Regulating Transcription, Epigenetics, and Stemness During Embryonic Male Germ Cell Development

Myles Austin Bartholomew

B.S., Xavier University of Louisiana, 2020

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Division of Biology of Medicine at Brown University:

Molecular Biology, Cell Biology, and Biochemistry

Providence, Rhode Island

May 2026

Signature Page

This dissertation by Myles A Bartholomew is accepted in its present form by the Department of Molecular Biology, Cell Biology, and Biochemistry as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Richard N. Freiman, Ph.D., Advisor

Recommended to the Graduate Council:

Date _____ _____

Mark Johnson, Ph.D., Chair

Date _____ _____

Alison DeLong Ph.D., Reader

Date _____ _____

Nicola Neretti Ph.D., Reader

Date _____ _____

Diana Laird Ph.D., Reader, UCSF

Approved by the Graduate Council:

Date _____ _____

David P. Lindstrom Ph.D., Dean of the Graduate School

Myles Bartholomew

Curriculum Vitae

myles_bartholomew@brown.edu

Education

Brown University, *Ph.D. in Molecular Bio, Cell Bio, & Biochemistry (MCB)* **May 2026**
GPA: 4.0/4.0, T32 training grant Awardee, 2-time department award winner
NSF GRFP for work in cell bio and genomics, 2022 (3 years full funding; top 12% of applicants)
HHMI Gilliam Fellowship for work in stem cell and disease, 2022 (3 years full funding; top 5% of applicants)

The Xavier University of Louisiana, *B.S. in Biology Minor in Chemistry, cum laude* **May 2020**
Student Government Association (SGA) President | Dean's List
Received the highest honor awarded to one graduate for academic and leadership excellence (1 in 370)
Amgen Scholar |NIH BUILD and LSAMP Scholar | HHMI Scholar | Leadership Alliance | McNair | TRIO program

Publications

Bartholomew MA, Kelsey M, Abt KM, Nixon AEK, Seymour KA, Freiman RN. Regulation of genomic integrity and novel spermatogonial lineage dynamics. *In preparation.*

Gura MA*, **Bartholomew MA***, Abt KM, Relovská S, Seymour KA, Freiman RN. Transcription and chromatin regulation by TAF4b during cellular quiescence of developing prospermatogonia. Front Cell Dev Biol. 2023Oct;12(11):1270408. *Co-first authorship

Abt KM, **Bartholomew MA,** Nixon AEK, Richman HE, Gura MA, Seymour KA, & Freiman RN. (2026). Transcriptional Integration of Meiotic Prophase I Progression and Early Oocyte Differentiation (*In press*) Development

Research Experience

Molecular Bio, Cell Bio, & Biochemistry | Brown University, HHMI & NSF Fellow **August 2020–Present**

- Employ multi-omics (Bulk RNA-Seq, scRNA-Seq, DNA methylation, CUT&RUN) to identify biomarkers and epigenetic signatures of spermatogonial stem cells (SSCs)
- Integrate five single-cell datasets to construct a cell atlas across pseudotime using Monocle and Seurat on HPC cluster
- Develop a ground-up R-based Bulk RNA-Seq analysis pipeline and code under that achieves high-quality control and eliminates batch effects
- Validate a GFP-tagged transcription factor transgenic mouse using microscopy, Western blot, FACS sorting, and transplantation
- Identified and isolated a novel cell population termed LIC germ cells based on biophysical parameters

Computational Genomics Intern | Vertex Pharmaceuticals **May 2025–August 2025**

- Employ multi-modality transcriptomics (Bulk RNA-Seq, scRNA-Seq, Spatial) to identify novel surface proteins for LNP gene therapy targeting in ADPKD
- Delivered presentations to cross-functional teams on new candidate markers for further therapeutic development

Targeting Cancer Vasculature | Xavier and Tulane Undergraduate Researcher **September 2016–May 2020**

- Utilized chick embryo CAM and wound healing assay to analyze the impact of rare ketohexose sugars on angiogenesis
- Discovered two ketohexoses with significant inhibition of angiogenesis and demonstrated their therapeutic potential
- Cultured cell lines to study how Human Herpes Virus-8 (HHV8/KSHV) shapes tumor microenvironment
- Utilized IPA for a Cytomegalovirus project with Tulane and the Louisiana Cancer Research Consortium
- Trained and mentored by NIH & NIA in sophisticated aging research at Frontiers in Aging and Regeneration course

Bioengineering Research | Amgen Scholar at Stanford University **June 2019–September 2019**

- Selected as one of 20 Amgen Scholars out of 750 applicants

- Live imaged and developed a functional assay for peptide-microbe interactions
- Assessed effects of physical perturbations on gut microbiome stability through fecal-derived culture and growth-curves
- Statistically quantified antagonism and synergistic relationships using MATLAB
- Conducted high-throughput biophysics co-culture experiments to determine if small peptides facilitated synergistic crosstalk between microbes and was acknowledged for my contributions in a Cell Press paper by KC Huang's lab
- Chosen to speak at the Amgen conference on behalf of the Stanford cohort about summer work and experience

Immunology Research | Brown University Research Fellow **June 2018–August 2018 | May 2019–June 2019**

- Investigated sepsis-induced immune dysfunction by examining the PD-1/SHP-1 pathway in mouse macrophage models and cecal ligation and puncture models
- Utilized the SHP-1 antagonist TPI-1 to assess effects on cytokine profile and therapeutic capabilities
- Contributed data and insights to enhance understanding of immunological alterations in sepsis and disease

Microbiology and Immunology Research| Dartmouth Research Scholar **June 2017–August 2017**

- Investigated *Pseudomonas aeruginosa's* flagellar motility and its effect on immune system evasion, using cellular and genomic techniques for functional validation
- Engineered motile and non-motile isolates via transposon mutagenesis that were then used in macrophage co-culture.
- Analyzed data through flow cytometry and binding assays, demonstrating a statistically significant decrease in adhesion of non-motile strains

Technical Proficiency

Wet Lab Skills Tissue Culture: Embryonic stem cells | Human cancer lines | Bone marrow-derived macrophages | HUVEC | Molecular Techniques: Fluorescence & confocal microscopy | Genetic manipulations | Transfection & transformation | Live cell | Flow cytometry/FACS | CUT&RUN | Western | qPCR | ELISA | Anaerobic chamber | Mouse Models: Dissection E.14-Adult | Cecal ligation& puncture | Lavage & organ harvest

Data Analysis and Visualization Bioinformatics: High proficiency in Batch & R using ggplot2, DESeq2, Monocle3, Seurat, and more | Familiarity with MATLAB & Ingenuity Pathway Analysis (IPA) Software: FlowJo, Fiji, Adobe Illustrator, Microsoft Office

Leadership Experience

Instructor for Frontiers of Stem Cell and Regenerative Biology **July 2022 | 2023 | 2024**
Created and executed course materials, lectures, computational experiments, and wet labs (Westerns, IF, PCR) for 60+

Dr. Samuel M. Nabrit Conference Organizer **March 2023–June 2023**
Committee member and organizer for conference showcasing research achievements of outstanding & diverse molecular life scientists with keynotes from 2 Howard Hughes Medical Institute Investigators

Graduate Teaching Assistant| Cell and Molecular Biology at Brown University **January 2022–May 2022**
Designed materials, exam prep, and review sessions for 160 students and 9 undergraduate teaching assistants for Cell and Molecular Biology course

HBCU-Stanford Planning Committee **February 2021–February 2022**
Served as sole graduate student on a 6 person committee with faculty and staff to create programming for inter-institutional research collaboration for students and faculty, which secured 3 years of funding

Presentations & Conferences

Talks

Mammalian Reproduction Gordon Research Conference | Barcelona, Spain **July 2024**
Invited Talk "Quiescence and Chromatin Regulation During Male Embryonic Germ Cell Development"
Full travel award |130+ attendees

Department of Biology Seminar Series | New Orleans, Louisiana **October 2023**

Invited Talk "Fertility, Quiescence, and Chromatin Regulation During Germline Stem Cell Development"
80+ attendees, including the University President and Dean of Arts and Science

Germinal Stem Cell Biology Gordon Research Conference | Barcelona, Spain **May 2023**
Invited Talk "Cell Cycle and Chromatin Regulation During Male Embryonic Germ Cell Development"
Served as a discussion leader | Full travel award | 130+ attendees

Leadership Alliance National Symposium | Hartford, Connecticut **July 2022 | 2023**
Speaking engagement on attaining a National Science Foundation Graduate Research Fellowship | 300+ attendees

Advancing Diversity in Aging Research ADAR Summit | Boston, Massachusetts **November 2019**
Presentation on "Frontiers in Aging and Regeneration Research Advanced Training Course" as a representative and past participant in the National Institute on Aging at NIH course

Posters

Bartholomew MA, Seymour KA, Gura MA, Abt KM, Freiman RN, The Birds and the Taf4b-ees: Male Germ Cell Transcriptional Regulation and Cell Dynamics, Poster. HHMI Gilliam Annual Meeting, September 2024 Washington DC
Bartholomew MA, Gura MA, Abt KM, Seymour KA, Freiman RN, Integrating Transcription, Chromatin, and Quiescence During Male Germline Stem Cell Development, Poster. HHMI Science meeting, March 2024, Chevy Chase, Maryland
Bartholomew MA, Seymour KA, Gura MA, Abt KM, Freiman RN, Quiescence and Chromatin Regulation During Male Germline Stem Cell Development, Poster. ABRCMS Graduate Symposium American Society for Microbiology, November 2023, Phoenix, Arizona
Bartholomew MA, Seymour KA, Gura MA, Abt KM, Freiman RN, Cell Cycle and Chromatin Regulation During Male Germline Stem Cell Development, Poster. HHMI Gilliam Annual Meeting, September 2023, Chevy Chase, Maryland
Bartholomew MA, Gura MA, Abt KM, Seymour KA, Freiman RN, Transcriptional Regulation During Male Germline Stem Cell Development, Poster. HHMI Gilliam Annual Meeting, March 2023, Chevy Chase, Maryland
Bartholomew MA, Gura MA, Abt KM, Seymour KA, Freiman RN, Quiescence and Chromatin Regulation During Male Embryonic Germline Stem Cell Development, Poster. Mammalian Reproduction Gordon Research Conference, August 2022, West Dover, Vermont
Bartholomew MA, Gura MA, Abt KM, Seymour KA, Freiman RN, The Regulatory Role of Transcription Factor TAF4b During Spermatogonial Stem Cell Development, Poster. Cold Spring Harbor meeting: Systems Biology: Global Regulation of Gene Expression, March 2022, Cold Spring Harbor, New York
Bartholomew MA, Chung CS, Chen Y, Ayala A, Effects of SHP-1 Signaling on Macrophage Function and Overall Survival During Sepsis" Immunology. Poster. ABRCMS American Society for Microbiology, November 2019, Anaheim, California
Bartholomew MA, Chung CS, Chen Y, Ayala A, Characterization of SHP-1 Signaling to Sepsis-Induced Changes in Macrophages Function and Overall Survival. Poster. SACNAS, October 2019, Honolulu, Hawaii
Bartholomew MA, Chung CS, Chen Y, Ayala A, Contribution of SHP-1 Signaling to Sepsis-Induced Changes in Not Only Immune Cell Function, but Overall Survival. Poster. Leadership Alliance National Symposium, July 2018, Hartford, Connecticut
Bartholomew MA, McFerrin H, Muniruzzaman S, Anti-Angiogenic Actions of Rare Ketohexoses in a Chick Embryo and Scratch Models. Poster. Conference ERN AAAS, February 2018, Washington DC
Bartholomew MA, Berwin B, Mutagenesis of Clinical Pseudomonas Aeruginosa Isolates to Assess Phagocytic Susceptibility. Poster. Icahn School of Medicine at Mount Sinai Research Symposium, September 2017, Manhattan, New York

Service

The Leadership Alliance Executive Director Search Committee
Brown University Vice President for Institutional Equity and Diversity Search Committee
Brown Center for Students of Color Student Leadership Grad Coordinator **January 2023–May 2024**
Led specific programming supporting the Associate Dean with events, strategic planning, and organization
Nabrit BGSA Chair of Socials | Brown University **June 2022–June 2023**
Coordinated all gatherings for Brown's Black graduate community and organized the largest event for graduate students of color, encompassing 100+ students in collaboration with SACNAS and other organizations
Mission to Honduras | Honduras Outreach Inc. **May 2018–June 2018**
Served in community by building, praying, and teaching while living in Ranchero el Paraiso, Agalta Valley/Olancho

"I will make discoveries and strongly impact the biological sciences while also excelling in a position that empowers me to improve human health"-Myles Bartholomew

Acknowledgements

I begin by thanking God, through whom all things are made possible. I pray that anyone reading this sees the work here as a testament to His love for us and the beauty of His design. I am in awe of His grace and faithfulness.

I would like to thank my wife, Praise. You are the light of my life. There isn't a world I would want to live in where you are not there with me. Through your consistent love, support, and encouragement, I've become the man I am today. You have given me the greatest blessing in life, and I thank you for our children. To Mylah & Trust, you all are my greatest source of joy and inspiration. Daddy's love for you is immeasurable, and I can't wait to see the places you will go. The Lord will bless you both and keep you, and show you great favor in this life.

To my amazing parents, thank you for teaching me everything I know. I thank the Lord for placing you both in my life, and I owe everything to y'all's boundless sacrifice and love. No kid has been as lucky as I am. To my siblings, Ania, Brazil, Rhyan, and Niguel, thank you all for being a constant source of comfort and grounding conversation throughout my life. We are literally built-in best friends, and I couldn't imagine my life any other way than being you all's big brother.

To my best friends. Collin, I literally would not have made it through grad school without you by my side. Our constant conversations, hangouts, and turn-ups got me to this point. Thank you for letting me share my obsession with my project and for your willingness to help in every aspect. I love you and your family and will always be here for you. Chauncey, you are truly my brother, the embodiment of a good man and friend. I thank God for your love and loyalty over the years. George, look how far we've come. I

can't believe we followed each other from undergrad all the way to the East Coast. We've had so much fun over the past ten years, our bond is unbreakable, and I can't wait to see all that you achieve and all we celebrate.

To my lab. Kim, you are the crème de la crème: a fantastic scientist, colleague, collaborator, and friend. I could never repay the hours of conversation or the number of snacks you let me "borrow." Our good mornings and your kindness got me through some of my toughest days. Megan and Kimberly, y'all are in a class all your own; some of the best mentors, writers, and scientists I am sure I will ever meet. I can't thank you both enough for teaching me everything you knew with patience, kindness, and humility. You set the bar high. Anna and Nicole, the energy you bring to the lab is contagious. I'm grateful for the joy you shared through spontaneous conversations, musical outbursts, and trips to Xaco. Kevin, you were a wonderful extended lab member. We spent countless hours together, and your conversation and company made a world of difference.

Chyna, Brandon, Maria, Tracy, Beth, Dr. Ayala and all the people that I first met years ago as a sophomore. Thank you for being part of my research journey to Brown. You have always treated me like family, and my experiences and interactions with you over the past eight years have surely helped me get here today. I also have tremendous gratitude for my Howard Hughes Medical Institute Gilliam peers and leadership. Josh, Adrielle, and Mel, and the crew at HHMI, I miss our time together. The energy you all bring is awe-inspiring, and the work you all have done will live on forever. Also, I sincerely thank the National Science Foundation for its support through the Graduate Research Fellowship Program (GRFP).

Also, a sincere thank you, Diana, for not only agreeing to be my outside reader but also for always being a kind and inspiring voice. Since the day Rich introduced us at my first Gordon to now, you have been nothing but supportive and encouraging.

To my Internal Committee, Wow. Mark and Alison, I've known you both for nearly nine years. Throughout that time, you have been a constant source of support, encouragement, and scientific insight. Through your open-door policies (including when you all were clearly busy) and the classes you taught, I've learned so much, not only about science, but about myself. Thank you both for encouraging me to apply, believing in me throughout, and seeing me to the finish line. Nicola, I have always thought you were one of the coolest scientists I know. Thank you not only for inspiring me, but for teaching me as a computational scientist; you imparted to me a skill set I will never take for granted. God bless you all.

And to cap it all off, Rich, thank you. Thank you for choosing me. Thank you for teaching me. Thank you for showing up for me on good days and on hard days. You inspired me then, and you continue to inspire me now. Every day we've shared has been a blessing. The hours you have invested in me are innumerable in value and gratitude. You are a lifelong mentor, teacher, and friend, the best to ever do it. I would not be here without you, and to God be the glory that I have you in my life. I truly will always cherish our time together; it has been the opportunity of a lifetime. I am proud to be a part of your legacy. I love and adore you, and may the Lord continue to bless you and everything you touch.

In the words of C. L. Mayfield, "Keep On Keeping On."

Table of Contents

List of Figures

CHAPTER 1: MALE FERTILITY AND THE GENETICS OF SPERMATOGONIAL STEM CELL ESTABLISHMENT

I wrote this chapter with editorial feedback from Dr. Richard Freiman.

Preface

Chapter 1 establishes the foundation and context for this thesis by examining striking trends in declining male fertility and introducing a timeline of the most critical stages of male germ cell development. It then transitions to the molecular regulators of transcription and epigenetic reprogramming that shape germ cell fate, heterogeneity, and overall cell survival. Together, these sections integrate transcriptional control with epigenetic integrity across late embryonic germ cell development and provide a framework for understanding how transcriptional machinery, and particularly TAF4b, contributes to spermatogonial stem cell establishment and long-term male fertility.

The Emerging Crisis in Male Reproductive Health

Bridging from past generations and carrying life into the future, germ cells occupy the most consequential role in biology. They are the cells responsible for ensuring the continuation of the species and life as we know it. In humans, there are two instruments that play in this symphony and bring forth future generations. In females, it is the egg, and in males, it is the sperm. In both sexes, there is tight regulation of the biological and molecular events that give rise to healthy gametes, and the smallest deviations from this canonical path can be detrimental not only to future progeny but also to species survival (Ellis et al., 2007; Yatsenko & Rajkovic, 2019). The role of male organisms is often overlooked, as most of the focus on fertility has been placed on female contributions (Arya & Dibb, 2016; Ravitsky & Kimmins, 2019). However, unexplained male infertility makes up 7.5% of infertility cases in couples globally, and male fertility has shown significant declines over the past few decades. Some reports indicate that sperm count has decreased by more than 60% and sperm concentration by 51.6% (Levine et al., 2023). Many have speculated that by the turn of the century, a significantly larger fraction of men will require some form of reproductive assistance (Greenberg et al., 2024). The difficulty in solving this growing problem is that many of its determinants remain to be uncovered. There is increasing evidence that the environment significantly affects reproductive health, with chemicals like phthalates, pesticides, and other endocrine-disrupting compounds linked to male reproductive toxicity (Krzastek et al., 2020). Conversely, research also shows that obesity and age greatly influence men's reproductive health, highlighting notable changes in spermatogonial dynamics in individuals with higher BMI and significant epigenetic alterations in older men (Ashapkin et al., 2023; Sermondade et

al., 2013). Primary forms of male infertility are obstructive azoospermia, where there is a physical barrier causing a disruption to the normal path of sperm, or non-obstructive azoospermia (NOA), where there is no blockage to the release of the male germ cells; instead, they are absent or dysfunctional (Alkandari & Zini, 2021). In the case of NOA, an interesting paradox is called into question: why have the germ cells, also called gametes, responsible for fertility failed? While there are many diverse causes of NOA, one focused on here is the embryonic development of the male germ line. This focus on early development arose from increasing evidence that genes and corresponding mutations can lead to premature failure in cells considered fundamental to male reproductive health (Nguyen et al., 2020).

From Specification to Immortality: Germ Cell Foundations

To understand what can go wrong in gametogenesis, we must first understand the origin and development of gametes. In a poetic and generational sense, the cells that will produce our offspring arise while we are still developing in our mother's womb, meaning that the potential for future life exists even before an individual is born (Nguyen et al., 2019). During this time, developing gametes must relinquish their mortal somatic identity, traverse their own unique path, and prepare for immortality (Ben Maamar et al., 2022). Their developmental journey, though complex and treacherous, equips them to carry life forward into the next generation.

Primordial Germ Cell Specification and Migration

Functional studies on the human fetal testis are largely prohibited by access and ethical concerns. Accordingly, the mouse and its genetic tool box, provides a comparable model organism used to study these elegant events (Ewen & Koopman, 2010; Law et al.,

2019; Saitou & Yamaji, 2012). In mice, a founding population of cells begins to express well-known markers of primordial germ cell (PGC) identity, and by embryonic day (E)7, approximately 40 PGCs are present (Ginsburg et al., 1990). These PGCs, along with those in other organisms such as humans, must then undergo the first major germ cell migration event (Jaszczak et al., 2025). This process exhibits mixed fidelity as some PGCs do not reach or occupy the gonadal ridge, thus constituting the first major selection event. This is particularly interesting because, speculatively, this migration could serve as a test run for future germ cells, providing an assessment of the founder cells' ability to respond to signals similar to those involved in chemotaxis in sperm as they pass through the female reproductive tract (Eisenbach, 2025). By E12.5, the final PGCs have migrated from the allantois to the gonadal ridge (Jaszczak et al., 2025). Throughout this stage and in subsequent development, the concepts of fitness, asynchronous development, and a spectrum of transitory cell states frequently appear in male germ cell biology.

Cell Cycle Exit and the Establishment of Quiescence

Around this same point, sex determination takes place with XX female PGCs taking one trajectory and XY male PGCs taking another (Capel, 2017). Although the cells have found their resident organs, the subsequent steps in both trajectories and timing are highly dynamic and sex-specific. For the purposes of this review, we will delve further into these dynamics in male germ cell development, which from E12.5-E16.5 is characterized as predominantly proliferative, with cells asynchronously transitioning into a state of quiescence (Du et al., 2021). Quiescence, or G_0, is best understood as a temporary exit from the cell cycle, allowing cells to divert time and energy to processes that protect their maintenance and renewal capacity. Research indicates that quiescence is essential for

maintaining fertility, as disruptions in cell-cycle regulators or failure to arrest during mitosis can lead to infertility (Du et al., 2021). It is believed that by E16.5, most germ cells enter a quiescent state. Our data also highlight the heterogeneity among these quiescent germ cells during late embryonic development (Chapter 2) (Gura et al., 2023). During this phase, the cells are often called prospermatogonia or gonocytes, which are further classified into M prospermatogonia (mitotic), T1 (transition 1), I for (intermediate), and T2 (transition 2) (McCarrey, 2013). It is thought that spermatogonial stem cells originate immediately afterward from T2 prospermatogonia. However, recent studies indicate that cells with stemness potential, characterized by their ability to restore spermatogenesis after transplantation, are present as early as E18.5 (Law et al., 2019).

Spermatogonial Stem Cell Establishment and Identification

With the establishment of the SSC pool during the perinatal period, there is also an indication that in mice, there is a round of directly differentiating spermatogenesis that can occur independently of SSCs beginning around postnatal day 10 (Sasaki & Sangrithi, 2023). It is thought that only subsequent rounds of spermatogenesis and the prolonged fertility window rely on the strong initial SSC pool. These cells remain an elusive population for research and discovery. In addition to their rarity, estimated at only 0.01–0.02% of cells in the adult tubule, evidence indicates that some cells, at least in mice, can regain stemness even after seemingly initiating differentiation programs (Nagano & Yeh, 2013). This heterogeneous and highly dynamic cell type has thus far been defined primarily by putative markers that, when selected for, confer spermatogenic rescue. These markers have included CD90 (also called THY1), POU5F1 (OCT4), inhibitor of differentiation 4 (ID4), neurogenin 3 (NGN3), GDNF family receptor alpha 1 (GFRα1), and

others (Abbasi et al., 2013; Valli et al., 2014). However, while these markers may enrich for SSCs, it remains unclear whether they are comprehensive or even exclusive, as the extent of enrichment depends on the co-expression or absence of other regulatory genes (Song et al., 2024).

Thus far, many aspects of germ cell development are conserved between humans and mice, although the timing and duration of events differ substantially. In humans, PGC specification occurs around 2 weeks post-fertilization and is followed by migration to the gonads by week 5. By week 14, M prospermatagonia transition into quiescent T1 prospermatagonia and, as in mice, undergo extensive epigenetic reprogramming throughout the fetal period (Figure 1.1). The identification of progenitor SSCs in mice prior to birth highlights the importance of the late embryonic period for later SSC establishment. However, the molecular and cellular mechanisms that underlie this foundational window for lifelong fertility remain to be fully understood. We and others have worked to uncover the transcriptional, epigenetic, and lineage determinants during this critical stage and have recently found mechanisms that integrate these biological processes.

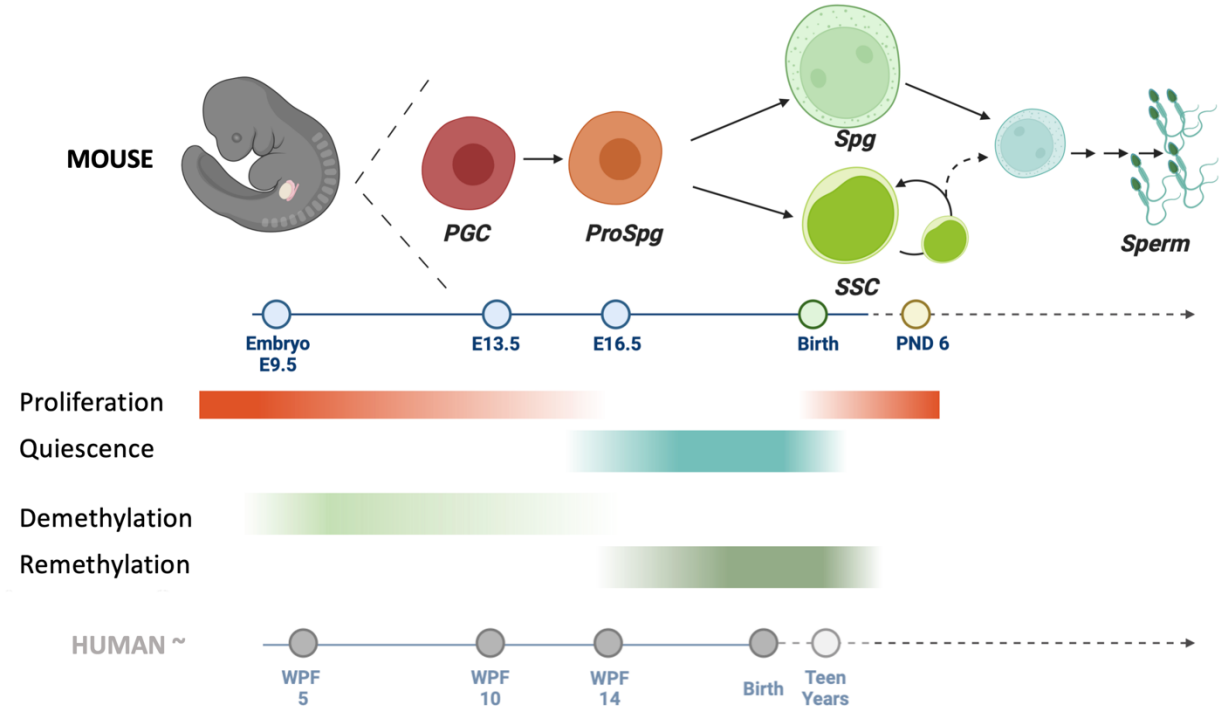


Figure 1.1 Male Germ Line Development
A timeline of germ cell development and key biological processes that overlap with the perinatal period in mice, as well as an approximate timeline for human male germ cell development, weeks post fertilization (WPF).

Transcriptional Regulation as a Driver of Identity

Cell fate decisions, governed by gene expression networks, are a fundamental principle of biology and reflect the transcriptional basis of life. Since the discovery of transcription and RNA polymerase II, scientists have been exploring how specific genes are activated or repressed (Jacob & Monod, 1961; Nikolov & Burley, 1997; Roeder, 2019). A key component in germ cell development is the basal transcription complex TFIID, which performs multiple roles: binding to the promoter, recruiting RNA polymerase II, and even influencing chromatin modifications (Gura et al., 2020; Mizzen et al., 1996). TFIID canonically consists of TATA-binding protein (TBP) and 8 to 14 TAF subunits, whose

combinations dictate which genes are expressed or silenced, as well as have varying degrees of cell specificity (Le et al., 2019; Pugh & Tjian, 1991). On a molecular level, the diversity of TFIID flavors can go beyond canonical TATA boxes and interact with other enhancers, complexes, and a plethora of genomic motifs.

TBP Associated Factor 4b (TAF4b): A Germline-Specific Regulator and TFIID

The primary focus of this study, TBP-Associated Factor 4b (TAF4B), was initially identified in lymphocytes as TAFII105; soon after, it was shown that its deficiency in mice resulted in female infertility (Dikstein et al., 1996a; Freiman et al., 2001). Our lab then demonstrated significant impairment of spermatogenesis in Taf4b's absence, supporting the conclusion that it is critical for fertility in both sexes (Falender et al., 2005). Continued investigation has relied primarily on the original transgenic mouse line harboring a homologous recombination-mediated mutation in Taf4b, referred to as Taf4b-deficiency. (Freiman et al., 2001). This model also mimics the phenotypes observed in human studies of idiopathic male infertility. The first study involved a Turkish population and found that men with nonsynonymous mutations were infertile but had no other impairments (Ayhan et al., 2014). The second was a Chinese GWAS study that identified single-nucleotide polymorphisms that lead to missense mutations, which were enriched among individuals diagnosed with azoospermia (Xi et al., 2020).

On a molecular level, TAF4b's presence or absence in the TFIID complex has been shown to specifically alter the conformation and structure of TFIID, thereby regulating gene activity through the kinetics of the larger preinitiation complex (PIC) (Liu et al., 2008). Transcriptional regulation is further complicated because a single factor can control hundreds of others; for instance, TAF4b influences the transcription factor c-Jun,

affecting its binding at its own promoter (Liu et al., 2008). Moreover, c-Jun expression varies significantly in TAF4b-deficient cells (Gura et al., 2023). Changes like this and in the hundreds of other transcription factors thereby have a compounding impact on cell identity, fate, and overall gene regulation.

Functional Roles of TFIID Subunits and Potential Contributions to Germ Cell Fate

As noted, these regulatory networks are influenced by diverse TFIID subunits, which can alter pre-initiation complex functions and gene regulation in ways that are beyond current understanding. Other TAFs that determine cell fate and behavior that are particularly relevant to TAF4b's role in germ cell development include its paralog TAF4, as well as TAF1, TAF7, TAF8, TAF10, and TAF12. Some of these have been shown to interact directly with TAF4b and/or share similar functions in promoting stemness or fertility (Luna-Arias & Castro-Muñozledo, 2024; Zhou et al., 2013).

TAF4, also known as TAF4a in mice, is a paralog of TAF4b with a conserved C-terminal histone fold domain and a nonconserved N-terminus (Gustafson et al., 2020). TAF4 is considered the most critical subunit for TFIID stability (Wright et al., 2006). Within the TFIID complex, TAF4a supports differentiation across various cell types. Evidence shows that its expression increases as cells mature into adulthood, with lower levels in undifferentiated tissues and early developmental stages (Bahat et al., 2013). Additionally, in Taf4a knockout embryonic stem cells (ESC), TAF4b becomes enriched in TFIID, implying that TAF4a may sometimes act antagonistically to TAF4b depending on the context (Langer et al., 2016). Studies ranging from neuronal to hepatic differentiation, including our own, indicate that, unlike TAF4b, TAF4a's expression pattern aligns more with basal transcriptional machinery and differentiation processes (Luna-Arias & Castro-

Muñozledo, 2024). Through further in-depth comparison and transcriptional profiling, we may elucidate the regulatory mechanisms underlying both.

TAF1 was among the first TAFs identified to interact with Taf4b and is the largest member of the TFIID complex (Dikstein et al., 1996b; Luna-Arias & Castro-Muñozledo, 2024). One of its most interesting functions is its ability to modify chromatin (Wassarman & Sauer, 2001). This is particularly intriguing because it provides a direct link between transcription and epigenetic regulation, potentially promoting gene expression via TFIID and the PIC while altering H3 and H4 acetylation and chromatin accessibility (Mizzen et al., 1996). This is also important to mention because it establishes the paradigm that other TAFs or alternative complexes may have a direct transcriptional component in addition to an epigenetic component. Furthermore, there are reasons to believe that, like Taf4b, Taf1 is involved in germ cell development. The first is that it is enriched in premeiotic male germ cells (Metcalf & Wassarman, 2007). Also like Taf4b, Taf1's expression correlates with Oct4, self-renewal, and stemness in mice. (Wang et al., 2022) Evidence of interaction and verified presence in germ cells could indicate a strong Taf4b co-regulatory mechanism or point to alternative contributions to germ cell development.

To gain insight into Taf4b's potential cooperative or counterbalancing role with its paralog Taf4a, it is worth noting other groups of paralogous TAFs, including TAF7 and Taf7l. TAF7, similar to TAF4b, directly interacts with TAF1 (Bhattacharya et al., 2014; Luna-Arias & Castro-Muñozledo, 2024). It also plays a crucial role in promoter pausing, which has a repressive effect (Bhattacharya et al., 2014; Gegonne et al., 2006). This highlights a nuanced aspect of TAFs: they are not only involved in transcription via POL2 but also in the kinetics of gene expression, thereby making the process tightly regulated

and context-dependent. Both TAF7 and Taf4b have been implicated in promoting proliferation in undifferentiated cells; however, TAF7 is also present in differentiating cells (Cheng et al., 2021; Gegonne et al., 2012; Huang et al., 2024a). The first line of evidence for its role in proliferation showed that it acted in an inverse manner to Transforming Growth Factor-beta (TGFb), which, in most contexts, acts as an antiproliferative tumor suppressor. TAF7L is particularly interesting due to its direct role in spermatogenesis. *Taf7l*-deficient mice show a notable reduction in sperm count and severe spermiogenesis abnormalities. Additionally, evidence suggests that *Taf7* may compensate for *Taf7l* deficiency in expression, although the two factors may drive opposing programs in certain contexts (Cheng et al., 2007). Our lab has elucidated the tight correlation in germ cell expression between our factor of interest, TAF4b and TAF7l (Chapters 3 and 4). Moreover, this relationship was found in both quiescent transitioning prospermatogonia and meiosis-initiated embryonic oogonia (Gura et al., 2020). Given the many parallels, further evidence is needed to determine that the relationships between TAF7 and TAF7l are analogous to those between TAF4b and TAF4, but it does offer insights into other paralogous TAFs.

