THE UNIVERSITY *of* EDINBURGH

Understanding the role and function of SPOCD1 in piRNA-directed *de novo* DNA methylation

Madeleine Dias Mirandela

A thesis submitted for the degree of

Doctor of Philosophy

The University of Edinburgh
2024

HOPE OVER FEAR.

UNITY OVER DIVISION.

SCIENCE OVER FICTION.

TRUTH OVER LIES.

FRANCES H. ARNOLD

Lay summary

The key for survival of every species is to transmit its genetic material, the DNA, to the next generation. This DNA must be of highest quality as damage can leave an individual sterile. Unfortunately, in mammals, nearly half of the DNA consists of “jumping genes” or transposons, which, as the name suggest, have the ability to move around and cause DNA damage. If these transposons are not properly controlled, it can lead to male infertility. Therefore, a defence mechanism safeguarding the germline against transposons exists which is known as the piRNA pathway. This system puts a unique marker called DNA methylation on the transposon genes, which helps keeping them in check. The pathway relies on the small machines, or proteins, specifically MIWI2 and SPOCD1, which are necessary for placing DNA methylation at these sited. While it is known how MIWI2 finds transposons in our DNA, little is known about SPOCD1. In this study, I investigated how SPOCD1 interacts with three other proteins: SPIN1, C19ORF84, and TPR. I discovered that each of the proteins sticks to a defined part of SPOCD1. Additionally, I examined a specific part of SPOCD1 called the SPOC domain. While I was able to get a very detailed picture of its structure, its exact job is not known yet. We then focused on understanding how SPOCD1 and SPIN1 work together and demonstrated that they communicate directly. We observed that some transposons bear a specific histone signature, which SPIN1 recognizes. This could indicate how SPOCD1 locates transposons within our DNA. To understand the importance of the SPOCD1-SPIN1 interaction, we created a mouse with a version of SPOCD1 that is unable to communicate with SPIN1. These male mice are infertile and do not produce any mature sperm because some transposons are not controlled. In summary, we not only define how SPOCD1 teams up with SPIN1, C19ORF84, and TPR, but also sheds light on the mysterious SPOC domain. We show that SPOCD1 and SPIN1 need to communicate to control transposons. Moreover, we discovered that transposons bear a specific signature to be recognized by the SPIN1-SPOCD1 complex. Our findings challenge the current understanding of the process and suggests a novel mechanism that ensures the control of transposons.

Abstract

The survival of the germline is essential for the survival of a species. Yet, the integrity of those cells is threatened by transposons during foetal development. In fact, failure to silence transposons results in male infertility in mammals. The PIWI-interacting RNA (piRNA) pathway is the primary germline defence system which places CpG DNA methylation at transposon loci. piRNAs recruit the PIWI protein MIWI2 to nascent transposon transcripts *via* complementary base pairing and instruct DNA methylation through SPOCD1. In this work, I defined the interaction of SPOCD1 with three downstream factors, SPIN1, C19ORF84 and TPR. C19ORF84 interacts with a c-terminal α-helix of SPOCD1 and TPR interacts with the SPOCD1 TFIIS-M domain. In addition, I solved the structure of the SPOCD1 SPOC domain, whose function remains to be determined, at 1.7 Å. I then focused on the SPOCD1-SPIN1 interaction. I showed that SPOCD1 directly interacts with SPIN1, a chromatin reader that recognises H3K4me3 and H3K9me3. I found that SPIN1 and SPOCD1 are interacting before MIWI2 is expressed which challenges the current view of all molecular events required for piRNA-directed DNA methylation occurring after the engagement of MIWI2. In addition, I saw that young LINE1 transposon copies are marked by the SPIN1-associated chromatin marks before the initiation of piRNA-directed DNA methylation. I then generated a *Spocd1* separation-of-function allele in the mouse encoding a SPOCD1 variant that can no longer interact with SPIN1. I showed that the SPOCD1-SPIN1 interaction is in fact essential for spermatogenesis and piRNA-directed DNA methylation of young LINE1 transposons. In summary, this work defines the interaction of SPOCD1 with SPIN1, C19ORF84 and TPR, gives an insight into the SPOCD1 SPOC domain structure and defines the importance of the SPOCD1-SPIN1 interaction which challenges the current model and proposes a novel two-factor authentication system to form the basis of precision.

Acknowledgements

First and foremost, I want to thank my supervisor Prof. Dónal O'Carroll for giving me the opportunity to work on this very exciting project in his lab. Thank you very much for all your support throughout the different challenges of my PhD.

I also want to thank my second supervisor, Prof. Atlanta G. Cook for the time she invested in teaching and helping me. It was an honour to learn from you. Thank you for your patience and (almost) making me a real biochemist.

I would also like to express my gratitude to my additional committee members, Prof. Tilo Kunath and Marcus Wilson. Thank you for making every meeting enjoyable, challenging me, and providing very valuable scientific insights.

Of course, I want to thank all the collaborators that are mentioned throughout the chapters, especially Tania Auchynnikava, Rebecca V. Berrens, Shaun Webb and Juan Zou who provided key data for my project. Thank you for sharing your expertise and bringing the science to the next level. I want to give a special thanks also to the people in the facilities in IRR and the Wellcome Centre, specifically the FACS facility in IRR and the MS facility and EPPF in the Wellcome Centre. Thank you for always helping.

I also want to thank all the present and former members of the O'Carroll lab. You made every bad experiment feel less bad and every good experiment even better. Thank you for making going to the lab fun. Ansgar, Azzurra, David, Gabriela, Margot, Martina, Ola, Shaunak, Tamoghna, Theresa, Xinyu and Yuka - thank you for creating a great working environment and always being supportive in- and outside of work. A special thanks to Ansgar, who had the honour/ burden of teaching me most techniques in the beginning.

I also want to thank all the friends I made in Edinburgh. Especially Akanksha, for being on the PhD journey with me and listening to my complains over countless

walks and coffees. Thanks to all my friends back in Germany and around the world for checking in and supporting me along the way.

A special thank you goes to my family, specifically my parents for always supporting me. Finally, I want to thank my husband Gaëtan for his support. Thank you for always putting a smile on my face no matter how bad my day was.

I am incredibly grateful for my time in Edinburgh, and it would have not been the same without any of you.

Declaration

I declare that the thesis has been composed entirely by myself and that the work has not be submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where stated otherwise. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

Madeleine Dias Mirandela

Table of Content

List of Figures

List of Tables

1 Introduction

1.1 Germ cell development

The only way for a species to survive is to pass its genetic information, the Deoxyribonucleic acid (DNA) to the next generation. Therefore, the transmission of high-quality and errorless genetic material is crucial. Consequently, the Germ cells (GCs) that give rise to gametes and transfer genetic material to the next generation must be of the highest quality. The development of primordial germ cells (PGCs) into spermatozoa in males or oocytes in females involves a carefully orchestrated program that includes genetic, epigenetic, and morphogenetic changes. Finally, these cells must perform the challenging task of fertilisation because only successful fertilisation will ensure the survival of the species. Therefore, complex quality control mechanisms are required to protect germ line integrity. How the complex development of the male germ line is orchestrated will be discussed in detail in the next Chapters.

1.1.1 Germ cell specification, migration, and colonization

During embryonic development, PGCs are derived from somatic cells (Lawson and Hage 1994). Around Embryonic day 6.5 (E6.5), several primed cells in the mouse embryo become PGCs (Lawson and Hage 1994; Ohinata et al. 2005). Signalling molecules from the extraembryonic ectoderm initiate reprogramming of cells to produce the PGC lineage (Ewen and Koopman 2010). Simultaneously, the somatic development program is repressed in PGC precursors (Ewen and Koopman 2010). One key factor is the transcription repressor B-lymphocyte-induced maturation protein 1 (BLIMP1), which is expressed from E6.25 and is an important factor in somatic transcriptional program repression (Ohinata et al. 2005; Kurimoto et al. 2008; Hayashi, de Sousa Lopes, and Surani 2007). At the same time, PGCs

continue to express pluripotency markers to maintain their pluripotency potential (Ewen and Koopman 2010).

The migration of PGCs is the next crucial step which is guided by signals from the surrounding somatic cells (Wylie 1999). PGCs first migrate from the posterior primitive streak at E7.5 into the subjacent endoderm, which develops into the hindgut (Anderson et al. 2000; Clark and Eddy 1975). Then, PGCs are moved passively into the embryo at E7.5 – 8.5 while the hindgut invaginates (Clark and Eddy 1975; Lawson and Hage 1994). From E8.5, PGCs actively migrate from the hindgut epithelium through the midline into two separate clusters, colonising the nascent genital ridges by E12.5 (Molyneaux et al. 2001; Godin, Wylie, and Heasman 1990) (Figure 1). With colonisation of the genital ridges, PGCs (now called germ cells or gonocytes) integrate into the developing gonads and initiate changes in morphology and gene expression, including the loss of their pluripotent potential (Ewen and Koopman 2010) (Figure 1).

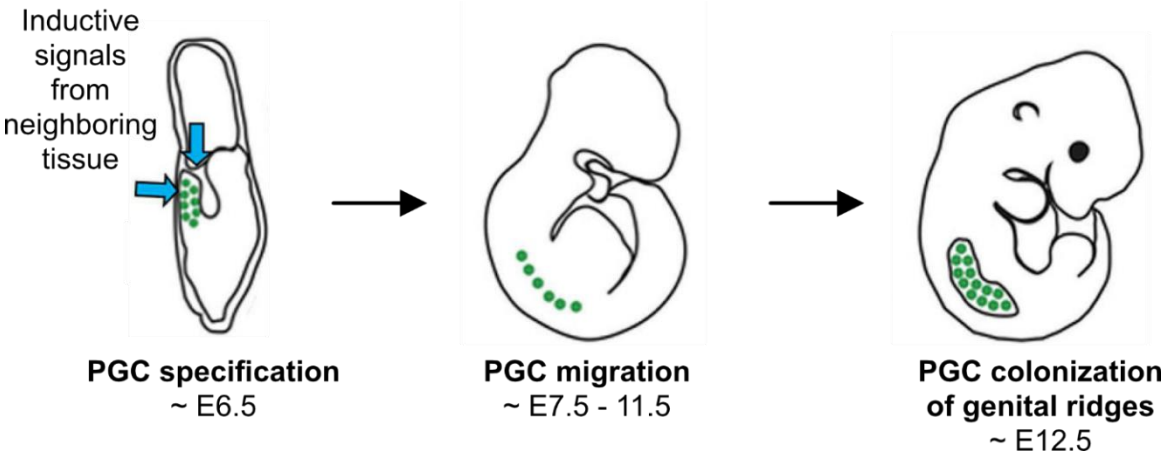

Figure 1: Germ cell development.

Illustration showing the key events of germ cell development: specification, migration, and colonization in the mouse embryo. Primordial germ cells (PGC) shown in green. Figure adapted from (Hansen and Pelegri 2021).

Until E12.5, germ cells retain their bipotential capacity, and male and female germ cells are indistinguishable (Nakatsuji and Chuma 2001; Adams and McLaren 2002; McLaren and Southee 1997). Germ cell differentiation starts around E12.5 and is controlled by the somatic environment and not by the cells autonomously (Ewen and Koopman 2010). The signature of differentiation is the entrance into meiosis

in female germ cells and mitotic arrest in male germ cells (McLaren and Southee 1997; Adams and McLaren 2002; Hilscher et al. 1974). Male germ cells arrest and exit the cell cycle at the G1/G0 transition as prospermatogonia in an unsynchronized manner from E12.5 – 14.5 which will be resumed only after birth (Hilscher et al. 1974; McLaren and Southee 1997; Western et al. 2008). Since my study focuses on the male germline, I will focus only on the male germline in the following Chapters.

1.1.2 Epigenetic changes in the male developing germline

In the development of the zygote to adulthood, the cellular potential is progressively restricted to give rise to all types of differentiated somatic cells (Hackett, Zylicz, and Surani 2012). The unique exception is the germline which is derived from somatic cells (Hackett, Zylicz, and Surani 2012). Therefore, epigenetic barriers have to be reversed to allow the development of the germ cell (Hackett, Zylicz, and Surani 2012). Since the number of early PGCs is highly restricted to approximately 40 cells by E7.5, this process must be very robust and very efficient (McLaren and Lawson 2005). Furthermore, it must be completed with mitotic arrest in male germ cells and meiotic arrest in female germ cells around E13.5 (Hackett, Zylicz, and Surani 2012). Therefore, it is likely that a combination of multiple mechanisms, including active and passive events ensure the successful completion of this process within the short time window of E7.5 – E13.5 (Hackett, Zylicz, and Surani 2012).

To achieve reprogramming, PGCs must repress the ongoing somatic programme and activate the germ cell specific programme (Saitou, Barton, and Surani 2002; Yabuta et al. 2006). This reprogramming in PGCs includes chromatin remodelling, extensive changes in DNA methylation and the erasure of genomic imprints representing the most comprehensive epigenetic resetting in the mammalian life cycle (Surani, Hayashi, and Hajkova 2007).

1.1.2.1 Histone modification

One of the epigenetic mechanisms that influences the reprogramming of PGCs is chromatin remodelling (Surani and Hajkova 2010). These chromatin changes are influenced by histone proteins which are essential for DNA packing in eukaryotes (Bannister and Kouzarides 2011). Different types of histones exist which can be modified on their tails by numerous modifications such as methylation and acetylation (Seah and Messerschmidt 2018). These tails serve as docking platforms for regulatory proteins with context-dependent regulatory outcomes (Bannister and Kouzarides 2011). For example, histone modifications Histone 3 lysine 9 di- or tri-methylation (H3K9me2/3) and H3 lysine 27 trimethylation (H3K27me3) are associated with repression and play an important role during germline development (Seah and Messerschmidt 2018).

1.1.2.2 DNA methylation

A second mechanism that plays a key role in PGC reprogramming is DNA methylation (Surani, Hayashi, and Hajkova 2007). DNA 5-methylcytosine (5mC) modification is the most direct way to alter DNA readouts in mammals (Bird 2002). Mechanistically, 5mC DNA methylation within a CpG context is primarily associated with transcriptional repression by restricting the binding of DNA-interacting proteins and by recruiting methylation-binding proteins (MBPs) (Borgel et al. 2010; Wu and Zhang 2010).

The enzymes that methylate DNA are known as DNA methyltransferases (DNMTs). The enzymes DNMT1, DNMT3A, DNMT3B and the cofactor DNMT3L belong to the group of DNA methyltransferases (Li, Beard, and Jaenisch 1993; Aapola et al. 2000; Okano, Xie, and Li 1998). In rodents, an additional DNA methyltransferase, DNMT3C, has been discovered (Barau et al. 2016). Each protein contains an ATRX-DNMT3L-DNMT3A (ADD) domain that recognises unmethylated histone 3 lysine 4 (H3K4), and a methyltransferase (MTase) domain (Lyko 2018; Otani et al. 2009). However, the MTase domain of DNMT3L is catalytically inactive (Chédin, Lieber, and Hsieh 2002). Additionally, DNMT3A and

DNMT3B contain a Pro-Trp-Trp-Pro (PWWP) domain which binds histones with an Histone 3 lysine 36 tri-methylation (H3K36me3) methylation pattern (Lyko 2018). While DNMT3A, DNMT3B, DNMT3C, and DNMT3L play important roles in *de novo* DNA methylation in mice, DNMT1 is responsible for methylation maintenance in the developing embryo (Sasaki and Matsui 2008; Barau et al. 2016; Li, Beard, and Jaenisch 1993; Lyko 2018). It has been demonstrated that *Dnmt1* knockout embryos are developmentally delayed and not viable and show de-repression of Intracisternal A particle (IAP) transposons (Walsh, Chaillet, and Bestor 1998; Greenberg and Bourc'his 2019; Li, Bestor, and Jaenisch 1992). DNMT3C is expressed only in male foetal germ cells, whereas DNMT3A, DNMT3B, and DNMT3L are expressed in various developmental contexts in both sexes (Barau et al. 2016; Aapola et al. 2000; Okano et al. 1999). It has been shown that *Dnmt3b* mice are not viable, whereas *Dnmt3a* knockout mice develop to term but die at approximately four weeks of age (Okano et al. 1999).

1.1.2.3 Mechanism of epigenetic changes

The epigenetic reprogramming in PGCs starts from E7.75 with the loss of H3K9me2 and an increase of H3K27me3 by E9.5 (Seki et al. 2005; Seki et al. 2007; Hajkova et al. 2008) (Figure 2). The loss of H3K9me2 is likely the consequence of a downregulation of the methyltransferase GLP which forms a complex with G9a (also EHMT2) which is responsible for deposition of Histone 3 lysine 9 mono- and di-methylation (H3K9me1 and H3K9me2) (Tachibana et al. 2005; Hackett, Zylicz, and Surani 2012). Simultaneously, specific lysine demethylases are upregulated (Hackett, Zylicz, and Surani 2012). The increasing H3K27me3 levels could be deposited by the upregulated methyltransferase enhancer of zeste homologue 1 (EZH1) (Shen et al. 2008). Interestingly, both H3K9me2 and H3K27me3 are repressive marks (Sasaki and Matsui 2008). Therefore, an increase in H3K27me3 may be necessary to maintain a repressive chromatin state in PGCs to complement H3K9me2 loss (Sasaki and Matsui 2008). At the same time H3K9me3 levels are maintained in migrating PGCs (Hajkova et al. 2008).

In parallel from E8.0 on, DNA methylation is removed in the developing germline (Hackett et al. 2013). The extensive reprogramming in PGCs results in the loss of approximately 90% of 5mC methylation in PGCs (Ben Maamar, Nilsson, and Skinner 2021) (Figure 2). Upregulation of the transcription regulators BLIMP1 and PR-domain containing protein 14 (PRDM14) results in the repression of essential components of the *de novo* DNA methylation machinery (DNMT3A and DNMT3B) as well as a key component of the methylation maintenance machinery (ubiquitin-like containing PHD and RING finger domains 1, UHRF1) (Kurimoto et al. 2008; Yamaji et al. 2008; Kato et al. 2007a; Kaneda et al. 2004; Sharif et al. 2007; Bostick et al. 2007). This repression of the DNA methylation machinery induces the first wave of DNA erasure which is a passive event (Ben Maamar, Nilsson, and Skinner 2021). Furthermore, the essential cofactor Np95 of the maintenance methyltransferase DNMT1 is repressed and excluded from the nucleus which affects DNMT1 activity from E9.5 (Seisenberger et al. 2012; Kurimoto et al. 2008; Kagiwada et al. 2013). Additional mechanisms, such as the repression of the methyltransferase GLP, may affect DNA methylation in an H3K9me2 dependent and independent manner (Dong et al. 2008; Tachibana et al. 2008).

However, only the second wave of demethylation which starts around E10.5 with entry into the genital ridges, results in demethylation of most genomic regions, including transposable elements, imprinted loci and single copy genes (Hajkova et al. 2002; Lee et al. 2002; Lane et al. 2003; Maatouk et al. 2006). This is believed to be an active event caused by removal of DNA methylation by Ten eleven translocation 1 and 2 (TET1 and TET2) (Hackett et al. 2013; Yamaguchi et al. 2013; Vincent et al. 2013). TET enzymes oxidize 5mC to 5-hydroxymethylcytosine (5hmC) thereby promoting the reversal of DNA methylation (Surani and Hajkova 2010) (Figure 2). The 5hmC can be further deaminated to 5-hydroxymethyluracil (5hmU) or oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Melamed et al. 2018). The last two of these modified cytosines, 5fC and 5caC, are targeted by the Thymine DNA glycosylase (TDG) followed by the Base excision repair (BER) which ultimately leads to repair-driven demethylation (He et al. 2011; Ito et al. 2011; Melamed et al. 2018). At the same time, PGCs undergo rapid amplification which results in replication-dependent dilution of DNA methylation (hemi methylated DNA) due to the absence of the DNA methylation maintenance machinery (Hackett et al. 2013) (Figure 2). Interestingly, some sequences, such as

IAP retrotransposons, are only partially demethylated during this process (Maatouk et al. 2006). The DNA methylation of repetitive sequences and non-promoter intergenic regions is completed before birth, while the majority of the remaining methylation is deposited before postnatal day three (P3) in mice and is finally completed at the age of five days (Li et al. 2004; Lees-Murdock, De Felici M Fau - Walsh, and Walsh 2003; Oakes et al. 2007).

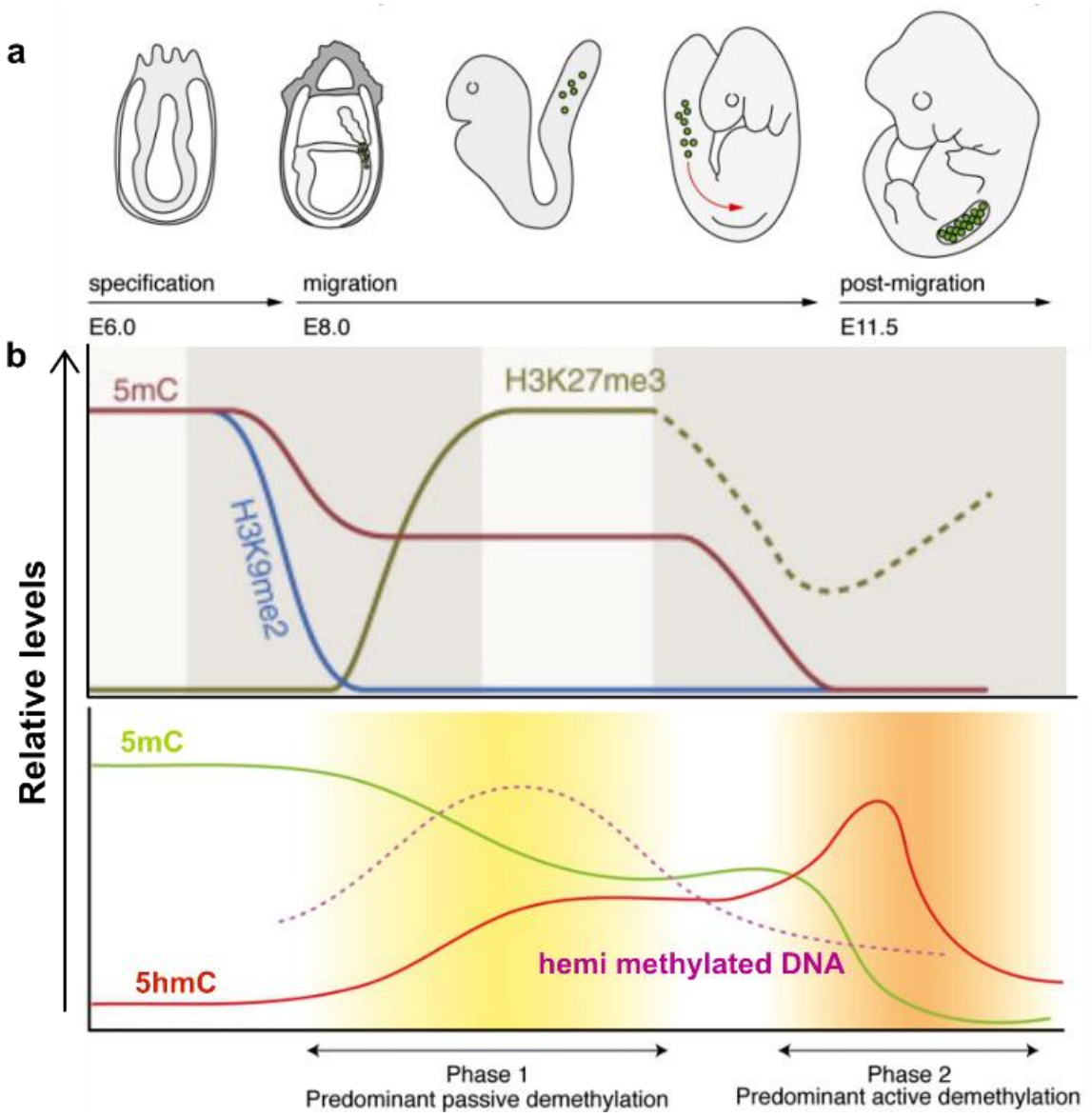

Figure 2: DNA and chromatin changes during germ cell development.

a, Different stages of primordial germ cell development. Stages are indicated with developing germ cell show in green. **b**, Epigenetic changes in PGCs. Histone H3 lysine 9 dimethylation (H3K9me2), H3 lysine 27 trimethylation (H3K27me3), 5mC (DNA methylation) and 5-hydroxymethylcytosine (5hmC). Phases of demethylation are indicated below. Figure adapted from (Hackett et al. 2013) and (Messerschmidt, Knowles, and Solter 2014).

In summary, reprogramming of PGCs is achieved by DNA demethylation and chromatin remodelling which are controlled by several complementary parallel pathways (Hackett et al. 2013). This provides the robustness, flexibility, and efficiency necessary to ensure successful germline development (Hackett et al. 2013).

1.1.3 Spermatogenesis

Approximately five days after birth, the arrested male germ cells, the prospermatogonia, resume mitosis and enter the prophase of meiosis I (Bellve et al. 1977; Hilscher et al. 1974). During spermatogenesis, diploid spermatogonial stem cells produce male haploid germ cells within the seminiferous tubule boundaries of the testis (Hess and de Franca 2008). This seminiferous epithelium contains numerous concentric layers of germ cells which are penetrated by a specific type of somatic cell, the Sertoli cells (Hess and França 2005) (Figure 3). Sertoli cells orchestrate germ cell development and differentiation and are key in providing the right environment and maintaining the correct timing during spermatogenesis (Ye et al. 1993).

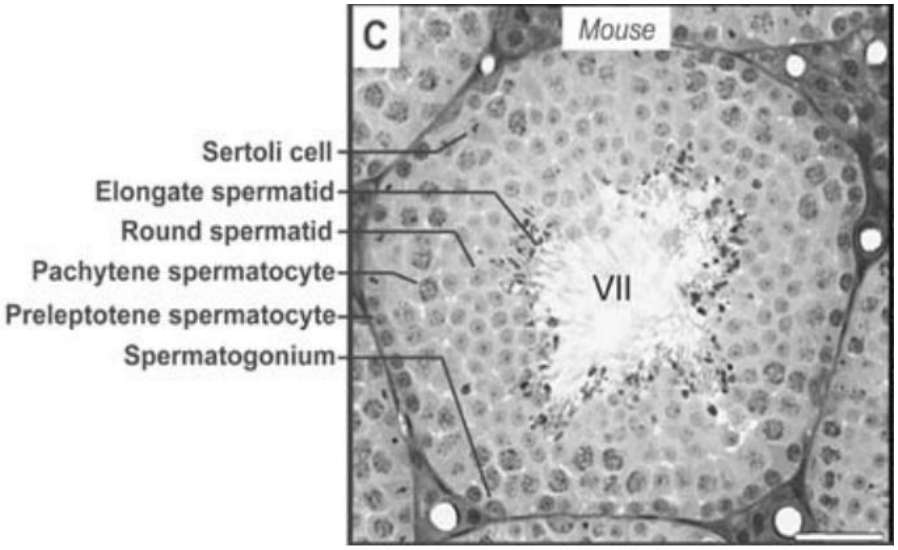

Figure 3: Cross-section of seminiferous tubules from mice.
Tubule in stage VII with cell types labelled in the image. Scale bar = 40 µm. Image adapted from (Hess and de Franca 2008)

Germ cells first undergo repeated mitotic divisions, followed by meiosis and spermiogenesis (Jan et al. 2012). The meiosis step includes duplication of chromosomes, genetic recombination, and reduction of chromosomes to obtain haploid spermatids that mature into spermatozoa for release into the tubule lumen (Hess and de Franca 2008) (Figure 4).

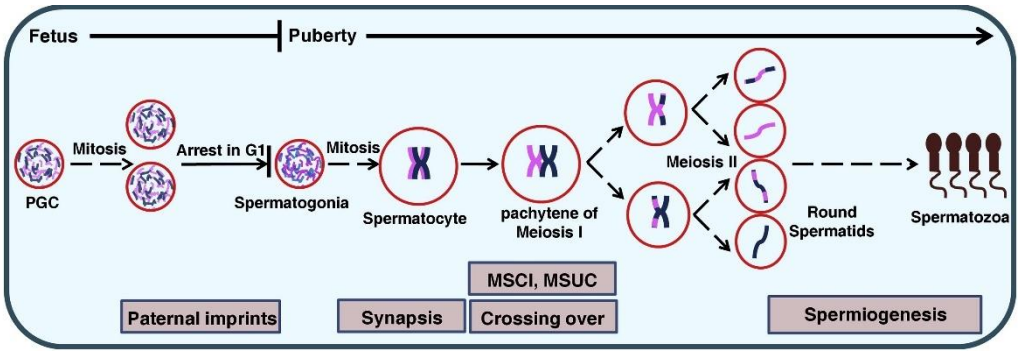

Figure 4: Developmental transition in male gametogenesis.
The top part indicates the timing of development. Key events are labelled in black overview. To simplify, meiosis is only shown for one pair of homologous chromosomes (pink and blue). Figure from (Kota and Feil 2010).

The process of spermatogenesis can be separated into different stages (Leblond and Clermont 1952). These stages are defined based on periodic acid-Schiff (PAS)-stained sections which allows to see changes in the Golgi region of spermatids (Hess and de Franca 2008) (Figure 5).

First, diploid spermatogonia undergo mitosis (Hess and de Franca 2008). Based on changes in heterochromatin, they can be divided into four classes: undifferentiated type A spermatogonia (A single (A_s), A paired (A_{pr}), A aligned (A_{al})), differentiated type A spermatogonia (A_1, A_2, A_3, A_4), intermediate spermatogonia (In), and type B spermatogonia (B) (de Rooij and Russell 2000). B-spermatogonia then divide by successive mitosis until the formation of preleptotene spermatocytes (Lara et al. 2018). This starts the prolonged meiosis step which takes approximately 14 days in the mouse and includes the following stages: preleptotene, leptotene, zygotene, pachytene and diplotene (Hess and de Franca 2008). DNA synthesis, chromosome pairing, genetic recombination, and chromosome separation occur during these steps (Lara et al. 2018). Subsequently, the cell undergoes meiotic division in stage XII (Hess and de Franca 2008). The

division step takes approximately one day and can be separated into three stages: first, meiosis I (division of 4n cells); second, the formation of secondary spermatocytes (2n cells); third, meiosis II, the division of the 2n spermatocytes in 1n round spermatids (Hess and de Franca 2008) (Figure 6).

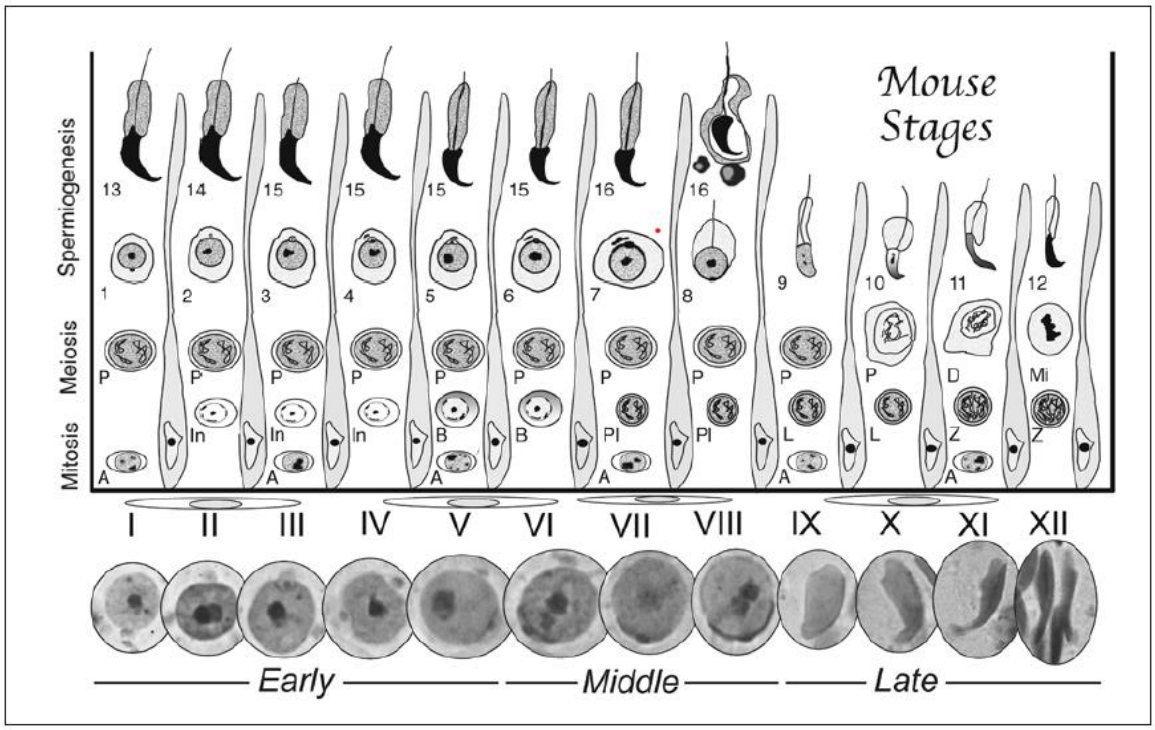

Figure 5: Overview of mouse spermatogenesis stages (I–XII).
Different stages are shown in layers, with the Sertoli cells separating each stage. Labelling of each stage with letters or numbers. Bottom, periodic acid-Schiff (PAS) and haematoxylin stained early, middle, and late spermatid nuclei. Spermatogonia (A, In, B); spermatocytes (Pl: preleptotene, L: leptotene, Z: zygotene, P: pachytene, D: diakinesis, Mi: meiotic division); round spermatids (1-8); and elongated spermatids (9-16). Figure from (Hess and de Franca 2008)

Finally, the transformation of 1n spermatids into elongated and mature spermatozoa is called spermiogenesis (Hess and de Franca 2008). During steps 1–3 of spermiogenesis, the Golgi apparatus is visible by PAS staining (Leblond and Clermont 1952). In step 2-3 a single, large acrosomal granule is observed (Leblond and Clermont 1952). The acrosomal granule begins to flatten and forms a cap over the surface of the nucleus (capping) during stages 4 – 8 (Hess and de Franca 2008). In late step 8, when the acrosome covers approximately 1/3 of the nuclear surface, the nucleus begins to change shape (Hess and de Franca 2008). In the following acrosomal steps 9 – 14, the acrosomal system migrates over the

surface of the elongating spermatid nucleus (Hess and de Franca 2008). At the same time, chromatin is packed more tightly which can be visualised by more intense haematoxylin staining (Hess and de Franca 2008). The maturation steps 15 – 16 occur during stages III – VIII, where the nucleus condensation continues and the acrosome develops into a thin structure (de França et al. 1995). The extra cytoplasm is removed and can be seen as cytoplasmic lobes and residual bodies (de França et al. 1995) (Figure 5).

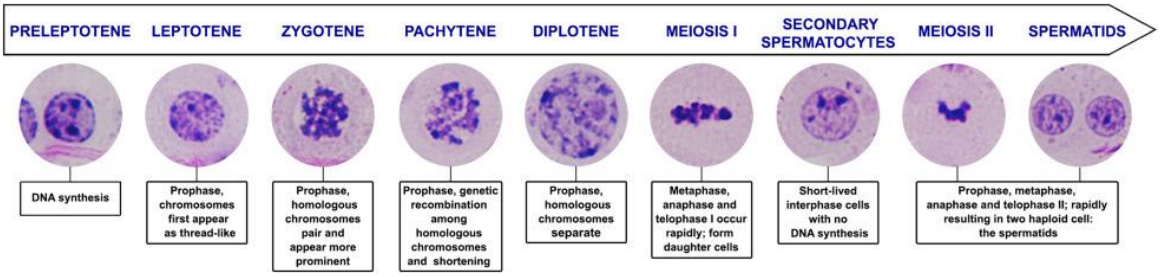

Figure 6: Morphology of mouse spermatocytes during the prolonged meiosis step.

Different stages are labelled at the top, with arrows indicating the direction of development. Periodic acid-Schiff (PAS) and haematoxylin staining of spermatocytes at the respective stages are shown below. The main characteristics of each phase are presented in boxes below the spermatocytes. Figure adapted from (Lara et al. 2018).

During spermatogenesis, the cell number and quality are constantly controlled (Lara et al. 2018). Significant germ cell loss is observed in the spermatogonial phase to regulate cell density (Murphy and Richburg 2014). Apoptosis, probably caused by quality issues related to DNA integrity, is frequently observed during meiosis and is slightly less frequent in spermatids (Murphy and Richburg 2014). In fact, a major threat to DNA integrity is the activity of specific DNA sequences, the transposons (Fu and Wang 2014).

1.2 The piRNA pathway

Transposons, also called jumping genes, account for approximately 40 – 50 % of the mammalian genome (Sasaki and Matsui 2008). As the name “jumping gene” suggests, these elements are mobile and can threaten genomic integrity (Sasaki and Matsui 2008). However, while transposons are commonly called invaders and

cells are described as fighting against them, transposons are also major regulatory units and drivers of evolution and adaptation (Casacuberta and Gonzalez 2013). Throughout most of our lives, transposon sequences are methylated and thus incapable of moving and causing damage (Slotkin and Martienssen 2007). However, during germline reprogramming, the genome is broadly demethylated which puts germ cells under the threat of a burst of transposon activity (Guibert, Forné, and Weber 2012; Seisenberger et al. 2012; Molaro et al. 2014) (Figure 7).

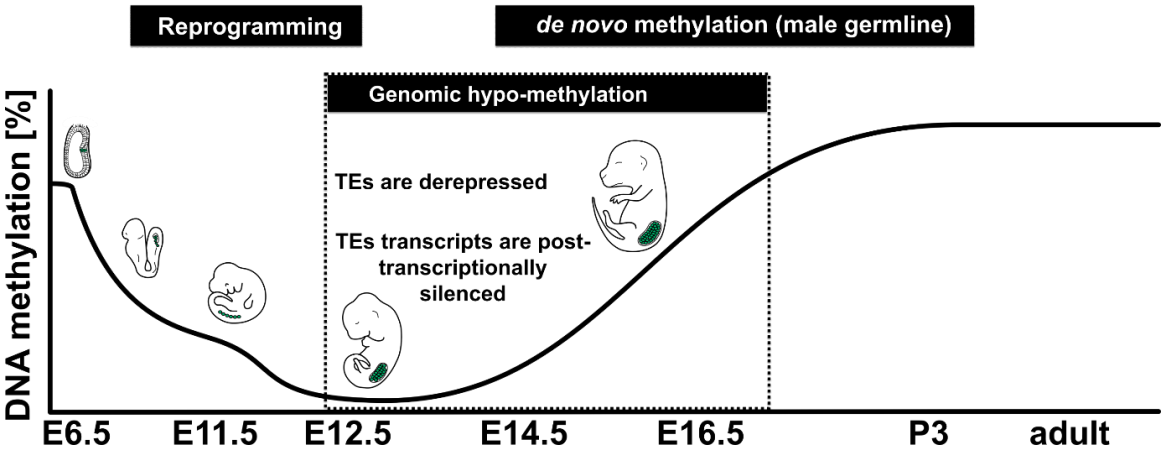

Figure 7: DNA demethylation and methylation during embryonic development.

The figure shows DNA methylation levels (y-axis) from embryonic day 6.5 (E6.5) to adulthood (x-axis). TE = transposable element.

1.2.1 Young transposable elements (transposons)

Mammals do have several types of transposons, including retrotransposons (also Class I elements) and DNA transposons (also Class II elements) (Mouse Genome Sequencing et al. 2002; Deininger and Batzer 2002). DNA transposons “cut and paste” themselves *via* a DNA intermediate that is associated with a transposase and occupy less than 3 % of the mammalian genome (Wicker et al. 2007). In fact, most mammals lack DNA transposons with the ability to move, because older elements accumulated mutations (Pace and Feschotte 2007; Nagamori et al. 2015).

The more common retrotransposons follow a “copy and paste” mechanism for propagation *via* an ribonucleic acid (RNA) intermediate (Platt, Vandewege, and

Ray 2018). Retrotransposons can be divided into three subclasses: Long interspersed nuclear elements (LINEs), Short interspersed nuclear elements (SINEs), and Long terminal repeats (LTRs) (Nagamori et al. 2015). LINEs and SINEs are non-LTR elements characterised by a repetitive tail and the absence of long terminal repeats (Platt, Vandewege, and Ray 2018). LINEs are typically four to seven kilo bases (kb) long and encode the enzymatic machinery necessary for mobilisation (Platt, Vandewege, and Ray 2018). In contrast, SINEs are nonautonomous; they lack machinery for self-mobilisation and are only 150 – 500 base pairs (bp) in length (Platt, Vandewege, and Ray 2018). Overall, LINE and SINE elements are the most common types of transposons (Platt, Vandewege, and Ray 2018). LTRs are a few hundred to ten kilobases, are self-autonomous, and are structured similar to retrotransposons (Platt, Vandewege, and Ray 2018). In mammals, LTRs occupy approximately 4 – 10 % of the genome (Mikkelsen, Wakefield, et al. 2007). They encode at least one gag and one pol protein which are surrounded by the 5' and 3' LTR (Platt, Vandewege, and Ray 2018). Some LTRs contain an envelope protein gene as well (Eickbush and Malik 2007) (Figure 8).

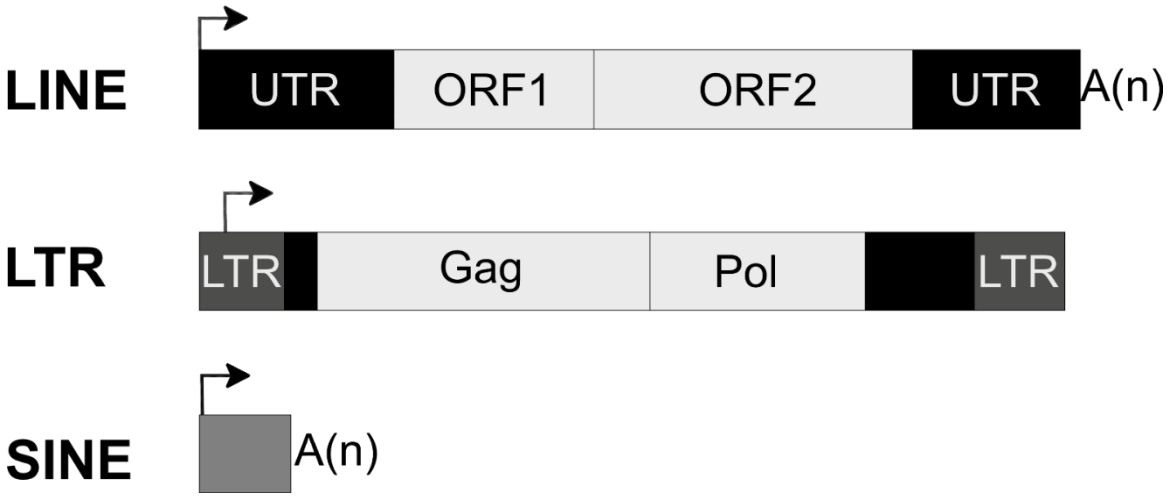

Figure 8: Overview of common retrotransposon structures.

All structural features are labelled accordingly. LINE, long interspersed nuclear elements; LTR, long terminal repeats; SINE, short interspersed nuclear elements; UTR, untranslated region**;** ORF1, nuclear chaperone protein; ORF2 reverse transcriptase; A(n), poly A repeat; LTR, long terminal repeat; GAG, GAG protein; POL, polyprotein

Most transposons accumulate mutations over time, and become inactive (Nagamori et al. 2015). Nevertheless, mammals usually have at least one active transposon family in their genomes (Platt, Vandewege, and Ray 2018). For example, mice mainly battle IAP and LINE1 transposons (Dewannieux et al. 2004; Goodier et al. 2001; Naas et al. 1998). Although these active transposons can drive evolution, transposon movement often causes irreversible DNA damage (Platt, Vandewege, and Ray 2018). In fact, transposon insertions have been associated with several diseases and various forms of cancer (Chenais 2013; Hancks and Kazazian 2016; Morse et al. 1988). At the same time, transposon expression is essential in some settings, for example, for the development of neurons (Li et al. 2012; Reilly et al. 2013).

However, owing to their ability to be deleterious and compromise genome integrity, transposons are tightly controlled in mammals, particularly in the germline (Platt, Vandewege, and Ray 2018). For example, Krüppel-associated box domain containing zinc finger proteins (KRAB-ZFPs) recognise the sequence of transposons *via* their zinc finger (Ecco, Imbeault, and Trono 2017). This recruits KRAB-associated protein 1 (KAP1) which in turn recruits a series of epigenetic regulators, including histone and DNA methyltransferases (Ecco, Imbeault, and Trono 2017; Wolf et al. 2015). In addition, the cytidine deaminase Apolipoprotein B mRNA editing enzyme catalytic polypeptide (APOBEC) family is involved in editing complementary DNA (cDNA) transcripts of LTR and non-LTR transposons in mammals. (Schumann 2007; Kinomoto et al. 2007; Richardson et al. 2014). This transposon editing causes either direct destruction or high levels of mutations in the cDNA, likely restricting its ability to cause further damage (Harris et al. 2003). However, the main transposon defence mechanism is the piRNA pathway, which silences through transcriptional and post-transcriptional mechanisms *via* P-element induced wimpy (PIWI) proteins (Ozata et al. 2018).

1.2.2 PIWI proteins

PIWI proteins are a subclade of Argonaute proteins that are primarily expressed in gonads (Lin and Spradling 1997; Fu and Wang 2014). In 2006, they were first found to associate with a specific class of small RNAs, the PIWI-interacting RNAs (piRNAs) in mammals (Lau et al. 2006; Grivna et al. 2006; Aravin et al. 2006; Girard

et al. 2006). PIWI proteins catalyse the endonucleolytic cleavage of RNA targets, play a key role in piRNA maturation, and are essential for piRNA-directed transposon silencing (Gunawardane et al. 2007; Saxe et al. 2013; Ozata et al. 2018).

In mammals, three PIWI proteins have been identified: MIWI (PIWIL1), MILI (PIWIL2), and MIWI2 (PIWIL4) (Aravin et al. 2007; Carmell et al. 2007; Reuter et al. 2011). Each PIWI protein contains three domains, the PAZ, MID and PIWI domain (Ozata et al. 2018). When bound to piRNA, the PAZ domain harbours the 3' of the piRNA containing the 2'-O-methyl modification, while the MID domain accommodates the 5' end with the phosphate modification (Parker, Roe, and Barford 2005; Ma et al. 2005; Tian et al. 2011). The PIWI domain contains an endonuclease that hydrolyses target RNA (Parker, Roe, and Barford 2004). However, only MILI and MIWI do have slicer activity on their targets (Reuter et al. 2011; De Fazio et al. 2011). It has been shown that mice piRNA biogenesis happens in specific areas in the cell, the so called nuage (Ozata et al. 2018). MILI is found in nuage located close to the mitochondria (pi-body), whereas MIWI2 is found in the perinuclear region (piP-body) (Aravin et al. 2009; Wang et al. 2020).

PIWI proteins are expressed during different stages of spermatogenesis (Ozata et al. 2018). MIWI2 is only expressed during the gonocyte stage (Carmell et al. 2007). MILI is expressed longest from the gonocyte until the round spermatid stage while MIWI expression only starts after birth in the late pachytene stage and stops in round spermatids (Kuramochi-Miyagawa et al. 2008; Deng and Lin 2002). Therefore, embryonic transposon silencing is facilitated by MILI and MIWI2, whereas MILI and MIWI are responsible for silencing in the adult testis (Ozata et al. 2018).

Mice without MILI or MIWI2 show deregulation of LINE1 and IAPs and a meiotic arrest phenotype (Kuramochi-Miyagawa et al. 2008; Carmell et al. 2007). Interestingly, animals with catalytically inactive MILI showed upregulation of LINE1 only (De Fazio et al. 2011). In contrast to that, MIWI knockout animals show a spermatogenic arrest phenotype due to the upregulation of transposons in the animal after birth (Deng and Lin 2002; Reuter et al. 2011).

1.2.3 piRNA biogenesis

The germline-specific PIWI proteins associate with piRNAs to form a complex called the piRNA-induced silencing complex (piRISC) (Siomi et al. 2011). In mice, piRISCs are involved in transcriptional and post-transcriptional silencing of transposons (Siomi et al. 2011). piRNAs were first identified in *Drosophila melanogaster* (Aravin et al. 2001). To date, piRNAs that guide PIWI proteins have been found in the gonads of insects, mammals, nematodes, and fish (Aravin et al. 2006; Girard et al. 2006; Vagin et al. 2006; Grivna et al. 2006; Houwing et al. 2007; Lau et al. 2006; Saito et al. 2006; Batista et al. 2008; Das et al. 2008).

piRNAs are usually 24 - 35 nucleotides (nt) in length (Sun, Lee, and Li 2022). In mice, piRNAs that associate with the PIWI proteins MILI, MIWI2 and MIWI do have a distinct size of ~26 nt, ~28 nt and ~30 nt, respectively (Siomi et al. 2011). While other small RNAs, such as Small interfering RNAs (siRNAs) and Micro RNAs (miRNAs), require the endonuclease Dicer for processing, piRNAs are processed *via* a Dicer-independent mechanism (Vagin et al. 2006; Grivna et al. 2006; Aravin et al. 2006; Girard et al. 2006; Houwing et al. 2007). piRNAs contain a 5’ monophosphate and a 2-O-methyl modification at the 3’ end which has been hypothesised to increase piRNA stability (Kirino and Mourelatos 2007; Vagin et al. 2006; Gunawardane et al. 2007; Ruby et al. 2006; Horwich et al. 2007; Saito et al. 2007; Ohara et al. 2007; Montgomery et al. 2012; Kamminga et al. 2010; Lim et al. 2015).

piRNAs can be further separated into prepachytene and pachytene piRNAs according to the developmental time point at which they are expressed (Girard et al. 2006). Prepachytene piRNAs are expressed from the spermatogonia until the prepachytene stage of meiosis (Ozata et al. 2018) (Figure 9). They are primarily derived from transposons and associate with MILI and MIWI2 to transcriptionally and post-transcriptionally silence transposons during embryonic development (Kuramochi-Miyagawa et al. 2008; Aravin et al. 2008). The expression of pachytene piRNAs begins during the pachytene stage of meiosis during adult spermatogenesis (Li et al. 2013). Most pachytene piRNAs are derived from long non-coding RNA (lncRNA) genes (Li et al. 2013). They associate with the PIWI proteins MILI and MIWI and are involved in post-transcriptional silencing in adult animals (Vourekas et al. 2012; Di Giacomo et al. 2013; Reuter et al. 2011).

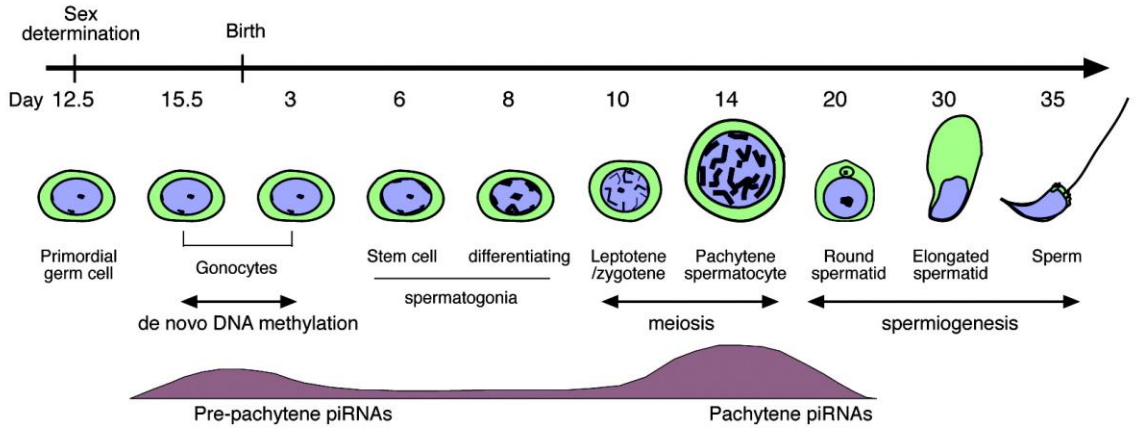

Figure 9: Expression of piRNAs during development.

Developmental expression of piRNAs during mouse spermatogenesis. The embryonic and postnatal developmental stages are indicated in the upper part. Pre-pachytene and pachytene piRNA expressions are shown below. Figure adapted from (Fu and Wang 2014).

piRNA biogenesis occurs in the so-called nuage within the cell which is in the Intramitochondrial cement (IMC, pi-bodies) or perinuclear granules (piP-bodies) (Wang et al. 2020; Aravin et al. 2009; Eddy 1974). piRNAs are made by two pathways, the primary processing in the intramitochondrial cement and the secondary processing which happens in the region close to the nucleus (Fu and Wang 2014; Wang et al. 2020). In this process, piRNA biogenesis always follows the order of 5' formation, PIWI binding, and 3' formation (Ozata et al. 2018). For primary piRNA biogenesis, long piRNA precursors are generated by RNA polymerase II (Pol II) from piRNA cluster (Siomi et al. 2011). These piRNA precursors have a 5' 7-methylguanosine cap and are 3' polyadenylated (Fu and Wang 2014). The long precursors are cut into shorter piRNA intermediates by an endonuclease in a Mov10 like RISC complex RNA helicase 1 (MOV10L1) dependent matter, which allows the piRNA intermediates to associate with MILI (Nishimasu et al. 2012; Ipsaro et al. 2012; Vourekas et al. 2012; Saxe et al. 2013; Vourekas et al. 2015). The 5' end of these intermediates already has the signature of a mature piRNA with a preference for uridine (Vourekas et al. 2012; Saxe et al. 2013). While bound to PIWI proteins, the piRNA intermediates are trimmed by the exonuclease PARN like ribonuclease domain containing exonuclease 1 (PNLDC1) and its cofactor Tudor and KH domain containing (TDRKH) to reach the length of a mature piRNA (Izumi et al. 2016; Kawaoka et al. 2011; Nishimura et al. 2018;

Saxe et al. 2013). Finally, the HEN methyltransferase 1 (HENMT1) modifies the 3' end with a 2-O-methyl modification, generating mature primary piRNAs (Kirino and Mourelatos 2007).

Primary processing produces the initial piRNA pool which is further amplified by secondary processing, also called the ping-pong amplification loop (Fu and Wang 2014). During secondary processing, initiator piRNAs associate with PIWI proteins and guide them to RNA targets through complementary base pairing (Brennecke et al. 2007; Gunawardane et al. 2007). The PIWI protein cleaves the target RNA between the 10^{th} and 11^{th} nucleotides, generating the 5' end of the new secondary piRNA intermediate with a 10 adenine bias (Gunawardane et al. 2007; Brennecke et al. 2007; Aravin et al. 2008). The secondary piRNA intermediate is loaded onto a PIWI protein where the 3' end is established, generating mature secondary piRNAs (Kirino and Mourelatos 2007; Fu and Wang 2014). This newly generated secondary piRNA can associate with MILI and support post-transcriptional silencing and piRNA amplification *via* the ping-pong cycle or associate with MIWI2, localise to the nucleus, and facilitate transcriptional transposon silencing (Ozata et al. 2018) (Figure 10).

Another essential pathway for piRNA generation is phased biogenesis (Ozata et al. 2018). Once an RNA target enters the piRNA pathway, MILI and MIWI2 can bind to it at regular distances, cleaving and thereby destroying the transposon transcript (Yang et al. 2016; Ozata et al. 2018). The resulting piRNAs show a 1 uracil signature (Yang et al. 2016).

These different mechanisms allow for the generation of a broad range of piRNAs that associates with PIWI proteins which are necessary for efficient transcriptional and post-transcriptional transposon silencing in germ cells (Watanabe et al. 2018). While post-transcriptional silencing relies on the endonuclease activity of PIWI proteins, transcriptional silencing is executed by the DNA methylation machinery (Ozata et al. 2018).

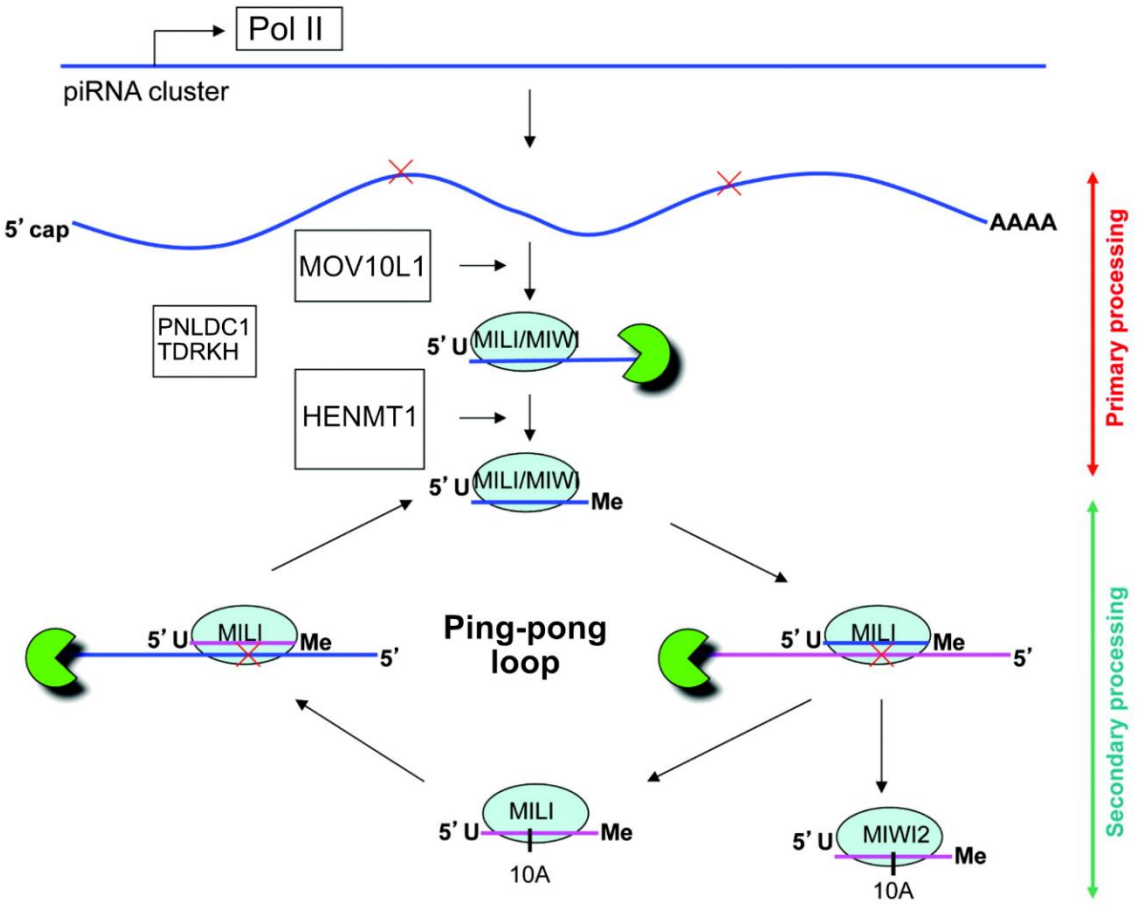

Figure 10: piRNA biogenesis.

