SI Appendix

MRG15 is required for pre-mRNA splicing and spermatogenesis

Naoki Iwamori[a,b,c,d], Kaoru Tominaga[e], Tetsuya Sato[f], Kevin Riehle[g,h], Tokuko Iwamori[a,b,c], Yasuyuki Ohkawa[i], Cristian Coarfa[h,j], Etsuro Ono[a,b], and Martin M. Matzuk[c,d,g,j,k,l]

[a]Laboratory of Biomedicine, Division of Pathobiology, Department of Basic Medicine, Faculty of Medicine, Kyushu University, Fukuoka 812-8582, Japan; [b]Center of Biomedical Research, Research Center for Human Disease Modeling, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan; [c]Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX 77030; [d]Center for Reproductive Medicine, Baylor College of Medicine, Houston, TX 77030; [e]Division of Functional Biochemistry, Department of Biochemistry, Jichi Medical University, Tochigi 329-0498, Japan; [f]Division of Bioinformatics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan; [g]Department of Molecular Human Genetics, Baylor College of Medicine, Houston, TX 77030; [h]Bioinformatics Research Laboratory, Baylor College of Medicine, Houston, TX 77030; [i]Division of Transcriptomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan; [j]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030; [k]Department of Pharmacology, Baylor College of Medicine, Houston, TX 77030; and [l]Center for Drug Discovery, Baylor College of Medicine, Houston, TX 77030

To whom correspondence should be addressed. E-mail: iwamori@anim.med.kyushu-u.ac.jp or mmatzuk@bcm.edu

