University of Nevada, Reno

Functions of X-linked *miR-506* family in Reproduction

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy in Cellular and Molecular Biology

by

Zhuqing Wang

Dr. Wei Yan, Dissertation Advisor

August, 2020

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

entitled

be accepted in partial fulfillment of the
requirements for the degree of

Advisor

Committee Member

Committee Member

Committee Member

Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

ABSTRACT

MicroRNAs (miRNAs) are ~22nt small non-coding RNAs that play a pivotal role in both development and adulthood physiology. Our lab has previously discovered that the *miR-506* family, which consists of 22 miRNAs existing in 5 clusters expanding a ~62kb region near *Slitrk2* and 1 cluster expanding a ~22kb region close to *Fmr1* on the X chromosome in mice. Derived from the same ancestors, these miRNAs are preferentially expressed in the testis and have undergone rapid evolution with variable seed sequences, whereas the two protein-coding genes (*Slitrk2* and *Fmr1*) flanking the 6 clusters remain highly conserved among all mammalian species. The quick evolution and testis preferential expression of these X-linked miRNAs strongly suggest that they are under selective pressure and play an important role in fine-tuning certain molecular events to ensure the production of high quality sperm. However, these hypotheses remain untested. My dissertation projects aimed to unveil the physiological roles of the *miR-506* family miRNAs. Using the CRISPR-Cas-based genome editing technology, we deleted individual or 5 clusters of the *miR-506* family in the mouse genome to analyzed potential phenotype. Interestingly, we found that ablation of one or two miRNAs, or even one cluster of several miRNAs did not generate any discernable defects in either spermatogenesis or male fertility. When we deleted four or five of the 6 miRNA clusters, we started to see some phenotype in these KO males, characterized by slightly reduced litter sizes and longer litter intervals despite normal testis weight and normal sperm counts. When the typical one male-one female mating scheme was used, the KO males were only subfertile. However,

when a female was mated sequentially with either a wild type male first and a KO second, or a KO first followed by a WT male, no or much fewer pups were derived from the KO sperm, suggesting the KO sperm are less fit in fertilizing eggs. Indeed, *in vitro* fertilization assays showed that the KO sperm were less competitive in fertilizing wild type eggs. Our data suggest that sperm produced in the absence of these X-linked miRNAs are less fertile/competitive than wild-type sperm, suggesting that the *miR-506* family miRNAs function to fine-tune certain molecular processes that render sperm better fitness. Among the *miR-506* family, 6 miRNAs of the *miR-465* cluster have similar seed sequences and are preferentially expressed in the testis. Surprisingly, loss of *miR-465* cluster did not cause male fertility defects, but led to sex ratio bias among KO offspring, and the distorted sex ratio was found to result from female-biased lethality as early as embryonic day (E) 8.5. It turns out that *miR-465* cluster miRNAs are also highly expressed in extra-embryonic tissues at E7.5. The female-biased lethality is likely due to dysregulation numerous target genes known critical for the survival of the female embryos, e.g., *Alkbh1*, a tRNA demethylation enzyme. Taken together, the data presented in this dissertation uncovered the physiological roles of the *miR-506* family. Consistent with our previous reports, miRNAs tend to act as a group, and loss of function of one miRNA can be easily compensated by other members of the same cluster or the same family. Thus, inactivation of a single miRNA rarely leads to a discernable phenotype in mice. To reveal the true physiological role of miRNAs within one cluster or one family, simultaneous inactivation of most or all of the miRNAs is required.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my deepest gratitude to Dr. Yan for inviting me to the US and having me pursue my Ph.D. degree in his lab. Toward the end of my MSc program, I applied for several Ph.D. programs, but was rejected several times largely due to my lack of first-authored papers. While I was feeling hopeless, it was Dr. Yan who brought me the silver lining. Whenever I ran into problems in research, Dr. Yan was always the best person to talk with. He is knowledgeable, always standing at the forefront of technologies, giving us space to explore. I have published 5 first-authored papers so far, and 3 more on the way; this wouldn't have been possible without Dr. Yan's tremendous support.

I would also like to thank Dr. Huili Zheng and other lab members, especially Dr. Chong Tang, Yue Wang, Shawn Wang, Dr. Tong Zhou, Hayden McSwiggin, Dr. Tian Yu, Dr. Sheng Chen, Dr. Jiaxin (Jason) Xing, Kimberly Castaneda-Garcia, and Dr. Hetan (Hedy) Wang for their support inside and outside lab.

I would like to thank my committee members, Drs. Qi Chen, Seungil Ro, Yumei Feng Earley, and Yong Zhang, for their constructive suggestions and criticism on my comprehensive exam and my dissertation. The knowledge and experiences learned from them will be beneficial forever to my career development.

Finally, I would like to thank the supports from my family and friends, especially Wenye Ye for his enormous help in my daily life.

Table of Contents

ABSTRACT ..... **i**

ACKNOWLEDGMENTS ..... **iii**

Table of Contents ..... **iv**

List of Tables ..... **viii**

List of figures ..... **ix**

CHAPTER I: Introduction **1**

Spermatogenesis **1**

SSCs self-renewal 3

Meiosis 5

Spermiogenesis 6

The cycle of the seminiferous epithelium 6

Placental development **9**

MicroRNA (miRNA) **11**

miRNA is transcribed by RNA pol II 12

Primary miRNA processing 13

Pre-miRNA processing 14

miRNA targeting 14

Methods for identifying miRNA targets 15

Genome editing **20**

Structure-based mechanisms of CRISPR-Cas9 in targeting DNA 26

Applications of CRISPR-Cas9/Cas12a 27

Major obstacles in utilizing CRISPR-Cas9/Cas12a technology 35

References 51

CHAPTER II: Rapidly evolving X-linked *miR-506* family miRNAs control sperm fitness through fine-tuning spermatogenesis 86

Abstract 87

Instruction 88

Results 90

The majority of X-linked *miR-506* family are located within the *Slitrk2-Fmr1* locus across species 90

LINE retrotransposons seem to drive the expansion of X linked *miR-506* family derived from MER91C DNA transposon 93

The X linked *miR-506* family is highly abundant in testis and sperm 96

Ablation of X linked *miR-506* compromise sperm fitness 98

X linked *miR-506* family miRNAs target genes throughout spermatogenesis and compensate with each other 100

X linked *miR-506* family gains targets during evolution 102

Discussion 104

Coevolution of 3'UTR and miRNAs 104

Divergent roles of these miRNAs in modulating the same genes 106

ACKNOWLEDGEMENTS 107

AUTHOR CONTRIBUTIONS 107

COMPETING FINANCIAL INTERESTS 107

Materials and Methods 107

Animal use and generation of global knockout mice 107

In vitro fertilization (IVF) 108

Serial mating 109

Mouse genotyping 109

Purification of germ cells 109

RNA extraction, libraries construction and qPCR analyses 110

Large and small RNA-Seq data analysis 110

Statistical analyses 111

References 112

CHAPTER III. Loss of X linked *miR-465* cluster induces male-biased sex ratio 140

Abstract 141

Introduction 142

Results 143

The X-linked *miR-465* cluster is preferentially expressed in extraembryonic tissue at E7.5. 143

Ablation of *miR-465* cluster lead to male-biased sex ratio 145

Loss of the *miR-465* cluster impaired female placental development. 146

Discussion 148

Materials and Methods 149

Animal use and generation of global knockout mice 149

Mouse genotyping 150

In situ hybridization 150

RNA extraction, libraries construction, and qPCR analyses 151

Large and small RNA-Seq data analysis 152

qPCR and Western Blot (WB). 152

Luciferase assay 152

Statistical analyses 153

References 154

CHAPTER IV: Conclusion and future direction. 167

References 169

List of Tables

Table 1. Commonly used miRNA target prediction tools……..…...15

Table 2. Websites and properties of several in silico design tools……...……..41

Table 3. Methods for CRISPR delivery……..……50

List of figures

Figure 1-1. Murine spermatogenesis and testicular cell types…..……..……2

Figure 1-2. Three SSCs self-renewal models…….……..4

Figure 1-3. Key steps in meiotic prophase I……..…….5

Figure 1-4. Schematic illustration of and histology of cross sections from the 12 stages of seminiferous epithelial cycle in mice …..……..…….8

Figure 1-5. The detailed structure of mouse fully functional placenta……..……..9

Figure 1-6. Key steps during placental development in mice and humans…….10

Figure 1-7. miRNA biogenesis…….……12

Figure 1-8. Canonical structure and motifs of pri-miRNA……..……13

Figure 1-9. miRNA target pattern…….……15

Figure 1-10. Overview of CLIP…….……18

Figure 1-11. Gene targeting procedure…….……..21

Figure 1-12. Gene editing strategies…….……...22

Figure 1-13. Structure-based mechanisms of CRISPR-Cas9 in target DNA recognition and cleavage…….……..27

Figure 1-14. Quaternary structure of SpCas9 upon binding to sgRNA…...…….30

Figure 1-15. Comparisons of specificity and efficiency among different Cas9 variants…….39

Figure 2-1. Evolution divergence, genomic location and sequence alignment of X linked *miR-506* family……..…….……121

Figure 2-2. Expression of X linked *miR-506* family in mice, rat, humans, monkeys and horses……..……...……...123

Figure 2-3. KO of X linked *miR-506* family compromised sperm fitness…..…..125

Figure 2-4. Target genes of X linked *miR-506* family……..…..127

Figure 2-5. Targets of X linked *miR-506* family are conserved among different species……...128

Figure 2-S1. Multiz Alignment & Conservation analysis of human X linked *miR-506* family across 100 species……..….130

Figure 2-S2. Evolution of X linked *miR-506* family based on Multiz Alignment & Conservation and Dot plot analysis…….……..132

Figure 2-S3. Phylogram tree of X linked *miR-506* family…..……..…..…134

Figure 2-S4. Genetic composition of X linked *miR-506* family, and sequences alignment of precursor *miR-891* and *miR-892*……..……..……..135

Figure 2-S5. The purity of germ cells after STA-PUT and the phenotype of dKO……..……..137

Figure 2-S6. Dysregulated targets and genetic compensation of X linked *miR-506* family……..……138

Figure 3-1. Expression pattern and conservation of *miR-465*…….…….…….160

Figure 3-2. Generation of *miR-465* cluster KO mice…….……..162

Figure 3-3 *MiR-465*$^{-/-}$ homozygous inbreeding showed highly skewed sex ratio…….……163

Figure 3-4 RNA-seq analysis of *miR-465* KO and WT mice……165

CHAPTER I: Introduction

Spermatogenesis

Spermatogenesis refers to the differentiation of spermatogonia stem cells to spermatids within the seminiferous epithelium in the testis (Leblond and Clermont, 1952). The process includes three major phases: mitotic (self-renewal and differentiation of spermatogonial stem cells (SSCs)), meiotic (spermatocyte differentiation and division), and haploid (also known as spermiogenesis, differentiation of round spermatids into spermatozoa) phases (Oakberg, 1956b). Spermatogenesis takes ~35 days in mice (Oakberg, 1956b) and ~64 days in humans (Heller and Clermont, 1963).

In the mitotic phase, SSCs can both self-renew and divide followed by differentiation into type A spermatogonia. Once committed to differentiation, type A spermatogonia further divide and differentiate into intermediate, and then type B spermatogonia. Derived from type B spermatogonia, the primary spermatocytes embark on a long meiosis prophase I, going through the states of leptotene, zygotene, pachytene, diplotene, and diakinesis, followed by two consecutive cell divisions (MI and MII) to form haploid round spermatids. Round spermatids then undergo spermiogenesis, which involves nuclear condensation and elongation, acrosome formation, flagellar growth, etc., resulting in spermatozoa (Turner, 2007) (Figure 1-1A).

In mice, the seminiferous epithelium in postnatal day 6 (P6) testes only contains primitive type A spermatogonia and Sertoli cells. By P8, both type A and type B spermatogonia are present. Meiotic prophase starts at P10, and primary

spermatocytes reach to early and late pachytene stages by P14 and P18, respectively. The populations of secondary spermatocytes and haploid spermatids enrich on days 18 and 20, respectively (Bellve et al., 1977).

In addition to germ cells, the testis also contains several types of somatic cells, including Sertoli cells, peritubular myoid cells, macrophages, and Leydig cells (Figure 1-1B). Sertoli cells are situated on the basal lamina and have large cytoplasm extending from the basal membrane to the lumen of the seminiferous tubules enveloping all developing germ cells within the seminiferous tubules. Sertoli cells function as supporting cells to provide signaling molecules essential for normal spermatogenesis. In between the seminiferous tubules are Leydig cells, peritubular myoid cells, and macrophages. Leydig cells in the interstitial tissue produce testosterone that plays an important role in modulating Sertoli cell functions (Russell, 1990).

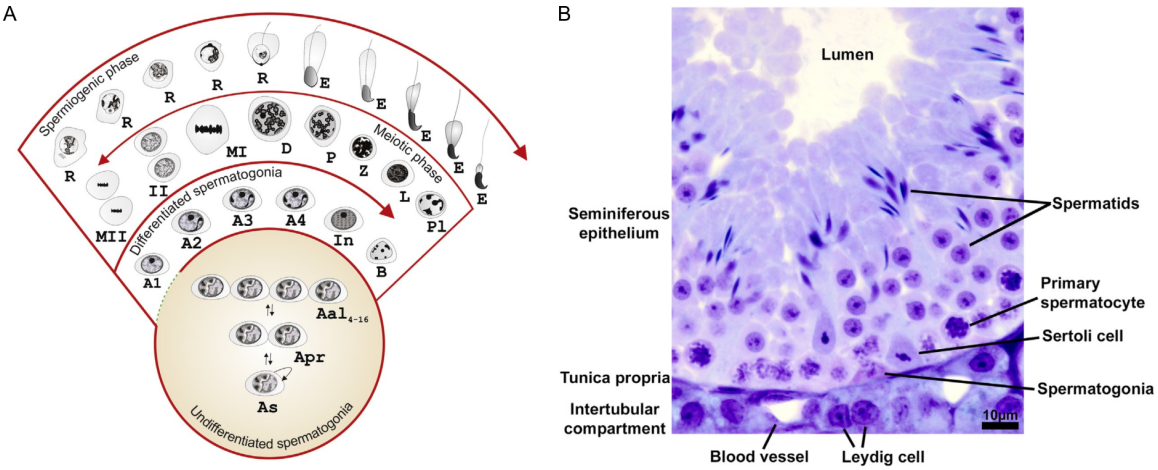

Figure 1-1. Murine spermatogenesis (A) and testicular cell types (B), modified from (Lara et al., 2018). A. Spermatogenesis consists of three phases: SSC self-renewal and differentiation in mitotic phase; meiotic phase and haploid or

spermiogenic phase. Spermatogonia: undifferentiated type A spermatogonia (A_s, single; A_{pr}, paired; A_{al}4-16, aligned); A1-4, differentiated type A spermatogonia; In, intermediate spermatogonia; B, type B spermatogonia. Primary spermatocytes: Pl, preleptotene spermatocytes; L, leptotene spermatocytes; Z, zygotene spermatocytes; P, pachytene spermatocytes; D, diplotene spermatocytes. MI, meiosis I. MII, meiosis II. II, Secondary spermatocyte. R, Round. E, elongating/elongated spermatids. B. Cell types inside the testis. In addition to germ cells (spermatogonia, spermatocytes, and spermatids), the testis also contains somatic cells including Sertoli cells and Leydig cells.

SSCs self-renewal

Spermatogonia include undifferentiated type A spermatogonia (A_s, A_{pr}, and A_{al}4-16), differentiated type A spermatogonia (A1-4), intermediate and type B spermatogonia. Among them, the undifferentiated type A spermatogonia have the ability to self-renew or differentiate into other types of spermatogonia. Three models of SSCs self-renewal have been proposed: A_s model (Huckins, 1971; Oakberg, 1971), fragmentation model (Nakagawa et al., 2007), and hierarchical model (Helsel et al., 2017) (Figure 1-2, for detailed reviews, see (de Rooij, 2017)) The A_s model was first proposed in 1971 by Huckins (Huckins, 1971) and Oakberg (Oakberg, 1971). In this model (Figure 1-2A), the A_s spermatogonia are the SSCs, and can divide into either two daughter SSCs that migrate away from each other or A_{pr} spermatogonia that are bridged by incomplete cytokinesis. The A_{pr}

spermatogonia further divide into aligned spermatogonia (A_{al}4, 8, and 16). The A_{al}4-16 spermatogonia then differentiate into type B spermatogonia, which are destined to enter meiotic prophase. The fragmentation model (Figure 1-2B) suggests that all A_s, A_{pr}, and A_{al} have the potential to self-renew, where A_{pr} and A_{al} reverse to the self-renewal potential by fragmentation of pairs or chains (Nakagawa et al., 2007). In the hierarchical model (Figure 1-2C), only some A_s spermatogonia have the self-renewal ability, while the rest of them barely do (Helsel et al., 2017), implying a hierarchy with in the A_s spermatogonia.

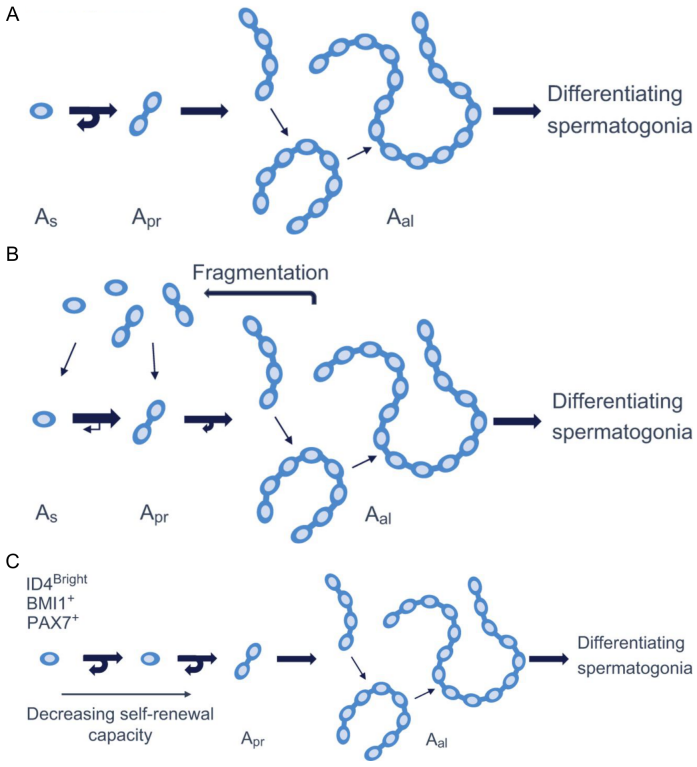

Figure 1-2. Three SSCs self-renewal models, modified from (de Rooij, 2017). A. A_s model. B. Fragmentation model. C. Hierarchical model. A_s, single; A_{pr}, paired; A_{al}, aligned.

Meiosis

After differentiation from type B spermatogonia, the primary spermatocytes enter meiosis S phase (preleptotene), then undergo the long meiosis prophase I, including leptotene, zygotene, pachytene, diplotene, and diakinesis, followed by two rounds of rapid meiosis (MI and MII) to form haploid round spermatids. During meiosis prophase I, double-stranded breaks (DSBs) form in leptotene, synapsis begins and DNA recombination repair occurs in zygotene, followed by synapsis ending and DNA resolution in pachytene (Figure 1-3) (Morgan et al., 2017; Turner, 2007). During meiosis, the X and Y chromosomes undergo meiotic sex chromosomal inactivation, resulting in inactivated X and Y chromosomes (Turner, 2007).

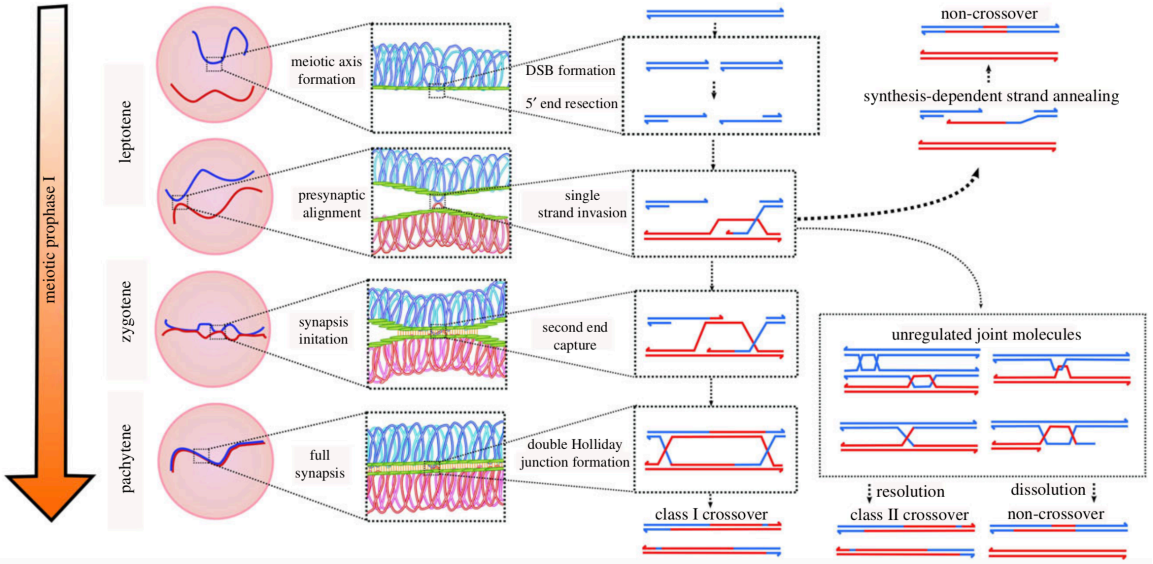

Figure 1-3. Key steps in meiotic prophase I (Morgan et al., 2017).

Spermiogenesis

Spermiogenesis refers to the development of round spermatids to spermatozoa. Several key events occur during spermiogenesis: 1) Formation of the chromatoid body that enriched with RNA binding proteins and RNA (Kotaja et al., 2006; Kotaja and Sassone-Corsi, 2007); 2) Nuclear chromatin condensation, during which histone is replaced with protamine (Carrell et al., 2007); 3) Formation of the acrosome that derived from the Golgi apparatus (Abou-Haila and Tulsiani, 2000); 4) Flagellum formation (Inaba, 2011); 5) Development of ectoplasmic specializations (Grove and Vogl, 1989) in between Sertoli cells and elongating/elongated spermatids; 6) Formation of tubulobulbar complexes between Sertoli cells and elongating/elongated spermatids, and elimination of cytoplasm (residual body) of spermatids (Russell, 1979); 7) Release of spermatids from the seminiferous epithelium to the lumen (spermiation) (O'Donnell et al., 2011).

The cycle of the seminiferous epithelium

The development of germ cells inside the testis follows a spatiotemporal order. Therefore, inside the seminiferous tubule of the testis, germ cells associate with each other successively in a specific pattern in any given area, and the pattern repeats itself. This phenomenon is termed as “cycle of the seminiferous epithelium” or the “cycle” (Leblond and Clermont, 1952) (Figure 1-4). Periodic acid-Schiff (PAS), which stains the acrosome derived from the Golgi apparatus, was used to

identify the stages of the cycle (Leblond and Clermont, 1952). Using this strategy, 6 stages of the cycle were identified in humans (Clermont, 1963), 14 in rats (Leblond and Clermont, 1952), and 12 in monkeys (Clermont and Leblond, 1959) and mice (Oakberg, 1956a; Russell, 1990). In stage I seminiferous tubules in mice, the acrosomal system is undetectable. In stage II, the acrosomal system starts to show up. In stage III, the acrosomal system remains round and locates on the nuclear surface. In stage IV, the acrosomal system flattens on the nuclear surface, the angle subtended by acrosomes from the center of the nucleus is around 40 degrees. In stage V, the angle subtended by acrosomes is greater than 40 degrees and less than 90 degrees. In stage VI, the angle subtended by acrosomes is from 95 degrees to 120 degrees, and the elongated spermatids are still embraced by Sertoli cells. In stage VII, the angle subtended by acrosomes is greater than 120 degrees, the nuclei of round spermatids are not in contact with the plasma membrane, and the elongated spermatids move to the lumen of the seminiferous epithelium. In stage VIII, the nuclei of round spermatids are in contact with the plasma membrane. In stage IX, the spermatids nuclei change into ovoid shape, the dorsal and ventral surfaces begin to show up. In stage X, the spermatids show a ventral angle, but not a dorsal angle. In stage XI, the spermatids show a dorsal angle. In stage XII, the elongating spermatids co-exist with meiotically dividing spermatocytes.

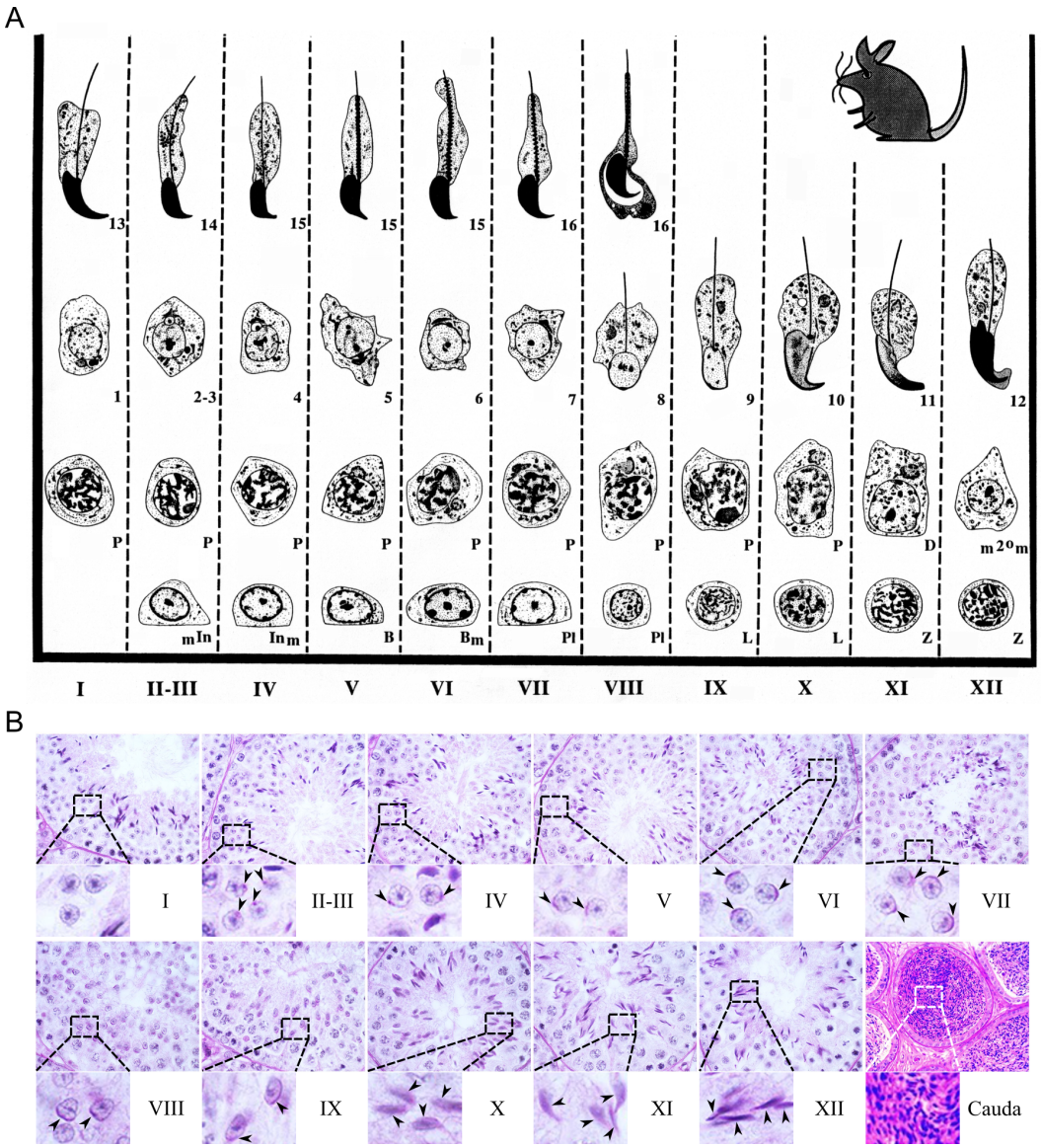

Figure 1-4. Schematic illustration (A) of and histology of cross-sections (B) from the 12 stages of the seminiferous epithelial cycle in mice, modified from Russell, 1990 and Yuan et al., 2015. Arrowheads indicate the location of the acrosomal system.

Placental development

The placenta is the transient interface between mother and fetus during pregnancy, which plays an essential role in exchanging gases, nutrients, and wastes, providing growth factors and hormones, as well as protecting babies from the immune response (Rossant and Cross, 2001). Fully developed placenta in humans and mice contain three main layers: the maternal decidua, the spongiotrophoblast (the junctional zone connecting fetoplacenta and maternal placenta), and the labyrinth (mainly villi derived from the fetus) (Figure 1-5) (Rai and Cross, 2014; Rossant and Cross, 2001; Watson and Cross, 2005).

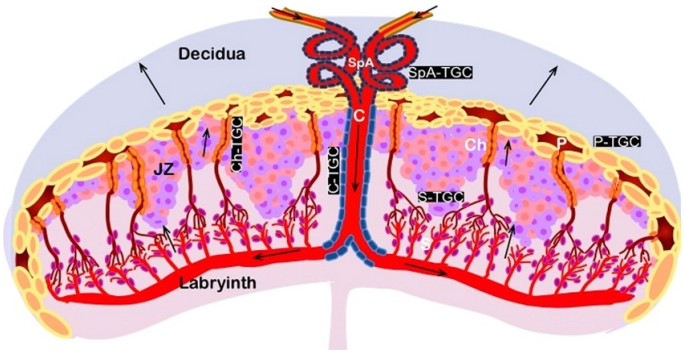

Figure 1-5. The structure of a fully-developed murine placenta (Rai and Cross, 2014), which consists of the maternal decidua, the junctional zone (JZ), and the labyrinth. SpA, spiral artery; TGC, trophoblast giant cells; C, maternal canal; Ch, channel; P, parietal venous spaces; S, sinusoidal spaces.

In mice, the placenta starts to develop during the blastocyst stage at embryonic day (E) 3.5. At the blastocyst stage, the blastocyst is composed of the inner cell mass (ICM) and the trophectoderm (TE) (Figure 6). The ICM will differentiate into epiblast (EPI) and primitive endoderm (PE), which will further form the embryo and yolk sac, respectively. The TE away from ICM (mural TE) will stop dividing but still

replicate their DNA to develop into trophoblast giant cells (TGC), whereas the TE in contact with ICM (polar TE) will form the extra-embryonic ectoderm (ExE) and the ectoplacental cone (EPC) at ~E6.5 (Hemberger et al., 2020; Rossant and Cross, 2001). The ExE further develops into the chorion. At ~E8.5, the allantois (All, extraembryonic mesoderm) derived from the posterior end of the embryo touches with chorion (Ch), and this process is termed chorioallantoic attachment (Watson and Cross, 2005). After the chorioallantoic attachment, the villi start to form at E9.0~E10.5. At E10.5, the initial structure of the placenta is already formed, whereas it is not fully functional until E14.5 (Watson and Cross, 2005).

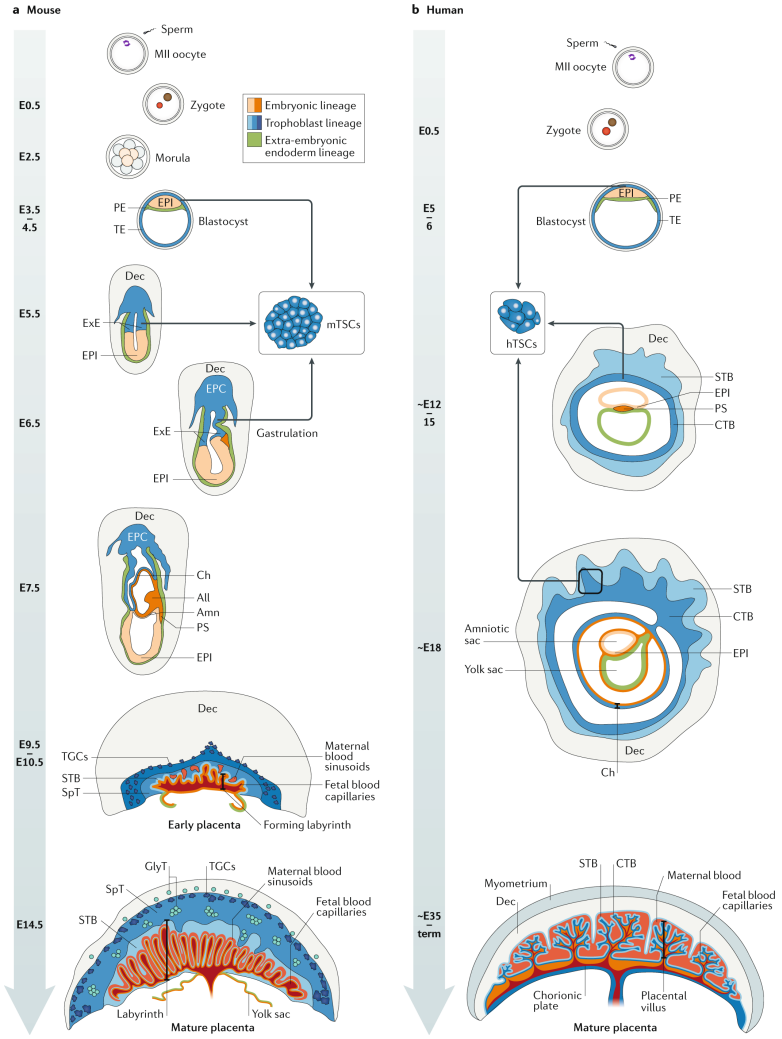

Figure 1-6. Key steps during placental development in mice (left) and humans (right) (Hemberger et al., 2020). Amn, amnion; PS, primitive streak; STB, syncytiotrophoblast; SpT, spongiotrophoblast; GlyT, glycogen trophoblast; CTB, cytotrophoblast.