Schematic of primary and secondary (ping-pong amplification) piRNA biogenesis. The top part shows the primary processing, and the bottom part depicts the secondary processing. All proteins were labelled accordingly. PARN like ribonuclease domain containing exonuclease 1 (PNLDC1), Tudor and KH domain containing (TDRKH), HEN methyltransferase 1 (HENMT1), Mov10 like RISC complex RNA helicase 1 (MOV10L1). Image adapted from (Fu and Wang 2014)

1.2.4 The DNA methyltransferases in transposon methylation

Previous studies give some insight into which of the DNA methyltransferases are necessary for *de novo* transposon methylation (Lyko 2018). A conditional knockout (with Cre recombinase activation at E9.5) showed that DNMT3A, is necessary for *de novo* methylation and normal spermatogenesis (Kaneda et al. 2004). However, *Dnmt3b* conditional knockout mice did not exhibit this phenotype (Kaneda et al. 2004). It has been concluded that DNMT3B plays an essential role in the methylation of a broad spectrum of sequences in the early embryo, whereas DNMT3A broadly methylates the genome to ensure spermatogonial stem cell

(SSC) commitment to spermatogenesis (Dura et al. 2022; Kaneda et al. 2004). *Dnmt3l* knockout mice showed high transcription of LINE and IAP transposons in spermatogonia and spermatocytes (Webster et al. 2005; Bourc'his and Bestor 2004; Hata et al. 2006). This mutation causes meiotic failure in male mice, resulting in germ cell loss by the mid-pachytene stage, and complete azoospermia in older animals (Sasaki and Matsui 2008). The same phenotype has been observed in mice without DNMT3C (Barau et al. 2016).

It has been shown that DNMT3A and DNMT3C with the adapter protein DNMT3L are essential for *de novo* methylation during germline development (Bourc'his and Bestor 2004; Kaneda et al. 2004; Dura et al. 2022). A recent study investigated this further and found that DNMT3A methylates the whole genome of prenatal male germ cells with the exception of young transposon promoters, whereas DNMT3C specifically methylates the promoter regions of young LTR or non-LTR elements (Dura et al. 2022). At the same time DNMT3L is essential for the methylation of both, the DNMT3A and DNMT3C genomic targets (Dura et al. 2022; Kato et al. 2007a). However, how the methylation machinery is associated with the piRNA pathway has remained elusive until recently.

1.2.5 The nuclear piRNA pathway

It is known that MIWI2 is essential for nuclear piRNA-directed *de novo* DNA methylation, but no direct connection between MIWI2 and the DNA methylation machinery has been established (Kuramochi-Miyagawa et al. 2008).

MIWI2 does not exhibit slicer activity; it is the only nuclear PIWI protein and is indispensable for the *de novo* methylation of active transposons (Aravin et al. 2008). MIWI2-directed *de novo* DNA methylation targets and controls evolutionary young LINE1 and IAP elements (Pezic et al. 2014; Kojima-Kita et al. 2016). Specifically, it associates with transposon promoter regions and guides the DNA methylation machinery to these loci (Barau et al. 2016; Pastor et al. 2014).

Around E15.5, MIWI2 expression starts to be visible in the cytoplasm of germ cells of the testis (Aravin et al. 2008). MIWI2 binds to piRNA which allows the piRNA-MIWI2 complex to localise to the nucleus (Kuramochi-Miyagawa et al. 2008; Aravin et al. 2008). By E16.5, MIWI2 is mainly nuclear and associates with the nascent transposon transcript through complementary base pairing *via* piRNA (Zoch et al.

2020; Ozata et al. 2018). This results in the recruitment of the DNA methylation machinery and the placement of methylation (Kuramochi-Miyagawa et al. 2008) (Figure 11). However, a direct link between MIWI2 and the DNA methylation machinery has not been shown, and it remains unclear how DNA methyltransferases are guided to these loci (Zoch et al. 2020).

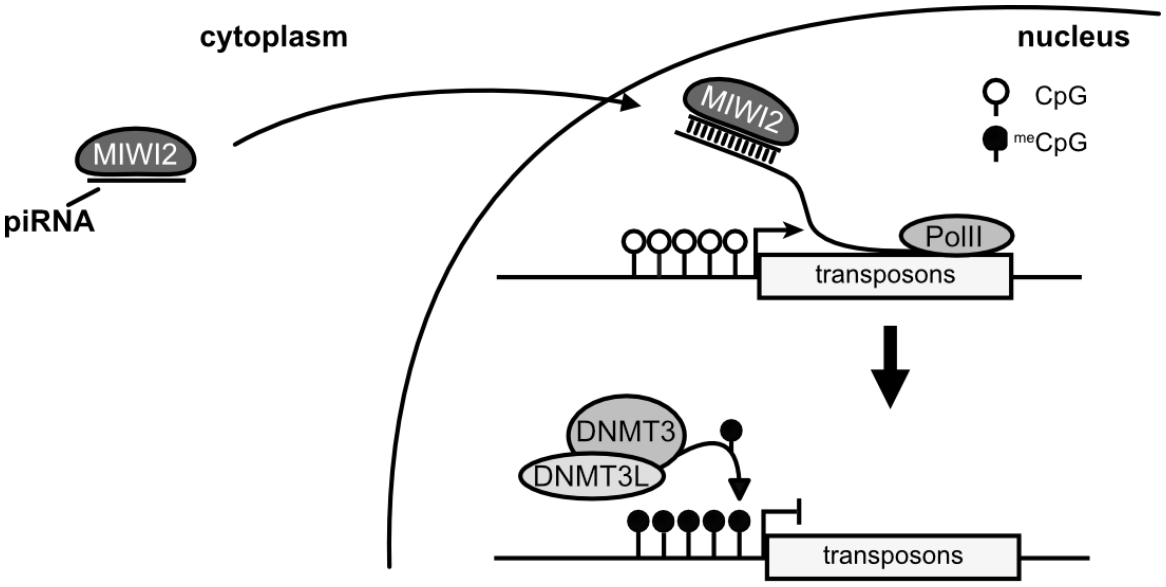

Figure 11: MIWI2-guided *de novo* DNA methylation.
Cytoplasmic MIWI2 (left) binds to piRNA and locates to the nucleus (right). Nascent transposon transcripts are recognised by the piRNA-MIWI2 complex *via* complementary base pairing. This results in recruitment of DNA methyltransferases (DNMT3s) and CpG methylation. Unmethylated CpG, white; methylated CpG, black.

In addition to MIWI2, the nuclear proteins MORC family CW-type zinc finger 1 (MORC1), Tudor domain containing 9 (TDRD9), Testis expressed 15 (TEX15) and SPOC domain-containing 1 (SPOCD1) are essential for transcriptional transposon silencing and are candidates for linking to the *de novo* DNA methylation machinery (Pastor et al. 2014; Shoji et al. 2009; Schopp et al. 2020; Zoch et al. 2020).
Deletion of MORC1 causes a phenotype very similar to that of the *Mili* or *Dnmt3c* knockout (Barau et al. 2016; Pastor et al. 2014; Inoue et al. 1999; Watson et al. 1998). MORC1 belongs to the GHKL ATPase family and is present in undifferentiated cells of blastocysts and gonocytes (Hruz et al. 2008; Inoue et al. 1999; Watson et al. 1998). MORC1 is essential for the methylation of transposon promoter but does not contribute to piRNA biogenesis (Pastor et al. 2014). Recent

discoveries have suggested that MORC1 re-establishes heterochromatin on active H3K9me3-marked transposons (Uneme et al. 2023). However, the exact role of MORC1 in MIWI2-guided *de novo* DNA methylation in the male germline remains unknown (Onishi, Yamanaka, and Siomi 2021; Uneme et al. 2023). Two proteins that are known to associate with MIWI2 are TDRD9 and TEX15 (Schopp et al. 2020; Shoji et al. 2009). TDRD9 is a DExH-box helicase/ATPase containing a Tudor domain (Wenda et al. 2017; Shoji et al. 2009). Although TDRD9 is not essential for piRNA biogenesis, it is essential for LINE1 silencing (Shoji et al. 2009; Wenda et al. 2017). TEX15 contains a DUF3715 domain and two TEX15 domains and is expressed in adult and foetal testes (Schopp et al. 2020; Yang et al. 2020). While TEX15 is not involved in piRNA biogenesis, it is essential for mammalian piRNA-directed *de novo* DNA methylation (Schopp et al. 2020; Yang et al. 2020). However, the functional mechanisms underlying the effects of TDRD9 and TEX15 remain unclear (Shoji et al. 2009; Wenda et al. 2017; Schopp et al. 2020). Finally, the discovery of SPOCD1 provided a link between the piRNA pathway and the *de novo* DNA methylation machinery (Zoch et al. 2020).

1.2.6 SPOCD1

SPOCD1 is a novel nuclear protein that associates with MIWI2, and is expressed only during the time of *de novo* DNA methylation (Zoch et al. 2020). The SPOCD1 interactome revealed an association with the DNA methyltransferases DNMT3L and DNMT3A, linking MIWI2 and the piRNA pathway to the methylation machinery (Zoch et al. 2020). In addition, SPOCD1 co-purifies *in vivo* with components of the Nucleosome remodelling deacetylase (NURD) and Barrier-to-autointegration factor (BAF) chromatin remodelling complex (Zoch et al. 2020) (Figure 12).

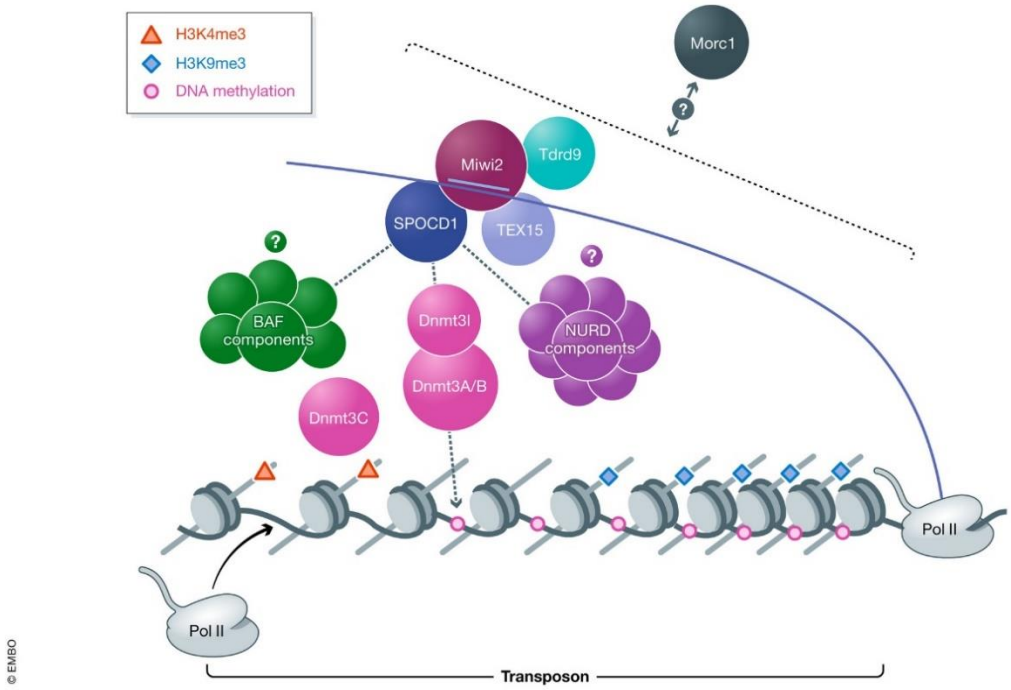

Figure 12: MIWI2-mediated *de novo* DNA methylation.

The MIWI2-piRNA complex recognises nascent transposon transcripts. MIWI2 interacts with TDRD9, SPOCD1, and TEX15. SPOCD1 recruits BAF and NURD components as well as DNMT3L and DNMT3A. DNMT3C and MORC1 are also important in *de novo* DNA methylation; however, the crosstalk between these components is not yet known. Deposition of histone methylation patterns or DNA methylation, as indicated. Image from (Onishi, Yamanaka, and Siomi 2021).

Male *Spocd1* deficient mice were infertile due to the upregulation of young LINE1 and IAP elements resembling the knockout phenotype of other nuclear piRNA effectors such as *Miwi2*, *Dnmt3c*, *Dnmt3l* or *Tex15* (Barau et al. 2016; Schopp et al. 2020; Zoch et al. 2020; Carmell et al. 2007; Bourc'his and Bestor 2004). In summary, SPOCD1 is essential for piRNA-directed *de novo* DNA methylation and transposon silencing, and provides the first linkage of the mammalian piRNA pathway to the DNA methylation machinery (Zoch et al. 2020).

In addition, studies have revealed that SPOCD1 plays a role in different cancer types such as colorectal cancer, gastric cancer, ovarian cancer, glioma, osteosarcoma, and bladder cancer (Gan et al. 2023; Zhu et al. 2017; Liu et al. 2020; Liu et al. 2018; Liang et al. 2018; van der Heijden et al. 2016). SPOCD1 mutations have also been associated with bipolar and major depressive disorders (Takamatsu et al. 2022). Furthermore, it has also been shown in adult human testes that early SSCs express SPOCD1 while SPOCD1 is downregulated in some non-obstructive azoospermia (NOA) patients (Zhou et al. 2022). This study

concluded that SPOCD1 regulates the proliferation and apoptosis of human SSCs *via* Adenylate kinase 4 (AK4) (Zhou et al. 2022).

Phylogenetic analysis revealed that SPOCD1 is closely related to the transcriptional regulator PHD finger protein 3 (PHF3) and death inducer obliterator (DIDO) and likely originated from a duplication of PHF3 (Zoch et al. 2020). In fact, all three proteins contain a transcription factor II S mid (TFIIS-M) and Spen paralogue and orthologue c-terminal (SPOC) domain, but SPOCD1 seems to have lost the chromatin-binding Plant homeo domain (PHD) domain in the gene duplication event (Zoch et al. 2020) (Figure 13).

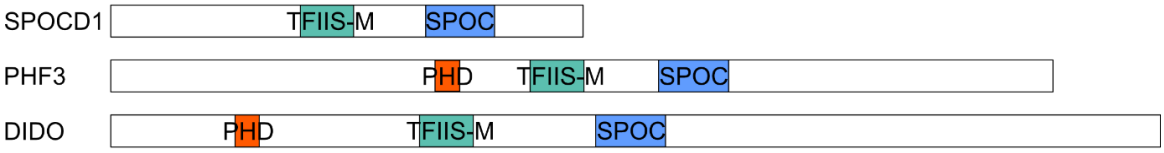

Figure 13: Domain structures of SPOCD1, PHF3, and DIDO.
TFIIS-M (transcription factor II S mid) domain in green, SPOC (Spen paralogue and orthologue C-terminal) domain in blue, and PHD finger in orange. The mouse sequences of SPOCD1 (B1ASB6), PHF3 (B2RQG2), and DIDO (Q8C9B9) were used.

The TFIIS-M domain was first discovered as the mid/central domain of the Transcription factor II S (TFIIS) (Kettenberger, Armache, and Cramer 2003). The TFIIS-M domain of TFIIS forms a three-helix bundle that is constructed around a hydrophobic core (Morin et al. 1996; Olmsted et al. 1998) (Figure 14). It has been shown that this domain is essential for binding to RNA Pol II (Kettenberger, Armache, and Cramer 2003). The TFIIS-Pol II interaction enables faster and more efficient transcription by the polymerase (Wind and Reines 2000). Specifically, TFIIS is important for inducing structural changes that allow cleavage and release of trapped RNA thereby reactivating arrested Pol II (Kettenberger, Armache, and Cramer 2004; Kettenberger, Armache, and Cramer 2003; Cheung and Cramer 2011).

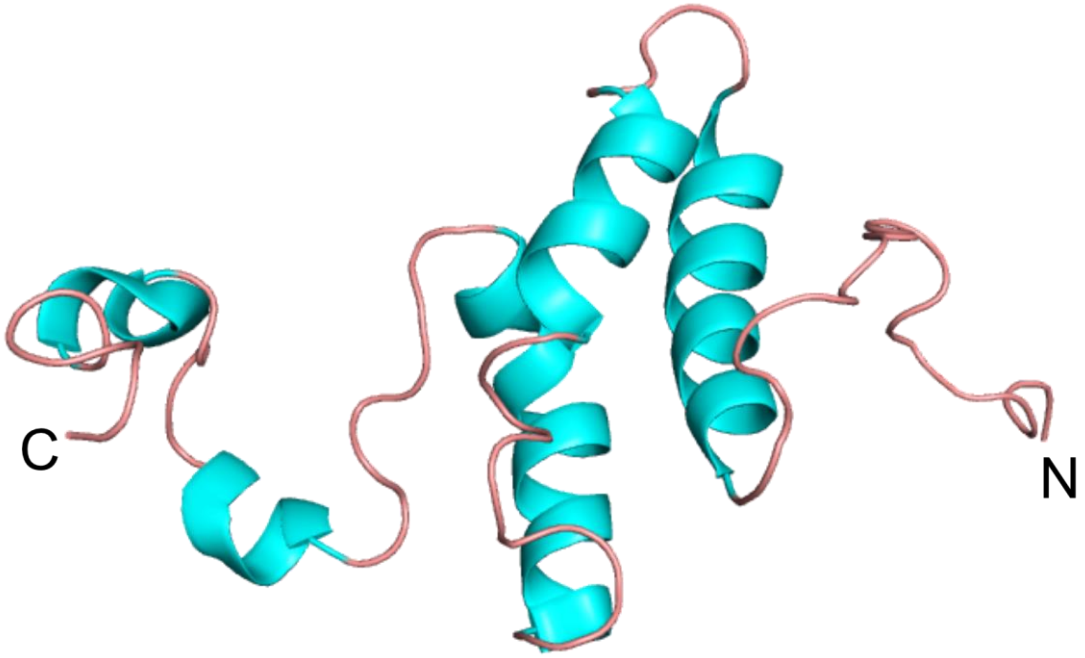

Figure 14: Structure of the TFIIS-M domain of TFIIS.
NMR structure of the mid-domain of transcription factor TFIIS. Secondary structure elements (α-helix) are indicated in blue. The C- and N-termini are labelled in the figure. PDB: 1ENW (Morin et al. 1996)

The SPOC domain is a small protein-protein interaction domain that can be found in different eukaryotic species (Appel, Benedum, et al. 2023). In addition to SPOCD1, DIDO, PHF3, SMRT/HDAC1-associated repressor protein (SPEN, also SHARP or MINT), RNA binding motif protein 15 (RBM15), and its paralogue, RBM15B, contain a SPOC domain (Appel, Benedum, et al. 2023). The SPOC domains form a distorted β-barrel that contains around seven β-stands and a variable number of α-helices (Ariyoshi and Schwabe 2003; Mikami et al. 2014; Zhang et al. 2016; Appel et al. 2021) (Figure 15). The surface of SPOC structures shows distinct basic patches that allow the anchoring of phosphorylated serine residues in protein-protein interactions (Appel et al. 2021; Mikami et al. 2014). PHF3 and DIDO have two of these basic patches, SPEN, RBM15, and RBM15B, do have one basic patch, whereas none is present in the SPOCD1 structure prediction (Appel, Franke, et al. 2023). In fact, no interaction of the SPOCD1 SPOC domain has been reported to date (Appel, Franke, et al. 2023). It is known that the SPOCD1 SPOC domain does not interact with the C-terminal domain (CTD) of Pol

II, likely because of the absence of basic patches in the predicted model (Appel, Franke, et al. 2023).

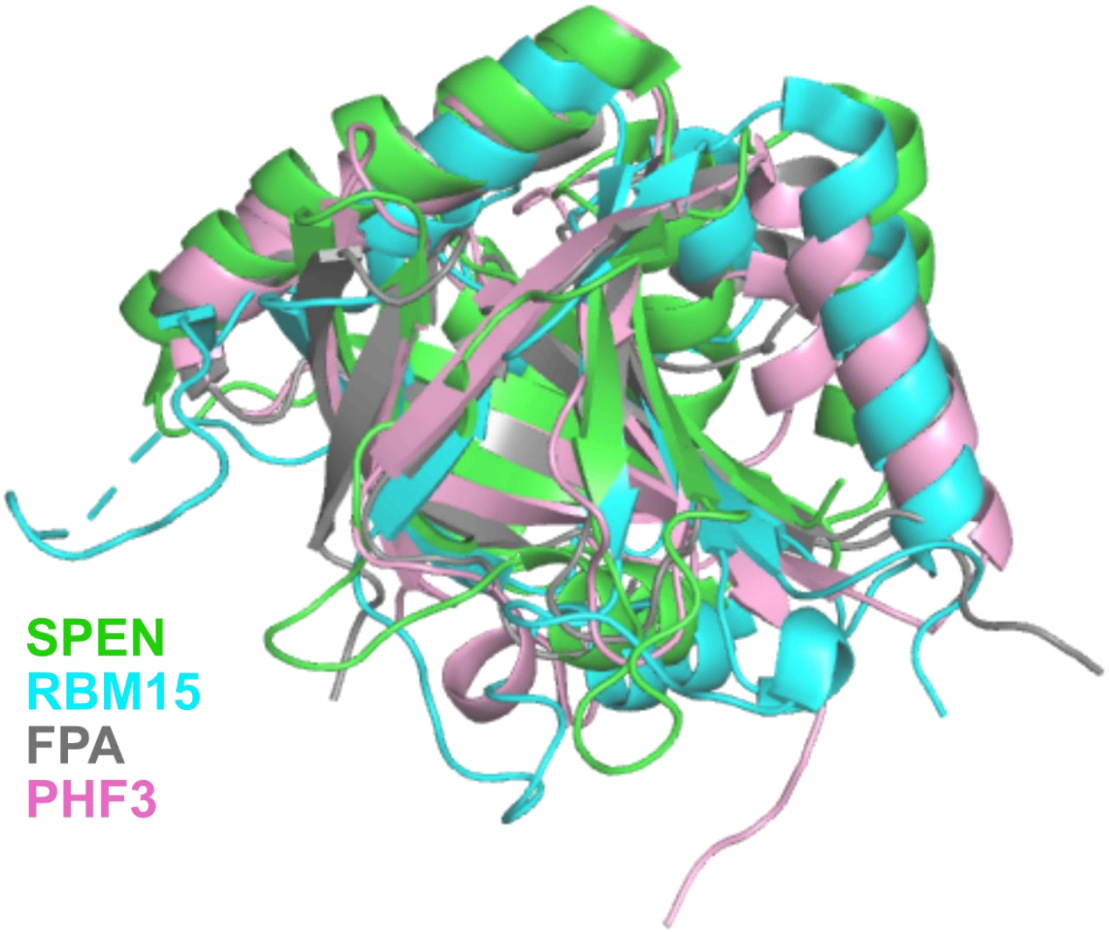

Figure 15: SPOC domain structure.
Overlay of the crystal/NMR structures of the SPOC domains of SPEN (2RT5, green); RBM15 (7Z27, blue); FPA (5KXF, grey) and PHF3 (6Q2V, pink).

Drosophila SPEN is involved in many signalling pathways that modulate gene regulation during development, such as the development of the nervous system and eye (Kolodziej, Jan, and Jan 1995; Jemc and Rebay 2006). Furthermore, SPEN is important for cell remodelling during metamorphosis, and is a key factor in Xist-mediated X-chromosome inactivation (Zhao et al. 2008; Gu et al. 2017; McHugh et al. 2015). *Drosophila*, as well as mammalian SPEN, regulate transcription *via* the Notch pathway (Oswald et al. 2002; Doroquez, Orr-Weaver, and Rebay 2007). SPEN interacts with Nuclear receptor corepressor (NCoR),

Silencing mediator for retinoid and thyroid receptors (SMRT), and RNA Pol II *via* its SPOC domain in a phosphoserine-dependent manner (Shi et al. 2001; Mikami et al. 2014; Appel, Franke, et al. 2023). More precisely, the basic surface of the SPEN SPOC domain recognises the phosphorylated serine residue within the LSD motif of SMRT or NCoR (Mikami et al. 2014; Oswald et al. 2016). Furthermore, the SPEN SPOC domain interacts with the m^6A reader Fragile X mental retardation 1 (FMR1) *via* a phosphorylated LSD motif (Appel, Franke, et al. 2023).

The *Arabidopsis* protein FPA also contains a SPOC domain and regulates flowering time *via* alternative polyadenylation of the Flowering locus C (FLC) Messenger RNA (mRNA) (Hornyik, Terzi, and Simpson 2010). However, the FPA SPOC domain does not appear to be essential for polyadenylation (Zhang et al. 2016). In contrast, the role of the SPOC domain-containing yeast protein Bypass of ess-1 (BYE1) is well known. It is a negative regulator of early transcriptional elongation, likely by SPOC domain-dependent tethering of BYE1 onto the CTD of Pol II (Appel et al. 2021; Appel, Benedum, et al. 2023; Harlen and Churchman 2017).

The homologous RBM15 and RBM15B share 51 % sequence identity and are post-transcriptional regulators that modulate alternative splicing, m^6A modification, and mRNA export (Appel, Benedum, et al. 2023; Lee and Skalnik 2012; Uranishi et al. 2009; Zhang et al. 2015). In each of these cases RBM15/ RBM15B interact in a phosphorylation-dependent manner *via* the SPOC domain with the LPDSD motif of SET domain bifurcated histone lysine methyltransferase 1 (SETDB1), with the m^6A writer complex *via* the LSETD motif in WT1 associated protein (WTAP), or with the nuclear export promoter End-binding 2 (EB2) (Lee and Skalnik 2012; Appel, Benedum, et al. 2023; Gu et al. 2017; Appel, Franke, et al. 2023).

The *Dido1* gene gives rise to three alternative splicing isoforms: DIDO1, DIDO2, and DIDO3 (Appel, Benedum, et al. 2023). DIDO3 is the longest isoform that contains all the domains shown in Figure 13 and is the only isoform containing the complete SPOC domain (Appel, Benedum, et al. 2023). DIDO is important for embryonic stem cell (ESC) gene regulation, differentiation, and self-renewal (Liu, Kim, et al. 2014; Futterer et al. 2017). It has been found to be associated with several cancer types and is an essential host factor for the replication of dengue and Zika viruses in mosquito cells (Xiao et al. 2020; Li et al. 2020; Forghanifard et al. 2020; Braig and Bosserhoff 2013; Caraballo et al. 2022). DIDO3 promotes self-

renewal in ESCs by PHD domain association with H3 lysine 4 trimethylation (H3K4me3), and in a SPOC-dependent manner with the Pol II CTD (Gatchalian et al. 2013; Appel, Franke, et al. 2023). The two phosphor-serine residues of the Pol II CTD anchor on the basic patches on the surface of the SPOC domain (Appel, Franke, et al. 2023; Appel et al. 2021). A similar binding mechanism has been shown for the DIDO paralogue PHF3 which is a transcription regulator that associates with Pol II by binding to pS2 marks in tandem CTD repeats *via* its SPOC domain (Appel et al. 2021). It has been shown that PHF3 competes with TFIIS for binding to Pol II (Appel et al. 2021). Once PHF3 associates with Pol II, thereby displacing TFIIS, the polymerase stalls and cannot continue transcribing (Appel et al. 2021). This mechanism is important for the regulation and differentiation of neuronal genes (Appel et al. 2021). Furthermore, PHF3 is a transcription and mRNA stability regulator and plays a role in tumour suppression (Appel et al. 2021; Fischer et al. 2001).

In conclusion, SPOC domains, apart from SPOCD1, can be described as phosphor-serine binding modules that mediate protein-protein interactions in transcription, RNA processing, and chromatin regulation (Appel, Benedum, et al. 2023).

1.3 Aim of this study

SPOCD1 connects the piRNA pathway with the DNA methylation machinery for the first time, indicating its key role in this pathway (Zoch et al. 2020). While it is known that SPOCD1 is important in piRNA-directed *de novo* DNA methylation, we do not fully understand its function and role (Zoch et al. 2020).

In the present study, we aim to elucidate the role of SPOCD1 in the nuclear piRNA pathway. Using structural, biochemical, and *in vitro* mouse model techniques as a combined approach, we aim to understand the role of SPOCD1. In Chapter 3, we defined three novel interactors of SPOCD1 and solved the crystal structure of one of the SPOCD1 domains. In Chapter 4, we study the association of SPOCD1 with one of its interactors in detail and conclude by proposing a new 2-factor authentication model to form the basis of precision in the nuclear piRNA pathway.

During my PhD studies, I contributed to the publication that discovered SPOCD1 and was published in *Nature* (Zoch et al. 2020). The work in Chapter 3.5 and 3.6,

is part of the manuscript “C19ORF84 connects piRNA and DNA methylation machineries to defend the mammalian germline”**,** which has been submitted for publication in *Molecular Cell*. Furthermore, the manuscript “Two-factor authentication underpins the precision of piRNA-directed LINE1 DNA methylation” which is included in Chapter 4, is currently in revision in *Nature*.

2 Material and methods

2.1 Protein structure prediction with AlphaFold2

The mouse SPOCD1 AlphaFold2 protein structure prediction model (Jumper et al. 2021; Tunyasuvunakool et al. 2021) was downloaded from the AlphaFold Protein Structure Database (https://www.alphafold.ebi.ac.uk/). Models for the C19ORF84, for the SPOCD1-SPIN1 interaction as well as the single SPOCD1 proteins from anolis, xenopus, and coelacanth were generated with AlphaFold2 without template and using mmseqs2 for the multiple sequence alignment generation (Jumper et al. 2021; Tunyasuvunakool et al. 2021) on ColabFold (Mirdita et al. 2022).

2.2 Cloning

To clone the constructs shown in Table 1, several different techniques were used based on the available materials and DNA sequences.

2.2.1 Polymerase chain reaction (PCR)

PCR for cloning was performed using Phusion hot start II DNA-polymerase (Thermo Scientific™ F549L) according to the manufacturer's instructions.

2.2.2 PCR and gel cleanup

PCR cleanup was performed using the Qiagen QIAquick PCR Purification Kit (Qiagen, 28104) or the MinElute PCR purification kit (Qiagen, 28004) for lower concentrations. In both cases, all steps were performed according to the manual. For DNA extraction from agarose gel, the QIAquick Gel extraction kit (Qiagen,

28704) was used according to the manufacturer's instructions. For very weak bands, the MinElute Gel extraction kit (Qiagen, 28604) was used to elute in smaller volumes.

2.2.3 Restriction digest and ligation

For restriction digestion, the appropriate NEB enzyme with compatible buffer was used to digest the backbone (1 µg) and insert (100 – 1000 ng) for 1-2 hours at digestion temperature. Backbones were additionally dephosphorylated with Alkaline phosphatase calf intestinal (CIP, NEB, M0525S), according to the manufacturer's protocol. For the ligation of DNA fragments, T4 DNA ligase (NEB, M0202) was used. Usually, a backbone : insert ratio of 1:3 is used, and the ligation mix is incubated for 30 – 60 minutes at room temperature.

2.2.4 Gibson assembly and ligase independent cloning (LIC)

Some constructs were generated by Gibson assembly (Gibson et al. 2009) using the NEB Gibson assembly master mix (NEB, E2611). Ligase-independent cloning (LIC) has been used for some bacterial expression vectors (Stols et al. 2002; Aslanidis and de Jong 1990).

2.2.5 Transformation into *E.coli*

50 µl $CaCl_2$ competent XL-1 *E.coli* were thawed on ice and 2-10 µl of the ligation mix was added. The cells were incubated for 30 minutes on ice. After heat shock at 42 °C for 45 seconds the cells were kept on ice for 10 minutes. Super optimal broth (SOC) medium (200 µl) was added, and the mixture was incubated for 30-60 minutes at 37 °C. The cells were plated on LB plates containing the respective antibiotics and incubated overnight at 37 °C.

2.2.6 Plasmid preparation from *E.coli*

Single colonies were picked and incubated in 4 ml LB media (5 g/l yeast extract, 10 g/l tryptone, 10 g/l NaCl, pH 7.0) with the appropriate antibiotic in a 14 ml tube for 12-16 hours at 37 °C and 180 rpm. The culture was pelleted, and plasmid isolation was performed according to the manufacturer's instructions using the QIAprep spin miniprep kit (Qiagen, 27104). The plasmid was eluted in 30-50 µl, control digested with the appropriate restriction enzymes (NEB) and loaded on a 1-2% agarose gel. If the fragment size was as expected, the plasmid was sent for sequencing using Source Bioscience. For sequencing analysis, ApE (A plasmid editor, http://jorgensen.biology.utah.edu/wayned/ape/) was used.

2.3 Plasmids used in this study

Table 1: Plasmids used in this study.

HEK293T cells			
Plasmid	**Encoding**	**Origin**	**amount used per transfection in HEK293T cells**
pcDNA3.1-GFP-HA	Carboxy-terminal HA epitope-tagged eGFP	O'Carroll laboratory stock	350 µg
pcDNA3.1-mmSPOCD1-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1	Zoch et al. *Nature* 2020	1000 µg
pcDNA3.1-hsSPOCD1-HA	Carboxy-terminal HA epitope-tagged human SPOCD1	GenScript clone ID OHu28102	1000 µg

pcDNA3.1-mmSPOCD1-L792R-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 with L792R point mutation	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-hsSPOCD1-L971R-HA	Carboxy-terminal HA epitope-tagged human SPOCD1 with L971R point mutation	Mutation of pcDNA3.1-hsSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-ΔTFIIS-M-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 410-519	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-ΔSPOC-M-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 680-824	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-1-HA	Carboxy-terminal HA epitope-tagged containing mouse SPOCD1 amino acid 1-409	Mutation of pcDNA3.1-mmSPOCD1-HA	400 µg
pcDNA3.1-mmSPOCD1-fragment-1b-HA	Carboxy-terminal HA epitope-tagged containing mouse SPOCD1 amino acid 205-409	Mutation of pcDNA3.1-mmSPOCD1-HA	350 µg
pcDNA3.1-mmSPOCD1-fragment-2-HA	Carboxy-terminal HA epitope-tagged containing mouse SPOCD1 amino acid 410-679	Mutation of pcDNA3.1-mmSPOCD1-HA	370 µg
pcDNA3.1-mmSPOCD1-fragment-3-HA	Carboxy-terminal HA epitope-tagged containing mouse SPOCD1 amino acid 680-1015	Mutation of pcDNA3.1-mmSPOCD1-HA	750 µg

pcDNA3.1-mmSPOCD1-fragment-4-HA	Carboxy-terminal HA epitope-tagged containing mouse SPOCD1 amino acid 830-1015	Mutation of pcDNA3.1-mmSPOCD1-HA	700 µg
pcDNA3.1-mmSPOCD1-fragment-Δ831-1015-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 831-1015	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ831-922-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 831-922	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ923-1015-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 932-1015	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ923-969-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 923-969	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ970-1015-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 970-1015	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δhelix-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 942-964	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ308-317-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 308-317	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg

pcDNA3.1-mmSPOCD1-fragment-Δ318-327-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 318-327	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ328-337-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 328-337	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ338-347-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 338-347	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ348-357-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 348-357	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ358-367-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 358-367	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ368-387-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 368-387	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-mmSPOCD1-fragment-Δ388-409-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 lacking amino acid 388-409	Mutation of pcDNA3.1-mmSPOCD1-HA	1000 µg
pcDNA3.1-SV40-NLS-HA-HIS-GST-GFP-mmSPOCD1-327-346-HA	eGFP fused carboxy-terminal with SPOCD1 amino acid 327-346 with amino-terminal SV40-NLS-HA-HIS-GST tag	Mutation of pcDNA3.1-mmSPOCD1-HA	820 µg

pcDNA3.1-mmSPOCD1-8alanine-HA	Carboxy-terminal HA epitope-tagged mouse SPOCD1 with amino acid 338-347 mutated to alanine	Mutation of pcDNA3.1-SV40-NLS-HA-HIS-GST-GFP	1000 µg
pcDNA3.1-mmC19ORF84-Flag	Carboxy-terminal FLAG epitope-tagged mouse C19ORF84	synthesized by GenScript	300 µg for IP-WB, 500 µg for IP-MS
pcDNA3.1-mmC19ORF84-Δ61-81-Flag	Carboxy-terminal FLAG epitope-tagged mouse C19ORF84 lacking amino acid 61-81	Mutation of pcDNA3.1-mmC19ORF84-Flag	300 µg
pcDNA3.1-mmC19ORF84-Δ61-90-Flag	Carboxy-terminal FLAG epitope-tagged mouse C19ORF84 lacking amino acid 61-90	Mutation of pcDNA3.1-mmC19ORF84-Flag	300 µg
pcDNA3.1-mmSPIN1-Flag	Carboxy-terminal FLAG epitope-tagged mouse SPIN1	GenScript clone ID OMu23362	725 µg

Recombinant protein

Plasmid	**Encoding**	**Origin**	**amount used per transfection in HEK293T cells**

pET-HIS-mmSPIN1-49-262	Amino-terminal HIS epitope-tagged mouse SPIN1 amino acids 49-262	Cloning of amino acid 49-262 (from pcDNA3.1-mmSPIN1-Flag) in pET-HIS	NA
pET-HIS-GST-mmSPOCD1-F1b	Amino-terminal HIS-GST epitope-tagged mouse SPOCD1 amino acids 205-409	Cloning of amino acid 205-409 (from pcDNA3.1-mmSPOCD1-HA) in pET-HIS	NA
pET-HIS-GST-anolis-SPOCD1-F1b	Amino-terminal HIS-GST epitope-tagged anolis SPOCD1 amino acids	amino acids 457-748 of codon optimized XM_008117905.1 cloned in pET-HIS-GST	NA
pET-HIS-GST-xenopus-SPOCD1-F1b	Amino-terminal HIS-GST epitope-tagged xenopus SPOCD1 amino acids	IDT genblock (amino acid 1-229 of XM_018091875) cloned in pET-HIS-GST	NA
pET-HIS-GST-coelacanth-SPOCD1-F1b	Amino-terminal HIS-GST epitope-tagged coelacanth SPOCD1 amino acids	IDT genblock (amino acid 510-1009 of XM_014492850.1) cloned in pET-HIS-GST	NA
pET-HIS-GST-mmSPOCD1-8alanine-F1b	Amino-terminal HIS-GST epitope-tagged mouse SPOCD1 amino acids 205-409 with amino acid 338-347 mutated to alanine	Mutation of pET-HIS-GST-mmSPOCD1-F1b	NA

pET-HIS-GST-mmSPOCD1-SPOC	Amino-terminal HIS-GST epitope-tagged mouse SPOCD1 SPOC domain	Cloning of amino acid 687-830 (from pcDNA3.1-mmSPOCD1-HA) in pET-HIS	NA
pET-HIS-GST-mmSPOCD1-SPOC-L792R	Amino-terminal HIS-GST epitope-tagged mouse SPOCD1 SPOC domain with L792R point mutation	Mutation of pET-HIS-GST-mmSPOCD1-SPOC	NA

2.4 Cell lines and maintenance

Human embryonic kidney 293T (HEK293T) cells were obtained from the O'Carroll lab stock, University of Edinburgh, and were regularly tested for mycoplasma contamination. The cells were cultured in Glasgow minimum essential medium (GMEM, Gibco, 11710035) that was supplemented with 10% (v/v) Foetal bovine serum (FBS, Gibco, 10500064), 2 mM L-glutamine (Gibco, 25030024), and 1 mM Sodium Pyruvate (Thermo Scientific, 11360039) and grown at 37 °C, 5% CO_2. Cells were grown until ~90% confluency in 10 cm dishes and split every 2 - 4 days. For splitting, the old media was removed, cells were incubated for 2 minutes in trypsin-EDTA solution (Sigma-Aldrich, SM-2003-C) and resuspended in an additional 8 ml of fresh media. The cells were then split in a 1:5 ratio or used for seeding.

For freezing, HEK293T cells were grown to ~90% confluency in a 10 cm dish. The media was carefully removed, and 2 ml trypsin-EDTA solution (Sigma-Aldrich, SM-2003-C) was added. 8 ml of fresh media was added, and the cell suspension was resuspended, transferred to a 15 ml falcon, and spun at 200 g for 5 minutes. The supernatant was discarded, and the pellet was resuspended in 2 ml Cell Culture Freezing Medium (Gibco, 12648010). One millilitre of the cell suspension was aliquoted into cryotubes and frozen at -80 °C.

2.5 HEK293T cell transfection

On day 0, the cells were detached with trypsin-EDTA (Sigma-Aldrich, SM-2003-C) for 2 minutes, resuspended in fresh media, and counted. 5×10^5 HEK293T cells were seeded in 6-well plates containing 2 ml of medium. 24 hours latern on day 1, the medium was changed, and cells were transfected with jetPRIME (Polyplus, 101000046). Briefly, up to 2 µg of plasmid (exact amount in Table 1) was mixed with 3 µl jetPRIME reagent in 200 µl jetPRIME buffer. The mixture was vortexed, incubated for 10 minutes at room temperature, and carefully added to the cells in fresh medium. The cells were incubated for another 2 days and then harvested for immunoprecipitation (IP).

2.6 Immunoprecipitation (IP)

On day 3 of the transfection protocol (2 days post-transfection), the cells were harvested. First, the cells were removed from the incubator, placed on ice, and washed twice with ice-cold phosphate-buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4). After the last washing step, cells were resuspended in 1 ml lysis buffer (IP buffer: 150 mM KCl, 2.5 mM $MgCl_2$, 0.5% (v/v) Triton X-100, 50 mM Tris-HCl pH 8.0, supplemented with 1× cOmplete ULTRA EDTA-free protease inhibitors (Roche, 05892791001) and 37 U/ml Benzonase (Millipore, 71205)) and incubated for 30 minutes on the rotating wheel at 4 °C. The lysate was cleared by centrifugation for 10 minutes at 21,000 x g. Anti-HA magnetic beads (Pierce, 88837) were calibrated in lysis buffer and aliquoted to obtain an equivalent volume of 20 µl beads in each tube. 800 µl of cleared lysate was added to the beads and incubated for 1 hour at 4 °C on a rotating wheel. The beads were washed four times with 1 ml of lysis buffer and resuspended in 20 µl of elution buffer (0.1% (w/v) sodium dodecyl sulphate (SDS), 50 mM Tris-HCl pH 8.0) after the last washing step. The protein was eluted by incubating the beads in elution buffer at 50 °C for 10 minutes. Finally, the eluate was transferred to a fresh tube and stored at -20 °C until further processing.

2.7 Western blot (WB)

For the Western blot (WB), 4 µl of 6x Lämmli buffer (15% (w/v) SDS, 0.075% (w/v) bromophenol blue, 50% (v/v) glycerol, 75 mM Tris-HCl pH 7.0, 0.75 M Dithiothreitol (DTT, added fresh just before usage)) was added to 20 µl of the lysate and eluate each. The samples were loaded onto a NuPage 4 – 12% Bis-Tris protein gel (Invitrogen, NP0323BOX) and separated for 40 minutes at 200 V. Proteins were blotted onto an Amersham Protran NC nitrocellulose membrane (Cytiva, 10600013) *via* wet transfer for 1 - 2 hours at 30 V. After Ponceau staining, the membrane was blocked for 45 - 60 minutes in blocking buffer (4% (w/v) skimmed milk powder (Sigma-Aldrich, 70166-500G) in PBS-T (PBS with 0.1% (v/v) Tween-20)). The membrane was then incubated with the primary antibody (anti-HA, anti-FLAG, anti-SPOCD1, or anti-TUBULIN, details in Chapter 2.29) in blocking buffer for 1 hour at room temperature or overnight at 4 °C. After three 10 minutes washes with PBS-T, the membrane was incubated for 1 hour with the appropriate secondary antibody (IRDye 680RD donkey anti-rabbit or IRDye 800CW donkey anti-mouse details in Chapter 2.29) diluted in blocking buffer. The membrane was again washed three times with PBS-T and imaged on an LI-COR Odyssey CLx system. The images were analysed using Image Studio Lite (LI-COR), and areas of interest were cropped for presentation.

2.8 Mouse strains and experimentation

Mice were maintained at the University of Edinburgh, UK, in accordance with the regulations of the UK Home Office. Animals were kept at 19 - 23°C and 30 - 70% humidity in open-top cages with 12 hours light/dark cycle (7 am - 7 pm/ 7 pm - 7 am).

2.8.1 Mice used in this study

Spocd1-HA (B6CBAF1/Crl;C57BL/6N;Hsd:ICR (CD1)): male embryo E14.5 and E16.5

Miwi2-tdTomato (B6CBAF1/Crl;C57BL/6N;Hsd:ICR (CD1)): male embryo E16.5

Spocd1-dSpin1 (B6CBAF1/Crl;C57BL/6N): male embryo E16.5, male postnatal day 14, male postnatal day 20, male and female 8-16 weeks
Hsd:ICR (CD1): male embryo E14.5, male and female 8-16 weeks

2.8.2 Allele generation

The *Spocd1*^{HA} and *Miwi2*^{tdTomato} (*Miwi2*^{tdTom}) mouse alleles have been previously described. The *Miwi2*^{tdTom} allele is a *Miwi2* null allele and is used as such. Both lines were maintained on a mixed B6CBAF1/Crl;C57BL/6N;Hsd:ICR (CD1) genetic background.
The *Spocd1*$^{\Delta SPIN1}$ allele was generated by CRISPR-Cas9 gene editing, as previously described (Wang et al. 2013; Yang et al. 2013). The single guide RNA (sgRNA) (GGGTCAGGAATCAGGCTTGT) was denatured at 72°C for 2 minutes and the single-stranded DNA (ssDNA) oligo containing the 8-alanine mutation flanked by an 85 bp homology arm on each side (AGATGGTAAACAGTTGAAGCCAAGGCAGGGAGGATTTCAGGCAGAGCCTT GCCATACTCTCTCTCAGCAGGTCTACACTGGGTCAGCTGCCGCAGCGGCCG CTGCCGCCGCTGCAAGTCAGCCAGGACAAATTGAACCTCTGGAGGAGTTGG ACACCAACTCAGCCAGAAGGAAGAGAAGGCCCACAACTGCTCACCCTA) was heated to 95°C for 5 minutes for denaturation.
The injection mix containing sgRNA (6.5 ng/ul, homemade in the lab), ssDNA oligo (12 ng/ul, IDT Ultramer) and Cas9 mRNA (25 ng/ul, homemade in the lab) was prepared and injected into the cytoplasm of fertilised 1-cell zygotes (B6CBA F1/Crl).
F_0 offspring were screened by PCR and the *Spocd1*$^{\Delta SPIN1}$ allele was confirmed by Sanger sequencing. The allele was established from one founder animal and backcrossed several times to a C57BL/6N genetic background. Thus, the *Spocd1*$^{\Delta SPIN1}$ mice were on a mixed B6CBAF1/Crl;C57BL/6N genetic background.

2.8.3 DNA isolation

Mouse ear biopsies or tails were incubated overnight at 55°C in 500 µl TAIL buffer (50 mM Tris-HCl pH 8.0, 100 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0,

100 mM NaCl, 1% (w/v) SDS) supplemented with 0.5 mg/ml Proteinase K (Roche, 3115852001). The mixture was centrifuged for 10 minutes at 21,000 × g. The supernatant (approximately 450 µl) was transferred to a fresh tube and one volume of isopropanol was added. The samples were centrifuged for 30 min at 21,000 × g, and the supernatant was discarded. The DNA pellet was washed with 70% ethanol (EtOH), dried at room temperature, and resuspended in 50-100 µl of PCR-grade water. Next, 1 µl of the DNA was used for the genotyping PCR.

2.8.4 Genotyping PCRs

Spocd1$^{\Delta SPIN1}$ animals were genotyped by PCR using four primers (F: GACCCTGTATTTATTGAAGTCACTG, R: CCTCAGTGACATCAGGCGGA, WT-F: CACTGGGTCAGGAATCAGGC, ΔSpin-R: GTCCTGGCTGACTTGCAGC). The wild-type allele generated bands at 331 and 216 bp, and the ΔSpin1 allele resulted in bands at 331 and 166 bp. The composition of the 25 µl PCR reaction was as follows:

2.500 µl 10x DreamTaq Green Buffer (B71, Life technologies)
0.250 µl 10 mM dNTPs (733-1854, VWR)
0.025 µl 100 mM primer each
0.250 µl Taq polymerase (EMBL Heidelberg Facility)
1.000 µl DNA from ears/ tails
21.90 µl water

The reaction was mixed in a PCR tube and run on a thermocycler according to the protocol in Table 2.
The PCR reaction was loaded on a 4% (w/v) agarose gel (A9539, Sigma-Aldrich) containing 4% (v/v) Ethidium bromide (EtBr, 09-0617, Sigma-Aldrich) and the bands were resolved by running at 180 V for 20 - 40 minutes in 1x TAE buffer (0.4 M Tris-Acetate pH8.3, 0.01 M EDTA) (Figure 16)

Table 2: Genotyping PCR setting for *Spocd1*$^{\Delta SPIN1}$ animals.

Time (seconds)	Temperature (in °C)	Step
120	95	1
15	95	2
30	61.5	3
13	72	4
Go to 2 and repeat step 2-4 35x		
120	72	5
infinite	20	6

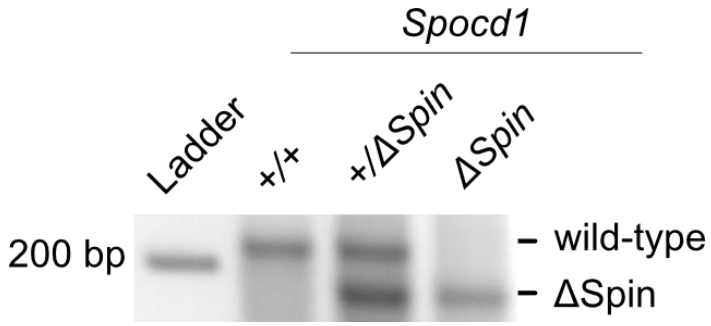

Figure 16: Genotyping of *Spocd1*$^{\Delta SPIN1}$ mice.
Representative image of genotyping of *Spocd1*$^{+/+}$, *Spocd1*$^{\Delta SPIN1/+}$ and *Spocd1*$^{\Delta SPIN1/\Delta SPIN1}$. The wild-type band at 216 bp and ΔSPIN1 band at 166 bp are shown.

The *Spocd1*^{HA} animals were genotyped using two primers (Spocd1^{HA}-Fw: CCCCATCCACTGTAGTATCTGC, Spocd1^{HA}-R: ATACAAACTAGGCAGGTGTGGG). The *Spocd1*^{HA} allele generates a 238 bp band, whereas the *wild-type* results in a 196 bp amplicon. The PCR settings are listed in Table 3.

The *Miwi2*^{tdTomato} mice were genotyped with a set of 3 primers (Miwi2-TOM_GenoF1: TACTCCCAAACTCCGAGTCAC, Miwi2-TOM_GenoR1: TCTTGCGTTTCTGATAGGCAC, Miwi2_Gen-R2: AAAAGGCACTCTGGCTAGGAG). The *Miwi2*^{tdTomato} mouse generates a 193 bp band while the *wild-type* allele results in a 262 bp band (Figure 17). The PCR settings are listed in Table 4.

Table 3: Genotyping PCR setting for *Spocd1*^{HA} animals.

Time (seconds)	Temperature (in °C)	Step
180	90	1
30	90	2
45	61.5	3
30	72	4
Go to 2 and repeat step 2-4 35x		
300	72	5
infinite	4	6

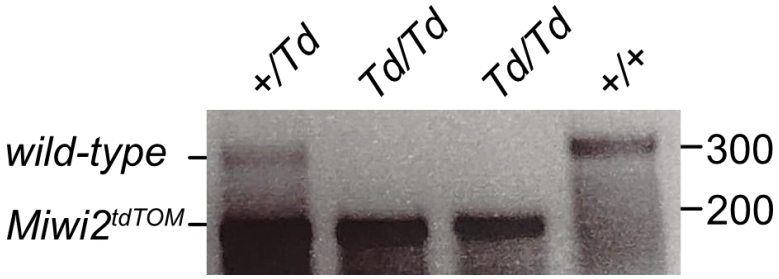

Figure 17: Genotyping of *Miwi2*^{tdTomato} mice.
Representative image of genotyping of *Miwi2*$^{+/tdTomato}$, *Miwi2*$^{tdTomato/tdTomato}$, and *Miwi2*$^{+/+}$. Wild-type band at 262 bp, tdTomato band at 193 bp.

Table 4: Genotyping PCR setting for *Miwi2*^{tdTom} animals.

Time (seconds)	Temperature (in °C)	Step
180	94	1
30	94	2
30	62	3
40	72	4
Go to 2 and repeat step 2-4 35x		
300	72	5
infinite	20	6

2.9 Foetal testes collection

For foetal testis collection, mice were mated for four days and checked for plugs every day. Once a plug was detected, females were separated from males, and the plug day was counted as E0.5. 14 (for E14.5) days or 16 days (for E16.5) after the detected plug the pregnant female was killed, and testes were collected from all male embryos. Depending on further processing, the testes were flash-frozen in liquid nitrogen or embedded in OCT (Cell Path, KMA-0100-00A) and then frozen in liquid nitrogen.

2.10 Immunoprecipitation coupled with mass spectrometry (IP-MS)

For Immunoprecipitation coupled with mass spectrometry (IP-MS), the IP conditions differ between HEK293T cells and foetal testes, but the STAGE tips and Mass spectrometry (MS) analysis were the same for both.

2.10.1 HEK293T cell immunoprecipitation (IP)

The transfection of HEK293T cells for MS was performed as described in Chapter 2.5. Before harvesting, the cells were washed three times with ice-cold PBS, and 1 ml hypotonic lysis buffer (10 mM KCl, 10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 0.1% (v/v) IGEPAL CA-630 with 1× cOmplete ULTRA EDTA-free protease inhibitors (Roche, 05892791001)) was added to the wells. The cells were resuspended, transferred to a glass cell homogeniser, and lysed with 20 strokes. 50 U/ml Benzonase (Milipore, E8263-5KU) was added to each sample, and the tubes were incubated for 30 minutes at 4°C on a rotating wheel. The lysate was cleared for 10 minutes at 21,000 × g, and 900 µl of the cleared lysate was added to 50 µl of anti-HA magnetic beads (Pierce, 88837, additionally crosslinked with 20 mM dimethyl pimelidate in borate buffer, pH 9.0, and equilibrated in lysis buffer). The tubes were incubated for 1 hour at 4°C on a rotating wheel. Afterwards, beads were washed 5 times with hypotonic wash buffer (100 mM KCl, 50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 0.1% (v/v) IGEPAL CA-630) and eluted for 15 minutes at 50 °C in elution buffer (0.1% (w/v) Rapigest (Waters, 186001861) in 50 mM Tris-HCl, pH 8.0)

2.10.2 Foetal testis immunoprecipitation (IP)

For the foetal testis IP-MS, testes were collected from E16.5 embryos and snap frozen in liquid nitrogen. Per replicate, 25 or 50 testes were lysed and homogenised in hypotonic lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM KCl, 5 mM $MgCl_2$, 0.1% (v/v) IGEPAL CA-630 with 1x cOmplete ULTRA EDTA-free protease inhibitors (Roche, 05892791001) with the help of 20 strokes in a glass cell homogeniser. 50 U/ml Benzonase (Milipore, E8263-5KU) was added to each sample, and the tubes were incubated for 30 minutes at 4°C on a rotating wheel. The lysate was cleared for 10 minutes at 21,000 × g, and 900 µl of the cleared lysate was added to 50 µl of anti-HA magnetic beads (Pierce, 88837, additionally crosslinked with 20 mM dimethyl pimelidate in borate buffer, pH 9.0, and equilibrated in lysis buffer). The tubes were incubated for 30 minutes at 4°C on a rotating wheel. Afterwards, beads were washed 4 times with hypotonic wash buffer (100 mM KCl, 50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 0.1% (v/v) IGEPAL CA-630) and eluted for 15 minutes at 50°C in elution buffer (0.1% (w/v) Rapigest (Waters, 186001861) in 50 mM Tris-HCl pH 8.0).

