Critical roles of long noncoding RNAs in *Drosophila* spermatogenesis

Kejia Wen[1, #], Lijuan Yang[1, 2, #], Tuanlin Xiong[1, #], Chao Di[1], Danhui Ma[1], Menghua Wu[1], Zhaoyu Xue[1], Xuedi Zhang[1], Li Long[1], Weimin Zhang[1], Jiaying Zhang[1], Xiaolin Bi[3], Junbiao Dai[1, 4], Qiangfeng Zhang[1, 4, 5], Zhi John Lu[1, 4], and Guanjun Gao[1, 4, *]

[1] MOE Key Laboratory of Bioinformatics, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

[2] College of Animal Science, Tarim University, Xinjiang, 843300, China

[3] Institute of Cancer Stem Cell, Dalian Medical University, Dalian, 116044, China

[4] Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China

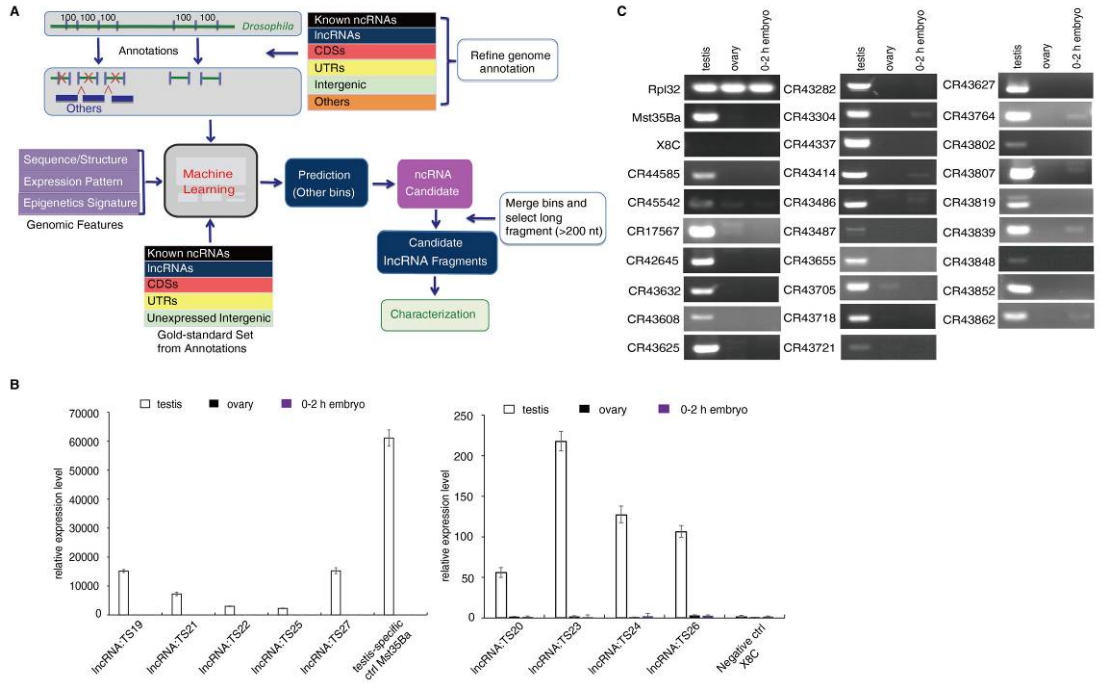
[5] Beijing Advanced Innovation Center for Structural Biology, Tsinghua University, Beijing 100084, China

[#]Equal contribution

[*]Corresponding author:

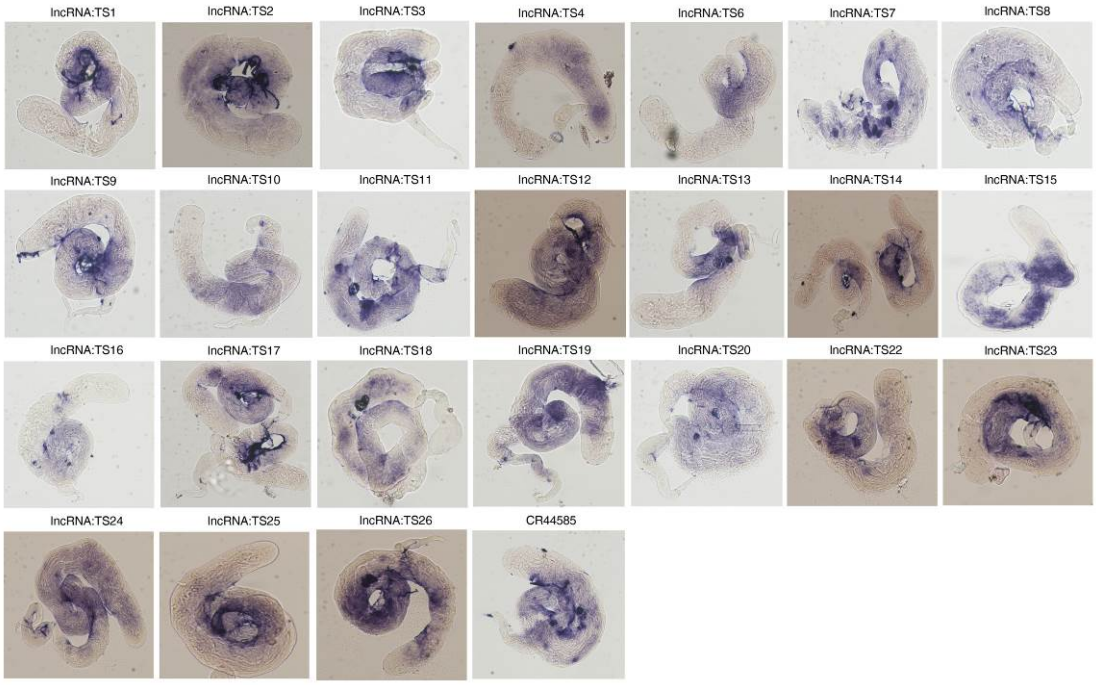
Guanjun Gao, School of Life Sciences, Tsinghua University, Yuanmingyuan Road 1, Beijing 100084, China. E-mail: gaogu@mail.tsinghua.edu.cn