MicroRNA (miRNA)

MiRNAs are ~22nt small non-coding RNAs that play a pivotal role in post-transcriptional regulation. Similar to mRNA, the primary miRNA (pri-miRNA) is transcribed by RNA polymerase II (pol II), and therefore contains 5' cap and 3' poly (A) tail (Lee et al., 2004). After transcription, the pri-miRNA is processed in the nucleus by a microprocessor containing the RNase III enzyme Drosha and the DGCR8 (DiGeorge critical region 8, a double-stranded RNA binding protein), forming a ~70nt precursor miRNA (pre-miRNA) (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Lee et al., 2003). The pre-miRNA is then transported to the cytoplasm by Exp5 (exportin 5) (Bohnsack et al., 2004; Lund et al., 2004; Yi et al., 2003) and further processed by the RNase III enzyme Dicer to form ~22nt miRNA duplex (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). Then the miRNA duplex is handed over to Argonaute (AGO) proteins, one strand of the miRNA duplex is selected as the main miRNA, and the complementary one is destined for degradation (Liu et al., 2004; Meister et al., 2004) (Figure 1-7). The chosen miRNA is further loaded into miRISC (miRNA induced silencing complex) to target mRNA. In all four AGO proteins (AGO1~4) in the human genome, AGO2 is the only protein that has the catalytic domain in cleaving mRNA (Liu et al., 2004; Meister et al., 2004).

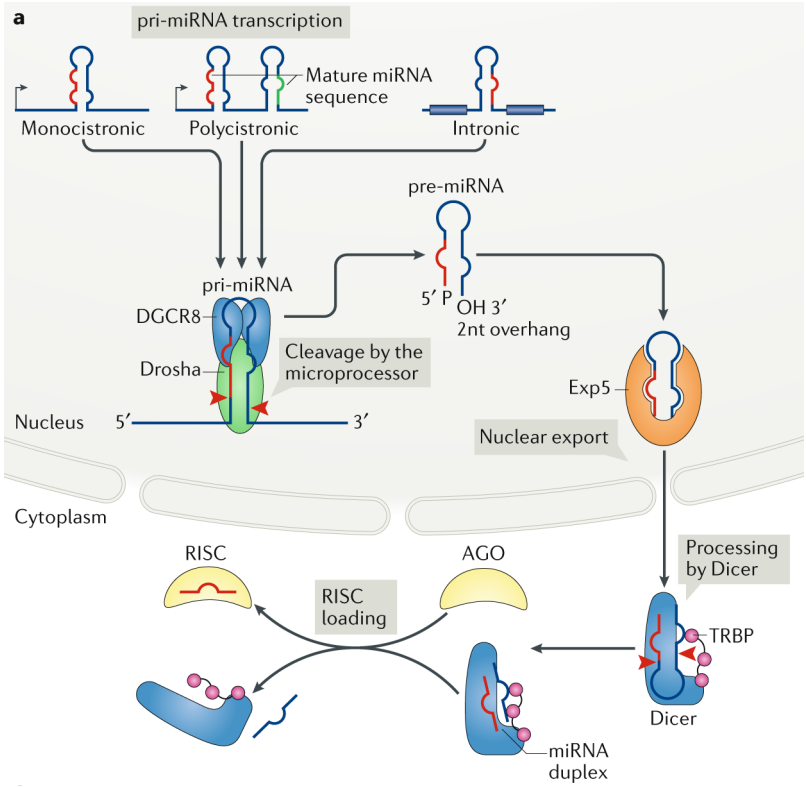

Figure 1-7. miRNA biogenesis (Treiber et al., 2019). DGCR8, DiGeorge critical region 8; Exp5, exportin 5; AGO, Argonaute; RISC, RNA induced silencing complex.

miRNA is transcribed by RNA pol II

To test which RNA polymerase is responsible for miRNA transcription, a team led by V. Narry Kim carried out several experiments (Lee et al., 2004). First, they showed that eIF4E, which has a high affinity to 7-methyl guanosine mRNA cap, binds to pri-miRNAs, indicating that the pri-miRNAs contain the 5’ cap as the mRNA does. Oligo d(T) enriched reverse transcription (RT)-PCR showed that the pri-miRNAs contain poly (A). Treating cells with α-amanitin, an inhibitor of RNA pol II, reduced the expression of pri-miRNAs. Insertion of the miRNA promoter region to the 5’ end of a luciferase reporter enhanced the luciferase activity, whereas the addition of α-amanitin reduced the activity. Chromatin immunoprecipitation (ChIP) further confirmed that the RNA pol II, especially the unphosphorylated RNA pol II,

binds to pri-miRNAs, indicating the transcription initiation of pri-miRNA is mediated by RNA pol II.

Primary miRNA processing

Despite the hairpin structure, several motifs on the pri-miRNA play important roles in pri-miRNA processing (Figure 1-8). The microprocessor consists of a DGCR8 dimer and a Drosha endonuclease. Two DGCR8 form a dimer and interact with the UGU motif at the >10nt apical loop, whereas the Drosha interact with the UG motif at the end of the ~35nt stem (Auyeung et al., 2013; Fang and Bartel, 2015). Furthermore, the GHG mismatched motif and the CNNC motif that bind by SRSF3 (serine/arginine-rich splicing factor 3) also promote pri-miRNA processing (Auyeung et al., 2013; Fang and Bartel, 2015).

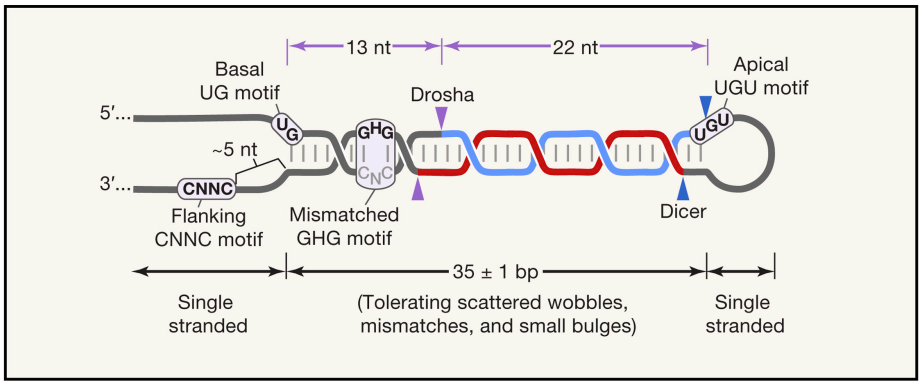

Figure 1-8. Canonical structure and motifs of pri-miRNA (Bartel, 2018).

Pre-miRNA processing

The Dicer contains a PiWi-Argonaute-Zwille (PAZ)-like domain, a platform, two RNase III domains (RIIID), and a helicase domain. The human Dicer forms an L-shaped structure, where the PAZ, platform, and two RIIID are along the axis, and the helicase domain branches out. The pre-miRNA is embedded in the Dicer protein, the 5' phosphate and the 3' end are within the platform 5' pocket and the PAZ 3' pocket, respectively. Two RIIID forms a dimer, and each RIIID cleaves one strand of the pre-miRNA. The human Dicer also interacts with TRBP that contains three double-stranded RNA binding domains (dsRBDs), in which two of them bind to the pre-miRNA, and the last one anchors to the Dicer helicase domain (Treiber et al., 2019). After cleavage by Dicer, the PAZ domain of Dicer hand over 3' overhang of the miRNA duplex to the PAZ domain of AGO. Then the 5' end of the chosen mature miRNA bind to MID domain of AGO; the passenger miRNA is degraded by AGO (Treiber et al., 2019).

miRNA targeting

With the guidance of miRNA, miRISC is able to target mRNA. In mammals, ~6bp matches (usually 2~7 nt of the 5' end of miRNA) between miRNA and target mRNA, termed the miRNA seed region, is sufficient for inducing mRNA repression or activation (Bartel, 2009), while the supplementary region (13~16) can increase target affinity (Sheu-Gruttadauria et al., 2019) (Figure 9).

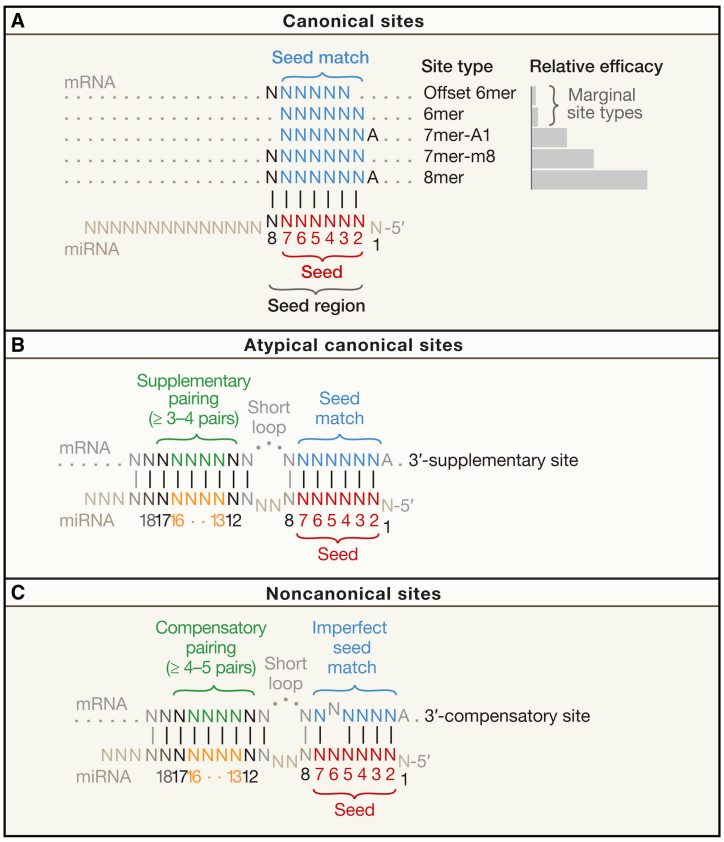

Figure 1-9. miRNA target pattern(Bartel, 2018).

Methods for identifying miRNA targets

The seed region is regarded as the crucial role in miRNA target prediction. Currently, several miRNA prediction tools have been developed based on this feature.

Table 1. Commonly used miRNA target prediction tools.

Name	Format	Link	Reference
TargetScan	Online/download	http://www.targetscan.org/vert_72/	(Agarwal et al., 2015)

microrna.org	Online/download	http://www.microrna.org/	(Betel et al., 2008)
miRWalk	Online/download	http://mirwalk.umm.uni-heidelberg.de/	(Sticht et al., 2018)
RNAhybrid	Online/download	https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid	(Rehmsmeier et al., 2004)
miRDB	Online/download	http://mirdb.org/	(Chen and Wang, 2020)
DIANA-microT	Online	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=microT_CDS/index	(Paraskevopoulou et al., 2013)
PicTar	Online	https://pictar.mdc-berlin.de/cgi-bin/PicTar_vertebrate.cgi	(Krek et al., 2005)

Despite the advances in computational prediction tools, whether the potential targets are bound by miRNAs still depends on the given physiological conditions. Therefore, as a complementary strategy, experimental verification is needed, such as cross-linking and immunoprecipitation (CLIP), miRNA overexpression/KO followed by RNA-seq/proteomics, and luciferase assay. CLIP (Chi et al., 2009) and its derivates (e.g. individual-nucleotide CLIP (iCLIP) (Konig et al., 2010), enhanced CLIP (eCLIP) (Van Nostrand et al., 2016), infrared-CLIP (irCLIP) (Zarnegar et al., 2016), and crosslinking, ligation, and sequencing of hybrids (CLASH) (Helwak et al., 2013)) could characterize the relationship between miRNA and mRNA *in vivo*. However, CLIP is not widely used due to its complicated steps. Therefore, some other alternative strategies were used to identify miRNA targets. One of these strategies is to overexpress/KO specific miRNA, followed by RNA-seq/proteomics. However, this strategy is unable to distinguish indirect targets from direct targets, and it may not reflect the *in vivo* situation. Another strategy is using luciferase as a reporter to reflect the function of miRNA to the target. Similar to overexpression/KO of specific miRNA followed by RNA-seq/proteomics, this strategy still can't mimic the *in vivo* situation, and can't perform high-throughput experiments, yet this system is easy to manipulate. Each system has its own advantages and disadvantages, and therefore, these systems are usually used together to identify the true targets.

1. Direct *in vivo* evidence (CLIP/CLASH)

Direct contacts between amino acids and nucleotides within cells can be persevered by UV crosslinking (Lee and Ule, 2018). After cell lysis, protein-RNA complexes are fragmented using RNase, followed by antibody pull-down and adaptor ligation. Following electrophoresis and transfer onto the membrane, the protein-RNA complexes are extracted from the membrane. The RNA is released from the protein-RNA complexes and followed by library construction. After next-generation sequencing, the protein-bound RNA can be analyzed (Figure 1-10).

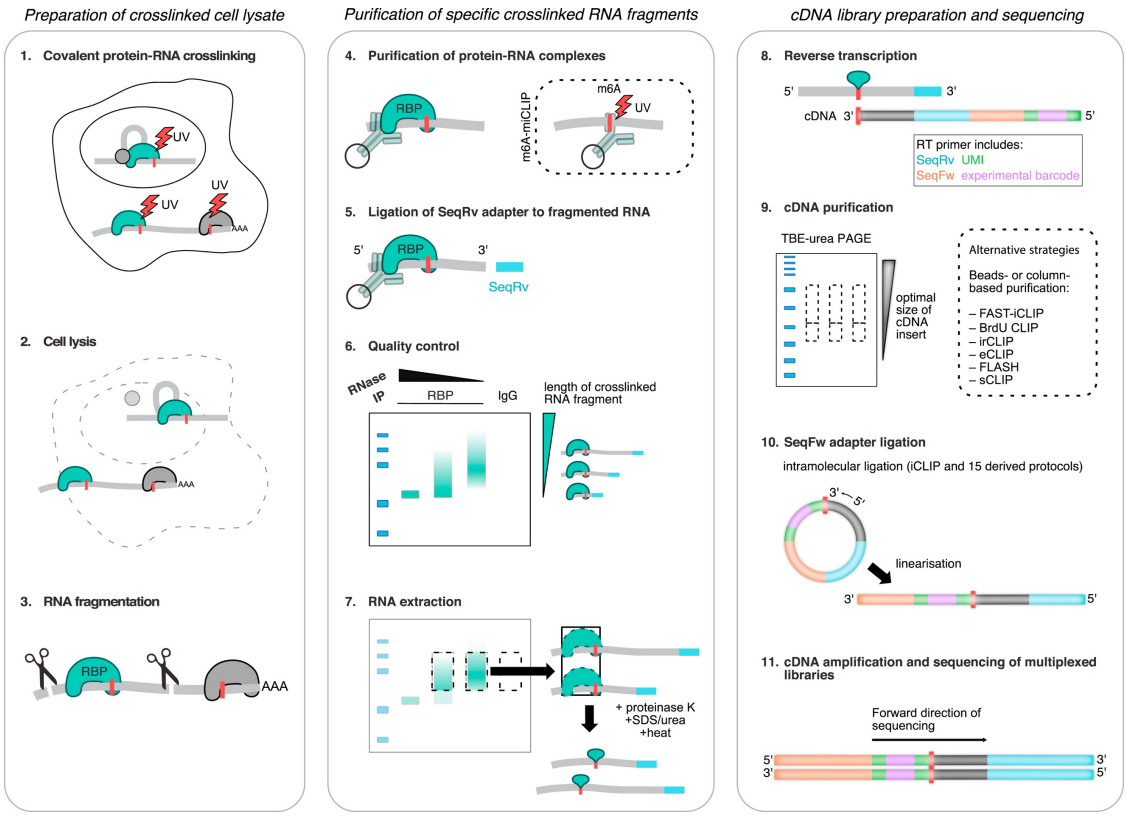

Figure 1-10. Overview of CLIP (Lee and Ule, 2018).

CLIP has been used to identify the *in vivo* targets; however, this technique is not widely used largely due to its complicated procedures (RNA pull-down + Western blot + library construction). First, the efficiency of UV crosslinking is very low, usually < 5% for each RNP, and even less when applying to tissues (Castello et

al., 2012). Second, during RNA fragmentation, the condition of RNase always needs to be optimized due to sample differences. Third, the availability and affinity of the antibodies are pivotal for RNP pull-down. What's more, the efficiency of ligation of adaptors to RNA is very low. Finally, the ability of the reverse transcriptase to process the UV crosslinked region is not very efficient. Recently, one group generated Halo tagged Ago2, termed Halo-enhanced Ago2 pull-down (HEAP), in which HaloTag can be efficiently pulled down by HaloTag ligand-bound beads (Li et al., 2020). This strategy could bypass the problems with antibodies, and this mouse line will certainly facilitate the target identification among different tissues under certain physiological conditions. However, the Halo tagged Ago2 appears to have a reduced activity, which may miss some targets during the pull-down.

2. Indirect *in vivo* evidence (Overexpression/KO of miRNA followed by RNA-seq/proteomics)

As a complementary strategy, several studies involved overexpression/KO of miRNAs followed by RNA-seq/proteomics (Yuan et al., 2019). By combining the dysregulated genes with the computational predicted targets, massively parallel "potential" targets can be identified, although further validation is needed (e.g. qPCR/Western blot). The repression/activation in the RNA level is rare, and the majority of the targets are reflected in protein level (Bartel, 2004; Wang et al., 2020b). With the advances of CRISPR Cas9/Cas12a genome editing systems

(Wang et al., 2018; Wang et al., 2020a; Wang et al., 2020b) and proteomics, the KO mice can be obtained, and the dysregulated proteins can be identified.

3. Direct *in vitro* evidence (Luciferase assay)

The readout of a specific gene can be difficult; therefore, the usage of a reporter is quite useful in quantifying the direct relationship between miRNA and targets. Since the majority of miRNA binding sites are in the 3'UTR of target genes, the 3' UTR of target genes are usually inserted into the 3' end of luciferase (e.g. Renilla luciferase) (Yuan et al., 2019). When the inserted construct is co-transfected with the corresponding miRNA, the luciferase activity can be used as the readout to define the miRNA-target relationship.

Genome editing

Loss-of-function and gain-of-function are the major strategies to elucidate gene functions in development, physiology, and disease. Gene targeting was the first developed genome modification tool, which involves homology-directed repair (HDR, also known as homologous recombination (HR)), a cell repair mechanism, in mouse embryonic stem (ES) cells to generate knockout or knock-in mice (Capecchi, 1989; Capecchi, 2005) (Figure 1-11). Gene targeting has played an essential role in generating animal models to study gene functions (Capecchi, 1989; Capecchi, 2005), and Drs. Mario R. Capecchi, Martin J. Evans and Oliver Smithies won the 2007 Nobel prize due to the invention of this technique. However, the use of this technology is largely hindered due to its low efficiency in correctly

inserting into the genome and the time-consuming screening procedures (Capecchi, 1989; Capecchi, 2005; Gaj et al., 2013; Hsu et al., 2014; Shen et al., 2014b).

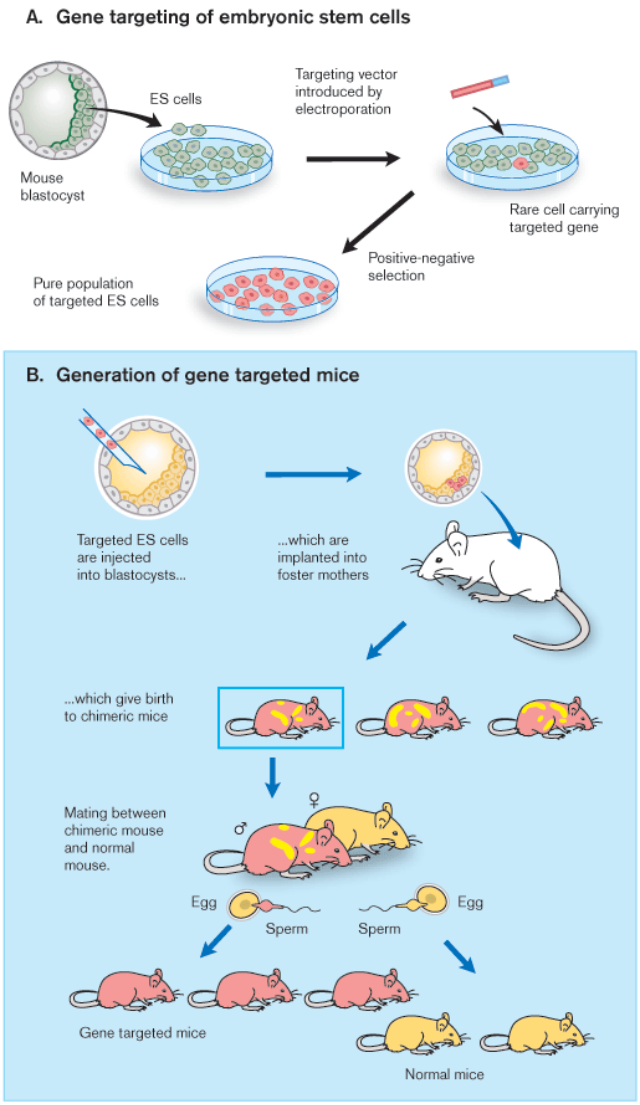

Figure 1-11. Gene targeting procedure. A. HDR selections in ES cells. B. Generation of chimeric founder mice.

(https://www.nobelprize.org/prizes/medicine/2007/advanced-information/)

Pioneer work has shown that the efficiency of HDR could be highly improved when double-stranded breaks (DSBs) were introduced in the yeast and mammalian genomes (Rouet et al., 1994; Rudin et al., 1989). Since then, several endonucleases have been harnessed to efficiently edit genes via introducing DSBs,

which can be further repaired in cells either by HDR or nonhomologous end joining (NHEJ) (Lin et al., 2014a; Weterings and Chen, 2008; Wyman and Kanaar, 2006). HDR can precisely modify a genome locus when a homologous donor template is provided, whereas NHEJ forms insertions or deletions (indels) at the breakpoint. These endonucleases include homing endonucleases (HEs, also called meganucleases) (Rouet et al., 1994; Smith et al., 2006), zinc finger (ZF) nucleases(ZFN) (Durai et al., 2005; Gaj et al., 2013), transcription activator-like effector (TALE) nucleases (TALEN) (Bogdanove and Voytas, 2011; Gaj et al., 2013), and the most recent CRISPR-Cas (clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins (Cas)) (Figure 1-12) (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013b).

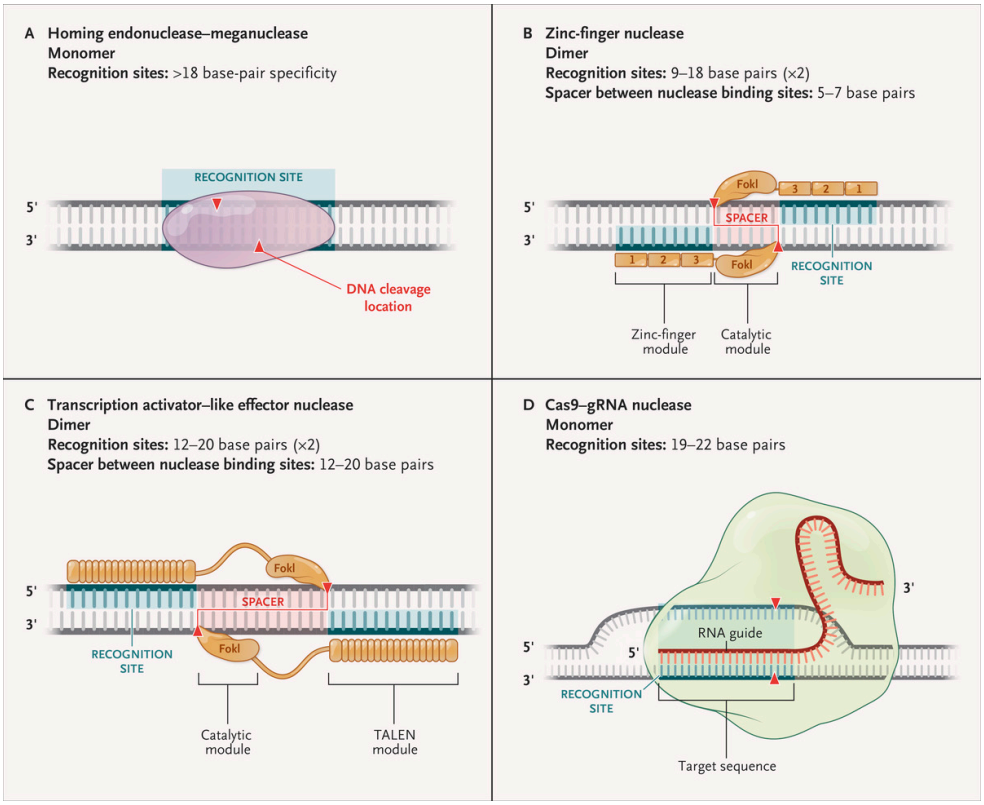

Figure 1-12. Gene editing strategies. Adapted from a review (Porteus, 2019).

HEs, ZFNs, and TALENs have their own limitations compared to the CRISPR-Cas system due to their intrinsic features. HEs, ZFNs and TALENs all recognize target DNA through protein-DNA interactions, and each DNA binding domain of HEs, ZFNs, and TALENs recognizes 20~30, 3 and 1 bp(s) of target DNA, respectively (Durai et al., 2005; Hsu et al., 2014; Joung and Sander, 2013; Stoddard, 2011). By contrast, CRISPR-Cas systems navigate their target DNA via the Watson-Crick base pairing between the guide RNA (gRNA) and target DNA (Doudna and Charpentier, 2014; Gasiunas et al., 2012; Hsu et al., 2014; Jinek et al., 2012). Although HEs harbor their own DNA-binding and nuclease domains, and have the strongest specificity (Hendel et al., 2015b), their corresponding residues to the target DNA are still unclear (Hsu et al., 2014; Sander and Joung, 2014). ZFNs and TALENs are recombinant proteins consisting of a DNA recognition domain (ZF or TALE) and a FokI endonuclease domain. ZFNs used to be the most developed system and the first approved genome editing tool in clinical trials (Hendel et al., 2015b; Laufer and Singh, 2015) and have immunogenicity advantages (Kungulovski and Jeltsch, 2015), but the limited resource of ZFs, context-dependent manner of specificity of ZFNs in an array and difficulty in designing and verifying proteins constrain the utilization of ZFNs (Bogdanove and Voytas, 2011; Doudna and Charpentier, 2014; Hsu et al., 2014). TALENs harbor versatile residues that recognize distinct base pairs (Bogdanove and Voytas, 2011; Joung and Sander, 2013); however, their applications were hindered by overall cumulative mismatches, the positional bias of repeat-variable di-residue

nucleotides (RVDs) associations within an array of TALENs, and difficulty in design and verification (Bogdanove and Voytas, 2011; Doudna and Charpentier, 2014; Juillerat et al., 2014; Sander and Joung, 2014). The simple design and verification, the capacity of multiplex targeting, and diverse available CRISPR-Cas systems have rendered the CRISPR-Cas system a feasible and efficient technology in editing genomes accurately (Cong et al., 2013; Doudna and Charpentier, 2014; Jinek et al., 2012).

The CRISPR-Cas systems are adaptive immunity systems that protect archaea and bacteria from the invasion of viruses or plasmids (Makarova et al., 2011). Based on the latest version of the classification, these systems have been classified into 2 classes (class1 and class 2), 6 types (type I, II, III, IV, V and VI), and 33 subtypes (Makarova et al., 2020). Among them, CRISPR-Cas9 and Cas12a (also known as Cpf1), a class 2 type II and a type V endonuclease, respectively, have emerged as the most commonly used genome editing tools because their RNA-guided DNA recognition and cleavage only require a single protein rather than a bulky complex of Cas proteins (Barrangou and Marraffini, 2014; Doudna and Charpentier, 2014; Jinek et al., 2012). So many genomic modified animal models have been generated since the first editing in mammalian cells (Cong et al., 2013; Mali et al., 2013b). These animals include fruit flies, zebrafish, nematodes, silkworms, salamanders, frogs, mice, rats, rabbits, pigs (for recent reviews, see (Doudna and Charpentier, 2014; Hsu et al., 2014; Ishii, 2015;

Sander and Joung, 2014)), goats (Ni et al., 2014), monkeys (Chen et al., 2015; Niu et al., 2014; Wan et al., 2015) and even human zygotes (Liang et al., 2015).

Following transcription of precursor CRISPR RNA (pre-crRNA) and trans-activating crRNA (tracrRNA), tracrRNA hybridizes with pre-crRNA to direct CRISPR RNA (crRNA) maturation in the presence of endogenous RNase III and Cas9 (Deltcheva et al., 2011). The mature tracrRNA: crRNA dual-RNA remains bounded to the Cas9 protein, forming the Cas9-tracrRNA-crRNA complex. With the guidance of the dual-RNA (Jinek et al., 2012), the protospacer adjacent motif (PAM) interacting domain of the Cas9 endonuclease recognize and associate with PAM within the genome, then the Cas9 initiates DNA separation (Nishimasu et al., 2014; Sternberg et al., 2014). After R-loop formation, crRNA couples to one strand of the target DNA that is complementary to the crRNA, forming an RNA-DNA duplex (Nishimasu et al., 2014; Sternberg et al., 2014). The Cas9 HNH nuclease domain nicks the DNA strand of the duplex (target DNA strand), while the Cas9 RuvC domain cleaves the non-complementary DNA strand (non-target DNA strand); therefore, a double-stranded break (DSB) forms (Jinek et al., 2012). The DSB is further repaired in cells either by NHEJ or HDR. The dual-tracrRNA: crRNA has been engineered as a single guide RNA (sgRNA) chimera to perform similar functions (Jinek et al., 2012).

Structure-based mechanisms of CRISPR-Cas9 in targeting DNA

Based on recent structural analysis of SpCas9, Jiang and Doudna proposed a model to explain the mechanisms of Cas9 in target DNA recognition and cleavage (Figure 1-13) (Jiang and Doudna, 2017). The Cas9 enzyme is converted from an inactive state into a DNA recognition conformation upon binding to sgRNA. The PI domain of Cas9 is prepositioned for PAM interrogation, and the seed sequence of sgRNA is preordered for DNA targeting and invasion. Cas9 unwinds DNA when it finds a potential target with a proper PAM. The phosphate lock loop in the CTD domain stabilizes the target DNA strand and facilitates gRNA-DNA duplex formation. The NUC lobe of Cas9 embeds the non-target DNA to stabilize the unwinding structure. The first two PAM-proximal seed nucleotides are exposed to the solvent to initiate the R-loop formation. The HNH domain undergoes conformation change from the inactivation state to the activation state; meanwhile, two linkers between the HNH and RuvC domains also involve conformation change to direct the nontarget strand close to the RuvC catalytic center. Cas9 remains bound to the cleaved target DNA after cleavage until it is displaced by other cellular factors (Jiang and Doudna, 2017).

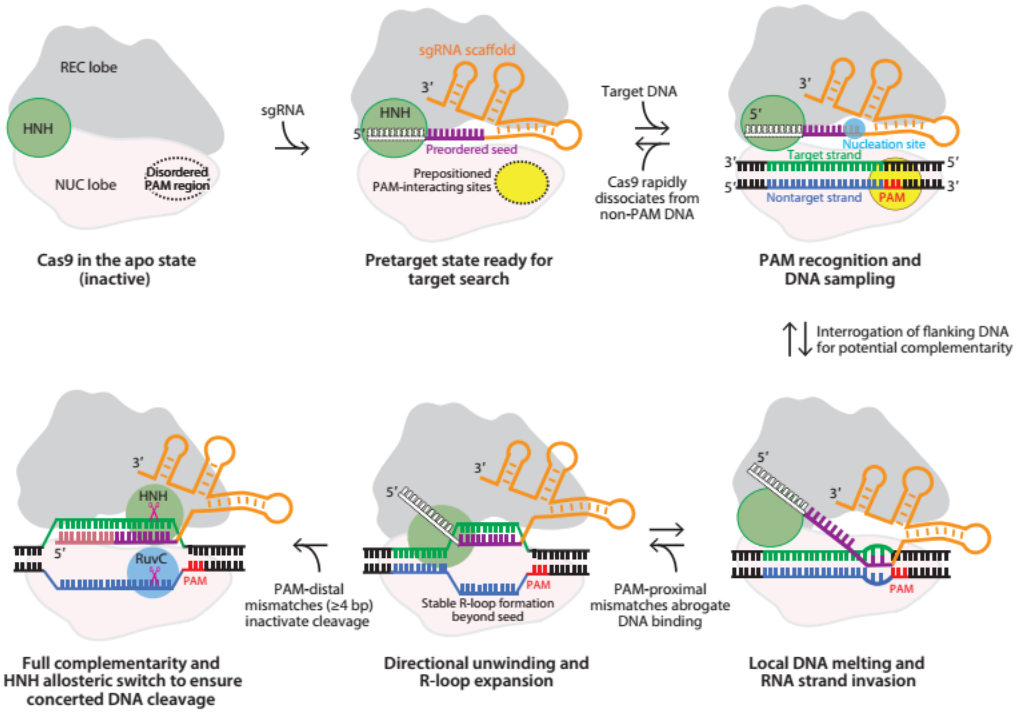

Figure 1-13. Structure-based mechanisms of CRISPR-Cas9 in target DNA recognition and cleavage (Jiang and Doudna, 2017).