2.10.3 STAGE tips

After the IP, the eluates were prepared for STAGE tips. An appropriate amount of urea (Sigma, U5378-100g) was added to each eluate to obtain a final concentration of 8 M and mixed until dissolved. The sample was spun down at 21,000 x g, transferred to a Vivacon 500 spin column 30 K cartridge (Sartorius, VN01H22), and centrifuged for 15 minutes. All centrifugation steps were performed at 13,800 g. 100 µl of 0.05 M Iodacetamide (IAA, Sigma, I1149-5g) in 8 M urea, 0.1 M Tris-HCl, pH 8.2 was added to the column, mixed at 600 rpm on the shaker for 1 minute at room temperature and incubated another 20 minutes at room temperature in the dark. After centrifugation for 10 minutes, 100 µl of 8 M urea in 0.1 M Tris-HCl, pH 8.2 was added to the column and centrifuged for an additional 15 minutes. This was followed by two washes with Ammonium bicarbonate (ABC) buffer (NH_4HCO_3 Sigma, A6141-500G-D) in ultrapure water) of 50 µl and 100 µl with 15 minutes centrifugation each. The old collection tubes were then removed and replaced with new ones. For trypsin digestion, 200 ng trypsin (Thermo Fisher, 90057) in 100 µl

0.05 M ABC buffer was added to the spin column, wrapped with parafilm, and incubated at 37 °C overnight after a brief mixing for 1 minute at 800 rpm on a shaker. The next day, the column was centrifuged for 15 minutes and 100 µl of ABC buffer was added. The tube was mixed at 800 rpm for 1 minute on a shaker, incubated for a further 10 minutes and then centrifuged for 15 minutes. The flow-through was transferred to a new tube, and the trypsin digest was stopped with 15 µl of 10% (v/v) Trifluoroacetic acid (TFA).

Now, the STAGE tips were prepared by inserting three layers of the C18 matrix into 200 µl tips. The STAGE tips were activated by adding 50 µl methanol to the STAGE tips and centrifuge. All centrifugation steps were performed at 1000 g for 3 - 5 minutes until all liquid passed through the matrix. Next, 50 µl of 80% (v/v) acetonitrile (ACN) in 0.1% (v/v) TFA was added, centrifuged, and washed with 50 µl 0.1% (v/v) TFA. The sample was loaded onto the stage tip and centrifuged, followed by a final wash with 70 µl of 0.1% (v/v) TFA. The STAGE tips were stored at -20°C until elution for mass-spectrometry analysis.

2.10.4 Mass-spectrometry analysis

After eluting peptides from the STAGE tips (Rappsilber, Ishihama, and Mann 2003), they were loaded on an ultra-high resolution nano-flow liquid chromatography nanoLC Ultimate 3000 unit fitted with an Easyspray column (50 cm, 2 µm particles) coupled to a high-resolution/ accurate-mass mass spectrometer Orbitrap Fusion Lumos operating in data acquisition (DDA) mode (Thermo Fisher Scientific). Separation was performed with a 2% - 40% - 95% 190 minutes gradient using mobile phase A (0.1% (v/v) aqueous formic acid) and mobile phase B (80% (v/v) acetonitrile in 0.1% (v/v) formic acid). The parameters for the MS acquisition were set as follows: cycle time of 3 seconds, MS1 scan Orbitrap resolution at 120,000, RF lens to 30%, AGC target to 4×10^{-5}, and maximum injection time to 50 milliseconds with the detected intensity threshold at 5×10^{-3}. For the MS2 scan, an Ion Trap with rapid scan settings was used. The maximum injection time was 50 milliseconds with the AGC target set to 2×10^{-4}. This setup has been used in large proteome and interactome screens and allows detection in the low attomole (10^{-18}) range (Hein et al. 2015; Richards et al. 2015). The raw data were processed using MaxQuant version 1.6.1.0, and the MaxQuant

label-free quantitation (LFQ) algorithm was used for LFQ (Cox et al. 2014). Common contaminants (e.g. trypsin and keratins) were removed during Perseus analysis with the peptides searched against the mouse UniProt database for testis samples and against the human UniProt database with specific mouse sequences added for HEK293T cell samples (Tyanova et al. 2016; Cox et al. 2014; The UniProt 2017). The LFQ intensities were imported into Perseus version 1.6.0.2 (Tyanova et al. 2016) for visualization and processed as described previously (Hubner and Mann 2011). Data for the IP-MS experiment were deposited in ProteomeXchange under the accession number PXD041214.

2.11 Multiple sequence alignment

Multiple sequence alignments of SPOCD1, C19ORF84, and SPIN1 were generated using ClustalW (Sievers et al. 2011) and edited using Jalview (Waterhouse et al. 2009). The alignment of the SPOC domains from SPOCD1, PHF3, DIDO1, SPEN, and RBM15 was generated with the following sequences: B1ASB6, Q6ZMY3, XP_008116112.1, XP_031752218.1, JH127468.5, B2RQG2, Q92576, H9GF02, A0A8J0R465, H3A2H7, B8A483, Q8C9B9, Q9BTC0, G1KE55, A0A1B8XZM4, H3BAT1, F1QQA3, Q96T58, Q62504, XP_016853340.1, A0A803J536, H3A1Y7, XP_003201252.2, Q0VBL3, Q96T37, A0A803TZ95, F6QUY1, H3BG89, F1RCY7. For the SPOCD1 alignment SPOCD1 sequences from different species with the following identifiers were used: B1ASB6, Q6ZMY3, F7FFW6, F1SV96, XP_008116112.1, F1MG39, A0A151MMW3, H0XFY4, A0A485PAI7, W5NRM3, H2R1B9, XP_023490296.1, XP_049473573.1, XP_044853461.1, XP_040829971.1, XP_039185503.1, XP_034976513.1, XP_036112709.1, XP_030727203.1, XP_032245124.1, XP_026965126.1, XP_031752218.1, XP_025303405.3, JH127468.5, XP_007493079.1. Additionally, all SPOCD1 alignments were edited based on the SPOC crystal structures or secondary structure elements of the AlphaFold2 model (B1ASB6) using Jalview (Waterhouse et al. 2009). The C19 alignment was generated with the following sequences: NP_001357781.1, NP_001180552.1, XP_008115509.1. For the SPIN1 alignment the following sequences were used: NP_001269957.1, NP_006708.2, XP_003216493.1, NP_001182299.1, XP_005200980.1, XP_027819975.1, XP_025058450.1, XP_028912020.1, XP_023372017.1,

XP_032451534.1, JAA37469.1, XP_023483342.1, XP_049495778.1, XP_044877341.1, XP_040819636.1, XP_039217345.1, XP_034955689.1, XP_036117258.1, XP_030688348.1, XP_032283569.1, XP_026947908.1, NP_001017097.1, XP_048949241.1, XP_001331783.2, AMEX60DD201042989.3, BAB59130.1, XP_035642328.1, XP_052634446.1, XP_006001418.1, XP_007497879.1.

2.12 Protein purification

The plasmids used for the protein purification are listed in
Table 1. All constructs were expressed in *E. coli* BL21 (DE3) cells and re-transformed freshly for purification. A few colonies were picked for 25 ml pre-culture in 2xTY media (16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl) supplemented with kanamycin and grown at 37 °C overnight. The following day, 1 ml of pre-culture was used to inoculate 250 ml of 2xTY media supplemented with kanamycin (bacteria were grown in a total of 2 L for each purification) and grown at 37°C until OD_{600} = 0.8. The temperature was reduced to 18°C and Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1mM. Bacteria were harvested after 16 - 18 hours by pelleting at 4,000 × g for 20 minutes. The pellets were frozen in the -80°C freezer and stored until purification.
Proteins were purified in 1 - 3 different steps as specified in Table 5 for each construct. Table 6 lists all buffers that were used. First, the pellet was resuspended in lysis buffer and lysed with the Constant systems 1.1 kW TS cell disruptor at 25 kpsi. The lysate was then cleared at 50,000 × g for 45 minutes and the supernatant was filtered through a 0.22 µm syringe filter.
For Glutathione S-transferase (GST) tagged proteins, the filtered supernatant was loaded onto 5 - 7 ml of Glutathione Sepharose High Performance beads (Cytica, 17-5279-02) equilibrated in wash buffer. After incubating for 2 hours rotating at 4 °C, the supernatant was discarded, and the beads were washed three times with wash buffer. The protein was eluted with 5 × 10 ml of elution buffer.
For Polyhistidine (HIS) tagged proteins, the filtered supernatant was loaded onto an equilibrated cOmplete His-Tag Purification Column (Roche, 6781535001) attached to an Äkta Start system to perform Metal ion affinity chromatography

(IMAC). The column was washed with five column volumes and eluted with an increasing imidazole gradient over 10 – 20 column volumes.

All fractions containing the protein of interest based on Coomassie staining of SDS-page were pooled and prepared for dialysis overnight using SnakeSkin Dialysis Tubing (Thermo Fisher, 68700 or 88244). If the tag needed to be removed, TEV or 3C protease (homemade from Atlanta Cook's lab) was added during overnight dialysis. If the protein was clean enough, it was concentrated (Sartorius Vivacon or Pierce PES protein concentrator was appropriate molecular weight cut-off), flash frozen in liquid nitrogen, and stored in the -80 °C freezer.

If further purification was required, the protein was loaded onto an ion exchange column (IEX, see Table 5), washed, and eluted with a gradient of NaCl (100 – 1000 mM, 0 – 100% buffer B). The fractions of interest were pooled, concentrated, and further purified using size exclusion chromatography (SEC). Depending on the size of the protein and contamination, a Superdex S200 or S75 was used. Fractions containing the desired protein were pooled, concentrated and frozen as described above.

Table 5: Overview of recombinant protein purification steps.

Tag	Organism	Protein	Purification step 1	Tag removal?	Purification step 2	Purification step 3
GST	mouse	SPOC	GST Pull-down	yes	IEX: Resource S	SEC
GST	mouse	SPOC L792R	GST Pull-down	yes	IEX: Resource S	
HIS	mouse	SPIN1 49-262	IMAC	yes	IEX: Resource Q	SEC
GST	mouse	SPOCD1 F1b	GST Pull-down	no		
GST	mouse	SPOCD1 F1b - 8 alanine	GST Pull-down	no		
GST	xenopus	SPOCD1 F1b	GST Pull-down	no		
GST	anolis	SPOCD1 F1b	GST Pull-down	no		
GST	coelacanth	SPOCD1 F1b	GST Pull-down	no		

Table 6: Buffer used for protein purification.

Name of buffer	Composition for HIS-tag	Composition for GST-tag
	Purification Step 1 - IMAC	
Lysis buffer	- 20 mM Tris-HCl or HEPES pH 7.5 - 200 mM NaCl - 2.5 mM imidazole - 0.5 mM beta-mercaptoethanol - cOmplete ULTRA EDTA-free protease inhibitors (Roche, 05892791001) - 0.01 g/l DNaseI (Sigma, D4527) - 2 mM Pefabloc (Supelco, 76307)	- 20 mM Tris-HCl or HEPES pH 7.5 - 150 mM NaCl - 1 mM dithiothreitol (DTT) - cOmplete ULTRA EDTA-free protease inhibitors (Roche, 05892791001) - 0.01 g/l DNaseI (Sigma, D4527) - 2 mM Pefabloc (Supelco, 76307)
Wash buffer	- 20 mM Tris-HCl or HEPES pH 7.5 - 200 mM NaCl - 2.5 mM imidazole - 0.5 mM beta-mercaptoethanol	- 20 mM Tris-HCl or HEPES pH 7.5 - 150 mM NaCl - 1 mM dithiothreitol (DTT)
Elution buffer	- 20 mM Tris-HCl or HEPES pH 7.5 - 200 mM NaCl - 500 mM imidazole - 0.5 mM beta-mercaptoethanol	- 20 mM Tris-HCl or HEPES pH 7.5 - 150 mM NaCl - 20 mM reduced Glutathione (Sigma, G4251) - 1 mM dithiothreitol (DTT)
Dialysis buffer	- 20 mM Tris-HCl or HEPES pH 7.5 - 100 mM NaCl	- 20 mM Tris-HCl or HEPES pH 7.5 - 100 mM NaCl

	- 1 mM dithiothreitol (DTT)	- 1 mM dithiothreitol (DTT)
Purification Step 2 - IEX		
Buffer A	- 20 mM Tris-HCl or HEPES pH 7.5 - 100 mM NaCl - 1 mM dithiothreitol (DTT)	- 20 mM Tris-HCl or HEPES pH 7.5 - 100 mM NaCl - 1 mM dithiothreitol (DTT)
Buffer B	- 20 mM Tris-HCl or HEPES pH 7.5 - 1000 mM NaCl - 1 mM dithiothreitol (DTT)	- 20 mM Tris-HCl or HEPES pH 7.5 - 1000 mM NaCl - 1 mM dithiothreitol (DTT)
Purification Step 3 - SEC		
SEC buffer	- 20 mM Tris-HCl or HEPES pH 7.5 - 150 mM NaCl - 1 mM dithiothreitol (DTT)	- 20 mM Tris-HCl or HEPES pH 7.5 - 150 mM NaCl - 1 mM dithiothreitol (DTT)

2.13 Protein crystallography, data collection and structure determination

For crystallisation, the SPOCD1 SPOC domain was concentrated to 9.5 mg/ml. Initially, a 96 well INDEX crystal screen (Hampton research, HR2-144) was set up using 0.2 µl protein plus 0.2 µl buffer in a sitting-drop vapour diffusion setting. The first crystal was observed at 293 K (20°C) in well B9 containing 1.8 M ammonium citrate tribasic, pH 7.0. More crystals were generated with the optimised condition of 1.6 M ammonium citrate tribasic, pH 7.5. The crystals were flash cooled in liquid nitrogen prior to data collection. The dataset was collected at I04 (DLS, Didcot Oxfordshire) at 100 K using a wavelength of 0.97949 Å.

The data were processed using autoPROC (Vonrhein et al. 2011) (v.1.0.5) including XDS (March 15, 2019, built 20191211) (Kabsch 2010), pointless (1.11.21) (Evans 2006), Aimless (0.7.4) (Evans and Murshudov 2013) and CCP4 (7.0.078) (Winn et al. 2011). Scaling was performed using SCALA (3.3.22) (Evans 2006). The structure was determined by molecular replacement with PHASER (v.2.8.3.) (McCoy et al. 2007) using the flowering protein FPA (PDB code: 5KXF) as a search model. The model was refined using PHENIX (v.1.17.1_3660)

(Liebschner et al. 2019) and rebuilt with Coot (v.0.8.9.2) (Emsley et al. 2010). Images were generated using the PyMol software (v.2.5.4, Schrödinger, LLC). The coordinates were submitted to the Protein Data Bank under the accession code 8OU1.

2.14 Protein surface charge and conservation prediction

The surface charge was calculated using APBS (Jurrus et al. 2018) in PyMol. The surface conservation was determined with Consurf (Yariv et al. 2023) inputting an alignment with SPOCD1 SPOC sequences from 25 different species (B1ASB6, Q6ZMY3, XP_008116112.1, XP_031752218.1, JH127468.5, XP_039185503.1, XP_034976513.1, XP_007493079.1, XP_040829971.1, W5NRM3, XP_036112709.1, XP_023490296.1, XP_030727203.1, XP_026965126.1, XP_025303405.3, XP_032245124.1, XP_049473573.1, H0XFY4, H2R1B9, XP_044853461.1, A0A485PAI7, F1MG39, KYO25862.1, XP_007668675.2, XP_020953261.1).

2.15 Immunofluorescence (IF)

For IF, freshly isolated testes were embedded in OCT embedding matrix (Cell Path, KMA-0100-00A), frozen in liquid nitrogen, and stored at the -80°C at least 24 hours before cutting. 6 – 8 µm tissue sections were cut using a cryostat (Thermo Fisher) at -16 °C and transferred to an adhesive microscopy glass slide. The slides were then dried for several hours at room temperature. For fixing, 3.5% (v/v) formaldehyde (15512, Sigma-Aldrich) diluted in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) was added for 10 minutes. The slides were washed twice with PBS, permeabilised for 10 minutes with 0.1% (v/v) Triton X-100 in PBS, and blocked for 1 hour with blocking buffer (10% (v/v) natural donkey serum (Merck, D9663), 1% (v/v) bovine serum albumin (BSA, Sigma-Aldrich, B6917), and 0.1% (v/v) glycine (Sigma-Aldrich, 50046) in PBS). Primary antibodies were diluted in blocking buffer and incubated overnight at 4 °C. The exact dilutions of each antibody are shown in Chapter 2.29. The following morning, the sections were washed three times with PBS and incubated with Alexa Fluor secondary antibody (donkey anti-rabbit or donkey anti-mouse 488, 568, or 647) diluted in

blocking buffer, as indicated in Chapter 2.29. DAPI (Invitrogen, D1306) stain at a concentration of 5 µg/ml was also added to the secondary antibody. After one hour incubation at room temperature, the slides were washed twice with PBS and mounted using Prolong Gold (Invitrogen, P36930). The slides were dried overnight and visualised on a Zeiss Observer or Zeiss LSM880 with an Airyscan module. If the images were acquired with the Airyscan module, they were deconvoluted using the Zeiss Zen software "Airyscan processing" with settings "3D" and a strength of 6. The Zeiss Zen software and ImageJ (ImageJ2; version 2.3.0/1.53q) were used to process and analyse the images.

2.16 Immunofluorescence (IF) with RNase A and Triton X-100 treatment

IF of MIWI2 or HA-SPOCD1 was performed on *Spocd1*$^{HA/+}$ E16.5 foetal testes as described in Chapter 2.29. Before fixing, the 8 µm section were either treated with 25 µg/ml RNase A (10109142001, Sigma-Aldrich) or PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) as the untreated control for 10 minutes at room temperature. The sections were washed three times for 5 minutes each and subsequently fixed with 3.7% (v/v) formaldehyde (15512, Sigma-Aldrich) diluted in PBS for 10 minutes. Slides were washed twice with PBS. Permeabilization, blocking, antibody incubation, and mounting were performed as described in Chapter 2.29.

2.17 Pull-down assay with recombinant protein

Untagged SPIN1 (amino acid 49-262) was mixed with GST-tagged recombinant SPOCD1 F1b fragments to a final volume of 60 µl. As a binding buffer for the mouse SPOCD1 assays, 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 1 mM DTT was used. For assays containing anolis, frog, and coelacanth fragments, 50 mM Tris-HCl pH 8.0, 150 mM KCl, 5 mM $MgCl_2$, and 0.5% (v/v) Triton X-100 was used. 50 µl of the protein mixture was added to 15 µl of washed and equilibrated Glutathione Sepharose High Performance beads (Cytica, 17-5279-02) and incubated for 1 hour on a rotating wheel at 4°C. Afterwards, the beads were washed four times with 500 µl binding buffer and eluted by boiling in 10 µl of 6x Lämmli buffer (15% (w/v) SDS, 0.075% (w/v) Bromophenol blue, 50% (v/v)

Glycerol, 75 mM Tris-HCl pH 7.0, 0.75 M DTT (added fresh just before usage)). Finally, 5 µl of the input and bead samples each were separated on a NuPage 4 – 12% Bis-Tris protein gel (Invitrogen, NP0323BOX) and stained with homemade Coomassie stain (10% (v/v) ethanol, 5% (v/v) acetic acid, 0.001% (w/v) Brilliant Blue R-250) .

2.18 Analytical size exclusion chromatography

For analytical size exclusion chromatography, SPIN1 (125 µg) and/or mouse GST-SPOCD1-F1b (500 µg) was used for each run. The proteins were mixed in SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT) to a final volume of 250 µl. After one hour incubation on ice, the sample was spun down at 21,000 x g for 10 minutes to remove precipitation and injected on a Superdex 200 10/300 GL column (Cytiva, 17-5175-01) that has been calibrated in SEC buffer. Fractions (500 µl) were collected, loaded on a NuPage 4 – 12% Bis-Tris protein gel (Invitrogen, NP0323BOX), and visualised with a homemade Coomassie stain (10% (v/v) ethanol, 5% (v/v) acetic acid, 0.001% (w/v) Brilliant Blue R-250).

2.19 Crosslink mass spectrometry (CL-MS)

2.19.1 Protein crosslinking

For crosslink mass spectrometry, 25 µg of recombinant fragments of Spocd1 (GST-F1b) and Spin1 were incubated in a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT. Bis(sulfosuccinimidyl)suberate (BS3, Thermo Fisher Scientific, 21580) was added for crosslinking at the BS3: protein ratios of 1:1, 2:1, and 4:1 w/w on ice. After 2 hours, the reaction was stopped by adding 2 µl of 2 M ABC. The mixture was separated on a NuPage 4 – 12% Bis-Tris protein gel (Invitrogen, NP0323BOX) for 15 minutes. The gel was stained with InstantBlue Protein Stain (Abcam, ab119211), and bands above 150 kDa were cut and processed for MS.

2.19.2 In-gel digestion of crosslinked protein

The gel piece was cut into small cubes and carefully transferred to an eppendorf tube containing 150 µl of 50 mM ABC. 150 µl of ACN was added to the tube, followed by a 30-minute incubation gently shaking at 37 °C. The supernatant was discarded, and another 150 µl ABC and ACN was added each, incubated for 30 minutes, and discarded. Finally, 50 µl ACN was added for 10 minutes to shrink the gel. For reduction, 150 µl of 10 mM DTT in 50 mM ABC buffer was added to the gel pieces and incubated for 30 minutes at 37°C in a shaker. The supernatant was removed, and the gel was shrunk by incubating for 10 minutes with 150 µl ACN. All supernatants were discarded, 150 µl of 55 mM iodoacetamide (IAA) in 50 mM ABC buffer was added, incubated for 20 minutes at room temperature, and all supernatants were removed. For digestion, the gel pieces were then covered in 150 µl trypsin solution (10 mM ABC containing 10% (v/v) ACN with 13 ng/µl trypsin (Thermo Scientific, 90057), from trypsin stock of 1 µg/µl in 0.1% (v/v) TFA). The samples were incubated overnight at 37°C. The next day, 10% (v/v) TFA was added to a final concentration of 0.1 – 0.5% (v/v) and incubated for 15 minutes in the shaker. All supernatants were transferred to a fresh tube for loading on the STAGE tip. The gel pieces were washed twice with 100 µl of 0.1% (v/v) TFA and incubated for 15 minutes in the shaker each time. All supernatants were collected for the STAGE tip.

2.19.3 STAGE tips of crosslinked protein

STAGE tips were prepared by inserting three layers of the C18 matrix into 200 µl tips. The STAGE tips were activated by adding 20 µl methanol to the STAGE tips and centrifuging. All centrifugation steps were performed at 1000 g for 3 - 5 minutes until all liquid passed through the matrix. Next, the STAGE tips were washed twice with 30 µl 0.1% (v/v) TFA, and the sample was loaded. This was followed by three final washes with 30 µl of 0.1% (v/v) TFA. The STAGE tips were stored at -20 °C until elution for mass-spectrometry analysis.

2.19.4 Mass-spectrometry analysis of crosslinked protein

The digested peptides were eluted from the STAGE tips for MS analysis (Rappsilber, Mann, and Ishihama 2007). The analysis was performed with the Orbitrap Fusion Lumos (Thermo Fisher Scientific) with a “high/high” acquisition strategy. An EASY-Spray column (50 cm × 75 µm i.d., PepMap C18, 2 µm particles, 100 Å pore size, Thermo Fisher Scientific) was used for separation, with mobile phase A consisting of water and 0.1% (v/v) formic acid and mobile phase B containing 80% (v/v) acetonitrile and 0.1% (v/v) formic acid. The peptides were loaded at a flow rate of 0.3 µl/min and eluted at 0.25 µl/min with a linear gradient from 2% mobile phase B to 40% mobile phase B over 102 or 132 minutes (each sample was run twice with different gradients). This was followed by a linear gradient increase from 40% to 95% mobile phase B over 11 minutes and a direct introduction of the eluted peptides to the mass spectrometer. Data acquisition was performed in a data-dependent mode with a 3 second acquisition cycle. In the Orbitrap, precursor spectra with a resolution of 120,000 and m/z range of 350-1700 were recorded, and ions with a precursor charge state between 3+ and 8+ were isolated with a window size of 1.6 m/z and fragmented using high-energy collision dissociation (HCD) with a collision energy of 30. The fragmentation spectra were recorded with a resolution of 15,000 with dynamic exclusion enabled (single repeat count and 60 seconds exclusion duration). ProteoWizard (version 3.0) (Kessner et al. 2008) was used for processing the mass spectrometric raw files into peak lists and Xi software (version 1.7.6.4) (Mendes et al. 2019) was used to match cross-linked peptides to spectra with in-search assignment of monoisotopic peaks (Lenz et al. 2018). The search parameters were as follows: MS accuracy, 3 ppm; MS/MS accuracy, 5 ppm; enzyme, trypsin; cross-linker, BS3; max missed cleavages, 4; fixed modification, carbamidomethylation on cysteine; variable modifications, oxidation on methionine; fragments, b and y ions with loss of H_2O, NH_3, and CH_3SOH. The BS3 linkage specificity was assumed to be at lysine, serine, threonine, tyrosine, and protein N-termini. Of the candidate cross-linked peptides, only cross-linked peptides validated by Xi software were used (Mendes et al. 2019). Crosslink MS data was deposited under PXD041135.

2.20 Chromatin immunoprecipitation (ChIP) sequencing analysis

For the ChIP sequencing (ChIP-seq) analysis, previously published raw fastq.gz files of H3K4me3 and H3K9me3 were downloaded from the Sequence Read Archive record SRP165187 (Yamanaka et al. 2019). For paired end reads, pre-processing was performed by removing adapter sequences and trimming low-quality bases with Trimmomatic v0.35 (Bolger, Lohse, and Usadel 2014). For ChIP-seq samples, tru-seq adapter sequences were used. Bwa mem v0.7.16 (Li 2013) using the -M parameter was used to align the trimmed reads to the mouse mm10 genome. Duplicates were removed by filtering the alignments with Picard MarkDuplicates v2.24.0 (http://broadinstitute.github.io/picard/) while improper alignments were removed with Samtools view v1.11 -F 260 -f 3 (Li et al. 2009). If multimapping reads were present, a single alignment, marked as primary by bwa, was selected for downstream analysis. For visualisation, the BAM files were converted to normalised bigWig files and plotted using deepTools (Ramirez et al. 2016) bamCoverage v3.5.0, with the following parameters: -bs 1 --normalizeUsing BPM.

2.21 Chromatin immunoprecipitation (ChIP) heatmaps and average profile plots

The genomic annotations for the repetitive elements L1Md_A, L1Md_T, L1Md_F, L1Md_Gf, IAPEy, and MMERVK_10C were extracted from Repeat Masker using the UCSC table browser. DeepTools v3.5.0 computeMatrix was used to compute normalised read coverage across these elements. The length of the central regions was normalised to 5 kb, with the flanking regions +/- 2 kb from the start and end positions. DeepTools v3.5.0 plotHeatmap was used to draw heatmaps while separating each repetitive element and sorting the rows in descending order of the total signal. The heatmaps of LINE1 elements (L1Md_A, L1Md_T, L1Md_F) were further separated into old (>38 nucleotides) and young (<=38 nucleotides) categories based on their divergence from the consensus sequence (Zoch, Auchynnikava, Berrens, Kabayama, Schopp, et al. 2020). Next, average profiles were generated by calculating the mean signal between replicate samples for each experiment and each category of repetitive elements. Calculations were performed

using the seqplots package (Stempor and Ahringer 2016) in R, with bins of 50 bases, flanking regions of 2 kb, and the central region normalised to a length of 5 kb. Finally, plots were generated and edited using tidyverse packages (Wickham 2019).

2.22 Fertility assessment

Male fertility of *Spocd1*$^{\Delta SPIN1}$ mice was assessed by mating *Spocd1*$^{\Delta SPIN1/\Delta SPIN1}$ studs with C57BL6/6N wild-type females and counting the number of pups born. As a control, wild-type male littermates were mated with C57BL6/6N wild-type females, and pups were counted.

2.23 Histology

Testes and epididymis were isolated from adult animals (6 - 8 weeks) and testes were weighed and photographed. Afterwards, the testes and epididymis were fixed overnight in Bouin's fluid (Sigma-Aldrich, HT10132) and washed at least three times in 70 % EtOH. The testes and epididymis were embedded by the Histology Facility in QMRI, University of Edinburgh. Sections were cut to 6 µm thickness using a microtome (Leica).

The sections were de-paraffinised and rehydrated using xylene (2 × 5 minutes) and a decreasing alcohol series (2 × 100% ethanol for 20 seconds, 95%, 80%, and 70% EtOH for 20 seconds each). Periodic-acid-Schiff (PAS) staining was performed on these slides according to the manufacturer's instructions (TCS Biosciences, HS462). The stained slides were then dehydrated using an increasing alcohol series (70%, 80%, 95% and 2 x 100% EtOH for 20 seconds each) (Table 7) and mounted on coverslips with Pertex mounting media (Pioneer Research Chemicals, PRC/R/750). Scanning was performed using a Zeiss AxioScan scanning microscope with a 40x objective. The images of the scans were exported and processed using ImageJ for the final figure.

Table 7: PAS staining of mouse tissue sections.

Step	Substance	Time	
1	Xylene	5 minutes	**Deparaffinization and rehydration**
2	Xylene	5 minutes	
3	Absolute EtOH	20 seconds	
4	Absolute EtOH	20 seconds	
5	95% EtOH	20 seconds	
6	80% EtOH	20 seconds	
7	70% EtOH	20 seconds	
Wash in ddH_2O			
8	Periodic acid	10 minutes, 1 ml per slide	**Staining**
Wash in tap water for 3 minutes			
9	Schiff Reagent	20 minutes, 1 ml per slide	
Wash in tap water			
10	70% EtOH	20 seconds	**Dehydrating**
11	80% EtOH	20 seconds	
12	95% EtOH	20 seconds	
13	Absolute EtOH	20 seconds	
14	Absolute EtOH	20 seconds	

2.24 Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay

TUNEL assay was performed on paraffin-embedded adult tissues. The sections were cut, de-paraffinised, and rehydrated, as described in Chapter 2.23. The slides were then treated with proteinase K (10 µg/ml in 10 mM Tris-HCl pH 8.0) and labelled using the Click-iT TUNEL assay (Invitrogen, C10617) with Alexa Fluor 647 dye according to the manufacturer's instructions. Sections were counterstained with DAPI (Invitrogen, D1306), embedded, imaged, and analysed as described in Chapter 2.15.

2.25 RNA sequencing and analysis

For RNA sequencing from P20 animals, RNA was extracted from one testis using the Qiagen RNeasy Mini kit (Qiagen, 74104). The protocol was performed according to the manufacturer's instructions, including an on-column DNase digest. The NEBrRNA depletion kit v2 (NEB, E7400) was used for ribosomal RNA (rRNA) depletion, and libraries were generated using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7760) with eight PCR cycles following the manufacturer's protocol. Samples were sequenced on a NextSeq 500 (Illumina) in 150 bp single-end read mode.

For analysis, the adapter sequence was removed from the FASTQ files using cutadapt version 1.18 (Martin 2011), and reads less than 50 nt were discarded using the -m50 option. The trimmed and filtered reads were mapped against the mouse consensus sequences of repetitive elements from RepBase44 (24.01) using bowtie2 (2.4.2) (Langmead and Salzberg 2012) with the options --qc-filter --no-unal. The number of reads mapping to each consensus sequence were counted and analysed using a custom Perl script. The RNA-seq data generated in this study has been deposited at the Gene Expression Omnibus under GSE228294. The scripts used for the RNA sequencing analysis are available on github (https://github.com/rberrens/SPOCD1-piRNA_ directed_DNA_met).

2.26 Fluorescence-activated cell sorting (FACS) of P14 spermatogonia

Spocd1$^{\Delta SPIN1/\Delta SPIN1}$ or *Spocd1* $^{+/+}$ testes were isolated from P14 animals, dealbulginated, and digested with 0.5 mg/ml Collagenase (Sigma-Aldrich, C7657) and 50 µg/ml DNase (Sigma-Aldrich, DN25) in 2 ml GONI-MEM ((DMEM (Gibco, 41965039) supplemented with 1x penicillin-streptavidin (Gibco, 15140122), 1x non-essential amino acids (NEAA, Gibco, 11140035), 1 mM sodium pyruvate (Gibco, 11360070) and 1.5 mM sodium lactate (Sigma-Aldrich, L4263)) for 20 minutes at 32 °C on a shaker. Cells were carefully pelleted at 100 - 200g for 3 - 5 minutes and the supernatant was carefully removed. One millilitre of 0.05% trypsin-EDTA (Gibco, 25200056) supplemented with 100 µg/ml DNase (Sigma-Aldrich, DN25) was added and incubated for 12 minutes at 32 °C on a shaker. The digestion

was stopped by adding 200 µl FBS (Gibco, 10270106) with 500 µg/ml DNase and 3 - 5 minutes incubation at 32 °C. The supernatant was removed, and the pellet was resuspended in 100 µl GONI-MEM with 1000 µg/ml DNase. After counting the cells, they were blocked by resuspending in 100 µl FACS buffer (Dulbecco's phosphate buffered saline, Sigma-Aldrich, D8537 with 2% (v/v) FBS) containing Fc block (anti-CD16/32, eBioscience, clone 93, 1:50) and incubating for 20 minutes on ice. Then, 100 µl FACS buffer containing anti-CD45 (eBioscience, clone 30-F11, 1:200) and anti-CD51 (Biolegend, clone RMV-7, 1:50) biotin-conjugated antibodies was added for 20 minutes on ice. After two washes with 100 µl FACS buffer, the cells were stained with anti-CD9^{APC} (clone eBioKMC8, eBioscience, 1:200), anti-c-Kit$^{PE\text{-}Cy7}$ (clone 2B8, eBioscience, 1:1600), streptavidin^{V450} (BD bioscience, 1:250), and 1 µg/ml DAPI (Invitrogen, D1306). After the final wash step, cells were resuspended in 400 µl FACS buffer, filtered, and sorted on a BD Fusion or BD Aria II. Single, living CD45$^-$, CD51$^-$, c-Kit$^-$, and CD9$^+$ cells were sorted in GONI-MEM at 4 °C, pelleted for 5 minutes at 500 g, snap frozen in liquid nitrogen, and stored at -80°C until used for methylation sequencing (Figure 18).

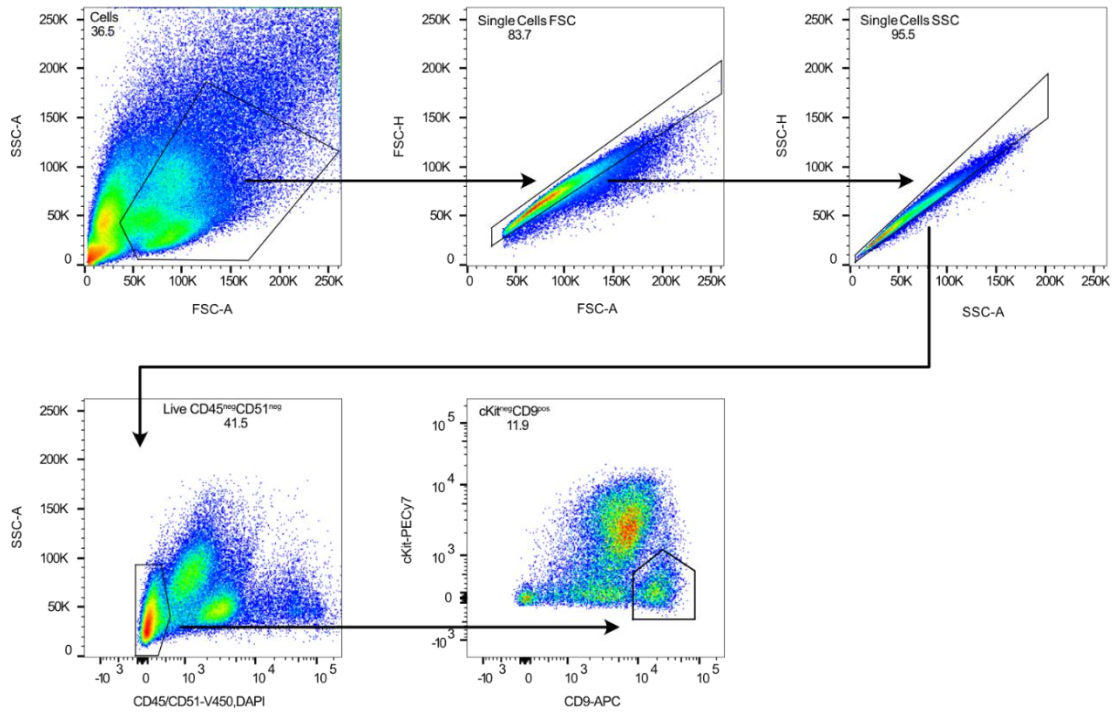

Figure 18: FACS gating strategy for P14 spermatogonia.

Representative gating strategy (n=3 mice) for sorting P14 undifferentiated spermatogonia. The gates select single, living, undifferentiated spermatogonia from a single-cell suspension of P14 testes that are CD45$^-$, CD51$^-$, c-Kit$^-$, and CD9$^+$.

2.27 Whole genome methylation sequencing (EM-seq) and analysis

For EM-seq, DNA was isolated from P14 FACS sorted P14 spermatogonial stem cells. First, cells were digested overnight at 55 °C with 0.2 mg/ml proteinase K (Roche, 3115852001) in 10 mM Tris-HCl pH 8.0, 5 mM EDTA, 1% (w/v) SDS, 0.3 M sodium acetate. This was followed by two rounds of phenol : chloroform : isoamyl alcohol (PCI, 25:24:1, Sigma-Aldrich, 77617) extraction and one round of chloroform extraction. For DNA precipitation, 1/10 volume of 3 M sodium acetate, 10 µg linear acrylamide (Invitrogen, AM9520), and 1 volume of isopropanol were added and incubated overnight at -20 °C. Finally, the pellet was washed twice with 70% ethanol and solubilised in 5 mM Tris-HCl at pH 8.0. Libraries for EM-seq were prepared with the NEBnext Enzymatic Methyl-seq kit (E7120, NEB) according to the manufacturer's protocol and sequenced on a NextSeq 2000 (Illumina) in 150 bp paired-end read mode.

Data for *Spocd1*$^{-/-}$ and corresponding *wild-type* P14 spermatogonia were retrieved from ArrayExpress under accession number E-MTAB-7997 (Zoch, Auchynnikava, Berrens, Kabayama, Schopp, et al. 2020).

First, Trim Galore (v0.4.1, www.bioinformatics.babraham.ac.uk/projects/trim_galore/, Cutadapt (Martin 2011) version 1.8.1, parameters:–paired–length 25–trim-n–clip_R2 5) was used to trim the raw sequences by removing poor-quality calls and adapters. To use overlapping read parts only once, trimmed reads were aligned to the mouse mm10 genome in paired-end mode with Bismark v0.22.1 (Krueger and Andrews 2011) with bismark–score_min L,0,-0.4–paired settings. CpG methylation calls were made based on the deduplicated mapping output from the Bismark methylation extractor (v0.22.1). The SeqMonk (www.bioinformatics.babraham.ac.uk/projects/seqmonk/) datastore summary report of aligned deduplicated bam files were used to calculate mapping statistics. All reads were mapped to the spiked-in CpG unmethylated lambda and CpG methylated pUC19 DNA to calculate the methylation conversation rate using the Bismark pipeline, as outlined above. For probes containing at least 10 reads, 50 adjacent CpG running window probes were generated and the mean percentage of methylation was calculated for each probe.

For the analysis, specific genome features were defined; probes overlapping genes, were defined as genic regions; probes overlapping 2,000 bp upstream of annotated transcripts (as annotated by Ensembl, GRCm38.p6) were defined as promoters; CpG islands (CGI) were defined as overlapping the Ensembl (GRCm38.p6) CGI annotation. Reads overlapping transposons were removed from the analysis of genetic, promoter, and CGI features. Transposons were defined by UCSC repeat masker annotations (downloaded from the table browser (https://genome.ucsc.edu/cgi-bin/hgTables, 02/2019). Simple repeats and small non-coding RNAs were removed from the transposon annotation (containing retrotransposons and DNA transposons). Transposable elements reads were mapped using unique mapping, and any repeats overlapping gene bodies were removed. Only mapping to full-length elements, defined as >5 kb for LINE1 elements, >6 kb for IAP families and >4.5 kb for MMERVK10C, were assessed. Regions that did not overlap with genes or transposons were defined as intergenic regions. The methylation level was determined by calculating the mean percentage of individual CG sites. Reads overlapping the respective genomic regions were extracted from SeqMonk and plotted in RStudio to generate metaplots, scatterplots, boxplots, and correlation analysis. Divergence (milliDiv) from the consensus sequences was used to perform the methylation difference analysis. Imprinted control regions (ICR) were extracted from https://atlas.genetics.kcl.ac.uk/. SeqMonk and RStudio were used to generate graphics and calculate statistics. The EM-seq data generated in this study has been deposited on ArrayExpress under accession number E-MTAB-12713. The scripts for the EM-seq analysis are available on github (https://github.com/rberrens/SPOCD1-piRNA_ directed_DNA_met).

2.28 Statistical information

Data was plotted in R (version 2022.07.01+554 running R version 4.0.3 (2020-10-10)) using the dplyr, ggplot2, tidyr, , cowplot, reshape2, ggrepel, ggpubr, scales and RColorBrewer packages (versions dplyr_1.0.4, ggplot2_3.3.3, tidyr_1.1.2, cowplot_1.1.1, scales_1.1.1, reshape2_1.4.4, ggrepel_0.9.1, ggpubr_0.4.0, scales_1.1.1, RColorBrewer_1.1-2) or Microsoft Excel for Mac (version 16). Statistical testing was performed with R (version 4.0.3 (2020-10-10)) using the R

Studio software and with Perseus version 1.6.5.0 (Tyanova et al. 2016) for the mass spectrometry data. Unpaired two-tailed Student's t-tests were used to compare differences between groups and adjusted for multiple testing using Bonferroni correction where indicated, except for RNA-seq data analysis, where Wald's tests and Benjamini–Hochberg correction were used. Averaged data are presented as mean ± s.e.m. (standard error of the mean), unless otherwise indicated.

2.29 Antibodies used in this study

2.29.1 Antibody generation

To generate anti-SPOCD1 rabbit serum, the TFIIS-M domain of mouse SPOCD1 was purified and injected into rabbits. The TFIIS-M construct contained an N-terminal GST-tag and SPOCD1 residue 407 - 568. It was purified in a three-step purification *via* GST batch pull-down, 3C digestion and dialysis, IEX (Resource Q), and SEC (Superdex S200) (Table 5).
Biotem performed the immunization of the rabbits for us. Two New Zealand white rabbits were injected on days 0, 14, 28, 56, and 70, according to the immunisation protocol. Antibody-containing sera were collected on day 82. After validation, rabbit 175 was found to be the most potent and was used for all further analyses.

2.29.2 Primary antibodies

anti-HA (C29F4s, Cell Signaling Technologies, IF: 1:200, WB: 1:1000);
anti-FLAG (M2, Sigma-Aldrich, WB: 1:1000);
anti-SPOCD1 rabbit serum rb175 (O'Carroll laboratory antibody, IF: 1:500, WB: 1:500);
anti-α-TUBULIN (T9026, Sigma-Aldrich, WB: 1:10000);
anti-HA (6E2, Cell Signaling Technologies IF: 1:200);
anti-LINE1-ORF1p (di Giacomo et all, 2014, IF: 1:500);
anti-IAP-GAG (a kind gift from B. Cullen, Duke University, Durham, NC, USA, IF: 1:500);
anti-γH2AX (IHC-00059, Bethyl Laboratories, IF: 1:500);

anti-MIWI2 (a kind gift from Ramesh Pillai, Université de Genève, Switzerland IF: 1:500);
anti-SPIN1 (E6R1Z, Cell Signaling Technologies, IF: 1:500 (of a custom preparation of 1.1 µg/ul in PBS));
anti-CD16/32 (clone 93, eBioscience, FACS: 1:50);
anti-CD45 (clone 30-F11, eBioscience, FACS: 1:400);
anti-CD51 (clone RMV-7, Biolegend, FACS: 1:100);
anti-$CD9^{APC}$ (clone eBioKMC8, eBioscience, FACS: 1:200);
anti-$cKit^{PE\text{-}Cy7}$ (clone 2B8, eBioscience, FACS: 1:1600);
anti-HA beads (88837, Pierce, IP-MS: 50 µl, IP-WB: 20 µl).

2.29.3 Secondary antibodies

IRDye® 680RD Anti-Rabbit (LI-COR, 926-68073, WB: 1:1000);
IRDye® 800CW Anti-Mouse (LI-COR, 926-32212, WB: 1:1000);
Alexa Fluor donkey anti rabbit 488, 568, 647 (Life Technologies, 1:1000);
Alexa Fluor donkey anti mouse 488, 568, 647 (Life Technologies, 1:1000);

2.29.4 Antibody validation

The anti-HA antibodies were validated and used for IF and WB previously (Zoch et al. 2020).
The anti-FLAG antibody is a widely used antibody and was used for WB previously (Zoch et al. 2020).
The anti-SPOCD1 rabbit serum rb175 was validated for IF by staining Spocd1-null testis sections and was validated for WB by staining samples without SPOCD1.
The anti-α-tubulin antibody is widely used and validated for WB using recombinant protein samples.
The anti-LINE1-ORF1p (Di Giacomo et al. 2014) and anti-IAP-GAG (Mietz et al. 1987) antibodies have been previously validated for IF in mouse sections.
The anti-γH2AX antibody has been previously used for IF on mouse sections (Vasiliauskaitė et al. 2018; Zoch et al. 2020; Much et al. 2016; Comazzetto et al. 2014).

The anti-MIWI2 antibody was previously validated for IF in mouse sections (Pandey et al. 2013).

The anti-CD16/32, anti-CD45, anti-CD51, anti-CD9, and anti-cKit antibodies have been validated for FACS analysis of undifferentiated spermatogonia (Vasiliauskaitė et al. 2017; Zoch et al. 2020).

Anti-HA beads were used for IP in previous studies (Zoch et al. 2020).

3 The interactome of SPOCD1

3.1 Contribution

Chapter 3.5 and 3.6 are part of a paper “C19ORF84 connects piRNA and DNA methylation machineries to defend the mammalian germline”, which has been submitted for publication. I have contributed to the design, execution, and analysis of most experiments in this chapter. I performed the IP-MS experiments with the help of Dr. Ansgar Zoch and Dr. Tania Auchynnikava under the guidance of Prof. Juri Rappsilber and Prof. Robin C. Allshire. Crystallographic data were collected and analysed under the guidance of Prof. Atlanta G. Cook. Human genetic data analysis was performed by Nadja Rotte and Dr. Birgit Stallmeyer under the guidance of Prof. Frank Tüttelmann. Prof. Donal O’Carroll conceived and supervised this study.

3.2 Introduction

The male germline is derived from somatic cells (Gan et al. 2013). During early development in mammals, this necessitates erasing and resetting of DNA methylation in these cells (Ginsburg, Snow, and McLaren 1990). This DNA demethylation results in young, active LINE1 and IAP copies being derepressed, threatening the integrity of the germline (Molaro et al. 2014). The piRNA pathway is the primary germline defence system against transposons (Ozata et al. 2018). The PIWI protein, MIWI2 binds to piRNA and instructs DNA methylation through the DNMT3 complex which contains DNMT3A, DNMT3C, and DNTM3L (Chédin, Lieber, and Hsieh 2002; Bourc'his and Bestor 2004; Webster et al. 2005; Barau et al. 2016; Kato et al. 2007b). Recently, it was shown that SPOCD1 links MIWI2 and the *de novo* DNA methylation machinery which ultimately induces DNA methylation

(Zoch et al. 2020). While SPOCD1 does have two annotated domains, a TFIIS-M and a SPOC domain (Zoch et al. 2020), the exact function of SPOCD1 and its binding proteins remain unknown.

3.3 Research objective and preliminary work

Dr. Ansgar Zoch was the first to successfully determine the interactome of SPOCD1 and MIWI2 in E16.5 testes (Zoch et al. 2020). He performed HA IP-MS from N-terminal HA-tagged MIWI2. In addition to SPOCD1 coming up as an interactor, he identified that the proteins Spindlin1 (SPIN1), Translocated promoter region nuclear basket protein (TPR), and C19ORF84 were enriched above our cut-off (>4-fold, $P < 0.05$) (Figure 19a). The same experiment was repeated with C-terminal HA-tagged SPOCD1, and SPIN1, TPR, and C19ORF84 were observed as the interactors with the highest enrichment and confidence (Figure 19b). In this work, I investigate all three interactors in Chapter 3. The SPOCD1-SPIN1 interaction is introduced and studied in detail in Chapter 4.

TPR is a 267 kilo Dalton (kDa) component of the nuclear pore complex (NPC) (Mitchell and Cooper 1992; Cordes et al. 1997). It is part of the nuclear interior with the N-terminal part forming coiled-coiled structures and harbouring a short stretch that is necessary for anchoring to the NPC. (Hase, Kuznetsov, and Cordes 2001; Cordes, Hase, and Muller 1998; Bangs et al. 1998). The acidic c-terminus is largely unfolded and flexible (Hase, Kuznetsov, and Cordes 2001). In different mammalian cells, it has been shown that the NPC-binding part is located at the NPC, while parts of the coiled-coiled and the c-terminus are found deeper within the nucleus (Cordes et al. 1997; Frosst et al. 2002; Krull et al. 2004; Huve et al. 2008). TPR is involved in mRNA and protein export (Bangs et al. 1998; Frosst et al. 2002). It plays a key role in the formation of heterochromatin exclusion zones (HEZs) around the NPC (Krull et al. 2010), is involved in mitotic spindle checkpoint signalling (Lee et al. 2008), and is associated with maintaining genomic stability (Kosar et al. 2021). However, it remains unknown which part of SPOCD1 interacts with TPR.

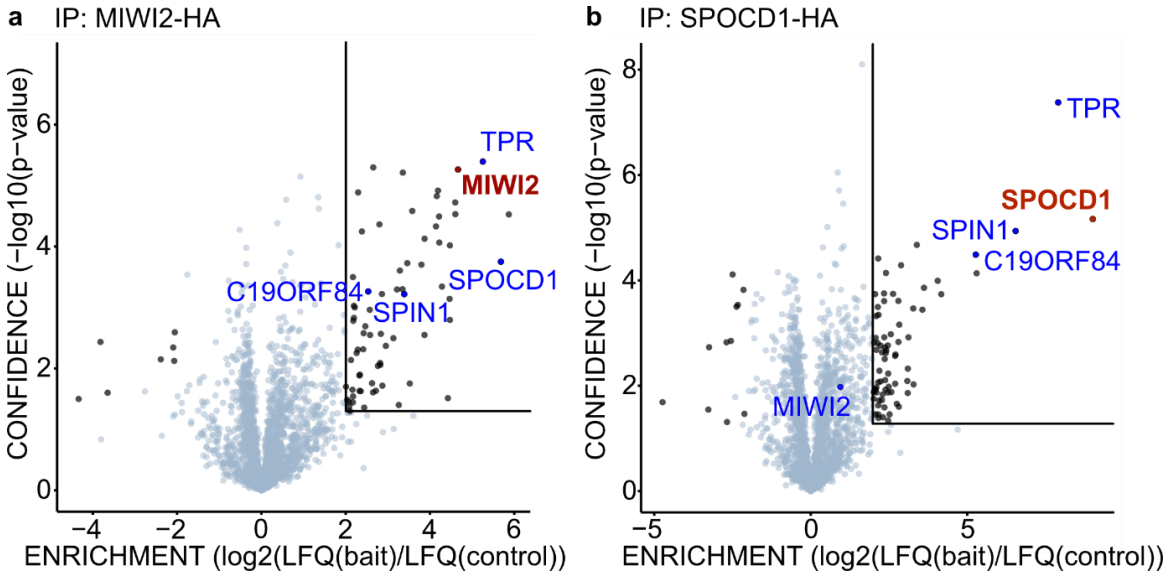

Figure 19: The MIWI2-HA and SPOCD1-HA interactome in foetal testis.
a, Volcano plot showing enrichment (log_2(mean LFQ ratio of anti-HA immunoprecipitates from n = 4 Spocd1$^{HA/+}$/wild-type) E16.5 foetal testes) and statistical confidence (log_{10}(P-value of two-sided Student's t-test)) of proteins co-purified with SPOCD1-HA. The line indicates enrichment >4-fold and a significance of $P < 0.05$. **b**, Volcano plot showing enrichment and confidence as in **a** from n = 4 Miwi2$^{HA/+}$/wild-type E16.5 foetal of proteins co-purified with MIWI2-HA. Data sourced from (Zoch et al. 2020).

C19ORF84 is an uncharacterized protein with unknown function. Like many key piRNA factors, it is exclusively expressed during *de novo* DNA methylation. Gabriela Konieczny generated a C19ORF84-null allele and showed that C19ORF84 is required for piRNA-directed DNA methylation and male fertility (Konieczny 2022). However, the modes of interaction between SPOCD1 and C19ORF84 remain unknown.
In this chapter, I mapped the interaction of SPOCD1 with TPR and C19ORF84. In addition, I solved the structure of the SPOC domain of SPOCD1, whose function still remains elusive.

3.4 The interactome of SPOCD1

SPOCD1 is a 1015 amino acids protein that contains two predicted domains: the TFIIS-M and SPOC domain (Figure 20a). Recent advances in protein structure prediction have allowed to examine the secondary structure of SPOCD1 predicted by AlphaFold (Jumper et al. 2021; Tunyasuvunakool et al. 2021) (Figure 20b). The

SPOCD1 model reveals that the protein is highly unstructured apart from its two domains (Figure 20b). Within the unstructured areas, small secondary structure elements such as α-helices or a β-hairpin can be found (Figure 20b).

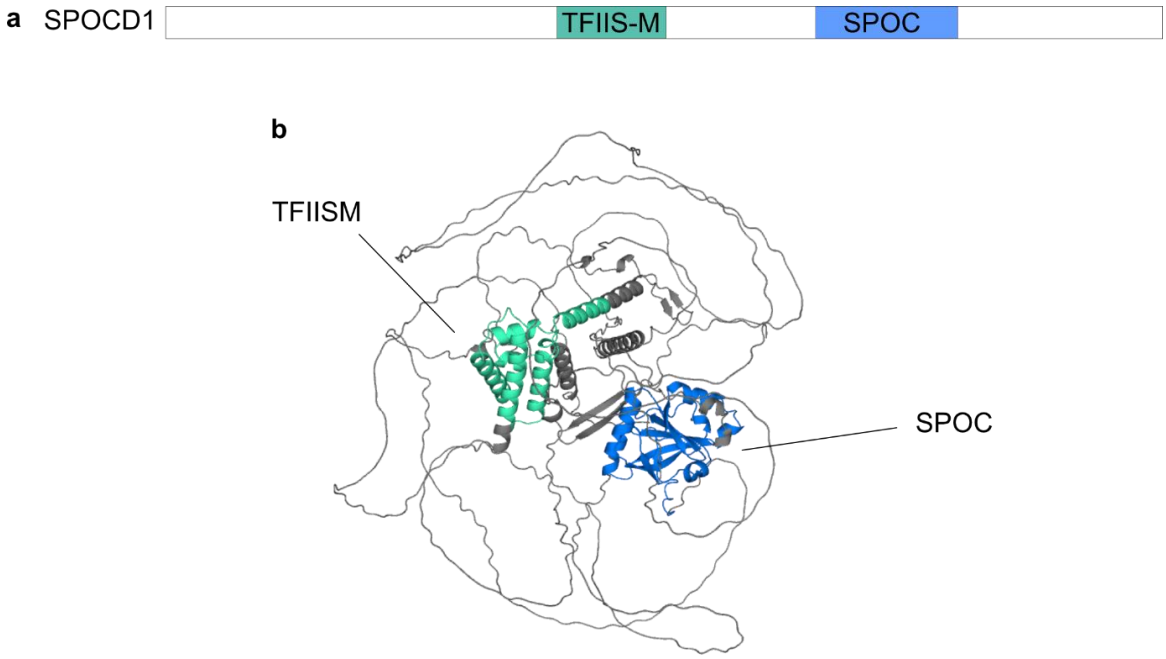

Figure 20: Overview of domain structure of mouse SPOCD1.
a, Schematic overview of SPOCD1 with its two domain indicated. **b**, AlphaFold2 prediction of SPOCD1 (B1ASB6) with secondary structure elements labelled and highlighted by colour. TFIIS-M domain, green; SPOC domain, blue.

First, I wanted to understand the functions of the two domains that are present in SPOCD1. Since all experiments carried out in mice are timely, expensive, and have ethical challenges, I decided to establish an IP-MS protocol in HEK293T cells. HEK293T cells were transfected with the protein(s) I wanted to study, and IP was performed using the HA-tag of my protein of interest. In this context, it must be noted that HEK293T cells do not express SPOCD1 or C19ORF84; however, human TPR and SPIN1 are expressed at endogenous levels in these cells.
In the first experiment, I transfected HEK293T cells with mouse SPOCD1-HA full-length and saw endogenous HEK293T cell SPIN1 and TPR coprecipitating in the mass-spectrometry analysis. Both proteins were clearly associated with SPOCD1 above our set enrichment cutoff (>4-fold, $P < 0.05$) (Figure 21a, b).

