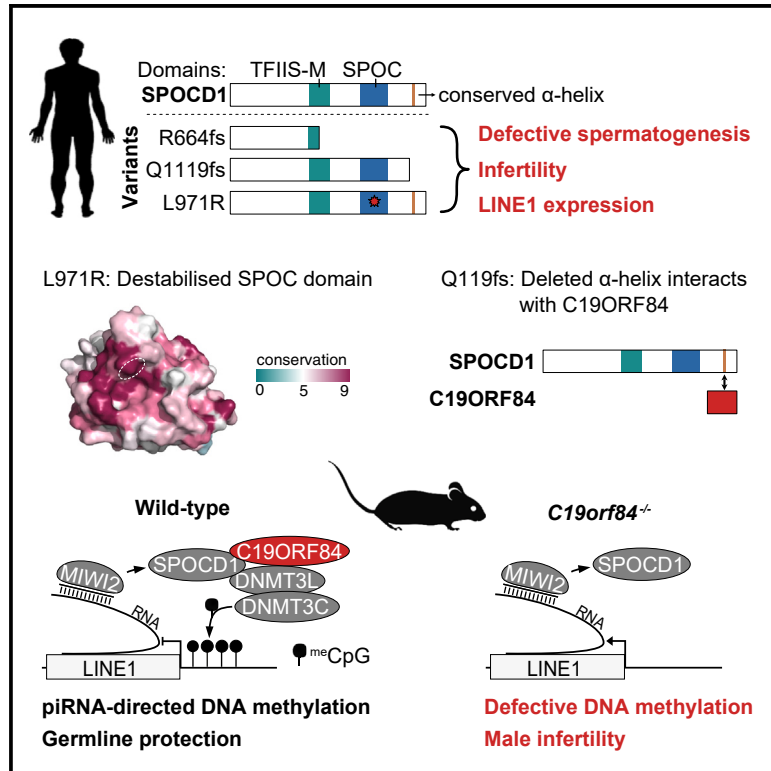


C19ORF84 connects piRNA and DNA methylation machineries to defend the mammalian germ line

Graphical abstract



Authors

Ansgar Zoch, Gabriela Konieczny, Tania Auchynnika, ..., Atlanta G. Cook, Frank Tüttelmann, Dónal O'Carroll

Correspondence

ansgar.zoch@ed.ac.uk (A.Z.), donal.ocarroll@ed.ac.uk (D.O.)

In brief

Zoch et al. revealed a functional conservation between the mouse and human nuclear piRNA pathways. Analysis of human *SPOCD1* loss-of-function variants led to the discovery of C19ORF84, an uncharacterized protein that connects the piRNA apparatus to the DNA methylation machinery *in vivo* ensuring the integrity of the germ line.

Highlights

- SPOCD1 is required for human male fertility and transposon silencing
- SPOCD1 associates with the uncharacterized protein C19ORF84
- C19ORF84 is essential for piRNA-directed transposon methylation in mice
- C19ORF84 bridges the piRNA and *de novo* methylation machineries *in vivo*



Article

C19ORF84 connects piRNA and DNA methylation machineries to defend the mammalian germ line

Ansgar Zoch,^{1,2,11,*} Gabriela Konieczny,^{1,2,11} Tania Auchynnika,^{2,11} Birgit Stallmeyer,³ Nadja Rotte,³ Madeleine Heep,^{1,2} Rebecca V. Berrens,⁴ Martina Schito,^{1,2} Yuka Kabayama,^{1,2} Theresa Schöpp,^{1,2} Sabine Kliesch,⁵ Brendan Houston,⁶ Liina Nagimaja,⁷ Moira K. O'Bryan,⁶ Kenneth I. Aston,⁸ Donald F. Conrad,^{7,9} Juri Rappsilber,^{2,10} Robin C. Allshire,² Atlanta G. Cook,² Frank Tüttelmann,³ and Dónal O'Carroll^{1,2,12,*}

¹Centre for Regenerative Medicine, Institute for Regeneration and Repair, Institute for Stem Cell Research, University of Edinburgh, 5 Little France Drive, Edinburgh EH16 4UU, UK

²Wellcome Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Max Born Crescent, Edinburgh EH9 3BF, UK

³Institute of Reproductive Genetics, University of Münster, Münster, Germany

⁴Institute for Developmental and Regenerative Medicine, University of Oxford, IMS-Tetsuya Nakamura Building, Old Road Campus, Roosevelt Drive, Oxford OX37TY, UK

⁵Centre of Reproductive Medicine and Andrology, Department of Clinical and Surgical Andrology, University Hospital Münster, Münster, Germany

⁶School of BioSciences and Bio21 Institute, Faculty of Science, The University of Melbourne, Parkville, VIC, Australia

⁷Division of Genetics, Oregon National Primate Research Center, Oregon Health and Science University, Beaverton, OR, USA

⁸Andrology and In Vitro Fertilization Laboratory, Department of Surgery (Urology), University of Utah School of Medicine, Salt Lake City, UT, USA

⁹Center for Embryonic Cell and Gene Therapy, Oregon Health and Science University, Portland, OR, USA

¹⁰Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, Gustav-Meyer-Allee 25, 13355 Berlin, Germany

¹¹These authors contributed equally

¹²Lead contact

*Correspondence: ansgar.zoch@ed.ac.uk (A.Z.), donal.ocarroll@ed.ac.uk (D.O.)

<https://doi.org/10.1016/j.molcel.2024.01.014>

SUMMARY

In the male mouse germ line, PIWI-interacting RNAs (piRNAs), bound by the PIWI protein MIWI2 (PIWIL4), guide DNA methylation of young active transposons through SPOCD1. However, the underlying mechanisms of SPOCD1-mediated piRNA-directed transposon methylation and whether this pathway functions to protect the human germ line remain unknown. We identified loss-of-function variants in human *SPOCD1* that cause defective transposon silencing and male infertility. Through the analysis of these pathogenic alleles, we discovered that the uncharacterized protein C19ORF84 interacts with SPOCD1. DNMT3C, the DNA methyltransferase responsible for transposon methylation, associates with SPOCD1 and C19ORF84 in fetal gonocytes. Furthermore, C19ORF84 is essential for piRNA-directed DNA methylation and male mouse fertility. Finally, C19ORF84 mediates the *in vivo* association of SPOCD1 with the *de novo* methylation machinery. In summary, we have discovered a conserved role for the human piRNA pathway in transposon silencing and C19ORF84, an uncharacterized protein essential for orchestrating piRNA-directed DNA methylation.

INTRODUCTION

Transposons pose a major threat to the integrity and continuity of the germ line. In mammals, transposon expression is restrained by DNA methylation.¹ However, in the developing germ line the genome undergoes demethylation followed by *de novo* methylation.² In the male mouse germ line, the threat posed by young, active transposons during the period of genome hypomethylation is neutralized by RNA-based silencing mechanisms.³ Through base-complementarity, piRNAs guide the PIWI protein MILI (PIWIL2) to destroy cytoplasmic transposon mRNAs and tether the nuclear PIWI protein MIWI2 (PIWIL4) to nascent trans-

poson transcripts.^{3,4} This tethering initiates gene silencing through the recruitment of factors to the nascent RNA and culminates in DNA methylation. Recently, proteins associated with MIWI2 in fetal gonocytes undergoing *de novo* genome methylation have been defined.^{5,6} Among these, TEX15 and SPOCD1 have been showed to be essential for piRNA-directed transposon methylation.^{5,6} Furthermore, SPOCD1 links MIWI2 to the *de novo* DNA methylation machinery.⁵ However, the underlying mechanisms of SPOCD1 in this process are not fully understood.

The *de novo* DNA methylation machinery is composed of the adaptor protein DNMT3L and the DNA methyltransferases



DNMT3A, DNMT3B, and DNMT3C.² DNMT3L interacts with and stimulates the *de novo* DNA methyltransferases and is required in the developing germ line for the methylation of the entire genome.^{7–10} The division of labor between DNMT3A, DNMT3B, and DNMT3C in the *de novo* genome methylation of fetal male germ cells has recently been defined. DNMT3A methylates the vast majority of the genome except for the promoter elements of young long interspersed nuclear element 1 (LINE1) and endogenous retrovirus (ERV) intra-cisternal A-type particle (IAP) transposons, which is the task of DNMT3C.^{11–13} Genetic studies have shown that DNMT3C, MIWI2, and SPOCD1 share the same target loci, and it is acknowledged that DNMT3C is the methyltransferase downstream of the piRNA pathway.^{2,5,11} Despite this central role in piRNA-directed DNA methylation DNMT3C remains enigmatic; visualization of the protein or a physical link to the piRNA pathway in fetal germ cells has not been achieved.

Homozygosity of loss-of-function variants in human piRNA pathway genes has been compellingly associated with male infertility.^{14–24} However, whether the genetic inactivation of the human pathway leads to transposon deregulation remains unknown. The murine piRNA pathway utilizes both the cytoplasmic RNAi-like mechanism and DNA methylation to silence transposons.³ One contrasting feature between the human and mouse pathways is that the human genome lacks a *Dnmt3c* orthologue.^{11,12} Indeed, DNMT3C is rodent-specific and derived from duplication of the *Dnmt3b* gene.^{11,12,25} This fact brings into question the role of the nuclear branch of the piRNA pathway in mediating transposons silencing in the human germ line.² Thus, a role for the human piRNA pathway in transposon silencing remains undetermined.

RESULTS

SPOCD1 is required for human male fertility and transposon silencing

To explore a possible role for SPOCD1 in human male fertility, we screened exome sequencing data from the Male Reproductive Genomics (MERGE) and Genetics of Male Infertility Initiative (GEMINI) studies. We identified three infertile men with non-obstructive azoospermia and elevated follicle-stimulating hormone (FSH) levels indicative of spermatogenic failure that carried homozygous variants in the *SPOCD1* gene (Figures 1A and 1B), which are absent in gnomAD.²⁶ Individual M3457 comes from a consanguineous family and carries the homozygous missense variant *SPOCD1* NM_144569.6:c.2912T>G (p.Leu971Arg) that results in substitution of leucine 971 to an arginine in the SPEN ortholog and paralog C-terminal (SPOC) domain of SPOCD1 (Figures 1B, 1C, and S1A). No testicular tissue of this patient was available. The individual GEMINI-88 was also from a consanguineous family, and homozygous for the *SPOCD1* NM_144569.6:c.3354_3355insA (*SPOCD1*^{c3354_3355insA}) allele, characterized by a single base insertion in the last exon. This insertion results in a frameshift variant that encodes the SPOCD1 p.Gln1119ThrfsTer66 (*SPOCD1*-Q1119fs) mutant protein, in which the carboxy-terminal 97 amino acids are deleted (Figures 1B, 1D, and S1B). Seminiferous tubules in the testes of Gemini-88 individual presented as Sertoli cell only lacking all germ cells

(Figure 1E), a known consequence of defective transposon methylation in mice.^{4,8,27,28} Individual M2021 carried an allele that encoded a frame-shift variant due to a two base pair deletion, *SPOCD1* NM_144569.6:c.1991_1992del (*SPOCD1*^{c1991_1992del}), in exon 8 (Figures 1B, 1F, and S1C). Further sequencing revealed that the parents and a fertile brother of individual M2021 were heterozygous carriers of the *SPOCD1*^{c1991_1992del} allele (Figure 1F). Histological analyses of testis biopsies from individual M2021 revealed aberrant seminiferous tubules that either had a meiotic arrest or were devoid of germ cells (Sertoli cell only) (Figure 1G). In rare tubules, we also observed the presence of individual round spermatids (Figure 1G). Furthermore, LINE1 ORF1p expression was detected in spermatogonia from individual M2021 but not from a healthy individual (Figure 1H). In summary, SPOCD1 is essential for human male fertility and transposon silencing.

Structural integrity of the SPOC domain is required for SPOCD1 function

These pathogenic alleles associated with human male infertility could provide insight into the molecular function of SPOCD1. *SPOCD1*^{c1991_1992del} is likely a null allele given that our *Spocd1*⁻ mouse allele contains a similar missense mutation in exon 7,⁵ which is syntenic to human exon 8 (Figures 1B and S1C). Should the transcript escape nonsense-mediated decay,²⁹ this would lead to a severely truncated SPOCD1 p.Arg664GlnfsTer57 (*SPOCD1*-R664fs) protein (Figures 1B and S1C). The SPOCD1 c.2912T>G allele results in a Leu971Arg substitution in the SPOC domain (Figures 1B and S1A). This residue is highly conserved within the SPOCD1 SPOC domain (Figure 2A). In fact, it is also conserved with the related PHF3 and DIDO1 SPOC domains, but not in the more distant SPOC domains of SPEN and RBM15 (Figure S2A). To understand the contribution of Leu971 to the SPOC domain, we solved the crystal structure of the mouse SPOCD1 SPOC domain at 1.7 Å resolution (Figure 2B; Table 1). SPOC domains are distorted β-barrels usually composed of 7 β strands and a variable number of surrounding α helices.³⁰ The β-barrel of the SPOCD1 SPOC domain is composed of 8 β strands surrounded by 5 α helices (Figure 2B). The SPOC domain of PHF3 interacts with phosphorylated serine 2 of the RNA polymerase II (RNAPII) carboxy-terminal domain (CTD).³¹ The two phospho-serine residues are bound by two highly positively charged patches on PHF3³¹ (Figure 2C). The SPOC domain of SPOCD1 does not interact with the RNAPII CTD,³² nor does SPOCD1 associate with RNAPII in fetal gonocytes.⁵ The structure of the mouse SPOCD1 SPOC domain revealed two smaller, closely spaced, positively charged patches and notably a highly negatively charged patch (Figure 2B), which would be incompatible with phospho-peptide binding in the same manner as PHF3. Furthermore, the conservation analysis showed that the SPOCD1 SPOC domain has conserved residues clustered on a different surface (Figures 2B and 2D), consistent with a different function to PHF3. Leucine 971 in human SPOCD1 corresponds to Leucine 792 in the mouse. In our crystal structure of the mouse SPOCD1 SPOC domain, Leucine 792 is located at the end of the seventh β strand, and the side chain contributes to the hydrophobic core of this region of the

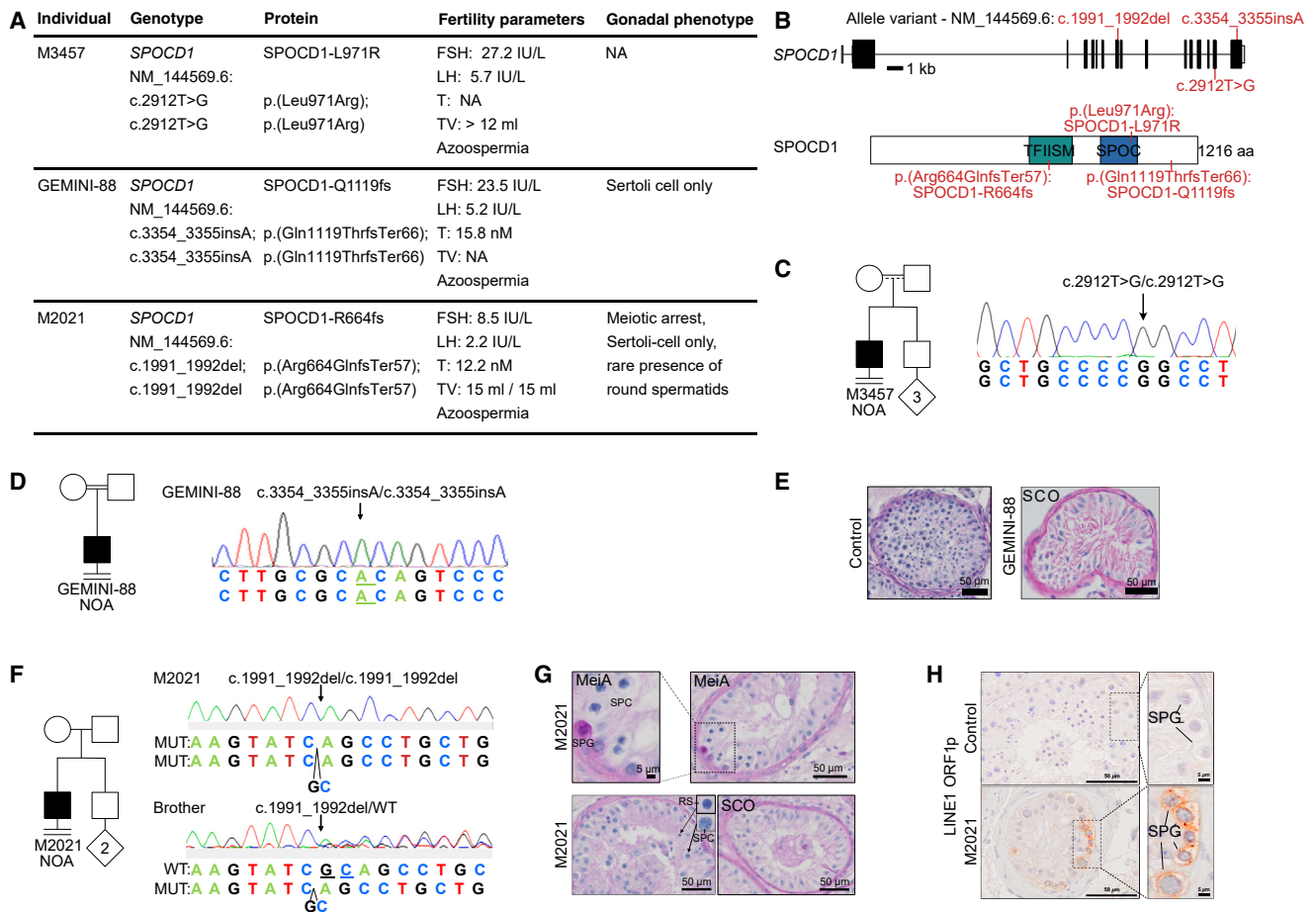


Figure 1. SPOCD1 is essential for human fertility and transposon silencing

(A) Table of clinical data for three identified infertile men carrying homozygous *SPOCD1* variants. FSH, follicle-stimulating hormone; LH, luteinizing hormone; T, testosterone; TV, testicular volume right/left; IU, international unit; NA, not available. Reference values: FSH 1–7 IU/L, LH 2–10 IU/L, T > 12 nmol/L, TV > 12 mL.

