alanine, results in sperm motility defects and infertility. (A) Schematic of modifications made to the mouse P1 sequence and corresponding Sanger sequencing traces show successful mutation of S43/T45 to both alanine and glutamic acid. Note that synonymous mutations were incorporated into the donor DNA to introduce either an NdeI site (for S43A/T45A) or an NruI site (for S43E/T45E) for genotyping purposes. **(B)** Total epididymal sperm count from a minimum of n=3 males per genotype. **(C)** Epididymal sperm progressive motility after 1 hour of incubation at 37 °C for a minimum of n=3 males per genotype. **(D)** Representative Hematoxylin and eosin-stained mature sperm from either a P1^{S43AT45A} male or a P1^{S43ET45E} male, highlighting the major observed abnormalities. Scale bar: 20 μm. **(E)** Quantification of major abnormalities observed in P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} mature sperm. Sperm was assessed from n=2 males per genotype. **(F)** Mitotracker staining of P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} mature sperm highlights a decrease as well as an abnormal distribution of mitochondria in P1^{S43ET45E} sperm. **(G)** Fertility assessment of 3 adult males per genotype as measured by percent of females impregnated (fecundity). Each dot represents measurement from a single animal. Statistical test was performed using a Kruskal-Wallis test and was adjusted for multiple comparisons. For B, C, and G, center line represents the mean and error bars represent standard deviation.

A

Genomic location	Number of mismatches	Sequence (including mismatches)	Genomic location
Chr2: 14571077	0	ACATTATTGTGTATGAGCAGTGG	intergenic:Gm13266-Gm10849
Chr17: 47494774	0	ACCTTACAGTGTATGAGTGGTGG	intron:Taf8
Chr18: 32840066	0	ACATCAGGGTGTATGAGCTGTGG	intron:Wdr36
Chr17: 72118878	0	ATCTTATGGAATATGAGCAGTGG	intron:Alk
Chr9: 45763594	0	ACTCCATGGTGTATGTGCGGAGG	intergenic:Dscaml1-Cep164
Chr14: 43986507	0	CCCTTCTGGTTAATGAGCGGTGG	intergenic:Ear-ps7-Ear-ps10
Chr5: 116003598	0	ACCTTAACATGTCTGAGCGGTGG	intron:Cit
Chr6: 99007405	0	ACTTAATAGTGTAGGAGCGGTGG	intron:Foxp1
Chr15: 83684134	0	GCCTTCTGGTCTCTGAGCGGAGG	intron:Scube1
Chr 6: 34086217	0	ACTTTATGATTTATGATCGGTGG	intron:Lrguk-T10

Genomic location	Number of mismatches	Sequence (including mismatches)	Genomic location
Chr2: 14571077	0	ACATTATTGTGTATGAGCAGTGG	intergenic:Gm13266-Gm10849
Chr17: 47494774	0	ACCTTACAGTGTATGAGTGGTGG	intron:Taf8
Chr18: 32840066	0	ACATCAGGGTGTATGAGCTGTGG	intron:Wdr36
Chr17: 72118878	0	ATCTTATGGAATATGAGCAGTGG	intron:Alk
Chr9: 45763594	0	ACTCCATGGTGTATGTGCGGAGG	intergenic:Dscaml1-Cep164
Chr14: 43986507	0	CCCTTCTGGTTAATGAGCGGTGG	intergenic:Ear-ps7-Ear-ps10
Chr5: 116003598	0	ACCTTAACATGTCTGAGCGGTGG	intron:Cit
Chr6: 99007405	0	ACTTAATAGTGTAGGAGCGGTGG	intron:Foxp1
Chr15: 83684134	0	GCCTTCTGGTCTCTGAGCGGAGG	intron:Scube1
Chr 6: 34086217	0	ACTTTATGATTTATGATCGGTGG	intron:Lrguk-T10

Figure 3-4 Neither P1^{S43AT45A} nor P1^{S43ET45E} males exhibit abnormal testis parameters or loss of germ cell populations (A) List of potential off-target genetic modifications and corresponding Sanger sequencing results verifying the absence of any of these off-target mutations generated by CRISPR/Cas9. **(B)** Testes/body weight ratio of P1^{S43AT45A} and P1^{S43ET45E} (a minimum of n=3 per genotype) suggests no loss of germ cell populations due to either set of substitutions. Statistical test was performed using a one-way ANOVA and adjusted for multiple comparisons, p=0.9. Center line represents the mean and error bars represent standard deviation. **(C)** Periodic acid Schiff (PAS) staining of adult testes cross sections shows normal testis morphology in both P1^{S43AT45A} and P1^{S43ET45E} males. **(D)** Quantification of the percent of sperm cells with a "comet" using a Comet assay of DNA damage highlights an increase in DNA damage in P1^{S43ET45E} sperm. **(E)** Immunofluorescence of adult testes cross sections from P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} males stained for P1 illustrates no loss of P1 expression for either set of substitutions. Scale bars: 50 μm. **(F)** Immunoblot of protein lysates from P1$^{+/+}$, P1^{S43AT45A}, and P1^{S43ET45E} testes shows comparable P1 expression across all genotypes.
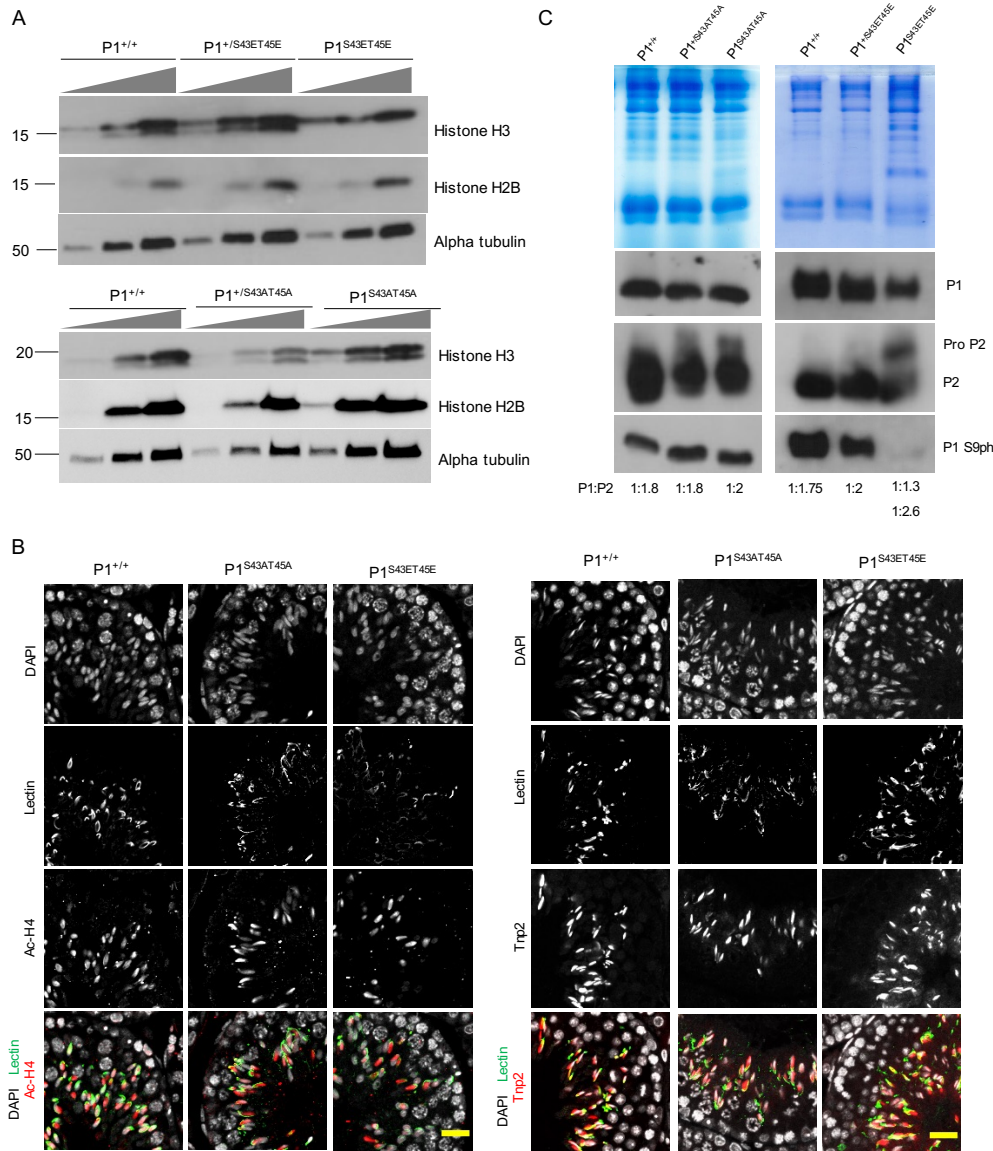
Figure 3-5 Both P1^S43AT45A and P1^S43ET45E males undergo proper histone-to-protamine exchange, yet P1^S43ET45E males exhibit altered sperm chromatin composition (A) Immunoblotting of sperm protein extracts reveals no increase in histone retention in either P1^S43AT45A or P1^S43ET45E sperm. **(B)** Immunofluorescence staining of adult P1^+/+, P1^S43AT45A, or P1^S43ET45E testes cross sections for ac-H4 (left panels) or Tnp2 (right panels). Scale bar: 20 μm. **(C)** Acid urea gel electrophoresis of sperm basic proteins reveals no difference in P1:P2 ratio in P1^S43AT45A males but a shift in P1:P2 ratio in P1^S43ET45E males by Coomassie blue staining (top). Immunoblotting reveals normal P1 and P2 levels in P1^S43AT45A sperm but a significant decrease in P1 in P1^S43ET45E sperm as well as a slight increase in pro P2 (bottom panels). Additionally, S9 phosphorylation is lost in P1^S43ET45E sperm but not P1^S43AT45A sperm.
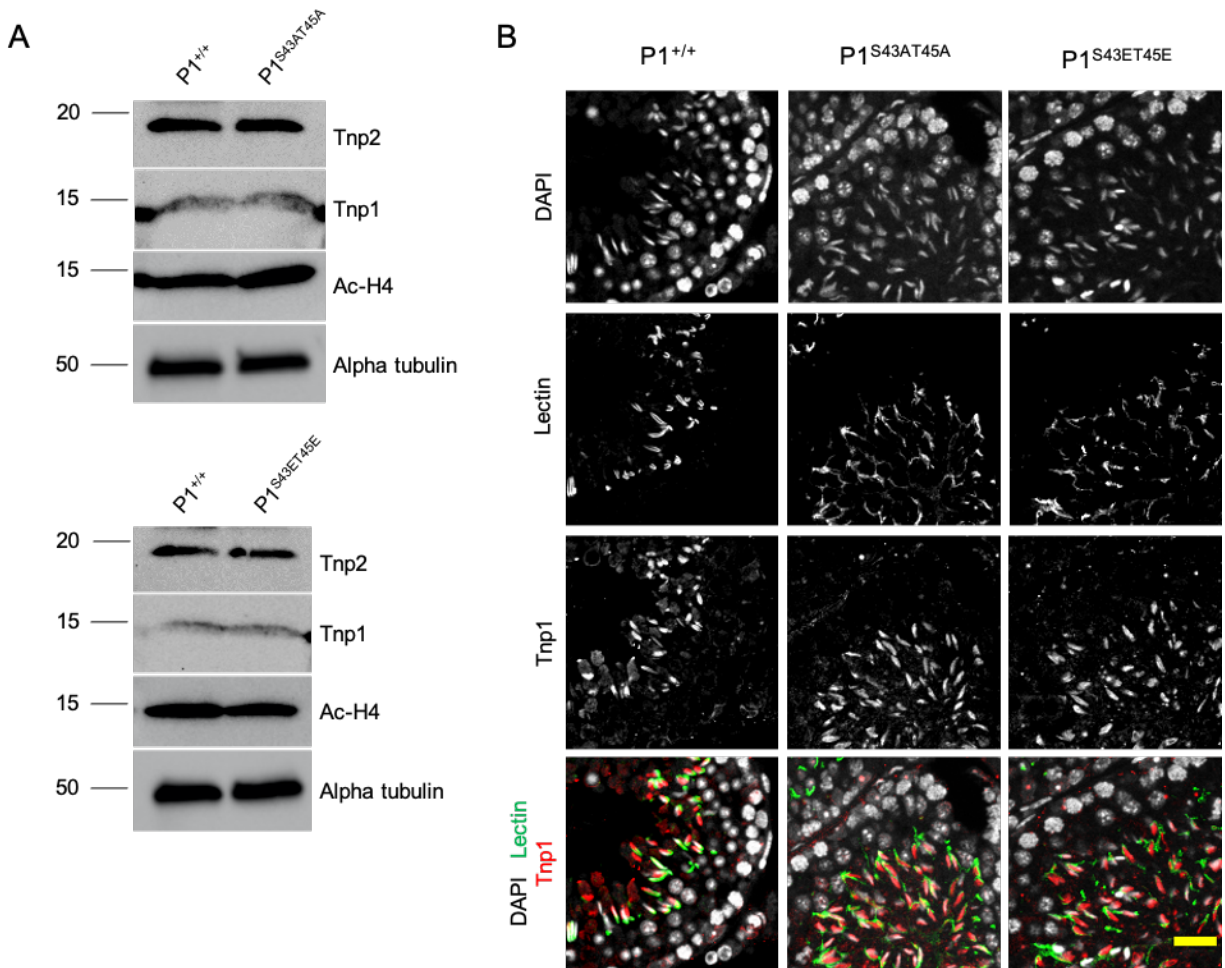
Figure 3-6 Both P1^S43AT45A and P1^S43ET45E males undergo proper histone-to-protamine exchange. (A) Immunoblots of testes lysates from P1^+/+, P1^S43AT45A, or P1^S43ET45E males reveals no difference in ac-H4, TNP1, or TNP2 levels. **(B)** Immunofluorescence staining of adult P1^+/+, P1^S43AT45A, or P1^S43ET45E testes cross sections reveals no difference in TNP1 levels. Scale bar: 20 μm.
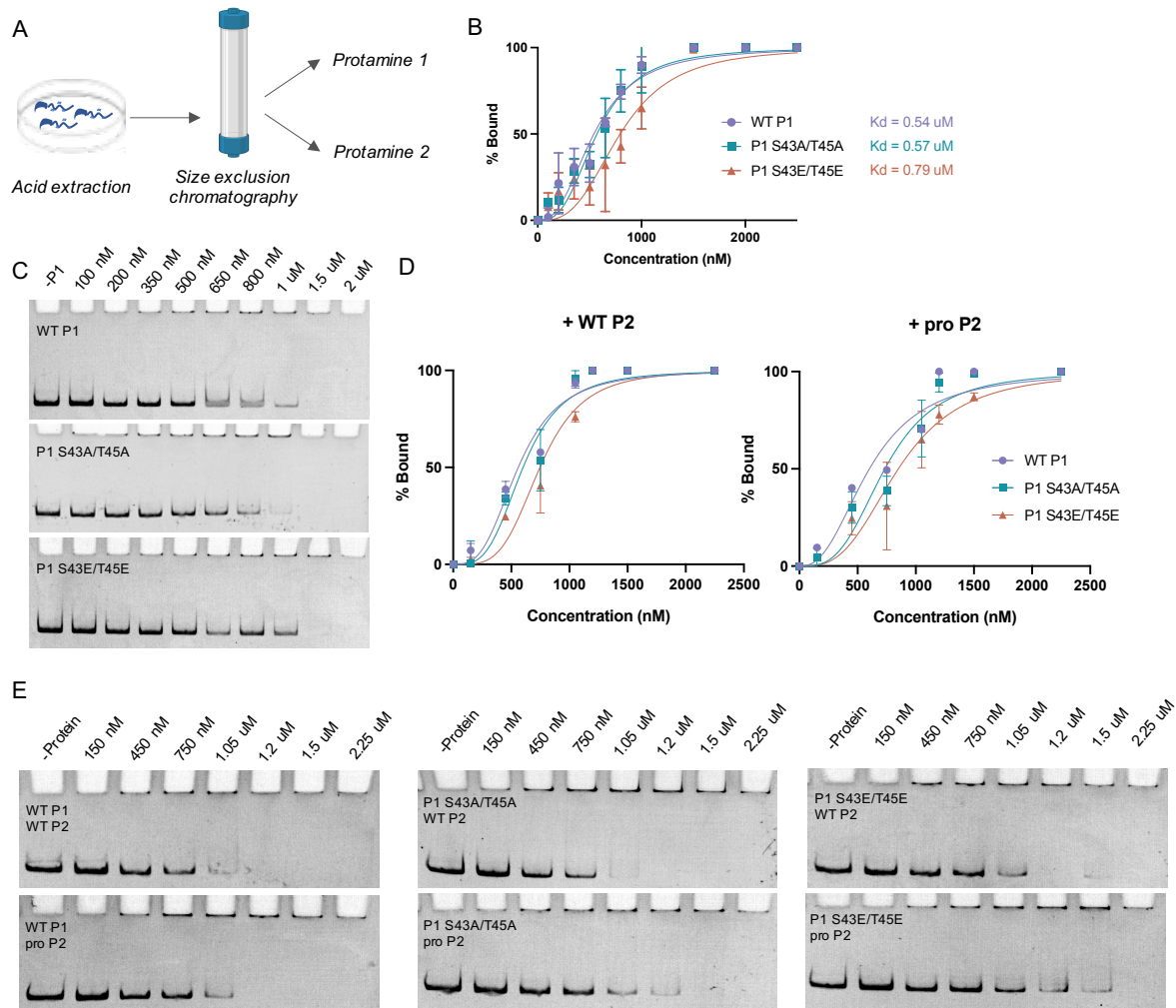
Figure 3-7 P1 S43E/T45E substitutions negatively impact DNA binding (A) Schematic of purification scheme used to purify P1 and P2 from mature mouse sperm. **(B)** Quantification of the binding affinities of WT P1, P1 S43A/T45A, and P1 S43E/T45E to a linear 280 bp DNA fragment. $K_{d,app}$ values were calculated using the Hill equation and were taken from at least 3 technical replicates per protein. Error bars represent standard deviation. **(C)** Representative EMSAs of a titration of increasing amounts of WT P1, P1 S43A/T45A, or P1 S43E/T45E. **(D)** Quantification of the binding affinities of WT P1, P1 S43A/T45A, or P1 S43E/T45E combined with either WT P2 or pro P2. **(E)** Representative EMSAs of a titration of increasing amounts of indicated P1 and P2 mixed in a 1:2 ratio. $K_{d,app}$ values were calculated using the Hill equation and were taken from at least 3 technical replicates per protein. Error bars represent standard deviation.

103
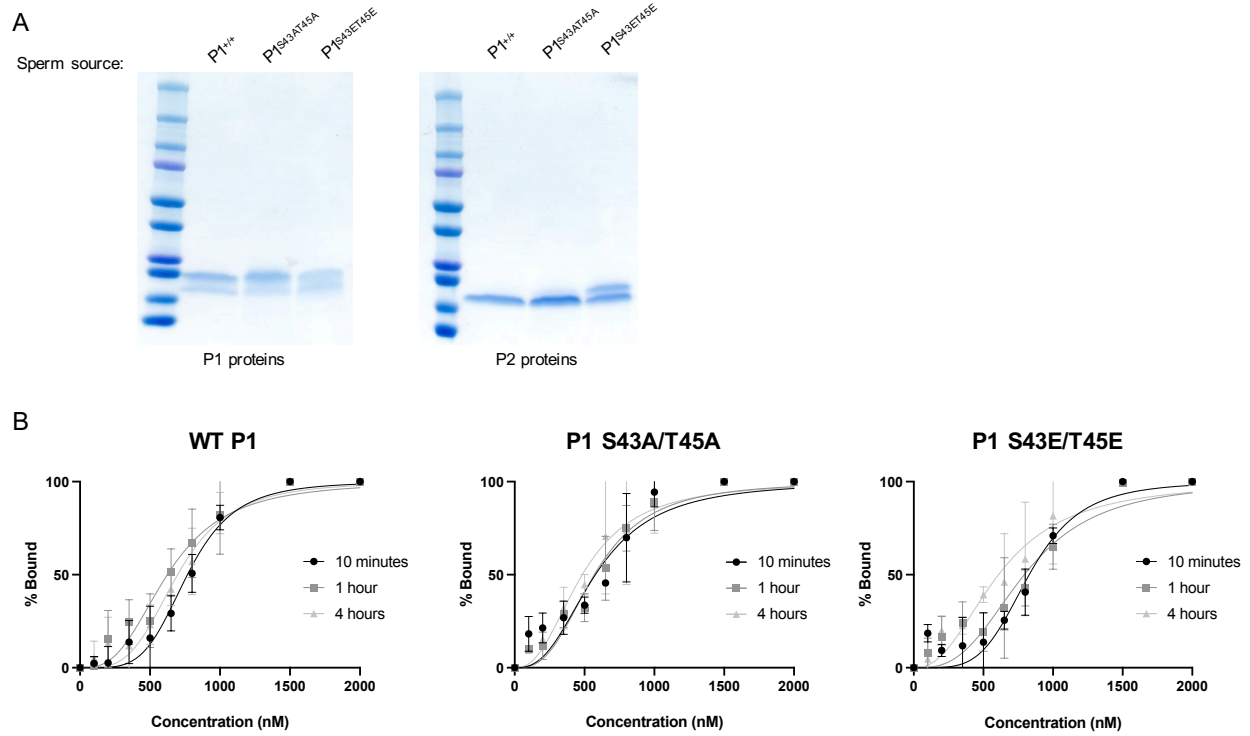
Figure 3-8 P1 S43E/T45E substitutions negatively impact DNA binding (A) Coomassie stained SDS-PAGE gel of purified WT and mutant protamines highlighting their purity. **(B)** Quantification of binding affinities of WT P1, P1 S43A/T45A, and P1 S43E/T45E after 10 minutes, 1 hour, or 4 hours of equilibration with DNA. Data were fit to the Hill equation and error bars represent standard deviation.

Figure 3-9 Model of potential structural alterations caused by S43E/T45E substitutions (A) Model of WT P1 protein (top), highlighting DNA binding via interactions with the central arginine segment, hairpin formation by the N- and C-terminal tails, and intramolecular disulfide bonds (dotted lines). (B) Potential structural alterations caused by the S43E/T45E substitutions. Top model representative of lost hairpin formation and bottom panel representative of premature attractive electrostatics between glutamic acid and neighboring arginine residues, causing alterations to position of DNA binding and intramolecular disulfides.

3.7 References

1. Yan, W., Ma, L., Burns, K. H. & Matzuk, M. M. HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. *Proc Natl Acad Sci U S A* **100**, 10546–10551 (2003).
2. Barral, S. *et al.* Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. *Mol Cell* **66**, 89-101 e8 (2017).
3. Govin, J. *et al.* Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol* **176**, 283–294 (2007).
4. Montellier, E. *et al.* Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev* **27**, 1680–1692 (2013).
5. Shinagawa, T. *et al.* Disruption of Th2a and Th2b genes causes defects in spermatogenesis. *Development* **142**, 1287–1292 (2015).
6. Tachiwana, H., Osakabe, A., Kimura, H. & Kurumizaka, H. Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins in vitro. *Nucleic Acids Res* **36**, 2208–2218 (2008).
7. Tachiwana, H. *et al.* Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc Natl Acad Sci U S A* **107**, 10454–10459 (2010).
8. Shirakata, Y., Hiradate, Y., Inoue, H., Sato, E. & Tanemura, K. Histone h4 modification during mouse spermatogenesis. *J Reprod Dev* **60**, 383–387 (2014).
9. Meistrich, M. L., Trostle-Weige, P. K., Lin, R., Allis, C. D. & Bhatnagar, Y. M. Highly acetylated H4 is associated with histone displacement in rat spermatids. *Molecular Reproduction and Development* **31**, (1992).
10. Lu, L. Y. *et al.* RNF8-dependent histone modifications regulate nucleosome removal during spermatogenesis. *Dev Cell* **18**, 371–384 (2010).
11. Yu, Y. E. *et al.* Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proceedings of the National Academy of Sciences* **97**, (2000).
12. Shirley, C. R., Hayashi, S., Mounsey, S., Yanagimachi, R. & Meistrich, M. L. Abnormalities and Reduced Reproductive Potential of Sperm from Tnp1- and Tnp2-Null Double Mutant Mice1. *Biology of Reproduction* **71**, (2004).
13. Cho, C. *et al.* Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nat Genet* **28**, 82–86 (2001).
14. Schneider, S. *et al.* Re-visiting the Protamine-2 locus: deletion, but not haploinsufficiency, renders male mice infertile. *Sci Rep* **6**, 36764 (2016).
15. Takeda, N. *et al.* Viable offspring obtained from Prm1-deficient sperm in mice. *Scientific Reports* **6**, (2016).
16. Pienta, K. J. & Coffey, D. S. A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome. *J Cell Sci Suppl* **1**, 123–135 (1984).
17. Ward, W. S. & Coffey, D. S. DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol Reprod* **44**, 569–574 (1991).
18. Green, G. R., Balhorn, R., Poccia, D. L. & Hecht, N. B. Synthesis and processing of mammalian protamines and transition proteins. *Mol Reprod Dev* **37**, 255–263 (1994).
19. Yelick, P. C. *et al.* Mouse protamine 2 is synthesized as a precursor whereas mouse protamine 1 is not. *Mol Cell Biol* **7**, 2173–2179 (1987).
20. Wykes, S. M. & Krawetz, S. A. The structural organization of sperm chromatin. *J Biol Chem* **278**, 29471–29477 (2003).

21. Corzett, M., Mazrimas, J. & Balhorn, R. Protamine 1: Protamine 2 stoichiometry in the sperm of eutherian mammals. *Molecular Reproduction and Development* **61**, (2002).

22. AOKI, V. *et al.* Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. *Fertility and Sterility* **86**, (2006).

23. de Mateo, S. *et al.* Protamine 2 precursors (Pre-P2), protamine 1 to protamine 2 ratio (P1/P2), and assisted reproduction outcome. *Fertil Steril* **91**, 715–722 (2009).

24. Cho, C. *et al.* Protamine 2 deficiency leads to sperm DNA damage and embryo death in mice. *Biol Reprod* **69**, 211–217 (2003).

25. Ingles, C. J. & Dixon, G. H. Phosphorylation of protamine during spermatogenesis in trout testis. *Proc Natl Acad Sci U S A* **58**, 1011–1018 (1967).

26. Seligman, J., Zipser, Y. & Kosower, N. S. Tyrosine phosphorylation, thiol status, and protein tyrosine phosphatase in rat epididymal spermatozoa. *Biol Reprod* **71**, 1009–1015 (2004).

27. Pruslin, F. H., Imesch, E., Winston, R. & Rodman, T. C. Phosphorylation state of protamines 1 and 2 in human spermatids and spermatozoa. *Gamete Res* **18**, 179–190 (1987).

28. Itoh, K. *et al.* Dephosphorylation of protamine 2 at serine 56 is crucial for murine sperm maturation in vivo. *Sci Signal* **12**, (2019).

29. Pirhonen, A., Linnala-Kankkunen, A. & Menpaa, P. H. P2 protamines are phosphorylated in vitro by protein kinase C, whereas P1 protamines prefer cAMP-dependent protein kinase. A comparative study of five mammalian species. *Eur J Biochem* **223**, 165–169 (1994).

30. Chira, F. *et al.* Phosphorylation of human sperm protamines HP1 and HP2: identification of phosphorylation sites. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1203**, (1993).

31. Brunner, A. M., Nanni, P. & Mansuy, I. M. Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics Chromatin* **7**, 2 (2014).