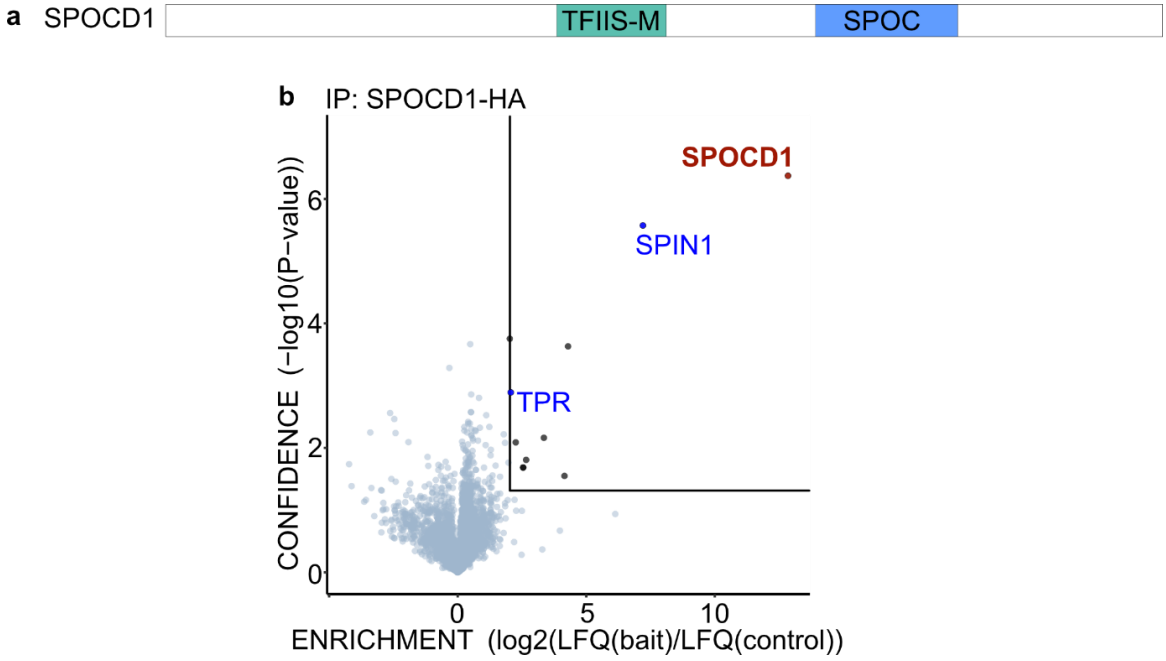

Figure 21: SPOCD1 interacts with SPIN1 and TPR in HEK293T cells.

a, Schematic overview of SPOCD1 with its two domains; TFIIS-M, green; SPOC, blue, indicated. **b**, Volcano plot showing enrichment and confidence (log_2(mean LFQ ratio of anti-HA immunoprecipitates from n = 3 Spocd-HA transfected/wild-type) HEK cells) and statistical confidence (log_{10}(P-value of two-sided Student's t-test)) of proteins co-purified with SPOCD1-HA. The line indicates enrichment >4-fold and a significance of $P < 0.05$.

Knowing that SPOCD1, TPR, and SPIN1 associate together in HEK293T cells, I wanted to understand the roles of the TFIIS-M and SPOC domains. For this purpose, I generated constructs with a deletion of each of the domains (Figure 22a). First, I generated a SPOCD1-HA construct with a deletion of the TFIIS-M domain and determined its interactome in HEK293T cells (Figure 22b). Although the SPOCD1-SPIN1 interaction was unaffected, the TPR interaction was lost in the TFIIS-M deletion (Figure 22b). This indicated that TPR interacts with the TFIIS-M domain. This experiment was repeated with an SPOCD1-HA construct that did not have the SPOC domain (Figure 22c). In this case, both the interaction of SPOCD1 with TPR and the interaction of SPOCD1 with SPIN1 were retained compared to the SPOCD1 full-length IP-MS (Figure 22c). The SPOC domain did not seem to play a role in the interaction with either TPR or SPIN1.

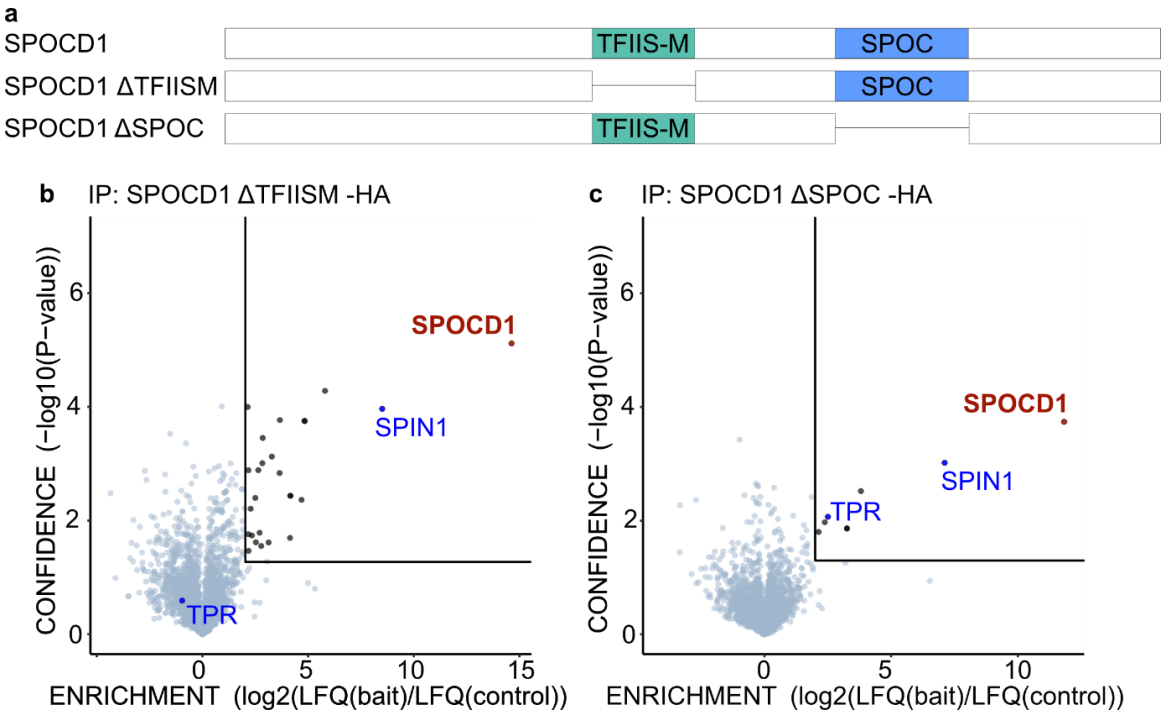

Figure 22: TPR interacts with the TFIIS-M domain of SPOCD1.

a, Schematic of the SPOCD1 ΔTFIISM and SPOCD1 ΔSPOC constructs. TFIIS-M, green; SPOC, blue. **b,** Volcano plot depicting enrichment and confidence (log_2(mean LFQ ratio of anti-HA immunoprecipitates from n = 3 SPOCD1 ΔTFIISM-HA transfected/wild-type) HEK293T cells) and statistical confidence (log_{10}(P-value of two-sided Student's t-test)) of proteins co-purified with SPOCD1 ΔTFIISM-HA. The line indicates enrichment by >4-fold and a significance level of $P < 0.05$. **c** Volcano plot (as in **b**) showing proteins co-purified with SPOCD1 ΔSPOC-HA.

Since C19ORF84 is not expressed in HEK293T cells, I repeated the same experiment with C19ORF84 added in addition. I could demonstrate that when co-transfecting SPOCD1-HA and C19ORF84 in HEK293T cells, I can see C19ORF84, SPIN1, and TPR co-precipitated with SPOCD1 (Figure 23a), which resembles the associations observed in foetal testes (Figure 19b). The C19ORF84 interaction was unaffected by deletion of the TFIIS-M domain of SPOCD1 (Figure 23b). Consistent with the previous TFIIS-M deletion experiment (Figure 23b), TPR interaction was lost. (Figure 23b). I then tested the SPOCD1-HA construct with a SPOC deletion while co-transfecting C19ORF84, and again, all three interactions seemed to be retained (Figure 23c).

In summary, I could recapitulate the SPOCD1 complex containing SPIN1, TPR and C19ORF884 in HEK293T cells and could show that the SPIN1 and C19ORF84 interactions are independent of any of the annotated domains in SPOCD1, while TPR interacts with the TFIIS-M domain.

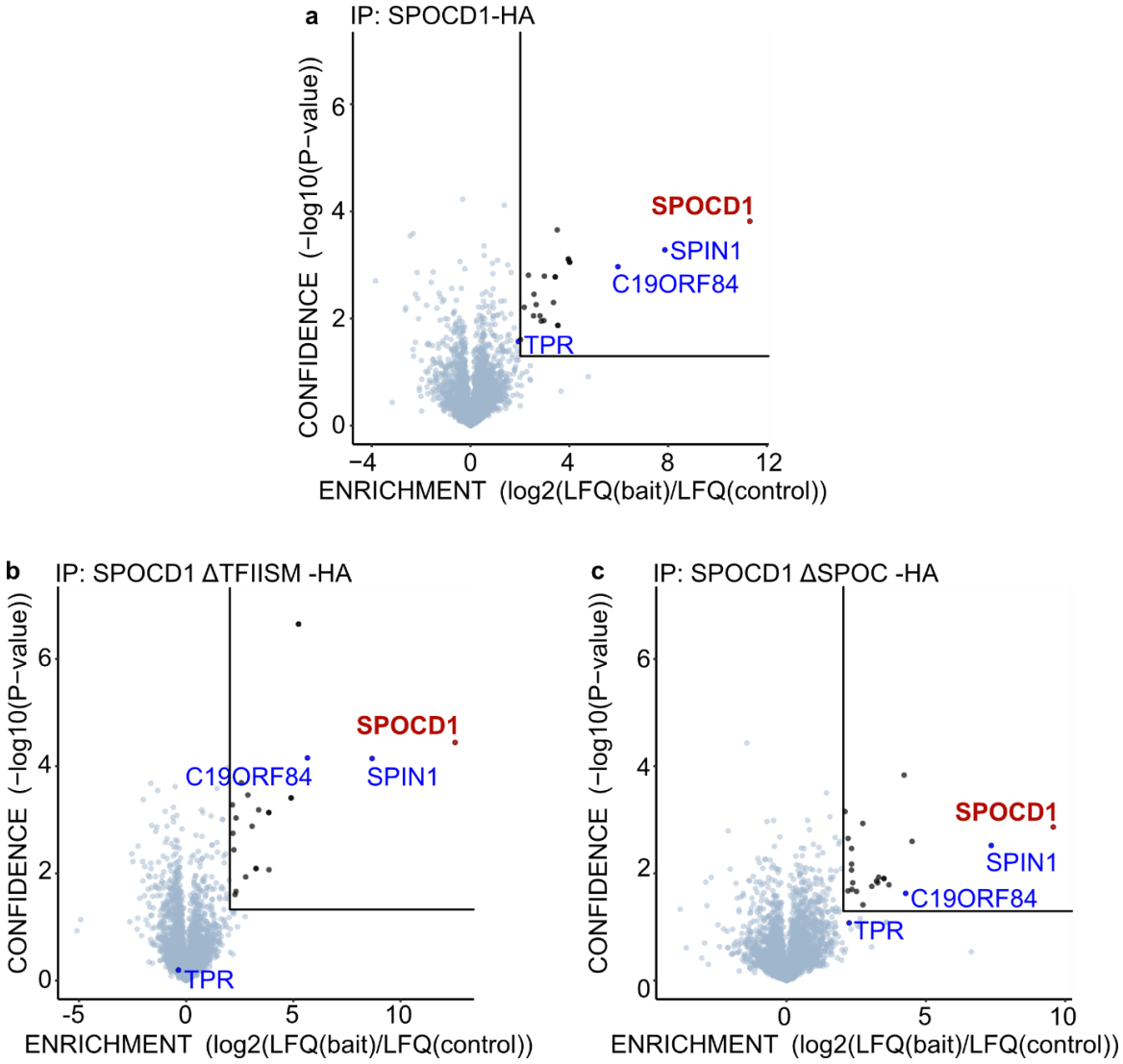

Figure 23: SPOCD1 interacts with SPIN1, TPR, and C19ORF84 in HEK293T cells.

a, Volcano plot showing enrichment and confidence (log_2(mean LFQ ratio of anti-HA immunoprecipitates from n = 3 SPOCD1-HA and C19ORF84 transfected/wild-type) HEK293T cells) and statistical confidence (log_{10}(P-value of two-sided Student's t-test)) of proteins co-purified with SPOCD1-HA. The line indicates enrichment by >4-fold and a significance level of $P < 0.05$. **b,** Volcano plot (as in **a**) of proteins co-purified with SPOCD1 ΔTFIISM-HA from HEK293T cells transfected with SPOCD1 ΔTFIISM-HA and C19ORF84. **c,** Volcano plot (as in **a**) showing proteins co-purified with SPOCD1 ΔSPOC-HA from HEK293T cells transfected with SPOCD1 ΔSPOC-HA and C19ORF84.

3.5 The interaction of SPOCD1 with C19ORF84

Next, the SPOCD1-C19ORF84 interaction was investigated. To confirm the IP-MS experiments (Figure 23), I repeated the IP-WB experiment with the SPOCD1 ΔTFIISM-HA and SPOCD1 ΔSPOC-HA constructs.

As expected, C19ORF84 is co-precipitating with SPOCD1 full-length as well as with the two deletion constructs (Figure 24). This confirmed the IP-MS results which showed that the SPOCD1-C19ORF84 interaction is independent of the TFIIS-M or SPOC domain (Figure 23).

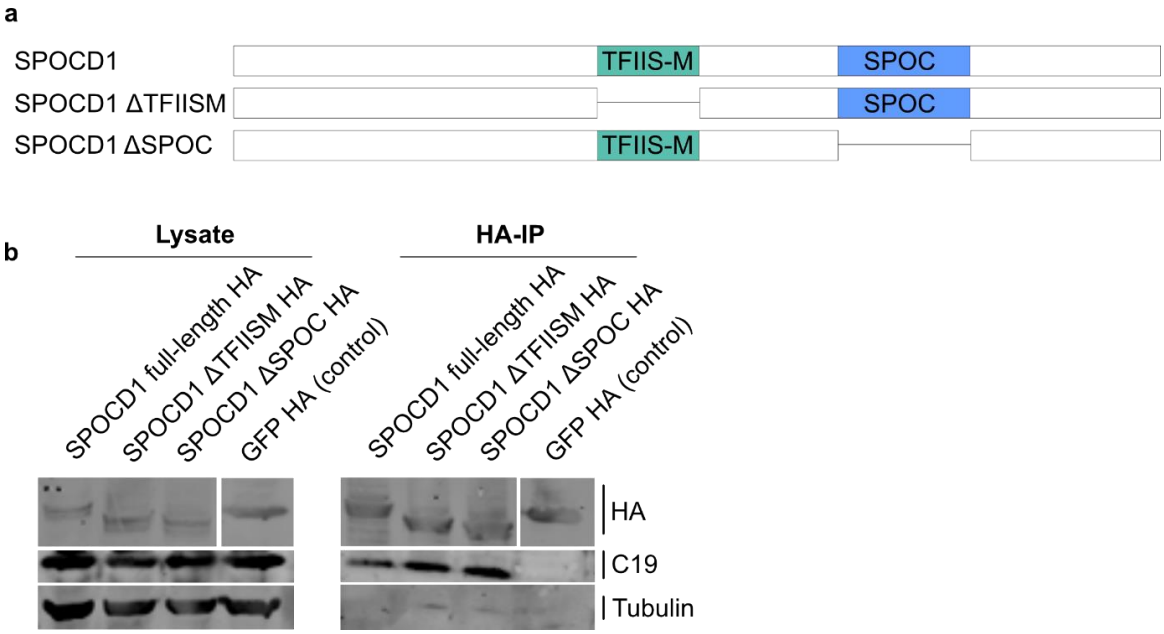

Figure 24: The SPOCD1-C19ORF84 interaction is independent of the TFIIS-M or SPOC domains.

a, Schematic of the SPOCD1 ΔTFIISM and SPOCD1 ΔSPOC constructs. TFIIS-M, green; SPOC, blue. **b**, Representative western blot analyses of n = 3 immunoprecipitations of the indicated SPOCD1 constructs with C19ORF84 in HEK293T cells.

To further map this interaction, I split SPOCD1 into four fragments and tested each for interaction with C19ORF84 in HEK293T cells (Figure 25a). I could show that the N-terminal (F1) and middle fragment (F2) did not interact with C19ORF84 (Figure 25b). However, the two C-terminal fragments (F3 and F4) pulled down C19ORF84 (Figure 25b). The minimal fragment for co-precipitation of C19ORF84 appears to be fragment F4 which spans residues 830 – 1015 of SPOCD1 (Figure 25a,b).

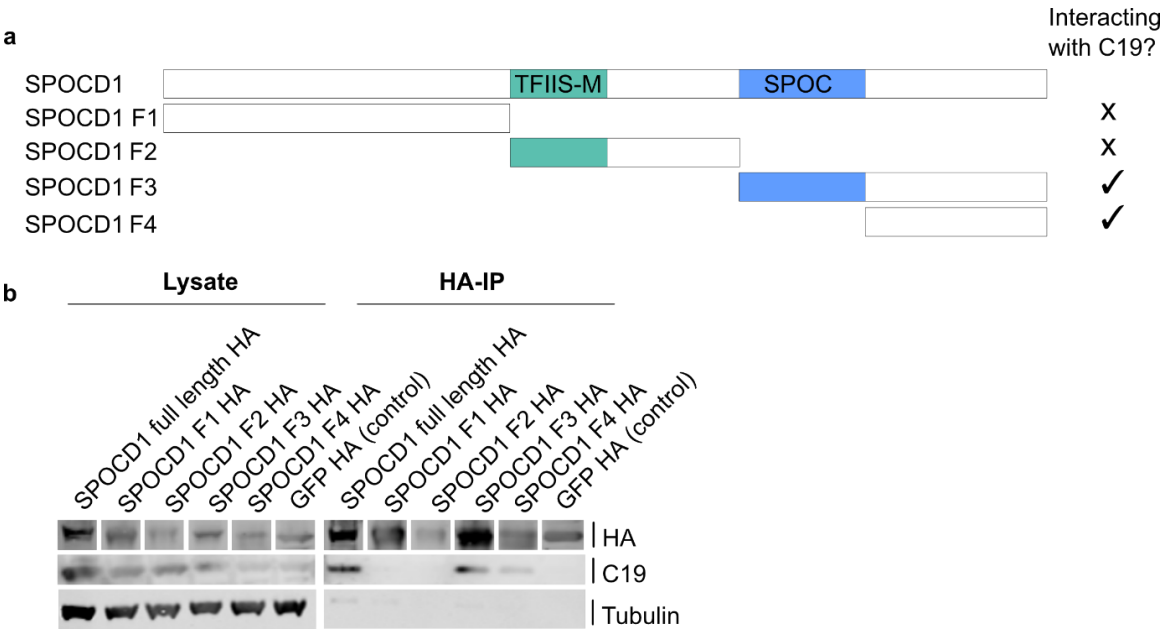

Figure 25: C19ORF84 interacts with the C-terminus of SPOCD1.

a, Schematic of the SPOCD1 fragment tested for interaction. TFIIS-M, green; SPOC, blue **b**, Representative western blot analyses of n = 3 immunoprecipitations of the indicated SPOCD1 fragments with C19ORF84 in HEK293T cells.

Following this, I generated SPOCD1 full-length mutants with deletions in the C-terminal region that interacted with C19ORF84 (Figure 26a). Deleting the whole fragment F4 (831-1015) from SPOCD1 resulted in loss of interaction (Figure 26b). This stretch was then divided into two equally sized pieces (Figure 26a). In a SPOCD1 deletion of residue 831-922 I maintain interaction with C19ORF84, while the SPOCD1 construct that stops at residue 922 (Δ923-1015) is not able to interact (Figure 26b). Again, I split the interacting stretch, amino acids 923 – 1015, into two parts and tested it for interaction with C19ORF84 (Figure 26a). In the SPOCD1 construct with a deletion of amino acid 923-969 I lose interaction, while the interaction is retained in the SPOCD1 construct with a deletion of the c-terminal residues 970-1015 (Figure 26b). In conclusion, SPOCD1 amino acid 923-969 are essential for the interaction with C19ORF84 (Figure 26c).

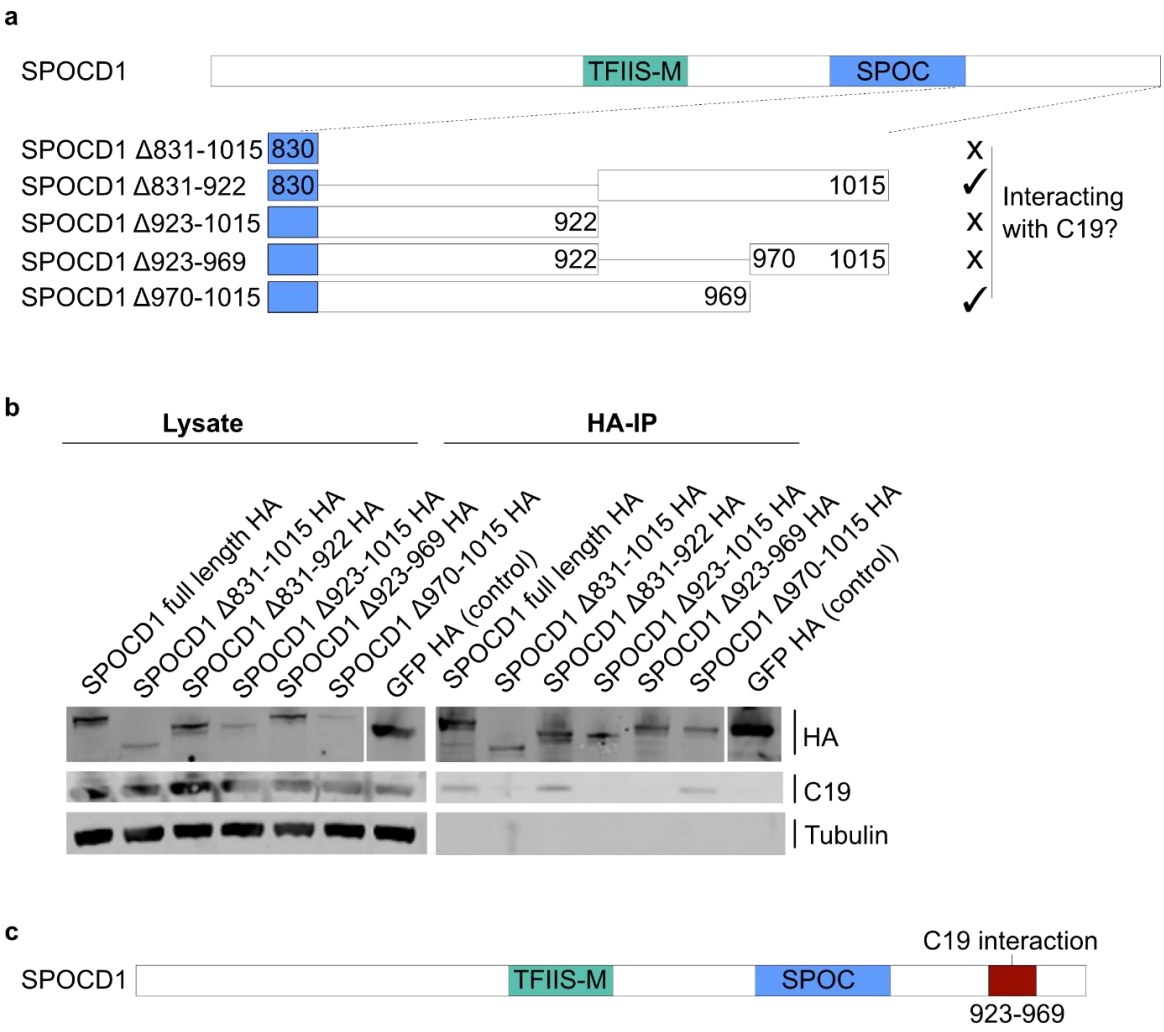

Figure 26: C19ORF84 interacts with SPOCD1 residue 923-969.

a, Schematic of the SPOCD1 deletions tested for interaction. **b**, Representative western blot analyses of n = 3 immunoprecipitations of the indicated SPOCD1 deletions with C19ORF84 in HEK293T cells. **c**, Schematic indicating residue 923-969 (red) in SPOCD1. **a**, **c**: TFIIS-M, green; SPOC, blue

Having an AlphaFold SPOCD1 protein structure prediction, I investigated if there is any secondary structure present in the C19ORF84-interacting part. Indeed, in the interacting part (residue 923-969), a well-defined α-helix (residue 942-964) can be found (Figure 27a). This helix is conserved among mammals but becomes less conserved in more distant animals such as amphibians (Figure 27b).

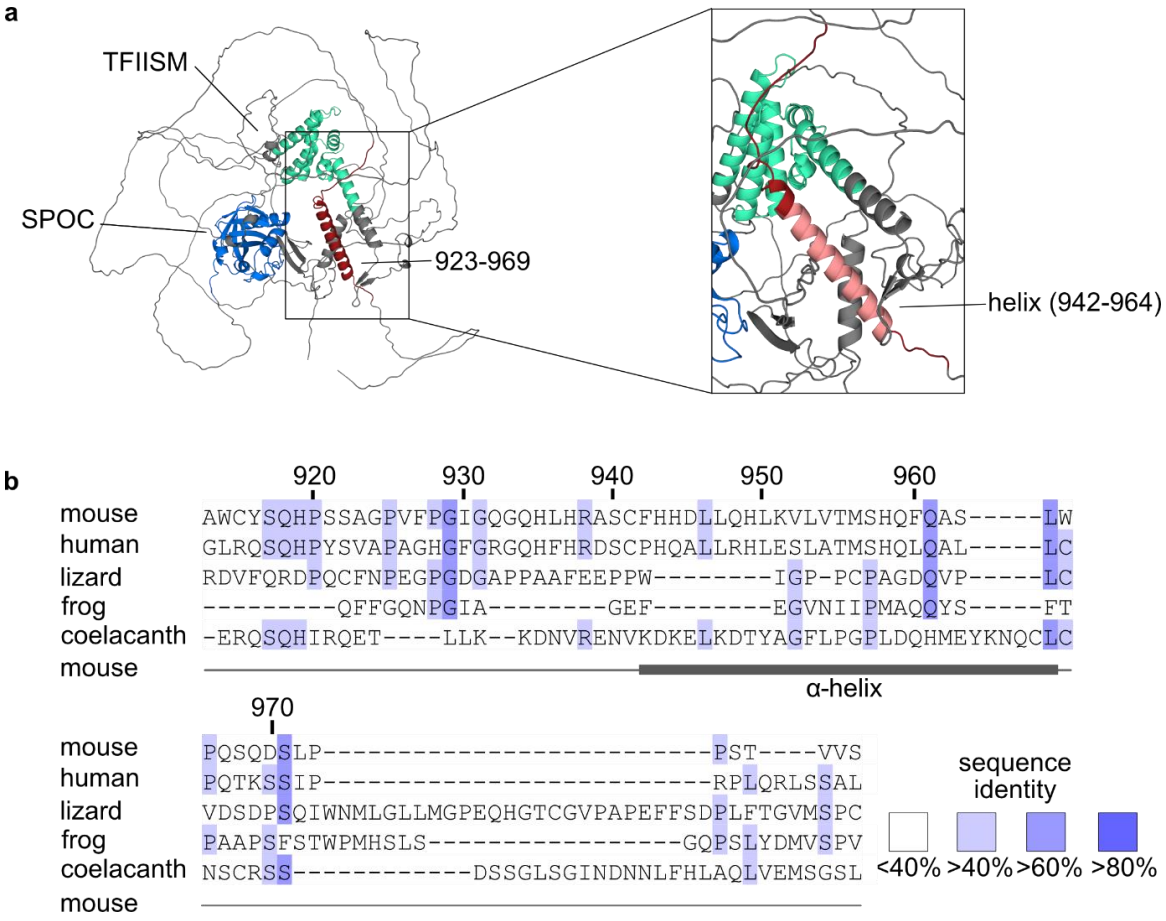

Figure 27: The SPOCD1 α-helix is conserved among mammals.

a, AlphaFold2 prediction of SPOCD1 (B1ASB6) with secondary structure elements highlighted by colour. TFIIS-M domain, green; SPOC domain, blue; residue 923-969, red; α-helix, pink. **b**, Multiple sequence alignment of the SPOCD1 C19ORF84-interacting region from different species. Numbering of mouse SPOCD1 above the sequences. The secondary structure elements of mouse SPOCD1 are shown below. Sequences are coloured according to sequence identity.

To understand whether this α-helix plays an important role in the SPOCD1-C19ORF84 interaction, I generated a SPOCD1 construct with a deletion of only the α-helix (942-964). Indeed, this helix was shown to be essential for the interaction of SPOCD1 with C19ORF84 in HEK293T cells (Figure 28).

Having defined the interaction of SPOCD1 with C19ORF84, I wanted to understand which part of C19ORF84 was responsible for interacting with SPOCD1. Using the AlphaFold model (Jumper et al. 2021; Tunyasuvunakool et al. 2021) I could show that C19ORF84 is a very unstructured protein and contains only one structural element, an α-helix spanning residue 64 – 81 (Figure 29a). This helix is highly conserved in mammals and lizards (Figure 29b).

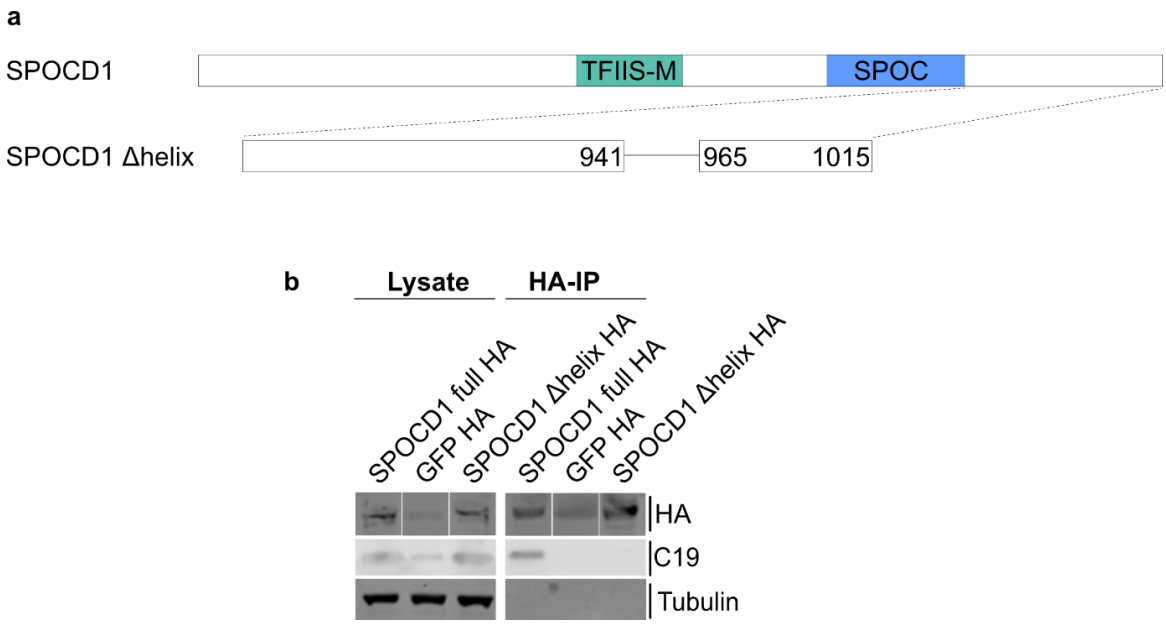

Figure 28: C19ORF84 interacts with the SPOCD1 C-terminal SPOCD1 α-helix.

a, Schematic of the SPOCD1 Δhelix construct tested for interaction. TFIIS-M, green; SPOC, blue. **b**, Representative western blot analyses of n = 3 immunoprecipitations of the SPOCD1 Δ helix construct with C19ORF84 in HEK293T cells.

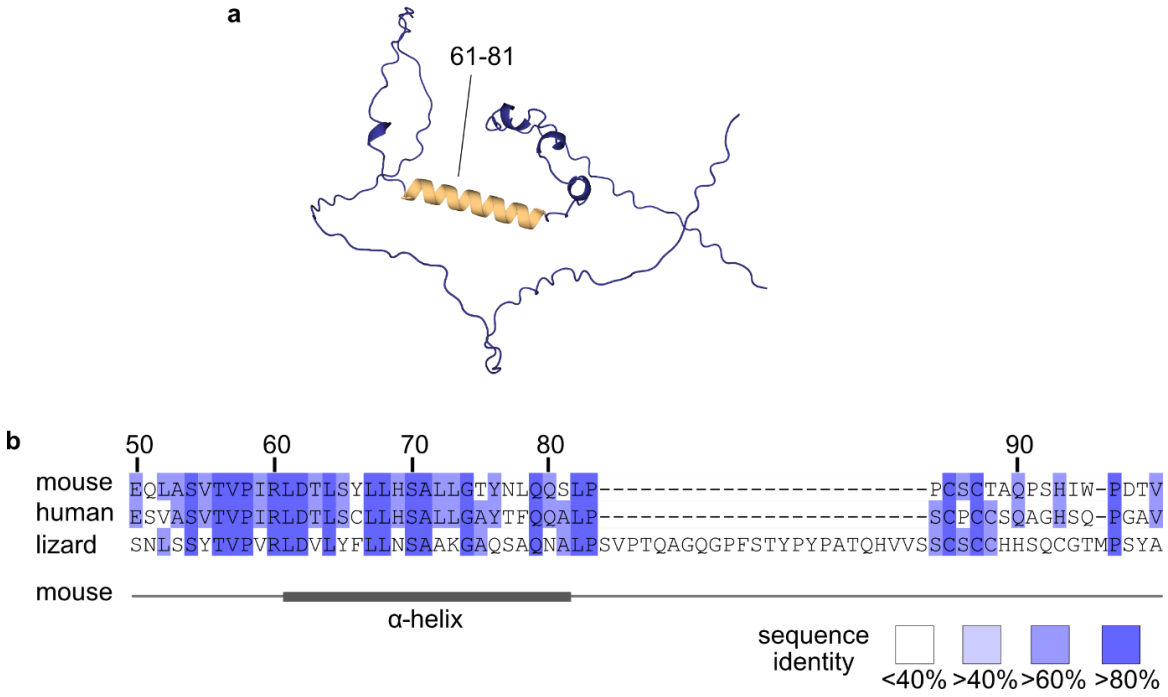

Figure 29: C19ORF84 contains a conserved α-helix.

a, AlphaFold2 prediction of C19ORF84 (H3BKT1), with the conserved α-helix highlighted in yellow. **b**, Multiple sequence alignment of the C19ORF84 α-helix in different species. Numbering for mouse C19ORF84 above the sequences. The secondary structure elements of mouse C19ORF84 are described below. Sequences are coloured according to sequence identity.

To determine if this C19ORF84 helix is important, I generated a SPOCD1 construct with a deletion of only the α-helix (61-81) (Figure 30a). With this construct, I still observed a weak interaction with C19ORF84 (Figure 30b). However, in the multiple sequence alignment, the residues in front of the helix were also conserved (Figure 30b). Therefore, I extended the C19ORF84 helix deletion to amino acid 61-90 which resulted in loss of interaction with SPOCD1 (Figure 30b). Although those additional nine amino acids are not part of the α-helix, they seem to form a secondary structure (Figure 30c).

In summary, my work showed that SPOCD1 interacts *via* an α-helix spanning residues 942 – 964 with the structured area of C19ORF84 within amino acids 61 – 90.

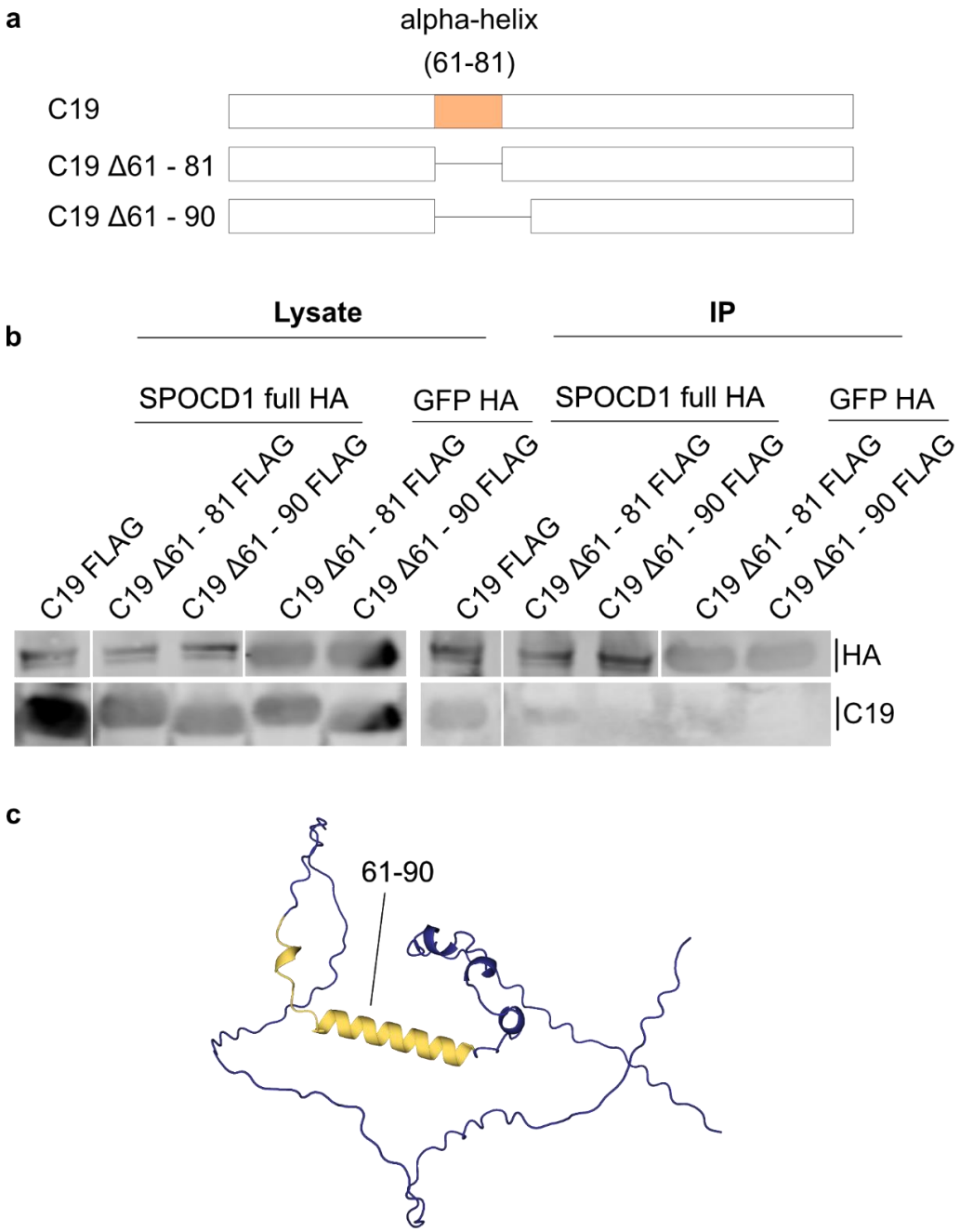

Figure 30: SPOCD1 interacts with the conserved α-helix of C19ORF84.

a, Schematic of the C19ORF84 deletion constructs tested for interaction. Residue 61-81 in orange. **b**, Representative western blot analyses of n = 3 immunoprecipitations of the C19ORF84 deletion constructs with SPOCD1 in HEK293T cells. **c**, AlphaFold2 C19ORF84 (H3BKT1) prediction with amino acids 61-90 highlighted in yellow.

3.6 The SPOCD1 SPOC domain

SPOC domains are known as protein-protein binding modules (Appel, Franke, et al. 2023). However, I could not define an interactor of the SPOCD1 SPOC domain *via* my HEK293T cell IP-MS approaches in Section 3.3.1. Therefore, I decided to follow a new strategy and studying the conservation and structure of the SPOCD1 SPOC domain.

The SPOC domain of SPOCD1 is highly conserved within its orthologue, extending as far as the coelacanth (Figure 31). Even vertebrate sequences from PHF3, DIDO1, SHARP/SPEN and RBM15 orthologues showed conservation of key residues within the SPOC domain (Figure 31).

The SPOC domain of mouse SPOCD1 was purified using a three-step purification procedure (Figure 32). Briefly, I purified the domain *via* its GST-tag using affinity batch purification (Figure 32a). After the 3C-digest to remove the tag and dialysis overnight (Figure 32b), the sample was loaded on a Resource S IEX column (Figure 32c, d). Finally, the domain was purified by size exclusion chromatography (Superdex 200pg column) (Figure 32e, f). The protein was concentrated, and an initial 96-well screen was set up. Shortly after, the first crystals were observed in one condition (Figure 33).

I optimised the crystal condition slightly and collected the crystallographic data. Finally, I solved the structure of the SPOC domain at 1.7 Å (Table 8). This allowed me to compare it to the published SPOC domains of other proteins.

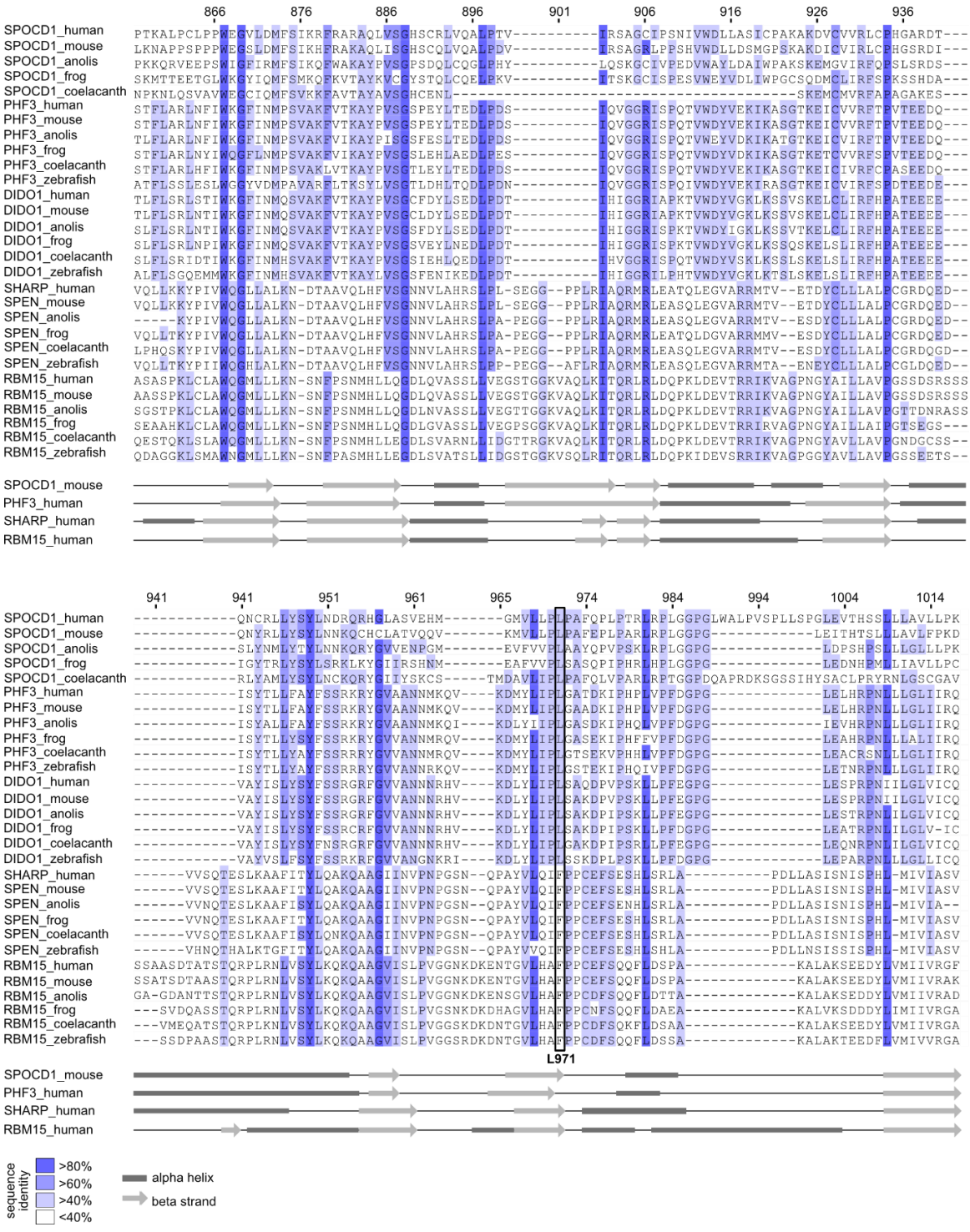

Figure 31: Multiple sequence alignment of SPOC domains.

Structure-based multiple sequence alignment of the SPOC domain from SPOCD1 with representative vertebrate sequences from the PHF3, DIDO1, SPEN, and RBM15 orthologues. The numbering of human SPOCD1 is shown above. The secondary structure elements of the mouse SPOCD1 SPOC domain (PDB:8OU1), human PHF3 SPOC domain (PDB:6Q2V), human SHARP SPOC domain (PDB:1OW1; SHARP is the human SPEN orthologue), and human RBM15 SPOC domain (PDB:7Z27) are shown below the sequence, with dark rectangles for α-helices and lighter arrows for β-strands. Sequences are coloured according to sequence identity. The position of the human residue L971 is indicated by a black box.

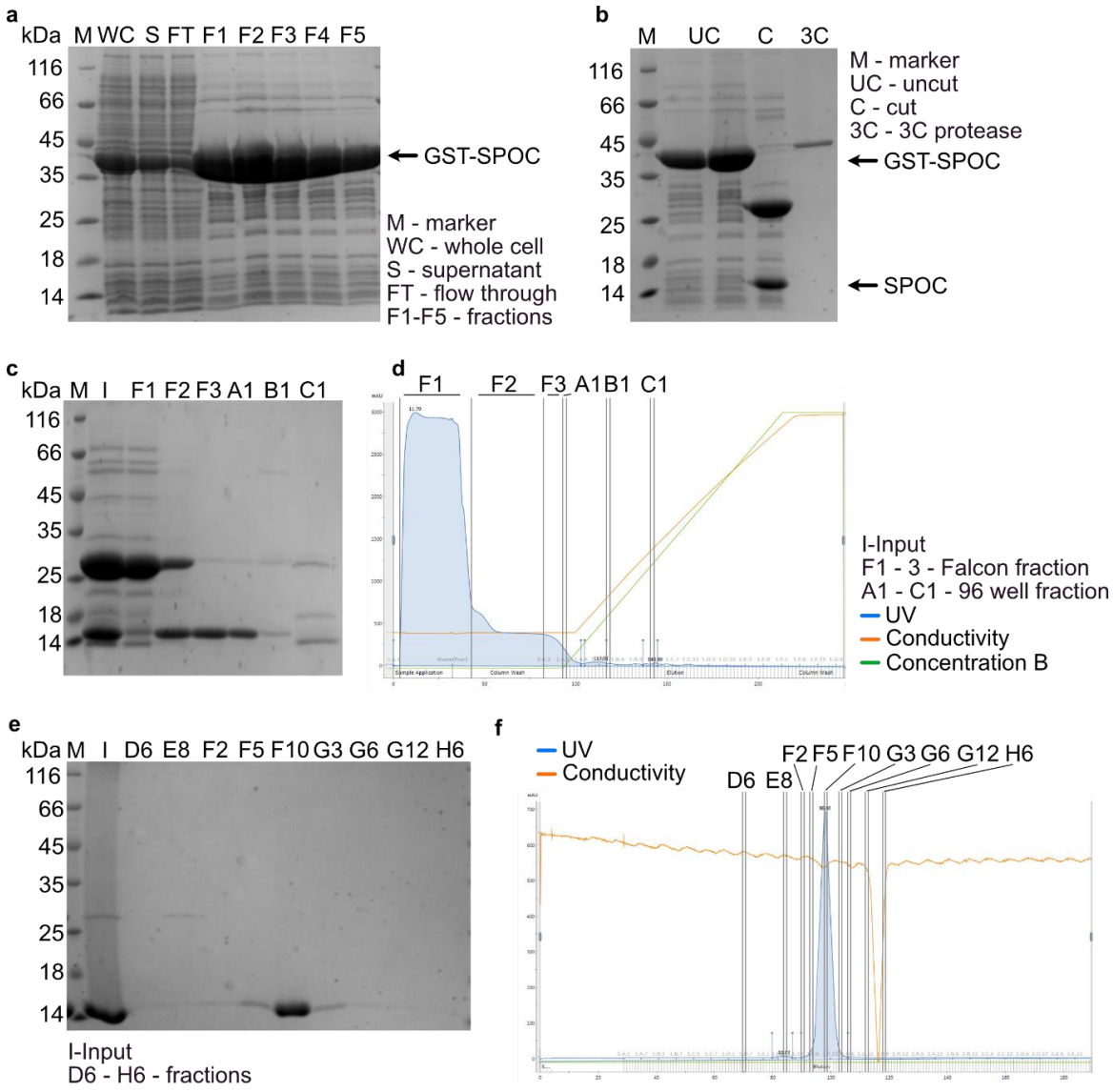

Figure 32: Purification of SPOCD1 SPOC domain.

a, Gel of the GST batch purification. GST-SPOC at 44 kDa. **b.** Gel of samples after dialysis and 3C protease digestion. GST-SPOC at 44 kDa and SPOC (without tag) at 16 kDa. **c**, The Gel of the Resource S run. **d,** Chromatograph of Resource S run shown in **c**. The UV, blue; conductivity, orange; and concentration of buffer B, green; are shown. Fractions loaded onto the gels are shown. **e**, Gel of SEC run. **f,** Chromatograph of the SEC run. The UV, blue; and conductivity, orange; are shown. The fractions loaded onto the gel are indicated. M – marker, WC – whole cell, S – supernatant, A1-H6 – fractions, UC – uncut, C – cut, 3C - 3C protease, I – input

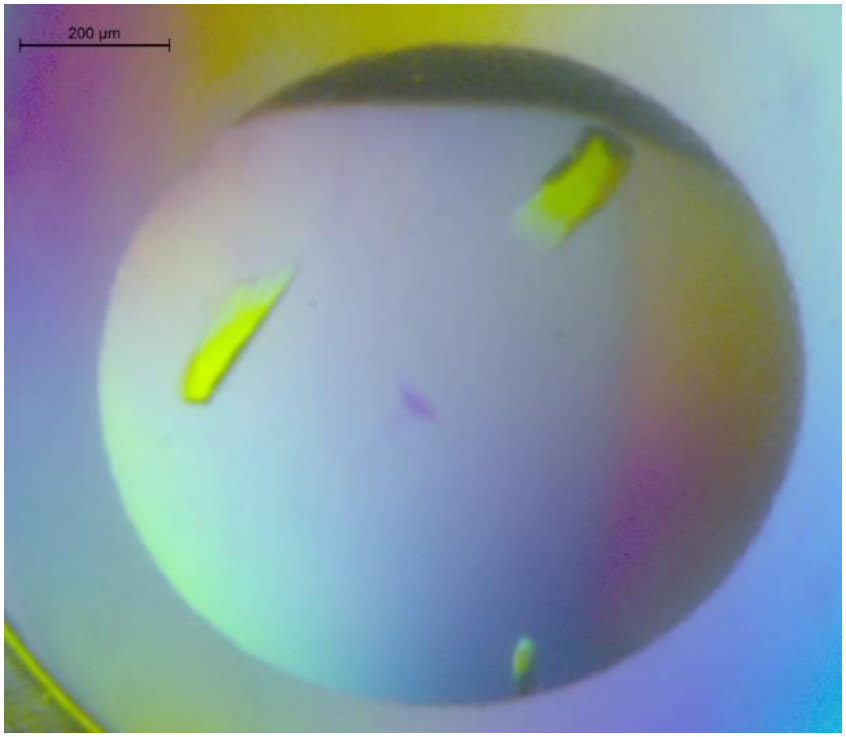

Figure 33: Crystal of the SPOCD1 SPOC domain.
Image of the crystal of the SPOCD1 SPOC domain in a 96-well INDEX screen (hit: B9). Scale bar: 200 µm.

Table 8: X-ray data collection and refinement of the SPOCD1 SPOC domain.

	SPOCD1 SPOC domain
Data Collection	
Beamline	Diamond beamline I04
Wavelength (Å)	0.97949
Space group	$P2_12_12_1$
Unit Cell	a = 54.52 Å, b = 56.62 Å, c = 97.66 Å α =90 °,β = 90 °,γ = 90 °
Resolution (Å)	54.52 – 1.70 (1.79 – 1.7)
Reflections	160141 (16403)
Unique Reflections	33752 (4799)
R_{meas} (%)	0.062 (0.513)
CC (1/2)	0.999 (0.802)
Completeness (%)	99.2 (98.7)
Mean I/σI	13.2 (2.1)
Multiplicity	4.7 (3.4)
Refinement	
R_{work}/R_{free}	0.201/ 0.237
r.m.s Bonds	0.007
r.m.s. Angles	1.037
Ramachandran (%)	0
Allowed	0
Partially allowed	0
Disallowed	0
Total number of atoms	4679
Protein atoms	4403
Water/ligands	276
Average B factor (protein/solvent) (Å^2)	32.0

Like known SPOC crystal structures, the SPOCD1 SPOC domain is composed of eight β-strands that form a distorted β-barrel surrounded by five α-helices (Figure 34a). SPOC domains that do have a defined interactor, show distinct surface charges that facilitate these interactions (Appel, Franke, et al. 2023). Therefore, I calculated the surface charge of the SPOCD1 SPOC domain. However, I could not find any strongly charged areas (Figure 34b). In addition to that, I could not detect any conserved areas on the surface of the SPOCD1 SPOC domain which would indicate a functional importance (Figure 34c).

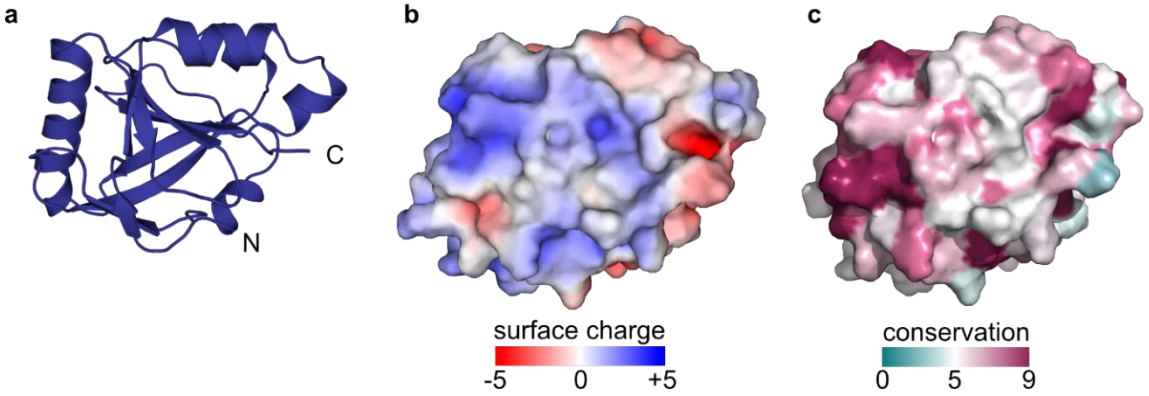

Figure 34: SPOCD1 SPOC Crystal Structure.

a, Cartoon view of the crystal structure of the SPOCD1 SPOC domain (8OU1). The C-and N – terminus was labelled with C or N, respectively. **b**, Surface charge of the SPOCD1 SPOC domain. Colouring according to the surface charge. **c,** Surface conservation of the SPOCD1 SPOC domain. Colouring according to conservation. All panels show the domains in the same orientation.

Since several SPOC domains from different proteins have been structurally characterised, I compared my SPOCD1 SPOC structure with those from PHF3, RBM15, and SHARP. Overall, all SPOC structures had similar folds (Figure 35). Since SPOC domains are known for being phosphor-serine binding modules (Appel, Benedum, et al. 2023), I examined whether I could identify a potential phosphor-serine binding area in the SPOCD1 SPOC structure (Figure 35a). However, based on the surface charge, the SPOCD1 SPOC domain looks different from all other known SPOC structures in the phosphor-serine binding area (Figure 35). PHF3 clearly showed two positively charged patches (Figure 35b). RBM15 and SHARP exhibited a single, very strongly charged, basic patch (Figure 35c, d). In contrast, the SPOCD1 SPOC domain did not have any strongly positively charged patches on its surface (Figure 35a).

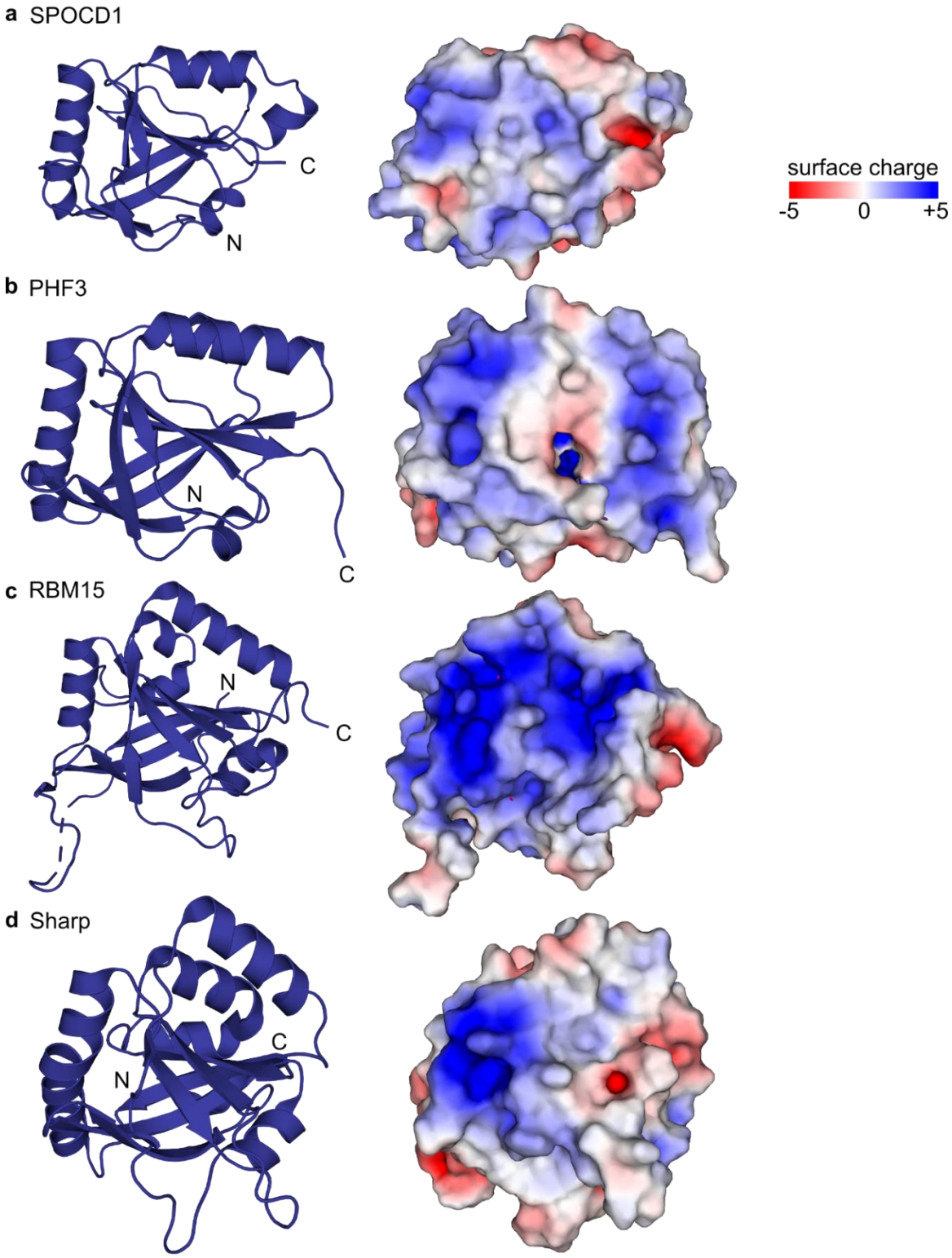

Figure 35: SPOCD1 SPOC domain does not have distinct basic patches.
a, SPOCD1 SPOC domain (PDB:8OU1) in cartoon view and surface charge in the same orientation. **b**, Cartoon and surface charge views of the SPOC domain of PHF3 (PDB:6Q2V). **c**, RBM15 SPOC domain (PDB:7Z27) in the cartoon and surface charge view. **d**, Domain structure and surface charge of the SPOC domain of SHARP (PDB:1OW1). All structures are in the same orientation. Colouring of the surface according to surface charge is indicated.

