Main Manuscript for

The Rapidly Evolving X-linked *miR-506* Family Finetunes Spermatogenesis to Enhance Sperm Competition

Zhuqing Wang[1,2], Yue Wang[1,2], Tong Zhou[1], Sheng Chen[1,2], Dayton Morris[2], Rubens Daniel Miserani Magalhães[2], Musheng Li[1], Shawn Wang[1], Hetan Wang[1,2], Yeming Xie[1], Hayden McSwiggin[1,2], Daniel Oliver[1], Shuiqiao Yuan[1], Huili Zheng[1,2], Jaaved Mohammed[3], Eric C. Lai[3], John R. McCarrey[4], Wei Yan[1,2,5]*

[1]Department of Physiology and Cell Biology, University of Nevada, Reno School of Medicine, Reno, NV 89557, USA.
[2]The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.
[3]Department of Developmental Biology, Memorial Sloan-Kettering Institute, 1275 York Ave, Box 252, New York, NY 10065, USA.
[4]Department of Biology, University of Texas at San Antonio, San Antonio, TX, USA.
[5]Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA.
*Corresponding author: Wei Yan **Email:** wei.yan@lundquist.org

Author Contributions: Z.W. and W.Y. designed the research. Z.W., Y. W., S. C., D.M., S. W., H. W., H.M., J. R. M., and H.Z., performed bench experiments. Z.W., T.Z., R.D.M.M., M.L., Y.X., J.M., and E. C. L. performed bioinformatics analysis. Z.W., and W.Y. wrote the manuscript.

Competing Interest Statement: The authors declare no conflict of interest.

Classification: Biological Sciences; Genetics; Physiology.

Keywords: *reproductive fitness; sperm competition; evolution; selective pressure; microRNAs; spermatogenesis; male fertility*

This PDF file includes:

Main Text

Figures 1 to 6

Abstract

Despite rapid evolution across eutherian mammals, the X-linked *miR-506* family miRNAs are located in a region flanked by two highly conserved protein-coding genes (*Slitrk2* and *Fmr1*) on the X chromosome. Intriguingly, these miRNAs are predominantly expressed in the testis, suggesting a potential role in spermatogenesis and male fertility. Here, we report that the X-linked *miR-506* family miRNAs were derived from the MER91C DNA transposons. Selective inactivation of individual miRNAs or clusters caused no discernable defects, but simultaneous ablation of five clusters containing nineteen members of the *miR-506* family led to reduced male fertility in mice. Despite normal sperm counts, motility and morphology, the KO sperm were less competitive than wild-type sperm when subjected to a polyandrous mating scheme. Transcriptomic and bioinformatic analyses revealed that these X-linked *miR-506* family miRNAs, in addition to targeting a set of conserved genes, have more targets that are critical for spermatogenesis and embryonic development during evolution. Our data suggest that the *miR-506* family miRNAs function to enhance sperm competitiveness and reproductive fitness of the male by finetuning gene expression during spermatogenesis.

Significance Statement

The X-linked *miR-506* family has rapidly evolved in mammals, but their physiological significance remains elusive. Given their abundant and preferential expression in the testis and sperm, these X-linked miRNAs likely play a functional role in spermatogenesis and/or early embryonic development. However, the deletion of either individual miRNA genes or all of the five miRNA clusters encoding 38 mature miRNAs did not cause major fertility defects in mice. When these mutant males were subjected to conditions resembling polyandrous mating, the mutant sperm were much less competitive than the wild-type sperm, rendering the mutant males “functionally sub-fertile”. Our data suggest that the *miR-506* family of miRNAs regulates sperm competition and the reproductive fitness of the male.

Main Text

Introduction

Spermatogenesis is highly conserved among all vertebrates. Although it generally consists of three phases (mitotic, meiotic, and haploid), many characteristics appear to be species-specific, *e.g.,* the duration of each of the three phases, the seminiferous epithelial organization, and the shape and length of spermatozoa, likely reflecting the adaptive changes during evolution (1-3). Several cellular events are unique to the male germ cells, *e.g.,* postnatal formation of the adult male germline stem cells (i.e., spermatogonial stem cells), pubertal onset of meiosis, and haploid male germ cell differentiation (i.e., spermiogenesis) (4). Unique cellular processes are often accompanied by a more complex yet unique transcriptome, which may explain why the testis expresses more genes than any other organs in the body, with the possible exception of the brain (5). Regulation of gene expression during spermatogenesis occurs at both transcriptional and post-transcriptional levels (6). As a post-transcriptional regulator, miRNAs are abundantly expressed in the testis and are required for spermatogenesis (7-11). miRNAs typically function at post-transcriptional levels by binding the complementary sequences in the untranslated regions (UTRs) of mRNAs – particularly in the 3'UTRs through the "seed sequence" (2nd-7th nucleotides)(12). Numerous miRNAs are subject to rapid evolution, probably in response to the accelerated rate of divergence of UTRs compared to the exonic sequences (13). Divergence of genomic sequences can be mediated by transposable elements (TEs), which are known as building blocks of the genome and mostly map to UTRs and intronic regions of protein-encoding genes (14). Each miRNA can bind numerous target mRNAs, and one mRNA can be targeted by multiple different miRNAs. This "one-to-multiple" relationship between miRNAs and mRNAs amplifies their potential to coordinate gene expression in the cell (12). Moreover, miRNA genes often exist in clusters, which are transcribed as a unit followed by nuclear and cytoplasmic cleavage events to generate individual miRNAs (12).

Multiple clusters of miRNA genes containing the same seed sequences are categorized as a miRNA family, and miRNAs within the same family likely evolved from a common ancestor sequence (15). Of great interest, many of the testis-enriched miRNA clusters map to the X chromosome (16). Sex-linked genes are generally subject to the male germline-specific phenomenon called meiotic sex chromosome inactivation (MSCI), which silences transcription during most, if not all, of meiosis (16). Indeed, prior to 2009, there were no confirmed reports of any sex-linked genes escaping the repressive effects of MSCI. Surprisingly, however, we found that many X-linked miRNA genes do escape MSCI, suggesting that they may contribute to particularly important functions during spermatogenesis (16). This notion has since been tested by the generation of knockouts (KOs) of individual miRNA genes or individual clusters of miRNA genes normally expressed during spermatogenesis. However, these KOs resulted in minimal, if any, phenotypic effects and did not appear to impede normal spermatogenesis or male fertility (15). This left the field facing multiple unanswered questions, including 1)

why do so many X-linked miRNAs express uniquely or preferentially during spermatogenesis and escape MSCI, 2) what's their origin, and 3) how and why did they evolve rapidly?

To address these questions and better understand the functional role played by these X-linked miRNAs, we investigated the evolutionary history of this unique miRNA family and also generated KOs of individual, paired, triple, quadruple, or quintuple sets of miRNA clusters within this family and tested the effects on male fertility, initially by standard monandrous mating assays. Consistent with previous efforts to inactivate miRNA genes in *C. elegans* and mice (15, 17-19), KOs of either individual members or individual clusters of the *miR-506* family induced no discernable phenotypes and did not impact male fertility. This may reflect the level of functional redundancy inherent within the members and clusters of this miRNA family. It was only when four or more clusters of the *miR-506* family were ablated that relevant phenotype became detectable, which was manifested in the form of reduced litter size despite normal sperm counts, motility, and morphology. Interestingly, the most common male fertility testing for lab rodents is based on a monandrous mating scheme, *i.e.,* one fertility-proven female mated with one male. However, there are additional aspects of male fertility in many mammalian species, particularly those that are normally litter-bearing. In the wild, litter-bearing females often mate with multiple different males such that a single litter may include pups sired by more than one male (20, 21). This polyandrous mating introduces the potential for additional aspects of male reproductive fitness to accrue, one of which involves sperm competition (20, 21). Sperm competition can occur when sperm from more than one male are present in the female reproductive tract simultaneously, such that they then compete to fertilize each oocyte (22). Sperm competition has been now recognized as a significant evolutionary force directly impacting male reproductive success (22). Using experiments that mimic polyandrous mating, we found that the quinKO male mice indeed displayed compromised sperm competition. Hence, the X-linked *miR-506* family miRNAs appear to function to finetune spermatogenesis to enhance sperm competition and, consequently, male reproductive fitness.

Results

X-linked *miR-506* family miRNAs flanked by two highly conserved protein-coding genes *Slitrk2* and *Fmr1* rapidly evolved across species. X-linked genes are generally more divergent between species than autosomal ones, a phenomenon known as the "faster-X effect" (23). However, despite a high degree of conservation of two protein-coding genes, *Slitrk2* and *Fmr1*, on the X chromosome across species, the miRNA genes located between these two loci are divergent among clades across the eutherian mammals (15, 24). Through tracing the evolution of this genomic region, we found that *Slitrk2* and *Fmr1* mapped to chromosome 4 (syntenic with mammalian X chromosome) in zebrafish and birds, but to the X chromosome in most mammals, with divergence and multiplication of numerous miRNA genes that belong to the *miR-506* family in between (Fig. 1A and Table S1). By mapping these miRNAs of various species using the UCSC genome

167 browser (25), we found that all members of the *miR-506* family are located in a
168 region flanked by *Slitrk2* and *Fmr1* (Fig. 1A). Consistent with previous reports (15,
169 24), *Slitrk2* and *Fmr1* are usually on the positive strand, whereas the *miR-506*
170 family miRNAs are in the reverse orientation (Fig. 1A). Based on the location of
171 these miRNAs, we named the miRNAs proximal to *Slitrk2* (*miR-892c ~ miR-891a*
172 in humans) and *Fmr1* (*miR-513c ~ miR-514a3* in humans) SmiRs (*Slitrk2*-proximal
173 miRNAs) and FmiRs (*Fmr1*-proximal miRNAs), respectively.
174 To evaluate the sequence conservation of these miRNAs across species,
175 we adopted the Multiz Alignment and Conservation pipeline, which utilizes
176 PhastCons and PhyloP algorithms (25), to search miRNA datasets from 100
177 different species using the human genome as a reference (Fig. 1B and Fig. S1).
178 The mean values of phyloP and phastCons of the *Fmr1* and *Slitrk2* coding
179 sequences (CDS) were ~4.9 and ~0.97, respectively (Fig. 1C and D), indicating
180 that these regions are highly conserved. In contrast, the mean values of phyloP
181 and phastCons of SmiRs were ~1.0 and ~0.002, respectively, and those of FmiRs
182 are ~0.03 and ~0.11, respectively (Fig. 1 C and D). The phyloP and phastCons
183 values of the CDS, FmiRs, and SmiRs are significantly different from each other
184 (adjusted p-value < 0.05, Kruskal-Wallis test) (Fig. 1 C and D), indicating that
185 FmiRs and SmiRs are highly divergent, and SmiRs are more divergent than
186 FmiRs.
187 We then assessed the genomic sequence similarity among various species
188 using D-GENIES-based dot plot analyses (26). Although the *Slitrk2-Fmr1* genomic
189 regions were highly variable among different species, the sequences within some
190 clades shared a high degree of similarities, e.g., primates (rhesus monkeys,
191 chimpanzees, and humans), *cetartiodactyla* (sheep and cows), *rodentia* (e.g., mice
192 and rats) and *carnivora* (e.g., dogs and cats) (Fig. S2A). Although the *miR-506*
193 family miRNAs were highly divergent, some orthologs displayed a higher degree of
194 sequence conservation (Fig. S2 B and C), e.g., the SmiRs within primates were
195 similar (Fig. S2B); *miR-891a* and *miR-891b* in rhesus monkeys were similar to
196 *miR-891b* and *miR-891a* in humans and chimpanzees, respectively, and *miR-892b*
197 in chimpanzees was homologous to *miR-892c* in humans and rhesus monkeys in
198 terms of sequence and location (Figs. S2B). The FmiRs, including *miR-201*
199 (assigned as *Mir-506-P1* (paralogue 1) in MirGeneDB (27)), *miR-547* (*Mir-506-P2*),
200 and *miR-509* (*Mir-506-P7*) in mice and rats, are orthologues of *miR-506*, *miR-507*
201 and *miR-509* in humans, respectively (Fig. S2C, and Table S1). Of interest,
202 although the *miR-506* family miRNAs are highly divergent, the seed sequences of
203 some miRNAs, such as *Mir-506-P6* and *Mir-506-P7*, remain conserved (Fig. S2C),
204 and these miRNAs represent the dominant mature miRNAs (28). It is noteworthy
205 that the majority of the substitutions among the *miR-506* family are U→C and
206 A→G (Fig. S2 B and C). Furthermore, we analyzed the conservation of the *miR-*
207 *506* family in modern humans using data from the 1000 Genomes Project (1kGP)
208 (Fig. 1 E and F) (29). We compared SmiRs, FmiRs, and all miRNAs with
209 pachytene piRNAs, which are known to be highly divergent in modern humans but
210 barely exert any biological functions (30). Of interest, the derived allele frequency
211 (DAF) and mean nucleotide diversity (MND) of FmiRs and all miRNAs were
212 significantly smaller than that of the pachytene piRNAs, whereas SmiRs were

213 significantly smaller than all miRNAs (Fig. 1 E and F) (adjusted p-value < 0.05,
214 Kruskal-Wallis test), suggesting that the *miR-506* family miRNAs are more
215 conserved than pachytene piRNAs in modern humans. Taken together, these data
216 indicate that the X-linked *miR-506* family, although rapidly evolving as a whole,
217 contains both divergent and conserved miRNAs, suggesting both conserved and
218 novel functions across species.
219
220 **X-linked *miR-506* family miRNAs are derived from MER91C DNA**
221 **transposons.** To visualize the family history of these miRNAs, we built a
222 phylogram for the *miR-506* family (Fig. S3). The phylogram suggests that these
223 miRNAs shared a common ancestor and that the FmiRs emerged earlier than the
224 SmiRs, which is also supported by the fact that some FmiRs exist in green sea
225 turtles (Figs. S1 and S3). These data suggest that the *miR-506* family miRNAs
226 arose much earlier than previously thought (25). The two subfamilies, FmiRs and
227 SmiRs, despite their common ancestors, may have evolved at different paces and
228 thus, might be functionally divergent.
229 Studies have shown that transposable elements (TEs) drive evolution
230 through transpositions (31). CRISPR-Cas9/Cas12a genome editing can induce
231 irreversible small indels at the cutting sites (also called "scars")(32), which have
232 been used for lineage tracing (33). Inspired by this strategy, we attempted to trace
233 the evolution of the *miR-506* family miRNAs by searching the transposon database
234 for the transpositional "scars" (partial TE sequences) after transposition. To search
235 the potential TE sources of the *miR-506* family miRNAs, we downloaded all
236 transposons in the human, horse, dog, and guinea pig genomes and aligned them
237 to their corresponding *miR-506* family miRNAs using BLAST (Basic Local
238 Alignment Search Tool) (34). The nonautonomous MER91C DNA transposon
239 (~100-150 million years) (35, 36) was the only transposon that aligned to FmiRs of
240 the *miR-506* family (> 94% identical matches) in all four species (Table S2).
241 Given that the FmiRs (e.g., *miRs-506~509*) emerged much earlier than the
242 SmiRs (fig. S1 and fig. S3) and that *miR-513* (belonging to FmiRs) and SmiRs
243 (including *miR-891a* and *miR-891b*) share a common ancestor (Fig. S3), we
244 reasoned that the X-linked *miR-506* family might be derived from the *MER91C*
245 DNA transposon. To test this hypothesis, we first aligned the X-linked *miR-506*
246 family miRNAs from several species to a human MER91C DNA transposon.
247 Indeed, numerous FmiRs of almost all species analyzed aligned to the MER91C
248 DNA transposon despite few mismatches (Fig. 2A). The phylogenetic tree further
249 confirmed that the MER91C is the sister group of the *miR-506* family miRNAs (Fig.
250 2B and Fig. S4), suggesting that the MER91C DNA transposon is the likely source
251 of the older *miR-506* family miRNAs. Further supporting this notion, the MER91C
252 DNA transposons could form hairpin structures, which is a prerequisite for miRNA
253 biogenesis (Fig. 2C and Fig. S5A). Moreover, analyses of the testis small RNA
254 datasets from humans, marmosets, dogs, and horses revealed the peaks
255 corresponding to these miRNAs (Fig. 2D and Fig. S5A). Finally, by overexpressing
256 several MER91C DNA regions randomly selected from humans, dogs, and horses
257 in HEK293T cells, we found that these DNA regions were indeed capable of
258 producing miRNAs (Fig. 2 E and F, and Fig. S5 B and C). Co-expression of

MER91C DNA regions and AGO2 significantly increased the abundance of human MER91C miRNAs, as compared with overexpression of MER91C DNA regions alone (Fig. 2 E and F), suggesting that these miRNAs could be loaded onto and protected by AGO2. Taken together, these results indicate that the X-linked *miR-506* miRNAs were originally derived from the MER91C DNA transposon.