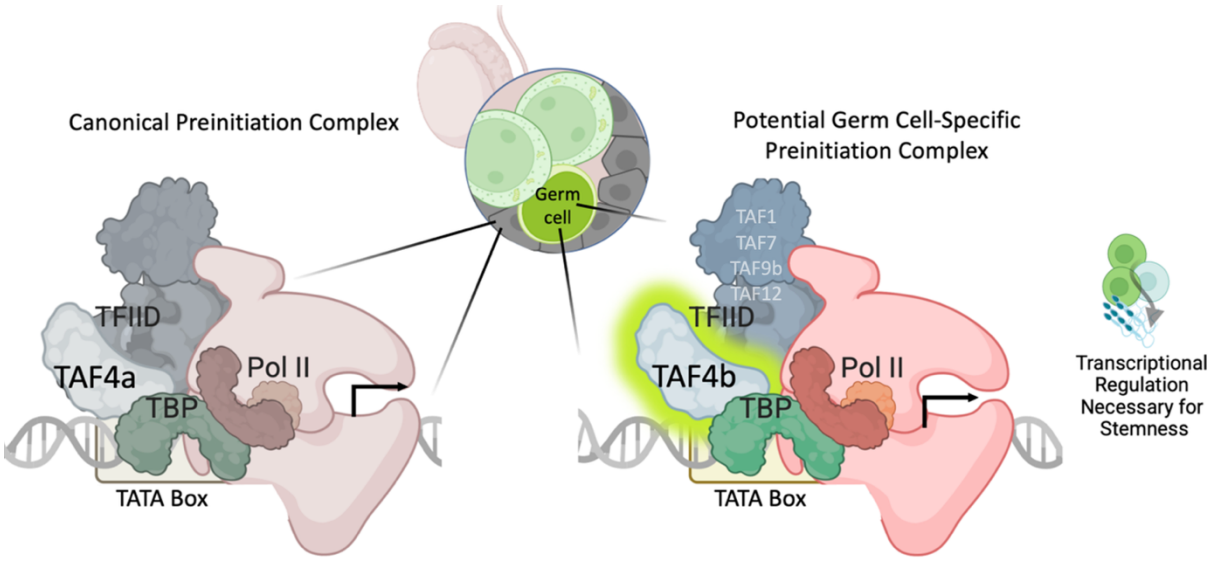


Figure 1.2 Potential germ-line specific mechanism of transcriptional regulation
The canonical preinitiation complex that consists of TAF4a and other basal TFIID components and the proposed germline stem cell-specific pre-initiation complex with TAF4b driving stem cell specific transcription.

While TAF4b and TAF7l have been studied specifically in the germ cell context and in TAFopathies, much remains to be learned from the less widely studied, yet more basal TAFs. These are the TAFs considered to be less cell-type specific and instead critical for general cellular transcription and viability. Here, our particular focus is applied to those that are highly dynamic during germ cell development and or differentially expressed in the absence of TAF4b (Chapter 3). The first is TAF8. This TAF is critical for gene expression programs through its heterodimerization with TAF10. It is reported to then recruit the remaining TFIID subunits. This synergism extends beyond transcription initiation and illuminates protein dynamics, as TAF8 is also responsible for the nuclear shuttling of its partner, TAF10 (Trowitzsch et al., 2015). The nuclear localization signals of specific transcription factors are often overlooked in regulation, yet this work reveals the breadth of modulatory mechanisms. Once in the nucleus, TAF10 can then carry out

its function as a driver of proliferation and differentiation. TAF10 is also another example of how either individual members of or the greater TFIID complex can drive chromatin modification, as TAF10 functions in a histone acetylase complex. Lastly, as with the heterodimerization of TAF8 and TAF10, TAF4b has been shown to directly interact with TAF12 via highly conserved histone-fold domains. (Gazit et al., 2009; Werten et al., 2002). This interaction is also intriguing because of the significant dysregulation of TAF12 in TAF4b-deficient mice during late embryonic germ cell development (Chapters 3 and 4). This TAF exhibits one of the most pronounced upregulations among TFIID members and may indicate a feedback mechanism. Furthermore, TAF12 has been shown to drive dramatic changes in gene expression; its dysregulation disrupts promoter-mediator-enhancer communication as well as causes severe failure of the PIC (Sun et al., 2021). Although not all TAFs have been specifically studied in spermatogenesis, these examples demonstrate the many ways TAFs influence each other's subcellular and chromatin localization. Taken together, these observations underscore the exquisite coordination of transcription required for correct cell fate.

Goals and significance.

This work primarily focuses on the development of late embryonic prospermatogonia and explores SSC establishment. It details how TAF4b, a key factor for maintaining fertility, influences transcription, epigenetics, and germ cell fate. It also shows conserved aspects of the Taf4b regulatory circuit in humans. Here, we produced a comprehensive assessment, using experimental and computational techniques to fulfill the following goals:

Characterize direct and indirect genomic regulatory effects of TAF4b via integration of Taf4b-dependent transcriptional changes, chromatin modification, and TAF4b occupancy in genome localization analysis.

Characterize TAF4bs' contribution to germ cell dynamics, including cell cycle, survival, and spermatogonial stem cell development.

This work demonstrates that TAF4b is crucial as an integrator of transcription, genomic stability, and male germ cell fate determination. We show that Taf4b expression is markedly increased during the late embryonic stages of male germ cell development. Interestingly, in Taf4b-deficient mice, we observed significant disturbances in entry into quiescence and the degree of quiescence around E16.5. We also found substantial TAF4b occupancy at genes involved in covalent chromatin modification, DNA repair, and cell cycle, all pathways that were differentially expressed between WT and Def. Most of these sites, as well as the majority of TAF4b binding sites, were CpG-rich promoters upregulated in the absence of TAF4b (**Chapter 2**). We next found that, beyond regulating quiescence entry, TAF4b influences cell fate decisions during this period. Using a TAF4b-GFP model, we identified a new germ cell subset that appears specifically at E18.5 and

is completely absent in TAF4b-deficient testes. We also observed that as TAF4b expression peaks and this bifurcation occurs, much of the differential gene expression remains unresolved in categories related to the cell cycle, epigenetic modification, and DNA damage repair. This supports that critical programs do not recover or reach canonical levels and that the effects of TAF4b-deficient go beyond a temporary delay. These changes are associated with epigenomic alterations, including decreased CpG methylation, increased Line1 retrotransposition, and ultimately, apoptosis. Finally, we present evidence that TAF4b-enriched cells possess stemness capacity (**Chapter 3**). Overall, the data provide new insights into the essential function of TAF4b and suggest new paradigms for its regulatory roles in mammalian spermatogonial stem cell development. (**Chapter 4**).

CHAPTER 2: TRANSCRIPTION AND CHROMATIN REGULATION BY TAF4B DURING CELLULAR QUIESCENCE OF DEVELOPING PROSPERMATOGONIA

Megan A. Gura[1†], **Myles A. Bartholomew[1†],** Kimberly M. Abt[1] ,Soňa Relovská[2], Kimberly A. Seymour[2] and Richard N. Freiman[1,2*]

I generated Figures 2.1, 2.2, 2.3, 2.4, and 2.7e,f, and wrote the corresponding results and methods sections. This chapter was published as

Gura, M. A. †, Bartholomew, M. A. †, Abt, K. M., Relovská, S., Seymour, K. A., & Freiman, R. N. (2023). Transcription and chromatin regulation by TAF4b during cellular quiescence of developing prospermatogonia [Original Research]. *Frontiers in Cell and Developmental Biology*, *11*. https://doi.org/10.3389/fcell.2023.1270408

† indicates co-first authors

[1]MCB Graduate Program, Providence, RI, United States

[2]Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI, United States

Abstract

Prospermatogonia (ProSpg) link the embryonic development of male primordial germ cells to the healthy establishment of postnatal spermatogonia and spermatogonial stem cells. While these spermatogenic precursor cells undergo the characteristic transitions of cycling and quiescence, the transcriptional events underlying these developmental hallmarks remain unknown. Here, we investigated the expression and function of TBP-associated factor 4b (Taf4b) in the timely development of quiescent mouse ProSpg using an integration of gene expression profiling and chromatin mapping. We find that Taf4b mRNA expression is elevated during the transition of mitotic-to-quiescent ProSpg and Taf4b-deficient ProSpg are delayed in their entry into quiescence. Gene ontology, protein network analysis, and chromatin mapping demonstrate that TAF4b is a direct and indirect regulator of chromatin and cell cycle-related gene expression programs during ProSpg quiescence. Further validation of these cell cycle mRNA changes due to the loss of TAF4b was accomplished via immunostaining for proliferating cell nuclear antigen (PCNA). Together, these data indicate that TAF4b is a key transcriptional regulator of the chromatin and quiescent state of the developing mammalian spermatogenic precursor lineage.

Introduction

Male fertility is dependent on a highly regulated series of developmental events that begin with primordial germ cell (PGC) specification in early fetal development and end with the exhaustion of an adult unipotent spermatogonial stem cell (SSC) population during old age (De Rooij and Grootegoed, 1998; Oatley and Brinster, 2008; Phillips et al., 2010). In

mammals, PGC specification is achieved via BMP signals secreted from a small group of dorsally localized extra-embryonic cells to the posterior epiblast. This then induces the expression of the master transcription regulators PRDM1, PRDM14, and TFAP2C (Ohinata et al., 2005; Yamaji et al., 2008; Weber et al., 2010). While more heterogeneous in nature, several transcription factors mark and/or support adult SSC identity, including ID4, RHOX10, PAX7, and SALL4 (Oatley et al., 2011; Gassei and Orwig, 2013; Aloisio et al., 2014; Song et al., 2016; Fayomi and Orwig, 2018; Green et al., 2018; Shami et al., 2020). Prospermatogonia (ProSpg, also called 'gonocytes') are male germ cells that have differentiated past PGC specification and have the potential to differentiate into adult spermatogonia (Spg) or become part of the SSC pool which provides the long-term renewing capabilities of the testis. Several critical molecular events occur during ProSpg development. First, epigenetic marks that were erased in PGCs are re-established in ProSpg. Second, an initial pool of unipotent SSCs is thought to arise from the heterogeneous pool of developing ProSpg during this developmental window (Kubo et al., 2015; Hill et al., 2018). These events are necessary to initiate spermatogenesis in a timely manner and maintain it properly throughout adulthood. However, the underlying gene expression networks that correctly time and integrate these complex events are currently unknown.

Sequence-specific transcription factors are thought to recruit RNA Polymerase II (RNAPII) to specific core promoters via coactivators and the general transcription machinery. How the regulation of RNAPII transcription is achieved in ProSpg and how it is integrated with epigenetic reprogramming of the male germline genome at these

developmental stages is unknown. We have identified a unique germ cell-enriched subunit of the general transcription factor TFIID, called TAF4b, that is required for proper long-term spermatogenesis in the mouse. Male mice lacking TAF4b become infertile after undergoing a first round of spermatogenesis; however, spermatogenesis can be re-established in the Taf4b-deficient testis following transplant of wildtype SSCs reflecting a functional adult SSC niche (Falender et al., 2005). More recent characterization of spermatogenesis in our Taf4b-deficient mice revealed that TAF4b is required for embryonic male germ cell development and helps regulate the delicate balance of SSC self-renewal and differentiation (Lovasco et al., 2015). In addition to studies of TAF4b in mice, human males who express a truncated version of the TAF4b protein are infertile owing to progressive azoospermia (Ayhan et al., 2014), and single nucleotide polymorphisms in human TAF4b have been linked to nonobstructive azoospermia (Xi et al., 2020). Thus, deciphering the developmental and molecular mechanisms of TAF4b in the mouse can improve our understanding and model important aspects of male reproductive development and fertility in humans.

To contextualize the progressive loss of Taf4b-deficient male germ cells during embryogenesis, we examined the spatial and temporal expression of TAF4b during embryonic testis development. We analyzed published RNA-seq data sets derived from E9.5-E18.5 Oct4-eGFP mice, in which germ cells were separated from somatic cell populations through fluorescence-activated cell sorting (FACS (Gura et al., 2020)). We determined that levels of Taf4b mRNA become progressively and significantly elevated in a germ cell-specific manner from embryonic day (E) 11.5 to E18.5. Germ cell-enriched

TAF4b protein expression at E13.5 suggests it plays an important germ cell-specific role during ProSpg development (Gura et al., 2020). Mitotic stage (M) ProSpg are proliferative between E13.5 and E16.5, at which they then progress to transitional 1 (T1) ProSpg and enter a period of cellular quiescence. Soon after birth, T1 ProSpg become transitional 2 (T2) ProSpg by reentering the cell cycle until they become Spg at approximately postnatal day (P) 8 (McCarrey, 2013). Recent studies by Law et al., 2019, suggest that a subpopulation of T1 ProSpg acquires high Id4 expression, which marks a potential initial pool of SSCs (Law et al., 2019). However, the transcriptional logic underlying these ProSpg transitions remains unknown.

Here, we tested if and how TAF4b regulates critical transcriptional programs and cell states during early ProSpg development. First, we examined Taf4b expression in published single-cell (sc) RNA-seq data and confirmed it is enriched in quiescent T1 ProSpg. Next, we performed bulk RNA-seq on Taf4b-deficient quiescent ProSpg at E16.5 to reveal cell cycle and chromatin structure as top gene ontology (GO) categories modulated in the absence of TAF4b. For genomic mapping of TAF4b localization in T1 ProSpg, we employed cleavage under targets and release using nuclease (CUT&RUN), which identified 617 TAF4b peaks just upstream of numerous transcription start sites (TSSs). Distinctive GO categories from the CUT&RUN included RNA processing, chromatin modification, and cell cycle regulation. The most consistent DNA motifs within the TAF4b-bound promoters included Sp/Klf family and NFY binding sites; a remarkable similarity to our findings in embryonic mouse oocytes, and for the first time linking these ubiquitously-expressed transcription factors to the precise regulation of ProSpg

development (Gura et al., 2022). At the cellular level, we determined that Taf4b-deficient ProSpg are delayed in their entry to quiescence, noting significant perturbation from controls at E14.5 and that these cells also display key cell cycle protein changes during ProSpg quiescence. Together, these data suggest that TAF4b is a dynamic and vital integrator of male germline transcription and chromatin states that promotes the timely quiescence of ProSpg required for further mammalian spermatogenic development.

Results

scRNA-seq reveals peak Taf4b mRNA expression coincides with ProSpg quiescence

Analyzing single-cell RNA-seq data allowed us to examine the gene expression within ProSpg at finer resolution, which we reprocessed from published publicly available data sets (Tan et al., 2020; Gura et al., 2022; Quatredeniers et al., 2023). In our analysis, we examined the expression profile of Taf4b over time to evaluate what genes and biological processes are possibly modulated concurrently or downstream of TAF4b expression. After applying our computational workflow (see Methods), the complete data set consisted of 17,310 germ cells in our uniform manifold approximation and projection (UMAP) (Figure 2.1A). We then assessed the expression of Taf4b compared to its paralog Taf4a, which is ubiquitously expressed in somatic cells. Across all time points, Taf4a mRNA expression was notably lower than Taf4b mRNA expression (Figures 2.1B,C). When observing the expression of Taf4b mRNA over the E12.5-P7 time course, we detected that Taf4b expression peaks at P0 (Figure 2.1C). The dynamic range of

Taf4b-expressing cells across time suggests Taf4b may act more akin to a dimmer switch than on/off regulation. Because previous experiments demonstrated that the Taf4b-deficient gonad exhibited changes in germ cell proliferation, we compared this newly characterized Taf4b expression profile to known proliferation markers, such as Mki67 (Sun and Kaufman, 2018). We saw an inverse expression pattern of Mki67 and Taf4b, where the significantly reduced expression of Mki67 between E16.5-P0 points to the characteristic quiescent period of T1 ProSpg (Figure 2.1D). The absence of Mki67 transcripts helps illustrate quiescence because its mRNA is expressed during S phase up until mitotic exit. Plk1 mRNA, another marker of cell proliferation, is decreased similarly from E16.5 to P2 (Figure 2.1E) (Liu et al., 2017). The combined decreased markers of proliferation and increased Taf4b expression led us to hypothesize that Taf4b may play a role in the cell cycle transitions during ProSpg development.

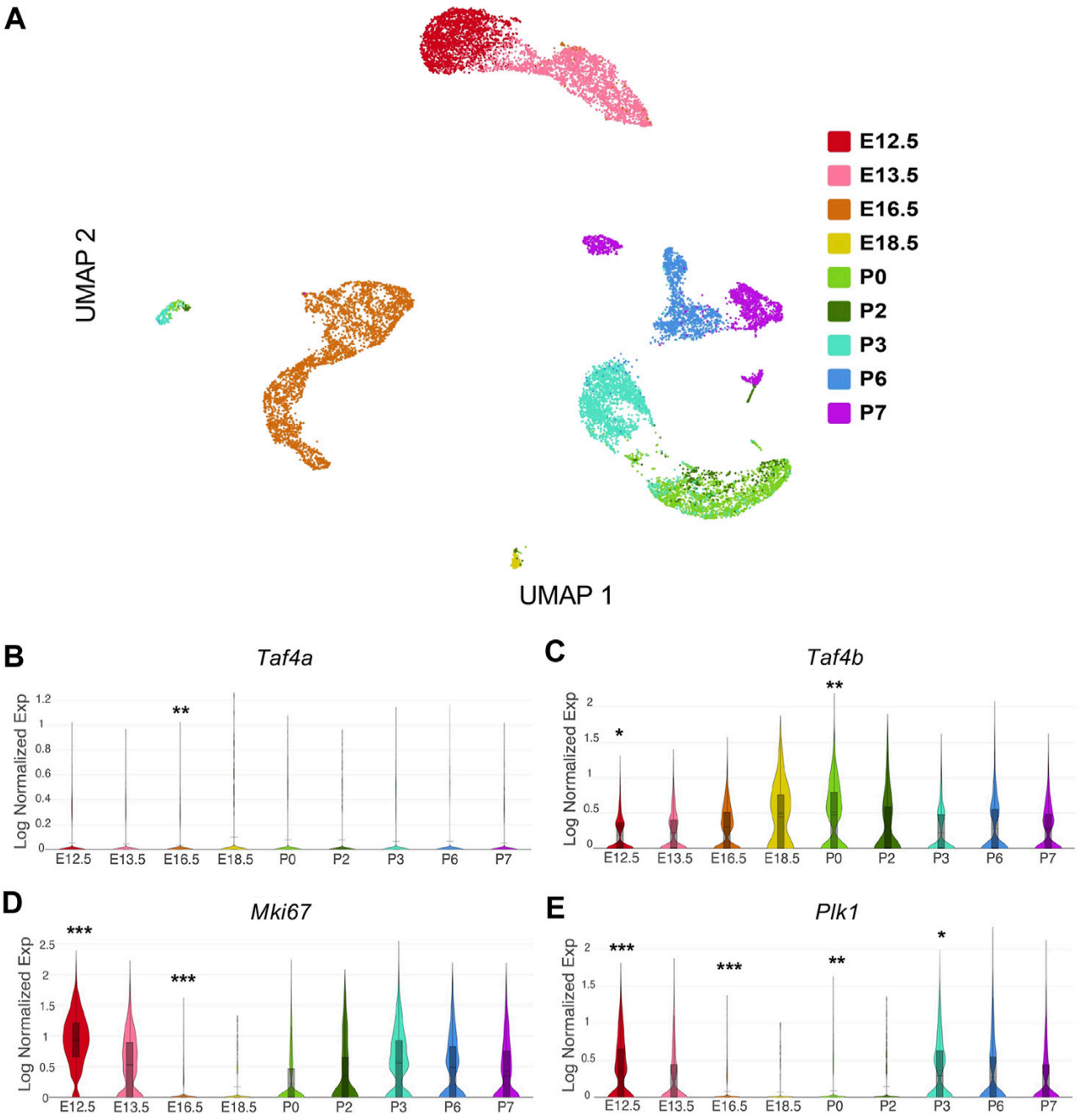


Figure 2.1. Characterizing Taf4b mRNA expression across male germ cell development in mouse.
(A) UMAP of E12.5-P7 germ cells colored by time point. (B–E) Expression of Taf4a, Taf4b, Mki67, and Plk1 across the developmental time course. p-values are denoted (*** = p< 0.001, ** = p< 0.01, * = p< 0.05) and are adjusted using the Benjamini-Hochberg correction method for multiple tests.

Global ProSpg RNA-seq reveals cell cycle and chromatin gene expression affected by TAF4b

We were intrigued to observe a clear increase in Taf4b mRNA during the ProSpg transition to quiescence. To better understand the potential function of TAF4b during this transition, we compared global mRNA differences between Taf4b-wildtype, -heterozygous, and -deficient ProSpg at E16.5. We first generated mouse embryos that were transgenic for Oct4-eGFP and either wildtype, heterozygous, or deficient for Taf4b, then sorted for GFP+ germ cells from the male embryonic testis. This provided germ cell-isolated samples from all three genotypes to perform bulk RNA-seq. Our ProSpg RNA-seq samples consisted of 7 Taf4b-wildtype, 9 Taf4b-heterozygous, and 8 Taf4b-deficient E16.5 embryos from six different litters (Supplementary Table S2.1). Although Taf4b-heterozygous male mice do not display overt signs of impaired spermatogenesis (Falender et al., 2005), including them as a unique hypomorphic and genotypic group was important. A PCA plot of these samples indicates that samples cluster primarily by genotype. While Taf4b-wildtype (WT) replicates display the tightest clustering, they do overlap with the Taf4b-heterozygous (Het) replicates, indicating a shared transcriptional profile (Figure 2.2A). This was expected as both genotypes are fertile. In contrast, the Taf4b-deficient (Def) replicates were much more dispersed and non-overlapping with the other two genotypes, suggesting large-scale transcriptional dysregulation. There were three Def samples that clustered more closely with the WT and Het samples. However, we completed the downstream analysis with all replicates because there were subtle but statistically significant transcriptional differences in these three samples compared to WT/HET. Overall, we detected 4551 DEGs between the Taf4b-wildtype and -deficient

ProSpg; 3820 were defined as protein-coding with an adjusted p-value <0.05 (Figure 2.2B). Some genes noted to be dysregulated were Mki67 and Plk1 (Liu et al., 2017; Sun and Kaufman, 2018). These markers of proliferation should decrease during this quiescent time point but are significantly upregulated in the Taf4b-deficient ProSpg (Figure 2.2B). Using gene ontology (GO) of the DEGs, we determined which biological processes are disrupted in the absence of Taf4b. This GO analysis indicated cell cycle regulation is the primary disruption driven by Taf4b ablation, followed by chromatin modification and gland development. (Figures 2.2C,D).

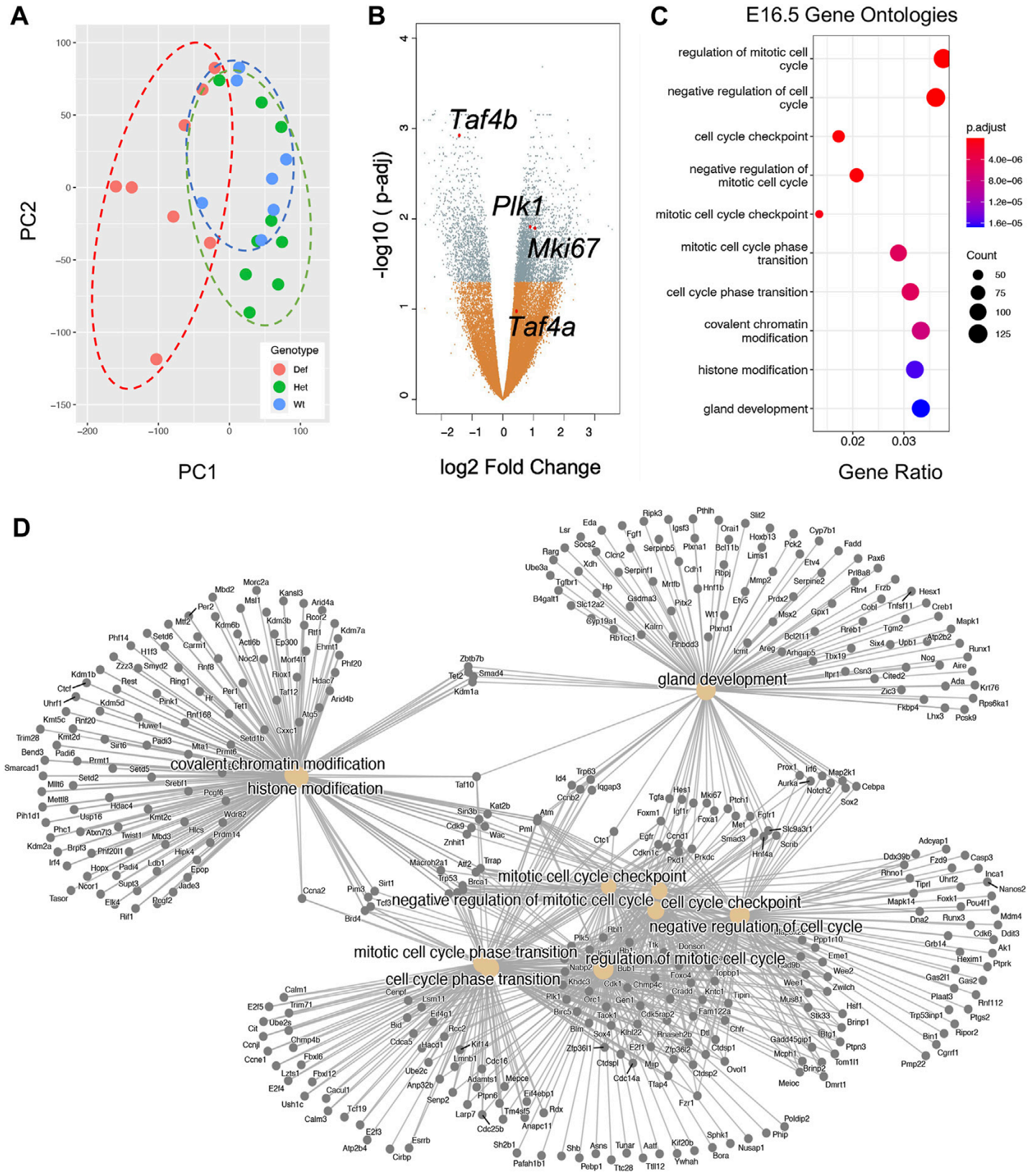


Figure 2.2. E16.5 ProSpg RNA-seq experiment reveals significant regulation of mitotic gene programs.

(A) PCA plot of the E16.5 samples labeled based on Taf4b genotype as indicated by color and dotted lines. (B) Volcano plot of genes; the 4551 significant genes (p-adj <0.05) are labeled in grey, DEGs of interest plus Taf4b are specified and nonsignificant gene changes are labeled in orange. (C) Dotplot of GO biological process analysis of all 3820 protein-coding DEGs from (B). (D) Category netplot showing the relationships between the top ten GO biological processes from (C) and the individual DEGs within each ontology.

After observing global transcriptional dysregulation with 1,474 downregulated genes and 3077 upregulated genes, we sought to parse out what direction particular biological processes are due to Taf4b ablation. Of the 1,259 downregulated genes that were protein-coding, the top three biological process gene ontology categories were cilium movement, negative regulation of the immune system process, and cell killing (Figures 2.3A,C). From the upregulated genes, 2,561 were protein-coding, and we saw that the biological process categories are primarily related to cell cycle, DNA repair, and covalent chromatin modification (Figures 2.3B,D). Surprisingly, there were almost twice as many upregulated genes than downregulated genes in the absence of TAF4b, many of which encoded critical components of the cell cycle and chromatin regulatory machinery.

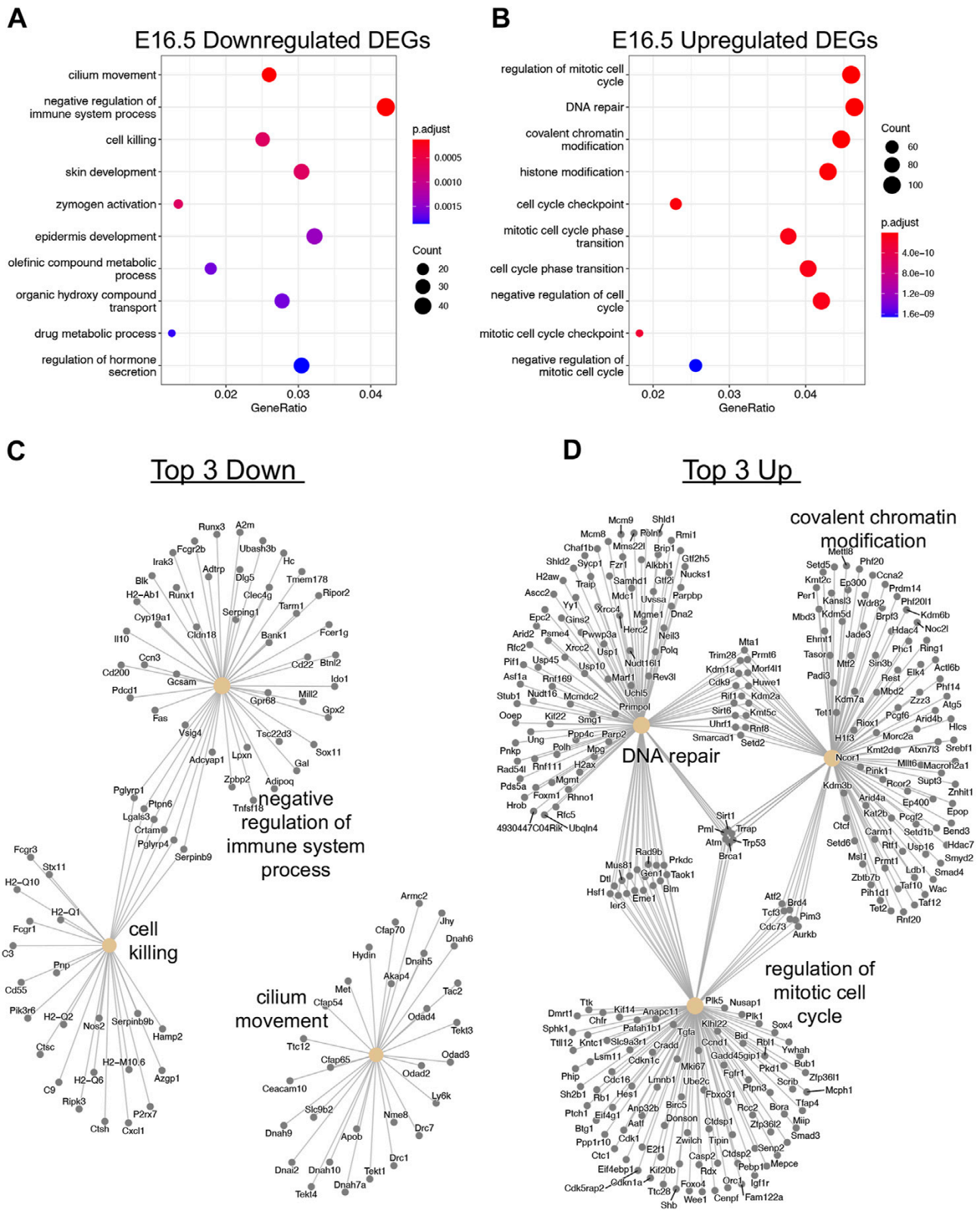


Figure 2.3. Transcriptional dysregulation of cell cycle and chromatin modification genes.

(A) Dotplot of GO biological process analysis of 1,259 protein-coding Down DEGs. (B) Dotplot of GO biological process analysis of 2,561 protein-coding Up DEGs. (C) A category netplot composed of the top three downregulated biological processes from (A) and the individual DEGs within each ontology. (D) A category netplot composed of the top three upregulated biological processes from (B) and the individual DEGs within each ontology.

Taf4b-deficient ProSpg are delayed in their entry into quiescence

Mitotic ProSpg enter a characteristic period of quiescence between E13.5 and E16.5. Based upon our scRNA-seq and bulk RNA-seq data, we hypothesized that TAF4b modulated some aspect of ProSpg quiescence. We performed immunofluorescence on Taf4b-wildtype and Taf4b-deficient testis sections to detect TRA98 (a germ cell marker) and MKi67 (a marker of cell proliferation) from E13.5 to E18.5 (Grive et al., 2014; Sun and Kaufman, 2018). Representative images of each genotype and time point are shown (Figure 2.4A), as well as the quantification of the MKi67- quiescent wildtype and Taf4b-deficient ProSpg (indicated by the TRA98-only signal) at each timepoint (Figure 2.4B). At E13.5, the majority of Taf4b-deficient M ProSpg were still proliferating as indicated by the nuclear localization of MKi67 (Figure 2.4A). At E14.5, there is a significant reduction of ProSpg entering quiescence in the Taf4b-deficient compared to wildtype controls (Figure 2.4B). Interestingly, the percentage of quiescent cells at both E16.5 to E18.5 did not differ significantly between controls and Taf4b-deficient Pro-Spg. These data indicate that although Taf4b-defcient ProSpg can largely achieve quiescence, they struggle with its timely entry and this transition. We also compared staining for the proliferating cell nuclear antigen (PCNA), which is a DNA replication and repair protein, across matched Taf4b-wildtype and Taf4b-deficient testis sections (Figure 2.4C). At E16.5, there is a significant increase in the percent of Taf4b-deficient ProSpg expressing PCNA compared to

matched wildtype controls and this difference is non-significant at E18.5 (Figure 2.4D). Taken together, these data indicate that Taf4b-deficient ProSpg are altered in their timing of ProSpg quiescence entry.

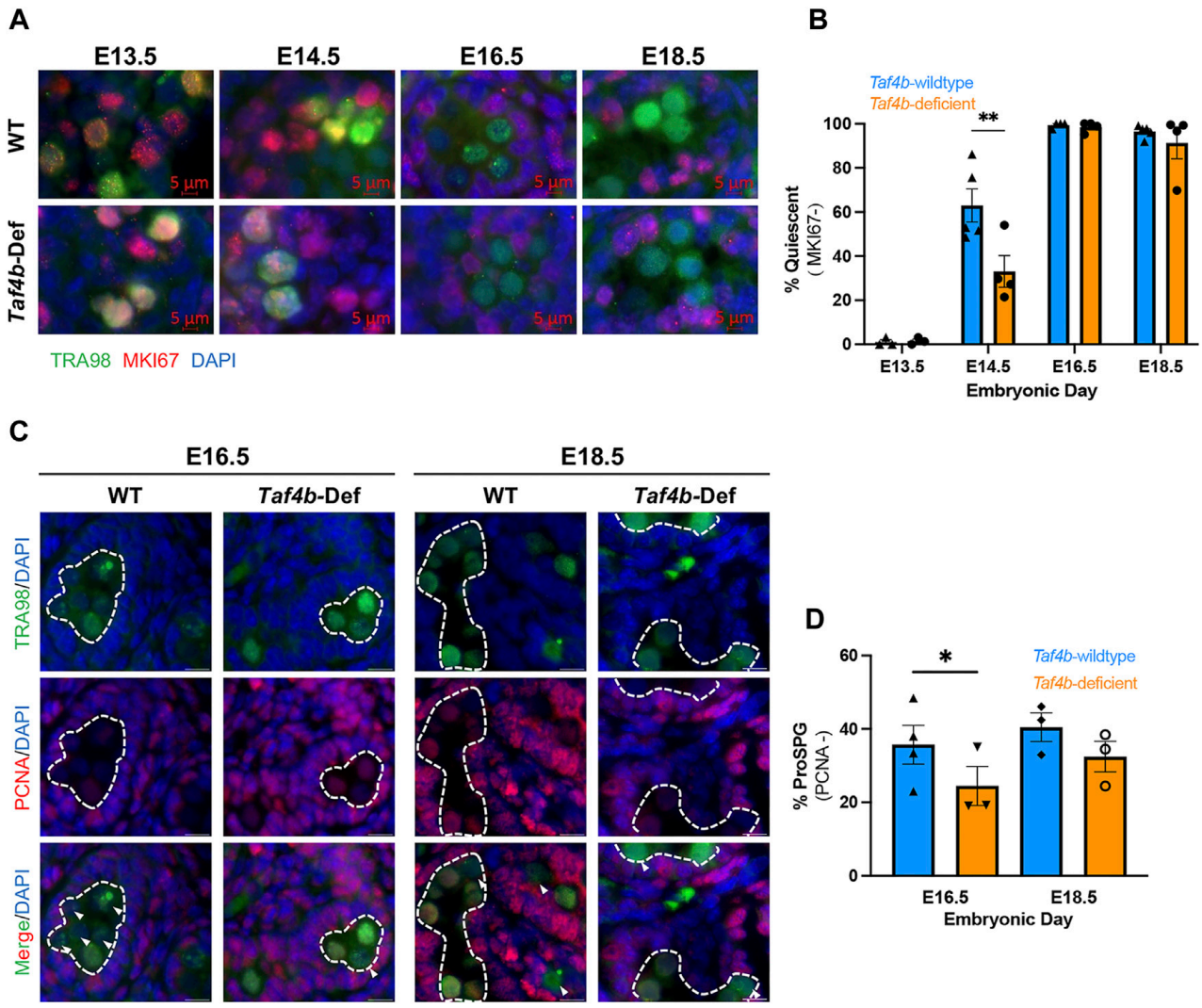


Figure 2.4. Taf4b-deficient ProSpg have increased MKI67 at E14.5.
(A) Immunofluorescence of E13.5-E18.5 testis sections comparing Taf4b-wildtype (TAF4b +/+) to Taf4b-deficient ProSpg (TAF4b −/−). Antibodies for TRA98 (green), MKI67 (red) were used, and DAPI labels DNA. (B) Quantification of MKI67-/TRA98+ cells in Taf4b-wildtype (blue) and Taf4b-deficient (orange) testis sections. Taf4b-deficient ProSpg have significantly fewer MKI67- ProSpg than wildtype at E14.5 (n = 3-5 mice, ** = p-adj< 0.01 as determined by two-way ANOVA). (C) Immunofluorescence of E16.5-E18.5 testis sections comparing Taf4b-wildtype (TAF4b +/+) to Taf4b-deficient ProSpg (TAF4b −/−). Antibodies for TRA98 (green), PCNA (red) were used, and DAPI (blue) labels DNA. (D) Quantification of PCNA−/TRA98+ cells in Taf4b-wildtype (blue) and

Taf4b-deficient (orange) testis sections. Taf4b-deficient ProSpg have significantly fewer PCNA− ProSpg than wildtype at E16.5 (n = 3-4 mice, * = p-value< 0.05 as determined by paired t-test).