Additionally, I looked at the binding mode of the RNA polymerase II CTD diheptapeptide phosphorylated on Ser2 with PHF3 (Appel et al. 2021) and at human SHARP with the phosphorylated SMRT peptide (Mikami et al. 2014) (Figure 36). In the first instance, phosphorylated Pol II CTD peptides were anchored within two prominent basic patches (Figure 36b). Similarly, for the binding of the phosphorylated SMRT peptide to SHARP, a strongly positively charged patch was necessary for anchoring the phosphorylated residues (Figure 36c). In SPOCD1, none of the positively charged patches were present (Figure 36a). Based on the SPOCD1 SPOC crystal structure, the known phosphorylated peptides were unlikely to bind to the domain (Figure 36a).

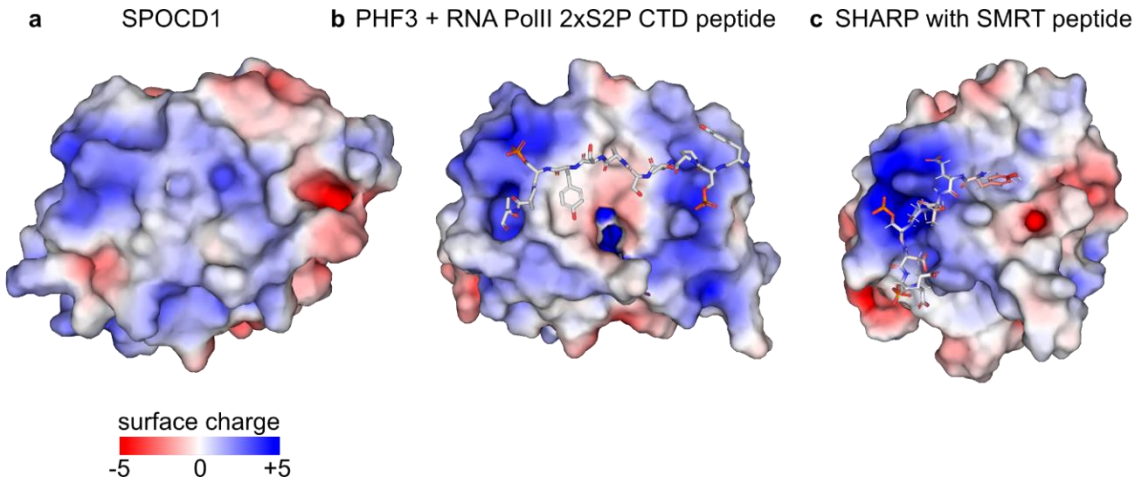

Figure 36: SPOCD1 SPOC domain is unlikely to interact with phosphorylated peptides.

a, SPOCD1 SPOC domain with colour according to its surface charge. **b**, Surface charge of the PHF3 SPOC domain with an RNA polymerase II CTD diheptapeptide phosphorylated on Ser2 (PDB:6IC8). **c**, SPOC domain of SHARP with the phosphorylated SMRT peptide (PDB:2RT5). Colouring according to the surface charge is indicated. All structures are in the same orientation.

Based on this, I can conclude that the SPOCD1 SPOC is different from known SPOC structures and unlikely to facilitate a protein-protein interaction in the same fashion as is known for SPOC domains.

Nonetheless, our collaborators Nadja Rotte, Dr. Birgit Stallmeyer and Prof. Frank Tüttelmann obtained some data that could help us understanding the domain better. They studied men with fertility issues and found one individual with a mutation in the SPOCD1 gene. This individual, M3457, does have an Azoospermia phenotype and is sterile. He carried a homozygous *SPOCD1*$^{c2912T>G}$ variant in exon

15 which resulted in a SPOCD1 variant with leucin 971 mutated to arginine (Leu971Arg). This residue is highly conserved in SPOCD1 as well as in its homologues DIDO1 and PHF3. This mutation is of particular interest to us, as it is located in the SPOC domain.

Since I solved the structure of the mouse SPOCD1 SPOC domain, I was able to inspect this residue in the domain (Figure 37). The human residue L971 corresponds to residue L792 in mice. This mutated leucin was located at the end of β-strand 7 (Figure 37a, b). Its side chain faced inward and was buried within a hydrophobic pocket (Figure 37b). In addition, the surface charge around the L792 residue was not strong (Figure 37c). However, the surface around this residue is highly conserved, indicating functional significance (Figure 37d). I modelled the L792R mutation in the crystal structure and found that the arginine residue clashed with the other residues in the pocket (Figure 37e). In addition, the replacement of the hydrophobic leucin with a positively charged arginine might result in problems during the folding and formation of secondary structures. However, analysis of the mutation based on the crystal structure is only speculative, and only experimental data will provide a clear answer.

Before conducting further experiments, I tested whether the lysine to arginine SPOCD1 mutation affects the expression of the full-length protein. HEK293T cells were transfected with cDNA of L971R human SPOCD1 and L792R mouse SPOCD1. I could see that in both cases, the SPOCD1 mutant was expressed at the wild-type level (Figure 38). This indicates that the full-length mutated SPOCD1 was stable, and that the observed phenotype was likely caused by defects within the SPOC domain of SPOCD1.

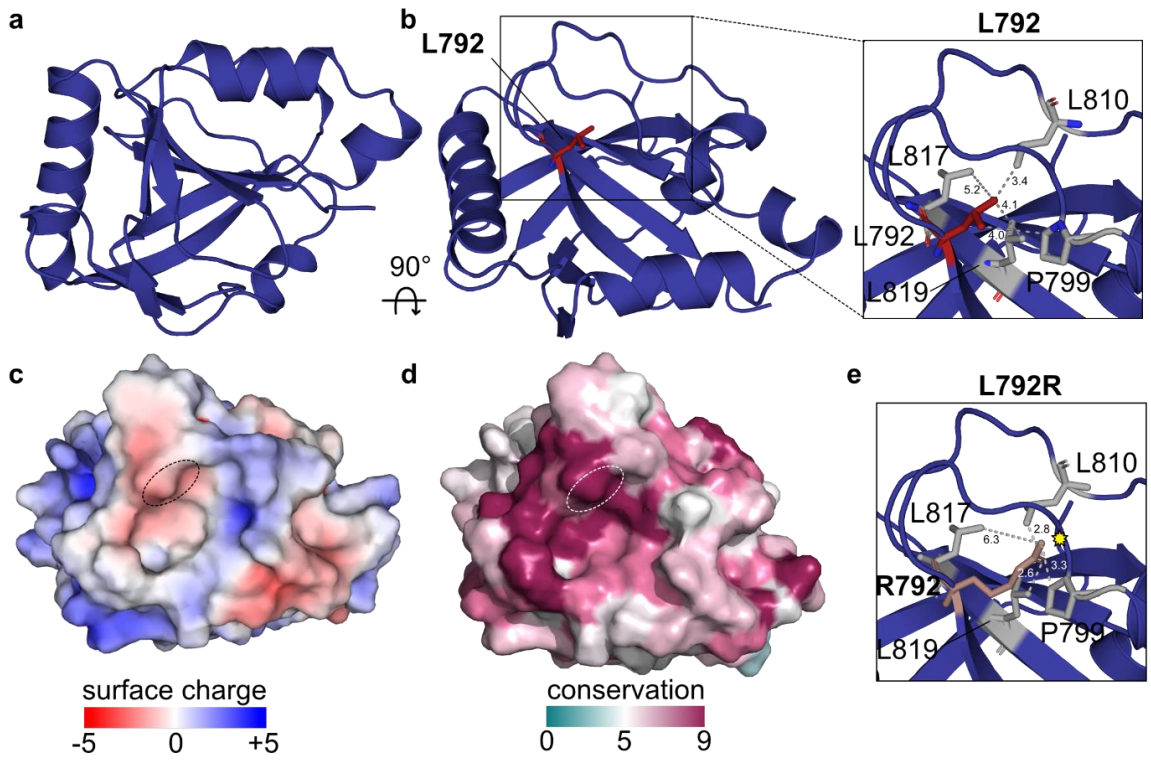

Figure 37: L792R mutation in the SPOCD1 SPOC domain.

a, Cartoon view of the SPOCD1 SPOC domain indicating the C - and N - terminus. **b**, Cartoon view highlighting leucine 792 in red. On the right, zoom in into the region around residue 792. Residues interacting with L792 are shown in grey, with the distance indicated by dotted lines. **c**, Surface charge of the SPOCD1 SPOC domain in the same orientation as **b**. The position of L792 is indicated by the small oval circle. **d,** Surface conservation of the SPOCD1 SPOC domain at position L792, indicated by an oval circle. The orientation is the same as in **b** and **c**. **e**, The L792R mutation modelled in the SPOCD1 SPOC crystal structure. Arginine is shown in pink, interacting residues in grey with distances indicated. Yellow stars indicate clashes between the residues. The same orientation is shown in **b**.

However, it is still unknown whether the SPOCD1 L792R mutation affects folding of the SPOC domain. Also, given that I was able to crystalize the wild-type SPOCD1 SPOC domain, it would be very informative to solve the structure of the L792R SPOC domain. The mutated domain was purified to start the preparation for crystallography. I started with a small-scale purification, where I saw that the SPOC wild-type and mutant were expressed to a similar extent, but the mutant failed to solubilise properly (Figure 39a). However, in the bead sample, I observed a band of the mutant despite it being weaker than the wild-type band (Figure 39a). The WT and L792R constructs were purified in a large-scale purification to get enough for crystallography. Again, I processed both in parallel. After the first purification step, the GST pull-down, I purified decent amounts of the wild-type

SPOC domain, but failed to do so for the L792R SPOC domain (Figure 39b). I also attempted to continue the purification using IEX. While the wild-type SPOC domain behaved well, I lost everything of the L792R SPOC domain (data not shown) making it possible to generate any crystallographic data.

In conclusion, the mutant SPOC L792R behaves differently than the wild-type SPOC domain. While full-length human and mouse mutant SPOCD1 proteins are expressed in HEK293T cells, mutations within the SPOC domain affect the solubility, folding, and/or stability of the domain.

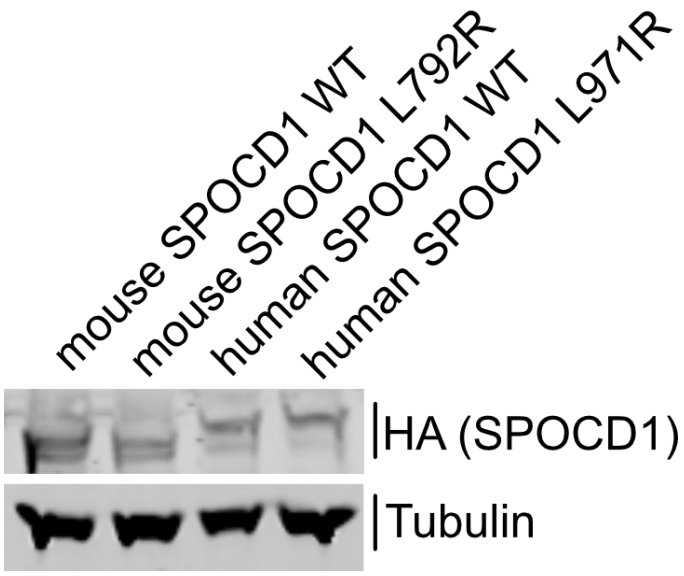

Figure 38: Full-length SPOCD1 with leucine to arginine mutation is expressed in HEK293T cells.

WB of lysates from HEK293T cells transfected with the indicated constructs.

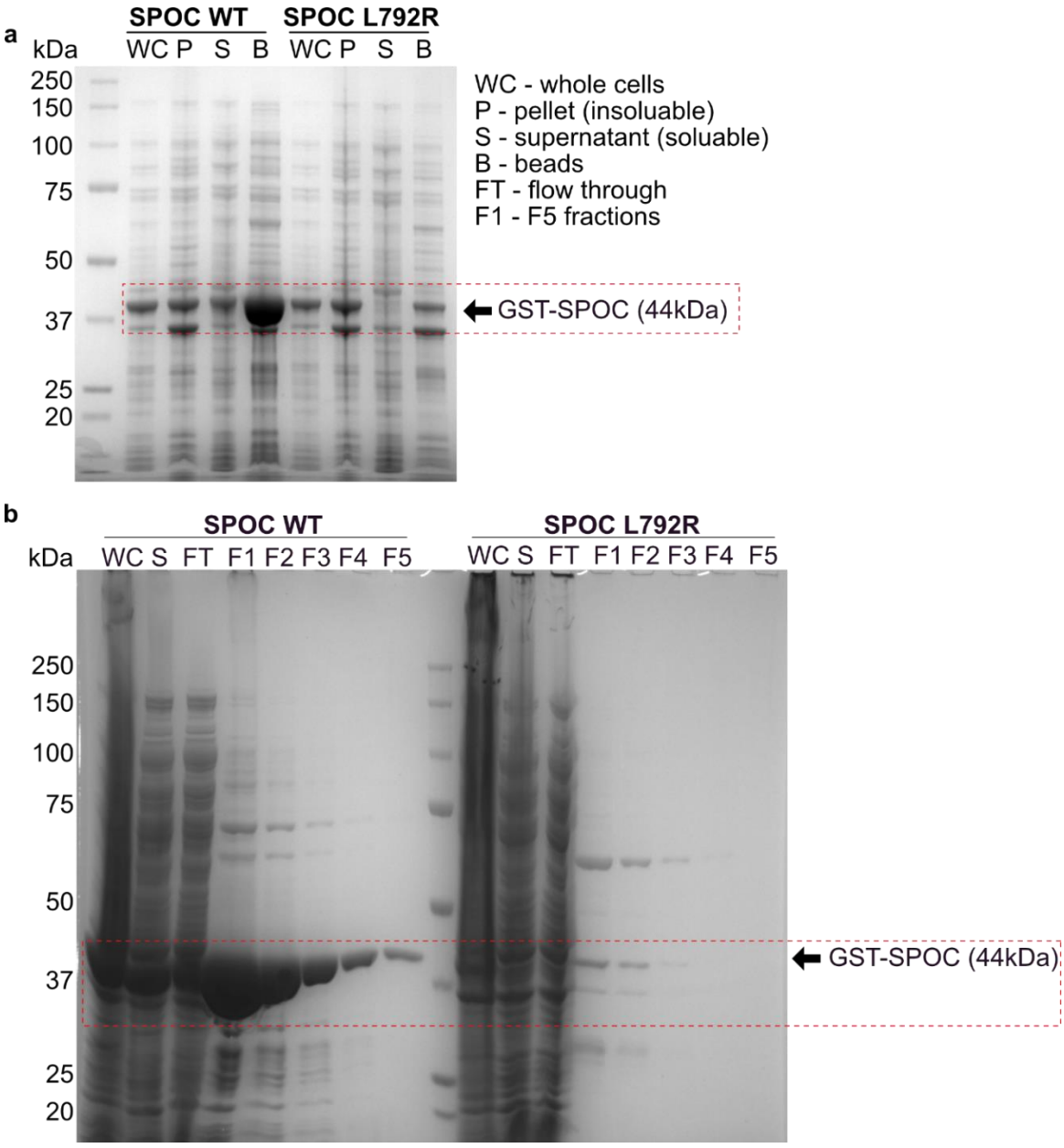

Figure 39: SPOCD1 SPOC domain cannot be purified.

a, Coomassie gel of small-scale pull-down of SPOCD1 SPOC wild-type (WT) and L792R. **b**, Coomassie gel for large-scale purification of wild-type and L792R SPOCD1 SPOC. The GST-SPOC was expected to be 44 kDa. WC – whole cell, P – pellet (=insoluble), S – supernatant (=soluble), B – beads, FT – flow through, F1 – F5 – fractions. n=2 for **a** and **b**

3.7 Conclusion

In summary, this chapter studies some of the interactors of SPOCD1 and defines the functions of different areas of SPOCD1. I could show that SPOCD1 interacts with overexpressed C19ORF84 and endogenous levels of TPR and SPIN1 in HEK293T cells. In addition, IP-MS demonstrated that TPR interacts with the TFIIS-M domain of SPOCD1. However, the interaction between SPIN1 and C19ORF84 was independent of the SPOCD1 TFIIS-M or SPOC domains. More precisely, the C19ORF84's residue 61-90 interacts with SPOCD1 *via* an α-helix spanning amino acid 942-964. In addition, I solved the structure of the SPOCD1 SPOC domain at 1.7 Å. Although the domain is different from all known SPOC structures and its function remains unknown, the human mutation indicates its importance.

4 Two-factor authentication underpins the precision of piRNA-directed LINE1 DNA methylation

4.1 Contribution

This chapter is a manuscript that has been submitted to *Nature* and is currently under revision. The co-authors of this study are Dr. Ansgar Zoch (A.Z.), Dr. Rebecca V. Berrens (R.V.B.), Shaun Webb (S.W.), Dr. Yuka Glinka (Y.G.), Dr. Tania Auchynnikava (T.A.), Martina Schito (M.S.), Tamoghna Chowdhury (T.C.), Dr. Juan Zou (J.Z.), Prof. Juri Rappsilber (J.R.), Prof. Robin C. Allshire (R.C.A.), Dr. Joan Barau (J.B.), Prof. Atlanta G. Cook (A.G.C.) and Prof. Dónal O'Carroll (D.O'C).

The contributions of this study are as follows: I contributed to the design, execution, and analysis of most experiments. A.Z. performed the IF and the IP-MS experiments, under the guidance of T.A., J.R. and R.C.A.. R.V.B., S.W. and Y.G. performed the bioinformatic analysis of the Methyl-seq, ChIP and RNA-seq data, respectively. M.S. and T.C. performed molecular biology and histological experiments. J.Z. performed the analysis of the CL-MS data. J.B. contributed to analyses. D.O'C. and A.G.C. supervised this study. D.O'C. conceived this study. D.O'C. and I wrote the final version of the manuscript.

4.2 Introduction

In Chapter 3, I defined two novel interactors of SPOCD1 and solved the structure of the SPOCD1 SPOC domain. However, it remains unknown how SPOCD1 is recruited to the sites of *de novo* DNA methylation in the genome of the developing germline (Zoch et al. 2020). Although the prevailing view is that MIWI2 is solely

responsible for guiding the *de novo* DNA methylation apparatus, we questioned this hypothesis (Ozata et al. 2018). Transposon methylation must be very precise to methylate every single transposon copy and at the same time avoid off-target methylation. This makes it very unlikely that the entire system depends on piRNA-MIWI2 recruitment to the nascent transposon transcripts.

In Chapter 3, we identified SPIN1 as another SPOCD1 interactor. This interaction was observed in foetal testes as well as in HEK293T cells (Figure 19, Figure 21). Could SPIN1 be the missing part that establishes the precision of piRNA-directed *de novo* DNA methylation?

SPIN1 is a chromatin reader and member of the spindlin/Y-linked spermiogenesis-specific transcript repeat (Spin/Ssty) family, which is important for gamete-to-embryo transition (Staub, Mennerich, and Rosenthal 2002). SPIN1 is involved in WNT/TCF-4 signalling (Wang et al. 2012), in RET signalling in liposarcoma (Franz et al. 2015), stimulates the expression of rRNA genes (Wang et al. 2011) and plays a role in antiviral defence (Ducroux et al. 2014). Additionally, SPIN1 is important for chromosome segregation during cell division (Zhang et al. 2008; Sun, Li, and Gui 2010; Choi et al. 2017).

SPIN1 is highly conserved across different vertebrate species (Staub, Mennerich, and Rosenthal 2002; Vallee et al. 2006; de Kloet and de Kloet 2005) and is ubiquitously expressed (Thul et al. 2017). In some cancers, SPIN1 is overexpressed, for example in breast cancer, in non-small-cell lung cancer (NSCLC) and in liver cancer (Li et al. 2017; Song et al. 2018; Chen et al. 2016). It has been shown that overexpression of SPIN1 affects the cell cycle progression and stability of genomic materials (Yuan et al. 2008; Zhang et al. 2008).

SPIN1 knockout mice are not viable and die within two days of birth (Chew et al. 2013). Although mouse oocytes lacking SPIN1 undergo normal folliculogenesis, they are unable to resume meiosis (Chew et al. 2013). This phenotype was observed because SPIN1 is involved in meiotic resumption and regulation of maternal transcripts post-transcriptionally through PDE3A and SERBP1, respectively (Chew et al. 2013).

SPIN1 associates with several histone marks, such as H3K4me3, Histone 4 lysine 20 tri-methylation (H4K20me3), and the dual mark H3K4me3 - Histone 3 arginine 8 asymmetric di-methylation (H3R8me2a) (Wang et al. 2011; Su et al. 2014; Wang et al. 2018). It has been shown that SPIN1 does have high affinity for H3K4me3

(Wang et al. 2011). However, the addition of H3K9me3, resulting in a bivalent histone H3K4me3-H3K9me3 methylation pattern, resulted in the highest affinity (Zhao et al. 2020; Du et al. 2021). Although the unusual dual histone marks H3K4me3 (activating) and H3K9me3 (repressing) are perceived to be mutually exclusive, the bivalent mark has been found in several types of lineage-committed stem cells (Matsumura et al. 2015; Mikkelsen, Ku, et al. 2007; Rugg-Gunn et al. 2010; Bilodeau et al. 2009).

SPIN1 is a 262 amino acid protein that contains three almost identical Tudor-like domains (short: Tudor) (Oh et al. 1997; Zhao et al. 2007). Its crystal structure has been solved, revealing that each of the three Tudor domains adapts a β-barrel-like fold (Zhao et al. 2007) (Figure 40). It has been shown that Tudor 1 associates with H3R8me2a or H3K9me3, Tudor 2 associates with H3K4me3 or H4K20me3 while Tudor 3 is not involved in the recognition of chromatin marks (Su et al. 2014; Yang et al. 2012; Zhao et al. 2020; Wang et al. 2018).

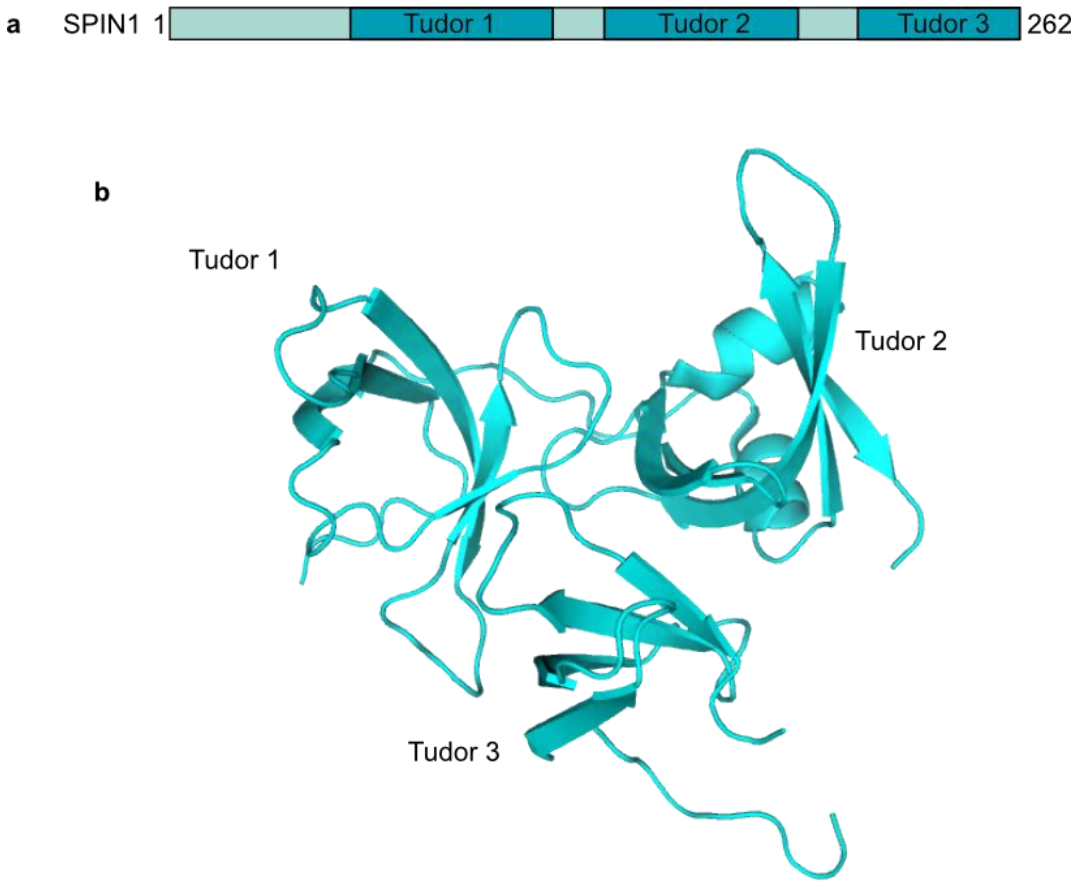

Figure 40: Structure of SPIN1.

a, Schematic overview of SPIN1 with the three Tudor domains indicated. **b**, Crystal structure of mouse SPIN1 with the three Tudor domains indicated, PDB: 2NS2

However, SPINDOC (also C11ORF84) has been identified as interacting with SPIN1 *via* its Tudor domain 3 (Du et al. 2021; Devi et al. 2019) (Figure 41). In addition, the SPIN1-SPINDOC complex exhibits high affinity for the H3K4me3-H3K9me3 histone methylation pattern (Du et al. 2021). It has been shown that the SPIN1-SPINDOC complex reading H3K4me3-H3K9me3 can displace HP1 proteins from ribosomal DNA (rDNA) loci, resulting in the activation of rRNA transcription (Zhao et al. 2020; Du et al. 2021).

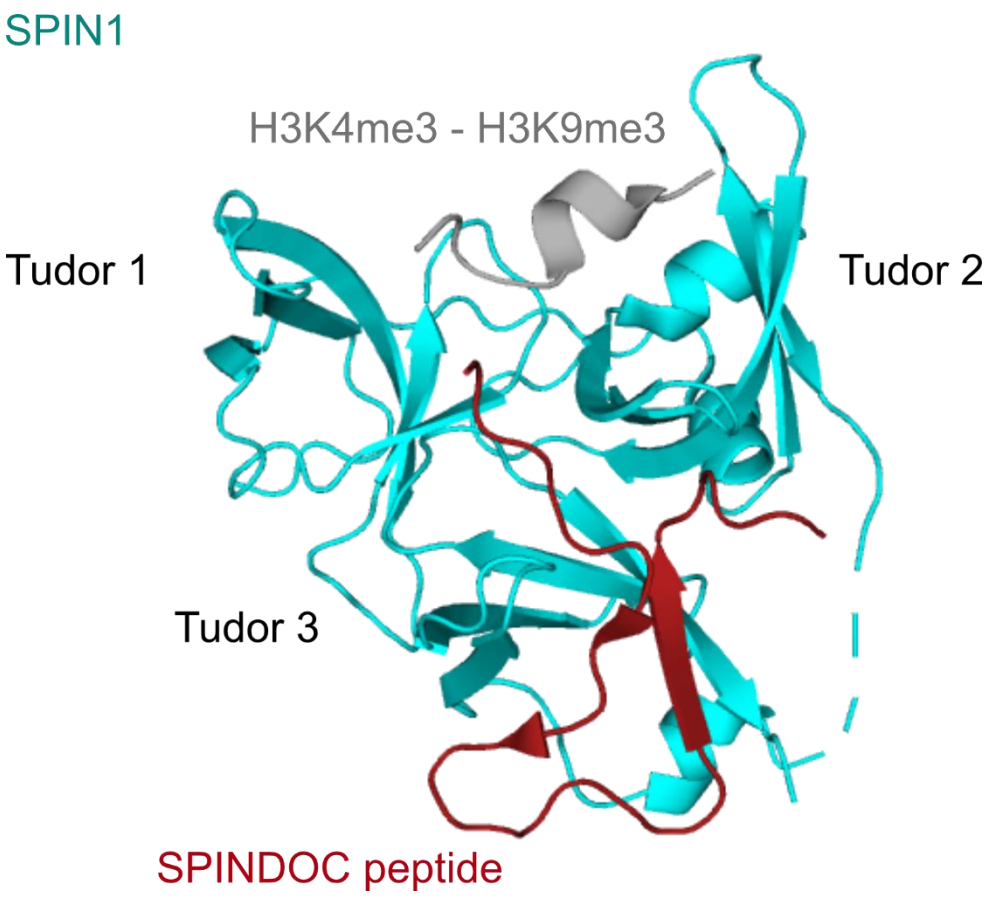

Figure 41:Crystal structure of SPIN1 with SPINDOC (C11ORF84) and H3K4me-H3K9me3 peptide.

SPIN1 is coloured in blue, with the Tudor domains indicated. The SPINDOC (C11ORF84) peptide is shown in red, and the histone peptide (H3K4me3-H3K9me3) is highlighted in grey. PDB: 7CNA

Recently, another SPIN1 interactor has been discovered (Liu et al. 2023). SPIN1 interacts *via* its Tudor domain 3 with HBx, a regulatory protein of the hepatitis B virus (HBV), while concurrently recognising bivalent H3K4me3-H3K9me3 *via* Tudor 1 and 2 (Liu et al. 2023). Functionally, this allows the virus to switch H3K9me3-enriched repressed covalently closed circular (ccc) DNA into an H3K4me3-marked active state, thus promoting HBV transcription (Liu et al. 2023). In this study, we defined SPIN1 as a direct interactor of SPOCD1. We investigated the role of the SPIN1-SPOCD1 interaction during embryonal development and proposed a novel two-factor authentication model for piRNA-directed *de novo* DNA methylation in mice, including SPIN1.

4.3 Manuscript: Two-factor authentication underpins the precision of piRNA-directed LINE1 DNA methylation

4.3.1 Abstract

The PIWI-interacting RNA (piRNA) pathway guides the DNA methylation of young active transposons during male mouse germline development. piRNAs tether the PIWI protein MIWI2 (PIWIL4) to the nascent transposon transcript that results in DNA methylation through SPOCD1. Transposon methylation requires exacting precision: all copies need to be methylated yet, at the same time, off-target methylation must be avoided. However, the underlying mechanisms that ensure this precision remain unknown. Here, we show that SPOCD1 directly interacts with SPIN1, a chromatin reader that primarily binds H3K4me3 and this association is augmented by H3K9me3. The prevailing assumption is that all molecular events required for piRNA-directed DNA methylation occur after the engagement of MIWI2. Interestingly, we find that SPIN1 expression precedes that of both SPOCD1 and MIWI2. Furthermore, we demonstrate that young LINE1s, but not old copies are marked by H3K4me3 and H3K9me3 prior to the initiation of piRNA-directed DNA methylation. We generated a *Spocd1* separation-of-function allele in the mouse that encodes a SPOCD1 variant that no longer interacts with SPIN1. We found that the SPOCD1-SPIN1 interaction is essential for spermatogenesis and piRNA-directed DNA methylation of young LINE1 elements. We propose that young LINE1 elements require a two-factor authentication process for DNA methylation, the first being the recruitment of SPIN1-SPOCD1 to licence the locus and the second is MIWI2 engagement with the nascent transcript, which is the trigger for methylation. In summary, independent events that licence, and trigger methylation underpin the basis of precision.

4.3.2 Results and Discussion

Young active transposable elements are a fundamental threat to the germline. The mouse genome is currently battling LINE1 and IAP elements (Dewannieux et al. 2004; Goodier et al. 2001; Naas et al. 1998). Failure to silence transposons in the germline results in infertility (Aravin et al. 2007; Bourc'his and Bestor 2004; Carmell

et al. 2007). DNA methylation is a key mechanism that represses transposons (Walsh, Chaillet, and Bestor 1998). However, this presents a major vulnerability to the germline because DNA methylation is erased and reset during germ cell development (Greenberg and Bourc'his 2019). The piRNA pathway defends the germline during this period of hypomethylation (Ozata et al. 2018). It post-transcriptionally silences young active transposons and directs their DNA methylation (Ozata et al. 2018). piRNAs are small RNAs that are bound to PIWI proteins (Ozata et al. 2018). Through base complementarity, piRNAs identify active transposons and tether MIWI2 to these nascent transcripts (Ozata et al. 2018). This sets in motion events that culminate in the recruitment of the *de novo* methylation machinery. We recently defined MIWI2 interactomes from foetal gonocytes and discovered SPOCD1, an essential factor that connects the piRNA and *de novo* methylation machineries *in vivo* (Schopp et al. 2020; Zoch et al. 2020). piRNA-directed transposon methylation requires precision. Failing to methylate every active transposon would be detrimental to the genomic integrity of the germline whereas aberrant off-target methylation could result in germline-transmitted epimutations. piRNAs endow MIWI2 with the specificity to identify active transposon loci and, through tethering, trigger silencing. However, whether other mechanisms contribute to identifying active transposons and the exacting precision of the pathway remains unknown.

SPOCD1 accumulates in the nucleus prior to the expression of MIWI2 during male germ cell development (Zoch et al. 2020). SPOCD1 expression is first observed in foetal gonocytes at embryonic day 14.5 (E14.5), whereas MIWI2 follows a day later where it is mostly cytoplasmic until E16.5 when a significant nuclear fraction is observed (Zoch et al. 2020). This pattern of successive accumulation could indicate that the recruitment of SPOCD1 to transposons may occur independently of MIWI2. The treatment of unfixed foetal gonocytes with RNase A results in the loss of nuclear MIWI2 staining (Schopp et al. 2020). Interestingly, the nuclear localization of SPOCD1 is insensitive to RNase A treatment (Figure 42a and Figure 47a). Furthermore, MIWI2-deficiency does not impact SPOCD1 nuclear localization (Figure 42b and Figure 47b). Together, these observations indicate that SPOCD1's recruitment to chromatin is independent of MIWI2. SPOCD1 does not contain any known chromatin-binding domains. We therefore interrogated the SPOCD1 IP-MS data from E16.5 foetal testis for chromatin-binding proteins and

found SPIN1 as a highly enriched, high-confidence associated protein (Figure 42c) (Zoch et al. 2020). SPIN1 is a chromatin reader that is comprised of three Tudor-like domains (TLD1-3). SPIN1 is principally a high-affinity H3K4me3 reader and TLD2 recognises this transcription-associated chromatin mark (Wang et al. 2011; Yang et al. 2012; Santos-Rosa et al. 2002). TLD1 binds heterochromatin-associated H3K9me3 and this interaction increases the overall affinity of SPIN1 chromatin binding (Bannister et al. 2001; Lachner et al. 2001; Du et al. 2021). The third TLD (TLD3) does not contain an aromatic cage and mediates interactions with other proteins. We next sought to identify which portion of SPOCD1 is required to associate with SPIN1. To this end, we co-expressed SPOCD1 or fragments of it with SPIN1 in HEK293T cells and tested their ability to interact. Full-length SPOCD1 and the N-terminal 1-409 amino acids (fragment 1) co-precipitated SPIN1 (Figure 42d). The interaction was further mapped to amino acids 205-409 (fragment 1b) (Figure 42d). By sequentially deleting 10 to 20 amino acids segments of the interacting fragment of SPOCD1, we mapped the SPIN1 association region to 20 amino acids (328-347) (Figure 42e). These amino acids of SPOCD1 are predicted to fold into a β-hairpin in the AlphaFold2 model (Figure 42f) (Jumper et al. 2021; Tunyasuvunakool et al. 2021). Furthermore, fusion of 20 amino acids (327-346) with GFP was sufficient to confer SPIN1 interaction (Figure 42g). Finally, we demonstrated that the SPOCD1-SPIN1 interaction can be recapitulated using recombinant proteins. (Figure 42h-i).

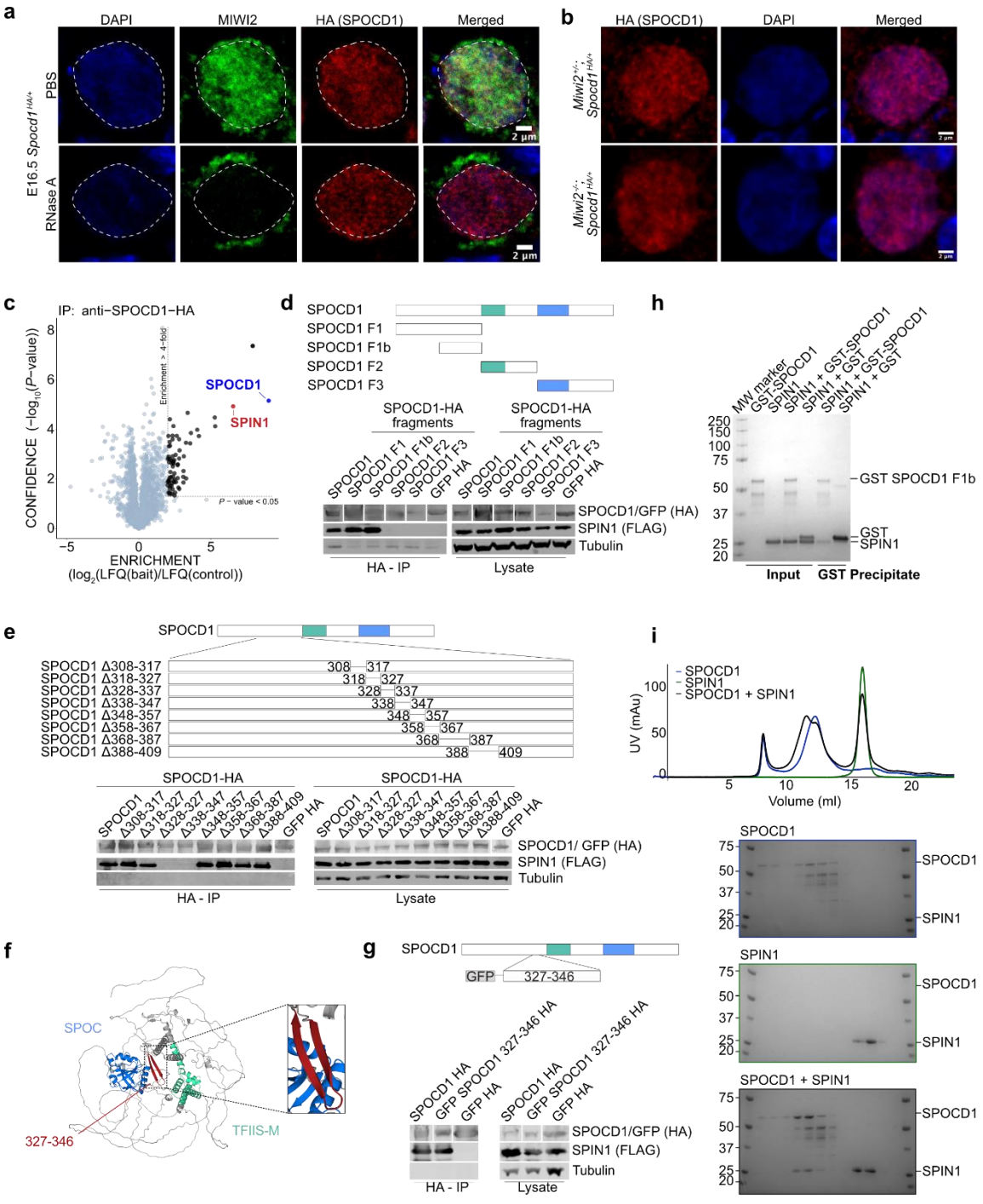

Figure 42: SPOCD1 directly interacts with the chromatin reader SPIN1.

a, MIWI2 (green), HA (red) and DAPI (blue) staining of E16.5 foetal testis sections from *Spocd1*$^{HA/+}$ mice treated with PBS or RNase A prior to fixation. **b**, HA (red) and DAPI (blue) staining of E16.5 foetal testis sections from E16.5 *Miwi2*$^{-/-}$*;Spocd1*$^{HA/+}$ and *Miwi2*$^{+/-}$*;Spocd1*$^{HA/+}$ mice. Images of (**a**) and (**b**) are representative of *n* = 3 biological replicates. Scale bars in (**a**) and (**b**) are 2 µm. **c**, Volcano plot showing enrichment (log_2(mean LFQ ratio of anti-HA immunoprecipitates from *n* = 4 *Spocd1*$^{HA/HA}$*/wild-type*) E16.5 foetal testes) and statistical confidence (log_{10}(*P*-value of two-sided Student's t-test)) of proteins co-purifying with HA-SPOCD1 (data from (Zoch, Auchynnikava, Berrens, Kabayama, Schopp, et al. 2020)). **d-e**, Representative western blot analyses of *n* = 3 immunoprecipitations of the indicated SPOCD1 constructs with SPIN1 in HEK293T cells. **f**,

AlphaFold2 structure prediction of mouse SPOCD1 (B1ASB6) with indicated domains highlighted. **g,** Representative western blot analyses of *n* = 3 immunoprecipitations of the indicated mouse SPOCD1 constructs with SPIN1 in HEK293T cells. **h**, Representative Coomassie gel image of *n* = 3 co-precipitation experiments with the indicated recombinant proteins. **i**, Analytical size exclusion chromatography of the SPOCD1-SPIN1 complex. The top panel shows a representative chromatogram for each of the runs superposed. Below, separate Coomassie gels of each run are shown with samples from the same set of fractions on each gel (*n* = 2).

AlphaFold2 co-folding of SPOCD1 with SPIN1 predicts that the SPOCD1 β-hairpin interacts with the TLD3 of SPIN1 (Figure 43a) (Jumper et al. 2021; Tunyasuvunakool et al. 2021). Cross-linking mass spectrometry (CL-MS) confirmed AlphaFold2's prediction with crosslinks found adjacent to SPOCD1's β-hairpin and TLD3 of SPIN1 (Figure 43b). SPIN1 is an extremely conserved protein (Figure 48a) and is found throughout vertebrates (Figure 43c). We could retrieve full-length SPOCD1 coding sequences for coelacanth (*Latimeria chalumnae*), the green anole lizard (*Anolis carolinensis*), the tropical clawed frog (*Xenopus tropicalis*) but not from salamanders (*Axolotl mexicanum*). SPOCD1 apparently first arose in vertebrates with orthologs found in coelacanth but not ray-finned fish, and was later lost in birds (Figure 43c) (Zoch et al. 2020). The SPOCD1 orthologs are predicted to have a similar overall fold to mouse SPOCD1 but only lizard and frog retain the conserved sequence and predicted β-hairpin structure that mediates SPIN1 interaction in mice (Figure 43d). Indeed, the sequence alignment revealed the coelacanth sequence to be less related to the other orthologs in this region (Figure 43e and Figure 48b). Finally, we demonstrated that the frog and lizard SPOCD1-SPIN1 interaction can be reconstituted using recombinant proteins (Figure 43f). In summary, SPOCD1 interacts with the chromatin reader SPIN1, and this association is conserved from amphibians to mammals.

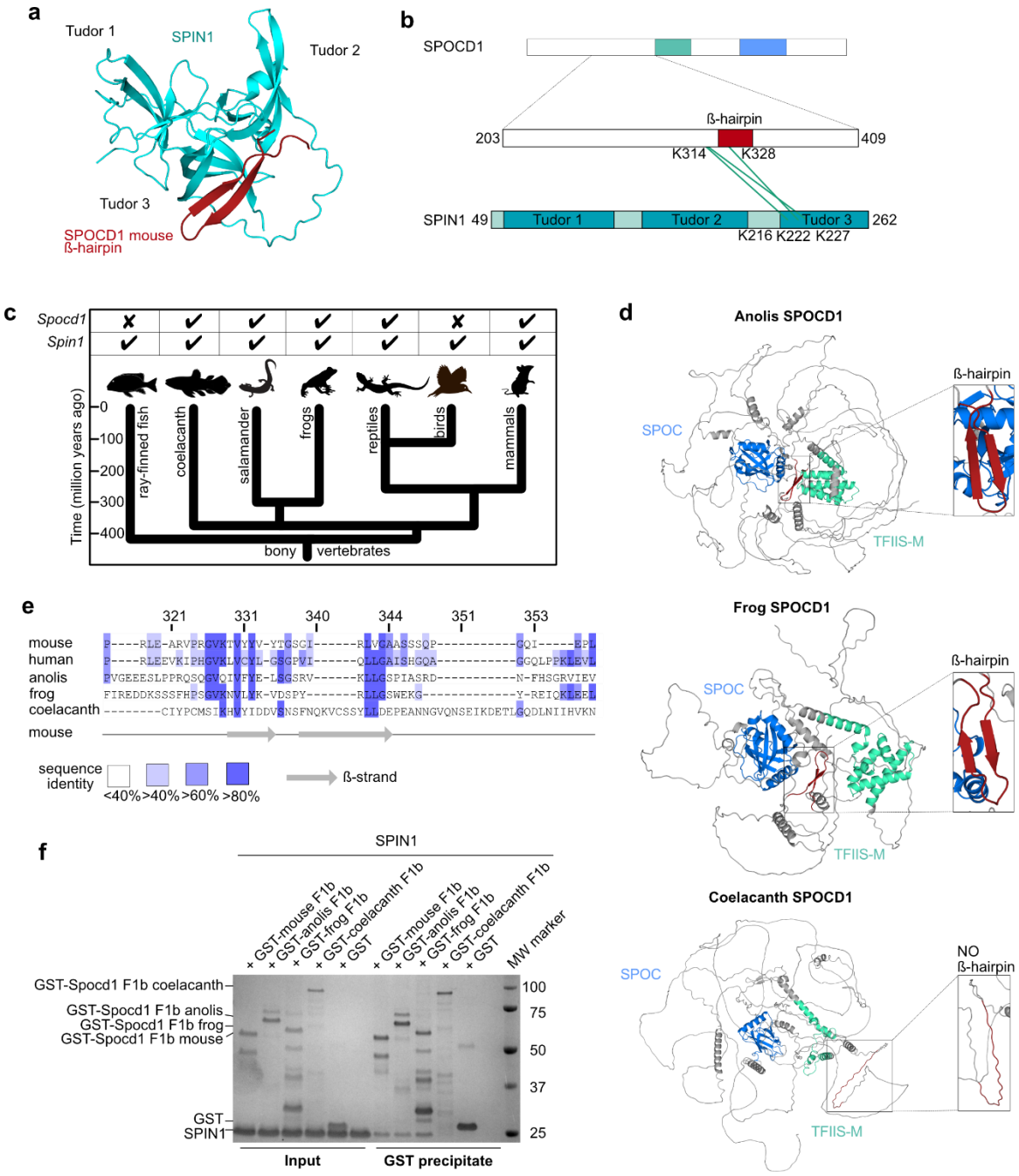

Figure 43: The SPOCD1-SPIN1 interaction is conserved.

a, AlphaFold2 prediction of SPIN1 (Q61142) and SPOCD1 (B1ASB6, only amino acid 326-348 are shown) interaction. **b**, Cross-link mass spectrometry of mouse SPOCD1 fragment F1b (amino acid 203-409) with mouse SPIN1 (amino acid 49-262). Crosslinks are shown in green. **c**, Phylogenic tree from mammals to ray-finned fish showing the presence of SPOCD1 and SPIN1 in the indicated animal clades. **d**, AlphaFold2 prediction of SPOCD1 from *Anolis carolinensis* (anolis, XP_008116112.1, amino acid 183 - 1397), *Xenopus tropicalis* (frog, XP_031752218.1) and *Latimeria chalumnae* (coelacanth, JH127468.5). SPOC domain, TFIIS-M domain and SPIN1-interacting β-hairpin are highlighted. **e,** Multiple sequence alignment of the SPOCD1 SPIN1-interacting β-hairpin region from different species. Numbering for mouse SPOCD1 above the sequences. Secondary structure elements of mouse SPOCD1 are shown below. Sequences are coloured according to

sequence identity. **f**, Representative Coomassie gel image of *n* = 3 co-precipitation experiments with the indicated recombinant SPOCD1 from different species with mouse SPIN1.

We next sought to understand if H3K4me3 and/or H3K9me3 mark young active transposons, the targets of the piRNA pathway. We reanalysed ChIP-seq data from foetal gonocytes purified from several developmental timepoints (Yamanaka et al. 2019). At E13.5, prior to the expression of the piRNA pathway and *de novo* methylation, the genome is fully demethylated. The process of genome and transposon methylation is occurring at E17.5, whereas by E19.5 and P2 the bulk of genomic methylation has been completed (Kato et al. 2007a; Molaro et al. 2014). We examined H3K4me3 at both young and old transposon families. We found that the young LINE1 families (L1M_T, L1M_Gf and L1M_A) but not the old family L1M_F were enriched in H3K4me3 at E13.5 prior to the onset of *de novo* methylation (Figure 44a and Figure 49a). This enrichment was diminished but still present at E17.5 but lost thereafter (Figure 44a and Figure 49a). H3K4me3 enrichment was not observed for the IAPEz and IAPEy families (Figure 44a and Figure 49a). Next, we analysed H3K9me3, for which the IAP families showed a high level of enrichment throughout all timepoints (Figure 44b and Figure 49b). Both young and old LINE1 families showed a peak of H3K9me3 across the promoter region at E13.5, thereafter the enrichment extended across the body of the element (Figure 44b and Figure 49b). The young transposons families contain both young active elements and older inactive copies, these can be roughly distinguished by their divergence from their consensus sequence. We segregated young and old copies within LINE1 families and analysed H3K4me3 and H3K9me3 enrichment. Strikingly, H3K4me3 promoter enrichment is only observed at young LINE1 copies at E13.5 (Figure 44c and Figure 49c). H3K9me3 enrichment was greater in older copies than their younger counterparts at E13.5 but thereafter old and young elements showed similar levels of enrichment (Figure 44d and Figure 49d). In summary, young LINE1 elements are marked by both H3K4me3 and H3K9me3 prior to piRNA-directed DNA methylation. The prevailing view is that engagement of MIWI2 with the nascent transcript is the trigger for all downstream processes that culminate in DNA methylation. However, the fact that young LINE1 elements show a distinct chromatin modification pattern prior to *de novo* genome methylation challenges this view. We hypothesised that H3K4me3, augmented by

H3K9me3, recruits SPIN1 and in turn SPOCD1 to young LINE1 elements prior to the engagement of MIWI2 and this event licences the element for methylation. In support of this hypothesis, we find that SPIN1 is expressed in foetal gonocytes at E13.5 (earliest timepoint analysed) and throughout the process of *de novo* genome methylation (Figure 44e and Figure 50). We tried to perform CUT&RUN for SPIN1 in E13.5 foetal testis but failed likely reflecting a limitation of the antibody. However, we show that SPOCD1 associates with SPIN1 in E14.5 foetal gonads (Figure 44f). In summary, we show that young LINE1 elements are enriched for H3K4me3-K9me3 and SPOCD1 associates with SPIN1 prior to the expression of MIWI2.

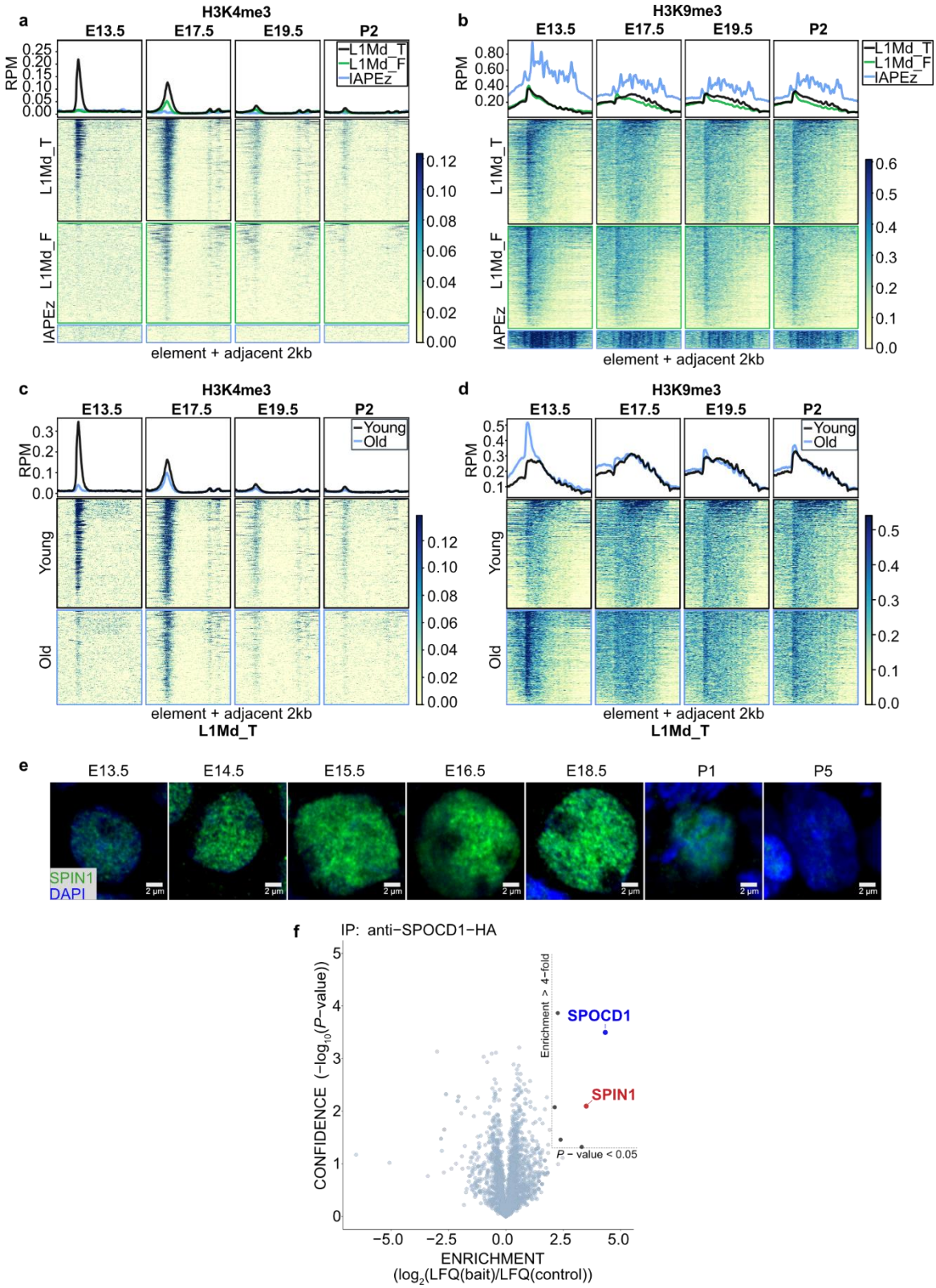

Figure 44: H3K4me3 and H3K9me3 mark young LINE1 elements prior to *de novo* genome methylation.

a-d, Metaplot and heat map of H3K4me3 (**a**, **c**) and H3K9me3 (**b**, **d**) ChIP from foetal gonocytes at the indicated timepoints during mouse development are shown. **a-b** shows data for different transposon families. **c-d** shows young and old elements within the L1Md_T family. Data reanalysed

from (Yamanaka et al. 2019) **e**, The panel shows SPIN1 (green) and DAPI (blue) staining of wild-type foetal testis sections from indicated developmental timepoints. Images are representative of *n* = 3 biological replicates. Scale bars are 2 µm. **f**, Volcano plot showing enrichment (log_2(mean LFQ ratio of anti-HA immunoprecipitates from *Spocd1$^{HA/HA}$/wild-type*)) and statistical confidence (log_{10}(*P*-value of two-sided Student's t-test)) of proteins co-purifying with HA-SPOCD1 from E14.5 foetal testes. *n* = 3.