Applications of CRISPR-Cas9/Cas12a

1. Study of gene function

There are two major types of CRISPR-Cas9/Cas12a systems in evaluating gene function: the commonly used gene knockout (KO)/knock-in (KI) (Cong et al., 2013; Jinek et al., 2012; Wang et al., 2018; Wang et al., 2019b; Wang et al., 2020a; Wang et al., 2020b) and the transcriptional repression/activation of genes (Cheng et al., 2013; Gilbert et al., 2013; Konermann et al., 2015; Maeder et al., 2013b; Mali et al., 2013a; Perez-Pinera et al., 2013). The former strategy involves the

cutting activity of Cas9, whereas the latter one fused the catalytically inactive Cas9 (dCas9) to a repressor or an activator domain (termed as CRISPR interference/inhibition (CRISPRi) or CRISPR activation (CRISPRa)).

Several studies have employed global knockout, including deletion of a part / the entire gene or insertion of fragments to disturb interested genes, to study gene functions in humans, mice, rat, sheep, zebrafish, silkworm and so on. However, ablation of some genes may be lethal to animals. To tackle this issue, conditional KO could be utilized in analyzing gene functions in specific tissues, for examples, Cre-Loxp conditional knockout (Gu et al., 1994), tissue-specific mutagenesis (including nerve (Shen et al., 2014b) and liver (Yin et al., 2015)), as well as inducing systems. Conditional alleles (Loxp) have been generated by CRISPR-Cas9 system in mice (Yang et al., 2013a), rats (Ma et al., 2014), nematode (Shen et al., 2014b), zebrafish (Yin et al., 2015), fruit fly (Gratz et al., 2014; Xue et al., 2014). What's more, inducing systems have been achieved, including chemically control of CRISPR-Cas9 (Dow et al., 2015; Gonzalez et al., 2014; Zetsche et al., 2015b), heat-shock inducible expression (Shen et al., 2014b; Yin et al., 2015) and optical control of CRISPR-Cas9 (Hemphill et al., 2015; Nihongaki et al., 2015).

Alternatively, dCas9 can be directed to promoters with the guide of gRNAs to inhibit or active gene of interest without changing the DNA sequence. Two main strategies were used: 1. Fusing dCas9 with transcriptional repressor domain KRAB or activation domain VP64 (or other variant tandem copies of VP16 motifs) (Cheng et al., 2013; Gilbert et al., 2013; Hu et al., 2014; Maeder et al., 2013b; Mali et al., 2013a; Perez-Pinera et al., 2013); 2. Recruiting the activator to dCas9 binding site

by gRNA (Konermann et al., 2015). The latter strategy, e.g. synergistic activation mediator (SAM), is based on that the REC1, the BH, and the PI domains of Cas9 interact directly with repeat: anti-repeat duplex, stem-loop 1 and linker region of sgRNA, while not with stem-loop 2 and stem-loop 3 (Nishimasu et al., 2014) (Figure 1-14). Mutation experiments showed that mutation in the linker, stem-loop 2 and 3 are more tolerant than repeat: anti-repeat duplex and stem-loop 1, which indicate that the repeat: anti-repeat duplex and stem-loop 1 are indispensable for Cas9 function, while the linker, stem-loop 2 and 3 are more likely to stabilize sgRNA as well as Cas9-gRNA-DNA complex (Nishimasu et al., 2014). In the SAM system, a minimal hairpin aptamer MS2, which could specifically recruit MS2 bacteriophage coat proteins (MSP) that tethered with VP64, or p65 and HSF1, is added to tetraloop or stem-loop 2 of gRNA. After the expression of the gRNA, dCas9 (D10A and H840A double mutation), and MSP-SAM, more active transcription can be observed with no effect on Cas9 binding (Konermann et al., 2015).

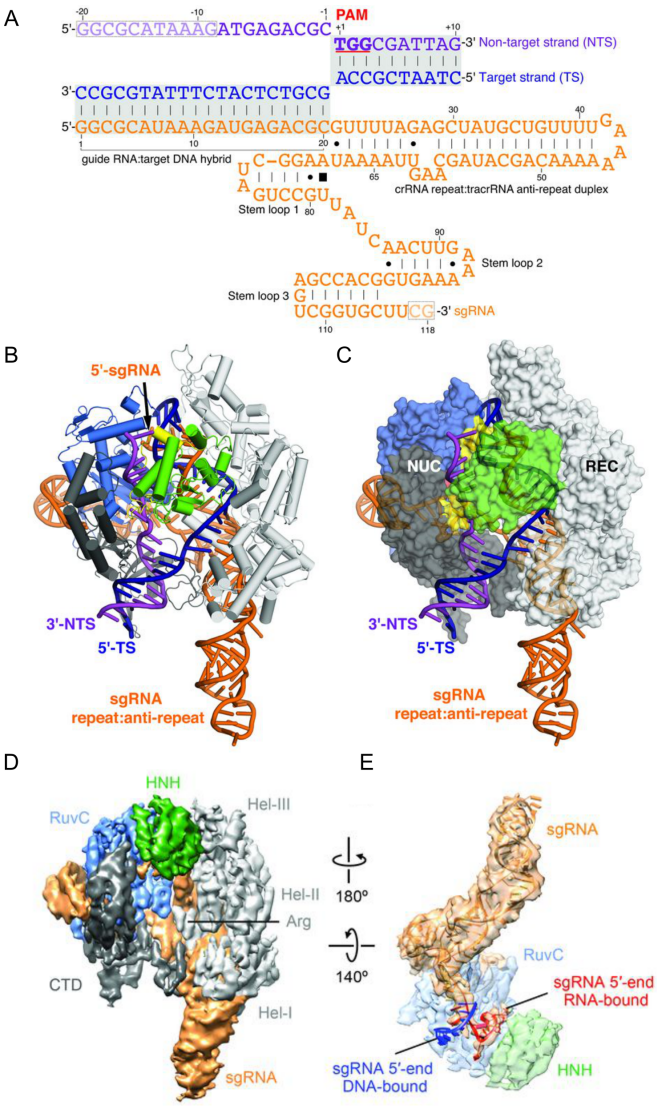

Figure 1-14. Quaternary structure of SpCas9 upon binding to sgRNA. A. Structure of gRNA. B and C (Jiang et al., 2015). Structure of SpCas9-gRNA determined by cryo-EM (Jiang et al., 2015). D and E. Structure of SpCas9-gRNA determined by X-ray crystallography (Jiang et al., 2016).

2. Correction of gene mutations

Corrections of gene mutations have been made in many cell types and species. In cell models, correction of deletion of phenylalanine in the CFTR gene that led to cystic fibrosis was performed in cultured primary intestinal stem cells from cystic fibrosis patients (Schwank et al., 2013). Combining CRISPR-Cas9 with the piggyBac transposon technology, iPSCs generated from fibroblasts of a β-

thalassemia patient with a -28A/G point mutation in the promoter and a 4-bp (TCTT) deletion in the exon 2 of human hemoglobin beta (*HBB*) gene were seamlessly corrected (Xie et al., 2014). Correction of Duchenne muscular dystrophy (DMD) patient-derived iPSCs had been harnessed by exon skipping, frameshifting, and conditional exon knock-in (Li et al., 2015). These corrected iPSCs retain pluripotency, differentiation potential, and normal karyotypes with restored expressions of *HBB* or *Dmd* (Li et al., 2015; Xie et al., 2014). Myoblasts from DMD patients with a deletion of exons 48-50 were also corrected by deleting exon 51 or exons 45-55 (Ousterout et al., 2015). When transplanted corrected myoblasts into mice, cryosections of injected tissue shows a restored expression of *Dmd* protein (Ousterout et al., 2015).

In addition, corrections of mutant animals such as mice and rats have been achieved. Mutant mice, with a dominant 1 bp deletion in the *Crygc* gene and a nonsense mutation in the *Dmd* gene that are responsible for cataract disorder and DMD, respectively, were corrected by zygotic injection of Cas9 and sgRNA with or without donor template (Long et al., 2014; Wu et al., 2013). This approach plays an indispensable role in correcting the genetic defects in the offspring, although the patients remain to be the carriers. Spermatogonia stem cells (SSCs) lines established from homozygous cataract males carrying a mutation in Crygc could be easily corrected by CRISPR/Cas9-induced NHEJ or HDR (Wu et al., 2015). SSC could be isolated and expanded in vitro for a long run with unchanged fertility, allowing to make the correction in vitro by using CRISPR-Cas9. When transplanted the engineered SSCs back to the testes after verification and cloning, these cell

lines could undergo spermatogenesis and generate round spermatids after transplantation into infertile mouse testes. Fertilization of the round spermatid generated from preselected corrected SSCs could give birth to offspring with the corrected phenotype (Wu et al., 2015). However, no good strategies are available for females due to the lack of methods to identify oogonial stem cells and culture *in vitro*.

3. Epigenome editing using CRISPR-Cas9

The epigenome is defined as all chemical modifications in DNA and histone proteins (Kungulovski and Jeltsch, 2015). Therefore, the fusion of DNA binding domain (ZFs, TALEs, and dCas9) to the various chromatin-modifying enzyme is termed the epigenome editing (Kungulovski and Jeltsch, 2015; Laufer and Singh, 2015). DNA methylation is mainly induced by DNA methyltransferases (Dnmt1, Dnmt3a, and Dnmt3b) (Reik et al., 2001), and demethylation by Ten-Eleven Translocation family (TET1, TET2, and TET3) and Thymidine DNA glycosylase (TDG) (Kohli and Zhang, 2013). Histone chemical modifications including histone acetylation, deacetylation, methylation, and demethylation are mediated by histone acetyltransferases, deacetylases, lysine methyltransferases, and demethylases, respectively (Kungulovski and Jeltsch, 2015).

In mammalian cells, through tethering ZF or TALE with the catalytic domain of DNMT3a, site-specific methylations are induced in mitochondria DNA (Minczuk et al., 2006) and nuclear DNA (Bernstein et al., 2015; Kungulovski et al., 2015; Nunna et al., 2014; Rivenbark et al., 2012; Siddique et al., 2013; Stolzenburg et al., 2015);

When ZF or TALE was fused with combined DNMT3a-DNMT3L, in which DNMT3L serves as an activator for DNMT3a (Chedin et al., 2002; Hata et al., 2002; Jia et al., 2007), a higher methylation level was harnessed (Bernstein et al., 2015; Siddique et al., 2013). Targeted DNA demethylation has also been achieved by fusing ZF or TALE to TET family or TDG (Chen et al., 2014; Gregory et al., 2013; Maeder et al., 2013a). The DNA demethylations of promoters of Nos2, ICAM-1, EpCAM, KLF4, RHOXF2, and HBB have been harnessed, and accompanied by upregulated expression (Chen et al., 2014; Gregory et al., 2013; Maeder et al., 2013a).

As for histone modification, repression of promoters in VEGF-A, E-cadherin, HER2/neu, Fosb, stem cell leukemia (SCL) locus, Oct4, Tbx3, ZFP42, KLF5, SALL3, KLF2, FOXP4, IL1RN, and MYOD have been fulfilled by H3K9 methylation (Cho et al., 2015; Falahi et al., 2013; Heller et al., 2014; Kungulovski et al., 2015; Snowden et al., 2002) and H3K4 demethylation (Kearns et al., 2015; Mendenhall et al., 2013). On the other hand, activating promoters and enhancers of IL1RN, MYOD, and OCT4 can be obtained by H3K27 acetylation (Hilton et al., 2015). CRISPR mediated histone modifications and DNA methylation were also developed (Hilton et al., 2015; Kearns et al., 2015; Liu et al., 2016; McDonald et al., 2016; Vojta et al., 2016). However, whether this epigenome editing status, either in DNA or histone, could be maintained is still controversial (Kungulovski et al., 2015; Rivenbark et al., 2012; Stolzenburg et al., 2015), which may result from variant delivery systems and mechanism between different vectors as well as

diverse chromosome context between target genes (Kungulovski and Jeltsch, 2015).

4. Lineage tracing

Cell lineage refers to the path of development of one-cell zygote to multicellular organisms. The traditional method for identifying cell lineage usually involves tagging known cell markers, and determine the tag throughout the development (Kester and van Oudenaarden, 2018). However, the known cell markers are limited, and hence, this method hinders tracing lineage precisely. Since the CRISPR genome editing preferentially generates indels once cut the genome, leaving scars that can be used for lineage tracing. Several studies have applied Cas9 genome editing for lineage tracing in zebrafish (McKenna et al., 2016; Spanjaard et al., 2018) and mice (Kalhor et al., 2018). In these studies, Cas9 and gRNAs were microinjected one-cell zygotes, the scars were then analyzed using single-cell RNA-seq at the latter stages.

5. Genomic loci imaging

Mutations in D10 and H840 render Cas9 a dead Cas9, which has no cleaving activity, while still maintain the ability to bind DNA. When co-transfecting gRNA and dCas9 tethered with eGFP into cells, genomic loci can be imaged (Chen et al., 2013). A dCas9-eGFP knock-in mice strain was further developed to enable in vivo genomic loci imaging (Duan et al., 2018). A much higher signal/noise ratio can be obtained when transfecting cells with dCas9 and gRNAs labeled with fluorophores,

such as Cy3 or Atto565 (Wang et al., 2019a).

6. Nucleic acids detection

Both Cas12a and Cas13 have collateral activity in vitro, which releases the universal ssDNA and RNA cleavage activity, respectively, after cutting the targets (Chen et al., 2018; Gootenberg et al., 2017). Based on this feature, Specific High-Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK) and DNA Endonuclease Targeted CRISPR Trans Reporter (DETECTR) were developed for detecting nucleic acids. These methods greatly facilitate pathogens nucleic acids detection, such as Zika, Dengue virus and current pandemic SARS-CoV-2 (Ackerman et al., 2020; Broughton et al., 2020; Chen et al., 2018; Gootenberg et al., 2017).

Major obstacles in utilizing CRISPR-Cas9/Cas12a technology

There are several main barriers in utilizing the CRISPR-Cas9/Cas12a engineering tool: 1) mosaicism; 2) off-target effects; 3) Large on target deletion/insertion; 4) Low efficiency in HDR-mediated insertion; 5) PAM sequence restriction; 6) Immune response; 7) Delivery.

1. The mosaicism

One major obstacle in utilizing CRISPR-Cas9 is the genetic mosaicism (Long et al., 2014; Oliver et al., 2015; Yang et al., 2013a) that may result from the delayed

translation of Cas9 and/or insufficient nucleases in cutting both copies of target DNA (Guo and Li, 2015; Hsu et al., 2014; Lanphier et al., 2015), variant DNA repair activities in zygotes and divided embryos (Guo and Li, 2015; Hsu et al., 2014). Mammal zygotes need several days to specify into germline may also account for mosaicism (Chapman et al., 2015). However, direct injection of Cas9 protein and sgRNA also results in mosaic mutants (Sung et al., 2014), suggesting that the delayed translation may not be the major force for the mosaicism. Therefore, with current technologies, mating the founders with wild type animals is needed to generate pure mutant progeny.

2. The off-target effects

Although the CRISPR-Cas9 system has been used feasibly and efficiently in genome engineering, mismatches (termed “off-target”) between sgRNA and target DNA can be tolerated (Fu et al., 2013; Hsu et al., 2013; Lin et al., 2014b). This phenomenon also can be found in the latest developed Cas12a/Cpf1 (CRISPR from Prevotella and Francisella 1) system (Wang et al., 2020a; Zetsche et al., 2015a). Hsu et al. demonstrated that the CRISPR-Cas9 system has off-target effects, and proposed to reduce this effect by limiting the SpCas9-sgRNA concentration, however, on-target cleavage using this approach also reduced (Hsu et al., 2013). Several strategies have been developed to reduce off-target effects by improving the specificity and efficiency of Cas9, optimizing sgRNA, and identifying potential off-target sites.

2.1 Optimizing Cas9/Cas12a proteins

Three major strategies were used to optimize Cas9/Cas12a proteins: fusion of Cas9 protein with other proteins or domains, Cas9/Cas12a orthologues, and Cas9 variants. The first strategy includes dCas9-FokI nuclease (Guilinger et al., 2014; Tsai et al., 2014), Cas9-pDBD chimera (Bolukbasi et al., 2015), base editing (Gaudelli et al., 2017; Komor et al., 2016) and prime editing (Anzalone et al., 2019). Orthologues include SaCas9 (Staphylococcus aureus Cas9) (Ran et al., 2015) and Cas12a/Cpf1 (Zetsche et al., 2015a). Cas9 variants include Cas9 nickase (Mali et al., 2013b; Ran et al., 2013; Shen et al., 2014a), “enhanced specificity” SpCas9 (eSpCas9) (Slaymaker et al., 2015), SpCas9-HF1 (Kleinstiver et al., 2016a), HypaCas9(Chen et al., 2017), HiFiCas9 (Vakulskas et al., 2018), evoCas9 (Casini et al., 2018), and Sniper-Cas9 (Lee et al., 2018).

In the Cas9 fusion proteins, targeted double-stranded breaks can be introduced by paired guide RNAs together with dCas9-FokI nuclease (Guilinger et al., 2014; Tsai et al., 2014). These approaches significantly minimized off-target mutagenesis (Guilinger et al., 2014; Mali et al., 2013b; Ran et al., 2013; Shen et al., 2014a; Tsai et al., 2014), however, one concern is that one additional gRNA might induce further off-target mutations (Fu et al., 2014; Sander and Joung, 2014). Precision and targeting range can be drastically enhanced by fusing a programmable DNA-binding domain (pDBD) to Cas9 and reducing the inherent DNA-binding affinity of Cas9 (Bolukbasi et al., 2015). The base editing fused Cas9 nickase (D10A) to deaminase (either cytidine or deoxyadenosine) (Gaudelli et al., 2017; Komor et al.,

2016), whereas the prime editing fused Cas9 nickase (H840A) to MMLV reverse transcriptase (Anzalone et al., 2019).

In the Cas9 orthologues, a smaller Cas9 orthologue, SaCas9, has been developed to efficiently and specifically edit the Pcsk9 gene in mouse liver (Ran et al., 2015). Cas12a itself displays reduced off-target effects compared to SpCas9 due to its irreversible binding to the target region and strong discrimination against the off-target sequences (Kim et al., 2017; Kleinstiver et al., 2016b; Strohkendl et al., 2018; Zetsche et al., 2015a).

In the Cas9 variants, paired guide RNAs together with Cas9 nickase (Mali et al., 2013b; Ran et al., 2013; Shen et al., 2014a) have lower off-target effects. eSpCas9, SpCas9-HF1 (Kleinstiver et al., 2016a), HypaCas9, and HiFiCas9 are structure-based rational design, while evoCas9 and Sniper-Cas9 are designed through direct evolution.

eSpCas9 could significantly reduce the off-target effect without minimizing the on-target efficiency through neutralizing positive charges in the non-target strand groove (nt-groove), which localized in Cas9 HNH, RuvC, and PAM-interacting domains (Slaymaker et al., 2015). Similarly, Cas9-HF1 reduced the direct contact between Cas9 and target DNA (Kleinstiver et al., 2016a). Further evidence showed that both eSpCas9 and Cas9-HF1 have difficulties in changing to active endonuclease in the mismatches (Chen et al., 2017). Based on that REC3 domain recognizes target and modulates the HNH nuclease domain conformation changes, the HypaCas9 was developed (Chen et al., 2017). The Zhang group systematically

analyzed the Cas9 variants, and found that there are tread offs between specificity and efficiency (Schmid-Burgk et al., 2020) (Figure 1-15).

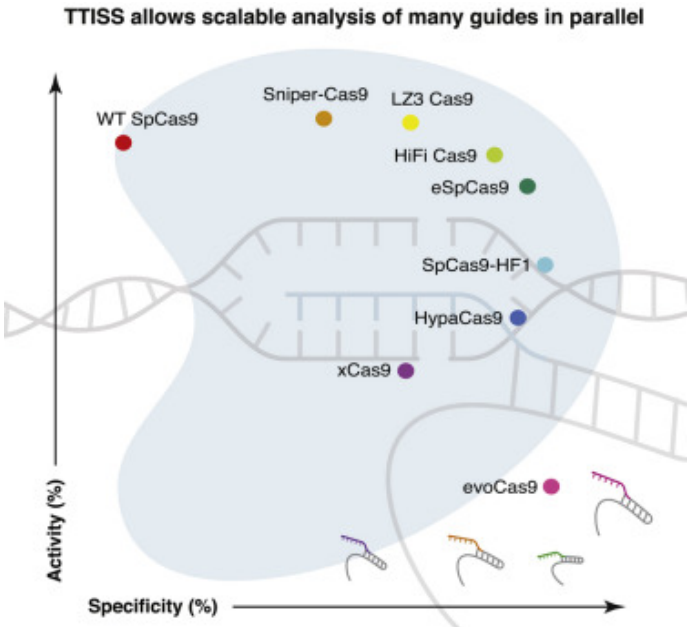

Figure 1-15. Comparisons of specificity and efficiency among different Cas9 variants (Schmid-Burgk et al., 2020)

2.2 Optimizing gRNA

A vast majority of investigators focus on mitigating the off-target effect by optimizing gRNAs design. These approaches involve gRNAs design using online tools with new algorithm (Doench et al., 2014; Hsu et al., 2013; Xu et al., 2015) and modified gRNAs (Fu et al., 2014; Hendel et al., 2015a; Sander and Joung, 2014).

As for the newest algorithm for gRNAs design, Genome Engineering 3.0 (Hsu et al., 2013), ZiFiT Targeter (Zinc Finger Targeter) (Fu et al., 2014), E-CRISP (Heigwer et al., 2014), CHOPCHOP (Montague et al., 2014), sgRNA Designer (Doench et al., 2014), SSC (Spacer Scoring for CRISPR) (Xu et al., 2015) and CRISPR-ERA(CRISPR-mediated editing, repression, and activation) (Liu et al.,

2015) are available for genome engineering. The websites and properties of these in silico design tools are list as Table 2.

Apart from designing gRNAs with online tools or software, several groups have utilized modified gRNAs with reduced off-target efficiency, including truncated gRNAs (tru-gRNAs) (Fu et al., 2014), chemically modified gRNAs (Hendel et al., 2015a), and hairpin gRNA (Kocak et al., 2019). Tru-gRNAs, shortened gRNAs with 17 or 18 bps, have been demonstrated could reduce the off-target cleavage of both individual Cas9 nucleases and paired Cas9 nickases (Fu et al., 2014). A combination of FokI-dCas9 and tru-gRNAs has also achieved high genome editing specificity (Wyvekens et al., 2015). Chemically modified gRNAs could significantly improve genome editing efficiency with enhanced on-target: off-target ratios in human cell lines by co-delivering the modified sgRNAs with Cas9 mRNA or protein (Hendel et al., 2015a). Although high efficiency in inducing targeted indels also observed in human primary T cells and CD34+ hematopoietic stem and progenitor cells (HSPCs), the specificity of chemically modified gRNAs in these cells remains unclear (Hendel et al., 2015a). Hairpin gRNAs require higher energy for opening up the secondary structure of gRNA so that the mismatches can be much more reduced (Kocak et al., 2019).

Table 2. Websites and properties of several in silico design tools.

Name	Website	Properties	Reference
Genome Engineering 3.0	http://www.genome-engineering.org/	A scoring algorithm that concerning the effect of mismatch location, density and identity on SpCas9 cleavage	(Hsu et al., 2013)
ZiFiT Targeter	http://zifit.partners.org/	Identify shorter gRNA sites with 17 or 18 base pairs of target complementarity that could minimize off-target effects	(Fu et al., 2014)
E-CRISP	http://www.e-crisp.org/	Design insertion of tags into N terminal or C terminal of genes interested or gene knockout, and annotations of targets	(Heigwer et al., 2014)
CHOPCHOP	https://chopchop.rc.fas.harvard.edu	Allows PAM variations, generates automatic primer restriction site	(Montague et al., 2014)

		for genotyping and provides annotations of the gene	
sgRNA Designer	http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design	Be designed based upon features of experimentally verified high active sgRNA sequence (eg. nucleotide preferences) to enhance on-target efficacy	(Doench et al., 2014)
SSC	http://sourceforge.net/projects/spacerscoringcrispr/ (software version) or http://crispr.dfci.harvard.edu/SSC (online version)	Integrally analyzed the contribution of sequence context to sgRNA efficiency based on 6 published datasets and developed Spacer Scoring for CRISPR that could design gRNAs either for Cas9-mediated knockout or	(Xu et al., 2015)

		for dCas9 fusion-mediated inhibition (CRISPRi) or activation (CRISPRa)	
CRISPR-ERA	http://CRISPR-ERA.stanford.edu	Determine efficiency and specificity of sgRNAs using rules summarized from published data for gene editing and regulation (ie. CRISPRi/a)	(Liu et al., 2015)

2.3. Off-target prediction

In addition, a series of approaches have been developed to identify potential off-target sites, such as in silico prediction tools, dCas9-ChIP, and experimentally identification technologies (Ran et al., 2015). In silico prediction tools includes RGEN Tools (RNA-guided engineered nucleases, http://www.rgenome.net/) (Bae et al., 2014), CasOT (CRISPR/Cas system (Cas9/gRNA) Off-Targeter, http://eendb.zfgenetics.org/casot/) (Xiao et al., 2014), COSMID (CRISPR Off-target Sites with Mismatches, Insertions, and Deletions, http://crispr.bme.gatech.edu) (Cradick et al., 2014), which significantly facilitate the improvement of specificity. Although online tools and dCas9-ChIP could predict off-target sites precisely, the nuclease activity of Cas9 cannot be analyzed. Thus,

experimentally identification technologies can be used, these technologies include BLESS (direct in situ breaks labeling, enrichment on streptavidin and next-generation sequencing) (Ran et al., 2015), HTGTS (high-throughput, genome-wide, translocation sequencing) (Frock et al., 2015), Digenome-seq (Kim et al., 2015), GUIDE-seq (genome-wide, unbiased identification of DSBs enabled by sequencing) (Tsai et al., 2015), IDLVs (integrase-defective lentiviral vectors) (Wang et al., 2015) and FISH-based method to detect CRISPR-Cas9 off-targets (Paulis et al., 2015). With the use of these technologies that could predict and verify the off-target effect, further refinements of sgRNAs design and reduction of off-target effects will be achieved.

3. Large on target deletion/insertion

Cas9 has been shown to cause on-target large deletions/insertions (Adikusuma et al., 2018; Kosicki et al., 2018; Lee and Kim, 2018). The incidences of large deletions induced by Cas9 with one single gRNA were 35.7%, 36.5%, and 45% in mouse ES cells, progenitor cells, and zygotes, respectively, whereas the incidence of large insertions was 26.3% in mouse ES cells (Adikusuma et al., 2018; Kosicki et al., 2018). By contrast, Cas12a-based genome editing predominantly generates indels, which may because that the MMEJ repair mechanism is preferentially adopted in fixing the staggered DNA ends that Cas12a generated (Wang et al., 2020a).

4. Low efficiency in HDR-mediated insertion

DSBs can be precisely repaired by HDR when a donor template is provided, however, the efficiency of this mechanism is extremely low (Chu et al., 2015; Maruyama et al., 2015). This phenomenon may be due to that HDR occurs only during S and G2 phase, while NHEJ can happen throughout the whole cell cycle (Lin et al., 2014a; Maruyama et al., 2015). Indeed, arresting cell cycle at G2/M through the addition of nocodazole enhanced HDR in HEK293T cells, human primary neonatal fibroblast, and human embryonic stem cells (Lin et al., 2014a). Several studies have focused on enhancing the efficiency of HDR, including using Cas9 variants, additional molecules, and optimizing design parameters for gRNAs and donor template.

4.1. Cas9/Cas12a variants

Using a double nicking strategy, a higher HDR efficiency can be achieved in human embryonic stem cells (hESCs) and mouse zygotes (Ran et al., 2013). Deliver Cas9/Cas12a tethered with streptavidin, gRNA, and biotinylated donor DNA into cells highly improved the HDR efficiency in multiple cell lines and mouse zygotes (Carlson-Stevermer et al., 2017; Gu et al., 2018; Savic et al., 2018; Wang et al., 2020a). When Cas9 protein modified with an azide-containing noncanonical amino acid (ncAA) that can bind to the donor DNA containing dibenzylcyclooctyne (DBCO) was delivered to human cell lines and mouse zygotes, high HDR efficiency can be obtained (Ling et al., 2020). Fusing Cas9 nickase to deaminase (either cytidine deaminase or deoxyadenosine deaminase, termed base editing) (Gaudelli et al., 2017; Komor et al., 2016) or MMLV reverse transcriptase (prime editing)

(Anzalone et al., 2019) could also induce high efficiency of precise KI by bypassing HDR.

4.2. Small molecules treatment

High efficiency of precise genome editing by HDR has been achieved by treating cells with small molecules. HDR can be improved in mammalian cells or mouse zygotes when suppressing NHEJ by the administration of the ligase IV inhibitor SCR7, by knocking down KU70, KU80 or DNA ligase IV, or by co-expressing adenovirus 4 (Ad4) E1B55K or E4orf6 proteins (Chu et al., 2015; Maruyama et al., 2015). RS-1 is an HDR stimulator by activating RAD51 binding (Jayathilaka et al., 2008) in rabbit embryos (Song et al., 2016). i53 represses 53BP1 to favor HDR over NHEJ in U2OS and 293T cells (Canny et al., 2018). Administration of XL413 or PHA-767491, which can inhibit CDC7 that repress HDR, could increase the HDR efficiency for both single-stranded donor and double-stranded donor in HEK293T, U-251, Hela, iPSCs (induced pluripotent stem cells), K562 cell line and primary human T cells (Wienert et al., 2020). Addition of L755507, an activator of β3-adrenergic receptor, or Brefeldin A, which inhibits protein transport from Endoplasmic Reticulum to Golgi apparatus, also enhanced the HDR efficiency (Yu et al., 2015). On the contrary, an anti-HIV drug AZT, could inhibit HDR efficiency and facilitate NHEJ efficiency (Yu et al., 2015).

4.3. Optimizing donor template

In addition, optimizing design parameters could also improve HDR efficiency.

In the double nicking system, HDR level could be enhanced by design sgRNA pairs that could producing 5'overhangs and having at least one nick within 22 bp of the homology arm (Ran et al., 2013).

In the wild-type Cas9 system, single-stranded oligonucleotide DNA (ssODNA) with 60-70 base pair of homology arm flanking the Cas9 cleavage site have achieved optimal HDR, either for substitutions or insertion of a short fragment, in human-induced pluripotent cells (hiPSCs), HEK293T or mouse zygotes (Gurumurthy et al., 2019; Lin et al., 2014a; Quadros et al., 2017; Yang et al., 2013b). Using asymmetric single-stranded donors have even better HDR efficiencies (Richardson et al., 2016). However, longer single-stranded donors are hard to obtain. The maximum length for commercially available megamer ssDNA (from IDT) is 2kb. Currently, three ways are efficient to generate large single-stranded donors: 1. In vitro transcription of DNA template followed by reverse transcription and RNase H digestion (Li et al., 2019); 2. asymmetric PCR (Veneziano et al., 2018); 3. PCR with a 5' phosphorylated primer and a regular primer (with a 5' hydroxyl group) followed by removal of the 5' phosphorylated strand by Lambda Exonuclease (Little, 1981). However, errors could be introduced during RT and PCR.

The length of homology arms for PCR-generated double-stranded donors are controversial, which may due to the different type of ends generated by the endonucleases (i.e. Cas9 induced blunt end and Cas12a induced stagger end). One study showed that homology arms of >1 kb have a higher HDR efficiency in HEK293 cells (Chu et al., 2015), however, one recent study using Cas12a showed

that ~30 bp homology arm is enough, and the 55bp ~ 90 bp is optimal (Fueller et al., 2020). However, linearized PCR products have the potentials of random insertions as well as concatemers of PCR products. Chemical modifications, such as 5' C6-PEG10 (Yu et al., 2020), Biotin and SpC3 (Gutierrez-Triana et al., 2018) of double-stranded donors also increase HDR efficiency and decrease concatemers of PCR products.

5. PAM sequence restriction

PAM sequence adjacent to target DNA is required to distinguish the target sequences from the gRNA itself in bacteria. For examples, the most commonly used SpCas9 requires NGG PAM sequence, whereas the other widely used AsCas12a and LbCas12a utilize TTTV PAM sequences for efficient genome editing (Cong et al., 2013; Jinek et al., 2012; Wang et al., 2013; Zetsche et al., 2015a). However, the PAM sequences restrict the application of CRISPR genome editing. Several studies have explored minimum PAM sequence Cas orthologues, both from nature and lab-based. Using phage-assisted continuous evolution (PACE), David R. Liu lab developed an expanded PAM SpCas9 variant (xCas9), which can recognize NG PAM sequence (Hu et al., 2018), and further NRRH, NRTH and NRCH PAM sequence (Miller et al., 2020). Based on structure analysis, Osamu Nureki group developed Cas9-NG, which can also recognize NG PAM sequence (Nishimasu et al., 2018), Benjamin P. Kleinstiver group generated SpG and SpRY (where R stands for A/G, and Y for C/T) (Walton et al., 2020). Besides, based on the Cas12a structure, enAsCas12a with broader PAM targeting ranges

was developed (Kleinstiver et al., 2019). Cas12a with less stringent PAM requirements from nature were developed, such as Mb3Cas12a (Wang et al., 2020a), HkCas12a (Teng et al., 2019), and so on. Until now, the PAM sequences seem not to be a limitation anymore, although some of the enzymes have lower efficiency compared to the wild-type version.

6. Immune response

Despite the promising CRISPR genome editing in clinical trials, several independent groups showed that in human bodies, antibodies against two wildly used SpCas9 and SaCas9 proteins exist (Charlesworth et al., 2019; Simhadri et al., 2018; Wagner et al., 2019), which raised concerns for utilizing these two endonucleases in clinical trials. To date, the pre-existing antibodies against Cas12a and Cas13 in human bodies have yet to be reported.

7. Delivery

Several methods have been developed for delivering CRISPR systems, including microinjection, viral vectors (e.g. AAV), nanoparticles, and electroporation of RNP complexes (Doudna, 2020) (Table 3). Each system has its own advantages and limitations. Microinjection has high efficiency, but it can’t be applied to clinical application. AAV has a high transduction efficiency, yet the gene size is limited to ~4.7kb. Nanoparticles are flexible in size, but they are probably toxic to cells. Electroporation can be used to deliver CRISPR components to zygotes, yet similar to microinjection, the in vivo application is hindered.