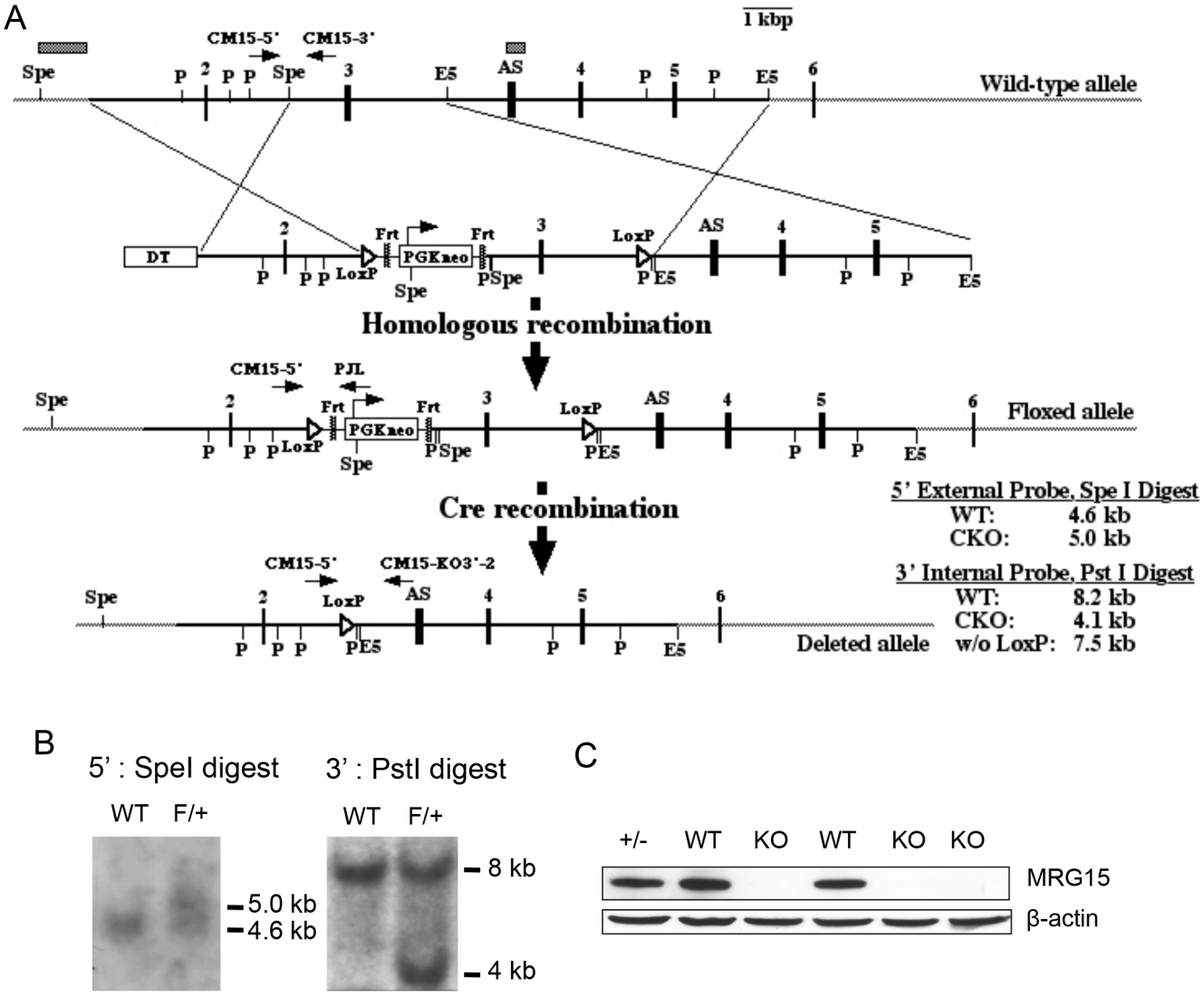

Fig. S1. Generation of the *Mrg15* targeted allele. A. Strategy of targeting conditional allele of the *Jmjd3* gene. An *Frt - PGK1-Neo – Frt - loxP* cassette and a *loxP* sequence were inserted between exons 2 and 3 and between exons 3 and 4, respectively. DT, diphtheria toxin fragment A. B. Southern blot analyses using 5' (left panel) and 3' (right panel) external probes. SpeI (Left) and PstI (Right) digested ES cell DNA were hybridized with probes that detect a 5.0 kb WT and 4.6 kb targeted alleles at the 5' end (Left), or a 8.2 kb WT and a 4.1 kb targeted alleles at the 3' end (Right). C. Confirmation of the *Mrg15* deletion by western blot. Proteins extracted from mouse embryonic fibroblast (MEF) of each genotype were analyzed with anti-MRG15 antibody.

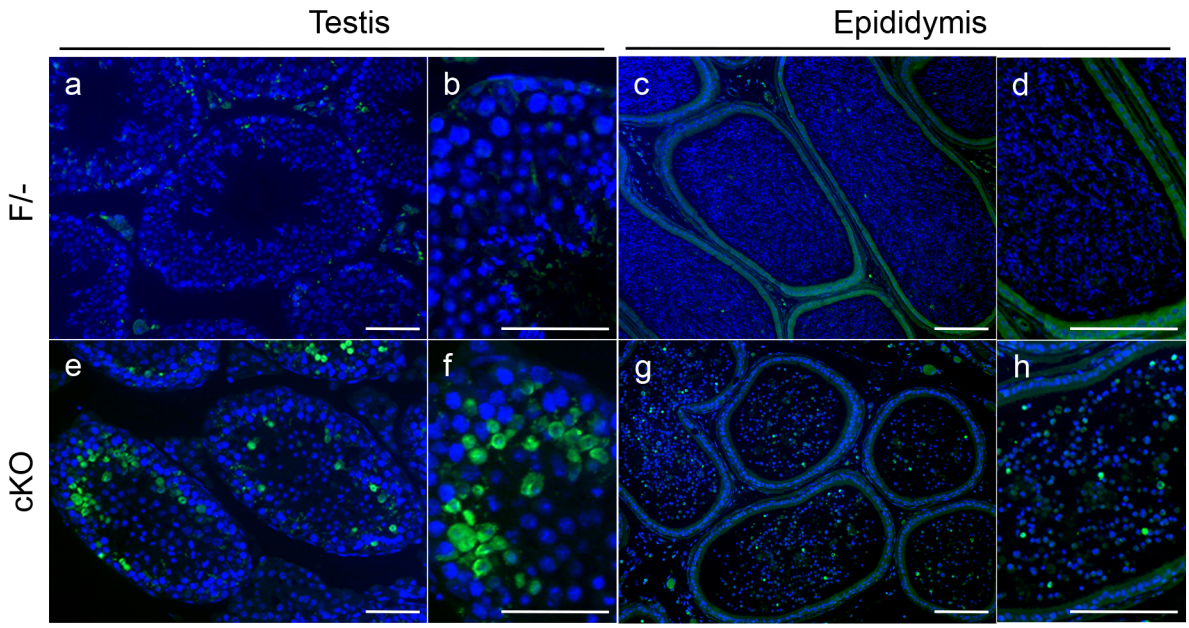

Fig. S2. Apoptotic cell death in *Mrg15* null testes. Detection of apoptotic cells in the testes. DNA fragmentation in the testes (a, c, e, g) and the epididymis (b, d, f, h) of MRG15 F/- (a, b, e, f) and cKO (c, d, g, h) mice was visualized by TUNEL labeling assay. Scale bar: 25 μm.