SPIN1 is expressed beyond the germline, involved in other cellular processes, and required for mouse viability (Chew et al. 2013). Therefore, we decided to identify a SPOCD1 separation-of-function mutation that uncouples SPOCD1 from SPIN1 to understand the importance of this interaction. Mutation of eight amino acids to alanine in one strand of the predicted β-hairpin that mediates SPIN1 binding abrogated the ability of SPOCD1 to co-precipitate SPIN1 when expressed in HEK293T cells (Figure 45a). Furthermore, a recombinant SPOCD1 F1b fragment (Figure 42d) with the 8 alanine mutation no longer interacts with recombinant SPIN1 (Figure 45b). We termed this separation-of-function SPOCD1 mutant SPOCD1-ΔSPIN1. We next engineered the *Spocd1$^{\Delta SPIN1}$* mouse allele (Figure 51). As is the case with *Spocd1$^{-/-}$* mice, *Spocd1$^{\Delta SPIN1/\Delta SPIN1}$* (termed *Spocd1$^{\Delta SPIN1}$*) mice are born in mendelian ratios from *Spocd1$^{\Delta SPIN1/+}$* intercrosses and are indistinguishable from their wild-type litter mates (data not shown) (Zoch et al. 2020). The separation-of-function mutation did not affect the levels or localization of the SPOCD-ΔSPIN1 protein in *Spocd1$^{\Delta SPIN1}$* E16.5 foetal testes compared to wild-type controls (Figure 45c and Figure 51e). Furthermore, both SPIN1 as well as MIWI2 levels and localization in *Spocd1$^{\Delta SPIN1}$* E16.5 foetal testes were indistinguishable from wild-type foetal testes (Figure 45d-e and Figure 51f-g). In summary, the *Spocd1$^{\Delta SPIN1}$* allele encodes a stably expressed SPOCD1 mutant protein and does not impact on SPIN1 or MIWI2 expression. *Spocd1$^{\Delta SPIN1}$* male mice are infertile and have atrophic testes (Figure 45f-g). Detailed histological analyses revealed a complex spermatogenic arrest (Figure 45h). The vast majority of *Spocd1$^{\Delta SPIN1}$* seminiferous tubules show a meiotic arrest that is typical of mutations that affect piRNA-directed transposon methylation (Figure 45h). However, a small number of tubules show cells that have further developed to the round or elongated spermatid stage (Figure 45h). The loss of the SPOCD1-SPIN1 interaction also results in DNA damage and apoptosis (Figure 45i-j). In summary,

the interaction of SPOCD1 with SPIN1 is essential for normal spermatogenesis and male fertility.

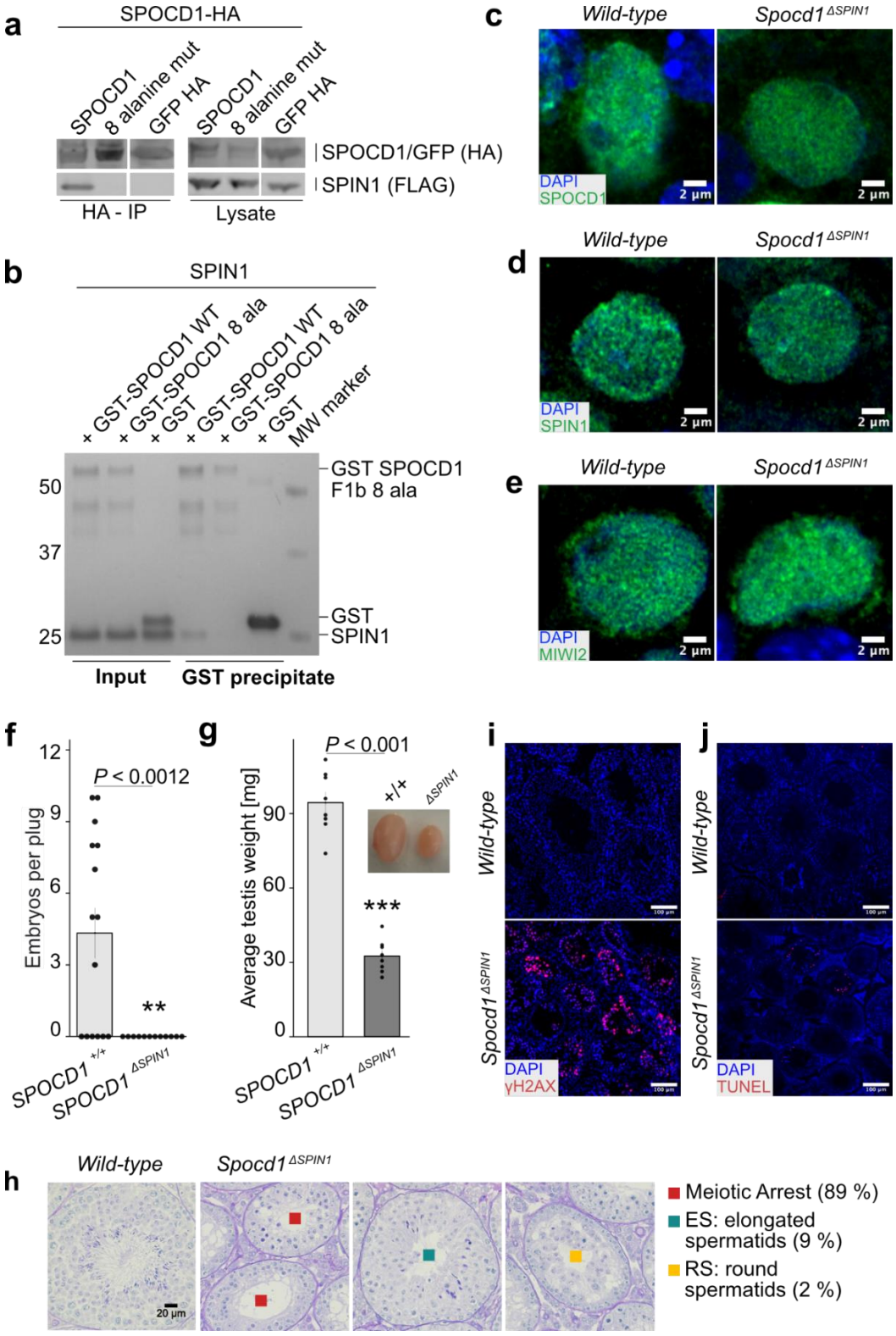

Figure 45: The SPOCD1-SPIN1 interaction is essential for spermatogenesis.

a, Representative western blot analyses of *n* = 3 immunoprecipitations of the mouse wild-type and SPOCD1 8 alanine mutated proteins with SPIN1 in HEK 293T cells. **b**, Representative Coomassie gel image of *n* = 3 co-precipitation experiments with the indicated recombinant proteins. **c-e**, Representative images of E16.5 gonocytes from *n* = 3 wild-type and *Spocd1*$^{\Delta SPIN1}$ mice stained for DNA (blue) and SPOCD1 (**c**), SPIN1 (**d**) or MIWI2 (**e**) in green. Scale bars are 2 µm. **f**, Number of embryos per plug fathered by studs with the indicated genotype mated to wild-type females. Data are mean and s.e.m. from *n* = 6 wild-type (15 plugs total) and *n* = 6 *Spocd1*$^{\Delta SPIN1}$ studs (12 plugs total). **g**, Testis weight of adult mice with the indicated genotype. Data are mean and s.e.m. from *n* = 8 wild-type and *n* = 8 *Spocd1*$^{\Delta SPIN1}$ mice. Insert shows a representative image of testes from wild-type (left) and *Spocd1*$^{\Delta SPIN1}$ (right) mice. **h**, Representative images of PAS & haematoxylin-stained and testes sections of wild-type and *n* = 5 *Spocd1*$^{\Delta SPIN1}$ adult animals, with different types of spermatogenic arrest observed in tubules of *Spocd1*$^{\Delta SPIN1}$ testes indicated. The percentage of each type of tubule is noted on the right. Scale bar is 20 µm. **i-j**, Adult testis sections stained for DNA damage marker γH2AX (red) (**i**) and apoptotic cells (red) by TUNEL assay (**j**) from wild-type and *Spocd1*$^{\Delta SPIN1}$ mice (representative of *n* = 3 mice per genotype for γH2AX and *n* = 2 wild-type plus *n* = 3 *Spocd1*$^{\Delta SPIN1}$ mice for TUNEL). DNA was stained with DAPI (blue). Scale bars, 100 µm. *P* in **f-g** determined by two-sided Student's *t*-test.

The spermatogenic arrest in *Spocd1*$^{\Delta SPIN1}$ mice is indicative of defective transposon silencing and DNA methylation. In agreement with the selective marking of young LINE1 families with H3K4me3-K9me3 prior to the expression of MIWI2, we found the expression of LINE1 ORF1p but not IAP GAG in *Spocd1*$^{\Delta SPIN1}$ adult testis (Figure 46a-b). Furthermore RNA-seq from postnatal day 20 (P20) testis confirmed that the same LINE1 families are deregulated in *Spocd1*$^{\Delta SPIN1}$ and *Spocd1*$^{-/-}$ mice (Figure 46c). The lack of deregulated expression of evolutionary young IAP families (IAPEz) in *Spocd1*$^{\Delta SPIN1}$ mice was also confirmed by this analysis (Figure 46c). We next analysed genome methylation from purified P14 spermatogonia, a timepoint used in previous analyses because it is prior to the onset of the spermatogenic arrest but after completion of *de novo* genome methylation (Schopp et al. 2020; Zoch et al. 2020). The piRNA pathway and SPOCD1 are specifically required for *de novo* DNA methylation of young LINE1 and IAP elements (Schopp et al. 2020; Zoch et al. 2020; Aravin et al. 2007; Carmell et al. 2007; Barau et al. 2016; Jain et al. 2017; Kuramochi-Miyagawa et al. 2008). Accordingly, general genome *de novo* methylation is normal in *Spocd1*$^{\Delta SPIN1}$ adult testis (Figure 46d). Indeed, the loss of the SPOCD1-SPIN1 interaction did not impact genic, intergenic, CpG island and gene promoter regions, or collective transposon DNA methylation levels (Figure 46d). The young LINE1 families L1Md_A, L1Md_Gf and L1Md_T failed to be fully methylated in *Spocd1*$^{\Delta SPIN1}$ spermatogonia, whereas near normal levels of methylation were observed for the

young IAPEz family (Figure 46e). The piRNA pathway directs DNA methylation at the promoters of transposons (Barau et al. 2016; Pastor et al. 2014). A metaplot analysis of methylation levels from *Spocd1*$^{\Delta SPIN1}$ spermatogonia revealed defective *de novo* promoter methylation specifically in young LINE1 families such as L1Md_T, L1Md_A and L1Md_Gf compared to the older L1Md_F family and the IAPEz family (Figure 46f). The overall reduction in promoter methylation in *Spocd1*$^{\Delta SPIN1}$ cells is similar to that observed in *Spocd1*$^{-/-}$ mice (Figure 46f) (Zoch et al. 2020). We next looked at the loss of methylation at individual transposon loci as a function of their divergence from the consensus sequence, which is a proxy for age. This analysis confirmed that the SPOCD1-SPIN1 interaction is required for the methylation of young LINE1 elements within the respective families (Figure 46g). IAPEz element methylation was unaffected in *Spocd1*$^{\Delta SPIN1}$ spermatogonia (Figure 46g).

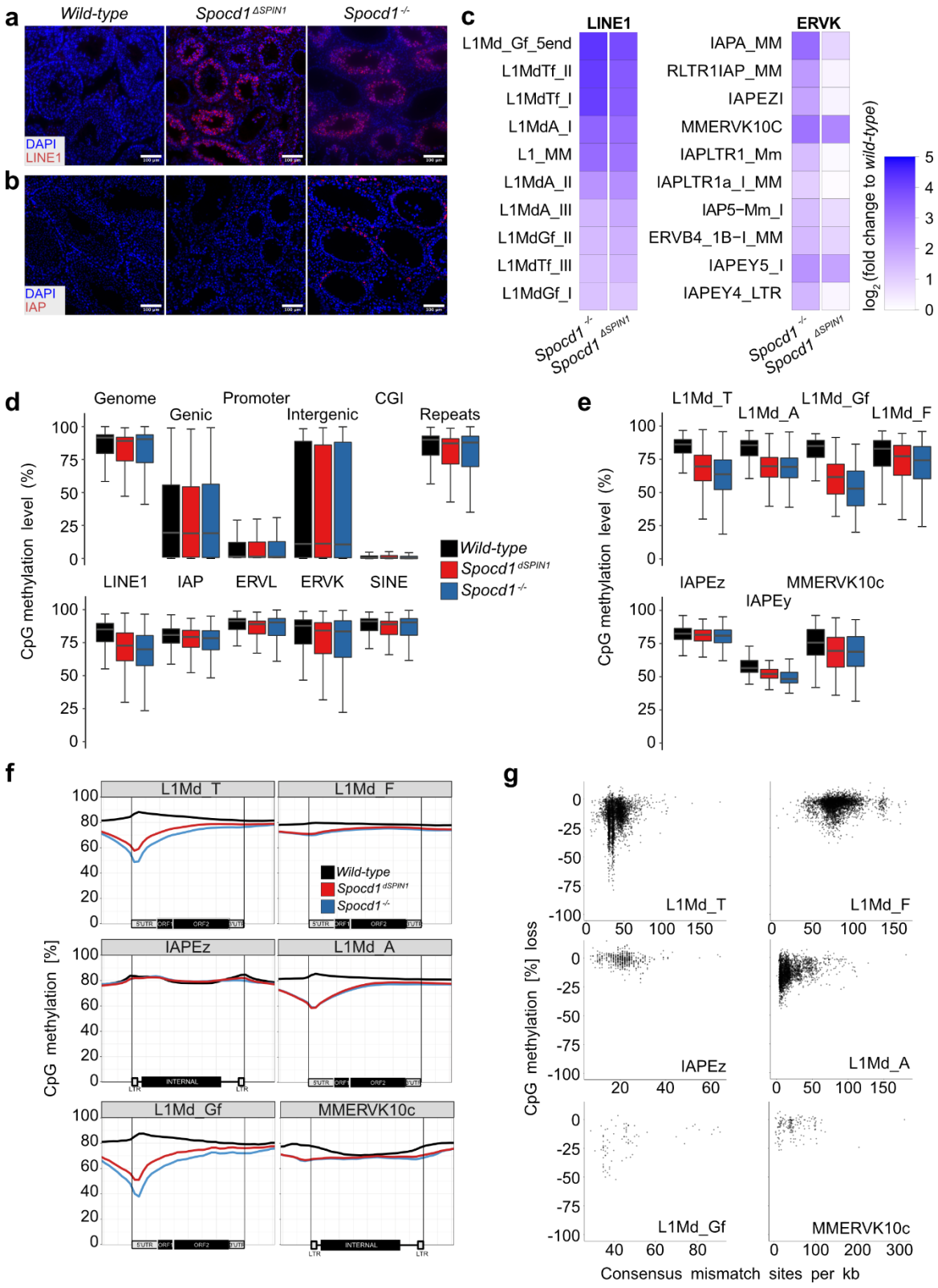

Figure 46: The SPOCD1-SPIN1 interaction is required for *de novo* DNA methylation of young LINE1 elements.

a-b, Representative testis sections of *n* = 3 wild-type, *Spocd1*$^{\Delta SPIN}$ and *Spocd1*$^{-/-}$ stained in red for the LINE1 ORF1p (**a**) or IAP GAG protein (**b**). DNA was stained with DAPI (blue). Scale bars are 100 µm. **c**, RNA-seq heat maps showing fold changes in expression relative to wild-type for the 10

most upregulated LINE and ERVK transposable elements in *Spocd1*$^{-/-}$ P20 testes (n = 3 from each genotype). **d-g**, Genomic CpG methylation analysis of P14 undifferentiated spermatogonia from wild-type (n = 6), *Spocd1*$^{\Delta SPIN}$ (n = 4) and *Spocd1*$^{-/-}$ mice (n = 3) . **d-e**, Percentages of CpG methylation levels of the indicated genomic features (with genic, promoter and CpG island (CGI) regions defined as those not overlapping transposable elements, and intergenic regions as those not overlapping transposable elements or genes) or transposable elements (not overlapping genes) are shown as box plots. Boxes represent interquartile range from 25th to 75th percentile, the horizontal line the median, whiskers the data range of the median ± 2× interquartile range. **f**, Metaplots of mean CpG methylation over indicated transposable element. **g**, Correlation analysis of mean CpG methylation loss relative to wild-type for individual transposable elements of the indicated LINE1 and ERVK families in relation to their divergence from the consensus sequence in *Spocd1*$^{\Delta SPIN}$ spermatogonia.

In summary, the SPOCD1-SPIN1 interaction is predominantly required for the piRNA-directed DNA methylation of young LINE1 elements. Here we show that SPOCD1 is a direct interactor of the chromatin reader SPIN1 and this interaction arose early during tetrapod evolution. H3K4me3, the key determinant of SPIN1 chromatin association specifically marks young LINE1 elements (Wang et al. 2011; Yang et al. 2012). H3K9me3, which biochemically augments SPIN1 chromatin binding, is also found at the same elements (Du et al. 2021). These events as well as the SPOCD1-SPIN1 interaction occur prior to the nuclear localisation of MIWI2 and the process of *de novo* DNA methylation itself. We demonstrate that this interaction is required for spermatogenesis and piRNA-directed LINE1 methylation. The spermatogenic phenotype in *Spocd1*$^{\Delta SPIN1}$ mice differs from *Spocd1*- or *Miwi2*-deficiency where a strict meiotic arrest is observed (Carmell et al. 2007; Zoch et al. 2020). The basis of this difference could lie in the fact that only LINE1s are deregulated in *Spocd1*$^{\Delta SPIN1}$ mice whereas defective LINE1 and IAP silencing are observed in *Spocd1*$^{-/-}$ or *Miwi2*$^{-/-}$ mice (Zoch et al. 2020; Carmell et al. 2007; Kuramochi-Miyagawa et al. 2008). Interestingly, in mice where the PIWI protein MILI has lost its endonuclease activity a similar spermatogenic arrest is observed and only LINE1s are deregulated (De Fazio et al. 2011). How SPOCD1 is recruited to IAPs remains unknown, we speculate that another SPOCD1-associated protein could mediate this recruitment through the recognition of a distinct chromatin signature or sequence motif. The prevailing notion is that all molecular events required for piRNA-directed DNA methylation occur after the engagement of the piRNA-MIWI2 ribonucleoprotein complex with the nascent transcript. Here, we demonstrate that multiple independent and developmentally choreographed events are required for LINE1 piRNA-directed DNA methylation. Our revised model

posits that the recruitment of SPIN1-SPOCD1 through chromatin modification to young LINE1 elements constitutes a first licencing step. The engagement of MIWI2 with the nascent transcript is the second licencing event that triggers the methylation process. In summary, we propose that a two-factor authentication system ensures the precision of LINE1 piRNA-directed methylation.

4.3.3 Extended data figures

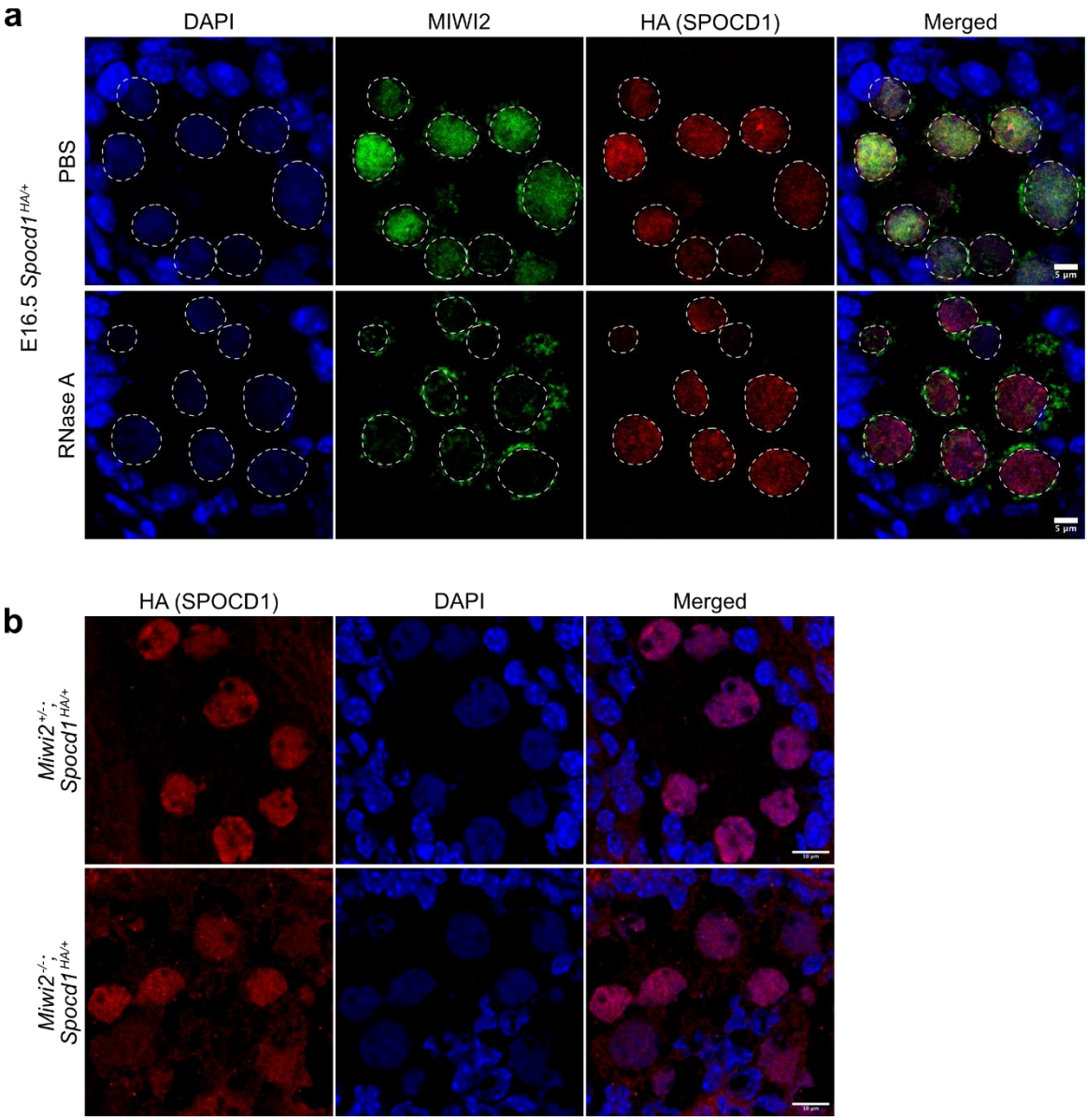

Figure 47: SPOCD1's recruitment to chromatin is independent of MIWI2.
a, MIWI2 (green), HA (red) and DAPI (blue) staining of E16.5 foetal testis sections from *Spocd1*$^{HA/+}$ mice treated with PBS or RNase A prior to fixation. **b,** HA (red) and DAPI (blue) staining of E16.5 foetal testis sections from E16.5 *Miwi2*$^{-/-}$*;Spocd1*$^{HA/+}$ and *Miwi2*$^{+/-}$*;Spocd1*$^{HA/+}$ mice. Images of (**a**) and (**b**) are representative of *n* = 3 biological replicates. Scale bars are 5 µm (**a**) and 10 µm (**b**).

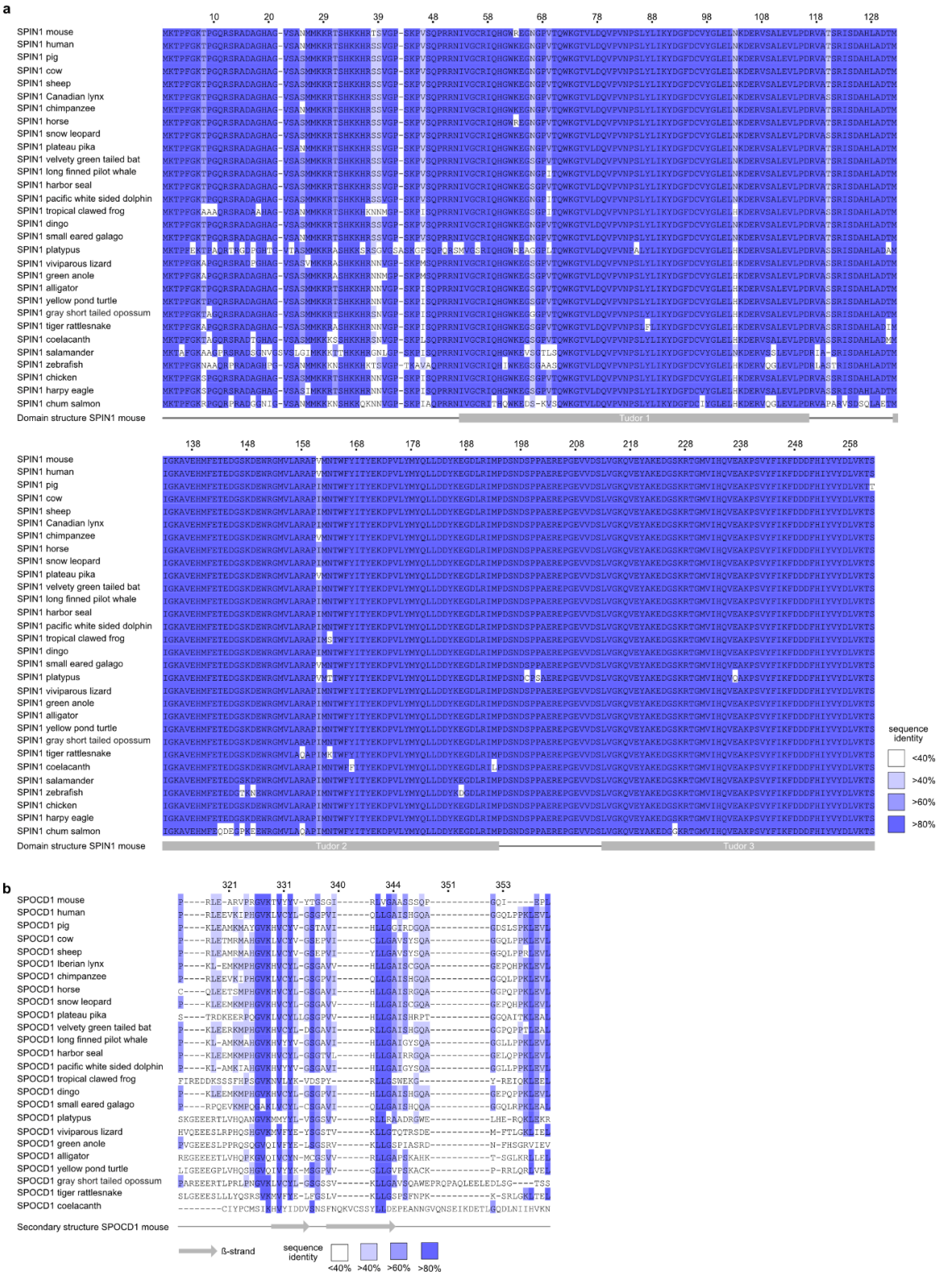

Figure 48: Multiple sequence alignment of SPIN1 and the SPOCD1 ß-hairpin region.

a, Multiple sequences alignment of SPIN1 from representative vertebrates. The domain structure of mouse SPIN1 (Q61142) is indicated underneath the alignment in grey. **b,** Multiple sequence alignment of the SPOCD1 ß-hairpin region with representative vertebrate SPOCD1 sequences.

Secondary structure elements from the AlphaFold2 model of mouse SPOCD1 (B1ASB6) are shown below with grey arrows representing a ß-strand. **a**-**b**, sequences are coloured according to sequence identity. Numbering above according to mouse sequence.

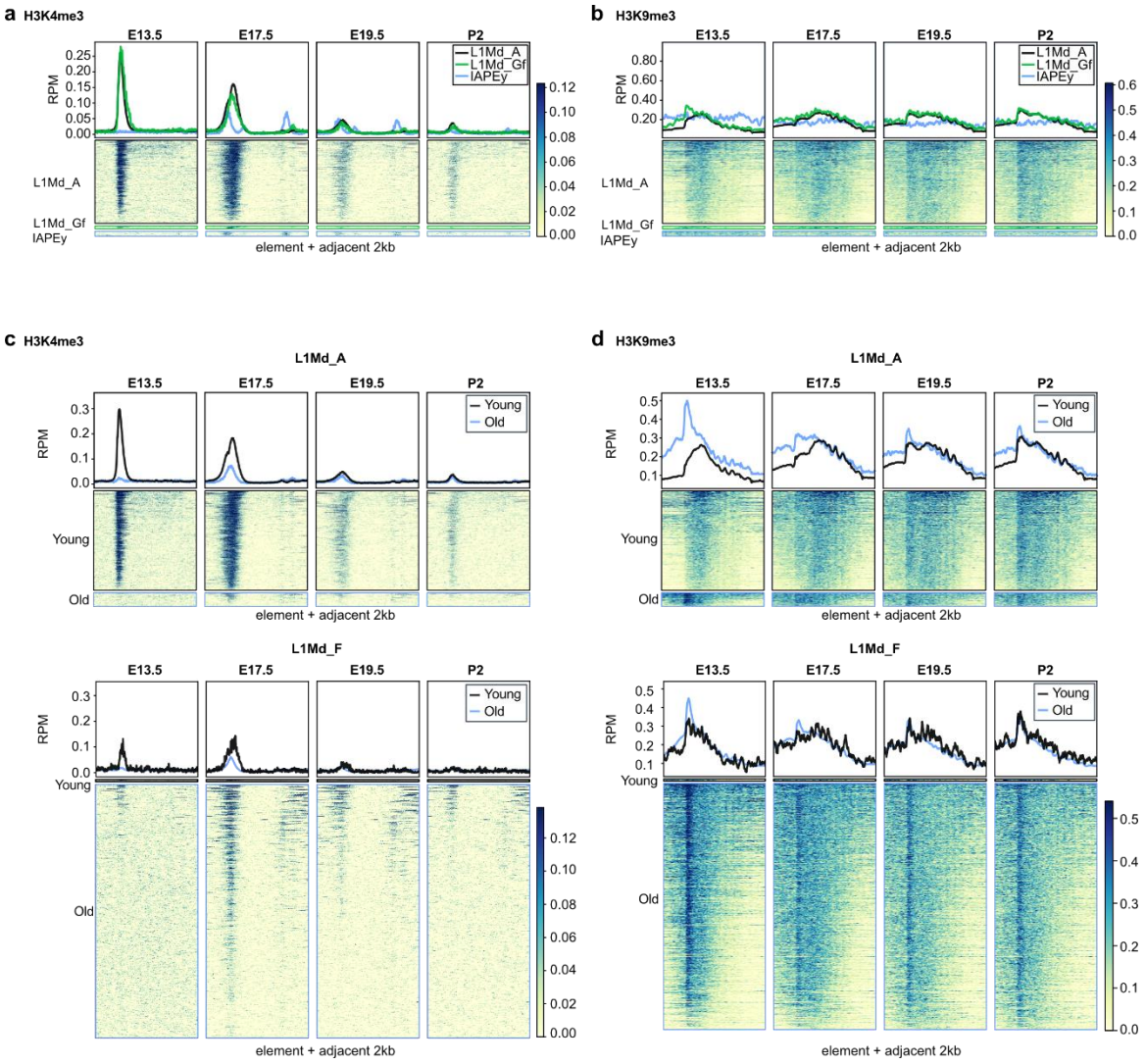

Figure 49: H3K4me3 and H3K9me3 mark young LINE1 elements prior to *de novo* genome methylation.

Metaplot and heat maps of H3K4me3 (**a, c**) and H3K9me3 (**b, d**) ChIP from foetal gonocytes at the indicated timepoints during mouse development. **a**-**b**, Panel shows data for different transposon families as indicated. **c**-**d**, Top panels show young and old elements within the L1Md_A family, and the bottom panels show young and old elements within the L1Md_F family. **a**-**d**, Data depicts element plus adjacent 2 kb for each of the transposon families indicated. Data reanalysed from (Yamanaka et al. 2019)

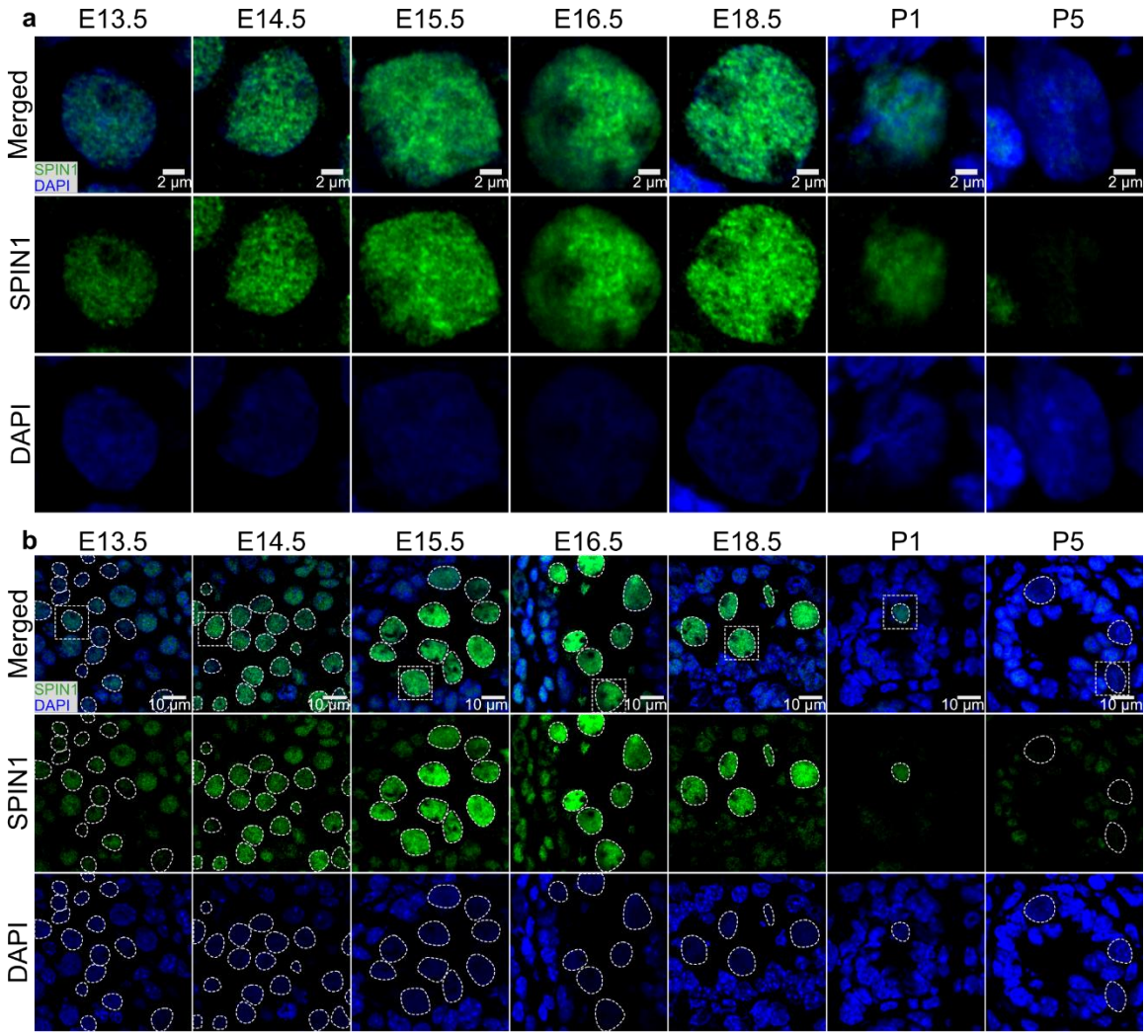

Figure 50: SPIN1 expression and localization in the developing mouse germline.

a, **b**, Representative images of sections from *n* = 3 wild-type foetal testis stained for SPIN1 (green) and DAPI (blue) from indicated timepoints. Cell shown in (**a**) is highlighted with a white box in (**b**). Scale bars are 2 µm (**a**) and 10 µm (**b**).

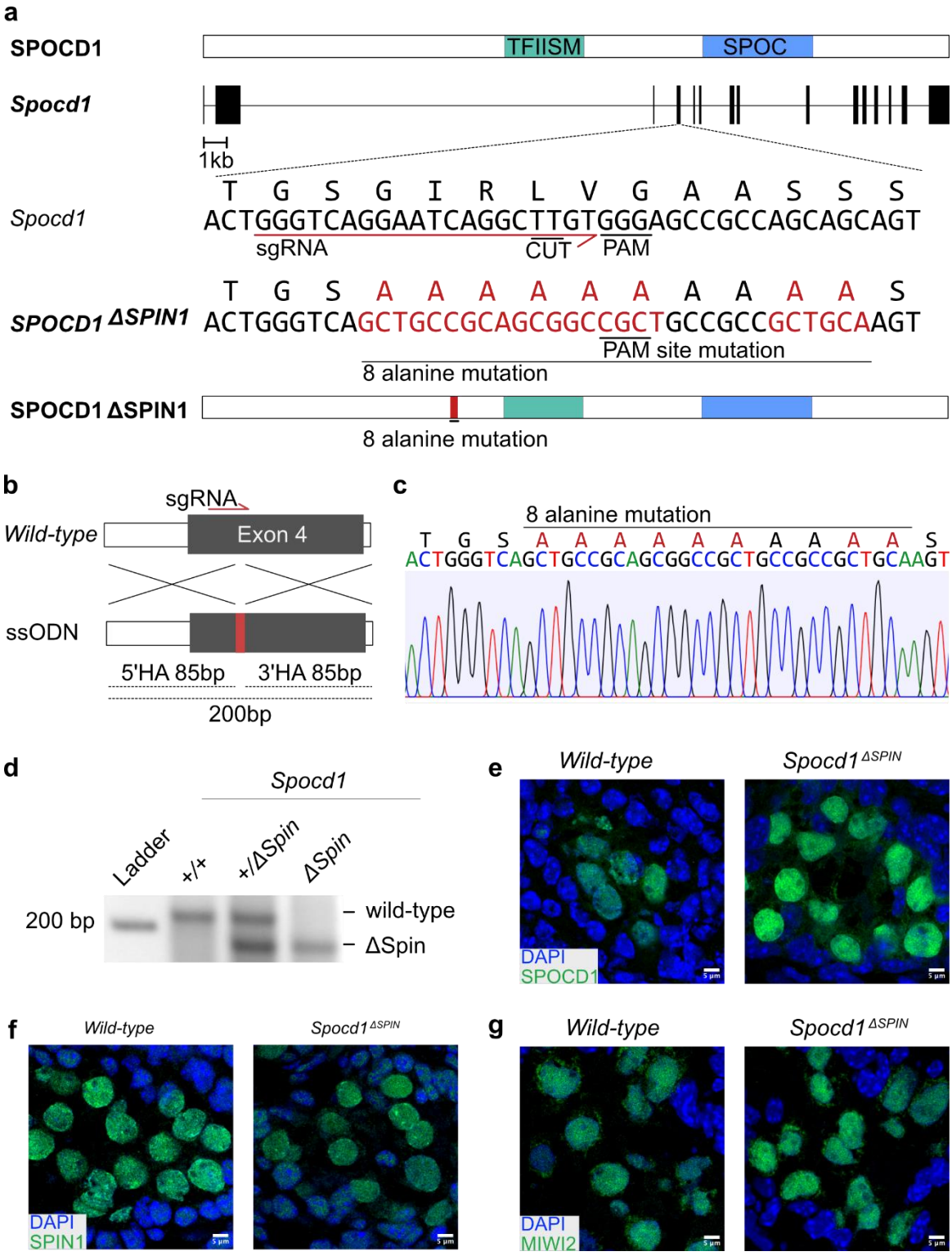

Figure 51: Generation of the *Spocd1*$^{\Delta SPIN1}$ mouse allele.

a, Schematic representations of the mouse *Spocd1* locus and encoded 1015 amino acid protein are shown. sgRNA used for generation of the *Spocd1*$^{\Delta SPIN1}$ allele and adjacent PAM site are indicated. **b,** Schematic of CRISPR targeting strategy showing the location of single-stranded oligo DNA donor (ssODN) and homology arms (HA) used. **c**, Schematic representation, and sequencing trace of the part of *Spocd1*$^{\Delta SPIN1}$ exon 4 harbouring the mutation sites, a 30 bp sequence creating the 8 alanine mutation is highlighted in red. Sequencing was performed on *n* = 3 animals. **d,**

Representative image of genotyping result for *Spocd1*$^{+/+}$, *Spocd1*$^{+/\Delta SPIN1}$ and *Spocd1*$^{\Delta SPIN1}$ mice. **e-g**, Representative images of E16.5 gonocytes from *n* = 3 *Spocd1*$^{\Delta SPIN1}$ and wild-type control mice stained for SPOCD1 (**e**), SPIN1 (**f**) or MIWI2 (**g**,) in green. DNA was stained with DAPI (blue). Scale bars are 5 µm.

4.4 Conclusion

In the present study, we elucidated the role of SPIN1 in piRNA-directed *de novo* DNA methylation. First, we showed that SPOCD1 chromatin association was independent of RNA and MIWI2. We demonstrated that SPOCD1 directly engages with the chromatin reader SPIN1. We mapped this interaction and showed that an antiparallel β-hairpin of SPOCD1 (residue 327-346) interacts with Tudor domain 3 of SPIN1. Furthermore, we recapitulated this interaction with SPOCD1 from frogs, showing that it is an ancient interaction. Next, we discovered that specifically young LINE1 elements are marked by SPIN1-associated chromatin marks, H3K4me3 and H3K9me3. This distinct methylation peak was observed at E13.5, immediately before the onset of *de novo* DNA methylation and disappeared after birth. In fact, we demonstrated that SPIN1-expression is upregulated in germ cells at E13.5, before other key factors, such as MIWI2 or SPOCD1, are expressed. In addition, we showed that SPOCD1 and SPIN1 interact in E14.5 foetal testes before MIWI2 is present in these cells. We generated a separation of function allele in mice that uncouples SPOCD1 and SPIN1 and observed the typical piRNA phenotype of male mice which is infertility and reduced testis size due to defects in spermatogenesis. We detected upregulation of young LINE1 elements, but not IAPs, in these animals. Methylation sequencing confirmed this observation and revealed that young LINE1s, but not IAPs, were hypomethylated. In summary, SPIN1 was defined as a new interactor of SPOCD1 that plays an essential role in the piRNA-directed *de novo* methylation of young LINE1 elements.

5 Discussion and Conclusion

Since the discovery of piRNAs, PIWI proteins, and the piRNA pathway, much knowledge has been gained regarding the pathway in flies (Ozata et al. 2018). However, the foetal piRNA pathway in mammals remains poorly understood mainly because of the difficulty in studying this pathway and the absence of an orthologous tissue culture system. Studying MIWI2 is particularly difficult because it only becomes nuclear once bound to a germline-specific small RNA, the piRNA. (Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008). Therefore, the use of foetal testes from mice, which is a very limited material, is inevitable. The O'Carroll laboratory recently overcame some of these problems by establishing an IP-MS protocol with only 25-50 foetal testes targeting HA-tagged proteins (Zoch et al. 2020; Schopp et al. 2020). This technical advancement has allowed us to obtain a high-quality MIWI2 interactome (Zoch et al. 2020; Schopp et al. 2020). Among the MIWI2 interactors, two new essential executors of piRNA-directed transposon silencing and methylation were discovered: TEX15 and SPOCD1 (Schopp et al. 2020; Zoch et al. 2020). It was shown that SPOCD1 is a nuclear protein that is essential for piRNA-directed *de novo* DNA methylation in the germline (Zoch et al. 2020). Furthermore, the SPOCD1 interactome revealed association to the DNA methyltransferases DNMT3A and DNMT3L, as well as the chromatin remodeller complexes BAF and NURD (Zoch et al. 2020). However, the exact function of SPOCD1 remains unclear. This study aimed to understand the role of SPOCD1 in transposon methylation and gain deeper insight into the mechanism of piRNA-directed *de novo* methylation in mammals.

First, I studied the interactome of SPOCD1. In foetal testis, three proteins, TPR, C19ORF84, and SPIN1, were associated with MIWI2 and SPOCD1 (Chapter 3). Even in HEK293T cells overexpressed SPOCD1 associates with TPR, C19ORF84, and SPIN1. I could show that TPR interacts with the TFIIS-M domain of SPOCD1.

Furthermore, I mapped the interaction between SPOCD1 and C19ORF84 in more detail and defined a small stretch in the c-terminus of SPOCD1 as the interacting part. This stretch is conserved and formed an α-helix in the SPOCD1 AlphaFold2 prediction. I also mapped this interaction from the C19ORF84 side and could identify a short, conserved stretch that folds into an α-helical structure.
Having discovered SPOCD1 and three of its interactors, TPR, C19ORF84 and SPIN1, one can speculate about the nature of the complex. We know that SPOCD1 interacts directly with C19ORF84 and SPIN1, but we do not know if it binds to TPR or MIWI2 directly. This could be tested by pull-down assays with recombinant protein. Based on some earlier IP-MS HEK experiments, I know that there is no direct interaction between TPR, C19ORF84 and SPIN1 (data not shown). SPOCD1 appears to be the central connection and docking platform for those three proteins. I hypothesis that the complex builds up gradually. SPIN1 is expressed first and will associate with chromatin. Once SPOCD1 is expressed, it will bind to SPIN1, with TPR and C19ORF84 being added stepwise to the complex. At the moment we do not know if SPOCD1 and MIWI2 are interacting directly. However, I hypothesis that this is not the case. I postulate that the piRNA-MIWI2 complex and the complex around SPOCD1 containing TPR, C19ORF84 and SPIN1 form separately. The DNA methylation machinery might function as a bridge between those two complexes, making contact with MIWI2 and SPOCD1, which results in the formation of one big complex. This hypothesis could be tested with interaction studies using recombinant proteins or with density gradient centrifugation to see which complexes are present in foetal testis.

5.1 The SPOCD1-TPR interaction

While the TFIIS-M domain of TFIIS is known to be a Pol II interacting module and promoting transcription, I showed here that the SPOCD1 TFIIS-M domain is interacting with the nuclear pore component TPR (Cheung and Cramer 2011). It has been shown that components of the NPC are involved in transcription and chromatin organization (Capelson et al. 2010; Gozalo et al. 2020). More specifically, TPR controls DNA accessibility around the nuclear pore by the formation of heterochromatin exclusion zones (Boumendil et al. 2019; Krull et al. 2010). Interestingly, transposons are associated with H3K9me3 enriched

heterochromatin before *de novo* methylation (E13.5) and in early adulthood (P10) (Slotkin and Martienssen 2007; Liu, Brind'Amour, et al. 2014; Pezic et al. 2014). In addition, a recent study has highlighted the role of MORC1-dependent heterochromatin formation in transposon silencing in gonocytes, indicating the importance of heterochromatin in transposon control (Uneme et al. 2023; Pastor et al. 2014). Therefore, one hypothesis is that the formation of heterochromatin exclusion zones is important for piRNA-directed transposon silencing in mammals. In fact, in *Drosophila*, components of the NPC play an essential role in piRNA-guided heterochromatin formation (Batki et al. 2019). Furthermore, *Drosophila* NPC components are involved in the export of piRNA precursors and there seems to even be a specific pathway for heterochromatin-associated piRNAs (ElMaghraby et al. 2019; Munafò et al. 2021). Additionally, TPR is involved in mRNA export (Bangs et al. 1998; Lee et al. 2020; Aksenova et al. 2020). Therefore, it is possible that TPR in the piRNA pathway exports piRNA precursors.

There are indications that TPR associated RNA export or heterochromatin regulation may play a role in the piRNA pathway. However, its exact role requires further investigation. First, the role of TPR in transposon silencing must be studied. This can be achieved by generating a *Tpr* deletion in mice. However, deletion of TPR results in preweaning lethality (Blake et al. 2021). An elegant way to circumvent this limitation would be to generate a separation-of-function allele. Once the interaction is better understood and the interacting parts of SPOCD1 and TPR are defined, small mutations that would uncouple these two proteins could be tested in HEK293T cells and once a promising candidate is found, generated in mice. This mouse model would show whether the typical piRNA phenotype of male infertility and upregulated transposons is observed. Furthermore, the mouse model would allow to study chromatin formation by examining the heterochromatin stage *via* Assay for transposase-accessible chromatin using sequencing (ATAC-seq). Small RNA/ piRNA analysis could reveal whether TPR is important in the export of piRNA precursors in germ cells, as has been done in previous *Drosophila* studies (ElMaghraby et al. 2019).

In summary, the discovery of the SPOCD1-TPR interaction opens a new field of research. It links the mammalian piRNA pathway to the nuclear pore complex and will help in understanding the mechanisms of mammalian piRNA-directed *de novo* DNA methylation during development.

5.2 The SPOCD1-C19ORF84 interaction

In my screen, I defined C19ORF84 as an interactor of SPOCD1. The interaction appears to occur between two α-helices, one from SPOCD1 and one from C19ORF84. C19ORF84 is a very unstructured protein, and the only secondary structural element present is this α-helix (Chapter 3.5). In SPOCD1, the interacting α-helix is surrounded by long stretches of unstructured sequences. The two α-helices could somehow interact with each other and facilitate this interaction; but the exact nature of this interaction remains unclear. However, attempts to map this interaction to a smaller minimal interaction area have not been successful.

Unfortunately, almost no information is available on C19ORF84 in the literature. Yet, it is known that Intrinsically disordered proteins (IDP) play an important role in many biological processes (Dyson and Wright 2005; Yoon et al. 2012). Therefore, it is possible that C19ORF84 plays a central role in the piRNA pathway, even if it is unstructured in most areas. From the CAGE dataset, we know that the mRNA expression pattern is almost identical to that of SPOCD1 in the germline (Consortium et al. 2014). Furthermore, a recent study investigating SPOCD1 in SSCs showed that SPOCD1 and C19ORF84 had the same expression pattern in Single-cell RNA sequencing (scRNA-seq) (Zhou et al. 2022). A former PhD student in the O'Carroll lab, Gabriela Konieczny, investigated the importance of this interaction in mice (Konieczny 2022). She showed that the loss of C19ORF84 in mice results in male infertility with early meiotic arrest (Konieczny 2022). She detected upregulation of active young LINE1 and IAP copies due to impaired methylation of these elements which is similar to the *Miwi2* or *Spocd1* knockout mice (Konieczny 2022; Carmell et al. 2007; Zoch et al. 2020). It is possible that C19ORF84 is an adapter protein that allows the interaction of SPOCD1 with MIWI2 or the DNA methylation machinery. This hypothesis could be tested by performing IP-MS of MIWI2 or components of the DNA methylation machinery in the absence of C19ORF84 from foetal testis. However, this approach is time- and cost-intensive. An alternative method is the use of recombinant proteins. Although, the purification of the highly unstructured proteins C19ORF84 and SPOCD1, as well as the purification of MIWI2, is likely to be a major challenge. If the purification of at least the interacting parts of C19ORF84 and SPOCD1 succeeded, it would allow

further study of this interaction by determining whether it is a direct interaction and gaining some structural information by crosslink-MS.

In summary, the importance of C19ORF84 in piRNA-directed *de novo* DNA methylation is unquestionable, as it shows the same phenotype as *Spocd1* or *Miwi2* knockout mice. It remains to be discovered what the exact role of C19ORF84 in this process is and how the two α-helices facilitate the interaction between SPOCD1 and C19ORF84.

5.3 The SPOCD1 SPOC domain

The SPOC domain is a well-characterised domain that has been found in several different proteins, including SPOCD1, PHF3, DIDO, SPEN/SHARP, and RBM15 (Appel, Benedum, et al. 2023). SPOC domains have been defined as phosphor-dependent protein-protein interaction domains (Appel, Benedum, et al. 2023; Appel, Franke, et al. 2023). In fact, most known SPOC domain interactions are phosphorylation-dependent (Appel, Benedum, et al. 2023). This includes the interaction of SPEN with SMRT, NCoR, Pol II, and FMR1, of RBM15 with WTAP and Pol II, of DIDO with Pol II, and of PHF3 with Pol II (Appel, Franke, et al. 2023; Appel et al. 2021; Oswald et al. 2016; Mikami et al. 2014; Shi et al. 2001). The interaction of SPEN with Lysine Methyltransferase 2D (KMT2D) or C-terminal binding protein 1 interacting protein (CtIP), as well as RBM15 with SETD1B, has not been tested for phosphor-dependency (Oswald et al. 2016; Lee and Skalnik 2012; Oswald et al. 2005).

While I could not define an interactor of the SPOCD1 SPOC domain in my HEK293T cells experiments, I solved the structure of the SPOCD1 SPOC domain in Chapter 3.6. Overall, the structure is very similar to known SPOC domain structures, but the surface charges are different. While the PHF3, RBM15, and SHARP SPOC domain structures have distinct basic patches on their surface, they are not present in the SPOC domain of SPOCD1. Some of the well-characterised interactors of the SPOC domains, such as the CTD of Pol II or the SMRT peptide, would not be able to bind to SPOCD1. Furthermore, I studied a point mutation in the SPOCD1 SPOC domain found in a sterile man. The residue that is mutated from leucine to arginine faces inward and likely causes some steric issues within the domain. The mutation was in an uncharged but highly conserved area. I

showed that full-length SPOCD1 with the point mutation is expressed, but recombinant expression of the SPOC domain alone failed, highlighted that the mutation affects stability of the domain.

The SPOCD1 SPOC domain surface charge is different from that of most SPOC domains. My analysis is in line with the conclusions based on the SPOCD1 AlphaFold2 model in a recent study (Appel, Franke, et al. 2023). The SPOC domain does not have the distinct negatively charged patches that are present on other SPOC domains (Appel, Franke, et al. 2023). In fact, the SPOCD1 SPOC domain does not bind to the unphosphorylated, neither to the phosphorylated Pol II CTD peptides (Appel, Franke, et al. 2023). It might have completely lost its phosphorylation dependency, or it may be associated with different targets in a phosphorylation-dependent manner. The SPOC domain must have changed during or after the gene duplication event from PHF3 to adapt to a new role and function.

Based on this structure, the role of the SPOCD1 SPOC domain in *de novo* DNA methylation is unknown. However, the human point mutation showed that it is important. Furthermore, the high conservation of the area around the mutated residue indicated that this specific region has an important function. In the infertile man, full-length SPOCD1 was most likely expressed but did have a non-functional SPOC domain. Because attempts to use an orthologous system (HEK293T cells) to determine the SPOC domain interactor have been unsuccessful, a mouse model must be generated to understand the role of the SPOCD1 SPOC domain. If we wanted to generate a SPOCD1 allele with a SPOC domain deletion, we would have to use a minigene strategy which is technically, but also in the result interpretation quite challenging. A more elegant strategy would be to use the pathological L792R mutation to generate the *Spocd1*^{L792R} allele in mice. Performing IP-MS on foetal testes from *Spocd1*$^{+/+}$ and *Spocd1*$^{L792R/L792R}$ mice would allow us to determine which interaction is lost in the SPOC domain point mutation.

In summary, we know that the SPOCD SPOC domain is important for male fertility, but its exact role remains unknown. These results highlight the differences between this SPOC domain and other known SPOC domains. It is likely that the SPOC domain of SPOCD1 changed and adapted after the gene duplication event to fulfil its unique new function in the *de novo* DNA methylation of transposons which makes it an interesting domain to study further. Overall, it is likely that the domain

is a protein-protein interaction module that plays a role in transcription, RNA processing, or chromatin regulation, like other known SPOC domains (Appel, Benedum, et al. 2023).

5.4 The SPOCD1-SPIN1 interaction

In Chapter 4, we studied the SPOCD1-SPIN1 interaction in detail. We demonstrated that SPOCD1 association with chromatin is independent of MIWI2 and RNA. In fact, we showed that SPOCD1 interacts directly with the chromatin reader SPIN1 *via* a small stretch in SPOCD1 that folds into a β-hairpin according to the SPOCD1 AlphaFold2 model. Using crosslink-MS, we showed that this β-hairpin interacts with Tudor-like domain three of SPIN1, an interaction also predicted by AlphaFold2. We observed that the SPOCD1 β-hairpin and the SPOCD1-SPIN1 interaction is conserved in green anole lizard (*Anolis carolinensis*) and tropical clawed frog (*Xenopus tropicalis*), but not in coelacanth (*Latimeria chalumnae*). We found that specifically young LINE1 elements, but not IAPs, were modified with H3K4me3 chromatin methylation prior to *de novo* DNA methylation. Simultaneously, H3K9me3 was present in these elements. This observation was even more pronounced if young and old elements within the young LINE1 family, L1Md_T, were separated. We saw that SPIN1 is expressed as early as E13.5 before *de novo* DNA methylation and SPOCD1 or MIWI2 expression. In fact, SPOCD1 associated with SPIN1 at E14.5, before MIWI2 localised to the nucleus. Finally, we generated a *Spocd1* separation-of-function allele that uncouples SPOCD1 and SPIN1, termed *Spocd1*^{dSPIN1}. We observed that male mice are infertile, germ cells do have DNA damage, and apoptotic cells which results in a complex meiotic arrest phenotype in adult testes. This phenotype was caused by derepressed young LINE1 but not IAP elements. Methylation sequencing revealed that only young LINE1s were hypomethylated in these animals, whereas IAPs and old transposons showed normal methylation levels.

In this study, we identified SPIN1 as a new key component of the piRNA-directed DNA methylation machinery. SPIN1 is a well-characterised chromatin reader that associates with different chromatin marks *via* its Tudor-like domain one and two (Zhao et al. 2020; Wang et al. 2018; Su et al. 2014). It was shown that it exhibits

the highest affinity for the bivalent H3K4me3-H3K9me3 histone modification (Du et al. 2021; Zhao et al. 2020). Tudor-like domain three is a protein-protein interaction domain, and two interactors, SPINDOC (C11ORF84) and HBx, from hepatitis B virus are known (Du et al. 2021; Devi et al. 2019; Liu et al. 2023). Interestingly, in both cases, SPIN1 associates with the interactor and H3K4me3-H3K9me3 methylated histone tail simultaneously (Liu et al. 2023; Du et al. 2021). In both instances, SPIN1 associates with a β-hairpin structure that complements the Tudor-like three domain fold (Liu et al. 2023; Du et al. 2021). According to the AlphaFold2 model, which is supported by the crosslink-MS and mapping data, the interaction of the SPOCD1 β-hairpin with Tudor-like domain three from SPIN1 appears to be the same. This indicates that the SPOCD1-SPIN1 complex would have high affinity for the H3K4me3-H3K9me3 histone tail as well. However, only recombinant histone or histone peptide binding assays can answer this question.