(B) Schematic representation of human *SPOCD1* variants identified in the three infertile men.

(C) Pedigree of individual M3457 and sequencing trace of *SPOCD1* exon 15 harboring the *SPOCD1*^{c.2912T>G} variation with substituted nucleotide highlighted. NOA, non-obstructive azoospermia.

(D) Pedigree of individual GEMINI-88 and sequencing trace of *SPOCD1* exon 16 harboring the *SPOCD1*^{c.3354_3355insA} variation with inserted nucleotide highlighted from GEMINI-88.

(E) Periodic acid-Schiff (PAS) and hematoxylin-stained testis sections of GEMINI-88 and control individual. SCO, Sertoli cell only. Scale bars, 50 μm.

(F) Pedigree of individual M2021 and sequencing trace of part of *SPOCD1* exon 8 harboring the *SPOCD1*^{c.1991_1992del} variation from M2021 and brother; missing nucleotides are indicated. WT, canonical *SPOCD1* allele; MUT, *SPOCD1*^{c.1991_1992del} allele.

(G) PAS and hematoxylin-stained testis sections of M2021. SPG, spermatogonia; SPCs, spermatocytes; RSs, round spermatids; MeiA, meiotic arrest; SCO, Sertoli cell only. Scale bars, 5 and 50 μm.

(H) Testis section of M2021 and control individual stained for LINE1 ORF1p (brown 3,3'-diaminobenzidine [DAB] stain). SPG, spermatogonia. Scale bars, 50 and 5 μm.

See also Figure S1.

SPOC domain (Figure 2D). The substitution of the non-polar, aliphatic side chain of Leucine 792 with that of a charged arginine in the SPOCD1 L792R protein is likely incompatible with domain folding (Figure 2D). Indeed, although the recombinant wild-type mouse SPOCD1 SPOC domain expressed well in *E. coli*, the L792R mutant did not (Figure S2B). This suggests that the *SPOCD1* c.2912T>G allele encodes a protein with a misfolded SPOC domain. Although the domain is unfolded, it did not affect the expression of the mouse *SPOCD1* L792R protein or human *SPOCD1* L971R protein in HEK293T (293T)

cells (Figure 2E). Although the molecular function of the *SPOCD1* SPOC domain remains unknown, our data indicates that it is essential for *SPOCD1* function.

The carboxy-terminal region of SPOCD1 associates with C19ORF84

The *SPOCD1*^{c.3354_3355insA} pathogenic allele was intriguing because the variant resides in the last exon, sparing the transcript from nonsense-mediated decay and potentially expressing a truncated protein. Indeed, the *SPOCD1*-Q1119fs mutant protein was stable

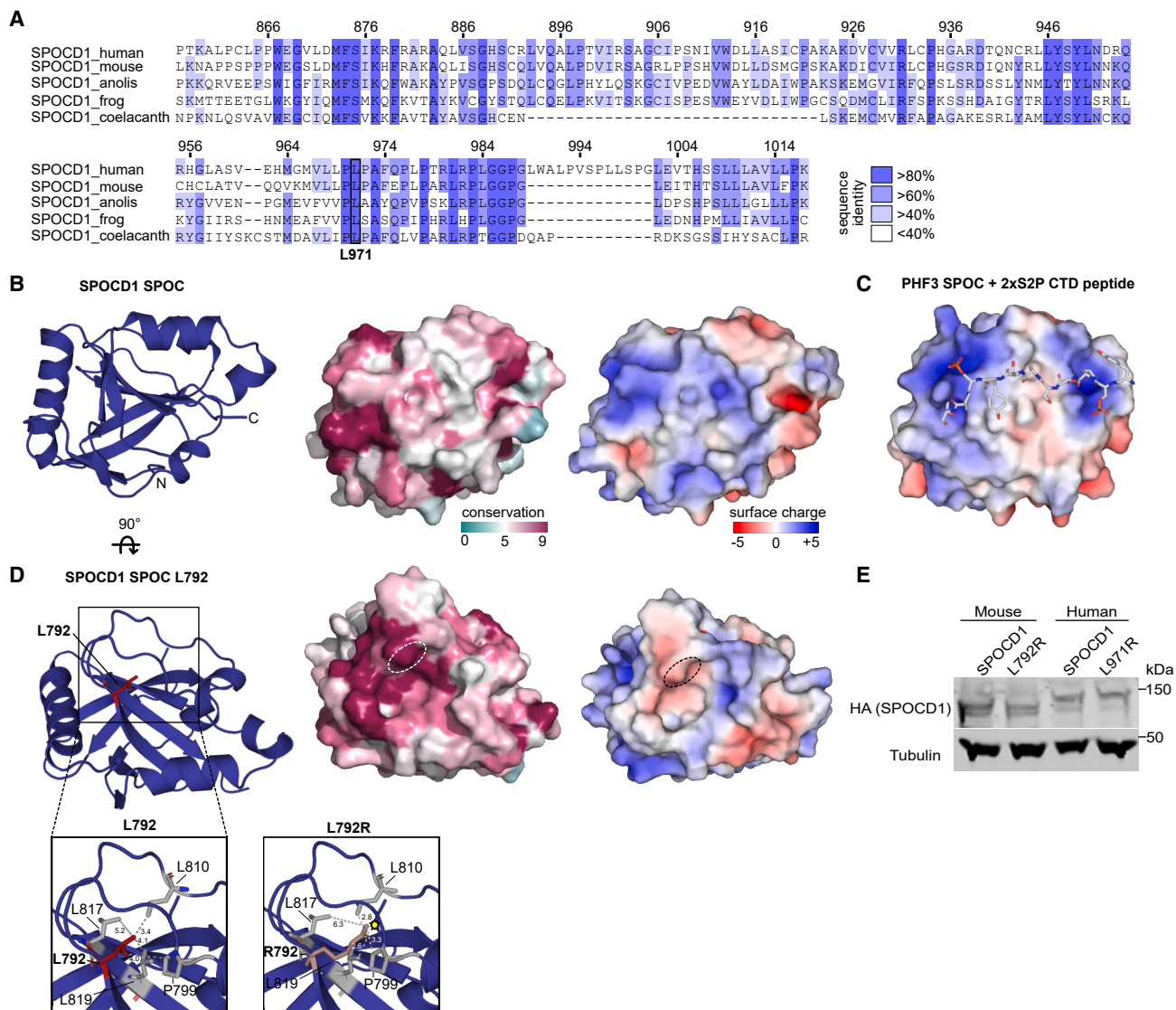


Figure 2. Crystal structure of SPOCD1 SPOC domain

(A) Multiple sequence alignment of the SPOCD1 SPOC domains from indicated species. Sequence identity conservation is denoted by depth of color. Red box highlights homologous position of human L971.

(B) Crystal structure of the mouse SPOCD1 SPOC domain, displayed from left to right as cartoon view indicating C and N terminus, surface conservation, and surface charge.

(C) Surface charge of PHF3 SPOC domain with RNA polymerase II CTD diheptapeptide phosphorylated on Ser2 (2xS2PCTD). PDB: 6IC8. Structures in (B) and (C) are shown in the same orientation.

(D) SPOCD1 SPOC domain shown in alternative orientation with residue L792 highlighted in red in the cartoon ribbon display and circled in the surface conservation and charge displays. Below is a zoom in on region around L792 with interacting residues highlighted in gray and distances between residues indicated. A similar view shows the modeling of the L792R mutation, indicating clashes with the surrounding residues.

(E) Western blot of HEK293T cell expression of indicated full-length human and mouse SPOCD1. See also Figure S2.

when expressed in 293T cells (Figure 3A). Protein sequence analysis predicts that the deleted 98 amino acids contain a conserved and ordered region (Figure 3B). The AlphaFold2 model of human and mouse SPOCD1^{33,34} also predicts the first part of the deleted C terminus to be an ordered α helix, with the remainder being unstructured (Figures 3C and S3A). Collectively, these observations indicated that the deleted polypeptide region in SPOCD1-Q1119fs is essential for SPOCD1 function and could mediate

an interaction through the conserved region. To explore this possibility, we tested two SPOCD1-associated factors from mouse fetal testis⁵ (Figure S3B), DNMT3L, and the uncharacterized C19ORF84 protein. C19ORF84 was of interest because, like other key factors, its expression is restricted to the period of *de novo* DNA methylation (Figure 3D). C19ORF84 is also an uncharacterized protein of unknown function. We found that the C terminus of mouse SPOCD1 (amino acids 825–1,015) mediated an interaction with mouse

Table 1. SPOCD domain data collection and refinements

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 a = 90°, b = 90°, g = 90°
Resolution (Å)	54.52 – 1.70 (1.79 – 1.7)
Reflections	160,141 (16,403)
Unique reflections	33,752 (4,799)
R _{meas} (%)	0.062 (0.513)
CC (1/2)	0.999 (0.802)
Completeness (%)	99.2 (98.7)
Mean I/sI	13.2 (2.1)
Multiplicity	4.7 (3.4)
Refinement	
R _{work} /R _{free}	0.201 / 0.237
RMS bonds	0.007
RMS angles	1.037
Ramachandran outliers	0%
Allowed	0
Partially allowed	0
Disallowed	0
Total number of atoms	4,679
Protein atoms	4,403
Water/ligands	276
Average B factor (protein/solvent) (Å ²)	32.0

C19ORF84 when expressed in 293T cells (Figure 3E). We confirmed that full-length SPOCD1 and C19ORF84 could reciprocally co-precipitate each other in 293T cells, and the interaction was independent of SPOCD1's conserved SPOC and transcription factor S-II central (TFIISM) domains (Figures 3F, S3C, and S3D). Indeed, the conserved residues 942–964 of SPOCD1, which are predicted to form an α helix, were required to co-precipitate C19ORF84 (Figures 3G, S3E, and S3F). The AlphaFold2 model predicted that C19ORF84 is predominantly disordered with a central conserved α -helical region^{33,34} (Figures S3G and S3H). We mapped the mouse SPOCD1-C19ORF84 association to amino acids 81–90 of mouse C19ORF84 (Figures 3H, S3G, and S3H). Finally, we demonstrated that the human SPOCD1 and C19ORF84 reciprocally co-precipitated when expressed in 293T cells but that the human SPOCD1-Q1119fs protein failed to interact with human C19ORF84 (Figures 3I and 3J). Next, we raised antibodies against mouse C19ORF84 (Figures S4A–S4C). Using these antibodies for immunoprecipitation mass spectrometry (IP-MS) from E16.5 fetal testis extracts, we found C19ORF84 associated with SPOCD1 (Figure 3K; Table S1). The overlap of C19ORF84- and epitope tagged SPOCD1-associated factors from fetal testis was modest (Figures 3K and S3I; Table S1). This is likely due to technical reasons. Firstly, the custom antibodies give higher background obscuring some associations. Secondly, the binding of antibodies to

C19ORF84 epitopes could displace factors from the complex, which is especially relevant given that C19ORF84 is a small protein. Recombinantly expressed C19ORF84 fragment encompassing the first 100 amino acids could pull down a C-terminal SPOCD1 fragment (amino acids 900–1,000) (Figure 3L). Analytical size exclusion chromatography (SEC) further demonstrated that the SPOCD1 fragment elutes at an earlier volume when co-injected with C19ORF84 fragment, as compared with SPOCD1 fragment alone (Figure 3M). In summary, we demonstrated that the SPOCD1-C19ORF84 interaction can be recapitulated using recombinant fragments of the respective mouse proteins. In conclusion, we have identified the uncharacterized C19ORF84 as an interactor of SPOCD1 that is likely to function in transposon silencing.

DNMT3C associates with the SPOCD1 and C19ORF84 in fetal gonocytes

Our mouse C19ORF84 antibodies allowed us to demonstrate that C19ORF84 is restricted to germ cell nuclei during the period of *de novo* methylation (Figures 4A, 4B, S4C, and S4D). Not only does the expression of C19ORF84 coincide with that of SPOCD1, but both proteins co-localize in gonocytes (Figures 4C and S4E). This is most apparent at E18.5 when both SPOCD1 and C19ORF84 are found throughout the nucleoplasm and in foci, which we termed C19ORF84 foci (Figures 4C and S4E). Approximately 30% of fetal gonocytes exhibited C19ORF84 foci at E18.5 with a median number of 3 foci per cell (Figure 4D). Almost all C19ORF84 foci co-localized with SPOCD1 foci (Figure 4E). Although the function and importance of these foci remain unknown, they allow us to explore which other factors co-localize to C19ORF84 foci. To this end, we used endogenously tagged hemagglutinin (HA) epitope alleles for *Miw12* (*Miw12^{HA}*),²⁸ *Dnmt3l* (*Dnmt3l^{HA}*) (Figures S5A–S5C, S5G, and S5H), and *Dnmt3c* (*Dnmt3c^{HA}*) (Figures S5D–S5F, S5I, and S5J). *MIW12* was also found to localize to C19ORF84 foci (Figures 4F, 4G, and S4F). We next analyzed DNMT3L and found that it was present in foci but it did not form foci itself (Figures 4H, 4I, and S4G). Intriguingly, DNMT3C co-localized to C19ORF84 foci and formed foci (Figures 4J, 4K, and S4H). We observed that DNMT3C is a lowly expressed protein compared with the other factors based on the strength of signal in the immuno-fluorescence experiments. We performed western blotting from compound heterozygous *Dnmt3l^{HA/+}*; *Dnmt3c^{HA/+}* E16.5 fetal testis using anti-HA antibodies to simultaneously detect both proteins. This revealed that DNMT3C is expressed approximately 14-fold less expressed than DNMT3L (Figure 4L). Next, we used the *Dnmt3l^{HA}* and *Dnmt3c^{HA}* alleles to perform IP-MS studies. For comparative purposes, we employed the previously established conditions used for SPOCD1 IP-MS from the same developmental time point, E16.5 fetal testis.⁵ Among the highly enriched and confident (>4-fold enrichment and p value < 0.05) factors that associated with DNMT3L, we found DNMT3C and DNMT3A but neither SPOCD1 nor C19ORF84 (Figure 4M; Table S2). DNMT3L was confidently detected in the DNMT3C precipitates but with 2-fold enrichment (Figure 4N; Table S3). The reason for this modest enrichment stems from the fact that under the conditions used DNMT3L weakly bound to the

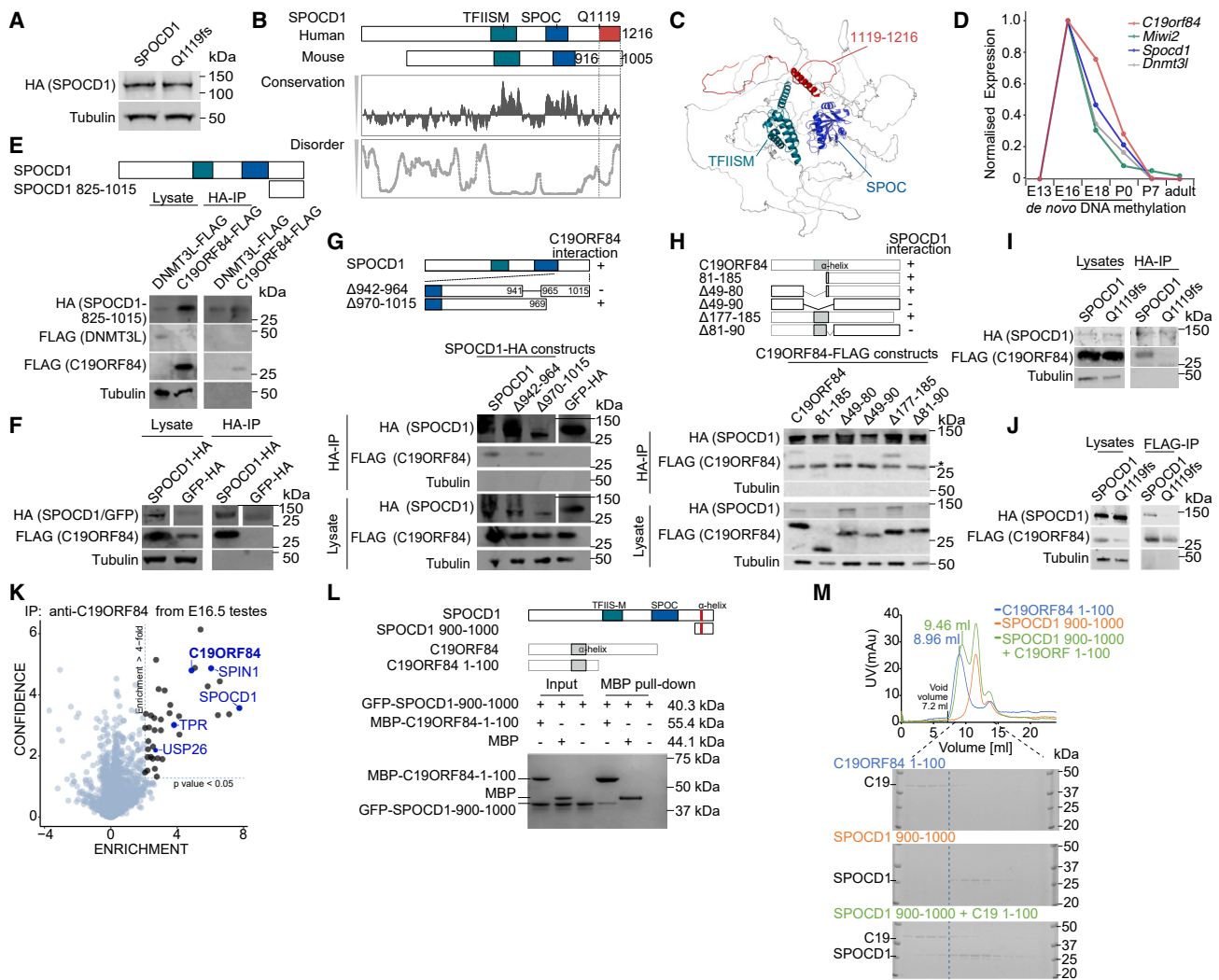


Figure 3. SPOCD1 interacts with C19ORF84 via a carboxy-terminal helix

(A) Representative western blot protein expression analysis of $n = 3$ 293T cell lysates after transfection with human SPOCD1-HA and SPOCD1-Q1119fs-HA. (B) Conservation analysis and structure disorder prediction for human SPOCD1 is displayed together with protein schematic showing the region homologous to the amino acids mutated in human SPOCD1-Q1119fs. (C) AlphaFold2 structure prediction of human SPOCD1 (Q6ZMY3) is shown, and indicated domains are highlighted. (D) Gene expression analysis from CAGE (cap analysis of gene expression) data³⁵ of whole testes from mice of the indicated age is shown. (E–J) Representative western blot analyses of $n = 3$ immunoprecipitations of the indicated mouse (E–H) or human (I and J) proteins expressed in 293T cells. *, Ig light chain eluting from beads. (K) Volcano plot showing enrichment (\log_2 [mean label-free quantification [LFQ] ratio of $n = 3$ anti-C19ORF84 immunoprecipitates / immunoprecipitates of $n = 3$ rabbit serum controls]) and statistical confidence (\log_{10} [p value of two-sided Student's t test]) of proteins co-purifying with C19ORF84 from E16.5 testes. Previously identified SPOCD1-associated proteins⁵ more than 4-fold enriched highlighted in blue. (L) Representative Coomassie blue gel image of $n = 3$ co-precipitation experiments with indicated recombinant mouse C19ORF84 and SPOCD1 fragments. (M) Analytical size exclusion chromatography of the indicated C19ORF84 (blue) fragment, the SPOCD1 (orange) fragment, and the combined protein fragments (green). The top panel shows superposed representative chromatograms ($n = 2$) for each sample, with void (black; 7.2 mL) and peak elution volumes indicated. Below, separate Coomassie blue gels of each run are shown, with samples from the same set of fractions used on each gel. Dashed line indicates the lowest volume that SPOCD1 elutes in on its own. See also Figure S3 and Table S1.

beads in the control IP. Most importantly, DNMT3C robustly associated with SPOCD1 and C19ORF84 in fetal testis (Figure 4N; Table S3). In summary, this is the first time the DNMT3C enzyme has been visualized *in vivo* and a direct connection to the piRNA pathway has been made.