32. Wu, J. Y. *et al.* Spermiogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. *Nature Genetics* **25**, (2000).

33. Gou, L. T. *et al.* Initiation of Parental Genome Reprogramming in Fertilized Oocyte by Splicing Kinase SRPK1-Catalyzed Protamine Phosphorylation. *Cell* **180**, 1212-1227 e14 (2020).

34. Soler-Ventura, A. *et al.* Characterization of Human Sperm Protamine Proteoforms through a Combination of Top-Down and Bottom-Up Mass Spectrometry Approaches. *J Proteome Res* **19**, 221–237 (2020).

35. Elbashir, S. *et al.* Relationship between sperm progressive motility and DNA integrity in fertile and infertile men. *Middle East Fertility Society Journal* **23**, 195–198 (2018).

36. Rarani, F. Z., Golshan-Iranpour, F. & Dashti, G. R. Correlation between sperm motility and sperm chromatin/DNA damage before and after cryopreservation and the effect of folic acid and nicotinic acid on post-thaw sperm quality in normozoospermic men. *Cell and Tissue Banking* **20**, 367–378 (2019).

37. Piomboni, P., Focarelli, R., Stendardi, A., Ferramosca, A. & Zara, V. The role of mitochondria in energy production for human sperm motility. *International Journal of Andrology* **35**, 109–124 (2012).

38. Simon, L. & Carrell, D. T. Sperm DNA damage measured by comet assay. *Methods in molecular biology (Clifton, N.J.)* **927**, (2013).

39. Castillo, J., Simon, L., de Mateo, S., Lewis, S. & Oliva, R. Protamine/DNA ratios and DNA damage in native and density gradient centrifuged sperm from infertile patients. *Journal of andrology* **32**, 324–32.

40. Simon, L., Castillo, J., Oliva, R. & Lewis, S. E. M. Relationships between human sperm protamines, DNA damage and assisted reproduction outcomes. *Reproductive biomedicine online* **23**, 724–34 (2011).

41. Lewis, S. E. M. & Aitken, R. J. DNA damage to spermatozoa has impacts on fertilization and pregnancy. *Cell and tissue research* **322**, 33–41 (2005).

42. Balhorn, R., Corzett, M. & Mazrimas, J. A. Formation of intraprotamine disulfides in vitro. *Archives of biochemistry and biophysics* **296**, 384–93 (1992).

43. Balhorn, R. A model for the structure of chromatin in mammalian sperm. *Journal of Cell Biology* **93**, 298–305 (1982).

44. Nishi, H., Shaytan, A. & Panchenko, A. R. Physicochemical mechanisms of protein regulation by phosphorylation. *Frontiers in Genetics* **5**, (2014).

45. Nadeau, P. J., Charette, S. J., Toledano, M. B. & Landry, J. Disulfide Bond-mediated multimerization of Ask1 and its reduction by thioredoxin-1 regulate H(2)O(2)-induced c-Jun NH(2)-terminal kinase activation and apoptosis. *Molecular biology of the cell* **18**, 3903–13 (2007).

46. Kim, J. -w., Tang, Q.-Q., Li, X. & Lane, M. D. Effect of phosphorylation and S-S bond-induced dimerization on DNA binding and transcriptional activation by C/EBPbeta. *Proceedings of the National Academy of Sciences* **104**, 1800–1804 (2007).

47. Piomboni, P., Focarelli, R., Stendardi, A., Ferramosca, A. & Zara, V. The role of mitochondria in energy production for human sperm motility. *International journal of andrology* **35**, (2012).

48. Mylonis, I. *et al.* Temporal Association of Protamine 1 with the Inner Nuclear Membrane Protein Lamin B Receptor during Spermiogenesis. *Journal of Biological Chemistry* **279**, (2004).

49. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148 (2016).

50. de Yebra, L. & Oliva, R. Rapid analysis of mammalian sperm nuclear proteins. *Anal Biochem* **209**, 201–203 (1993).

51. Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24**, 1285–1292 (2006).

52. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, (2004).

53. Nakata, H., Wakayama, T., Takai, Y. & Iseki, S. Quantitative analysis of the cellular composition in seminiferous tubules in normal and genetically modified infertile mice. *J Histochem Cytochem* **63**, 99–113 (2015).

Chapter 4 Conclusions and Future Directions

These studies represent the first efforts to conduct an in depth molecular and genetic investigation into the roles of individual P1 residues outside the arginine rich core, which has for decades been presumed to be the functional segment of the P1 protein. Our work demonstrated that substitution of individual residues outside this central arginine core can have profound impacts on sperm chromatin reorganization during spermiogenesis as well as developmental outcomes. Our studies have generated numerous molecular and genetic tools to demonstrate stage-specific establishment of multiple PTMs on P1, as well as alterations in the histone-to-protamine exchange, sperm chromatin composition, sperm morphology/motility, fertility, and dynamics of interactions with DNA caused by substitution of single residues that bear PTMs. Thus, the studies presented in this dissertation provide insight into not only the contribution of individual residues in the P1 protein to reproductive success, but more broadly a complex, informational, and instructive role for protamine proteins and their PTMs in sperm chromatin structure and function.

Importantly, our efforts towards understanding P1 residues/modifications have allowed us to gain essential insights into characteristics and behaviors of protamine proteins generally, which will be invaluable in future studies and may aid in the development of new biomarkers in cases of unexplained infertility–such as detectable changes in protamine PTM levels–leading to more successful reproductive outcomes. The data presented here have also paved the way for numerous new questions regarding protamine protein function both in the male germline and in the early embryo. In this chapter, I describe these outstanding questions and observations made by our work and suggest areas for future study.

Lineage specific functions of protamine residues/modifications

It is interesting to note that modified residues in P1 are lineage specific, except for S9 which is highly conserved across species. Specifically, K49 and T45 are highly conserved across the rodent lineage, but are occupied by alternative residues in Primates, whereas S43 is conserved in a small subset of the rodent lineage and is completely absent in Primates. Together, these observations

suggest that these residues and/or their PTMs may exhibit a lineage specific function in the mouse, such as promoting interaction with P2 and maintenance of the species-specific P1:P2 ratio or controlling protamine removal timing in the early embryo.

The observation that both the P1 K49A and P1 S43E/T45E substitutions result in disrupted P2 processing suggests that C-terminal P1 residues or their PTMs may interact directly with P2. Given that the P1 K49A protein binds equally as well to processed and unprocessed P2, it cannot be ruled out, however, that the lack of P2 processing in this case is simply a consequence of the processing event not being needed for efficient compaction, as is the case with WT P1. It is possible that the mechanism by which P1 and P2 interact (i.e. heterodimers, heterotrimers, etc.) may be influenced by P1 PTMs in a manner that aims to maintain the species-specific P1:P2 ratio. Currently, it is not known how P1 and P2 interact and furthermore, whether P1 and P2 bind together uniformly throughout the genome or if they are programmatically placed in distinct genomic locations. Future studies performing ChIP-seq for P1 and P2 as well as their PTMs in both wild type and mutant P1 sperm will be essential for understanding the baseline genomic localization of P1, P2 and their PTMs and whether these locations are altered upon substitution of individual P1 residues. In addition to providing essential insight into the genomic locations of P1 and P2 in the paternal genome, these studies also have potential implications for understanding their removal in the fertilized zygote.

Removal of protamines from the paternal genome following fertilization, as well as embryonic developmental timing overall are highly species specific.[1,2] Zygotes derived from several heterologous *in vitro* fertilization (IVF) experiments (combining sperm from one species and eggs from a different species) either fail to form pronuclei altogether, form pronuclei at atypical rates, or are unable to decondense the paternal genome while the maternal genome undergoes proper restructuring.[3–6] It is possible, therefore, that the P1:P2 ratio, the genomic locations of P1 and P2, or protamine PTMs may modulate species-specific removal timing of protamines from paternal chromatin and therefore when sperm from one species is injected into a distant and therefore incompatible species, paternal and maternal pronuclear remodeling timing becomes asynchronous, and the embryo fails to develop. This hypothesis is supported by our observation that P1 localizes to the female pronucleus after fertilization, highlighting potential crosstalk between the two

pronuclei that has also been suggested by previous studies.[7] Furthermore, our *in vitro* data from DNA curtains showed that DNA that was initially compacted by P2 decompacts at a significantly slower rate than DNA initially compacted by P1, suggesting that in the zygote, P1 may be removed first followed by P2. Whether species-specific PTMs aid in modulating P1 and P2 removal timing from paternal chromatin will be an exciting area of exploration. It is especially curious that P1 maintains acetylation in sperm, given that when added to histones, acetylation weakens interaction with DNA and promotes chromatin opening.[8,9] This raises the possibility that K49 acetylation may persist in sperm to ultimately allow for accessible remodeling after fertilization. A recent study reported that P1 S9 and S43 phosphorylation are absent in the testis and in sperm and are acquired in the zygote to initiate P1 removal from paternal chromatin, likely by introducing negative charges that weaken the interaction of P1 with DNA.[10] These results conflict with both our findings as well as previously published mass spectrometry data that S9 and S43/T45 phosphorylation are abundant in mature mouse sperm, and therefore additional studies are needed to corroborate this discrepancy.[11]

Our mass spectrometry analysis also identified numerous modifications on P2, many of which were also previously identified in a recent analysis.[11] Notably, the majority of these modifications are present on the N-terminal portion of P2 that is ultimately cleaved off and include phosphorylation at S10, S33, and T43, methylation at R23 and R36, and acetylation at S56, K58, and K65. While conservation of these residues across species is variable, all serine residues we found to be modified are highly conserved across species (**Figure 1-3**). Given that phosphorylation of P2 has been observed and appreciated for decades in several species, cycling phosphorylation/dephosphorylation at these sites likely serves a conserved function in spermiogenesis across species. Additionally, the mouse P2 T43 residue is highly conserved across species and is directly adjacent to the cleavage site, therefore making it possible that phosphorylation at this site controls P2 cleavage.

A longstanding question in the field is what function the cleaved N-terminal portion of P2 serves and further, whether it serves any function after it is removed. A recent study illustrated that loss of the cleaved portion results in abnormal retention of histones and transition proteins and is essential for fertility, highlighting a clear and essential role in histone-to-protamine exchange.[12]

Because full length P2 (pro P2), and not processed P2, is initially translated, shuttled into the cytoplasm, and deposited onto DNA, it is possible that the N-terminal portion, which contains a much broader spectrum of amino acids than C-terminal processed portion, is important for protein-protein interactions, including those with chaperones. Future studies will be needed to investigate the validity of this hypothesis and to further determine whether P2 PTMs play a role in P2 interactions or genome packaging.

The potential role of protamines in the morphological transformation of spermatids

While we do observe a high proportion of sperm with abnormal head morphology in P1$^{K49A/K49A}$ and P1^{S43ET45E} sperm, many of the morphological defects we observe are curiously concentrated outside of the sperm nucleus where chromatin is contained and instead localize to either the midpiece (K49A) or the entire tail (S43E/T45E). This phenotype has previously been reported for mice that lack P1, P2, or TNP1 and TNP2, suggesting a relationship between remodeling of the chromatin landscape and cytoskeletal remodeling/general spermatid reshaping.[13–16] Additionally, our observation that the P1 S43E/T45E substitutions result in decreased mitochondrial content in the sperm midpiece as well as a large proportion of sperm with broken tails also raises the possibility that these substitutions cause disrupted or dysregulated cytoskeletal remodeling or manchette formation. Given that the manchette is essential for both nuclear shaping and protein trafficking for proper sperm tail formation, it is possible that manchette formation is disrupted in P1$^{K49A/K49A}$ and/or P1^{S43ET45E} males. However, an in depth understanding of the mechanisms behind potential manchette disruption will first rely on a basic understanding of if and how protamine incorporation is linked to this process.

An additional possible explanation for the involvement of protamines in spermatid reshaping/cytoskeletal remodeling involves the previously reported finding that phosphorylated P1 interacts with Lamin B receptor (LBR), which resides in the inner nuclear membrane and is a critical component of the nuclear lamina.[17] LBR is also known to interact directly with DNA, free histone proteins, and the CBX heterochromatin proteins, therefore raising the possibility that these interactions physically link chromatin and the nuclear lamina.[18–20] The nuclear lamina then in turn interacts with the cytoskeleton through the linker of nucleoskeleton and cytoskeleton (LINC) complex.[21] Hence, these interactions could provide a direct mechanism to physically link

chromatin to active cytoskeletal remodeling to either provide mechanical force to override the repulsive forces of protamines, increase local concentration of protamines to enhance cooperativity, or potentially to ensure that sperm DNA compacts in a tightly controlled, yet energetically favorable manner. Alternatively, given that P1 appears in the nucleus around the same time as TNPs, it is possible that tethering P1 to the nuclear lamina prevents it from binding to DNA before TNPs.

Identification of chromatin remodeling factors for the histone-to-protamine exchange

Sperm chromatin remodeling during spermiogenesis is essential for reproductive success, yet we lack a complete understanding of the proteins and machinery involved, largely due to a lack of both *in vitro* and *in vivo* tools. As described in Chapter 1, several candidate remodeling factors have been proposed to function during this process, but conclusive evidence for such a role was precluded by confounding upstream functions for these proteins in spermatogenesis. Furthermore, whether the same set of protein factors are responsible for the transition from histones to transition proteins as are responsible for the transition from transition proteins to protamines also remains unclear. Two recent studies performed an in-depth assessment of sperm chromatin remodeling upon loss of CHD5, and interestingly many of the observed phenotypes overlap with those we observed in P1$^{K49A/K49A}$ males, including abnormal retention of histones, an accumulation of pro P2, and decreased sperm motility. Together, these results strongly support a role for CHD5 in sperm chromatin remodeling. However, broader identification of remodeling factors and chaperones utilizing targeted genetic knockouts without *a priori* knowledge of candidate factors likely will be cost and time inefficient and a more unbiased approach is needed to first identify candidate factors.

Because high quality, immunoprecipitation grade antibodies against mouse P1 and P2 are not commercially available, our lab has generated transgenic mice containing epitope tagged P1 and P2 to identify candidate interacting factors. Using elongating spermatids from the testes of these mice, we performed immunoprecipitation followed by mass spectrometry (IP-MS) to gain an unbiased perspective of protamine-interacting proteins that potentially may play a role in chromatin remodeling/histone-to-protamine exchange. These preliminary experiments identified several known chromatin remodeling factors including BRDT, SNF2H, and BAZ1A, among

others. Direct interaction with BRDT is especially intriguing given that BRDT binds directly to acetylated histone H4 to initiate remodeling and in its absence, proper histone eviction and protamine deposition cannot take place.[22] Future studies will elucidate whether BRDT may bind simultaneously to acetylated H4 and acetylated P1 to modulate remodeling.

The identification of SNF2H (an ISWI homolog, also known as SMARCA5) and BAZ1A (bromodomain adjacent to zinc finger domain 1A, also known as ACF1) is additionally interesting given that these two proteins function in a complex together to remodel chromatin.[23,24] While the function of either protein alone or in combination during spermiogenesis is not fully understood, one study has generated a BAZ1A[-/-] mouse and while profound defects in sperm development and fertility were observed, these alterations are caused by perturbed gene expression in meiotic spermatocytes and round spermatids, ultimately precluding the ability to investigate potential functions in the histone-to-protamine exchange.[25] Additionally, while several studies have investigated the remodeling function of SNF2H in other tissues, limited studies to date have focused on its role in the testis, except one study that reported that SNF2H forms a complex with the bromodomain containing protein CECR2, and loss of CECR2 results in decreased fertility, abnormal sperm morphology, and decreased sperm motility–suggesting a potential role in chromatin remodeling during spermiogenesis.[26]

The data gained from these experiments will be invaluable not only for identifying protamine-interacting proteins, but more broadly for understanding factors involved and mechanisms underlying chromatin remodeling and the histone-to-protamine exchange. The generation of knockout mouse models of the candidate factors described above will undoubtedly inform our understanding of chromatin reorganization during spermiogenesis, which to date has been a significant bottleneck in understanding how disruption of this process contributes to cases of infertility.

Structure determination of P1 and P1 mutants

In addition to a fundamental lack of knowledge regarding mechanisms underlying chromatin remodeling during spermiogenesis, the physical structures of P1 and P2 either alone or bound to DNA are unknown. Attempts at structural determination of protamines by X-ray crystallography

have all been unsuccessful most likely due to the predicted high level of inherent disorder of protamines due to their high arginine content. *In vitro* studies have instead focused on understanding the higher order structure of sperm chromatin imposed by protamines, which is thought to be a toroid containing ~50 kb of DNA.[27–30] In this model of protamine-based packaging, protamines likely induce bending of the DNA into a more favorable conformation before progressively condensing the DNA through several intermediate states.[31] Further *in vitro* analysis using salmon sperm protamine revealed that toroids induced by protamine-DNA binding are inherently stable, although stability of the toroid is directly impacted by arginine content.[32] While these studies provide a foundational framework for the general structure of sperm chromatin, the use of more modern approaches for structure determination of these toroids, such as Cryo electron microscopy (cryo EM) will provide a much higher resolution picture of these structures. Furthermore, use of cryo EM could be applied in the context of toroids formed by P1 K49A, P1 S43A/T45A, and P1 S43E/T45E which would enable our understanding of how higher order chromatin structure is perturbed by these substitutions.

Because of the inherent disorder of protamines alone, it is unlikely that any X-ray crystallographic conditions will allow us to solve the crystal structure of protamines, however it is possible that binding to DNA confers additional order/structure to the protein that may permit this type of analysis.[33] Additionally, alternative techniques such as nuclear magnetic resonance (NMR) or circular dichroism may represent feasible methods for structure determination of protamines or at least provide some level of insight into structural characteristics.[34,35]

The role of protamines in the early embryo

One of the most striking and unexpected results of our analysis was the localization of P1 in the maternal pronucleus following fertilization. While unexpected, we are not the first to observe this phenomenon, as protamines have previously been shown to localize to the maternal pronucleus, but this localization was assumed to be an antibody artifact and its significance largely dismissed.[36] We further confirmed that this pattern is not an artifact of our antibody, as we knocked in a V5 tag to the endogenous P1 locus to create V5 tagged P1 and further observed V5-P1 in both pronuclei (**Figure 2-8C**). Together, these results raise the question of whether the presence of P1 in the maternal pronucleus serves a biological function in the embryo and if so, what this function is.

Interestingly, we observe that P1 appears to be removed from DNA and subsequently accumulates in the prenucleolar body (**Figure 2-7D**), a pattern we have additionally observed with other antibodies against protamines and protamine PTMs (data not shown). Due to the high arginine content of protamines, it is possible that their movement into the prenucleolar body provides a mechanism for their destruction, as highly positively charged proteins are known to accumulate in the nucleoli of somatic cells.[37] Therefore, the maternal pronucleus and maternal prenucleolar body may represent an additional location for protamines to accumulate before they are degraded. Additional analysis quantifying the overlap of P1 on chromatin vs. in the prenucleolar body in the maternal pronucleus of P1$^{+/+}$ and P1$^{K49A/K49A}$ zygotes will also be informative for understanding its specific location and potential alterations caused by the K49A substitution.

Alternatively, it is possible that trafficking of protamines from the paternal pronucleus to the maternal pronucleus serves as a level of crosstalk between the two pronuclei to allow for coordination of the two genomes until the first mitotic division. Several studies have observed crosstalk between the maternal and paternal pronuclei of both mouse and rat zygotes, largely in response to DNA damage in the injected sperm. Specifically, one study found that the maternal pronucleus of the mouse zygote responds to irradiation-induced DNA damage in sperm by activating the p53 DNA damage checkpoint and significantly decreasing DNA synthesis output.[7] Interestingly, the embryos derived from irradiated mouse sperm did not die immediately, but rather much later in development (after implantation), suggesting that marking of the pronuclei with p53 marks the embryos for delayed death. Several additional studies have shown that paternal exposure to the genotoxicant cyclophosphamide (a potent anticancer drug) results in the accumulation of PARP1 (involved in late DNA damage response) in the maternal pronucleus as well as both hyperacetylation of H4 and hypomethylation of DNA, indicative of potential destabilization of the maternal pronucleus.[38,39] Altogether, these studies strongly support the presence of crosstalk between the paternal and maternal genomes, likely serving to ensure that the genomes remain coordinated while they are physically separated in the one cell zygote.

In order to understand the functional significance of protamine localization to the maternal pronucleus, it will be essential to (1) follow protamine movement in real time using live cell

116

imaging and (2) investigate remodeling of the two genomes and developmental outcomes in the absence of protamines, which can be accomplished via targeted degradation of protamines using Trim-Away.[40,41] The ability to follow protamine movement in real time using live cell imaging will provide invaluable information regarding both the kinetics of protamine removal from paternal chromatin as well as infiltration of protamines into the maternal pronucleus. Our current mouse models will not accommodate live cell imaging as conventional antibodies against epitope tags are large, severely limiting resolution.[42] Anecdotally, we have also found that signal intensity is insufficient using commercially available single domain nanobodies against the specific epitope tags we employed. Therefore, possibly the generation of a fluorescently labeled nanobody against P1 may represent a viable method for live cell imaging. Alternatively, several recent reports have illustrated the utility of short peptide tags for live cell imaging, against which nanobodies have been generated (SPOT tag and ALFA tag). The incorporation of these short peptide tags therefore may be used as well.[43]

In addition to live cell imaging to monitor the kinetics of protamine movement in the zygote, the use of Trim-Away to degrade protamines will also be essential for understanding the biological function of protamine localization to the maternal pronucleus. Trim-Away is based on the ability of the ubiquitin ligase TRIM21 to specifically recognize antibody-bound proteins and targets them for degradation. Previous optimization experiments by others have shown that endogenous levels of TRIM21 in oocytes/embryos is not sufficient to drive protein degradation; therefore, overexpression of TRIM21 via microinjection of *Trim21* mRNA into oocytes is necessary to achieve complete depletion.[41] In the context of protamine depletion, oocytes could be injected with both *Trim21* mRNA and an antibody against P1, subsequently fertilized with sperm using either conventional IVF or ICSI, and downstream events such as pronuclear formation, pronuclear remodeling, re-establishment of epigenetic landscapes, and overall development can be monitored.

Protamine PTMs in human sperm and implications in infertility
A logical extension of our findings of mouse P1 residues and PTMs would be to consider whether a similar catalog of modifications exists on human protamines, and whether alterations in their levels may correlate with infertility. Mass spectrometry analysis from our lab and others has in fact identified several PTMs on both P1 and P2 and interestingly, it appears that levels of PTMs

on P2 vary by isoform (whereas mouse contains a processed and an unprocessed form of P2, humans possess multiple cleavage products of P2).[44] Whether higher or lower levels of specific PTMs may have an association with male factor infertility generally, or specific subclasses of sperm defects (based on our analysis in the mouse, possibly asthenospermia or low sperm motility could be associated with altered protamine PTM levels) remains to be determined.

Concluding remarks

In summary, much remains to be understood about the dramatic protamine-induced remodeling of the chromatin landscape that occurs during the final stage of spermatogenesis and how this mode of packaging is reversed in the early embryo. Our work has overturned the previously held assumption that protamines are passive, static structural proteins serving a sole function of non-specifically packaging the genome and instead, we propose a novel model in which protamine sequences and/or protamine PTMs outside of the central arginine core have evolved to confer specific structural characteristics and/or carry out species-specific and regulated sperm genome compaction and unpackaging in the embryo. The work presented here therefore provides a novel layer of knowledge regarding how protamine-based sperm genome packaging is fine tuned to safeguard male reproductive fitness and may inform our understanding of protamine-related cases of infertility outside of altered P1:P2 ratio.

References

1. Lequarre, A. S., Marchandise, J., Moreau, B., Massip, A. & Donnay, I. Cell Cycle Duration at the Time of Maternal Zygotic Transition for In Vitro Produced Bovine Embryos: Effect of Oxygen Tension and Transcription Inhibition1. *Biology of Reproduction* **69**, (2003).
2. Ciemerych, M. A. & Sicinski, P. Cell cycle in mouse development. *Oncogene* **24**, (2005).
3. Naish, S. J., Perreault, S. D. & Zirkin, B. R. DNA synthesis following microinjection of heterologous sperm and somatic cell nuclei into hamster oocytes. *Gamete research* **18**, (1987).
4. Hanada, A. & Chang, M. C. Penetration of zone-free eggs by spermatozoa of different species. *Biology of reproduction* **6**, (1972).
5. Perreault, S. D., Barbee, R. R., Elstein, K. H., Zucker, R. M. & Keefer, C. L. Interspecies differences in the stability of mammalian sperm nuclei assessed in vivo by sperm microinjection and in vitro by flow cytometry. *Biology of reproduction* **39**, (1988).
6. Barnetova, I., Fulka, H. & Fulka, J. Epigenetic characteristics of paternal chromatin in interspecies zygotes. *The Journal of reproduction and development* **56**, (2010).
7. Shimura, T. *et al.* p53-dependent S-phase damage checkpoint and pronuclear cross talk in mouse zygotes with X-irradiated sperm. *Mol Cell Biol* **22**, 2220–2228 (2002).
8. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, (1997).
9. Eberharter, A. & Becker, P. B. Histone acetylation: a switch between repressive and permissive chromatin. *EMBO reports* **3**, (2002).
10. Gou, L. T. *et al.* Initiation of Parental Genome Reprogramming in Fertilized Oocyte by Splicing Kinase SRPK1-Catalyzed Protamine Phosphorylation. *Cell* **180**, 1212-1227 e14 (2020).
11. Brunner, A. M., Nanni, P. & Mansuy, I. M. Epigenetic marking of sperm by post-translational modification of histones and protamines. *Epigenetics Chromatin* **7**, 2 (2014).
12. Lena Arvola *et al.* Loss of the cleaved-protamine 2 domain leads to incomplete histone-to-protamine exchange and infertility in mice. *BioRxiv* (2021).
13. Takeda, N. *et al.* Viable offspring obtained from Prm1-deficient sperm in mice. *Scientific Reports* **6**, (2016).
14. Schneider, S. *et al.* Re-visiting the Protamine-2 locus: deletion, but not haploinsufficiency, renders male mice infertile. *Sci Rep* **6**, 36764 (2016).
15. Yu, Y. E. *et al.* Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proceedings of the National Academy of Sciences* **97**, (2000).
16. Shirley, C. R., Hayashi, S., Mounsey, S., Yanagimachi, R. & Meistrich, M. L. Abnormalities and Reduced Reproductive Potential of Sperm from Tnp1- and Tnp2-Null Double Mutant Mice1. *Biology of Reproduction* **71**, (2004).
17. Mylonis, I. *et al.* Temporal Association of Protamine 1 with the Inner Nuclear Membrane Protein Lamin B Receptor during Spermiogenesis. *Journal of Biological Chemistry* **279**, (2004).
18. Makatsori, D. *et al.* The Inner Nuclear Membrane Protein Lamin B Receptor Forms Distinct Microdomains and Links Epigenetically Marked Chromatin to the Nuclear Envelope. *Journal of Biological Chemistry* **279**, 25567–25573 (2004).

19. Courvalin, J. C., Segil, N., Blobel, G. & Worman, H. J. The lamin B receptor of the inner nuclear membrane undergoes mitosis-specific phosphorylation and is a substrate for p34cdc2-type protein kinase. *The Journal of biological chemistry* **267**, 19035–8 (1992).

20. Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J. C. & Worman, H. J. Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. *The Journal of biological chemistry* **272**, 14983–9 (1997).

21. Gob, E., Schmitt, J., Benavente, R. & Alsheimer, M. Mammalian sperm head formation involves different polarization of two novel LINC complexes. *PLoS One* **5**, e12072 (2010).

22. Gaucher, J. *et al.* Bromodomain-dependent stage-specific male genome programming by Brdt. *The EMBO Journal* **31**, (2012).