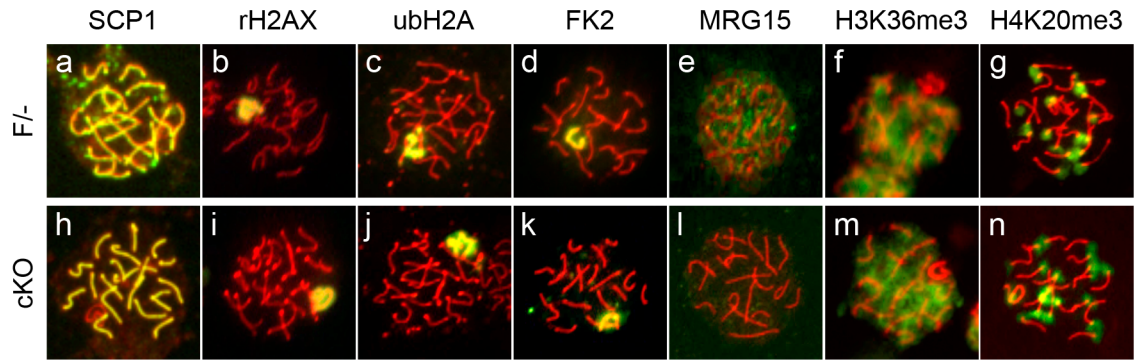

Fig. S3. Analyses of meiotic homologous recombination. MRG15 F/- (a-g) and cKO (h-n) pachytene spermatocytes were stained with anti-SCP1 (a, h), anti-rH2AX (b, i), anti-ubH2A (c, j), anti-FK2 (d, k), anti-MRG15 (e, l), anti-H3K36me3 (f, m), or anti-H4K20me3 (g, n) (green) and anti-SCP3 (red) antibodies. Merged images are shown.