C19ORF84 is required for mouse spermatogenesis and transposon silencing

To understand if C19ORF84 functions in piRNA-directed DNA methylation, we generated a mouse *C19orf84* null allele (*C19orf84*^{-/-}) that resulted in the loss of C19ORF84 protein

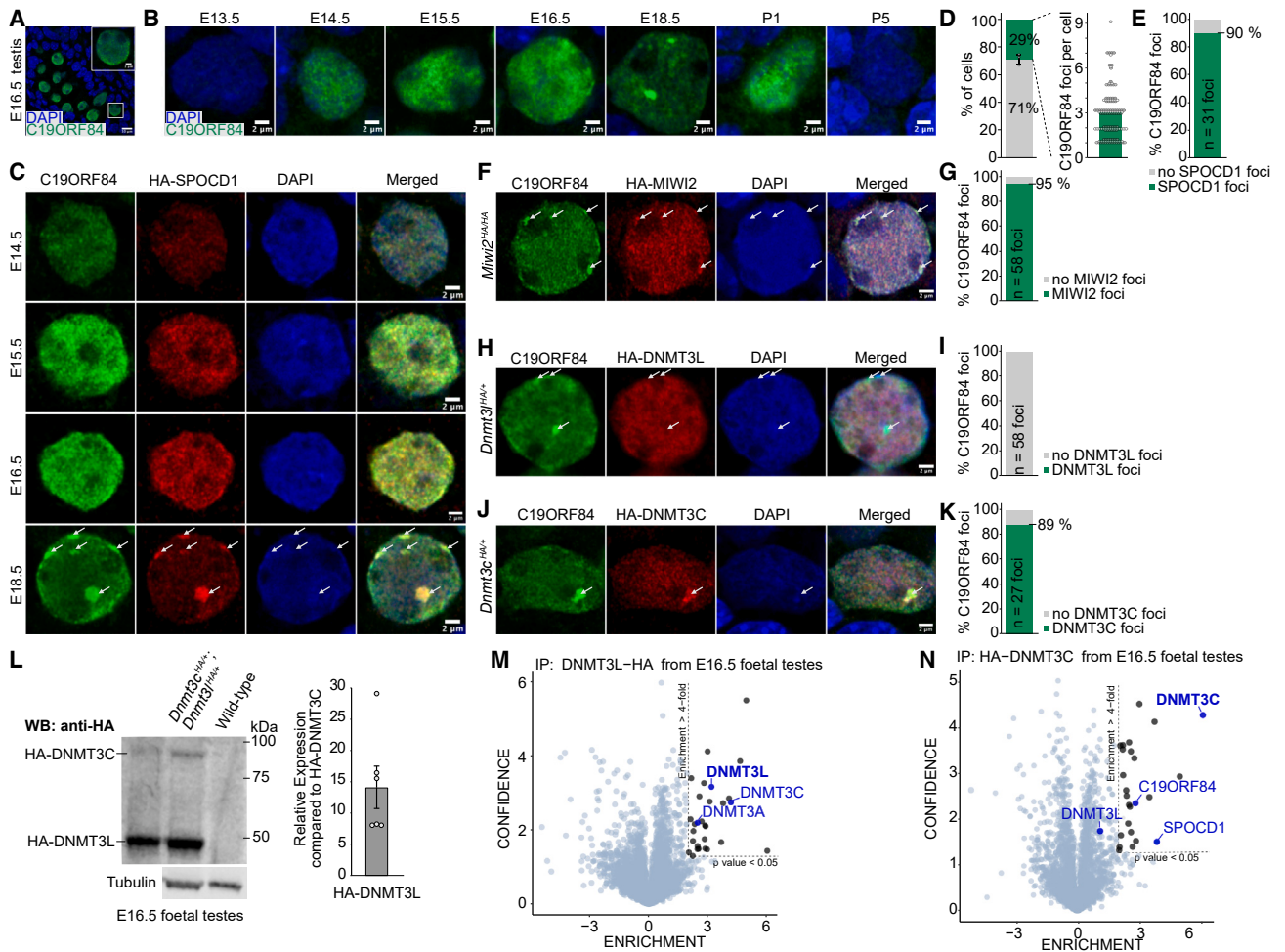


Figure 4. C19ORF84 is a gonocyte-specific protein that associates with nuclear piRNA pathway and *de novo* DNA methylation factors

(A) C19ORF84 localization in male germ cells at E16.5 with zoom-in of indicated gonocyte shown in insert.
 (B) C19ORF84 protein expression in gonocytes for the indicated developmental stages.
 (C) C19ORF84 (green) co-localization with HA-SPOCD1 (red) in *Spocd1*^{HA/+} fetal testes shown for the indicated time points.
 (D) Left plot shows percentage of germ cells at E18.5 with at least one C19ORF84 focus. Right plot datapoints show number of foci in foci-containing cells with median foci count displayed as green bar. Data are mean and SD on left plot and median and datapoints on right for n = 3 biological replicates with datapoints offset and shaded differently for each replicate in right plot.
 (E) Percentage of C19ORF84 foci overlapping with HA-SPOCD1 foci (n = 31 foci).
 (F–K) C19ORF84 (green) co-localization in gonocytes at E18.5 with in red (F) HA-MIWI2, (H) HA-DNMT3L, and (J) HA-DNMT3C and quantification of C19ORF84 foci overlap with (G) HA-MIWI2 (n = 58 foci), (I) HA-DNMT3L (n = 58 foci), and (K) HA-DNMT3C (n = 27 foci) foci. White arrows denote localization of C19ORF84 foci. Images in (A)–(C), (F), (H), and (J) are representative for n = 3 biological replicates of the indicated genotype or time point. Scale bars are 10 μ m (A) and 2 μ m (A [insert], B, C, F, H, and J).
 (L) Representative anti-HA western blot from *Dnmt3c*^{HA/+}; *Dnmt3l*^{HA/+} and wild-type E16.5 testes lysates. Quantification of relative expression of DNMT3L to DNMT3C shown to the right. Data are mean and SEM of n = 6 biological replicates.
 (M) Volcano plot showing enrichment (\log_2 [mean LFQ ratio of anti-HA immunoprecipitates from *Dnmt3l*^{HA/+} / wild-type testes lysates]) and statistical confidence (\log_{10} [p value of two-sided Student's t test]) of proteins co-purifying with HA-DNMT3L from E16.5 fetal testes (n = 3).
 (N) Volcano plot as in (M) showing proteins co-purifying with HA-DNMT3C in anti-HA IP-MS from E16.5 *Dnmt3c*^{HA/+} fetal testes (n = 3). Proteins of interest highlighted in blue in (M) and (N).
 See also Figures S4 and S5 and Tables S2 and S3.

expression (Figures 5A, 5B, and S6A–S6C). The loss of factors required for *de novo* transposon methylation in mice results in male infertility, meiotic arrest, and deregulation of both LINE1 and IAP transposons.^{4–6,8,12,27,36} *C19orf84*-deficiency resulted in male-specific infertility, atrophic testes, and the absence of

spermatozoa in the epididymis (Figures 5C–5E). Histological analysis of *C19orf84*^{−/−} testes revealed aberrant seminiferous tubules that presented an early pachytene meiotic arrest (Figures 5F, S6D, and S6E). The LINE1 ORF1 (ORF1p) and IAP GAG proteins were expressed in *C19orf84*^{−/−} spermatocytes

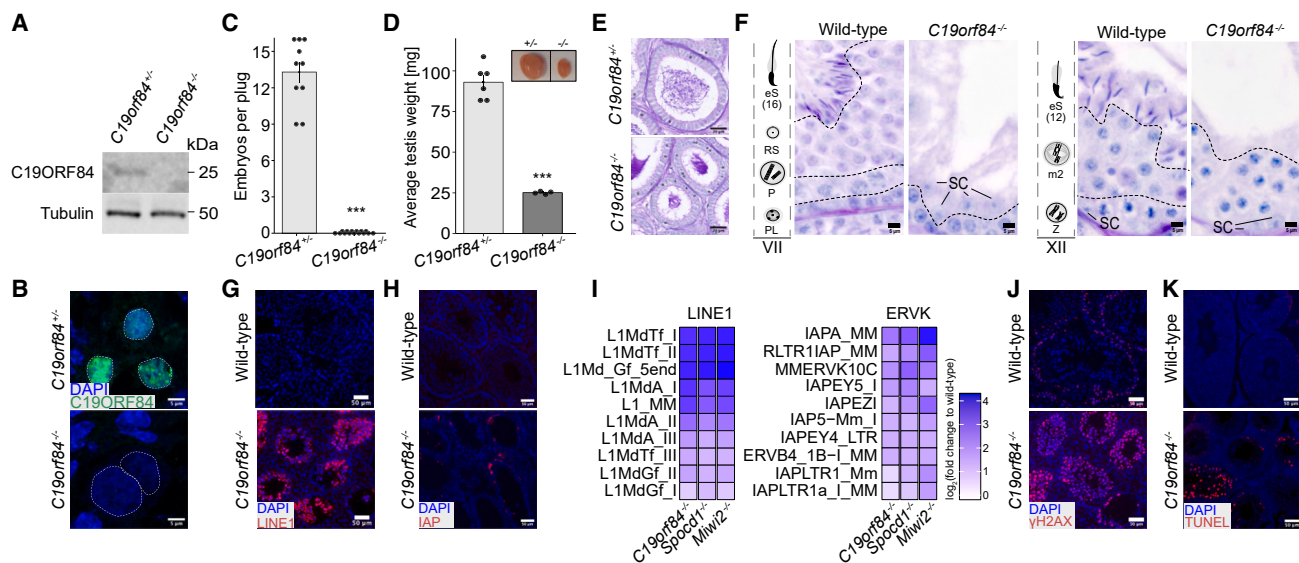


Figure 5. C19ORF84 is required for male fertility and transposon silencing

(A) Western blot analysis of C19ORF84 protein abundance in E16.5 testes from *C19orf84*^{+/-} and *C19orf84*^{-/-} mice.
 (B) C19ORF84 staining of E16.5 fetal testis sections from *C19orf84*^{+/-} and *C19orf84*^{-/-} mice.
 (C) Number of embryos per plug fathered by studs with the indicated genotype mated to wild-type females. Data are mean and SEM from n = 3 *C19orf84*^{+/-} (10 plugs total) and n = 4 *C19orf84*^{-/-} studs (15 plugs total), ***p < 0.001.
 (D) Testis weight of adult mice with the indicated genotype. Data are mean and SEM from n = 6 *C19orf84*^{+/-} and n = 4 *C19orf84*^{-/-} mice, ***p < 0.001. Insert shows a representative image of testes from *C19orf84*^{+/-} (left) and *C19orf84*^{-/-} (right) mice.
 (E and F) Representative images of PAS and hematoxylin-stained epididymis (E) and testes (F) sections of two stages of the seminiferous cycle, indicating a germ cell arrest at the early pachytene stage. VII, Stage 7; PL, pre-leptotene; P, pachytene; RSs, round spermatids; eS(16), step 16 elongated spermatids; SC, Sertoli cell; Z, zygotene; m2, secondary meicytes.
 (G and H) Adult testes from wild type and *C19orf84*^{-/-} mice stained for LINE1 ORF1p (G) or IAP GAG (H) protein.
 (I) RNA-seq heatmap analysis showing fold-change of expression (relative to wild type) of the 10 most upregulated LINE1 and ERV transposons in P20 testes from *C19orf84*^{-/-}, *Spocd1*^{-/-}, and *Miwi2*^{-/-} mice. n = 3 for all genotypes.
 (J and K) Adult testis sections stained for DNA damage marker γ H2AX (J) or for apoptotic cells by terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay (K) from wild-type and *C19orf84*^{-/-} mice. Scale bars, 5 μ m (B and F), 20 μ m (E), 50 μ m (G, H, J, and K). Germ cell nuclei are indicated with a dashed line in (B). Images shown in (B), (E)–(H), (J), and (K) are representative of data from n = 3 mice per genotype.
 See also Figure S6.

and spermatogonia, respectively (Figures 5G and 5H), as is the case in *Spocd1*^{-/-}, *Miwi2*^{-/-}, or *Dnmt3l*^{-/-} mice.^{4,5,8,11,27} We employed RNA sequencing (RNA-seq) of postnatal day 20 (P20) to explore the full repertoire of deregulated transposons. Indeed, the same young transposon families were derepressed as in *Spocd1*^{-/-} or *Miwi2*^{-/-} mice (Figure 5I). A hallmark of transposon expression is DNA double-strand breaks, typically marked by phosphorylation of the histone variant H2AX (γ H2AX).³⁷ Strong γ H2AX staining, indicative of extensive DNA damage, was observed in *C19orf84*^{-/-} meicytes as distinct from characteristic foci observed in meiotic cells in control *C19orf84*^{+/-} testes (Figure 5J). Indeed, this extensive DNA damage correlated with widespread apoptosis of meiotic cells in *C19orf84*^{-/-} testes (Figure 5K). In summary, C19ORF84 safeguards spermatogenesis by mediating transposon silencing.

C19ORF84 is essential for piRNA-directed transposon methylation

The tissue changes observed in mouse testis, activation of transposons, and DNA damage presented by *C19orf84*^{-/-} mice are the defining phenotype associated with defective piRNA-directed

DNA methylation.^{27,38–40} To understand the role of C19ORF84 in transposon silencing, we analyzed genome methylation from purified P14 spermatogonia, a time point prior to the phenotypic onset but after completion of *de novo* genome methylation. The piRNA pathway is specifically required for *de novo* DNA methylation of some IAP elements, several LINE1 sub-families, and the imprinted *Rasgrf1* locus.^{27,38,39,41,42} Accordingly, C19ORF84 was not required for general genome *de novo* methylation (Figure 6A). Indeed, *C19orf84*-deficiency did not impact genic, intergenic, CpG island, and gene promoter regions, or collective transposon DNA methylation levels (Figure 6A). The young LINE1 families L1Md_A, L1Md_Gf, and L1Md_T were hypomethylated, whereas this was less pronounced in the older L1Md_F family (Figure 6B). The ERV IAPEy also failed to be fully methylated in *C19orf84*^{-/-} spermatogonia, unlike IAPEz, ERVL, or short interspersed nuclear element (SINE) transposons that are not under the control of the piRNA pathway (Figure 6B). The piRNA pathway is required for promoter DNA methylation of young active transposons.^{5,6,11,43} A metaplot analysis of methylation levels across LINE1 families revealed defective *de novo* methylation specifically at their promoters in *C19orf84*^{-/-} spermatogonia (Figure 6C). Hypomethylation was particularly evident in young LINE1 families

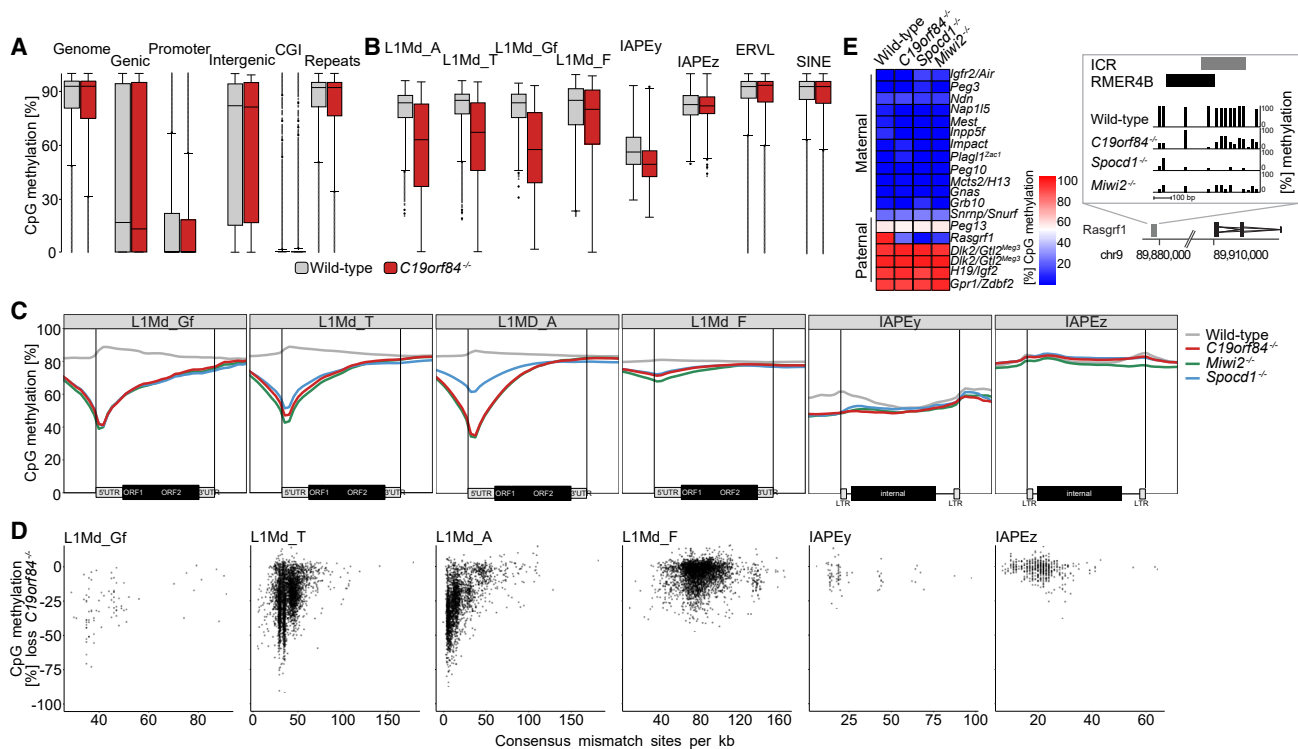


Figure 6. C19ORF84 is essential for piRNA-directed DNA methylation

Analysis of whole genome CpG methylation of P14 spermatogonia from $n = 3$ $C19orf84^{-/-}$ and wild type (A–E) as well as $Spocd1^{-/-}$ and $Miwi2^{-/-}$ (C and E) mice. (A and B) Percentage of CpG methylation levels for the indicated genomic features and transposon families (genic, promoter, intergenic, and CpG islands [CGIs] being defined as non-overlapping with transposons) shown as boxplots. The horizontal line represents the median, boxes the 25th to 75th interquartile range, and dots datapoints outside that range.