Supplemental Figures



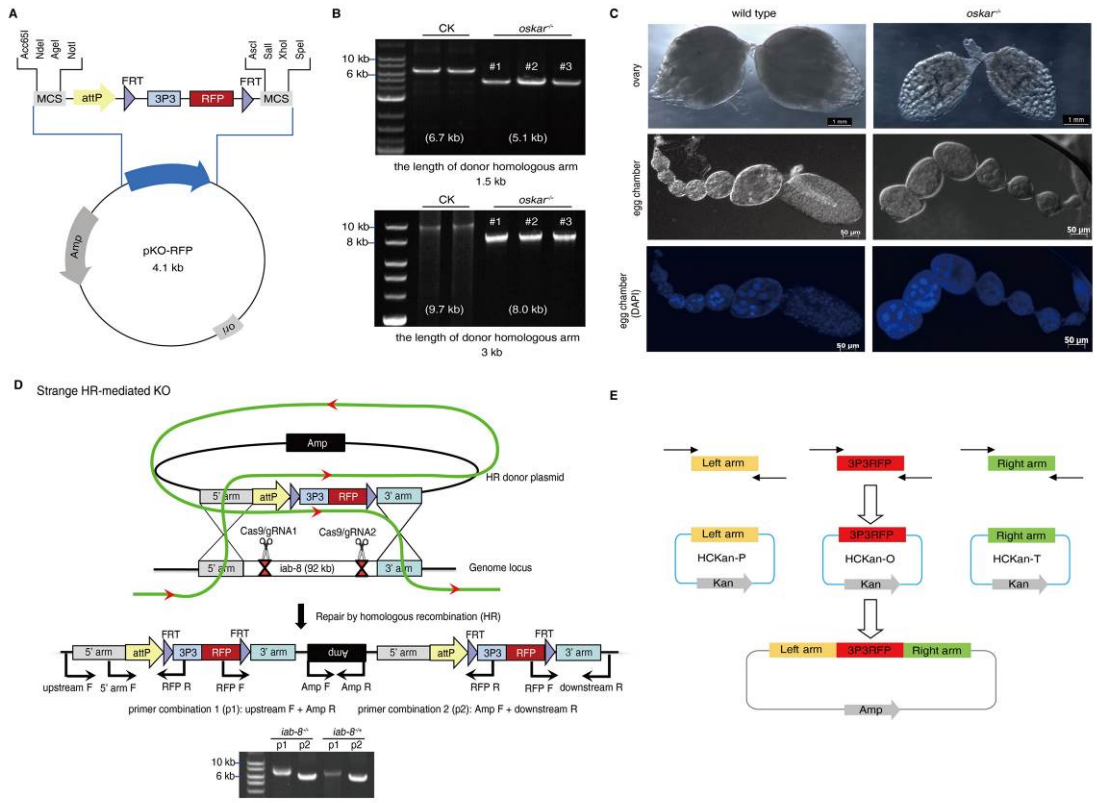
Supplemental Figure S1. Systematic identification of *Drosophila* lncRNAs involved in spermatogenesis.

(A) Flowchart of *incRNA* (integrated ncRNA finder) for prediction and characterization of novel lncRNA candidates in *Drosophila*. (B) Testis-specific expression screen of novel predicted lncRNA by quantitative RT-PCR. Rpl32 was used as an internal control. Results are representative of means ± SEM of triplicate determinations. *Mst35Ba* was used as a testis-specific control. *X8C* (a chromosome X-linked intergenic region that is transcriptionally silent) was used as a negative control to rule out contamination of RNA by genomic DNA. (C) Testis-specific expression screen of annotated lncRNA in FlyBase by semi-quantitative RT-PCR. *Rpl32* was used as a positive control, and *Mst35Ba* was used as a testis-specific control.



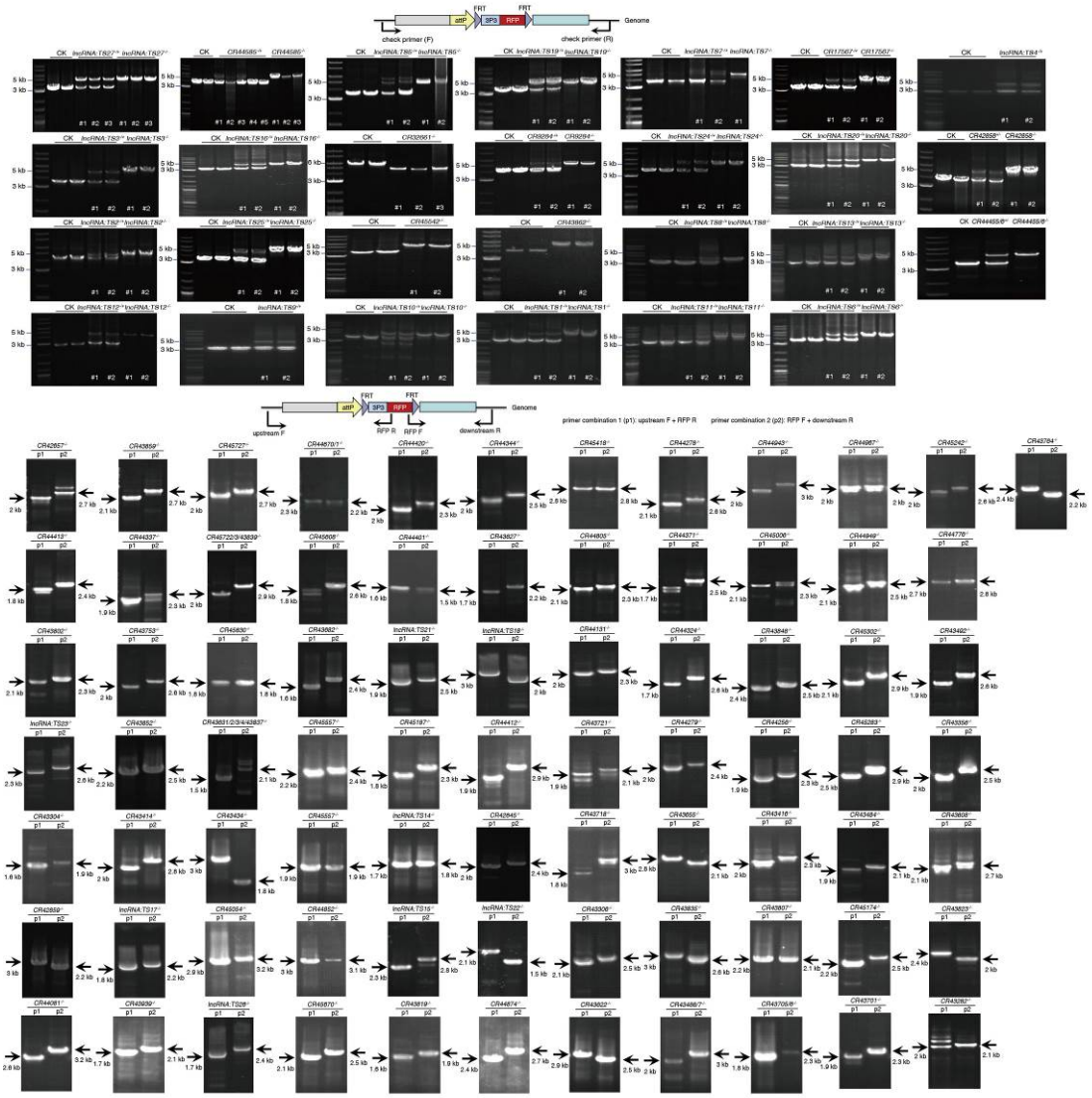
Supplemental Figure S2. Whole-mount lncRNA *in situ* hybridization of *Drosophila* testis.