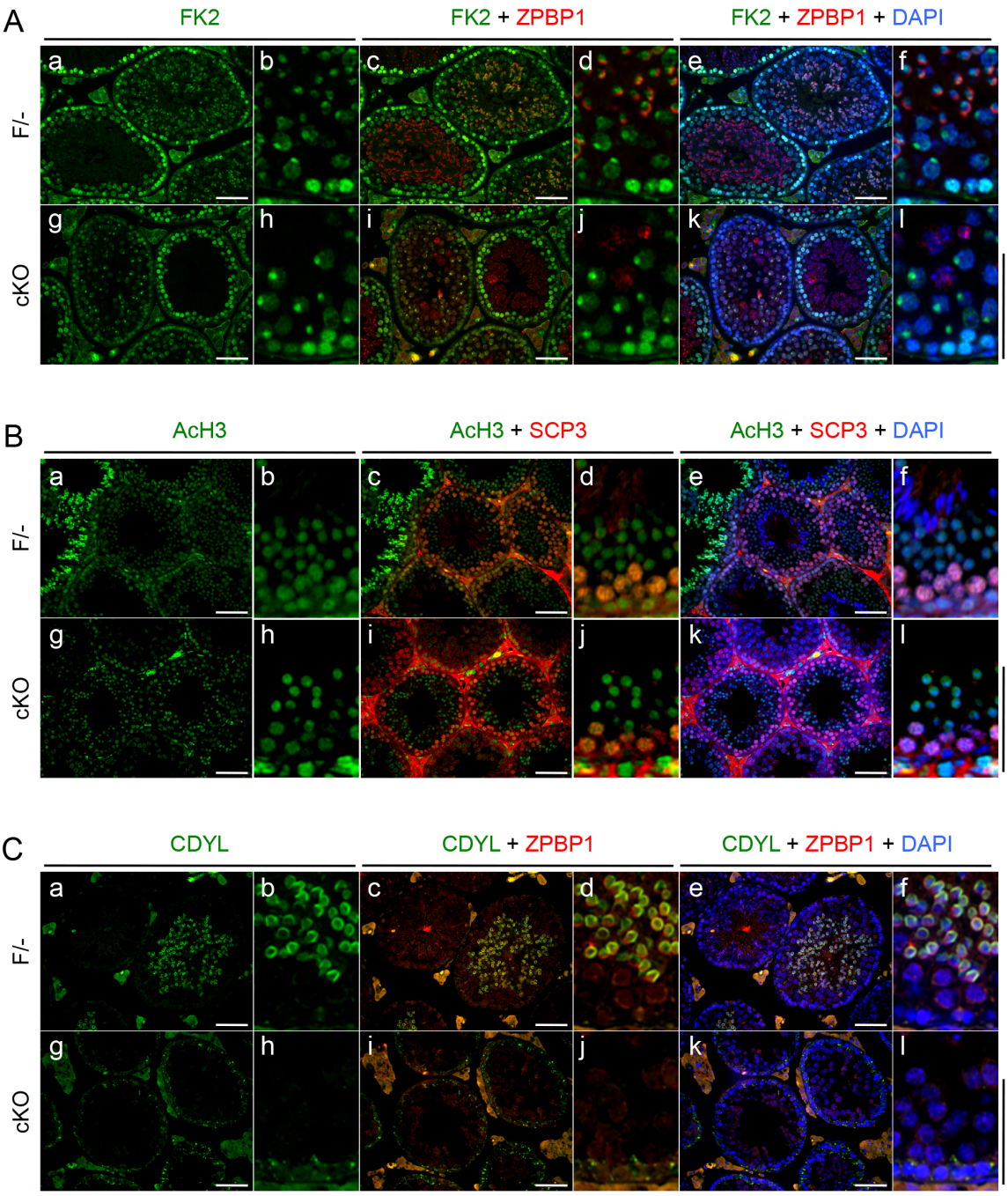

Fig. S4. Testicular distribution of meiosis and histone acetylation related proteins. Testicular distribution of (A) ubiquitin (FK2: green) and ZPBP1 (red), (B) acetylated histone H3 (green) and SCP3 (red), and (C) CDYL (green) and ZPBP1 (red) are shown. The nuclei were counter stained with DAPI (blue). Pictures of green signals (a, b, g, h), merged with red signals (e, d, I, j), and merged with red and blue signals (e, f, k, l) of MRG15 F/- (a-f) and cKO (g-l) are shown. Scale bar: 50μm. Scale bar for high magnification images (b, d, f, h, j, l) were located on the right (black bar).

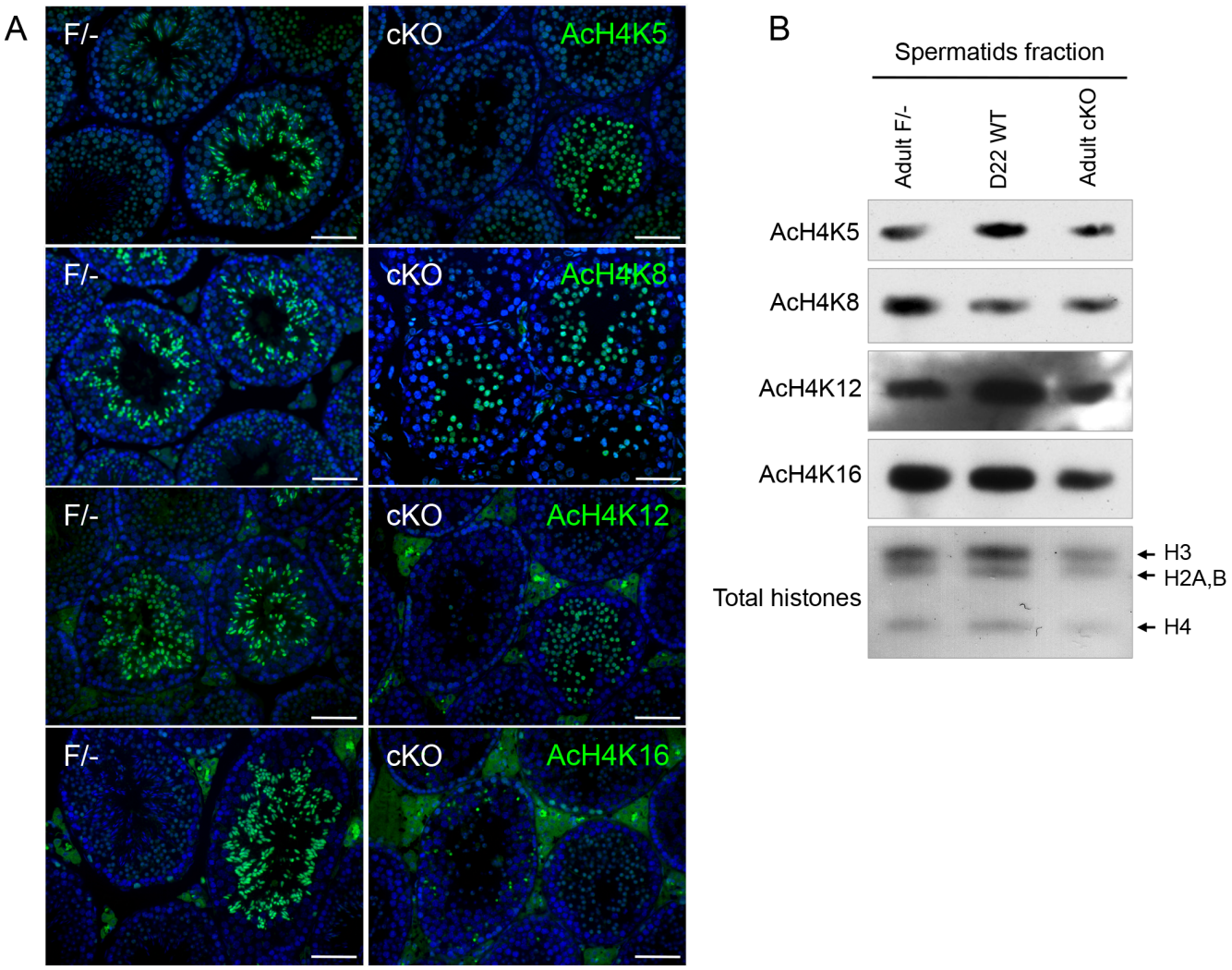

Fig. S5. Acetylation of histone H4 in the testes. A. Immuno-localization of acetylated histone H4 (AcH4K5, AcH4K8, AcH4K12, and AcH4K16) in the F/- and the cKO testes. The nuclei were stained with DAPI. Scale bar: 50μm. B. Quantitative changes of AcH4K5, AcH4K8, AcH4K12, and AcH4K16 in the adult the F/- and the cKO testes and 22 day-old testes, in which spermatogenesis developed to the round spermatid stage. Total histones are shown as loading control.