(C) Metaplots of mean CpG methylation over the consensus sequence for the indicated LINE1 and IAP families and adjacent 2 kb.

(D) Correlation analysis for individual elements of the specified LINE1 and IAP families of mean CpG methylation loss in $C19orf84^{-/-}$ spermatogonia (relative to wild type) in relation to the element’s sequence divergence from the consensus.

(E) Mean CpG methylation of imprinted loci is presented as a heatmap. The imprinted control region (ICR) of *Rasgrf1* is shown in detail on the right.

(L1Md_A, L1Md_Gf, and L1Md_T) compared with the older L1Md_F family (Figure 6C). In addition, the methylation of young LINE1s within the respective families requires C19ORF84 function (Figure 6D). Defective promoter DNA methylation was evident in IAPEy but less so in IAPEz families from $C19orf84^{-/-}$ spermatogonia (Figure 6C). However, defective methylation (> 25% CpG methylation loss) within both families was observed when viewed at the single locus levels (Figure 6D). Collectively, the defective transposon methylation in $C19orf84^{-/-}$ spermatogonia is the same as described for $Miwi2^{-/-}$, $Spocd1^{-/-}$, and $Dnmt3c^{-/-}$ mice.^{5,11} Finally, among imprinted loci, the methylation of only *Rasgrf1* is dependent upon C19ORF84 (Figure 6E). In summary, C19ORF84 is essential for piRNA-directed *de novo* transposon methylation.

C19ORF84 connects piRNA and DNA methylation machineries *in vivo*

The fact that C19ORF84 is a SPOCD1-interacting nuclear protein and is required for piRNA-directed methylation indicates a clear function in *de novo* genome methylation. To exclude the remote possibility that C19ORF84 participates in piRNA biogenesis, we sequenced small RNAs from $C19orf84^{+/-}$ and $C19orf84^{-/-}$

E16.5 fetal testes. The loss of C19ORF84 did not impact piRNA length distribution, annotation, or amplification (Figures 7A–7C). The loading of MIWI2 with piRNA licenses its entry to the nucleus; thus, disruption of piRNA processing results in the dramatic reduction of MIWI2’s nuclear localization.^{28,36} The fact that MIWI2 was nuclear in the absence of C19ORF84 (Figures 7D and S7A) indicates that piRNA processing occurs normally and that C19ORF84 must act downstream of MIWI2 in the DNA methylation process. Normal levels and localization of SPOCD1 in $C19orf84^{-/-}$ fetal gonocytes (Figures 7E and S7B) also excluded a role for the expression or nuclear localization of SPOCD1.

We next explored the impact of C19ORF84-deficiency on the formation and composition of the C19ORF84 foci in E18.5 fetal gonocytes. The loss of C19ORF84 dramatically reduced the frequency of cells that had SPOCD1 localized to foci from 34% in wild type to 6% in $C19orf84^{-/-}$ fetal testes (Figures 7F, 7G, and S7C). A similar reduction, from 19% to 7%, was observed in fetal gonocytes with foci-containing DNMT3C in $C19orf84^{-/-}$ fetal testes (Figures 7H, 7I, and S7D). The frequency of cells with foci differed when using SPOCD1 (34% in wild type) and DNMT3C (19% wild type) in identifying foci, which could be due to different antibodies used. However, the very low expression levels of

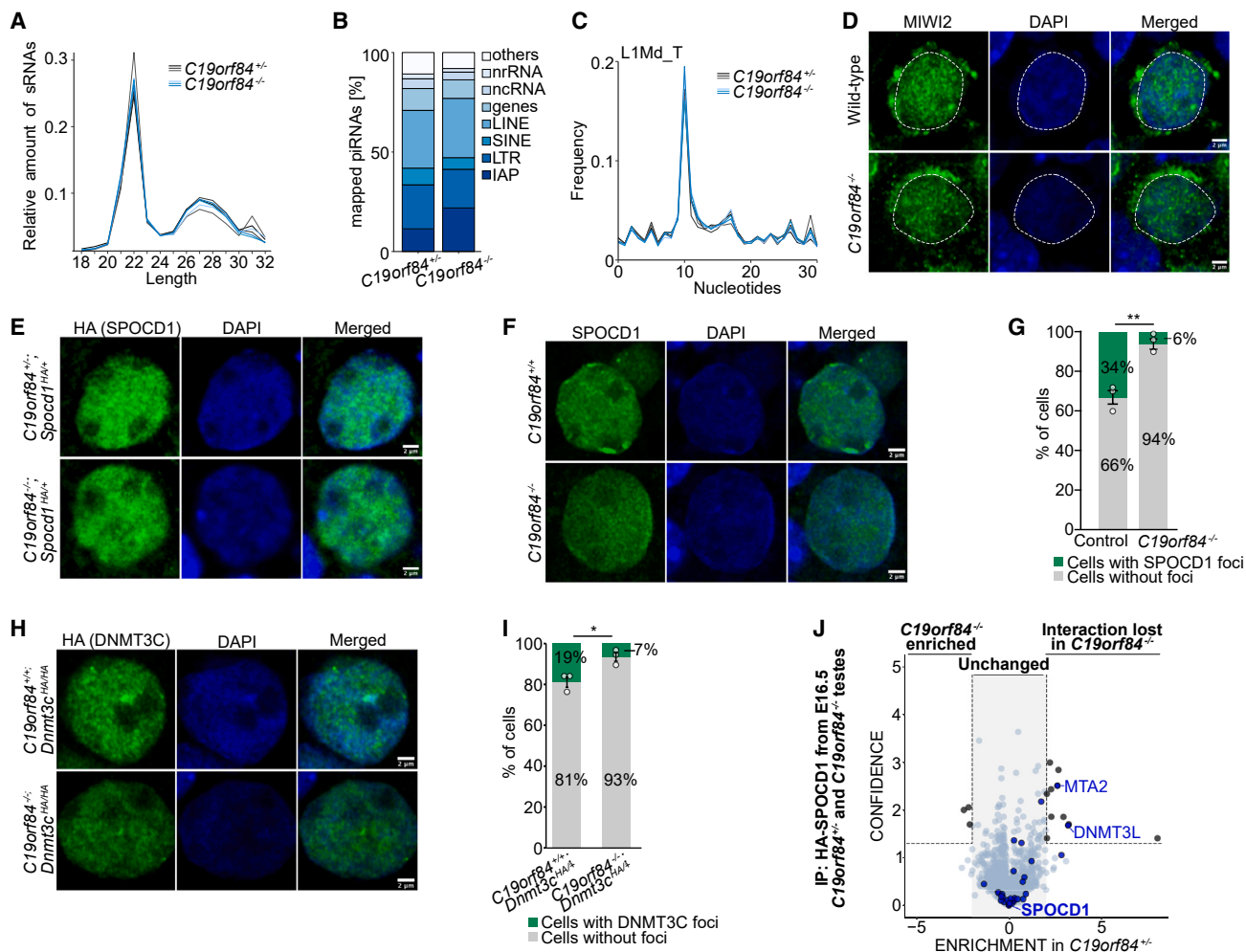


Figure 7. C19ORF84 connects the piRNA and DNA methylation machineries *in vivo*

(A–C) piRNA analysis of small RNA-seq data from $n = 3$ $C19orf84^{+/+}$ and $C19orf84^{-/-}$ E16.5 testes, showing nucleotide (nt) length distribution of small RNAs in (A) (no significant differences were observed in Bonferroni-adjusted Student's t test, $p \approx 0.99$), annotation of piRNA targets from merged replicates in (B), and piRNA ping-pong analysis presented as relative frequency of the distance between 5' ends of complementary piRNAs mapping to the LINE1 L1Md_T consensus in (C). (D) MIW12 staining of fetal testis sections from $n = 3$ wild type and $C19orf84^{-/-}$ E16.5 mice. Germ cell nuclei are indicated with dashed line. Scale bars, 2 μm . (E) HA-SPOCD1 staining of fetal testis sections from $n = 3$ $Spocd1^{HA/HA}; C19orf84^{+/+}$ and $Spocd1^{HA/HA}; C19orf84^{-/-}$ mice. Scale bars, 2 μm .

(F and G) E18.5 $C19orf84^{+/+}$ and $C19orf84^{-/-}$ gonocytes stained for SPOCD1 (F) and quantification of germ cell percentage presenting with at least one SPOCD1 foci (G), $**p \approx 0.0035$.

(H and I) E18.5 $C19orf84^{+/+}; Dnmt3c^{HA/HA}$ and $C19orf84^{-/-}; Dnmt3c^{HA/HA}$ gonocytes stained for HA-DNMT3C (H) and quantification of germ cell percentage presenting with at least one HA-DNMT3C foci (I), $*p \approx 0.022$. Representative images of gonocytes in (F and H) and data (mean and SEM) in (G) and (I) from $n = 3$ biological replicates per genotype. Scale bars, 2 μm (F and H).

(J) Volcano plot showing enrichment ($\log_2[\text{mean LFQ ratio of anti-HA immunoprecipitates from } n = 2 \text{ } Spocd1^{HA/HA}; C19orf84^{+/+} / n = 3 \text{ } Spocd1^{HA/HA}; C19orf84^{-/-}$ fetal testes lysates]) and statistical confidence ($-\log_{10}[p \text{ value of two-sided Student's } t \text{ test}]$) of proteins co-purifying with HA-SPOCD1 from E16.5 testes. Previously identified SPOCD1-associated proteins⁵ highlighted in blue. See also Figure S7 and Table S4.

DNMT3C (Figure 4L) could additionally explain the discrepancy, as the sensitivity of foci staining is likely relative to the abundance of the target protein being analyzed. Due to the necessity of C19ORF84 for SPOCD1 and DNMT3C to form foci, we hypothesized that C19ORF84 may play a role in recruiting *de novo* methylation or other factors to SPOCD1. To test this hypothesis, we performed IP-MS of HA-SPOCD1 from $C19orf84^{+/+}; Spocd1^{HA/HA}$ and $C19orf84^{-/-}; Spocd1^{HA/HA}$ fetal testes (Figure 7J; Table S4). Because of the limitations associated with obtaining fetal testes

of complex genotypes, we optimized the SPOCD1 IP-MS procedure to start with lower amounts of input sample. C19ORF84 itself was not detected (Figure 7J; Table S4) due to reduced input and the fact that C19ORF84 generates only four quantifiable peptides under optimal conditions.⁵ Of the proteins that were previously shown to associate with SPOCD1, we found only two proteins, DNMT3L and MTA2, whose enrichment with SPOCD1 is dependent upon C19ORF84 (Figure 7J; Table S4). MTA2 is part of the NURD chromatin remodeling complex,⁴⁴ which does not have a

proven role in piRNA-directed transposon methylation. However, DNMT3L is a core component of the *de novo* methylation machinery and an essential factor in this process.^{7–10,45} Collectively, our data reveal that C19ORF84 connects SPOCD1 to the *de novo* methylation machinery *in vivo*.

DISCUSSION

The discovery of C19ORF84 identifies an essential component in the pathway and reveals a hidden complexity of the final stages of piRNA-directed transposon methylation. The recruitment of DNMT3L to SPOCD1 in fetal germ cells requires C19ORF84. This suggests that this small, principally unstructured protein could act as an adaptor protein that licenses transposon methylation through the recruitment of the *de novo* methylation machinery. Given that SPOCD1 requires additional factors to execute transposon methylation, it suggests that SPOCD1 acts as a scaffold, coordinating events that culminate in DNA methylation. This multifactorial approach may underlie the precision of piRNA-directed DNA methylation ensuring against aberrant off-target methylation that could result in germline epimutations.

Here, we identify foci in E18.5 fetal gonocytes that comprise of piRNA pathway components and the *de novo* methylation machinery. We have termed these C19ORF84 foci. That only a subset of cells at this time point have C19ORF84 foci could arise from our observations providing only a snapshot analysis. There may be heterogeneity with regard to developmental timing, with some cells forming foci slightly earlier or later in development. Alternatively, focus formation may be stochastic, so not every cell will form foci. Given the bulk of transposon methylation is completed by E18.5,^{39,46} it is unlikely the foci play a central role in the process. That said, their identification has proved extremely insightful. It has allowed us to confidently observe piRNA factors (MIWI2, SPOCD1, and C19ORF84) and the *de novo* methylation machinery (DNMT3C and DNMT3L) co-localizing. Interestingly, DNMT3L is present in foci but does not form brightly staining foci like the other factors. We demonstrated that DNMT3L associates with both DNMT3A and DNMT3C methyltransferases in fetal testis (Figure 4M). However, SPOCD1 and C19ORF84 were not found in the DNMT3L IP which likely reflects that only a tiny fraction of DNMT3L is associated with these factors. These data are consistent with the role of DNMT3L in the methylation of the entire genome, whereas the DNMT3C-piRNA pathway is focused on only active transposons, constituting less than 1% of the genome.^{5,11} Previously, we never observed DNMT3C in our IP-MS experiments.^{5,6} We now believe that this could be due to its low abundance, which is substantiated by our *Dnmt3^{H1A}* and *Dnmt3c^{H1A}* alleles. Using the same antibody for detection of both proteins, we demonstrated that DNMT3C is drastically less abundant than DNMT3L (Figure 4L). Finally, we demonstrated that DNMT3C associates with SPOCD1 and C19ORF84 in fetal testes (Figure 4N). Together, these studies have allowed us to link piRNA pathway factors with DNMT3C, the methyltransferase responsible for piRNA-directed transposon methylation in mice.¹³ DNMT3C is a murid rodent-specific innovation.^{11,12} The human genome encodes two *de novo* DNA methyltransferases, DNMT3A and DNMT3B.⁴⁷ Although *Dnmt3c* evolved from a duplication of the *Dnmt3b* locus,^{11,12} the N terminus of murid DNMT3C and primate DNMT3A have both evolved

under strong diversifying selection.²⁵ The human *de novo* DNA methyltransferases that realizes human piRNA-directed transposon methylation remain unknown. We favor the hypothesis that C19ORF84 acts a key adaptor protein required for recruitment of the *de novo* methylation machinery. Several lines of evidence support this hypothesis. Firstly, both SPOCD1 and MIWI2 associates with it *in vivo*.⁵ Secondly, C19ORF84 is essential for the process of piRNA-directed *de novo* transposon methylation (Figures 5 and 6). Thirdly, the absence of C19ORF84 does not interfere with the expression of MIWI2, SPOCD1, and DNMT3C (Figures 7D–7F and 7H). Fourthly, the ability of SPOCD1 and DNMT3C to form foci is dependent upon C19ORF84 (Figures 7F and 7H). Fifthly, DNMT3C associates with both SPOCD1 and C19ORF84 *in vivo* (Figure 4N). Finally, SPOCD1 fails to associate with DNMT3L in fetal gonocytes in the absence of C19ORF84 (Figure 7J). In summary, we can confidently assign a critical function to this uncharacterized protein C19ORF84 in the final stages of piRNA-directed DNA methylation.