CUT&RUN identifies direct targets of TAF4b in E16.5 ProSpg

To distinguish which DEGs identified in our E16.5 bulk RNA-seq experiment were direct targets of TAF4b, we performed Cleavage Under Targets and Release Using Nuclease (CUT&RUN), a genome mapping technique to identify enrichment of specific proteins and histone modifications in the genome (Skene and Henikoff, 2017; Hainer and Fazzio, 2019). We isolated two replicates of wildtype Oct4-eGFP E16.5 male germ cells using FACS and examined the genomic localization of TAF4b. H3K4me3 served as a positive control and marked active promoters, and IgG served as a negative control. CUT&RUN data analysis using the program Homer identified 64,891 H3K4me3 peaks and 7,861 TAF4b peaks in Replicate 1 and 1,730 H3K4me3 peaks and 848 TAF4b peaks in Replicate 2. The differences in the number of peaks called can be attributed to stochastic differences in sample preparation and starting material. However, it is necessary to note that these differences did not affect the assay's specificity because we still see the same top motifs and the greater than 80% overlap between each replicate. We also found that 73% and 88% of TAF4b peaks were classified as localizing to promoters/transcription start site ("promoter-TSS") for Replicate 1 and 2, respectively (Figure 2.5A). Importantly, 617 TAF4b peaks overlapped between the two replicates allowing us to perform GO analysis of the shared TAF4b-bound gene promoter-TSSs between the two replicates. These top five categories include mRNA processing, proteasomal degredation, RNA splicing, covalent chromatin modification, and negative regulation of the cell cycle (Figure 2.5B). When plotting the enrichment profile of TAF4b relative to TSSs, we found the

highest TAF4b enrichment was located just upstream of the TSS (Figure 2.5C). This initial chromatin mapping of TAF4b indicates that the majority of the binding sites for TAF4b are just upstream of the TSS which is consistent with where TFIID is known to bind to core promoters.

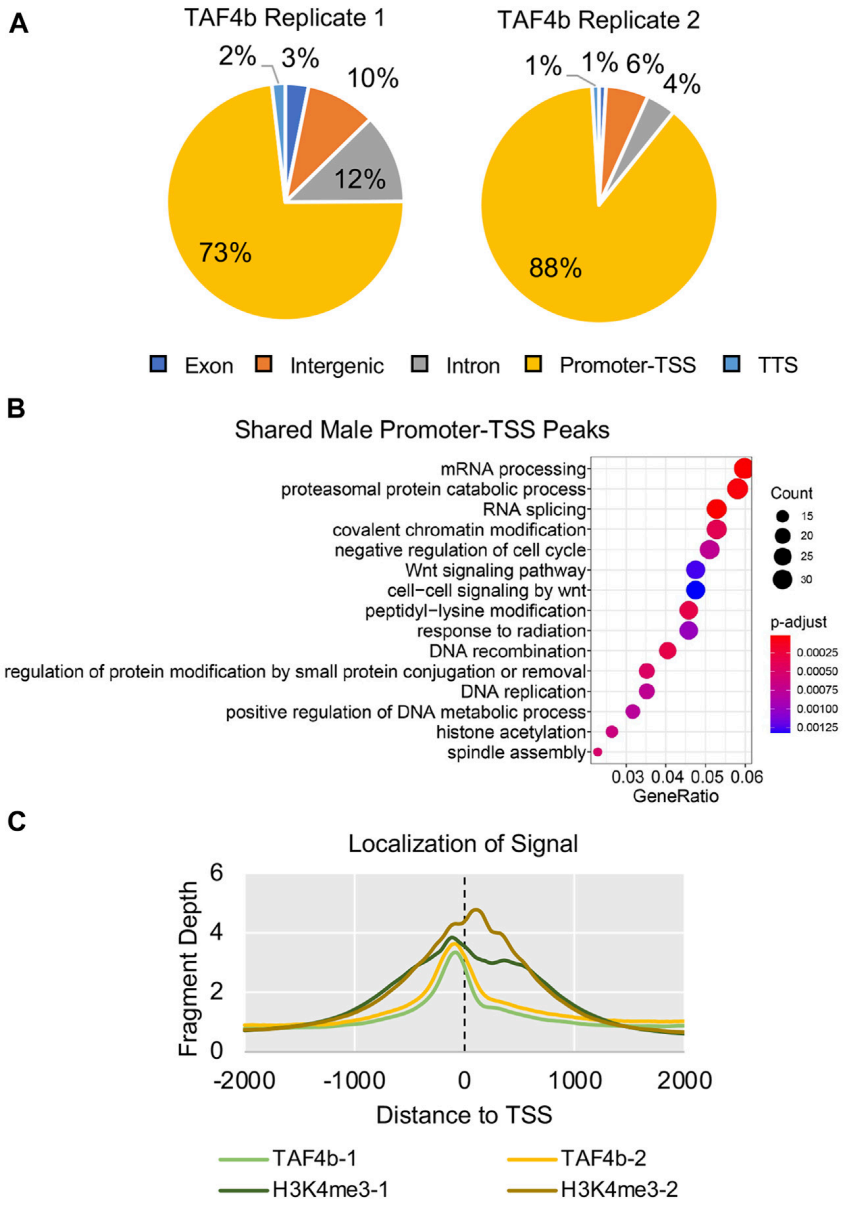


Figure 2.5. Genome-wide Occupancy of H3K4me3 and TAF4b in E16.5 ProSpg. CUT& RUN of E16.5 ProSpg using antibodies for H3K4me3 and TAF4b.
(A) Pie charts of TAF4b peak locations across the E16.5 ProSpg genome in two independent CUT&RUN replicates. (B) GO biological process dotplot for the shared CUT&RUN peaks categorized as "promoter-TSS". (C) Average enrichment of TAF4b and H3K4me3 signal near TSSs (dotted line) for each replicate.

We then explored the motifs associated with TAF4b-enriched peaks within our CUT&RUN data and found consistent enrichment of CCAAT and GC motifs, which are associated with the NFY complex and the Sp/Klf transcription factor family, respectively (Figures 2.6A,B) (Suske, 2017). Importantly, these motifs were shared between the two CUT&RUN replicates. We then compared DEGs from our RNA-seq experiment to direct targets of TAF4b from the CUT&RUN to identify a subset of genes directly modulated by TAF4b at E16.5. When comparing our DEGs to the "promoter-TSS" peaks of TAF4b, we found 876 DEGs that had at least one peak near their TSS. GO analysis of these 876 TAF4b bound and differentially expressed genes reveals that the most notable biological process affected is "regulation of the mitotic cell cycle" (Figure 2.6C). A volcano plot of these TAF4b-bound E16.5 DEGs revealed far more upregulated DEGs (674 genes) than downregulated DEGs (202 genes) in Taf4b-deficient ProSpg (Figure 2.6D). As examples of TAF4b-bound DEGs, we present TSS gene tracks for Nuclear Transcription Factor Y Subunit Alpha (Nfya), Protein Regulator of cytokinesis 1 (Prc1), H2A histone family member X (H2afx), Kruppel-like factor 6 (Klf6), and Cyclin Dependent Kinase 20 (Cdk20), with TAF4b and H3K4me3 enrichment compared with control IgG (Figure 2.6E). Nfya is a component of the NFY transcription factor complex and interestingly TAF4b peaks were located between the TSS of Nfya and another gene Oard1 which was not a DEG. Prc1 encodes a microtubule-binding protein that plays a role in mitosis and binds to Plk1

another critical mitosis gene (Hernández-Ortega et al., 2019). H2afx encodes an essential DNA damage response protein and has been previously found to be increased in Taf4b-deficient oocytes (Grive et al., 2016). Klf6 encodes a transcription factor that plays a role in several developmental processes such as cell proliferation and differentiation (Hu et al., 2021). Cdk20, decreased in Taf4b-deficient ProSpg, encodes a positive regulator of the cell cycle by activating CDK2 and Cyclin D (Lai et al., 2020). A protein-protein interaction network of these 876 TAF4b-bound DEGs revealed Cdk1 and other cell cycle-related genes as a major group of associated genes and interestingly the entire NFY complex (encoded by Nfya, Nfyb, and Nfyc) was present (Supplementary Figure S2.1).

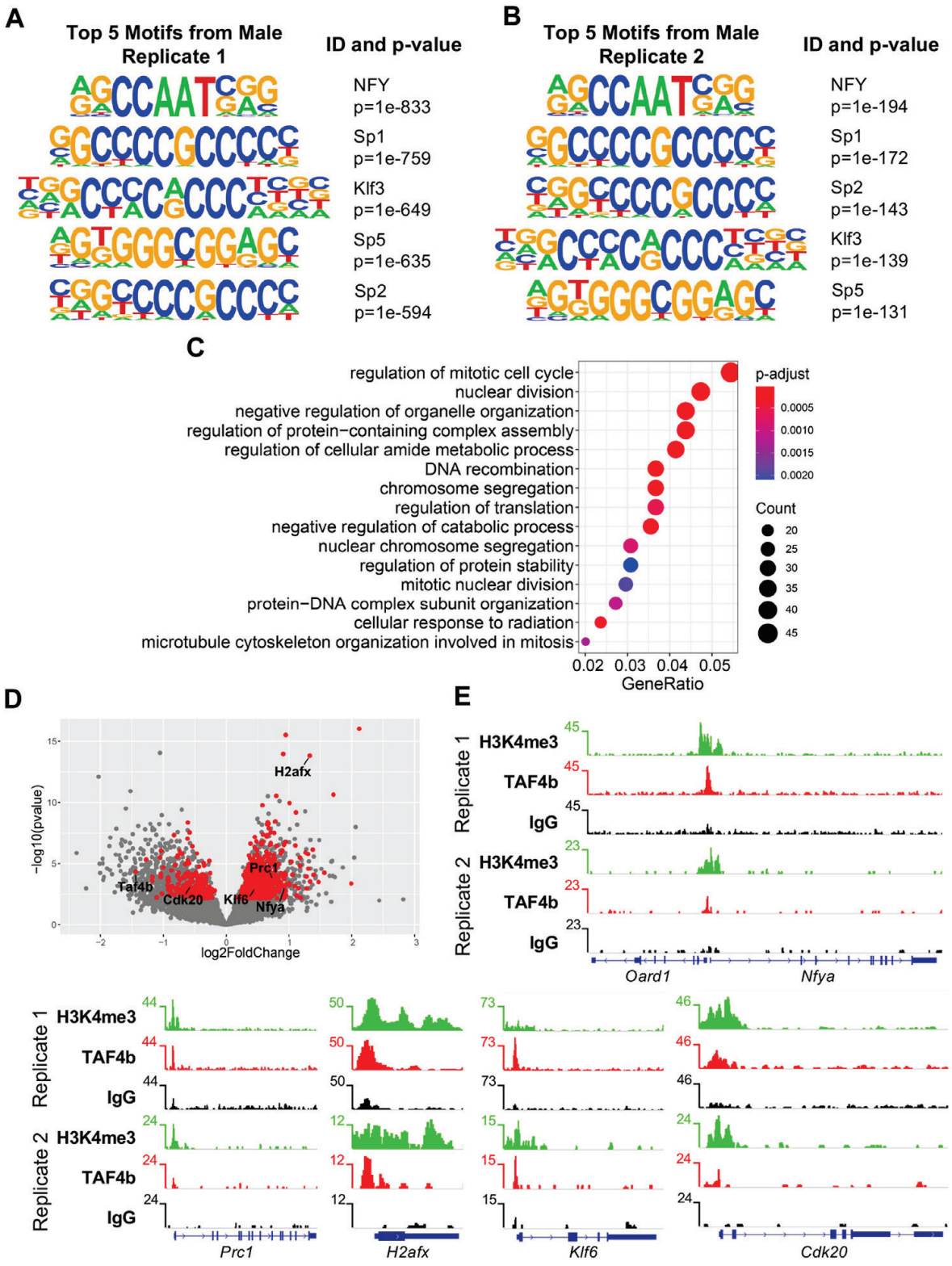


Figure 2.6. CUT&RUN and RNA-seq integration identify TAF4b bound and modulated cell cycle genes.

(A) Top five TAF4b motifs from “promoter-TSS” peaks in Top five TAF4b motifs from “promoter-TSS” peaks in male Replicate 1. (B) Top five TAF4b motifs from “promoter-

TSS" peaks in male Replicate 2. (C) Biological process GO dotplot of the 876 genes that are DEGs and had a TAF4b promoter-TSS peak in at least one of the two germ cell samples. (D) Volcano plot of the 876 DEGs that had at least one TAF4b promoter peak (red dots). (E) Gene track of Oard1 and Nfya, in which Nfya was a DEG that had a TAF4b promoter-TSS called in both replicates. Gene tracks of Prc1, H2afx, Klf6, and Cdk20, which were DEGs (labeled red dots in D) that had a TAF4b promoter-TSS called in both replicates.

TAF4b-bound and differentially expressed ProSpg gene promoters are enriched for related specificity protein and kruppel-like family (SP/KLF) of transcription factor binding motifs

When integrating our CUT&RUN "promoter-TSS" peaks with TAF4b DEGs, we again found an enrichment for SP/KLF transcription factor binding motifs (Figure 2.7A). However when we looked at the integrated data we no longer observed NFY in the top five overall binding motifs. The surprising absence of NFY sites in this analysis prompted us to further parse these data into TAF4b peaks whose levels of mRNA decreased (Downregulated DEGs) or increased (Upregulated DEGs) in TAF4b-deficient ProSpg. We first looked at the occupancy of TAF4b and noted that it centered around the −50 bp region upstream of the TSS, further supporting a role for TAF4b in core promoter recognition and transcription initiation. It was also observed that there was a slight difference in the median being approximately 5 bases futher upsteam in the Up DEGs as compared to the Down DEGs (Figure 2.7B). Interestingly, we discovered that NFY is actually the top Down DEG motif while Sp1 is the top Up DEG motif (Figures 2.7C,D). The differences in top motifs between decreased and increased mRNA levels may allude to the multiple modes of transcriptional modulation by TAF4b. As follow-up, we mapped TAF4b targets to the GO analysis from bulk RNA-seq to determine which GO categories were directly modulated. Most notable in this analysis were the categories of cell cycle

regulation, covalent chromatin modification, and DNA repair which had increased mRNA expression levels in the absence of TAF4b and also contained 7 genes that had TAF4b-bound promoters (Figure 2.7E). Finally, we quantified the overall fraction of CUT&RUN peaks that were also DEGs and found significantly more mRNAs that were increased in the absence of TAF4b to contain a TAF4b-bound promoter (Figure 2.7F). Overall, these data suggest that TAF4b specifically controls the levels of cell cycle and chromatin regulating gene expression programs required for ProSpg to properly navigate the transition from cycling to quiescence during embryonic mammalian testes development.

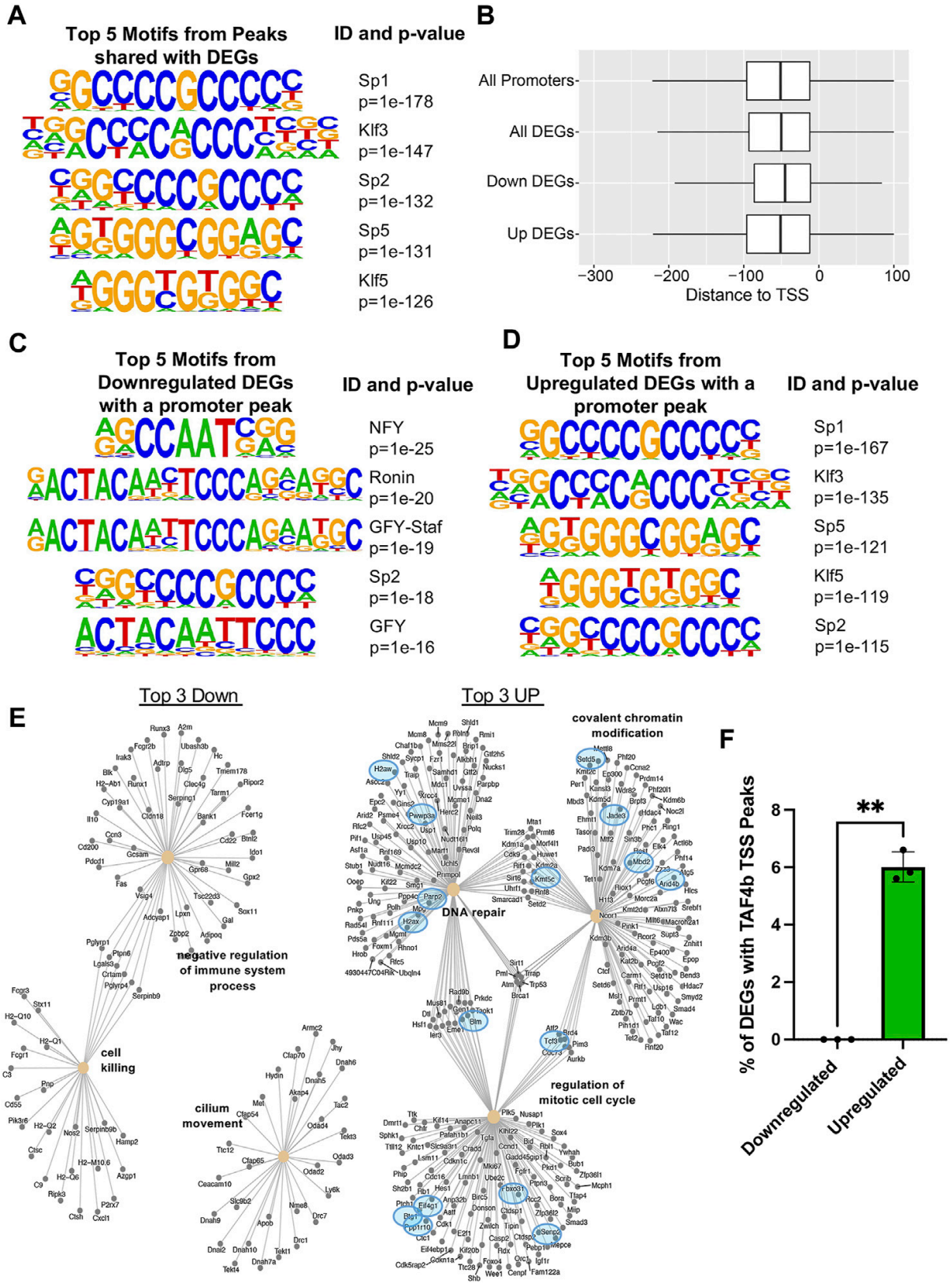


Figure 2.7. Sp/Klf motif enrichment at cell cycle and chromatin regulating TAF4b bound gene promoters.

(A) Top five motifs enriched at TAF4b “promoter-TSS” peaks for genes that were also DEGs, the promoter ID, and the associated p-value. (B) Boxplots (no outliers included) of promoter TAF4b peaks relative to the TSS with a median of −53 bp, all TAF4b “promoter-TSS” peaks for genes that were also DEGs with a median of −51 bp, TAF4b “promoter-TSS” peaks for genes that were only Downregulated DEGs with a median of −49 bp and TAF4b “promoter-TSS” peaks for genes that were only Upregulated DEGs with a median of −53 bp. (C) Top five TAF4b associated binding motifs from “promoter-TSS” peaks in Down DEGs and (D) Up DEGs. (E) Category netplots composed of the top three downregulated (left) and upregulated (right) biological processes from (Figures 3C,D) along with DEGs that contained TAF4b peaks at their TSS (blue circles). (F) Comparison of TAF4b target DEG percentages from top three upregulated and downregulated GO categories. Significance determined by Welch’s t-test, ** = p-value <0.01.

Discussion

Soon after sex determination in mammals, embryonic female and male germ cells receive distinct signals to initiate a sex-specific germ cell differentiation program (Capel, 2019). However, it is unclear how these newly defined germ cells integrate external signals to execute precise and sex-specific gene expression programs and cell cycle states required for germ cell development. We have discovered that the embryonic germline-enriched TFIID subunit, TAF4b, is required for male and female germ cell development and maintenance in mice. Here, we identify many ProSpg genes whose expression are affected by the loss of TAF4b and whose core promoter regions display TAF4b occupancy just upstream of the TSS. TAF4b enrichment at ProSpg core promoters correlates with the binding sites for Sp1 and NFY, two well-characterized and ubiquitous promoter-proximal and sequence-specific transcription factors. The integration of these high throughput genomic sequencing and mapping data indicate that TAF4b fine tunes male-specific cell cycle gene expression required for the timely entry of ProSpg into quiescence via complex and unexpected mechanisms.

Several notable regulators of post-transcriptional gene expression play similar developmental functions as TAF4b when ProSpg transition towards quiescence. The most notable is the RNA-binding protein DND1 which displays intriguing parallels with the functions of TAF4b shown here. DND1 deficiency in mice leads to defects of prospermatogonial entry into quiescence and increased germ cell loss (Cook et al., 2011). However, the Ter mutation in the Dnd1 gene on the 129 genetic background strikingly results in spontaneous teratoma formation at E16.5 (Cook et al., 2011). While Taf4b-deficient testes display similar kinetics in male germ cell loss and fertility disruption, our Taf4b-deficiency presented here is on a C57Bl/6 background where we might not expect to see these teratomas (Falender et al., 2005). Remarkably, TAF4b and DND1 share common regulation of mitotic cell cycle and chromatin modification gene expression, suggesting they are two distinct modes of regulating a common ProSpg quiescence and survival program ((Ruthig et al., 2019), this study). A second RNA-binding protein, NANOS2, is also responsible for ProSpg development and has been recently shown to be bound to and work with DND1 in regulating specific RNA loading into the CNOT deadenylase complex (Suzuki et al., 2016; Ruthig et al., 2019). Interestingly, Nanos2 is an upregulated DEG, and Dnd1 is a downregulated DEG in E16.5 Taf4b-deficient ProSpg. The developmental similarities of perturbed quiescence, increased germ cell loss, and infertility in Nanos2, Dnd1, and Taf4b mouse mutants suggest that regulating the gene expression of prospermatogonia transitions is critical to properly set up the future spermatogonia that arise from this lineage in the postnatal testis. The dysregulation of Nanos2 and Dnd1 in the absence of TAF4b also indicates cross-talk between these crucial gene expression regulators and the sensitivity of germline development.

There are also several key transcription factors that direct cell cycle progression and transcription during prospermatogonial development. The master regulator of the G1 phase of the cell cycle, RB1, is required for the timely mitotic arrest of early prospermatogonia (Spiller et al., 2010). A recent germ cell-specific knockout approach for RB1 linked this cell cycle defect to disrupting the metabolic nature of these developing prospermatogonia and their corresponding ability to become SSCs (Du et al., 2021). While the timing of RB1 and TAF4b functions in promoting quiescence are similar, RB1 deficiency leads to much more severe cell cycle defects than Taf4b-deficiency, suggesting they play more complementary roles in SSC development (Du et al., 2021). RHOX10 is a homeodomain family sequence-specific DNA-binding protein found in the reproductive homeobox cluster of the X chromosome and is required for prospermatogonial and SSC development. A recent genomic study using similar tools implemented here uncovered a network of transcription factozrs that are direct targets of RHOX10, most notably DMRT1 and ZBTB16 (Tan et al., 2021). While there are minor differences in the timing and cells used in these two studies, Tan et al., 2021 uncovered a critical CCAAT box bound by RHOX10 in the Dmrt1 promoter and predicted it to be a binding site for the NFY transcription factor, which was the most correlated binding site at TAF4b bound core promoters (Tan et al., 2021). It will be interesting to determine how these transcriptional regulators work together and separately to promote prospermatogonial development required for the proper establishment of SSCs and longterm mammalian spermatogenesis.

The precise molecular mechanism of TAF4b in promoting ProSpg transitions remains unknown, but several interesting new twists are revealed here. TFIID was first discovered as a ubiquitous multiprotein core promoter binding and coactivator complex consisting of the TATA-box binding protein and 14 TBP-associated factors (TAFs) required for Sp1-dependent transcriptional activation in vitro (Pugh, 1990). However, determining Sp1 as the top binding motif at the cell cycle and chromatin modification gene promoters bound by TAF4b in ProSpg was surprising, given this unique cell type-specificity. Moreover, mRNA expression levels of these gene programs primarily increase in the absence of TAF4b, suggesting TAF4b plays some role in limiting their expression as ProSpg transitions to quiescence. Future studies will determine whether TAF4b works in a chromatin modification capacity along with its canonical function in the TFIID complex. As TAF4b regulates chromatin modification genes, we also plan on determining if changes in chromatin accessibility are a strong driver of our observed gene expression changes. This non-canonical role of TAF4b may contribute to proper silencing and timely repression of specific genes during spermatogenesis. We also know now that other members of the TFIID complex have been found in chromatin-modifying complexes like that of the Spt/Ada/Gcn5 acetyltransferase (SAGA) complex. Members such as TAF5L, TAF6L, and TAF9B are just some with documented dual activity (Soffers and Workman, 2020). Another possibility is that as later T2 ProSpg return to proliferation quickly after birth, TAF4b helps mark the cell cycle and chromatin modifying genes for future activation during later stages of postnatal germ cell expansion. Lastly, the TATA-box that marks a subset of TFIID-dependent promoters is usually found between −25 bp and −30 bp upstream of the TSS, and our peaks of TAF4b binding center around the −50 bp region

suggesting that TAF4b might extend the upstream boundary of the TFIID footprint at the core promoter by 20 bp (Sloutskin et al., 2021). Future studies of ProSpg core promoter architecture and TAF4b′s transcriptional mechanism with individual SP/KIF and NFY family members will be required to understand these novel aspects of ProSpg transcription regulation.

While we focus here on TAF4b's role in promoting the early development of the male germline, we have recently reported that at identical embryonic time points in ovarian development, TAF4b plays a similar critical role in early oocyte differentiation (Gura et al., 2022). As female and male embryonic germ cells express unique sets of genes and navigate different biological processes, i.e., meiotic initiation vs. mitotic arrest, it is surprising that the same transcription factor would have distinct functions in the early life of these future gametes. We propose that one common function of TAF4b, and its associated regulatory complex, is to promote the embryonic germ cell identity and survival after PGC specification, analogous to the germ cell licensing function of DAZL (Gill et al., 2011). In addition, TAF4b has likely evolved sex-specific functions that allow it to promote context-specific transcription events in both male and female developing germ cells. One commonality of these two populations is that even though they enter meiosis at different times, they both exit the mitotic cell cycle at the end of PGC development, and TAF4b, along with many other unknown regulatory proteins, helps direct these transitions. Finally, genome-wide epigenetic modifications are erased in both female and male germ cell populations at the end of PGC development. In the context of a more open genome, TAF4b and its related regulators could help maintain the appropriate chromatin

environment for controlling transcription-coupled DNA repair mechanisms unique and critical to the germline genome of both sexes and, ultimately, the next-generation.

Materials and methods

Mice

Mice, homozygous for an Oct4-eGFP transgene (The Jackson Laboratory: B6; 129S4-Pou5f1tm2Jae/J), were mated for CUT&RUN collections. Mice, homozygous for an Oct4-eGFP transgene (The Jackson Laboratory: B6; 129S4-Pou5f1tm2Jae/J) and heterozygous for the Taf4b-deficiency mutation (in exon 12 of the 15 total exons of the Taf4b gene that disrupts the endogenous Taf4b gene), were mated for mRNA collections. Timed matings were estimated to begin at day 0.5 by evidence of a copulatory plug. The sex of the embryos was identified by confirming the presence or absence of testicular cords. Genomic DNA, isolated from tail biopsies using QIAGEN DNeasy Blood & Tissue Kits (Cat #: 69506), was used for PCR genotyping assays.

All animal protocols were reviewed and approved by Brown University Institutional Animal Care and Use Committee and were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals. Gonads were dissected out of embryos into cold PBS.

Immunofluorescence

Testes were gathered from embryonic mice (E13.5-E18.5), briefly fixed with 10% formalin and paraffin-embedded. The Molecular Pathology Core at Brown University completed

tissue preparation and performed sectioning. After the serial sections were prepared, antigen retrieval was performed using a 0.1% antigen unmasking solution (Vector Laboratories, H-3300) in a steamer for 20 min. Samples were then permeabilized in 0.1% sodium citrate and Triton X-100 for 10 min, washed with 0.1% PBST (Triton X-100), and incubated in blocking buffer (PBS with 3% goat serum, 1% BSA, and 0.1% Triton X-100) for 1 hour. This was then followed by primary antibody incubation for 24 h at 4°C prepared in the blocking buffer. The samples were then washed in PBST, incubated with secondary antibody for 1 h at 37°C, washed in PBST, and mounted in DAPI-containing Vectashield Mounting Medium (Vector Laboratories H-1200, Burlingame, CA). Primary antibodies used were rat anti-TRA98 (1:100, abcam, ab82527), anti-KI67 rabbit mAb (1:100, Cell Signaling Technology D3B5, 9129S) and anti-PCNA (1:100, Cell Signaling Technology, 13110T). For immunofluorescence, the secondary antibodies used were Alexa Fluor 555-conjugated anti-rat IgG (1:500; abcam ab150158) and Alexa Fluor 488-conjugated anti-rabbit IgG (1:500; abcam ab150081). Images were all acquired at the same exposures and received the same immunofluorescence preparation. Visualization and counting were completed via blinded assessment of TRA98-expressing cells colocalized with MKI67 on a Zeiss Axio Imager M1 and Zeis software. A two-way ANOVA was used to determine significance using GraphPad Prism 9 version 9.3.1. for MacOS (GraphPad Software, Boston, Massachusetts United States, www.graphpad.com) with an n of 3-5 mice.

Embryonic gonad dissociation and fluorescence-activated cell sorting

To dissociate gonadal tissue into a single-cell suspension, embryonic gonads were harvested and placed in 0.25% Trypsin/EDTA and incubated at 37°C for 15 and 25 min

for E14.5 and E16.5 gonads, respectively, as previously described (Gura et al., 2020). Eppendorf tubes were flicked to dissociate tissue halfway through and again at the end of the incubation. Trypsin was neutralized with FBS. Cells were pelleted at 1,500 RPM for 5 min, the supernatant was removed, and cells were resuspended in 100 µL PBS. The cell suspension was strained through a 35 µm mesh cap into a FACS tube (Gibco # 352235). Propidium iodide (1:500) was added to the cell suspension as a live/dead distinguishing stain. Fluorescence-activated cell sorting (FACS) was performed using a Becton Dickinson FACSAria III in the Flow Cytometry and Cell Sorting Core Facility at Brown University. A negative control of a non-transgenic mouse gonad was used for each experiment to establish an appropriate GFP signal baseline. Dead cells were discarded and the remaining cells were sorted into GFP+ and GFP− samples in PBS at 4°C for each embryo.

For RNA-seq analysis, GFP+ cells from each individual embryo were kept in separate tubes and were then spun down at 1,500 RPM for 5 min, had PBS removed, and were then resuspended in Trizol (ThermoFisher # 1556026). If samples had roughly less than 50 µL of PBS in the tube, Trizol was added immediately. The number of cells for each sample can be found in Supplementary Table S2.1. Samples were stored at −80°C.

For CUT&RUN, germ cells from all the gonads were pooled prior to FACS. Sorted germ cells were then spun down at 1,500 RPM for 5 min and were resuspended in 300 µL of PBS, then split into three Eppendorf tubes. These three tubes of germ cells were then used for CUT&RUN. The number of cells for each sample were as follows: Replicate 1

germ cell samples had approximately 56,000 cells per tube (obtained from 22 embryos) and Replicate 2 germ cell samples had approximately 131,000 cells per tube (obtained from 28 embryos).

Single cell RNA-seq data analysis

All computational scripts regarding single cell RNA-seq (scRNA-seq) used in this publication are available to the public: https://github.com/mg859337/Gura_et_al._TAF4b_male_transcription. SRP194420, SRP158811, and SRP178196 were downloaded from NCBI SRA onto Brown University's high-performance computing cluster at the Center for Computation and Visualization. We used Law et al. and Nguyen et al., which enriched for germ cells via FACS sorting (Law et al., 2019; Nguyen et al., 2020). We then combined the data with Tan et al., who collected the whole testis, thus establishing our comprehensive observation window from E12.5 to P7 (Tan et al., 2020). The fastq files were aligned using Cell Ranger (v 6.0.0) count and then aggregated using Cell Ranger aggr. The cloupe file created from Cell Ranger aggr was used as input for Loupe Cell Browser (v 5.0). We selected the germ cells by first assessing the log-normalized expression of germ cell marker Dazl within the complete 71,584 cell data set. Then, using unbiased clustering, we eliminated somatic cell clusters based on their mean Dazl expression being less than 0.85. The data set underwent further quality control and filtering based on fitting the metrics of 12.5–15.8 log2-normalized unique molecular identifiers, 10.6–12.8 log2-normalized features, and lastly, less than 24% mitochondrial transcript ratio (Quatredeniers et al., 2023).