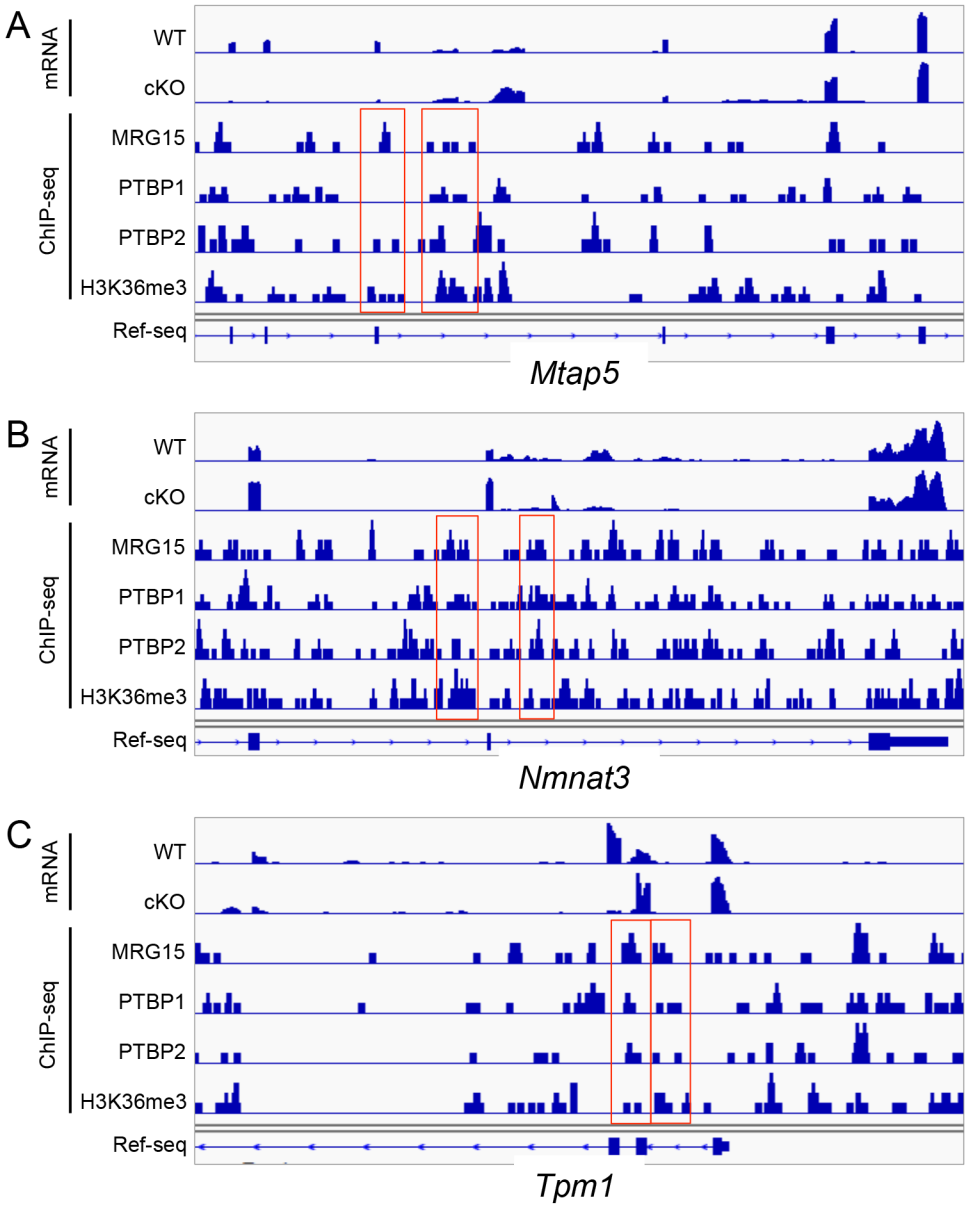

Fig. S6. Localization of MRG15 complex on the genes that have retained intron. Genome browser views show the alignment of the read sequences of mRNA wild-type and MRG15 cKO spermatid and ChIP with MRG15, PTBP1, PTBP2, and H3K36me3 of wild-type spermatid on the mouse genomic regions of *Mfap5* (A), *Nmnat3* (B), and *Tpm1* (C).

Table S1. Primer sequences used in the study.