X-linked *miR-506* family miRNAs are predominantly expressed in spermatogenic cells and sperm. Several previous studies have shown that the X-linked *miR-506* family miRNAs are predominantly expressed in the testis of multiple species (15, 16, 24, 37-39). By analyzing the publicly available small RNA sequencing (sRNA-seq) datasets from multiple species, including humans, rhesus monkeys, mice, rats, rabbits, dogs, and cows (27, 40, 41), we further confirmed that *miR-506* family miRNAs were indeed highly abundant in the testis, but barely expressed in other organs (Fig. S6, and Table S3). To further determine whether these miRNAs are expressed in male germ cells in rodent testes, we conducted sRNA-seq using pachytene spermatocytes, round spermatids, and sperm purified from adult mice (Fig. S7A and Table S3). Consistent with previous data (15, 16, 24), these miRNAs were abundantly expressed in spermatogenic cells in murine testes (Fig. 3A). ~80% of these miRNAs were significantly upregulated (FDR < 0.05) when pachytene spermatocytes developed into round spermatids, and ~83.3% were significantly upregulated (FDR < 0.05) when developed into cauda sperm (Fig. 3A). By analyzing the publicly available sRNA-seq datasets from humans (42), marmosets (37) and horses (43), we determined the expression patterns of the *miR-506* family miRNAs in the testes of these species (Fig. 3 B, C and D, Table S3). The significantly increasing abundance of the SmiRs from immature to mature testes in horses (43) supports the elevated expression in haploid male germ cells (round, elongating/elongated spermatids, and sperm) compared to meiotic male germ cells (spermatocytes) (Fig. 3D). More interestingly, the SmiRs and FmiRs appear to be differentially expressed in the testes of various species, e.g., relative levels of the SmiRs were greater than those of the FmiRs in mice (Fig. 3 A and Table S3). Both the SmiRs and FmiRs were highly expressed in horses (Fig. 3D and Table S3). Levels of the FmiRs in marmoset (37) and human testes (42) were much greater than in those of the SmiRs (Fig. 3 B and C and Table S3). Overall, the *miR-506* family miRNAs are abundant in the testis and predominantly expressed in haploid male germ cells, *i.e.,* spermatids and spermatozoa.

Ablation of X-linked *miR-506* family miRNAs compromises male fertility due to reduced sperm competitiveness. To define the physiological role of the *miR-506* family, we sequentially deleted these miRNA genes using CRISPR-Cas9-based genome editing (Fig. 4A) (15). We first generated the KO mice lacking either the *miR-883* single cluster (*miR-883* sKO) or the *miR-465* single cluster (*miR-465* sKO) (Fig. 4A), as these two clusters are the most abundantly expressed in the mouse testes (Fig. 3 A and Fig. S6). No discernable defects were observed, and these KO males developed normally and were fertile (Fig. 4 B and C). On the *miR-883* sKO background, we further deleted the *miR-741* cluster, which we termed

305 double KO (dKO) (Fig. 4 A), but no discernable abnormalities were observed in the
306 dKO males either (Fig. S7 B-E). On the dKO background, we next deleted either
307 the *miR-465*, termed triple KO (tKO), or the *miR-471* and *miR-470* clusters, termed
308 quadruple KO (quadKO) (Fig. 4 A). Lastly, we ablated the *miR-465* cluster on the
309 quadKO background, named quintuple KO (quinKO) (Fig. 4 A). To reduce potential
310 off-target effects due to multiple rounds of CRISPR-Cas9 targeting, we also
311 generated a KO mouse line with only 4 guide RNAs (gRNAs) flanking the SmiR
312 region, named X-linked SmiR KO (XS) (Fig. 4 A). The XS mice were genetically
313 equivalent to the quinKO mice, and phenotypically identical to quinKOs (Fig. 4 B
314 and C), suggesting the phenotype observed was not due to the accumulating off-
315 target effects. To further exclude the potential off-target effect, all KO mouse
316 strains were backcrossed with WT C57BL/6J mice for at least 5 generations before
317 data collection. In addition, T7 endonuclease I (T7EI) assays showed no
318 discernable off-target effects in the quinKO mice (Fig. S7F).
319 While the litter size was still comparable between the tKO and WT
320 control mice, the quadKO, quinKO, and XS males produced significantly
321 smaller litters (~5 *vs.* ~8 pups/litter) (adjusted p-value < 0.05, One-Way
322 ANOVA) (Fig. 4 B). Of interest, no significant changes were detected in litter
323 interval, testis weight or histology in any of the four types of KOs, as
324 compared to WT controls (Fig. 4 C, D and G). Computer-assisted sperm
325 analyses (CASA) revealed no significant differences in sperm counts and
326 motility parameters among the four types of KOs (Fig. 4 E and F and Fig.
327 S7H). Overall, there appears to be an inverse correlation ($R^2 = 0.9139$, $p <$
328 0.05, F-test) between the number of miRNAs inactivated and the litter size
329 (Fig. S7G). Interestingly, several human studies have correlated the
330 dysregulated *miR-506* family miRNAs with impaired male fertility due to
331 maturation arrest and oligo-asthenozoospermia (Table S4)(44-48). These
332 data suggest that the *miR-506* family may play an important role in
333 spermatogenesis and male fertility.
334 Most of the protein-coding genes that are exclusively or preferentially
335 expressed in the testis with an essential role in spermatogenesis are highly
336 conserved across species (49). Despite their male germ cell-predominant
337 expression, the *miR-506* family miRNAs appear to have evolved rapidly to
338 diverge their sequences, suggesting that these miRNAs might control
339 certain "non-conserved" aspects of spermatogenesis, leading to enhanced
340 sperm competitiveness for male reproductive success. Supporting this
341 hypothesis, previous reports have documented that females of most species
342 throughout the animal kingdoms mate with multiple males before
343 pregnancy, suggesting that sperm competition may serve as a selection
344 mechanism to bias the birth of offspring sired by the males with more
345 competitive sperm (20, 21). Studies have also shown female rodents in the
346 wild mate with multiple males and produce litters of mixed paternity, and
347 that pups born to the females following such polyandrous mating display
348 greater survival rates than those produced from females following
349 monandrous mating (50). Given that CASA detected no difference in
350 swimming patterns between quinKO and WT sperm (Fig. S7H), we next

351 carried out sperm competition experiments that mimic polyandrous mating in the
352 wild. Since the *miR-506* family miRNAs are X-linked, the Y sperm from the quinKO
353 mice are genetically indistinguishable from those of WT controls. We, therefore,
354 adopted the mTmG male mice (51) for sperm competition experiments because
355 the embryos or offspring fathered by the mTmG males can be easily identified
356 based on the constitutively expressed membrane-tagged tomato red (mT)
357 fluorescence and/or PCR genotyping.
358 We first conducted sequential mating with two mating events ~6-8 hours
359 apart. Interestingly, all of the pups born were fathered by mTmG males (n=8) when
360 the WT females were mated first with mTmG males and subsequently with the
361 quinKO males. In contrast, when the WT females were mated first with quinKO
362 males and subsequently with mTmG males, ~89% of the pups born were fathered
363 by quinKO males, and the remaining ~11% of pups were from mTmG males
364 (n=28) (Fig. 4 H-I). It is noteworthy that in the sequential mating experiments, the
365 two coituses occurred ~6-8 hours apart due to practical reasons, whereas in the
366 wild, polyandrous mating may take place much faster. To better mimic polyandrous
367 mating *in vitro*, we mixed the WT and quinKO sperm in different ratios and used
368 the mixed sperm to perform *in vitro* fertilization (IVF). MII oocytes fertilized by
369 mTmG, quinKO, or a mixture of two types of sperm at three ratios (mTmG: quinKO
370 = 1:1, 4:1 and 1:4) all displayed comparable rates at which fertilized oocytes
371 developed into blastocysts (Fig. S7I). Interestingly, when a 1:1 ratio (mTmG
372 sperm: quinKO sperm) was used, ~73% of the resulting blastocysts were derived
373 from mTmG sperm, whereas the remaining ~27% were from quinKO sperm
374 (n=179) ($p < 0.0001$, Chi-squared test) (Fig. 4 J). When a 4:1 sperm ratio (mTmG:
375 quinKO) was used, ~92% of the blastocysts were from mTmG sperm and only 8%
376 were from quinKO sperm (n=170) ($p < 0.05$, Chi-squared test) (Fig. 4 J). In
377 contrast, when a 1:4 sperm ratio (mTmG: quinKO) was used, blastocysts derived
378 from mTmG and quinKO sperm represented ~28% and ~72% of the total,
379 respectively (n=135) (Fig. 4 J). We also performed co-artificial insemination (AI)
380 using mTmG and quinKO sperm. When a 1:1 sperm ratio (mTmG: quinKO) was
381 used, ~62.5% of the embryos (n=96) were derived from mTmG sperm ($p< 0.05$,
382 Chi-squared test) (Fig. 4 K). When a 1:4 ratio was used, ~35.3% of the embryos
383 (n=17) were from the mTmG mice (Fig. 4K and Fig. S7J). Together, these results
384 indicate that the quinKO sperm are less competitive than the control mTmG sperm
385 both *in vivo* and *in vitro*. Previous studies suggest that sperm aggregation and
386 midpiece size might be involved in sperm competitiveness (52, 53), but no
387 changes in these two parameters were observed in the quinKO sperm (Fig. S7 K
388 and L). Although the blastocyst rate (out of 2-cell embryos) of quinKO was
389 comparable to that of the mTmG mice, the 2-cell rates (out of zygotes) were
390 significantly reduced ($p< 0.05$, paired t-test) in the quinKO mice (~39%, n=67)
391 when compared to the mTmG mice (~88%, n=58) (Fig. S7M), implying that the
392 quinKO sperm is indeed less efficient in fertilizing eggs and/or supporting early
393 embryonic development, especially the first cleavage of the zygotes.
394
395 **X-linked *miR-506* family miRNAs mostly target the genes involved in**
396 **spermatogenesis and embryonic development and compensate for each**

397 **other**. To identify the target genes of these X-linked *miR-506* family miRNAs, we
398 performed RNA-seq analyses using testis samples from the five types of KOs
399 (*miR-465* sKO, dKO, tKO, quadKO, and quinKO) (Fig. S8A and Table S5).
400 Comparisons between the KO and WT testes revealed thousands of differentially
401 expressed genes (DEGs) (fold change ≥ 2, FDR< 0.05, Fig. S8A and Table S5).
402 The DEGs identified were then compared with the predicted *miR-506* target genes
403 using four different databases, including TargetScan (54), microrna.org (55),
404 miRWalk (56) and mirDB (57), to predict the differentially expressed targets (DETs)
405 of the *miR-506* family miRNAs (Fig. S8A and Table S5). We obtained 2,692, 2,028,
406 1,973, 3,405, and 1,106 DETs from *miR-465* sKO, dKO, tKO, quadKO and quinKO
407 testes, respectively. GO terms of DETs from each KO testis revealed that the
408 DETs were mostly involved in embryonic development, response to stimulus,
409 centrosome cycle, epithelium morphogenesis, organelle organization, cell
410 projection, RNA metabolic process, and DNA repair (Fig. S8B). The 431 DETs
411 identified to be shared across all five KO testes were also enriched in similar
412 pathways (Fig. 5 An and B). Several genes, including *Crisp1*, *Egr1*, and *Trpv4*,
413 were selected for validation using qPCR, Western blots and luciferase-based
414 reporter assays. Consistent with the RNA-seq data, qPCR showed that *Crisp1*,
415 *Egr1*, and *Trpv4* were significantly downregulated in the quinKO testes (Fig. S8C).
416 *CRISP1* is enriched in the sperm principle piece and head (Fig. S8D). Western
417 blots also confirmed that *CRISP1* is downregulated in the quinKO testis when
418 compared to the WT testis (Fig. S8E). Luciferase assays further confirmed that
419 *Egr1* and *Crisp1* are targets of the *miR-506* family members (Fig. S8 F and G).
420 *Egr1* 3'UTR luciferase activity was upregulated by *miR-465c*, while downregulated
421 by *miR-743b* (Fig. S8F). *miR-465a*, *miR-465c*, *miR-470*, *miR-741*, and *miR-743a*
422 upregulated *Crisp1* 3'UTR luciferase activity, while *miR-743b* exerted the opposite
423 effect (Fig. S8G). Of interest, KO of *Crisp1* in mice or inhibition of *CRISP1* in
424 human sperm appears to phenocopy the quinKO mice (58, 59). Specifically, sperm
425 motility in the *Crisp1* KO mice is comparable to that in WT mice, but their ability to
426 penetrate the eggs was reduced in the *Crisp1* KO mice (58); a similar effect was
427 also observed in human sperm treated with anti-hCRISP1 antibody (59).
428 The inverse correlation between the number of miRNAs inactivated and the
429 severity of the phenotype strongly hints that these miRNAs compensate for each
430 other (Fig. 4B and Fig. S7G). To test this hypothesis, we performed sRNA-seq on
431 four KO (*miR-465* sKO, tKO, quadKO and quinKO) testes. The sRNA-seq data
432 showed that these miRNAs were no longer expressed in the corresponding KOs,
433 confirming the successful deletion of these miRNAs in these KOs (Fig. 5C and
434 Table S6). Interestingly, in *miR-465* sKO testes, *miR-201, miR-547, miR-470*, *miR-*
435 *471*, *miR-742, miR-871*, *miR-881, miR-883a,* and *miR-883b* were all significantly
436 upregulated (FDR < 0.05). Similarly, *miR-201, miR-547, miR-470*, *miR-471*, *miR-*
437 *871,* and *miR-883b* were all significantly upregulated in the tKO testes (FDR <
438 0.05); *miR-201* and *miR-547* were all significantly upregulated in the quadKO and
439 the quinKO testes (FDR < 0.05) (Fig. 5 C and Table S6). These results support the
440 notion that genetic compensation exists among the X-linked *miR-506* family
441 miRNAs.
442

443 **Rapid evolution of the *miR-506* family is not driven by the increased**
444 **complexity of 3'UTRs of the conserved targets but rather adaptive to**
445 **targeting more genes.** To minimize false positives, the DETs in mice were
446 selected using the following criteria: (1) dysregulated by fold change ≥ 2 and FDR
447 < 0.05. (2) Falling within the predicted targets. (3) Intersected with at least two
448 different KO mouse samples. Using the 3,043 DETs identified in mice as a
449 reference, we searched the predicted targets of the *miR-506* family miRNAs in rats
450 and humans to determine if these target genes were shared across species (Fig.
451 5A). While 2,098 (~69%) target genes were shared among all three species, 2,510
452 (~82%) were common to both humans and mice, and 2,202 (~72%) were shared
453 between mice and rats (Fig. 6 A and Table S7). To test the accuracy of the
454 predicted targets, we selected several genes in humans and performed luciferase
455 assays using human *miR-506* family miRNAs and their corresponding target genes
456 (Fig. S9 A and B). Among these targets, *Crisp1* and *Fmr1* were shared among
457 humans, mice, and rats, and confirmed to be targeted by the *miR-506* family
458 miRNAs in mice (Fig. S8C, S8E, S8G, and ref 11 and 15)(11, 15). Luciferase
459 assays also confirmed that human *CRISP1* (*hCRISP1*) and *FMR1*(*hFMR1*) were
460 targets of the *miR-506* family, and *miR-510* and *miR-513b* both could activate
461 *hCRISP1* 3'UTR luciferase activity (Fig. S9A), whereas *miR-509-1*, *miR-509-2*,
462 *miR-509-3*, *miR-513b*, *miR-514a*, and *miR-514b* could enhance *hFMR1* 3'UTR
463 luciferase activity (Fig. S9B). These results confirmed the accuracy of our
464 predicted targets in humans.
465
466 We considered two likely explanations for the paradox where the majority of their
467 target genes were shared across species despite the rapid evolution of the *miR-*
468 *506* family miRNAs: 1) The 3'UTR sequences in extant target genes became
469 increasingly divergent during evolution such that the *miR-506* family miRNAs had
470 to adapt to maintain their ability to bind these 3'UTRs; or 2) that the *miR-506* family
471 miRNAs evolved rapidly in a manner that allowed them to target mRNAs encoded
472 by additional genes involved in spermatogenesis. To distinguish the two
473 possibilities, we first compared the extent of similarities among the 3'UTR
474 sequences of the 2,510 shared target genes between humans and mice (Fig. 6 A
475 and Table S7). We adopted the PhyloP scores to measure the evolutionary
476 conservation at individual nucleotide sites in the 3'UTRs of the shared target
477 genes. The overall conservation appeared to be greater in the regions targeted by
478 *miR-506* family miRNAs than in the non-target regions in the 3'UTRs of the shared
479 target genes in both mice and humans (Fig. 6 B and C) with few exceptions (Fig. 6
480 D and E) ($p < 0.05$, t-test). These data suggest that the regions targeted by the X-
481 linked *miR-506* family miRNAs are under relatively stronger purifying, rather than
482 adaptive, selection. We then tested the second hypothesis that the rapid evolution
483 of the *miR-506* family resulted in more extant mRNAs being targeted by these
484 miRNAs. We first compared the average target numbers of each *miR-506* family
485 miRNA between humans and mice using the 2,510 shared targets between
486 predicted targets in humans and the dysregulated targets in mice (Fig. 6 A and F).
487 Among these shared targets, the human *miR-506* family members could target
488 ~1,268 unique transcripts per miRNA, whereas the murine *miR-506* family