Here, we show that SPOCD1 is a guardian of human male fertility and is required for transposon silencing. We found three infertile men with three distinct homozygous pathogenic *SPOCD1* alleles. According to the ClinGen guidelines that govern gene-disease association in humans,⁴⁸ our data establishes a strong validity for the gene-disease relationship and, thus, mutations in the *SPOCD1* gene as a cause of infertility in men (Data S1). Furthermore, under these guidelines, sequencing of the *SPOCD1* gene can be used for diagnostic purposes. Each of the disease associated *SPOCD1* alleles gives important insight into human SPOCD1 function and the piRNA pathway. The deregulation of LINE1 in the spermatogonia of infertile man M2021 demonstrates a conserved role for the piRNA pathway in regulating human germline transposon silencing. The SPOCD1 L971R substitution that lies within the SPOC domain in combination with our structural and molecular analyses reveals that an intact SPOC domain is required for human SPOCD1 function. The fact that the pathogenic human SPOCD1-Q1119fs protein can no longer bind C19ORF84 also indicates a potential role for C19ORF84 in human fertility. In summary, we have defined a role for human SPOCD1 in transposon silencing and discovered C19ORF84, an essential factor that mediates piRNA-directed DNA methylation.

Limitations of the study

Here, we identify C19ORF84 as an essential constituent of the piRNA pathway. We demonstrate that C19ORF84 interacts with SPOCD1 and present multiple lines of evidence indicating that C19ORF84 acts as an adaptor protein in the recruitment of the *de novo* methylation machinery *in vivo*. We were unable to purify full-length recombinant proteins to biochemically test C19ORF84's adaptor function. There could also be additional unidentified components that are essential for complex assembly. Future biochemical studies will be critical to understanding how these factors assemble to mediate the final steps of piRNA-directed DNA methylation. The SPOCD1 and C19ORF84 fragments formed a high molecular weight complex; we cannot be sure if this could be due to trace amounts of nucleic acid after protein purification or represents a true biochemical aggregation or oligomerization characteristic of the protein complex. We

identified three infertile men with distinct homozygous pathogenic *SPOCD1* alleles. We found deregulation of LINE1 in the patient that had germ cells remaining in the seminiferous tubules. A key and challenging future goal will be to determine the impact of *SPOCD1* or other piRNA-factor mutation on the establishment of human germ cell genome methylation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - *In vivo* mouse studies
 - Human participants
 - Cell lines and bacteria strains
- **METHOD DETAILS**
 - Study cohorts
 - Exome sequencing and bioinformatics analysis
 - Sanger sequencing
 - Histology of human testis samples
 - Protein conservation and disorder analysis
 - Protein purification
 - X-ray crystallography
 - Cell culture, transfection, immuno-precipitation and western blotting
 - Gene expression analysis
 - Antibody generation
 - Immunoprecipitation and mass spectrometry (IP-MS)
 - Pull-down assay
 - Analytical size exclusion chromatography
 - Mouse strains and experimentation
 - Immuno-fluorescence and foci count
 - Foetal testes extract preparation for western blotting
 - Histology of mouse samples
 - Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay
 - RNA sequencing and analysis
 - Fluorescence-activated cell sorting (FACS)
 - Whole-genome methylation sequencing and analysis
 - Small RNA sequencing and analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2024.01.014>.

ACKNOWLEDGMENTS

We acknowledge the EMBL GeneCore facility in Heidelberg, Germany, for preparing the EM-seq dataset of P14 spermatogonia and sequencing all next-generation sequencing libraries. The authors thank David Kelly at the Well-

come Centre for Cell Biology Centre for Optical Instrumentation Laboratory (COIL) and Matthieu Vermeren at the Centre for Regenerative Medicine Imaging Core Facility for microscopy and instrumentation support; Christos Spanos at the Wellcome Centre for Cell Biology Proteomics Facility for support with mass spectrometry analysis; the Shared University Research Facilities (SURF) histology and histological imaging facility for support with histology of mouse samples; Fiona Rossi at the Centre for Regenerative Medicine Flow Cytometry Facility for handling cell sorting of spermatogonia; and Theresa O'Connor at the Centre for Regenerative Medicine Tissue Culture Facility for providing cell culture facilities. This project was funded by Wellcome Trust grant 106144 and 225237 (D.O'C.), Wellcome Trust grant 213612 (R.V.B.), Wellcome Trust grant 200898 (A.G.C.), Wellcome Trust grant 103139 (J.R.), Wellcome Trust grant 095021 (R.C.A.), Wellcome Trust grant 200885 (R.C.A.), Wellcome Trust Core grant 203149 (Wellcome Centre for Cell Biology), Wellcome Trust multi-user equipment grant 108504 (Wellcome Centre for Cell Biology), German Research Foundation fellowship DFG ZO 376/1-1 (A.Z.), German Research Foundation Clinical Research Unit "Male Germ Cells" DFG CRU326, National Institutes of Health R01HD07864 (D.F.C. and K.I.A.), and National Health and Medical Research Project grant APP1120356 (M.K.O'B., D.F.C., and K.I.A.).

AUTHOR CONTRIBUTIONS

A.Z., G.K., and T.A. contributed to the design, execution, and analysis of most or key experiments. N.R., B.S., S.K., and F.T. recruited, identified, and characterized individuals M2021 and M3457. B.H., M.K.O'B., D.F.C., K.I.A., and L.N. recruited, identified, and characterized individual GEMINI-88. T.A. helped establish IP conditions and performed mass spectrometry analysis under the guidance of J.R. and R.C.A. M.H. solved the *SPOCD1* SPOC crystal structure under the guidance of A.G.C. M.H. and G.K. performed the interaction mapping of C19ORF84 to *SPOCD1* and together with A.G.C. analyzed protein conservation and disorder predictions of *SPOCD1* and C19ORF84. M.S. performed the human *SPOCD1* interaction studies with human C19ORF84. R.B.V. and Y.K. performed bioinformatic analysis of Methyl-seq data or sRNA-seq as well as RNA-seq data, respectively. T.S. prepared sRNA-seq libraries and, with Y.K., RNA-seq libraries of P20 testes. A.Z. and D.O'C. conceived and supervised this study and wrote the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 1, 2023

Revised: December 1, 2023

Accepted: January 17, 2024

Published: February 14, 2024

REFERENCES

1. Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–117.
2. Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607.
3. Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108.
4. De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., and O'Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**, 259–263.
5. Zoch, A., Auchynnikava, T., Berrens, R.V., Kabayama, Y., Schöpp, T., Heep, M., Vasiliauskaitė, L., Pérez-Rico, Y.A., Cook, A.G., Shkumatava, A., et al. (2020). *SPOCD1* is an essential executor of piRNA-directed de novo DNA methylation. *Nature* **584**, 635–639.

6. Schöpp, T., Zoch, A., Berrens, R.V., Auchynnikava, T., Kabayama, Y., Vasiliauskaitė, L., Rappsilber, J., Allshire, R.C., and O'Carroll, D. (2020). TEX15 is an essential executor of MIWI2-directed transposon DNA methylation and silencing. *Nat. Commun.* *11*, 3739.
7. Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc. Natl. Acad. Sci. USA* *99*, 16916–16921.
8. Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* *431*, 96–99.
9. Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J. Biol. Chem.* *279*, 27816–27823.
10. Webster, K.E., O'Bryan, M.K., Fletcher, S., Crewther, P.E., Aapola, U., Craig, J., Harrison, D.K., Aung, H., Phutikanit, N., Lyle, R., et al. (2005). Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc. Natl. Acad. Sci. USA* *102*, 4068–4073.
11. Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Héroult, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* *354*, 909–912.
12. Jain, D., Meydan, C., Lange, J., Claeys Bouuaert, C., Lailler, N., Mason, C.E., Anderson, K.V., and Keeney, S. (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 Genet.* *13*, e1006964.
13. Dura, M., Teissandier, A., Armand, M., Barau, J., Lapoujade, C., Fouchet, P., Bonneville, L., Schulz, M., Weber, M., Baudrin, L.G., et al. (2022). DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis. *Nat. Genet.* *54*, 469–480.
14. Okutman, O., Muller, J., Baert, Y., Serdarogullari, M., Gultomruk, M., Piton, A., Rombaut, C., Benkhalifa, M., Teletin, M., Skory, V., et al. (2015). Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. *Hum. Mol. Genet.* *24*, 5581–5588.
15. Arafat, M., Har-Vardi, I., Harlev, A., Levitas, E., Zeadna, A., Abofoul-Azab, M., Dyomin, V., Sheffield, V.C., Lunenfeld, E., Huleihel, M., et al. (2017). Mutation in TDRD9 causes non-obstructive azoospermia in infertile men. *J. Med. Genet.* *54*, 633–639.
16. Colombo, R., Pontoglio, A., and Bini, M. (2017). Two Novel TEX15 Mutations in a Family with Nonobstructive Azoospermia. *Gynecol. Obstet. Invest.* *82*, 283–286.
17. Gou, L.T., Kang, J.Y., Dai, P., Wang, X., Li, F., Zhao, S., Zhang, M., Hua, M.M., Lu, Y., Zhu, Y., et al. (2017). Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. *Cell* *169*, 1090–1104.e13.
18. Wang, X., Jin, H.R., Cui, Y.Q., Chen, J., Sha, Y.W., and Gao, Z.L. (2018). Case study of a patient with cryptozoospermia associated with a recessive TEX15 nonsense mutation. *Asian J. Androl.* *20*, 101–102.
19. Tan, Y.Q., Tu, C., Meng, L., Yuan, S., Sjaarda, C., Luo, A., Du, J., Li, W., Gong, F., Zhong, C., et al. (2019). Loss-of-function mutations in TDRD7 lead to a rare novel syndrome combining congenital cataract and nonobstructive azoospermia in humans. *Genet. Med.* *21*, 1209–1217.
20. Alhathal, N., Maddirevula, S., Coskun, S., Alali, H., Assoum, M., Morris, T., Deek, H.A., Hamed, S.A., Alsuhaibani, S., Mirdawi, A., et al. (2020). A genomics approach to male infertility. *Genet. Med.* *22*, 1967–1975.
21. Araujo, T.F., Friedrich, C., Grangeiro, C.H.P., Martelli, L.R., Grzesiuk, J.D., Emich, J., Wyrwoll, M.J., Kliesch, S., Simões, A.L., and Tüttelmann, F. (2020). Sequence analysis of 37 candidate genes for male infertility: challenges in variant assessment and validating genes. *Andrology* *8*, 434–441.
22. Nagiraj, L., Mørup, N., Nielsen, J.E., Stakaitis, R., Golubickaite, I., Oud, M.S., Winge, S.B., Carvalho, F., Aston, K.I., Khani, F., et al. (2021). Variant *PNLDC1*, Defective piRNA Processing, and Azoospermia. *N. Engl. J. Med.* *385*, 707–719.
23. Kherraf, Z.E., Cazin, C., Bouker, A., Fourati Ben Mustapha, S., Hennebicq, S., Septier, A., Coutton, C., Raymond, L., Nouchy, M., Thierry-Mieg, N., et al. (2022). Whole-exome sequencing improves the diagnosis and care of men with non-obstructive azoospermia. *Am. J. Hum. Genet.* *109*, 508–517.
24. Wyrwoll, M.J., Gaasbeek, C.M., Golubickaite, I., Stakaitis, R., Oud, M.S., Nagiraj, L., Dion, C., Sindi, E.B., Leitch, H.G., Jayasena, C.N., et al. (2022). The piRNA-pathway factor FKBP6 is essential for spermatogenesis but dispensable for control of meiotic LINE-1 expression in humans. *Am. J. Hum. Genet.* *109*, 1850–1866.
25. Molaro, A., Malik, H.S., and Bourc'his, D. (2020). Dynamic Evolution of De Novo DNA Methyltransferases in Rodent and Primate Genomes. *Mol. Biol. Evol.* *37*, 1882–1892.
26. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* *581*, 434–443.
27. Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* *12*, 503–514.
28. Vasiliauskaitė, L., Vitsios, D., Berrens, R.V., Carrieri, C., Reik, W., Enright, A.J., and O'Carroll, D. (2017). A MILI-independent piRNA biogenesis pathway empowers partial germline reprogramming. *Nat. Struct. Mol. Biol.* *24*, 604–606.
29. Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* *20*, 406–420.
30. Appel, L.M., Benedum, J., Engl, M., Platzer, S., Schleiffer, A., Strobl, X., and Slade, D. (2023). SPOC domain proteins in health and disease. *Genes Dev.* *37*, 140–170.
31. Appel, L.M., Franke, V., Bruno, M., Grishkovskaya, I., Kasiliauskaitė, A., Kaufmann, T., Schoeberl, U.E., Puchinger, M.G., Kostrhon, S., Ebenwaldner, C., et al. (2021). PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC. *Nat. Commun.* *12*, 6078.
32. Appel, L.M., Franke, V., Benedum, J., Grishkovskaya, I., Strobl, X., Polyansky, A., Ammann, G., Platzer, S., Neudolt, A., Wunder, A., et al. (2023). The SPOC domain is a phosphoserine binding module that bridges transcription machinery with co- and post-transcriptional regulators. *Nat. Commun.* *14*, 166.
33. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* *596*, 583–589.
34. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* *596*, 590–596.
35. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., et al. (2014). A promoter-level mammalian expression atlas. *Nature* *507*, 462–470.
36. Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* *31*, 785–799.
37. Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* *273*, 5858–5868.
38. Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* *316*, 744–747.

39. Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22**, 908–917.
40. Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839–849.
41. Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A., et al. (2011). Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. *Science* **332**, 848–852.
42. Molaro, A., Falcatori, I., Hodges, E., Aravin, A.A., Marran, K., Rafii, S., McCombie, W.R., Smith, A.D., and Hannon, G.J. (2014). Two waves of de novo methylation during mouse germ cell development. *Genes Dev.* **28**, 1544–1549.
43. Manakov, S.A., Pezic, D., Marinov, G.K., Pastor, W.A., Sachidanandam, R., and Aravin, A.A. (2015). MIWI2 and MILI Have Differential Effects on piRNA Biogenesis and DNA Methylation. *Cell Rep.* **12**, 1234–1243.
44. Kloet, S.L., Baymaz, H.I., Makowski, M., Groenewold, V., Jansen, P.W., Berendsen, M., Niazi, H., Kops, G.J., and Vermeulen, M. (2015). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS Journal* **282**, 1774–1785.
45. Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. (2005). Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J. Biol. Chem.* **280**, 13341–13348.
46. Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). 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**, 2272–2280.
47. Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219–220.
48. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen—the Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242.
49. Di Giacomo, M., Comazzetto, S., Sampath, S.C., Sampath, S.C., and O'Carroll, D. (2014). G9a co-suppresses LINE1 elements in spermatogonia. *Epigenetics Chromatin* **7**, 24.
50. Pandey, R.R., Tokuzawa, Y., Yang, Z., Hayashi, E., Ichisaka, T., Kajita, S., Asano, Y., Kunieda, T., Sachidanandam, R., Chuma, S., et al. (2013). Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. *Proc. Natl. Acad. Sci. USA* **110**, 16492–16497.
51. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572.
52. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. j.* **17**, 3.
53. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.
54. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv.
55. Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987.
56. Magi, A., Tattini, L., Palombo, F., Benelli, M., Gialluisi, A., Giusti, B., Abbate, R., Seri, M., Gensini, G.F., Romeo, G., et al. (2014). H3M2: detection of runs of homozygosity from whole-exome sequencing data. *Bioinformatics* **30**, 2852–2859.
57. Scrucca, L., Fraley, C., Murphy, B.T., and Raftery, A.E. (2023). Model-Based Clustering, Classification, and Density Estimation Using mclust in R (Chapman and Hall/CRC).
58. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164.
59. Wilfert, A.B., Chao, K.R., Kaushal, M., Jain, S., Zöllner, S., Adams, D.R., and Conrad, D.F. (2016). Genome-wide significance testing of variation from single case exomes. *Nat. Genet.* **48**, 1455–1461.
60. Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641.
61. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview, Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.
62. Pei, J., and Grishin, N.V. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700–712.
63. Walsh, I., Martin, A.J., Di Domenico, T., and Tosatto, S.C. (2012). ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509.
64. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302.
65. Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.
66. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.
67. Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214.
68. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242.
69. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
70. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877.
71. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
72. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128.
73. Yariv, B., Yariv, E., Kessel, A., Masrati, G., Chorin, A.B., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2023). Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. *Protein Sci.* **32**, e4582.
74. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526.
75. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740.
76. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