23. Collins, N. *et al.* An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nature genetics* **32**, (2002).

24. He, X., Fan, H.-Y., Narlikar, G. J. & Kingston, R. E. Human ACF1 alters the remodeling strategy of SNF2h. *The Journal of biological chemistry* **281**, (2006).

25. Dowdle, J. A. *et al.* Mouse BAZ1A (ACF1) is dispensable for double-strand break repair but is essential for averting improper gene expression during spermatogenesis. *PLoS genetics* **9**, (2013).

26. Thompson, P. J., Norton, K. A., Niri, F. H., Dawe, C. E. & McDermid, H. E. CECR2 is involved in spermatogenesis and forms a complex with SNF2H in the testis. *Journal of molecular biology* **415**, (2012).

27. Hud, N. V., Allen, M. J., Downing, K. H., Lee, J. & Balhorn, R. Identification of the Elemental Packing Unit of DNA in Mammalian Sperm Cells by Atomic Force Microscopy. *Biochemical and Biophysical Research Communications* **193**, (1993).

28. Widom, J. & Baldwin, R. L. Cation-induced toroidal condensation of DNA. *Journal of Molecular Biology* **144**, (1980).

29. Watanabe, F. & Schwarz, G. Thermodynamics and kinetics of co-operative protein-nucleic acid binding. *Journal of Molecular Biology* **163**, (1983).

30. Allen, M. AFM analysis of DNA-protamine complexes bound to mica. *Nucleic Acids Research* **25**, (1997).

31. Ukogu, O. A. *et al.* Protamine loops DNA in multiple steps. *Nucleic Acids Research* **48**, (2020).

32. Balhorn, R., Brewer, L. & Corzett, M. DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. *Molecular reproduction and development* **56**, (2000).

33. Hud, N. v, Milanovich, F. P. & Balhorn, R. Evidence of novel secondary structure in DNA-bound protamine is revealed by Raman spectroscopy. *Biochemistry* **33**, 7528–7535 (1994).

34. Purslow, J. A., Khatiwada, B., Bayro, M. J. & Venditti, V. NMR Methods for Structural Characterization of Protein-Protein Complexes. *Frontiers in Molecular Biosciences* **7**, (2020).

35. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **1**, 2876–90 (2006).

36. McLay, D. W. & Clarke, H. J. Remodelling the paternal chromatin at fertilization in mammals. *Reproduction* **125**, 625–633 (2003).

37. Martin, R. M. *et al.* Principles of protein targeting to the nucleolus. *Nucleus* **6**, 314–325 (2015).

38. Barton, T. S., Robaire, B. & Hales, B. F. DNA damage recognition in the rat zygote following chronic paternal cyclophosphamide exposure. *Toxicological sciences : an official journal of the Society of Toxicology* **100**, 495–503 (2007).

39. Barton, T. S., Robaire, B. & Hales, B. F. Epigenetic programming in the preimplantation rat embryo is disrupted by chronic paternal cyclophosphamide exposure. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7865–70 (2005).

40. Xenopoulos, P., Nowotschin, S. & Hadjantonakis, A.-K. Live imaging fluorescent proteins in early mouse embryos. *Methods in enzymology* **506**, 361–89 (2012).

41. Clift, D., So, C., McEwan, W. A., James, L. C. & Schuh, M. Acute and rapid degradation of endogenous proteins by Trim-Away. *Nature Protocols* **13**, 2149–2175 (2018).

42. de Beer, M. A. & Giepmans, B. N. G. Nanobody-Based Probes for Subcellular Protein Identification and Visualization. *Frontiers in cellular neuroscience* **14**, 573278 (2020).

43. Götzke, H. *et al.* The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nature Communications* **10**, 4403 (2019).

44. Soler-Ventura, A. *et al.* Characterization of Human Sperm Protamine Proteoforms through a Combination of Top-Down and Bottom-Up Mass Spectrometry Approaches. *J Proteome Res* **19**, 221–237 (2020).

Appendix: TCF21[+] Mesenchymal Cells Contribute to Testis Somatic Cell Development, Homeostasis, and Regeneration in Mice[3]

A.1 Abstract

Testicular development and function rely on interactions between somatic cells and the germline, but similar to other organs, regenerative capacity declines in aging and disease. Whether the adult testis maintains a reserve progenitor population remains uncertain. Here, we characterize a recently identified mouse testis interstitial population expressing the transcription factor Tcf21. We found that TCF21[lin] cells are bipotential somatic progenitors present in fetal testis and ovary, maintain adult testis homeostasis during aging, and act as potential reserve somatic progenitors following injury. In vitro, TCF21[lin] cells are multipotent mesenchymal progenitors which form multiple somatic lineages including Leydig and myoid cells. Additionally, Tcf21[+] cells resemble resident fibroblast populations reported in other organs having roles in tissue homeostasis, fibrosis, and regeneration. Our findings reveal that the testis, like other organs, maintains multipotent mesenchymal progenitors that can be potentially leveraged in development of future therapies for hypoandrogenism and/or infertility.

A.2 Introduction

Sexual reproduction relies on the generation of distinct sexes to increase biological diversity. The core of this strategy rests with a bipotential gonadal primordium that supports development of sex-specific reproductive organs with functionally distinct gonadal cell types. The gonadal primordium is comprised of primordial germ cells and mesenchymal cells that originate from two sources: the coelomic epithelium[1–5] and the mesonephros[6]. In early development, the coelomic epithelial cells give rise to multiple cell lineages, including interstitial cells and Sertoli cells[1], but later become restricted to the interstitial compartment of the testis. In contrast, mesonephric-derived cells migrating into the gonad contribute only to fetal/adult Leydig and interstitial cells, but not Sertoli cells[7,8]. Hence, the gonadal mesenchyme is heterogeneous on the molecular and cellular scales, and certain somatic lineages (e.g. Leydig and myoid) have multiple cells of origin[9], pointing to a complex developmental programming of reproductive organs (reviewed in[10–12]).

Establishment of a functional somatic microenvironment is essential for continuous sperm production, germ cell homeostasis, and regeneration[13–15]. Disruptions in somatic cell populations can dramatically alter germ cell development and testis function. For instance, genetic ablation of macrophages in the adult testis leads to disruption of spermatogonial proliferation and differentiation[16]. Others have shown that testicular endothelial cells[14] as well as lymphatic endothelial cells[15] support human and mouse spermatogonial stem cell (SSC) survival and expansion, and can also modulate spermatogonial cell homeostasis and regeneration[15]. Therefore, these studies underscore the importance of defining the interstitial cell composition and function.

The testis interstitial cells are believed to be postmitotic (not actively proliferating). However, studies in rats have demonstrated that adult Leydig cells can regenerate after ethane dimethane sulfonate (EDS)-induced cell death (reviewed in[17]). Multiple interstitial cell (CD90/PDGFRA/COUPTFII/NESTIN) populations were shown to re-enter the cell cycle upon EDS-induced Leydig cell death, suggesting that the interstitial compartment contains a reserve Leydig or a general somatic cell progenitor population that can be activated in response to damage[18–26]. Without single cell RNA-seq analysis it remains difficult to tease apart whether these markers are observed in a single homogenous population of Leydig stem cells or if these cells are

a heterogenous pool of progenitors. Furthermore, it is also unclear if a common somatic progenitor could give rise to additional cell types other than Leydig, and what the role of such stem/progenitors would be in the normal adult testis.

Using single cell RNA sequencing (scRNA-seq) we previously identified a mesenchymal cell population in the adult mouse testis that expresses the transcription factor *Tcf21* and appears *in vivo* as rare spindle-shaped cells surrounding the seminiferous tubule[13]. The *Tcf21* gene has known roles in the development of multiple organs, including the testis[27–29]. Loss of *Tcf21* promotes feminization of external genitalia in karyotypically male mice[30], while overexpression of *Tcf21* in primary embryonic ovary cells leads to *in vitro* sex-reversal via aberrant anti-Mullerian hormone expression[31]. These results provide evidence for a role of TCF21in male sex determination and testis somatic cell differentiation. Recent reports of single-cell sequencing during sex determination and cell linage specification also identified *Tcf21* expression among subsets of gonadal somatic cells in both male and female, although these experiments were limited to NR5A1-eGFP cells[32,33]. The similarity between the adult *Tcf21*[+]population and fetal somatic progenitors[32,33] suggests that the *Tcf21*[+] population that persists in adulthood may retain fetal developmental or functional properties.

In this study, we examine the role of *TCF21*[+] cells in the developing and adult mouse testis. To answer this question, we utilize genetic lineage tracing to mark and follow the potential and fate of the TCF21[lin] population both *in vitro* and *in vivo*. First, in directed *in vitro* differentiation paradigms, we find that flow-sorted TCF21[lin] cells possess mesenchymal stem cell (MSC)-like properties and can be directed to differentiate to either myoid or Leydig cell lineages, thus acting as true multipotent progenitors *in vitro*. *In vivo,* fetal TCF21[lin] cells are bipotential somatic progenitors, contributing to all known somatic cell populations in the fetal and adult testis and ovary. In the adult testis, the TCF21[lin] cells replenish somatic populations in response to injury as well as in normal aging. Furthermore, the adult testis *Tcf21*[+] cells resemble resident fibroblast populations in multiple organs which have been implicated in tissue homeostasis, fibrosis, and regeneration. In summary, our work demonstrates the first evidence for a reserve somatic cell population in the adult testis, representing a potential targetable cell population for development of treatments for gonad-related defects and disease.

A.3 Results

A.3.1 The *Tcf21*[+]population is a molecularly heterogenous mesenchymal cell population that is transcriptomically similar to myoid and Leydig cells

We recently employed single-cell RNA-seq (scRNA-seq) to generate a cell atlas of the mouse testis[13,34]. Our analysis of ~35,000 cells identified all known somatic cell types as well as an unexpected *Tcf21*[+] population (**Figure A1-A**)[13]. To better understand its potential function, we first examined if the *Tcf21*[+] population has molecular similarity to other known somatic cells in the testis. By using a pairwise dissimilarity matrix for somatic cell centroids, we find that the *Tcf21*[+] population was distinct from macrophage and Sertoli lineages but transcriptomically similar to myoid, Leydig, and endothelial cells (**Figure A-1B**). In contrast with these cell types, *Tcf21*[+] cells did not express any of their terminally differentiated markers (**Figure A-1C**). Rather, this population was uniquely demarcated by the expression of *Tcf21*, *Pdgfra*, *CoupTFII*, and mesenchymal progenitor cell (*MP*) markers including *Sca1* (**Figure A-1C**), *Arx*, and *Vim*[13]. To define molecular properties of *Tcf21*[+] cells more broadly we identified differentially expressed genes between the *Tcf21*[+] population and its most similar cell types (using >2-fold change and FDR <5%). Gene ontology analysis suggested that the *Tcf21*[+] population is of mesenchymal origin, and likely involved in extracellular matrix (ECM) biology, tissue injury and repair processes (**Figure A-1D**). Although the ECM was once considered a passive support scaffold, a wealth of data now suggests an active role for ECM in many aspects of biology, from tissue maintenance, regeneration, cell differentiation, to fibrosis and cancer[35,36].

Given the identification of multiple mesenchymal progenitor markers in the *Tcf21*[+] population, we next sought to validate expression of these markers *in vivo*. We took advantage of previously generated *Tcf21*[mCrem] mice, in which a tamoxifen inducible Cre recombinase inserted at the *Tcf21* locus enables the long-term tracing of the descendant populations of the *Tcf21*-expressing cells, whether in the gonads, heart, kidney and cranial muscle[37]. Specifically, testes were collected from *Tcf21*[mCrem]:*R26R*[tdTom] mice after 3 doses of tamoxifen (tdTom[+] labeled cells referred to as TCF21[lin]), dissociated, stained for a comprehensive panel of mouse MSC markers (SCA1, CD73, CD29, CD34 and THY1) while excluding mature Leydig (cKIT[+]) or immune (CD45[+]) cells, and analyzed using flow cytometry (**Figure A-2A-F**). Since our *Tcf21*[+] cells were initially discovered

by enriching SCA1$^+$ cells in the testis, we examined the heterogeneity of the SCA1$^+$ and TCF21^{lin} cells. As expected, the SCA1$^+$ population has a broader representation in the testis than TCF21^{lin} (approximately ~3-5% vs. ~1-2%), with about 45% of SCA1$^+$cells being also positive for TCF21^{lin} (**Figure A-2-G**) – a proportion consistent with our estimates from scRNA-seq data. Within the SCA1$^+$ population, we identified two TCF21^{lin} subpopulations (blue and purple in **Figure A-2B-C**) based on co-expression of mesenchymal stem cell markers: a larger TCF21^{lin} population (13.8%, blue) that strongly expressed CD29 (ITGB1), CD73, and CD34, and a smaller discrete TCF21^{lin} population (0.77%, purple) that expressed all MSC markers. In contrast, 80% of the TCF21^{lin} population expresses SCA1 (**Figure A-2G**), and within the TCF21^{lin} populations we identify multiple subtypes that are molecularly heterogenous with respect to MSC marker expression (**Figure A-2D-F**) (see below for the functional assessment of TCF21^{lin} heterogeneity *in vitro*).

Given the heterogeneity observed within the TCF21^{lin} population, we asked whether TCF21^{lin} subtypes could be further defined by co-expression of previously described interstitial markers including PDGFRA, COUPTFII, CD34, and FGF5, respectively. To answer this question, we co-stained testes from *Tcf21^{mCrem}:R26R^{tdTom}* or *Tcf21^{mCrem}:R26R^{tdTom}:PDGFRA^{GFP}* mice with COUPTFII, CD34, and FGF5 antibodies (**Figure A-2H**), but we did not detect a preferred segregation of TCF21^{lin} cells within the PDGRA, COUPTFII, CD34 or FGF5 subpopulation (**Figure A-2H**), suggesting that these populations are heterogeneous on both the cellular and molecular level. Furthermore, for all markers analyzed we find co-stained cells both in the interstitium or surrounding tubules – suggesting that the cellular heterogeneity is not a result of spatial location.

A.3.2 The TCF21^{lin}/SCA1$^+$population has mesenchymal progenitor properties *in vitro* and can be differentiated to Leydig and myoid cell fates *in vitro*.

Mesenchymal progenitors are typically characterized by their capacity for forming adherent fibroblast-like colonies on plastic (measured as CFU-F: fibroblastic colony forming units) and for differentiating into adipocytes, osteocytes, and chondrocytes *in vitro* (Reviewed in [38]). Such populations have been isolated from multiple human and mouse organs, including juvenile or adult testes[18,39–41].To examine whether the adult SCA1$^+$ and TCF21^{lin} cells have mesenchymal

progenitor-like properties *in vitro*, we sorted four types of cells: SCA1$^-$/cKIT$^+$ interstitial cells (control) , SCA1$^+$/cKIT$^-$, SCA1$^+$/TCF21^{lin}, or TCF21^{lin} cells from *Tcf21^{mCrem}:R26R^{tdTom}* animals (**Figure A-4A)** and plated these cells at single cell density to measure clonogenic potential *in vitro*. Although SCA1$^+$/cKIT$^-$, SCA1$^+$/TCF21^{lin}, and TCF21^{lin}populations all formed colonies *in vitro,* clonogenic potential differed across populations. The TCF21^{lin} and SCA1$^+$/TCF21^{lin} double positive cells formed the highest number of colonies - ~100 colonies per 1000 plated cells, and this was followed by SCA1$^+$/cKIT$^-$cells (regardless of the TCF21^{lin} status) and cKIT$^+$ cells with ~40 and 10 colonies, respectively (**Figure A-4B**). Altogether, these data demonstrate that SCA1$^+$and SCA1$^+$/TCF21^{lin} populations have characteristics of mesenchymal progenitors *in vitro*, but selecting for the TCF21^{lin} population significantly increases colony formation potential.

In addition to expanding in culture, mesenchymal progenitors, under appropriate conditions, can be directed to differentiate into adipocytes, chondrocytes and osteoblasts. To examine if TCF21^{lin} cells maintain such properties *in vitro,* we sorted either SCA1$^+$/cKIT$^-$ or cKIT$^+$ cells from *Tcf21^{mCrem}:R26R^{tdTom}* animals and examined whether the TCF21^{lin} within the SCA1$^+$ population contribute to all three lineages (**Figure A-4A)**. Importantly, we found that SCA1$^+$cells robustly differentiated into adipocytes, chondrocytes, and osteocytes as shown by Oil red, Alizarin red, and Alcian blue staining, respectively (**Figure A-4C**). Furthermore, co-immunofluorescence of Osterix (osteocytes), Perilipin (adipocytes), and SOX9 (chondrocytes) with tdTom$^+$ (TCF21^{lin}) cells demonstrates that the TCF21^{lin} cells within the SCA1$^+$ population can contribute to all three lineages *in vitro* (**Figure A-4D**).

Given the ability of the SCA1$^+$/TCF21^{lin} to generate multiple mesenchymal cell types *in vitro* and the transcriptomic relationship of *Tcf21$^+$* cells with both Leydig and myoid cells (**Figure A-1B**), we next asked whether the TCF21^{lin}/SCA1$^+$ population can be directed to differentiate to both Leydig and myoid cells *in vitro* and if the TCF21^{lin}/SCA1$^+$population serves as a multipotent progenitor for Leydig and myoid lineages. To this end, we sorted SCA1$^+$/cKIT$^-$ (regardless of TCF21 status) or TCF21^{lin}/SCA1$^+$/cKIT$^-$cellsfrom*Tcf21^{mCrem}:R26R^{tdTom}* adult male testes and directed their differentiation to either myoid or Leydig cells using a set of growth factors based on the repertoire of receptors expressed in our scRNA-seq datasets and earlier *in vivo* genetic findings of Leydig or myoid cell specification (**Figure A-2E-K**)[42–44]. After treating the bulk SCA1$^+$or SCA1$^+$/TCF21^{lin} cells with a myoid differentiation cocktail which includes Smoothened agonist

(SAG, an activator of Desert Hedgehog), PDGFAA, PDGFBB, ACTIVINA, BMP2, BMP4, and Valproic Acid (outlined in **Figure A-4E**), we observed a morphological conversion of spindle-shaped cells to flattened and striated cells resembling smooth muscle cells (**Figure A-4F**). This conversion was confirmed by expression of smooth muscle cell markers such as smooth muscle actin (**Figure A-4G, H**).

Previous *in vivo* and *in vitro* experiments uncovered that Desert Hedgehog (DHH), FGF, and PDGF signaling are stimulatory to Leydig cell differentiation, while Notch signaling and other factors are inhibitory[42–44]. By incorporating these findings from the literature with our scRNA-seq data, we developed a 14-day differentiation protocol that includes PDGFAA, PDGFBB, SAG, FGF2, $LiCl_2$ and DAPT (Notch inhibitor) (detailed in **Figure A-4I**). This protocol enables successful differentiation of SCA1$^+$/cKIT$^-$ or SCA1$^+$/TCF21^{lin} cells to Leydig cells (**Figure A-4J-K**). Importantly, the *in vitro*-derived Leydig cells expressed steroidogenic factor 1 (SF1) (**Figure A-4K, L**) and secreted testosterone (**Figure A-4M, N**). Notably, Leydig cell generation and testosterone secretion was achieved *in vitro* independent of LH control (**Figure A-4M, N**). Consistent with absence of LH regulation, we detected only low levels of luteinizing hormone/choriogonadotropin receptor (*Lhcgr*) transcripts in day 14 Leydig cells (**Table A-2**).

As the SCA1/TCF21^{lin} cells were labeled as a population, the results described above could not distinguish the two alternative scenarios: (1) each TCF21^{lin}/SCA1$^+$ cell is a multipotent progenitor, capable of adopting any of multiple fates, or (2) these cells are heterogeneous, comprised of multiple subtypes of progenitors that each have been cryptically committed to differentiate into a different somatic lineage. To distinguish between these two scenarios, we sorted individual TCF21^{lin}/SCA1$^+$ cells into separate wells on 96-well plates and allowed each to expand into individual clones. Following clonal expansion, clones were randomly assigned to either the myoid or Leydig cell differentiation protocol (**Figure A-3A, B**). Therefore, if these cells have already been primed for one or the other lineage, we would expect that some clones, but not all, that undergo the myoid-inducing treatment will differentiate to the myoid lineage, and likewise some clones in the Leydig treatment group will follow the Leydig lineage. Co-immunostaining with either SMA for myoid cells or SF1 for Leydig cells reveals that all TCF21^{lin}/SCA1$^+$ clones in either group differentiate to adopt the induced fate 100% of the time (number of clones provided in

Figure A-3E), suggesting that the TCF21^lin^/SCA1^+^cells are individually multipotent *in vitro* (**Figure A-3C-E**). However, since the TCF21^lin^/SCA1^+^ population can be further stratified by the expression of CD105 *in vivo,* we examined whether TCF21^lin^/SCA1^+^ multipotency may be restricted to a specific subset of TCF21^lin^/SCA1^+^ cells. To address this, we sorted and differentiated TCF21^lin^/SCA1^+^/CD105^+^ or TCF21^lin^/SCA1^+^/CD105^-^cells to Leydig or myoid cells. Consistent with our earlier finding using TCF21^lin^/SCA1^+^ cells, both the TCF21^lin^/SCA1^+^/CD105^+^ and the TCF21^lin^/SCA1^+^/CD105^-^ clonal cells differentiated to both lineages with 100% efficiency (**Figure A-3C-E**). These observations suggest that multipotency is not due to heterogeneous subpopulations of TCF21^lin^/SCA1^+^ cells identified by flow cytometry (**Figure A-2E-F**). We caution, however, since these clonal cell experiments were performed *in vitro*, the multipotency of these cells *in vivo* remains to be confirmed in future studies.

A.3.3 Single-cell time-course analysis reveals timing and diverging trajectories during Leydig cell differentiation.

We next characterized the Leydig cell differentiation process *in vitro* using scRNA-seq. To this end, we collected and analyzed ~6500 cells across four time-points along the *in vitro* differentiation process (d0 sorted SCA1^+^/cKIT^-^, d4, d7 and d14 in culture). By clustering the merged dataset and using gene expression and marker gene analysis we identified seven distinct clusters (**Figure A-5B**). When overlaying time points on the different clusters, we find that freshly sorted SCA1^+^/cKIT^-^ cells at d0 contribute to clusters 1 and 2 (**Figure A-5C**). A small number of cells in cluster 1 express von Willebrand factor (*Vwf*) and the receptor tyrosine kinase *Tie-1*, indicating that some endothelial cells also express *Sca1* (**Figure A-1C, A-5B; Table A-2**). However, cluster 2, which constitutes the majority of SCA1^+^/cKIT^-^ cells, is the interstitial progenitor population that expresses *Tcf21*, *Pdgfra*, and *CoupTFII* (**Figure A-5B; Table A-2**).

Four days after exposure to expansion media, cells dominate in clusters 3 and 5 with a smaller number of cells appearing in clusters 4 and 6 (**Figure A-5C**). Cells in cluster 3 are actively proliferating, as reflected by expression of proliferative marker *Mki67*, cyclins *Cdk1* and *Cdc20*, and mitotic microtubule associated proteins (**Figure A-5B; Table S2**). Cells in cluster 5 are post-mitotic myofibroblasts expressing genes involved in actin cytoskeleton dynamics and remodeling (e.g. *S100a4*, *Actn1*, *Lmod1*, *Nexn*, *Acta2*) (**Figure A-5B; Table A-2**). Three days (d7) after

transition to differentiation media, cells have become directed to an ECM-depositing myofibroblast cell state (expressing: *Clu*, *Postn*, *Tnc*, *Col5A2*; cluster 4) or a progenitor Leydig cell state (cluster 6) (**Figure A-5B-C; Table A-2**). Although cells in cluster 4 do not ultimately contribute to Leydig differentiation, they express fibroblast markers (*Col5a2*,*Postn*, *Tnc*) as well as extracellular-matrix related proteins involved in tissue remodeling (**Figure A-5B-C; Table A-2**). Interestingly, this population also expresses *Pdgfa*, raising the possibility that cluster 4 cells serve as an intermediate supportive cell population required to promote continued differentiation of Leydig cells.

By day 14 (10 days of exposure to differentiation media), cells in clusters 6 and 7 have become more differentiated (**Figure A-5C**). Cells in cluster 6 appear to prepare for steroidogenesis by increasing expression of lysosome/exosome genes, likely employing autophagy to degrade cellular components into steroid building blocks like cholesterol and fats (**Figure A-5B; Table A-2**). Previously, autophagy in Leydig cells was shown to be a rate-limiting step for testosterone synthesis[45]. Apolipoprotein E (*ApoE*) is also expressed at this time, indicating that LDL uptake is occurring which is critical for steroidogenesis, in line with genetic evidence in *ApoE/Ldlr* knockout mice[46]. Finally, in cluster 7, steroidogenic enzymes *Cyp17A1*, *Hsd3B1*, *StAR*, *Cyp11A1*are expressed, as well as the mature Leydig cell factor *Insl3*, indicating a cellular state with functional steroidogenesis (**Figure A-5B,F; Table A-2**). The *in vitro* developmental progression based on time point sampling also confirmed Monocle3 pseudotime analysis (**Figure A-5D-F**), where sorted progenitors give rise to a cycle of proliferating and differentiating intermediates. Cells then branch into two differentiation trajectories, one aborting in cluster 4, which is an ECM producing myofibroblast, possibly a support intermediate cell, while the remaining cells proceed through clusters 6 and 7 which lead to differentiated Leydig cells (**Figure A-5D-F**).

Given our success with generating molecularly functional (testosterone secreting) Leydig cells *in vitro*, we next asked whether the *in vitro* derived Leydig cells bear resemblance to *in vivo* Leydig cells by comparing to previously published adult and fetal somatic cell states[13,33]. Notably, the *in vitro* intermediate states (clusters 2-5) in Leydig cell differentiation correlate with early interstitial progenitors in the fetal gonad, whereas clusters 6 and 7 have a higher correlation to fetal Leydig cells (**Figure A-6A**). When comparing to the adult testis, the *in vitro* intermediate states (Clusters

2-5) correlate more closely with the *Tcf21*-expressing interstitial population, whereas clusters 6 and 7 have the highest correlation to adult Leydig cells (**Figure A-6B**). Interestingly, overall the *in vitro* derived Leydig cells have higher correlation to adult Leydig cells than fetal Leydig cells (r= 0.84 vs. 0.58, respectively) (**Figure A-6A,B**).

A.3.4 TCF21^{lin} cells contribute to somatic lineages in the male gonad *in vivo*.

Given our ability to differentiate TCF21^{lin} cells to Leydig or myoid cells *in vitro*, we next asked if the *Tcf21* lineage can serve as a somatic progenitor *in vivo*. To this end, we performed lineage-tracing from early developmental time points in *Tcf21^{mCrem}:R26R^{tdTom}* mice and analyzed fully formed testes at E17.5 and adult testes at 10 weeks. Specifically, timed pregnant females were given a single dose of tamoxifen at either gestational days E9.5, E10.5, E11.5 orE12.5 (**Figure A-7A**). We verified that TCF21^{lin} labeling was absent in embryonic gonads harvested from vehicle-treated timed pregnant females, confirming tight regulation of the tamoxifen inducible Cre (**Figure A-8A**). Additionally, lineage traced cells did not overlap with Vasa, a germ cell marker, confirming specificity of the *Tcf21^{mCrem}* line (**Figure A-8B)**.