This figure is an extension of Fig. 1C.



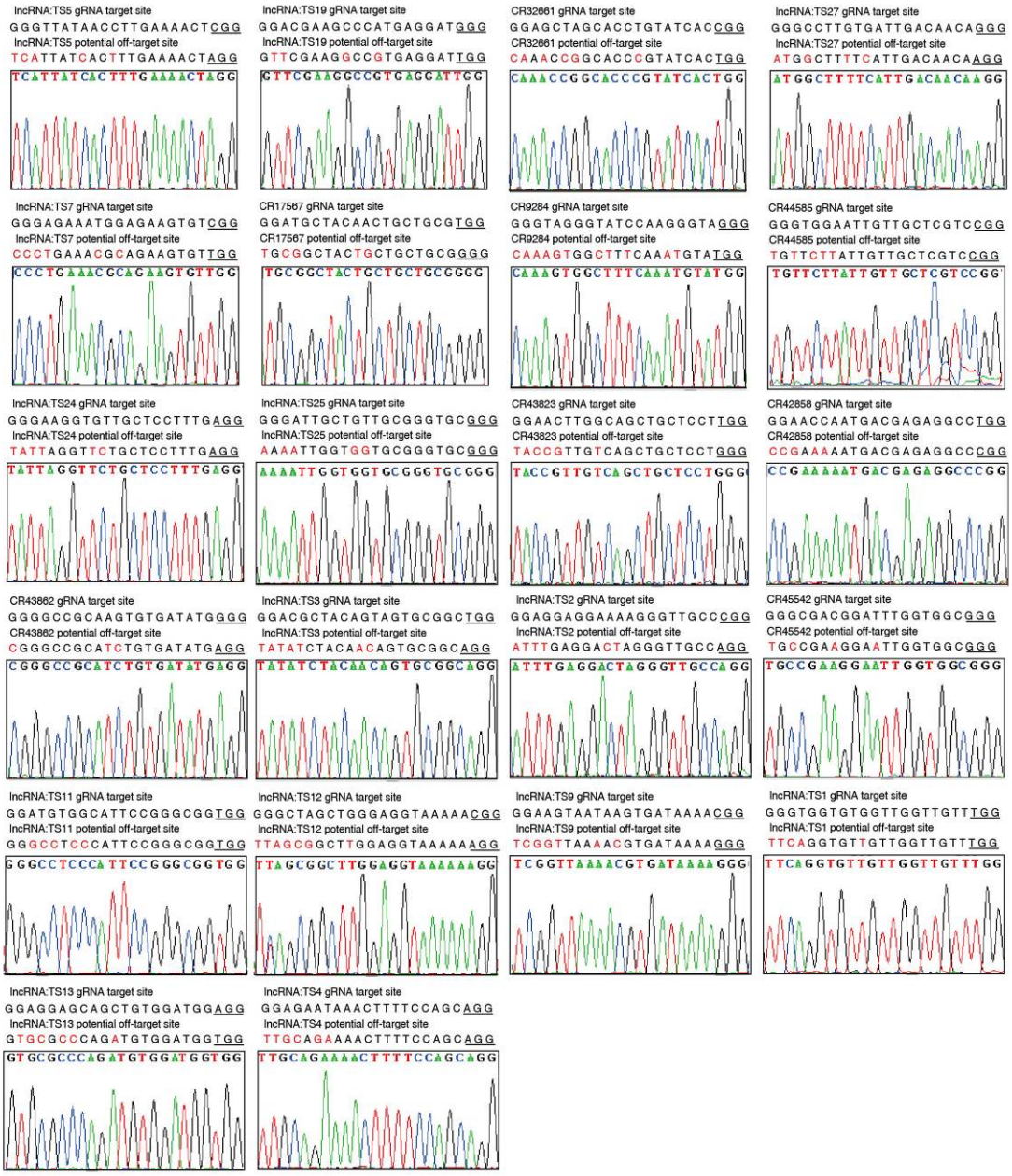
Supplemental Figure S3. Optimization of CRISPR/Cas9 system for systematic knockout.

(A) Map of donor plasmid for lncRNA knockouts. The 3P3-RFP selective marker is flanked by two FRT sites to facilitate marker removal, when necessary. The attP site was designed for *in cis* rescue. (B) Genotyping of w1118 and *oskar*$^{-/-}$ flies to confirm deletion of the *oskar* gene. Mutants with a 5 kb deletion of *oskar* DNA were obtained using the donor plasmids with 1.5 kb and 3 kb homologous sequences. (C) Characterization of the *oskar*$^{-/-}$ phenotype. Deletion of *oskar* caused defects in the ovary and egg chamber. (D) When a 92 kb *iab-8* lncRNA knockout mutant was generated using CRISPR, one line had the RFP reporter instead of the *iab-8* lncRNA gene, as expected. (E) Schematic representation of modular construction of donor plasmid for different targets. Left and right HR arms can be cloned in parallel into two separate vectors and assembled together with the marker plasmid to generate the final donor plasmid. This strategy allowed us to make a large quantity of donor plasmids in a short time.