Next, we attempted to understand the conservation of the SPOCD1-SPIN1 interaction. Interestingly, SPOCD1 is not only present in mammals but appears to have first developed in vertebrates. While we did find a SPOCD1 sequence in coelacanth, this SPOCD1 orthologue did not interact with SPIN1. Therefore, SPOCD1 must have acquired SPIN1 binding ability after the coelacanth lineage and before or with the emergence of frogs. Unfortunately, we could not obtain a full-length sequence of SPOCD1 from salamander, as this would allow us to understand exactly where SPOCD1 acquired its SPIN1 binding capacity. Hopefully, advancements in next generation sequencing technologies and analyses will allow for this in the future. Furthermore, it would be interesting to determine if additional piRNA factors are present in vertebrates, indicating where the piRNA pathway first arose. Germline-specific piRNAs have been identified in frog oocytes (Armisen et al. 2009). Interestingly, birds seem to have lost SPOCD1, while they do have piRNAs (Rengaraj et al. 2014). This could be because germ cell development differs between mammals and other vertebrates (Strome and Updike 2015). While the mammalian germ line is induced, germ cell specification in vertebrates is preformed in early embryonal development (Strome and Updike 2015). Further investigation is necessary to understand the role of SPOCD1, and specifically the SPOCD1-SPIN1 interaction, in less studied species.

SPIN1 binds strongly to H3K4me3 which can be augmented by H3K9me3 (Zhao et al. 2020). Interestingly, young LINE1 elements but not IAPs were decorated with

this mark. A very distinct H3K4me3 peak was observed at E13.5, before the onset of *de novo* DNA methylation, indicating that SPIN1 could be one of the first factors to be recruited. This peak gets smaller at E17.5, during DNA methylation, until the peak becomes barely visible from E19.5, when DNA methylation is finished. This could indicate that SPIN1 chromatin association is particularly important in the initial steps of *de novo* DNA methylation. Interestingly, repressing H3K9me3 histone methylation was present in all transposons over the entire length of the element. This is in line with previous observations that H3K9me3 is enriched in piRNA-dependent regions (Nagamori et al. 2018). At the same time, H3K4me3 is a chromatin methylation that is associated with active transcription (Santos-Rosa et al. 2002). The recognition of both H3K4me3 and H3K9me3 may allow SPIN1 to specifically target actively transcribed transposons. Furthermore, it is known that *de novo* DNA methylation of the male germline is inefficient when histones are methylated at Histone 4 lysine 4 (H3K4) (Singh et al. 2013; Ooi et al. 2007). Therefore, the SPIN1-SPOCD1 association with H3K4me3-H3K9me3 might be necessary to overcome this obstacle by direct recruitment of the *de novo* DNA methylation machinery through SPOCD1.

SPIN1 and SPOCD1 co-precipitate together as early as E14.5, shortly after the distinct H3K4me3-H3K9me3 methylation pattern is observed and before the expression of MIWI2 (Zoch et al. 2020). It is highly probable that at this time the SPOCD1-SPIN1 complex associates with the promoter regions of young LINE1 elements, exhibiting this methylation pattern. This timing of the expression of key piRNA factors and the change in chromatin signatures seem to be meticulously regulated and are likely to hold functional relevance.

Interestingly, the phenotype that we observed in *Spocd1*^{dSpin1} animals is similar to mice with a slicer inactive Mili which shows upregulation of LINE1, but not IAP elements (De Fazio et al. 2011). In these animals, secondary piRNA biogenesis is affected which results in de-repression of specifically LINE1 elements (De Fazio et al. 2011). Interestingly, the same phenotype was observed in mice without TDRD9 and in mice with ATPase-mutated TDRD9 (Shoji et al. 2009; Wenda et al. 2017). In both cases, the deletion or mutation of TDRD9 resulted in the upregulation of LINE1 elements due to hypomethylation (Shoji et al. 2009; Wenda et al. 2017). However, TDRD9 appears to be a nuclear effector and the ATPase domain is essential for its function, although its exact role remains unknown (Shoji et al. 2009;

Wenda et al. 2017). We know that SPOCD1 is not involved in piRNA biogenesis (Zoch et al. 2020). Therefore, the observed phenotype in *Spocd1*^{dSpin1} animals must be caused by a different mechanism. Since SPIN1 and TDRD9 are only essential for LINE1 silencing, a second piRNA-dependent pathway that controls IAP elements must be present. Interestingly, most nuclear piRNA factors are necessary for LINE1 and IAP silencing, such as MIWI2, SPOCD1, TEX15, DNMT3C, MORC1 and C19ORF84 and only TDRD9 and SPIN1 are LINE1-specific (Pastor et al. 2014; Barau et al. 2016; Schopp et al. 2020; Zoch et al. 2020; Carmell et al. 2007; Konieczny 2022; Shoji et al. 2009). This highlights the complexity of the entire mechanism and emphasises how much remains to be discovered. In addition, it remains to be understood how SPOCD1 is recruited to active IAP elements, if not by SPIN1.

In other settings, SPIN1 is known as a transcriptional activator, for example, by promoting the activation of rRNA genes or the Wnt signalling pathway (Du et al. 2021; Wang et al. 2011; Su et al. 2014). In addition to recruiting SPOCD1, SPIN1 may be important for promoting the transcription of LINE1 sequences. This might be necessary to quickly boost the number of transposon copies in cells to promote the production of a wide range of secondary piRNAs which are necessary for MIWI2 recruitment and LINE1 silencing. Furthermore, SPIN1 plays a role in facilitating epigenetic changes (Liu et al. 2023). SPIN1 interacts with the hepatitis B virus protein HBx and promotes the expression of the chromatinized viral genome (Liu et al. 2023). In this process, SPIN1 supports the switch of the viral chromatin from a repressive H3K9me3-enriched state to an active, H3K4me3-enriched state (Liu et al. 2023). Already, an earlier study hypothesised that SPIN1 acts as a downstream effector of H3K4me3, which is found in H3K9me3-enriched heterochromatin regions (Zhao et al. 2020). In the piRNA pathway, SPIN1 could play a role in epigenetic changes of young LINE1 sequences. This could fit the purpose of making LINE1 elements more accessible to the methylation machinery or simply boost piRNA production. It is known that the piRNA pathway is necessary for some chromatin changes, such as the establishment of H3K9me3 in germ cells (Pezic et al. 2014). However, the mechanism by which the SPIN1-SPOCD1 complex affects transcription or chromatin dynamics is yet to be elucidated.

Based on our findings, we propose a novel two-factor authentication mechanism to form the basis of precision. We believe that the first key factor is SPIN1

recruitment to transposon loci *via* H3K4me3-H3K9me3 as early as E13.5. As soon as SPOCD1 is expressed, at approximately E14.5, SPIN1 guides it to these loci. When piRNA-bound MIWI2 is located at the locus around E16.5, it independently finds nascent transposon transcripts by complementary RNA base pairing. Only the presence of both SPOCD1-SPIN1 and the piRNA-MIWI2 complex will result in the recruitment of the DNA methylation machinery and *de novo* methylation of young LINE1 elements (Figure 52).

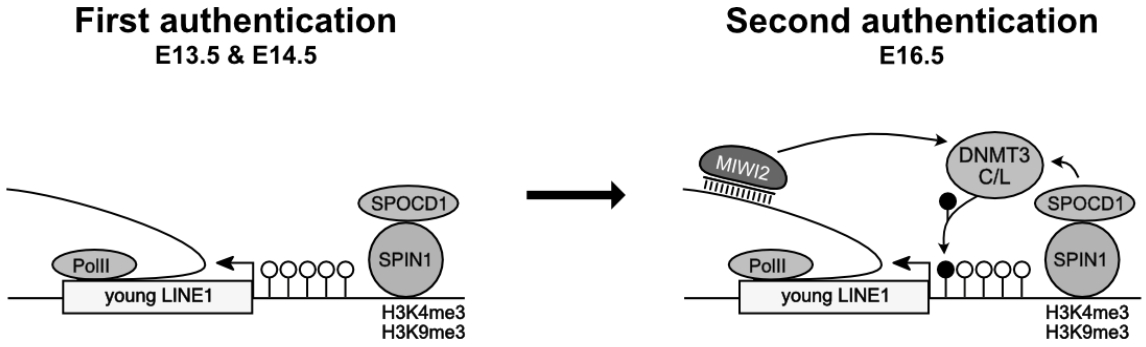

Figure 52: Two-factor authentication model.
A model of piRNA-directed *de novo* DNA methylation including SPIN1. All steps and proteins are labelled accordingly. Unmethylated DNA is indicated by white circles, methylated DNA is indicated by a black circle.

This mechanism could have developed to ensure precision of *de novo* transposon methylation. One aspect of precision is avoiding off-targeting. The two-factor authentication model decreases the risk of off-targeting due to its necessity of two independent complexes to trigger methylation. However, one can speculate if off-target methylation or failure to methylate a transposon would be more harmful to the germline. There is a high chance that one single unmethylated, active transposon could cause long-lasting damage to germ cells, and thus threaten fertility. Alternatively, the SPOCD1-SPIN1 complex could be necessary to target sequences that are not reliably detected by the piRNA-MIWI2 complex alone. A piRNA-MIWI2 complex with lower degrees of base complementarity to the transposon is likely to have different on/off rates. The SPOCD1-SPIN1 complex would still detect those targets in a chromatin-dependent matter and might be able to lock MIWI2 at the target to induce methylation. This would ensure a more

sequence-independent mechanism that is able to tolerate transposon mutations and therefore reliably silence transposons.

Another key aspect of the two-factor authentication model could be to ensure completion of DNA methylation within the short time window before birth. The recruitment of SPIN1-SPOCD1 via chromatin allows the assembly of the complex before the start of *de novo* DNA methylation. As soon as the piRNA-MIWI2 complex moves to the nucleus and finds its targets, it can trigger methylation and ensure immediate transposon silencing. Weather the two-factor authentication model is necessary to avoid off-targeting, to ensure silencing of targets with a low complementarity to the piRNA or to allow a completion of the process in a short time window, remains to be discovered.

5.5 Impact and Outlook

Overall, the work presented in this thesis provides a big advancement in the understanding of mammalian piRNA-directed *de novo* DNA methylation. We identified and mapped three interactors of SPOCD1: TPR, C19ORF84, and SPIN1. This finding highlights the importance and key role of SPOCD1 in this pathway. Owing to the number of interactions, SPOCD1 seems to be a central platform for the recruitment of factors that are essential for *de novo* methylation. Furthermore, the SPOCD1-TPR interaction indicates a connection of the nuclear pore complex with the piRNA pathway in mammals. Given that we could not attribute a function to the SPOC domain but showed that it is essential for human fertility, there are more potential interactors or functions of SPOCD1 to be discovered. In fact, SPOCD1 contains more conserved and ordered regions outside its domains which could indicate additional protein interaction areas (Figure 53).

The discovery of the SPOCD1-SPIN1 interaction and its importance is a big step towards better understanding the mechanisms of piRNA-directed *de novo* DNA methylation in mammals. This study provides the first insight into the involvement of a chromatin reader in recruiting piRNA factors during germline development which could be a general mechanism adapted by piRNA effector proteins. Furthermore, it highlights the importance and interplay between histone methylation and the piRNA pathway. If the importance of SPIN1 solely depends on SPOCD1 recruitment or also on its ability to promote transcription and induce chromatin changes, remains to be discovered. In addition, the proposed two-factor authentication model shows how precision could be ensured for the first time. Given that the SPOCD1-SPIN1 interaction is only important for LINE1 silencing, the pathway that is silencing IAPs is yet to be discovered. Our findings highlight the complex mechanism of piRNA-directed *de novo* DNA methylation and provide a starting point for a better understanding of the system.

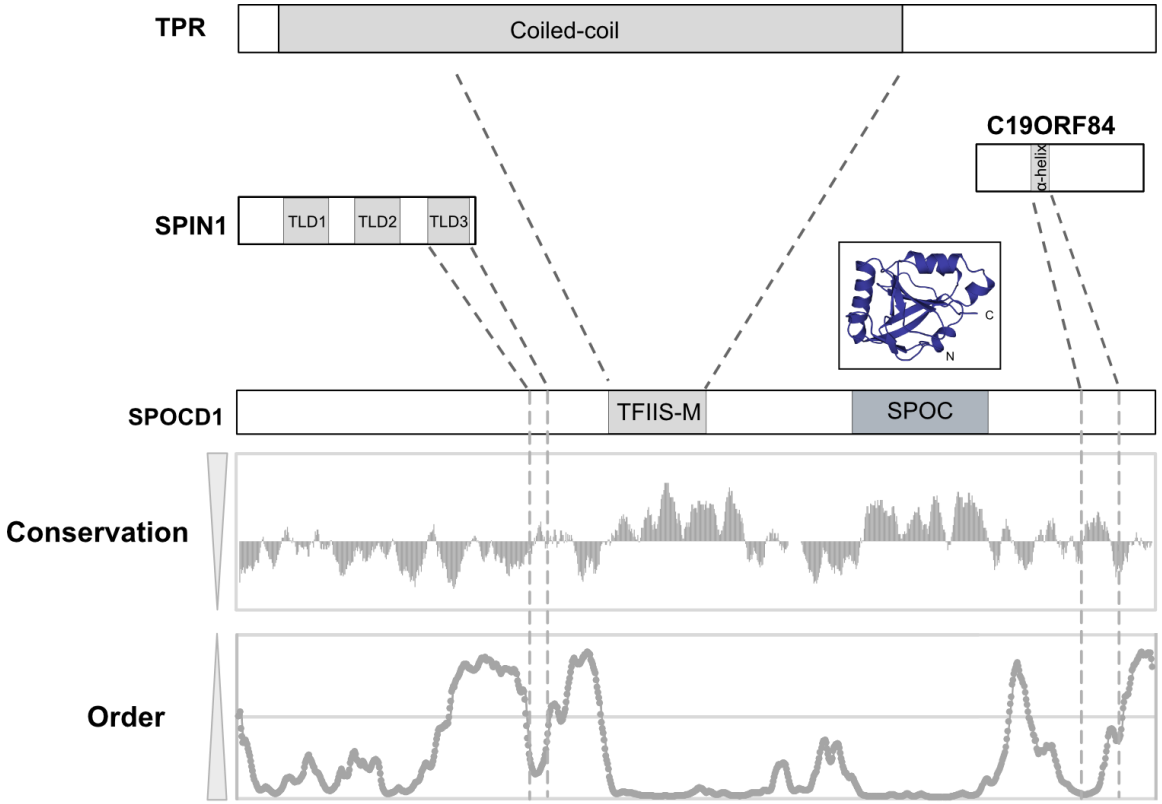

Figure 53: SPOCD1 interacts with TPR, SPIN1, and C19ORF84.

The figure shows which part of SPOCD1 interacts with TPR, SPIN1, or C19ORF84. Below SPOCD1, the conservation and order of the respective SPOCD1 regions are shown. In addition, the structure of the SPOC domain is presented.

6 Appendix

6.1 Appended tables

Table 9: Proteins identified in SPOCD1-HA IP-MS from HEK293T cells.
Table lists all statistically significant identified proteins with at least 4-fold enrichment in the HA-SPOCD1 immunoprecipitation from HEK293T cells. Enrichment and p-value are shown as indicated. P<0.05, two-sided Student's t-test.

Protein.ID	Protein	Gene	Enrichment	p-value
B1ASB6	SPOCD1	Spocd1	7356.9	4.25E-07
Q9Y657	SPIN1	SPIN1	147.3	2.67E-06
Q9NX47	MARCH5	MARCH5	19.6	2.34E-04
Q96EY1	DNAJA3	DNAJA3	17.8	2.83E-02
Q96BR5	COA7	COA7	10.2	6.89E-03
P36873	PPP1CC	PPP1CC	6.3	1.56E-02
Q9BPZ2	SPIN2B	SPIN2B	5.8	2.07E-02
P08574	CYC1	CYC1	4.8	8.15E-03
P12270	TPR	TPR	4.2	1.28E-03
P62136	PPP1CA	PPP1CA	4.1	1.76E-04

Table 10: Proteins identified in SPOCD1-ΔTFIISM-HA IP-MS from HEK293T cells.

Table lists all statistically significant identified proteins with at least 4-fold enrichment in the HA-SPOCD1-ΔTFIISM immunoprecipitation from HEK293T cells. Enrichment and p-value are shown as indicated. P<0.05, two-sided Student's t-test.

Protein.ID	Protein	Gene	Enrichment	p-value
B1ASB6	SPOCD1	Spocd1	25252.2	7.64E-06
Q9Y657	SPIN1	SPIN1	362.9	1.08E-04
Q9NX47	MARCHF5	MARCHF5	55.7	5.21E-05
Q9BPZ2	SPIN2B	SPIN2B	28.5	1.77E-04
Q96EY1	DNAJA3	DNAJA3	25.8	4.32E-03
Q5JUX0	SPIN3	SPIN3	17.9	3.66E-03
P36873	PPP1CC	PPP1CC	17.7	2.01E-02
P62136	PPP1CA	PPP1CA	12.7	1.71E-04
Q56A73	SPIN4	SPIN4	12.6	1.46E-03
P62140	PPP1CB	PPP1CB	9.7	7.49E-04
Q13131	PRKAA1	PRKAA1	8.8	2.43E-02
Q9H936	GC-1	SLC25A22	7.2	3.52E-04
Q99943	AGPAT1	AGPAT1	7.1	9.78E-04
Q13576	IQGAP2	IQGAP2	6.9	2.80E-02
P41236	PPP1R2	PPP1R2	6.5	1.64E-02
Q9BT22	ALG1	ALG1	6.3	1.30E-03
O14735	CDIPT	CDIPT	5.8	2.41E-02
Q15386	UBE3C	UBE3C	5.7	4.00E-03
O15091	PRORP	PRORP	5.1	1.83E-02
Q04727	TLE4	TLE4	4.9	6.18E-03
Q9Y5Y2	NUBP2	NUBP2	4.5	3.41E-02
O60884	DNAJA2	DNAJA2	4.5	1.30E-03
Q969X5	ERGIC1	ERGIC1	4.5	1.74E-02
Q8IXB1	DNAJC10	DNAJC10	4.4	1.00E-04

Table 11: Proteins identified in SPOCD1-ΔSPOC-HA IP-MS from HEK293T cells.

Table lists all statistically significant identified proteins with at least 4-fold enrichment in the HA-SPOCD1-ΔSPOC immunoprecipitation from HEK293T cells. Enrichment and p-value are shown as indicated. P<0.05, two-sided Student's t-test.

Protein.ID	Protein	Gene	Enrichment	p-value
B1ASB6	SPOCD1	Spocd1	3632.6	1.83E-04
Q9Y657	SPIN1	SPIN1	138.3	9.61E-04
Q96BR5	COA7	COA7	14.0	3.02E-03
Q9BPZ2	SPIN2B	SPIN2B	9.5	1.37E-02
P12270	TPR	TPR	5.7	8.55E-03
O14735	CDIPT	CDIPT	5.2	1.07E-02
Q56A73	SPIN4	SPIN4	4.4	1.58E-02

Table 12: Proteins identified in SPOCD1-HA IP-MS from HEK293T cells transfected with C19ORF84.

Table lists all statistically significant identified proteins with at least 4-fold enrichment in the HA-SPOCD1 immunoprecipitation from HEK293T cells transfected with C19ORF84. Enrichment and p-value are shown as indicated. P<0.05, two-sided Student's t-test.

Protein.ID	Protein	Gene	Enrichment	p-value
B1ASB6	SPOCD1	Spocd1	2531.3	1.53E-04
Q9Y657	SPIN1	SPIN1	233.8	5.25E-04
H3BKT1	C19ORF84	Gm38999	62.6	1.08E-03
Q13555	CAMK2G	CAMK2G	16.1	8.92E-04
Q9BPZ2	SPIN2B	SPIN2B	15.6	7.79E-04
Q13557	CAMK2D	CAMK2D	11.6	1.35E-02
O14879	IFIT3	IFIT3	11.4	2.22E-04
Q5JUX0	SPIN3	SPIN3	10.8	1.68E-03
P09914	IFIT1	IFIT1	10.3	5.04E-03
P20674	COX5A	COX5A	7.9	1.62E-03
P42224	STAT1	STAT1	7.8	1.09E-02
Q96EY1	DNAJA3	DNAJA3	7.3	1.12E-02
Q8WZ42	TTN	TTN	7.0	8.91E-03
Q14240	EIF4A2	EIF4A2	6.3	5.50E-03
P61960	UFM1	UFM1	5.9	3.53E-03
P05161	ISG15	ISG15	5.9	8.93E-03
P62136	PPP1CA	PPP1CA	5.1	1.55E-03
Q9NX47	MARCHF5	MARCHF5	4.5	6.17E-03
Q14139	UBE4A	UBE4A	4.1	2.48E-02

Table 13: Proteins identified in SPOCD1-ΔTFIISM-HA IP-MS from HEK293T cells transfected with C19ORF84.

Table lists all statistically significant identified proteins with at least 4-fold enrichment in the HA-SPOCD1-ΔTFIISM immunoprecipitation from HEK293T cells transfected with C19ORF84. Enrichment and p-value are shown as indicated. P<0.05, two-sided Student's t-test.

Protein.ID	Protein	Gene	Enrichment	p-value
B1ASB6	SPOCD1	Spocd1	5983.6	3.61E-05
Q9Y657	SPIN1	SPIN1	408.9	7.20E-05
H3BKT1	C19ORF84	Gm38999	50.5	7.03E-05
Q9BPZ2	SPIN2B	SPIN2B	37.9	2.24E-07
Q13555	CAMK2G	CAMK2G	29.8	3.91E-04
Q96EY1	DNAJA3	DNAJA3	14.5	8.56E-03
Q5JUX0	SPIN3	SPIN3	14.5	7.35E-04
P08134	RHOC	RHOC	10.4	6.54E-04
Q13557	CAMK2D	CAMK2D	9.6	8.14E-03
Q9NX47	MARCHF5	MARCHF5	8.4	1.32E-03
P62136	PPP1CA	PPP1CA	7.4	3.45E-04
P09914	IFIT1	IFIT1	6.9	1.17E-02
Q14240	EIF4A2	EIF4A2	6.0	2.04E-04
P36873	PPP1CC	PPP1CC	5.0	9.28E-04
P42224	STAT1	STAT1	5.0	2.18E-02
Q15392	DHCR24	DHCR24	4.9	2.50E-02
O14879	IFIT3	IFIT3	4.7	3.65E-03
Q9UBM7	DHCR7	DHCR7	4.5	1.79E-03
P62140	PPP1CB	PPP1CB	4.5	5.28E-04

Table 14: Proteins identified in SPOCD1-ΔSPOC-HA IP-MS from HEK293T cells transfected with C19ORF84.

Table lists all statistically significant identified proteins with at least 4-fold enrichment in the HA-SPOCD1-ΔSPOC immunoprecipitation from HEK293T cells transfected with C19ORF84. Enrichment and p-value are shown as indicated. P<0.05, two-sided Student's t-test.

Protein.ID	Protein	Gene	Enrichment	p-value
B1ASB6	SPOCD1	Spocd1	759.9	1.37E-03
Q9Y657	SPIN1	SPIN1	162.5	3.02E-03
O14879	IFIT3	IFIT3	22.7	2.53E-03
H3BKT1	C19ORF84	Gm38999	19.3	2.36E-02
P09914	IFIT1	IFIT1	18.6	1.47E-04
P42224	STAT1	STAT1	12.8	1.64E-02
Q13555	CAMK2G	CAMK2G	11.3	1.25E-02
P05161	ISG15	ISG15	9.9	1.18E-02
Q01650	LAT1	SLC7A5	9.6	1.50E-02
P20674	COX5A	COX5A	9.3	1.39E-02
Q14240	EIF4A2	EIF4A2	8.3	1.75E-02
P62328	TYB4	TMSB4X	6.7	3.87E-02
P83916	CBX1	CBX1	6.7	1.17E-03
P18031	PTPN1	PTPN1	5.7	2.18E-02
Q9Y314	NOSIP	NOSIP	5.2	1.50E-02
P61769	B2MG	B2M	5.1	1.99E-02
P0DN76	U2AF	U2AF1L5	5.0	6.71E-03
Q9UK45	LSM7	LSM7	5.0	3.44E-03
Q9NWV4	CZIB	CZIB	5.0	8.70E-03
O00217	NDUFS8	NDUFS8	4.6	2.23E-03
O75352	MPDU1	MPDU1	4.6	2.12E-02
Q9NXH9	TRMT1	TRMT1	4.3	7.02E-04

Table 15: Peptides of Crosslink-MS between SPOCD1 and SPIN1.

PepSeq: Peptide sequence, LinkPos: Linker position, ExpMz: expected Mz value, ExpMass: expected mass, CalcMz, calculated Mz value, CalcMass: Calculated mass.

Protein1	PepSeq1	LinkPos1	Protein2	PepSeq2	LinkPos2	Score
SPIN1	EPGEVVDSLVGKQVEYAK	12	SPOCD1	KDFPR	1	11.37
SPIN1	EPGEVVDSLVGKQVEYAK	12	SPOCD1	KDFPR	1	10.69
SPOCD1	GVKTVYYVYTGSGIR	3	SPIN1	QVEYAKEDGSK	6	9.49
SPOCD1	GVKTVYYVYTGSGIR	3	SPIN1	QVEYAKEDGSKR	6	8.75
SPIN1	EDGSKR	5	SPOCD1	KDFPR	1	8.92
SPOCD1	KDFPR	1	SPIN1	EDGSKR	5	7.61

Table 16: Proteins identified in E14.5 IP-MS from foetal testis.

Table lists all statistically significant identified proteins with at least 4-fold enrichment in the HA-SPOCD1 immunoprecipitation from E14.5. Enrichment and p-value are shown as indicated. P<0.05, two-sided Student's t-test.

Protein.ID	Protein	Gene	Enrichment	p-value
B1ASB6	SPOCD1	Spocd1	20.2	3.17E-04
Q61142	SPIN1	Spin1	11.4	7.99E-03
Q3UZQ3	EEF1A1	Eef1a1	9.9	4.78E-02
Q9JI39	ABCB10	Abcb10	5.2	3.48E-02
Q9CQE3	MRPS17	Mrps17	4.8	1.36E-04
Q91VT0	POLD1	Pold1	4.4	8.39E-03

Table 17: Deregulated transposons in *Spocd1*$^{-/-}$ P20 testis.

All significantly deregulated transposons in Spocd1$^{-/-}$ P20 testis compared to wild type are shown. Log_2 fold change and p-value are shown as indicated. P<0.01, Benjamini-Hochberg adjusted two-sided Wald test, >2-fold change.

TE	Fold Change	p-value
L1Md_Gf_5end	4.313397147	1.07E-120
L1MdTf_II	4.188874068	6.22E-161
L1MdTf_I	4.154862378	4.07E-157
L1MdA_I	3.376421483	1.77E-08
IAPA_MM	3.25518228	4.25E-25
L1_MM	3.222265738	1.46E-08
GSAT_MM	3.105330597	5.70E-07
MMERVK10C	3.09248506	1.23E-43
IAPEY5_I	2.393241767	6.63E-16
L1MdA_II	2.336343028	4.95E-06
RLTR1IAP_MM	2.236978286	1.82E-19
IAPEZI	2.035161072	8.17E-29
IAPEYI	1.763579261	2.90E-12
IAPEY4_LTR	1.613067821	2.97E-12
IAPEY4_I	1.605771134	7.67E-13
L1MdA_III	1.578347008	1.99E-05
ERVB4_1B-I_MM	1.52996832	4.34E-16
L1MdGf_II	1.485955625	1.37E-10
IAPLTR1_Mm	1.44624431	1.20E-17
IAP5-Mm_I	1.436609386	4.87E-17
L1MdTf_III	1.383782621	1.26E-13
RLTR10B2	1.359400966	0.000106574
L1MdGf_I	1.320545901	5.27E-06
SRV_MM-int	1.313437606	1.50E-06

IAPLTR4	1.286526944	0.001910298
RLTR11B	1.195482295	0.008079329
ERVB4_1B-LTR_MM	1.117933848	5.98E-07
IAPLTR1a_I_MM	1.067636827	7.12E-14
IAP-d	1.057592293	1.35E-07
IAPEY_LTR	1.023802039	0.000650549
RLTR27_MM	-1.058447755	0.00295282
RLTR10F	-1.076573502	5.97E-05
RLTR1D_MM	-1.194358123	0.000268012
RLTR44E_LTR	-1.198956301	0.009312125
RLTR30_MM	-1.207043131	0.00920941
RLTR19B_MM	-1.270478325	0.000130374
RMER2	-1.362146877	0.000130374
ERVB5_2-LTR_MM	-1.406550807	1.12E-06
RLTR13D2	-1.470015488	1.00E-11
RLTR19A_MM	-1.553120483	0.00773014
RLTR43C	-1.66463772	9.93E-05
RLTR9D	-1.752353336	8.65E-10
RLTR10D2	-1.796240835	3.93E-07
ERVB7_3-LTR_MM	-2.094260281	0.000139046
RLTR50B	-2.221295866	2.07E-05
LTRIS5	-2.261079985	2.68E-06
RLTR41A2	-2.389136075	0.00018098
RLTR1F_Mm	-2.42550902	2.07E-07
RLTR9D2	-2.66877229	4.66E-10
IAPEY3_LTR	-2.740075348	3.23E-20
RLTR41B	-4.421002565	1.41E-07
RLTR30B_MM	-4.607736872	5.87E-05
IAPLTR3	-4.864531996	1.93E-34
RLTR9A2	-5.815360486	1.46E-05
IAPLTR3_I	-6.220278044	5.54E-130

Table 18: Deregulated transposons in *Spocd1*$^{dSpin1/dSpin1}$ P20 testis.

All significantly deregulated transposons in Spocd1$^{dSpin1/dSpin1}$ P20 testis compared to wild type are shown. Log_2 fold change and p-value are shown as indicated. P<0.01, Benjamini-Hochberg adjusted two-sided Wald test, >2-fold change.

TE	Fold Change	p-value
L1Md_Gf_5end	3.891044694	6.42E-107
L1MdTf_II	3.609445572	3.09E-120
L1MdTf_I	3.578740966	2.51E-117
L1MdA_I	3.19212338	1.51E-07
L1_MM	3.047286404	1.26E-07
MMERVK10C	2.682364915	7.76E-33
L1MdA_II	2.467950832	1.78E-06
IAPEY5_I	2.049196676	3.22E-12
RLTR10B2	1.970640636	2.06E-09
L1MdA_III	1.897733797	2.81E-07
L1MdGf_II	1.693428808	1.69E-13
IAPEYI	1.594396192	1.64E-10
L1MdTf_III	1.568184198	3.38E-17
L1MdGf_I	1.242621255	2.65E-05
L1MdA_IV	1.161493765	1.97E-17
RLTR43C	-1.214152088	1.64E-05
RLTR9F	-1.282363971	0.00258532
RLTR50B	-1.283257663	0.00081605
IAPLTR3_I	-1.308278093	5.35E-08
RLTR20D_MM	-1.333841431	7.33E-05
RLTR30_MM	-1.398168259	8.57E-07
RLTR10D2	-1.403272988	2.49E-09
RLTR49	-1.46997044	0.001757721
RLTR9D2	-1.476592744	3.61E-06
RLTR41B	-1.608733971	5.53E-05
IAPEY3_LTR	-1.79230907	4.81E-23
RLTR1F_Mm	-1.891237133	5.38E-11
RLTR41A2	-2.032497312	1.69E-13
RLTR30B_MM	-2.060714439	2.64E-16
RLTR9A2	-3.076100531	4.27E-13
RLTR9A3B	-3.245467578	0.002927817

6.2 List of Abbreviations

5caC	**5-carboxylcytosine**
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylctosine
ABC	Ammonium bicarbonate
ACN	Acetonitrile
ADD	ATRX-DNMT3L-DNMT3A
AK4	Adenylate kinase 4
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
BAF	Barrier-to-autointegration factor
BER	Base excision repair
BLIMP1	B-lymphocyte-induced maturation protein 1
bp	Base pair
BS3	Bis(sulfosuccinimidyl)suberate
BSA	Bovine serum albumin
BYE1	Bypass of Ess-1
cccDNA	Covalently closed circular DANN
cDNA	Complementary DNA
CGI	CpG islands
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CIP	Alkaline phosphatase calf intestinal
CL-MS	Crosslink mass spectrometry
CpG	Cytosine-phosphate-Guanine
CTD	C-terminal domain
CtIP	C-terminal binding protein 1 (CtBP1) interacting protein
DIDO	Death inducer obliterator
DNA	Deoxyribonucleic acid
DNMT	DNA methyl transferase
DNMT1	DNA methyltransferase 1
DNMT3A/B/C/L	DNA methyltransferase 3A/B/C/L
DTT	Dithiothreitol
E6.5	Embryonic day 6.5
EB2	End-binding 2 protein
EDTA	Ethylenediaminetetraacetic acid
EM-seq	Whole genome methylation sequencing
ESC	Embryonic stem cells
EtBr	Ethidium bromide
EtOH	Ethanol

EZH1	Enhancer of zeste homologue 1
FACS	Fluorescence-activated cell sorting
FBS	Foetal bovine serum
FLC	Flowering locus C
FMR1	Fragile X mental retardation 1
GC	Germ cell
GST	Glutathione S-transferase
H3K27me3	H3 lysine 27 trimethylation
H3K36me3	Histone 3 lysine 36 tri-methylation
H3K4	Histone 4 lysine 4
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9me1	Histone 3 lysine 9 -methylation
H3K9me2	Histone 3 lysine 9 di-methylation
H3K9me3	Histone 3 lysine 9 tri-methylation
H3R8me2a	Histone 3 arginine 8 asymmetric di-methylation
H4K20me3	Histone 4 lysine 20 tri-methylation
HBV	Hepatitis B virus
HCD	High-energy collision dissociation
HEK293T	Human embryonic kidney 293T
HENMT1	HEN methyltransferase 1
HEZ	Heterochromatin exclusion zones
HIS	Polyhistidine
IAA	Iodoacetamide
IAP	Intracisternal A particle
ICR	Imprinted control regions
IDP	Intrinsically disordered proteins
IEX	Ion exchange
IF	Immunofluorescence
IMAC	Metal ion affinity chromatography
IMC	Intramitochondrial cement
IP	Immunoprecipitation
IP-MS	Immunoprecipitation coupled with mass spectrometry
IPTG	Isopropyl β-d-1-thiogalactopyranoside
KAP1	KRAB-associated protein 1
kb	Kilo base
kDA	Kilo Dalton
KMT2D	Lysine Methyltransferase 2D
KRAB-ZFP	Krüppel-associated box domain containing zinc finger protein
LFQ	Label-free quantitation
LIC	Ligase-independent cloning
LINE	Long interspersed nuclear element
lncRNA	Long non-coding RNA

LTR	Long terminal repeat
MBP	Methylation-binding protein
MILI	Murine Piwi-like protein 2, or PIWIL2
miRNA	MicroRNA
MIWI	Murine Piwi-like protein 1, or PIWIL1
MIWI2	Murine Piwi-like protein 4 or PIWIL4
MORC1	MORC Family CW-Type Zinc Finger 1
MOV10L1	Mov10 like RISC complex RNA helicase 1
mRNA	Messenger RNA
MS	Mass spectrometry
MTase	Methyltransferase
NA	Not applicable
NCoR	Nuclear receptor corepressor
NOA	Non-obstructive azoospermia
NPC	Nuclear pore complex
NSCLC	Non-small-cell lung cancer
nt	Nucleotide
NURD	Nucleosome remodelling deacetylase
P3	Postnatal day 3
PAS	Periodic acid-Schiff stain
PBS	Phosphate-buffered saline
PBS-T	Phosphate-buffered saline with 0.1% (v/v) Tween-20
PCI	Phenol, chloroform, isoamyl alcohol
PCR	Polymerase chain reaction
PGC	Primordial germ cell
PHD	Plant homeo domain
PHF3	PHD finger protein 3
piRISC	piRNA- induced silencing complex
piRNA	PIWI-interacting RNA
PIWI	P-element induced wimpy
PNLDC1	PARN like ribonuclease domain containing exonuclease 1
Pol II	RNA polymerase II
PRDM14	PR-domain containing protein 14
PWWP	Pro-Trp-Trp-Pro
RBM15	RNA binding motif protein 15
RBM15B	RNA binding motif protein 15B
rDNA	Ribosomal deoxyribonucleic acid
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
scRNA-seq	Single-cell RNA sequencing
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography

SETDB1	SET Domain Bifurcated Histone Lysine Methyltransferase 1
sgRNA	Single guide RNA
SINE	Short interspersed nuclear element
siRNA	Small interfering RNA
SMRT	Silencing mediator for retinoid and thyroid receptors
SOC	Super optimal broth
SPEN/ Sharp	SMRT/HDAC1-associated repressor protein
Spin/Ssty	Spindlin/Y-linked spermiogenesis-specific transcript repeat
SPIN1	Spindlin1
SPOC	Spen paralog and ortholog c-terminal
SPOCD1	SPOC domain-containing 1
SSC	Spermatogonial stem cells
ssDNA	Single-stranded DNA
TDG	Thymine DNA glycosylase
TDRD9	Tudor Domain Containing 9
TDRKH	Tudor and KH domain containing
TE	Transposable element
TET1/2	Ten eleven translocation 1 and 2
Tex15	Testis expressed 15
TFA	Trifluoroacetic acid
TFIIS	Transcription factor II S
TFIIS-M	Transcription factor II S mid domain
TPR	Translocated promoter region, nuclear basket protein
TUNEL	Terminal deoxynucleotidyl transferase dUTP nick end labelling
UHRF1	Ubiquitin like with PHD and ring finger domains 1
WTAP	WT1 Associated Protein

7 References

Aapola, U., K. Kawasaki, H. S. Scott, J. Ollila, M. Vihinen, M. Heino, A. Shintani, K. Kawasaki, S. Minoshima, K. Krohn, S. E. Antonarakis, N. Shimizu, J. Kudoh, and P. Peterson. 2000. 'Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family', *Genomics*, 65: 293-8.

Adams, I. R., and A. McLaren. 2002. 'Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis', *Development*, 129: 1155-64.

Aksenova, V., A. Smith, H. Lee, P. Bhat, C. Esnault, S. Chen, J. Iben, R. Kaufhold, K. C. Yau, C. Echeverria, B. Fontoura, A. Arnaoutov, and M. Dasso. 2020. 'Nucleoporin TPR is an integral component of the TREX-2 mRNA export pathway', *Nat Commun*, 11: 4577.

Anderson, R., T. K. Copeland, H. Scholer, J. Heasman, and C. Wylie. 2000. 'The onset of germ cell migration in the mouse embryo', *Mech Dev*, 91: 61-8.

Appel, L. M., J. Benedum, M. Engl, S. Platzer, A. Schleiffer, X. Strobl, and D. Slade. 2023. 'SPOC domain proteins in health and disease', *Genes & development*, 37: 140-70.

Appel, L. M., V. Franke, J. Benedum, I. Grishkovskaya, X. Strobl, A. Polyansky, G. Ammann, S. Platzer, A. Neudolt, A. Wunder, L. Walch, S. Kaiser, B. Zagrovic, K. Djinovic-Carugo, A. Akalin, and D. Slade. 2023. 'The SPOC domain is a phosphoserine binding module that bridges transcription machinery with co- and post-transcriptional regulators', *Nat Commun*, 14: 166.

Appel, L. M., V. Franke, M. Bruno, I. Grishkovskaya, A. Kasiliauskaite, T. Kaufmann, U. E. Schoeberl, M. G. Puchinger, S. Kostrhon, C. Ebenwaldner, M. Sebesta, E. Beltzung, K. Mechtler, G. Lin, A. Vlasova, M. Leeb, R. Pavri, A. Stark, A. Akalin, R. Stefl, C. Bernecky, K. Djinovic-Carugo, and D. Slade. 2021. 'PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC', *Nat Commun*, 12: 6078.

Aravin, A. A., R. Sachidanandam, A. Girard, K. Fejes-Toth, and G. J. Hannon. 2007. 'Developmentally regulated piRNA clusters implicate MILI in transposon control', *Science*, 316: 744-7.

Aravin, A. A., G. W. van der Heijden, J. Castaneda, V. V. Vagin, G. J. Hannon, and A. Bortvin. 2009. 'Cytoplasmic compartmentalization of the fetal piRNA pathway in mice', *PLoS Genet*, 5: e1000764.

Aravin, Alexei A, Natalia M Naumova, Alexei V Tulin, Vasilii V Vagin, Yakov M Rozovsky, and Vladimir A Gvozdev. 2001. 'Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline', *Current Biology*, 11: 1017-27.

Aravin, Alexei A., Ravi Sachidanandam, Deborah Bourc'his, Christopher Schaefer, Dubravka Pezic, Katalin Fejes Toth, Timothy Bestor, and Gregory J. Hannon. 2008. 'A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice', *Molecular cell*, 31: 785-99.

Aravin, Alexei, Dimos Gaidatzis, Sébastien Pfeffer, Mariana Lagos-Quintana, Pablo Landgraf, Nicola Iovino, Patricia Morris, Michael J Brownstein, Satomi Kuramochi-Miyagawa, and Toru Nakano. 2006. 'A novel class of small RNAs bind to MILI protein in mouse testes', *Nature*, 442: 203.

Ariyoshi, Mariko, and John W.R. Schwabe. 2003. 'A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling', *Genes & development*, 17: 1909-20.

Armisen, J., M. J. Gilchrist, A. Wilczynska, N. Standart, and E. A. Miska. 2009. 'Abundant and dynamically expressed miRNAs, piRNAs, and other small RNAs in the vertebrate Xenopus tropicalis', *Genome Res*, 19: 1766-75.

Aslanidis, C., and P. J. de Jong. 1990. 'Ligation-independent cloning of PCR products (LIC-PCR)', *Nucleic Acids Res*, 18: 6069-74.

Bangs, Peter, Brian Burke, Christine Powers, Roger Craig, Aruna Purohit, and Stephen Doxsey. 1998. 'Functional Analysis of Tpr: Identification of Nuclear Pore Complex Association and Nuclear Localization Domains and a Role in mRNA Export', *Journal of Cell Biology*, 143: 1801-12.

Bannister, A. J., and T. Kouzarides. 2011. 'Regulation of chromatin by histone modifications', *Cell Res*, 21: 381-95.

Bannister, A. J., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides. 2001. 'Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain', *Nature*, 410: 120-4.

Barau, Joan, Aurélie Teissandier, Natasha Zamudio, Stéphanie Roy, Valérie Nalesso, Yann Hérault, Florian Guillou, and Déborah Bourc'his. 2016. 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity', *Science*, 354: 909-12.

Batista, P. J., J. G. Ruby, J. M. Claycomb, R. Chiang, N. Fahlgren, K. D. Kasschau, D. A. Chaves, W. Gu, J. J. Vasale, S. Duan, D. Conte, Jr., S. Luo, G. P. Schroth, J. C. Carrington, D. P. Bartel, and C. C. Mello. 2008. 'PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans', *Mol Cell*, 31: 67-78.

Batki, Julia, Jakob Schnabl, Juncheng Wang, Dominik Handler, Veselin I. Andreev, Christian E. Stieger, Maria Novatchkova, Lisa Lampersberger, Kotryna Kauneckaite, Wei Xie, Karl Mechtler, Dinshaw J. Patel, and Julius Brennecke. 2019. 'The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation', *Nature Structural & Molecular Biology*, 26: 720-31.

Bellve, A. R., J. C. Cavicchia, C. F. Millette, D. A. O'Brien, Y. M. Bhatnagar, and M. Dym. 1977. 'Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization', *J Cell Biol*, 74: 68-85.

Ben Maamar, M., E. E. Nilsson, and M. K. Skinner. 2021. 'Epigenetic transgenerational inheritance, gametogenesis and germline developmentdagger', *Biol Reprod*, 105: 570-92.

Bilodeau, S., M. H. Kagey, G. M. Frampton, P. B. Rahl, and R. A. Young. 2009. 'SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state', *Genes & development*, 23: 2484-9.

Bird, A. 2002. 'DNA methylation patterns and epigenetic memory', *Genes & development*, 16: 6-21.

Blake, J. A., R. Baldarelli, J. A. Kadin, J. E. Richardson, C. L. Smith, C. J. Bult, and Group Mouse Genome Database. 2021. 'Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology', *Nucleic Acids Res*, 49: D981-D87.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30: 2114-20.

Borgel, J., S. Guibert, Y. Li, H. Chiba, D. Schubeler, H. Sasaki, T. Forne, and M. Weber. 2010. 'Targets and dynamics of promoter DNA methylation during early mouse development', *Nat Genet*, 42: 1093-100.

Bostick, M., J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan, and S. E. Jacobsen. 2007. 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science*, 317: 1760-4.

Boumendil, C., P. Hari, K. C. F. Olsen, J. C. Acosta, and W. A. Bickmore. 2019. 'Nuclear pore density controls heterochromatin reorganization during senescence', *Genes & development*, 33: 144-49.

Bourc'his, Déborah, and Timothy H Bestor. 2004b. 'Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L', *Nature*, 431: 96.

Braig, S., and A. K. Bosserhoff. 2013. 'Death inducer-obliterator 1 (Dido1) is a BMP target gene and promotes BMP-induced melanoma progression', *Oncogene*, 32: 837-48.

Brennecke, Julius, Alexei A Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, and Gregory J Hannon. 2007. 'Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila', *Cell*, 128: 1089-103.

Capelson, M., Y. Liang, R. Schulte, W. Mair, U. Wagner, and M. W. Hetzer. 2010. 'Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes', *Cell*, 140: 372-83.

Caraballo, G. I., R. Rosales, M. Viettri, J. M. Castillo, R. Cruz, S. Ding, H. B. Greenberg, and J. E. Ludert. 2022. 'The Dengue Virus Nonstructural Protein 1 (NS1) Interacts with the Putative Epigenetic Regulator DIDO1 to Promote Flavivirus Replication in Mosquito Cells', *J Virol*, 96: e0070422.

Carmell, Michelle A., Angélique Girard, Henk J. G. van de Kant, Deborah Bourc'his, Timothy H. Bestor, Dirk G. de Rooij, and Gregory J. Hannon. 2007. 'MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline', *Developmental Cell*, 12: 503-14.

Casacuberta, E., and J. Gonzalez. 2013. 'The impact of transposable elements in environmental adaptation', *Mol Ecol*, 22: 1503-17.

Chédin, Frédéric, Michael R Lieber, and Chih-Lin Hsieh. 2002. 'The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a', *Proceedings of the National Academy of Sciences*, 99: 16916-21.

Chen, X., Y. W. Wang, A. Y. Xing, S. Xiang, D. B. Shi, L. Liu, Y. X. Li, and P. Gao. 2016. 'Suppression of SPIN1-mediated PI3K-Akt pathway by miR-489 increases chemosensitivity in breast cancer', *J Pathol*, 239: 459-72.

Chenais, B. 2013. 'Transposable elements and human cancer: a causal relationship?', *Biochim Biophys Acta*, 1835: 28-35.

Cheung, A. C., and P. Cramer. 2011. 'Structural basis of RNA polymerase II backtracking, arrest and reactivation', *Nature*, 471: 249-53.

Chew, Ting Gang, Anne Peaston, Ai Khim Lim, Chanchao Lorthongpanich, Barbara B. Knowles, and Davor Solter. 2013. 'A Tudor Domain Protein SPINDLIN1 Interacts with the mRNA-Binding Protein SERBP1 and Is Involved in Mouse Oocyte Meiotic Resumption', *PLOS ONE*, 8: e69764.

Choi, J. W., M. H. Zhao, S. Liang, J. Guo, Z. L. Lin, Y. H. Li, Y. J. Jo, N. H. Kim, and X. S. Cui. 2017. 'Spindlin 1 is essential for metaphase II stage maintenance and chromosomal stability in porcine oocytes', *Mol Hum Reprod*, 23: 166-76.

Clark, J. M., and E. M. Eddy. 1975. 'Fine structural observations on the origin and associations of primordial germ cells of the mouse', *Dev Biol*, 47: 136-55.

Comazzetto, S., M. Di Giacomo, K. D. Rasmussen, C. Much, C. Azzi, E. Perlas, M. Morgan, and D. O'Carroll. 2014. 'Oligoasthenoteratozoospermia and infertility in mice deficient for miR-34b/c and miR-449 loci', *PLoS Genet*, 10: e1004597.

Consortium, Fantom, Riken Pmi the, Clst, A. R. Forrest, H. Kawaji, M. Rehli, J. K. Baillie, M. J. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jorgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. M. Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drablos, A. S. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furino, J. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. P. Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-Sim, R. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, S. Noma, T. Noazaki, S. Ogishima, N. Ohkura, H. Ohimiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. Prendergast, O. J. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama,

M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. t Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyodo, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verado, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci, and Y. Hayashizaki. 2014. 'A promoter-level mammalian expression atlas', *Nature*, 507: 462-70.

Cordes, V. C., M. E. Hase, and L. Muller. 1998. 'Molecular segments of protein Tpr that confer nuclear targeting and association with the nuclear pore complex', *Exp Cell Res*, 245: 43-56.

Cordes, V. C., S. Reidenbach, H. R. Rackwitz, and W. W. Franke. 1997. 'Identification of protein p270/Tpr as a constitutive component of the nuclear pore complex-attached intranuclear filaments', *J Cell Biol*, 136: 515-29.

Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann. 2014. 'Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ', *Mol Cell Proteomics*, 13: 2513-26.

Das, P. P., M. P. Bagijn, L. D. Goldstein, J. R. Woolford, N. J. Lehrbach, A. Sapetschnig, H. R. Buhecha, M. J. Gilchrist, K. L. Howe, R. Stark, N. Matthews, E. Berezikov, R. F. Ketting, S. Tavare, and E. A. Miska. 2008. 'Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline', *Mol Cell*, 31: 79-90.

De Fazio, S., N. Bartonicek, M. Di Giacomo, C. Abreu-Goodger, A. Sankar, C. Funaya, C. Antony, P. N. Moreira, A. J. Enright, and D. O'Carroll. 2011. 'The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements', *Nature*, 480: 259-63.

de França, L. R., L. Ye Sj Fau - Ying, M. Ying L Fau - Sandberg, L. D. Sandberg M Fau - Russell, and L. D. Russell. 1995. 'Morphometry of rat germ cells during spermatogenesis'.

de Kloet, R. S., and S. R. de Kloet. 2005. 'The evolution of the spindlin gene in birds: sequence analysis of an intron of the spindlin W and Z gene reveals four major divisions of the Psittaciformes', *Mol Phylogenet Evol*, 36: 706-21.

de Rooij, D. G., and L. D. Russell. 2000. 'All you wanted to know about spermatogonia but were afraid to ask'.

Deininger, P. L., and M. A. Batzer. 2002. 'Mammalian retroelements', *Genome Res*, 12: 1455-65.

Deng, W., and H. Lin. 2002. 'miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis', *Dev Cell*, 2: 819-30.

Devi, M. S., R. Meiguilungpou, A. L. Sharma, C. Anjali, K. M. Devi, L. S. Singh, and T. R. Singh. 2019. 'Spindlin docking protein (SPIN.DOC) interaction with SPIN1 (a histone code reader) regulates Wnt signaling', *Biochem Biophys Res Commun*, 511: 498-503.

Dewannieux, M., A. Dupressoir, F. Harper, G. Pierron, and T. Heidmann. 2004. 'Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells', *Nat Genet*, 36: 534-9.

Di Giacomo, M., S. Comazzetto, S. C. Sampath, S. C. Sampath, and D. O'Carroll. 2014. 'G9a co-suppresses LINE1 elements in spermatogonia', *Epigenetics Chromatin*, 7: 24.

Di Giacomo, Monica, Stefano Comazzetto, Harpreet Saini, Serena De Fazio, Claudia Carrieri, Marcos Morgan, Lina Vasiliauskaite, Vladimir Benes, Anton J Enright, and Dónal O'Carroll. 2013. 'Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis', *Molecular cell*, 50: 601-08.

Dong, K. B., I. A. Maksakova, F. Mohn, D. Leung, R. Appanah, S. Lee, H. W. Yang, L. L. Lam, D. L. Mager, D. Schubeler, M. Tachibana, Y. Shinkai, and M. C. Lorincz. 2008. 'DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity', *EMBO J*, 27: 2691-701.

Doroquez, D. B., T. L. Orr-Weaver, and I. Rebay. 2007. 'Split ends antagonizes the Notch and potentiates the EGFR signaling pathways during Drosophila eye development', *Mech Dev*, 124: 792-806.

Du, Yongming, Yinxia Yan, Si Xie, Hao Huang, Xin Wang, Ray Kit Ng, Ming-Ming Zhou, and Chengmin Qian. 2021. 'Structural mechanism of bivalent histone H3K4me3K9me3 recognition by the Spindlin1/C11orf84 complex in rRNA transcription activation', *Nature Communications*, 12: 949.

Ducroux, A., S. Benhenda, L. Riviere, O. J. Semmes, M. Benkirane, and C. Neuveut. 2014. 'The Tudor domain protein Spindlin1 is involved in intrinsic antiviral defense against incoming hepatitis B Virus and herpes simplex virus type 1', *PLoS Pathog*, 10: e1004343.

Dura, M., A. Teissandier, M. Armand, J. Barau, C. Lapoujade, P. Fouchet, L. Bonneville, M. Schulz, M. Weber, L. G. Baudrin, S. Lameiras, and D. Bourc'his. 2022. 'DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis', *Nat Genet*, 54: 469-80.

Dyson, H. J., and P. E. Wright. 2005. 'Intrinsically unstructured proteins and their functions', *Nat Rev Mol Cell Biol*, 6: 197-208.

Ecco, G., M. Imbeault, and D. Trono. 2017. 'KRAB zinc finger proteins', *Development*, 144: 2719-29.

Eddy, E. M. 1974. 'Fine structural observations on the form and distribution of nuage in germ cells of the rat', *Anat Rec*, 178: 731-57.

Eickbush, Thomas H., and Harmit S. Malik. 2007. 'Origins and Evolution of Retrotransposons.' in, *Mobile DNA II*.

ElMaghraby, Mostafa F., Peter Refsing Andersen, Florian Pühringer, Ulrich Hohmann, Katharina Meixner, Thomas Lendl, Laszlo Tirian, and Julius Brennecke. 2019. 'A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production', *Cell*, 178: 964-79.e20.

Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. 'Features and development of Coot', *Acta Crystallogr D Biol Crystallogr*, 66: 486-501.

Evans, P. 2006. 'Scaling and assessment of data quality', *Acta Crystallogr D Biol Crystallogr*, 62: 72-82.

Evans, P. R., and G. N. Murshudov. 2013. 'How good are my data and what is the resolution?', *Acta Crystallogr D Biol Crystallogr*, 69: 1204-14.

Ewen, K. A., and P. Koopman. 2010. 'Mouse germ cell development: from specification to sex determination', *Mol Cell Endocrinol*, 323: 76-93.

Fischer, U., A. K. Struss, D. Hemmer, A. Michel, W. Henn, W. I. Steudel, and E. Meese. 2001. 'PHF3 expression is frequently reduced in glioma', *Cytogenet Cell Genet*, 94: 131-6.

Forghanifard, M. M., P. Naeimi Khorasanizadeh, M. R. Abbaszadegan, A. Javdani Mallak, and M. Moghbeli. 2020. 'Role of DIDO1 in Progression of Esophageal Squamous Cell Carcinoma', *J Gastrointest Cancer*, 51: 83-87.

Franz, H., H. Greschik, D. Willmann, L. Ozretic, C. A. Jilg, E. Wardelmann, M. Jung, R. Buettner, and R. Schule. 2015. 'The histone code reader SPIN1 controls RET signaling in liposarcoma', *Oncotarget*, 6: 4773-89.

Frosst, P., T. Guan, C. Subauste, K. Hahn, and L. Gerace. 2002. 'Tpr is localized within the nuclear basket of the pore complex and has a role in nuclear protein export', *J Cell Biol*, 156: 617-30.

Fu, Qi, and P. Jeremy Wang. 2014. 'Mammalian piRNAs', *Spermatogenesis*, 4: e27889.

Futterer, A., J. de Celis, R. Navajas, L. Almonacid, J. Gutierrez, A. Talavera-Gutierrez, C. Pacios-Bras, I. Bernascone, F. Martin-Belmonte, and A. C. Martinez. 2017. 'DIDO as a Switchboard that Regulates Self-Renewal and Differentiation in Embryonic Stem Cells', *Stem Cell Reports*, 8: 1062-75.

Gan, Haiyun, Lu Wen, Shangying Liao, Xiwen Lin, Tingting Ma, Jun Liu, Chun-xiao Song, Min Wang, Chuan He, Chunsheng Han, and Fuchou Tang. 2013. 'Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis', *Nature Communications*, 4: 1995.

Gan, L., C. Yang, L. Zhao, S. Wang, Z. Gao, and Y. Ye. 2023. 'Prognostic Biomarker SPOCD1 and Its Correlation with Immune Infiltrates in Colorectal Cancer', *Biomolecules*, 13.

Gatchalian, J., A. Futterer, S. B. Rothbart, Q. Tong, H. Rincon-Arano, A. Sanchez de Diego, M. Groudine, B. D. Strahl, A. C. Martinez, K. H. van Wely, and T. G. Kutateladze. 2013. 'Dido3 PHD modulates cell differentiation and division', *Cell Rep*, 4: 148-58.

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith. 2009. 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nat Methods*, 6: 343-5.

Ginsburg, Malka, MH Snow, and Anne McLaren. 1990. 'Primordial germ cells in the mouse embryo during gastrulation', *Development*, 110: 521-28.

Girard, Angélique, Ravi Sachidanandam, Gregory J Hannon, and Michelle A Carmell. 2006. 'A germline-specific class of small RNAs binds mammalian Piwi proteins', *Nature*, 442: 199.