RNA-sequencing

Embryonic germ cells resuspended in Trizol were shipped to GENEWIZ (GENEWIZ Inc., NJ) on dry ice. Sample RNA extraction, sample QC, library preparation, sequencing, and initial bioinformatics were done at GENEWIZ. RNA was extracted following the Trizol Reagent User Guide (ThermoFisher Scientific). Glycogen was added (1 µL, 10 mg/mL) to the supernatant to increase RNA recovery. RNA was quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, United States) and RNA integrity was checked with TapeStation (Agilent Technologies, Palo Alto, CA, United States) to see if the concentration met the requirements.

SMART-Seq v4 Ultra Low Input Kit for Sequencing was used for full-length cDNA synthesis and amplification (Clontech, Mountain View, CA), and Illumina Nextera XT Library Preparation Kit was used for library preparation. The sequencing libraries were multiplexed and clustered on a lane of a flowcell. After clustering, the flowcell was loaded onto an Illumina HiSeq 4000 according to manufacturer's instructions. The samples were sequenced using a 2 × 150 Paired End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS) on the HiSeq instrument. Raw sequence data (.bcl files) generated from Illumina HiSeq were converted into fastq files and de-multiplexed using bcl2fastq (v 2.17). One mismatch was allowed for index sequence identification.

RNA-seq data analysis

All computational scripts regarding RNA-seq used in this publication are available to the public: https://github.com/mg859337/Gura_et_al._TAF4b_male_transcription. All raw fastq files were initially processed on Brown University's high-performance computing cluster. Reads were quality-trimmed and had adapters removed using Trim Galore! (v 0.5.0) with the parameters–nextera -q 10. Samples before and after trimming were analyzed using FastQC (v 0.11.5) for quality and then aligned to the Ensembl GRCm38 using HiSat2 (v 2.1.0) (Andrews, 2010; Pertea et al., 2016). Resulting sam files were converted to bam files using Samtools (v 1.9) (Li et al., 2009).

To obtain TPMs for each sample, StringTie (v 1.3.3b) was used with the optional parameters -A and -e. A gtf file for each sample was downloaded and, using RStudio (R v 4.0.2), TPMs of all samples were aggregated into one comma separated (csv) file using a custom R script. To create interactive Microsoft Excel files for exploring the TPMs of each dataset: the csv of aggregated TPMs was saved as an Excel spreadsheet (Supplementary Table S2.2). Colored tabs were added to set up different comparisons, and a flexible Excel function was created to adjust to gene name inputs. To explore the Excel files, please find the appropriate tab (named "Quick_Calc") and type in the gene name of interest into the highlighted yellow boxes.

To obtain count tables, featurecounts (Subread v 1.6.2) was used (Liao et al., 2014). Metadata files for each dataset were created manually in Excel and saved as a csv. These count tables were used to create PCA plots by variance-stabilizing transformation (vst) of the data in DESeq2 (v 1.22.2) and plotting by ggplot2 (v 3.1.0) (Love et al., 2014;

Wickham, 2016). DESeq2 was also used for differential gene expression analysis, where count tables and metadata files were used as input. We accounted for the litter batch effect in our mouse germ cells by setting it as a batch parameter in DESeq2. For the volcano plot, the output of DESeq2 was used and plotted using ggplot2. DEG lists were used for ClusterProfiler (v 3.16.1) input to create dotplots of significantly enriched gene ontology (GO) categories for all DEGs. The highest-confidence protein-protein interactions were identified using STRING, with unconnected proteins not shown in the images (Szklarczyk et al., 2019). Net plots were constructed using Bioconductor packages in R after normalization and QC. The top three GO processes category nets for both Up and Downregulated DEGs were cross-referenced with the TAF4b-bound gene TSS list.

CUT&RUN

The CUT&RUN performed in E16.5 germ cells followed the protocol in Hainer and Fazzio, 2019 (Hainer and Fazzio, 2019). CUT&RUN antibodies were as follows: polyclonal rabbit TAF4b (as previously described (Grive et al., 2016)), monoclonal rabbit H3K4me3 (EMD Millipore # 05-745R), rabbit IgG (ThermoFisher # 02–6102), pA-MNase was a generous gift from Dr. Thomas Fazzio, UMASS Med. For library preparation, the KAPA HyperPrep Kit (Roche Cat. No 07962363001) was used with New England Biolabs NEBNext Multiplex Oligos for Illumina (NEB #E7335). After library amplification through PCR, libraries were size-selected through gel extraction (~150–650 bp) and cleaned up using the QIAGEN QIAquick Gel Extraction Kit (Cat. # 28704). CUT&RUN libraries in EB buffer

were shipped to GENEWIZ (GENEWIZ Inc., NJ) on dry ice. Sample QC, sequencing, and initial bioinformatics were done at GENEWIZ.

The sequencing libraries were validated on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, United States), and quantified by using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) as well as by quantitative PCR (KAPA Biosystems, Wilmington, MA, United States). The sequencing libraries were clustered on flowcells. After clustering, the flowcells were loaded onto the Illumina HiSeq instrument (4,000 or equivalent) according to manufacturer's instructions. The samples were sequenced using a 2 × 150bp Paired End (PE) configuration. Raw sequence data (.bcl files) generated from Illumina HiSeq were converted into fastq files and de-multiplexed using bcl2fastq (v. 2.20). One mismatch was allowed for index sequence identification.

CUT&RUN data analysis

All computational scripts regarding CUT&RUN data analysis used in this publication are available at: https://github.com/mg859337/Gura_et_al._TAF4b_male_transcription and based on other CUT&RUN publications (Hainer and Fazzio, 2019). All raw fastq files were initially processed on Brown University's high-performance computing cluster. Reads were quality-trimmed and had adapters removed using Trim Galore! (v 0.5.0) with the parameter -q 10 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Samples before and after trimming were analyzed using FastQC (v 0.11.5) for quality and then aligned to the Ensembl GRCm39 using Bowtie2 (v 2.3.0). Fastq screen (v 0.13.0) was used to determine the percentage of reads uniquely mapped to the mouse genome

in comparison to other species. Resulting sam files were converted to bam files, then unmapped, duplicated reads, and low quality mapped were removed using Samtools (v 1.9). Resulting bam files were split into size classes using a Unix script. For calling peaks, annotating peaks, and identifying coverage around TSSs, Homer (v 4.10) was used (Heinz et al., 2010). For gene track visualization, the final bam file before splitting into size classes was used as input to Integrative Genomics Viewer (IGV) (Robinson et al., 2011). A custom genome was created using a genome fasta and gtf file for Ensembl GRCm39.

Pie charts were created using data from Homer output and Venn diagrams were created using BioVenn. Dotplots of Promoter-TSS peaks were made using ClusterProfiler. TSS plots were created using the “tss” function of Homer and plotted using Microsoft Excel. All plots produced in RStudio were saved as an EPS file type and then opened in Adobe Illustrator in order to export a high-quality JPEG image. Excel file of the composite CUT&RUN replicates is presented as several unique tabs in Supplementary Table S2.3.

Data availability statement

The male mouse E16.5 RNA-sequencing data are available from NCBI GEO under accession number GSE188351. The male mouse E16.5 CUT&RUN sequencing data are available from NCBI GEO under accession number GSE188701. The sequencing datasets accessed in this research are from the follow accession numbers: the scRNA-seq mouse data from E12.5 to P7 male sorted Oct4-eGFP gonads used for Figure 1 was obtained through NCBI GEO: Q9 GSE119045, GSE124904, and GSE130593.

Ethics statement

This study was approved by Brown University IACUC protocol #21-02-0005. The primary method of euthanasia is CO2 inhalation and the secondary method used is cervical dislocation both as per American Veterinary Medical Association (AVMA) guidelines on euthanasia. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

MB: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. MG: Conceptualization, Formal Analysis, Funding acquisition, Validation, Writing–review and editing, Data curation, Investigation, Methodology, Resources, Software, Visualization, Writing–original draft. KA: Writing–review and editing, Data curation, Investigation, Methodology, Software. SR: Data curation, Investigation, Methodology, Writing–review and editing. KS: Data curation, Investigation, Methodology, Writing–review and editing, Conceptualization, Project administration, Resources, Visualization. RF: Conceptualization, Project administration, Writing–review and editing, Formal Analysis, Funding acquisition, Supervision, Validation.

Support

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The Brown University Flow Cytometry and Sorting Facility has

received generous support in part by the National Institutes of Health (NCRR Grant No. 1S10RR021051) and the Division of Biology and Medicine, Brown University. As some of our insights were gained by reprocessing publicly available scRNA-seq datasets, we greatly appreciate both the researchers that generated and shared the data initially and the respective repositories for making them available. We are grateful to the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) for their generous support through awards 1F31HD097933 to MG, 1F31HD105340 to KA, and 1R01HD091848 to RF. We thank the generous support of the Howard Hughes Medical Institute for the Gilliam Fellowship to MB. We also thank the US-Israel Binational Science Foundation (BSF) for their support to RF (2019285). We thank the Rhode Island Foundation and the Advance-CTR at Brown University supported by the NIH IDeA-CTR grant (U54GM115677) for support for these studies and its publication.

Acknowledgments

We thank the Center for Computation and Visualization at Brown University for computational resources for scRNA-seq, RNA-seq, and CUT&RUN data analysis. We thank Kevin Carlson and the Brown University Flow Cytometry and Sorting Facility for expertise completing the flow sorting. We also thank Joseph Aguilera and Abigail Brown for their assistance with ATAC-seq data analysis.

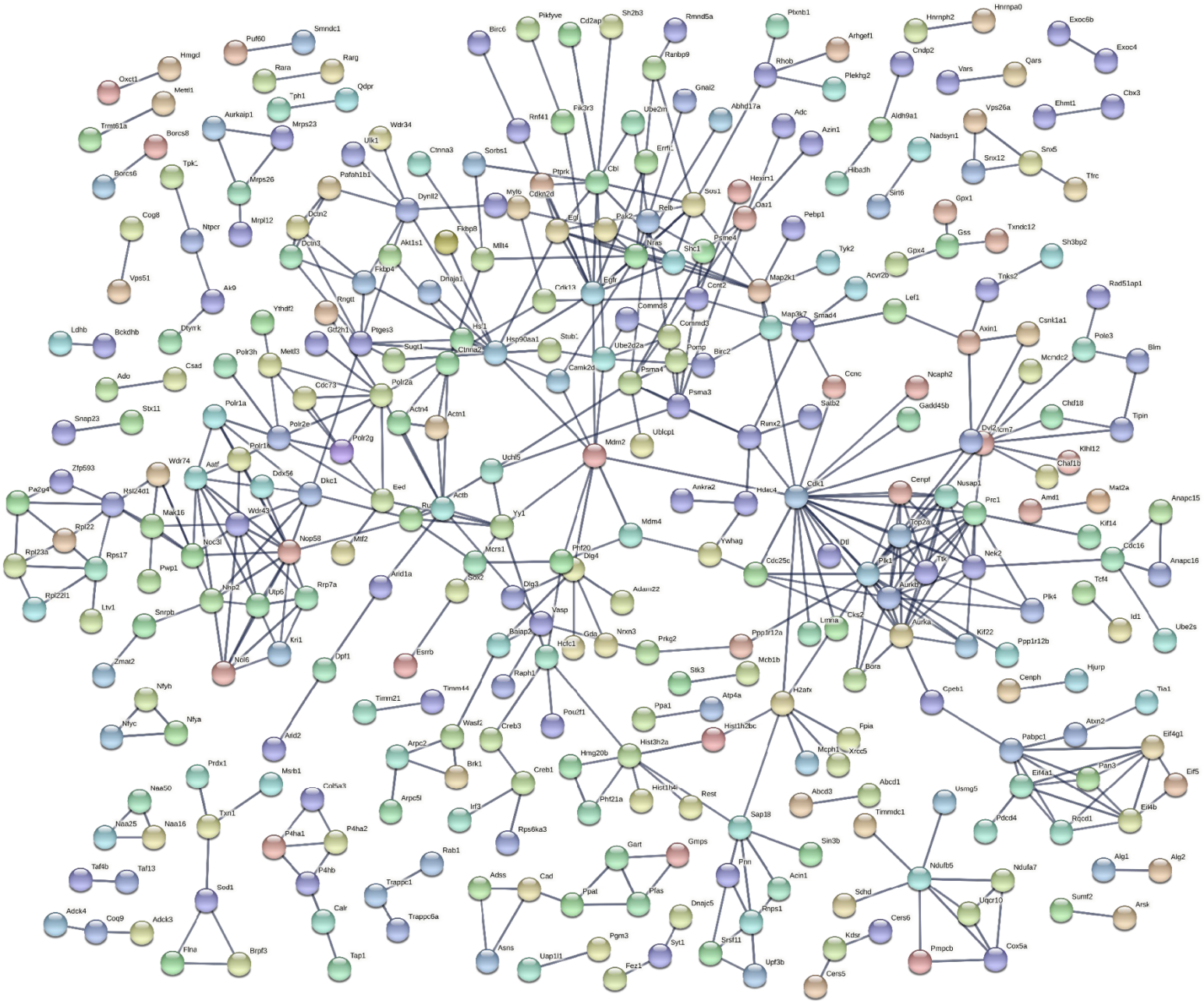
Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material

Experiment	Age	Genotype	Sample #	Cell #
RNA-seq	E16.5	Wildtype	1	13904
			2	13983
			3	34749
			4	23013
			5	10719
			6	11479
			7	27413
		Heterozygous	1	15818
			2	11364
			3	14041
			4	17917
			5	18862
			6	16384
			7	12084
			8	11075
			9	14498
		Deficient	1	16106
			2	17238
			3	7582
			4	19818
			5	22754
			6	10188
			7	8173
			8	12635
CUT&RUN	E16.5	Wildtype	Replicate 1-TAF4b	55829
			Replicate 1-H3K4me3	55829
			Replicate 1-IgG	55829
			Replicate 2-TAF4b	131000
			Replicate 2-H3K4me3	131000
			Replicate -IgG	131000

Supplementary Table S2.1 | Cell numbers for RNA-seq samples.



Supplementary Figure S2.1 | Protein-protein interactions of TAF4b bound DEGs in E16.5 ProSpg.

CHAPTER 3: REGULATION OF GENOMIC INTEGRITY AND NOVEL SPERMATOGONIAL LINEAGE DYNAMICS

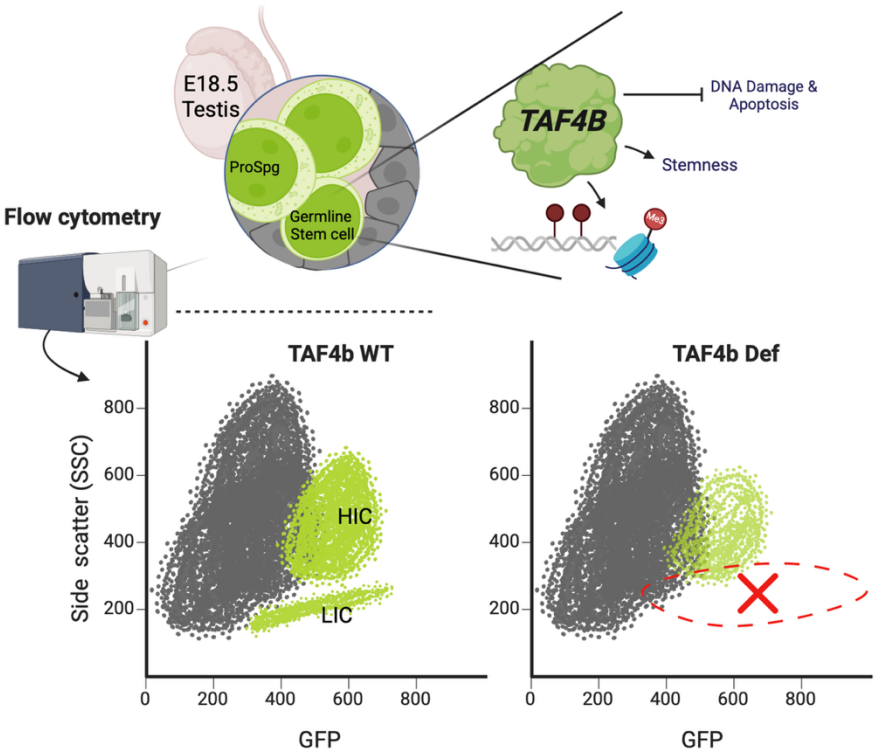
Myles A. Bartholomew[1], Maxfield MG. Kelsey[1], Kimberly M. Abt[1], Anna E. Nixon[1], Kimberly A. Seymour[2] and Richard N. Freiman[1,2*]

I wrote this chapter with editorial feedback from Dr. Richard N. Freiman

[1]MCB Graduate Program, Providence, RI, United States

[2]Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI, United States

Abstract



Reproductive health sits as the cornerstone and origin of every human life, yet infertility is on the rise. With male factor infertility becoming increasingly prevalent, it is critical that we gain an understanding of how the spermatogonial stem cell (SSC) pool is established, the regulatory dynamics governing its maintenance, and its defining developmental characteristics. Although the lineage dynamics preceding germ-line stem cell establishment remain obscure, this study identifies a novel cell-fate bifurcation and heterogeneity event at E18.5. Most intriguingly, the previously unobserved germ cell population is uniquely defined by intrinsic biophysical properties rather than by marker expression alone. Using an integrated single-cell RNA-sequencing atlas from E12.5 to P6, we first observed dynamic Taf4b expression during late embryogenesis, correlating with quiescence, epigenetic reprogramming, and changes in germ cell lineage. These biological processes have been well described as the building blocks of stemness and

fertility. From these observations, we isolated Taf4b-expressing prospermatogonia and identified a novel low intracellular complexity (LIC) cell population emerging at E18.5, alongside a high intracellular complexity (HIC) population. This bifurcated population was consistently observed across TAF4b-GFP, Oct4-GFP, and Id4-GFP reporter lines, indicating a discrete and conserved germ cell subset arising during late embryonic development. TAF4b deficiency resulted in specific ablation of the LIC population and a reduction in HIC cells, demonstrating that TAF4b regulates germ cell population dynamics during late embryogenesis. In our initial transplantation assays, E18.5 TAF4b-GFP–positive germ cells show promise in restoring spermatogenesis in germ cell–deficient hosts, indicating that a subset of TAF4b-expressing cells display germline stem cell activity as early as E18.5. At the molecular level, loss of TAF4b led to significant transcriptional disruption, with enrichment of pathways related to cell cycle checkpoints, double-strand break repair, and epigenetic regulation during HIC–LIC specification. Whole-genome bisulfite sequencing revealed altered DNA methylation dynamics, and Taf4b-deficient prospermatogonia exhibited increased LINE-1 expression, elevated ORF1p levels, and increased apoptosis. These findings indicate that genomic integrity is severely compromised in the absence of TAF4b. Collectively, our work identifies a novel prospermatogonial population, defines TAF4b as a molecular safeguard of SSC establishment, and shifts the focus of SSC establishment and infertility risk to late fetal development where failures in genomic integrity may irreversibly shape later reproductive capacity.

Introduction

Sexual reproduction is a fundamental biological process. Over the past century, researchers have sought to elucidate the molecular mechanisms underlying the development of sperm and eggs. Despite these efforts, infertility rates continue to rise. In men, both sperm concentration and total sperm count have declined by more than 50% over the past 40 years, with male factor infertility becoming increasingly prevalent (Levine et al., 2017). Understanding how the spermatogonial stem cell (SSC) pool is established, the regulatory dynamics governing its maintenance, and the defining characteristics of this cell type may be crucial for reversing these trends, developing new treatments, and restoring male fertility.

Current models indicate that the spermatogenic lineage originates from primordial germ cells (PGCs). In mice, PGCs migrate through the genital ridge and colonize the embryonic testes by embryonic day 9.5 (Nguyen et al., 2019). These cells subsequently undergo sex determination, differentiating into prospermatogonia (Capel, 2019). Prospermatogonia are dynamic, transitioning from a proliferative state to quiescence around embryonic day 16.5 (Du et al., 2021). During this developmental window, they experience extensive epigenetic reprogramming, including the reestablishment of histone and DNA methylation (Capel, 2019; Hill et al., 2018). These robust transcriptional and epigenetic changes are thought to facilitate the establishment of the spermatogonial stem cell (SSC) pool. However, the molecular regulators of the proliferative-to-quiescent transition, epigenetic remodeling, and SSC establishment remain largely unidentified.

A major obstacle to identifying these regulators has been the heterogeneity in the prospermatogonia believed to give rise to SSC and prolonged fertility. During the perinatal period, prospermatogonia differ in morphology, epigenetic state, and marker expression, highlighting their asynchronous developmental progression (Hermann et al., 2015; Kluin & de Rooij, 1981). This heterogeneity also complicates efforts to define SSC precursors; however, it raises the possibility that one or more prospermatogonial subsets are fated to give rise to the SSC pool.

To elucidate these dynamics and the regulation of early germ cell development, we used a mouse model lacking the transcription factor TATA-box binding protein associated factor 4b (TAF4b). Notably, human studies have already implicated TAF4b in male infertility. The first evidence of a strong translational application came from a Turkish population study linking TAF4b mutations to infertility in otherwise physiologically healthy men of reproductive age (Ayhan et al., 2014). A subsequent analysis of a human nonobstructive azoospermia cohort also identified men with single-nucleotide polymorphisms that cause nonsynonymous mutations in TAF4b (Xi et al., 2020). Consistent with these findings, our Taf4b-deficient mice are otherwise healthy but males exhibit severely compromised long-term reproductive capacity. While germ cells are present at birth and some males can initially sire litters, fertility is rapidly lost, indicating a defect in SSC establishment or maintenance (Falender et al., 2005; Lovasco et al., 2015).

Together, these studies and the evidence we present demonstrate that TAF4b is critical for the robust establishment of the SSC pool. Here, we identify a novel bifurcation and

potentially critical germ cell subpopulation as well as reveal how its establishment, along with overall DNA integrity and germ cell survival are Taf4b dependent. Our findings define TAF4b as a molecular safeguard of SSC establishment, provide mechanistic insight into the origins of male infertility, and point to new avenues for therapeutic intervention.

Results

Dynamic *Taf4b* expression in late embryonic mouse prospermatogonia

Recent evidence suggests that multiple male germ cell populations may become primed for specific germ cell fates, such as differentiation or stemness, as early as E18.5 (Law et al., 2019). To investigate this, we integrated five previously published single-cell RNA-sequencing datasets (Figure 3.1a). Collectively, these datasets span embryonic development from E12.5 through postnatal day 6 (P6) and comprise more than 17,000 cells (Figure 3.1b). Following quality control and initial alignment, a largely linear transcriptional trajectory from embryonic stages into the early postnatal period was observed (Figure 3.1c). We observed that while *Taf4b*-expressing cells were present throughout the time course, expression was markedly elevated from E16.5 to P2, peaking at birth (Figures 3.1d, e). Notably, *Taf4b* expression was inversely associated with proliferation-related genes, including *Mki67*, *Plk1*, and *Pcna* (Figure 3.1f). This corroborated our earlier evidence that *Taf4b* is required for proper initiation of quiescence and also related Taf4b's expression to other transcriptional regulators like *Cxxc1* and *Id4* implicated in epigenetic reprogramming and SSC development, respectively (Figures 3.1g,h). This integrated dataset was a launching point to further understand the heterogeneity of prospermatogonial development.

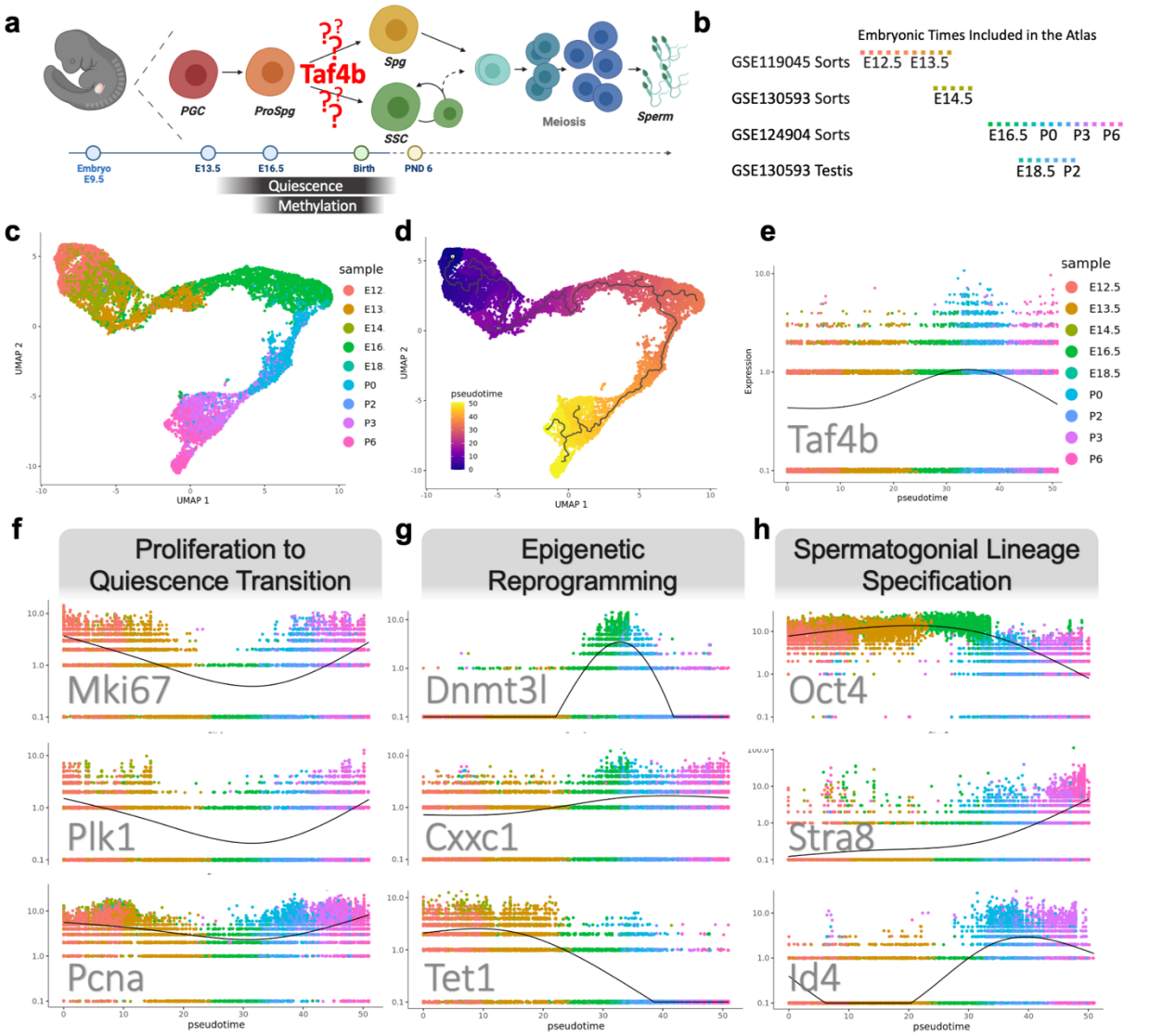


Fig. 3.1 | Single-cell integration defines TAF4b expression during embryonic male germ cell development.

a, Model of mouse male germline development and the potential timing and functions of TAF4b. **b,** Outline of data sets integrated for perinatal germline development and their corresponding timepoints. **c-d,** Uniform manifold approximation and projection (UMAP) of integrated single-cell transcriptomes spanning embryonic day (E)12.5 to postnatal day (P)6 and Monocle pseudotime trajectory showing continuous lineage progression. **e,** Dynamic expression of *Taf4b* over pseudotime **f-h,** Gene expression dynamics for markers associated with proliferation/quiescence, epigenetics, and prospermatgonial lineage specification.

A novel low intracellular complexity prospermatogonial population emerges at E18.5

To isolate and characterize Taf4b-expressing prospermatogonia, we generated a GFP–TAF4b knock-in mouse expressing an amino-terminal fusion from the endogenous promoter. (Figure 3.2a). We used confocal microscopy to detect the prospermatogonial-specific and robust expression of TAF4b in the E17.5 embryonic testis (Figure 3.2b). We then implemented fluorescence-activated cell sorting (FACS) for GFP expressing cells to identify and characterize GFP-TAF4b expressing cells in the embryonic testis. In addition, we prepared whole cell extracts from male germ cells sorted from TAF4b-GFP and Oct4-GFP embryonic testes and used immunoprecipitation followed by western blot to confirm the specific expression of a 132 kD GFP-TAF4b fusion protein in the *eGfp-Taf4b* mouse line. We then isolated cells based on GFP fluorescence by FACS at E16.5 and E18.5. At E18.5, there was a clear emergence of two distinct GFP-positive cell populations parsed by differences in their side scatter characteristics (SSC-A) (Figures 3.2d-f). As side scatter is commonly used as a proxy for intracellular complexity, we identified the emergence of a novel and distinct low intracellular complexity (LIC) compared to high intracellular complexity (HIC) population of E18.5 prospermatogonia (Shinohara et al., 2000).

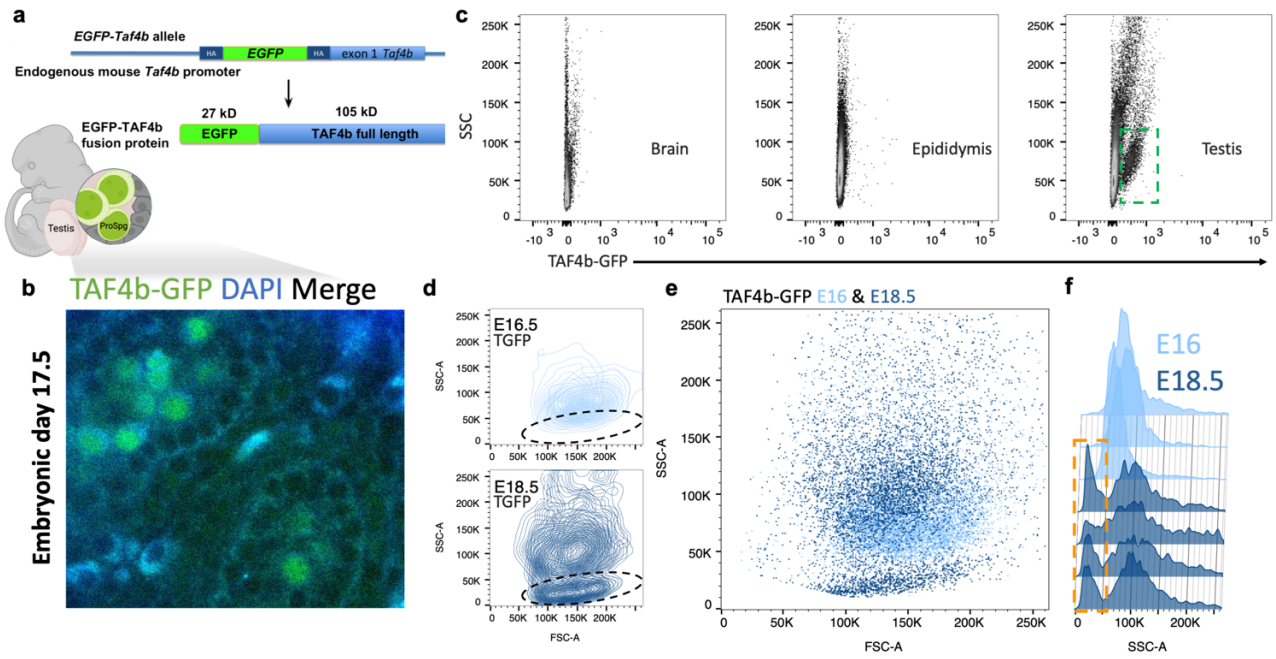


Fig. 3.2 | Characterization of TAF4b-expressing germ cells and lineage dynamics during late embryonic development

a, Schematic of GFP-TAF4b transgenic mouse model. **b,** Confocal microscopy of seminiferous tubules from E17.5 GFP-TAF4b testis whole mounts with endogenous GFP-TAF4b expression. **c,** Fluorescence-activated cell sorting (FACS) profiles of side scatter area (SSC-A) versus GFP fluorescence from male GFP-TAF4b tissues including the brain and epididymis as negative controls compared to a testis sample with a GFP positive cell population (green box). **d,** FACS sorts of GFP-TAF4b+ cells at E16.5 (n = 3) and E18.5 (n = 4). All GFP_TAF4b positive (TGFP+) are shown as contour plots depicting biophysical parameters of SSC-A vs forward scatter area (FSC-A) contour lines set at 5% threshold. Dashed line denotes a secondary cell population that is only observable at E18.5 **e-f** Overlapped FACS profiles of SSC-A vs FSC-A followed by overlapping histograms of SSC-A with novel E18.5 subpopulation (orange box).

Oct4-GFP-expressing mice display two distinct E18.5 prospermatogonial populations

To determine whether this LIC population was specific to TAF4b-GFP-expressing prospermatogonia, we compared FACS profiles between embryonic prospermatogonia from *Oct4*-GFP reporter mice (Gura et al., 2023). We observed that *Oct4*-GFP positive prospermatogonia at E15.5 and E16.5 exhibited a unimodal distribution, whereas E18.5 *Oct4*-GFP cells displayed a bimodal distribution that matched the distribution observed in

E18.5 TAF4b-GFP sorts (Figure 3.3a). We assessed whether these changes were specific to the germ cell population by comparing somatic whole testis intracellular complexity across embryonic time. This showed no corresponding differences. In addition to the consistent bifurcated population at E18.5 germ cells, we note a broader trend of increasing intracellular complexity over the course of late embryonic development (Figures 3.3b, c). Lastly, we found that the distinct population of low complexity GFP-expressing prospermatogonia are also detected in *Id4-GFP* reporter mice (Supplementary Figure 3. S2). Together these data indicate that E18.5 LIC prospermatogonia are a novel feature of male embryonic germ cells isolated by multiple reporter lines.