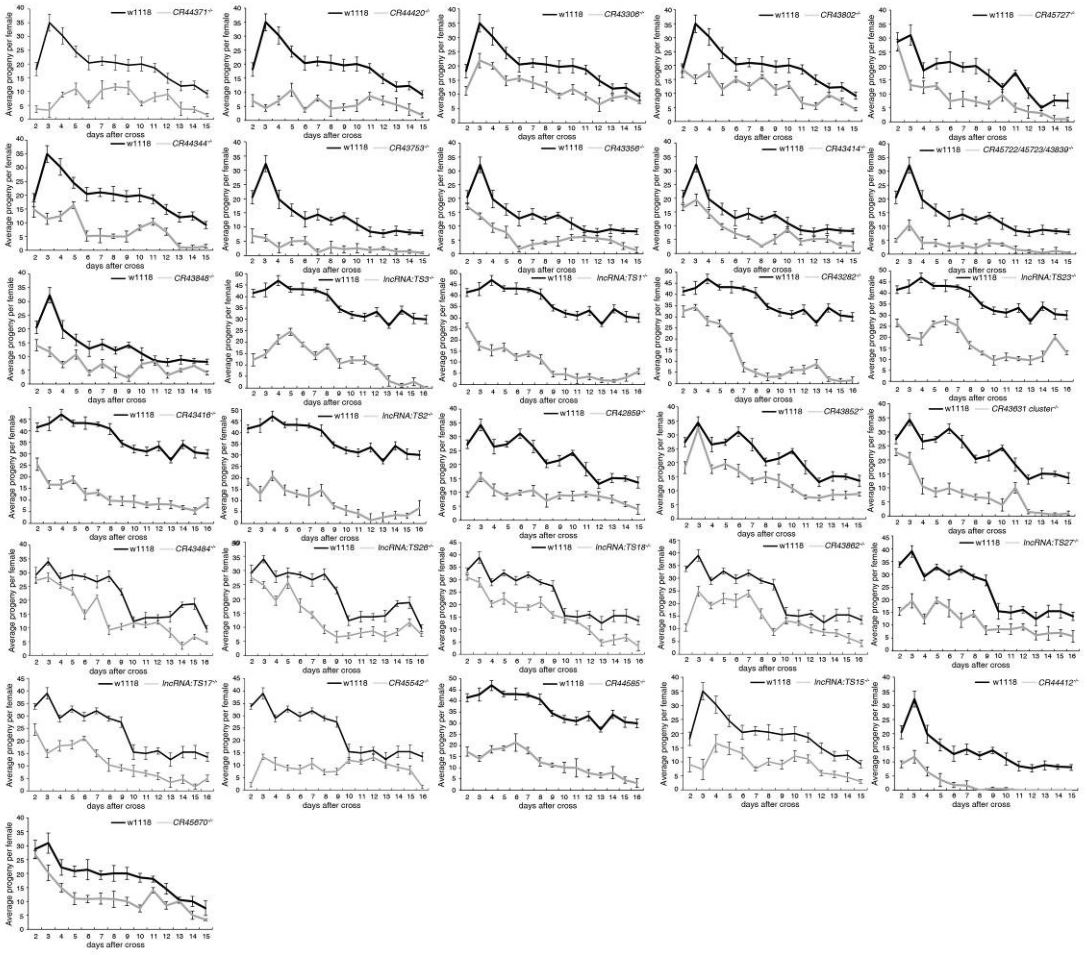
Supplemental Figure S4. Genotyping of w1118 and *lncRNA*$^{-/-}$ flies to confirm deletion of the remaining lncRNAs mentioned in Fig. 2E.

Insertion of the reporter cassette results in a size increase in the PCR products, which is visualized only in the heterozygous and homozygous flies.



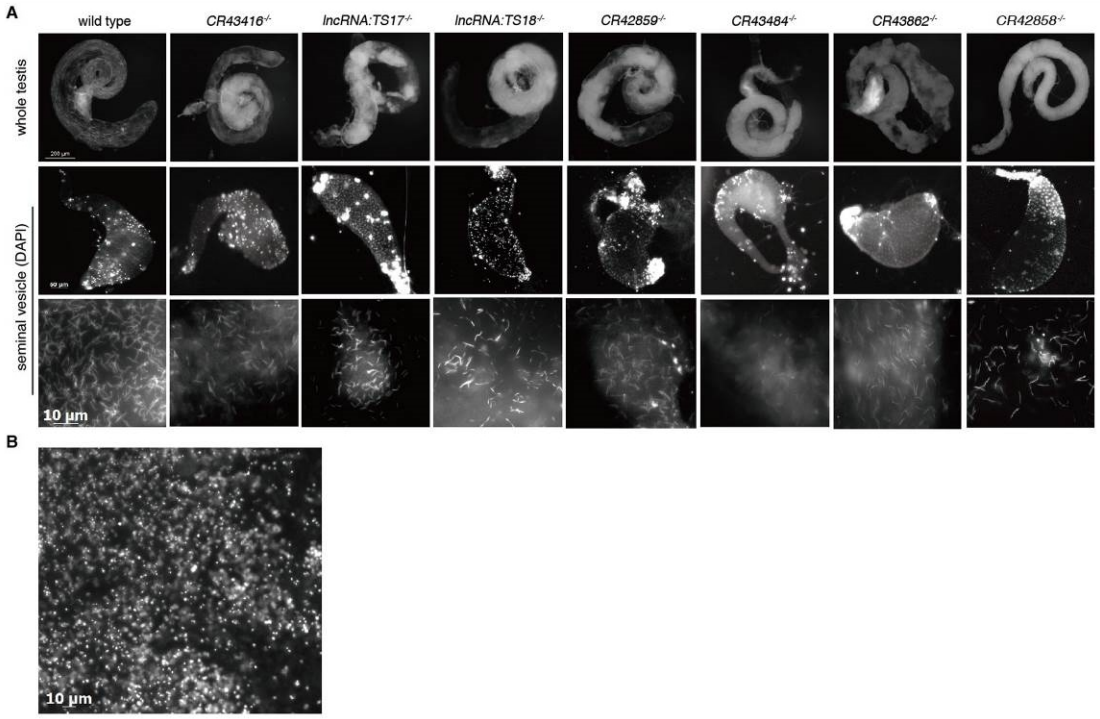
Supplemental Figure S5. Survey of potential sites for off-target effects of the CRISPR/Cas9 system.

The sequencing results of one representative off-target site are presented for each gRNA. One potential off-target site was tested for each lncRNA mutant. None of the sites were modified as a result of an off-target effect.



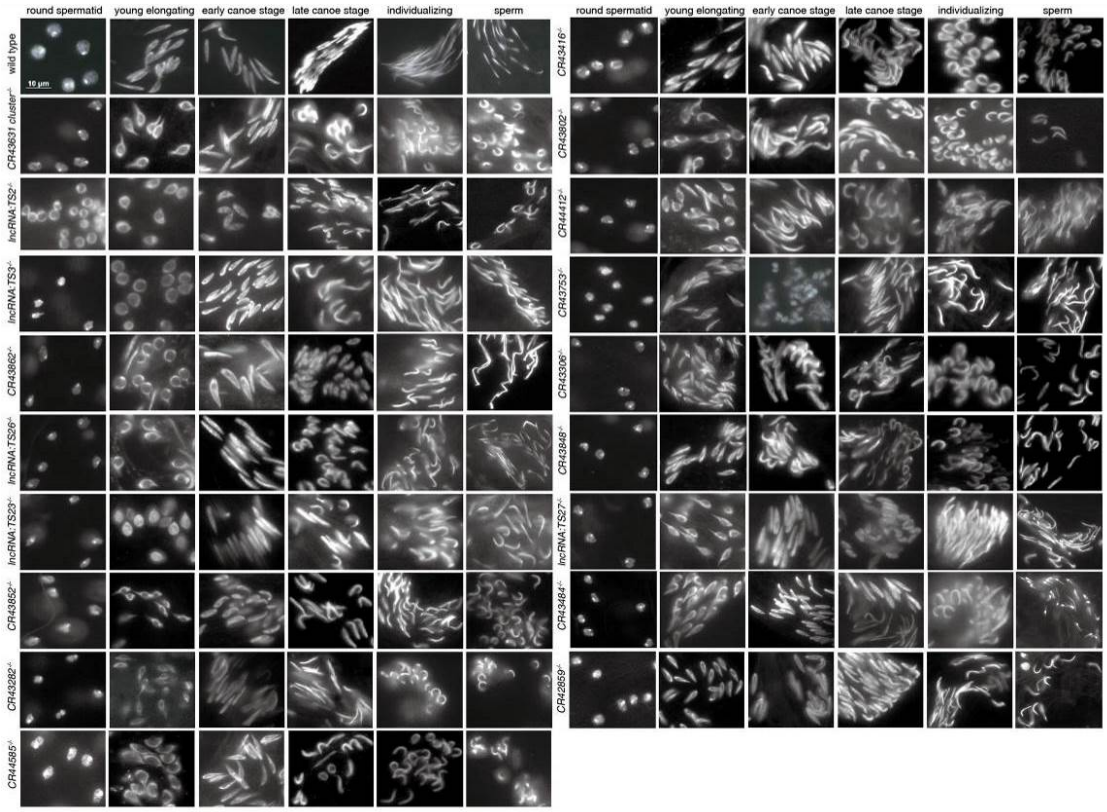
Supplemental Figure S6. Qualitative fertility assay of the remaining lncRNA knockout mutants mentioned in Fig. 3C.