Immunofluorescence and histological analysis of E17.5 male gonads revealed marked differences in the extent of labeling across the different tamoxifen injection time points (**Figure A-7B**). We observed relatively fewer TCF21^{lin}cells from E9.5 injected animals, yet those cells co-localized with multiple somatic lineages, as evidenced by co-localization with markers for Sertoli cells (SOX9), interstitial cells (COUPTFII) and, to a lower extent, fetal Leydig cells (3BHSD), and myoid (SMA) cells (**Figure A-7C-F, A-8C-F)**.In contrast, a greater number of TCF21^{lin}cells are observed in the E17.5 gonads collected from animals treated with tamoxifen at E10.5, E11.5 and E12.5, possibly due to broader expression of *Tcf21* in multiple somatic progenitors at these later injection timepoints (**Figure A-7C-F, A-8C-F**). The TCF21^{lin}cells at these different timepoints again contributed to Sertoli (SOX9), fetal Leydig (3BHSD), interstitial (COUPTFII, PDGFRA, and GLI), and myoid (SMA) cells (**Figure A-7C-F, A-8C-H**). Curiously, a significant fraction of E9.5 or E10.5 TCF21^{lin} cells contribute to Sertoli cells, but these cells account for only a fraction of all SOX9$^+$ cells in the fetal and postnatal testis (**Fig A-8C**). This suggests either incomplete labeling of TCF21^{lin}cells or Sertoli cells arising from multiple somatic progenitor populations.

To determine the origin ofTCF21$^+$cells in the embryonic gonad and potential overlap with the WT1$^+$ population – a somatic progenitor previously shown to give rise to Sertoli and interstitial populations including adult Leydig cells[47] – we injected timed pregnant *Tcf21^{mCrem}:R26R^{tdTom};Oct4-eGFP* females with a single dose of tamoxifen at E10.5 and collected and stained whole mount gonads with WT1 at E11.5. In the E11.5 gonads, theTCF21^{lin}cells localize both to the coelomic epithelium and the mesonephros, making it difficult to determine if TCF21^{lin} cells truly originate from the coelomic epithelium or mesonephros, or are present in either location. However, we find that TCF21^{lin} cells partially overlap with the WT1$^+$ cells in the coelomic epithelium, but many cells are either TCF21^{lin} or WT1$^+$, suggesting these are possibly two distinct populations (**Figure A-7G**).

Unlike most somatic cell populations in the testis, the steroid producing Leydig cells are unique in that they arise in two distinct waves. To ascertain whether the fetal-derivedTCF21^{lin} cells persist in the postnatal testis and give rise to adult Leydig cells, *Tcf21^{mCrem}:R26R^{tdTom}* time pregnant females received a single injection of tamoxifen at E10.5 and the pups were fostered and matured to adulthood (see Methods; **Figure A-7A**). In 10-week old male testes, we found that a fraction of adult Sertoli, peritubular myoid, endothelial cells, and interstitial cells are tdTom$^+$ (i.e., derived from TCF21^{lin}) (**Figure A-7C-F**). Furthermore, we observe overlap between TCF21^{lin} and 3BHSD in the adult testis, suggesting that the fetal TCF21^{lin} population gives rise to adult Leydig cells (**Figure A-7D**). Taken together, we demonstrate that the fetal TCF21^{lin} contributes to all adult somatic lineages of the testis. However, since the overlap is incomplete in all somatic populations this raises two possibilities: incomplete labeling or an alternative progenitor source population.

Although the fetal TCF21^{lin} cells *in vivo* contribute to multiple somatic progenitor populations, the limitations of labeling a single somatic progenitor in the fetal testis has hampered our ability to conclusively determine whether the fetal TCF21 population is a heterogeneous population already committed to different fates, or if each cell is individually multipotent.

A.3.5 The TCF21^{lin} gives rise to multiple fetal and adult ovarian somatic cell types

Prior to sex determination in mammals, the gonadal primordium is bipotential, meaning that the gonadal mesenchyme has the ability to give rise to either male or female support cell and

steroidogenic cell lineages[32]. Once *Sry* expression is turned on during a critical window of fetal development, testis differentiation is initiated[48]. Despite an early commitment to either male or female somatic cell types, genetic studies in mouse have shown that terminally differentiated cell types must be actively maintained throughout life. For example, loss of either *Dmrt1 in Sertoli cells* or *Foxl2 in granulosa cells* can trigger reciprocal cell fate conversions (from Sertoli to granulosa cell fate or granulosa to Sertoli cell fate, respectively)[49–51]. Taking into account the known gonadal mesenchyme plasticity and the ability of TCF21[lin] cells to give rise to multiple somatic lineages in the testis, we next asked if TCF21[lin] cells are present in the fetal ovary and whether they might give rise to analogous cell types in females. For our female gonad experiments, we injected timed pregnant female *Tcf21^(mCrem):R26R^(tdTom);Oct4-eGFP* mice with a single dose of tamoxifen at E10.5 and collected female embryonic gonads at E11.5. Our analysis shows that TCF21[lin] cells are present in the coelomic epithelium and mesonephros, similarly to what we had observed in male embryonic gonads (**Figure A-8K**). We then injected tamoxifen in timed-pregnant females at E10.5, E11.5, or E12.5 *Tcf21^(mCrem):R26R^(tdTom)* mice and found broad somatic cell labeling in the female gonads at E17.5 (**Figure A-8L**). Co-staining ovarian cross-sections with terminally differentiated markers shows that TCF21[lin] cells overlap with markers for granulosa cells (FOXL2[+]), interstitial cells (COUPTFII[+]), and smooth muscle cells (SMA[+]), and does not overlap with the germ cell markers VASA or OCT4 (**Figure A-8M-P**).

In the postnatal mouse ovary, the fetal TCF21[lin] population contributes to multiple adult somatic lineages including granulosa cells (FOXL2[+] and WT1[+]) of primordial and growing follicles, and endothelial cells (PECAM[+]), but not germ cells (**Figure A-8Q).** Furthermore, we find that the TCF21[lin] cells surround the follicles and co-expresses the theca cell marker 3BHSD[+52]. Previous studies have shown that theca cells are derived postnatally from GLI1[+] and/or WT1[+] populations[53,54]. Therefore, we asked if the fetal TCF21[lin] cells co-express WT1 . To this end, we examined E11.5 embryonic gonads from timed pregnant *Tcf21^(mCrem):R26R^(tdTom);Oct4-eGFP* mice and co-stained gonad sections for WT1 (**Figure A-8K**). WT1 and TCF21[lin] appear to be largely in distinct populations with the exception of a few cells in the coelomic epithelium (**Figure A-8K**). Therefore, the fetal TCF21[lin] population is largely distinct from the WT1 population, yet, it contributes to all somatic lineages in the adult ovary including: granulosa cells (FOXL2[+] and WT1[+]) of primordial and growing follicles, endothelial cells (PECAM[+]), and theca cells

(3BHSD[+]), suggesting that the female gonad, like the male gonad, may have multiple somatic progenitors.

A.3.6 Tcf21[lin] cells regenerate Leydig cells in the adult testis in response to chemical ablation.
Although somatic cells of the testis are considered to be post-mitotic and do not naturally turnover, our steady state scRNA-seq datasets identified rare *Tcf21[+]* cells that express low levels of either steroidogenic acute regulatory protein (*StAR*) or *Sma*, suggesting that a rare subset of Tcf21[+] cells transition to either Leydig or myoid cells, respectively. We then examined if the adult TCF21[lin] is capable of serving as a somatic progenitor at least for adult Leydig cells in the testis. Specifically, animals were treated with ethane dimethane sulfonate (EDS) to reduce Leydig cell numbers and the testes were collected and assessed to determine (1) if regeneration does occur and (2) if TCF21[lin] contributes to the regeneration.

EDS treatment was previously used in rats to selectively ablate Leydig cells[55–58]. Although there is strong species specificity and animal-to-animal variation in Leydig cell sensitivity to EDS[59], two 300 mg/kg EDS injections in C57BL/6 mice spaced 48 hours apart (**Figure A-10A**) resulted in a reduction of mature Leydig cells in the adult testis as evidenced by cell death and CYP17A1 protein levels (**Figure A-10B-C**). At 12 hours post final injection (hpfi), we detected TUNEL positive Leydig cells, but given the spacing of 48 hours between the first and second injection, the majority of apoptotic events likely preceded the time of analysis and could therefore not be quantified (**Figure A-10B**). However, consistent with Leydig cell loss, we find that CYP17A1 protein expression decreased as early as 12hpfi and had largely recovered by 14dpfi (**Figure A-10C**). We confirmed that the recovery of CYP17A1 protein expression was not simply due to Leydig cell hypertrophy, as the Leydig cell diameter is similar between the EDS and vehicle treated animals at 14dpfi (**Figure A-10D**).

To determine if regenerating Leydig cells were derived from the adult testis TCF21[lin] we injected 8-week old *Tcf21[mCrem]:R26R[tdTom]* mice with three 2mg tamoxifen injections, and then treated the mice with two 300 mg/kg EDS injections spaced 48 hours apart (**Figure A-9A**). Similar to EDS-treated C57BL/6 animals, CYP17A1 protein levels in *Tcf21[mCrem]:R26R[tdTom]* EDS-treated animals decreased at 3dpfi and recovered by 24dpfi **(Figure A-9B)**. To ensure that the reduction in

CYP17A1 is due to Leydig cell loss, we quantified the number of SF1$^+$ cells in EDS and vehicle-treated animals at 3dpfi. In 3 dpfi animals, ~20% of all cells in testis cross sections are SF1$^+$ in the vehicle-treated animal, whereas the number of SF1$^+$ cells is reduced to 5% in the EDS-treated animals, suggesting many Leydig cells were lost (n= 3 vehicle and n=3 EDS animals). To test whether the TCF21^{lin} population contributes to Leydig cell regeneration, we examined if TCF21^{lin} cells proliferate in response to injury and contribute to the regenerated Leydig cells. At 3dpfi, TCF21^{lin} cells surrounding the tubules re-entered the cell cycle as detected by co-localization of BrdU and TCF21^{lin} (**Figure A-9C**). Importantly, by 24dpfi in EDS-treated animals, ~18% of SF1$^+$ Leydig cells are Tcf21^{lin} positive as compared to ~5% in the vehicle treated animals (**Figure A-9D, E**). Furthermore, we find that the EDS-treated animals have a higher number of SF1$^+$ cells as compared to controls (**Figure A-9F),** which is consistent with an overcompensation in CYP17A1 protein levels observed in EDS-treated animals **(Figure A-9B)** and the absence of Leydig cell hypertrophy (**Figure A-9G**).

To independently validate the role of TCF21^{lin} cells in testis somatic cell regeneration, we performed allogenic transplants. Specifically, we sorted adult SCA1$^+$ cells from *Tcf21^{mCrem}:R26R^{tdTom}* animals and transplanted them into the testis interstitium of EDS-treated C57BL/6 animals where no Leydig cells in the host animal will be TdTom$^+$ (**Figure A-10E**). We reasoned that if the transplanted TCF21^{lin} population contributed to Leydig cell regeneration, then we should detect TCF21^{lin}/SF1 double positive cells in the C57BL/6EDS-treated animal. By 24hpfi, we found TCF21^{lin} cells homed to the basement membrane, and by 7dpfi, we began detecting TCF21^{lin}/SF1$^+$ cells (**Figure A-10F**), suggesting that the Tcf21^{lin} cells engrafted in the ablated C57BL/6 mouse testis and gave rise to SF1$^+$ cells.

Therefore, by using the *Tcf21^{mCrem}:R26R^{tdTom}*EDS model and the TCF21^{lin} transplant approach in EDS-treated C57BL/6 mice, we demonstrate that the adult Tcf21^{lin} population in the testis can at least serve as a reserve Leydig progenitor in response to Leydig cell loss.

A.3.7 Peritubular Myoid cells of the testis can regenerate after injury in adult testis.

While we demonstrated that Leydig cells can be regenerated in response to tissue injury, it is unclear whether additional somatic cells in the testis can do the same. Previous studies have shown

that Sertoli cells can be replaced by transplantation but cannot be regenerated following targeted diphtheria toxin (DTX) treatment[60–62], but the regenerative ability of peritubular myoid cells has not been assessed. To examine whether (1) peritubular myoid cells regenerate and (2) if TCF21^{lin} cells contribute to the regeneration, we treated 6-13 week old *Myh11$^{cre-eGFP}$; Rosa26$^{iDTR/+}$* mice (referred to as *MYH11-cre:iDTR* hereafter) with multiple low doses of DTX to balance animal survival and myoid cell ablation and collected testes at 12hpfi and 4dpfi (**Figure A-12A**). By 12hpfi, TUNEL positive cells were present on the tubule basement membrane in *Myh11$^{cre-eGFP}$; Rosa26$^{iDTR/+}$* mice but absent in control mice but were no longer detectable by 4dpfi (**Figure A-12B**). However, at 4dpfi the testis cross-sections of *Myh11$^{cre-eGFP}$; Rosa26$^{iDTR/+}$* animals continue to display vacuoles and disordered tubules, whereas, these histological features were absent from controls (**Figure A-12C**). By 4dpfi, we detect BrdU positive smooth muscle cells surrounding the basement membrane (**Figure A-12D, yellow arrow**) as well as BrdU positive cells on the tubule surface which lacked SMA expression (**Figure A-12D, white arrow**), indicating both neighboring peritubular smooth muscle cells and nearby interstitial cell progenitors re-enter the cell cycle to regenerate the basement membrane in the DTX-treated *Myh11$^{cre-eGFP}$; Rosa26$^{iDTR/+}$*mice. However, since our TCF21 antibody did not yield clear immunofluorescence staining, it could not be determined whether proliferating cells are TCF21^{+}.

In an attempt to overcome these limitations, we sorted SCA1^{+}/TCF21^{lin} cells from *Tcf21^{mCrem}:R26R^{tdTom}* animals and transplanted these cells into the interstitial space of*Myh11$^{cre-eGFP}$; Rosa26$^{iDTR/+}$*diphtheria toxin treated animals. Within 24 hours after transplant, we were able to detect TCF21^{lin} cells surrounding the damaged tubules (**Figure A-12E-F**), but failed to observe any TCF21^{lin} become SMA^{+} at this time point (**Figure A-12G**). Therefore, unlike with Leydig cells, for which we relied on EDS to induce cell-specific ablation, our effort to ablate myoid cells relied on the use of *MYH11$^{cre-egfp}$; Rosa26$^{iDTR/+}$* which is expressed in multiple organs, leading to lower animal viability and preventing the analysis of the fate of TCF21 transplanted cells at later time-points.

In summary, we have demonstrated that peritubular myoid cells can be regenerated. Unlike in the case of Leydig cells described above, whether myoid cells can be derived from TCF21^{lin} cells could not be ascertained at this time and will require additional tools.

A.3.8 Adult TCF21[lin] cells contribute to somatic turnover in the testis during natural aging.

Once established, the somatic cells of the testis are maintained throughout a male's reproductive age. However, a study in rats using [3H]thymidine labeling to detect proliferation suggested that Leydig and possibly peritubular myoid cells may undergo rare events of cellular turnover during an animal's natural lifespan[63]. We then asked to what extent adult mouse testis somatic cells turnover during natural aging, and secondly, whether new cells would be derived from TCF21[lin] progenitors. To answer this question, we performed a long-term lineage tracing experiment where 8-week-old (adult) *Tcf21[mCrem]:R26R[tdTom]* animals were injected with a single dose of 2mg tamoxifen. Animals were euthanized either one-week past final injection or 1 year past final injection (aged mice) (**Figure A-11A**). In animals with one-week labeling we detected rare TCF21[lin] cells surrounding the basement membrane, and these cells do not significantly overlap with Leydig cell markers such as SF1 (**Figure A-11B**). In aged mice, we detected a significant increase in TCF21[lin] and SF1 double positive cells (**Figure A-11C**) as well as more labeling around peritubular cells suggesting that peritubular cells may also be replenished (**Figure A-11B**). This extent of labeling varied by individual animal which could be due to tamoxifen injection efficiency or true natural biological variability in inbred mice. Interestingly, while there is no significant difference in Leydig cell number between short-term labeled and aged individuals, there is more variability among aged individuals. Nevertheless, these data indicate that somatic cells of the testis naturally turnover, possibly at low rates, and the TCF21[lin] population contributes to the replenishment and maintenance of the adult Leydig cell population, and likely peritubular myoid cells as well.

A.3.9 Additional intercellular interactions suggested by scRNA-seq data.

To gain a sense of cellular crosstalk between *Tcf21[+]* cells and germ/somatic cells, we focused on previously documented ligand-receptor (L-R) pairs that are highly variable among major cell types in our scRNA-seq data, and calculated Interaction Scores between germ cells and somatic cells, or among somatic cells of the testis (**Figure A-11E**). We previously showed that the *Tcf21[+]* population has more potential interactions with spermatogonial populations than other germ cells[34] (**Figure A-11E**, right panel), Similarly, we find that spermatogonia can potentially signal back to

the *Tcf21*[+] population (**Figure A-11E**, left panel) via *Pdgfa*, various ADAMs, calmodulins, FGFs, and guanine nucleotide binding proteins.

Within the somatic compartment, *Tcf21*[+] cells have the greatest potential interactions with endothelial, myoid, and macrophages (**Figure A-11E,** middle panel) involving receptor-ligand signaling systems such as *Lrp1* (*Cd91*), a multifunctional, endocytic receptor capable of binding a vast array of ligands[64] and known to regulate the levels of signaling molecules by endocytosis, as well as directly participate in signaling for cell migration, proliferation, and vascular permeability (reviewed in[65]). Like other tissue mesenchymal progenitors, *Tcf21*[+] cells may also modulate local inflammatory responses, as they express Thrombomodulin (*Tbhd*) which interacts with and can proteolytically cleave the pro-inflammatory molecule *Hmgb1* (reviewed in[66]). Additionally, there are several more cell-specific interactions with myoid cells (*Tgfb2-Tgfbr3* ;*Gpc3-Cd81*), endothelial cells (*Cxcl12-Itgb1*; *Pdgfa-Pdgfra*; and *Vegfa-Itgb1*), and macrophages (*Igf1-Igf1r; F13a1-Itgb1*). Several of these putative interactions are involved in growth, wound healing, phagocytosis, and matrix remodeling in various mesenchymal cell types[67]. These putative interactions will need to be validated by spatial analysis and/or functional perturbations of individual signaling pathways.

A.3.10 The *Tcf21*[+] population in the testis resembles resident fibroblast populations in other tissues.

Finally, given the essential role of *Tcf21*[+] cells in mesenchymal development of many tissues, including the heart, lung, and kidney[27,29,37,68], we sought to understand whether the adult testis *Tcf21*[+] population resembles other *Tcf21*[+] mesenchymal populations found in single-cell analyses of other tissues. Our comparison of testis somatic cells with the publicly available scRNA-seq datasets from coronary artery, heart, lung, and liver[69–73] find that the testis *Tcf21*[+] population most closely resembled the resident fibroblast or myofibroblast populations (**Figure A-13, left**), as well as the fibroblast/myofibroblast populations that appear transiently after tissue injury (**Figure A-13, right**). These diverse fibroblast cell types have been documented across tissues and implicated in fibrotic damage and tissue regeneration, even when they may differ with respect to cellular markers or nomenclature (reviewed in[74]).Our results indicate that they are transcriptomically similar to the testicular *Tcf21*[+] population characterized here, suggesting that they collectively

represent an emerging class of resident adult progenitor cells playing similar roles in tissue maintenance and repair across multiple organ systems.

A.4 Discussion

Tcf21 is a basic helix-loop-helix transcription factor, known to have roles in the development of numerous organs, including the testis[27–29]. During testis development, TCF21 is expressed in the bipotential gonadal ridge at E10.5 similar to other key transcription factors, including WT1 and GATA4[75], and loss of *Tcf21* results in gonadal dysgenesis[29,30]. Here, our lineage tracing data show that the fetal TCF21[lin] population is a bipotential gonadal progenitor giving rise to most somatic cell types including steroid-producing cells (fetal and adult), stromal/interstitial cells, and supporting cells, as well as vasculature, consistent with a common bipotential progenitor model recently described[32,33]. Gonadal organogenesis is a complex process with multiple somatic progenitors have been described in both males and females including: WT1, COUPTFII, NESTIN, and PDGFRA[18,21,23,47,76]. To a certain extent, our data reconcile these findings by showing that TCF21[lin] overlaps with many of these markers, but our data supports a multi- progenitor model for both interstitial and Sertoli cell populations that are molecularly and cellularly heterogeneous.

Furthermore, we demonstrate that the TCF21[lin] cells share characteristics with adult mesenchymal progenitors (MPs) for example: TCF21[lin] /SCA1[+]cells can self-renew *in vitro* and possess numerous traits like homing ability, secretion of molecules with anti-inflammatory and immunoregulatory effects, and have multi-lineage potential[77,78]. Furthermore, we show that the SCA1[+]/TCF21[lin] cells can be directed to differentiate to myoid and Leydig cells *in vitro.* Importantly*,* the *in vitro* derived Leydig cells secrete testosterone and highly resemble *in vivo*-derived Leydig cells. The availability of such a robust differentiation protocol makes it possible to use *in vitro*-derived cells to study how environmental toxicants effect steroidogenesis or Leydig cell function/biology.

Testosterone is essential for the development and maintenance of male characteristics and fertility. Reduced serum testosterone affects millions of men and is associated with numerous pathologies including infertility, cardiovascular diseases, metabolic syndrome, and decreased sexual function.

Although exogenous replacement therapies are largely successful in ameliorating these symptoms, they carry increased risks of cardiovascular and prostate disease or infertility[79–81]. Therefore, identifying a progenitor population and/or natural mechanism to restore testosterone levels *in vivo* and combat hypogonadism or age-related decline in testosterone levels is critical[82,83]. Here, we show that resident TCF21[lin] cells or TCF21[lin] allogenic transplants can be activated to support Leydig cell regeneration and replenish Leydig cells upon injury or aging. To date there have been several putative stem Leydig cell populations described in the fetal and early postnatal testis of multiple species. These likely analogous populations are demarcated by diverse cell surface markers that are often species specific, like CD90 for rat and CD51 for mouse (reviewed in[82,84–86]) and exhibit species specific properties. For example, PDGFRA[+] cells isolated from rat can be differentiated to Leydig cells, whereas, although the PDGFRA+ isolated cells from the human testes exhibit aspects of MSC characteristics *in vitro*, but they are unable to fully differentiate into Leydig cells, nor can they produce testosterone[41].

Finally, given the critical role of *Tcf21* in the development of other tissues, as well as in aging and disease models[87,88], we examined commonalities between testis *Tcf21*[+] cells and similar populations in other organ systems and disease states[69–73]. The *Tcf21*[+] population in the adult testis molecularly resembles *Tcf21*[+]fibroblast or fibroblast-like populations that have functional roles in normal tissue maintenance, injury, or disease in other organs such as the heart, lung and liver. While the response to tissue injury is often context dependent, resulting in fibrosis vs. regeneration, it remains unknown if a single resident mesenchymal population is activated to promote either response depending on the levels of damage or signaling pathways activated or if multiple populations respond and leading to divergent outcomes. Here in response to EDS, the TCF21[lin] restores Leydig cells but in many cases, fibrosis is often observed in men with impaired spermatogenesis[89–93]. Therefore, whether dysregulation of the TCF21[+] population may be involved in the pathogenesis of testis fibrosis in certain contexts of infertility remains to be examined. A greater understanding of this population and its regulation may contribute to more informed strategies to restore testis function, as well as the tissue regeneration and tissue repair therapies for other organs[94].

A.5 Methods

Mice

All animal experiments were carried out with prior approval of the University of Michigan Institutional Committee on Use and Care of Animals (Animal Protocols: PRO00006047, PRO00008135), in accordance with the guidelines established by the National Research Council Guide for the Care and Use of Laboratory Animals. Mice were housed in the University of Michigan animal facility, in an environment controlled for light (12 hours on/off) and temperature (21 to 23°C) with ad libitum access to water and food (Lab Diet #5008 for breeding pairs, #5LOD for non-breeding animals).

Colony founders for Rosa26^{iDTR} (Stock #007900), Oct4-eGFP (Stock #004654), *PDGFRA^{EGFP}* (Stock007669), *R26R^{tdTom}* mice (Stock #007909), and *Myh11$^{cre-egfp}$* (Stock #007742, maintained on a B6 background) were obtained from Jackson Labs. The *Tcf21^{mCrem} and the Gli1^{EGFP}* were generously provided by Michelle Tallquist and Deb Gumuccio, respectively. EDS injection studies were performed on age matched C57BL/6 mice obtained from Jackson Labs (Stock #000664). For detailed mouse strain information, see below. All primers used for genotyping are provided in Table S4.

Interstitial Populations Single Cell Data Analysis

Single cell RNA-sequencing analysis comparing somatic cell types in the adult mouse testis.
Somatic cells (N=3,622) and their cell type classifications were defined by Green et. al.[13]. To compare somatic cell populations of the testis we obtained the Euclidean distances for the somatic cell centroids then ordered the cell types using the optimal leaf ordering (OLO) algorithm in R Package Seriation. Based on this analysis, we discovered that the *Tcf21$^+$* population is highly correlated to endothelial, myoid and Leydig cells, but distinct from immune cells and Sertoli cells. To get a better understanding of the functional role of the Tcf21 population, we called differentially expressed genes in each somatic cell type using a nonparametric binomial test. The differentially expressed genes have: (1) At least 20% difference in detection rate; (2) a minimum of 2-fold change in average expression levels, and (3) p-value< 0.01 in the binomial test. Pathway Enrichment analysis for the differentially expressed genes was performed with PANTHER tool v.15 (http://www.pantherdb.org)[95]. Significance of the over- and under representation of GO

Complete Biological Process categories was calculated using Fisher's exact test and multiple testing correction with the false discovery rate.

Ligand-receptor analysis

We used a previously published curated list of ligand-receptor (LR) pairs[34]. We limited the analysis to LR pairs that have either highly variable ligand genes among the 7 somatic cell centroids, or highly variable receptor gene among the 26 germ cell centroids (6 SPG and 20 non-SPG clusters). Thresholds were set for both genes mean and variance across the clusters according to the density. For each ligand-receptor pair we calculated its apparent signaling strength as an "Interaction Score", defined as the product of the mean expression level of the ligand in one cell type and that of the receptor in another cell type. In all, we calculated such an Interaction Score matrix of cell type pairs for germ (ligand)-soma (receptor) interaction, soma (ligand)-soma (receptor) interaction and soma (ligand)-germ (receptor) interaction (reproduced with permission from[34], respectively. To extract the general signaling pattern for each interaction matrix, we defined "strong interactions" for each matrix by keeping the highest 5% Interaction Scores for each matrix. We then calculated the number of such strong L-R interactions for each pair of cell types as their overall interaction strength and displayed them as the line width of arrows in the pairwise interaction plots.

Cell type correlations across tissues

We downloaded the single cell counts data from GEO for artery, lung, heart and two liver datasets[69–73]. For the datasets providing cluster information including our testis datasets, we generated expression centroid for each cell type. We then calculated the spearman rank correlation for all cell type pairs between testis and other tissues. For the artery and the liver datasets, the cell type clusters were not provided, so for these datasets we re-analyzed the raw data using Seurat - following the analysis descriptions from the original papers. For parameters that are not specified, we either used default values or set accordingly. To regenerate the clusters for these raw datasets, we used Louvain clustering in Seurat and assigned cell types according to markers listed in the two papers. We then followed the same procedure to calculate cell type expression centroid and spearman rank correlations with cell types from testis. Summary of all correlations are illustrated in the heatmap.