Name	Forward primer	Reverse primer
Prm1 qPCR	ATGTCGCAGACGGAGGAG	ACGCAGGAGTTTTGATGGAC
Prm2 qPCR	GCTGCTCTCGTAAGAGGCTACA	AGTGATGGTGCCTCCTACATTT
Tnp1 qPCR	GAGAGGTGGAAGCAAGAGAAAA	CCCACTCTGATAGGATCTTTGG
Tnp2 qPCR	GAGGGAAAGTGAGCAAGAGAA	GCATAGAAATTGCTGCAGTGAC
Prm1 1kb up	ACCCAGTACTTTGGAGGCAGAG	GGCAGAAGGGTGATAGGCACC
Prm1 pro	CCATATTTGGACATGGTACAGG	TCGAGCCAGCCAACCTGTGAGCA
Prm1 Ex1-Int	GGCGAGATGCTCTTGAAGTC	GGGAGTCCCTCTCACCACTT
Prm2 1kb up	GAACTTTCCTCTGGGCTTCC	AGAGAGAGAATGCCTCACTGG
Prm2 pro	CACGGTGTTGTTGAAAAGGA	CCATTCCATGTCGATGTCTG
Prm2 Ex1-Int	GTGCTTACCTCTGCGATGC	GAGCGCGTAGAGGACTATGG
Tnp1 1kb up	AGATAGGCAACAGCTTTTGGAA	GGGATCTCAGAGACTGAACCAC
Tnp1 pro	TAGGTACCAGCAAGGTGTAGCA	CTGGTCGACATGGTACTTTCT
Tnp1 Ex1-Int	CCCTCTGATGTCCTCATGCT	TCCCAAGTGCTGGGATTAAG
Tnp2 1kb up	TGAGACCATCTCAGCAGGTAAA	CTGAAGAATTCCCTCTCACCAC
Tnp2 pro	ACAATCTGACCTGACCCACAG	CAGAGACTTCCTCCTCCTCCTC
Tnp2 Ex1-Int	GAGTGCGTGTTGCTCAGGTA	TCGACACTCACCTGCAAGAC

SI Materials and Methods

Generation of Germ Cell-Specific *Mrg15* Null Mice and Fertility Analysis. About 3 kb and 5 kb of genomic region of *Mrg15* were inserted into pBluescript SK containing diphtheria toxin A for negative selection (kindly provided by Dr. Ronald A. DePinho, Dana-Farber Cancer Institute, now at MD Anderson Cancer Center) (1). A *loxP* was inserted between exon 3 and 4, and an *Frt - Pgk1-Neo – Frt - loxP* cassette was inserted between exon 2 and 3 to frank exon containing a part of CHROMO domain with *loxP* sequences (Fig. S1A). The linearized targeting construct was electroporated into AB2.2 embryonic stem (ES) cells, which are derived from 129S5 strain mice. ES cell clones were selected in ES cell culture media supplemented with 0.18 mg/ml G418. Targeted clones were screened by Southern blot analysis using 5' and 3' probes (Fig. S1B). Three of the correctly targeted clones were expanded and injected into C57BL/6J blastocysts. Chimeric males were bred to C57BL/6J females to obtain heterozygous *Mrg15* floxed mice ($Mrg15^{F/+}$), followed by mating with *Stra8-Cre* mice (kindly provided by Dr. Robert E. Braun, Jackson Laboratory) (2) to generate $Mrg15^{+/-;Stra8\text{-}Cre+}$ mice. A germline specific mutant mice (*Mrg15* cKO; $Mrg15^{F/-;Stra8\text{-}Cre+}$) were produced by crossing male $Mrg15^{+/-;Stra8\text{-}cre+}$ and female $Mrg15^{F/F}$ mice. For fertility analysis, six-week-old control ($Mrg15^{F/-;Stra8\text{-}Cre-}$) and *Mrg15* cKO ($Mrg15^{F/-;Stra8\text{-}Cre+}$) male littermates were individually bred to WT females. The number of litters and pups born per litter were monitored over a 6-month period. All mouse experiments were performed on a C57BL/6J: 129S5 hybrid background in accordance with protocols approved by the Institutional Animal Care and Use Committee of Baylor College of Medicine and the Experimental Animal Care Committee of Kyushu University.

Northern Blot Analysis and Semi-quantitative and Quantitative RT-PCR. Total RNA from mouse adult tissue samples and developing testes samples, and cultured and isolated cell samples were extracted using RNeasy mini and micro kit (QIAGEN), respectively, according to the manufacturer's instruction. For northern blot analysis, twenty microgram of total RNA was separated by agarose gel, transferred to nylon membrane, and hybridized with *Mrg15* and *Cdyl* specific probes (3). For reverse transcription, one microgram of total RNA was reverse transcribed using Superscript III reverse transcriptase (Invitrogen) and an oligo-dT primer (Invitrogen). For semi-quantitative RT-PCR, PCR amplifications were performed using *Taq* DNA polymerase and visualized by gel electrophoresis. For quantitative PCR, PCR amplifications were performed using SYBR Green PCR Master Mix (ABI) and

analyzed by the ABI 7500 sequence detection system. The primer sequences used in the experiments are listed on *SI Appendix*, Table S1.