Results represent means ± SEM for 15 crosses each.



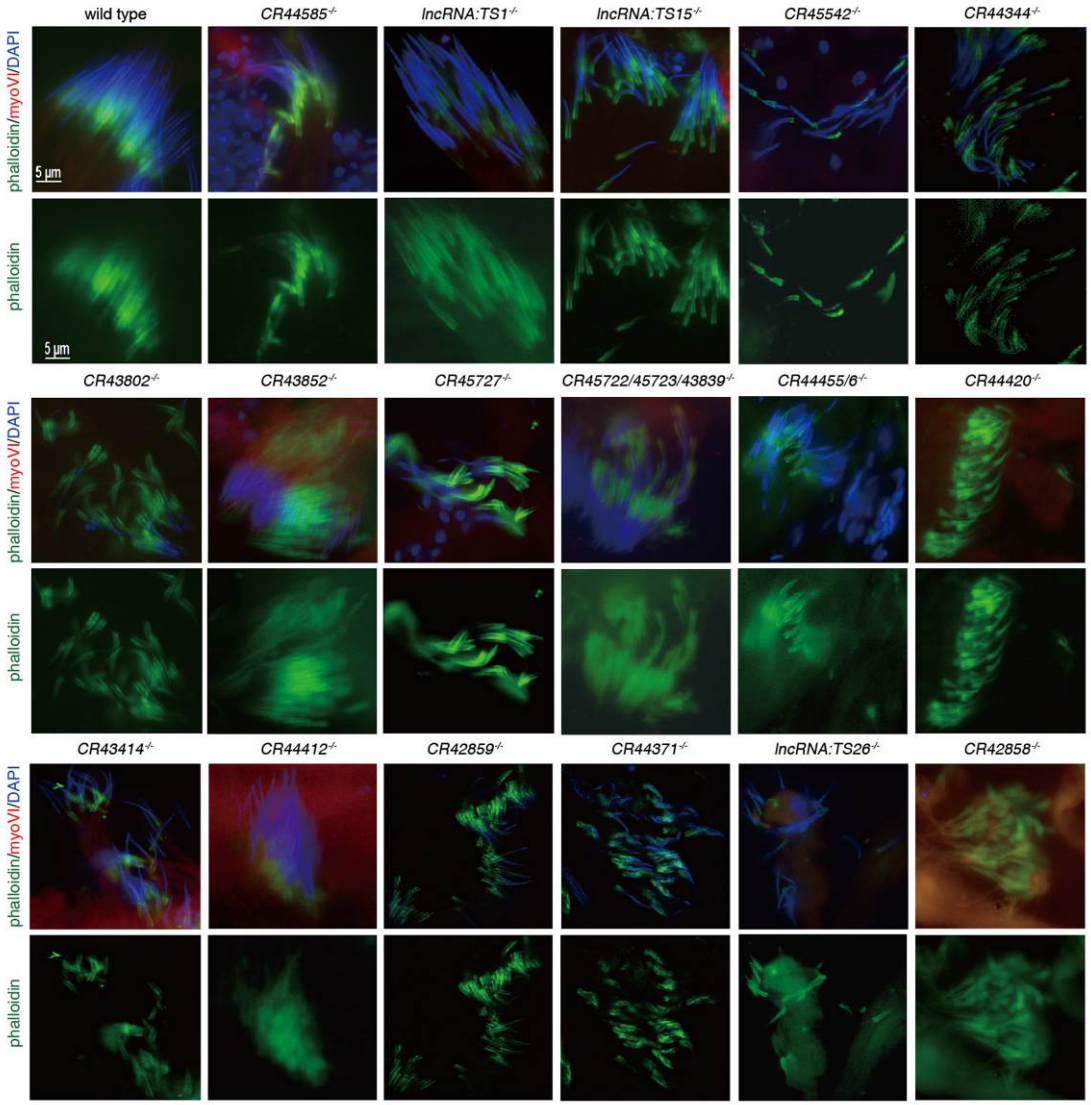
Supplemental Figure S7. LncRNA knockout mutants cause malformation and obstruction of testis.

LncRNA knockout mutants cause malformation and obstruction of testis. (A) The testes of seven lncRNA mutants exhibited an accumulation of abnormal white flocculus. Seminal vesicles from *lncRNA:TS18*$^{-/-}$, *CR43484*$^{-/-}$ and *CR42858*$^{-/-}$ contained reduced numbers of mature sperm, whereas the numbers in the other four were comparable to those in the wild type. Top rows, whole testis; middle rows, seminal vesicle stained with DAPI; bottom rows, sperm in seminal vesicle. Scale bars are noted. This figure is an extension of Fig. 4A. (B) White flocculus in lncRNA mutants contained a large number of tightly packed small cells.