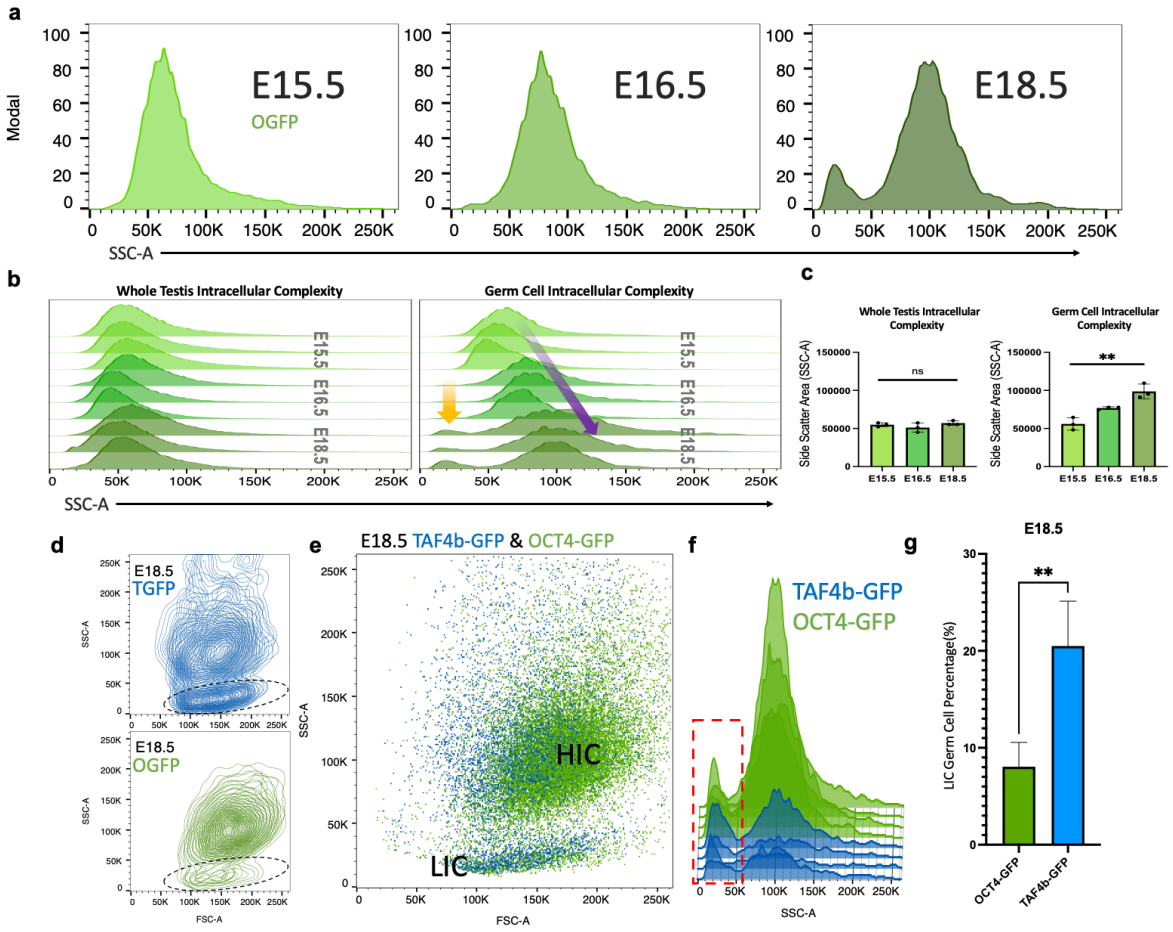


Fig. 3.3 | Identification of LIC population in male embryonic germ cells from *Oct4-GFP* mice

a, FACS histograms of side scatter area (SSC-A) from *Oct4-GFP* (OGFP) male testes at E15.5, E16.5, and E18.5. **b,** Histograms showing intracellular complexity of whole testes across E15.5, E16.5, and E18.5 in triplicate, as well as germ cell–specific intracellular complexity, showing a progressive increase over developmental time (purple arrow) and the emergence of a consistent secondary population at E18.5 in OGFP (yellow arrow). **c,** Corresponding quantification of whole-testis intracellular complexity and overall germ cell intracellular complexity over developmental time shown as mean ± SD (n = 3 mice per group). ** $P < 0.005$, ordinary one-way ANOVA. **d,** Biophysical parameters from FACS as contour plots of SSC-A vs FSC-A using a 5% contour threshold depicting germ cell populations from E18.5 TAF4b–GFP and OCT4–GFP sorts with a dotted outline denoting the novel LIC population. **e-f,** Overlapping FACS profiles and corresponding SSC-A histograms. (TGFP, blue, n = 4) and (OGFP, green, n = 4) **e,** Quantification of the LIC population in each FACS sort, shown as mean ± s.e.m. (n = 4 mice per group). $P < 0.05$, *t*-test.

Dependence of E18.5 LIC prospematogonia on TAF4b

Given that GFP-TAF4b sorting reveals the LIC prospermatogonia at E18.5, we next examined its potential dependence on TAF4b. We crossed our TAF4b-deficient mouse line with the *Oct4-GFP* reporter mouse line, both on a C57Bl/6 genetic background, to label littermate-matched prospermatogonia of all three possible *TAF4b* genotypes: wild-type (WT), heterozygous (Het), or deficient (Def). In biological quadruplicate, we observe that while both *Taf4b* WT and Hets exhibit two distinct male germ cell populations, the LIC prespermatogonia are largely absent in TAF4b-Def mice (Figures 3.4a-c). Further quantification confirmed this striking effect on the LIC population, but also revealed a smaller but significant reduction in the number of germ cells within the larger HIC population (Figures 3.4d, e). Together, these data suggest that the LIC prespermatogonia are more sensitive to the loss of TAF4b than the HIC population, yet TAF4b contributes to the population dynamics of both.

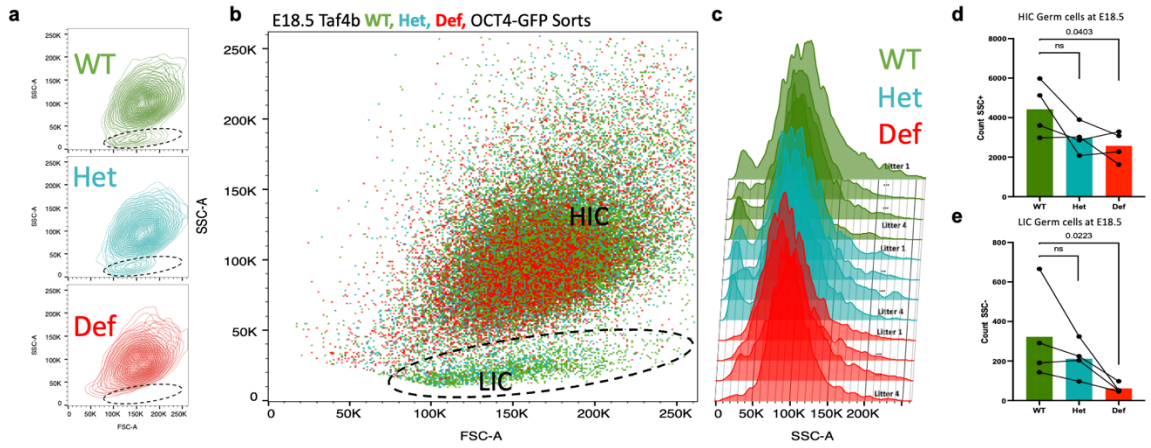


Fig. 3.4 | Loss of TAF4b greatly diminishes the LIC prospermatogonial population at E18.5

a, Contour plots of SSC-A versus FSC-A using a 5% contour threshold, depicting prospermatogonia from E18.5 *Oct4-GFP* sorts across three genotypes: wild-type TAF4b (WT), heterozygous (Het), and Taf4b-deficient (Def). A dotted outline denotes the LIC population in WT and Het samples and its absence in Def samples. **b,** Overlapping FACS profiles illustrating the loss of the LIC population in Taf4b-deficient samples. WT (green), Het (teal), and Def (red) **c,** Corresponding SSC-A histograms for litter matched TAF4b WT, Het, and Def samples. **d,** Quantification of the larger HIC population in each FACS sort, shown as mean (n = 4 mice per group); paired *t*-test. **e,** Quantification of the LIC population in each FACS sort, shown as mean (n = 4 mice per group); paired *t*-test.

TAF4b-GFP-expressing late embryonic prospermatogonia can support spermatogenesis

Recent work has shown that E18.5 *Id4-GFP* prospermatogonia contain SSC activity (Law et al., 2019) and TAF4b has been implicated in SSC expression and development, both in mouse and human contexts (Bahat et al., 2013; Huang et al., 2024b; Lovasco et al., 2015; Shinohara et al., 2000). Thus, we sought to determine whether TAF4b-expressing late embryonic male germ cells have the capacity to rescue spermatogenesis in W/Wv spermatogenesis-impaired hosts (Garbuzov et al., 2018; Law et al., 2019; Valli et al., 2014). Given the small numbers of TGFP cells in each embryo, we pooled E18.5 TGFP-expressing germ cells from multiple embryos and litters. Compared to TGFP-negative cells, TGFP-positive cells were able to support spermatogenic colony formation in W/W^{v}

hosts (Figures 3.5b,c). We note that we have only completed one biological replicate for this experiment, and several more replicates are currently underway.

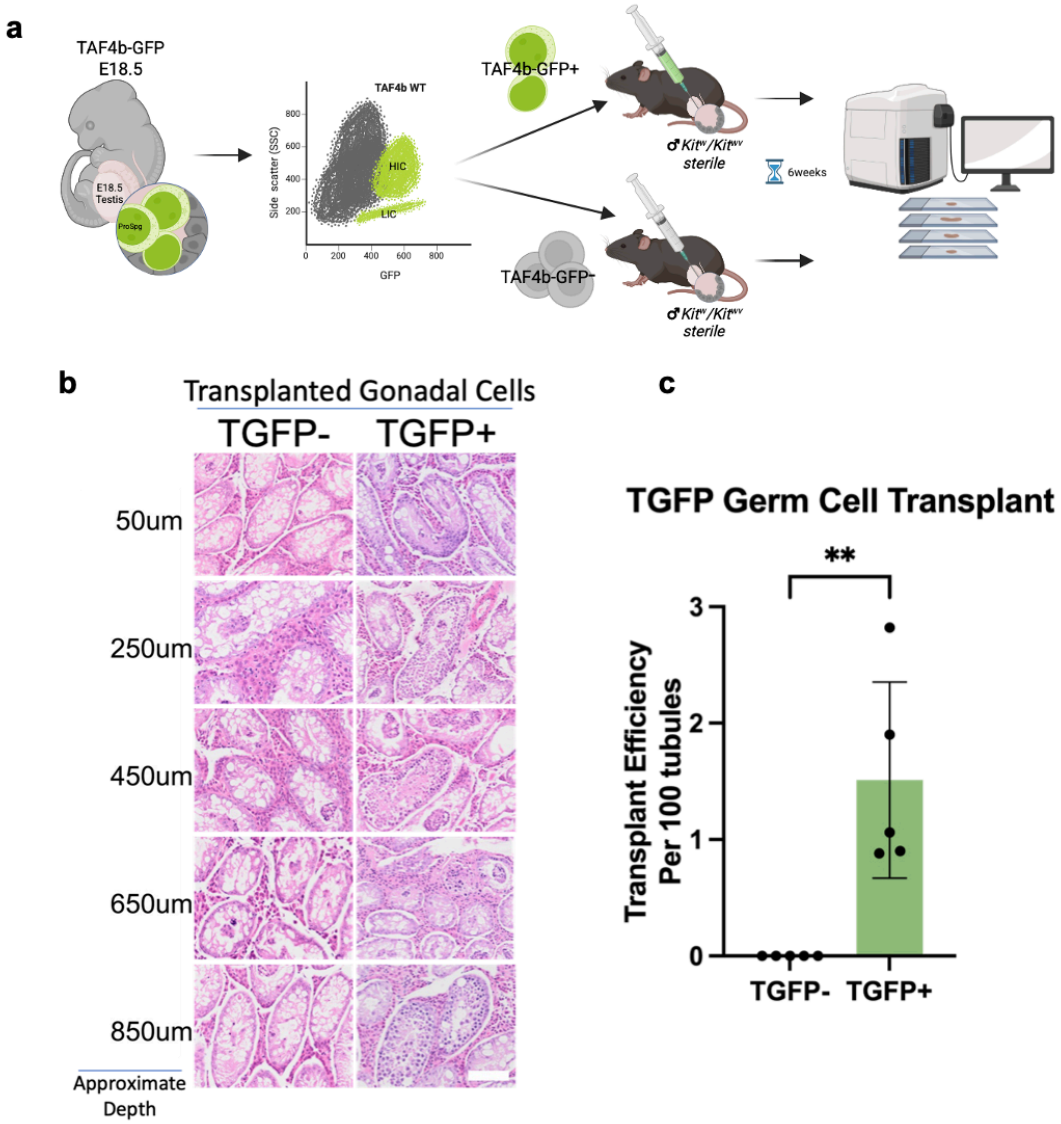


Fig. 3.5 | Functional rescue of spermatogenesis by TAF4b-expressing embryonic germ cells

a, Experimental strategy illustrating collections and FACS of E18.5 TAF4b–GFP–positive and -negative embryonic germ cells followed by transplantation into spermatogenesis-impaired W/Wᵛ recipient testes. **b,** Representative brightfield hematoxylin & eosin (H&E) staining of recipient testes showing donor-derived spermatogenesis following transplantation of TAF4b–GFP–positive cells spanning ~800um. **c,** Quantification of the number of seminiferous tubules exhibiting rescued spermatogenesis in recipient testes per 100 tubules counted. Data are shown as mean ± s.e.m.; *n* = 5 section. Significance was assessed using t-test.

TAF4b is essential for orchestrating late embryonic male germ cell transcription programs

Next, we aimed to understand potential transcriptional disturbances caused by TAF4b loss and how they might underlie defects in prospermatogonial development. Thus, we extended our previous Taf4b WT and Def gene expression profiling from E16.5 (Gura et al., 2023) to E18.5. This allowed us to observe which biological processes are disrupted between WT and Def during the period of newly characterized germ cell dynamics. When examining WT and Def across developmental timepoints by principal component analysis (PCA), it was observed that changes in transcriptional profile are more pronounced at E16.5 between WT and Def, and the transcriptional disruption appears less pronounced by E18.5 (Figure 3.6a). We then analyzed differential gene expression between Def and WT across time points, noting 1,098 genes upregulated in Def samples and 837 genes upregulated in WT samples. *Taf4b* was the most upregulated gene in WT, thereby substantiating DEG detection. *Caprin2*, a regulator of cell growth and WNT signaling, was the top increased DEG in Def (Figure 3.6b) (Acebron et al., 2014; Wang et al., 2016; Yang et al., 2016). GSEA analyses revealed significant enrichment for pathways related to genetic integrity and its readouts, including cell cycle checkpoints, double-strand break repair, and epigenetic modification. It also revealed interesting trends in pathways for cell elimination, with intrinsic apoptotic signaling being enriched in Def from E16.5 to E18.5, We found that, in contrast, negative regulation of immune response and cell-killing pathways were upregulated in WT. Together, these pathways possibly indicate increased apoptosis in Def, while WT is running protective programs and warding off immune responses. Lastly, examining DEGs across LIC–HIC specification and by genotype

revealed coordinated dysregulation of multiple pathway components between WT and Def cells. Notably, several of the DEGs in these dysregulated biological processes, including *H2ax, H2aw, Setd5, and Ctcf,* exhibit direct TAF4b transcription start site occupancy (Supplementary Figures 3.4) (Gura et al., 2023).

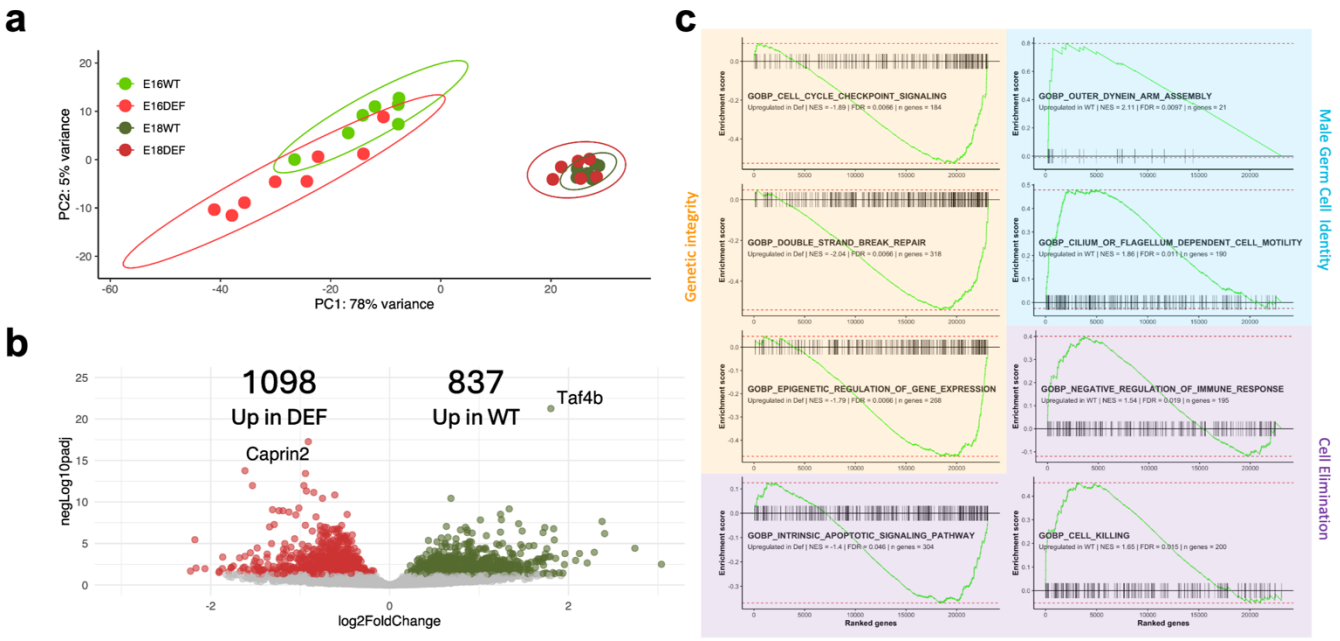


Fig. 3.6 |Cell cycle, chromatin, and genome integrity pathways are misexpressed in the absence of TAF4b
a, PCA plot comparing the bulk RNAseq profiles of Taf4b WT and Def across E16.5 and E18.5. **b,** Volcano plot of top 2 differentially expressed genes along with the total DEGs in either direction from bulk-RNAseq comparing *Taf4b*$^{+}$/$^{+}$; WT -and *TAF4b*$^{-}$/$^{-}$; Def, *Oct4-GFP* sorted germ cells. **c,** Results from gene set enrichment analysis (GSEA) showing enrichment of processes related to genomic integrity, cell elimination and male gamete generation.

Aberrant DNA methylation, retrotransposon activity and apoptosis in the absence of TAF4b

Epigenetic regulation is a well-established feature of prospermatogonial quiescence, and disruption of these processes has been linked to infertility phenotypes similar to those observed in Taf4b-Def mutants (Dura et al., 2022; Webster et al., 2005). Given the germ

cell loss, together with the epigenetic and apoptotic signatures identified above, we asked whether DNA methylation and genomic integrity were functionally altered in the absence of TAF4b. We therefore hypothesized that perturbations in DNA methylation may compromise genomic integrity and contribute to germ cell death. To address this, we first performed whole-genome bisulfite sequencing (WGBS) using fertile *Taf4b*-Het and infertile *Taf4b*-Def *Oct4-GFP*-positive sorted prospermatogonia at E16.5 and E18.5. *Taf4b*-Het embryos display normal establishment of the LIC germ cell population (Figures 3.4d, e). Consistent with canonical DNA methylation dynamics during late embryonic development, we observed widespread increases in CpG methylation from E16.5 to E18.5 in *Taf4b*-Het samples, as well as a similar increase in *Taf4b*-Def samples (Figure 3.7a). However, we did note that E16.5 Hets displayed more CpG erasure, as indicated by yellow, with approximately 34% of CpGs lacking methylation. This is compared to E16.5 Def, in which only approximately 24% of CpGs were fully erased (Figure 3.7a). These observations indicate that aside from subtle disruptions, global DNA methylation dynamics appear to proceed in the absence of TAF4b.

As DNA methylation is also tightly linked to the repression of transposable elements we examined CpG methylation at LINE-1 (L1) elements across the genome (Nagamori et al., 2018; Ponomaryova et al., 2020). This analysis revealed that methylation at L1 loci largely follows the trend of increasing DNA methylation for CpGs from E16.5 to E18.5 in both *Taf4b*-Het and -Def samples (Figure 3.7b). There was, however, a modest reduction in L1 methylation in *Taf4b*-Def samples by E18.5 that did not reach statistical significance. To determine whether this hypomethylation was accompanied by increased

retrotransposon expression, we analyzed bulk RNA-sequencing data for expression of full-length L1 elements using our TE-Seq pipeline (Kelsey et al., 2025). The changes in L1 mRNA closely correspond with the methylation dynamics detected by WGBS at the respective time points. WT showed a significant decrease in L1 RNA expression from E16.5 to E18.5, whereas Def showed the opposite effect, with L1 RNA increasing significantly between E16.5 and E18.5 (Figure 3.7c). Consistent with transcriptional and methylation data, *Taf4b*-deficient germ cells exhibited increased ORF1p expression at E18.5, confirming deregulated L1 elements in the absence of TAF4b (Figure 3.7d). We then compared TUNEL staining at E16.5 and E18.5 in *Taf4b*-WT and -Def testis sections and revealed a significant increase in germ cell apoptosis in *Taf4b*-deficient testes at E16.5 (Figure 3.7e). Together, these findings indicate that proper transcription and DNA methylation are compromised at E16.5 in *Taf4b*-deficient prospermatogonia leading to their excessive loss.

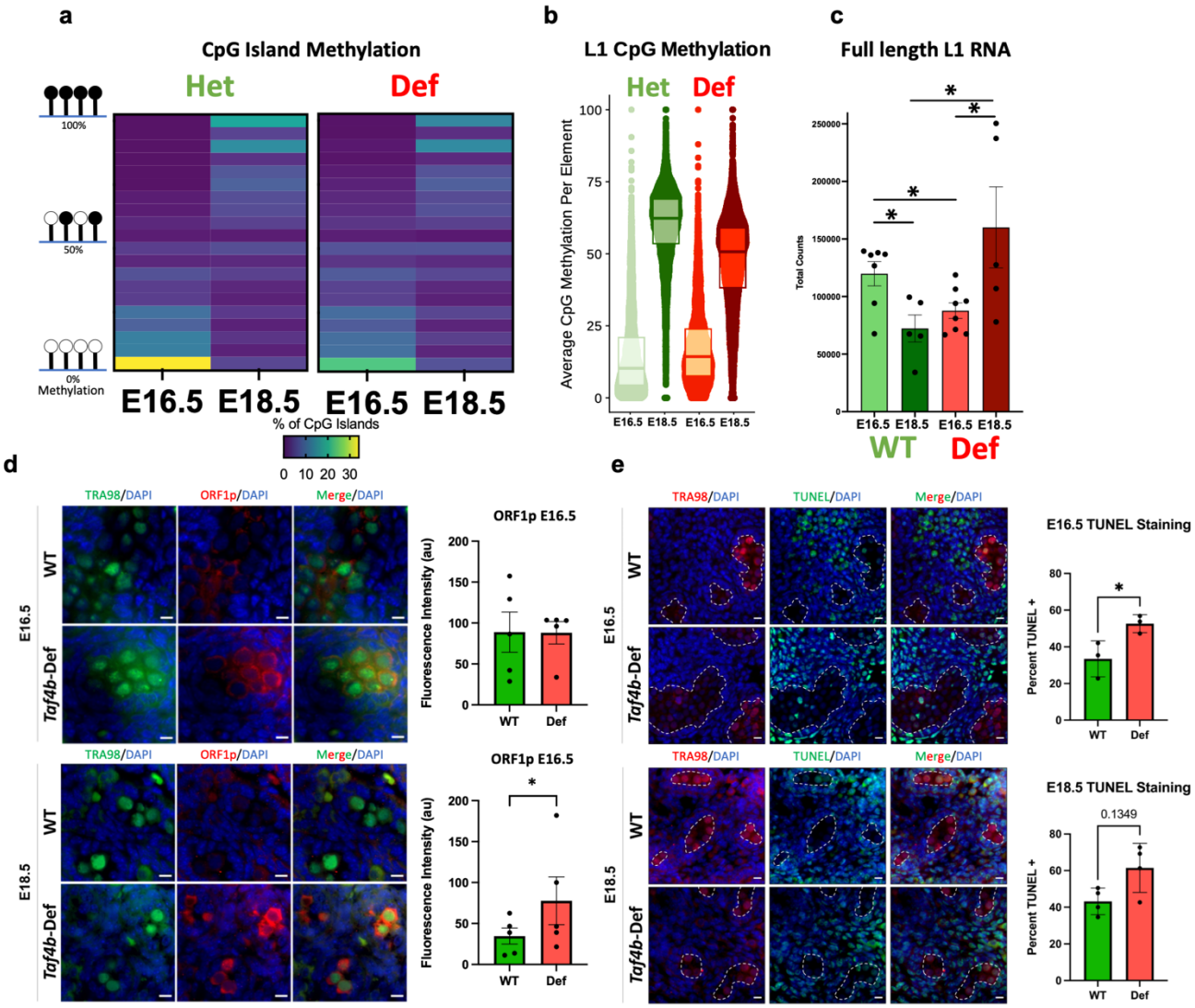


Fig. 3.7 | Transcriptional and epigenetic consequences of TAF4b deficiency during late embryonic germ cell development

a, Whole-genome bisulfite sequencing (WGBS) comparison of CpG island methylation levels between E16.5 and E18.5 *Taf4b* heterozygous and *Taf4b* Def prospermatogonia isolated by FACS. Each row represents a methylation bin (0–5%, 5–10%, …, 95–100%), with color intensity indicating the proportion of CpG islands falling within each bin **b,** Violin plot of percent CpG methylation for L1 elements annotated across the genome, as determined by WGBS. Horizontal lines indicate median and upper and lower quartiles. **c,** Quantification of full-length L1 elements from RNA-sequencing data, mean ± s.e.m.(n = 5–8 mice per group) ordinary one-way ANOVA followed by t-test. * $P < 0.05$ **d,** Immunofluorescence of E16.5 and E18.5 testis sections from WT and Taf4b-Def embryos with antibodies against germ cell marker, TRA98, LINE-1 ORF1p, and DAPI for nuclei; followed by quantification of L1 ORF1p fluorescence intensity in TRA98$^+$ germ cells from WT and Taf4b-Def shown as mean ± s.e.m. (n = 3–5 mice per group), * $P < 0.05$, paired

t-test. **f,** TUNEL staining of E16.5 and E18.5 testis sections comparing WT and Taf4b-Def samples. Germ cells were identified by TRA98 staining, apoptotic cells by TUNEL, and nuclei with DAPI; along with quantification of the percent $TUNEL^+$ germ cells shown as mean ± s.e.m.(n = 3 mice per group). *P < 0.05, t-test.

Discussion

The lineage dynamics preceding germline stem cell establishment have long remained poorly understood. In this study, we identify a previously unrecognized bifurcation and heterogeneity event at E18.5 defined by intrinsic biophysical properties. Our results demonstrate that TAF4b functions as a molecular regulator for a newly defined subset of prospermatogonia. While TAF4b supports the overall abundance of both LIC and HIC populations, LIC prospermatogonia are disproportionately depleted in its absence. This depletion correlates with significant transcriptional alterations and compromised genomic integrity, which may underlie the loss of prospermatogonia observed at E18.5. In parallel, transplantation assays were performed to assess whether cells expressing TAF4b possess functional stem cell capacity, further linking this transcriptional regulator to the establishment of the spermatogonial lineage. Together, our studies reveal the emergence of an intrinsically low-complexity prospermatogonial population that may represent a previously unrecognized stage of reproductive development and SSC establishment

Using an integrated and comprehensive single cell-seq atlas, we first identified TAF4b expression related to multiple aspects of perinatal germ cell development. Analysis revealed dynamic expression of *Taf4b* at E16.5 and peaking around E18.5. *Taf4b* mRNA expression correlated with both well-defined and lesser-known epigenetic regulators during late embryogenesis, suggesting a more specific role during quiescence. This

critical developmental period has long been characterized by significant epigenetic reprogramming (Nagamori et al., 2018). In addition, the initial establishment of prospermatogonia with stem cell properties occurs during this dynamic developmental window (Law et al., 2019).

The presence of two discrete subpopulations of prospermatogonia arising in the embryonic testis prior to birth is intriguing. Similar to the cytometric heterogeneity observed in hematopoietic cells, further insights may be gained by comparing LIC cells with other cell types of low intracellular complexity (Basford et al., 2010). The presence of this novel population across different germ cell isolation strategies may suggest their necessity for the long-term maintenance of adult spermatogenesis. Specific ablation of the LIC fraction in *Taf4b*-deficient embryonic testes further supports Taf4b's role in regulating germ cell lineage specification at late embryonic stages and underscores its necessity for coordinated germ cell dynamics. Concurrently, the reduction in HIC cells indicates that TAF4b also regulates the abundance of the major fraction of prospermatogonia. Collectively, these findings support a model in which TAF4b is required to establish appropriate germ cell populations and to drive their progression through distinct germline stages. We note that the observation of LIC cells could benefit from further transcriptional characterization. Thus far, we have attempted to retrospectively assess them using our integrated single-cell atlas from Figure 3.1b. The difficulty in identifying this population in existing datasets is that this study, to our knowledge, represents the first reported isolation of LIC cells. Many currently available single-cell datasets rely on FACS-enriched germ cell populations, which may have

inadvertently excluded this cell subset. In addition, the unsorted E18.5 testis dataset generated by Tan et al. contains only a few hundred germ cells among the total single-cell population. As a result, even if LIC cells were present, they would likely be represented by only a small number of cells and would remain difficult to distinguish without a definitive marker for their identification within the single-cell landscape.

However, because this novel population is defined by low intracellular complexity, we have begun exploring features that may help denote these cells based on factors known to influence intracellular complexity. One such factor is mitochondrial content. Remarkably, we observe significantly lower mitochondrial content in LIC cells isolated by FACS (Supplementary Figures 3.7b,c). With this in mind, we also focused our scRNA seq on the time points at which we identified LIC cells. Intriguingly, we noted a subset of cells with markedly reduced mitochondrial RNA via single-cell analysis (Supplementary Figures 3.7d). Although this observation remains preliminary, it may provide a useful avenue for further defining LIC cells and potentially characterizing their transcriptional lineage and expression programs.

Transplantation assays are considered the most comprehensive means of assessing regenerative potential (Kubota & Brinster, 2018). In our study, we found that E18.5 TGFP-positive germ cells restore spermatogenesis in germ cell-deficient hosts. These results indicate that, as early as E18.5, a subset of TAF4b-expressing cells displays germline stem cell activity. Future studies are needed to determine whether only one fraction, or both HIC and LIC fractions, comprise the progenitors of future stem cells. At the molecular

level, the drastic transcriptional disruption observed in Def during HIC/LIC specification events highlights a breadth of pathways and checkpoints affected by the loss of a single factor. Premature or persistent expression of DNA damage-associated genes indicates that cells expected to transition are delayed and transcriptionally compromised. The enrichment of pathways related to cell cycle regulation, DNA integrity, and methylation is particularly informative, as it enables the identification of checkpoints and factors previously associated with fertility defects, such as CXXC1, PIWIL4, and TET1 (Huang et al., 2020; Jiang et al., 2020; Muñoz et al., 2014; Sha et al., 2021).

In the context of existing models, this study further confirms the notion of germ cell heterogeneity arising through coordinated events just prior to birth in mice (Hargy & Sasaki, 2023; Law et al., 2019). Furthermore, it shifts the focus of SSC establishment and infertility risk to late fetal development, where early failures in propspermatogonial sub-population specification and genomic integrity may irreversibly shape much later reproductive capacity. Along with resolving HIC or LIC stemness, future work will determine which biophysical properties, such as nuclear-to-cytoplasmic ratio, mitochondrial content, or organelle assembly, contribute to the distinct characteristics of these populations and their ability to maintain stemness. By further parsing out and identifying early SSC precursors, the regulatory mechanisms that give rise to proper SSC development could lead to advances in male reproductive health.

Materials and methods

Mice

Mice used for TAF4b-GFP (TGFP) fluorescence-activated cell sorting (FACS) were homozygous for a *Taf4b*-eGFP transgene. Mice used for bulk *Taf4b* WT vs. Def RNA sequencing and FACS experiments were homozygous for an *Oct4*-eGFP transgene (The Jackson Laboratory: B6; 129S4-*Pou5f1*^{tm2Jae}/J) and heterozygous for a *Taf4b*-deficiency mutation located in exon 12 of the 15 total exons of the *Taf4b* gene, which disrupts endogenous *Taf4b* expression. Mice used for *ID4*-GFP (IGFP) FACS were homozygous for Id4-eGFP(Law et al., 2019) (gift J. Oatley, Washington State University). W/W^v (c-Kit mutant) mice were used as germ cell-deficient recipients for transplantation experiments (Garbuzov et al., 2018; Ogawa et al., 1997).

Timed matings were established, with embryonic day (E)0.5 defined by the presence of a copulatory plug. Embryonic sex was determined based on gonadal morphology and the presence or absence of testicular cords. Genomic DNA was isolated from tail biopsies using QIAGEN DNeasy Blood & Tissue Kits (Cat. #69506) and used for PCR genotyping assays.

All animal procedures were reviewed and approved by the Brown University IACUC and performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

Immunofluorescence

Testes were collected from embryonic mice (E16.5-E18.5), briefly fixed in 10% formalin, and paraffin-embedded. Tissue processing and sectioning were performed by the

Molecular Pathology Core at Brown University. Following sectioning, antigen retrieval was carried out using 0.1% antigen unmasking solution (Vector Laboratories, H-3300) with 0.5% Tween-20 in a steamer for 20 minutes. Sections were permeabilized in 0.1% sodium citrate with Triton X-100 for 5 minutes, washed in 0.1% PBST (Triton X-100), and incubated at room temperature in blocking buffer (PBS containing 5% goat serum, and 0.1% Triton X-100) for 1 hour. Primary antibodies were applied in blocking buffer and incubated for 24 hours at 4°C. Sections were then washed in PBST, incubated with secondary antibodies for 1 hour at 37°C, washed again in PBST, and mounted using Vectashield Mounting Medium (Vector Laboratories, H-1900).

Primary antibodies included rat anti-TRA98 (1:100; Abcam, ab82527) and rabbit anti-LINE-1 ORF1p (EPR21844-108; 1:100; Abcam, ab216324). Apoptosis was detected using the In situ cell death detection kit (Roche, 12156792910) Secondary antibodies used were Alexa Fluor 488-conjugated anti-rat IgG (1:500, abcam, ab150165) and Alexa Fluor 555-conjugated anti-rabbit IgG (1:500, abmcam, ab150086). In addition, DAPI was added during the secondary incubation.

Images were acquired under identical exposure conditions and processed uniformly. Imaging and fluorescence intensity measurements were performed using a Zeiss Axio Imager M1 microscope and Zeiss software. Statistical significance was assessed using GraphPad Prism 9 (v9.3.1) for macOS, with n = 3-5 mice per condition.

Validation of the TAF4b-GFP fusion protein

To validate the TAF4b-GFP fusion protein, testis lysates were prepared from TGFP mice and subjected to immunoprecipitation followed by Western blotting. Antibodies against GFP (rabbit, abcam, ab290) and TAF4b affinity purified chicken were used to confirm the presence of a GFP-tagged TAF4b species, identified by an expected molecular weight shift of approximately 27 kDa relative to endogenous TAF4b. Reciprocal immunoprecipitation with GFP antibodies bound to equal parts protein G Sepharose (cytiva, 17061801) and protein A Sepharose beads (cytiva ,17528001) followed by TAF4b immunoblotting further confirmed successful fusion and enrichment. Endogenous TAF4b-GFP expression was visualized using confocal microscopy. Dissected testes were whole-mount, live-imaged to confirm robust and cell-specific expression.

Embryonic gonad dissociation and fluorescence-activated cell sorting

Embryonic gonads were harvested and dissociated into single-cell suspensions by incubation in 1 mL of 0.25% Trypsin/EDTA at 37°C for 40-50 minutes, as previously described(Gura et al., 2020) . During incubation, tissues were mechanically dissociated by pipetting with a P200 every 20 minutes and once at the end of the incubation period. Trypsin was neutralized with 100 µL fetal bovine serum (FBS). Cells were pelleted at 1,500 rpm for 5 minutes, resuspended in 100 µL PBS, and passed through a 35-µm mesh cap into FACS tubes (Gibco, #352235). Propidium iodide (1:500) was added as a live/dead discrimination stain.