489 members could only target ~1,068 ($p < 0.05$, t-test) (Fig. 6 F), indicating that the
490 *miR-506* family miRNAs target more genes in humans than in mice. Furthermore,
491 we analyzed the number of all potential targets of the *miR-506* family miRNAs
492 predicted by the aforementioned four algorithms among humans, mice, and rats.
493 The total number of targets for all the X-linked *miR-506* family miRNAs among
494 different species did not show significant enrichment in humans (Fig. S9C),
495 suggesting the sheer number of target genes does not increase in humans. We
496 then compared the number of target genes per miRNA. When comparing the
497 number of target genes per miRNA for all the miRNAs (baseline) between humans
498 and mice, we found that on a per miRNA basis, human miRNAs have more targets
499 than murine miRNAs ($p<0.05$, t-test) (Fig. S9D), consistent with higher biological
500 complexity in humans. This became even more obvious for the X-linked *miR-506*
501 family ($p<0.05$, t-test) (Fig. S9D). In humans, the X-linked *miR-506* family, on a per
502 miRNA basis, targets a significantly greater number of genes than the average of
503 all miRNAs combined ($p<0.05$, t-test) (Fig. S9D). In contrast, in mice, we observed
504 no significant difference in the number of targets per miRNA between X-linked
505 miRNAs and all of the mouse miRNAs combined (mouse baseline) (Fig. S9D).
506 These results suggest that although the sheer number of target genes remains the
507 same between humans and mice, the human X-linked *miR-506* family targets a
508 greater number of genes than the murine counterpart on a per miRNA basis. We
509 also investigated the number of *miR-506* family miRNA targeting sites within the
510 individual target genes in both humans and mice, but no significant differences
511 were found between humans and mice (Fig. 6 G). To determine whether increased
512 target sites in humans were due to the expansion of the MER91C DNA
513 transposon, we analyzed the MER91C DNA transposon-containing transcripts and
514 associated them with our DETs. Of interest, 28 human and 3 mouse mRNAs
515 possess 3'UTRs containing MER91C DNA sequences, and only 3 and 0 out of
516 those 28 and 3 genes belonged to DETs in humans and mice, respectively (Fig.
517 S9E), suggesting a minimal effect of MER91C DNA transposon expansion on the
518 number of target sites. Taken together, these results suggest the human X-linked
519 *miR-506* family has been subjected to additional selective pressure, causing them
520 to exert additional regulatory functions by targeting more mRNAs expressed during
521 spermatogenesis (Fig. 6 H).
522
523
524 **Discussion**
525
526 Successful reproduction is pivotal for the perpetuation of species, and sperm are
527 constantly facing selective pressures (60). To enhance their chance to fertilize
528 eggs, sperm need to adapt accordingly, and miRNAs-mediated regulation of gene
529 expression in spermatogenesis provides a rapidly adaptable mechanism toward
530 this end. Although miRNAs were initially believed to be evolutionarily conserved,
531 the number of non-conserved miRNAs has been steadily increasing (61). Among
532 the non-conserved miRNAs, many are derived from TEs, suggesting that TEs may
533 serve as a major source of miRNA sequences (61). The delayed recognition of TE-
534 derived miRNAs, in part, results from the fact that repetitive sequences were

535 usually excluded during the computational annotation of miRNAs. In theory, TEs
536 can serve as a good donor of miRNA sequences for the following reasons: 1) TEs
537 are ubiquitous and abundant in the genome and are known to contribute to the
538 regulatory elements of the coding genes, e.g., UTRs (62). As one of the regulatory
539 factors that target mainly the 3'UTRs, TE-derived miRNAs can regulate a larger
540 number of mRNAs with multiple miRNA-targeting sites. 2) TEs are among the most
541 rapidly evolving sequences in the genome (63) and thus, can continuously produce
542 species/lineage-specific miRNA genes to diversify their regulatory effects.
543 Consistent with these notions, the present study provides evidence supporting that
544 the *miR-506* family miRNAs originated from the MER91C DNA transposons. Of
545 more interest, the *miR-506* family miRNAs, despite their rapid evolution, are all
546 expressed in spermatogenic cells in the testis and sperm, supporting a lineage-
547 specific functional diversification of TE-derived miRNAs. In fact, the *miR-506* family
548 miRNAs were among the first reported TE-derived miRNAs because of their
549 location in a small region of the X chromosome and their confined, abundant
550 expression in the testis (16). RNAs are much less abundant in sperm than in
551 somatic or spermatogenic cells (~1/100) (64). Sperm-borne small RNAs represent
552 a small fraction of total small RNAs expressed in their precursor spermatogenic
553 cells, including spermatocytes and spermatids (64). Therefore, when the same
554 amount of total/small RNAs are used for quantitative analyses, sperm-borne small
555 RNAs (e.g., *miR-506* family miRNAs) would be proportionally enriched in sperm
556 compared to other spermatogenic cells.
557 It has been demonstrated that under certain circumstances, genes that
558 evolve under sexual conflicts tend to move to the X chromosome, especially when
559 they are male-beneficial, female-deleterious, and act recessively (65, 66). The X
560 chromosome is enriched with genes associated with male reproduction (15, 16,
561 67). The rapid evolution of the X-linked *miR-506* family strongly suggests that
562 these miRNA genes were under selection to expand and diversify their regulatory
563 effects on spermatogenesis. Indeed, our data strongly suggest that targeting new
564 mRNAs was likely the driving force for the rapid evolution of the *miR-506* family of
565 miRNAs. However, expansion and sequence divergence of the X-linked *miR-506*
566 family may simply reflect natural drifting without functional significance, similar to
567 some of the pachytene piRNA clusters (30). We argue that the neutral drifting
568 theory may not be true to the *miR-506* family for the following reasons: 1) Despite
569 highly divergent overall sequences of the *miR-506* family, some miRNAs share the
570 same seed regions across multiple species, suggesting that these regions may
571 undergo strong selections. 2) The *miR-506* family miRNAs, especially the FmiRs,
572 are highly conserved in modern humans, implying a strong selection of these
573 miRNAs. 3) Knockout the *miR-506* family (either quinKO or XS) results in male
574 subfertility, reflecting a biological function. 4) The quinKO sperm are less
575 competitive than the WT sperm both in *vivo* and *in vitro*. 5) Several human studies
576 have linked the dysregulation of the *miR-506* family with male infertility/subfertility
577 (44-48). Similarly, one study in *Drosophila* also showed that the rapidly evolving
578 testis-restricted miRNAs underwent adaptive evolution rather than neutral drifting
579 (68).

580 Since TEs are abundant in UTRs, TE-derived miRNAs can target a much greater
581 number of mRNAs than those derived from distinct non-repetitive genomic loci
582 (61). Indeed, thousands of the dysregulated genes detected in the *miR-465* sKO,
583 dKO, tKO, quadKO and quinKO testes are involved in multiple pathways of
584 spermatogenesis. By analyzing the sequence divergence, we noticed that the
585 most common sequence substitutions among all of the *miR-506* miRNA sequences
586 were U-to-C and A-to-G, which were likely mediated by ADARs (adenosine
587 deaminases acting on RNA) that can change A to I (which is functionally equivalent
588 to G) (69). Interestingly, ~90% of the A-to-I editing appears to have occurred in Alu
589 elements (belonging to the SINE family), and some of the edits occurred in
590 miRNAs (69). Since G and U can form the so-called G-U wobble base pair (70),
591 those U-to-C or A-to-G substitutions can, in theory, target similar sequences and
592 exert regulatory functions (71), suggesting that the evolving miRNA sequences
593 could target not only the original sequences but also new sites with similar
594 sequences. Consistent with this notion, the predicted target genes of the *miR-506*
595 family in mice can also be found in rats and humans, suggesting the target genes
596 are shared across species despite the quick divergence of the miRNA sequences
597 across species. This is also supported by our data showing that the binding sites
598 for the *miR-506* family of miRNAs are more conserved than the surrounding, non-
599 targeting regions in the 3'UTRs of the predicted target mRNAs. Furthermore, seed
600 sequences among some *miR-506* family miRNAs remain the same despite the
601 high divergence of these miRNAs, and these conserved seed sequences appear to
602 be present in the dominant mature miRNAs. Thus, the seed region of these
603 miRNAs appears to have undergone strong selection. Supporting this notion,
604 previous studies have shown correlations between miRNA expression and the
605 evolution of miRNAs and target sites (72, 73). In general, miRNAs repress their
606 target gene expression. However, numerous studies have also shown that some
607 miRNAs, such as human *miR-369-3*, *Let-7*, and *miR-373*, mouse *miR-34/449* and
608 the *miR-506* family, and the synthetic miRNA *miRcxcr4*, activate gene expression
609 both *in vitro* (74, 75) and *in vivo* (11, 15, 76, 77). Earlier reports have shown that
610 these miRNAs can upregulate their target gene expression, either by recruiting
611 *FXR1*, targeting promoters, or sequestering RNA subcellular locations (11, 74, 75).
612 Of interest, miRNAs with the same seed sequences may exert divergent functions.
613 For example, the mature *miR-465a*, *miR-465b*, and *miR-465c* only have a few
614 mismatches outside of the seed region, but only *miR-465c* exerts functional
615 activation of *Egr1*. A similar effect has also been reported in the *miR-465* cluster on
616 the *Alkbh1* 3'UTR activity (18). Similarly, despite the same seed sequences in the
617 *miR-465* or *miR-743* cluster, *miR-465a* and *miR-465c* have differential activating
618 effects on the 3'UTR of *Crisp1*, and *miR-743a* and *miR-743b* exert opposite effects
619 on the *Crisp1* 3'UTR , further confirming their functional divergence. Therefore, the
620 sequences outside of the miRNA seed region may play an important role in their
621 functions, which have also been observed in *C. elegans* and human HEK293 cells
622 (18, 78, 79). To unequivocally demonstrate the physiological role of miRNAs, it
623 would be ideal to delete not only the miRNAs but also their binding sites in their
624 target transcripts *in vivo*. A few previous studies have established the miRNA:
625 target relationship by deleting the miRNA-binding sites in target transcripts in *C.*

626 *elegans*, *Drosophila*, and cell lines (80-82), but similar studies have not been
627 reported in mice or humans. The strategy may work in mRNAs with 3'UTRs
628 containing only one or two miRNA-binding sites, but for more complex 3'UTRs of
629 mRNAs in mice and humans that often contain multiple binding sites for the same
630 or different miRNAs, deletion of one miRNA binding site may not cause any
631 discernable effects as the loss of function can easily be compensated by other
632 miRNAs. Nevertheless, by deleting all five highly expressed clusters of the *miR-*
633 *506* family one by one, we were able to overcome the compensatory effects
634 among the family members/clusters and successfully revealed the physiological
635 role of this miRNA family. Based on small RNA-seq, some FmiRs, e.g., *miR-201*
636 and *miR-547*, were upregulated in the SmiRs KO mice, suggesting that this small
637 cluster may act in concert with the other 5 clusters and thus, worth further
638 investigation.
639 It is well-known that mice in the wild are promiscuous, and one female often
640 mates with multiple males sequentially, giving rise to polyandrous litters derived
641 from sperm from more than one sire (20, 21). Polyandrous mating establishes a
642 situation where sperm from multiple males co-exist in the female reproductive tract,
643 with the most competitive ones fertilizing eggs and producing offspring (22, 83).
644 Therefore, a male that may be fertile in the monandrous mating scheme may rarely
645 sire offspring in a polyandrous mating scenario, rendering this male functionally
646 "sub-fertile" or even "infertile." Therefore, sperm competitiveness reflects the
647 general reproductive fitness of the male (83). Although the quinKO males tend to
648 produce smaller litters under the monandrous mating scheme, their sperm counts,
649 sperm motility and morphology are indistinguishable from those of WT sperm. This
650 is not surprising given that miRNAs of the *miR-506* family most likely function to
651 control certain non-essential aspects of spermatogenesis. Sperm can be subject to
652 competition at multiple steps during fertilization, including their migration through
653 the female reproductive tract (cervix, uterine cavity and oviduct), binding the
654 cumulus-oocyte complexes, penetration of zona pellucida, etc. Therefore, IVF may
655 not be ideal for evaluating sperm competition in the real world as it bypasses
656 several key sites where sperm competition likely takes place. Artificial insemination
657 (AI) may represent a better way to assess sperm competition than IVF, but it is
658 probably less desirable than polyandrous mating for the following reasons: First, in
659 the wild, sperm from two males rarely, if not never, enter the female reproductive
660 tract simultaneously. We had tried to place two males into the cage with one
661 female, but the two males ended up fighting, and the submissive one never mated.
662 Second, sperm are delivered directly into the uterus or oviduct during AI (84, 85),
663 thus bypassing the potential sites for sperm competition (e.g., cervix and uterine
664 cavity). Although our breeding scheme also involves sperm competition, by
665 shortening the time between the two mating events in a laboratory setting, the
666 sequential mating method reported here may be further improved to better mimic
667 the natural polyandrous mating in the future. Moreover, future analyses of the
668 quinKO sperm may help identify biochemical or molecular biomarkers for sperm
669 competitiveness.
670 In summary, our data suggest that the *miR-506* family miRNAs are derived
671 from the MER91C DNA transposon. These miRNAs share many of their targets

672 and can compensate for each other's absence, and they work jointly through
673 regulating their target genes in spermatogenesis to ensure sperm competitiveness
674 and male reproductive fitness.
675
676 **Materials and Methods**

677 **Animal care and use**
678 All mice used in this study were on C57BL/6J background and housed in a
679 temperature- and humidity-controlled, specific pathogen-free facility under a light-
680 dark cycle (12:12 light-dark) with food and water *ad libitum*. Animal use protocol was
681 approved by the Institutional Animal Care and Use Committees (IACUC) of the
682 University of Nevada, Reno and The Lundquist Institute at Harbor-UCLA, and is in
683 accordance with the "Guide for the Care and Use of Experimental Animals"
684 established by the National Institutes of Health (1996, revised 2011).
685
686 **Generation of the knockout mice**
687 The single, double, triple, quadruple, and quintuple *miR-506* KO mice were
688 generated as previously described (15). Briefly, Cas9 mRNA (200 ng/μl) and gRNAs
689 flanking the *miR-506* family subclusters (100 ng/μl) were mixed and injected into the
690 cytoplasm of zygotes in the M2 medium. After injection, all embryos were cultured
691 for 1 h in KSOM+AA medium (cat. #MR-121-D, Millipore) at 37°C under 5% CO_2 in
692 the air before being transferred into 7–10-week-old female CD1 recipients. The *miR-*
693 *883* sKO or the *miR-465* sKO mice were first generated. After two rounds of
694 backcrossing with C57BL/6J mice, the *miR-741* cluster was knocked out on the *miR-*
695 *883* sKO background to generate dKO mice. After two rounds of backcrossing of the
696 dKO with C57BL/6J mice, the *miR-465* cluster or *miR-471* and *miR-470* clusters
697 were further deleted, thereby generating tKO or quadKO mice. Lastly, the *miR-465*
698 cluster was ablated on the quadKO background, leading to the production of quinKO
699 mice. The XS mice were generated by using only four gRNAs flanking the SmiR
700 region on the C57BL/6J background. All KO mice were backcrossed with the
701 C57BL/6J mice for at least five generations before collecting data.
702
703 **Sequential polyandrous mating**
704 Sequential polyandrous mating was carried out based on the ovulation time point
705 (10-13h after hCG) as previously described (84). Adult (8-12 weeks of age)
706 C57BL/6J females were injected (*i.p.*) with 7 IU PMSG at 8 p.m., followed by 7 IU
707 hCG 48h later. After hCG, the first male mouse was put into the cage of one female
708 from 8 p.m. to 6 a.m. the next day. The first plug was marked with a marker pen. A
709 second male mouse was then introduced into the cage of the plugged female, which
710 was checked every 30-40 min to identify a new plug (non-marked). Females that
711 were plugged twice were kept for producing pups for paternity analyses.
712
713 ***In vitro* fertilization (IVF)**
714 Adult (8-12 weeks) C57BL/6J female mice were first treated with 7 IU pregnant mare
715 serum gonadotropin (PMSG, Cat.# HOR-272, Prospecbio) through intraperitoneal
716 (*i.p.*) injection followed by *i.p.* injection of 7 IU hCG 48h later. Oocytes were collected
717 from the ampulla ~14 h after the hCG (Cat.# HOR-250, Prospecbio) treatment, and