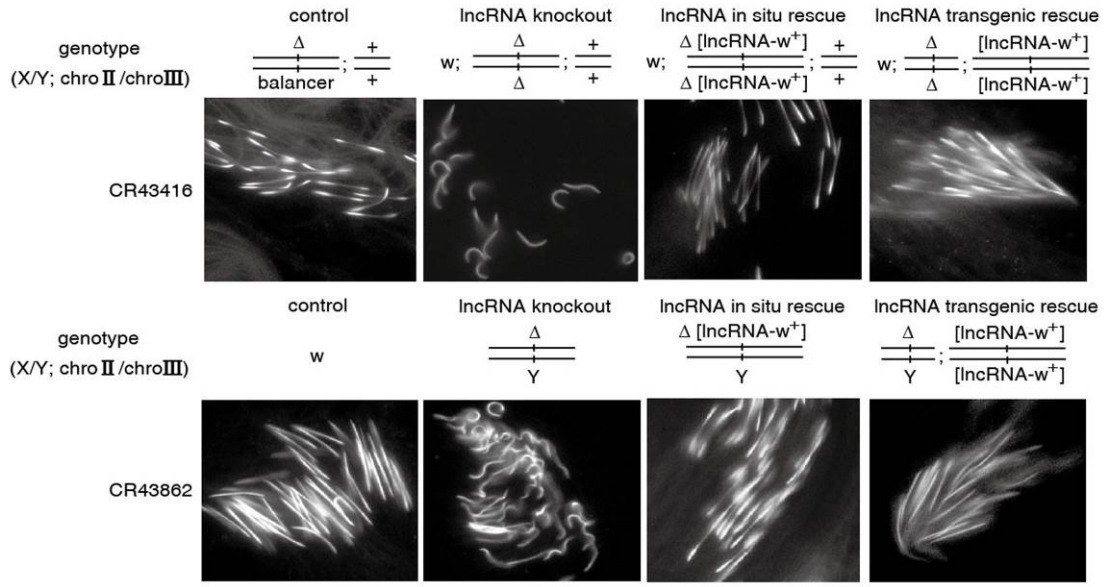
Supplemental Figure S8. LncRNA knockout mutants affect male germ cell development.

Testes squash preparations were stained with DAPI to visualize DNA of wild type and lncRNA mutants. In the wild type, the initially round spermatids nuclei elongated and condensed to form long, straight, needle-shaped mature sperm. In *CR43484*$^{-/-}$, some mature-stage sperms adopted a tadpole shape, in which sperm nuclei were concentrated at one end of the spermatid head. Deletion of other lncRNA mutants led to scattered or curled sperm in which some nuclei did not fully condense. This figure is an extension of Fig. 4B.



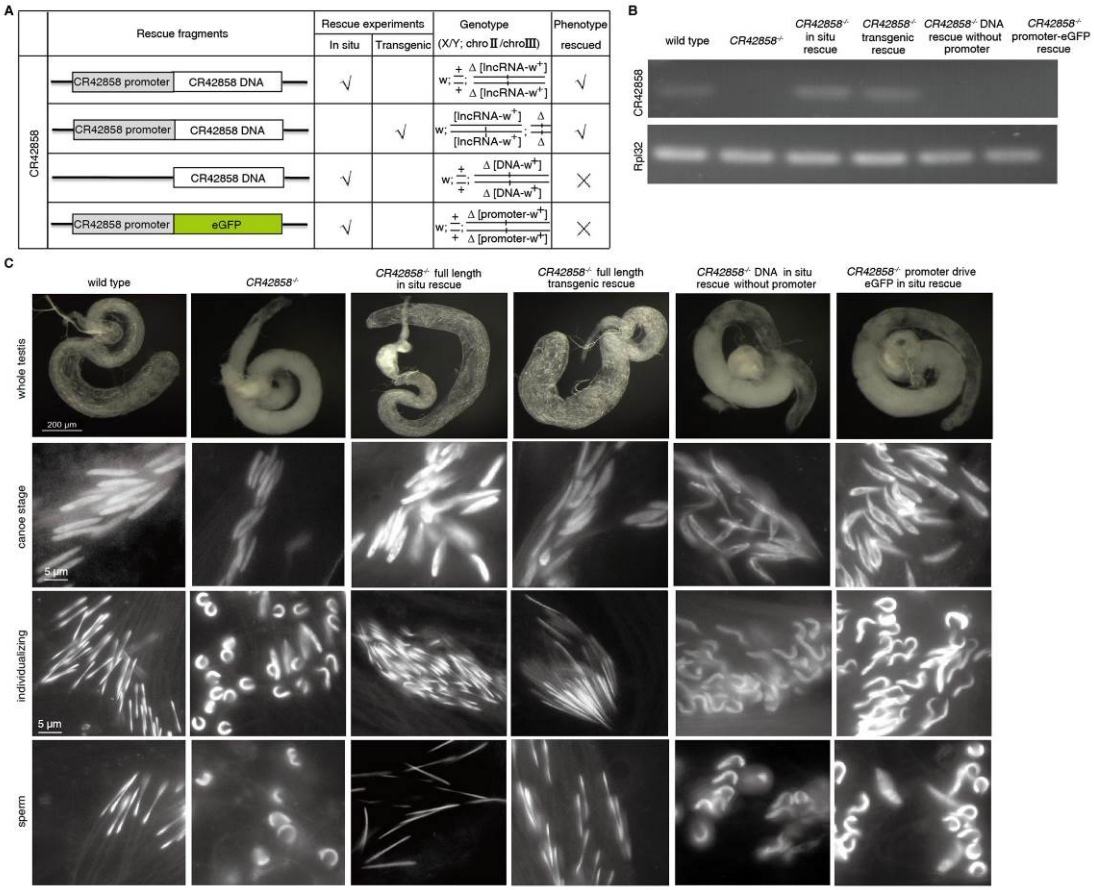
Supplemental Figure S9. Phalloidin staining of investment cones (ICs) in wild type and lncRNA mutants.

Wild type testes exhibited ordered and associated ICs, whereas lncRNA mutants had lagging or poorly aligned ICs. Scale bars are noted. This figure is an extension of Fig. 4E.



Supplemental Figure S10. The spermatogenesis defects of *CR43416* and *CR43862* knockouts were rescued by restoration of the corresponding lncRNAs *in cis* or *in trans*.