In Vitro Differentiation Assays

Flow cytometry

Testes were collected from adult C57BL/6 (JAX®mice, stock #000664) mice and enzymatically and mechanically dissociated into a single cell suspension. Briefly, testes from adult mice were excised and the tunica albuginea was removed. Seminiferous tubules were transferred to 10 ml of digestion buffer 1 (comprised of Advanced DMEM:F12 media (ThermoFisher Scientific), 200 mg/ml Collagenase IA (Sigma), and 400 units/ml DNase I (Worthington Biochemical Corp). Tubules were dispersed by gently shaking by hand and a five-minute dissociation at 35°C/215 rpm. To enrich for interstitial cells, tubules were allowed to settle, and the supernatant was collected, quenched with the addition of fetal bovine serum (FBS) (ThermoFisher Scientific), filtered through a 100 um strainer and pelleted at 600g for 5 minutes. The remaining tubules were then transferred to digestion buffer 2 (200 mg/ml trypsin (ThermoFisher Scientific) and 400 units/ml DNase I (Worthington Biochemical Corp) dissolved in Advanced DMEM:F12 media) and dissociated at 35°C / 215 rpm for 5 min each and quenched with the addition of fetal bovine serum (FBS) (ThermoFisher Scientific). The cell pellets from multiple digests were combined and filtered through a 100um strainer, washed in Phosphate-buffered saline (PBS), pelleted at 600g for 3min, and re-suspended in MACS buffer containing 0.5% BSA (Miltenyi Biotec).

The single cell suspensions were stained with a single antibody or combination of antibodies depending on the experiment. The antibodies used include anti-Ly6a-AlexaFluor 488 (1:100; Biolegend, Cat#108115), Biotinylated anti-Ly6a (1:200; Biolegend, Cat#108103), streptavidin conjugated AlexaFluor 488 (1:1000; Life Technologies Cat# S11223; RRID: AB_2336881), anti-CD73-APC (1:300; Biolegend, Cat#127209), anti-CD90.1-Brilliant Violet 650 (1:300; Biolegend, Cat#202533), anti-CD29-PE/Dazzle 594 (1:300; Biolegend, Cat#102231), anti-CD105-PerCP/Cy5.5 (1:300; Biolegend, Cat#120415), anti-CD45-Brilliant Violet 510 (1:300; Biolegend, Cat#), anti-CD117-PE/Cy7 (c-KIT) (1:300; Biolegend, Cat#105813), and anti-CD34-PE (1:300; Biolegend, Cat#128609).

Tri-lineage differentiation assay

Testes were collected from adult C57BL/6 or$Tcf21^{mCrem}$:$R26R^{tdTom}$ mice and dissociated into a single cell suspension and sorted for SCA1$^+$/cKit$^-$ or SCA1$^+$/TCF21^{lin}/cKit$^-$cells, respectively. For adipogenic differentiation, 3 x 10^4 cells were plated in a monolayer, cultured for 10 days (Stem Pro Adipogenic differentiation kit) and stained with either Oil Red O or Perilipin (1:250, Sigma, Cat#**P1873)**. For chondrogenic differentiation, 5 x 10^4 cells were plated in micromass, cultured for 21 days (Stem Pro Chondrogenic differentiation kit) and stained with either Alcian blue or SOX9 (1:250,EMD Millipore, Cat#ABE571). For osteogenic differentiation, 1 x 10^4 cells were plated in micromass, cultured for 14 days (StemPro Osteogenic differentiation kit), and stained with either Alizarin red or Osterix (1:250, Abcam, Cat#ab22552)[96].

CFU-F assay

Colony forming unit assays were performed as previously described[97]. Briefly, testes were collected fromC57BL/6 oradult $Tcf21^{mCrem}$:$R26R^{tdTom}$males following 3 injections of tamoxifen (2 mg) every other day. Following single cell dissociations, SCA1$^+$/cKit$^-$, SCA1$^+$/TCF21^{lin}/cKit$^-$, TCF21^{lin}/cKit$^-$, or cKIT$^+$/SCA1$^-$cells were plated into Corning Primaria 6 well plates at a density of 1000 cells/well. Cells were cultured for 14 days in Mesen Cult MSC medium (StemCell Technologies) and colonies were stained using Giemsa. Colonies were defined as clumps having either>20 or>50 cells.

In vitro directed differentiation to Leydig and myoid cells

Testes were collected from adult C57BL/6 or $Tcf21^{mCrem}$:$R26R^{tdTom}$ males, dissociated into a single cell suspension, and sorted for SCA1$^+$/cKit$^-$ cells. Cells were plated into a 24 well, Matrigel-coated plate at a density of 100,000 cells/well in DMEM/F12 supplemented with 10% FBS and 1X Normocin. For myoid cell differentiation, after 18 hours media was replaced with differentiation media- DMEM/F12 supplemented with 1X Penicillin/streptomycin, 10 ng/ml PDGFAA, 10 ng/ml PDGFBB, 0.5 uM SAG, 10 ng/ml BMP2, 10 ng/ml BMP4, 10 ng/ml Activin A, and 1 mM Valproic acid. After the 7-day differentiation protocol, cells were stained for SMA (1:200, Sigma, Cat#A5228). For Leydig cell differentiation, the FACs sorted cells were initially recovered for 18 hours in DMEM + 10% FBS and then the cells were expanded for 3 days in DMEM/F12 supplemented with 1X Normocin, 10 ng/ml PDGFAA, 10 ng/ml PDGFBB, 0.5 uM SAG, and 10 ng/ml FGF2. After 3 days, the expansion media was replaced with differentiation media-

DMEM/F12 supplemented with 1X Penicillin/streptomycin, 10 ng/ml PDGFAA, 10 ng/ml PDGFBB, 0.5 uM SAG, 10 ng/ml FGF2, 5 mM LiCl$_2$, and 10 uM DAPT. After 10 days in differentiation media, cells were stained for SF1 (1:100, CosmoBio, Cat#KAL-KO610). Media was collected every other day for testosterone measurements.

Clonal expansion and directed differentiation of TCF21[lin]cells

To assess multipotency of TCF21[lin] cells, testes were collected from adult *Tcf21[mCrem]:R26R[tdTom]* males and dissociated into a single cell suspension. Individual TCF21[lin]/cKit[-], SCA1[+]/TCF21[lin]/cKit[-], SCA1[+]/TCF21[lin]/cKit[-]/Cd105[+] or SCA1[+]/TCF21[lin]/cKit[-]/Cd105[-] cells were sorted into Corning Primaria 96 well plates and cultured in MesenCult mouse MSC media for~3 weeks to allow for colony formation. Individual colonies were then directed to differentiate to either a myoid or Leydig cell fate following the directed differentiation protocols described above. Cells were then stained for either SMA (for myoid cells, 1:200, Sigma, Cat#A5228) or SF1 (Leydig cells, 1:100,CosmoBio, Cat#KAL-KO610).

Drop-seq analysis of the Leydig cell differentiation time course analysis.

Drop-seq was performed on cells collected from various points of differentiation, where a single sample per timepoint was diluted to 280 cells/ul and processed as described previously[98]. Briefly, cells, barcoded microparticle beads (MACOSKO-2011-10, Lots 113015B and 090316, ChemGenes Corporation), and lysis buffer were co-flown into a microfluidic device and captured in nanoliter-sized droplets. After droplet collection and breakage, the beads were washed, and cDNA synthesis occurred on the bead using Maxima H-minus RT (ThermoFisher Scientific) and the Template Switch Oligo. Excess oligos were removed by exonuclease I digestion. cDNA amplification was done for 15 cycles from pools of 2,000 beads using Hot Start Ready Mix (Kapa Biosystems) and the SMART PCR primer. Individual PCRs were purified and pooled for library generation. A total of 600 pg of amplified cDNA was used for a NexteraXT library preparation (Illumina) with the New-P5-SMARTPCR hybrid oligo, and a modified P7 Nextera oligo with 10 bp barcodes. Sequencing was performed on a NovaSeq (Illumina) for read 2 length of 94 nt with the Read1 Custom Seq primer. Oligosequences are the same as previously described[13,98].

Single-cell RNA-seq Analysis Across Time points

The paired-end Drop-seq data from days 4, 7, and 14 of *in vitro* Leydig cell differentiation were sequenced in the same batch and were processed using *Drop-seq tools* v1.13 from the McCarroll laboratory as previously described[98,99]. Specifically, the reads were aligned to the mouse reference genome (GRCm38, version 38) using *STAR* v2.7.1a[100]. The pipeline generated digital gene expression matrices with genes as rows and cells as columns that served as the starting point for downstream analyses.

The cells from each time point were first filtered by cell size and integrity – cells with <500 detected genes or with >10% of transcripts corresponding to mitochondria-encoded genes were removed, resulting in 973 – 2980 pass-QC cells for the three time points, for a total of 6,124 cells. Among the retained cells, the average number of detected genes per cell was ~1,888, and the average number of UMIs was ~4,838. For each cell, we normalized transcript counts by (1) dividing by the total number of UMIs per cell and (2) multiplying by 10,000 to obtain a transcripts-per-10K measure, and then log-transformed it by E=ln(transcripts-per-10K+1).

For each time point, we standardized the expression level of each gene across cells by using (E-mean(E))/sd(E) and performed PCA using highly-variable genes (HVG). We obtained 4 clusters using Louvain-Jaccard clustering with top PCs by R package *Seurat* (v2.3.4). We calculated cluster centroids and ordered the clusters by minimizing pairwise Euclidean distance of cluster centroids in R package *Seriation*. We evaluated batch effect by comparing the top PC placements and rank correlation of ordered cluster centroids across the 3 time points. Differentially expressed markers for each cluster were obtained by comparing it against all other clusters using a nonparametric binomial test, requiring at least 20% higher detection rate, a minimum of 1.5-fold higher average expression level, and p-value < 0.01. Based on known markers, there is detected germ cell no in the data.

We extracted the somatic cells from the INT4 dataset from[13]. This dataset was enriched for the *Tcf21*+ interstitial population and was used as the starting time point: day0. We merged the 3 datasets of *in vitro* Leydig differentiation with the somatic cells of INT4, for 6,619 good-quality cells and 24,698 detected genes. These cells on average have 1,837 detected genes and 4,623 UMIs per cell. We selected 2,344 HVG genes in the merged dataset and did PCA using HVG. We performed t-SNE, UMAP and Louvain-Jaccard clustering using top PCs. We obtained 10 clusters

initially, and ordered the clusters as described above. Based on differentially expressed markers for each cluster and rank correlation across the cluster centroids, we decided to merge 3 clusters (clusters 4-6) and identified them as ECM myofibroblast. We merged 2 other clusters (clusters 7-8) as differentiating myofibroblast. This led to the identification of 7 cell types for *in vitro* Leydig differentiation– 1. Endothelial, 2. Interstitial progenitor, 3. Proliferating progenitor, 4. ECM myofibroblast, 5. Differentiating Myofibroblast, 6. Differentiating Leydig, and 7. Leydig. We did pseudotemporal ordering of cells from the 4 time points (days 0, 4, 7, and 14) by *Monocle3* and visualized the single-cell trajectory in UMAP space.

We compared our *in vitro* Leydig differentiation data with those of fetal mouse gonads and adult mouse testis somatic cells. Specifically, we calculated the rank correlation between our 7 cell type centroids of Leydig differentiation data with the 6 cell type centroids from NR5A1-eGFP$^+$progenitor cells from E10.5 – E16.5 fetal male mouse gonads[33] using markers present in both data (N=2,692). We also calculated the rank correlation between our 7 cell type centroids with the 7 cell type centroids from adult mouse testis (Green et al. 2018) using the union of markers present in both datasets (N=1,769).

TCF21LineageTracing

TCF21 lineage tracing analysis and co-localization with immunofluorescence in fetal and adult testis and ovary. *Tcf21^{mCrem}:R26R^{tdTom}* or *Tcf21^{mCrem}:R26R^{tdTom}*; OCT4-eGFP timed-pregnant females were administered with a single dose of 1 mg tamoxifen via gavage at E9.5, E10.5, E11.5 or E12.5. Embryos were obtained at E11.5, E17.5, or E19.5 via C-section. Tail clippings from the embryos were used to identify sex and genotype. Embryonic gonads were fixed in 4% paraformaldehyde (PFA) at 4°C for one hour, transferred to 30% sucrose in 1XPBS at 4°C overnight and embedded in OCT (Surgipath cryo-gel, Leica #39475237). To analyze the TCF21^{lin} contribution in adult testis and ovaries, the E19.5 pups obtained by C-section were fostered to CD1 females. The foster mice were euthanized at 10 weeks. Adult testes and ovaries were fixed in 4% PFA at 4°C overnight, transferred to 30% sucrose in 1XPBSat 4°C for overnight and embedded in OCT.

For immunofluorescence, 7-10 micron thick OCT sections were cut using a Leica CM3050S cryostat and sections were refixed with 4% PFA for 10 minutes and permeabilized by incubation in 0.1% Triton in PBS for 15 minutes. Sections were blocked in 1XPBS supplemented with 3% BSA and 500mM glycine for one hour at room temperature and co-stained with FGF5, SOX9, 3β-HSD, CD34, SMA, VASA, COUPTFII, CD31, SF1, WT1, FOXL2, and DsRed antibodies. The primary antibodies and concentrations used are summarized in the table below. All secondary antibodies (Alexa-488-, Alexa-568-, and Alexa-647-conjugated secondary antibodies; Life Technologies/Molecular Probes) were all used at a 1:1000 dilution. DAPI was used as a nuclear counterstain at 1:1000. Representative images were taken with a ZeissAX10 epifluorescence microscope, a Leica SP8 confocal microscope, or a Nikon A1R-HD25 confocal microscope and processed with ImageJ.

Quantification of immunofluorescence colocalization

Tissue sections were stained for immunofluorescence as described above and>20 images per testis were imaged with a 40X 1.2NA objective on a ZeissAX10 epifluorescence microscope, all at a single z-section. The percent overlap between the TCF21[lin] population and several marker proteins was done using accustom written ImageJ macro (available upon request). Briefly, nuclear regions of interest (ROIs) were created from DAPI staining by blurring with a Gaussian filter, making the image binary, separating over lapping nuclei with a watershed function, then saving the outline of each binary nucleus to ImageJ's ROI manager. The signal from TCF21[lin] and the immunostaining were made binary using ImageJ's automatic thresholding function and the overlap of the binary stain and the nuclear ROI was measured using Image J's Measurement function. Cells positive for each stain and double positive cells were sorted and identified in Microsoft Excel. The quantification from each testis was the sum of all quantified images taken from a single testis. The macro was optimized by contrasting its results to manual quantification from at least five images per immunostaining.

Lineage-tracing analysis of the Tcf21+ population in the aged testis.

Five or eight week-old *Tcf21[mCrem]:R26R[tdTom]* males were injected with a single dose of 2 mg tamoxifen intraperitoneally. The testes were collected one week (as control) or one year past injection. The testes were fixed in 4% paraformaldehyde at 4°C overnight, transferred to 30%

sucrose in 1XPBS at 4°C overnight and embedded in OCT. The sections were co-stained for SF1 (1:100, CosmoBio, Cat#KAL-KO610). Overlap of SF1 and tdTomato were counted manually per field; ~ 25-100 fields per biological replicate per condition were collected.

In vivo ablation of Leydig cells

Ethane Dimethane Sulfonate (EDS) Injections

Ethane dimethane sulfonate (EDS) (AABlocks, Cat. No: 4672-49-5) was dissolved in DMSO at a concentration of 150 mg/mL. Working solutions of EDS were further diluted in PBS to a final concentration of 50mg/mL. A total of 300mg/kg of EDS was injected intraperitoneally every other day for 2 days into C57BL/6 age-matched mice. Testes and sera were collected 12 hours, 24 hours, 4 days, 7 days, 14 days or 21 days past final injections. Negative controls were given vehicle treatment of DMSO: PBS without EDS.

Diphtheria Toxin Injections

Diphtheria toxin (DTX) (Sigma) was diluted in PBS at a concentration of 2 mg/ml. A final concentration of 150ng DTX was injected intraperitoneally every day for 3 or 4 days into *Myh11$^{cre-egfp}$;Rosa26$^{DTR/+}$* male mice. Testes were collected 12 hours, and 4 days past final injections. Littermates that are Myh11-Cre-recombinase negative were given DTX and served as controls.

TUNEL staining

Testes were collected, fixed in 4% PFA for approximately 16 hours at 4°C, dehydrated in ethanol wash series, and embedded in paraffin. Five-micron FFPE tissue sections were deparaffinized, rehydrated, and permeabilized in 20 μg/ml Proteinase K solution for 15 minutes at room temperature. Samples were further processed following the Promega Dead End Colorimetric TUNEL kit according to the manufacturer instructions. All images collected used a Leica Leitz DMRD microscope.

Hormone Measurements

Testosterone measurements were performed by the University of Virginia Center for Research in Reproduction Ligand Assay and Analysis Core.

FFPE Immunofluorescence

Whole testes were fixed in 4% PFA overnight at 4°C and processed for formalin fixed paraffin embedding as described in (Fisheretal 2006). Five-micron FFPE tissue sections were deparaffinized by incubation in Histoclear 3x for 5 minutes, followed by incubation in 100% EtOH 2x for 5 minutes, 95% EtOH 2x for 5 minutes, 80% EtOH 1x for 5 minutes, 70% EtOH 1x for 5 minutes, 50% EtOH 1x for 5 minutes, 30% EtOH 1x for 5 minutes, and deionized water 2x for 3 minutes each. Tissue sections were permeabilized by incubation in 0.1% Triton in PBS for 15 minutes. For all antibodies, antigen retrieval was performed by boiling in 10mM sodium citrate, pH 6.0 for 30 minutes. Sections were blocked in 1xPBS supplemented with 3% BSA and 500mM glycine for three hours at room temperature. Endogenous peroxidases and alkaline phosphatases were blocked by a ten-minute incubation in BloxAll solution (VectorLabs, Cat. No: SP-6000). The primary antibodies and concentrations used are listed below. Alexa-488-, Alexa- 555-, and Alexa-647-conjugated secondary antibodies (Life Technologies/MolecularProbes) were all used at 1:1000. DAPI was used as a nuclear counterstain. For quantification, the overlap of SF1 and tdTomato were counted manually per field; ~ 25-50 fields per biological replicate per condition were collected.

Cell diameter measurements

For all cell diameter measurements, FFPE slides were stained as described above and co-stained with SF1 to mark Leydig cells (1:100, CosmoBio, Cat#KAL-KO610) and 488-wheat germ agglutinin (WGA, Biotium Cat. No: 29022-1) to mark cell perimeter. A tubule was centered in the field of view at 40X magnification and an image was taken. At least 70 well-defined Leydig cells, marked by both SF1 and clear visible cell perimeter by WGA, were counted per condition. Cell diameter was measured manually using ImageJ's Line and Measure functions.

FFPE Immunohistochemistry

Whole testes were fixed in 4% PFA overnight at 4°C, processed and deparaffinized as described above. The primary antibodies and concentrations used are listed in table below. Horseradish peroxidase and alkaline phosphatase conjugated secondary antibodies (Abcam) were all used at a

1:100 concentration and left to incubate for 1-hour at room temperature. Slides were rinsed of developing solution under running DI water for 1 minute and mounted in permount.

Transplantation of TCF21^{lin}cells into the testis of WTEDS treated or Myh11$^{cre-egfp}$;Rosa26$^{DTR/+}$animals

6-18 week *Tcf21^{mCrem}:R26R^{tdTom}* male mice were injected with 1mg tamoxifen intraperitoneally every other day for a total of three injections. Testes were dissociated into a single cell suspension and the SCA1$^+$/cKit$^-$ stained cells were collected by FACs, as described above. For each animal~65,000-150,000 cells were diluted in 10 ul of MEM media plus Trypan Blue and were injected into the interstitium via the rete testes of either EDS-treated C57BL/6 animals 24 hpfi of EDS or into *Myh11$^{cre-egfp}$;Rosa26$^{DTR/+}$* mice 24 hpfi of DTX. As a control, 10 uL of MEM media plus Trypan Blue was injected into the contralateral testis of each experimental animal. Testes were collected at 24 hours pasttransplant (hpt), 4 days pasttransplant (dpt) and 7dpt for FPPE processing as described above.

Data and Code Availability

Single-cell RNA-seq data files are available in "GSE151337." Code used in this analysis were deposited onto GitHub: https://doi.org/10.5281/ zenodo.4743036.

A.5 Acknowledgements

Table A-1 Reagents used in this study

Antibody	Company	Product #	Application	Concentration
Ly6a-AlexaFluor 488	Biolegend	108115	Flowcytometry	1:200
CD73-APC	Biolegend	127209	Flowcytometry	1:300

CD90.1-BrillianatViolet 650	Biolegend	202533	Flowcytometry	1:300
CD29-PE/Dazzle 594	Biolegend	102231	Flowcytometry	1:300
CD105-PerCP/Cy5.5	Biolegend	120415	Flowcytometry	1:300
CD45-BrillianatViolet 510	Biolegend	103137	Flowcytometry	1:300
CD117-PE/Cy7 (c-KIT)	Biolegend	105813	Flowcytometry	1:300
CD34-PE	Biolegend	128609	Flowcytometry	1:300
FGF5	SantaCruz	sc-376264	immunofluorescence	1:100
SOX9	EMDMillipore	ABE571	immunofluorescence	1:500
SOX9	Abcam	ab3697	immunofluorescence	1:500
3BHSD	Santa Cruz	Sc-30820	immunofluorescence	1:200
CD34	Abcam	ab81289	immunofluorescence	1:200
Smoothmuscleactin	Sigma	A5228	immunofluorescence	1:200
VASA(Ddx4)	Abcam	ab13840	immunofluorescence	1:200
COUPTFII(Nr2F2)	R&DSystems	PPH7147-00	immunofluorescence	1:300
CD31 (PECAM)	BDBiosciences	553370	immunofluorescence	1:100
SF1 (Nr5A1)	Hammerlab		immunofluorescence	1:500
SF1 (Nr5A1)	CosmoBio	KAL-KO610	immunofluorescence	1:100
WT1	Santa Cruz	sc-192	immunofluorescence	1:100
FOXL2	Wilhelm lab		immunofluorescence	1:300
FOXL2	Novus Biologicals	NB100-1277	immunofluorescence	1:300
DsRed (fortdTomato)	TakaraBioUSAInc	632496	immunofluorescence	1:1000
CYP17A1	ProteinTech	14447-1-AP	Western	1:500
β-actin	Novus Biologicals	NBP1-47423	Western	1:2000
GAPDH	ProteinTech	10494-1-AP	Western	1:500

BrdU	Abcam	Ab6326	immunofluorescence	1:500
Osterix	Abcam	ab22552	Immunofluorescence	1:250
Perilipin	Sigma	**P1873**	Immunofluorescence	1:250
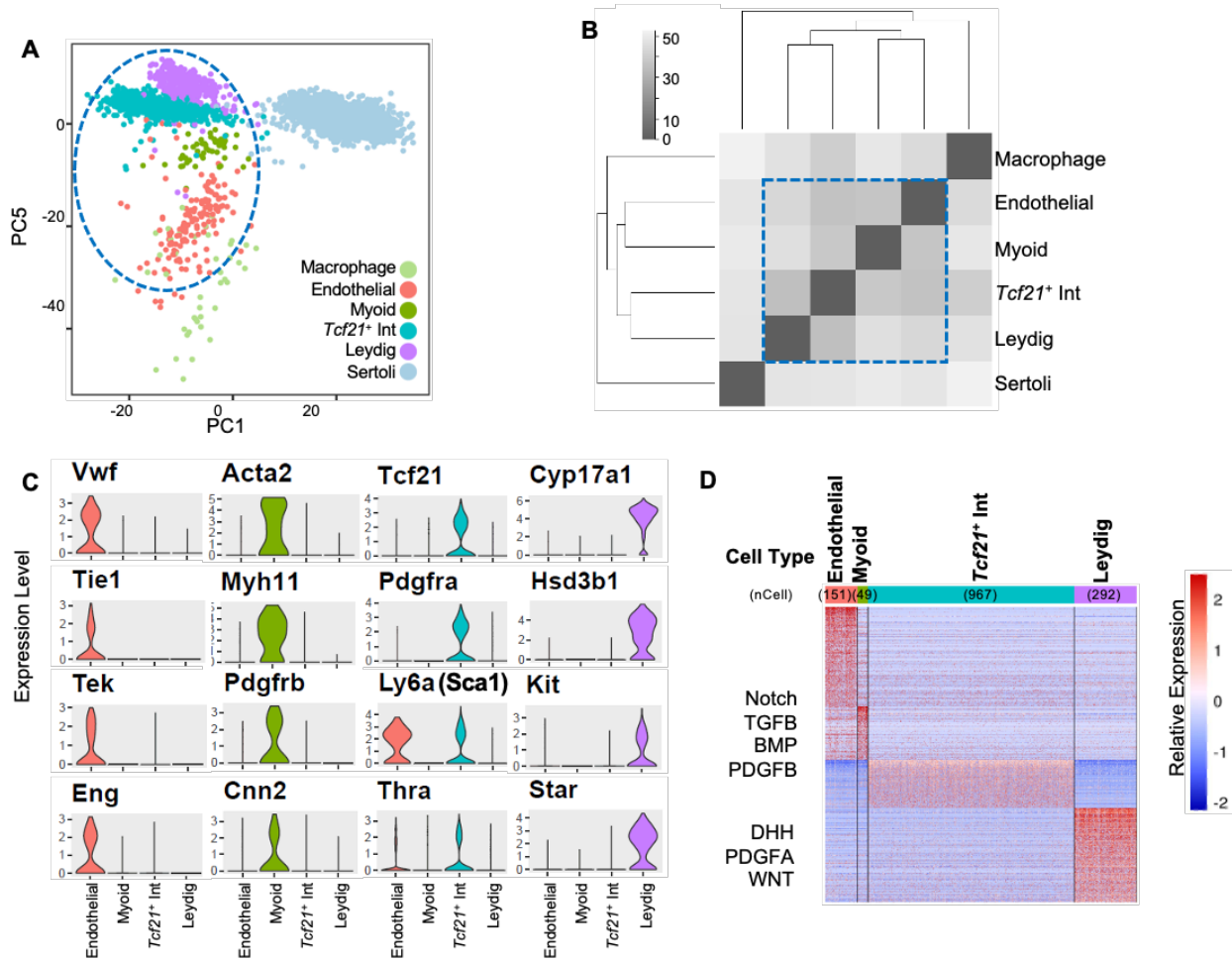
Figure A-1: Identification and characterization of a *Tcf21*-expressing interstitial somatic cell in the adult testis. (A) Visualization of 6 somatic cell types in PC space. Data reprocessed from Green et al.[13] Note somatic cell populations were enriched using a combination of cell surface markers or transgenic lines, therefore cell frequencies in PCA plot are not representative of *in vivo*. **(B)** Heatmap of dissimilarity matrix of somatic cell types illustrates high similarity for the *Tcf21*+ interstitial population with 3 somatic cell types – myoid, Leydig, and endothelial cells. Gray scale indicates Euclidean distance. **(C)** Violin plots of representative markers for endothelial, myoid, *Tcf21*+ interstitial population, and Leydig cells. **(D)** Heatmap of differentially expressed markers for each of the 4 closely related somatic cell types – endothelial, myoid, *Tcf21*+ interstitial population, and Leydig cells.