Table 3. Methods for CRISPR delivery (Doudna, 2020).

Property	Nanoparticles	Viruses	RNPs
Features and applications	Cationic lipid polymers can be used to encapsulate molecular cargo, facilitating cellular entry.	AAVs are the most commonly used clinical delivery vehicle for gene therapy.	Purified protein and guide RNA can be electroporated into stem cells extracted from patients to treat blood disorders such as sickle cell disease.
Size	50–500 nm	20 nm	12 nm
Payload	mRNA, DNA, RNP (from most to least commonly used)	DNA	Preformed enzyme complexes
Advantages	- Inexpensive and relatively easy to produce - No genomic integration - Low immunogenicity	- Broad tissue targeting possibilities - Clinically established method - Efficient	- No genomic integration - No long-term expression and fewer off-target effects
Disadvantages	- Limited capacity for tissue targeting	- Limited cargo size - Undesired integration risk - Sustained expression can lead to off-target effects - Immunogenicity - High cost and manufacturing challenges	- Will not enter cells without engineering or additional reagents - Potential immunogenicity in vivo - Unprotected RNPs are at risk of degradation
Targets	Liver	Liver, eyes, brain, lungs and muscle	Oocytes, stem cells and T cells

The three main delivery strategies that could be used for clinical genome-editing applications are nanoparticles, viruses and purified RNPs. The approaches vary in important ways, which generally limit their suitability for editing to specific cell or tissue types.

References

Abou-Haila, A. and Tulsiani, D. R. (2000). Mammalian sperm acrosome: formation, contents, and function. Arch Biochem Biophys 379, 173-82.

Ackerman, C. M., Myhrvold, C., Thakku, S. G., Freije, C. A., Metsky, H. C., Yang, D. K., Ye, S. H., Boehm, C. K., Kosoko-Thoroddsen, T. F., Kehe, J. et al. (2020). Massively multiplexed nucleic acid detection with Cas13. Nature 582, 277-282.

Adikusuma, F., Piltz, S., Corbett, M. A., Turvey, M., McColl, S. R., Helbig, K. J., Beard, M. R., Hughes, J., Pomerantz, R. T. and Thomas, P. Q. (2018). Large deletions induced by Cas9 cleavage. Nature 560, E8-E9.

Agarwal, V., Bell, G. W., Nam, J. W. and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife 4.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A. et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149-157.

Auyeung, V. C., Ulitsky, I., McGeary, S. E. and Bartel, D. P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell 152, 844-58.

Bae, S., Kweon, J., Kim, H. S. and Kim, J. S. (2014). Microhomology-based choice of Cas9 nuclease target sites. Nat Methods 11, 705-6.

Barrangou, R. and Marraffini, L. A. (2014). CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. Molecular Cell 54, 234-244.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-97.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. Cell 136, 215-33.

Bartel, D. P. (2018). Metazoan MicroRNAs. Cell 173, 20-51.

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

Bernstein, D. L., Le Lay, J. E., Ruano, E. G. and Kaestner, K. H. (2015). TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J Clin Invest 125, 1998-2006.

Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409, 363-6.

Betel, D., Wilson, M., Gabow, A., Marks, D. S. and Sander, C. (2008). The microRNA.org resource: targets and expression. Nucleic Acids Res 36, D149-53.

Bogdanove, A. J. and Voytas, D. F. (2011). TAL effectors: customizable proteins for DNA targeting. Science 333, 1843-6.

Bohnsack, M. T., Czaplinski, K. and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA 10, 185-91.

Bolukbasi, M. F., Gupta, A., Oikemus, S., Derr, A. G., Garber, M., Brodsky, M. H., Zhu, L. J. and Wolfe, S. A. (2015). DNA-binding-domain fusions enhance the targeting range and precision of Cas9. Nat Methods.

Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., Miao, X., Streithorst, J. A., Granados, A., Sotomayor-Gonzalez, A. et al. (2020). CRISPR-Cas12-based detection of SARS-CoV-2. Nature Biotechnology.

Canny, M. D., Moatti, N., Wan, L. C. K., Fradet-Turcotte, A., Krasner, D., Mateos-Gomez, P. A., Zimmermann, M., Orthwein, A., Juang, Y. C., Zhang, W. et al. (2018). Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. Nature Biotechnology 36, 95-102.

Capecchi, M. R. (1989). Altering the Genome by Homologous Recombination. Science 244, 1288-1292.

Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nature Reviews Genetics 6, 507-512.

Carlson-Stevermer, J., Abdeen, A. A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M. and Saha, K. (2017). Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. Nature Communications 8, 1711.

Carrell, D. T., Emery, B. R. and Hammoud, S. (2007). Altered protamine expression and diminished spermatogenesis: what is the link? Hum Reprod Update 13, 313-27.

Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F. et al. (2018). A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nature Biotechnology 36, 265-271.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M. et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149, 1393-406.

Chapman, K. M., Medrano, G. A., Jaichander, P., Chaudhary, J., Waits, A. E., Nobrega, M. A., Hotaling, J. M., Ober, C. and Hamra, F. K. (2015). Targeted Germline Modifications in Rats Using CRISPR/Cas9 and Spermatogonial Stem Cells. Cell Reports 10, 1828-1835.

Charlesworth, C. T., Deshpande, P. S., Dever, D. P., Camarena, J., Lemgart, V. T., Cromer, M. K., Vakulskas, C. A., Collingwood, M. A., Zhang, L., Bode, N. M. et al. (2019). Identification of preexisting adaptive immunity to Cas9 proteins in humans. Nat Med 25, 249-254.

Chedin, F., Lieber, M. R. and Hsieh, C. L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A 99, 16916-21.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S. et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155, 1479-91.

Chen, H., Kazemier, H. G., de Groote, M. L., Ruiters, M. H., Xu, G. L. and Rots, M. G. (2014). Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. Nucleic Acids Res 42, 1563-74.

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A. and Doudna, J. A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550, 407-410.

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M. and Doudna, J. A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science 360, 436-439.

Chen, Y. and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res 48, D127-D131.

Chen, Y., Zheng, Y., Kang, Y., Yang, W., Niu, Y., Guo, X., Tu, Z., Si, C., Wang, H., Xing, R. et al. (2015). Functional disruption of the dystrophin gene in rhesus monkey using CRISPR/Cas9. Hum Mol Genet 24, 3764-74.

Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., Rangarajan, S., Shivalila, C. S., Dadon, D. B. and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res 23, 1163-71.

Chi, S. W., Zang, J. B., Mele, A. and Darnell, R. B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460, 479-86.

Cho, H. S., Kang, J. G., Lee, J. H., Lee, J. J., Jeon, S. K., Ko, J. H., Kim, D. S., Park, K. H., Kim, Y. S. and Kim, N. S. (2015). Direct regulation of E-cadherin

by targeted histone methylation of TALE-SET fusion protein in cancer cells. Oncotarget 6, 23837-23844.

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nature Biotechnology 33, 543-U160.

Clermont, Y. (1963). The cycle of the seminiferous epithelium in man. Am J Anat 112, 35-51.

Clermont, Y. and Leblond, C. P. (1959). Differentiation and renewal of spermatogonia in the monkey, Macacus rhesus. Am J Anat 104, 237-73.

Cong, L., Ran, F. A., Cox, D., Lin, S. L., Barretto, R., Habib, N., Hsu, P. D., Wu, X. B., Jiang, W. Y., Marraffini, L. A. et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819-823.

Cradick, T. J., Qiu, P., Lee, C. M., Fine, E. J. and Bao, G. (2014). COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. Molecular Therapy-Nucleic Acids 3.

de Rooij, D. G. (2017). The nature and dynamics of spermatogonial stem cells. Development 144, 3022-3030.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-7.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature 432, 231-5.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J. and Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nature Biotechnology 32, 1262-U130.

Doudna, J. A. (2020). The promise and challenge of therapeutic genome editing. Nature 578, 229-236.

Doudna, J. A. and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258096.

Dow, L. E., Fisher, J., O'Rourke, K. P., Muley, A., Kastenhuber, E. R., Livshits, G., Tschaharganeh, D. F., Socci, N. D. and Lowe, S. W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nature Biotechnology 33, 390-4.

Duan, J., Lu, G., Hong, Y., Hu, Q., Mai, X., Guo, J., Si, X., Wang, F. and Zhang, Y. (2018). Live imaging and tracking of genome regions in CRISPR/dCas9 knock-in mice. Genome Biol 19, 192.

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M. H. and Chandrasegaran, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res 33, 5978-90.

Falahi, F., Huisman, C., Kazemier, H. G., van der Vlies, P., Kok, K., Hospers, G. A. and Rots, M. G. (2013). Towards sustained silencing of HER2/neu in cancer by epigenetic editing. Mol Cancer Res 11, 1029-39.

Fang, W. and Bartel, D. P. (2015). The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. Mol Cell 60, 131-45.

Frock, R. L., Hu, J. Z., Meyers, R. M., Ho, Y. J., Kii, E. and Alt, F. W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nature Biotechnology 33, 179-186.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K. and Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology 31, 822-6.

Fu, Y. F., Sander, J. D., Reyon, D., Cascio, V. M. and Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature Biotechnology 32, 279-284.

Fueller, J., Herbst, K., Meurer, M., Gubicza, K., Kurtulmus, B., Knopf, J. D., Kirrmaier, D., Buchmuller, B. C., Pereira, G., Lemberg, M. K. et al. (2020). CRISPR-Cas12a-assisted PCR tagging of mammalian genes. J Cell Biol 219.

Gaj, T., Gersbach, C. A. and Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology 31, 397-405.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive

immunity in bacteria. Proceedings of the National Academy of Sciences of the United States of America 109, E2579-E2586.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I. and Liu, D. R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature 551, 464-471.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z. R., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A. et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell 154, 442-451.

Gonzalez, F., Zhu, Z., Shi, Z. D., Lelli, K., Verma, N., Li, Q. V. and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell 15, 215-26.

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., Verdine, V., Donghia, N., Daringer, N. M., Freije, C. A. et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438-442.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M. and O'Connor-Giles, K. M. (2014). Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. Genetics 196, 961-+.

Gregory, D. J., Zhang, Y., Kobzik, L. and Fedulov, A. V. (2013). Specific transcriptional enhancement of inducible nitric oxide synthase by targeted promoter demethylation. Epigenetics 8, 1205-12.

Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature 432, 235-40.

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G. and Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell 106, 23-34.

Grove, B. D. and Vogl, A. W. (1989). Sertoli cell ectoplasmic specializations: a type of actin-associated adhesion junction? J Cell Sci 93 (Pt 2), 309-23.

Gu, B., Posfai, E. and Rossant, J. (2018). Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. Nature Biotechnology 36, 632-637.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H. and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science 265, 103-6.

Guilinger, J. P., Thompson, D. B. and Liu, D. R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature Biotechnology 32, 577-82.

Guo, X. and Li, X. J. (2015). Targeted genome editing in primate embryos. Cell Res 25, 767-8.

Gurumurthy, C. B. O'Brien, A. R. Quadros, R. M. Adams, J., Jr. Alcaide, P. Ayabe, S. Ballard, J. Batra, S. K. Beauchamp, M. C. Becker, K. A. et al. (2019).

Reproducibility of CRISPR-Cas9 methods for generation of conditional mouse alleles: a multi-center evaluation. Genome Biol 20, 171.

Gutierrez-Triana, J. A., Tavhelidse, T., Thumberger, T., Thomas, I., Wittbrodt, B., Kellner, T., Anlas, K., Tsingos, E. and Wittbrodt, J. (2018). Efficient single-copy HDR by 5' modified long dsDNA donors. Elife 7.

Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H. and Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev 18, 3016-27.

Hata, K., Okano, M., Lei, H. and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development 129, 1983-93.

Heigwer, F., Kerr, G. and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. Nat Methods 11, 122-3.

Heller, C. G. and Clermont, Y. (1963). Spermatogenesis in man: an estimate of its duration. Science 140, 184-6.

Heller, E. A., Cates, H. M., Pena, C. J., Sun, H., Shao, N., Feng, J., Golden, S. A., Herman, J. P., Walsh, J. J., Mazei-Robison, M. et al. (2014). Locus-specific epigenetic remodeling controls addiction- and depression-related behaviors. Nat Neurosci 17, 1720-7.

Helsel, A. R., Yang, Q. E., Oatley, M. J., Lord, T., Sablitzky, F. and Oatley, J. M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. Development 144, 624-634.

Helwak, A., Kudla, G., Dudnakova, T. and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153, 654-65.

Hemberger, M., Hanna, C. W. and Dean, W. (2020). Mechanisms of early placental development in mouse and humans. Nat Rev Genet 21, 27-43.

Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. and Deiters, A. (2015). Optical Control of CRISPR/Cas9 Gene Editing. J Am Chem Soc 137, 5642-5.

Hendel, A., Bak, R. O., Clark, J. T., Kennedy, A. B., Ryan, D. E., Roy, S., Steinfeld, I., Lunstad, B. D., Kaiser, R. J., Wilkens, A. B. et al. (2015a). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nature Biotechnology.

Hendel, A., Fine, E. J., Bao, G. and Porteus, M. H. (2015b). Quantifying on- and off-target genome editing. Trends in Biotechnology 33, 132-140.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E. and Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 33, 510-U225.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157, 1262-1278.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O. et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology 31, 827-32.

Hu, J. B., Lei, Y., Wong, W. K., Liu, S. Q., Lee, K. C., He, X. J., You, W. X., Zhou, R., Guo, J. T., Chen, X. F. et al. (2014). Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. Nucleic Acids Research 42, 4375-4390.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z. et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57-63.

Huckins, C. (1971). The spermatogonial stem cell population in adult rats. I. Their morphology, proliferation and maturation. Anat Rec 169, 533-57.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T. and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293, 834-8.

Inaba, K. (2011). Sperm flagella: comparative and phylogenetic perspectives of protein components. Mol Hum Reprod 17, 524-38.

Ishii, T. (2015). Germline genome-editing research and its socioethical implications. Trends Mol Med.

Jayathilaka, K., Sheridan, S. D., Bold, T. D., Bochenska, K., Logan, H. L., Weichselbaum, R. R., Bishop, D. K. and Connell, P. P. (2008). A chemical compound that stimulates the human homologous recombination protein RAD51. Proc Natl Acad Sci U S A 105, 15848-53.

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449, 248-51.

Jiang, F. and Doudna, J. A. (2017). CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys.

Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E. and Doudna, J. A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science 351, 867-71.

Jiang, F., Zhou, K., Ma, L., Gressel, S. and Doudna, J. A. (2015). STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science 348, 1477-81.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337, 816-821.

Joung, J. K. and Sander, J. D. (2013). INNOVATION TALENs: a widely applicable technology for targeted genome editing. Nature Reviews Molecular Cell Biology 14, 49-55.

Juillerat, A., Dubois, G., Valton, J., Thomas, S., Stella, S., Marechal, A., Langevin, S., Benomari, N., Bertonati, C., Silva, G. H. et al. (2014). Comprehensive analysis of the specificity of transcription activator-like effector nucleases. Nucleic Acids Res 42, 5390-5402.

Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P. and Church, G. M. (2018). Developmental barcoding of whole mouse via homing CRISPR. Science 361.

Kearns, N. A., Pham, H., Tabak, B., Genga, R. M., Silverstein, N. J., Garber, M. and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods 12, 401-403.

Kester, L. and van Oudenaarden, A. (2018). Single-Cell Transcriptomics Meets Lineage Tracing. Cell Stem Cell 23, 166-179.

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J. and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev 15, 2654-9.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., Hwang, J., Kim, J. I. and Kim, J. S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods 12, 237-43, 1 p following 243.

Kim, H. K., Song, M., Lee, J., Menon, A. V., Jung, S., Kang, Y. M., Choi, J. W., Woo, E., Koh, H. C., Nam, J. W. et al. (2017). In vivo high-throughput profiling of CRISPR-Cpf1 activity. Nat Methods 14, 153-159.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z. and Joung, J. K. (2016a). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490-5.

Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., Welch, M. M., Horng, J. E., Malagon-Lopez, J., Scarfo, I. et al. (2019). Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nature Biotechnology 37, 276-282.

Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., McCaw, Z. R., Aryee, M. J. and Joung, J. K. (2016b). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nature Biotechnology 34, 869-74.

Knight, S. W. and Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science 293, 2269-71.

Kocak, D. D., Josephs, E. A., Bhandarkar, V., Adkar, S. S., Kwon, J. B. and Gersbach, C. A. (2019). Increasing the specificity of CRISPR systems with engineered RNA secondary structures. Nature Biotechnology 37, 657-666.

Kohli, R. M. and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502, 472-9.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-4.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H. et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583-8.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M. and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17, 909-15.

Kosicki, M., Tomberg, K. and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nature Biotechnology 36, 765-771.

Kotaja, N., Bhattacharyya, S. N., Jaskiewicz, L., Kimmins, S., Parvinen, M., Filipowicz, W. and Sassone-Corsi, P. (2006). The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. Proc Natl Acad Sci U S A 103, 2647-52.

Kotaja, N. and Sassone-Corsi, P. (2007). The chromatoid body: a germ-cell-specific RNA-processing centre. Nat Rev Mol Cell Biol 8, 85-90.

Krek, A., Grun, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M. et al. (2005). Combinatorial microRNA target predictions. Nat Genet 37, 495-500.

Kungulovski, G. and Jeltsch, A. (2015). Epigenome Editing: State of the Art, Concepts, and Perspectives. Trends Genet.

Kungulovski, G., Nunna, S., Thomas, M., Zanger, U. M., Reinhardt, R. and Jeltsch, A. (2015). Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. Epigenetics Chromatin 8, 12.

Lanphier, E., Urnov, F., Haecker, S. E., Werner, M. and Smolenski, J. (2015). Don't edit the human germ line. Nature 519, 410-411.

Lara, N. L. M., Costa, G. M. J., Avelar, G. F., Lacerda, S. M. S. N., Hess, R. A. and de França, L. R. (2018). Testis Physiology—Overview and Histology. In Encyclopedia of Reproduction (Second Edition), (ed. M. K. Skinner), pp. 105-116. Oxford: Academic Press.

Laufer, B. I. and Singh, S. M. (2015). Strategies for precision modulation of gene expression by epigenome editing: an overview. Epigenetics Chromatin 8, 34.

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

Lee, F. C. Y. and Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. Mol Cell 69, 354-369.

Lee, H. and Kim, J. S. (2018). Unexpected CRISPR on-target effects. Nature Biotechnology 36, 703-704.

Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., Lee, K., Jung, I., Kim, D., Kim, S. et al. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. Nature Communications 9, 3048.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415-9.

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H. and Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23, 4051-60.

Li, H., Beckman, K. A., Pessino, V., Huang, B., Weissman, J. S. and Leonetti, M. D. (2019). Design and specificity of long ssDNA donors for CRISPR-based knock-in. bioRxiv, 178905.

Li, H. L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., Tanaka, M., Amano, N., Watanabe, A., Sakurai, H. et al. (2015). Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced

Pluripotent Stem Cells by TALEN and CRISPR-Cas9. Stem Cell Reports 4, 143-154.

Li, X., Pritykin, Y., Concepcion, C. P., Lu, Y., La Rocca, G., Zhang, M., King, B., Cook, P. J., Au, Y. W., Popow, O. et al. (2020). High-Resolution In Vivo Identification of miRNA Targets by Halo-Enhanced Ago2 Pull-Down. Mol Cell.

Liang, P. P., Xu, Y. W., Zhang, X. Y., Ding, C. H., Huang, R., Zhang, Z., Lv, J., Xie, X. W., Chen, Y. X., Li, Y. J. et al. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein & Cell 6, 363-372.

Lin, S., Staahl, B. T., Alla, R. K. and Doudna, J. A. (2014a). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3, e04766.

Lin, Y. N., Cradick, T. J., Brown, M. T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B. M., Vertino, P. M., Stewart, F. J. and Bao, G. (2014b). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res 42, 7473-7485.

Ling, X., Xie, B., Gao, X., Chang, L., Zheng, W., Chen, H., Huang, Y., Tan, L., Li, M. and Liu, T. (2020). Improving the efficiency of precise genome editing with site-specific Cas9-oligonucleotide conjugates. Sci Adv 6, eaaz0051.

Little, J. W. (1981). Lambda exonuclease. Gene Amplif Anal 2, 135-45.

Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X. and Qi, L. S. (2015). CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. Bioinformatics.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L. and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science 305, 1437-41.

Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R. A. and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell 167, 233-247 e17.

Long, C. Z., McAnally, J. R., Shelton, J. M., Mireault, A. A., Bassel-Duby, R. and Olson, E. N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science 345, 1184-1188.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. and Kutay, U. (2004). Nuclear export of microRNA precursors. Science 303, 95-8.

Ma, Y. W., Zhang, X., Shen, B., Lu, Y. D., Chen, W., Ma, J., Bai, L., Huang, X. X. and Zhang, L. F. (2014). Generating rats with conditional alleles using CRISPR/Cas9. Cell Res 24, 122-125.

Maeder, M. L., Angstman, J. F., Richardson, M. E., Linder, S. J., Cascio, V. M., Tsai, S. Q., Ho, Q. H., Sander, J. D., Reyon, D., Bernstein, B. E. et al. (2013a). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nature Biotechnology 31, 1137-42.

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H. and Joung, J. K. (2013b). CRISPR RNA-guided activation of endogenous human genes. Nat Methods 10, 977-9.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F. et al. (2011).

Evolution and classification of the CRISPR-Cas systems. Nature Reviews Microbiology 9, 467-477.

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P. et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18, 67-83.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L. and Church, G. M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology 31, 833-8.

Mali, P., Esvelt, K. M. and Church, G. M. (2013b). Cas9 as a versatile tool for engineering biology. Nature Methods 10, 957-963.

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R. and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nature Biotechnology 33, 538-U260.

McDonald, J. I., Celik, H., Rois, L. E., Fishberger, G., Fowler, T., Rees, R., Kramer, A., Martens, A., Edwards, J. R. and Challen, G. A. (2016). Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. Biol Open 5, 866-74.

McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F. and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science 353, aaf7907.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15, 185-97.

Mendenhall, E. M., Williamson, K. E., Reyon, D., Zou, J. Y., Ram, O., Joung, J. K. and Bernstein, B. E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. Nature Biotechnology 31, 1133-6.

Miller, S. M., Wang, T., Randolph, P. B., Arbab, M., Shen, M. W., Huang, T. P., Matuszek, Z., Newby, G. A., Rees, H. A. and Liu, D. R. (2020). Continuous evolution of SpCas9 variants compatible with non-G PAMs. Nature Biotechnology 38, 471-481.

Minczuk, M., Papworth, M. A., Kolasinska, P., Murphy, M. P. and Klug, A. (2006). Sequence-specific modification of mitochondrial DNA using a chimeric zinc finger methylase. Proc Natl Acad Sci U S A 103, 19689-94.

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res 42, W401-7.

Morgan, C. H., Zhang, H. and Bomblies, K. (2017). Are the effects of elevated temperature on meiotic recombination and thermotolerance linked via the axis and synaptonemal complex? Philos Trans R Soc Lond B Biol Sci 372.

Nakagawa, T., Nabeshima, Y. and Yoshida, S. (2007). Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. Developmental Cell 12, 195-206.

Ni, W., Qiao, J., Hu, S. W., Zhao, X. X., Regouski, M., Yang, M., Polejaeva, I. A. and Chen, C. F. (2014). Efficient Gene Knockout in Goats Using CRISPR/Cas9 System. PLoS One 9.

Nihongaki, Y., Kawano, F., Nakajima, T. and Sato, M. (2015). Photoactivatable CRISPR-Cas9 for optogenetic genome editing. Nature Biotechnology 33, 755-60.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell 156, 935-949.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H. et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361, 1259-1262.

Niu, Y. Y., Shen, B., Cui, Y. Q., Chen, Y. C., Wang, J. Y., Wang, L., Kang, Y., Zhao, X. Y., Si, W., Li, W. et al. (2014). Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos. Cell 156, 836-843.

Nunna, S., Reinhardt, R., Ragozin, S. and Jeltsch, A. (2014). Targeted methylation of the epithelial cell adhesion molecule (EpCAM) promoter to silence its expression in ovarian cancer cells. Plos One 9, e87703.

O'Donnell, L., Nicholls, P. K., O'Bryan, M. K., McLachlan, R. I. and Stanton, P. G. (2011). Spermiation: The process of sperm release. Spermatogenesis 1, 14-35.

Oakberg, E. F. (1956a). A description of spermiogenesis in the mouse and its use in analysis of the cycle of the seminiferous epithelium and germ cell renewal. Am J Anat 99, 391-413.

Oakberg, E. F. (1956b). Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. Am J Anat 99, 507-16.

Oakberg, E. F. (1971). Spermatogonial stem-cell renewal in the mouse. Anat Rec 169, 515-31.

Oliver, D., Yuan, S. Q., McSwiggin, H. and Yan, W. (2015). Pervasive Genotypic Mosaicism in Founder Mice Derived from Genome Editing through Pronuclear Injection. PLoS One 10.

Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Majoros, W. H., Reddy, T. E. and Gersbach, C. A. (2015). Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. Nature Communications 6.

Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T. and Hatzigeorgiou, A. G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res 41, W169-73.

Paulis, M., Castelli, A., Lizier, M., Susani, L., Lucchini, F., Villa, A. and Vezzoni, P. (2015). A pre-screening FISH-based method to detect CRISPR/Cas9 off-targets in mouse embryonic stem cells. Sci Rep 5, 12327.

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W. et al.

(2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods 10, 973-6.

Porteus, M. H. (2019). A New Class of Medicines through DNA Editing. N Engl J Med 380, 947-959.

Quadros, R. M., Miura, H., Harms, D. W., Akatsuka, H., Sato, T., Aida, T., Redder, R., Richardson, G. P., Inagaki, Y., Sakai, D. et al. (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biol 18, 92.

Rai, A. and Cross, J. C. (2014). Development of the hemochorial maternal vascular spaces in the placenta through endothelial and vasculogenic mimicry. Dev Biol 387, 131-41.

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X. B., Makarova, K. S. et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186-U98.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. et al. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell 154, 1380-1389.

Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. RNA 10, 1507-17.

Reik, W., Dean, W. and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science 293, 1089-93.

Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. and Corn, J. E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nature Biotechnology 34, 339-44.

Rivenbark, A. G., Stolzenburg, S., Beltran, A. S., Yuan, X., Rots, M. G., Strahl, B. D. and Blancafort, P. (2012). Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics 7, 350-60.

Rossant, J. and Cross, J. C. (2001). Placental development: lessons from mouse mutants. Nat Rev Genet 2, 538-48.

Rouet, P., Smih, F. and Jasin, M. (1994). Introduction of Double-Strand Breaks into the Genome of Mouse Cells by Expression of a Rare-Cutting Endonuclease. Mol Cell Biol 14, 8096-8106.

Rudin, N., Sugarman, E. and Haber, J. E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. Genetics 122, 519-34.

Russell, L. D. (1979). Spermatid-Sertoli tubulobulbar complexes as devices for elimination of cytoplasm from the head region late spermatids of the rat. Anat Rec 194, 233-46.

Russell, L. D. (1990). Histological and histopathological evaluation of the testis. Clearwater, Fl: Cache River Press.

Sander, J. D. and Joung, J. K. (2014). CRIRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology 32, 347-355.

Savic, N., Ringnalda, F. C., Lindsay, H., Berk, C., Bargsten, K., Li, Y., Neri, D., Robinson, M. D., Ciaudo, C., Hall, J. et al. (2018). Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. Elife 7.

Schmid-Burgk, J. L., Gao, L., Li, D., Gardner, Z., Strecker, J., Lash, B. and Zhang, F. (2020). Highly Parallel Profiling of Cas9 Variant Specificity. Mol Cell 78, 794-800 e8.

Schwank, G., Koo, B. K., Sasselli, V., Dekkers, J. F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C. K. et al. (2013). Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. Cell Stem Cell 13, 653-658.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X. et al. (2014a). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat Methods 11, 399-402.

Shen, Z. F., Zhang, X. L., Chai, Y. P., Zhu, Z. W., Yi, P. S., Feng, G. X., Li, W. and Ou, G. S. (2014b). Conditional Knockouts Generated by Engineered CRISPR-Cas9 Endonuclease Reveal the Roles of Coronin in C. elegans Neural Development. Developmental Cell 30, 625-636.

Sheu-Gruttadauria, J., Xiao, Y., Gebert, L. F. and MacRae, I. J. (2019). Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. EMBO J 38, e101153.

Siddique, A. N., Nunna, S., Rajavelu, A., Zhang, Y., Jurkowska, R. Z., Reinhardt, R., Rots, M. G., Ragozin, S., Jurkowski, T. P. and Jeltsch, A. (2013).

Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. Journal of Molecular Biology 425, 479-91.

Simhadri, V. L., McGill, J., McMahon, S., Wang, J., Jiang, H. and Sauna, Z. E. (2018). Prevalence of Pre-existing Antibodies to CRISPR-Associated Nuclease Cas9 in the USA Population. Mol Ther Methods Clin Dev 10, 105-112.

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X. and Zhang, F. (2015). Rationally engineered Cas9 nucleases with improved specificity. Science.

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J. C., Chames, P., Prieto, J., Redondo, P., Blanco, F. J., Bravo, J. et al. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34, e149.

Snowden, A. W., Gregory, P. D., Case, C. C. and Pabo, C. O. (2002). Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Current Biology 12, 2159-66.

Song, J., Yang, D., Xu, J., Zhu, T., Chen, Y. E. and Zhang, J. (2016). RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. Nature Communications 7, 10548.

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N. and Junker, J. P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nature Biotechnology 36, 469-473.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62-7.

Sticht, C., De La Torre, C., Parveen, A. and Gretz, N. (2018). miRWalk: An online resource for prediction of microRNA binding sites. PLoS One 13, e0206239.

Stoddard, B. L. (2011). Homing Endonucleases: From Microbial Genetic Invaders to Reagents for Targeted DNA Modification. Structure 19, 7-15.

Stolzenburg, S., Beltran, A. S., Swift-Scanlan, T., Rivenbark, A. G., Rashwan, R. and Blancafort, P. (2015). Stable oncogenic silencing in vivo by programmable and targeted de novo DNA methylation in breast cancer. Oncogene 34, 5427-35.

Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J. and Russell, R. (2018). Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. Mol Cell.

Sung, Y. H., Kim, J. M., Kim, H. T., Lee, J., Jeon, J., Jin, Y., Choi, J. H., Ban, Y. H., Ha, S. J., Kim, C. H. et al. (2014). Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Research 24, 125-131.

Teng, F., Li, J., Cui, T., Xu, K., Guo, L., Gao, Q., Feng, G., Chen, C., Han, D., Zhou, Q. et al. (2019). Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds. Genome Biol 20, 15.

Treiber, T., Treiber, N. and Meister, G. (2019). Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. Nat Rev Mol Cell Biol 20, 5-20.

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J. and Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature Biotechnology 32, 569-76.

Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P. et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology 33, 187-197.

Turner, J. M. (2007). Meiotic sex chromosome inactivation. Development 134, 1823-31.

Vakulskas, C. A., Dever, D. P., Rettig, G. R., Turk, R., Jacobi, A. M., Collingwood, M. A., Bode, N. M., McNeill, M. S., Yan, S., Camarena, J. et al. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. Nat Med 24, 1216-1224.

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K. et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods 13, 508-14.

Veneziano, R., Shepherd, T. R., Ratanalert, S., Bellou, L., Tao, C. and Bathe, M. (2018). In vitro synthesis of gene-length single-stranded DNA. Sci Rep 8, 6548.

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., Klasic, M. and Zoldos, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res 44, 5615-28.

Wagner, D. L., Amini, L., Wendering, D. J., Burkhardt, L. M., Akyuz, L., Reinke, P., Volk, H. D. and Schmueck-Henneresse, M. (2019). High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. Nat Med 25, 242-248.

Walton, R. T., Christie, K. A., Whittaker, M. N. and Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science 368, 290-296.

Wan, H. F., Feng, C. J., Teng, F., Yang, S. H., Hu, B. Y., Niu, Y. Y., Xiang, A. P., Fang, W. Z., Ji, W. Z., Li, W. et al. (2015). One-step generation of p53 gene biallelic mutant Cynomolgus monkey via the CRISPR/Cas system. Cell Res 25, 258-261.

Wang, H., Nakamura, M., Abbott, T. R., Zhao, D., Luo, K., Yu, C., Nguyen, C. M., Lo, A., Daley, T. P., La Russa, M. et al. (2019a). CRISPR-mediated live imaging of genome editing and transcription. Science 365, 1301-1305.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153, 910-8.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R. J. and Yee, J. K. (2015). Unbiased detection of off-target cleavage by

CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nature Biotechnology 33, 175-8.

Wang, Z., Lee, S., Oliver, D., Yuan, S., Tang, C., Wang, Y., Zheng, H. and Yan, W. (2018). Prps1l1, a testis-specific gene, is dispensable for mouse spermatogenesis. Mol Reprod Dev 85, 802-804.

Wang, Z., McSwiggin, H., Newkirk, S. J., Wang, Y., Oliver, D., Tang, C., Lee, S., Wang, S., Yuan, S., Zheng, H. et al. (2019b). Insertion of a chimeric retrotransposon sequence in mouse Axin1 locus causes metastable kinky tail phenotype. Mob DNA 10, 17.

Wang, Z., Wang, Y., Wang, S., Gorzalski, A. J., McSwiggin, H., Yu, T., Castaneda-Garcia, K., Prince, B., Wang, H., Zheng, H. et al. (2020a). Efficient genome editing by CRISPR-Mb3Cas12a in mice. J Cell Sci 133.

Wang, Z., Xie, Y., Wang, Y., Morris, D., Wang, S., Oliver, D., Yuan, S., Zayac, K., Bloomquist, S., Zheng, H. et al. (2020b). X-linked miR-506 family miRNAs promote FMRP expression in mouse spermatogonia. EMBO Rep 21, e49024.