77. Wickham, H., Vaughan, D., Girlich, M., and Ushey, K. (2023). tidy: Tidy Messy Data. <https://tidyr.tidyverse.org>.
78. Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. <https://dplyr.tidyverse.org/>.
79. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
80. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122.
81. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
82. Pemberton, T.J., and Szpiech, Z.A. (2018). Relationship between Deleterious Variation, Genomic Autozygosity, and Disease Risk: Insights from The 1000 Genomes Project. *Am. J. Hum. Genet.* **102**, 658–675.
83. Albert, S., Wistuba, J., Eildermann, K., Ehmcke, J., Schlatt, S., Gromoll, J., and Kossack, N. (2012). Comparative marker analysis after isolation and culture of testicular cells from the immature marmoset. *Cells Tissues Organs* **196**, 543–554.
84. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
85. Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H., and Kasukawa, T. (2017). FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Sci. Data* **4**, 170107.
86. Noguchi, S., Arakawa, T., Fukuda, S., Furuno, M., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Kaida, K., Kaiho, A., Kanamori-Katayama, M., et al. (2017). FANTOM5 CAGE profiles of human and mouse samples. *Sci. Data* **4**, 170112.
87. Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362.
88. Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670.
89. The UniProt Consortium (2017). UniProt: the universal protein knowledge-base. *Nucleic Acids Res.* **45**, D158–D169.
90. Hubner, N.C., and Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). *Methods* **53**, 453–459.
91. Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918.
92. Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370–1379.
93. Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-LINE1 ORF1p (for human samples)	abcam	Cat#ab245249; RRID: AB_2941773
Anti-LINE1 ORF1p (for mouse samples)	Di Giacomo et al. ⁴⁹	N/A
Anti-IAP GAG	B. Cullen, Duke University, Durham	N/A
Anti-gH2AX	Bethyl laboratories	Cat#IHC-00059; RRID: AB_533402
Rabbit anti-HA	Cell Signaling	Clone C29F4; RRID: AB_1549585
Mouse anti-HA	Cell Signaling	Clone 6E2; RRID: AB_10691311
Anti-Tubulin	Sigma Aldrich	Cat#T9026; RRID: AB_477593
ANTI-FLAG® M2	Sigma Aldrich	Cat#F3165; RRID: AB_259529
Anti-MIWI2	R. Pillai, University of Geneva Pandey et al. ⁵⁰	N/A
Anti-C19ORF84 rb632	This paper	N/A
Anti-C19ORF84 rb659	This paper	N/A
goat anti-rabbit biotin	Abcam	Cat#ab6012; RRID: AB_AB954766
IRDye 680RD donkey anti-rabbit	LI-COR	Cat#926-68073; RRID: AB_10954442
IRDye 800CW donkey anti-mouse	LI-COR	Cat#926-32212; RRID: AB_621847
donkey anti-rabbit ^{AlexaFluor568}	Invitrogen	Cat#A10042; RRID: AB_2534017
donkey anti-mouse ^{AlexaFluor647}	Invitrogen	Cat#A31571; RRID: AB_162542
anti-CD16/32	eBioscience	Clone 93; RRID: AB_467133
anti-CD45	eBioscience	Clone 30-F11; RRID: AB_466446
anti-CD51	Biolegend	Clone RMV-7; RRID: AB_466477
anti-CD9 ^{APC}	eBioscience	Clone eBioKMC8; RRID: AB_10669565
anti-cKit ^{PE-Cy7}	eBioscience	Clone 2B8; RRID: AB_647250
streptavidin ^{V450}	BD bioscience	RRID: AB_2033992
Anti-SPOCD1 rb175	O'Carroll group	N/A
Bacterial and virus strains		
<i>E. coli</i> 5-alpha	NEB	Cat#C2987H
<i>E. coli</i> XL1	O'Carroll group	XL-1
<i>E. coli</i> BL21	O'Carroll group	BL21
Biological samples		
Human DNA sample and testis biopsies	MERGE consortium	M3457
Human DNA sample and testis biopsies	MERGE consortium	M2021
Human DNA sample and testis biopsies	GEMINI consortium	GEMINI-88
Chemicals, peptides and recombinant proteins		
cOmplete ULTRA EDTA-free protease inhibitor	Roche	Cat#5892791001
Benzonase	Millipore	Cat#71206
Anti-HA magnetic beads	Thermo Fisher	Cat#88837
Anti-FLAG magnetic beads	Sigma Aldrich	Cat#M8823
Transfection (Jetprime)	Polyplus	Cat#101000015
3,3'-diaminobenzidine tetrahydrochloride	Merck	Cat#5905
Pefabloc	Merck	Cat#11429868001
Resource Q column	Cytiva	Cat#17117901
Dynabeads™ Protein G	Invitrogen	Cat#10004D

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dimethyl-pimelidate (DMP)	Thermo Fisher	Cat#21666
Amylose beads	NEB	Cat#E8021S
PhiX	Illumina	Cat#FC-110-3001
Critical commercial assays		
Click-iT TUNEL with Alexa Fluor 647	Invitrogen	Cat#C10247
NEBNext® Enzymatic Methyl-seq Kit	NEB	Cat#E7120L
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	NEB	Cat#E7760S
SureSelect ^{QXT} Target Enrichment kit	Agilent	Cat#G9683A
Twist Human Core Exome kit	Twist Bioscience	Cat#102027
SureSelect ^{XT} Human All Exon Kits V4, V5 and V6	Agilent	Cat#5190-4633, Cat#5190-6223, Cat#5190-8864
Human Core Exome plus RefSeq spike-ins	Twist Bioscience	Cat#102030
TruSeq SBS Kit v3 - HS	Illumina	N/A
NextSeq 500 V2 High-Output Kit	Illumina	N/A
HiSeq Rapid SBS Kit V2	Illumina	N/A
Glutathione Sepharose High Performance	Cytiva	Cat#17527902
16/600 Superdex200pg	Cytiva	Cat#28989335
16/600 Superdex75pg	Cytiva	Cat#28989333
Superdex 200 10/300 GL	Cytiva	Cat#17517501
NEBnext Multiplex Small RNA Library Prep Set for Illumina (NEB)	NEB	Cat#E7580
Deposited data		
EM-seq	This paper	ArrayExpress: E-MTAB-11612
RNA-seq	This paper; Zoch et al. ⁵	GEO: GSE199038
piRNA-seq	This paper; Zoch et al. ⁵	GEO: GSE199038, GSE131377
IP-MS	This paper	PRIDE: PXD047331
Human <i>SPOCD1</i> variants	This paper	ClinVar: SCV004098696, SCV004098697, SCV004098698
Mouse <i>SPOCD1</i> SPOC structure	This paper	PDB: 8OU1
Experimental models: Cell lines		
HEK 293T	O'Carroll lab stock	293T; RRID: CVCL_0063
Experimental models: Organisms/strains		
Mouse: Spocd1-HA	Zoch et al. ⁵	Spocd1-HA
Mouse: C19orf84-null	This paper	C19orf84-null
Mouse: Dnmt3c-HA	This paper	Dnmt3c-HA
Mouse: Dnmt3l-HA	This paper	Dnmt3l-HA
Mouse: Miwi2-HA	Vasiliauskaitė et al. ²⁸	Miwi2-HA
Oligonucleotides		
sgRNA C19orf84: GATGGACGAGCTAGAAGACG	This paper	N/A
sgRNA Dnmt3c: CTCCACAGACAACAATGAG	This paper	N/A
Donor oligo <i>Dnmt3c-HA</i> : GGTGAGCTGCTCTCATGTCCC TGTCTCCTTCTTCTCCTATCC ATTCTGGCCTTCTCCACAGACAA CAATGGGATCCTACCCATACGATGT TCCAGATTACGCTGGAGGCGGAGGA TCCAGGGGAGGTAGCAGACACCTCA GTAATGAGGAGGATGTCAGTGGATG TGAGGACTGTATTATCATCAGTGGGACCT	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
sgRNA Dnmt3l: TTCTAGCCGATTACATCAA	This paper	N/A
Donor oligo <i>Dnmt3l</i> -HA: GTCGCGTTTTAGGGTTCTGACGACCCTGCT GTCACACCCGCCATCCCTTGACGCAGAC CCTTCTAGCCGATTACATCAgcccaccATGgc gTCCTACCCATACGATGTTCCAGATTACG CTGGTTCCCGGGAGACACCTTCTTCTGCT CTAAGACCCTTGAACCTTGACCTGGA GACTTCCGACAGCTCTAGCCCTG	This paper	N/A
C19orf84-null WT-R: CATTACTCAAAGCCCCGTCTT	This paper	N/A
C19orf84-null NULL-F: GAGATGGAC GAGCTAGAAGCTA	This paper	N/A
C19orf84-null F: CTTTACTTTCAGGCCAGCC	This paper	N/A
C19orf84-null R: GGCTTGAAGTATCTTGTCTCAA	This paper	N/A
Dnmt3l-HA F: AGCCCTCTCTCTTCTATCCAA	This paper	N/A
Dnmt3l-HA R: AGAGCAAGAAGAAGGTGTCTCC	This paper	N/A
Dnmt3c-HA F: CTTTCTGAACACTACACAGAC	This paper	N/A
Dnmt3c-HA R: GAGCTGACTGTATAGTGAGAC	This paper	N/A
SPOCD1_Ex16_for: AGGCCCAAAGAGGAGTCAC	This paper	N/A
SPOCD1_Ex16_rev: CCAGATGACAGGAGGCCGAA	This paper	N/A
SPOCD1_Ex8_for: CTCTCAGGCCACCCACTCT	This paper	N/A
SPOCD1_Ex8_rev: TTTCCCTGAGCCCCAGTAAC	This paper	N/A
SPOCD1_Ex15_for: CAATGGCTGGCATGCAGTTC	This paper	N/A
SPOCD1_Ex15_rev: GCCTGGCTGGATATCTGGG	This paper	N/A
Recombinant DNA		
See Table S5 for plasmids.		N/A
Software and algorithms		
Bismark v0.22.1	Krueger and Andrews ⁵¹	https://github.com/FelixKrueger/Bismark
Cutadapt v1.15	Martin ⁵²	https://cutadapt.readthedocs.io/en/stable/
Bowtie2 v2.3.4.3	Langmead and Salzberg ⁵³	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
BWA Mem v0.7.17	Li ⁵⁴	https://bio-bwa.sourceforge.net/
GATK toolkit v3.8 HaplotypeCaller	Poplin et al. ⁵⁵	https://gatk.broadinstitute.org/hc/en-us
gnomAD database v2.1.1	Karczewski et al. ²⁶	https://gnomad.broadinstitute.org/
Picard	Broad Institute	http://broadinstitute.github.io/picard/
Genome Analysis Toolkit	Broad Institute	https://software.broadinstitute.org/gatk
H3M2 tool	Magi et al. ⁵⁶	https://sourceforge.net/projects/h3m2/
mclust R package v5.4.3	Scrucca et al. ⁵⁷	https://cran.r-project.org/web/packages/mclust/index.html
population sampling probability (PSAP) software	Wang et al. ⁵⁸ ; Wilfert et al. ⁵⁹	https://github.com/conradlab/PSAP
Clustal Omega	Madeira et al. ⁶⁰	https://www.ebi.ac.uk/Tools/msa/clustalo/
Jalview	Waterhouse et al. ⁶¹	https://www.jalview.org/
AL2CO	Pei and Grishin ⁶²	http://prodata.swmed.edu/al2co/al2co.php
Espritz	Walsh et al. ⁶³	http://old.protein.bio.unipd.it/espritz/
autoPROC v.1.0.5	Vonrhein et al. ⁶⁴	https://www.globalphasing.com/autoproc/
XDS (March 15, 2019, built 20191211)	Kabsch ⁶⁵	https://xds.mr.mpg.de/
pointless v1.11.21	Evans ⁶⁶	https://www.mrc-lmb.cam.ac.uk/harry/pre/pointless.html
Aimless v0.7.4	Evans and Murshudov ⁶⁷	https://www.ccp4.ac.uk/html/aimless.html

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CCP4 v7.0.078	Winn et al. ⁶⁸	https://www.ccp4.ac.uk/html/
SCALA v3.3.22	Evans ⁶⁶	https://www.mrc-lmb.cam.ac.uk/harry/pre/scala.html
PHASER v2.8.3.	McCoy et al. ⁶⁹	https://www.phaser.cimr.cam.ac.uk/index.php/Phaser_Crystallographic_Software
PHENIX v1.17.1_3660	Liebschner et al. ⁷⁰	https://phenix-online.org/
Coot v0.8.9.2	Emsley et al. ⁷¹	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/cool/
Py Mol, v2.5.4	Schrödinger, LLC	https://pymol.org/2/
APBS	Jurrus et al. ⁷²	https://github.com/Electrostatics/apbs-pdb2pqr
Consurf	Yariv et al. ⁷³	https://consurf.tau.ac.il
ImageStudioLite software	LICOR	https://www.licor.com/bio/image-studio-lite/
MaxQuant v1.6.1.0	Cox et al. ⁷⁴	https://www.maxquant.org/
Perseus v1.6.0.2	Tyanova et al. ⁷⁵	https://www.maxquant.org/perseus/
R v4.0.3 (2020-10-10)	The Comprehensive R Archive Network	https://cran.r-project.org/
ggplot2 v3.3.3	Wickham ⁷⁶	https://ggplot2.tidyverse.org/
tidyr v1.1.2	Wickham et al. ⁷⁷	https://tidyr.tidyverse.org/
dplyr v1.0.4	Wickham et al. ⁷⁸	https://dplyr.tidyverse.org/
Microsoft Excel for Mac (v16)	Microsoft	https://www.microsoft.com/en-gb/microsoft-365/p/excel/
DESeq2 v1.26.0	Love et al. ⁷⁹	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Custom code	This paper; Zoch et al. ⁵	https://doi.org/10.5281/zenodo.10509247

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dónal O'Carroll (donal.ocarroll@ed.ac.uk).

Materials availability

Mouse alleles and plasmids generated and used in this study are available from the lead author upon request.

Data and code availability

The described human *SPOCD1* variants were submitted to ClinVar under the accession numbers ClinVar: SCV004098696, SCV004098697 and SCV004098698. The coordinates of the mouse SPOC domain were submitted to the Protein Data Bank under accession code: PDB: 8OU1. The EM-seq data generated in this study have been deposited at ArrayExpress under the accession number ArrayExpress: E-MTAB-11612. The sRNA-seq and RNA-seq data generated in this study have been deposited at the Gene Expression Omnibus under the accession number GSE199038. Data for the IP-MS experiments were deposited at ProteomeXchange under the accession number PRIDE: PXD047331.

Code scripts used for the EM-seq, RNA-seq and sRNA-seq analysis are available on github (https://github.com/rberrens/SPOCD1-piRNA_directed_DNA_met) and a “version of record” archive has been deposited at Zenodo (<https://doi.org/10.5281/zenodo.10509247>).

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

In vivo mouse studies

Animals were maintained at the University of Edinburgh, UK, in accordance with the regulation of the UK Home Office. Ethical approval for the mouse experimentation has been given by the University of Edinburgh's Animal Welfare and Ethical Review Body and the work done under licence from the United Kingdom's Home Office. This sex and developmental time point used for each

experiment are stated in the figure legends. Detailed information on backcrossing status of each line is provided in the methods section.

Animals of the same sex from one or more litters per experiment were assigned to experimental groups according to genotype. No further randomisation or blinding was applied during data acquisition and analysis.

Human participants

All participants gave written informed consent for the evaluation of their clinical data and analysis of their DNA samples. The study protocol was approved by the respective Ethics Committees/Institutional Review Boards (Ref. No. Münster: 2010-578-f-S, GEMINI consortium: 201502059). Further information on the study cohorts and methodology is provided in the methods section.

Cell lines and bacteria strains

HEK 293T cells (RRID: CVCL_0063; female; sourced from the O'Carroll laboratory stock, University of Edinburgh, not further authenticated and regularly tested for mycoplasma contamination) were cultured at 37 °C, 5 % CO₂ in Glasgow minimum essential medium (Sigma Aldrich) supplemented with 10 % foetal calf serum (Gibco), 2 mM L-glutamine and 1 mM Na pyruvate (Invitrogen).

E. coli BL21 (DE3) cells (sourced from the O'Carroll laboratory stock, University of Edinburgh, and not further authenticated) were used to express recombinant proteins. Cells were grown at 37 °C in 2x TY media supplemented with kanamycin until OD₆₀₀ = 0.8 was reached. The temperature was then reduced to 18 °C and 1 mM IPTG was added to induce expression.

E. coli XL-1 (sourced from the O'Carroll laboratory stock, University of Edinburgh, not further authenticated) and 5-alpha (NEB) cells were used to produce plasmids. Cells were grown in LB medium or on LB agar plates supplemented with ampicillin.

METHOD DETAILS

Study cohorts

Exome data from 2,412 men included in the Male Reproductive Genomics (MERGE) study, comprising varying infertility phenotypes, were included in this study. The majority (n = 1,902) had pathomechanistically idiopathic azoo- or cryptozoospermia. Most of these men were recruited at the Centre of Reproductive Medicine and Andrology (CeRA), Münster. Exome data was likewise generated from the GEMINI cohort, which included 1,011 unrelated, infertile men recruited from eleven centres across seven countries. The vast majority of men represented in the GEMINI cohort presented with idiopathic azoospermia (n = 1,002) while 9 were severely oligozoospermic (sperm concentration <~1M/ml). Well-established causes for infertility including chromosomal aberrations and Y-chromosomal AZF microdeletions were excluded in all men.

All participants gave written informed consent for the evaluation of their clinical data and analysis of their DNA samples. The study protocol was approved by the respective Ethics Committees/Institutional Review Boards (Ref. No. Münster: 2010-578-f-S, GEMINI consortium: 201502059).

Age, at the time of initial presentation, and ancestry of the three individuals in the cohort carrying the *SPOCD1* allele variants analysed in this study were as follows: M3457: 29 year old male of Turkish ancestry; M2021: 31 year old male of German ancestry; GEMINI-88: Adult male from Australia with age and ancestry information not available.

Exome sequencing and bioinformatics analysis

MERGE cohort: Genomic DNA was extracted from peripheral blood leukocytes via local standard methods. WES sample preparation and enrichment were carried out in accordance with the protocols of either Agilent's SureSelect^{QXT} Target Enrichment kit or Twist Bioscience's Twist Human Core Exome kit. Agilent's SureSelect^{XT} Human All Exon Kits V4, V5 and V6 or Twist Bioscience's Human Core Exome plus RefSeq spike-ins were used to capture libraries. For multiplexed sequencing, the libraries were index tagged using appropriate pairs of index primers. Quantity and quality of the libraries were assessed with the ThermoFisher Qubit and Agilent's TapeStation 2200, respectively. Sequencing was conducted on the Illumina HiScan[®]SQ, NextSeq[®]500/550, or HiSeqX[®] systems using the TruSeq SBS Kit v3 - HS (200 cycles), the NextSeq 500 V2 High-Output Kit (300 cycles) or the HiSeq Rapid SBS Kit V2 (300 cycles), respectively. After trimming, Cutadapt v1.15⁵² was used to remove the remaining adapter sequences and primers. Sequence reads were aligned against the reference genome GRCh37.p13 using BWA Mem v0.7.17.⁵⁴ We excluded duplicate reads and reads that mapped to multiple locations in the genome from further analysis. Small insertions/deletions (indels) and single nucleotide variations were identified and quality-filtered by GATK toolkit v3.8 (<https://gatk.broadinstitute.org/hc/en-us>) with HaplotypeCaller⁵⁵ in accordance with the best practice recommendations. Ensembl Variant Effect Predictor was used to annotate called variants.⁸⁰ Exome data were screened focusing on rare biallelic variants in *SPOCD1* (minor allele frequency [MAF] < 0.01 in the gnomAD database [version v2.1.1],²⁶ which are predicted to affect protein function (stop-gain, frameshift and splice site variants as well as missense variants with a CADD score ≥ 20)).