Histology, Immunofluorescence, and TUNEL Staining. Testes were fixed in 4% paraformaldehyde or Bouin's fixative overnight, followed by paraffin embedding, which was performed by the Department of Pathology Histology Core, Baylor College of Medicine and the Laboratory for Technical Support, Medical Institute of Bioregulation, Kyushu University. For testis histology, five-micrometer sections were stained with periodic acid-Schiff reagent and counterstained with hematoxylin. For immunofluorescence of tissue sections, paraformaldehyde fixed sections were retrieved by microwave, and then incubated with primary antibodies overnight at 4˚C, followed by Alexa 488 and Alexa 594 conjugated secondary antibodies (Invitrogen) for 1 hour at room temperature. Fluorescent sections were mounted with VECTASHIELD containing DAPI (VECTOR Laboratories). To detect apoptotic cells, DNA fragmentation was analyzed by ApopTag Plus Fluorescein In Situ Apoptosis Detection Kit (Millipore) according to manufacturer's instruction. Primary antibodies used in the experiments were as follows: rabbit anti-MRG15 (4), goat anti-ZPBP1 (5), and goat anti-TEX14 (6), rabbit anti-AcH4K5, anti-AcH4K12, anti-AcH4K16, anti-CDYL (Abcam) and rabbit anti-AcH4K8, mouse anti-FK2 (anti-ubiquitin), rabbit anti-AcH3 (Upstate Biotechnology, now EMD Millipore), mouse anti-SCP3, rabbit anti-PTBP2 (Santa Cruz Biotechnology).

Histone Extraction and Immunoblot. Histone extraction and immunoblot were performed as previously described (7). Briefly, after the nuclear fraction was collected in NP-40 lysis buffer (100mM Tris, 150mM M NaCl, 1.5mM $MgCl_2$, 0.65% NP-40, and protease inhibitor cocktail), histones were extracted from the nuclear fraction by treatment of 0.2M H_2SO_4. The extracted histones were concentrated by trichloro acetic acid agglutination and subjected to immunoblot analysis. Individual membranes were used to detect the methylation status of histone H3. Antibodies used in the experiments were rabbit anti-AcH4K5, anti-AcH4K12, anti-AcH4K16 (Abcam), rabbit anti-AcH4K8, (Upstate Biotechnology, now EMD Millipore), and goat anti-TNP2 (Santa Cruz Biotechnology).

Chromosome Analysis. Chromosome spread analysis was performed as previously described (7). The following primary antibodies were used for this experiment: rabbit anti-MRG15 (4), rabbit anti-SCP3 (Thermo Scientific), mouse anti-SCP3 (Santa Cruz Biotechnology), rabbit anti-SCP1, mouse anti-

rH2AX, rabbit anti-H3K36me3, rabbit anti-H4K20me3 (Abcam), mouse anti-ubH2A, and mouse anti-FK2 (Upstate Biotechnology, now EMD Millipore). Alexa 488 and Alexa 594 conjugated secondary antibodies (Invitrogen) were used to visualize the signals.

Co-Immunoprecipitation. Crude spermatid enriched samples were collected as previously reported method with slight modifications (8). After removal of tunica, adult seminiferous tubules were treated with collagenase for 15 min at room temperature to remove interstitial cells. Next, tubules were minced by scissors and suspended by 5 mL pipette without any enzyme treatment. The majority of spermatogonia and some of spermatocyte are not released without enzymatic treatment. Cell suspension and crumps were filtrated through 70μm mesh filter, followed by 35μm mesh filter to remove cell crumps and sertoli cells and to collect crude spermatid-enriched cell suspension. Five million spermatids were lysed in NP-40 lysis buffer and pre-cleared with normal rabbit IgG (Santa Cruz Biotechnology). The pre-cleared lysates were immunoprecipitated with 2 μg of rabbit anti-MRG15 or 2 μg of rabbit anti-PTBP2 (Santa Cruz Biotechnology) for 4 hours at 4˚C. Precipitation with normal rabbit IgG was simultaneously performed as a control and 1% of protein lysate was kept as input. Immune complexes were captured with protein G sepharose (GE Healthcare Life Sciences) for overnight at 4˚C. After 5 times washing with NP-40 lysis buffer, immune complexes were eluted in SDS sample buffer and subjected to an immunoblot analysis with anti-PTBP2 and anti-MRG15.