Watson, E. D. and Cross, J. C. (2005). Development of structures and transport functions in the mouse placenta. Physiology (Bethesda) 20, 180-93.

Weterings, E. and Chen, D. J. (2008). The endless tale of non-homologous end-joining. Cell Res 18, 114-24.

Wienert, B., Nguyen, D. N., Guenther, A., Feng, S. J., Locke, M. N., Wyman, S. K., Shin, J., Kazane, K. R., Gregory, G. L., Carter, M. A. M. et al. (2020). Timed

inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. Nature Communications 11, 2109.

Wu, Y. X., Liang, D., Wang, Y. H., Bai, M. Z., Tang, W., Bao, S. M., Yan, Z. Q., Li, D. S. and Li, J. S. (2013). Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. Cell Stem Cell 13, 659-662.

Wu, Y. X., Zhou, H., Fan, X. Y., Zhang, Y., Zhang, M., Wang, Y. H., Xie, Z. F., Bai, M. Z., Yin, Q., Liang, D. et al. (2015). Correction of a genetic disease by CRISPR-Cas9-mediated gene editing in mouse spermatogonial stem cells. Cell Res 25, 67-79.

Wyman, C. and Kanaar, R. (2006). DNA double-strand break repair: All's well that ends well. Annual Review of Genetics 40, 363-383.

Wyvekens, N., Topkar, V. V., Khayter, C., Joung, J. K. and Tsai, S. Q. (2015). Dimeric CRISPR RNA-Guided FokI-dCas9 Nucleases Directed by Truncated gRNAs for Highly Specific Genome Editing. Hum Gene Ther 26, 425-31.

Xiao, A., Cheng, Z. C., Kong, L., Zhu, Z. Y., Lin, S., Gao, G. and Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics 30, 1180-1182.

Xie, F., Ye, L., Chang, J. C., Beyer, A. I., Wang, J., Muench, M. O. and Kan, Y. W. (2014). Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. Genome Res 24, 1526-33.

Xu, H., Xiao, T., Chen, C. H., Li, W., Meyer, C. A., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J. S. et al. (2015). Sequence determinants of improved CRISPR sgRNA design. Genome Res.

Xue, Z. Y., Wu, M. H., Wen, K. J., Ren, M. D., Long, L., Zhang, X. D. and Gao, G. (2014). CRISPR/Cas9 Mediates Efficient Conditional Mutagenesis in Drosophila. G3-Genes Genomes Genetics 4, 2167-2173.

Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L. and Jaenisch, R. (2013a). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154, 1370-9.

Yang, L., Guell, M., Byrne, S., Yang, J. L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A. W., Rios, X. et al. (2013b). Optimization of scarless human stem cell genome editing. Nucleic Acids Res 41, 9049-61.

Yi, R., Qin, Y., Macara, I. G. and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev 17, 3011-6.

Yin, L., Maddison, L. A., Li, M., Kara, N., LaFave, M. C., Varshney, G. K., Burgess, S. M., Patton, J. G. and Chen, W. (2015). Multiplex Conditional Mutagenesis Using Transgenic Expression of Cas9 and sgRNAs. Genetics 200, 431-41.

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S. et al. (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell 16, 142-7.

Yu, Y., Guo, Y., Tian, Q., Lan, Y., Yeh, H., Zhang, M., Tasan, I., Jain, S. and Zhao, H. (2020). An efficient gene knock-in strategy using 5'-modified double-stranded DNA donors with short homology arms. Nat Chem Biol 16, 387-390.

Yuan, S., Liu, Y., Peng, H., Tang, C., Hennig, G. W., Wang, Z., Wang, L., Yu, T., Klukovich, R., Zhang, Y. et al. (2019). Motile cilia of the male reproductive system require miR-34/miR-449 for development and function to generate luminal turbulence. Proc Natl Acad Sci U S A 116, 3584-3593.

Yuan, S., Stratton, C. J., Bao, J., Zheng, H., Bhetwal, B. P., Yanagimachi, R. and Yan, W. (2015). Spata6 is required for normal assembly of the sperm connecting piece and tight head-tail conjunction. Proc Natl Acad Sci U S A 112, E430-9.

Zarnegar, B. J., Flynn, R. A., Shen, Y., Do, B. T., Chang, H. Y. and Khavari, P. A. (2016). irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods 13, 489-92.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A. et al. (2015a). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell 163, 759-71.

Zetsche, B., Volz, S. E. and Zhang, F. (2015b). A split-Cas9 architecture for inducible genome editing and transcription modulation. Nature Biotechnology 33, 139-42.

CHAPTER II: Rapidly evolving X-linked *miR-506* family miRNAs control sperm fitness through fine-tuning spermatogenesis

Zhuqing Wang[1], Yue Wang[1], Shawn Wang[1], Tong Zhou[1], Hetan Wang[1], Yeming Xie[1], Hayden McSwiggin[1], Sheng Chen[1], Kevin J. Peterson[2], Eric Lai[3], Huili Zheng[1], Wei Yan[1,4,5]

[1]Department of Physiology and Cell Biology, University of Nevada, Reno School of Medicine, Reno, NV 89557, USA; [2] Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA; [3] Department of Developmental Biology, Sloan-Kettering Institute, 1275 York Ave, Box 252, New York, NY 10065, USA; [4]The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [5]Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

Correspondence:

Wei Yan M.D., Ph.D.

Investigator, The Lundquist Institute at Harbor-UCLA Medical Center

Professor, David Geffen School of Medicine at UCLA

1124 W. Carson St., Torrance, CA 90502

Email: wei.yan@lundquist.org

Abstract

The X-linked *miR-506* family consists of 22 miRNAs clustered in 5 sub-regions expanding a ~62kb region near *Slitrk2* and 1 cluster expanding a ~22kb region close to *Fmr1* on the X chromosome in mice. Although the X-linked *miR-506* family miRNAs were all derived from the same ancestors, they have undergone rapid evolution and are preferentially expressed in the testis among all eutherian mammals. To define the physiological roles of these rapidly evolving, testis-specific, X-linked miRNAs, we sequentially deleted 5 of the 6 miRNA clusters using CRISPR-Cas-based genome editing in the mouse genome. While the deletion of individual miRNAs or one or two clusters did not generate discernable phenotypes, we started to see abnormalities in the KO males lacking more than 4 clusters. Mice lacking the 18 miRNAs showed slightly reduced litter size and longer litter interval despite normal testis weight and normal sperm counts. When the typical one male-one female mating scheme was used, the KO males were only subfertile. However, when a female was mated sequentially with either a wild type male first and a KO second, or a KO first followed by a WT male, no or much fewer pups were derived from the KO sperm, suggesting the KO sperm are less fit in fertilizing eggs. Indeed, *in vitro* fertilization assays showed that the KO sperm were less competitive in fertilizing wild type eggs. Our data suggests that sperm produced in the absence of these X-linked miRNAs are less fertile/competitive than wild-type sperm, and that the *miR-506* family miRNAs function to fine-tune certain molecular processes that render sperm better fitness.

Instruction

Spermatogenesis is the differentiation of spermatogonia to spermatozoa in the male reproductive system, which involves three major phases: mitosis (including spermatogonia stem cell renewal and differentiation), meiosis and spermiogenesis (1). After dividing from type B spermatogonia, the primary spermatocytes enter into meiosis S phase (preleptotene), then undergo long meiosis prophase I, including leptotene, zygotene, pachytene, diplotene, and diakinesis, followed by two rounds of rapid meiosis (MI and MII) to form haploid round spermatids (2). During meiosis prophase I, double-stranded breaks (DSBs) form in leptotene, synapsis begins and DNA recombination repair occurs in zygotene, followed by synapsis ending and DNA resolution in pachytene (2). Round spermatids then undergo spermiogenesis, which involves degradation of histone and replacement with protamine that are composed of arginine- and cysteine-rich proteins compacting DNA, resulting in spermatozoa (2). Each phase involves tightly spatiotemporal gene expression control to ensure proper spermatogenesis.

As one type of the small non-coding RNAs, ~22nt microRNAs (miRNAs) regulate genes at the post-transcriptional level. After being transcribed by RNA polymerase II, primary miRNA (pri-miRNA) is cleaved by microprocessor containing DROSHA-DGCR8 (DiGeorge syndrome critical region 8) to form ~70 nt hairpin precursor miRNA (pre-miRNA) in the nucleus(3). Then the pre-miRNA is transported into the cytoplasm and further processed by DICER to form ~22 nt miRNA duplex. MiRNA duplex is unwound by Argonaute (Ago) proteins, one strand of miRNA is loaded

into miRNA induced silencing complex (miRISC) to target mRNA (4-6). In most cases, miRNA downregulates translation by repressing translation or degrading mRNA, sometimes miRNA could also upregulate translation (6-9).

Conditional knock-out (KO) of enzymes for miRNA biogenesis, Drosha or Dicer, impaired spermatogenesis (10), indicating the essential role of miRNAs in spermatogenesis. MiRNAs usually function together by either family that has similar seeds or cluster that transcribed from a single transcript. Indeed, single KO of either *miR-34b* or *miR-449* cluster had no discernable phenotype (11), whereas the ablation of *miR-34b* and *miR-449* clusters led to perinatal lethality, impaired spermatogenesis and motile cilia dysfunction (12-14). Previous studies have indicated that the X linked *miR-506* family is highly expressed in testis and germ cells(15, 16), and these miRNAs evolve fast across species(17-20). However, despite their high expression in testis and rapid evolution, their roles in spermatogenesis remain largely unknown. Several studies attempted to dissect these miRNAs' functions, yet these functional analyses were largely done either in vitro (20-22) or with insufficient KO of these miRNAs (23). Here we explored the evolution of the *miR-506* family, the expression pattern of these miRNAs, furthermore, we generated several different KOs of these miRNAs to dissect their functions in spermatogenesis.

Results

The majority of X-linked *miR-506* family are located within the *Slitrk2-Fmr1* locus across species

X linked genes are usually more divergent than autosomal genes, which is termed as the “faster-X effect” (24). These genes include protein-coding genes (25, 26) as well as miRNAs (17, 18, 20). Interestingly, on the X chromosome, despite the fact that both *Slitrk2* and *Fmr1* genes are relatively conserved across species, the miRNAs between these two loci, which have been cataloged into the *miR-506* family based on their common ancestors (similar sequences of genomic DNA and precursor miRNAs) (27, 28), are highly divergent among clades across all the eutherian mammals. The *miR-506* family miRNAs have also been called spermiR (22), Fx-mir (21) or XmiRs(23).

To identify the miRNA orthologs within the *Slitrk2-Fmr1* loci, a previous study used Blast and Exonerate to search the miRNA databases (22). However, this method may lose a lot of miRNAs due to the sequence divergence of the *miR-506* family. Therefore, we first compared the genomic DNA sequences within the *Slitrk2-Fmr1* loci across 100 different species using the Multiz Alignment & Conservation pipeline, which are based on PhastCons and PhyloP, on the UCSC genome browser with the human genome as a reference (Figure 2-1A, S1, and S2A) (29, 30). PhastCons takes the flanking sequences into consideration and doesn’t rely on fixed sliding windows; consequently, highly conserved short sequences and moderately conserved long sequences can have higher scores (31). By contrast,

PhyloP compares the conservation of individual nucleotides among all clades of phylogeny, and gives positive scores once the region is conserved and vice versa. Consistent with the previous reports (21, 22), the *SLITRK2* and *FMR1* loci are highly conserved, while the miRNAs within the *SLITRK2-FMR1* region are highly divergent, among all the species when using the human genome as the reference. Similar results were obtained when the mouse genome was used as a reference (Figure 2-S2A). Moreover, we found that the miRNAs (*miR-892c* - *miR-891a*) located proximal to *SLITRK2* (named as SLITRK2-associated miRNAs, SmiR) are much more divergent than the ones (*miR-513c* - *miR-514a3*) closer to *FMR1* (FmiR for *FMR1*- associated miRNAs) according to the PhyloP scores (Figure 2-1A). Interestingly, in contrast to what previously described that these miRNAs only exist in eutherian/placental mammals and one marsupial (Monodelphis domestica)(22), we found that the orthologues of FmiR, including *miR-514b*, *miR-509*, and *miR-510*, exist in green sea turtles, suggesting that these miRNAs may have emerged very early in animal kingdoms (Figure 2-S1). To take a deeper look at the *Slitrk2~Fmr1* genomic DNA region, we applied Dot-plot analysis using D-GENIES (32), which utilizes a fixed sliding window alignment and allowed us to examine sequence similarity between species. Similar to the PhyloP scores, the Dot-plot confirmed that the *Slitrk2* and *Fmr1* genomic DNA sequences are conserved across all species, yet the rest of the sequences are highly divergent among clades (Figure 2-S2B). Interestingly, the sequences within some clades are quite similar, for example, the primates (e.g. humans, chimpanzees and rhesus monkeys) and the cetartiodactyla (e.g. sheep, cows), although inversions happened within sheep

and cow (Figure 2-S2B). Some similar regions were also shared within the rodentia (e.g. mice and rats) and the carnivora (e.g. dog and cat) (Figure 2-S2B). The majority of the *Slitrk2-Fmr1* in mammals are on the positive strand, whereas miRNAs are in the reverse orientation (Figure 2-1B and S2B), which is consistent with what previously described (21, 22). However, we also observed some exceptions in rats, cows and sheep (Figure 2-1B), which were not included in the previous study (22). In rat, the *miR-201- Fmr1* loci (including *miR-201*, *miR-547*, *miR-509* and *Fmr1*) were inverted compared to the other genomes (Figure 2-1B). The *Slitrk2* and *miR-509~547* were inverted in cows and sheep, respectively (Figure 2-1B).

In order to see the relationship among these miRNAs, we built a phylogram for the *miR-506* family (Figure 2-S3). According to the phylogram, these miRNAs shared a common ancestor and the FmiR cluster emerged earlier than the SmiR, which is consistent with the aforementioned that some FmiRs exist in green sea turtles based on the analysis from Multiz Alignment & Conservation. Interestingly, the FmiR miRNA clusters tended to cluster together, the same scenario happened to the SmiR clusters, suggesting that these two clusters may have evolved separately. Although the miRNAs within the *miR-506* family are highly divergent, some similarities exist for some orthologues among different species. For example, the SmiRs within primates are quite similar. Although it seems the *miR-891a* and *miR-891b* were swapped in rhesus monkey compared to human and chimpanzee genomes (Figure 2-S4A), it turns out that this is due to the nomenclature since

sequence alignment showed that the *miR-891a* and *miR-891b* in rhesus monkey were more similar to *miR-891b* and *miR-891a*, in humans and chimpanzees, respectively (Figure 2-S4B). Similarly, when comparing humans and rhesus monkeys, the chimpanzee seems to have lost *miR-892c* and *miR-888*, but the *miR-892b* in the chimpanzee is more similar to *miR-892c* regarding the location as well as the sequences in humans and rhesus monkeys (Figure 2-S4A, S4B). The FmiR cluster, including *miR-201* (assigned as P1 (paralogue 1) in MirGeneDB(27)), *miR-547* (P2), and *miR-509* (P7) in mice and rats, are orthologues of *miR-506*, *miR-507* and *miR-509* in humans, respectively (Figure 2-1C). Although the orthologs of *miR-891a*, *miR-891b*, *miR-513* and *miR-510* in humans were not observed in mice, we noticed that *miR-513* and SmiR (including *miR-891a* and *miR-891b*) share a common ancestor (Fig.S3), suggesting that the FmiRs and SmiRs may be divergent from this common ancestor. Interestingly, *miR-513* has 4 copies in the human genome, which is derived from MER91C DNA transposon (Figure 2-S4A).

LINE retrotransposons seem to drive the expansion of X linked *miR-506* family derived from MER91C DNA transposon

Several studies have shown that transposable elements (TEs) drive evolution (33), and a lot of miRNAs were derived from TEs (34). We asked what TEs might be involved in driving the evolution of the X linked *miR-506* family. Interestingly, some X-linked *miR-506* family miRNAs are derived from TEs. Besides mentioned above that *miR-513s* are derived from MER91C DNA transposon, *miR-891a* and *miR-891b* are partially derived from MIRc SINE (short interspersed element) family in

the human genome. In the X linked *miR-506* locus, long interspersed elements (LINEs) are pervasive in mice and rats, whereas SINEs are more abundant than LINEs in dogs, horses, rhesus monkeys, chimpanzees and humans (Figure 2-S4A). LINE-1 (L1) is the most abundant and the only active autonomous TE in the human genome (35). Although SINEs can't mobilize by themselves, they can hijack the L1 mechanism (35). Results from both the phylogram tree and Multiz Alignment & Conservation analyses suggested that the FmiRs (e.g., *miRs-506-509*) emerged much earlier than the SmiRs. To test whether the X-linked *miR-506* family might be derived from MER91C DNA transposon and further expanded by LINE-mediated retrotransposition, we analyzed the region containing the *miR-506* family miRNAs in the human genome. Indeed, we observed that most of the FmiRs are flanked by L1 retrotransposons (Figure 2-S4A). For example, *miR-506*, *miR-507* and *miR-508* are flanked by L1MA6, whereas *miR-510* is accompanied by LTR22 and L1M4b. *miR-509* has 3 copies, which are all flanked by X6A_LINE. Three copies of *miR-514a* and one copy of *miR-514b* are flanked with L2a and L1MA9, respectively. By contrast, there are fewer transposons in SmiR locus in the human genome. The composition of transposons in the SmiR loci are almost identical among humans, rhesus monkeys and chimpanzees. The miRNAs location and order are all the same between humans and rhesus monkeys. AluY and AluYa5, the youngest active Alu family(36), present within *miR-891b* and *miR-892b*, and *miR-892a* and *miR-888* in the human genome, respectively. The AluY is also found within *miR-891b* and *miR-892a* in chimpanzees. However, there are no transposons in between *miR-891b* ~ *miR-892c* cluster in the rhesus monkey,

suggesting that AluY and AluYa5 may not be responsible for the transposition of this cluster. What's more, the THE1C LTR retrotransposon is always next to the *miR-892c* in humans and rhesus monkeys, and the *miR-892b* in chimpanzees, suggesting this retrotransposon may be responsible for the evolution of the *miR-891b ~ miR-892c* cluster in primates.

On the contrary to humans, the mice and rats SmiR have more TEs than the FmiR (Figure 2-S4A). The majority of the TEs in mice and rat are from Lx family, which is the oldest LINE1(37). In mice, the 6 copies of the *miR-465* cluster are largely flanked with Lx7 and Lx5c LINE1 retrotransposons, the *miR-883a* cluster is flanked with Lx7 and L1_Mur2, next to the *miR-741* cluster is Lx9. In FmiR, *miR-201* and *miR-547* are flanked with L1_Mus2, *miR-509* is flanked with Lx3_Mus. Similarly, in rat, the *miR-465* is flanked by Lx7 and Lx8, the miR-883a cluster was flanked by Lx7 and L1_Mur2, and next to the *miR-741* cluster is Lx9 and Lx5c. In FmiR, *miR-509* is flanked with Lx2, *miR-3585* is flanked with L1_Rat1 and Lx8.

In horses, the SmiRs are very compact, which are flanked with L1MA10 and L1MEc (Figure 2-S4A). The *miR-9142* and *miR-9143* are also derived from MIRc, similar to human *miR-891a* and *miR-891b*. The FmiRs, except *miR-509*, are all derived from DNA transposon MER91C, which is similar to human *miR-513*. The *miR-509a* and *miR-509b* are both flanked by ERE3B2 (SINE family) and L1M4b (Figure 2-S4A).

In dogs, the SmiR are flanked with SINEC_b1 and SINEC_b2, all FmiRs, including *miR-506*, *miR-507*, *miR-508* and *miR-514*, are all derived from MER91C DNA

transposon (Figure 2-S4A).

Taken together, these results suggest that the X linked *miR-506* miRNAs are largely derived from MER91C DNA transposon, and expanded by LINE family.

The X linked *miR-506* family is highly abundant in testis and sperm

The X linked *miR-506* family is highly expressed in testes in humans (21), rhesus monkeys (17), marmosets (38), mice (15, 21, 22, 28), rats (22), rabbits (22) and pigs (22). To define the expression pattern of the *miR-506* family during spermatogenesis, we performed STA-PUT cell sorting (1) on mice and rats, followed by small RNA sequencing (sRNA-seq). Consistent with what previously described, these miRNAs are highly abundant in high purity of germ cells (Figure 2-S5A) in either CD1 (Figure 2-2A) or C57BL/6J (Figure 2-2B) mice strain (15, 22, 28), suggesting that these miRNAs escape the meiotic sex chromosome inactivation (MSCI) in mice. Unlike mice, these miRNAs were not highly expressed in rat pachytene spermatocyte and round spermatids (Figure 2-2D), suggesting these miRNAs underwent MSCI in rats. We also analyzed the horse testis small RNA both in immature and mature horses from the previous data (39). We found that the *miR-506* family miRNAs are highly expressed in horse testis, and the abundances of these miRNAs are increasing during sexual maturation (Figure 2-2F).

Interestingly, despite their high expression in the germ cells in the testis, these miRNAs also have high abundances in mature sperm in both mice and rats (Figure

2-2A, 2B and 2D), suggesting that these miRNAs may have roles during early embryonic development. We observed some inconsistences with the previous study showing that these miRNAs are in low abundance in mature sperm(22). This is not surprising because the sperm RNAs seem to be highly compacted in sperm DNA and chromatins that are enriched with disulfide bonds, so that the RNAs cannot be easily released in the lysis buffer without reducing reagents and detergent (40), such as the TRIzol reagent that the authors used. Consistent with this expression pattern in mice, humans with lower expression of the X linked *miR-506* family miRNAs in the semen showed asthenozoospermia with reduced motility(41).

Despite the high abundance of these miRNAs among placental mammals, it seems that these miRNAs express in a linage specific pattern. In mice, the SmiR cluster tends to be highly expressed, whereas the expressions of FmiR cluster are very low. In rats, the two clusters all have high abundances. Analysis of human testis small RNA-seq data (42) showed the FmiR cluster tends to have high abundance, while SmiR cluster is barely expressed (Figure 2-2D). Similar phenomena happened in the other primate, such as marmoset (Figure 2-2E) (38) and rhesus monkey(17). In horses, both SmiR and FmiR are highly expressed in adult testis (Figure 2-2F). A recent study showed that highly expressed protein-coding genes in testis tend to have lower mutation rates and evolution rate due to the so-called transcription-coupled repair (TCR) (43). We observed the same phenomenon for the Fmr1 and Slitrk2, in which the conserved Fmr1 was highly expressed in

spermatogonia cells in either humans or mice testis, as previously described (28), and the conserved Slitrk2 was enriched in mouse elongating spermatids (43). However, in contrast to the protein-coding genes, the X linked miRNAs have high expression in testes despite their fast evolution.

Ablation of X linked *miR-506* compromise sperm fitness

To define the function of the *miR-506* family, we sequentially knocked out (KO) this family using CRISPR-Cas9 (Figure 2-3A) (28). We first KO the *miR-883* cluster (*miR-883* sKO) or the *miR-465* cluster (*miR-465* sKO) (Figure 2-3A), since these two clusters have the most abundance in the mouse testis (15, 28), unfortunately, we didn't observe any discernable phenotype regarding the testis size, morphology, sperm concentration and sperm motility (Figure 2-3B, 3C, 3D and 3G). On the *miR-883* sKO background, we KO *miR-741* cluster, termed double KO (dKO) (Figure 2-3A), still didn't find any phenotype in fertility and testis morphology as well as sperm motility and concentration (Figure 2-S5B-S5E). Then on the dKO background, we either KO the *miR-465* (termed triple KO (tKO)) or the *miR-471* & *miR-470* clusters (termed quadruple KO (qKO)) (Figure 2-3A), and lastly, we KO the *miR-465* cluster on the qKO background, named pentuple KO (pKO) (Figure 2-3A). We found that in tKO, the litter size is still comparable with the wild type, while in the qKO and pKO, the litter size is highly reduced, although no differences were found in the litter interval (Figure 2-3B and 3C). The testes sizes and morphology of these KOs are comparable to WT testis (Figure 2-3D and 3G). Then we measured the sperm quality by computer-assisted sperm analysis (CASA),

although the sperm concentrations were not significantly reduced among all the of KOs, sperm motility was highly compromised in the qKO and pKO (Figure 2-3E and 3F). In the natural world, the females of most species throughout the animal kingdoms undergo multiple matings with males before breeding, thus leading to sperm competition after mating(44). We set several experiments to test the sperm fitness between WT and pKO. Since these miRNAs are X linked, the sperm that carries Y chromosome from the pKO mice cannot be distinguished from the ones from WT. Therefore, we adapted a transgenic mTmG mice (45) for sperm competition experiments, which allows us to visualize the embryos as well as identify genotype. We first compared their ability in fertilizing eggs by performing in vitro fertilization (IVF) with several different sperm ratios. When 100% mTmG or 100% pKO sperm were used to fertilize WT eggs, the pKO has comparable blastocyst rate to the mTmG (~80%), the blastocyst rates were lower when different sperm ratio was used, especially in the 1:1 and 1:4 ratio, although no significant differences were observed (Figure 2-3H). Interestingly, when 1:1 sperm ratio (mTmG: pKO) was used for sperm competition, the number of blastocysts of mTmG is always ~ three-fold more than the pKO (Figure 2-3I). The blastocysts numbers of mTmG went up to ~ seven-fold compared to pKO when the 4:1 sperm ratio was used (Figure 2-3I). When the 1:4 sperm ratio was used, the mTmG can still reach up to a 1:1.9 blastocysts rate (Figure 2-3I). These results indicating that the pKO is less competitive than WT during IVF. Then we set a serial mating experiments to see the in vivo situation. Mice ovulation happens 10-13h after HCG administration. WT female mice were treated with HCG at 8 PM, then the first male

mice were introduced with the female and replaced with the second male mice at 6 AM the next morning. Interestingly, no pKO pups were found (n=8) when the mTmG male mice mated first with the WT female (mTmG V.S. pKO), by contrast, ~11% of mTmG pups were found (n=28) when the pKO male mice mated first (pKO V.S. mTmG) (Figure 2-3J and 3K). In summary, these results indicate that the sperm fitness was highly compromised in the pKO mice.

X linked *miR-506* family miRNAs target genes throughout spermatogenesis and compensate with each other

To identify the target genes of X linked *miR-506* family, we performed RNA-seq on the four KO mice testes (*miR-465* sKO, tKO, qKO and pKO) as well as the WT. To identify the direct targets of X linked *miR-506* family miRNAs, we overlapped the dysregulated genes with the predicted targets (dysregulated targets) across four different databases (i.e. TargetScan(46), microrna.org(47), miRWalk(48) and mirDB(49)) (Figure 2-S6A). TargetScan and miRWalk give the most overlapping genes, followed by microrna.org, whereas the mirDB gives the least dysregulated targets (Figure 2-S6A). To avoid bias among different algorisms, we choose the union of these four databases for the following analysis. These four KOs share a lot of dysregulated target genes, but also own their own dysregulated targets (Figure 2-4A). The *miR-465* sKO and qKO are mutually exclusive within the X linked *miR-506* family and have large datasets of dysregulated target genes, which allow us to perform pathway analysis. Interestingly, Ingenuity Pathway Analysis (IPA, QIAGEN) of the dysregulated target genes for *miR-465* KO and qKO showed

that these two KOs have distinct functions yet share some overlapping pathways (Figure 2-4B and 4C), for example, the ATM signaling, eIF4 and p70S6K signaling. Consistent with that the *miR-506* family is highly expressed in germ cells, these pathways are involved in spermatogonia stem cells (SSCs) self-renewal (e.g. MAPK (50, 51), PI3K-AKT pathway(51-53), and oxidative phosphorylation (54)), meiosis (ATM signaling (55), CHK proteins in cell cycle checkpoint control (55) and eIF4 and p70S6K Signaling (56)), as well as spermiogenesis (e.g. histamine, methionine, phenylalanine and aspartate degradation and cysteine biosynthesis(2)) (Figure 2-4B, and 4C). Interestingly, their functions seem to be complementary to each other, for example, although all belong to protein degradation, which plays an important role in degrading meiotic proteins, histones and nonessential organelles during spermatogenesis (57), the *miR-465* sKO shows upregulation of histamine degradation, whereas the qKO shows upregulation of methionine, phenylalanine and aspartate degradation (Figure 2-4B, 4C). Dysregulated targets in qKO also showed upregulated biogenesis of cysteine (Figure 2-4C), which is a key amino acid components of protamine (2). Further analysis of disease and function all showed that these pathways are highly responsible for DNA repair, reproductive development and early embryonic development (Figure 2-4D and 4E). Since the phenotype appears to be more severe when more X linked *miR-506* family miRNAs are KO, we reasoned that these miRNAs compensate with each other. To test this hypothesis, we performed small RNA-seq on the four KO testes. The sRNA sequencing confirmed that these miRNAs were no longer expressed in the corresponding KOs (Figure 2-4F). Interestingly, in *miR-465* sKO testes, *miR-*

201, miR-463, miR-471, *miR-741*, *miR-871*, *miR-883a,* and *miR-883b* are upregulated. Similarly, *miR-201, miR-470*, *miR-471*, *miR-871* and *miR-883b* are upregulated in the tKO testes, *miR-201, miR-465a, miR-465b, miR-465c*, and *miR-547* are upregulated in the qKO testes, and the *miR-201, miR-547* are upregulated in the pKO testes (Figure 2-4F). These results indicate that genetic compensation happens within the X linked *miR-506* family.

X linked *miR-506* family gains targets during evolution

When we use the union of the dysregulated targets in mice as a reference to overlapping with predicted targets of highly expressed X linked *miR-506* family members in humans and rats, the vast majority of the dysregulated targets in mice are shared with humans and rats (Figure 2-5A). These results suggest the coevolution of the X linked *miR-506* family with their targets.

There are several possibilities for the phenomenon that the targets are shared among species whereas the *miR-506* family is quickly evolved: 1. The 3'UTR sequences of the targets of the *miR-506* family are highly divergent so that the *miR-506* family needs to adapt their sequences to maintain the ability to target the shared targets. 2. The 3'UTR sequences of the targets remain conserved, whereas the *miR-506* family evolved fast to gain more targets. To test the possibilities, we first compared the conservation of the 3'UTR sequences of the targets (Figure 2-5B and 5C). We applied the phyloP scheme to measure the evolutionary conservation level at individual nucleotide sites in the 3'UTRs. Interestingly, the overall phyloP scores showed that the targets sites of X linked

miR-506 family are more conserved than the other regions, although some exceptions exist (Figure 2-5D and 5E), suggesting that regions targeted by X-linked *miR-506* are under relatively stronger purifying selection, instead of adaptive selection. This result leads us to the second hypothesis that the *miR-506* family evolved fast to gain more targets. We first compared all the potential targets, predicted by the aforementioned four databases, within the *miR-506* family among humans, mice, and rats. The overlapping numbers of targets of the pooled X-linked *miR-506* family among different species didn't show significant enrichment in humans (Figure 2-S6B). Then we compared the number of target genes per miRNAs. When comparing the number of target genes of all the miRNAs between humans and mice, we found that human miRNAs have more target genes than the counterpart of mouses ($p<0.05$, t-test) (Figure 2-5F), which suggests the higher biological complexity in humans. This phenomenon is even more obvious for the X-linked *miR-506* family ($p<0.05$, t-test) (Figure 2-5F). In humans, the X-linked *miR-506* family has more target genes compared with the human baseline (all the human miRNAs) ($p<0.05$, t-test) (Figure 2-5F). By contrast, in mice, we didn't observe a significant difference in target gene number between X-linked miRNAs and mice baseline (all the mouse miRNAs) (Figure 2-5F). These results indicate that the X-linked *miR-506* family gained more target genes during the evolution from mice to humans. We also investigated the target site numbers within individual target genes in both humans and mice, but no significant differences were found between humans and mice for the targets of the X-linked *miR-506*

family (Figure 2-5G), suggesting that the X-linked *miR-506* family didn't gain target sites in individual genes.

Taken together, these results suggest the human X-linked miRNAs have undergone stronger selective pressure and gained additional regulatory functions compared with mice, and these miRNAs gained more target genes instead of gaining more target sites within individual target genes.

Discussion

Coevolution of 3'UTR and miRNAs

We observed the dysregulated targets of the *miR-506* family were shared within humans, mice and rats despite their divergent sequences. Consistent with our conclusion, several studies have shown similar results for the *miR-506* family (21, 22). Interestingly, among all the shared targets, several key factors that play essential roles during meiosis prophase I. For example, *Brca1*, *Fmr1*, *Atm*, *Sycp2*, *Ago4*, *Meioc*, and *Ythdc2*. Mutation in *BRCA1* impaired DNA repair and homologs chromosomes crossover during spermatogenesis(58). *FMRP*, which was identified as a target of X linked *miR-506* family previously (28), was shown to function in DNA damage response by binding to chromatin during replication in spermatogenesis (59). ATM was shown to be responsible for activating recombination-dependent arrest with the accumulation of unrepaired DSBs(55). *SYCP2* and *SYCP3* heterodimers are fundamental components in forming axial/lateral elements, which form the synaptonemal complex together with

transverse filaments(60). Ablation of *Ago4* led to early meiosis in spermatogonia, aberrant sex body (silenced XY chromosome subdomain), and downregulated X linked *miR-506* family, including *miR-743a*, *miR-743b*, *miR-741*, and *miR-871*. Interestingly, *Ago3* and *Ago4* were both downregulated in the *miR-465* sKO and qKO, suggesting a positive feedback loop between these miRNAs and Ago proteins. *MEIOC* interacts with *YTHDC2*, and loss of *Meioc* or *Ythdc2* impaired meiosis (61-63). Although the exact underlining mechanism remains controversial, one study proposed that *MEIOC* stabilizes meiotic mRNAs at the beginning of meiotic prophase I, then *YTHDC2* promotes translation and degradation of these mRNAs at the end of the meiotic prophase I(63).