GEMINI cohort (<https://gemini.conradlab.org/>): Whole-exome sequencing was performed for 1,002 unrelated NOA (non-obstructive azoospermia) cases at the McDonnell Genome Institute of Washington University (genome.wustl.edu) on Illumina HiSeq 4000 and using an in-house exome targeting reagent which captures 39.1 Mb of exome at an average coverage of 80x. The sequence reads were aligned to hg38 using bwa-mem,⁵⁴ Picard (<http://broadinstitute.github.io/picard/>) and Genome Analysis Toolkit⁸¹ (GATK; <https://software.broadinstitute.org/gatk>). Genotype calling was performed jointly for all samples using GATK tools and

following their procedures of best practices. All sequenced cases were screened for known infertility causes, including Klinefelter syndrome, deleterious CFTR mutations, Y-chromosome microdeletions and large structural variation on sex chromosomes utilising the WES genotype dataset.

To prioritise deleterious variation among the NOA cases, a modified version of the population sampling probability (PSAP) software^{58,59} (<https://github.com/conradlab/PSAP>) was applied to the WES genotype callset. The list of prioritized variants was subsequently filtered by only including variations with PSAP $P < 0.001$ and minor allele frequency $< 1\%$ across all populations in the gnomAD database (v2.1.1, <https://gnomad.broadinstitute.org/>)²⁶ and excluded variation common in the cohort.

To determine the degree of consanguinity, long runs of homozygosity (ROH) were detected using the H3M2 tool.⁵⁶ Only the longest class of ROH regions (lower boundary > 6.7 Mb on average across populations) reflecting recent inbreeding⁸² and detected with mclust R package (v.5.4.3)⁵⁷ was considered for calculating the fraction of the autosome being homozygous.

Sanger sequencing

Sanger sequencing of PCR amplicons of parts of *SPOCD1* GenBank: NM_144569.6 from genomic DNA from individuals M3457 (primers: *SPOCD1_Ex15_for* CAATGGCTGGCATGCAGTTC *SPOCD1_Ex15_rev* GCCTGGCTTGGATATCTGGG), M2021 (primers: *SPOCD1_Ex8_for* CTCTCAGGCCACCCACTCT & *SPOCD1_Ex8_rev* TTTCCCTGAGCCCCAGTAAC) and GEMINI-88 (*SPOCD1_Ex16_for* AGGCCCAAGAGGAGTAC & *SPOCD1_Ex16_rev* CCAGATGACAGGAGGCCGAA) confirmed the reported variants and when DNA was available from family members, segregation analysis was performed.

Histology of human testis samples

Testicular biopsies of patients and control subjects were surgically obtained from individuals during the process of testicular sperm extraction or histological evaluation at the Department of Clinical and Surgical Andrology (University Hospital Münster, Germany) and made available for research purposes. Testis biopsies of patient GEMINI-88 were obtained from Monash University, Melbourne. Biopsies were fixed in Bouin's solution overnight, washed with 70 % ethanol and embedded in paraffin. Subsequently, 5 μm sections were stained with periodic acid-Schiff (PAS) for histological evaluation.

For immunohistochemical analyses, 3 μm sections of testicular tissue were deparaffinised and rehydrated as described previously.⁸³ After rinsing with tap water (15 min, RT) heat-induced antigen retrieval was performed in Tris-EDTA buffer (pH 9). This step was followed by cooling and washing with 1X Tris-buffered saline (TBS) before endogenous peroxidase activity was blocked using 3 % hydrogen peroxide (15 min, RT), washed in distilled water followed by TBS and finally blocked against unspecific antibody binding with 25 % goat serum (#ab7481, abcam) in TBS containing 0.5 % bovine serum albumin (BSA, #A9647, Merck, 30 min, RT). Sections were incubated overnight at 4 °C in primary antibody solution, anti-LINE-1 ORF1p (#ab245249, abcam) diluted 1:10 in 0.5 % BSA/TBS. The following day, sections were washed with 1x TBS and incubated with a corresponding secondary antibody for 1 hour at room temperature (goat anti-rabbit biotin, #ab6012, abcam or goat anti-mouse biotin, #ab5886, abcam – both diluted 1:100 in 0.5 % BSA/TBS). After washing with TBS, sections were incubated for 45 min at room temperature with streptavidin-horseradish peroxidase (#S5512, Merck – diluted 1:500 in 0.5 % BSA/TBS). Subsequently, sections were washed with TBS and incubated with 3,3'-diaminobenzidine tetrahydrochloride (DAB, #D5905, Merck) for visualisation of antibody binding. Staining was validated by microscopical acquisition and stopped with aqua bidest. Counterstaining was conducted using Mayer's haematoxylin (#109249, Merck). Finally, sections were rinsed with tap water, dehydrated with decreasing ethanol concentrations and mounted using M-GLAS® mounting medium (#103973, Merck). Slides were digitalised using the Olympus BX61VS microscope and scanner software VS-ASW-S6.

Protein conservation and disorder analysis

Multiple sequence alignment was done as described previously.⁵ Protein sequences used for alignments are shown in [Data S2](#), [S3](#), and [S4](#). Alignment of SPOC domains was edited using published structures: SPOCD1 (structure resolved in this study: PDB: 8OU1), PHF3 (PDB: 6Q2V), Sharp (PDB: 1OW1) and RBM15 (PDB: 7Z27).

For protein conservation scores, a multiple sequence alignment of sequences retrieved by blastp⁸⁴ search against SPOCD1 or C19ORF84 was generated with Clustal Omega.⁶⁰ All gaps were removed in Jalview⁶¹ and the conservation score were calculated using AL2CO.⁶² Data were plotted in Microsoft Excel and the data were smoothed by applying a sliding window over 10 amino acids.

Ordered and disordered regions for SPOCD1 were predicted with ESpritz.⁶³ The data were then imported and plotted in Microsoft Excel.

AlphaFold2^{33,34} protein structure prediction models were downloaded from the AlphaFold Protein Structure Database (<https://www.alphafold.ebi.ac.uk/>).

Protein purification

The mouse SPOCD1 SPOC domain (amino acids 687 – 830) was cloned via ligase independent cloning into a pET-based vector with an amino-terminal His-GST tag followed by a 3C cleavage site (LEVLFGQP). The SPOC domain was expressed in *E. coli* BL21 (DE3) cells. 1 ml of overnight culture was used to inoculate 250 ml of 2x TY media supplemented with kanamycin. Cells were grown at 37 °C until $\text{OD}_{600} = 0.8$ was reached. The temperature was then reduced to 18 °C and 1 mM IPTG was added to induce expression. The bacteria were harvested after 16-18 hours and stored at -80 °C until needed. GST-SPOC domain was first purified using Glutathione

Sepharose High Performance (Cytiva) equilibrated with 20 mM Tris-Cl pH 7.5, 200 mM NaCl, 1 mM DTT. After elution with 20 mM Tris-Cl pH 7.5, 200 mM NaCl, 1 mM DTT, 20 mM reduced glutathione the protein was cleaved with 3C protease (homemade) overnight at 4 °C and dialysed in 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT. The protein was further purified by ion exchange (Resource S) eluting over a gradient up to 20 mM Tris-HCl pH 7.5, 1000 mM NaCl, and 1 mM DTT was run. The final purity was achieved by size exclusion chromatography using either a 16/600 Superdex200pg or a 16/600 Superdex75pg column equilibrated in 20 mM Tris-Cl pH 7.5, 150 mM NaCl, 1 mM DTT. The protein was concentrated and stored at -80°C until needed for experiments.

The mouse C19ORF84-1-100 and SPOCD1-900-1000 constructs were cloned into pET-based vector as N-terminal hexahistidine-MBP or as N-terminal hexahistidine-GFP fusions, respectively. Proteins were expressed in BL21 bacteria as described above. Cells were lysed in a cell disruptor with 50 ml lysis buffer (20 mM Tris pH 7.5, 200 mM NaCl, 5 mM Imidazol, 0.5 mM beta-mercaptoethanol, cOmplete ULTRA EDTA-free protease inhibitor (Roche), 20 µg/ml DNase I (Sigma D4527) and 480 µg/ml Pefabloc (Merck 11429868001)). The lysate was cleared by centrifugation at 45,000 g for 1 hour. The supernatant was collected and incubated for two hours with 5 ml Ni-NTA resin (Sigma Aldrich P6611) on a rotator at 4 °C. The Ni-NTA resin was washed the once with washing buffer (lysis buffer without protease inhibitor, Pefabloc and DNaseI), and then four times with ATP washing buffer (20 mM Tris pH 7.5, 300 mM KCl, 1 mM MgCl₂ 1 mM ATP, 0.5 mM beta-mercaptoethanol). Protein was eluted in 5 fractions of 10 ml (first four fractions with 250 mM imidazole in lysis buffer and last fraction with 500 mM imidazole in lysis buffer). The first three fractions were pooled and overnight dialysis performed in a dialysis buffer (100 mM NaCl, 20 mM Tris pH 7.5, 1 mM DTT). Next day, samples were further purified by Ion Exchange Chromatography with Resource Q column 6 ml (17117901-CYTIVA) using a salt gradient from low salt buffer (20 mM Tris pH 7.5, 100 mM NaCl, 1 mM DTT) to high salt buffer (20 mM Tris pH 7.5, 1000 mM NaCl, 1 mM DTT) and eluted with a concentration of 150 mM NaCl. Fractions of interest were aliquoted and immediately frozen in liquid nitrogen.

X-ray crystallography

The mouse SPOCD1 SPOC domain was concentrated to 9.5 mg/ml for crystallisation. Using sitting-drop vapour diffusion technique, the best diffracting crystals were grown in 1.6 M ammonium citrate tribasic, pH 7.5 at 293 K. Crystals were flash-cooled in liquid nitrogen prior to data collection. The data set was collected at the I04 (DLS, Didcot Oxfordshire) at 100 K using a wavelength of 0.97949 Å. The data were processed using autoPROC⁶⁴ (v.1.0.5) including XDS⁶⁵ (March 15, 2019, built 20191211), pointless⁶⁶ (1.11.21), Aimless⁶⁷ (0.7.4) and CCP4⁶⁸ (7.0.078). Scaling was carried out with SCALA⁶⁶ (3.3.22). The structure was determined by molecular replacement with PHASER⁶⁹ (v.2.8.3.) using the flowering protein FPA (PDB code: 5KXF) as a search model. The model was refined using PHENIX⁷⁰ (v.1.17.1_3660) and rebuilt with coot⁷¹ (v.0.8.9.2). Images were generated using PyMol (v.2.5.4, Schrödinger, LLC). The surface charge was calculated with APBS⁷² and the conservation was determined with ConSurf.⁷³ The coordinates were submitted to the Protein Data Bank under accession code: PDB: 8OU1. Data collection and structure refinement statistics shown in Table 1.

Cell culture, transfection, immuno-precipitation and western blotting

HEK 293T cells were cultured and transfected as previously described⁵ with minor alterations using plasmids and concentrations listed in Table S5. HEK 293T cells (sourced from the O'Carroll laboratory stock, University of Edinburgh; not additionally authenticated and regularly tested for mycoplasma contamination) were cultured at 37 °C, 5 % CO₂ in Glasgow minimum essential medium (Sigma Aldrich) supplemented with 10 % foetal calf serum (Gibco), 2 mM L-glutamine and 1 mM Na pyruvate (Invitrogen). For transfection 4 × 10⁵ cells were seeded per well on a 6-well plate on day 0, and on day 1, 50-1500 ng of each plasmid was transfected by Jetprime transfection (Polyplus) using 4 µl Jetprime reagent according to the manufacturer's instructions. After 48 hours, cells were washed twice with ice-cold PBS, scrapped off the plate in 1 ml lysis buffer (IP buffer: 150 mM KCl, 2.5 mM MgCl₂, 0.5 % Triton X-100, 50 mM Tris pH 8, supplemented with 1 × protease inhibitors (cOmplete ULTRA EDTA-free, Roche) and 37 U/ml benzonase (Millipore)) and lysed for 30 min at 4 °C on a rotating wheel. Lysates were cleared for 5 min at 21,000 g, and 400 µl of each supernatant was incubated for 1-2 hours at 4 °C with 20 µl of anti-HA beads (Pierce) or anti-FLAG beads (Sigma) that had been pre-washed twice in PBS, 0.5% Triton X-100 and resuspended in 500 µl lysis buffer. Immunoprecipitates were eluted for 10 min at 50 °C in 20 µl 0.1% sodium dodecyl sulphate (SDS), 50 mM Tris-Cl pH 8. Lysates and eluates were separated on a 4–12% Bis-Tris acrylamide gel (Invitrogen) and blotted onto nitrocellulose membrane (Amersham Protran 0.45 NC). The membrane was stained for protein with 0.1% (w/v) Ponceau S in 5% (v/v) acetic acid solution for 5 min, blocked with blocking buffer (4 % (w/v) skimmed milk powder (Sigma-Aldrich) in TBS-T (Tris buffered saline, 0.1 % Tween-20)), incubated for 1 hour with primary antibodies (anti-HA (C29F4 or 6E2, Cell Signaling Technologies) 1:1000; anti-FLAG (M2, Sigma-Aldrich) 1:1000, anti-C19ORF84 rabbit serum 1:500 (rb632 and rb659) or as loading control anti- α -Tubulin (T9026, Sigma-Aldrich) 1:1000) in blocking buffer, washed four times for 5 min in TBS-T, incubated with secondary antibodies (IRDye 680RD donkey anti-rabbit IRDye 800CW donkey anti-mouse, LI-COR, 1:10,000) in Immobilon Block – PO blocking solution (Millipore), washed four times for 5 min in TBS-T and imaged on a LI-COR Odyssey Fc or CLx system. Exposure of entire images was adjusted in Image Studio Lite (LI-COR) and regions of interest cropped for presentation.

Uncropped images of western blots are shown in Data S5.

Gene expression analysis

Gene expression data from the FANTOM5 project^{35,85,86} was extracted for mouse testis samples of the annotated developmental time-points using the FANTOM5 Table Extraction Tool (<https://fantom.gsc.riken.jp/5/tet/#/>) on the 'CAGE peak level expression (RLE normalized) of robust phase 1 and 2 CAGE peaks for mouse samples with annotation (mm9)' dataset. Signal intensity for all annotated CAGE

peaks mapping to the 5' region of *C19orf84* (2 peaks originally assigned to Lim2 (ENSMUST00000164977): chr7:50683014..50683022,+; chr7:50683034..50683049,+), *Spocd1* (2 peaks assigned to uniprot ID B1ASB6: chr4:129606011..129606014,+; chr4:129606033..129606044,+), *Piwil4* (4 peaks assigned to ENSMUST00000164977: chr9:14545062..14545068,-; chr9:14545086..14545097,-; chr9:14545110..14545121,-; chr9:14545135..14545140,-) or *Dnmt3l* (4 peaks assigned to ENSMUST00000000746 or ENSMUST00000151242: chr10:77512581..77512592,+; chr10:77512605..77512612,+; chr10:77512624..77512641,+; chr10:77512679..77512707,+) was summed and normalized per gene to the timepoint of peak gene expression.

Antibody generation

Antibodies against mouse C19ORF84 were made with a commercial producer (Biotem). Two anti-C19ORF84 sera (rb632 and rb659) were raised in New Zealand white rabbits by immunising with a pool of two peptides (KLH-peptides C+SQRGP ERAEERER NMAGE and C+SQDQGKEAGGLSEADWEADY). The final bleed serum was used for all experiments.

Immunoprecipitation and mass spectrometry (IP-MS)

For anti-C19ORF84 IP-MS experiments protein G dynabeads (Thermo Fisher) were crosslinked with either anti-C19ORF84 rb659 serum or control commercial rabbit serum (Sigma Aldrich) in a 2:3 serum to beads ratio with 20 mM DMP (Thermo Fisher) in borate buffer pH 9 (0.25 g boric acid, 1.53 g sodium tetraborate decahydrate in 100 ml ddH₂O). IP was performed from 25 wild-type E16.5 foetal testes per replicate. 150 testes were pooled and lysed in hypotonic lysis buffer (10 mM Tris-HCl pH 8, 10 mM KCl, 5 mM MgCl₂, 0.1% IGEPAL CA-630, cOmplete™ protease inhibitor EDTA-free (Roche)) with 20 strokes in a glass douncer. Lysates were further incubated for 30 min at 4 °C after addition of 50 U/ml benzonase (Millipore). Lysate was then cleared for 5 min at 21,000 g and divided equally onto 3 x 50 µl cross-linked anti-C19ORF84 beads and 3x 50 µl cross-linked rabbit serum beads. Immuno-precipitation and mass spectrometry analysis then proceeded as previously described⁵: Briefly, the beads were washed 4 times with wash buffer (100 mM KCl, 5 mM MgCl₂, 0.1 % IGEPAL, 50 mM Tris pH 8) and proteins eluted in 0.1 % Rapigest (Waters), 50 mM Tris pH 8. Eluate was further processed as described,⁸⁷ followed by desalting using STAGE tips⁸⁸ and finally resuspended in 0.1 % tri-fluoro-acetic acid (v/v) for LC-MS. Peptides were analysed on a Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher) operated in data-dependent acquisition mode, after separation through nano-flow liquid chromatography on a nanoLC Ultimate 3000 unit fitted with an EasySpray (50 cm, 2 µm particles) column. Samples were separated using a 2 % - 40 % - 95 % 190 min gradient (mobile phase A - 0.1 % aqueous formic acid, B - 80 % acetonitrile in 0.1 % formic acid). Acquisition parameters were set to cycle time 3 s, MS1 scan Orbitrap resolution to 120,000, RF lens to 30 %, AGC target to 4.0e5, and maximum injection time to 50 ms. Detected intensity threshold was 5.0e3, MS2 scan was performed with the Ion Trap using rapid scan setting, the AGC target was set to 2.0e4, and maximum injection time was 50 ms. MaxQuant version 1.6.1.0 was used to process the raw data, and label-free quantification (LFQ) was performed using the MaxQuant LFQ algorithm.⁷⁴ Peptides were searched against the mouse UniProt database (date 21.07.2017) with commonly observed contaminants (e.g. trypsin, keratins, etc.) removed during Perseus analysis.^{74,75,89} LFQ intensities were visualised with Perseus version 1.6.0.2.^{75,90}

Anti-HA IP-MS experiments for HA-DNMT3L and HA-DNMT3C were performed from 50 E16.5 foetal testes (*Dnmt3l*^{HA/+} or *Dnmt3c*^{HA/+}, respectively as well as wild-type controls) per replicate as previously described.⁵ In brief, 50 testes were lysed in 1 ml hypotonic lysis buffer and treated with 50 U/ml benzonase, cleared for 5 min at 21,000 g and added to 50 µl cross-linked anti-HA magnetic beads (Pierce). Immuno-precipitation and mass spectrometry analysis then proceeded as described above.