Godin, I., C. Wylie, and J. Heasman. 1990. 'Genital ridges exert long-range effects on mouse primordial germ cell numbers and direction of migration in culture', *Development*, 108: 357-63.

Goodier, J. L., E. M. Ostertag, K. Du, and H. H. Kazazian, Jr. 2001. 'A novel active L1 retrotransposon subfamily in the mouse', *Genome Res*, 11: 1677-85.

Gozalo, A., A. Duke, Y. Lan, P. Pascual-Garcia, J. A. Talamas, S. C. Nguyen, P. P. Shah, R. Jain, E. F. Joyce, and M. Capelson. 2020. 'Core Components of the Nuclear Pore Bind Distinct States of Chromatin and Contribute to Polycomb Repression', *Mol Cell*, 77: 67-81 e7.

Greenberg, M. V. C., and D. Bourc'his. 2019. 'The diverse roles of DNA methylation in mammalian development and disease', *Nat Rev Mol Cell Biol*, 20: 590-607.

Grivna, Shane T, Ergin Beyret, Zhong Wang, and Haifan Lin. 2006. 'A novel class of small RNAs in mouse spermatogenic cells', *Genes & development*, 20: 1709-14.

Gu, T., T. Zhao, U. Kohli, and R. S. Hewes. 2017. 'The large and small SPEN family proteins stimulate axon outgrowth during neurosecretory cell remodeling in Drosophila', *Dev Biol*, 431: 226-38.

Guibert, Sylvain, Thierry Forné, and Michael Weber. 2012. 'Global profiling of DNA methylation erasure in mouse primordial germ cells'.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M. C. Siomi. 2007. 'A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila', *Science*, 315: 1587-90.

Hackett, J. A., R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down, and M. A. Surani. 2013. 'Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine', *Science*, 339: 448-52.

Hackett, J. A., J. J. Zylicz, and M. A. Surani. 2012. 'Parallel mechanisms of epigenetic reprogramming in the germline', *Trends Genet*, 28: 164-74.

Hajkova, P., K. Ancelin, T. Waldmann, N. Lacoste, U. C. Lange, F. Cesari, C. Lee, G. Almouzni, R. Schneider, and M. A. Surani. 2008. 'Chromatin dynamics during epigenetic reprogramming in the mouse germ line', *Nature*, 452: 877-81.

Hajkova, Petra, Sylvia Erhardt, Natasha Lane, Thomas Haaf, Osman El-Maarri, Wolf Reik, Jörn Walter, and M Azim Surani. 2002. 'Epigenetic reprogramming in mouse primordial germ cells', *Mechanisms of development*, 117: 15-23.

Hancks, D. C., and H. H. Kazazian, Jr. 2016. 'Roles for retrotransposon insertions in human disease', *Mob DNA*, 7: 9.

Hansen, C. L., and F. Pelegri. 2021. 'Primordial Germ Cell Specification in Vertebrate Embryos: Phylogenetic Distribution and Conserved Molecular Features of Preformation and Induction', *Front Cell Dev Biol*, 9: 730332.

Harlen, K. M., and L. S. Churchman. 2017. 'Subgenic Pol II interactomes identify region-specific transcription elongation regulators', *Mol Syst Biol*, 13: 900.

Harris, R. S., K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger, and M. H. Malim. 2003. 'DNA deamination mediates innate immunity to retroviral infection', *Cell*, 113: 803-9.

Hase, M. E., N. V. Kuznetsov, and V. C. Cordes. 2001. 'Amino acid substitutions of coiled-coil protein Tpr abrogate anchorage to the nuclear pore complex but not parallel, in-register homodimerization', *Mol Biol Cell*, 12: 2433-52.

Hata, K., M. Kusumi, T. Yokomine, E. Li, and H. Sasaki. 2006. 'Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells', *Mol Reprod Dev*, 73: 116-22.

Hayashi, K., S. M. de Sousa Lopes, and M. A. Surani. 2007. 'Germ cell specification in mice', *Science*, 316: 394-6.

He, Y. F., B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, and G. L. Xu. 2011. 'Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA', *Science*, 333: 1303-7.

Hein, M. Y., N. C. Hubner, I. Poser, J. Cox, N. Nagaraj, Y. Toyoda, I. A. Gak, I. Weisswange, J. Mansfeld, F. Buchholz, A. A. Hyman, and M. Mann. 2015. 'A human interactome

in three quantitative dimensions organized by stoichiometries and abundances', *Cell*, 163: 712-23.

Hess, Rex A., and Luiz Renato de Franca. 2008. 'Spermatogenesis and Cycle of the Seminiferous Epithelium.' in C. Yan Cheng (ed.), *Molecular Mechanisms in Spermatogenesis* (Springer New York: New York, NY).

Hess, Rex, and Luiz França. 2005. 'History of the Sertoli Cell Discovery.' in.

Hilscher, B., W. Hilscher, B. Bulthoff-Ohnolz, U. Kramer, A. Birke, H. Pelzer, and G. Gauss. 1974. 'Kinetics of gametogenesis. I. Comparative histological and autoradiographic studies of oocytes and transitional prospermatogonia during oogenesis and prespermatogenesis', *Cell Tissue Res*, 154: 443-70.

Hornyik, C., L. C. Terzi, and G. G. Simpson. 2010. 'The spen family protein FPA controls alternative cleavage and polyadenylation of RNA', *Dev Cell*, 18: 203-13.

Horwich, M. D., C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, and P. D. Zamore. 2007. 'The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC', *Curr Biol*, 17: 1265-72.

Houwing, S., L. M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D. V. Filippov, H. Blaser, E. Raz, C. B. Moens, R. H. Plasterk, G. J. Hannon, B. W. Draper, and R. F. Ketting. 2007. 'A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish', *Cell*, 129: 69-82.

Hruz, T., O. Laule, G. Szabo, F. Wessendorp, S. Bleuler, L. Oertle, P. Widmayer, W. Gruissem, and P. Zimmermann. 2008. 'Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes', *Adv Bioinformatics*, 2008: 420747.

Hubner, N. C., and M. Mann. 2011. 'Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC)', *Methods*, 53: 453-9.

Huve, J., R. Wesselmann, M. Kahms, and R. Peters. 2008. '4Pi microscopy of the nuclear pore complex', *Biophys J*, 95: 877-85.

Inoue, N., K. D. Hess, R. W. Moreadith, L. L. Richardson, M. A. Handel, M. L. Watson, and A. R. Zinn. 1999. 'New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis', *Hum Mol Genet*, 8: 1201-7.

Ipsaro, J. J., A. D. Haase, S. R. Knott, L. Joshua-Tor, and G. J. Hannon. 2012. 'The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis', *Nature*, 491: 279-83.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang. 2011. 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science*, 333: 1300-3.

Izumi, N., K. Shoji, Y. Sakaguchi, S. Honda, Y. Kirino, T. Suzuki, S. Katsuma, and Y. Tomari. 2016. 'Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms', *Cell*, 164: 962-73.

Jain, Devanshi, Cem Meydan, Julian Lange, Corentin Claeys Bouuaert, Nathalie Lailler, Christopher E. Mason, Kathryn V. Anderson, and Scott Keeney. 2017. 'rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline', *PLOS Genetics*, 13: e1006964.

Jan, S. Z., G. Hamer, S. Repping, D. G. de Rooij, A. M. van Pelt, and T. L. Vormer. 2012. 'Molecular control of rodent spermatogenesis', *Biochim Biophys Acta*, 1822: 1838-50.

Jemc, J., and I. Rebay. 2006. 'Characterization of the split ends-like gene spenito reveals functional antagonism between SPOC family members during Drosophila eye development', *Genetics*, 173: 279-86.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. 2021. 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596: 583-89.

Jurrus, E., D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G. W. Wei, M. J. Holst, J. A. McCammon, and N. A. Baker. 2018. 'Improvements to the APBS biomolecular solvation software suite', *Protein Sci*, 27: 112-28.

Kabsch, W. 2010. 'XDS', *Acta Crystallogr D Biol Crystallogr*, 66: 125-32.

Kagiwada, S., K. Kurimoto, T. Hirota, M. Yamaji, and M. Saitou. 2013. 'Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice', *EMBO J*, 32: 340-53.

Kamminga, L. M., M. J. Luteijn, M. J. den Broeder, S. Redl, L. J. Kaaij, E. F. Roovers, P. Ladurner, E. Berezikov, and R. F. Ketting. 2010. 'Hen1 is required for oocyte development and piRNA stability in zebrafish', *EMBO J*, 29: 3688-700.

Kaneda, M., M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, and H. Sasaki. 2004. 'Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting', *Nature*, 429: 900-3.

Kato, Y., M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki, and H. Sasaki. 2007a. 'Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse', *Hum Mol Genet*, 16: 2272-80.

Kato, Yuzuru, Masahiro Kaneda, Kenichiro Hata, Kenji Kumaki, Mizue Hisano, Yuji Kohara, Masaki Okano, En Li, Masami Nozaki, and Hiroyuki Sasaki. 2007b. 'Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse', *Human Molecular Genetics*, 16: 2272-80.

Kawaoka, S., N. Izumi, S. Katsuma, and Y. Tomari. 2011. '3' end formation of PIWI-interacting RNAs in vitro', *Mol Cell*, 43: 1015-22.

Kessner, D., M. Chambers, R. Burke, D. Agus, and P. Mallick. 2008. 'ProteoWizard: open source software for rapid proteomics tools development', *Bioinformatics*, 24: 2534-6.

Kettenberger, H., K. J. Armache, and P. Cramer. 2004. 'Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS', *Mol Cell*, 16: 955-65.

Kettenberger, Hubert, Karim-Jean Armache, and Patrick Cramer. 2003. 'Architecture of the RNA Polymerase II-TFIIS Complex and Implications for mRNA Cleavage', *Cell*, 114: 347-57.

Kinomoto, M., T. Kanno, M. Shimura, Y. Ishizaka, A. Kojima, T. Kurata, T. Sata, and K. Tokunaga. 2007. 'All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition', *Nucleic Acids Res*, 35: 2955-64.

Kirino, Y., and Z. Mourelatos. 2007. 'Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini', *Nat Struct Mol Biol*, 14: 347-8.

Kojima-Kita, Kanako, Satomi Kuramochi-Miyagawa, Ippei Nagamori, Narumi Ogonuki, Atsuo Ogura, Hidetoshi Hasuwa, Takashi Akazawa, Norimitsu Inoue, and Toru Nakano. 2016. 'MIWI2 as an Effector of DNA Methylation and Gene Silencing in Embryonic Male Germ Cells', *Cell Reports*, 16: 2819-28.

Kolodziej, P. A., L. Y. Jan, and Y. N. Jan. 1995. 'Mutations that affect the length, fasciculation, or ventral orientation of specific sensory axons in the Drosophila embryo', *Neuron*, 15: 273-86.

Konieczny, Gabriela. 2022. 'Role of C19ORF84H in MIWI2-mediated silencing of retrotransposons in the male germline', University of Edinburgh.

Kosar, M., M. Giannattasio, D. Piccini, A. Maya-Mendoza, F. Garcia-Benitez, J. Bartkova, S. I. Barroso, H. Gaillard, E. Martini, U. Restuccia, M. A. Ramirez-Otero, M. Garre, E. Verga, M. Andujar-Sanchez, S. Maynard, Z. Hodny, V. Costanzo, A. Kumar, A.

Bachi, A. Aguilera, J. Bartek, and M. Foiani. 2021. 'The human nucleoporin Tpr protects cells from RNA-mediated replication stress', *Nat Commun*, 12: 3937.

Kota, S. K., and R. Feil. 2010. 'Epigenetic transitions in germ cell development and meiosis', *Dev Cell*, 19: 675-86.

Krueger, F., and S. R. Andrews. 2011. 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics*, 27: 1571-2.

Krull, Sandra, Julia Dörries, Björn Boysen, Sonja Reidenbach, Lars Magnius, Helene Norder, Johan Thyberg, and Volker C. Cordes. 2010. 'Protein Tpr is required for establishing nuclear pore-associated zones of heterochromatin exclusion', *The EMBO Journal*, 29: 1659-73.

Krull, Sandra, Johan Thyberg, Birgitta Björkroth, Hans-Richard Rackwitz, and Volker C. Cordes. 2004. 'Nucleoporins as Components of the Nuclear Pore Complex Core Structure and Tpr as the Architectural Element of the Nuclear Basket', *Molecular Biology of the Cell*, 15: 4261-77.

Kuramochi-Miyagawa, Satomi, Toshiaki Watanabe, Kengo Gotoh, Yasushi Totoki, Atsushi Toyoda, Masahito Ikawa, Noriko Asada, Kanako Kojima, Yuka Yamaguchi, and Takashi W Ijiri. 2008. 'DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes', *Genes & development*, 22: 908-17.

Kurimoto, K., Y. Yabuta, Y. Ohinata, M. Shigeta, K. Yamanaka, and M. Saitou. 2008. 'Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice', *Genes & development*, 22: 1617-35.

Lachner, M., D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein. 2001. 'Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins', *Nature*, 410: 116-20.

Lane, N., W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter, and W. Reik. 2003. 'Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse', *Genesis*, 35: 88-93.

Langmead, B., and S. L. Salzberg. 2012. 'Fast gapped-read alignment with Bowtie 2', *Nat Methods*, 9: 357-9.

Lara, Nathália L. M., Guilherme M. J. Costa, Gleide F. Avelar, Samyra M. S. N. Lacerda, Rex A. Hess, and Luiz R. de França. 2018. 'Testis Physiology—Overview and Histology.' in Michael K. Skinner (ed.), *Encyclopedia of Reproduction (Second Edition)* (Academic Press: Oxford).

Lau, N. C., A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel, and R. E. Kingston. 2006. 'Characterization of the piRNA complex from rat testes', *Science*, 313: 363-7.

Lawson, K. A., and W. J. Hage. 1994. 'Clonal analysis of the origin of primordial germ cells in the mouse', *Ciba Found Symp*, 182: 68-84; discussion 84-91.

Leblond, C. P., and Y. Clermont. 1952. 'Definition of the stages of the cycle of the seminiferous epithelium in the rat', *Ann N Y Acad Sci*, 55: 548-73.

Lee, E. S., E. J. Wolf, S. S. J. Ihn, H. W. Smith, A. Emili, and A. F. Palazzo. 2020. 'TPR is required for the efficient nuclear export of mRNAs and lncRNAs from short and intron-poor genes', *Nucleic Acids Res*, 48: 11645-63.

Lee, J. H., and D. G. Skalnik. 2012. 'Rbm15-Mkl1 interacts with the Setd1b histone H3-Lys4 methyltransferase via a SPOC domain that is required for cytokine-independent proliferation', *PLOS ONE*, 7: e42965.

Lee, J., K. Inoue, R. Ono, N. Ogonuki, T. Kohda, T. Kaneko-Ishino, A. Ogura, and F. Ishino. 2002. 'Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells', *Development*, 129: 1807-17.

Lee, S. H., H. Sterling, A. Burlingame, and F. McCormick. 2008. 'Tpr directly binds to Mad1 and Mad2 and is important for the Mad1-Mad2-mediated mitotic spindle checkpoint', *Genes & development*, 22: 2926-31.

Lees-Murdock, D. J., C. P. De Felici M Fau - Walsh, and C. P. Walsh. 2003. 'Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage'.

Lenz, S., S. H. Giese, L. Fischer, and J. Rappsilber. 2018. 'In-Search Assignment of Monoisotopic Peaks Improves the Identification of Cross-Linked Peptides', *J Proteome Res*, 17: 3923-31.

Li, E., C. Beard, and R. Jaenisch. 1993. 'Role for DNA methylation in genomic imprinting', *Nature*, 366: 362-5.

Li, En, Timothy H. Bestor, and Rudolf Jaenisch. 1992. 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality'.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.

Li, Heng. 2013. 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', *arXiv*, 1303.3997v2.

Li, J., A. S. Wang, S. Wang, C. Y. Wang, S. Xue, W. Y. Li, T. T. Ma, and Y. X. Shan. 2020. 'Death-inducer obliterator 1 (DIDO1) silencing suppresses growth of bladder cancer cells through decreasing SAPK/JNK signaling cascades', *Neoplasma*, 67: 1074-84.

Li, Jing-Yu, Diane J Lees-Murdock, Guo-Liang Xu, and Colum Patrick Walsh. 2004. 'Timing of establishment of paternal methylation imprints in the mouse', *Genomics*, 84 6: 952-60.

Li, W., Y. Jin, L. Prazak, M. Hammell, and J. Dubnau. 2012. 'Transposable elements in TDP-43-mediated neurodegenerative disorders', *PLOS ONE*, 7: e44099.

Li, Xin Zhiguo, Christian K Roy, Xianjun Dong, Ewelina Bolcun-Filas, Jie Wang, Bo W Han, Jia Xu, Melissa J Moore, John C Schimenti, Zhiping Weng, and Phillip D Zamore. 2013. 'An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes', *Molecular cell*, 50: 67-81.

Li, Y., X. Ma, Y. Wang, and G. Li. 2017. 'miR-489 inhibits proliferation, cell cycle progression and induces apoptosis of glioma cells via targeting SPIN1-mediated PI3K/AKT pathway', *Biomed Pharmacother*, 93: 435-43.

Liang, J., H. Zhao, J. Hu, Y. Liu, and Z. Li. 2018. 'SPOCD1 promotes cell proliferation and inhibits cell apoptosis in human osteosarcoma', *Mol Med Rep*, 17: 3218-25.

Liebschner, D., P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, and P. D. Adams. 2019. 'Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix', *Acta Crystallogr D Struct Biol*, 75: 861-77.

Lim, S. L., Z. P. Qu, R. D. Kortschak, D. M. Lawrence, J. Geoghegan, A. L. Hempfling, M. Bergmann, C. C. Goodnow, C. J. Ormandy, L. Wong, J. Mann, H. S. Scott, D. Jamsai, D. L. Adelson, and M. K. O'Bryan. 2015. 'HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse', *PLoS Genet*, 11: e1005620.

Lin, H., and A. C. Spradling. 1997. 'A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary', *Development*, 124: 2463-76.

Liu, D., Y. Yang, A. Yan, and Y. Yang. 2020. 'SPOCD1 accelerates ovarian cancer progression and inhibits cell apoptosis via the PI3K/AKT pathway', *Onco Targets Ther*, 13: 351-59.

Liu, Q., X. Y. Wang, Y. Y. Qin, X. L. Yan, H. M. Chen, Q. D. Huang, J. K. Chen, and J. M. Zheng. 2018. 'SPOCD1 promotes the proliferation and metastasis of glioma cells by up-regulating PTX3', *Am J Cancer Res*, 8: 624-35.

Liu, S., J. Brind'Amour, M. M. Karimi, K. Shirane, A. Bogutz, L. Lefebvre, H. Sasaki, Y. Shinkai, and M. C. Lorincz. 2014. 'Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells', *Genes & development*, 28: 2041-55.

Liu, W., Q. Yao, X. Su, Y. Deng, M. Yang, B. Peng, F. Zhao, C. Du, X. Zhang, J. Zhu, D. Wang, W. Li, and H. Li. 2023. 'Molecular insights into Spindlin1-HBx interplay and its impact on HBV transcription from cccDNA minichromosome', *Nat Commun*, 14: 4663.

Liu, Y., H. Kim, J. Liang, W. Lu, B. Ouyang, D. Liu, and Z. Songyang. 2014. 'The death-inducer obliterator 1 (Dido1) gene regulates embryonic stem cell self-renewal', *The Journal of biological chemistry*, 289: 4778-86.

Lyko, F. 2018. 'The DNA methyltransferase family: a versatile toolkit for epigenetic regulation', *Nat Rev Genet*, 19: 81-92.

Ma, J. B., Y. R. Yuan, G. Meister, Y. Pei, T. Tuschl, and D. J. Patel. 2005. 'Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein', *Nature*, 434: 666-70.

Maatouk, D. M., L. D. Kellam, M. R. Mann, H. Lei, E. Li, M. S. Bartolomei, and J. L. Resnick. 2006. 'DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages', *Development*, 133: 3411-8.

Martin, Marcel. 2011. 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet. journal*, 17: 10-12.

Matsumura, Y., R. Nakaki, T. Inagaki, A. Yoshida, Y. Kano, H. Kimura, T. Tanaka, S. Tsutsumi, M. Nakao, T. Doi, K. Fukami, T. F. Osborne, T. Kodama, H. Aburatani, and J. Sakai. 2015. 'H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation', *Mol Cell*, 60: 584-96.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. 2007. 'Phaser crystallographic software', *J Appl Crystallogr*, 40: 658-74.

McHugh, C. A., C. K. Chen, A. Chow, C. F. Surka, C. Tran, P. McDonel, A. Pandya-Jones, M. Blanco, C. Burghard, A. Moradian, M. J. Sweredoski, A. A. Shishkin, J. Su, E. S. Lander, S. Hess, K. Plath, and M. Guttman. 2015. 'The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3', *Nature*, 521: 232-6.

McLaren, A., and K. A. Lawson. 2005. 'How is the mouse germ-cell lineage established?', *Differentiation*, 73: 435-7.

McLaren, A., and D. Southee. 1997. 'Entry of mouse embryonic germ cells into meiosis', *Dev Biol*, 187: 107-13.

Melamed, P., Y. Yosefzon, C. David, A. Tsukerman, and L. Pnueli. 2018. 'Tet Enzymes, Variants, and Differential Effects on Function', *Front Cell Dev Biol*, 6: 22.

Mendes, M. L., L. Fischer, Z. A. Chen, M. Barbon, F. J. O'Reilly, S. H. Giese, M. Bohlke-Schneider, A. Belsom, T. Dau, C. W. Combe, M. Graham, M. R. Eisele, W. Baumeister, C. Speck, and J. Rappsilber. 2019. 'An integrated workflow for crosslinking mass spectrometry', *Mol Syst Biol*, 15: e8994.

Messerschmidt, D. M., B. B. Knowles, and D. Solter. 2014. 'DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos', *Genes & development*, 28: 812-28.

Mietz, J. A., Z. Grossman, K. K. Lueders, and E. L. Kuff. 1987. 'Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function', *J Virol*, 61: 3020-9.

Mikami, Suzuka, Teppei Kanaba, Naoki Takizawa, Ayaho Kobayashi, Ryoko Maesaki, Toshinobu Fujiwara, Yutaka Ito, and Masaki Mishima. 2014. 'Structural Insights into the Recruitment of SMRT by the Corepressor SHARP under Phosphorylative Regulation', *Structure*, 22: 35-46.

Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein. 2007. 'Genome-wide maps of chromatin state in pluripotent and lineage-committed cells', *Nature*, 448: 553-60.

Mikkelsen, T. S., M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang, S. Duke, M. Garber, A. J. Gentles, L. Goodstadt, A. Heger, J. Jurka, M. Kamal, E. Mauceli, S. M. Searle, T. Sharpe, M. L. Baker, M. A. Batzer, P. V. Benos, K. Belov, M. Clamp, A. Cook, J. Cuff, R. Das, L. Davidow, J. E. Deakin, M. J. Fazzari, J. L. Glass, M. Grabherr, J. M. Greally, W. Gu, T. A. Hore, G. A. Huttley, M. Kleber, R. L. Jirtle, E. Koina, J. T. Lee, S. Mahony, M. A. Marra, R. D. Miller, R. D. Nicholls, M. Oda, A. T. Papenfuss, Z. E. Parra, D. D. Pollock, D. A. Ray, J. E. Schein, T. P. Speed, K. Thompson, J. L. VandeBerg, C. M. Wade, J. A. Walker, P. D. Waters, C. Webber, J. R. Weidman, X. Xie, M. C. Zody, Platform Broad Institute Genome Sequencing, Team Broad Institute Whole Genome Assembly, J. A. Graves, C. P. Ponting, M. Breen, P. B. Samollow, E. S. Lander, and K. Lindblad-Toh. 2007. 'Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences', *Nature*, 447: 167-77.

Mirdita, M., K. Schutze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. 2022. 'ColabFold: making protein folding accessible to all', *Nat Methods*, 19: 679-82.

Mitchell, P. J., and C. S. Cooper. 1992. 'The human tpr gene encodes a protein of 2094 amino acids that has extensive coiled-coil regions and an acidic C-terminal domain', *Oncogene*, 7: 2329-33.

Molaro, Antoine, Ilaria Falciatori, Emily Hodges, Alexei A. Aravin, Krista Marran, Shahin Rafii, W. Richard McCombie, Andrew D. Smith, and Gregory J. Hannon. 2014. 'Two waves of de novo methylation during mouse germ cell development', *Genes & development*, 28: 1544-49.

Molyneaux, K. A., J. Stallock, K. Schaible, and C. Wylie. 2001. 'Time-lapse analysis of living mouse germ cell migration', *Dev Biol*, 240: 488-98.

Montgomery, T. A., Y. S. Rim, C. Zhang, R. H. Dowen, C. M. Phillips, S. E. Fischer, and G. Ruvkun. 2012. 'PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1', *PLoS Genet*, 8: e1002616.

Morin, P E, D E Awrey, A M Edwards, and C H Arrowsmith. 1996. 'Elongation factor TFIIS contains three structural domains: solution structure of domain II', *Proceedings of the National Academy of Sciences*, 93: 10604-08.

Morse, B., P. G. Rotherg, V. J. South, J. M. Spandorfer, and S. M. Astrin. 1988. 'Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma', *Nature*, 333: 87-90.

Mouse Genome Sequencing, Consortium, R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O.

Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, and E. S. Lander. 2002. 'Initial sequencing and comparative analysis of the mouse genome', *Nature*, 420: 520-62.

Much, Christian, Tania Auchynnikava, Dinko Pavlinic, Andreas Buness, Juri Rappsilber, Vladimir Benes, Robin Allshire, and Dónal O'Carroll. 2016. 'Endogenous Mouse Dicer Is an Exclusively Cytoplasmic Protein', *PLOS Genetics*, 12: e1006095.

Munafò, M., V. R. Lawless, A. Passera, S. MacMillan, S. Bornelöv, I. U. Haussmann, M. Soller, G. J. Hannon, and B. Czech. 2021. 'Channel Nuclear Pore Complex subunits are required for transposon silencing in Drosophila', *Elife*, 10.

Murphy, Caitlin J., and John H. Richburg. 2014. 'Implications of Sertoli cell induced germ cell apoptosis to testicular pathology', *Spermatogenesis*, 4: e979110.

Naas, T. P., R. J. DeBerardinis, J. V. Moran, E. M. Ostertag, S. F. Kingsmore, M. F. Seldin, Y. Hayashizaki, S. L. Martin, and H. H. Kazazian. 1998. 'An actively retrotransposing, novel subfamily of mouse L1 elements', *EMBO J*, 17: 590-7.

Nagamori, I., H. Kobayashi, T. Nishimura, R. Yamagishi, J. Katahira, S. Kuramochi-Miyagawa, T. Kono, and T. Nakano. 2018. 'Relationship between PIWIL4-Mediated H3K4me2 Demethylation and piRNA-Dependent DNA Methylation', *Cell Rep*, 25: 350-56.

Nagamori, Ippei, Hisato Kobayashi, Yusuke Shiromoto, Toru Nishimura, Satomi Kuramochi-Miyagawa, Tomohiro Kono, and Toru Nakano. 2015. 'Comprehensive DNA methylation analysis of retrotransposons in male germ cells', *Cell Reports*, 12: 1541-47.

Nakatsuji, N., and S. Chuma. 2001. 'Differentiation of mouse primordial germ cells into female or male germ cells', *Int J Dev Biol*, 45: 541-8.

Nishimasu, H., H. Ishizu, K. Saito, S. Fukuhara, M. K. Kamatani, L. Bonnefond, N. Matsumoto, T. Nishizawa, K. Nakanaga, J. Aoki, R. Ishitani, H. Siomi, M. C. Siomi,

and O. Nureki. 2012. 'Structure and function of Zucchini endoribonuclease in piRNA biogenesis', *Nature*, 491: 284-7.

Nishimura, T., I. Nagamori, T. Nakatani, N. Izumi, Y. Tomari, S. Kuramochi-Miyagawa, and T. Nakano. 2018. 'PNLDC1, mouse pre-piRNA Trimmer, is required for meiotic and post-meiotic male germ cell development', *EMBO Rep*, 19.

Oakes, C. C., D. J. La Salle S Fau - Smiraglia, B. Smiraglia Dj Fau - Robaire, J. M. Robaire B Fau - Trasler, and J. M. Trasler. 2007. 'Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells'.

Oh, B., S.Y. Hwang, D. Solter, and B.B. Knowles. 1997. 'Spindlin, a major maternal transcript expressed in the mouse during the transition from oocyte to embryo', *Development*, 124: 493-503.

Ohara, T., Y. Sakaguchi, T. Suzuki, H. Ueda, K. Miyauchi, and T. Suzuki. 2007. 'The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated', *Nat Struct Mol Biol*, 14: 349-50.

Ohinata, Yasuhide, Bernhard Payer, Dónal O'Carroll, Katia Ancelin, Yukiko Ono, Mitsue Sano, Sheila C. Barton, Tetyana Obukhanych, Michel Nussenzweig, Alexander Tarakhovsky, Mitinori Saitou, and M. Azim Surani. 2005. 'Blimp1 is a critical determinant of the germ cell lineage in mice', *Nature*, 436: 207-13.

Okano, M., D. A. Bell Dw Fau - Haber, E. Haber Da Fau - Li, and E. Li. 1999. 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development'.

Okano, M., S. Xie, and E. Li. 1998. 'Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases', *Nat Genet*, 19: 219-20.

Olmsted, V. K., D. E. Awrey, C. Koth, X. Shan, P. E. Morin, S. Kazanis, A. M. Edwards, and C. H. Arrowsmith. 1998. 'Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region', *The Journal of biological chemistry*, 273: 22589-94.

Onishi, R., S. Yamanaka, and M. C. Siomi. 2021. 'piRNA- and siRNA-mediated transcriptional repression in Drosophila, mice, and yeast: new insights and biodiversity', *EMBO Rep*, 22: e53062.

Ooi, S. K., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng, and T. H. Bestor. 2007. 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature*, 448: 714-7.

Oswald, F., U. Kostezka, K. Astrahantseff, S. Bourteele, K. Dillinger, U. Zechner, L. Ludwig, M. Wilda, H. Hameister, W. Knochel, S. Liptay, and R. M. Schmid. 2002. 'SHARP

is a novel component of the Notch/RBP-Jkappa signalling pathway', *EMBO J*, 21: 5417-26.

Oswald, F., M. Winkler, Y. Cao, K. Astrahantseff, S. Bourteele, W. Knochel, and T. Borggrefe. 2005. 'RBP-Jkappa/SHARP recruits CtIP/CtBP corepressors to silence Notch target genes', *Molecular and cellular biology*, 25: 10379-90.

Oswald, Franz, Patrick Rodriguez, Benedetto Daniele Giaimo, Zeus A. Antonello, Laura Mira, Gerhard Mittler, Verena N. Thiel, Kelly J. Collins, Nassif Tabaja, Wiebke Cizelsky, Melanie Rothe, Susanne J. Kühl, Michael Kühl, Francesca Ferrante, Kerstin Hein, Rhett A. Kovall, Maria Dominguez, and Tilman Borggrefe. 2016. 'A phospho-dependent mechanism involving NCoR and KMT2D controls a permissive chromatin state at Notch target genes', *Nucleic Acids Research*, 44: 4703-20.

Otani, J., T. Nankumo, K. Arita, S. Inamoto, M. Ariyoshi, and M. Shirakawa. 2009. 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain', *EMBO Rep*, 10: 1235-41.

Ozata, Deniz M., Ildar Gainetdinov, Ansgar Zoch, Dónal O'Carroll, and Phillip D. Zamore. 2018. 'PIWI-interacting RNAs: small RNAs with big functions', *Nature Reviews Genetics*.

Pace, J. K., 2nd, and C. Feschotte. 2007. 'The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage', *Genome Res*, 17: 422-32.

Pandey, R. R., Y. Tokuzawa, Z. Yang, E. Hayashi, T. Ichisaka, S. Kajita, Y. Asano, T. Kunieda, R. Sachidanandam, S. Chuma, S. Yamanaka, and R. S. Pillai. 2013. 'Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice', *Proc Natl Acad Sci U S A*, 110: 16492-7.

Parker, J. S., S. M. Roe, and D. Barford. 2004. 'Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity', *EMBO J*, 23: 4727-37.

Parker, J. S., S. M. Roe, and D. Barford. 2005. 'Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex', *Nature*, 434: 663-6.

Pastor, William A., Hume Stroud, Kevin Nee, Wanlu Liu, Dubravka Pezic, Sergei Manakov, Serena A. Lee, Guillaume Moissiard, Natasha Zamudio, Déborah Bourc'his, Alexei A. Aravin, Amander T. Clark, and Steven E. Jacobsen. 2014. 'MORC1 represses transposable elements in the mouse male germline', *Nature Communications*, 5: 5795.

Pezic, D., S. A. Manakov, R. Sachidanandam, and A. A. Aravin. 2014. 'piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells', *Genes & development*, 28: 1410-28.

Platt, Roy N., Michael W. Vandewege, and David A. Ray. 2018. 'Mammalian transposable elements and their impacts on genome evolution', *Chromosome Research*, 26: 25-43.

Ramirez, F., D. P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dundar, and T. Manke. 2016. 'deepTools2: a next generation web server for deep-sequencing data analysis', *Nucleic Acids Res*, 44: W160-5.

Rappsilber, J., Y. Ishihama, and M. Mann. 2003. 'Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics', *Anal Chem*, 75: 663-70.

Rappsilber, J., M. Mann, and Y. Ishihama. 2007. 'Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips', *Nat Protoc*, 2: 1896-906.

Reilly, M. T., G. J. Faulkner, J. Dubnau, I. Ponomarev, and F. H. Gage. 2013. 'The role of transposable elements in health and diseases of the central nervous system', *J Neurosci*, 33: 17577-86.

Rengaraj, D., S. I. Lee, T. S. Park, H. J. Lee, Y. M. Kim, Y. A. Sohn, M. Jung, S. J. Noh, H. Jung, and J. Y. Han. 2014. 'Small non-coding RNA profiling and the role of piRNA pathway genes in the protection of chicken primordial germ cells', *BMC Genomics*, 15: 757.

Reuter, Michael, Philipp Berninger, Shinichiro Chuma, Hardik Shah, Mihoko Hosokawa, Charlotta Funaya, Claude Antony, Ravi Sachidanandam, and Ramesh S. Pillai. 2011. 'Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing', *Nature*, 480: 264.

Richards, A. L., A. S. Hebert, A. Ulbrich, D. J. Bailey, E. E. Coughlin, M. S. Westphall, and J. J. Coon. 2015. 'One-hour proteome analysis in yeast', *Nat Protoc*, 10: 701-14.

Richardson, S. R., I. Narvaiza, R. A. Planegger, M. D. Weitzman, and J. V. Moran. 2014. 'APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition', *Elife*, 3: e02008.

Ruby, J. G., C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D. P. Bartel. 2006. 'Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans', *Cell*, 127: 1193-207.

Rugg-Gunn, P. J., B. J. Cox, A. Ralston, and J. Rossant. 2010. 'Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo', *Proc Natl Acad Sci U S A*, 107: 10783-90.

Saito, K., K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi, and M. C. Siomi. 2006. 'Specific association of Piwi with rasiRNAs derived from

retrotransposon and heterochromatic regions in the Drosophila genome', *Genes & development*, 20: 2214-22.

Saito, K., Y. Sakaguchi, T. Suzuki, T. Suzuki, H. Siomi, and M. C. Siomi. 2007. 'Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends', *Genes & development*, 21: 1603-8.

Saitou, M., S. C. Barton, and M. A. Surani. 2002. 'A molecular programme for the specification of germ cell fate in mice', *Nature*, 418: 293-300.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides. 2002. 'Active genes are tri-methylated at K4 of histone H3', *Nature*, 419: 407-11.

Sasaki, H., and Y. Matsui. 2008. 'Epigenetic events in mammalian germ-cell development: reprogramming and beyond', *Nat Rev Genet*, 9: 129-40.

Saxe, J. P., M. Chen, H. Zhao, and H. Lin. 2013. 'Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline', *EMBO J*, 32: 1869-85.

Schopp, T., A. Zoch, R. V. Berrens, T. Auchynnikava, Y. Kabayama, L. Vasiliauskaite, J. Rappsilber, R. C. Allshire, and D. O'Carroll. 2020. 'TEX15 is an essential executor of MIWI2-directed transposon DNA methylation and silencing', *Nat Commun*, 11: 3739.

Schumann, G. G. 2007. 'APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition', *Biochem Soc Trans*, 35: 637-42.

Seah, M. K. Y., and D. M. Messerschmidt. 2018. 'From Germline to Soma: Epigenetic Dynamics in the Mouse Preimplantation Embryo', *Curr Top Dev Biol*, 128: 203-35.

Seisenberger, S., Felix Andrews S Fau - Krueger, Julia Krueger F Fau - Arand, Jörn Arand J Fau - Walter, Fátima Walter J Fau - Santos, Christian Santos F Fau - Popp, Bernard Popp C Fau - Thienpont, Wendy Thienpont B Fau - Dean, Wolf Dean W Fau - Reik, and W. Reik. 2012. 'The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells'.

Seki, Y., K. Hayashi, K. Itoh, M. Mizugaki, M. Saitou, and Y. Matsui. 2005. 'Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice', *Dev Biol*, 278: 440-58.

Seki, Y., M. Yamaji, Y. Yabuta, M. Sano, M. Shigeta, Y. Matsui, Y. Saga, M. Tachibana, Y. Shinkai, and M. Saitou. 2007. 'Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice', *Development*, 134: 2627-38.

Sharif, J., M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, and H.

Koseki. 2007. 'The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA', *Nature*, 450: 908-12.

Shen, X., Y. Liu, Y. J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G. C. Yuan, and S. H. Orkin. 2008. 'EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency', *Mol Cell*, 32: 491-502.

Shi, Y., M. Downes, W. Xie, H. Y. Kao, P. Ordentlich, C. C. Tsai, M. Hon, and R. M. Evans. 2001. 'Sharp, an inducible cofactor that integrates nuclear receptor repression and activation', *Genes & development*, 15: 1140-51.

Shoji, M., T. Tanaka, M. Hosokawa, M. Reuter, A. Stark, Y. Kato, G. Kondoh, K. Okawa, T. Chujo, T. Suzuki, K. Hata, S. L. Martin, T. Noce, S. Kuramochi-Miyagawa, T. Nakano, H. Sasaki, R. S. Pillai, N. Nakatsuji, and S. Chuma. 2009. 'The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline', *Dev Cell*, 17: 775-87.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. 2011. 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol Syst Biol*, 7: 539.

Singh, P., A. X. Li, D. A. Tran, N. Oates, E. R. Kang, X. Wu, and P. E. Szabo. 2013. 'De novo DNA methylation in the male germ line occurs by default but is excluded at sites of H3K4 methylation', *Cell Rep*, 4: 205-19.

Siomi, Mikiko C, Kaoru Sato, Dubravka Pezic, and Alexei A Aravin. 2011. 'PIWI-interacting small RNAs: the vanguard of genome defence', *Nature reviews Molecular cell biology*, 12: 246.

Slotkin, R. K., and R. Martienssen. 2007. 'Transposable elements and the epigenetic regulation of the genome', *Nat Rev Genet*, 8: 272-85.

Song, Q., Q. Ji, J. Xiao, F. Li, L. Wang, Y. Chen, Y. Xu, and S. Jiao. 2018. 'miR-409 Inhibits Human Non-Small-Cell Lung Cancer Progression by Directly Targeting SPIN1', *Mol Ther Nucleic Acids*, 13: 154-63.

Staub, E., D. Mennerich, and A. Rosenthal. 2002. 'The Spin/Ssty repeat: a new motif identified in proteins involved in vertebrate development from gamete to embryo', *Genome Biol*, 3: RESEARCH0003.

Stempor, P., and J. Ahringer. 2016. 'SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics', *Wellcome Open Res*, 1: 14.

Stols, L., M. Gu, L. Dieckman, R. Raffen, F. R. Collart, and M. I. Donnelly. 2002. 'A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site', *Protein Expr Purif*, 25: 8-15.

Strome, S., and D. Updike. 2015. 'Specifying and protecting germ cell fate', *Nat Rev Mol Cell Biol*, 16: 406-16.

Su, Xiaonan, Guixin Zhu, Xiaozhe Ding, Shirley Y. Lee, Yali Dou, Bing Zhu, Wei Wu, and Haitao Li. 2014. 'Molecular basis underlying histone H3 lysine–arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1', *Genes & development*, 28: 622-36.

Sun, Min, Zhi Li, and Jian-Fang Gui. 2010. 'Dynamic distribution of spindlin in nucleoli, nucleoplasm and spindle from primary oocytes to mature eggs and its critical function for oocyte-to-embryo transition in gibel carp', *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 313A: 461-73.

Sun, Yu H., Brent Lee, and Xin Zhiguo Li. 2022. 'The birth of piRNAs: how mammalian piRNAs are produced, originated, and evolved', *Mammalian Genome*, 33: 293-311.

Surani, M Azim, Katsuhiko Hayashi, and Petra Hajkova. 2007. 'Genetic and epigenetic regulators of pluripotency', *Cell*, 128: 747-62.

Surani, M. A., and P. Hajkova. 2010. 'Epigenetic reprogramming of mouse germ cells toward totipotency', *Cold Spring Harb Symp Quant Biol*, 75: 211-8.

Tachibana, M., Y. Matsumura, M. Fukuda, H. Kimura, and Y. Shinkai. 2008. 'G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription', *EMBO J*, 27: 2681-90.

Tachibana, M., J. Ueda, M. Fukuda, N. Takeda, T. Ohta, H. Iwanari, T. Sakihama, T. Kodama, T. Hamakubo, and Y. Shinkai. 2005. 'Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9', *Genes & development*, 19: 815-26.

Takamatsu, G., K. Yanagi, K. Koganebuchi, F. Yoshida, J. S. Lee, K. Toyama, K. Hattori, C. Katagiri, T. Kondo, H. Kunugi, R. Kimura, T. Kaname, and M. Matsushita. 2022. 'Haplotype phasing of a bipolar disorder pedigree revealed rare multiple mutations of SPOCD1 gene in the 1p36-35 susceptibility locus', *J Affect Disord*, 310: 96-105.

The UniProt, Consortium. 2017. 'UniProt: the universal protein knowledgebase', *Nucleic Acids Res*, 45: D158-D69.

Thul, P. J., L. Akesson, M. Wiking, D. Mahdessian, A. Geladaki, H. A. Blal, T. Alm, A. Asplund, L. Bjork, L. M. Breckels, A. Backstrom, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C. M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J. M. Schwenk, A. Sivertsson, E. Sjostedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Ponten, K. von Feilitzen, K. S. Lilley, M. Uhlen, and E. Lundberg. 2017. 'A subcellular map of the human proteome', *Science*, 356.

Tian, Y., D. K. Simanshu, J. B. Ma, and D. J. Patel. 2011. 'Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains', *Proc Natl Acad Sci U S A*, 108: 903-10.

Tunyasuvunakool, K., J. Adler, Z. Wu, T. Green, M. Zielinski, A. Zidek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, and D. Hassabis. 2021. 'Highly accurate protein structure prediction for the human proteome', *Nature*, 596: 590-96.

Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, and J. Cox. 2016. 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nat Methods*, 13: 731-40.

Uneme, Yuta, Ryu Maeda, Gen Nakayama, Haruka Narita, Naoki Takeda, Ryuji Hiramatsu, Hidenori Nishihara, Ryuichiro Nakato, Yoshiakira Kanai, Kimi Araki, Mikiko C. Siomi, and Soichiro Yamanaka. 2023. 'Morc1 re-establishes heterochromatin on activated transposons and shapes the host transcriptome in gonocytes', *bioRxiv*: 2023.06.26.546627.

Uranishi, H., A. S. Zolotukhin, S. Lindtner, S. Warming, G. M. Zhang, J. Bear, N. G. Copeland, N. A. Jenkins, G. N. Pavlakis, and B. K. Felber. 2009. 'The RNA-binding motif protein 15B (RBM15B/OTT3) acts as cofactor of the nuclear export receptor NXF1', *The Journal of biological chemistry*, 284: 26106-16.

Vagin, V. V., A. Sigova, C. Li, H. Seitz, V. Gvozdev, and P. D. Zamore. 2006. 'A distinct small RNA pathway silences selfish genetic elements in the germline', *Science*, 313: 320-4.

Vallee, M., C. Robert, S. Methot, M. F. Palin, and M. A. Sirard. 2006. 'Cross-species hybridizations on a multi-species cDNA microarray to identify evolutionarily conserved genes expressed in oocytes', *BMC Genomics*, 7: 113.

van der Heijden, A. G., L. Mengual, J. J. Lozano, M. Ingelmo-Torres, M. J. Ribal, P. L. Fernandez, E. Oosterwijk, J. A. Schalken, A. Alcaraz, and J. A. Witjes. 2016. 'A five-gene expression signature to predict progression in T1G3 bladder cancer', *Eur J Cancer*, 64: 127-36.

Vasiliauskaitė, Lina, Rebecca V. Berrens, Ivayla Ivanova, Claudia Carrieri, Wolf Reik, Anton J. Enright, and Dónal O'Carroll. 2018. 'Defective germline reprogramming rewires the spermatogonial transcriptome', *Nature Structural & Molecular Biology*, 25: 394-404.

Vasiliauskaitė, Lina, Dimitrios Vitsios, Rebecca V. Berrens, Claudia Carrieri, Wolf Reik, Anton J. Enright, and Dónal O'Carroll. 2017. 'A MILI-independent piRNA biogenesis pathway empowers partial germline reprogramming', *Nature Structural &Amp; Molecular Biology*, 24: 604.

Vincent, J. J., Y. Huang, P. Y. Chen, S. Feng, J. H. Calvopina, K. Nee, S. A. Lee, T. Le, A. J. Yoon, K. Faull, G. Fan, A. Rao, S. E. Jacobsen, M. Pellegrini, and A. T. Clark. 2013. 'Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells', *Cell Stem Cell*, 12: 470-8.

Vonrhein, C., C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack, and G. Bricogne. 2011. 'Data processing and analysis with the autoPROC toolbox', *Acta Crystallogr D Biol Crystallogr*, 67: 293-302.

Vourekas, A., K. Zheng, Q. Fu, M. Maragkakis, P. Alexiou, J. Ma, R. S. Pillai, Z. Mourelatos, and P. J. Wang. 2015. 'The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing', *Genes & development*, 29: 617-29.

Vourekas, A., Q. Zheng, P. Alexiou, M. Maragkakis, Y. Kirino, B. D. Gregory, and Z. Mourelatos. 2012. 'Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis', *Nat Struct Mol Biol*, 19: 773-81.

Walsh, Colum P, J Richard Chaillet, and Timothy H Bestor. 1998. 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nature genetics*, 20: 116.

Wang, C., L. Zhan, M. Wu, R. Ma, J. Yao, Y. Xiong, Y. Pan, S. Guan, X. Zhang, and J. Zang. 2018. 'Spindlin-1 recognizes methylations of K20 and R23 of histone H4 tail'.

Wang, Haoyi, Hui Yang, Chikdu S Shivalila, Meelad M Dawlaty, Albert W Cheng, Feng Zhang, and Rudolf Jaenisch. 2013. 'One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering', *Cell*, 153: 910-18.

Wang, J. X., Q. Zeng, L. Chen, J. C. Du, X. L. Yan, H. F. Yuan, C. Zhai, J. N. Zhou, Y. L. Jia, W. Yue, and X. T. Pei. 2012. 'SPINDLIN1 promotes cancer cell proliferation through activation of WNT/TCF-4 signaling', *Mol Cancer Res*, 10: 326-35.

Wang, Weixiang, Zhi Chen, Zhuo Mao, Huihui Zhang, Xiaojun Ding, She Chen, Xiaodong Zhang, Ruiming Xu, and Bing Zhu. 2011. 'Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes', *EMBO reports*, 12: 1160-66.

Wang, X., C. Lv, Y. Guo, and S. Yuan. 2020. 'Mitochondria Associated Germinal Structures in Spermatogenesis: piRNA Pathway Regulation and Beyond', *Cells*, 9.

Watanabe, T., X. Cui, Z. Yuan, H. Qi, and H. Lin. 2018. 'MIWI2 targets RNAs transcribed from piRNA-dependent regions to drive DNA methylation in mouse prospermatogonia', *EMBO J*, 37.

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp, and G. J. Barton. 2009. 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25: 1189-91.

Watson, M. L., A. R. Zinn, N. Inoue, K. D. Hess, J. Cobb, M. A. Handel, R. Halaban, C. C. Duchene, G. M. Albright, and R. W. Moreadith. 1998. 'Identification of morc (microrchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse', *Proc Natl Acad Sci U S A*, 95: 14361-6.

Webster, Kylie E, Moira K O'Bryan, Stephen Fletcher, Pauline E Crewther, Ulla Aapola, Jeff Craig, Dion K Harrison, Hnin Aung, Nawapen Phutikanit, and Robert Lyle. 2005. 'Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis', *Proceedings of the National Academy of Sciences*, 102: 4068-73.

Wenda, J. M., D. Homolka, Z. Yang, P. Spinelli, R. Sachidanandam, R. R. Pandey, and R. S. Pillai. 2017. 'Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis and Function', *Dev Cell*, 41: 623-37 e9.

Western, P. S., D. C. Miles, J. A. van den Bergen, M. Burton, and A. H. Sinclair. 2008. 'Dynamic regulation of mitotic arrest in fetal male germ cells', *Stem Cells*, 26: 339-47.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, and A. H. Schulman. 2007. 'A unified classification system for eukaryotic transposable elements', *Nat Rev Genet*, 8: 973-82.

Wickham, Hadley. 2019. 'Welcome to the Tidyverse.', *Journal of Open Source Software*, 1686.

Wind, M., and D. Reines. 2000. 'Transcription elongation factor SII', *Bioessays*, 22: 327-36.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. 2011. 'Overview of the CCP4 suite and current developments', *Acta Crystallogr D Biol Crystallogr*, 67: 235-42.

Wolf, G., P. Yang, A. C. Fuchtbauer, E. M. Fuchtbauer, A. M. Silva, C. Park, W. Wu, A. L. Nielsen, F. S. Pedersen, and T. S. Macfarlan. 2015. 'The KRAB zinc finger protein

ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses', *Genes & development*, 29: 538-54.

Wu, S. C., and Y. Zhang. 2010. 'Active DNA demethylation: many roads lead to Rome', *Nat Rev Mol Cell Biol*, 11: 607-20.

Wylie, C. 1999. 'Germ cells', *Cell*, 96: 165-74.

Xiao, J., R. Zhang, J. Peng, and Z. Yang. 2020. 'BAP1 maintains chromosome stability by stabilizing DIDO1 in renal cell carcinoma', *Am J Cancer Res*, 10: 1455-66.

Yabuta, Y., K. Kurimoto, Y. Ohinata, Y. Seki, and M. Saitou. 2006. 'Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling', *Biol Reprod*, 75: 705-16.

Yamaguchi, S., L. Shen, Y. Liu, D. Sendler, and Y. Zhang. 2013. 'Role of Tet1 in erasure of genomic imprinting', *Nature*, 504: 460-4.

Yamaji, M., Y. Seki, K. Kurimoto, Y. Yabuta, M. Yuasa, M. Shigeta, K. Yamanaka, Y. Ohinata, and M. Saitou. 2008. 'Critical function of Prdm14 for the establishment of the germ cell lineage in mice', *Nat Genet*, 40: 1016-22.

Yamanaka, Soichiro, Hidenori Nishihara, Hidehiro Toh, Luis Augusto Eijy Nagai, Kosuke Hashimoto, Sung-Joon Park, Aoi Shibuya, Ana Maria Suzuki, Yujiro Tanaka, Kenta Nakai, Piero Carninci, Hiroyuki Sasaki, and Haruhiko Siomi. 2019. 'Broad Heterochromatic Domains Open in Gonocyte Development Prior to De Novo DNA Methylation', *Developmental Cell*, 51: 21-34.e5.

Yang, F., Y. Lan, R. R. Pandey, D. Homolka, S. L. Berger, R. S. Pillai, M. S. Bartolomei, and P. J. Wang. 2020. 'TEX15 associates with MILI and silences transposable elements in male germ cells', *Genes & development*, 34: 745-50.

Yang, Hui, Haoyi Wang, Chikdu S Shivalila, Albert W Cheng, Linyu Shi, and Rudolf Jaenisch. 2013. 'One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering', *Cell*, 154: 1370-79.

Yang, Na, Weixiang Wang, Yan Wang, Mingzhu Wang, Qiang Zhao, Zihe Rao, Bing Zhu, and Rui-Ming Xu. 2012. 'Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1', *Proceedings of the National Academy of Sciences*, 109: 17954-59.

Yang, Zhaolin, Kuan-Ming Chen, Radha Raman Pandey, David Homolka, Michael Reuter, Bruno Kotska Rodino Janeiro, Ravi Sachidanandam, Marie-Odile Fauvarque, Andrew A McCarthy, and Ramesh S Pillai. 2016. 'PIWI slicing and EXD1 drive biogenesis of nuclear piRNAs from cytosolic targets of the mouse piRNA pathway', *Molecular cell*, 61: 138-52.

Yariv, B., E. Yariv, A. Kessel, G. Masrati, A. B. Chorin, E. Martz, I. Mayrose, T. Pupko, and N. Ben-Tal. 2023. 'Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf', *Protein Sci*, 32: e4582.

Ye, Shi-Jun, Li Ying, Sushmita Ghosh, Luiz Renato De França, and Lonnie D. Russell. 1993. 'Sertoli cell cycle: A re-examination of the structural changes during the cycle of the seminiferous epithelium of the rat', *The Anatomical Record*, 237: 187-98.

Yoon, M. K., D. M. Mitrea, L. Ou, and R. W. Kriwacki. 2012. 'Cell cycle regulation by the intrinsically disordered proteins p21 and p27', *Biochem Soc Trans*, 40: 981-8.

Yuan, H., P. Zhang, L. Qin, L. Chen, S. Shi, Y. Lu, F. Yan, C. Bai, X. Nan, D. Liu, Y. Li, W. Yue, and X. Pei. 2008. 'Overexpression of SPINDLIN1 induces cellular senescence, multinucleation and apoptosis', *Gene*, 410: 67-74.

Zhang, L., N. T. Tran, H. Su, R. Wang, Y. Lu, H. Tang, S. Aoyagi, A. Guo, A. Khodadadi-Jamayran, D. Zhou, K. Qian, T. Hricik, J. Cote, X. Han, W. Zhou, S. Laha, O. Abdel-Wahab, R. L. Levine, G. Raffel, Y. Liu, D. Chen, H. Li, T. Townes, H. Wang, H. Deng, Y. G. Zheng, C. Leslie, M. Luo, and X. Zhao. 2015. 'Cross-talk between PRMT1-mediated methylation and ubiquitylation on RBM15 controls RNA splicing', *Elife*, 4.

Zhang, P., B. Cong, H. Yuan, L. Chen, Y. Lv, C. Bai, X. Nan, S. Shi, W. Yue, and X. Pei. 2008. 'Overexpression of spindlin1 induces metaphase arrest and chromosomal instability', *J Cell Physiol*, 217: 400-8.

Zhang, Yinglu, Katarzyna Rataj, Gordon G. Simpson, and Liang Tong. 2016. 'Crystal Structure of the SPOC Domain of the Arabidopsis Flowering Regulator FPA', *PLOS ONE*, 11: e0160694.

Zhao, Fan, Yunan Liu, Xiaonan Su, Ji-Eun Lee, Yutong Song, Daliang Wang, Kai Ge, Juntao Gao, Michael Q. Zhang, and Haitao Li. 2020. 'Molecular basis for histone H3 “K4me3-K9me3/2” methylation pattern readout by Spindlin1', *Journal of Biological Chemistry*, 295: 16877-87.

Zhao, Qiang, Lipeng Qin, Fuguo Jiang, Beili Wu, Wen Yue, Feng Xu, Zhili Rong, Hongfeng Yuan, Xiaoyan Xie, Yanhong Gao, Cixian Bai, Mark Bartlam, Xuetao Pei, and Zihe Rao. 2007. 'Structure of Human Spindlin1: TANDEM TUDOR-LIKE DOMAINS FOR CELL CYCLE REGULATION', *Journal of Biological Chemistry*, 282: 647-56.

Zhao, T., T. Gu, H. C. Rice, K. L. McAdams, K. M. Roark, K. Lawson, S. A. Gauthier, K. L. Reagan, and R. S. Hewes. 2008. 'A Drosophila gain-of-function screen for candidate genes involved in steroid-dependent neuroendocrine cell remodeling', *Genetics*, 178: 883-901.

Zhou, D., F. Zhu, Z. H. Huang, H. Zhang, L. Q. Fan, and J. Y. Fan. 2022. 'SPOC domain-containing protein 1 regulates the proliferation and apoptosis of human

spermatogonial stem cells through adenylate kinase 4', *World J Stem Cells*, 14: 822-38.

Zhu, M., C. Yan, C. Ren, X. Huang, X. Zhu, H. Gu, M. Wang, S. Wang, Y. Gao, Y. Ji, X. Miao, M. Yang, J. Chen, J. Du, T. Huang, Y. Jiang, J. Dai, H. Ma, J. Zhou, Z. Wang, Z. Hu, G. Ji, Z. Zhang, H. Shen, Y. Shi, and G. Jin. 2017. 'Exome Array Analysis Identifies Variants in SPOCD1 and BTN3A2 That Affect Risk for Gastric Cancer', *Gastroenterology*, 152: 2011-21.

Zoch, A., T. Auchynnikava, R. V. Berrens, Y. Kabayama, T. Schopp, M. Heep, L. Vasiliauskaite, Y. A. Perez-Rico, A. G. Cook, A. Shkumatava, J. Rappsilber, R. C. Allshire, and D. O'Carroll. 2020. 'SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation', *Nature*, 584: 635-39.