Chromatin Immunoprecipitation (ChIP). Crude spermatid enriched samples of 22-25 days old mice were collected as above. Approximately ten million spermatids were crosslinked in 1% paraformaldehyde for 5 min at room temperature, followed by incubation with PBS supplemented with 200 mM glycine for 5 min at room temperature. After washing twice with PBS, cells were lysed in NP-40 lysis buffer (10 mM Tris, 10 mM NaCl, 0.5% NP-40, and protease inhibitors cocktail) for 10 min on ice. The nuclei were collected and resuspended in 1 ml sonication buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate, and protease inhibitor cocktail). After sonication of DNA to an average length 200-500bp, debris were remove by centrifugation. Supernatants were immunoprecipitated with pre-coated protein G Sepharose, which was incubated with the antibody of interest in advance, for overnight at 4˚C. Control immunoprecipitation was performed with normal rabbit IgG. Beads were sequentially washed for 5 min at 4˚C each in Low-salt (50mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% sodium

deoxycholate), High-salt (50mM Tris/HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate), and twice in TE. Beads were eluted in SDS elution Buffer (10 mM Tris-HCl pH8.0, 300 mM NaCl, 5mM EDTA, 0.5% SDS) for 8 hours at 65˚C to reserve crosslinking. After incubation with RNase A and proteinase K for 30 min at 37˚C and for 3 hours at 55˚C, respectively, chromatin was precipitated with ethanol, following phenol/chloroform extraction. Immunoprecipitated DNA and input DNA were analyzed by PCR using the primers listed on *SI Appendix*, Table S1.

RNA Deep-Sequencing and ChIP-seq Analysis. The 50-mer reads were generated using an Illumina GA starting from polyA+ RNA isolated from wildtype and MRG15-null spermatids. Four replicates were sequenced for each of the two conditions, MRG15 wild-type and MRG15 cKO, yielding an average of 21.8 million read pairs per replicate. Duplicate sequences were removed using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The sequence reads were trimmed using PoPoolation at the first base pair with a quality score below 20 using; reads with a length shorter than 40% of the original read-length were discarded (9). The remaining reads were mapped onto the mouse genome (UCSC mm9/NCBI build 37) using the TopHat progam (10), which also reported junction reads. To examine intron retention we used the RefSeq gene model (11). We further selected junctions reported by TopHat with a coverage of at least 5 reads and that are contained exclusively in introns, and that do not overlap with exons of the RefSeq genes. Recurrence across 3-4 out the 4 replicates for each condition was used as false-positive filter for the reported intron retentions.

For ChIP-seq, 50-mer reads were sequenced using an Illumina HiSeq 1500 starting from ChIP library prepared from ChIP samples described above using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB), according to the manufacturer's instruction. ChIP-seq reads were mapped to the reference mouse genome (mm10/GRCm38) using the Bowtie (12) software version 1.0.0, allowing up to two mismatches. The multiple-hit reads were excluded, and only the uniquely mapped reads to the reference mouse genome were kept for further analysis. Duplicated reads were also removed with the samtools (13) software version 0.1.19-44428cd. To identify peaks and regions of MRG15, PTBP1, PTBP2 and H3K36me3 enrichment, the reads were analyzed using the MACS (14) software version 2.0.10.20130528 with the option '--nomodel --to-large --qvalue 0.05'. The input reads were used to normalize the background noise.

Statistics. Data are presented as the mean ± SEM. Statistical significance was determined by Student's *t*-test. $P<0.05$ was considered statistically significant.

SI References

1. Bardeesy N, et al. (2002) Loss of the Lkb1 tumour suppressor provokes intestinal polyposis but resistance to transformation. *Nature* 419(6903):162–167.

2. Sadate-Ngatchou PI, Payne CJ, Dearth AT, Braun RE (2008) Cre recombinase activity specific to postnatal, premeiotic male germ cells in transgenic mice. *genesis* 46(12):738–742.

3. Tominaga K, Pereira-Smith OM (2002) The genomic organization, promoter position and expression profile of the mouse MRG15 gene. *Gene* 294(1-2):215–224.

4. Tominaga K, et al. (2005) MRG15 regulates embryonic development and cell proliferation. *Mol Cell Biol* 25(8):2924–2937.

5. Lin Y-N, Roy A, Yan W, Burns KH, Matzuk MM (2007) Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Mol Cell Biol* 27(19):6794–6805.

6. Greenbaum MP, et al. (2006) TEX14 is essential for intercellular bridges and fertility in male mice. *Proc Natl Acad Sci U S A* 103(13):4982–4987.

7. Iwamori N, Zhao M, Meistrich ML, Matzuk MM (2011) The testis-enriched histone demethylase, KDM4D, regulates methylation of histone H3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility. *Biol Reprod* 84(February):1225–1234.

8. Kimura Y and Tanagimachi R. (1995) Mouse oocytes injected with testicular spermatozoa or round spermatids can develop into normal offspring *Development* 121(8): 2397-2405.

9. Kofler R, et al. (2011) Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* 6(1). doi:10.1371/journal.pone.0015925.

10. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.

11. Pruitt KD, Tatusova T, Maglott DR (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33(Database issue):D501–504.

12. Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.

13. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

14. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137.