155

A

CD45-BV510 — Side Scatter — 98.2

SCA1-AF488 — cKIT-PE/Cy7 — 1.45 — 0.71

SCA1 FMO — 0.012

cKIT FMO — 0.070

B

SCA1+

tSNE1 / tSNE2

A:10.6 B:0.77 C:13.8 D:7.91 E: 66.7

C

SCA1

CD29-PE/D594 CD73-APC CD105-PerCP/Cy5.5

THY1-BV650 TCF21lin (tdT) CD34-PE

Percent

Cluster
- **A** CD29LoCD73Lo CD105-THY1-TCF21lin-CD34Lo
- **B** CD29+CD73LoCD105-THY1+TCF21linLoCD34Lo
- **C** CD29+CD73+CD105-THY1-TCF21lin+CD34+
- **D** CD29+CD73+CD105-THY1-TCF21lin- CD34-
- **E** CD29+CD73+CD105-THY1-TCF21lin-CD34+

D

TCF21lin (tdT) — cKIT-PE/Cy7 — 1.33

E

TCF21lin

tSNE1 / tSNE2

F:16.2 G:8.81 J:13.1 H:58.5 I:1.0

Cluster
- **F** SCA1+CD34HiCD29+CD73+CD105+THY1-
- **G** SCA1+CD34HiCD29+CD73+CD105-THY1-
- **H** SCA1+CD34HiCD29+CD73+CD105-THY1-
- **I** SCA1-CD34LoCD29HiCD73HiCD105+THY1+
- **J** SCA1-CD34LoCD29+CD73+CD105LoTHY1-

F

TCF21lin

SCA1-AF488 CD34-PE CD29-PE/D594

CD73-APC CD105-PerCP/Cy5.5 THY1-BV650

Percent

G

cKIT-PE/Cy7 — SCA1-AF488 — SCA1+ 3.9%

TCF21lin (tdT): tdT- 55% tdT+ 45%

cKIT-PE/Cy7 — TCF21lin (tdT) — tdT+ 2.0%

SCA1-AF488: SCA1- 20% SCA1+ 80%

H

DAPI TCF21lin Marker | DAPI | TCF21lin | Marker
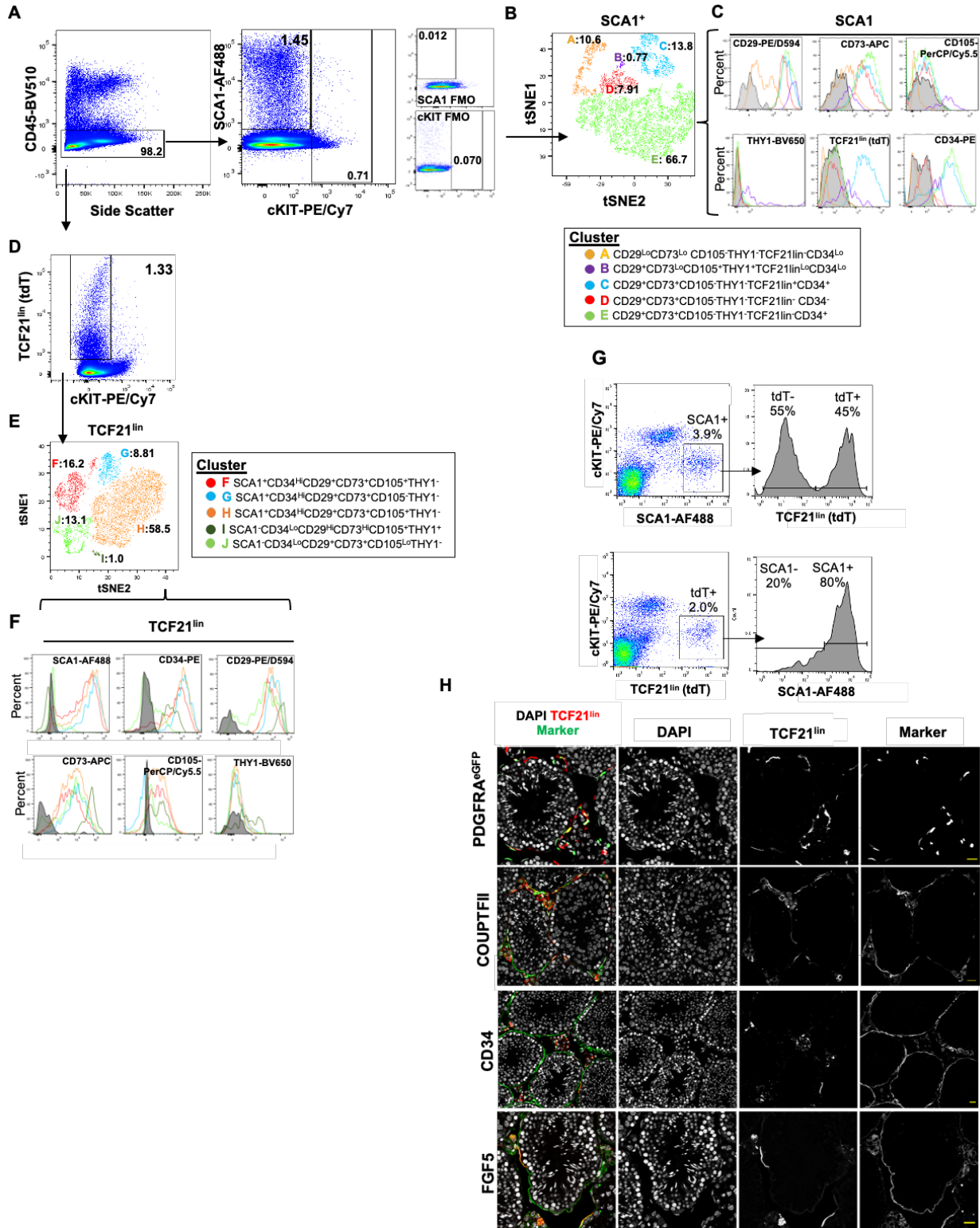
PDGFRAeGFP

COUPTFII

CD34

FGF5

Figure A-2: The SCA1$^+$ and TCF21^{Lin}populations are molecularly heterogeneous and express multiple mesenchymal progenitor markers. (A) Gating strategy used to identify interstitial SCA1$^+$ cells in the testis. The first plot shows side scatter area and CD45 fluorescent intensity used to eliminate immune cell populations (selection of SCA1$^+$**,** CD45$^-$). The second plot is a SCA1$^+$and cKIT$^+$ fluorescent intensity plot used to remove Leydig cells (selection of SCA1$^+$**,** CD45$^-$, cKIT$^-$). All positive gates were determined using FMO isotype controls. **(B)** t-SNE analysis (perplexity 35,300 iterations) of thousands of cells co-stained for TCF21^{lin}, CD73, CD105, Thy1, CD29, and CD34 identifies five SCA1$^+$subtypes (Clusters A-E) based on the signal intensity of various markers. **(C)** Fluorescent intensity histograms of the 6 markers for the 5 clusters defined in panel B, depicted with corresponding colors. Isotype controls for each marker are indicated with a grey shaded peak. **(D)** Gating strategy used to characterize interstitial TCF21^{lin}cells in the testis, similar to A. Note no FMO isotype control shown for Tcf21 due to genetic labeling strategy. **(E)** t-SNE analysis (perplexity 35,300 iterations) of thousands of cells co-stained for SCA1, CD73, CD105, THY1, CD29, and CD34 identifies five TCF21^{lin}clusters (Clusters F-J) based on the signal intensity of various markers. **(F)** Fluorescent intensity histogram for each of 6 markers across the 5 clusters defined in panel E, depicted with corresponding colors. Isotype controls for each marker are indicated with grey shaded peaks. Data was collected from a pool of 2-3 animals. **(G)** Characterizing the overlap between the TCF21^{lin} (Tdtomato) and SCA1.Data collected from a pool of 2-3 animals. **(H)** Immunofluorescence overlay of TCF21^{lin}with PDGFRA^{eGFP}(n=1), COUPTFII, CD34**,** and FGF5 (n=2) expression in the adult mouse testis. Scale bars: 20μm.

A Myoid Differentiation

FACS sorted individual cells → Clonal expansion → d1 ‖ d7

DMEM/F12 + PDGFAA, PDGFBB, SAG, BMP2, BMP4, ACTIVIN A, Valproic acid

B Leydig Differentiation

FACS sorted individual cells → Clonal expansion → d1 ‖ d4 ‖ d7 ‖ d14

DMEM/F12 + PDGFAA, PDGFBB, FGF2, SAG | DMEM/F12 + PDGFAA, PDGFBB, FGF2, SAG, LiCl₂, DAPT

Media collection (d4), Media collection (d7), Media collection (d14)

C

DAPI | TCF21lin | SMA | TCF21lin SMA

TCF21lin/cKIT⁻
TCF21lin/SCA1⁺/cKIT⁻
TCF21lin/SCA1⁺/CD105⁺/cKIT⁻
TCF21lin/SCA1⁺/CD105⁻/cKIT⁻

D

DAPI | TCF21lin | SF1 | TCF21lin SF1

TCF21lin/cKIT⁻
TCF21lin/SCA1⁺/cKIT⁻
TCF21lin/SCA1⁺/CD105⁻/cKIT⁻
TCF21lin SCA1⁺/CD105⁺/cKIT⁻

E

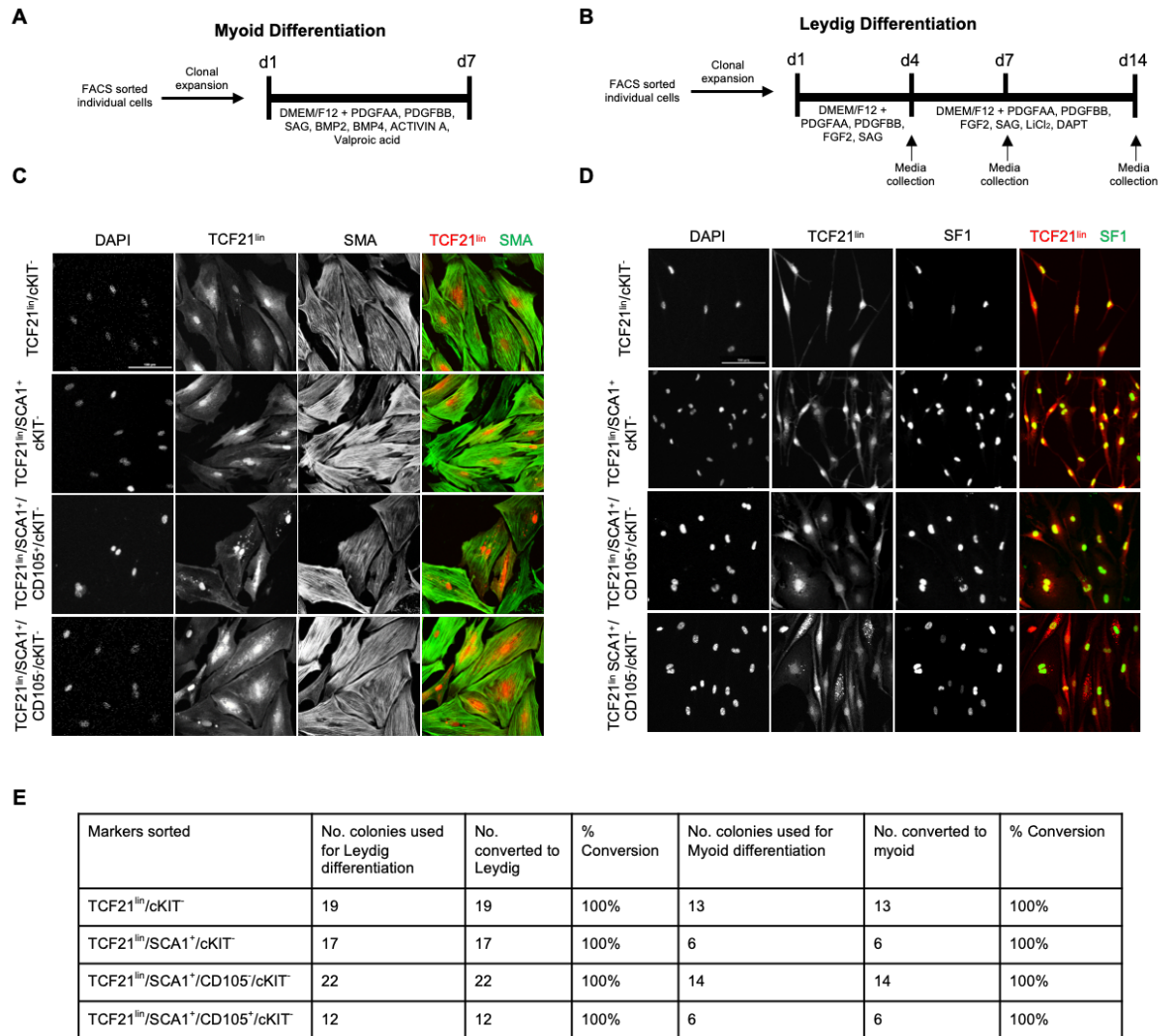| Markers sorted	No. colonies used for Leydig differentiation	No. converted to Leydig	% Conversion	No. colonies used for Myoid differentiation	No. converted to myoid	% Conversion
TCF21lin/cKIT⁻	19	19	100%	13	13	100%
TCF21lin/SCA1⁺/cKIT⁻	17	17	100%	6	6	100%
TCF21lin/SCA1⁺/CD105⁻/cKIT⁻	22	22	100%	14	14	100%
TCF21lin/SCA1⁺/CD105⁺/cKIT⁻	12	12	100%	6	6	100%

Figure A-3: The adult TCF21^{lin} cells are multipotent and can be directed to differentiate to Leydig and smooth muscle cells *in vitro*. (A-B) Schematic representation of experimental timelines for myoid **(A)** and Leydig cell **(B)** directed differentiation. Gating strategy used for FACS is presented in Supplemental Figure 2G. **(C)** Immunofluorescence staining of smooth muscle actin (SMA) in clonal TCF21^{lin} cells after 7 days of *in vitro* culture in the presence of differentiation media. Representative images of n=13 technical replicates of TCF21^{lin}/cKit⁻, n=6 technical replicates of TCF21^{lin}/SCA1⁺/cKit⁻, n=6 technical replicates of TCF21^{lin}/SCA1⁺/cKit⁻/CD105⁺, and n=14 technical replicates of TCF21^{lin}/SCA1⁺/cKit⁻/CD105. Scale bar: 100 μm **(D)** Immunofluorescence staining of steroidogenic factor 1 (SF1) in clonal TCF21^{lin} cells after 14 days of *in vitro* culture in the presence of differentiation media. Representative images of n=19 technical replicates of TCF21^{lin}/cKit⁻, n=17 technical replicates of TCF21^{lin}/SCA1⁺/cKit⁻, n=12 technical replicates of TCF21^{lin}/SCA1⁺/cKit⁻/CD105⁺, and n=22 technical replicates of TCF21^{lin}/SCA1⁺/cKit⁻/CD105⁻.Scale bar: 100 μm. **(E)** Table summarizing the conversion efficiency of various TCF21^{lin}subtypes to either myoid or Leydig cells.
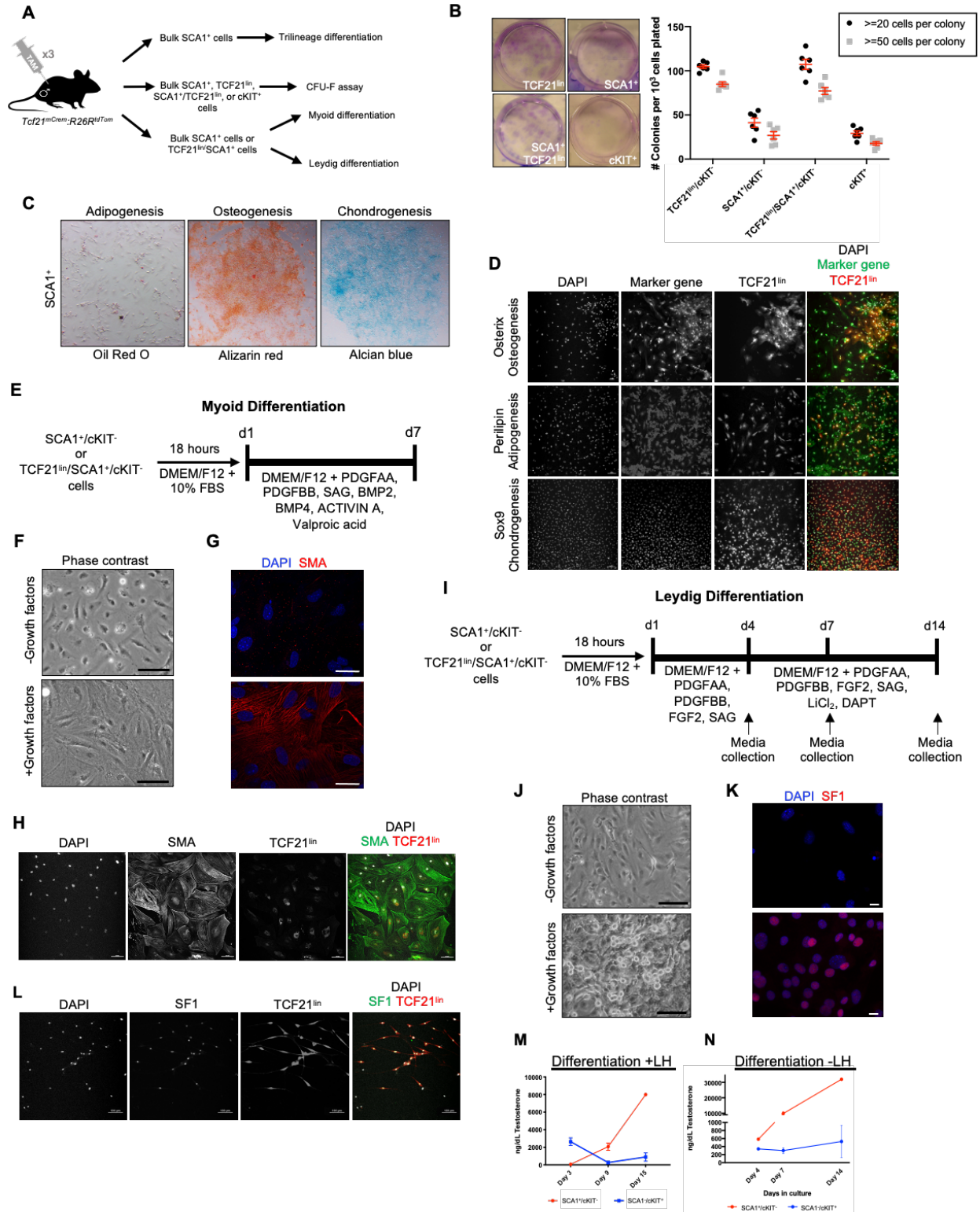
Figure A-4: The TCF21^{lin} population is a mesenchymal progenitor that can be differentiate to Leydig or smooth muscle cells. (A) Schematic representation of experimental outline for trilineage differentiation, CFU-F, and myoid/Leydig cell differentiation**. (B)** Enrichment of clonogenic cells in the TCF21^{lin}population. Left panels are CFU-F representative images and right panel is quantification of the number of colonies with either >20 or >50 cells (n=6 wells). Data are presented as mean ±SEM. **(C)** Representative images from SCA1^{+} trilineage differentiation. Adipogenic differentiation was analyzed at 10 days using Oil Red O, chondrogenic differentiation using Alcian blue at 14 days, and osteogenic differentiation at 21 days using Alizarin red. Representative of n=3 wells per condition per 4 biological replicates. **(D)** Representative images from SCA1^{+}/TCF21^{lin} trilineage differentiation (n=4 wells per condition). Scale bar: 100 μm.**(E)** Myoid cell differentiation scheme for bulk SCA1^{+}/cKIT^{-} or SCA1^{+}/TCF21^{lin}/cKIT^{-} populations. Cell morphology and expression of smooth muscle actin (SMA) in SCA1^{+} cells after 7 days of culture in the presence or absence of myoid differentiation media via phase contrast **(**Scale bar:100 μm, **F)** and immunofluorescence **(**Scale bar: 25 μm, **G).** Representative images from n=3 wells per condition. **(H)** Representative images of SMA in SCA1^{+}/TCF21^{lin}/cKIT^{-} cells after 7 days of culture. Scale bar: 100 μm. A total of n=3 independent experiments with n=8 technical replicates were performed. **(I)** Leydig cell differentiation scheme for bulk SCA1^{+}/cKIT^{-} or SCA1^{+}/TCF21^{lin}/cKIT^{-} populations. Cell morphology in the presence or absence of a differentiation media via phase contrast **(**Scale bar:100 μm, **J)** and expression of SF1 in SCA1^{+} cells **(K)** or SCA1^{+}/TCF21^{lin}/cKIT^{-} cells **(**Scale bar:10 μm, **L)** after 14 days of culture. Scale bar: 100 μm**.** A total of n=3 independent experiments with n=8 technical replicates were performed for J-L. **(M)** Media testosterone levels measured from SCA1^{+}/cKIT^{-} and cKIT^{+}/SCA1^{-} cells at 3 (n=6 per condition), 9, and 15 days of culture (n=4 per condition) in the presence of LH. Data are presented as mean ±SEM. **(N)** Media testosterone levels measured from SCA1^{+}/cKIT^{-} and cKIT^{+}/SCA1^{-}cells at 4, 7, and 14 days (n=3 per condition) of culture in the absence of LH. Data are presented as mean ±SEM.
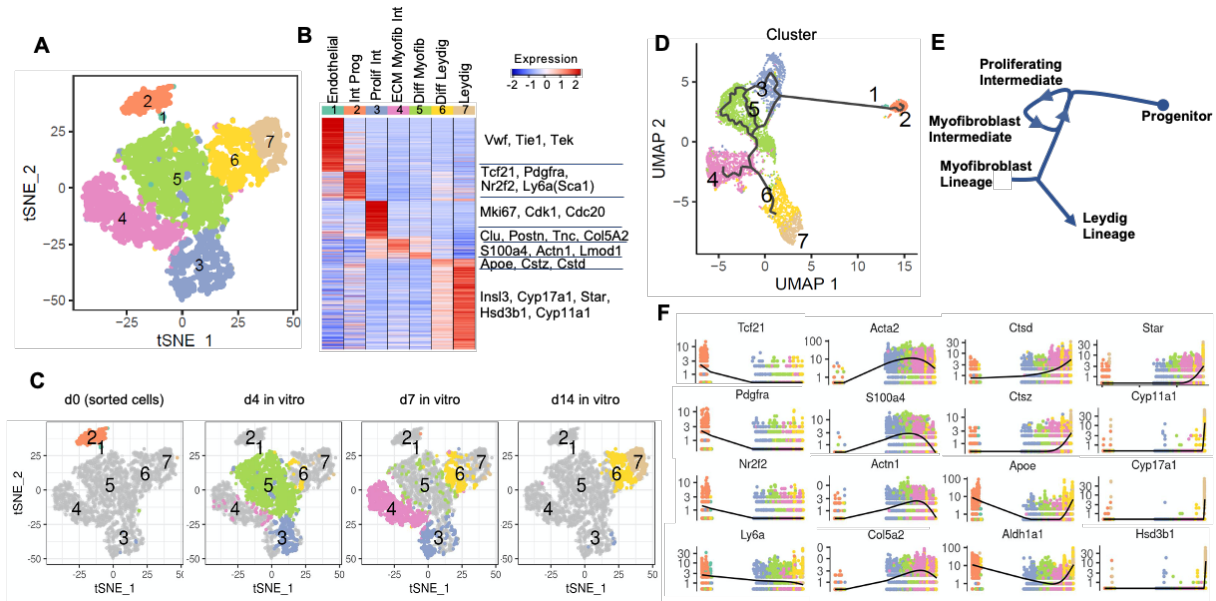
Figure A-5: scRNA-seq differentiation trajectory of *in vitro* derived Leydig cells. (A) Single cell RNA-seq time course analysis of *in vitro* Leydig differentiation (days 0, 4, 7, and 14) identifies seven clusters, as visualized in t-SNE space. Note: n=1 replicate per timepoint, but *in vitro* differentiation was successfully completed 8 times in prior experiments. **(B)** Heatmap of differentially expressed markers across the seven cluster centroids. **(C)** Visualization of the contribution of each individual time point at day 0, 4, 7, or 14 to the 7 clusters in t-SNE space. **(D)** Pseudotime ordering of cells from the 4 time points (days 0, 4, 7, and 14) by Monocle3 in UMAP space, colored by 7 clusters. **(E)** Schematic annotation of differentiation trajectory as defined by Monocle. **(F)** Expression profiles of selected markers across the differentiation pseudotime.