718 the cumulus cells surrounding oocytes were removed by treatment with bovine
719 testicular hyaluronidase (1.5 mg/ml; Cat.# H3506, Sigma,) in M2 (Cat.# MR-015-D,
720 Millipore) at 37°C for 2 min. The cumulus-free oocytes were washed and kept in
721 equilibrated HTF (Cat# MR-070-D, Millipore) at a density of 20-30 oocytes per 60 µl
722 HTF at 37°C in an incubator with air containing 5% CO_2 prior to IVF. Cauda
723 epididymal sperm were collected in 100 µl of equilibrated HTF medium, allowing
724 spermatozoa to capacitate for ~30 min at 37°C in an incubator containing 5% CO2
725 air. After capacitation, spermatozoa (2 µl) were diluted by 10 folds and subjected to
726 computer-assisted sperm analyses (CASA) using the Sperm Analyzer Mouse Traxx
727 (Hamilton-Thorne). Based on the sperm concentration, an aliquot of 2.5 x 10^8
728 spermatozoa was added into each HTF-oocytes drop (~60µl) for IVF. ~4 h later,
729 zygotes were washed and cultured in KSOM+AA (Cat.# MR-121-D, Millipore) until
730 the blastocyst stage at 37°C in an incubator with air containing 5% CO_2. The 2-cell
731 embryos were counted 24-26 h after IVF, and blastocysts were counted and
732 analyzed under a fluorescence microscope 70-72 h after IVF.
733
734 **Artificial insemination (AI)**
735 At least 2-month-old female CD1 or C57BL/6J mice were administrated with 2.5 IU
736 of PMSG (Cat.# HOR-272, Prospecbio) at 5:30 PM 3 days before artificial
737 insemination, followed by 2.5 IU of hCG (Cat.# HOR-250, Prospecbio) at 5:00 PM 1
738 day prior to AI. The next morning at 8:00 AM, ~2-month-old mTmG and quinKO male
739 mice were sacrificed, the cauda epididymis was dissected, and fat tissue and blood
740 were removed before placing the cauda epididymis into 500µl or 150µl of
741 EmbryoMax® Human Tubal Fluid (HTF) (1X) (Cat.# MR-070-D, Millipore Sigma)
742 containing 4mg/ml BSA (Cat.# 12659-250GM, EMD Millipore Corp.) (HTF-BSA)
743 covered with 4 ml of mineral oil (Cat.# M8410-500ML, Sigma). Three incisions were
744 made on the cauda to allow sperm to swim out and to get capacitated for at least 30
745 min. 25 µl of mTmG and 25 µl quinKO sperm suspensions were mixed, and 40µl (if
746 using 500 µl HTF-BSA) or 25µl (if using 150 µl HTF-BSA) of the mixed sperm were
747 delivered to superovulated females using C&I Device for Mice (Cat.# 60020,
748 Paratech) at 9:00 AM. Recipients were immediately paired with vasectomized males
749 overnight. The next day, the plug was checked and the female mice with plugs were
750 used for collecting embryonic day 10 (E10) embryos, and the ones without plugs
751 were used for collecting zygotes, 2-cell embryos, morulae, or blastocysts embryos.
752
753 **Mouse genotyping**
754 Mouse tail snips were lysed in a lysis buffer (40mM NaOH, 0.2mM EDTA, pH=12)
755 for 1h at 95°C, followed by neutralization with the same volume of neutralizing buffer
756 (40mM Tris-HCl, pH 5.0). PCR reactions were conducted using the 2×GoTaq Green
757 master mix (Promega, Cat. M7123). The primers used for genotyping are the same
758 as previously described (15). For single embryo genotyping (e.g. zygotes, 2-cell
759 embryos, 4-cell embryos, morulae, and blastocysts), each embryo was picked up by
760 mouse pipetting and transferred into a 200µl tube, and lysed in 10 µl of lysis buffer
761 (100 mM Tris–HCl (pH 8.0), 100 mM KCl, 0.02% gelatin, 0.45% Tween 20, 60 µg/mL
762 yeast tRNA, and 125 µg/mL proteinase K) at 55 °C for 30min followed by inactivation
763 at 95 °C for 10min. 2µl of the lysis was used as the template for the first round of

764 PCR (30 cycles) in a 10µl reaction using the PrimeSTAR® HS DNA Polymerase
765 (Cat.# R010B, Takara) or 2×GoTaq Green master mix (Cat.# M7123, Promega).
766 Then 2µl of the first PCR was used for the second round of PCR in a 10µl reaction
767 using 2×GoTaq Green master mix (Cat.# M7123, Promega) for 35 cycles. Primers
768 used for embryo genotyping were included in Table S8.
769
770 **openCASA**
771 Sperm parameters were assessed using openCASA (86). After sperm capacitation
772 in HTF containing 4%BSA at 37 °C for 30min, the video was recorded as an AVI
773 format at 60 frames per second (FPS) for 2 seconds with a resolution of 768*576
774 pixels using UPlan FL N 4x/0.13 PhP Objective Lens (Olympus) and DMK
775 33UP1300 camera (The imaging source). Motility module in openCASA was set with
776 the following parameters: 1.21 Microns per Pixel, the cell size of 10-200 µm^2,
777 Progressive motility (STR>50%, VAP>50), Minimum VCL of 10 µm/s, VCL threshold
778 of 30-200 µm/s, 60 Frame Rate (frames/S), 10 Minimum Track Length (frames), 20
779 µm Maximum displacement between frames, and Window Size (frames) of 4.
780
781 **Analysis of conservation of *miR-506* family in modern humans using the 1000**
782 **Genomes Project (1kgp)**
783 The vcf files from the 1000 Genomes Project covering 3202 samples were
784 downloaded from
785 http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_cove
786 rage/working/20220422_3202_phased_SNV_INDEL_SV/. The miRNA annotations
787 were obtained from UCSC genome browser, and pachytene piRNA hg19 genome
788 coordinance was obtained from (30) and converted to GRCh38 genome
789 coordinance using LiftOver. The derived allele frequency (DAF) was retrieved from
790 the vcf file, and mean nucleotide diversity (MND) was calculated as 2*DAF*(1-DAF)
791 as previously described (30). Kruskal-Wallis test was used for statistical analysis,
792 and adjusted $p < 0.05$ was identified as statistically significant.
793
794 **Overexpression of MER91C**
795 RNA structures for MER91C were predicted using RNAfold (87). MER91C DNA
796 transposons from humans (*hsa-miR-513a1*), dogs (*cfa-miR-507b*), and horses (*eca-*
797 *miR-514a*) were synthesized by IDT and inserted into pCI-Neo plasmid (Cat.# E1841,
798 Promega) using NEBuilder® HiFi DNA Assembly Master Mix (Cat.# E2621L, NEB).
799 150ng of pCI-Neo (negative control) or pCI-Neo-Mer91c were transfected with or
800 without 150ng of pcDNA3.1+_FH-AGO2-WT (Plasmid # 92006, Addgene) into
801 HEK293T cells at ~60% confluency in 24-well plates. 24h later, cells were harvested
802 followed by RNA extraction using mirVana™ miRNA Isolation Kit (Cat# AM1561,
803 Thermo Fisher Scientific), polyadenylation by E. coli Poly(A) Polymerase (Cat.#
804 M0276L, NEB), reverse transcription using SuperScript IV Reverse Transcriptase
805 (Cat.# 18090010, Thermo Fisher Scientific), and PCR using 2×GoTaq Green master
806 mix (Cat.# M7123, Promega) or qPCR by PowerUp™ SYBR™ Green Master Mix
807 (Cat.# A25742, Thermo Fisher Scientific). Primers used for RT-PCR, PCR, and
808 qPCR were included in Table S8.
809

810 **Bioinformatic analyses of transposable element (TE)**
811 Genomic regions and GFF3 files for transcript 3'UTR annotations were downloaded
812 from the UCSC genome browser, and GTF files for transposon annotations were
813 downloaded from https://labshare.cshl.edu/shares/mhammelllab/www-
814 data/TEtranscripts/TE_GTF/. For the TE containing transcripts, bedtools closest was
815 used to extract the closest TEs to transcripts. Genomes used were GRCh38
816 (humans), panTro6 (chimpanzees), equCab3 (horses), GRCm39 (mice), and rn7
817 (rats).
818
819 **Phylogenetic tree analysis of the MER91C DNA transposon and the *miR-506***
820 **family miRNAs**
821 The transposon fasta sequences from humans, dogs, horses, and guinea pigs were
822 downloaded from the UCSC genome browser and aligned to the *miR-506* family
823 miRNAs in their corresponding species using BLAST (Basic Local Alignment Search
824 Tool) (34). After retrieving the transposons that aligned to the *miR-506* family
825 miRNAs, the *miR-506* family miRNAs and the transposons were aligned using
826 ClustalW2 followed by phylogenetic tree building using IQ-TREE2 with default
827 parameters (88). The final figure was generated using Geneious software.
828
829 **Purification of germ cells**
830 Pachytene spermatocytes and round spermatids were purified from adult C57BL/6J
831 mice using the STA-PUT method. BSA gradients (2-4%) were prepared in EKRB
832 buffer with a pH of 7.2 containing 1× Krebs-Ringer Bicarbonate Buffer (Cat.# K4002,
833 Sigma), 1.26 g/L sodium bicarbonate (Cat.# S6761, Sigma), 1× GlutaMAX (Cat.#
834 35050061, Thermo Fisher Scientific), 1× Antibiotic-Antimycotic (Cat.# 15240062,
835 Thermo Fisher Scientific), 1× MEM Non-Essential Amino Acids (Cat.# 11140050,
836 Thermo Fisher Scientific), 1×MEM Amino Acids (Cat.# 11130051, Thermo Fisher
837 Scientific), and 100 ng/ml cycloheximide (Cat.# 01810, Sigma). After being removed
838 and decapsulated, testes were placed into 10 ml of EKRB buffer containing 0.5
839 mg/ml type IV collagenase (Cat.# C5138, Sigma) and digested at 33 °C for ~12min
840 to dissociate the seminiferous tubules. Once dissociated, the seminiferous tubules
841 were washed three times using EKRB buffer to remove the interstitial cells and red
842 blood cells followed by trypsin digestion by incubation at 33 °C for ~12 min with
843 occasional pipetting in 10 ml EKRB buffer containing 0.25 mg/ml trypsin (Cat.#
844 T9935, Sigma) and 20 µg/ml DNase I (Cat.# DN25, Sigma). 1 mL of 4% BSA-EKRB
845 was added to the 10 mL fully dispersed testicular cells to neutralize the trypsin
846 digestion followed by centrifuge at 800g for 5min at 4°C. Testicular cells were
847 washed two times with EKRB buffer and resuspended in 10 mL 0.5% BSA-EKRB.
848 The cell suspension was passed through 70 µm cell strainer (Cat.# 431751, Corning)
849 and loaded onto the STA-PUT apparatus containing 2-4% BSA-EKRB gradients for
850 sedimentation. After 2-3 h sedimentation at 4 °C, cell fractions were collected from
851 the bottom of the sedimentation chamber. Fractions containing the same cell types
852 were pooled and saved for RNA sequencing.
853
854 **Library construction and RNA-seq**

855 RNA was extracted using the mirVana™ miRNA Isolation Kit (Cat.# AM1561,
856 Thermo Fisher Scientific) following the manufacturer's instructions. Large RNA
857 (>200nt) and small RNA (<200nt) were isolated separately for library construction.
858 Small RNA libraries were constructed using NEBNext® Small RNA Library Prep Set
859 for Illumina® (Multiplex Compatible) (Cat.# E7330L, NEB) following the
860 manufacturer's instructions and sequenced using HiSeq 2500 system for single-end
861 50bp sequencing. Large RNA libraries were constructed using the KAPA Stranded
862 RNA-Seq Kit with RiboErase (Cat.# KK8483, Roche) and the adaptor from
863 NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1, Cat.# E7335L, NEB).
864 The indexed large RNA libraries were sequenced using Nextseq 500 with paired-
865 end 75bp sequencing.
866
867 **Large and small RNA-Seq data analysis**
868 For the large RNA-seq data, raw sequences were trimmed by Trimmomatic (89),
869 followed by alignment using Hisat2 (90), and assembly using StringTie (90). Reads
870 were summarized using featureCounts (91) and the differential gene expression was
871 compared using DESeq2 (92). For each KO mouse sample, the genes with a fold
872 change ≥ 2 and FDR < 0.05 were considered DEGs. The DEGs in each KO mouse
873 were then intersected with the corresponding miRNA targets predicted by four
874 different algorithms, including TargetScan (54), microrna.org (55), miRWalk (56),
875 and mirDB (57). The gene is considered a putative target (differentially expressed
876 target, DETs) as long as it intersects with the targets predicted by any method
877 mentioned above. The DETs identified in each KO mouse sample were intersected
878 with other KO mouse samples, and the DETs intersected at least two different KO
879 mouse samples were selected as the "pool" of DETs in mice. Then the DETs "pool"
880 in mice was intersected with *miR-506* family predicted targets in humans and rats to
881 determine the shared targets among humans, rats, and mice. A mRNA that any *miR-*
882 *506* family member is targeting is deemed as the shared target.
883 For the small RNA-seq data, we applied the AASRA (93) pipeline (for mice) or
884 SPORTS1.0 (94) (for humans, monkeys, rats, and horses) to parse the raw
885 sequencing data. The clean reads were mapped against miRbase (28). The DESeq2
886 (92) (for mice) or edgeR (95) (for humans, monkeys, rats, and horses) algorithm was
887 used to compare the groupwise miRNA expression levels. The RNAs with a false
888 discovery rate (FDR) < 5% were deemed differentially expressed. Cohen's d was
889 computed by the "cohensD" function within the "lsr" package.
890
891 **Luciferase assay**
892 For luciferase reporter assays, the 3′ UTR of *Crisp1*, *Egr1*, *hCRISP1*, and
893 *hFMR1* were amplified using C57BL/6J tail snips or HEK293 cells genomic DNA
894 template with Q5® Hot Start High-Fidelity 2X Master Mix (Cat.# M0494L, NEB). The
895 PCR products were inserted into psiCHECK-2 vector (Cat.# C8021, Promega) via
896 Xho I (Cat.# R0146S, NEB) and Not I (Cat.# R3189S, NEB) restriction enzymes
897 cutting sites downstream of the Renilla luciferase-coding sequence using either
898 NEBuilder® HiFi DNA Assembly Master Mix (Cat.# E2621L, NEB) or T4 DNA ligase
899 (Cat.# M0202L, NEB). For the miRNA overexpression plasmids, ~300bp upstream
900 and downstream of the precursor miRNA genomic region were amplified using

901 C57BL/6J tail snips or HEK293 cells genomic DNA with Q5® Hot Start High-Fidelity
902 2X Master Mix (Cat.# M0494L, NEB) or PrimeSTAR® HS DNA Polymerase (Cat.#
903 R010B, Takara), and inserted into pcDNA3.1 plasmids using NEBuilder® HiFi DNA
904 Assembly Master Mix (Cat.# E2621L, NEB). HEK293T cells were co-transfected
905 with 150ng pcDNA3.1-miRNA and 150 ng psiCHECK-2 containing the 3′ UTR of the
906 target gene using Lipofectamine 3000 (Cat.# L3000015, Thermofisher Scientific) in
907 a 24-well cell culture plate (Cat.# 3524, Corning) at ~60% confluency. After 24 h of
908 culture, cells were lysed and assayed with Dual Luciferase Assays (Cat.# E1910,
909 Promega) according to the manufacturer's instructions. Renilla luciferase signals
910 were normalized to Firefly luciferase signals to adjust the transfection efficiency.
911 pcDNA3.1-cel-miR-67 which has a minimal sequence identity to the miRNAs in
912 humans, mice and rats was used as a negative control miRNA. Primers used for
913 generating plasmids containing miRNAs or 3'UTR of the target genes were included
914 in Table S8.
915
916 **Immunofluorescence**
917 Cauda sperm were capacitated in HTF at 37 °C for half an hour followed by
918 spreading onto Superfrost Plus slides (Cat.# 22-037-246, Thermo Fisher Scientific).
919 The slides were air-dried, fixed in 4% paraformaldehyde for 15 min (Cat.# J19943-
920 K2, Thermo Fisher Scientific), then washed twice in 0.4% Photo-Flo 200 (Cat.#
921 1464510, Kodak) /1×PBS (5 min/wash), followed by a 5 min wash in 0.4% Photo-
922 Flo 200/ddH_2O, and stored in -80 °C after air-dried. The slides were equilibrated to
923 room temperature before immunofluorescence, followed by incubation in acetone
924 for 20 min at 4°C and rehydration in 95% ethanol twice (5 min/wash), 70% ethanol
925 twice (5 min/wash) and 1×PBS for 3 times (5 min/wash) sequentially. Heat-induced
926 antigen retrieval was performed in citrate buffer (pH 6.0) with high power for 4 min
927 once, and three times with low power for 4 min in microwave. Slides were cooled
928 down to room temperature and washed in 1×PBS twice (5 min/wash). Following
929 permeabilization with 0.25% Triton X-100 (Sigma-Aldrich, Cat.# T8787) in 1×PBS
930 for 20 min at room temperature, the slides were washed with 1× PBS three times (5
931 min/wash), and incubated at 3% H_2O_2 solution to block endogenous peroxidase
932 activity. After washing in 1× PBS twice (5 min/wash), the slides were blocked with
933 1× blocking solution (5% normal donkey serum, 5% fetal bovine serum, and 1%
934 bovine serum albumin in 1×PBS) at room temperature for 1h, then incubated with
935 the anti-CRISP-1 antibody (Cat.# AF4675-SP, R&D systems, 1: 100 in 1× blocking
936 solution) at 4 °C overnight. After primary antibody incubation, the slides were
937 washed in 1× PBS three times (10 min/wash), followed by incubation in the Donkey
938 Anti-Goat IgG H&L (HRP) (Cat.# ab97110, Abcam, 1: 250 in 1× blocking solution)
939 at room temperature for 1 h, and 3 times washes in 1× PBS (10 min/wash). Tyramide
940 signal amplification was performed and stopped using reagents from Invitrogen™
941 Alexa Fluor™ 488 Tyramide SuperBoost™ Kit (Cat. # B40941, Thermo Fisher
942 Scientific), followed by mounting and counterstaining in Antifade Mounting Medium
943 with DAPI (Cat.# H-1800, Vector Lab). Nail polish was applied on the edge of the
944 coverslips after 2h of mounting to prevent further evaporation and stored at 4°C
945 before taking images. Images were taken using the Nikon ECLIPSE Ti2 Confocal
946 microscope with the NIS-Elements Software.