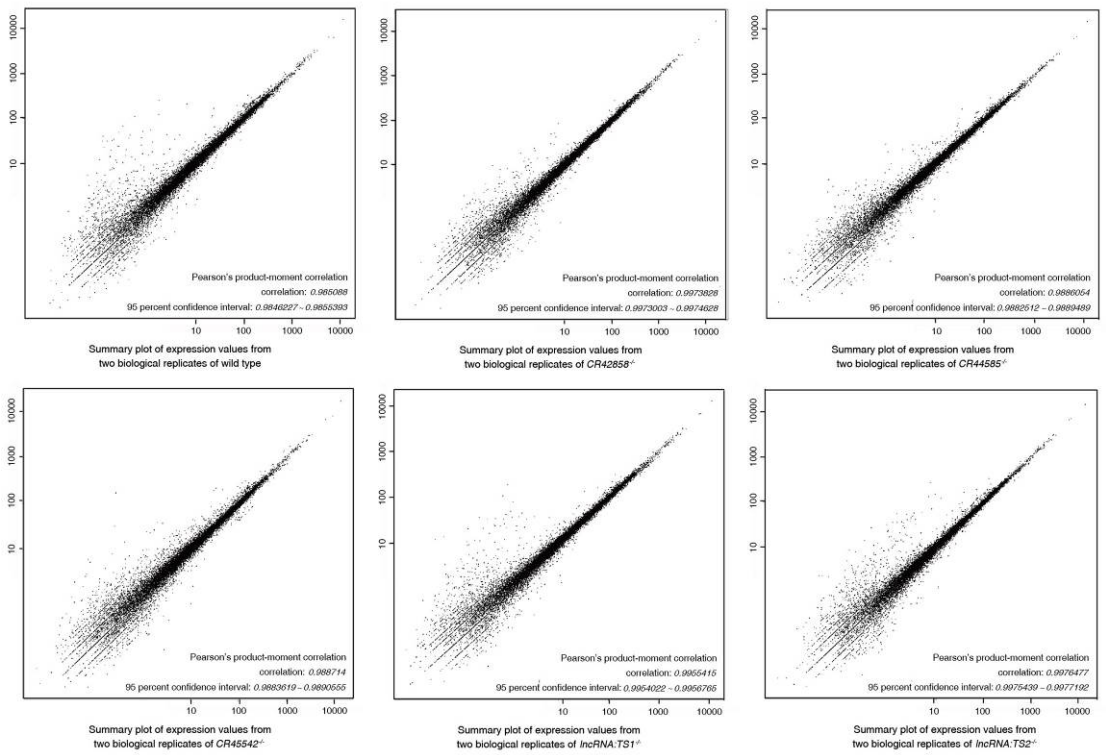
For *in cis* rescue, a construct bearing an lncRNA gene fragment under the control of the endogenous promoter was inserted back into the genomic locus from which the original lncRNA had been deleted. For rescue *in trans*, the lncRNA was placed on another chromosome (i.e., different from the original chromosomal locus of the lncRNA) through PhiC31-mediated attB/attP exchange. The abnormal morphological phenotype of scattered and curled sperm nuclei in late spermatogenesis of *CR43416*$^{-/-}$ and *CR43862*$^{-/-}$ were rescued by restoration of these lncRNAs either *in cis* or *in trans*.



Supplemental Figure S11. Different fragments rescue experiments on *CR42858*.

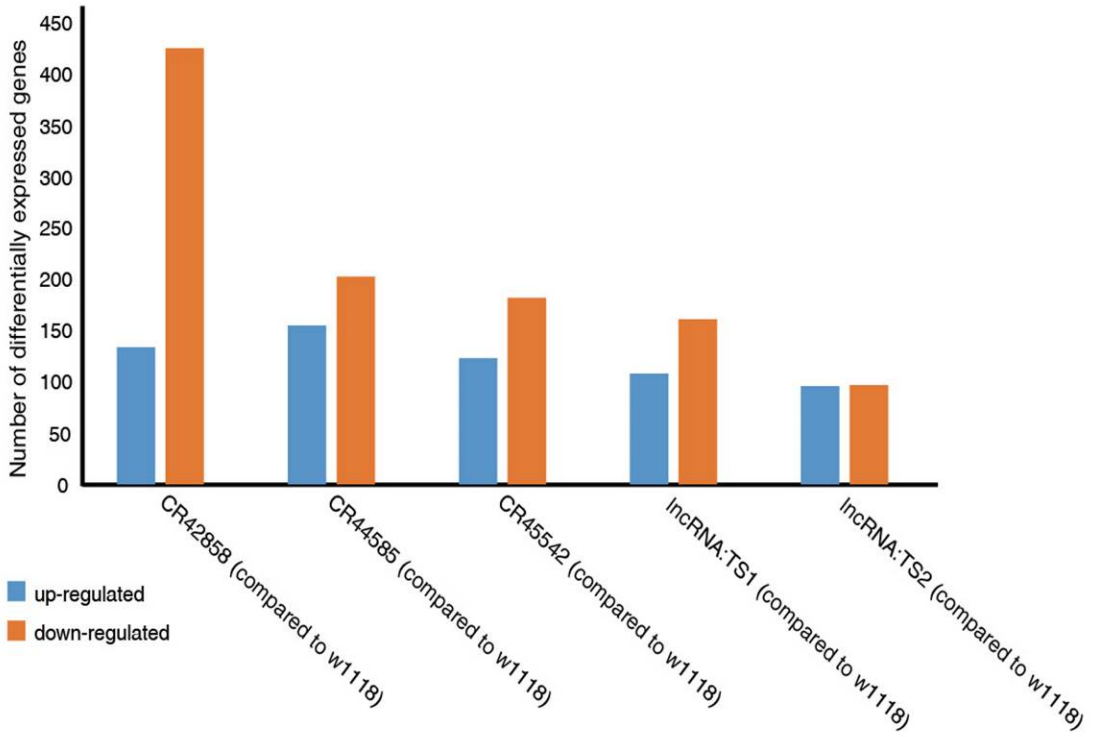
(A) Schematic illustration of rescue experiments on *CR42858*. For full-length rescue, a construct bearing an lncRNA gene fragment under the control of the endogenous promoter was inserted into the genomic locus from which the original lncRNA had been deleted. For DNA rescue, a construct bearing only the lncRNA DNA fragment without the promoter was inserted. For *CR42858* promoter rescue, a construct bearing an eGFP gene fragment under the control of the *CR42858* promoter was inserted. (B) The lncRNA *CR42858* was transcribed in flies rescued with full-length *CR42858 in cis* and *in trans*, but not in *CR42858*$^{-/-}$ or flies rescued with the *CR42858* DNA without promoter or *CR42858* promoter-driven eGFP. RNA was isolated from the testes of wild type, *CR42858*$^{-/-}$, flies rescued with full-length *CR42858 in cis* and *in trans*, flies rescued with *CR42858* DNA without the promoter, and flies rescued

with *CR42858* promoter-driven eGFP. The RNA was subjected to semi-quantitative RT-PCR. *Rpl32* was used as a control. (C) The abnormal morphological phenotype of testis and curled sperm nuclei in late spermatogenesis of $CR42858^{-/-}$ were rescued by restoration of full-length lncRNA either *in cis* or *in trans*, but not by *CR42858* DNA without the promoter or *CR42858* promoter-driven eGFP.