Germ cells were isolated using GFP-based reporter strategies (TAF4b-GFP, OCT4-GFP, or ID4-GFP). FACS was performed using a BD FACSAria III or Cytek Aurora CS system at the Brown University Flow Cytometry and Cell Sorting Core Facility. Dead cells were

excluded, and remaining cells were sorted into GFP-positive and GFP-negative populations in PBS at 4°C for each embryo. Negative control tissues, including brain and epididymis, were used to establish GFP gating thresholds and confirm testis-specific expression.

Cells with low side scatter were defined as low intracellular complexity (LIC) germ cells, whereas cells with higher side scatter values (SSC-A > 50,000) were defined as high intracellular complexity (HIC) germ cells. Subsequent analyses was performed with FlowJo (v10.10). For RNA-seq analysis, GFP-positive cells from individual embryos were collected separately, pelleted at 1,500 rpm for 5 minutes, PBS was removed, and samples were stored at −80°C until shipment.

Germ cell transplantation assays

To assess stem cell potential, germ cell transplantation assays were performed using E18.5 TAF4b-GFP-positive germ cells. Cells were collected from multiple independent embryo pools comprising more than 40 embryos in total. All GFP-positive germ cells were isolated by FACS and transplanted into the seminiferous tubules of infertile W/W^v recipient mice. Recipient testes were analyzed after a 2-month recovery period. Spermatogenic rescue was assessed by histological examination and quantified as the proportion of seminiferous tubules displaying active spermatogenesis.

Single-cell RNA-seq data analysis

Datasets SRP194420, SRP258129, SRP158811, and SRP178196 were downloaded from the NCBI Sequence Read Archive (SRA) to Brown University's high-performance

computing cluster at the Center for Computation and Visualization. Datasets from Law et al. and Nguyen et al, and Du et al. which enriched for germ cells by FACS, were used and combined with whole-testis datasets from Tan et al., establishing a developmental window spanning E12.5 to P7 (Du et al., 2021; Law et al., 2019; Nguyen et al., 2020; Tan et al., 2020). FASTQ files were aligned using Cell Ranger (v6.0.0) count and aggregated using Cell Ranger aggr. The resulting cloupe files were analyzed using Loupe Cell Browser (v5.0). Germ cells were identified based on log-normalized expression of the germ cell marker *Dazl*. Somatic cell clusters were removed based on mean *Dazl* expression below 0.85. Additional quality control filtering was applied based on 12.5-15.8 $\log_2$-normalized unique molecular identifiers, 10.6-12.8 $\log_2$-normalized features, and less than 20% mitochondrial transcript content.

RNA sequencing

Embryonic germ cells resuspended in PBS were shipped on dry ice to GENEWIZ (GENEWIZ Inc., NJ). RNA extraction, quality control, library preparation, sequencing, and initial bioinformatic processing were performed by GENEWIZ. Glycogen (1 µL, 10 mg/mL) was added to increase RNA recovery. RNA concentration was measured using a Qubit 2.0 Fluorometer, and RNA integrity was assessed using TapeStation. Full-length cDNA synthesis and amplification were performed using the SMART-Seq v4 Ultra Low Input Kit (Clontech), followed by library preparation using the Illumina Nextera XT Library Preparation Kit. Libraries were multiplexed and sequenced on an Illumina HiSeq 4000 using a 2 × 150 bp paired-end configuration. Base calling and demultiplexing were performed using HiSeq Control Software and bcl2fastq (v2.17), allowing one index mismatch.

RNA-seq data analysis

Raw FASTQ files were processed on Brown University's high-performance computing cluster. Reads were trimmed using Trim Galore! (v0.5.0; parameters: -nextera -q 10) and assessed with FastQC (v0.11.5). Reads were aligned to the Ensembl GRCm38 reference genome using HISAT2 (v2.1.0)(Andrews, 2010; Pertea et al., 2016). Gene-level counts were generated using featureCounts (Subread v1.6.2)(Liao et al., 2013). Metadata files were manually generated and used for downstream analysis in DESeq2 (v1.22.2). Variance-stabilized transformed counts were used for PCA visualization with ggplot2 (v3.1.0)(Love et al., 2014; Wickham, 2016). Differential gene expression analysis was performed using DESeq2. Volcano plots were generated using ggplot2. Gene set enrichment analysis (GSEA) and Gene Ontology enrichment were performed using ClusterProfiler (v3.16.1), and heatmaps were generated from differentially expressed gene sets.

Bisulfite Sequencing analysis

Whole genome bisulfite (WGBS) libraries were prepared by Azenta using the NEB EM-Seq protocol. Sequencing reads were trimmed with Trim Galore, deduplicated, and then aligned to the mm10 genome with Bowtie2. A methylation bedgraph was produced using Bismark(Krueger & Andrews, 2011). We examined the methylation levels of full length, young LINE-1 elements in the L1MdA, L1MdGf, and L1MdTf subfamilies. CpG methylation was modeled across genotype at each developmental time point using a binomial mixed-effects model with the glmmTMB package(Brooks et al., 2017).

LINE-1 RNA sequencing analysis

We analyzed these data using our TE-Seq pipeline (Kelsey et al., 2025). Full methods can be found at https://doi.org/10.1186/s13100-025-00381-w. Briefly, sequencing reads were trimmed using fastp (Chen et al., 2018) and aligned to the mm39 genome using STAR (Dobin et al., 2013). Repetitive element expression was quantified using the telescope(Bendall et al., 2019) tool in “multi” mode. Differential expression analysis of aggregated RTE subfamily counts was performed using a negative binomial GLM with the glmmTMB package (Brooks et al., 2017). For this analysis, raw counts were summed across group members, and sample size factors estimated by DESeq2 (Love et al., 2014) were incorporated as an offset in the model. At both the E16.5 and E18.5 time points, genotype was modeled as a predictor.

Ethics statement

This study was approved by Brown University IACUC protocol #21-02-0005. The primary method of euthanasia is CO2 inhalation and the secondary method used is cervical dislocation both as per American Veterinary Medical Association (AVMA) guidelines on euthanasia. The study was conducted in accordance with the local legislation and institutional requirements.

Support

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The Brown University Flow Cytometry and Sorting Facility has received generous support in part by the National Institutes of Health (NCRR Grant No. 1S10RR021051) and the Division of Biology and Medicine, Brown University. As some of

our insights were gained by reprocessing publicly available scRNA-seq datasets, we greatly appreciate both the researchers that generated and shared the data initially and the respective repositories for making them available. We are grateful to the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) for their generous support through 1R01HD091848 to RF. We thank the generous support of the Howard Hughes Medical Institute for the Gilliam Fellowship to MB. We also thank the U.S. National Science Foundation Graduate Research Fellowship Program for their support of MB and this work. We also thank the US-Israel Binational Science Foundation (BSF) for their support to RF (2019285). We thank the Rhode Island Foundation and the Advance-CTR at Brown University supported by the NIH IDeA-CTR grant (U54GM115677) for support for these studies and its publication.

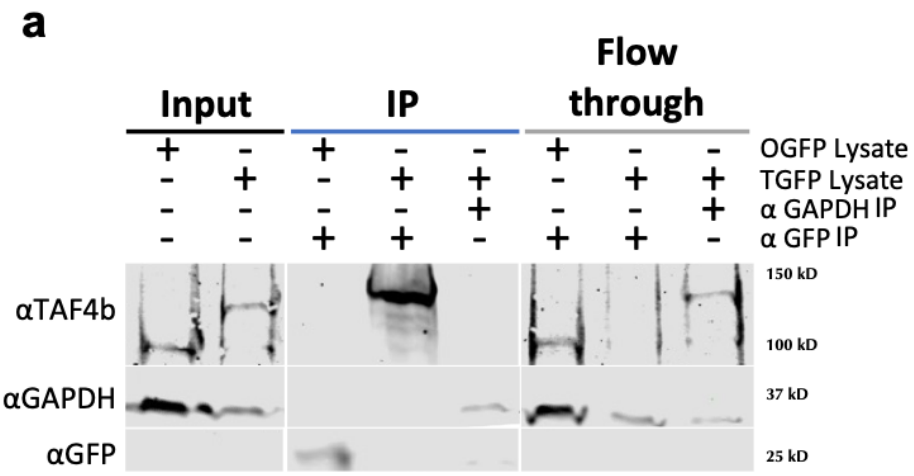
Acknowledgments

We thank Kevin Carlson and the Brown University Flow Cytometry and Sorting Facility for expertise completing the flow sorting. We thank the Center for Computation and Visualization at Brown University for computational resources for scRNA-seq and RNA-seq data analysis.

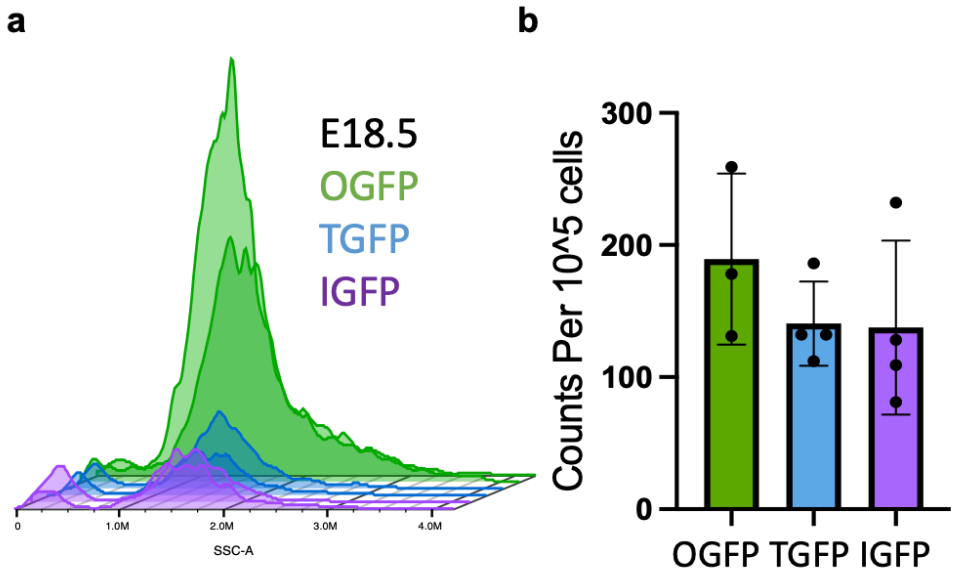
Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

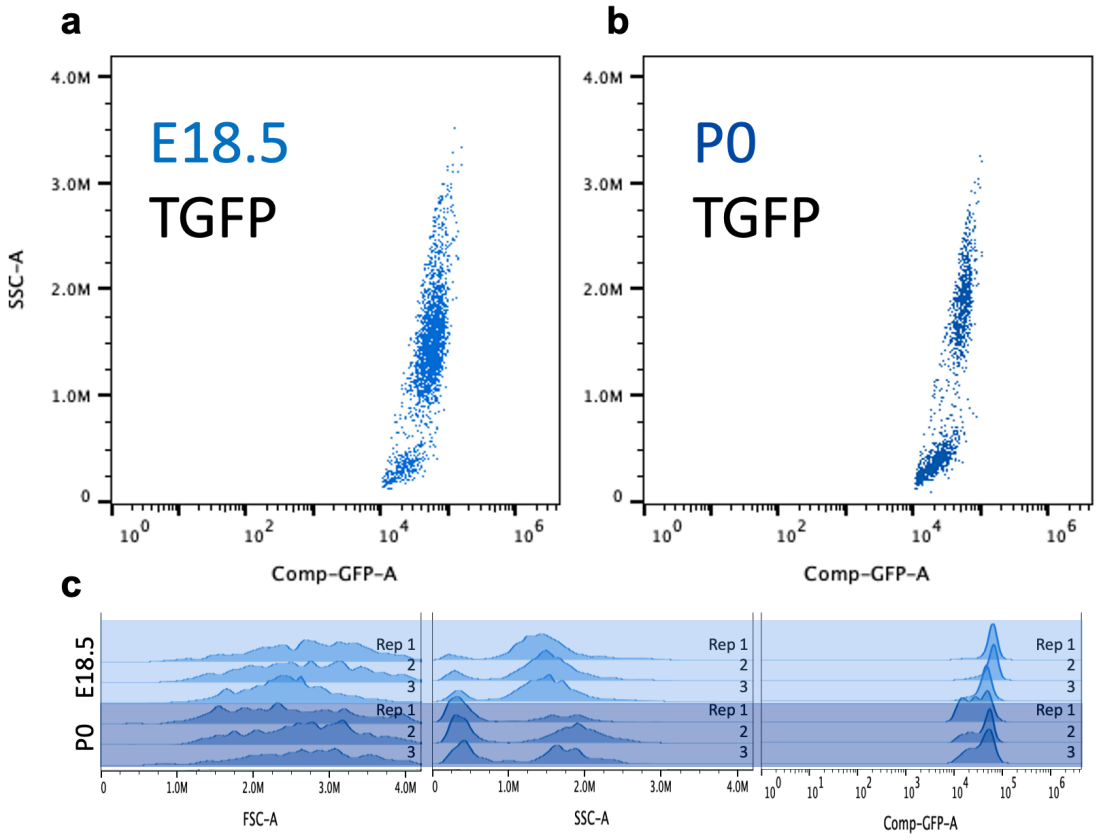
Supplementary material



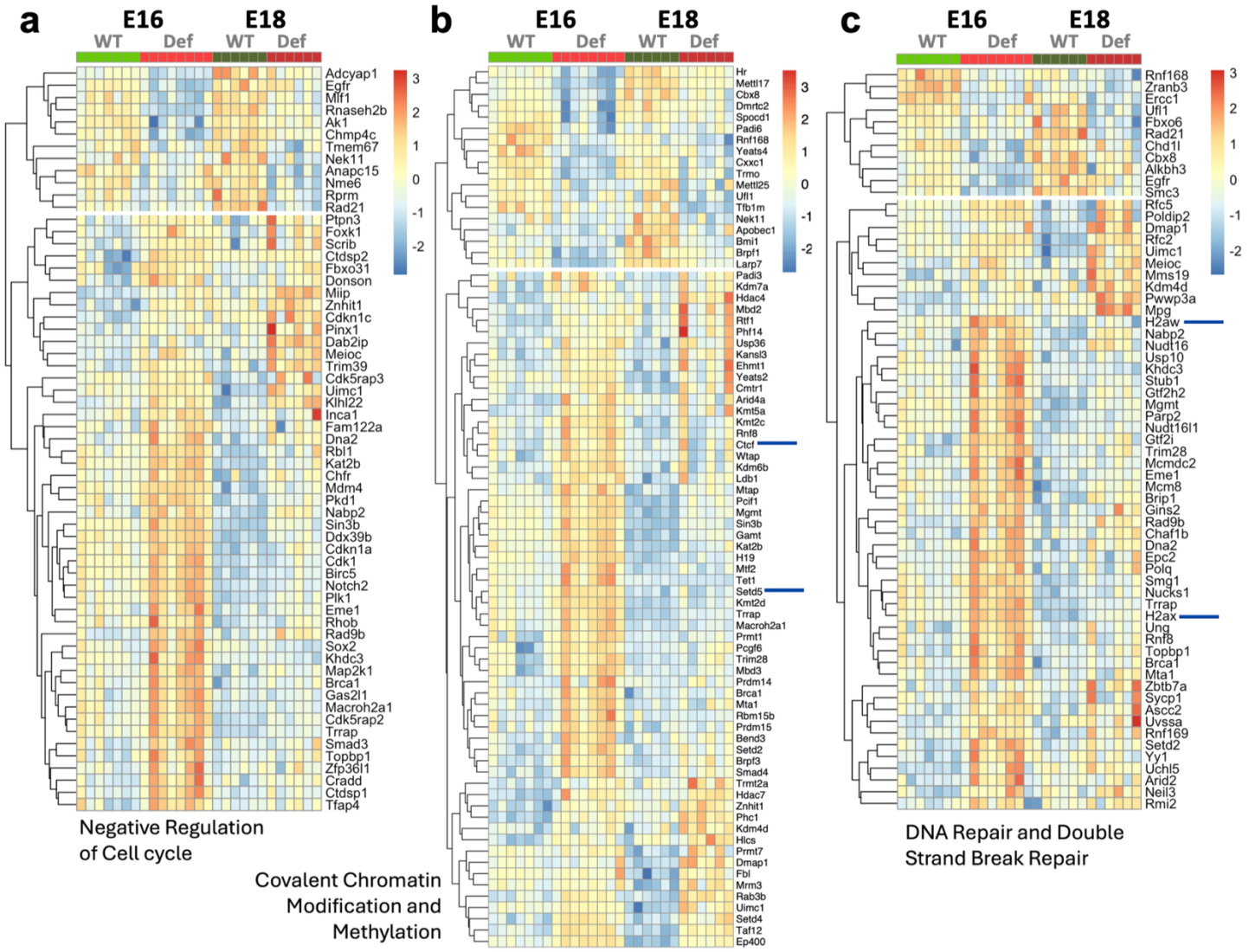
Supplementary Fig. 3.1 |Western blot analysis of protein extracts from Oct4-GFP and TAF4b-GFP testes validating TAF4b–GFP fusion knock in
a, Input lysates (left), anti-GFP or anti-GAPDH immunoprecipitations resulting bands (center), and corresponding flow-through fractions (right) are shown; antibodies used for detection are indicated adjacent to each blot.



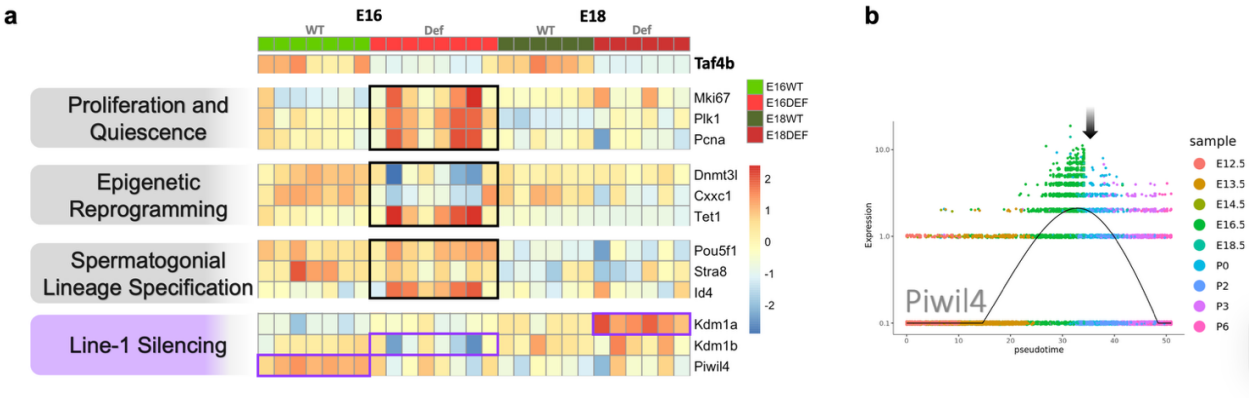
Supplementary Fig. 3.2 | Identification of LIC cells in multiple reporter lines
a, Comparative analysis of GFP expressing E18.5 prospermatogonia by FACS from several reporter strains including Oct-4-GFP (OGFP), TAF4b-GFP (TGFP), and Id4-Gfp (IGFP) at. **b,** Quantification of LIC cells showing no significant difference.



Supplementary Fig. 3.3 | From LIC cell establishment to birth
a,b TGFP sorts at E18.5 and P0, axes for GFP and side scatter. **c,** Cytometric parameters of FSC-A, SSC-A, and GFP across E18.5 and P0 TGFP sorts.

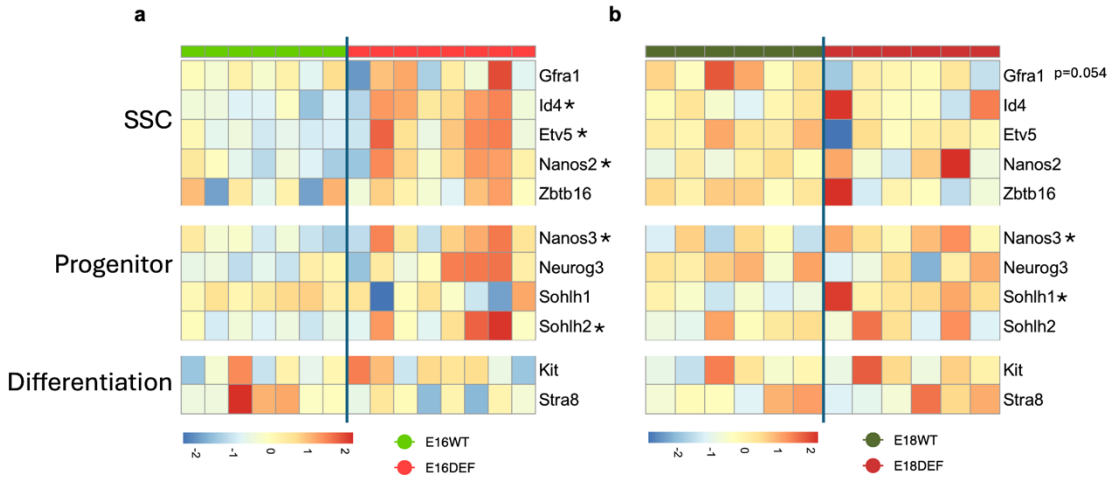


Supplementary Fig. 3.4 | DEGs in significantly disrupted GO pathways
a-c Heatmaps of GO biological processes and their DEGs from bulk RNA-seq across E16.5 to E18.5 and between Taf4b WT and Def. Selected Genes with TAF4b TSS occupancy (blue bars).



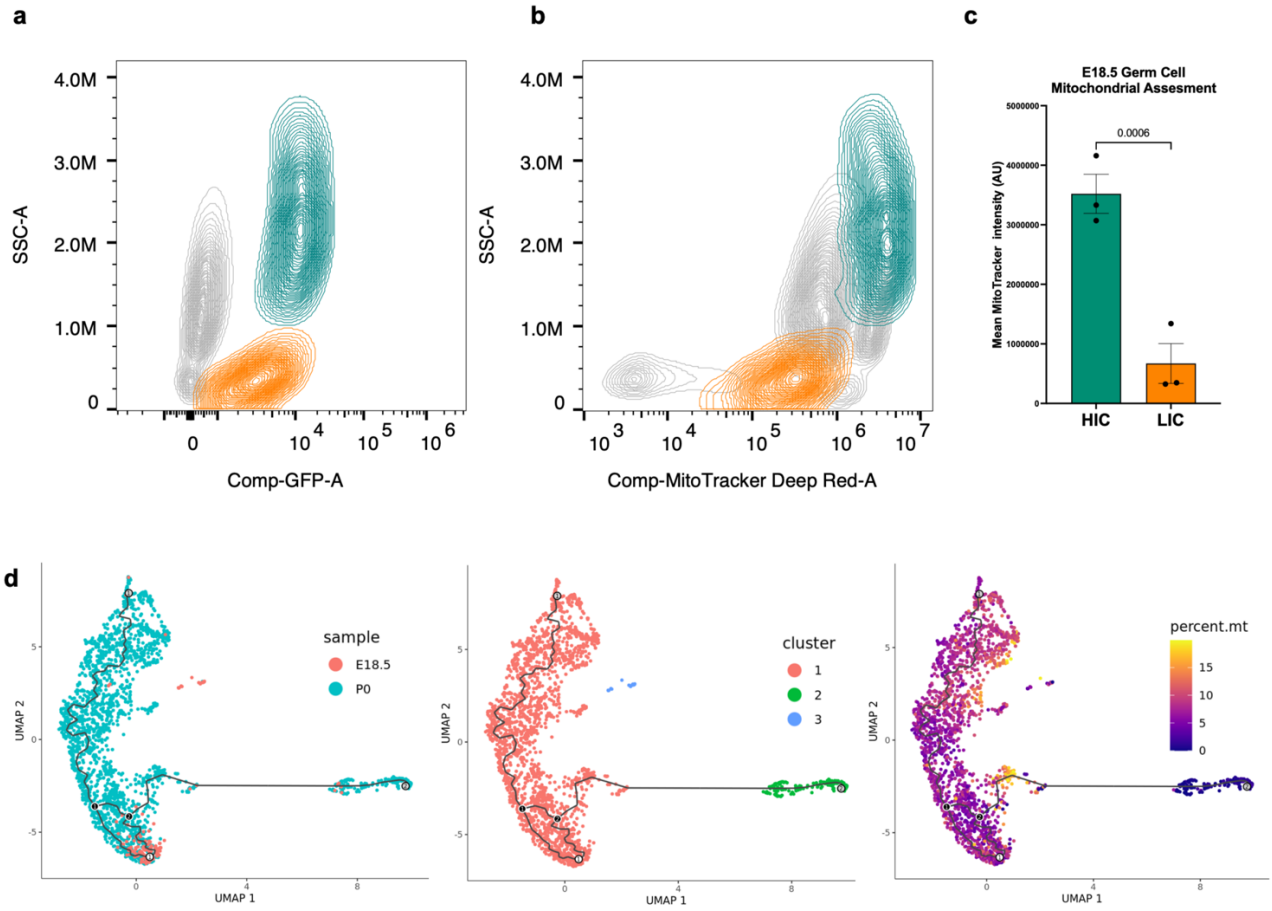
Supplementary Fig. 3.5 |Taf4b is necessary for proper PiwiL4 expression in addition to quiescence and other epigenetic programs

a, bulk-RNAseq heatmap of genes comparing WT Oct4-GFP+ and TAF4b Def Oct4-GFP+ prospermatogonia at both E16.5 and E18.5 **b,** Pseudotime gene expression dynamics for Piwil4 from E12.5 through P6 denoting the prominent increase at E16.5 followed by a sharp decline in expression immediately after E18.5 (black arrow).



Supplementary Fig. 3.6 |Taf4b is necessary for preserving canonical and finetuned SSC and differentiation programs

a-b, Heat map of markers and regulators of SSC, progenitor, and differentiation fates across WT and DEF at **a**, E16.5 and **b**, E18.5 from RNA seq. *P < 0.05, t-test.



Supplementary Fig. 3.7 |LIC fractions and subset of germ cells in scRNA-seq have significantly less mitochondrial presence

a, FACS of E18.5 germ cells by SSC-A and TGFP expression. **b**, FACS of E18.5 germ cells by SSC-A and Mito Tracker. **c**, Quantification of Mito Intensity, paired t-test. **d,** UMAPs of cells from E18.5 and P0 by sample timepoint, clustering, and mitochondrial ratio.

CHAPTER 4: Discussion and Future Directions

Myles A. Bartholomew[1]

I wrote this chapter incorporating feedback from Dr. Richard N. Freiman and my thesis committee members, Drs. Alison Delong, Diana Laird, Mark Johnson, and Nicola Neretti.

[1]MCB Graduate Program, Providence, RI, United States

The importance of this study

Across the globe, infertility has increased, with more and more reproductive-age couples finding themselves unable to conceive. This can be devastating for many hopeful families, and 50% of the increase is attributable to male factors (Vander Borght & Wyns, 2018). To stop or correct these disturbing trends in male fertility, we must understand its origins. This includes identifying the responsible cells, their developmental timelines, and key molecular regulators. There is a long history of focusing on sperm dysfunction, including motility and morphology. In addition, spermatogonial stem cells have also been studied and are regarded as the fundamental units for ongoing reproductive capacity. Our work, along with others, highlights that even their establishment and maintenance depend on molecular events that occur long before sexual maturity. By taking a focused look at late embryonic germ cell development, I not only identify regulatory nodes that are critical for future successful cell lineages, but also hope to identify possible targets for treating a subset of idiopathic male infertility

Summary

In Chapter 1, we laid the foundation for germ cell development by introducing a timeline from primordial germ cell (PGC) specification to the generation of spermatogonial stem cells (SSC). We discussed processes such as the quiescence period associated with exit from the cell cycle, dramatic changes in epigenetic status, and overall conservation

across many stem cell types. We also introduced TAF4b, which is shown as a key factor in human fertility. Given the hypothesized transcriptional regulatory role of TAF4b and the ubiquitous role of TFIID, we next review the various constituents, both basal and those proposed to intersect with TAF4b in directing germ cell development. We note that many of the members highlighted in Chapter 1 were identified as dynamic through single-cell pseudotime or as differentially expressed as a result of TAF4b ablation (Figure 4.1a). Throughout this review, it was noted that many of the TAFs have regulatory functions both through transcription preinitiation complex and other alternative mechanisms, such as histone modification or shaping protein localization. Many of these factors also play known roles in driving differentiation or preserving stemness (Luna-Arias & Castro-Muñozledo, 2024; Zhou et al., 2013). Together, the information imparted lends insights into potential germ cell relevant associations, novel regulatory mechanisms, and what downstream processes may be affected if these factors are disrupted in TAF4b's absence.

In both Chapter 2 and 3, significant work was completed, defining the once unknown roles and contributions of Taf4b to spermatogenic development and fertility. We began our studies by grounding our hypotheses in transcriptomics and pseudotime atlases. To this end, I created an integration spanning PGC colonization, the proliferative-to-quiescent transition, and prospermatogonial differentiation through birth. We were able to not only measure *Taf4b* dynamics but also infer which biological processes relied most on TAF4b regulation.

In Chapter 2, it was first apparent that *Taf4b* expression peaks during the well-established quiescence period. By examining markers of proliferation across via bulk RNAseq and immunofluorescence, we concluded that entry into quiescence was disturbed in *Taf4b*-deficient mice. When assessing proliferation, we noted that although Def cells lacked MKI67 at late embryonic time points, they showed increased PCNA expression. Upon further consideration and the completion of Chapter 3, this observation is even more intriguing because PCNA functions as a molecular tool belt, indicating not only proliferation but also DNA repair (Boehm et al., 2016). Thus, we further hypothesized that disruption of quiescence at this stage reflects underlying disturbances in genome maintenance during a critical developmental window. Additionally, in line with Du et al., 2021, timely entry into and progression through quiescence during embryogenesis may support the proper development of a strong SSC population.

In conjunction with disturbances in quiescence, we also gained insight into Taf4b's role in transcriptional regulation. In Chapter 2, GO analysis revealed substantial differential gene expression in pathways related to covalent chromatin modification. Motif analysis of TAF4b CUT&RUN data showed that TAF4b binds CpG-rich motifs outside of the TATA box, and many of these directly bound genes were upregulated in the absence of Taf4b. This was surprising, as TFIID and TAFs are more commonly associated with activation and transcription initiation. It also suggests that TAF4b may have an inhibitory function or work with alternative complexes. Beyond identifying direct Taf4b TSS occupancy in Chapter 2, Chapter 3 showed that many genes remained differentially regulated across the E16.5 to E18.5 time course. This reinforces that Taf4b-dependent transcriptional

control is sustained during late embryonic development. Since conducting this work, we have also examined the differential expression of *Taf4b* beyond embryonic stages using single-cell integration of adult mouse germ cells (Green et al., 2018; Nie et al., 2022). We observe that Taf4b remains elevated beyond birth and that adult spermatogonia and SSCs express TAF4b and members of its regulatory network that were differentially expressed at E16.5 and E18.5 (Figures 4.2b and 4.3b). It remains to be validated, but in some respects, this could indicate that similar processes are being regulated between late embryonic stages and the spermatogonial dynamics in mature males, or tangentially, that the same TAF4b-dependent mechanisms required to establish the SSC pool are also necessary to maintain it.

E18.5 represents a particularly critical time point for SSC establishment. This stage closely coincides with the initial formation of pioneering germ cells with stemness-like properties (Law et al., 2019). Building on this finding, and given that TAF4b expression is highly enriched at this timepoint, we investigated TAF4b's contributions to heterogeneity and stemness in Chapter 3. Through FACS analysis of GFP-TAF4b mice, prospermatogonia at E18.5 were found to form two distinct subpopulations based on side scatter, a direct measure of intracellular complexity. This identification of multiple germ cell subsets extends beyond heterogeneity defined by immunological markers such as GFRA1, ID4, or THY1. Ours is a cytometric distinction that reveals a major population of high intracellular complexity and a minor population of low intracellular complexity. This observed bifurcation in side-scatter also represents the first identification of two discrete

germ cell subsets arising in the embryonic gonad prior to birth and the first instance in which germ cell heterogeneity was distinguished by quantifiable cytometric properties.

The presence of this bifurcation across isolation strategies using different mouse reporter lines was reassuring. Given the time point at which it arises, we next examined the relationship between TAF4b and stemness. Chapter 3 provides evidence that, using the total TAF4b-GFP-positive (TGFP+) fraction, one subpopulation or both TGFP+ HIC and LIC fractions can restore spermatogenesis in infertile mouse models. These findings demonstrate that TAF4b-positive cells denote a spermatogonial stem cell-containing subpopulation as early as E18.5. Furthermore, the decrease in germ cell number and the more pronounced ablation of LIC cells in the E18.5 Taf4b-deficient testes indicate that TAF4b regulates the population dynamics of the late-fetal germ cell pool, and this population likely plays a yet-to-be-defined functional role in promoting SSC development. Together, the disturbances in quiescence and sustained transcriptional dysregulation converge at this critical window of SSC progenitor establishment and emerging heterogeneity.

Across Chapters 2 and 3, altered transcription and epigenetic dynamics in the Taf4b-Def state further reveal how genomic integrity and cell fate are interconnected. Many key regulators, including those directly controlled by TAF4b, such as *H2afx*, have been previously shown by CUT and RUN to influence multiple processes and can disrupt several pathways simultaneously, including DNA repair and the cell cycle (Gura et al., 2023; Podhorecka et al., 2010). Preliminary work supports this, as data indicate that

γH2AX, a phosphorylated form of H2AX that binds to double-strand DNA breaks (DSBs), is also elevated. We also observed changes in *Piwil4,* a critical driver of retrotransposable element silencing and epigenetic regulation (Supplementary Fig. 3.4) (Dias Mirandela et al., 2024). Given the multiple epigenetic nodes implicated in spermatogenic dysregulation, and in light of the ontologies of DNA damage repair and chromatin modification, we sought to further characterize TAF4b's role in genomic integrity.

One of the nodes that we further elucidated was DNA methylation. We note that it is normal for retrotransposable elements to temporarily lose DNA methylation during early development, similar to imprinted loci and other genomic regions (Cui et al., 2025; Dias Mirandela et al., 2024). However, in most prenatal germ cells, these elements are repressed after epigenetic reprogramming and before quiescence ends (Cui et al., 2025; Gkountela et al., 2015), as shown by our whole-genome bisulfite sequencing (WGBS) data and ORF1p staining in *Taf4b*-WT cells (Chapter 3). LINE1 elements can serve as indicators of incomplete or faulty epigenetic reprogramming and may also aid in removing malformed cells. We suggest that even modest hypomethylation of LINE1 could be particularly harmful, given that the expression of young, full-length transposable elements might cause insertion events and double-strand breaks (Dias Mirandela et al., 2024). The increased LINE1 expression observed in Chapter 3, combined with cell cycle arrest indicated by GSEA and GO analyses, may imply that embryonic germ cells are shifting toward a state more similar to senescence rather than quiescence. This could point to a coevolutionary mechanism where transposons serve as checks and balances in SSC establishment, signaling whether the germ cell genome has reached the correct

epigenetic state. In this paradigm, the epigenetically compromised cell does not silence regions of Line1 and other transposable elements, and thus the prolonged expression of these inflammatory and harmful elements drives cell elimination.