947
948 **Western blot**
949 Testes from adult WT and KO mice were collected and sonicated in 2 × Laemmli
950 buffer (Cat.# 1610737, Bio-Rad) supplemented with 2-Mercaptoethanol (Cat.#
951 M6250, Sigma-Aldrich) and cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail
952 (Cat.# 11836170001, Sigma-Aldrich) followed by incubating at 100°C for 10 min.
953 The proteins were separated on 4–20% Mini-PROTEAN® TGX™ Precast Protein
954 Gels (Cat.# 4561094, Bio-Rad) and then transferred onto Amersham™ Protran®
955 Premium Western blotting membranes, nitrocellulose (Cat.# GE10600003, Sigma-
956 Aldrich). The membranes were then stained with Ponceau S solution (Cat.# P7170,
957 Sigma-Aldrich) to check the samples' loading. After taking pictures, the membrane
958 was de-stained with 0.1 M NaOH, and washed with water and TBS. Then the
959 membrane was blocked with 5% skim milk in TBST (TBS containing 0.1% (v/v)
960 Tween-20) for 1 h at room temperature and incubated with the anti-CRISP-1
961 antibody (Cat.# AF4675-SP, R&D systems, 1: 2000 in TBST containing 5% skim
962 milk) and anti-GAPDH antibody (Cat.# G9545, Sigma, 1: 6000 in TBST) overnight
963 at 4°C. After washing with TBST three times, the membrane was incubated with the
964 Donkey Anti-Goat IgG H&L (HRP) (Cat.# ab97110, Abcam) or Goat Anti-Rabbit IgG
965 H&L (HRP) (Cat.# ab6721, Abcam) at room temperature for 1 h. Followed by three
966 washes with TBST, the bands were detected using the WesternBright ECL kit (Cat.#
967 K-12045-D20, Advansta).
968
969 **T7 endonuclease I (T7EI) assay**
970 The potential off-target sites that may be induced by CRISPR-Cas9 were predicted
971 using Alt-R Custom Cas9 crRNA Design Tool (IDT) and assessed by T7
972 endonuclease I (Cat.# M0302L, NEB) assay. The sequences were retrieved from
973 the UCSC genome browser, and the primers flanking the off-target sites were
974 designed to cover ~600bp. Genomic DNA from WT C57BL/6J or quinKO was
975 amplified using Q5® Hot Start High-Fidelity 2X Master Mix (Cat.# M0494L, NEB) or
976 PrimeSTAR® HS DNA Polymerase (Cat.# R010B, Takara) with the designed off-
977 target primers. 2μl of the unpurified PCR product was diluted in 1X NEBuffer 2 in a
978 9.5 μl volume and denatured at 95°C for 5min, followed by annealing at 95-85°C at
979 a -2°C/second rate, and 85-25°C at a -0.1°C/second rate. Then 0.5 μl of T7EI was
980 added to the 9.5 μl denatured PCR products and incubated at 37°C for 30min. The
981 T7EI-treated PCR products were run on 1X TBE gel and stained with SYBR™ Gold
982 Nucleic Acid Gel Stain (Cat.# S11494, Thermo Fisher Scientific) to detect the off-
983 target effects. Primers used for T7EI were included in Table S8.
984
985 **miRNA and 3' UTR conservation analysis**
986 The Multiz Alignment and Conservation method was used to measure miRNA
987 sequence conservation with either the human or the mice genome as the reference
988 (25, 96). 100 and 60 species were used for the human and mouse references,
989 respectively. PhastCons takes the flanking sequences into consideration and
990 doesn't rely on fixed sliding windows; consequently, both highly conserved short
991 sequences and moderately conserved long sequences can yield higher scores (97).
992 PhastCons gives a value between 0 to 1, the higher the value is, the more conserved

993 the region is. By contrast, PhyloP compares the conservation of individual
994 nucleotides among all phylogeny clades, giving positive scores once the region is
995 conserved and *vice versa*. PhyloP and PhastCons scores of all miRNAs, *miR-506*
996 family, *Fmr1* CDS, *Slitrk2* CDS, and the intergenic region (IGR) were retrieved from
997 the UCSC genome browser and quantified. Kruskal-Wallis test was used for
998 statistical analysis, and adjusted $p < 0.05$ was identified as statistically significant.
999 PhyloP scheme was used to measure the evolutionary conservation level at
1000 individual nucleotide sites in the 3'UTRs of the target genes. Positive PhyloP scores
1001 suggest higher conservation and stronger purifying selection, whereas negative
1002 PhyloP scores indicate accelerated evolution and potential adaptive selection. The
1003 genomic annotations and mRNA sequences were based on the hg38 (25)
1004 ('TxDb.Hsapiens.UCSC.hg38.knownGene' and 'BSgenome.Hsapiens.UCSC.hg38')
1005 and mm10 ('TxDb.Mmusculus.UCSC.mm10.knownGene' and
1006 'BSgenome.Mmusculus.UCSC.mm10') assemblies for human and mouse,
1007 respectively. The PhyloP scores were mapped to individual nucleotides in 3'UTRs
1008 based on the transcript coordinates (98).
1009
1010 **Data Availability**
1011 The sRNA-seq and RNA-seq datasets have been deposited into the SRA database
1012 with accession#: PRJNA558973 and PRJNA670945. The scripts for sRNA analysis
1013 can be found on GitHub (https://github.com/biogramming/AASRA and
1014 https://github.com/junchaoshi/sports1.1).
1015
1016 **Statistical analyses**
1017 Data are presented as mean ± SEM, and statistical differences between datasets
1018 were assessed by two samples t-test, F-test, Dunnett's multiple comparisons test as
1019 the posthoc test following one-way ANOVA, Chi-squared test, or Kruskal-Wallis test
1020 as described in the text or figure legends. Normal distribution was assessed by
1021 quantile-quantile (QQ) plot or density plot. $p < 0.05$, 0.01, 0.001, and 0.0001 are
1022 considered as statistically significant and indicated with *, **, ***, and ****
1023 respectively.
1024
1025 **Acknowledgments**
1026
1027 We would like to thank Dr. Kevin J. Peterson, Dartmouth College, Hanover, NH, for
1028 his help with phylogenetic analyses of the *miR-506* miRNA genes. This work was
1029 supported by grants from the NIH (HD071736, HD085506, HD098593, HD099924,
1030 and P30GM110767 to WY), NIH/National Center for Advancing Translational
1031 Science (NCATS) UCLA CTSI (UL1TR001881-01 to ZW and WY), and the
1032 Templeton Foundation (PID: 61174 to WY). Work in the Lai lab was supported by
1033 the NIH (R01-GM083300 and R01-HD108914) and Memorial Sloan-Kettering
1034 Institute Grant P30-CA008748.
1035

1036 **References**
1037

1038 1. E. F. Oakberg, Duration of spermatogenesis in the mouse. *Nature* **180**,
1039 1137-1138 (1957).
1040 2. B. Muciaccia *et al.*, Novel stage classification of human spermatogenesis
1041 based on acrosome development. *Biol Reprod* **89**, 60 (2013).
1042 3. N. H. Gu, W. L. Zhao, G. S. Wang, F. Sun, Comparative analysis of
1043 mammalian sperm ultrastructure reveals relationships between sperm
1044 morphology, mitochondrial functions and motility. *Reprod Biol Endocrinol*
1045 **17**, 66 (2019).
1046 4. L. Hermo, R. M. Pelletier, D. G. Cyr, C. E. Smith, Surfing the wave, cycle,
1047 life history, and genes/proteins expressed by testicular germ cells. Part 1:
1048 background to spermatogenesis, spermatogonia, and spermatocytes.
1049 *Microsc Res Tech* **73**, 241-278 (2010).
1050 5. P. Khaitovich, W. Enard, M. Lachmann, S. Paabo, Evolution of primate gene
1051 expression. *Nat Rev Genet* **7**, 693-702 (2006).
1052 6. R. K. Idler, W. Yan, Control of messenger RNA fate by RNA-binding
1053 proteins: an emphasis on mammalian spermatogenesis. *J Androl* **33**, 309-
1054 337 (2012).
1055 7. S. Ro, C. Park, K. M. Sanders, J. R. McCarrey, W. Yan, Cloning and
1056 expression profiling of testis-expressed microRNAs. *Dev Biol* **311**, 592-602
1057 (2007).
1058 8. Q. Wu *et al.*, The RNase III enzyme DROSHA is essential for microRNA
1059 production and spermatogenesis. *J Biol Chem* **287**, 25173-25190 (2012).
1060 9. M. D. Papaioannou *et al.*, Sertoli cell Dicer is essential for spermatogenesis
1061 in mice. *Dev Biol* **326**, 250-259 (2009).
1062 10. Y. Zhang *et al.*, MicroRNAs control mRNA fate by compartmentalization
1063 based on 3' UTR length in male germ cells. *Genome Biol* **18**, 105 (2017).
1064 11. M. Guo *et al.*, Uncoupling transcription and translation through miRNA-
1065 dependent poly(A) length control in haploid male germ cells. *Development*
1066 **149** (2022).
1067 12. D. P. Bartel, MicroRNAs: target recognition and regulatory functions. *Cell*
1068 **136**, 215-233 (2009).
1069 13. C. Mayr, Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell*
1070 *Biol* **26**, 227-237 (2016).
1071 14. P. J. Thompson, T. S. Macfarlan, M. C. Lorincz, Long Terminal Repeats:
1072 From Parasitic Elements to Building Blocks of the Transcriptional
1073 Regulatory Repertoire. *Mol Cell* **62**, 766-776 (2016).
1074 15. Z. Wang *et al.*, X-linked miR-506 family miRNAs promote FMRP expression
1075 in mouse spermatogonia. *EMBO Rep* **21**, e49024 (2020).
1076 16. R. Song *et al.*, Many X-linked microRNAs escape meiotic sex chromosome
1077 inactivation. *Nat Genet* **41**, 488-493 (2009).
1078 17. J. Bao *et al.*, MicroRNA-449 and microRNA-34b/c function redundantly in
1079 murine testes by targeting E2F transcription factor-retinoblastoma protein
1080 (E2F-pRb) pathway. *J Biol Chem* **287**, 21686-21698 (2012).

1081 18. Z. Wang *et al.*, Ablation of the miR-465 Cluster Causes a Skewed Sex Ratio
1082 in Mice. *Front Endocrinol (Lausanne)* **13**, 893854 (2022).
1083 19. E. A. Miska *et al.*, Most Caenorhabditis elegans microRNAs are individually
1084 not essential for development or viability. *PLoS Genet* **3**, e215 (2007).
1085 20. M. D. Dean, K. G. Ardlie, M. W. Nachman, The frequency of multiple
1086 paternity suggests that sperm competition is common in house mice (Mus
1087 domesticus). *Mol Ecol* **15**, 4141-4151 (2006).
1088 21. R. C. Firman, L. W. Simmons, Polyandry facilitates postcopulatory
1089 inbreeding avoidance in house mice. *Evolution* **62**, 603-611 (2008).
1090 22. G. A. Parker, Sperm competition and its evolutionary consequences in the
1091 insects. *Biological reviews* **45**, 525-567 (1970).
1092 23. R. P. Meisel, T. Connallon, The faster-X effect: integrating theory and data.
1093 *Trends Genet* **29**, 537-544 (2013).
1094 24. F. Zhang *et al.*, Evolution of an X-Linked miRNA Family Predominantly
1095 Expressed in Mammalian Male Germ Cells. *Mol Biol Evol* **36**, 663-678
1096 (2019).
1097 25. J. Casper *et al.*, The UCSC Genome Browser database: 2018 update.
1098 *Nucleic Acids Res* **46**, D762-D769 (2018).
1099 26. F. Cabanettes, C. Klopp, D-GENIES: dot plot large genomes in an
1100 interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).
1101 27. B. Fromm *et al.*, MirGeneDB 2.0: the metazoan microRNA complement.
1102 *Nucleic Acids Res* **48**, D1172 (2020).
1103 28. A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA
1104 sequences to function. *Nucleic Acids Res* **47**, D155-D162 (2019).
1105 29. M. Byrska-Bishop *et al.*, High-coverage whole-genome sequencing of the
1106 expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-
1107 3440 e3419 (2022).
1108 30. D. M. Ozata *et al.*, Evolutionarily conserved pachytene piRNA loci are highly
1109 divergent among modern humans. *Nat Ecol Evol* **4**, 156-168 (2020).
1110 31. N. V. Fedoroff, Presidential address. Transposable elements, epigenetics,
1111 and genome evolution. *Science* **338**, 758-767 (2012).
1112 32. Z. Wang *et al.*, Efficient genome editing by CRISPR-Mb3Cas12a in mice. *J*
1113 *Cell Sci* **133** (2020).
1114 33. A. McKenna *et al.*, Whole-organism lineage tracing by combinatorial and
1115 cumulative genome editing. *Science* **353**, aaf7907 (2016).
1116 34. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local
1117 alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
1118 35. J. Giordano *et al.*, Evolutionary history of mammalian transposons
1119 determined by genome-wide defragmentation. *PLoS Comput Biol* **3**, e137
1120 (2007).
1121 36. J. K. Pace, 2nd, C. Feschotte, The evolutionary history of human DNA
1122 transposons: evidence for intense activity in the primate lineage. *Genome*
1123 *Res* **17**, 422-432 (2007).
1124 37. T. Hirano *et al.*, Small RNA profiling and characterization of piRNA clusters
1125 in the adult testes of the common marmoset, a model primate. *RNA* **20**,
1126 1223-1237 (2014).