Anti-HA IP-MS experiment for HA-SPOCD1 in the presence or absence of C19ORF84 was performed as described above for HA-DNMT3L/HA-DNMT3C using 25 of either *Spocd1*^{HA/+}; *C19orf84*^{-/-} or *Spocd1*^{HA/+}; *C19orf84*^{+/-} control E16.5 foetal testes per replicate.

Statistically significant ($P < 0.05$) enriched (Enrichment > 4-fold) proteins are shown in [Tables S1–S4](#).

Pull-down assay

3 µg of recombinant MBP tagged mouse C19ORF84 (amino acids 1-100) or 3 µg His-MBP was mixed with the 6 µg GFP-tagged recombinant mouse SPOCD1 (amino acids 900-1000) fragment in binding buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.01 % Tween-20, 1 mM DTT) to a final volume of 100 µl and incubated for 10 min at 4 °C. 80 µl of the mixture was incubated with 50 µl of amylose beads (NEB E8021S) for 1.5 hours at 4 °C on the rotator. The resin was washed five times with 1 ml of binding buffer and eluted with 30 µl of LDS 4x Sample Buffer (NuPage, NP0007) with 100 mM DTT at 95 °C for 5 minutes. 17 µl from the input and 25 µl from the IP were loaded on a gradient from 4% to 12%, Bis-Tris SDS-PAGE gel (Invitrogen, NP0321BOX) with a protein molecular weight marker. Proteins were visualized by Coomassie brilliant blue staining.

Analytical size exclusion chromatography

For the analytical size exclusion chromatography (SEC), 98.6 µg mouse MBP-C19ORF84-1-100 and/ or 204.6 µg mouse GFP-SPOCD1-900-1000 were used in each run. Proteins were diluted in a final volume of 250 µl SEC buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT) and injected on a Superdex 200 10/300 GL column. Peak fractions were collected, loaded on an SDS-PAGE and visualized by Coomassie staining.

Mouse strains and experimentation

Generation of the *Spoed1^{HA}* and *Miw12^{HA}* alleles has been described previously^{5,28} and the lines were kept on a mixed B6CBAF1/Crl; C57BL/6N;Hsd:ICR (CD1) genetic background.

The *C19orf84⁻* allele was generated through CRISPR-Cas9 gene editing as previously described^{91,92} by injection of a single sgRNA (GATGGACGAGCTAGAAGACG) together with CAS9 mRNA into the cytoplasm of fertilised 1-cell zygotes of B6CBAF1/Crl genetic background. The *Dnmt3l^{HA}* allele was generated through CRISPR-Cas9 gene editing using a single sgRNA (TTCTAGCCGA TTACATCAA), a ssDNA donor (GTCGCGTTTTAGGGTTCTGACGACCCTGCTGTACACCCGCCATCCCTTGGACGCAGACCCTTCT AGCCGATTACATCAgcccaccATGgcgTCCTACCCATACGATGTTCCAGATTACGCTGGTTCCTGGGAGACACCTTCTTCTTGC TCTAAGACCCTTCAAACCTTGGACCTGGAGACTTCCGACAGCTCTAGCCCTG) and CAS9 mRNA. For generation of the *Dnmt3c^{HA}* allele we used the sgRNA (CTCCACAGACAACAATGAG) and ssDNA donor (GGTGAGCTGCTCTCTCATGTCCCTGTCTCCTTCTT CCTTCTATCCATTCTGGCCTTCTCCACAGACAACAATGGGATCCTACCCATACGATGTTCCAGATTACGCTGGAGGCGGAGGA TCCAGGGGAGGTAGCAGACACCTCAGTAATGAGGAGGATGTCAGTGGATGTGAGGACTGTATTATCATCAGTGGGACCT).

F₀ offspring were genotyped by PCR and Sanger sequencing and the *Dnmt3l^{HA}*, *Dnmt3c^{HA}*, *C19orf84⁻* alleles were established from one founder animal and back-crossed several times to a C57BL/6N genetic background. Thus, the *Dnmt3l^{HA}*, *Dnmt3c^{HA}*, *C19orf84⁻* mice were on a mixed B6CBAF1/Crl; C57BL/6N genetic background. *C19orf84⁻* animals were genotyped with a four primer PCR (F: CTTTACTTTTCAGGCCAGCC, R: GGCTTGAAGTATCTTGCTCAA, WT-R: CATTACTCAAAGCCCCGTCTT, NULL-F: GAGATGGACGAGCTAGAAGCTA). *Dnmt3l^{HA}* animals were genotyped with a two primer PCR (F: AGCCCTCTCTCT TCCTATCCAA, R: AGAGCAAGAAGAAGGTGTCTCC). *Dnmt3c^{HA}* animals were genotyped with a two primer PCR (F: CTTTC CTGAACACTACACAGAC, R: AGAGCTGACTGTATAGTGAGAC).

Assessment of male fertility was done by mating studs to Hsd:ICR (CD1) wild-type females counting the number of embryos at E16.5 for each plugged female.

Animal tissue samples were collected from one or more litters per experiment and allocated to groups according to genotype. No further randomisation or blinding was applied during data acquisition and analysis.

Animals were maintained at the University of Edinburgh, UK, in accordance with the regulation of the UK Home Office. Ethical approval for the mouse experimentation has been given by the University of Edinburgh's Animal Welfare and Ethical Review Body and the work done under licence from the United Kingdom's Home Office.

Immuno-fluorescence and foci count

Immuno-fluorescence experiments were done as previously described⁵ with the following primary antibodies incubated overnight in blocking buffer (10 % donkey serum, 1% BSA, 100 mM glycine in PBS): anti-HA (C29F4, Cell Signaling Technologies) 1:100 (HA-DNMT3C) or 1:500 (HA-DNMT3L); anti-HA (6E2, Cell Signaling Technologies) 1:1000 1.1 mg/ml in PBS custom formulation for HA-DNMT3C and 1:200 12 µg/ml catalogue formulation for all other HA tagged proteins; anti-LINE1-ORF1p⁴⁹ 1:500; anti-IAP-GAG (a kind gift from B. Cullen, Duke University, Durham, NC, USA) 1:500; anti-γH2AX (IHC-00059, Bethyl Laboratories) 1:500; anti-C19ORF84 rabbit serum (rb632 used for all experimental immuno-fluorescence experiments and rb659 where indicated in [Figure S4C](#)) 1:500; anti-SPOCD1 rabbit serum (rb175, raised against the mouse SPOCD1 TFISM domain, amino acids 407-568) 1:500; anti-MIW12⁵⁰ (a kind gift from Ramesh Pillai, Université de Genève, Switzerland) 1:500. Sections were then stained with 5 µg/ml DAPI and the appropriate donkey anti-rabbit^{AlexaFluor568} or donkey anti-mouse^{AlexaFluor647} antibody diluted 1:1000 in blocking buffer and mounted on coverslips with Prolong Gold (Invitrogen). Images were acquired on a Zeiss Observer or Zeiss LSM880 with Airyscan module. Airyscan images were deconvoluted with the Zeiss Zen software "Airyscan processing" with settings "3D" and a strength of 6. Images were processed and analysed with ImageJ and Zeiss Zen software.

Germ cells containing C19ORF84 foci and number thereof as well as SPOCD1 and HA-DNMT3C foci in *C19orf84^{+/+}* and *C19orf84^{-/-}* germ cells were counted manually through the eyepiece at the Zeiss Airyscan 880 microscope using the 100x objective.

Foetal testes extract preparation for western blotting

Foetal testis protein extracts for western blot analysis were made by lysing a pair of E16.5 foetal testes of the indicated genotype in RIPA buffer (150 mM NaCl, 50 mM Tris pH 8, 1 % IGEPAL, 0.5 % deoxycholate, 0.1 % SDS, cOmplete ULTRA EDTA-free protease inhibitor (Roche)) using micropestles for tissue homogenization (Sigma Aldrich). Lysates were then sonicated on a bioruptor pico sonicator (diagenode) for 10 cycles using the default 30 s ON / 30 s OFF programme. SDS loading buffer was added to a final concentration of 2 % SDS, 100 mM Tris pH 8, 8.3 mM DTT, samples heated to 95°C for 5 min and lysates cleared by centrifugation for 5 min at 21,000 g before being subjected to Western blot analysis as described above.

Signal intensity of anti-HA stained bands of interest were quantified using ImageStudioLite software (LICOR).

Uncropped images of western blots are shown in [Data S5](#).

Histology of mouse samples

Histology experiments on mouse samples were done as previously described.⁵ Briefly, Bouin's fluid (Sigma Aldrich) fixed testes were embedded in paraffin, sectioned to 3 µm and deparaffinised in a graded alcohol series. Sections were stained with Periodic Acid Schiff Stain Kit (CellPath). After dehydration through a reverse alcohol series, the sections were embedded in Pertex (Pioneer Research Chemicals).

Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay

TUNEL assay experiments were done as previously described.⁵ Briefly, Bouin's fluid (Sigma Aldrich) fixed and paraffin embedded adult testes were sectioned at 5 μm , rehydrated using a decreasing alcohol series, and TUNEL stained with the Click-iT TUNEL assay, Alexa Fluor 647 dye (Invitrogen) according to the manufacturer's instructions after pretreatment with proteinase K (10 $\mu\text{g}/\text{ml}$ in 10 mM Tris pH 8; Thermo Scientific). Sections were then stained with DAPI (1 $\mu\text{g}/\text{ml}$) and embedded in ProLong Gold Antifade Mountant (Invitrogen).

RNA sequencing and analysis

RNA sequencing experiments and analysis were done as previously described⁵ with data for *Spocd1*^{-/-}, *Miwi2*^{-/-} and wild-type samples retrieved from GEO: GSE131377.⁵ Briefly, we extracted total RNA from P20 testes with Qiagen RNeasy Kit according to the manufacturer's instructions and prepared libraries for RNA-seq using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. Libraries were sequenced on Illumina NextSeq 500 in 150 bp single-read mode.

Adapter sequences were removed from reads using cutadapt (1.8.1)⁵² with default settings and processed reads then mapped to the consensus sequence of rodent retrotransposons annotated by Repbase (24.01)⁹³ using bowtie2 (2.3.4.3)⁵³ with default settings. Mapped reads per retrotransposon were counted and significantly de-regulated species were analysed using DESeq2 (1.26.0).⁷⁹ Data was plotted in R studio.

Fluorescence-activated cell sorting (FACS)

CD9⁺ spermatogonia were sorted from P14 testes as previously described²⁸ with minor alterations: Testes were dealbuginated and digested with 0.83 mg/ml collagenase (Sigma-Aldrich), 50 $\mu\text{g}/\text{ml}$ DNase (Sigma-Aldrich) in goni-mem (DMEM (Life Technologies) supplemented with penicillin-streptomycin (Life Technologies), NEAA (Life Technologies), sodium pyruvate (Life Technologies) and sodium lactate (Sigma-Aldrich)) for 30 min at 32 °C 1000 rpm, followed by digestion with 0.05 % Trypsin (Gibco), 50 $\mu\text{g}/\text{ml}$ DNase in goni-mem. Fetal calf serum together with 100 $\mu\text{g}/\text{ml}$ DNase I (Sigma-Aldrich) was added to stop the trypsin digestion, and to facilitate complete degradation of released DNA the cell suspension was further incubated for 3 min at 32 °C. Cells were resuspended in 500 $\mu\text{g}/\text{ml}$ DNase in goni-mem and then blocked with Fc block (anti-CD16/32, clone 93, eBioscience, 1:50; then labelled with anti-CD45 (clone 30-F11, eBioscience, 1:400) and anti-CD51 (clone RMV-7, Biolegend, 1:100) biotin-conjugated antibodies. Cells were subsequently stained with anti-CD9^{APC} (clone eBioKMC8, eBioscience, 1:200), anti-cKit^{PE-Cy7} (clone 2B8, eBioscience, 1:1600), streptavidin^{V450} (BD bioscience, 1:400) together with 1 $\mu\text{g}/\text{ml}$ DAPI and finally sorted into goni-mem on a BD Aria II sorter. Sorted cells were pelleted for 5 min at 500 g and snap frozen in liquid nitrogen.

Whole-genome methylation sequencing and analysis

Whole genome methylation sequencing experiments and analysis were done as previously described⁵ with data for *Spocd1*^{-/-}, *Miwi2*^{-/-} and wild-type samples retrieved from ArrayExpress: E-MTAB-7997.⁵ Briefly, genomic DNA was extracted from sorted cells by overnight proteinase K digest (10 mM Tris-HCl pH 8, 5 mM EDTA, 1 % SDS, 0.3 M Na-acetate, 0.2 mg/ml proteinase K), followed by two rounds of phenol/chloroform/isoamylalcohol (25:24:1, Sigma-Aldrich) extraction, one round of chloroform extraction and purification through 2-propanol precipitation in 300 mM Na-acetate pH 5, 10 $\mu\text{g}/\text{ml}$ linear acrylamide (Invitrogen) followed by two washes in 70 % ethanol. Genomic DNA was then further processed using the NEBNext® Enzymatic Methyl-seq Kit (EM-seq; NEB) and sequenced on Illumina NextSeq in 150 bp paired-end read mode.

Raw sequence reads were trimmed using Trim Galore (v0.4.1, www.bioinformatics.babraham.ac.uk/projects/trim_galore/), Cutadapt version 1.8.1,⁵² parameters: -paired -length 25 -trim-n -clip_R2 5) and then aligned to the mouse genome (mm10) in paired-end mode with Bismark v0.22.1⁵¹ (parameters: bismark -score_min L,0,-0.4 -paired). CpG methylation calls were extracted from deduplicated output using the Bismark methylation extractor (v0.22.1).⁵¹ Mapping statistics were calculated with SeqMonk datastore summary report (www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Sequencing statistics are shown in Table S6.

Probes were generated from 50 adjacent CpG running windows containing at least 10 reads. We defined genome features as follows: Genic regions as probes overlapping genes, promoters as probes overlapping 2000 bp upstream of annotated transcripts (Ensembl (GRCm38.p6)), CpG islands (CGIs) as probes overlapping the Ensembl (GRCm38.p6) CGI annotation. Genic, promoters and CGIs regions overlapping transposons were filtered out. Transposons were annotated using the UCSC repeat masker annotation (<https://genome.ucsc.edu/cgi-bin/hgTables>, 02/2019). We excluded simple repeats as well as any small non-coding RNA annotations from the transposon list. Transposon analysis was done using only unique mapping reads and excludes elements overlapping gene bodies. Transposon families were assessed by analysis of full length elements (defined as > 5 kb for LINE1 elements, > 6 kb for IAP families and > 4.5 kb for MMERVK10C). Intergenic regions were defined as regions non-overlapping genes or transposons. Meta-plots, scatterplots and correlation analysis were done using SeqMonk and RStudio. The methylation difference analysis was performed using the divergence (milliDiv) from the consensus sequences. Annotation of imprinted control regions (ICR) was obtained from (<https://atlas.genetics.kcl.ac.uk/>). For CpG methylation of the *Rasgrf1* imprinted region we quantified individual CpGs with a minimum of 1 read mapping. SeqMonk and RStudio were used for graphing and statistics.

Small RNA sequencing and analysis

Small RNA sequencing experiments were done as previously described⁵ with minor alterations to the library preparation: RNA was precipitated overnight at -20 °C in 2.5 volumes 100% ethanol and 2 μl linear acrylamide (E7325AVIAL, NEB), washed with 80%

ethanol and dissolved in 10 μ l nuclease-free water. For library generation, the NEBnext Multiplex Small RNA Library Prep Set for Illumina (NEB) was used following the manufacturer's instructions with 5 μ l size-selected RNA per reaction, adaptors diluted 1:2 and 15 cycles of PCR amplification. The finished library was size-selected using a 6% TBE gel (Invitrogen) and DNA purified from the gel by addition of nuclease-free water and two successive 1 hour incubation steps at 37 °C, 1000 rpm, with a freeze-thaw step in between. Samples were then transferred onto spin columns (Corning) plugged with filter paper (Whatman) and centrifuged at maximum speed for 1 min. The DNA was then precipitated as described above with addition of 1 μ l linear acrylamide. Concentration was measured on a Qubit fluorometer (Life Technologies) using the Qubit high-sensitivity dsDNA kit, and the library was quality-controlled with an HSD1000 tape on a TapeStation 2200 instrument (Agilent). Equal nmole amounts of each library were pooled and diluted with RSB (Illumina) to 1 nM following the manufacturer's recommendations for MiniSeq (Illumina) before being chemically denatured using 0.1 M NaOH for 5 min at room temperature followed by the addition of 200 mM Tris-HCl, pH 7 to stop the reaction. The denatured library pool was diluted in hybridization buffer (Illumina) to 1.35 pM and mixed at equal concentration with 10 % PhiX (Illumina), denatured separately as described above. The library-PhiX mix was loaded onto a MiniSeq High Output Reagent Cartridge (75 cycles) and sequenced in 60-base single-end read mode.

Analysis of sequencing data was then performed as previously described.⁵

QUANTIFICATION AND STATISTICAL ANALYSIS

Data were plotted in R (version 4.0.3 (2020-10-10)) using the ggplot2,⁷⁶ tidyr⁷⁷ and dplyr⁷⁸ toolkits (versions ggplot2_3.3.3, tidyr_1.1.2, dplyr_1.0.4) 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 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 \pm s.e.m. (standard error of the mean) for comparisons or mean \pm SD (standard deviation) for descriptive statistics, unless otherwise indicated. No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.