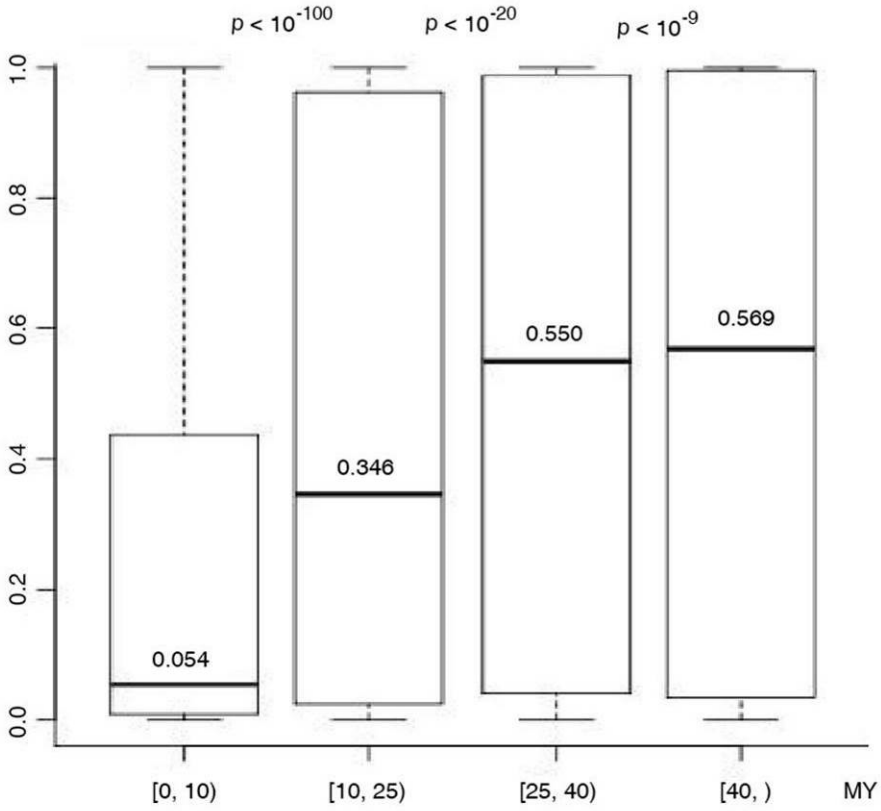
Supplemental Figure S12. Correlation of RNA-seq data between each of the two biological replicates for the following fly strains: wild type, *CR42858*$^{-/-}$, *CR44585*$^{-/-}$, *CR45542*$^{-/-}$, *lncRNA:TS1*$^{-/-}$ and *lncRNA:TS2*$^{-/-}$.

Summary plots show the expression values for both replicates (black). Using 0.9 as the recommended threshold (NOISeq manual) for probability of different expressions (significant expression variance), genes with significantly different expression levels will be highlighted in red. Here for our biological replicates no significant outlier genes can be observed for all the six comparisons. X-axis and Y-axis depict log scale of corrected expression data from Biological Replicate-1 and Replicate-2 [ref. to R package "NOISeq" (version 2.14.1)]. Pearson's correlation coefficients were assessed by R package "stats" (version 3.2.2).



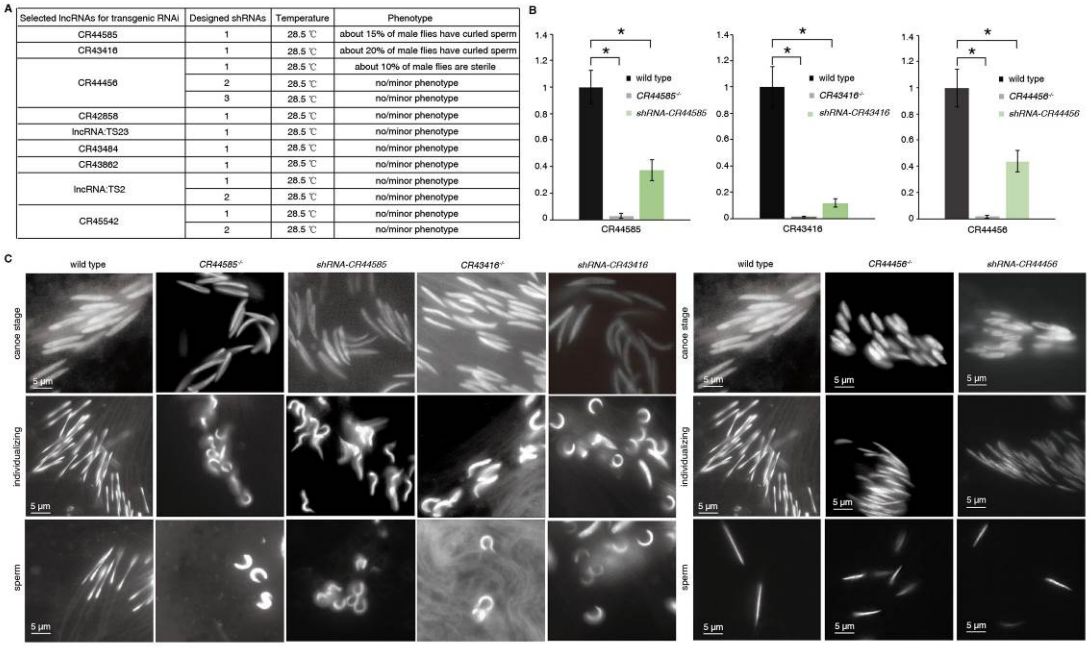
Supplemental Figure S13. The number of differentially expressed genes in five lncRNA knockout mutants, revealed by RNA sequencing.

RNA was from wild type, *CR42858*$^{-/-}$, *CR44585*$^{-/-}$, *lncRNA:TS2*$^{-/-}$, *lncRNA:TS1*$^{-/-}$, or *CR45542*$^{-/-}$.



Supplemental Figure S14. Evolutionary age of testis-specific lncRNAs.

Sequence conservation of lncRNAs of different age groups. All testis-specific lncRNAs were divided into different age groups: 0-10 MY, 10-25 MY, 25-40 MY, and >40 MY (MY: million years). Vertical axis represents PhastCons Scores, and thick horizontal lines in each box are median values. P-values were calculated for each pair of neighboring boxes by Mann-Whitney *U* test.



Supplemental Figure S15. Generation of 13 transgenic RNA interference lines targeting nine lncRNAs with clear knockout phenotypes.

(A) Analysis of spermatogenesis phenotypes in transgenic RNAi lines targeting lncRNAs. ShRNA females were crossed to Ubiqutin-Gal4 males at 28.5°C respectively. (B) Knockdown efficiencies of three lncRNA RNAi lines with observed phenotypes were determined using quantitative RT-PCR. *Rpl32* was used as an internal control. Values represent means ± SEM for triplicate determinations. (C) Three lncRNA transgenic RNAi lines had phenotypes in spermatogenesis. Knockdown of lncRNA *CR44585* (shRNA vs. *CR44585*, labeled *shRNA-CR44585*) and *CR43416* (shRNA vs. *CR43416*, labeled *shRNA-CR43416*) led to curled sperm in which some nuclei did not fully condense. Knockdown of lncRNA *CR44456* (shRNA vs. *CR44456*, labeled *shRNA- CR44456*) led to sterility in 10% of male flies. In these sterile male flies, the nuclei of mature sperm were half the size of those in wild type sperm.