It is this cell elimination that also supports that TAF4b is a key driver of prospermatogonial development. The apoptotic phenotypes observed by TUNEL staining, along with the transcriptional profile of the cells that do exist at e18.5, being less disturbed overall, even though some genes never reach canonical expression levels like Piwil4, all further support that while a delay may originally exist, the timing of the events is important. It is this timing that we now know is TAF4b dependent. Taf4b Def is not the only mutation that highlights this model. Multiple other studies have shown infertility phenotypes when the timing of quiescence entry and even exit is disrupted or when epigenetic mechanisms go awry during embryonic time (Du et al., 2021; Nguyen et al., 2020). This may also mean there are further mechanisms to uncover by comparing our Taf4b Def model with other transgenic mouse models. The first would be Rb1 knockout mice, which are also reported to complete the first wave of spermatogenesis, but, due to deregulation of quiescence, the spermatogonial stem cell population fails (Du et al., 2021). Other models that may contribute to our understanding of what is going on in Taf4b Def are Dnmt3a and Dnmt3l knockout mouse, in which spermatogonial stem cells lose canonical DNA methylation and the ability to progress properly through spermatogenesis. We note in particular that Dnmt3l is downregulated alongside increases in LINE1 elements and apoptosis (Chapters 2 and 3). These parallels and future integrations of the different models could reveal specific loci that are critical to SSC development and maintenance.

Altogether, as we observe late embryonic germ cell development and its dependence on TAF4b, we conclude that increased LINE1 accumulation, DNA damage, and elevated TUNEL staining in the Def indicate a substantial insult to genomic integrity in a cell population poised to establish its foundational pool of spermatogonial stem cells. The timing of this germ cell loss during the window of SSC progenitor establishment and heterogeneity provides a potential cellular explanation for the impairment of prolonged fertility observed in mice and individuals with similar Taf4b mutations.

New perspectives to consider from the data

Taf4b Dependent Epigenetic Regulation as a Driver of Integrity

Alongside TAFs as a driver of differential gene expression and the specification of diverse cell lineages, there are other regulators outside of transcription or genotype. These factors, which extend beyond the genetic code to influence cell identity and behavior, are often referred to as epigenetic regulators (Stephens et al., 2013). The two primary types of epigenetic regulation known to be critical in germ cell development, as well as in stemness, pluripotency, and cell survival, are DNA methylation and histone modification (Cui et al., 2025). The work completed in chapters 2 and 3 demonstrates that multiple nodes and molecular modulators of these processes are disrupted in TAF4b's absence. Thus, unraveling the dynamics of DNA methylation and histone modification is critical to understanding TAF4b regulation.

DNA methylation has long been recognized as a key driver of gene expression, genomic stability, and the silencing of harmful elements. In male germ cell development, DNA methylation is considered a biphasic process. The first phase involves erasing genomic methylation. This return to low baseline DNA methylation is essential for reestablishing pluripotency programs in many stem cell types including SSCs (Lee et al., 2014). Moreover, studies show that DNA methylation alone confers analog properties to gene expression, with the level of 5-methylcytosine correlating with transcriptional repression (Palacios et al., 2025). Therefore, in a hypomethylated state, such as in PGCs and undifferentiated cells, the absence of methylation leads to increased expression of

stemness and pluripotency genes that would otherwise be repressed (Gkountela et al., 2015). To accomplish this erasure, it is believed that PGCs, which are highly proliferative and increase from about 40 cells to thousands within days, partially lose methylation marks as DNA is being synthesized but not methylated (Kagiwada et al., 2013; Resnick et al., 1992). Additionally, chromatin-modifying enzymes like TET actively convert 5-methylcytosine to hydroxymethylcytosine, further promoting the loss of methylation at specific genes (Hill et al., 2018; Tan & Shi, 2012).

The second phase, following erasure, is de novo methylation (Molaro et al., 2014). In mice and humans, this phase is closely linked to the quiescence period during germ cell development (Smallwood & Kelsey, 2012). It is believed that altering the epigenetic landscape is a key function of quiescence, and many of the regulators involved are only beginning to be understood (Reik & Dean, 2001; Zhao et al., 2024). Among these, DNA methyltransferases such as DNMT1, DNMT3A, DNMT3B, and DNMT3L have been studied in germ cells (Dura et al., 2022; Liao et al., 2014). DNMT3A and DNMT3L are essential for prolonged fertility, as DNMT3A mutants fail to differentiate, and DNMT3L mutants experience germ cell loss through apoptosis and retrotransposon activation. (Dura et al., 2022; Webster et al., 2005). It is also important to note that DNA methylation, although generally considered repressive, can, in some cases, promote gene expression; a notable example is OCT4, which preferentially binds methylated motifs in stem cells (Lyko, 2018; Tan et al., 2023). Taken together, DNMTs modulate gene expression, chromatin structure, and the overall genomic integrity essential for cell survival. It is also important that we further characterize if and how they are regulated by TAF4b, as many

of them, as well as other members of their complexes, are differentially expressed in the infertile Taf4b Def males (Chapter 3).

Another important player linking transcription and epigenetics is CCCTC-binding factor (CTCF). Studies have shown that CTFC is critical for male fertility, and its differential expression level in the absence of Taf4b further implicates CTCF as a regulator of epigenetic status in germ cell development (Figure 3.6d) (Hernández-Hernández et al., 2016). This multifunctional transcription factor was found to inhibit Myc, a master regulator of stemness and proliferation. At the molecular level, CTCF functions as an insulator, altering chromatin architecture and maintaining the demethylated state of specific genes (Wei et al., 2023). In exchange, this factor has also been shown to be sensitive to changes in DNA methylation, specifically at the CTCF motif, which can inhibit CTCF binding (Cui et al., 2025; Wei et al., 2023). Future studies could explore where CTCF-bound insulators are situated within the Taf4b Def model. Together, these discoveries further emphasize the interdependence and sensitive integration of transcription, epigenetics, and stemness.

DNA methylation does not operate in isolation; it often correlates with and is directed by changes in histone states (Sasaki & Sangrithi, 2023). Histones, the fundamental components of nucleosomes, function as molecular spools within a protein complex (Luger et al., 1997). They are essential for organizing genomic material and making it accessible during various developmental stages. Each nucleosome can associate with over 140 bases of DNA, and each histone has a tail that can carry specific modifications

and marks (Cui et al., 2025; Luger et al., 1997). These marks interact with the broader genomic environment and influence the behavior of transcription factors and chromatin modifiers. It is important to recognize that histone modifications are highly context-specific, affecting gene expression, DNA integrity, and germ cell development (Cui et al., 2025).

Among the most studied marks are H3K4me3 and H3K4me2. Their presence at specific loci signals gene activation, often found in promoters and CpG islands (Barsoum et al., 2022). They influence gene expression through various mechanisms, including H3K4me3's ability to affect DNA methylation by preventing DNMT enzymes from accessing these regions (Li et al., 2011; Ooi et al., 2007). These histone marks are deposited by SETD1A/COMPASS and removed by KDM1b (Stewart et al., 2016). H3K36me3, on the other hand, reveals even more nuance. It is associated with both gene activation and repression of cryptic transcripts (Carrozza et al., 2005; Chantalat et al., 2011; Suzuki et al., 2016; Weinberg et al., 2019). This mark is highly enriched in the gene body, and its deposition is influenced by SET2 along with Pol II (Suzuki et al., 2016). Furthermore, H3K36me3 is noted to play a major role in facilitating DNA methylation by recruiting DNMT3a and DNMT3b (Weinberg et al., 2019).

There are also other factors that integrate DNA methylation, histone modification, and transcription regulation. For instance, the chromatin modifier CXXC1 plays an active role in the SETD1A/COMPASS complex and increases chromatin accessibility in developing germ cells (Yu et al., 2017). In another study, CXXC1 was recruited by SETD1A to

enhancers, where it correlated with increased gene expression, elevated H3K27ac, and H3K4me3 in a context-dependent manner (Greulich et al., 2021). Moreover, research into CXXC1's role in DNA methylation shows that it binds to unmethylated CpGs and can interact with DNMT1 outside of the COMPASS complex (Butler et al., 2008; Thomson et al., 2010; Yu et al., 2017). Further studies are needed to clarify its role in the establishment of spermatogonial stem cells. One study that may provide insight is observing CXXC1 genomic occupancy via CUT&RUN. This could indicate direct regulation of stemness-related genes or biological processes. It is also worth using Cxxc1 Def mouse models and assessing their FACS profiles to see whether germ cell populations are affected similarly to those in Taf4b Def. Thus far, data from the CXXC1 Def mouse models suggest that Cxxc1 could be equally important in genomic integrity in both spermatogenesis and oogenesis (Jiang et al., 2020; Park et al., 2025). Lastly, another key regulator of genomic integrity, acting through both DNA methylation and histone modification, is PIWIL4 (Nagamori et al., 2018). Defects in this gene and the piRNA pathway result in abnormal expression of L1 elements, and it is well established that repression via this pathway is essential for spermatogenesis (Hwang et al., 2020; Nagamori et al., 2018).

Throughout our study, we observed changes in the expression of DNMTs, CTCF, TETs, CXXC1, and PIWIL4 in the absence of TAF4b, suggesting that TAF4b-dependent transcriptional networks, as well as novel associations among multiple epigenetic architects, may influence genomic integrity and cell fate (Figure 4.1).

Germ Cell Heterogeneity and Lineage Determination as an Outcome

One of the most important perspectives highlighted in our work concerns TAf4b's contribution to heterogeneity. We define this heterogeneity as the variability among phenotypes, expression levels, and cell states within a given population. There are several metrics of heterogeneity that could be applied to our model of TAF4b-driven fertility. The first is asynchronous changes, defined as cells on a temporal spectrum, with some cells progressing further or lagging behind in a specific process. In the context of male germ cells during embryonic development, one aspect under this umbrella is positionality. The importance of this is most clearly observed in the cells that are capable of migrating or occupying the gonad. Studies have shown that cells leading the migration reach the gonad, whereas lagging cells remain outside the gonad at E13.5 (Jaszczak et al., 2025). Another example of positionality is where the cell is located within the developing tubule. Multiple studies and microscopic anatomy observations have confirmed that, like many stem cell populations, SSCs occupy a niche, and at some point during development, prospermatogonia must traverse the tubule and settle in the basement membrane (de Rooij et al., 2008; Kanatsu-Shinohara et al., 2008). These observations raise the question of whether there is a temporal node in Taf4b regulation, such as cells that lead migration having higher or lower Taf4b expression, or whether the heterogeneity we observe in intracellular complexity is related to position in the tubule.

Another key metric of heterogeneity is marker gene expression. This is a broad indicator that can reflect many aspects of spermatogonial development discussed earlier. Examples include the expression of OCT4 (POU5f1), ID4, and, as we show here, TAF4b.

These markers, though they show overlap in some cells, are highly variable, with a fraction of prospermatogonia having very high levels and others little to none. Our work demonstrates that TAF4b expression is heterogeneous, with only about 2,000 prospermatogonia expressing it at E18.5. Related to the heterogeneity observed here in TAF4b expression, OCT4 is a pluripotency marker and Yamanaka factor that also isolated multiple germ cell fractions by intracellular complexity (Chapter 3) (Shi & Jin, 2010). In mice, this protein is highly expressed in prospermatogonia, while in humans, evidence suggests that its expression remains enriched in undifferentiated SSCs. (Dann et al., 2008; Liu et al., 2011). Similarly, ID4 is recognized as a stem cell marker, with studies showing heterogeneous expression, including positive, negative, and further subgroups such as ID4 Bright, ID4 Dim, and ID4 Negative cells (Law et al., 2019). Together, the heterogeneity and selective expression observed during late embryonic stages, particularly during HIC and LIC specification at E18.5, support the broader concept that gene expression shapes cell identity and function. Building on this, we have begun examining the intersections of marker heterogeneity, regulatory networks, and their conservation between mice and humans (Figure 4.2c and 4.3). As we strengthen our appreciation for heterogeneity, I propose that a combinatorial perspective may be especially informative. For example, a (TAF4b+ CXXC1+ KI67-) cell may represent an SSC undergoing epigenetic remodeling during quiescence, whereas a (TAF4b+ CXXC1- KI67+) cell may reflect an SSC that has completed epigenetic reprogramming and is primed for self-renewal. Another example would be (TAF4b+ OCT4+ TCF3+), representing the least differentiated state of an SSC, whereas (TAF4b+ OCT4- TCF3-) represents the more differentiated SSC state. Here, the interpretation of cell identity is

more analog than discrete, and cells would fall more on a spectrum for any particular state. Currently, neither scenario is validated, and the exact trend may vary, but we have tools and models that may answer these questions. (Figure 4.3b,c).

Heterogeneity does not only apply to the lineages that survive but also to the lineages and cells that exhibit early signs of failure. In our search to elucidate the molecular underpinnings of spermatogonial stem cell establishment, we have found that several indicators of apoptosis are elevated in TAF4b-Deficient mice (Chapter 3). Germ cell apoptosis occurs throughout development and is typically indicative of a selective pressure (De Pol et al., 1997). Molecular cues for apoptosis align with aforementioned changes in DNA methylation, LINE1 expression, and unresolved double-strand breaks, which are all strong signals of compromised genomic integrity (De Zio et al., 2013). It is also well established that both males and females undergo apoptosis and culling (De Pol et al., 1997). Under normal conditions, this pathway protects against the development of teratomas from PGCs that migrate incorrectly or fail to colonize the gonad. In later male development, it protects against spermatogonial stem cells, which could lead to testicular cancer if transcriptional, epigenetic, or cell-cycle controls become dysregulated (Chen et al., 2012). Though it has many protective functions it is understood that dysregulation of apoptosis is directly linked to infertility. This has been demonstrated through manipulations in cell-cycle regulation mediated by TGF-β, TAf4b, γH2AX, and numerous other checkpoint proteins (Chapter 3) (Bejarano et al., 2018; Prabhu et al., 2024). Together with our own findings, we center the ability to respond to stress, possibly induced by transcriptional or epigenetic changes, as a key determinant of germ cell fate.

Effective treatment of male infertility may depend on understanding why certain cells bypass this pathway, either by maintaining higher genomic integrity or by possessing genes that confer survival benefits.

Future Directions

The question of which cells are being isolated in the LIC population still remains, and we propose two potential models. The first suggests that the LIC population specified at E18.5 indicates the formation of a spermatogonial stem cell lineage. This idea is supported by work showing that germ cells with low side scatter and intracellular complexity have higher stemness capacity (Shinohara et al., 2000). Still, it is important to note that no data exist in the field that isolates two distinct populations according to side scatter, and thus far, this study only demonstrates that spermatogonial stemness correlates with low side scatter (Shinohara et al., 2000).

The second model is equally intriguing due to the longstanding hypothesis of the first and second waves of spermatogenesis (Jones et al., 2021; Kluin & de Rooij, 1981). In this alternative model, LIC cells could represent cells that are in the process of differentiating. This suggests that as cells go through spermatogenesis, changes in their morphology and cytometric properties make the differentiating cells distinct from other germ cells. In this model, we would hypothesize that the cells that we observe in the LIC are either A1 spermatogonia or cells that are mitotically active, as described by Kluin's histological studies (Kluin & de Rooij, 1981). This is primarily due to the lack of a nuclear envelope and condensed chromatin, both of which may contribute to the

observed low intracellular complexity. This model is also consistent with a study in human pluripotent stem cells, which showed that cell side scatter area intensity decreased with differentiation (Ramirez et al., 2013).

In either case, the first step to identify the characteristics of the novel LIC population would be germ cell transplants of the LIC population, along with transplants of the more abundant HIC population. Then, we would evaluate which, if any, has a higher capacity for spermatogenic rescue. I also propose that we complete quantifications of total cells from the transplant that colonize the niche, even in tubules lacking visible spermatogenesis. In this experiment, we would stain for GFP along with stemness markers such as UCHL1 to identify SSCs that originated from transplantation. This is a critical experiment because it is still unknown whether all stem cells contribute to spermatogenesis when transplanted into the testes, as some cells may remain dormant or quiescent or have different proliferation rates, which could cause them to be missed in measurements of tubule spermatogenesis.

As stated in Chapter 3 these investigations would also greatly benefit from single-cell RNA sequencing of the two germ cell populations, as it is likely that, because this is the first reported isolation of LIC cells, the existing single-cell datasets that apply FACS enrichment of germ cells simply do not include LIC. Moreover, the non-sorted testis single-cell sample collected at E18.5 by Tan et al. consists of only a few hundred germ cells among the entire single-cell population. Thus, we would likely capture only a handful

of LIC germ cells and still lack a marker to definitively identify them within the single-cell structure.

In future studies, we also aim to directly identify the protein complexes that associate with TAF4b in embryonic germ cells. Currently, the transcription preinitiation complex stands at the center of gene regulation and, ultimately, cell differentiation. It drives the behavior of diverse cell types, including stem cells. While TAF4b has been shown to alter the conformation and structure of TFIID and regulate gene activity through the kinetics of the larger preinitiation complex (PIC), it remains unresolved which factors are physically in complex with TAF4b in embryonic germ cells and how these interactions drive the transcriptional and epigenetic programs required for spermatogonial stem cell establishment. Dikstein et al. showed that TAF4b can interact with the TFIID complex in B cells, yet TAF4b is not found as a TFIID constituent in HeLa cell lines, highlighting context-dependent incorporation.

Given that TFIID canonically consists of TBP and multiple TAF subunits whose combinations dictate gene expression and cell specificity, one possibility is that TAF4b participates in a germ cell specific TFIID variant during late embryogenesis. Other TAFs that determine cell fate and behavior, including TAF1, TAF7, TAF7L, TAF8, TAF10, and TAF12, are highly dynamic during germ cell development and/or differentially expressed in the absence of TAF4b (Figure 4.1). TAF1 provides a direct link between transcription and chromatin modification, TAF7 and TAF7L influence promoter pausing and proliferation, and TAF8/TAF10 heterodimerization regulates nuclear localization and

complex assembly. TAF4b itself has been shown to interact with TAF12 via conserved histone fold domains. Together, these observations underscore how TAFs influence each other's subcellular localization, chromatin engagement, and transcriptional output.

We have already begun investigating some of the possible TAF4b interactions via co-immunoprecipitation (co-IP), followed by Western blotting (Figure 4.1e,f). Furthermore, through unbiased proteomic approaches in purified germ cell populations, particularly at E16.5–E18.5 during the proliferative to quiescent transition and LIC emergence, we will determine whether TAF4b functions within canonical TFIID, a modified TFIID complex, or an alternative chromatin-associated complex.

Furthermore, our identification of a TAF4b-dependent low intracellular complexity (LIC) prospermatogonial population suggests that TAF4b-containing complexes may be uniquely assembled in this subset during late embryonic development. Characterizing the composition of TAF4b complexes specifically within LIC versus HIC populations may reveal whether differential PIC assembly underlies germ cell heterogeneity and lineage determination, though it is substantially more challenging.

Given that TAF4b deficiency alters DNA methylation dynamics, LINE1 repression, histone associated regulators, and apoptosis pathways, it is plausible that TAF4b dependent transcriptional networks influence genomic integrity through multi-layered mechanisms. TAF4b may regulate chromatin accessibility through interaction with histone acetylase containing complexes, influence DNA methylation by modulating DNMT expression, or

coordinate transcription with retrotransposon repression programs. We plan to also determine whether there are direct physical interactions with some of these less canonical complexes by completing co-IP mass specs with proteins identified in this study such as CXXC1, DNMT1 and others. Some of the preliminary investigations done so far have revealed that CXXC1 is associated with 9 TAFs that do not appear in any GAPDH IP controls. The list of TAFs was used as input to STRING; all were observed to have some degree of protein-protein interaction (Figure 4.1b).

Together, these experiments will clarify whether TAF4b drives transcriptional regulation through TFIID/PIC and/or through coordination with epigenetic architects under a biologically relevant context. The work here and to come will define which factors interact with TAF4b, provide mechanistic insights into how SSC programs are established, and may hold the answer to idiopathic male infertility.

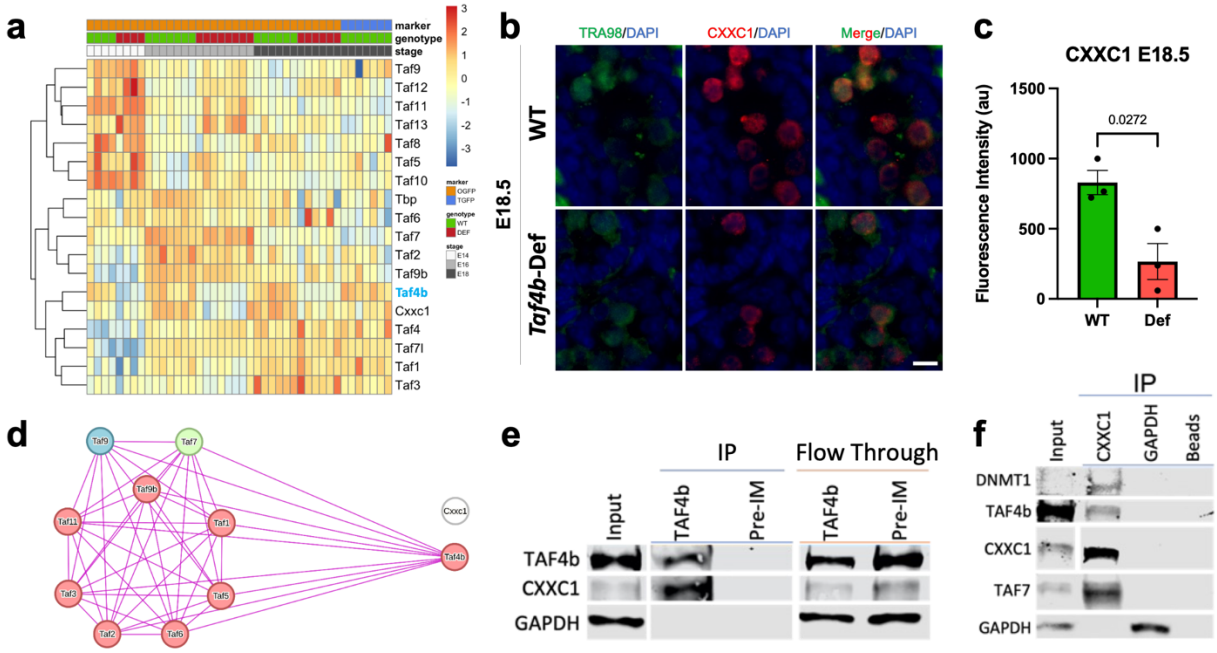


Figure 4.1 | TFIID subunits and CXXC1 exhibit coordinated transcriptional and biochemical association with TAF4b

a, Heatmap of bulk RNA-seq from E14.5, E16.5, and E18.5 germ cells. Samples are annotated by developmental stage (gray gradient), genotype (WT, green; Taf4b-deficient, red), and sort marker (OGFP or TGFP). Displayed genes include canonical TFIID subunits and Cxxc1. Hierarchical clustering identifies transcripts most closely aligned with Taf4b-dependent expression dynamics. **b,** Immunofluorescence of TRA98 and CXXC1 at E18.5 in WT and Taf4b-deficient testes. **c,** Quantification of CXXC1 immunofluorescence intensity from (b). **d,** Co-immunoprecipitation of CXXC1 followed by mass spectrometry. STRING clustering reveals enrichment of TFIID/pre-initiation complex components (pink nodes); pink edges denote experimentally validated protein–protein interactions. **e,** Co-immunoprecipitation for TAF4b in C18-4 cells using immune versus pre-immune serum, followed by western blot. **f,** Co-immunoprecipitation for CXXC1 along with GAPDH and bead-only controls.

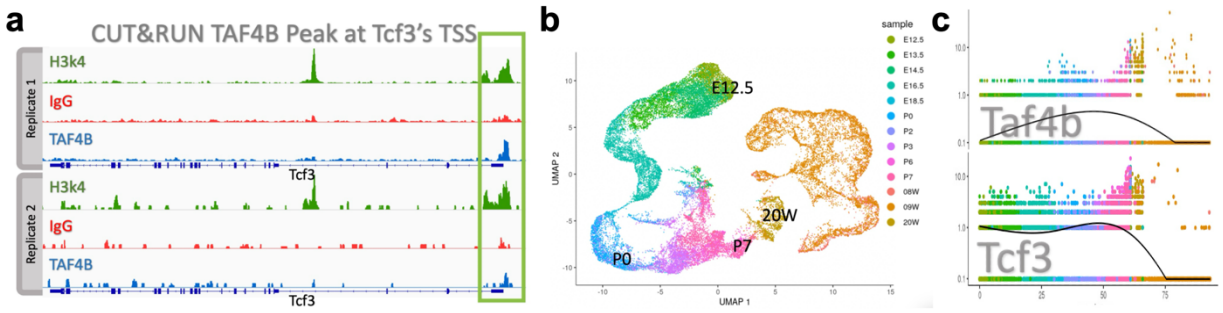


Figure 4.2| TAF4b occupies the Tcf3 promoter and exhibits coordinated dynamics in adult male mice.
a, CUT&RUN, TAF4b occupancy at the transcription start site (TSS) of Tcf3, green box. **b,** Integrated single-cell RNA-seq analysis of mouse germ cells across developmental time, incorporating embryonic datasets together with postnatal stages from Green et al. **c,** Pseudotime analysis showing coordinated expression dynamics of Taf4b and Tcf3 across germ cell development and intro adult hood.

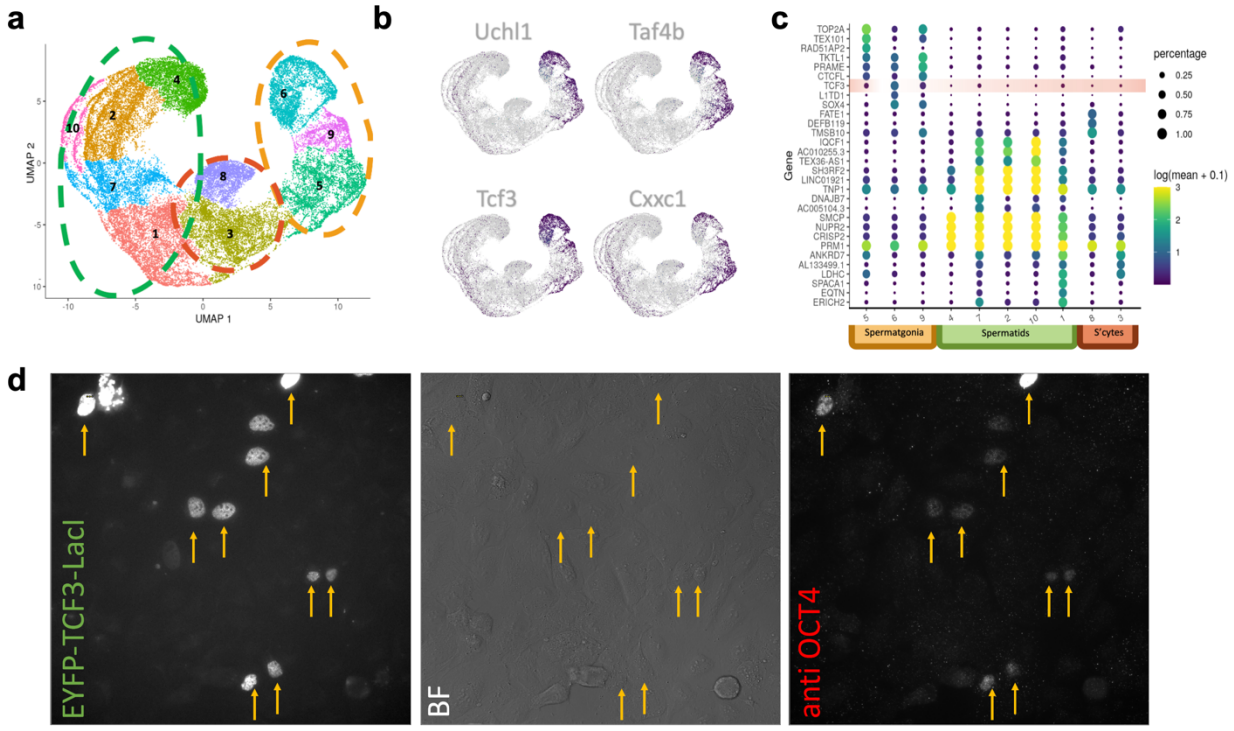


Figure 4.3| Stemness Marker expression and TAF4b regulatory network in human SSCs.

a, Reanalysis of human male single-cell RNA-seq data spanning spermatogenesis. unbiased clustering and UMAP. Major germ cell populations are annotated as spermatogonia (orange), spermatocytes (red), and spermatids (green). (Nei et al. GSE182786). **b,** Feature plots showing expression of UCHL1, TAF4B, TCF3, and CXXC1 across UMAP. **c,** Top marker analysis identifying TCF3 as the highest-ranked marker for the most undifferentiated spermatogonial cluster (cluster 6). **d,** TCF3 overexpression in a human cell line via EYFP-TCF3 reporter plasmid transfection. EYFP(left), brightfield image (middle), and OCT4 immunostaining (right). Arrows indicate successfully transfected cells.

BIBLIOGRAPHY

Abbasi, H., Tahmoorespur, M., Hosseini, S. M., Nasiri, Z., Bahadorani, M., Hajian, M., Nasiri, M. R., & Nasr-Esfahani, M. H. (2013). THY1 as a reliable marker for enrichment of undifferentiated spermatogonia in the goat. *Theriogenology*, *80*(8), 923–932. https://doi.org/https://doi.org/10.1016/j.theriogenology.2013.07.020

Acebron, Sergio P., Karaulanov, E., Berger, Birgit S., Huang, Y.-L., & Niehrs, C. (2014). Mitotic Wnt Signaling Promotes Protein Stabilization and Regulates Cell Size. *Molecular Cell*, *54*(4), 663–674. https://doi.org/https://doi.org/10.1016/j.molcel.2014.04.014

Alkandari, M. H., & Zini, A. (2021). Medical management of non-obstructive azoospermia: A systematic review. *Arab J Urol*, *19*(3), 215–220. https://doi.org/10.1080/2090598x.2021.1956233

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data* . http://www.bioinformatics.babraham.ac.uk/projects

Arya, S. T., & Dibb, B. (2016). The experience of infertility treatment: the male perspective. *Hum Fertil (Camb)*, *19*(4), 242–248. https://doi.org/10.1080/14647273.2016.1222083

Ashapkin, V., Suvorov, A., Pilsner, J. R., Krawetz, S. A., & Sergeyev, O. (2023). Age-associated epigenetic changes in mammalian sperm: implications for offspring health and development. *Hum Reprod Update*, *29*(1), 24–44. https://doi.org/10.1093/humupd/dmac033

Ayhan, Ö., Balkan, M., Guven, A., Hazan, R., Atar, M., Tok, A., & Tolun, A. (2014). Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J Med Genet*, *51*(4), 239–244. https://doi.org/10.1136/jmedgenet-2013-102102

Bahat, A., Kedmi, R., Gazit, K., Richardo-Lax, I., Ainbinder, E., & Dikstein, R. (2013). TAF4b and TAF4 differentially regulate mouse embryonic stem cells maintenance and proliferation. *Genes Cells*, *18*(3), 225–237. https://doi.org/10.1111/gtc.12030

Barsoum, M., Stenzel, A. T., Bochyńska, A., Kuo, C.-C., Tsompanidis, A., Sayadi-Boroujeni, R., Bussmann, P., Lüscher-Firzlaff, J., Costa, I. G., & Lüscher, B. (2022). Loss of the Ash2l subunit of histone H3K4 methyltransferase complexes reduces chromatin accessibility at promoters. *Scientific Reports*, *12*(1), 21506. https://doi.org/10.1038/s41598-022-25881-0

Basford, C., Forraz, N., & McGuckin, C. (2010). Optimized multiparametric immunophenotyping of umbilical cord blood cells by flow cytometry. *Nature Protocols*, *5*(7), 1337–1346. https://doi.org/10.1038/nprot.2010.88

Bejarano, I., Rodríguez, A. B., & Pariente, J. A. (2018). Apoptosis Is a Demanding Selective Tool During the Development of Fetal Male Germ Cells. *Front Cell Dev Biol*, *6*, 65. https://doi.org/10.3389/fcell.2018.00065

Ben Maamar, M., Beck, D., Nilsson, E., McCarrey, J. R., & Skinner, M. K. (2022). Developmental alterations in DNA methylation during gametogenesis from primordial germ cells to sperm. *iScience*, *25*(2), 103786. https://doi.org/10.1016/j.isci.2022.103786

Bendall, M. L., Mulder, M. d., Iñiguez, L. P., Lecanda-Sánchez, A., Pérez-Losada, M., Ostrowski, M. A., Jones, R. B., Mulder, L. C. F., Reyes-Terán, G., Crandall, K. A., Ormsby, C. E., & Nixon, D. F. (2019). Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLOS Computational Biology*, *15*(9), e1006453. https://doi.org/10.1371/journal.pcbi.1006453

Bhattacharya, S., Lou, X., Hwang, P., Rajashankar, K. R., Wang, X., Gustafsson, J., Fletterick, R. J., Jacobson, R. H., & Webb, P. (2014). Structural and functional insight into TAF1-TAF7, a subcomplex of transcription factor II D. *Proc Natl Acad Sci U S A*, *111*(25), 9103–9108. https://doi.org/10.1073/pnas.1408293111

Boehm, E. M., Gildenberg, M. S., & Washington, M. T. (2016). The Many Roles of PCNA in Eukaryotic DNA Replication. *Enzymes*, *39*, 231–254. https://doi.org/10.1016/bs.enz.2016.03.003

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R Journal*, *9*(2), 378–400. https://doi.org/Doi 10.32614/Rj-2017-066

Butler, J. S., Lee, J. H., & Skalnik, D. G. (2008). CFP1 interacts with DNMT1 independently of association with the Setd1 Histone H3K4 methyltransferase complexes. *DNA Cell Biol*, *27*(10), 533–543. https://doi.org/10.1089/dna.2007.0714

Capel, B. (2017). Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nature Reviews Genetics*, *18*(11), 675–689. https://doi.org/10.1038/nrg.2017.60

Capel, B. (2019). WOMEN IN REPRODUCTIVE SCIENCE: To be or not to be a testis. *Reproduction*, *158*(6), F101–F111. https://doi.org/10.1530/rep-19-0151

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M. P., & Workman, J. L. (2005). Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell*, *123*(4), 581–592. https://doi.org/10.1016/j.cell.2005.10.023

Chantalat, S., Depaux, A., Héry, P., Barral, S., Thuret, J. Y., Dimitrov, S., & Gérard, M. (2011). Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res*, *21*(9), 1426–1437. https://doi.org/10.1101/gr.118091.110

Chen, H., Lam Fok, K., Jiang, X., & Chan, H. C. (2012). New insights into germ cell migration and survival/apoptosis in spermatogenesis: Lessons from CD147. *Spermatogenesis*, *2*(4), 264–272. https://doi.org/10.4161/spmg.22014