1127 38. R. Zhang, Y. Peng, W. Wang, B. Su, Rapid evolution of an X-linked
1128 microRNA cluster in primates. *Genome Res* **17**, 612-617 (2007).
1129 39. E. M. Koenig *et al.*, The beagle dog MicroRNA tissue atlas: identifying
1130 translatable biomarkers of organ toxicity. *BMC Genomics* **17**, 649 (2016).
1131 40. A. Keller *et al.*, miRNATissueAtlas2: an update to the human miRNA tissue
1132 atlas. *Nucleic Acids Res* **50**, D211-D221 (2022).
1133 41. P. R. Bushel *et al.*, RATEmiRs: the rat atlas of tissue-specific and enriched
1134 miRNAs for discerning baseline expression exclusivity of candidate
1135 biomarkers. *RNA Biol* **17**, 630-636 (2020).
1136 42. I. V. Gainetdinov *et al.*, Assessment of piRNA biogenesis and function in
1137 testicular germ cell tumors and their precursor germ cell neoplasia in situ.
1138 *BMC Cancer* **18**, 20 (2018).
1139 43. B. Li *et al.*, Identification of piRNAs and piRNA clusters in the testes of the
1140 Mongolian horse. *Sci Rep* **9**, 5022 (2019).
1141 44. M. Abu-Halima *et al.*, Altered microRNA expression profiles of human
1142 spermatozoa in patients with different spermatogenic impairments. *Fertil*
1143 *Steril* **99**, 1249-1255 e1216 (2013).
1144 45. Z. Heidary, M. Zaki-Dizaji, K. Saliminejad, H. R. Khorram Khorshid,
1145 MicroRNA profiling in spermatozoa of men with unexplained
1146 asthenozoospermia. *Andrologia* **51**, e13284 (2019).
1147 46. H. Tian *et al.*, Semen-specific miRNAs: Suitable for the distinction of infertile
1148 semen in the body fluid identification? *Forensic Sci Int Genet* **33**, 161-167
1149 (2018).
1150 47. X. Qing *et al.*, Dysregulation of an X-linked primate-specific epididymal
1151 microRNA cluster in unexplained asthenozoospermia. *Oncotarget* **8**, 56839-
1152 56849 (2017).
1153 48. C. Wang *et al.*, Altered profile of seminal plasma microRNAs in the
1154 molecular diagnosis of male infertility. *Clin Chem* **57**, 1722-1731 (2011).
1155 49. B. Xia *et al.*, Widespread Transcriptional Scanning in the Testis Modulates
1156 Gene Evolution Rates. *Cell* **180**, 248-262 e221 (2020).
1157 50. R. C. Firman, Polyandrous females benefit by producing sons that achieve
1158 high reproductive success in a competitive environment. *Proc Biol Sci* **278**,
1159 2823-2831 (2011).
1160 51. M. D. Muzumdar, B. Tasic, K. Miyamichi, L. Li, L. Luo, A global double-
1161 fluorescent Cre reporter mouse. *Genesis* **45**, 593-605 (2007).
1162 52. H. S. Fisher, H. E. Hoekstra, Competition drives cooperation among closely
1163 related sperm of deer mice. *Nature* **463**, 801-803 (2010).
1164 53. H. S. Fisher, E. Jacobs-Palmer, J. M. Lassance, H. E. Hoekstra, The
1165 genetic basis and fitness consequences of sperm midpiece size in deer
1166 mice. *Nat Commun* **7**, 13652 (2016).
1167 54. V. Agarwal, G. W. Bell, J. W. Nam, D. P. Bartel, Predicting effective
1168 microRNA target sites in mammalian mRNAs. *Elife* **4** (2015).
1169 55. D. Betel, A. Koppal, P. Agius, C. Sander, C. Leslie, Comprehensive
1170 modeling of microRNA targets predicts functional non-conserved and non-
1171 canonical sites. *Genome Biol* **11**, R90 (2010).

1172 56. H. Dweep, N. Gretz, miRWalk2.0: a comprehensive atlas of microRNA-
1173 target interactions. *Nat Methods* **12**, 697 (2015).
1174 57. Y. Chen, X. Wang, miRDB: an online database for prediction of functional
1175 microRNA targets. *Nucleic Acids Res* **48**, D127-D131 (2020).
1176 58. V. G. Da Ros *et al.*, Impaired sperm fertilizing ability in mice lacking
1177 Cysteine-RIch Secretory Protein 1 (CRISP1). *Dev Biol* **320**, 12-18 (2008).
1178 59. J. A. Maldera *et al.*, Human fertilization: epididymal hCRISP1 mediates
1179 sperm-zona pellucida binding through its interaction with ZP3. *Mol Hum*
1180 *Reprod* **20**, 341-349 (2014).
1181 60. E. H. Morrow, How the sperm lost its tail: the evolution of aflagellate sperm.
1182 *Biol Rev Camb Philos Soc* **79**, 795-814 (2004).
1183 61. J. Piriyapongsa, L. Marino-Ramirez, I. K. Jordan, Origin and evolution of
1184 human microRNAs from transposable elements. *Genetics* **176**, 1323-1337
1185 (2007).
1186 62. I. K. Jordan, I. B. Rogozin, G. V. Glazko, E. V. Koonin, Origin of a
1187 substantial fraction of human regulatory sequences from transposable
1188 elements. *Trends Genet* **19**, 68-72 (2003).
1189 63. J. Gonzalez, D. A. Petrov, Evolution of genome content: population
1190 dynamics of transposable elements in flies and humans. *Methods Mol Biol*
1191 **855**, 361-383 (2012).
1192 64. H. Wang *et al.*, Small RNA shuffling between murine sperm and their
1193 cytoplasmic droplets during epididymal maturation. *Dev Cell* **58**, 779-790
1194 e774 (2023).
1195 65. W. R. Rice, Sex Chromosomes and the Evolution of Sexual Dimorphism.
1196 *Evolution* **38**, 735-742 (1984).
1197 66. J. R. Gibson, A. K. Chippindale, W. R. Rice, The X chromosome is a hot
1198 spot for sexually antagonistic fitness variation. *Proc Biol Sci* **269**, 499-505
1199 (2002).
1200 67. P. J. Wang, J. R. McCarrey, F. Yang, D. C. Page, An abundance of X-linked
1201 genes expressed in spermatogonia. *Nat Genet* **27**, 422-426 (2001).
1202 68. J. Mohammed *et al.*, Adaptive evolution of testis-specific, recently evolved,
1203 clustered miRNAs in Drosophila. *RNA* **20**, 1195-1209 (2014).
1204 69. K. Nishikura, A-to-I editing of coding and non-coding RNAs by ADARs. *Nat*
1205 *Rev Mol Cell Biol* **17**, 83-96 (2016).
1206 70. G. Varani, W. H. McClain, The G x U wobble base pair. A fundamental
1207 building block of RNA structure crucial to RNA function in diverse biological
1208 systems. *EMBO Rep* **1**, 18-23 (2000).
1209 71. J. G. Doench, P. A. Sharp, Specificity of microRNA target selection in
1210 translational repression. *Genes Dev* **18**, 504-511 (2004).
1211 72. A. Simkin, R. Geissler, A. B. R. McIntyre, A. Grimson, Evolutionary
1212 dynamics of microRNA target sites across vertebrate evolution. *PLoS Genet*
1213 **16**, e1008285 (2020).
1214 73. J. Meunier *et al.*, Birth and expression evolution of mammalian microRNA
1215 genes. *Genome Res* **23**, 34-45 (2013).
1216 74. S. Vasudevan, Y. Tong, J. A. Steitz, Switching from repression to activation:
1217 microRNAs can up-regulate translation. *Science* **318**, 1931-1934 (2007).

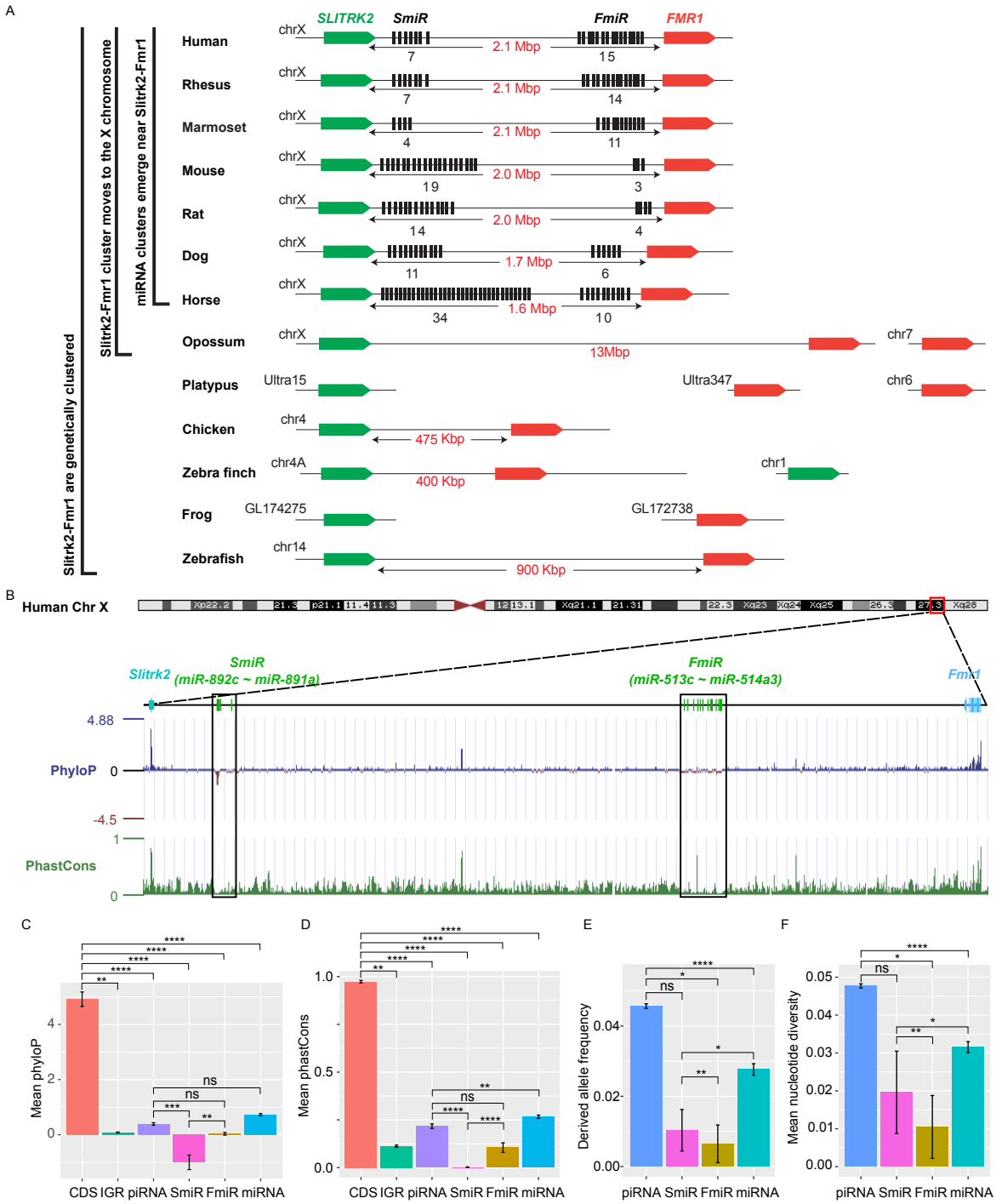
1218 75. R. F. Place, L. C. Li, D. Pookot, E. J. Noonan, R. Dahiya, MicroRNA-373
1219 induces expression of genes with complementary promoter sequences.
1220 *Proc Natl Acad Sci U S A* **105**, 1608-1613 (2008).
1221 76. S. Yuan *et al.*, Motile cilia of the male reproductive system require miR-
1222 34/miR-449 for development and function to generate luminal turbulence.
1223 *Proc Natl Acad Sci U S A* **116**, 3584-3593 (2019).
1224 77. S. Yuan *et al.*, Oviductal motile cilia are essential for oocyte pickup but
1225 dispensable for sperm and embryo transport. *Proc Natl Acad Sci U S A* **118**
1226 (2021).
1227 78. J. P. Broughton, M. T. Lovci, J. L. Huang, G. W. Yeo, A. E. Pasquinelli,
1228 Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol Cell*
1229 **64**, 320-333 (2016).
1230 79. A. Helwak, G. Kudla, T. Dudnakova, D. Tollervey, Mapping the human
1231 miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*
1232 **153**, 654-665 (2013).
1233 80. N. Pinzon *et al.*, microRNA target prediction programs predict many false
1234 positives. *Genome Res* **27**, 234-245 (2017).
1235 81. M. Ecsedi, M. Rausch, H. Grosshans, The let-7 microRNA directs vulval
1236 development through a single target. *Dev Cell* **32**, 335-344 (2015).
1237 82. D. L. Garaulet, B. Zhang, L. Wei, E. Li, E. C. Lai, miRNAs and Neural
1238 Alternative Polyadenylation Specify the Virgin Behavioral State. *Dev Cell* **54**,
1239 410-423 e414 (2020).
1240 83. M. Gomendio, J. Martin-Coello, C. Crespo, C. Magana, E. R. Roldan,
1241 Sperm competition enhances functional capacity of mammalian
1242 spermatozoa. *Proc Natl Acad Sci U S A* **103**, 15113-15117 (2006).
1243 84. A. Nagy, M. Gertsenstein, K. Vintersten, R. Behringer, *Manipulating the*
1244 *mouse embryo: a laboratory manual* (Cold Spring Harbor Laboratory Press
1245 Cold Spring Harbor, NY:, 2003), vol. 10.
1246 85. B. J. Stone, K. H. Steele, A. Fath-Goodin, A rapid and effective nonsurgical
1247 artificial insemination protocol using the NSET device for sperm transfer in
1248 mice without anesthesia. *Transgenic Res* **24**, 775-781 (2015).
1249 86. C. Alquezar-Baeta *et al.*, OpenCASA: A new open-source and scalable tool
1250 for sperm quality analysis. *PLoS Comput Biol* **15**, e1006691 (2019).
1251 87. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
1252 88. B. Q. Minh *et al.*, IQ-TREE 2: New Models and Efficient Methods for
1253 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534
1254 (2020).
1255 89. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for
1256 Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
1257 90. M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level
1258 expression analysis of RNA-seq experiments with HISAT, StringTie and
1259 Ballgown. *Nat Protoc* **11**, 1650-1667 (2016).
1260 91. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose
1261 program for assigning sequence reads to genomic features. *Bioinformatics*
1262 **30**, 923-930 (2014).

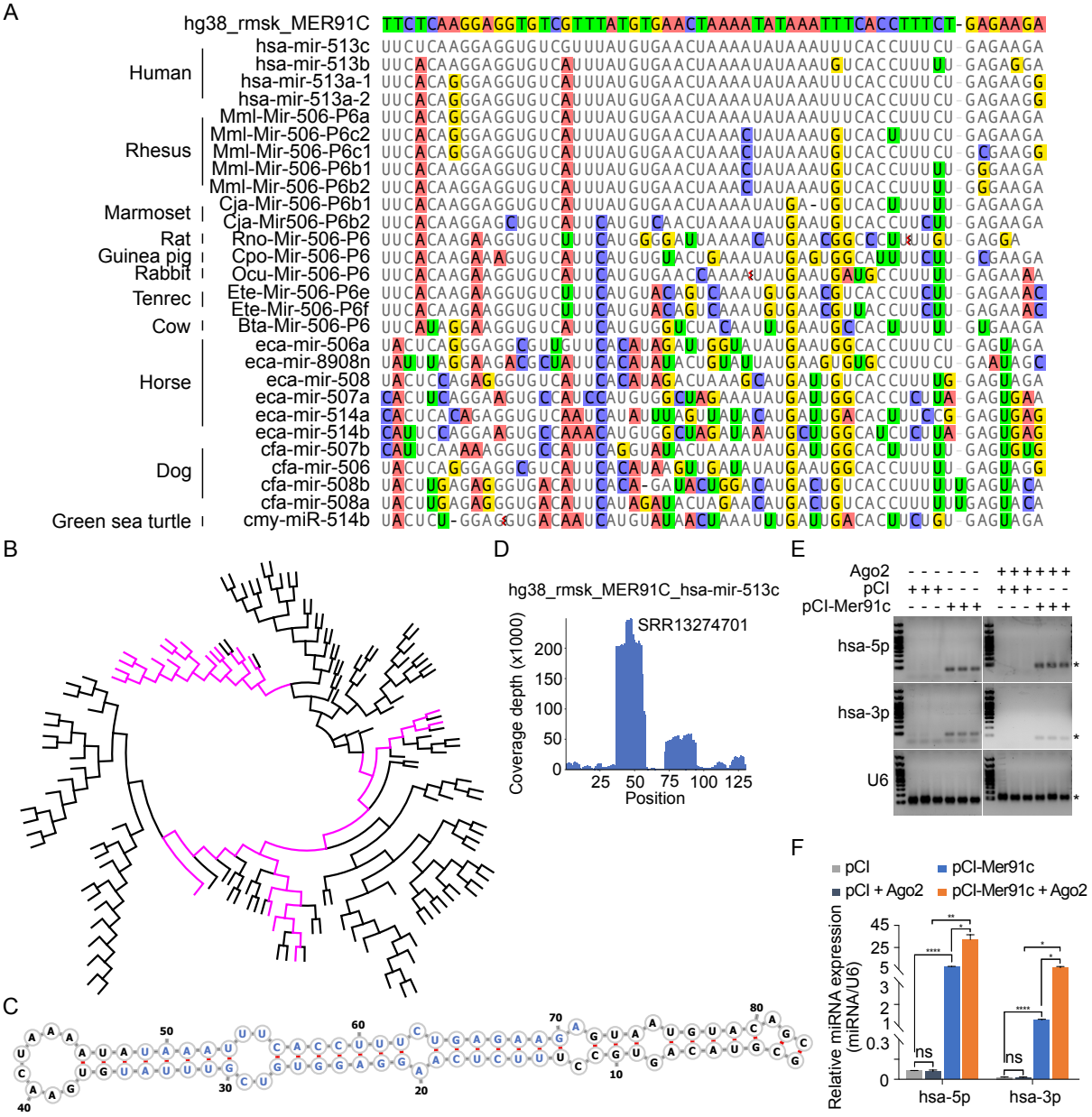
1263 92. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and
1264 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
1265 93. C. Tang, Y. Xie, M. Guo, W. Yan, AASRA: an anchor alignment-based small
1266 RNA annotation pipelinedagger. *Biol Reprod* **105**, 267-277 (2021).
1267 94. J. Shi, E. A. Ko, K. M. Sanders, Q. Chen, T. Zhou, SPORTS1.0: A Tool for
1268 Annotating and Profiling Non-coding RNAs Optimized for rRNA- and tRNA-
1269 derived Small RNAs. *Genomics Proteomics Bioinformatics* **16**, 144-151
1270 (2018).
1271 95. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor
1272 package for differential expression analysis of digital gene expression data.
1273 *Bioinformatics* **26**, 139-140 (2010).
1274 96. M. Blanchette *et al.*, Aligning multiple genomic sequences with the threaded
1275 blockset aligner. *Genome Res* **14**, 708-715 (2004).
1276 97. A. Siepel *et al.*, Evolutionarily conserved elements in vertebrate, insect,
1277 worm, and yeast genomes. *Genome Res* **15**, 1034-1050 (2005).
1278 98. M. Lawrence *et al.*, Software for computing and annotating genomic ranges.
1279 *PLoS Comput Biol* **9**, e1003118 (2013).
1280