Since we performed RNA-seq, some potential target genes that were changed at the protein level but not at the RNA level may be missed out (28). A previous study showed that 3'UTR of some genes avoid miRNA regulation by shorting their 3'UTR length(64), suggesting that the 3' UTR of mRNA and miRNA are co-evolved. There may have some explanations for that some miRNAs still target the same mRNA even though the sequence of the miRNAs is changed. First, when we aligned the sequences of *miR-506* family, we observed that some U-to-C or A-to-G substitutions exist in the seed sequences within the *miR-506* family (Figure 2-1C, S4B). G and U can match with each other, therefore some U-to-C or A-to-G substitutions may also be able to target the same sequences. Secondly, besides the "seed sequence", which is located on the 5' end of miRNAs (2-8nt) and the most commonly used method for classifying family, some data showed that the 3'

end of miRNA can also target mRNA, termed as "motif" (65). Indeed, in the sequence alignment of *miR-506* (Figure 2-1C, S4B), it seems that sometimes the 3' end is more conserved.

Divergent roles of these miRNAs in modulating the same genes

To our surprise, the pKO tends to cluster with WT when compared to other KO groups (Figure 2-S6C). Although the majority function of miRNAs is to silence genes, several studies showed that miRNA could upregulate gene expression via interactions between FXR1 (fragile X mental retardation related protein 1) and AGO2 (8, 9). What's more, in all four AGO proteins (AGO1~4) in the human genome, AGO2 is the only protein that has the catalytic domain in cleaving mRNA(66, 67). In the overlapping dysregulated genes, we observed some genes have upregulation in some KO groups, but downregulation in the rest of the groups, and all of them (10 of the 10) are predicted targets of the *miR-506* family. These results suggesting that the miRNAs within this *miR-506* family may exert antagonistic rather than the exact same roles in modulating the same gene expression, so that the same gene could be regulated more precisely. Consistent with this theory, a recent paper showed that *miR-883a* promotes, while the rest within the *miR-506* family repress the FMRP expression (21), although the overall role of this family is to promote FMRP expression(28). What's more, an early study on a subfamily of this miRNA cluster showed that *miR-513* have divergent roles in modulating GNG13, DR1, and BTG3 (20).

ACKNOWLEDGEMENTS

This work was supported by grants from the NIH (HD071736, HD085506, and P30GM110767 to WY) and the Templeton Foundation (PID: 61174 to WY).

AUTHOR CONTRIBUTIONS

Z.W. and W.Y. designed the research. Z.W., Y. W., S. W., T. Z., S. C., H. W., Y.X., H.M., and H.Z., performed bench experiments. Z.W., T.Z., Y.X., K.J.P., and E. L. performed bioinformatics analysis. Z.W., and W.Y. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no conflict of interest.

Materials and Methods

Animal use and generation of global knockout mice

The animal protocol for this study was approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Nevada, Reno. All mice were housed and maintained under specific pathogen-free conditions with a temperature- and humidity-controlled animal facility in the Department of Lab Animal Medicine, University of Nevada, Reno. Generation of global KO mice was performed as previously described (28). Briefly, 4-6 weeks of FVB/NJ female mice were super-ovulated and mated with C57BL/6J stud males; fertilized eggs were collected from the oviducts. Cas9 mRNA (200ng/μl) and gRNAs (100 ng/μl) were mixed and injected into the cytoplasm or pronucleus of the fertilized eggs in M2 medium (Millipore, Cat. MR-051-F). After injection, all zygotes were cultured for 1h

in KSOM+AA medium (Millipore, Cat. MR-121-D) at 37°C under 5% CO_2 in the air before being transferred into 7-10 week-old female CD1 foster mothers.

In vitro fertilization (IVF)

C57BL/6J female were intraperitoneal (IP) injected with 7 IU pregnant mare serum gonadotropin (PMSG), the mice were treated with 7 IU hCG 48h later. 14 h after hCG treatment, oocytes were collected from the ampulla, and cumulus cells surrounding oocytes were removed by treatment with 1.5 mg/ml bovine testicular hyaluronidase (Sigma, Cat. H3506) in M2 (Millipore, Cat# MR-015-D) at 37°C for 2 min. The cumulus-free oocytes were washed and kept in equilibrated HTF (Millipore, Cat# MR-070-D) with a ratio of 20-30 oocytes /60 µl HTF at 37°C in air containing 5% CO_2 incubator prior to IVF. One side of cauda epididymal sperm were collected in 100 µl of equilibrated HTF medium, allowing spermatozoa to capacitate for ~30 min at 37°C in air containing 5% CO_2 incubator before IVF. After capacitation, 2 µl spermatozoa were 10-fold diluted and analyzed by computer-assisted sperm analysis (CASA). Based on the sperm concentration, a total of 2.5 x 10^8 spermatozoa was added into each 60µl HTF-oocytes drop for IVF. 4 h later, zygotes were washed and cultured in KSOM+AA (Millipore, Cat. MR-121-D) until blastocyst stage at 37°C in air containing 5% CO_2 incubator. 2-cells embryos were counted 24-26 h after IVF, blastocysts were counted and analyzed under fluorescence microscope 70-72 h after IVF.

Serial mating

Serial mating experiment were carried out based on the ovulation time point (10-13h after hCG) as previously described (68). C57BL/6J females were IP injected with 7 IU PMSG at 8 p.m., followed with 7 IU hCG 48h later. After hCG, every female was introduced to the first male mice from 8 p.m. to 6 a.m. the next day. After checking the plug at 6 a.m., female with the plug was introduced to the second male mice. Then, plugs in females were checked every 30min. Females with two times of plugs were kept until delivery. Tail snips of 5 days postnatal pups were collected, and analyzed under a fluorescence microscope and by PCR genotyping.

Mouse genotyping

Mouse tail or ear snips were lysed in a lysis buffer (40mM NaOH, 0.2mM EDTA, pH=12) for 1h at 95°C, followed by neutralization with the same volume of neutralizing buffer (40mM Tris-HCl, pH=5). PCR reactions were conducted using the 2×GoTaq Green master mix (Promega, Cat. M7123). The primers used for genotyping are the same as previously described (28).

Purification of germ cells

Germ cells were purified using STA-PUT as previously described (1). Pachytene and round cells were purified from adult CD1 and C57BL/6J mice, and Sprague Dawley rat.

RNA extraction, libraries construction and qPCR analyses

RNA of WT and KO testes was extracted using the mirVana™ miRNA Isolation Kit (Thermo Fisher Scientific, Cat. AM1561) following the manufacturer's instruction. Large RNA (>200nt) and small RNA (<200nt) were isolated separately for libraries' construction or qPCR validation. Small RNA libraries were constructed using NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) (NEB, Cat. E7330L) following manufacturer's instruction, and sequenced using HiSeq 2500 system for single-end 50bp sequencing. Large RNA libraries were constructed using KAPA Stranded RNA-Seq Kits with RiboErase (Roche, Cat. KK8483) according to the manufacturer's instruction, and sequenced using Nextseq 500 with paired-end 75bp sequencing. For generating cDNAs for large RNA, reverse transcription was performed using SuperScript™ II Reverse Transcriptase (Thermo Fisher Scientific, Cat.18064014). qPCR analyses for large RNA were then conducted using Fast SYBR® Green Master Mix (Thermo Fisher Scientific, Cat.4385616). Large RNA expression was normalized to Gapdh.

Large and small RNA-Seq data analysis

For the large RNA-seq data, we applied the Sailfish pipeline (69) to quantify the mRNA expression from the raw sequencing data, using the Ensembl (70) mouse gene annotation (mm10). Transcript per million reads (TPM) was used as the unit of gene expression level. For the small RNA-seq data, we applied the AASRA (71) for mice and rats, or SPORTS1.0 (72) for humans, monkeys and horses to parse the raw sequencing data. The clean reads were mapped against miRbase (73).

The DESeq2 (74) (for mice and rats) or edgeR (75) algorithm (for humans, monkeys and horses) was used to compare the groupwise RNA expression pattern. The trimmed mean of M-values (TMM (76), for mice and rats) or counts per million (CPM, for humans, monkeys and horses) algorithm was applied for reads count normalization and effective library size estimation. Groupwise differential expression was estimated by the likelihood ratio test and the RNAs with a false discovery rate < 5% were deemed differentially expressed.

Statistical analyses

Data are presented as mean ± SEM, and statistical differences between datasets were assessed by two samples t-test unless stated otherwise. $p < 0.05$, 0.01, and 0.001 are considered as statistically significant and indicated with *, **, and ***, respectively.

References

1. Bellve AR, et al. (1977) Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. J Cell Biol 74(1):68-85.
2. Turner JM (2007) Meiotic sex chromosome inactivation. Development 134(10):1823-1831.
3. Nguyen TA, et al. (2015) Functional Anatomy of the Human Microprocessor. Cell 161(6):1374-1387.
4. Daugaard I & Hansen TB (2017) Biogenesis and Function of Ago-Associated RNAs. Trends Genet 33(3):208-219.
5. Krol J, Loedige I, & Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. Nature reviews. Genetics 11(9):597-610.
6. Lin S & Gregory RI (2015) MicroRNA biogenesis pathways in cancer. Nat Rev Cancer 15(6):321-333.
7. Pasquinelli AE (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nature reviews. Genetics 13(4):271-282.
8. Vasudevan S, Tong Y, & Steitz JA (2007) Switching from repression to activation: microRNAs can up-regulate translation. Science 318(5858):1931-1934.
9. Vasudevan S & Steitz JA (2007) AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. Cell 128(6):1105-1118.

10. Wu Q, et al. (2012) The RNase III enzyme DROSHA is essential for microRNA production and spermatogenesis. J Biol Chem 287(30):25173-25190.

11. Bao J, et al. (2012) MicroRNA-449 and microRNA-34b/c function redundantly in murine testes by targeting E2F transcription factor-retinoblastoma protein (E2F-pRb) pathway. J Biol Chem 287(26):21686-21698.

12. Wu J, et al. (2014) Two miRNA clusters, miR-34b/c and miR-449, are essential for normal brain development, motile ciliogenesis, and spermatogenesis. Proc Natl Acad Sci U S A 111(28):E2851-2857.

13. Song R, et al. (2014) miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110. Nature 510(7503):115-120.

14. Yuan S, et al. (2019) Motile cilia of the male reproductive system require miR-34/miR-449 for development and function to generate luminal turbulence. Proc Natl Acad Sci U S A 116(9):3584-3593.

15. Song R, et al. (2009) Many X-linked microRNAs escape meiotic sex chromosome inactivation. Nature genetics 41(4):488-493.

16. Sosa E, Flores L, Yan W, & McCarrey JR (2015) Escape of X-linked miRNA genes from meiotic sex chromosome inactivation. Development 142(21):3791-3800.

17. Zhang R, Peng Y, Wang W, & Su B (2007) Rapid evolution of an X-linked microRNA cluster in primates. Genome research 17(5):612-617.

18. Guo X, Su B, Zhou Z, & Sha J (2009) Rapid evolution of mammalian X-linked testis microRNAs. BMC Genomics 10:97.

19. Meunier J, et al. (2013) Birth and expression evolution of mammalian microRNA genes. Genome research 23(1):34-45.

20. Sun Z, Zhang Y, Zhang R, Qi X, & Su B (2013) Functional divergence of the rapidly evolving miR-513 subfamily in primates. BMC Evol Biol 13:255.

21. Ramaiah M, et al. (2019) A microRNA cluster in the Fragile-X region expressed during spermatogenesis targets FMR1. EMBO Rep 20(2).

22. Zhang F, et al. (2019) Evolution of an X-Linked miRNA Family Predominantly Expressed in Mammalian Male Germ Cells. Mol Biol Evol 36(4):663-678.

23. Ota H, Ito-Matsuoka Y, & Matsui Y (2019) Identification of the X-linked germ cell specific miRNAs (XmiRs) and their functions. PLoS One 14(2):e0211739.

24. Meisel RP & Connallon T (2013) The faster-X effect: integrating theory and data. Trends Genet 29(9):537-544.

25. Larson EL, et al. (2016) Contrasting Levels of Molecular Evolution on the Mouse X Chromosome. Genetics 203(4):1841-1857.

26. Kousathanas A, Halligan DL, & Keightley PD (2014) Faster-X adaptive protein evolution in house mice. Genetics 196(4):1131-1143.

27. Fromm B, et al. (2020) MirGeneDB 2.0: the metazoan microRNA complement. Nucleic acids research 48(D1):D1172.

28. Wang Z, et al. (2020) X-linked miR-506 family miRNAs promote FMRP expression in mouse spermatogonia. EMBO Rep 21(1):e49024.

29. Casper J, et al. (2018) The UCSC Genome Browser database: 2018 update. Nucleic acids research 46(D1):D762-D769.

30. Blanchette M, et al. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. Genome research 14(4):708-715.

31. Siepel A, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome research 15(8):1034-1050.

32. Cabanettes F & Klopp C (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6:e4958.

33. Fedoroff NV (2012) Presidential address. Transposable elements, epigenetics, and genome evolution. Science 338(6108):758-767.

34. Smalheiser NR & Torvik VI (2005) Mammalian microRNAs derived from genomic repeats. Trends Genet 21(6):322-326.

35. Wang Z, et al. (2019) Insertion of a chimeric retrotransposon sequence in mouse Axin1 locus causes metastable kinky tail phenotype. Mobile DNA 10:17.

36. Bennett EA, et al. (2008) Active Alu retrotransposons in the human genome. Genome research 18(12):1875-1883.

37. Sookdeo A, Hepp CM, McClure MA, & Boissinot S (2013) Revisiting the evolution of mouse LINE-1 in the genomic era. Mobile DNA 4(1):3.

38. Hirano T, et al. (2014) Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate. RNA 20(8):1223-1237.

39. Li B, et al. (2019) Identification of piRNAs and piRNA clusters in the testes of the Mongolian horse. Scientific reports 9(1):5022.

40. Schuster A, et al. (2016) SpermBase: A Database for Sperm-Borne RNA Contents. Biol Reprod 95(5):99.

41. Qing X, et al. (2017) Dysregulation of an X-linked primate-specific epididymal microRNA cluster in unexplained asthenozoospermia. Oncotarget 8(34):56839-56849.

42. Gainetdinov IV, et al. (2018) Assessment of piRNA biogenesis and function in testicular germ cell tumors and their precursor germ cell neoplasia in situ. BMC Cancer 18(1):20.

43. Xia B, et al. (2020) Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. Cell 180(2):248-262 e221.

44. Eberhard W (1996) Female control: sexual selection by cryptic female choice (Princeton University Press).

45. Muzumdar MD, Tasic B, Miyamichi K, Li L, & Luo L (2007) A global double-fluorescent Cre reporter mouse. Genesis 45(9):593-605.

46. Agarwal V, Bell GW, Nam JW, & Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. eLife 4.

47. Betel D, Koppal A, Agius P, Sander C, & Leslie C (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 11(8):R90.

48. Dweep H & Gretz N (2015) miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nat Methods 12(8):697.

49. Chen Y & Wang X (2020) miRDB: an online database for prediction of functional microRNA targets. Nucleic acids research 48(D1):D127-D131.

50. Ishii K, Kanatsu-Shinohara M, Toyokuni S, & Shinohara T (2012) FGF2 mediates mouse spermatogonial stem cell self-renewal via upregulation of Etv5 and Bcl6b through MAP2K1 activation. Development 139(10):1734-1743.

51. Morimoto H, et al. (2013) ROS are required for mouse spermatogonial stem cell self-renewal. Cell stem cell 12(6):774-786.

52. Oatley JM, Avarbock MR, & Brinster RL (2007) Glial cell line-derived neurotrophic factor regulation of genes essential for self-renewal of mouse spermatogonial stem cells is dependent on Src family kinase signaling. J Biol Chem 282(35):25842-25851.

53. Lee J, et al. (2007) Akt mediates self-renewal division of mouse spermatogonial stem cells. Development 134(10):1853-1859.

54. Hermann BP, et al. (2018) The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. Cell Rep 25(6):1650-1667 e1658.

55. Pacheco S, et al. (2015) The ATM signaling cascade promotes recombination-dependent pachytene arrest in mouse spermatocytes. PLoS genetics 11(3):e1005017.

56. Modzelewski AJ, Holmes RJ, Hilz S, Grimson A, & Cohen PE (2012) AGO4 regulates entry into meiosis and influences silencing of sex chromosomes in the male mouse germline. Dev Cell 23(2):251-264.

57. Zhang Q, Ji SY, Busayavalasa K, Shao J, & Yu C (2019) Meiosis I progression in spermatogenesis requires a type of testis-specific 20S core proteasome. Nat Commun 10(1):3387.

58. Xu X, Aprelikova O, Moens P, Deng CX, & Furth PA (2003) Impaired meiotic DNA-damage repair and lack of crossing-over during spermatogenesis in BRCA1 full-length isoform deficient mice. Development 130(9):2001-2012.

59. Alpatov R, et al. (2014) A chromatin-dependent role of the fragile X mental retardation protein FMRP in the DNA damage response. Cell 157(4):869-881.

60. Yang F, et al. (2006) Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis. J Cell Biol 173(4):497-507.

61. Abby E, et al. (2016) Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. Nat Commun 7:10324.

62. Soh YQS, et al. (2017) Meioc maintains an extended meiotic prophase I in mice. PLoS genetics 13(4):e1006704.

63. Hsu PJ, et al. (2017) Ythdc2 is an N(6)-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell research 27(9):1115-1127.

64. Stark A, Brennecke J, Bushati N, Russell RB, & Cohen SM (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell 123(6):1133-1146.

65. Helwak A, Kudla G, Dudnakova T, & Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153(3):654-665.

66. Meister G, et al. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15(2):185-197.

67. Liu J, et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science 305(5689):1437-1441.

68. Nagy A, Gertsenstein M, Vintersten K, & Behringer R (2003) Manipulating the mouse embryo: a laboratory manual (Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY:).

69. Patro R, Mount SM, & Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol 32(5):462-464.

70. Cunningham F, et al. (2015) Ensembl 2015. Nucleic Acids Res 43(Database issue):D662-669.

71. Tang C, Xie Y, & Yan W (2017) AASRA: An Anchor Alignment-Based Small RNA Annotation Pipeline. bioRxiv.

72. Shi J, Ko EA, Sanders KM, Chen Q, & Zhou T (2018) SPORTS1.0: A Tool for Annotating and Profiling Non-coding RNAs Optimized for rRNA- and tRNA-derived Small RNAs. Genomics Proteomics Bioinformatics 16(2):144-151.

73. Kozomara A & Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research 42(Database issue):D68-73.

74. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550.

75. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139-140.

76. Robinson MD & Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11(3):R25.

Figure 1

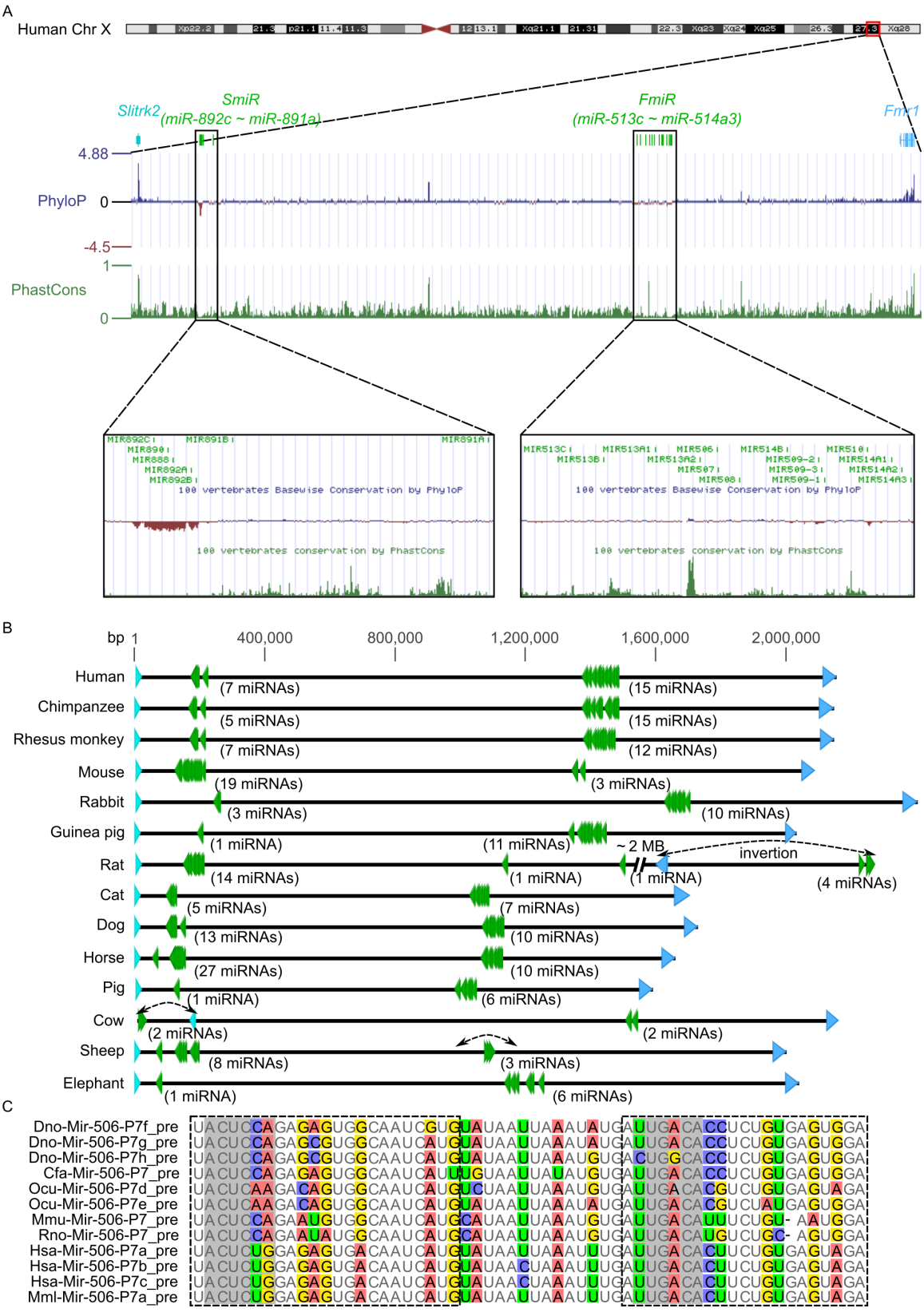
A
Human Chr X
Slitrk2
SmiR
(miR-892c ~ miR-891a)
FmiR
(miR-513c ~ miR-514a3)
Fmr1
4.88
PhyloP
0
-4.5
1
PhastCons
0
MIR892C
MIR891B
MIR891A
MIR890
MIR888
MIR892A
MIR892B
100 vertebrates Basewise Conservation by PhyloP
100 vertebrates conservation by PhastCons
MIR513C
MIR513B
MIR513A1
MIR513A2
MIR506
MIR507
MIR508
MIR514B
MIR509-2
MIR509-3
MIR509-1
MIR510
MIR514A1
MIR514A2
MIR514A3
100 vertebrates Basewise Conservation by PhyloP
100 vertebrates conservation by PhastCons
B
bp 1
400,000
800,000
1,200,000
1,600,000
2,000,000
Human
(7 miRNAs)
(15 miRNAs)
Chimpanzee
(5 miRNAs)
(15 miRNAs)
Rhesus monkey
(7 miRNAs)
(12 miRNAs)
Mouse
(19 miRNAs)
(3 miRNAs)
Rabbit
(3 miRNAs)
(10 miRNAs)
Guinea pig
(1 miRNA)
(11 miRNAs)
~2 MB
invertion
Rat
(14 miRNAs)
(1 miRNA)
(1 miRNA)
(4 miRNAs)
Cat
(5 miRNAs)
(7 miRNAs)
Dog
(13 miRNAs)
(10 miRNAs)
Horse
(27 miRNAs)
(10 miRNAs)
Pig
(1 miRNA)
(6 miRNAs)
Cow
(2 miRNAs)
(2 miRNAs)
Sheep
(8 miRNAs)
(3 miRNAs)
Elephant
(1 miRNA)
(6 miRNAs)
C
Dno-Mir-506-P7f_pre
Dno-Mir-506-P7g_pre
Dno-Mir-506-P7h_pre
Cfa-Mir-506-P7_pre
Ocu-Mir-506-P7d_pre
Ocu-Mir-506-P7e_pre
Mmu-Mir-506-P7_pre
Rno-Mir-506-P7_pre
Hsa-Mir-506-P7a_pre
Hsa-Mir-506-P7b_pre
Hsa-Mir-506-P7c_pre
Mml-Mir-506-P7a_pre

Figure 2-1. Evolution divergence, genomic location and sequence alignment of X linked *miR-506* family.

A. Evolution divergence of X linked *miR-506* family based on Multiz Alignment & Conservation using the human genome as a reference. The miRNA clusters nearby Slitrk2 and Fmr1 were named as SmiR and FmiR, respectively. Positive PhyloP scores indicate conservation and vice versa. PhastCons has a score between 0~1, the higher score the DNA region has, the more conserve the DNA region is. B. The genomic location of X linked *miR-506* family. *Slitrk2* and *Fmr1* are shown in cyan and blue, respectively, X linked *miR-506* family miRNAs are indicated as green. Three head arrows indicate the inversion of the DNA region compared to the other genomic DNA. C. Alignment of precursor miRNA of miR-506-P7 subfamily across species. Bases in the dotted square represent the mature miRNA. Seed sequences are shown in the grey background, alignment with mismatches are indicated in bases with the colored background.

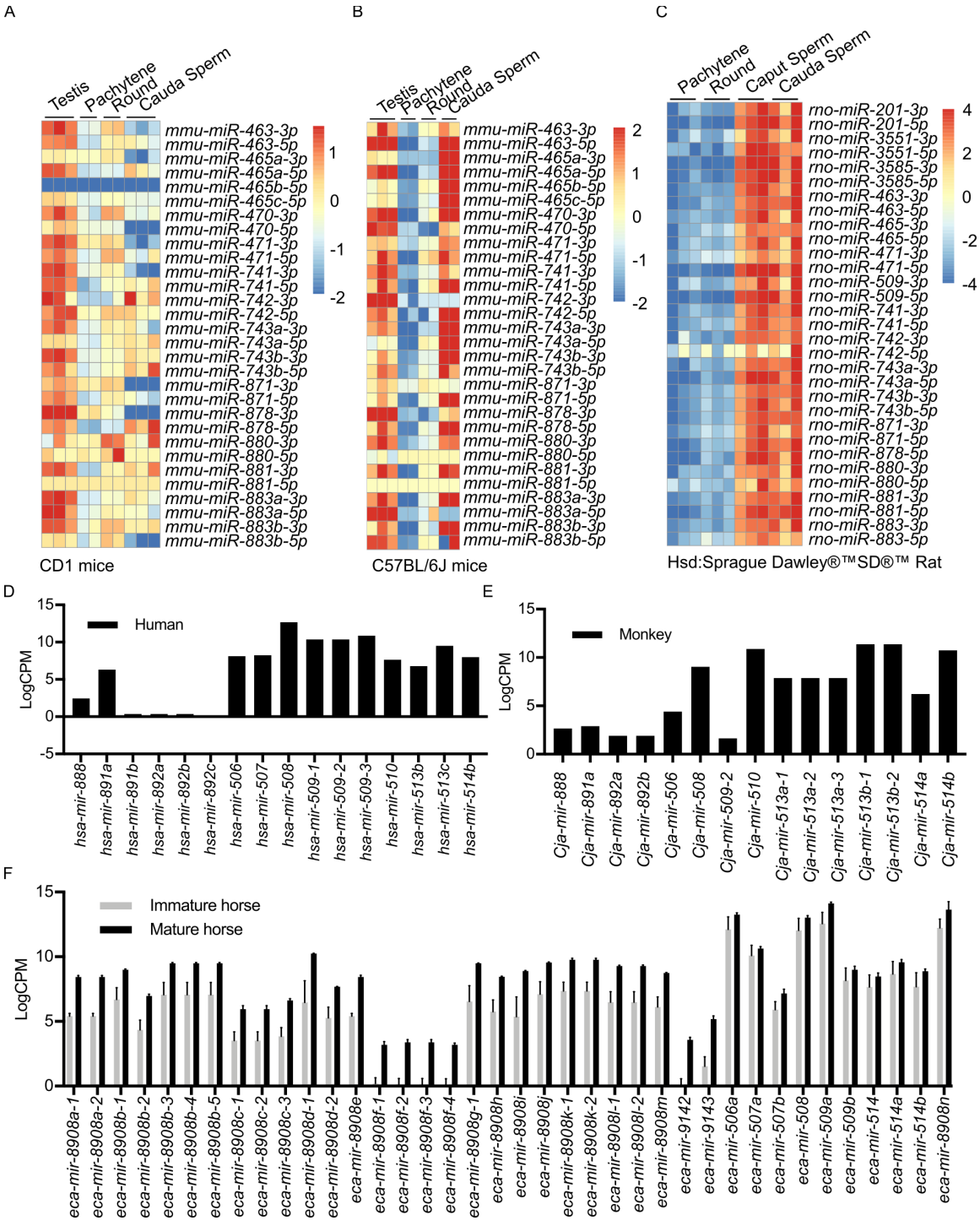

Figure 2-2. Expression of X linked *miR-506* family in mice, rats, humans, monkeys and horses.

A. Heatmap showing the *miR-506* family expression pattern in testis, pachytene spermatocytes, round spermatids and sperm in CD1 mice strain. B. Heatmap showing the *miR-506* family expression pattern in testis, pachytene spermatocytes, round spermatids and sperm in the C57BL/6J mice strain. C. Heatmap showing the *miR-506* family expression pattern in testis, pachytene spermatocytes, round spermatids and sperm in Sprague Dawley rat strain. D. LogCPM showing the *miR-506* family expression pattern in human testis. E. LogCPM showing the *miR-506* family expression pattern in marmoset monkey testis. F. LogCPM showing the *miR-506* family expression pattern in sexually immature and mature horse testis. n=3.

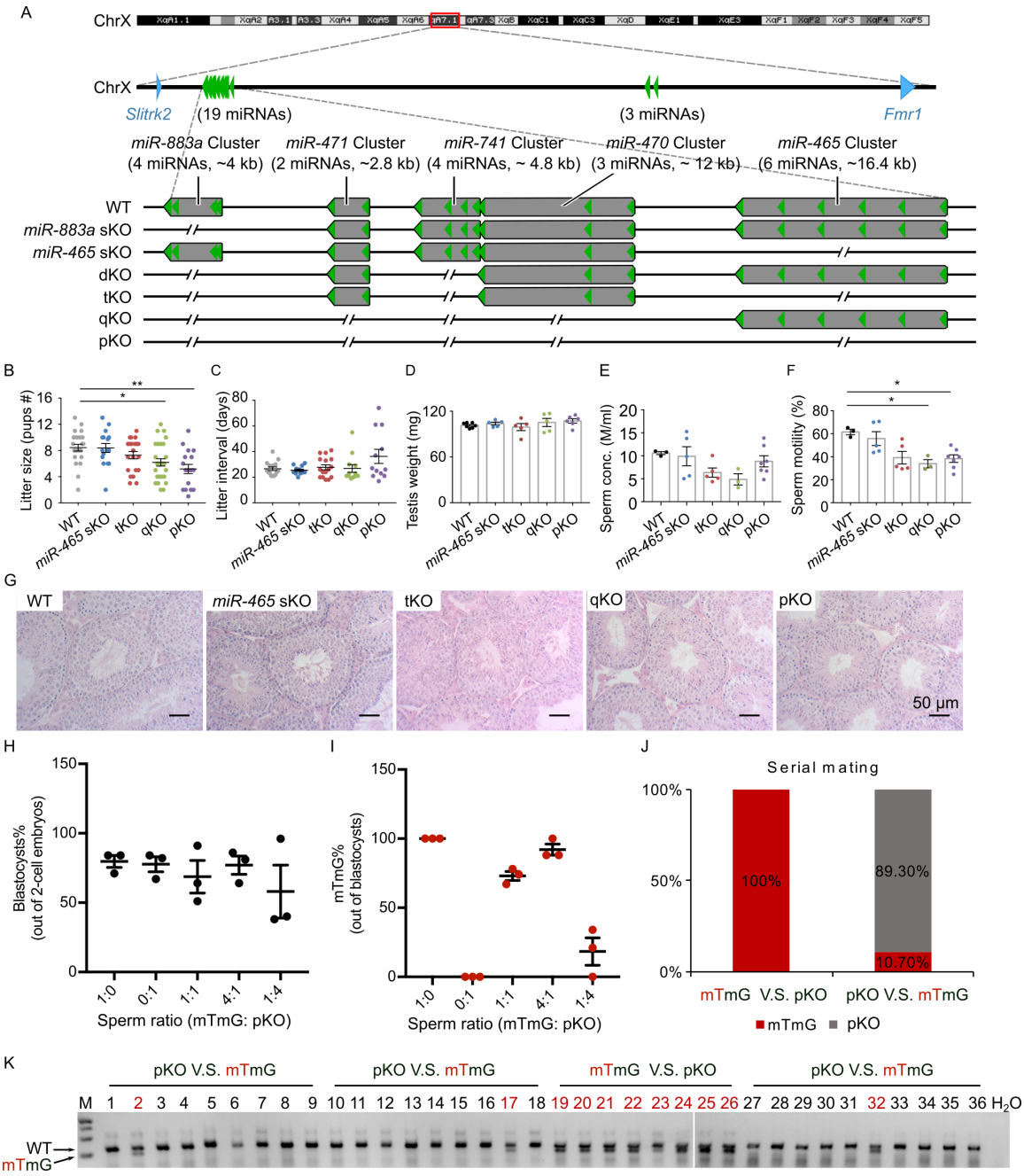

Figure 2-3. KO of X linked *miR-506* family compromised sperm fitness.