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cheng, D., Semmens, K., McManus, E., Chen, Q., Meerzaman, D., Wang, X., Hafner, M., Lewis, B. A., Takahashi, H., Devaiah, B. N., Gegonne, A., & Singer, D. S. (2021). The nuclear transcription factor, TAF7, is a cytoplasmic regulator of

protein synthesis. *Sci Adv*, *7*(50), eabi5751. https://doi.org/10.1126/sciadv.abi5751

Cheng, Y., Buffone, M. G., Kouadio, M., Goodheart, M., Page, D. C., Gerton, G. L., Davidson, I., & Wang, P. J. (2007). Abnormal sperm in mice lacking the Taf7l gene. *Mol Cell Biol*, *27*(7), 2582–2589. https://doi.org/10.1128/mcb.01722-06

Cui, Y., Deng, J., Zhang, Y., Du, L., Jiang, F., Li, C., Chen, W., Zhang, H., & He, Z. (2025). Epigenetic regulation by DNA methylation, histone modifications and chromatin remodeling complexes in controlling spermatogenesis and their dysfunction with male infertility. *Cell Mol Life Sci*, *82*(1), 343. https://doi.org/10.1007/s00018-025-05831-5

Dann, C. T., Alvarado, A. L., Molyneux, L. A., Denard, B. S., Garbers, D. L., & Porteus, M. H. (2008). Spermatogonial Stem Cell Self-Renewal Requires OCT4, a Factor Downregulated During Retinoic Acid-Induced Differentiation. *STEM CELLS*, *26*(11), 2928–2937. https://doi.org/https://doi.org/10.1634/stemcells.2008-0134

De Pol, A., Vaccina, F., Forabosco, A., Cavazzuti, E., & Marzona, L. (1997). Apoptosis of germ cells during human prenatal oogenesis. *Hum Reprod*, *12*(10), 2235–2241. https://doi.org/10.1093/humrep/12.10.2235

de Rooij, D. G., Repping, S., & van Pelt, A. M. (2008). Role for adhesion molecules in the spermatogonial stem cell niche. *Cell stem cell*, *3*(5), 467–468. https://doi.org/10.1016/j.stem.2008.10.010

De Zio, D., Cianfanelli, V., & Cecconi, F. (2013). New insights into the link between DNA damage and apoptosis. *Antioxid Redox Signal*, *19*(6), 559–571. https://doi.org/10.1089/ars.2012.4938

Dias Mirandela, M., Zoch, A., Leismann, J., Webb, S., Berrens, R. V., Valsakumar, D., Kabayama, Y., Auchynnikava, T., Schito, M., Chowdhury, T., MacLeod, D., Xiang, X., Zou, J., Rappsilber, J., Allshire, R. C., Voigt, P., Cook, A. G., Barau, J., & O'Carroll, D. (2024). Two-factor authentication underpins the precision of the piRNA pathway. *Nature*, *634*(8035), 979–985. https://doi.org/10.1038/s41586-024-07963-3

Dikstein, R., Zhou, S., & Tjian, R. (1996a). Human TAF_{II}105 Is a Cell Type–Specific TFIID Subunit Related to hTAF_{II}130. *Cell*, *87*(1), 137–146. https://doi.org/10.1016/S0092-8674(00)81330-6

Dikstein, R., Zhou, S., & Tjian, R. (1996b). Human TAFII105 Is a Cell Type–Specific TFIID Subunit Related to hTAFII130. *Cell*, *87*(1), 137–146. https://doi.org/https://doi.org/10.1016/S0092-8674(00)81330-6

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Du, G., Oatley, M. J., Law, N. C., Robbins, C., Wu, X., & Oatley, J. M. (2021). Proper timing of a quiescence period in precursor prospermatogonia is required for stem cell pool establishment in the male germline. *Development*, *148*(9). https://doi.org/10.1242/dev.194571

Dura, M., Teissandier, A., Armand, M., Barau, J., Lapoujade, C., Fouchet, P., Bonneville, L., Schulz, M., Weber, M., Baudrin, L. G., Lameiras, S., & Bourc'his, D. (2022). DNMT3A-dependent DNA methylation is required for spermatogonial

stem cells to commit to spermatogenesis. *Nat Genet*, *54*(4), 469–480. https://doi.org/10.1038/s41588-022-01040-z

Eisenbach, M. (2025). Sperm navigation in humans: a concerted action of multiple means. *Communications Biology*, *8*(1), 923. https://doi.org/10.1038/s42003-025-08358-4

Ellis, P. J., Furlong, R. A., Conner, S. J., Kirkman-Brown, J., Afnan, M., Barratt, C., Griffin, D. K., & Affara, N. A. (2007). Coordinated transcriptional regulation patterns associated with infertility phenotypes in men. *J Med Genet*, *44*(8), 498–508. https://doi.org/10.1136/jmg.2007.049650

Ewen, K. A., & Koopman, P. (2010). Mouse germ cell development: from specification to sex determination. *Mol Cell Endocrinol*, *323*(1), 76–93. https://doi.org/10.1016/j.mce.2009.12.013

Falender, A. E., Freiman, R. N., Geles, K. G., Lo, K. C., Hwang, K., Lamb, D. J., Morris, P. L., Tjian, R., & Richards, J. S. (2005). Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. *Genes Dev*, *19*(7), 794–803. https://doi.org/10.1101/gad.1290105

Freiman, R. N., Albright, S. R., Zheng, S., Sha, W. C., Hammer, R. E., & Tjian, R. (2001). Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science*, *293*(5537), 2084–2087. https://doi.org/10.1126/science.1061935

Garbuzov, A., Pech, M. F., Hasegawa, K., Sukhwani, M., Zhang, R. J., Orwig, K. E., & Artandi, S. E. (2018). Purification of GFRα1+ and GFRα1– Spermatogonial Stem Cells Reveals a Niche-Dependent Mechanism for Fate Determination. *Stem Cell*

Reports, *10*(2), 553–567. https://doi.org/https://doi.org/10.1016/j.stemcr.2017.12.009

Gazit, K., Moshonov, S., Elfakess, R., Sharon, M., Mengus, G., Davidson, I., & Dikstein, R. (2009). TAF4/4b·TAF12 Displays a Unique Mode of DNA Binding and Is Required for Core Promoter Function of a Subset of Genes*. *Journal of Biological Chemistry*, *284*(39), 26286–26296. https://doi.org/https://doi.org/10.1074/jbc.M109.011486

Gegonne, A., Tai, X., Zhang, J., Wu, G., Zhu, J., Yoshimoto, A., Hanson, J., Cultraro, C., Chen, Q.-R., Guinter, T., Yang, Z., Hathcock, K., Singer, A., Rodriguez-Canales, J., Tessarollo, L., Mackem, S., Meerzaman, D., Buetow, K., & Singer, D. S. (2012). The General Transcription Factor TAF7 Is Essential for Embryonic Development but Not Essential for the Survival or Differentiation of Mature T Cells. *Molecular and Cellular Biology*, *32*(10), 1984–1997. https://doi.org/10.1128/MCB.06305-11

Gegonne, A., Weissman, J. D., Zhou, M., Brady, J. N., & Singer, D. S. (2006). TAF7: a possible transcription initiation check-point regulator. *Proc Natl Acad Sci U S A*, *103*(3), 602–607. https://doi.org/10.1073/pnas.0510031103

Ginsburg, M., Snow, M. H. L., & McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development*, *110*(2), 521–528. https://doi.org/10.1242/dev.110.2.521

Gkountela, S., Zhang, K. X., Shafiq, T. A., Liao, W. W., Hargan-Calvopiña, J., Chen, P. Y., & Clark, A. T. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell*, *161*(6), 1425–1436. https://doi.org/10.1016/j.cell.2015.05.012

Green, C. D., Ma, Q., Manske, G. L., Shami, A. N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S. J., Moore, B. B., Tallquist, M. D., Li, J. Z., & Hammoud, S. S. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental Cell*, *46*(5), 651–667.e610. https://doi.org/https://doi.org/10.1016/j.devcel.2018.07.025

Greenberg, D. R., Nam, C. S., & Halpern, J. A. (2024). The increasing demand for reproductive urologists and male fertility care. *F S Rep*, *5*(2), 125–126. https://doi.org/10.1016/j.xfre.2024.05.001

Greulich, F., Wierer, M., Mechtidou, A., Gonzalez-Garcia, O., & Uhlenhaut, N. H. (2021). The glucocorticoid receptor recruits the COMPASS complex to regulate inflammatory transcription at macrophage enhancers. *Cell Reports*, *34*(6), 108742. https://doi.org/https://doi.org/10.1016/j.celrep.2021.108742

Gura, M. A., Bartholomew, M. A., Abt, K. M., Relovská, S., Seymour, K. A., & Freiman, R. N. (2023). Transcription and chromatin regulation by TAF4b during cellular quiescence of developing prospermatogonia [Original Research]. *Frontiers in Cell and Developmental Biology*, *11*. https://doi.org/10.3389/fcell.2023.1270408

Gura, M. A., Mikedis, M. M., Seymour, K. A., de Rooij, D. G., Page, D. C., & Freiman, R. N. (2020). Dynamic and regulated TAF gene expression during mouse embryonic germ cell development. *PLOS Genetics*, *16*(1), e1008515. https://doi.org/10.1371/journal.pgen.1008515

Gustafson, E. A., Seymour, K. A., Sigrist, K., Rooij, D. G. D. E., & Freiman, R. N. (2020). ZFP628 Is a TAF4b-Interacting Transcription Factor Required for Mouse

Spermiogenesis. *Molecular and Cellular Biology*, *40*(7), e00228–00219. https://doi.org/10.1128/MCB.00228-19

Hargy, J., & Sasaki, K. (2023). The developmental dynamics of the human male germline. *Development*, *150*(17). https://doi.org/10.1242/dev.202046

Hermann, B. P., Mutoji, K. N., Velte, E. K., Ko, D., Oatley, J. M., Geyer, C. B., & McCarrey, J. R. (2015). Transcriptional and translational heterogeneity among neonatal mouse spermatogonia. *Biol Reprod*, *92*(2), 54. https://doi.org/10.1095/biolreprod.114.125757

Hernández-Hernández, A., Lilienthal, I., Fukuda, N., Galjart, N., & Höög, C. (2016). CTCF contributes in a critical way to spermatogenesis and male fertility. *Scientific Reports*, *6*(1), 28355. https://doi.org/10.1038/srep28355

Hill, P. W. S., Leitch, H. G., Requena, C. E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., Bagci, H., Dharmalingham, G., Haberle, V., Lenhard, B., Zheng, Y., Pradhan, S., & Hajkova, P. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature*, *555*(7696), 392–396. https://doi.org/10.1038/nature25964

Huang, G., Liu, L., Wang, H., Gou, M., Gong, P., Tian, C., Deng, W., Yang, J., Zhou, T. T., Xu, G. L., & Liu, L. (2020). Tet1 Deficiency Leads to Premature Reproductive Aging by Reducing Spermatogonia Stem Cells and Germ Cell Differentiation. *iScience*, *23*(3), 100908. https://doi.org/10.1016/j.isci.2020.100908

Huang, Z., Li, N., Ji, X., Zhou, D., Wang, R., Zhao, X., Wang, S., Zhang, H., Huang, C., & Lin, G. (2024a). EEF1B2 regulates the proliferation and apoptosis of human

spermatogonial stem cell lines through TAF4B. *Heliyon*, *10*(17). https://doi.org/10.1016/j.heliyon.2024.e36467

Huang, Z., Li, N., Ji, X., Zhou, D., Wang, R., Zhao, X., Wang, S., Zhang, H., Huang, C., & Lin, G. (2024b). EEF1B2 regulates the proliferation and apoptosis of human spermatogonial stem cell lines through TAF4B. *Heliyon*, *10*(17), e36467. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e36467

Hwang, Y. S., Suzuki, S., Seita, Y., Ito, J., Sakata, Y., Aso, H., Sato, K., Hermann, B. P., & Sasaki, K. (2020). Reconstitution of prospermatogonial specification in vitro from human induced pluripotent stem cells. *Nature Communications*, *11*(1), 5656. https://doi.org/10.1038/s41467-020-19350-3

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, *3*, 318–356. https://doi.org/10.1016/s0022-2836(61)80072-7

Jaszczak, R. G., Zussman, J. W., Wagner, D. E., & Laird, D. J. (2025). Comprehensive profiling of migratory primordial germ cells reveals niche-specific differences in non-canonical Wnt and Nodal-Lefty signaling in anterior vs posterior migrants. *Elife*, *14*, RP103074. https://doi.org/10.7554/eLife.103074

Jiang, Y., Zhang, H. Y., Lin, Z., Zhu, Y. Z., Yu, C., Sha, Q. Q., Tong, M. H., Shen, L., & Fan, H. Y. (2020). CXXC finger protein 1-mediated histone H3 lysine-4 trimethylation is essential for proper meiotic crossover formation in mice. *Development*, *147*(6). https://doi.org/10.1242/dev.183764

Jones, S. L., Styles, J., & Geyer, C. B. (2021). A new translation and reader's guide to the first detailed description of the first wave of spermatogenesis in the mouse. *Mol Reprod Dev*, *88*(7), 473–478. https://doi.org/10.1002/mrd.23519

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., & Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *Embo j*, *32*(3), 340–353. https://doi.org/10.1038/emboj.2012.331

Kanatsu-Shinohara, M., Takehashi, M., Takashima, S., Lee, J., Morimoto, H., Chuma, S., Raducanu, A., Nakatsuji, N., Fässler, R., & Shinohara, T. (2008). Homing of mouse spermatogonial stem cells to germline niche depends on beta1-integrin. *Cell stem cell*, *3*(5), 533–542. https://doi.org/10.1016/j.stem.2008.08.002

Kelsey, M. M. G., Kalekar, R. L., & Sedivy, J. M. (2025). TE-Seq: a transposable element annotation and RNA-Seq pipeline. *Mob DNA*, *16*(1), 44. https://doi.org/10.1186/s13100-025-00381-w

Kluin, P. M., & de Rooij, D. G. (1981). A comparison between the morphology and cell kinetics of gonocytes and adult type undifferentiated spermatogonia in the mouse. *Int J Androl*, *4*(4), 475–493. https://doi.org/10.1111/j.1365-2605.1981.tb00732.x

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Krzastek, S. C., Farhi, J., Gray, M., & Smith, R. P. (2020). Impact of environmental toxin exposure on male fertility potential. *Transl Androl Urol*, *9*(6), 2797–2813. https://doi.org/10.21037/tau-20-685

Kubota, H., & Brinster, R. L. (2018). Spermatogonial stem cells†. *Biology of Reproduction*, *99*(1), 52–74. https://doi.org/10.1093/biolre/ioy077

Langer, D., Martianov, I., Alpern, D., Rhinn, M., Keime, C., Dollé, P., Mengus, G., & Davidson, I. (2016). Essential role of the TFIID subunit TAF4 in murine embryogenesis and embryonic stem cell differentiation. *Nature Communications*, *7*(1), 11063. https://doi.org/10.1038/ncomms11063

Law, N. C., Oatley, M. J., & Oatley, J. M. (2019). Developmental kinetics and transcriptome dynamics of stem cell specification in the spermatogenic lineage. *Nature Communications*, *10*(1), 2787. https://doi.org/10.1038/s41467-019-10596-0

Le, S. N., Brown, C. R., Harvey, S., Boeger, H., Elmlund, H., & Elmlund, D. (2019). The TAFs of TFIID Bind and Rearrange the Topology of the TATA-Less RPS5 Promoter. *Int J Mol Sci*, *20*(13). https://doi.org/10.3390/ijms20133290

Lee, H. J., Hore, T. A., & Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. *Cell stem cell*, *14*(6), 710–719. https://doi.org/10.1016/j.stem.2014.05.008

Levine, H., Jørgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Jolles, M., Pinotti, R., & Swan, S. H. (2023). Temporal trends in sperm count: a systematic review and meta-regression analysis of samples collected globally in the 20th and 21st centuries. *Hum Reprod Update*, *29*(2), 157–176. https://doi.org/10.1093/humupd/dmac035

Levine, H., Jorgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Mindlis, I., Pinotti, R., & Swan, S. H. (2017). Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum Reprod Update*, *23*(6), 646–659. https://doi.org/10.1093/humupd/dmx022

Li, B.-Z., Huang, Z., Cui, Q.-Y., Song, X.-H., Du, L., Jeltsch, A., Chen, P., Li, G., Li, E., & Xu, G.-L. (2011). Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. *Cell Research*, *21*(8), 1172–1181. https://doi.org/10.1038/cr.2011.92

Liao, H.-F., Chen, W. S. C., Chen, Y.-H., Kao, T.-H., Tseng, Y.-T., Lee, C.-Y., Chiu, Y.-C., Lee, P.-L., Lin, Q.-J., Ching, Y.-H., Hata, K., Cheng, W. T. K., Tsai, M.-H., Sasaki, H., Ho, H.-N., Wu, S.-C., Huang, Y.-H., Yen, P., & Lin, S.-P. (2014). DNMT3L promotes quiescence in postnatal spermatogonial progenitor cells. *Development*, *141*(12), 2402–2413. https://doi.org/10.1242/dev.105130

Liao, Y., Smyth, G. K., & Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu, S., Tang, Z., Xiong, T., & Tang, W. (2011). Isolation and characterization of human spermatogonial stem cells. *Reprod Biol Endocrinol*, *9*, 141. https://doi.org/10.1186/1477-7827-9-141

Liu, W. L., Coleman, R. A., Grob, P., King, D. S., Florens, L., Washburn, M. P., Geles, K. G., Yang, J. L., Ramey, V., Nogales, E., & Tjian, R. (2008). Structural changes in TAF4b-TFIID correlate with promoter selectivity. *Mol Cell*, *29*(1), 81–91. https://doi.org/10.1016/j.molcel.2007.11.003

Lovasco, L. A., Gustafson, E. A., Seymour, K. A., de Rooij, D. G., & Freiman, R. N. (2015). TAF4b is Required for Mouse Spermatogonial Stem Cell Development. *STEM CELLS*, *33*(4), 1267–1276. https://doi.org/https://doi.org/10.1002/stem.1914

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Luna-Arias, J. P., & Castro-Muñozledo, F. (2024). Participation of the TBP-associated factors (TAFs) in cell differentiation. *J Cell Physiol*, *239*(2), e31167. https://doi.org/10.1002/jcp.31167

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, *19*(2), 81–92. https://doi.org/10.1038/nrg.2017.80

McCarrey, J. R. (2013). Toward a more precise and informative nomenclature describing fetal and neonatal male germ cells in rodents. *Biol Reprod*, *89*(2), 47. https://doi.org/10.1095/biolreprod.113.110502

Metcalf, C. E., & Wassarman, D. A. (2007). Nucleolar colocalization of TAF1 and testis-specific TAFs during Drosophila spermatogenesis. *Developmental Dynamics*, *236*(10), 2836–2843. https://doi.org/https://doi.org/10.1002/dvdy.21294

Mizzen, C. A., Yang, X.-J., Kokubo, T., Brownell, J. E., Bannister, A. J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S. L., Kouzarides, T., Nakatani, Y., & Allis, C. D. (1996). The TAF_{II}250 Subunit of TFIID Has Histone Acetyltransferase Activity. *Cell*, *87*(7), 1261–1270. https://doi.org/10.1016/S0092-8674(00)81821-8

Molaro, A., Falciatori, I., Hodges, E., Aravin, A. A., Marran, K., Rafii, S., McCombie, W. R., Smith, A. D., & Hannon, G. J. (2014). Two waves of de novo methylation during mouse germ cell development. *Genes Dev*, *28*(14), 1544–1549. https://doi.org/10.1101/gad.244350.114

Muñoz, X., Navarro, M., Mata, A., Bassas, L., & Larriba, S. (2014). Association of PIWIL4 genetic variants with germ cell maturation arrest in infertile Spanish men. *Asian J Androl*, *16*(6), 931–933. https://doi.org/10.4103/1008-682x.131069

Nagamori, I., Kobayashi, H., Nishimura, T., Yamagishi, R., Katahira, J., Kuramochi-Miyagawa, S., Kono, T., & Nakano, T. (2018). Relationship between PIWIL4-Mediated H3K4me2 Demethylation and piRNA-Dependent DNA Methylation. *Cell Reports*, *25*(2), 350–356. https://doi.org/10.1016/j.celrep.2018.09.017

Nagano, M. C., & Yeh, J. R. (2013). Chapter Three - The Identity and Fate Decision Control of Spermatogonial Stem Cells: Where Is the Point of No Return? In P. M. Wassarman (Ed.), *Current Topics in Developmental Biology* (Vol. 102, pp. 61–95). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-416024-8.00003-9

Nguyen, D. H., Jaszczak, R. G., & Laird, D. J. (2019). Heterogeneity of primordial germ cells. *Curr Top Dev Biol*, *135*, 155–201. https://doi.org/10.1016/bs.ctdb.2019.04.009

Nguyen, D. H., Soygur, B., Peng, S. P., Malki, S., Hu, G., & Laird, D. J. (2020). Apoptosis in the fetal testis eliminates developmentally defective germ cell clones. *Nat Cell Biol*, *22*(12), 1423–1435. https://doi.org/10.1038/s41556-020-00603-8

Nie, X., Munyoki, S. K., Sukhwani, M., Schmid, N., Missel, A., Emery, B. R., DonorConnect, Stukenborg, J. B., Mayerhofer, A., Orwig, K. E., Aston, K. I., Hotaling, J. M., Cairns, B. R., & Guo, J. (2022). Single-cell analysis of human testis aging and correlation with elevated body mass index. *Dev Cell*, *57*(9), 1160–1176.e1165. https://doi.org/10.1016/j.devcel.2022.04.004

Nikolov, D. B., & Burley, S. K. (1997). RNA polymerase II transcription initiation: A structural view. *Proceedings of the National Academy of Sciences*, *94*(1), 15–22. https://doi.org/doi:10.1073/pnas.94.1.15

Ogawa, T., Arechaga, J. M., Avarbock, M. R., & Brinster, R. L. (1997). Transplantation of testis germinal cells into mouse seminiferous tubules. *The International Journal of Developmental Biology*, *41*(1), 111–122.

Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, *448*(7154), 714–717. https://doi.org/10.1038/nature05987

Palacios, S., Bruno, S., Weiss, R., Salibi, E., Goodchild-Michelman, I., Kane, A., Ilia, K., & Del Vecchio, D. (2025). Analog epigenetic memory revealed by targeted chromatin editing. *Cell Genomics*, *5*(11). https://doi.org/10.1016/j.xgen.2025.100985

Park, C., Choi, Y., Yoo, S., La, H., & Hong, K. (2025). Analysis of DNA methylation changes following Cfp1 knockout in mouse spermatocytes. *Anim Biosci*, *38*(8), 1570–1579. https://doi.org/10.5713/ab.24.0807

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650–1667. https://doi.org/10.1038/nprot.2016.095

Podhorecka, M., Skladanowski, A., & Bozko, P. (2010). H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy. *J Nucleic Acids*, *2010*. https://doi.org/10.4061/2010/920161

Ponomaryova, A. A., Rykova, E. Y., Gervas, P. A., Cherdyntseva, N. V., Mamedov, I. Z., & Azhikina, T. L. (2020). Aberrant Methylation of LINE-1 Transposable Elements: A Search for Cancer Biomarkers. *Cells*, *9*(9). https://doi.org/10.3390/cells9092017

Prabhu, K. S., Kuttikrishnan, S., Ahmad, N., Habeeba, U., Mariyam, Z., Suleman, M., Bhat, A. A., & Uddin, S. (2024). H2AX: A key player in DNA damage response and a promising target for cancer therapy. *Biomedicine & Pharmacotherapy*, *175*, 116663. https://doi.org/https://doi.org/10.1016/j.biopha.2024.116663

Pugh, B. F., & Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev*, *5*(11), 1935–1945. https://doi.org/10.1101/gad.5.11.1935

Ramirez, J. M., Bai, Q., Péquignot, M., Becker, F., Kassambara, A., Bouin, A., Kalatzis, V., Dijon-Grinand, M., & De Vos, J. (2013). Side scatter intensity is highly heterogeneous in undifferentiated pluripotent stem cells and predicts clonogenic self-renewal. *Stem Cells Dev*, *22*(12), 1851–1860. https://doi.org/10.1089/scd.2012.0658

Ravitsky, V., & Kimmins, S. (2019). The forgotten men: rising rates of male infertility urgently require new approaches for its prevention, diagnosis and treatment. *Biol Reprod*, *101*(5), 872–874. https://doi.org/10.1093/biolre/ioz161

Reik, W., & Dean, W. (2001). DNA methylation and mammalian epigenetics. *Electrophoresis*, *22*(14), 2838–2843. https://doi.org/10.1002/1522-2683(200108)22:14<2838::Aid-elps2838>3.0.Co;2-m

Resnick, J. L., Bixler, L. S., Cheng, L., & Donovan, P. J. (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature*, *359*(6395), 550–551. https://doi.org/10.1038/359550a0

Roeder, R. G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat Struct Mol Biol*, *26*(9), 783–791. https://doi.org/10.1038/s41594-019-0287-x

Saitou, M., & Yamaji, M. (2012). Primordial germ cells in mice. *Cold Spring Harb Perspect Biol*, *4*(11). https://doi.org/10.1101/cshperspect.a008375

Sasaki, K., & Sangrithi, M. (2023). Developmental origins of mammalian spermatogonial stem cells: New perspectives on epigenetic regulation and sex chromosome function. *Molecular and Cellular Endocrinology*, *573*, 111949. https://doi.org/https://doi.org/10.1016/j.mce.2023.111949

Sermondade, N., Faure, C., Fezeu, L., Shayeb, A. G., Bonde, J. P., Jensen, T. K., Van Wely, M., Cao, J., Martini, A. C., Eskandar, M., Chavarro, J. E., Koloszar, S., Twigt, J. M., Ramlau-Hansen, C. H., Borges, E., Jr., Lotti, F., Steegers-Theunissen, R. P., Zorn, B., Polotsky, A. J.,…Czernichow, S. (2013). BMI in relation to sperm count: an updated systematic review and collaborative meta-

analysis. *Hum Reprod Update*, *19*(3), 221–231. https://doi.org/10.1093/humupd/dms050

Sha, Q. Q., Zhu, Y. Z., Xiang, Y., Yu, J. L., Fan, X. Y., Li, Y. C., Wu, Y. W., Shen, L., & Fan, H. Y. (2021). Role of CxxC-finger protein 1 in establishing mouse oocyte epigenetic landscapes. *Nucleic Acids Res*, *49*(5), 2569–2582. https://doi.org/10.1093/nar/gkab107

Shi, G., & Jin, Y. (2010). Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res Ther*, *1*(5), 39. https://doi.org/10.1186/scrt39

Shinohara, T., Orwig, K. E., Avarbock, M. R., & Brinster, R. L. (2000). Spermatogonial stem cell enrichment by multiparameter selection of mouse testis cells. *Proceedings of the National Academy of Sciences*, *97*(15), 8346–8351. https://doi.org/doi:10.1073/pnas.97.15.8346

Smallwood, S. A., & Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. *Trends in Genetics*, *28*(1), 33–42. https://doi.org/10.1016/j.tig.2011.09.004

Song, Y., Zhang, X., Desmarais, J. A., & Nagano, M. (2024). Postnatal development of mouse spermatogonial stem cells as determined by immunophenotype, regenerative capacity, and long-term culture-initiating ability: a model for practical applications. *Scientific Reports*, *14*(1), 2299. https://doi.org/10.1038/s41598-024-52824-8

Stephens, K. E., Miaskowski, C. A., Levine, J. D., Pullinger, C. R., & Aouizerat, B. E. (2013). Epigenetic regulation and measurement of epigenetic changes. *Biol Res Nurs*, *15*(4), 373–381. https://doi.org/10.1177/1099800412444785

Stewart, K. R., Veselovska, L., & Kelsey, G. (2016). Establishment and functions of DNA methylation in the germline. *Epigenomics*, *8*(10), 1399–1413. https://doi.org/10.2217/epi-2016-0056

Sun, F., Sun, T., Kronenberg, M., Tan, X., Huang, C., & Carey, M. F. (2021). The Pol II preinitiation complex (PIC) influences Mediator binding but not promoter-enhancer looping. *Genes Dev*, *35*(15-16), 1175–1189. https://doi.org/10.1101/gad.348471.121

Suzuki, S., Kato, H., Suzuki, Y., Chikashige, Y., Hiraoka, Y., Kimura, H., Nagao, K., Obuse, C., Takahata, S., & Murakami, Y. (2016). Histone H3K36 trimethylation is essential for multiple silencing mechanisms in fission yeast. *Nucleic Acids Res*, *44*(9), 4147–4162. https://doi.org/10.1093/nar/gkw008

Tan, D. S., Cheung, S. L., Gao, Y., Weinbuch, M., Hu, H., Shi, L., Ti, S. C., Hutchins, A. P., Cojocaru, V., & Jauch, R. (2023). The homeodomain of Oct4 is a dimeric binder of methylated CpG elements. *Nucleic Acids Res*, *51*(3), 1120–1138. https://doi.org/10.1093/nar/gkac1262

Tan, K., Song, H. W., & Wilkinson, M. F. (2020). Single-cell RNAseq analysis of testicular germ and somatic cell development during the perinatal period. *Development*, *147*(3). https://doi.org/10.1242/dev.183251

Tan, L., & Shi, Y. G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*, *139*(11), 1895–1902. https://doi.org/10.1242/dev.070771

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R., Deaton, A., Andrews, R., James, K. D., Turner, D. J., Illingworth, R., & Bird, A.

(2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, *464*(7291), 1082–1086. https://doi.org/10.1038/nature08924

Trowitzsch, S., Viola, C., Scheer, E., Conic, S., Chavant, V., Fournier, M., Papai, G., Ebong, I. O., Schaffitzel, C., Zou, J., Haffke, M., Rappsilber, J., Robinson, C. V., Schultz, P., Tora, L., & Berger, I. (2015). Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat Commun*, *6*, 6011. https://doi.org/10.1038/ncomms7011

Valli, H., Sukhwani, M., Dovey, S. L., Peters, K. A., Donohue, J., Castro, C. A., Chu, T., Marshall, G. R., & Orwig, K. E. (2014). Fluorescence- and magnetic-activated cell sorting strategies to isolate and enrich human spermatogonial stem cells. *Fertility and Sterility*, *102*(2), 566–580.e567. https://doi.org/https://doi.org/10.1016/j.fertnstert.2014.04.036

Wang, X., Jia, Y., Fei, C., Song, X., & Li, L. (2016). Activation/Proliferation-associated Protein 2 (Caprin-2) Positively Regulates CDK14/Cyclin Y-mediated Lipoprotein Receptor-related Protein 5 and 6 (LRP5/6) Constitutive Phosphorylation*. *Journal of Biological Chemistry*, *291*(51), 26427–26434. https://doi.org/https://doi.org/10.1074/jbc.M116.744607

Wang, Y., Cheng, Y., Yang, Q., Kuang, L., & Liu, G. (2022). Overexpression of FOXD2-AS1 enhances proliferation and impairs differentiation of glioma stem cells by activating the NOTCH pathway via TAF-1. *Journal of Cellular and Molecular Medicine*, *26*(9), 2620–2632. https://doi.org/https://doi.org/10.1111/jcmm.17268

Wassarman, D. A., & Sauer, F. (2001). TAFII250: a transcription toolbox. *Journal of Cell Science*, *114*(16), 2895–2902. https://doi.org/10.1242/jcs.114.16.2895

Webster, K. E., O'Bryan, M. K., Fletcher, S., Crewther, P. E., Aapola, U., Craig, J., Harrison, D. K., Aung, H., Phutikanit, N., Lyle, R., Meachem, S. J., Antonarakis, S. E., de Kretser, D. M., Hedger, M. P., Peterson, P., Carroll, B. J., & Scott, H. S. (2005). Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A*, *102*(11), 4068–4073. https://doi.org/10.1073/pnas.0500702102

Wei, Z., Wang, S., Xu, Y., Wang, W., Soares, F., Ahmed, M., Su, P., Wang, T., Orouji, E., Xu, X., Zeng, Y., Chen, S., Liu, X., Jia, T., Liu, Z., Du, L., Wang, Y., Chen, S., Wang, C.,…Guo, H. (2023). MYC reshapes CTCF-mediated chromatin architecture in prostate cancer. *Nature Communications*, *14*(1), 1787. https://doi.org/10.1038/s41467-023-37544-3

Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., Horth, C., McGuire, J. T., Xu, X., Nikbakht, H., Lemiesz, A. E., Marchione, D. M., Marunde, M. R., Meiners, M. J., Cheek, M. A., Keogh, M. C., Bareke, E., Djedid, A., Harutyunyan, A. S.,…Lu, C. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*, *573*(7773), 281–286. https://doi.org/10.1038/s41586-019-1534-3

Werten, S., Mitschler, A., Romier, C., Gangloff, Y. G., Thuault, S., Davidson, I., & Moras, D. (2002). Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J Biol Chem*, *277*(47), 45502–45509. https://doi.org/10.1074/jbc.M206587200

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wright, K. J., Marr, M. T., 2nd, & Tjian, R. (2006). TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci U S A*, *103*(33), 12347–12352. https://doi.org/10.1073/pnas.0605499103

Xi, Q., Zhang, H., Zhang, X., Jiang, Y., Wang, R., Liu, R., & Zhang, H. (2020). Analysis of TATA-box binding protein associated factor 4b gene mutations in a Chinese population with nonobstructive azoospermia. *Medicine (Baltimore)*, *99*(23), e20561. https://doi.org/10.1097/MD.0000000000020561

Yang, K., Wang, X., Zhang, H., Wang, Z., Nan, G., Li, Y., Zhang, F., Mohammed, M. K., Haydon, R. C., Luu, H. H., Bi, Y., & He, T.-C. (2016). The evolving roles of canonical WNT signaling in stem cells and tumorigenesis: implications in targeted cancer therapies. *Laboratory Investigation*, *96*(2), 116–136. https://doi.org/10.1038/labinvest.2015.144

Yatsenko, S. A., & Rajkovic, A. (2019). Genetics of human female infertility†. *Biol Reprod*, *101*(3), 549–566. https://doi.org/10.1093/biolre/ioz084

Yu, C., Fan, X., Sha, Q.-Q., Wang, H.-H., Li, B.-T., Dai, X.-X., Shen, L., Liu, J., Wang, L., Liu, K., Tang, F., & Fan, H.-Y. (2017). CFP1 Regulates Histone H3K4 Trimethylation and Developmental Potential in Mouse Oocytes. *Cell Reports*, *20*(5), 1161–1172. https://doi.org/10.1016/j.celrep.2017.07.011

Zhao, T., Hong, Y., Yan, B., Huang, S., Ming, G.-l., & Song, H. (2024). Epigenetic maintenance of adult neural stem cell quiescence in the mouse hippocampus via

Setd1a. *Nature Communications*, *15*(1), 5674. https://doi.org/10.1038/s41467-024-50010-y

Zhou, H., Grubisic, I., Zheng, K., He, Y., Wang, P. J., Kaplan, T., & Tjian, R. (2013). Taf7l cooperates with Trf2 to regulate spermiogenesis. *Proc Natl Acad Sci U S A*, *110*(42), 16886–16891. https://doi.org/10.1073/pnas.1317034110