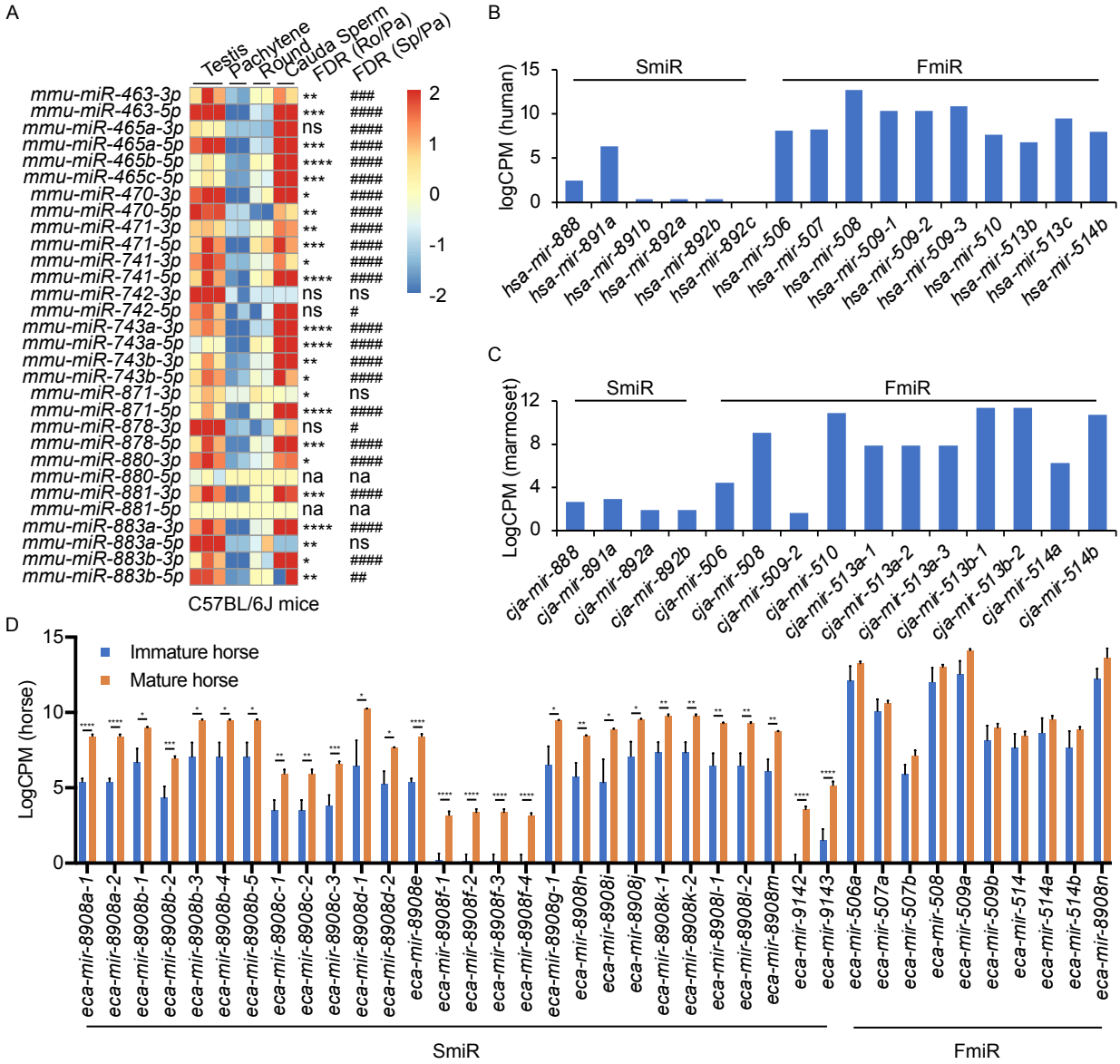
1281 **Figures and Tables**
1282
1283 **Figure 1. Genomic location, sequence alignment and evolution conservation**
1284 **of the X-linked *miR-506* family.**
1285 A. Genomic location of the X-linked *miR-506* family miRNAs (black bars) and the
1286 two flanking coding genes, *Slitrk2* (green blocks) and *Fmr1* (red blocks). The
1287 number of miRNAs within each cluster is indicated underneath the miRNA clusters.
1288 B. Evolution conservation of X-linked *miR-506* family based on Multiz Alignment &
1289 Conservation using the human genome as a reference. Positive PhyloP scores
1290 indicate conservation and vice versa. PhastCons has a score between 0~1, and
1291 the higher the score, the more conserved the DNA region is. C-D. Comparison of
1292 mean phyloP (C) phastCons (D) scores among CDS of *Slitrk2* and *Fmr1*,
1293 intergenic region (IGR), pachytene piRNAs, SmiR, FmiR, and all miRNAs. **, ***,
1294 **** indicate adjusted p-value < 0.01, 0.001, and 0.0001, respectively. ns, not
1295 significant. Kruskal-Wallis test was used for statistical analyses. E-F. Comparison
1296 of derived allele (E) and mean nucleotide (F) frequencies among pachytene
1297 piRNAs, SmiR, FmiR, and all miRNAs. *, ** and **** indicate adjusted p-value <
1298 0.05, 0.01 and 0.0001, respectively. ns, not significant. Kruskal-Wallis test was
1299 used for statistical analyses.
1300
1301 **Figure 2. Evolutionary history of the X-linked *miR-506* family.**
1302 A. Sequences alignment of FmiRs from various species using human MER91C
1303 DNA transposon as the reference. The first line is the human MER91C DNA
1304 transposon, and below are the miRNAs of various species. Mismatched
1305 nucleotides are highlighted with various colors. B. A phylogenetic tree of the
1306 MER91C DNA transposons and the X-linked *miR-506* family miRNAs. The
1307 MER91C DNA transposons are labeled in purple. C. RNA structure of the MER91C
1308 DNA transposon-derived miRNA (hsa-miR-513c). D. sRNA-seq reads (lower
1309 panel) of the MER91C DNA transposon-derived miRNA, hsa-miR-513c. E.
1310 Representative gel images showing expression levels of the MER91C DNA
1311 transposon-derived miRNA (hsa-miR-513a1) in HEK293T cells. The asterisk (*)
1312 indicates the expected miRNA size. U6 was used as the loading control. F. qPCR
1313 analyses of expression levels of MER91C DNA transposon-derived miRNA (hsa-
1314 miR-513a1) in HEK293T cells. *, **, **** indicate adjusted p-value < 0.05, 0.01,
1315 0.0001, respectively. One-way ANOVA was used for statistical analyses.
1316
1317
1318 **Figure 3. Expression profiles of X-linked *miR-506* family in mammalian testes**
1319 **and male germ cells.**
1320 A. Heatmaps showing the *miR-506* family expression in the testis, pachytene
1321 spermatocytes, round spermatids and sperm in mice. Biological triplicates of the
1322 testis samples (n=3) and duplicates of pachytene spermatocytes, round
1323 spermatids and sperm samples isolated from 2-4 mice were used for sRNA-seq.
1324 *, **, ***, and **** indicate FDR < 0.05, 0.01, 0.001, and 0.0001, respectively, when
1325 comparing round spermatids to pachytene spermatocytes. #, ##, ###, and ####
1326 indicate FDR < 0.05, 0.01, 0.001, and 0.0001, respectively, when comparing cauda

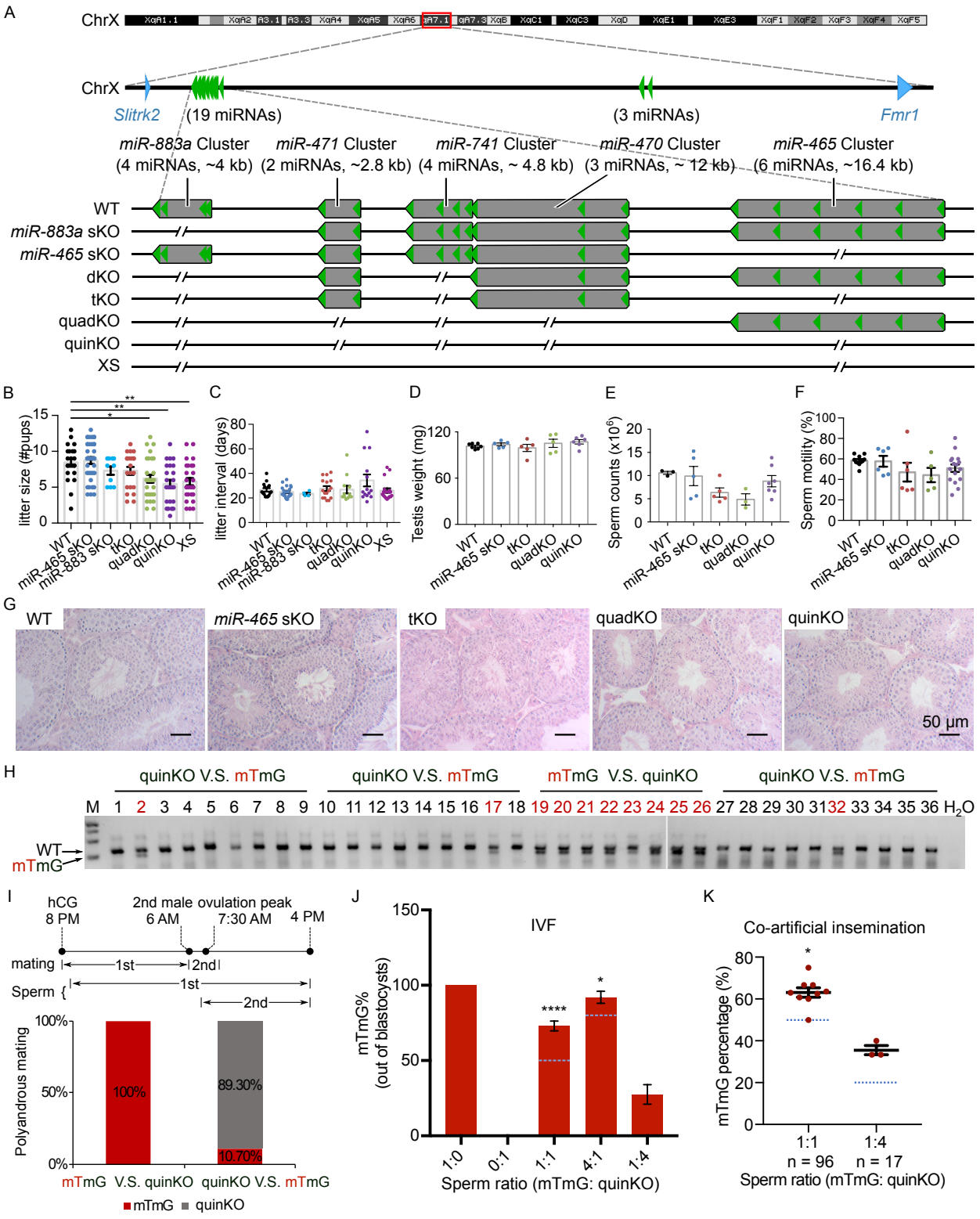
1327 sperm to pachytene spermatocytes. ns and na indicate not significantly and not not
1328 applicable, respectively.B-C. LogCPM bar graphs showing the *miR-506* family
1329 expression in the testis of humans (B) and marmosets (C). D. LogCPM bar graph
1330 showing the *miR-506* family expression in sexually immature and mature horse
1331 testes. n=3. *, **, ***, and **** indicate FDR < 0.05, 0.01, 0.001, and 0.0001,
1332 respectively.
1333
1334
1335 **Figure 4. Ablation of X-linked *miR-506* family miRNAs compromised sperm**
1336 **competitiveness and reproductive fitness in male mice.**
1337 A. Schematics showing the strategy used to generate six lines of KO mice lacking
1338 individual or combined miRNA clusters within the *miR-506* family using CRISPR-
1339 Cas9. B-C. Litter size (B) and litter interval (C) of six *miR-506* KO lines, at least 10
1340 litters from 3 different breeding pairs for each KO line were counted. Dunnett's
1341 multiple comparisons test as the posthoc test following one-way ANOVA was used
1342 for the statistical analysis. ns, not significant. * and **indicate adjusted p-value
1343 <0.05 and 0.01, respectively. D-F. Analyses of testis weight (D), sperm counts (E)
1344 and sperm motility (F) in four *miR-506* KO lines. n ≥ 3 and Dunnett's multiple
1345 comparisons test as the posthoc test following one-way ANOVA was used for the
1346 statistical analysis. G. Testicular histology of WT and four *miR-506* KO lines
1347 showing largely normal spermatogenesis. Scale bars = 50 μm. H. Representative
1348 genotyping results of the sequential mating experiments. I. Sequential mating of
1349 WT female mice with mTmG and quinKO males. Upper panel, an overview of the
1350 polyandrous mating scheme. "mTmG V.S. quinKO": mTmG male mice mated first;
1351 "quinKO V.S. mTmG": quinKO male mice mated first. J. Percentage of mTmG
1352 blastocysts obtained from IVF using WT MII oocytes and mixed sperm from mTmG
1353 (control) and quinKO males at different ratios. Data were based on three
1354 independent IVF experiments. The expected ratio was indicated as the blue line.
1355 Chi-squared test was used for statistical analyses. * and **** indicate p<0.05 and
1356 0.0001, respectively. K. Percentage of mTmG embryos obtained from co-artificial
1357 insemination using different ratios of mTmG and quinKO sperm. Data were based
1358 on 9 and 3 independent AI experiments for the 1:1 and 1:4 sperm ratio (mTmG:
1359 quinKO), respectively. The expected ratio was indicated as the blue line. Chi-
1360 squared test was used for statistical analyses. *, p<0.05.
1361
1362
1363 **Fig. 5. Target genes and genetic compensation of the X-linked *miR-506***
1364 **family miRNAs.** A. Intersections of the DETs among different KO testes. B. GO
1365 term enrichment analyses of the 431 DETs shared among the four different miR-
1366 506 KO testes. C. MA plots showing the expression levels of the miR-506 family
1367 miRNAs in WT, sKO, tKO, quadKO, and quinKO testes. Three biological replicates
1368 (n=3) were used for sRNA-seq analyses.
1369
1370
1371 **Figure 6. Rapid evolution of the X-linked *miR-506* family miRNAs correlates**
1372 **with increased complexity of genetic networks that regulate**