A

Stevant *et al.*, 2018
Somatic cell clusters of the fetal mouse testis

In vitro differentiation Leydig cell clusters	C1.Endothelial	C2.EarlyProg	C3.IntProg	C4.Pre.Sertoli	C5.FetalLeydig	C6.Sertoli
1.Endothelial	0.49	0.35	0.39	0.35	0.35	0.36
2.IntProg	0.42	0.41	0.56	0.40	0.46	0.39
3.ProlifProg	0.46	0.54	0.64	0.39	0.48	0.45
4.ECMmyofibroblast	0.46	0.49	0.58	0.43	0.45	0.41
5.DiffMyofibroblast	0.46	0.49	0.60	0.41	0.45	0.40
6.DiffLeydig	0.43	0.46	0.61	0.45	0.53	0.46
7.Leydig	0.39	0.42	0.55	0.49	0.58	0.50

B

Green *et al.*, 2018
Somatic cell clusters of the adult mouse testis

In vitro differentiation Leydig cell clusters	InnateLymphoid	Macrophage	Endothelial	Myoid	*Tcf21*+Int	Leydig	Sertoli
1.Endothelial	0.26	0.36	0.89	0.59	0.54	0.43	0.35
2.IntProg	0.32	0.61	0.58	0.63	0.98	0.73	0.41
3.ProlifProg	0.39	0.50	0.56	0.63	0.74	0.58	0.37
4.ECMmyofibroblast	0.33	0.50	0.56	0.66	0.77	0.62	0.41
5.DiffMyofibroblast	0.37	0.53	0.58	0.67	0.80	0.61	0.36
6.DiffLeydig	0.28	0.54	0.51	0.61	0.85	0.73	0.42
7.Leydig	0.24	0.47	0.44	0.53	0.76	0.84	0.47

Figure A-6: *In vitro* **derived Leydig cells resemble adult Leydig cells present in vivo. (A)** Rank correlation of the seven centroids from the *in vitro* Leydig time-course differentiation with 6 cluster centroids derived from previously published Nr5a1+GFP+ progenitor cells from E10.5 – E16.5 male mouse gonads[33], and **(B)** previously published adult mouse somatic cells[13]
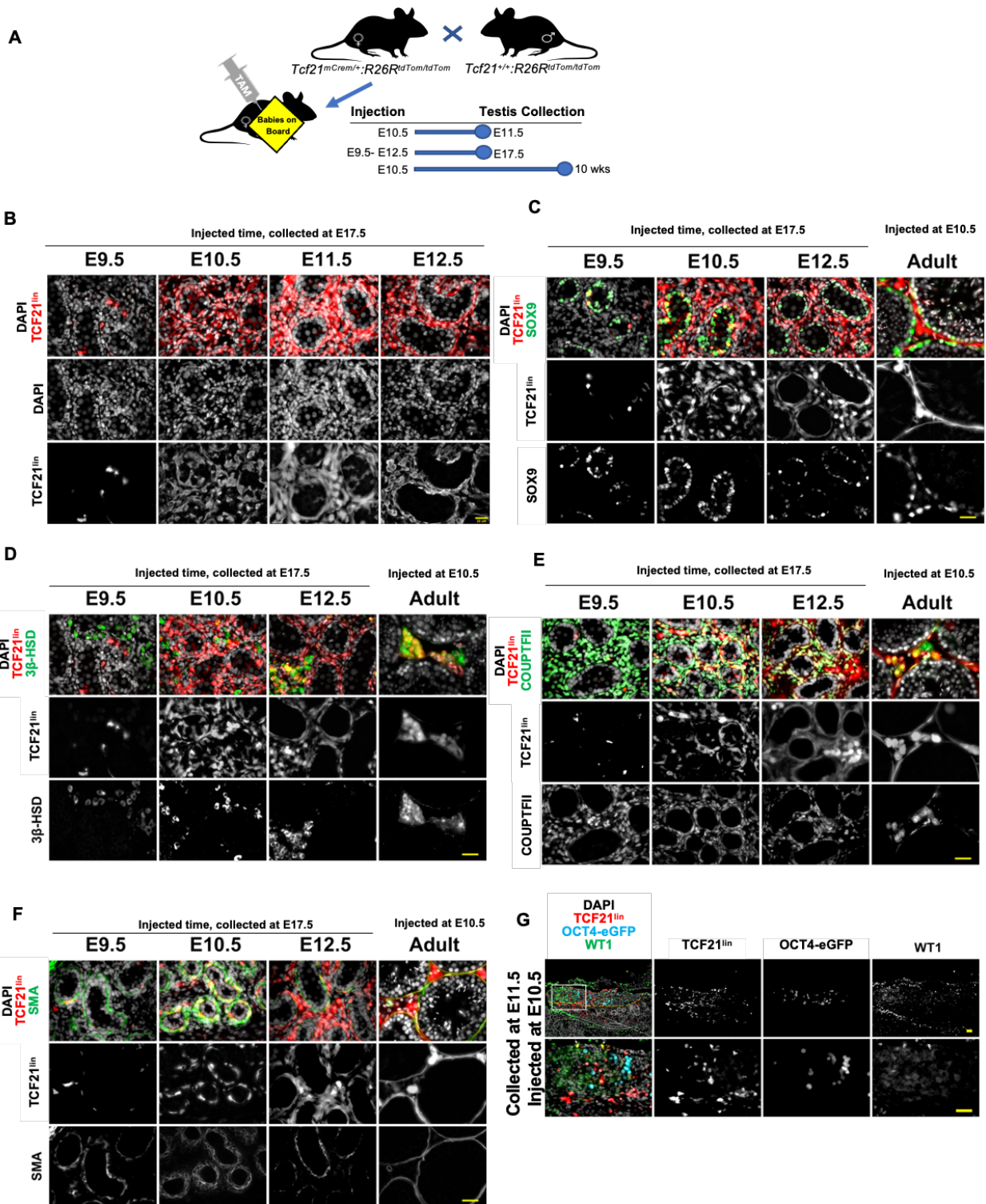
163

Figure A-7: The TCF21^{lin} population contributes to multiple somatic lineages in the fetal and adult testis. (A) Experimental timeline used for *Tcf21* lineage tracing analysis. *Tcf21^{mCrem}:R26R^{tdTom}* timed pregnant females were injected with a single dose of Tamoxifen at E9.5, 10.5, E11.5 or 12.5 and testes were analyzed at E17.5 or 10 weeks. **(B)** The TCF21^{lin}cells at E10.5 (n=17) and E11.5 (n=11) contribute to all major somatic cell populations in the fetal gonad, whereas the E12.5 (n=11) TCF21^{lin} cells give rise only to testis interstitial cells. **(C-F)** Co-immunostaining of fetal or fostered adult TCF21^{lin}testis cross-sections with Sertoli cell marker SOX9**(green; C**,n=5 for E9.5 and E12.5, n=3 for E10.5**),** Leydig cell marker 3β-HSD **(green; D**,n=5 for E9.5, n=3 for E10.5 and E12.5**)**, interstitial cell marker COUPTFII **(NR2F2; green in E**,n=5 for E9.5, E10.5 and E12.5**),** and a myoid cell marker alpha smooth muscle actin **(SMA, green in F**,n=5 for E9.5, n=3 for E10.5, n=4 for E12.5**)**. **(G)** The TCF21^{lin} in the fetal testis of *Tcf21^{mCrem}; R26R^{tdTom}*; Oct4-eGFP embryos are present in both the coelomic epithelium and mesonephros. Colocalization of WT1^{+} **(Green)** cells in the E11.5 gonad with**TCF21^{lin}**cells(n=2). In all panels the nuclear counterstain is DAPI **(white B-G,** n=2**)**. Scale bars for all panels: 20 μm.

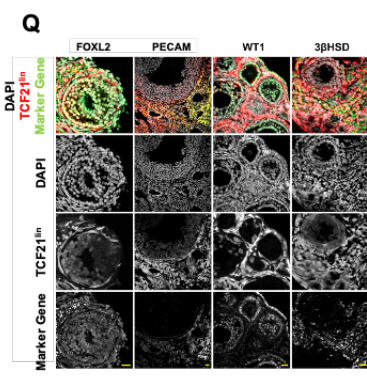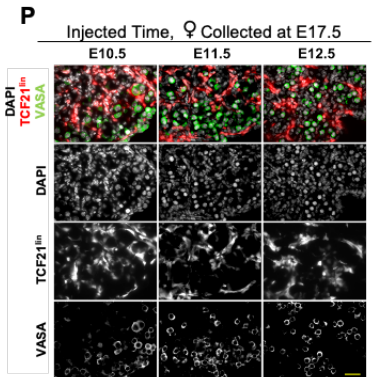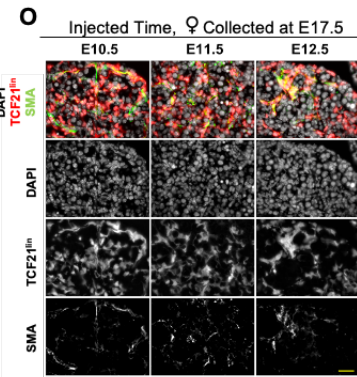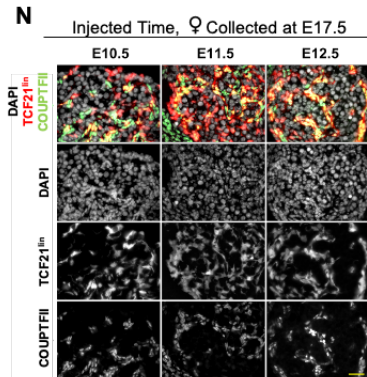Injected Time, ♂ Collected at E17.5

A E10.5 *Tcf21+/+:R26R^tdTom* Tamoxifen | E10.5 *Tcf21^mCrem:R26R^tdTom* Oil

B E9.5 | E10.5 | E11.5 | E12.5

I E10.5

J E10.5

♂ Co-Stain Marker

C SOX9 **D** 3BHSD **E** COUPTFII **F** SMA **G** GLI1 **H** PDGFRA

- % Marker+ Cells Expressing tdTomato
- % tdTomato+ Cells Expressing Marker

K Injected E10.5, ♀ Collected at E11.5

DAPI TCF21^lin OCT4-eGFP WT1

L Injected Time, ♀ Collected at E17.5
E10.5 | E11.5 | E12.5

M Injected Time, ♀ Collected at E17.5
E10.5 | E11.5 | E12.5

N Injected Time, ♀ Collected at E17.5
E10.5 | E11.5 | E12.5

O Injected Time, ♀ Collected at E17.5
E10.5 | E11.5 | E12.5

P Injected Time, ♀ Collected at E17.5
E10.5 | E11.5 | E12.5
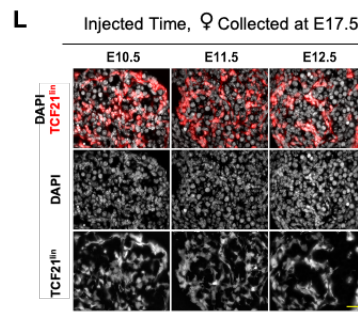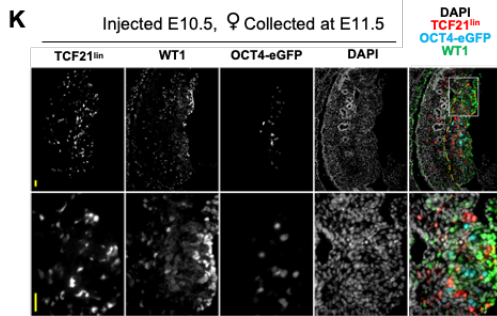
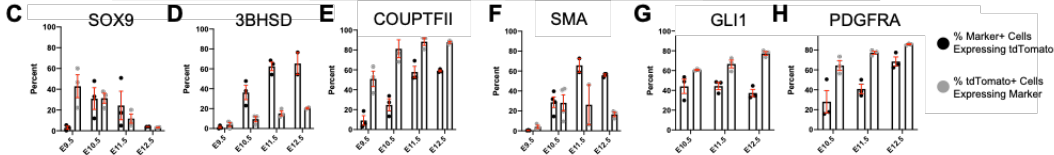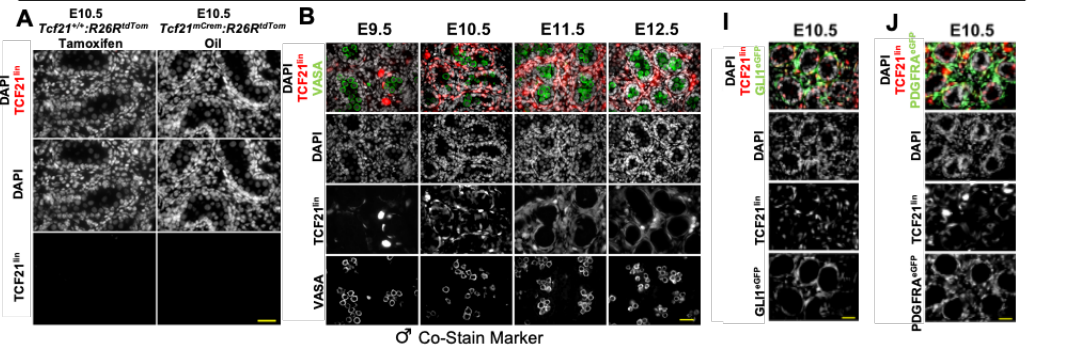**Q** FOXL2 | PECAM | WT1 | 3βHSD

165

Figure A-8: The fetal TCF21^{lin} population is a bipotential somatic progenitor. (A) *Tcf21$^{+/+}$:R26R^{tdTom}*or *Tcf21^{mCrem}:R26R^{tdTom}* time pregnant females (E10.5) were treated with Tamoxifen (left panels, n=3) or corn oil (right panels, n=3), respectively, confirming specificity and tightness of Cre expression. **(B)** Co-immunostaining of the TCF21^{lin} with germ cell marker VASA **(DDX4; green)** in the male fetal testis (collected at different injection timepoints, n=2 for E9.5, n=3 for E10.5, n=2 for E11.5, n=2 for E12.5). **(C-H)** Quantification of the percentage of various testis somatic cells co-labeled with TCF21^{lin} (black bars) and percentage of cells in the TCF21^{lin} population co-expressing testis somatic markers (gray bars) at different embryonic days (n=2 for E12.5 **C,D,** E11.5 **F**; n=3 for E9.5 **C-F**, E10.5 and E11.5 **C-E,G,H**, E12.5 **E-H**; n=4 for E10.5 **F**) or (n=3 for all, except n=2 for E12.5 **C,D,** E11.5 **F**;n=4 for E10.5 **F**); A minimum of 500 cells were counted per marker per timepoint. Data are presented as mean ±SEM. **(I, J)** Colocalization of TCF21^{lin}and GLI1-egfp **(I, n=6)**, PDGFRA^{eGFP}**(J, n=7)**, and the nuclear counterstain DAPI. **(K)** TCF21^{lin} cells in the E11.5 fetal ovary are present in the coelomic epithelium and mesonephros (n= 1 whole ovary). Overlaying TCF21^{lin} (Tdtom$^+$) cells with WT1$^+$ cells **(green; (L)** E10.5-E12.5 Tcf21^{lin} cells label somatic cells broadly in the fetal ovary at E17.5 (n=12 for E10.5, n=7 for E11.5, n=6 for E12.5). **(M-P)** Overlay of the TCF21^{lin} with granulosa cell marker FOXL2 **(green; M,** n=2 for E10.5, n=3 for E11.5 and E12.5**)**, the interstitial cell marker COUPTFII (**Nr2F2, green; N,** n=3 for E10.5, n=2 for E11.5, and n=1 for E12.5); the smooth muscle marker SMA **(green; O,** n=2 for E10.5, n=4 for E11.5, n=3 for E12.5**)**; the germ cell marker VASA**(DDX4, green; P.** n=3 for E10.5 and 11.5, n=2 for E12.5**)**. **(Q)** The E10.5 fetal TCF21^{lin} contributes to multiple somatic cell types in the adult ovary (n=4 for FOXL2, PECAM, and 3BHSD, n=3 for WT1). In all panels the nuclear counterstain is DAPI. Scale bars for all panels: 20 μm.
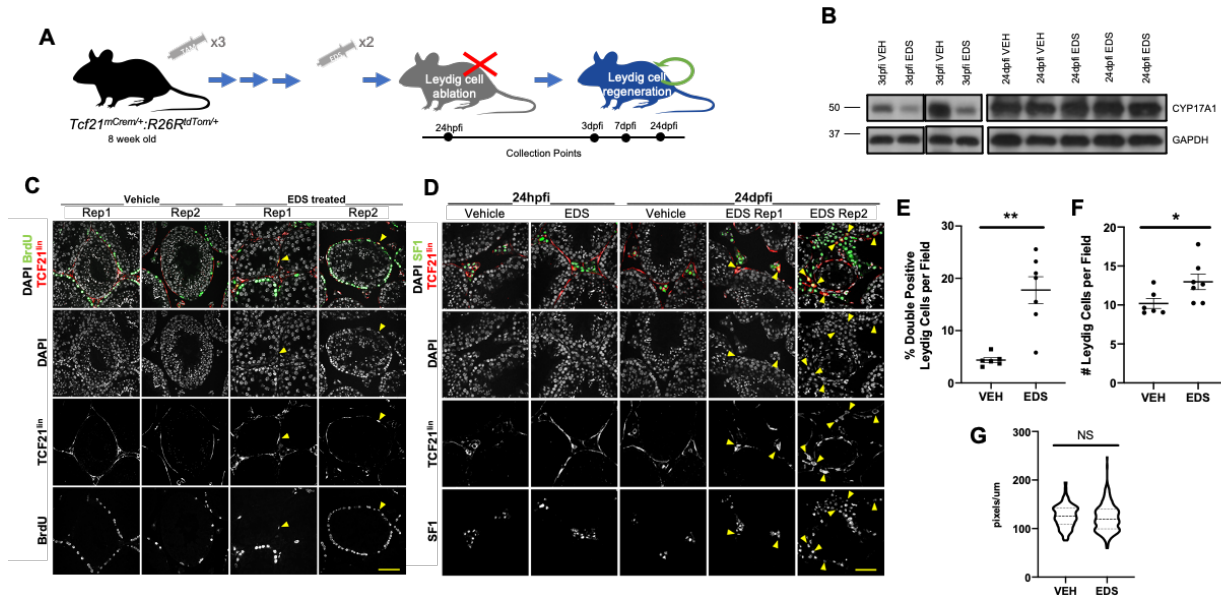
Figure A-9: The TCF21^{lin} population regenerates Leydig cells in vivo after injury. (A) Schematic representation of the experimental method used to ablate Leydig cells in the *Tcf21^{mCrem}:R26R^{tdTom}* background. **(B)** Representative CYP17A1 protein immunoblots from EDS and vehicle treated animals at 3dpfi and 24dpfi. A total of n=8 independent experiments were performed. **(C)** Representative images of BrdU$^+$ cells in the testis of vehicle and EDS treated animals (representative from n=2 vehicle, n=3 EDS). **(D)** Colocalization of TCF21^{lin}(tdTom) and Leydig cell marker SF1 in EDS or vehicle-treated *Tcf21^{mCrem}:R26R^{tdTom}* animals at 24 hours post final injection (24hpfi) or 24 days post final injection (24dpfi). Representative images from n=6 vehicle, n=7 EDS animals. **(E)** Quantification of the percentage of Leydig cells expressing both SF1 and TCF21^{lin}(tdTom) per field at 24dpfi (Note- each circle represents the percentage per animal. A total of n=6 vehicle, n=7 EDS were analyzed). Data are presented as mean ±SEM. P= 0.0018. **(F)** Quantification of Leydig cell number per field of view at 24dpfi (n=6 vehicle, n=7 EDS). Data are presented as mean ±SEM. P= 0.042. **(G)** Average Leydig cell diameter measurements in *Tcf21^{mCrem}:R26R^{tdTom}* EDS or vehicle injected animals, after Leydig cell recovery at 24dpfi (n=3 per condition). Lines indicate mean and quartiles. P=0.092. All statistical tests were performed using Welch's unpaired, two-sided t-tests. Scale bars: 20 μm.
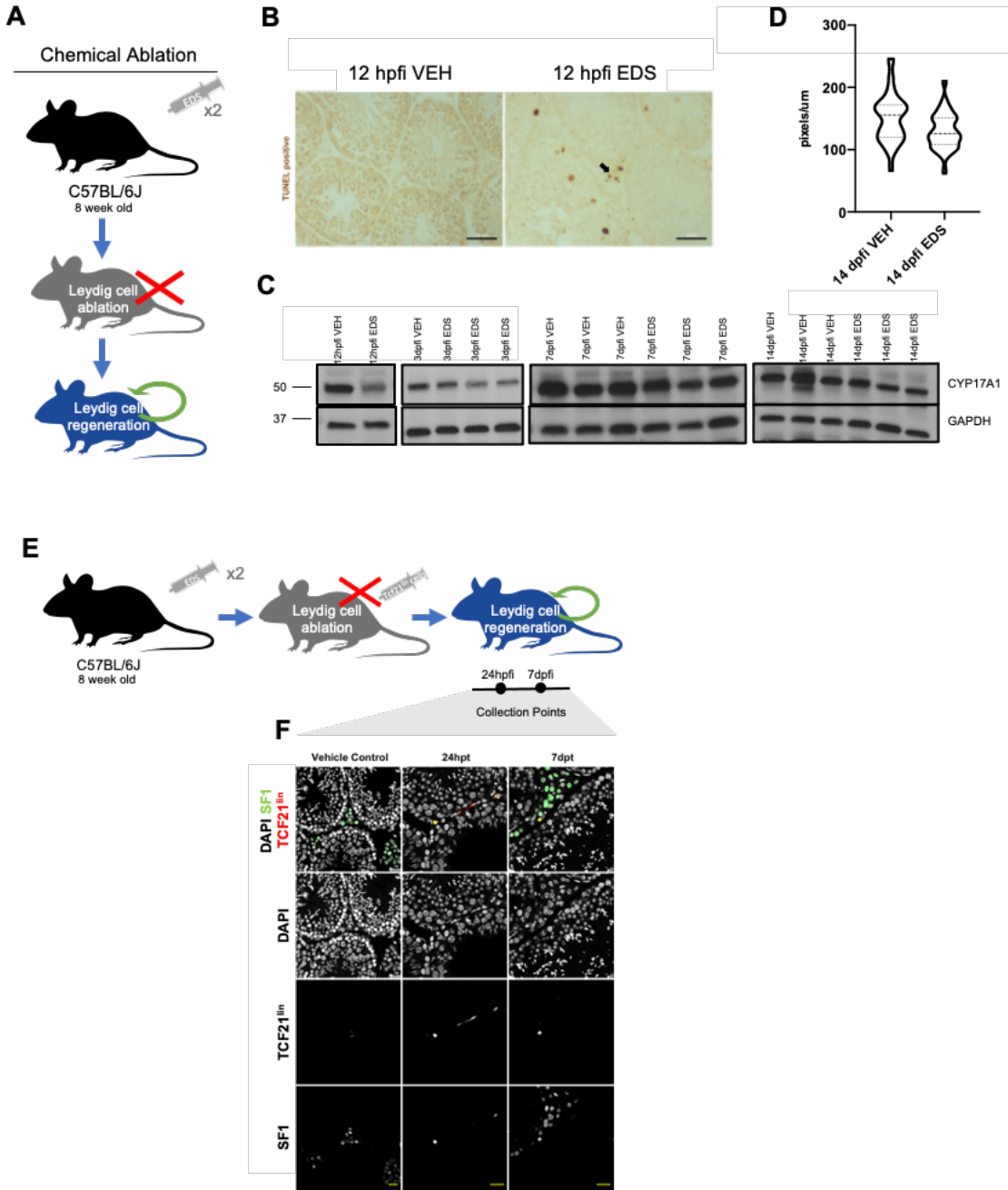
A Chemical Ablation

C57BL/6J
8 week old

Leydig cell ablation

Leydig cell regeneration

B

12 hpfi VEH   12 hpfi EDS

TUNEL positive

C

| | 12hpfi VEH | 12dpfi EDS | 3dpfi VEH | 3dpfi EDS | 3dpfi EDS | 3dpfi EDS | 7dpfi VEH | 7dpfi VEH | 7dpfi VEH | 7dpfi EDS | 7dpfi EDS | 7dpfi EDS | 14dpfi VEH | 14dpfi VEH | 14dpfi VEH | 14dpfi EDS | 14dpfi EDS | 14dpfi EDS |

CYP17A1

GAPDH

50 —
37 —

D

300

pixels/um

200

100

0

14 dpfi VEH   14 dpfi EDS

E

C57BL/6J
8 week old

Leydig cell ablation

Leydig cell regeneration

24hpfi   7dpfi

Collection Points

F

| | Vehicle Control | 24hpt | 7dpt |

DAPI SF1 TCF21lin

DAPI

TCF21lin

SF1

Figure A-10: Ethane Dimethane Sulfonate (EDS) treatment ablates Leydig cells in the adult testis. (A) Schematic representation of the chemical method used to ablate Leydig cells. **(B)** Representative images of apoptotic Leydig cells by TUNEL staining in adultC57BL/6 testes collected 12hpfi of EDS (vehicle n=3, EDS n=4) **(C)** Western blot illustrating CYP17A1protein expression levels at 12hpfi, 3dpfi, 7dpfi, and 14dpfi in C57BL/6 animals after EDS or vehicle (VEH) treatment (representative of n=3-4 per condition per timepoint). **(D)** Average Leydig cell diameter measurements 14dpfi of EDS or vehicle in C57BL/6 animals, after Leydig cell recovery (n=3).Lines indicate mean and quartiles.**(E)** Schematic representation of transplantation of TCF21^{lin}cells following chemical ablation of Leydig cells in C57BL/6 animals. **(F)** Representative images of EDS treated C57BL/6 testes collected 24 hours and 7 days after transplant of SCA1^{+}/cKIT^{-} cells from *Tcf21^{mCrem}:R26R^{tdTom}*animals, co-immunostained with SF1/TdTom (24 hrs n=4. 7 days n=6 per condition) Scale bar: 20 μm.

A *Tcf21^mCrem/+:R26R^tdTom/tdTom* 8 weeks — normal physiological aging — 1 year

B Injection Time: 8 Weeks, Collected at:

C % Double Positive Leydig Cells per Field

D # Leydig Cells per Field

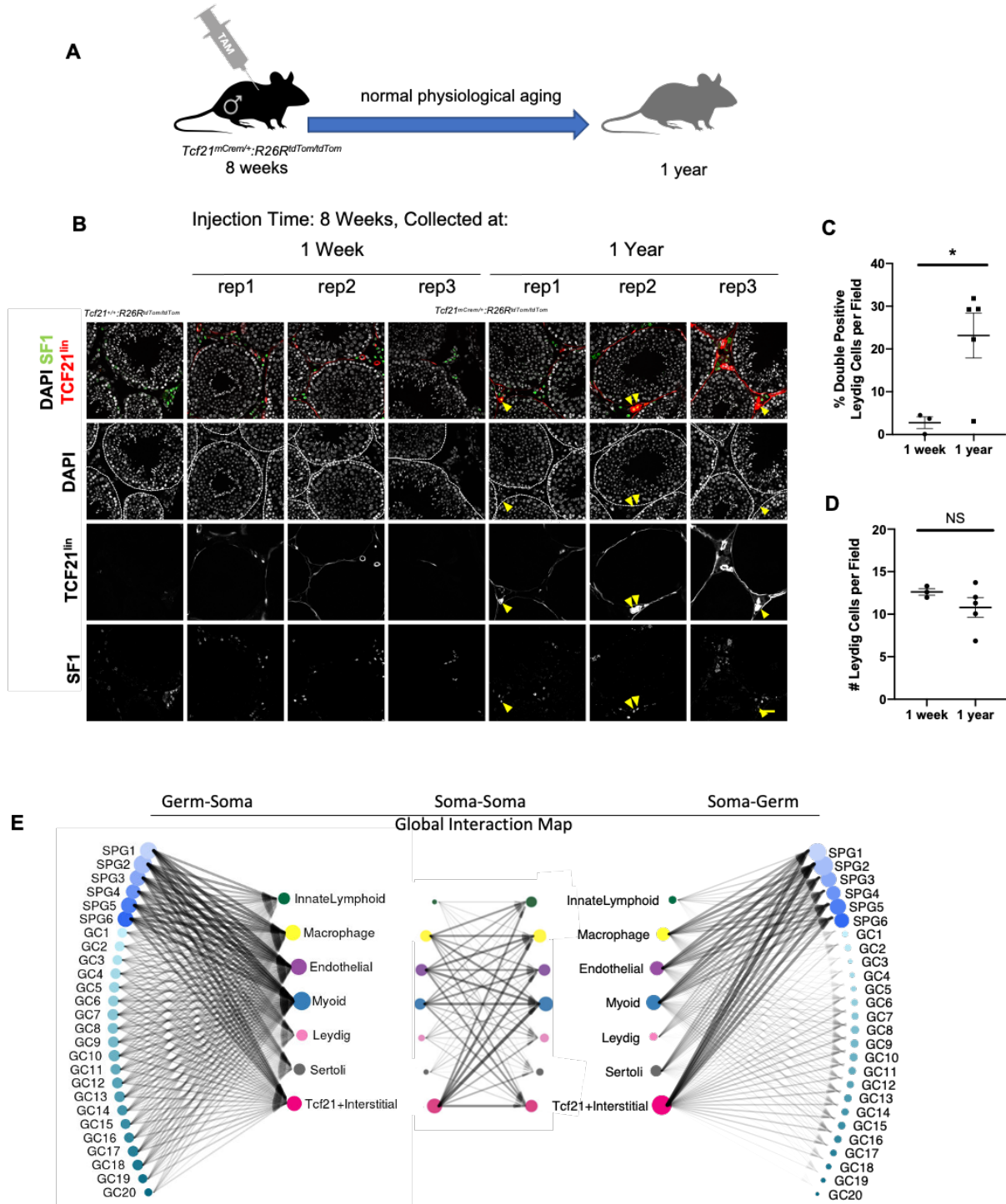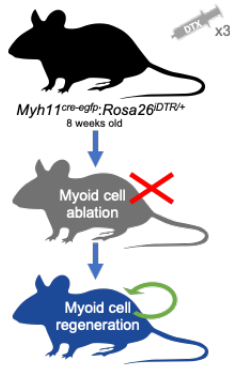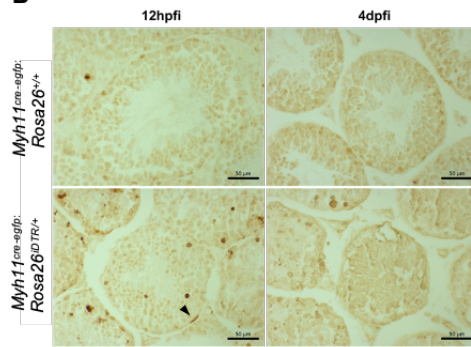**E** Global Interaction Map — Germ-Soma, Soma-Soma, Soma-Germ