A. A cartoon shows the strategies in generating the miRNA clusters KO within the X linked *miR-506* family. B. Litter size of different KOs of X linked *miR-506* family. $p < 0.05$, 0.01, and 0.001 are considered as statistically significant and indicated

with *, **, and ***, respectively. C. Litter interval of different KOs of X linked *miR-506* family. D. Testis size of different KOs of X linked *miR-506* family. E. Sperm concentration of different KOs of X linked *miR-506* family. F. Sperm motility of different KOs of X linked *miR-506* family. $p < 0.05$ is considered as statistically significant and indicated with *. G. HE stains of different KOs of X linked *miR-506* family. H. IVF blastocysts rate of mTmG and pKO with different sperm ratio. Blastocysts were normalized to 2-cell embryos. I. mTmG percentage in blastocysts from IVF with different sperm ratio. mTmG were normalized to total blastocysts. J. Serial mating of mTmG and pKO. WT female mice were injected with HCG at 8 PM, then the first male mice were introduced with the female and replaced with the second male mice at 6 AM the next morning. mTmG V.S. pKO, mTmG male mice mate first; pKO V.S. mTmG, pKO male mice mate first. K. Genotyping results of serial mating. mTmG V.S. pKO, mTmG male mice mate first; pKO V.S. mTmG, pKO male mice mate first.

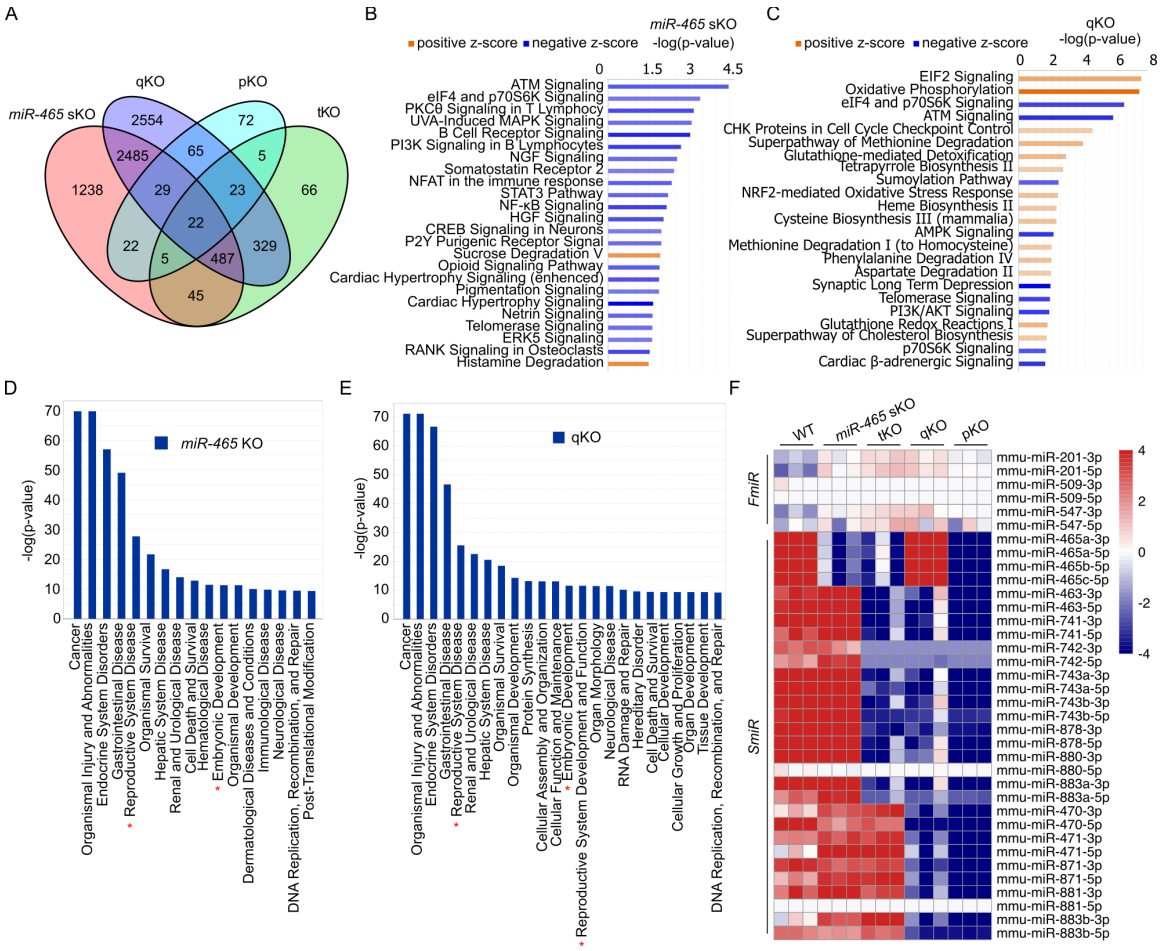

Figure 2-4. Target genes of X linked *miR-506* family. A. Overlap of dysregulated targets among different KOs in mice. B. IPA Canonical Pathways analysis of *miR-465* sKO testis. Similar pathways were shown in the same background color. C. IPA Canonical Pathways analysis of qKO testis. D. IPA diseases and functions analysis of *miR-465* sKO testis. E. IPA diseases and functions analysis of qKO testis. Reproductive and embryonic development were indicated in red stars. F. Small RNA-seq of *miR-465* sKO, dKO, tKO, qKO, and pKO.

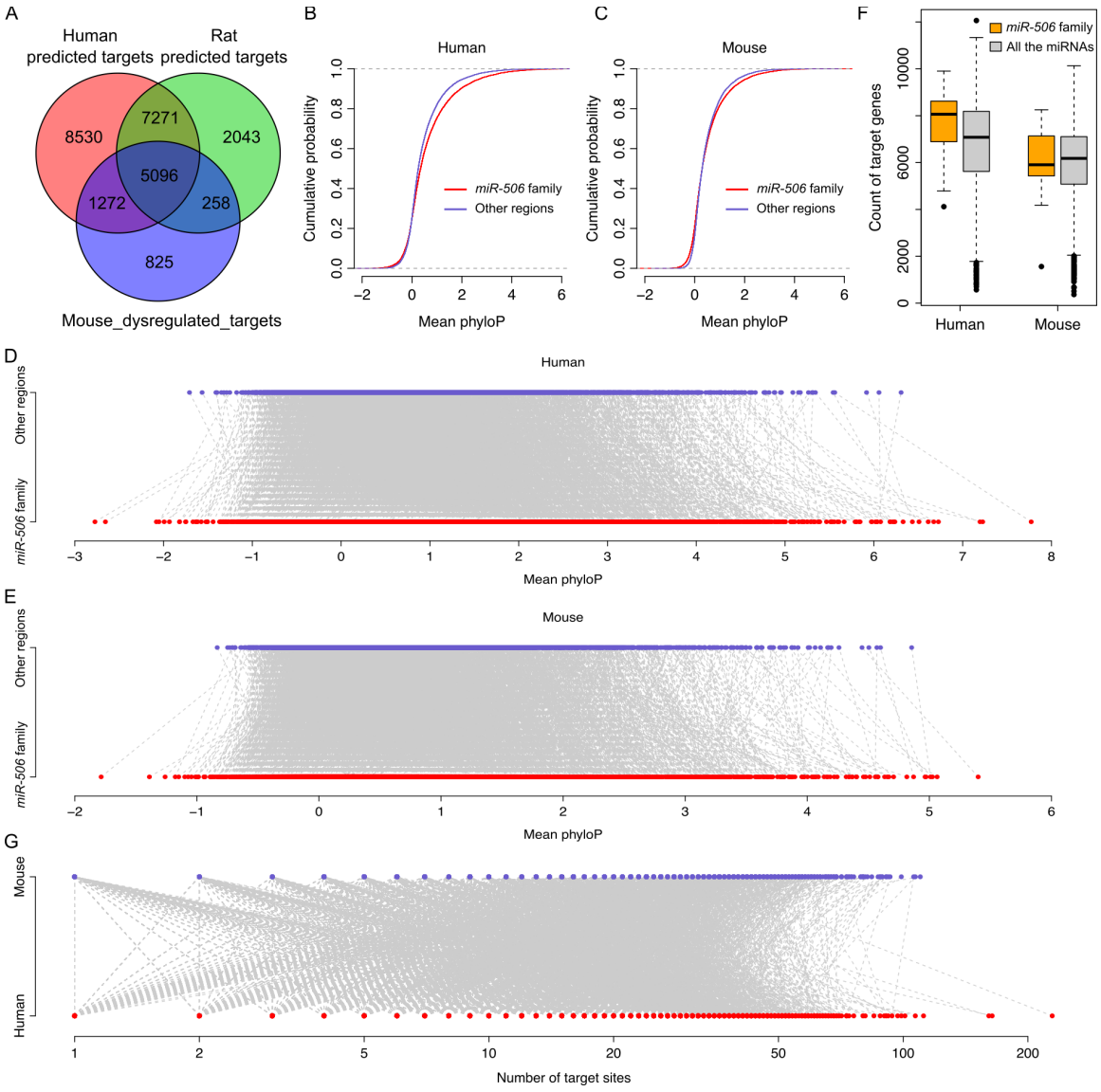

Figure 2-5. Targets of X linked *miR-506* family are conserved among different species. A. Overlap of dysregulated targets in mice with predicted targets of humans and rats. B. Comparison in the cumulative distribution of the *miR-506* family target sites and the other regions in humans. C. Comparison in the cumulative distribution of the *miR-506* family target sites and the other regions in mouse. D. Paired comparison in the phyloP score of the *miR-506* family target sites and the other regions in humans. E. Paired comparison in the phyloP score

of the *miR-506* family target sites and the other regions in mouse. F. Comparison of target numbers per miRNA for the X-linked *miR-506* family and all the miRNAs (baseline) in mice and humans. G. The target site numbers within individual target genes in both humans and mice.


SmiR
FmiR
100 kb
SLITRK2
SLITRK2
MIR892C
MIR891A
MIR890
MIR888
MIR892A
MIR892B
MIR891B
MIR513C
MIR513A1
MIR506
MIR514B
MIR510
MIR513B
MIR513A2
MIR509-2
MIR514A1
MIR507
MIR509-3
MIR514A2
MIR508
MIR509-1
MIR514A3
4.88
100 vertebrates Basewise Conservation by PhyloP
Cons 100 Verts
0
-4.5
1
100 vertebrates conservation by PhastCons
Cons 100 Verts
0
Multiz Alignments of 100 Vertebrates
Chimp
Gorilla
Orangutan
Gibbon
Rhesus
ab-eating_macaque
Marmoset
Squirrel_monkey
Bushbaby
hinese_tree_shrew
Squirrel
r_Egyptian_jerboa
Prairie_vole
Chinese_hamster
Golden_hamster
Mouse
Naked_mole-rat
Guinea_pig
Chinchilla
Brush-tailed_rat
Rabbit
Pika
Pig
Alpaca
Bactrian_camel
Dolphin
Killer_whale
Tibetan_antelope
Sheep
Domestic_goat
White_rhinoceros
Dog
Ferret_
Panda
Pacific_walrus
Weddell_seal
Black_flying-fox
Megabat
id's_myotis_(bat)
Microbat
Big_brown_bat
Hedgehog
Shrew
Star-nosed_mole
Elephant
pe_elephant_shrew
Manatee
Cape_golden_mole
Tenrec
Aardvark
Armadillo
Opossum
Tasmanian_devil
Wallaby
Platypus
Saker_falcon
Peregrine_falcon
llared_flycatcher
-throated_sparrow
dium_ground_finch
Zebra_finch
ibetan_ground_jay
Budgerigar
Parrot
Scarlet_macaw
Rock_pigeon
Mallard_duck
Chicken
Turkey
merican_alligator
Green_seaturtle
_softshell_turtle
_softshell_turtle
Lizard
X_tropicalis
Coelacanth
Tetraodon
Fugu
wbelly_pufferfish
Nile_tilapia
incess_of_Burundi
on's_mouthbreeder
Zebra_mbuna
ndamilia_nyererei
Medaka
outhern_platyfish
Stickleback
Atlantic_cod
Zebrafish
tetra(cavefish)
Spotted_gar
Lamprey
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Repeating Elements by RepeatMasker

Figure 2-S1. Multiz Alignment & Conservation analysis of human X linked *miR-506* family across 100 species.

The human genome is used as a reference. MiRNAs are in pink background.

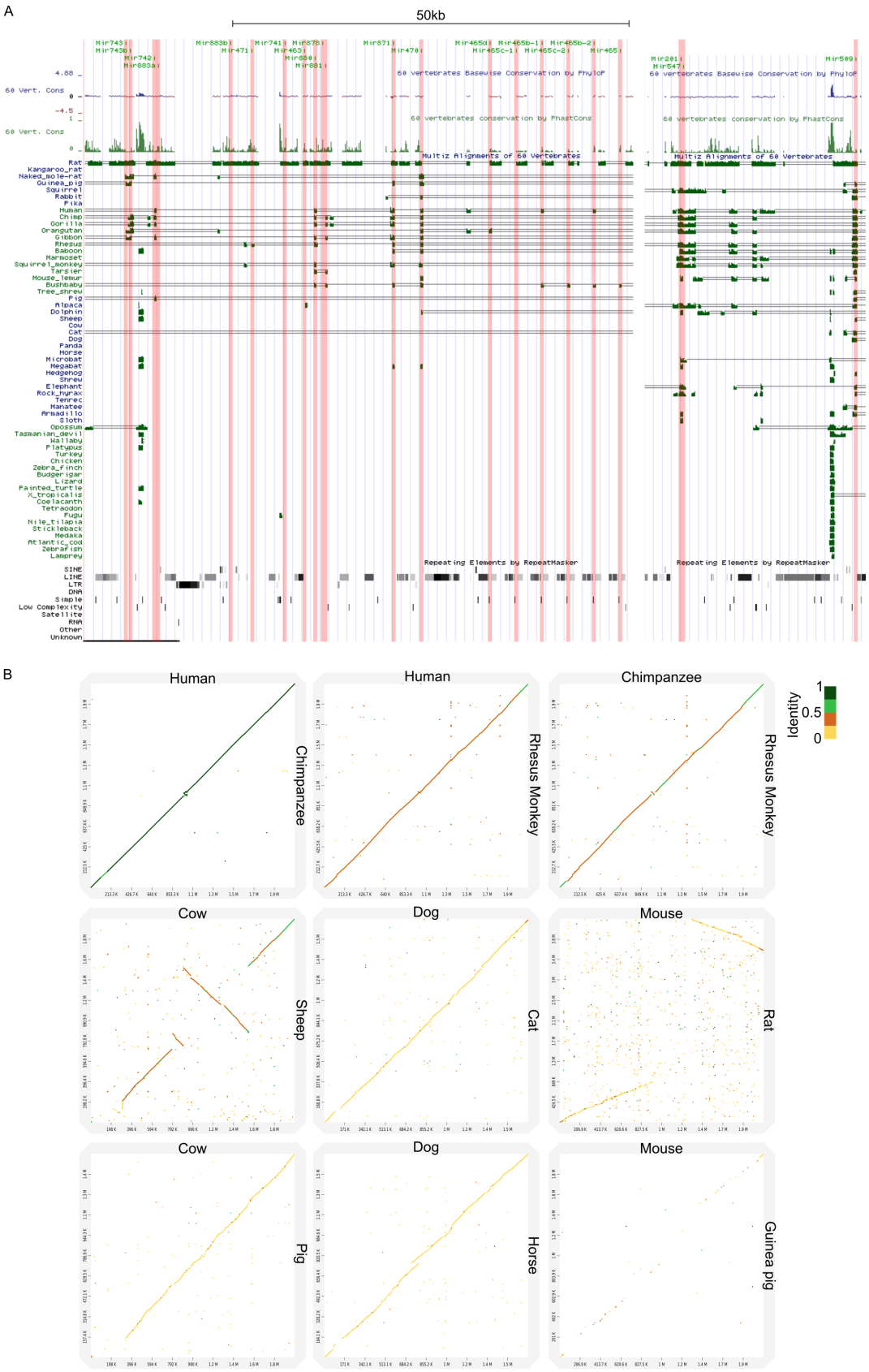
A
50kb
60 vertebrates Basewise Conservation by PhyloP
60 vertebrates conservation by PhastCons
Multiz Alignments of 60 Vertebrates
Repeating Elements by RepeatMasker
B
Human
Chimpanzee
Rhesus Monkey
Cow
Sheep
Dog
Cat
Mouse
Rat
Pig
Horse
Guinea pig
Identity
1
0.5
0

Figure 2-S2. Evolution of X linked *miR-506* family based on Multiz Alignment & Conservation and Dot plot analysis.

A. Detailed evolution divergence of X linked *miR-506* family across species based on Multiz Alignment & Conservation using mouse genome as a reference. miRNAs are in pink background. B. Dot plot analysis of different species.

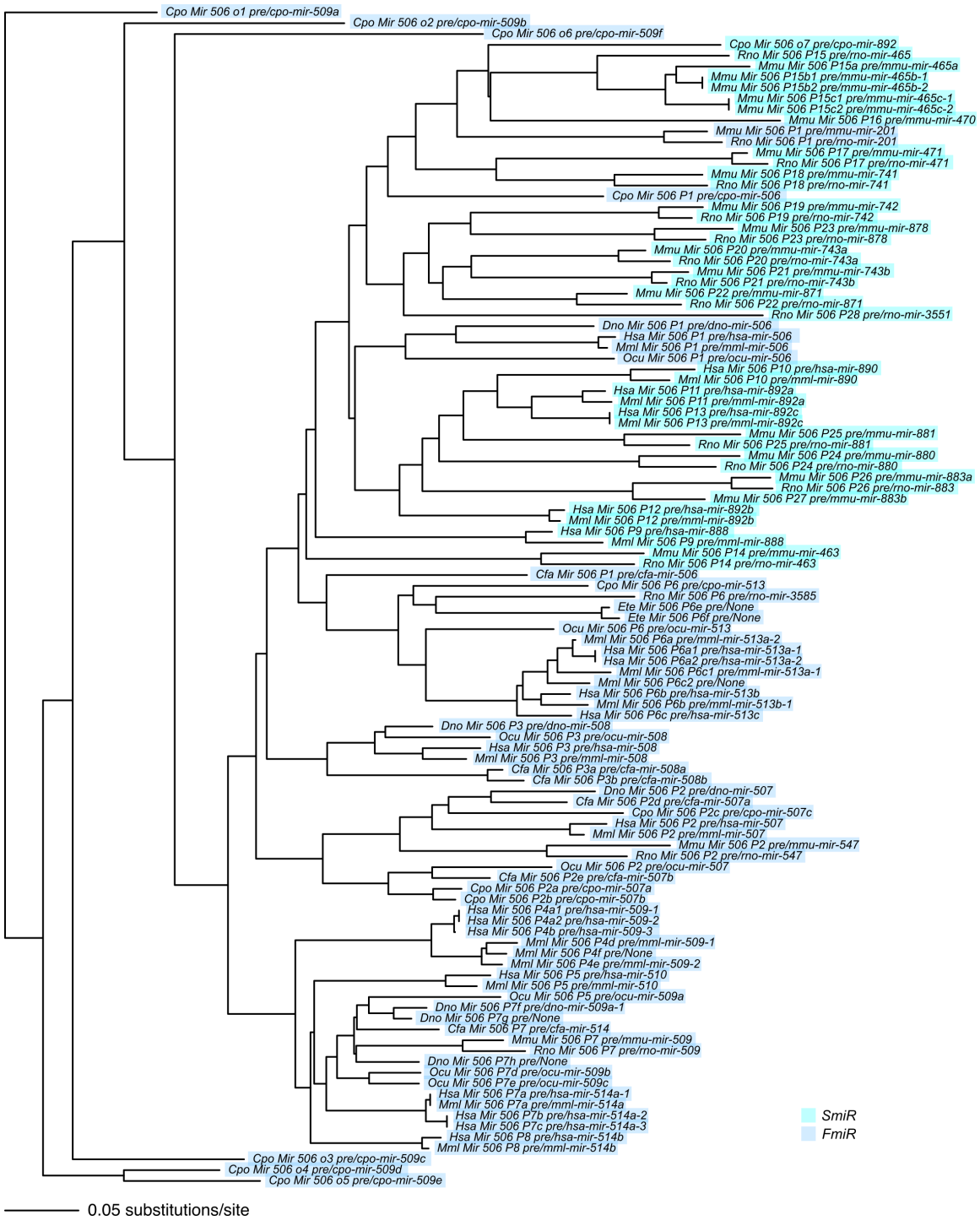

Figure 2-S3. Phylogram tree of X linked *miR-506* family.

SmiR and FmiR are shown in cyan and blue, respectively.

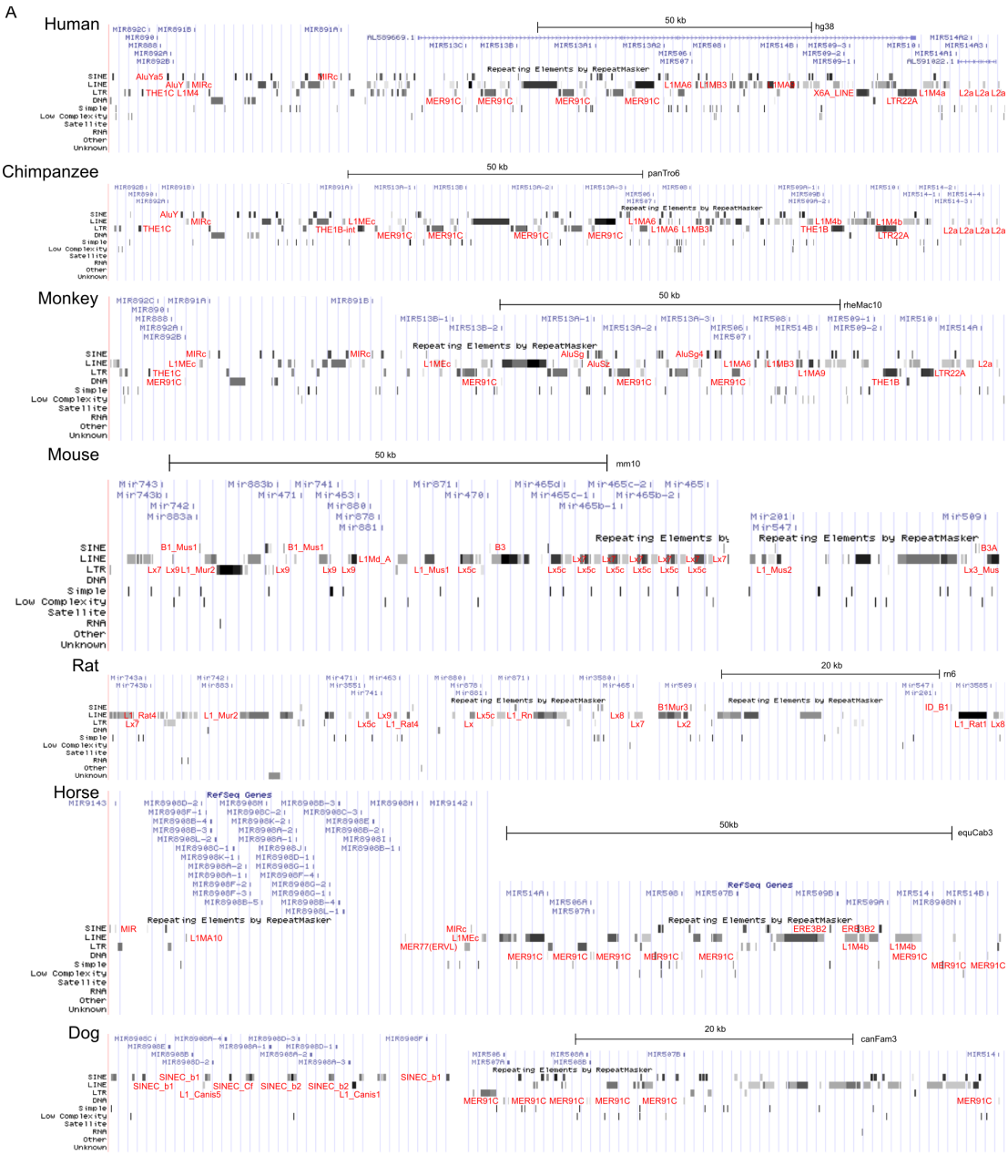
A
Human
50 kb
hg38
Repeating Elements by RepeatMasker
Chimpanzee
50 kb
panTro6
Repeating Elements by RepeatMasker
Monkey
50 kb
rheMac10
Repeating Elements by RepeatMasker
Mouse
50 kb
mm10
Repeating Elements by RepeatMasker
Rat
20 kb
rn6
Repeating Elements by RepeatMasker
Horse
RefSeq Genes
50kb
equCab3
Repeating Elements by RepeatMasker
Dog
20 kb
canFam3
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
B
Identity
hsa-mir-891a
mml-mir-891b
ppy-mir-891a
ptr-mir-891a
hsa-mir-891b
mml-mir-891a
ppy-mir-891b
ptr-mir-891b
Identity
hsa-mir-892a
mml-mir-892a
ppy-mir-892a
ptr-mir-892a
hsa-mir-892c
mml-mir-892c
ptr-mir-892b
hsa-mir-892b
mml-mir-892b
ppy-mir-892b

Figure 2-S4. Genetic composition of X linked *miR-506* family, and sequences alignment of precursor *miR-891* and *miR-892*.

A. Genetic composition of X linked *miR-506* family. The family of TEs are labeled in red. B. Upper panel, sequences alignment of precursor *miR-891* among primate. Lower panel, sequences alignment of precursor *miR-892* among primate. Bases in the dotted square represent the mature miRNA. Seed sequences are shown in the grey background, alignment with mismatches are indicated in bases with the colored background.

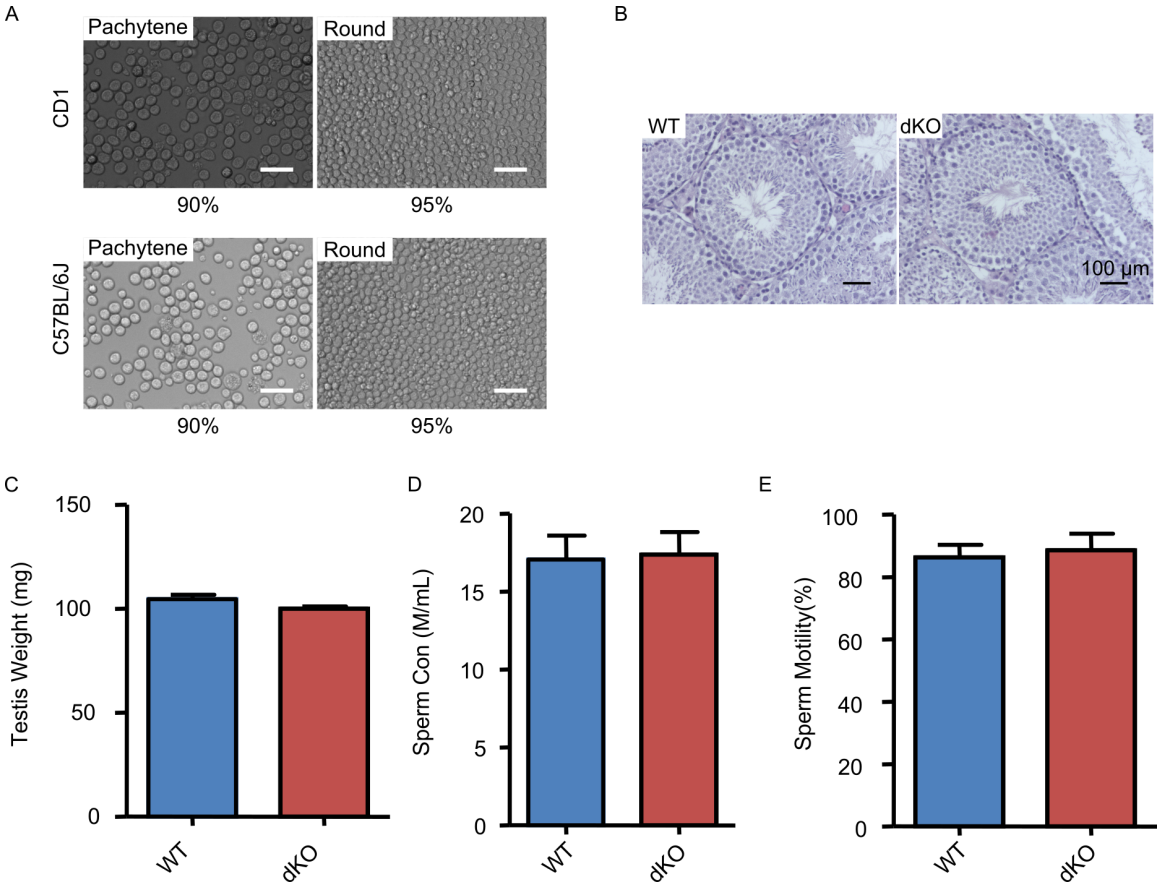

Figure 2-S5. The purity of germ cells after STA-PUT and the phenotype of dKO.

A. The purity of pachytene spermatocytes and round spermatids after STA-PUT cell sorting in CD1 mice (upper panel) and C57BL/6J (lower panel) mice. B. HE stains of WT and dKO testes. C. Testis weight of WT and dKO. D. Sperm concentration of WT and dKO. E. Sperm motility of WT and dKO.

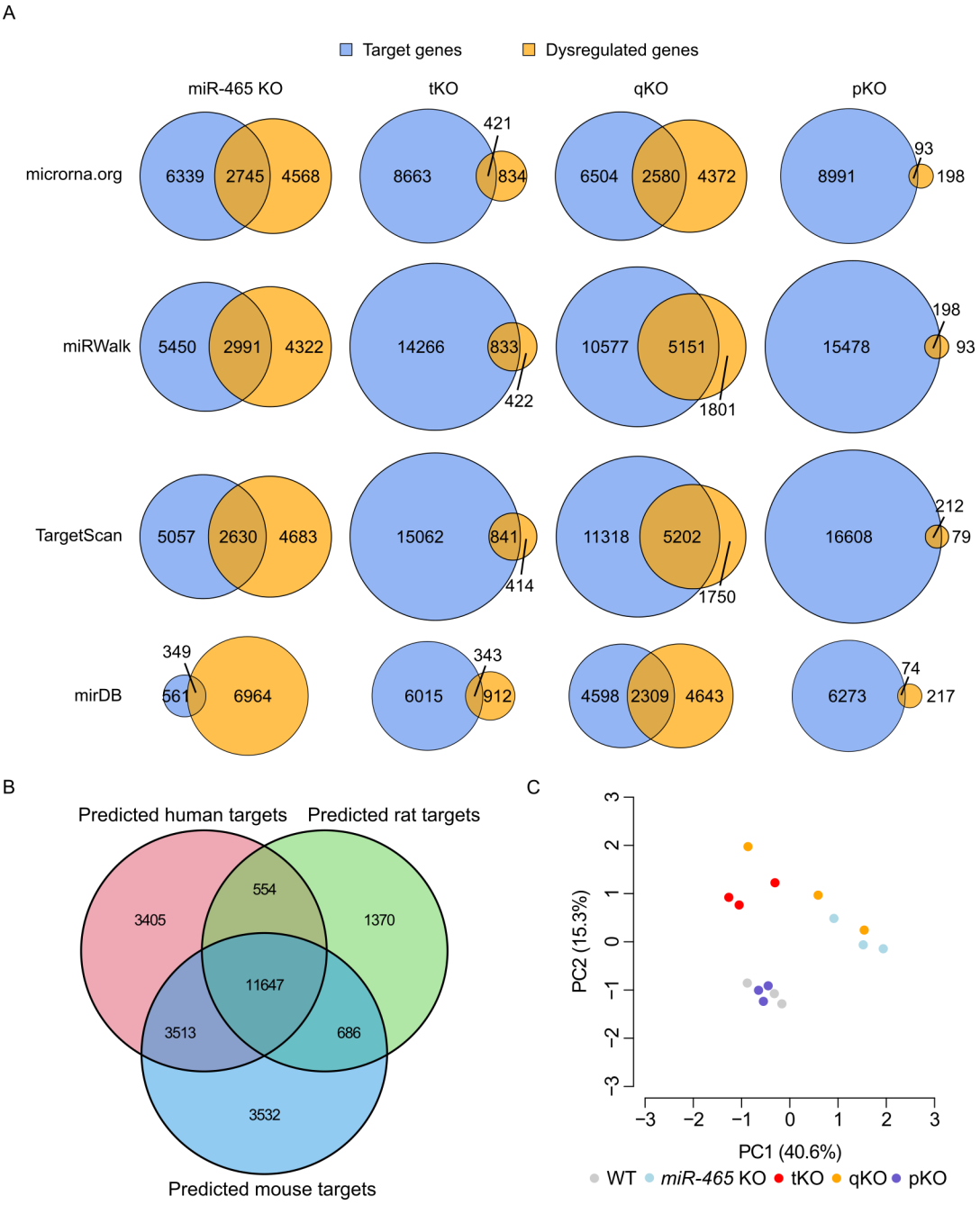

Figure 2-S6. Dysregulated targets and genetic compensation of X linked *miR-506* family.

A. Overlaps of dysregulated genes and predicted target genes from microrna.org, miRWalk, TargetScan, and mirDB four databases. B. Overlaps of predicted targets of all the X linked *miR-506* family from microrna.org, miRWalk, TargetScan, and mirDB databases. C. PCA analysis of different KOs in mice.