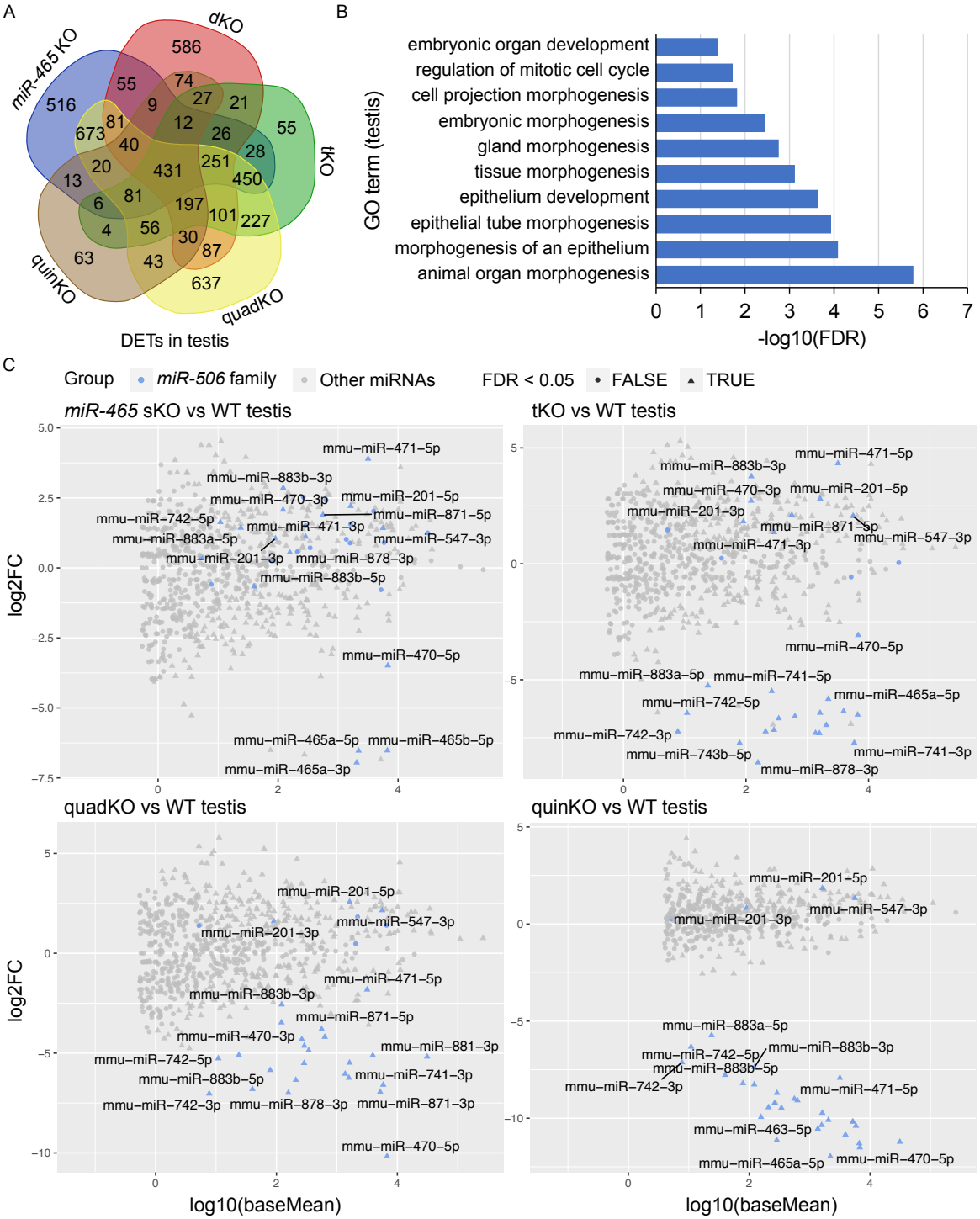
1373 **spermatogenesis across mammalian species.** A. Overlap between the
1374 dysregulated targets in mice and the predicted targets in humans and rats. B.
1375 Comparison of the cumulative distribution between the *miR-506* family targeting
1376 sites and the other regions in humans. *: $p < 0.05$; t-test was used for statistical
1377 analyses. C. Comparison of the cumulative distribution between the *miR-506*
1378 family targeting sites and the other regions in mice. *: $p < 0.05$; t-test was used for
1379 statistical analyses. D. Paired comparison of the PhyloP score between the *miR-*
1380 *506* family targeting sites and the other regions in humans. E. Paired comparison
1381 of the PhyloP score between the *miR-506* family targeting sites and the other
1382 regions in mice. F. Comparison of the number of the targets per miRNA for the X-
1383 linked *miR-506* family in mice and humans. *: $p < 0.05$; t-test was used for
1384 statistical analyses. G. The number of target sites within individual target mRNAs
1385 in both humans and mice. H. Schematics show that human *miR-506* family
1386 miRNAs have more targets relative to those of mice during evolution.
1387

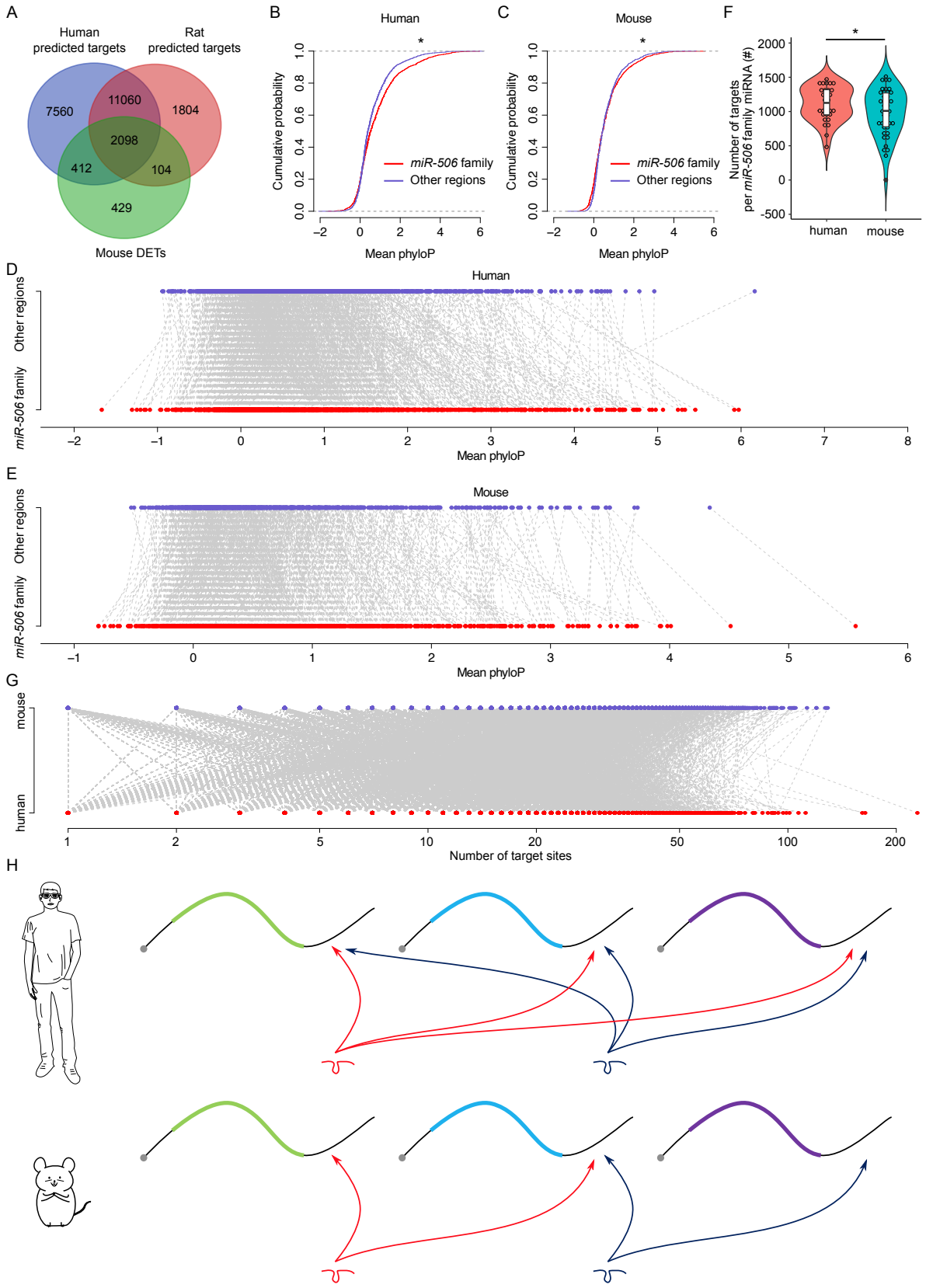

A
Slitrk2-Fmr1 are genetically clustered
Slitrk2-Fmr1 cluster moves to the X chromosome
miRNA clusters emerge near Slitrk2-Fmr1
SLITRK2
SmiR
FmiR
FMR1
Human
Rhesus
Marmoset
Mouse
Rat
Dog
Horse
Opossum
Platypus
Chicken
Zebra finch
Frog
Zebrafish
chrX
2.1 Mbp
2.0 Mbp
1.7 Mbp
1.6 Mbp
13Mbp
475 Kbp
400 Kbp
900 Kbp
chr7
chr6
chr4
chr4A
chr1
chr14
Ultra15
Ultra347
GL174275
GL172738
B
Human Chr X
Slitrk2
SmiR
(miR-892c ~ miR-891a)
FmiR
(miR-513c ~ miR-514a3)
Fmr1
PhyloP
PhastCons
C
Mean phyloP
CDS IGR piRNA SmiR FmiR miRNA
D
Mean phastCons
CDS IGR piRNA SmiR FmiR miRNA
E
Derived allele frequency
piRNA SmiR FmiR miRNA
F
Mean nucleotide diversity
piRNA SmiR FmiR miRNA


A
hg38_rmsk_MER91C
Human
hsa-mir-513c
hsa-mir-513b
hsa-mir-513a-1
hsa-mir-513a-2
Rhesus
Mml-Mir-506-P6a
Mml-Mir-506-P6c2
Mml-Mir-506-P6c1
Mml-Mir-506-P6b1
Mml-Mir-506-P6b2
Marmoset
Cja-Mir506-P6b1
Cja-Mir506-P6b2
Rat
Rno-Mir-506-P6
Guinea pig
Cpo-Mir-506-P6
Rabbit
Ocu-Mir-506-P6
Tenrec
Ete-Mir-506-P6e
Ete-Mir-506-P6f
Cow
Bta-Mir-506-P6
Horse
eca-mir-506a
eca-mir-8908n
eca-mir-508
eca-mir-507a
eca-mir-514a
eca-mir-514b
Dog
cfa-mir-507b
cfa-mir-506
cfa-mir-508b
cfa-mir-508a
Green sea turtle
cmy-miR-514b
B
C
D
hg38_rmsk_MER91C_hsa-mir-513c
SRR13274701
Coverage depth (x1000)
Position
E
Ago2
pCI
pCI-Mer91c
hsa-5p
hsa-3p
U6
F
pCI
pCI-Mer91c
pCI + Ago2
pCI-Mer91c + Ago2
Relative miRNA expression (miRNA/U6)
ns
hsa-5p
hsa-3p


A
Testis
Pachytene
Round
Cauda Sperm
FDR (Ro/Pa)
FDR (Sp/Pa)
mmu-miR-463-3p
mmu-miR-463-5p
mmu-miR-465a-3p
mmu-miR-465a-5p
mmu-miR-465b-5p
mmu-miR-465c-5p
mmu-miR-470-3p
mmu-miR-470-5p
mmu-miR-471-3p
mmu-miR-471-5p
mmu-miR-741-3p
mmu-miR-741-5p
mmu-miR-742-3p
mmu-miR-742-5p
mmu-miR-743a-3p
mmu-miR-743a-5p
mmu-miR-743b-3p
mmu-miR-743b-5p
mmu-miR-871-3p
mmu-miR-871-5p
mmu-miR-878-3p
mmu-miR-878-5p
mmu-miR-880-3p
mmu-miR-880-5p
mmu-miR-881-3p
mmu-miR-881-5p
mmu-miR-883a-3p
mmu-miR-883a-5p
mmu-miR-883b-3p
mmu-miR-883b-5p
C57BL/6J mice
B
SmiR
FmiR
logCPM (human)
hsa-mir-888
hsa-mir-891a
hsa-mir-891b
hsa-mir-892a
hsa-mir-892b
hsa-mir-892c
hsa-mir-506
hsa-mir-507
hsa-mir-508
hsa-mir-509-1
hsa-mir-509-2
hsa-mir-509-3
hsa-mir-510
hsa-mir-513b
hsa-mir-513c
hsa-mir-514b
C
SmiR
FmiR
LogCPM (marmoset)
cja-mir-888
cja-mir-891a
cja-mir-892a
cja-mir-892b
cja-mir-506
cja-mir-508
cja-mir-509-2
cja-mir-510
cja-mir-513a-1
cja-mir-513a-2
cja-mir-513a-3
cja-mir-513b-1
cja-mir-513b-2
cja-mir-514a
cja-mir-514b
D
Immature horse
Mature horse
LogCPM (horse)
eca-mir-8908a-1
eca-mir-8908a-2
eca-mir-8908b-1
eca-mir-8908b-2
eca-mir-8908b-3
eca-mir-8908b-4
eca-mir-8908b-5
eca-mir-8908c-1
eca-mir-8908c-2
eca-mir-8908c-3
eca-mir-8908d-1
eca-mir-8908d-2
eca-mir-8908e
eca-mir-8908f-1
eca-mir-8908f-2
eca-mir-8908f-3
eca-mir-8908f-4
eca-mir-8908g-1
eca-mir-8908h
eca-mir-8908i
eca-mir-8908j
eca-mir-8908k-1
eca-mir-8908k-2
eca-mir-8908l-1
eca-mir-8908l-2
eca-mir-8908m
eca-mir-9142
eca-mir-9143
eca-mir-506a
eca-mir-507a
eca-mir-507b
eca-mir-508
eca-mir-509a
eca-mir-509b
eca-mir-514
eca-mir-514a
eca-mir-514b
eca-mir-8908n
SmiR
FmiR


A
ChrX
Slitrk2
(19 miRNAs)
(3 miRNAs)
Fmr1
miR-883a Cluster (4 miRNAs, ~4 kb)
miR-471 Cluster (2 miRNAs, ~2.8 kb)
miR-741 Cluster (4 miRNAs, ~ 4.8 kb)
miR-470 Cluster (3 miRNAs, ~ 12 kb)
miR-465 Cluster (6 miRNAs, ~16.4 kb)
WT
miR-883a sKO
miR-465 sKO
dKO
tKO
quadKO
quinKO
XS
B
litter size (#pups)
C
litter interval (days)
D
Testis weight (mg)
E
Sperm counts (x10^6)
F
Sperm motility (%)
G
WT
miR-465 sKO
tKO
quadKO
quinKO
50 μm
H
quinKO V.S. mTmG
mTmG V.S. quinKO
M
H2O
WT
mTmG
I
hCG 8 PM
2nd male ovulation peak 6 AM
7:30 AM
4 PM
mating
1st
2nd
Sperm
Polyandrous mating
100%
89.30%
10.70%
mTmG V.S. quinKO
quinKO V.S. mTmG
mTmG
quinKO
J
IVF
mTmG% (out of blastocysts)
1:0
0:1
1:1
4:1
1:4
Sperm ratio (mTmG: quinKO)
K
Co-artificial insemination
mTmG percentage (%)
1:1
1:4
n = 96
n = 17
Sperm ratio (mTmG: quinKO)


A
miR-465 KO
dKO
tKO
quinKO
quadKO
586
55
74
27
21
516
9
12
26
55
673
81
40
28
20
431
251
450
13
6
81
197
101
227
4
56
30
87
63
43
637
DETs in testis
B
GO term (testis)
embryonic organ development
regulation of mitotic cell cycle
cell projection morphogenesis
embryonic morphogenesis
gland morphogenesis
tissue morphogenesis
epithelium development
epithelial tube morphogenesis
morphogenesis of an epithelium
animal organ morphogenesis
0 1 2 3 4 5 6 7
-log10(FDR)
C
Group
miR-506 family
Other miRNAs
FDR < 0.05
FALSE
TRUE
miR-465 sKO vs WT testis
tKO vs WT testis
quadKO vs WT testis
quinKO vs WT testis
log2FC
log10(baseMean)
mmu−miR−471−5p
mmu−miR−883b−3p
mmu−miR−470−3p
mmu−miR−201−5p
mmu−miR−742−5p
mmu−miR−871−5p
mmu−miR−471−3p
mmu−miR−883a−5p
mmu−miR−547−3p
mmu−miR−201−3p
mmu−miR−878−3p
mmu−miR−883b−5p
mmu−miR−470−5p
mmu−miR−465a−5p
mmu−miR−465b−5p
mmu−miR−465a−3p
mmu−miR−741−5p
mmu−miR−742−3p
mmu−miR−743b−5p
mmu−miR−878−3p
mmu−miR−881−3p
mmu−miR−463−5p


A
Human predicted targets
Rat predicted targets
7560
11060
1804
2098
412
104
429
Mouse DETs
B
Human
Cumulative probability
miR-506 family
Other regions
Mean phyloP
C
Mouse
Cumulative probability
miR-506 family
Other regions
Mean phyloP
F
Number of targets per miR-506 family miRNA (#)
human
mouse
D
Human
Other regions
miR-506 family
Mean phyloP
E
Mouse
Other regions
miR-506 family
Mean phyloP
G
mouse
human
Number of target sites
H