Figure A-11: The TCF21^{lin} population maintains testis tissue homeostasis during aging.
(A) Schematic representation of TCF21lineage tracing in the aging testis. **(B)** Colocalization of TCF21^{lin} (tdTom) and Leydig cell marker SF1 (green) at one week or one year following a single injection of tamoxifen in 8 week-old adults. Scale bars: 20 μm. DAPI is used as the nuclear counterstain (white). (Oil only control n=2, demonstrating the tight control of TCF21-Cre, n=3 for 1 week, n=5 for 1 year) **(C)** Quantification of the percent of Leydig cells expressing both SF1 and TCF21^{lin}(tdTom) per field (Note- each circle/square represents the percentage per animal; n=3 for 1 week, n=5 for 1 year). P= 0.016. Data are presented as mean ± SEM. All statistical tests were performed using Welch's unpaired, two-sided t-tests. **(D)** Quantification of Leydig cell number per field (Note- each circle / square represents the percentage per animal; n=3 for 1 week, n=5 for 1 year). P=0.19. Data are presented as mean ± SEM. All statistical tests were performed using Welch's unpaired, two-sided t-tests.**(E)** Summary of putative ligand – receptor interactions in the mouse testis between the germline and soma (left), within soma (center), and soma and germline (right). Arrows summarize top 5% of all interactions. ScRNA-seq data from [42], processed as described in[34]. Symbol size indicates the number of receptor-ligand interactions contributed by a cell type and line width shows the number of interactions between the two cell types.
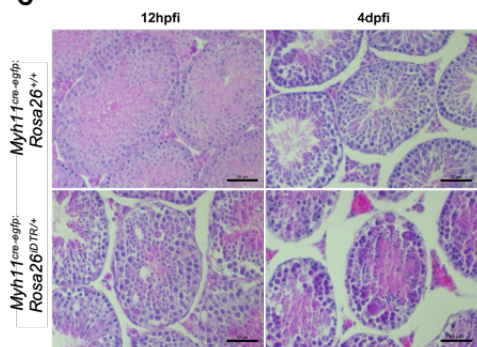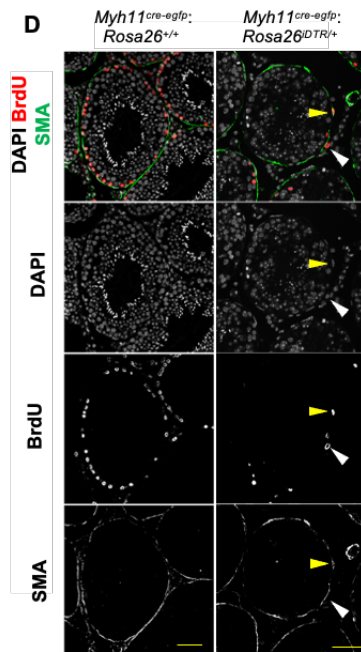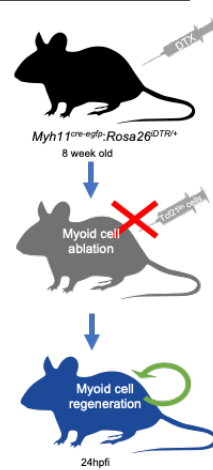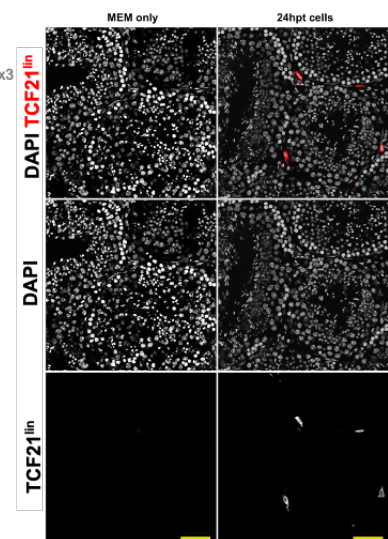
A Genetic Ablation

B 12hpfi 4dpfi

C 12hpfi 4dpfi

D

E Transplantation

F MEM only 24hpt cells

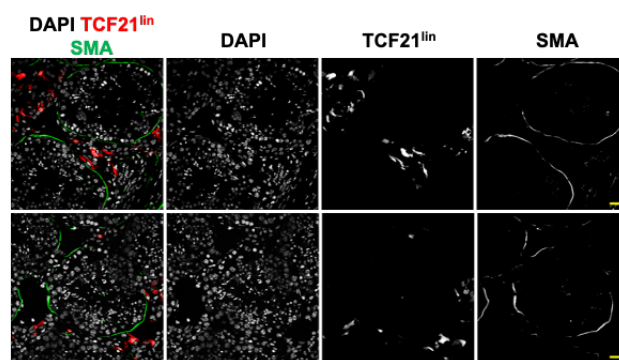G DAPI TCF21lin SMA | DAPI | TCF21lin | SMA

172

Figure A-12: Peritubular myoid cells can regenerate after Diphtheria toxin treatment and transplanted TCF21^{lin} cells home to the seminiferous tubule basement membrane after injury. (A) Schematic representation of the genetic method used to ablate myoid cells. **(B-E)**Efficacy of diphtheria toxin treatment on *Myh11$^{cre-egfp}$:Rosa26$^{iDTR/+}$*or *Myh11$^{cre-egfp}$:Rosa26$^{+/+}$* animals. **(B)** Representative images of apoptotic myoid cell death by TUNEL staining in testes collected 12hpfi (n=4 per genotype)or 4dpfi (n=2 per genotype) of DTX. Scale bar: 50 μm. **(C)** Representative H&E staining at 12hpfi (n=3) or 4dpfi (n=2) of DTX. Scale bar: 100 μm**. (D)** Representative images of BrdU^{+}/SMA^{+} (yellow arrows) or BrdU^{+}/SMA^{-} (white arrows)at 4dpfi in *Myh11$^{cre-egfp}$:Rosa26$^{+/+}$*(n=3) or *Myh11$^{cre-egfp}$:Rosa26$^{iDTR/+}$*(n=2)animals.Scale bar: 20 μm**. (E)**Schematic representation of the TCF21^{lin}cell transplant following genetic ablation of myoid cells in *Myh11$^{cre-egfp}$:Rosa26$^{iDTR/+}$*animals. **(F)** Representative immunofluorescence images of TCF21^{lin} in *Myh11$^{cre-egfp}$:Rosa26$^{iDTR/+}$*animals at 24hpt (n=1 for MEM vehicle, n=3 for cells). **(G)** Two representative immunofluorescence images of TCF21^{lin} (red) and SMA (green) in *Myh11$^{cre-egfp}$:Rosa26$^{iDTR/+}$* animals at 24hpt (n=3). Scale bar: 20 μm.
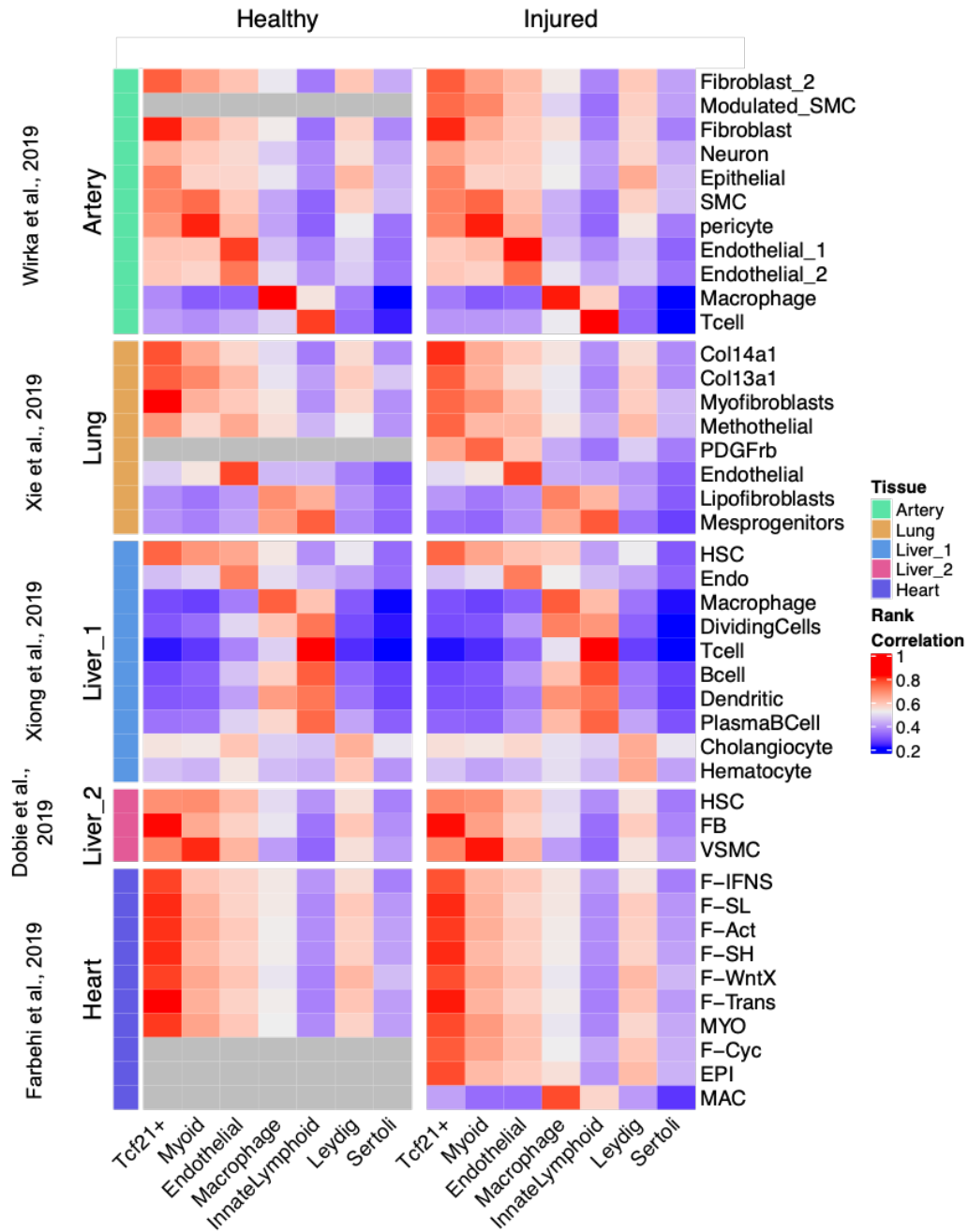
Figure A-13: The adult *Tcf21⁺* interstitial population resembles fibroblasts in other tissues. Rank correlation of scRNA-seq cluster centroids of somatic cells from wild type adult mouse testes with other tissues in healthy adult mice (left panels) or after injury (right panels). Grey blocks indicate cell types not present in healthy datasets.

A.7 References

1.   Lin, Y.-T., Barske, L., DeFalco, T. & Capel, B. Numb regulates somatic cell lineage commitment during early gonadogenesis in mice. *Development* (2017) doi:10.1242/dev.149203.
2.   Ikeda, Y., Shen, W. H., Ingraham, H. A. & Parker, K. L. Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases. *Molecular Endocrinology* **8**, (1994).
3.   Morohashi, K. The ontogenesis of the steroidogenic tissues. *Genes to Cells* **2**, (1997).
4.   Hatano, O., Takakusu, A., Nomura, M. & Morohashi, K. Identical origin of adrenal cortex and gonad revealed by expression profiles of Ad4BP/SF-1. *Genes to Cells* **1**, (1996).
5.   Karl, J. & Capel, B. Sertoli Cells of the Mouse Testis Originate from the Coelomic Epithelium. *Developmental Biology* **203**, (1998).
6.   Martineau, J., Nordqvist, K., Tilmann, C., Lovell-Badge, R. & Capel, B. Male-specific cell migration into the developing gonad. *Current Biology* **7**, (1997).
7.   Tilmann, C. & Capel, B. Mesonephric cell migration induces testis cord formation and Sertoli cell differentiation in the mammalian gonad. *Development (Cambridge, England)* **126**, (1999).
8.   Buehr, M., McLaren, A., Bartley, A. & Darling, S. Proliferation and migration of primordial germ cells in We/We mouse embryos. *Developmental Dynamics* **198**, (1993).
9.   DeFalco, T., Takahashi, S. & Capel, B. Two distinct origins for Leydig cell progenitors in the fetal testis. *Developmental Biology* **352**, (2011).
10.   Cederroth, C. R., Pitetti, J.-L., Papaioannou, M. D. & Nef, S. Genetic programs that regulate testicular and ovarian development. *Molecular and Cellular Endocrinology* **265–266**, (2007).
11.   Rotgers, E., Jørgensen, A. & Yao, H. H.-C. At the Crossroads of Fate—Somatic Cell Lineage Specification in the Fetal Gonad. *Endocrine Reviews* **39**, (2018).
12.   Brennan, J. & Capel, B. One tissue, two fates: molecular genetic events that underlie testis versus ovary development. *Nature Reviews Genetics* **5**, (2004).
13.   Green, C. D. *et al.* A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental Cell* **46**, (2018).
14.   Bhang, D. H. *et al.* Testicular endothelial cells are a critical population in the germline stem cell niche. *Nature Communications* **9**, (2018).
15.   Kitadate, Y. *et al.* Competition for Mitogens Regulates Spermatogenic Stem Cell Homeostasis in an Open Niche. *Cell Stem Cell* **24**, (2019).
16.   DeFalco, T. *et al.* Macrophages Contribute to the Spermatogonial Niche in the Adult Testis. *Cell Reports* **12**, (2015).
17.   Smith, L. B., O'Shaughnessy, P. J. & Rebourcet, D. Cell-specific ablation in the testis: what have we learned? *Andrology* **3**, (2015).
18.   Jiang, M. H. *et al.* Characterization of Nestin-positive stem Leydig cells as a potential source for the treatment of testicular Leydig cell dysfunction. *Cell Research* **24**, (2014).
19.   Davidoff, M. S. *et al.* Progenitor cells of the testosterone-producing Leydig cells revealed. *Journal of Cell Biology* **167**, (2004).
20.   Kilcoyne, K. R. *et al.* Fetal programming of adult Leydig cell function by androgenic effects on stem/progenitor cells. *Proceedings of the National Academy of Sciences* **111**, (2014).

21.   Qin, J., Tsai, M.-J. & Tsai, S. Y. Essential Roles of COUP-TFII in Leydig Cell Differentiation and Male Fertility. *PLoS ONE* **3**, (2008).

22.   Ge, R.-S. *et al.* In search of rat stem Leydig cells: Identification, isolation, and lineage-specific development. *Proceedings of the National Academy of Sciences* **103**, (2006).

23.   Landreh, L., Stukenborg, J.-B., Söder, O. & Svechnikov, K. Phenotype and steroidogenic potential of PDGFRα-positive rat neonatal peritubular cells. *Molecular and Cellular Endocrinology* **372**, (2013).

24.   Chen, H., Huhtaniemi, I. & Zirkin, B. R. Depletion and repopulation of Leydig cells in the testes of aging brown Norway rats. *Endocrinology* **137**, (1996).

25.   Chen, H., Stanley, E., Jin, S. & Zirkin, B. R. Stem Leydig cells: From fetal to aged animals. *Birth Defects Research Part C: Embryo Today: Reviews* **90**, (2010).

26.   Kumar, D. L. & DeFalco, T. A perivascular niche for multipotent progenitors in the fetal testis. *Nature Communications* **9**, (2018).

27.   Lu, J. *et al.* The basic helix-loop-helix transcription factor capsulin controls spleen organogenesis. *Proceedings of the National Academy of Sciences* **97**, (2000).

28.   Lu, J. *et al.* Control of Facial Muscle Development by MyoR and Capsulin. *Science* **298**, (2002).

29.   Quaggin, S. E. *et al.* The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. *Development (Cambridge, England)* **126**, (1999).

30.   Cui, S. *et al.* Disrupted gonadogenesis and male-to-female sex reversal in *Pod1* knockout mice. *Development* **131**, (2004).

31.   Bhandari, R. K., Sadler-Riggleman, I., Clement, T. M. & Skinner, M. K. Basic Helix-Loop-Helix Transcription Factor TCF21 Is a Downstream Target of the Male Sex Determining Gene SRY. *PLoS ONE* **6**, (2011).

32.   Stévant, I. *et al.* Dissecting Cell Lineage Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell Transcriptomics. *Cell Reports* **26**, (2019).

33.   Stévant, I. *et al.* Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing. *Cell Reports* **22**, (2018).

34.   Shami, A. N. *et al.* Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Developmental Cell* **54**, (2020).

35.   Hynes, R. O. The Extracellular Matrix: Not Just Pretty Fibrils. *Science* **326**, (2009).

36.   Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nature Reviews Molecular Cell Biology* **15**, (2014).

37.   Acharya, A., Baek, S. T., Banfi, S., Eskiocak, B. & Tallquist, M. D. Efficient inducible Cre-mediated recombination in Tcf21cell lineages in the heart and kidney. *genesis* **49**, (2011).

38.   Uccelli, A., Moretta, L. & Pistoia, V. Mesenchymal stem cells in health and disease. *Nature Reviews Immunology* **8**, (2008).

39.   Gonzalez, R. *et al.* A putative mesenchymal stem cells population isolated from adult human testes. *Biochemical and Biophysical Research Communications* **385**, (2009).

40.   Ahmed, M., Ghabriel, M. & Amleh, A. Enrichment, Propagation, and Characterization of Mouse Testis-Derived Mesenchymal Stromal Cells. *Cellular Reprogramming* **19**, (2017).

41.   Eliveld, J. *et al.* Primary human testicular PDGFRα+ cells are multipotent and can be differentiated into cells with Leydig cell characteristics in vitro. *Human Reproduction* **34**, (2019).

42.	Li, X. *et al.* Regulation of seminiferous tubule-associated stem Leydig cells in adult rat testes. *Proceedings of the National Academy of Sciences* **113**, (2016).

43.	Clark, A. M., Garland, K. K. & Russell, L. D. Desert hedgehog (Dhh) Gene Is Required in the Mouse Testis for Formation of Adult-Type Leydig Cells and Normal Development of Peritubular Cells and Seminiferous Tubules. *Biology of Reproduction* **63**, (2000).

44.	Tang, H. *et al.* Notch signaling maintains Leydig progenitor cells in the mouse testis. *Development* **135**, (2008).

45.	Gao, F. *et al.* Autophagy regulates testosterone synthesis by facilitating cholesterol uptake in Leydig cells. *Journal of Cell Biology* **217**, (2018).

46.	Steinfeld, K. *et al.* Low testosterone in ApoE/LDL receptor double-knockout mice is associated with rarefied testicular capillaries together with fewer and smaller Leydig cells. *Scientific Reports* **8**, (2018).

47.	Liu, C., Rodriguez, K. & Yao, H. H.-C. Mapping lineage progression of somatic progenitor cells in the mouse fetal testis. *Development* (2016) doi:10.1242/dev.135756.

48.	Koopman, P., Münsterberg, A., Capel, B., Vivian, N. & Lovell-Badge, R. Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* **348**, (1990).

49.	Matson, C. K. *et al.* DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* **476**, (2011).

50.	Uhlenhaut, N. H. *et al.* Somatic Sex Reprogramming of Adult Ovaries to Testes by FOXL2 Ablation. *Cell* **139**, (2009).

51.	Matson, C. K. & Zarkower, D. Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nature Reviews Genetics* **13**, (2012).

52.	Edson, M. A., Nagaraja, A. K. & Matzuk, M. M. The Mammalian Ovary from Genesis to Revelation. *Endocrine Reviews* **30**, (2009).

53.	Honda, A. *et al.* Isolation, characterization, and in vitro and in vivo differentiation of putative thecal stem cells. *Proceedings of the National Academy of Sciences* **104**, (2007).

54.	Liu, C., Peng, J., Matzuk, M. M. & Yao, H. H.-C. Lineage specification of ovarian theca cells requires multicellular interactions via oocyte and granulosa cells. *Nature Communications* **6**, (2015).

55.	Kerr, J. B., Donachie, K. & Rommerts, F. F. G. Selective destruction and regeneration of rat Leydig cells in vivo. *Cell and Tissue Research* **242**, (1985).

56.	MOLENAAR, R., ROOIJ, D. G. de, ROMMERTS, F. F. G. & MOLEN, H. J. van der. Repopulation of Leydig Cells in Mature Rats after Selective Destruction of the Existent Leydig Cells with Ethylene Dimethane Sulfonate Is Dependent on Luteinizing Hormone and Not Follicle-Stimulating Hormone. *Endocrinology* **118**, (1986).

57.	MORRIS, I. D., PHILLIPS, D. M. & BARDIN, C. W. Ethylene Dimethanesulfonate Destroys Leydig Cells in the Rat Testis*. *Endocrinology* **118**, (1986).

58.	JACKSON, H. ANTISPERMATOGENIC AGENTS. *British Medical Bulletin* **26**, (1970).

59.	Ewing, L. & Keeney DS. Leydig cell: structure and function. *Cell and Molecular Biology of the Testis* 137–165 (1993).

60.	Yokonishi, T., McKey, J., Ide, S. & Capel, B. Sertoli cell ablation and replacement of the spermatogonial niche in mouse. *Nature Communications* **11**, (2020).

61.	Rebourcet, D. *et al.* Sertoli Cells Maintain Leydig Cell Number and Peritubular Myoid Cell Activity in the Adult Mouse Testis. *PLoS ONE* **9**, (2014).

62.	Rebourcet, D. *et al.* Sertoli cells control peritubular myoid cell fate and support adult Leydig cell development in the prepubertal testis. *Development* **141**, (2014).

63. Teerds, K. J., de Rooij, D. G., Rommerts, F. F. G., van der Tweel, I. & Wensing, C. J. G. Turnover Time of Leydig Cells and Other Interstitial Cells in Testes of Adult Rats. *Archives of Andrology* **23**, (1989).

64. May, P. & Herz, J. LDL Receptor-Related Proteins in Neurodevelopment. *Traffic* **4**, (2003).

65. LILLIS, A. P., MIKHAILENKO, I. & STRICKLAND, D. K. Beyond endocytosis: LRP function in cell migration, proliferation and vascular permeability. *Journal of Thrombosis and Haemostasis* **3**, (2005).

66. Li, Y.-H., Kuo, C.-H., Shi, G.-Y. & Wu, H.-L. The role of thrombomodulin lectin-like domain in inflammation. *Journal of Biomedical Science* **19**, (2012).

67. Mitchell, J. L. & Mutch, N. J. Let's cross-link: diverse functions of the promiscuous cellular transglutaminase factor XIII-A. *Journal of Thrombosis and Haemostasis* **17**, (2019).

68. Robb, L. *et al.* epicardin: A novel basic helix-loop-helix transcription factor gene expressed in epicardium, branchial arch myoblasts, and mesenchyme of developing lung, gut, kidney, and gonads. *Developmental Dynamics* **213**, (1998).

69. Xie, T. *et al.* Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis. *Cell Reports* **22**, (2018).

70. Wirka, R. C. *et al.* Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nature Medicine* **25**, (2019).

71. Farbehi, N. *et al.* Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *eLife* **8**, (2019).

72. Dobie, R. *et al.* Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme during Liver Fibrosis. *Cell Reports* **29**, (2019).

73. Xiong, X. *et al.* Landscape of Intercellular Crosstalk in Healthy and NASH Liver Revealed by Single-Cell Secretome Gene Analysis. *Molecular Cell* **75**, (2019).

74. Swonger, J. M., Liu, J. S., Ivey, M. J. & Tallquist, M. D. Genetic tools for identifying and manipulating fibroblasts in the mouse. *Differentiation* **92**, (2016).

75. Barsoum, I. B., Bingham, N. C., Parker, K. L., Jorgensen, J. S. & Yao, H. H.-C. Activation of the Hedgehog pathway in the mouse fetal ovary leads to ectopic appearance of fetal Leydig cells and female pseudohermaphroditism. *Developmental Biology* **329**, (2009).

76. Brennan, J. Pdgfr-alpha mediates testis cord organization and fetal Leydig cell development in the XY gonad. *Genes & Development* **17**, (2003).

77. Dominici, M. *et al.* Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **8**, (2006).

78. Nauta, A. J. & Fibbe, W. E. Immunomodulatory properties of mesenchymal stromal cells. *Blood* **110**, (2007).

79. Yeap, B. B., Page, S. T. & Grossmann, M. Testosterone treatment in older men: clinical implications and unresolved questions from the Testosterone Trials. *The Lancet Diabetes & Endocrinology* **6**, (2018).

80. Baburski, A. Z., Andric, S. A. & Kostic, T. S. Luteinizing hormone signaling is involved in synchronization of Leydig cell's clock and is crucial for rhythm robustness of testosterone production†. *Biology of Reproduction* **100**, (2019).

81.   Patel, A. S., Leong, J. Y., Ramos, L. & Ramasamy, R. Testosterone Is a Contraceptive and Should Not Be Used in Men Who Desire Fertility. *The World Journal of Men's Health* **37**, (2019).

82.   Stanley, E. *et al.* Identification, Proliferation, and Differentiation of Adult Leydig Stem Cells. *Endocrinology* **153**, (2012).

83.   Ge, R.-S. *et al.* Gene Expression in Rat Leydig Cells During Development from the Progenitor to Adult Stage: A Cluster Analysis1. *Biology of Reproduction* **72**, (2005).

84.   Chen, P., Zirkin, B. R. & Chen, H. Stem Leydig Cells in the Adult Testis: Characterization, Regulation and Potential Applications. *Endocrine Reviews* **41**, (2020).

85.   Ye, L., Li, X., Li, L., Chen, H. & Ge, R.-S. Insights into the Development of the Adult Leydig Cell Lineage from Stem Leydig Cells. *Frontiers in Physiology* **8**, (2017).

86.   Chen, H., Wang, Y., Ge, R. & Zirkin, B. R. Leydig cell stem cells: Identification, proliferation and differentiation. *Molecular and Cellular Endocrinology* **445**, (2017).

87.   Maezawa, Y. *et al.* Loss of the Podocyte-Expressed Transcription Factor *Tcf21/Pod1* Results in Podocyte Differentiation Defects and FSGS. *Journal of the American Society of Nephrology* **25**, (2014).

88.   Kikuchi, K. *et al.* Epicardial cells adopt non-myocardial fates during zebrafish heart development and regeneration. *Development* **138**, (2011).

89.   Adashi EY, Rock JA & Rosenwaks Z. Reproductive endocrinology, surgery, and technology. (1996).

90.   Meineke, V., Frungieri, M. B., Jessberger, B., Vogt, H.-J. & Mayerhofer, A. Human testicular mast cells contain tryptase: increased mast cell number and altered distribution in the testes of infertile men. *Fertility and Sterility* **74**, (2000).

91.   Frungieri, M. B. *et al.* Number, distribution pattern, and identification of macrophages in the testes of infertile men. *Fertility and Sterility* **78**, (2002).

92.   Frungieri, M. B., Weidinger, S., Meineke, V., Kohn, F. M. & Mayerhofer, A. Proliferative action of mast-cell tryptase is mediated by PAR2, COX2, prostaglandins, and PPAR : Possible relevance to human fibrotic disorders. *Proceedings of the National Academy of Sciences* **99**, (2002).

93.   Mayer, C. *et al.* Sterile inflammation as a factor in human male infertility: Involvement of Toll like receptor 2, biglycan and peritubular cells. *Scientific Reports* **6**, (2016).

94.   Li, L. *et al.* Directing differentiation of human induced pluripotent stem cells toward androgen-producing Leydig cells rather than adrenal cells. *Proceedings of the National Academy of Sciences* **116**, (2019).

95.   Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* **47**, (2019).

96.   Rux, D. R. *et al.* Regionally Restricted Hox Function in Adult Bone Marrow Multipotent Mesenchymal Stem/Stromal Cells. *Developmental Cell* **39**, (2016).

97.   Park, D. *et al.* Endogenous Bone Marrow MSCs Are Dynamic, Fate-Restricted Participants in Bone Maintenance and Regeneration. *Cell Stem Cell* **10**, (2012).

98.   Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, (2015).

99.   Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166**, (2016).

100. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, (2013).