CHAPTER III. Loss of X linked *miR-465* cluster induces male-biased sex ratio

Zhuqing Wang[1], Shawn Wang[1], Nan Meng[2,3], Shuangbo Kong[2,3], Yue Wang[1], Tong Zhou[1], Huili Zheng[1], Haibin Wang[2,3], Wei Yan[1,4,5]

[1]Department of Physiology and Cell Biology, University of Nevada, Reno School of Medicine, Reno, NV 89557, USA; [2]Reproductive Medical Center, The First Affiliated Hospital of Xiamen University, 361003 Xiamen, Fujian, People's Republic of China; [3]Fujian Provincial Key Laboratory of Reproductive Health Research, School of Medicine Xiamen University, 361102 Xiamen, Fujian, People's Republic of China;[4]The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [5]Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

Correspondence:

Wei Yan M.D., Ph.D.

Professor

The Lundquist Institute at Harbor-UCLA Medical Center

David Geffen School of Medicine at UCLA

1124 W. Carson St., Torrance, CA 90502

Email: wei.yan@lundquist.org

Abstract

miRNA plays a pivotal role in early embryonic development, however, the role of miRNA in sexual dimorphism during early embryonic development is still in its infancy. Here we describe that *miR-465* cluster is preferentially expressed in extra-embryonic tissues at E7.5. Loss of *miR-465* cluster lead to female-biased lethality as early as embryonic day (E) 8.5. The female-biased lethality appears to result from upregulation of *Alkbh1*, a tRNA demethylation enzyme.

Introduction

Sexual dimorphism is differential characteristics between males and females within the same species, including physiology, morphology, and so on [1]. A recent study from 14,250 wild-type and 40,192 mutant mice demonstrated that 9.9% of categorical (qualitative) and 56.6% of continuous (quantitative) traits were sexual dimorphisms in wild-type mice, but most of the published data didn't take sex as a variable factor [2]. Sexual dimorphism happens as early as embryonic development. In most mammals, female contains two X chromosomes, while male contains one X chromosome and one Y chromosome, the female embryo undergoes X chromosome inactivation (XCI) to compensate dosage differences between males and females [3]. Female and male early embryos respond differently to the same stimulus, resulting in different sex ratios in some cases.

Recent studies have identified numerous sex-biased genes, including mRNAs, lncRNAs and microRNAs (miRNAs), that contribute to sexual dimorphism in the fetus, placenta as well as adult tissues [4-7]. As an important player in post-transcriptional regulation, ~22 nucleotides (nt) miRNA plays an essential role in early embryonic development [8-14]. After being transcribed by RNA polymerase II, primary miRNA (pri-miRNA) is cleaved by the microprocessor containing DROSHA-DGCR8 (DiGeorge syndrome critical region 8) to form ~70 nt hairpin precursor miRNA (pre-miRNA) in the nucleus [15]. Then the pre-miRNA is transported into the cytoplasm and further processed by DICER to form ~22 nt miRNA duplex. The miRNA duplex is unwound by Argonaute (Ago) proteins; one

strand of miRNA is loaded into miRNA induced silencing complex (miRISC) to target mRNA[8, 16, 17]. In most cases, miRNA downregulates translation by repressing translation or degrading mRNA, sometimes miRNA could also upregulate translation[8, 18]. The deletion of the miRNA biogenesis processors in mice led to early embryonic lethality, indicating an important role of miRNA in early embryonic development [9-12].

However, little is known regarding the miRNAs' function in sexual dimorphism during early embryonic development. One study showed that maternal obesity-induced upregulation of *miR-210* was mediated by NFκB1 in a sex-biased manner, leading to dysfunction of the placenta in the female[19]. However, the samples that they used were human placentas at term, when the placentas were already totally matured. In another study, the deletion of the *miR-290* cluster (miR-290~295) in mice leads to partially penetrant embryonic lethality, and *miR-290* cluster KO adult male mice are fertile while adult females are infertile[13]. A third study showed that the deletion of *miR-34b/c* and *miR-449* clusters lead to higher male mortality in mice at the perinatal stage[20]. Here we show that the ablation of the X linked *miR-465* cluster led to male-biased sex ratio due to the impaired placental formation in female embryos as early as E8.5.

Results

The X-linked *miR-465* cluster is preferentially expressed in extraembryonic tissue at E7.5.

Previous studies showed that the X-linked *miR-465* cluster is preferentially

expressed in testis, sperm, newborn ovary as well as in 8-16-cell embryos and blastocysts in mice [5, 21-25]. These results suggested that the *miR-465* cluster may be involved in spermatogenesis, early folliculogenesis and/or early embryonic development. When we compare the expression pattern of the *miR-465* between WT male and female enbryos during early embryonic development, we found that the *miR-465* cluster was highly expressed in extraembryonic tissues at E7.5 (Figure 3-1A), suggesting that this cluster escapes from iXCI and its important role in the formation of the placenta. Interestingly, *miR-465b-5p* has a higher abundance in the female extraembryonic tissues when compared with male extraembryonic tissues at E7.5 (Figure 3-1A), suggesting sexual dimorphism in *miR-465* expression. Although other *miR-465-5p* from this cluster also showed preferential expression in female extraembryonic tissues at E7.5, no significant differences were observed (Figure 3-1A). Using RNA in situ hybridization (ISH) with *miR-465* probes, we obtained similar results showing that the *miR-465*, including *miR-465b-5p*, *miR-465-3p*, and *miR-465d-3p*, are highly expressed in extraembryonic tissues, especially in the ectoplacental cone (EPC) (Figure 3-1B). Besides, we found that the *miR-465* is highly expressed in maternal decidua (Figure 3-1B). When comparing the *miR-465* sequences among different species, we found that the orthologs of the *miR-465* cluster, which are annotated as the *miR-892b* in humans, monkeys and chimpanzees in miRBase [26], are highly divergent among primate and rodent in the similar X chromosome locus (Figure 3-1C)[26, 27]. Since G-U wobble base pairing can also occur, these variations including some U-to-C or A-to-G substitutions may suggest a coevolution of the X

linked miRNAs or species-specific functions.

Ablation of *miR-465* cluster lead to male-biased sex ratio

To elucidate the function of *miR-465* cluster, we knocked out the *miR-465* cluster using CRISPR-Cas9, and generated three founders after microinjection (Figure 3-2A and 2B). PCR and Sanger sequencing showed this cluster is deleted precisely (Figure 3-2B and 2C). The *miR-465*$^{-/-}$ mice are fertile, and the testis size, litter size and litter interval of *miR-465*$^{-/-}$ mice were comparable to wild type (Figure 3-2D, 2E and 2F), suggesting this cluster is dispensable for spermatogenesis. Interestingly, the sex ratio (proportion of males) is highly skewed towards male in *miR-465* KO homozygous inbreeding (-/Y × -/-), when compared to wild-type inbreeding (+/Y × +/+) (Figure 3-3A).

The skewed sex ratio could result from either skewed X/Y sperm ratio or loss of female fetuses during embryonic development. We reasoned that if the sex ratio is already skewed in X/Y sperm stage, the bias should be observed in homozygous male (-/Y) outcross with wild-type females (+/+), rather than in homozygous female (-/-) outcross with wild-type male (+/Y). After outcrossing homozygous male or female with wild-type female or male, respectively, we found that the sex ratios of both breeding schemes are slightly, but not significantly, skewed (Figure 3-3A), suggesting that the significantly skewed sex ratio happened during early embryonic development. When collecting the KO early embryos, the sex ratio was still 51% at E3.5 and E7.5, but turned into 60% at E10.5 (Figure 3-3B). This further confirmed that the skewed sex ratio happened during early embryonic

development rather than prior to fertilization or in the postnatal day. When collecting E8.5-E10.5 embryos, some embryos were already reabsorbed (Figure 3-3C). Genotyping results of E9.5 showed that the majority of the reabsorbed embryos are females (Figure 3-3D). In E10.5, the reabsorbed embryos were too disintegrated to be separated from maternal decidua and were therefore not genotyped.

Loss of the *miR-465* cluster impaired female placental development.

Our qPCR data and breeding data showed that the *miR-465* cluster is preferentially expressed in extraembryonic tissues at E7.5, and some of the *miR-465*$^{-/-}$ embryos die during E8.5-E10.5, when the placenta forms[28]. These results suggest that the *miR-465* cluster is functional during the formation of the placenta. The extraembryonic tissues start from blastocyst stage (E3.5), chorioallantoic attachment occurs at E8.0, followed by villi formation during E8.0-E10.5[28]. The initial shape of the placenta is already formed at E10.5, while the placenta is not totally functional until E13.5[28]. Since the sex ratio is skewed during E8.5-E10.5, while not at E7.5, suggesting that the transcriptome and/or proteome are drastically changed at the E7.5 stage. Therefore, we used embryos and extraembryonic tissues at E7.5 both in WT and KO, including male and female, to determine the changes.

Principal Component Analysis (PCA) indicated that all embryos from WT males and females as well as KO males are clustered together, the same situation happened to the extraembryonic tissues (Figure 3-4A). By contrast, although most

of KO female embryos and extraembryonic tissues are clustered with WT embryos and KO male embryos (termed as normal-like embryos), some KO females didn't cluster with the rest of them (outlier embryos) (Figure 3-4A). Interestingly, when compared the extraembryonic tissue of outlier *miR-465*$^{-/-}$ female to the counterparts of normal-like *miR-465*$^{-/-}$ female, we observed some dysregulated genes related to sex bias and placental development, gene ontology (GO) term of the dysregulated genes also showed enrichment in extraembryonic development, ectoderm development and placental development (Figure 3-4B, 4C).

Among the dysregulated genes, two genes that influence placental development in a sex-dependent manner caught our eyes. Among the dysregulated genes, Rlim (also known as Rnf12) is an X linked gene that is responsible for imprinted XCI by maintaining Xist expression [29], and is highly expressed in extraembryonic tissues at E7.5 [30], and its dysregulation led to female-biased lethality[31, 32]. *Alkbh1* is a tRNA demethylation enzyme [33] that highly expressed in chorion and the ectoplacental cone at E8.5 [34], and its ablation induced female-biased lethality[35], whereas overexpression of *Alkbh1* in Hela cells inhibits translation elongation [33]. We performed luciferase assay to identify the direct target of *miR-465*. The *Rlim* is not the direct target of *miR-465* (Figure 3-4D), suggesting the dysfunction of the female placenta may not due to the impaired XCI. By contrast, the luciferase assay showed that *miR-465* directly target *Alkbh1* 3'UTR (Figure 3-4D). Consistent with this result, the *miR-465*$^{-/-}$ abnormal female showed upregulation of the *Alkbh1* when compared to the *miR-465*$^{-/-}$ normal female (Figure 3-4C).

Discussion

The male-biased sex ratio has been observed in humans, mice, and rats [32, 36-41], and has been associated with the increased proportion of single man, violent behavior, and trafficking of women in the human population[41, 42]. In humans, sex ratio bias towards males during natural fertilization has been reported in Asia and North Africa, with a range of 103-107 or more males per 100 females[40, 41]. Analysis of data from assisted reproductive technology (ART) showed that the sex ratio at conception is 50%, while the proportion of males increases during pregnancy due to overall higher female mortality [39]. In mice, IVF offspring showed a 57.2% male-biased sex ratio due to the loss of the female fetuses [32, 36]. The sex ratio of pups from mothers that were fed with high-fat and low-fat diets are 0.67 and 0.39, respectively [37]. In rats, mothers fed with high fructose diet before and during gestation have a skewed sex ratio of 0.6 in their offspring [38].

Despite the phenomena observed above, the molecular mechanisms underlying the skewed sex ratio is still in its infancy. One study showed that the loss of females was due to insufficient iXCI in the IVF induced skewed sex ratio[32]. Knock-down of Rnf12 in *in vivo* fertilization mimicked the IVF phenotype, while overexpression of Rnf12 or the treatment of the embryos with RA, which can induce Rnf12 expression, rescued the skewed sex ratio in IVF[32]. Another study showed that inflammation in the placenta caused by a deficiency of minichromosome maintenance complex (MCM) (e.g. $Mcm4^{C3/C3}$ $Mcm2^{Gt/+}$) led to the loss of female embryos since the males are protected from the anti-inflammation role of

testosterone [43, 44]. Treat the mice with testosterone or ibuprofen, a non-steroidal anti-inflammatory drug, rescued the skewed sex bias [43]. Here we show that loss of *miR-465* led to the the upregulation of *Alkbh1*, which was shown to catalyze m^1A demethylation in tRNA and inhibit translation elongation [33], and preferential female lethality as early as E8.5. Our study provides a new mechanism for the male-biased sex ratio.

Materials and Methods

Animal use and generation of global knockout mice

The animal protocol for this study was approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Nevada, Reno. All mice were housed and maintained under specific pathogen-free conditions with a temperature- and humidity-controlled animal facility in the Department of Lab Animal Medicine, University of Nevada, Reno. Generation of global KO mice was performed as previously described [25]. Briefly, 4-6 weeks of FVB/NJ female mice were super-ovulated and mated with C57BL/6J stud males; fertilized eggs were collected from the oviducts. Cas9 mRNA (200ng/μl) and gRNAs (100 ng/μl) were mixed and injected into the cytoplasm or pronucleus of the fertilized eggs in M2 medium (Millipore, Cat. MR-051-F). After injection, all zygotes were cultured for 1h in KSOM+AA medium (Millipore, Cat. MR-121-D) at 37°C under 5% CO_2 in the air before being transferred into 7-10-week-old female CD1 foster mothers.

Mouse genotyping

Mouse tail or ear snips were lysed in a lysis buffer (40mM NaOH, 0.2mM EDTA, pH=12) for 1h at 95°C, followed by neutralization with the same volume of neutralizing buffer (40mM Tris-HCl, pH=5). PCR reactions were conducted using the 2×GoTaq Green master mix (Promega, Cat. M7123). The primers used for genotyping are the same as previously described [25].

In situ hybridization

Frozen sections (10 μm) were adhered to poly-L-lysine-coated slides and fixed in 4 % paraformaldehyde (Sigma-Aldrich, Cat. NO. P6148) solution in PBS for 1 h at room temperature. Then, the sections were washed 3 times in PBS for 5 min, acetylated for 10 minutes (0.25% acetic anhydride), washed 2 times in PBS for 5 min, and hybridized with DIG-labeled probes overnight at 50°C. Hybridization buffer contained 1×salts (200 mM NaCl, 13 mM Tris, 5 mM sodium phosphate monobasic, 5mM sodium phosphate dibasic, 5 mM EDTA), 50% formamide, 10% (w/v) dextran sulfate, 1 mg/ml yeast tRNA (Roche, Cat. No. 10109509001), 1×Denhardt's [1% (w/v) bovine serum albumin, 1% (w/v) Ficoll, 1% (w/v) polyvinylpyrrolidone], and RNA probe (final concentration: 1 μM). Post-hybridization washes were followed by an RNase treatment (20 μg/ml RNase A). After blocking in 20% heat-inactivated sheep serum (Beijing Zhongshan Jinqiao Biotechnology Company, Cat. NO. ZLI-9021) and 2% blocking reagent (Roche, Cat. NO. 12039672910) for 1 h, sections were incubated overnight in blocking solution containing anti-DIG antibody (1:2500 dilution; Roche, Cat. No.

11093274910) at room temperature. After antibody washes, the color was developed using NBT/BCIP according to the manufacturer's instructions (Gentihold, NBT Cat. NO. N1332, BCIP Cat. NO. B1360). Sections were counterstained in Nuclear Fast Red (Solarbio, Cat. NO. G1321), dehydrated in gradient alcohol, cleared in xylene, and mounted in neutral resins.

RNA extraction, libraries construction, and qPCR analyses

RNA of WT and KO testes was extracted using the mirVana™ miRNA Isolation Kit (Thermo Fisher Scientific, Cat. AM1561) following the manufacturer's instruction. Large RNA (>200nt) and small RNA (<200nt) were isolated separately for libraries' construction or qPCR validation. Small RNA libraries were constructed using NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) (NEB, Cat. E7330L) following manufacturer's instruction, and sequenced using HiSeq 2500 system for single-end 50bp sequencing. Large RNA libraries were constructed using KAPA Stranded RNA-Seq Kits with RiboErase (Roche, Cat. KK8483) according to the manufacturer's instruction, and sequenced using NovaSeq SP with paired-end 50bp sequencing. For generating cDNAs for large RNA, reverse transcription was performed using SuperScript™ II Reverse Transcriptase (Thermo Fisher Scientific, Cat.18064014). qPCR analyses for large RNA were then conducted using Fast SYBR® Green Master Mix (Thermo Fisher Scientific, Cat.4385616). Large RNA expression was normalized to Gapdh.

Large and small RNA-Seq data analysis

For the large RNA-seq data, we applied the Sailfish pipeline [45] to quantify the mRNA expression from the raw sequencing data, using the Ensembl [46] mouse gene annotation (mm10). Transcript per million reads (TPM) was used as the unit of gene expression level. For the small RNA-seq data, we applied the SPORTS1.0 tool [47] to parse the raw sequencing data. The clean reads were mapped against miRbase [27]. The edgeR algorithm [48] was used to compare the groupwise RNA expression pattern. The TMM algorithm was applied for reads count normalization and effective library size estimation. Groupwise differential expression was estimated by the likelihood ratio test and the RNAs with a false discovery rate < 5% were deemed differentially expressed.

qPCR and Western Blot (WB).

The differentially expressed genes that fall into the predicted target genes were selected for qPCR and WB validation. WB was performed as previously described[32, 49].

Luciferase assay

The 3'UTR of target genes were amplified from mouse genomic DNA and inserted into multiple cloning site (MCS) of psiCHECK-2 plasmid (Promega) located downstream of the Renilla luciferase-coding sequence. *MiR-465* pre-miRNA along with ~300 bp of flanking genomic DNA were amplified from mouse genomic DNA and inserted into the pcDNA3.1 plasmid. HEK293 cells were co-transfected with psiCHECK-2-target containing 3′ UTR of target genes and pcDNA3.1-miR-465

using Lipofectamine 2000 (Thermo Fisher Scientific) in a 24 well cell culture plate (Corning). HEK293 cells co-transfected with psiCHECK-2-target containing 3′ UTR of target genes and pcDNA3.1-miRNA-NC were used as a negative control. 24 h later, cells will be lysed and assayed with Dual Luciferase Assay (Promega) according to the instructions of manufacturer. Renilla luciferase signal were normalized to Firefly luciferase signal to correct the transfection efficiency.

Statistical analyses

Data are presented as mean ± SEM, and statistical differences between datasets were assessed by two samples t-test unless stated otherwise. $p < 0.05$, 0.01, and 0.001 are considered as statistically significant and indicated with *, **, and ***, respectively.

References

1. Williams, T.M. and S.B. Carroll, Genetic and molecular insights into the development and evolution of sexual dimorphism. Nat Rev Genet, 2009. 10(11): p. 797-804.
2. Karp, N.A., et al., Prevalence of sexual dimorphism in mammalian phenotypic traits. Nat Commun, 2017. 8: p. 15475.
3. Lee, J.T. and M.S. Bartolomei, X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell, 2013. 152(6): p. 1308-23.
4. Buckberry, S., et al., Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal-maternal interface. Mol Hum Reprod, 2014. 20(8): p. 810-9.
5. Warnefors, M., et al., Sex-biased microRNA expression in mammals and birds reveals underlying regulatory mechanisms and a role in dosage compensation. Genome Res, 2017. 27(12): p. 1961-1973.
6. Capel, B., Vertebrate sex determination: evolutionary plasticity of a fundamental switch. Nat Rev Genet, 2017. 18(11): p. 675-689.
7. Lin, Y.T. and B. Capel, Cell fate commitment during mammalian sex determination. Curr Opin Genet Dev, 2015. 32: p. 144-52.
8. Lin, S. and R.I. Gregory, MicroRNA biogenesis pathways in cancer. Nat Rev Cancer, 2015. 15(6): p. 321-33.
9. Bernstein, E., et al., Dicer is essential for mouse development. Nat Genet, 2003. 35(3): p. 215-7.

10. Wang, Y., et al., DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet, 2007. 39(3): p. 380-5.
11. Alisch, R.S., et al., Argonaute2 is essential for mammalian gastrulation and proper mesoderm formation. PLoS Genet, 2007. 3(12): p. e227.
12. Morita, S., et al., One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. Genomics, 2007. 89(6): p. 687-96.
13. Medeiros, L.A., et al., Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. Proc Natl Acad Sci U S A, 2011. 108(34): p. 14163-8.
14. Parchem, R.J., et al., Two miRNA clusters reveal alternative paths in late-stage reprogramming. Cell Stem Cell, 2014. 14(5): p. 617-31.
15. Nguyen, T.A., et al., Functional Anatomy of the Human Microprocessor. Cell, 2015. 161(6): p. 1374-87.
16. Daugaard, I. and T.B. Hansen, Biogenesis and Function of Ago-Associated RNAs. Trends Genet, 2017. 33(3): p. 208-219.
17. Krol, J., I. Loedige, and W. Filipowicz, The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet, 2010. 11(9): p. 597-610.
18. Pasquinelli, A.E., MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat Rev Genet, 2012. 13(4): p. 271-82.

19. Muralimanoharan, S., et al., Sexual dimorphism in miR-210 expression and mitochondrial dysfunction in the placenta with maternal obesity. Int J Obes (Lond), 2015. 39(8): p. 1274-81.

20. Wu, J., et al., Two miRNA clusters, miR-34b/c and miR-449, are essential for normal brain development, motile ciliogenesis, and spermatogenesis. Proc Natl Acad Sci U S A, 2014. 111(28): p. E2851-7.

21. Song, R., et al., Many X-linked microRNAs escape meiotic sex chromosome inactivation. Nat Genet, 2009. 41(4): p. 488-93.

22. Ohnishi, Y., et al., Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. Nucleic Acids Res, 2010. 38(15): p. 5141-51.

23. Yuan, S., et al., Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. Development, 2016. 143(4): p. 635-47.

24. Ahn, H.W., et al., MicroRNA transcriptome in the newborn mouse ovaries determined by massive parallel sequencing. Mol Hum Reprod, 2010. 16(7): p. 463-71.

25. Wang, Z., et al., X-linked miR-506 family miRNAs promote FMRP expression in mouse spermatogonia. EMBO Rep, 2020. 21(1): p. e49024.

26. Fromm, B., et al., A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu Rev Genet, 2015. 49: p. 213-42.

27. Kozomara, A. and S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res, 2014. 42(Database issue): p. D68-73.

28. Watson, E.D. and J.C. Cross, Development of structures and transport functions in the mouse placenta. Physiology (Bethesda), 2005. 20: p. 180-93.

29. Wang, F., et al., Regulation of X-linked gene expression during early mouse development by Rlim. Elife, 2016. 5.

30. Shin, J., et al., RLIM is dispensable for X-chromosome inactivation in the mouse embryonic epiblast. Nature, 2014. 511(7507): p. 86-9.

31. Shin, J., et al., Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. Nature, 2010. 467(7318): p. 977-81.

32. Tan, K., et al., Impaired imprinted X chromosome inactivation is responsible for the skewed sex ratio following in vitro fertilization. Proc Natl Acad Sci U S A, 2016. 113(12): p. 3197-202.

33. Liu, F., et al., ALKBH1-Mediated tRNA Demethylation Regulates Translation. Cell, 2016. 167(3): p. 816-828 e16.

34. Pan, Z., et al., Impaired placental trophoblast lineage differentiation in Alkbh1(-/-) mice. Dev Dyn, 2008. 237(2): p. 316-27.

35. Nordstrand, L.M., et al., Mice lacking Alkbh1 display sex-ratio distortion and unilateral eye defects. PLoS One, 2010. 5(11): p. e13827.

36. Tan, K., et al., IVF affects embryonic development in a sex-biased manner in mice. Reproduction, 2016. 151(4): p. 443-53.

37. Rosenfeld, C.S., et al., Striking variation in the sex ratio of pups born to mice according to whether maternal diet is high in fat or carbohydrate. Proc Natl Acad Sci U S A, 2003. 100(8): p. 4628-32.

38. Gray, C., et al., Maternal fructose and/or salt intake and reproductive outcome in the rat: effects on growth, fertility, sex ratio, and birth order. Biol Reprod, 2013. 89(3): p. 51.

39. Orzack, S.H., et al., The human sex ratio from conception to birth. Proc Natl Acad Sci U S A, 2015. 112(16): p. E2102-11.

40. Zhu, W.X., L. Lu, and T. Hesketh, China's excess males, sex selective abortion, and one child policy: analysis of data from 2005 national intercensus survey. BMJ, 2009. 338: p. b1211.

41. Hesketh, T. and Z.W. Xing, Abnormal sex ratios in human populations: causes and consequences. Proc Natl Acad Sci U S A, 2006. 103(36): p. 13271-5.

42. Dyson, T., Causes and Consequences of Skewed Sex Ratios. Annual Review of Sociology, 2012. 38(1): p. 443-461.

43. McNairn, A.J., et al., Female-biased embryonic death from inflammation induced by genomic instability. Nature, 2019. 567(7746): p. 105-108.

44. Wang, Z. and W. Yan, Inflammation induced by faulty replication during embryonic development causes skewed sex ratio. Biol Reprod, 2019. 101(2): p. 259-261.

45. Patro, R., S.M. Mount, and C. Kingsford, Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol, 2014. 32(5): p. 462-4.

46. Cunningham, F., et al., Ensembl 2015. Nucleic Acids Res, 2015. 43(Database issue): p. D662-9.

47. Shi, J., et al., SPORTS1.0: A Tool for Annotating and Profiling Non-coding RNAs Optimized for rRNA- and tRNA-derived Small RNAs. Genomics Proteomics Bioinformatics, 2018. 16(2): p. 144-151.

48. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010. 26(1): p. 139-40.

49. Sui, L., et al., Dynamic proteomic profiles of in vivo- and in vitro-produced mouse postimplantation extraembryonic tissues and placentas. Biol Reprod, 2014. 91(6): p. 155.

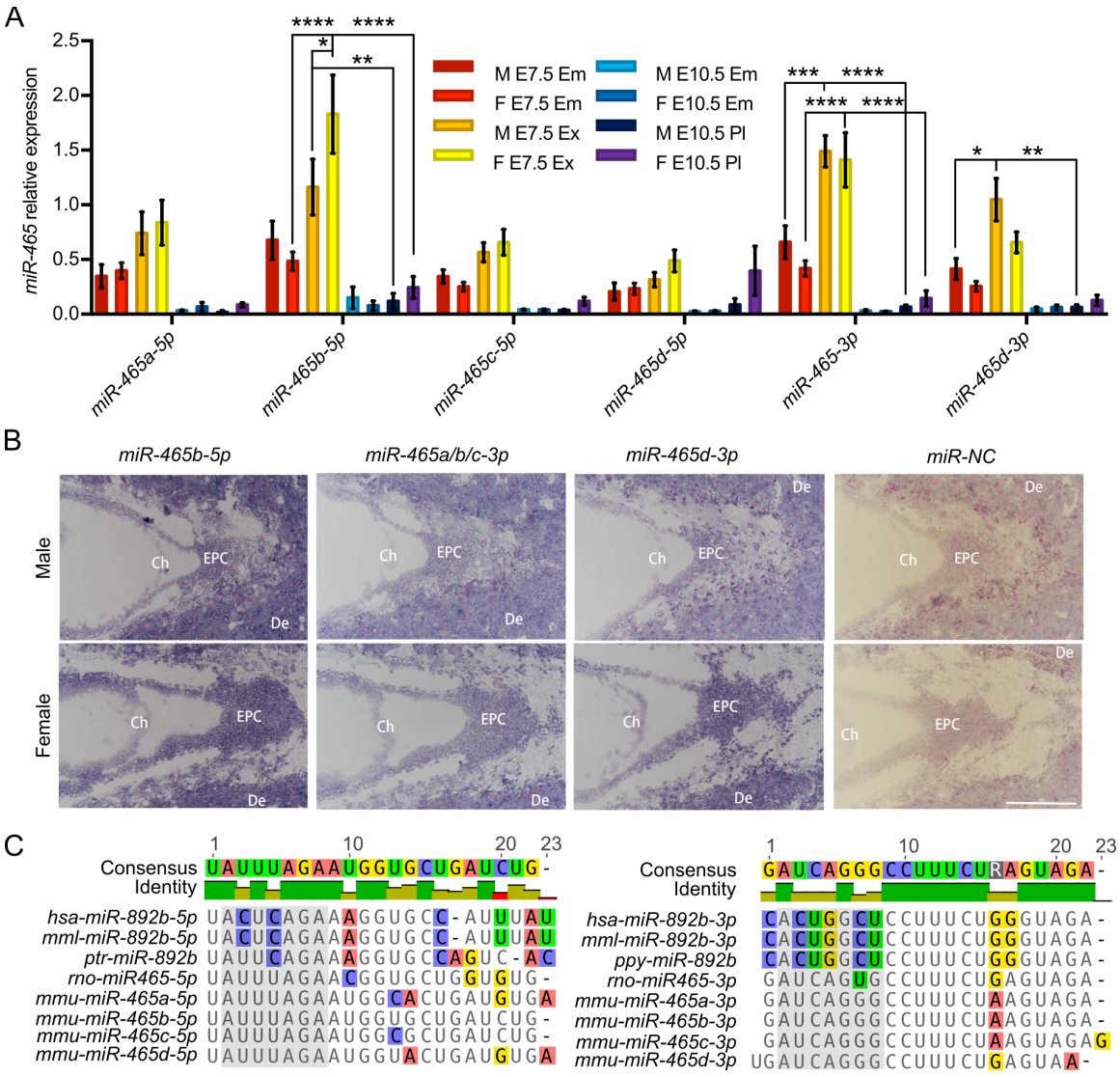

Figure 3-1. Expression pattern and conservation of *miR-465*.

A. *MiR-465* expression pattern in embryonic tissues (or placenta) and fetus at E7.5 and E10.5. M, male; F, female; Em, embryo; Ex, extraembryonic tissue; Pl, placenta. * indicates p<0.05, ** indicates p<0.01, ***indicates p<0.001. Two-way ANOVA was used for statistical analysis. B. RNA-ISH of the *miR-465b-5p*, *miR-465a/b/c-3p*, and *miR-465d-3p* at E7.5. Ch, chorion; EPC, ectoplacental cone; De, maternal decidua; Scale bars: 200 μm. The *miR-NC* is used as a negative control

probe. C. The orthologs of the *miR-465* cluster are relatively conserved among primate and rodent. Bases in grey are potential seed regions. C. The orthologs of the *miR-465* are in the similar X chromosome locus.

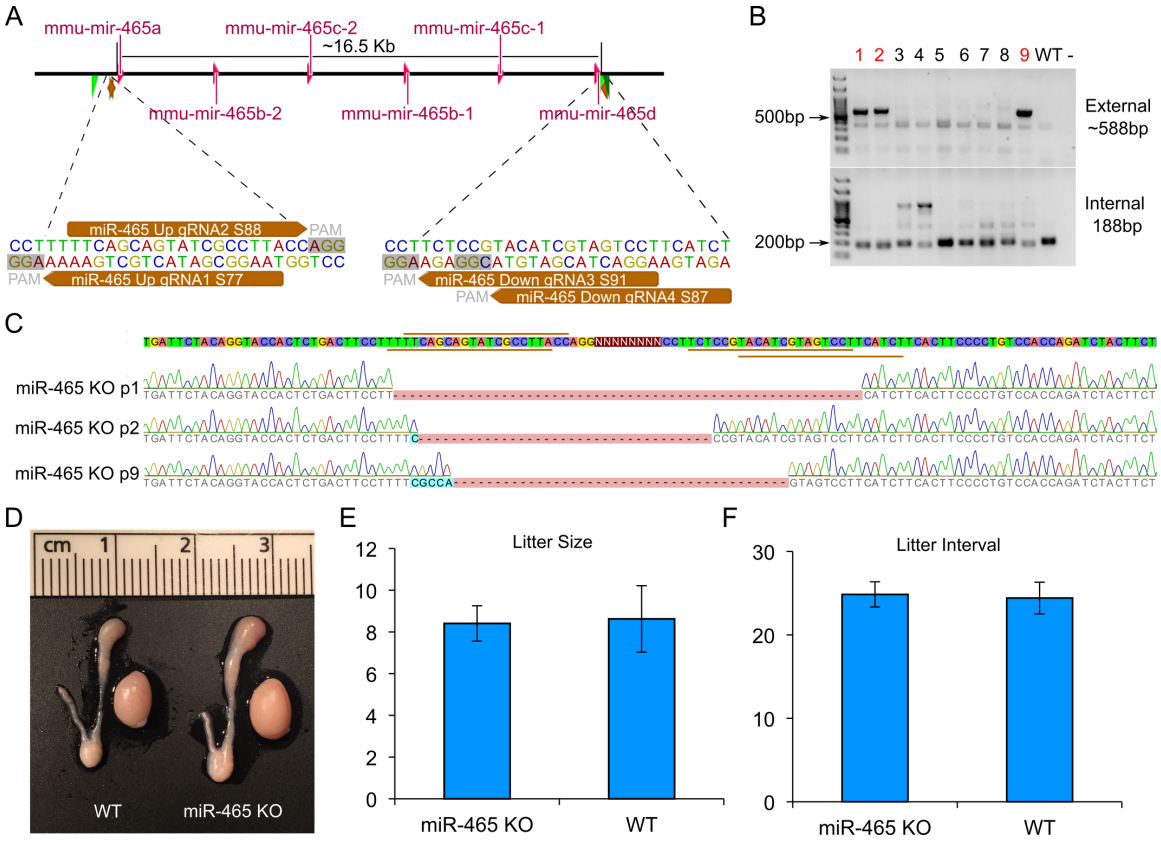

Figure 3-2. Generation of *miR-465* cluster KO mice.

A. The cartoon shows the strategy for generating *miR-465* cluster KO mice. Dark red indicates miRNA cluster, brown indicates gRNAs used, green indicates primers used, grey indicates PAM (protospacer adjacent motif) sequence. B. PCR shows the genotyping of founder mice. C. Sanger sequence showed that the *miR-465* cluster was deleted precisely. D. Testis size of *miR-465*$^{-/-}$ was comparable to WT. E. Litter size of *miR-465*$^{-/-}$ was comparable to WT. F. Litter interval of *miR-465*$^{-/-}$ was comparable to WT.

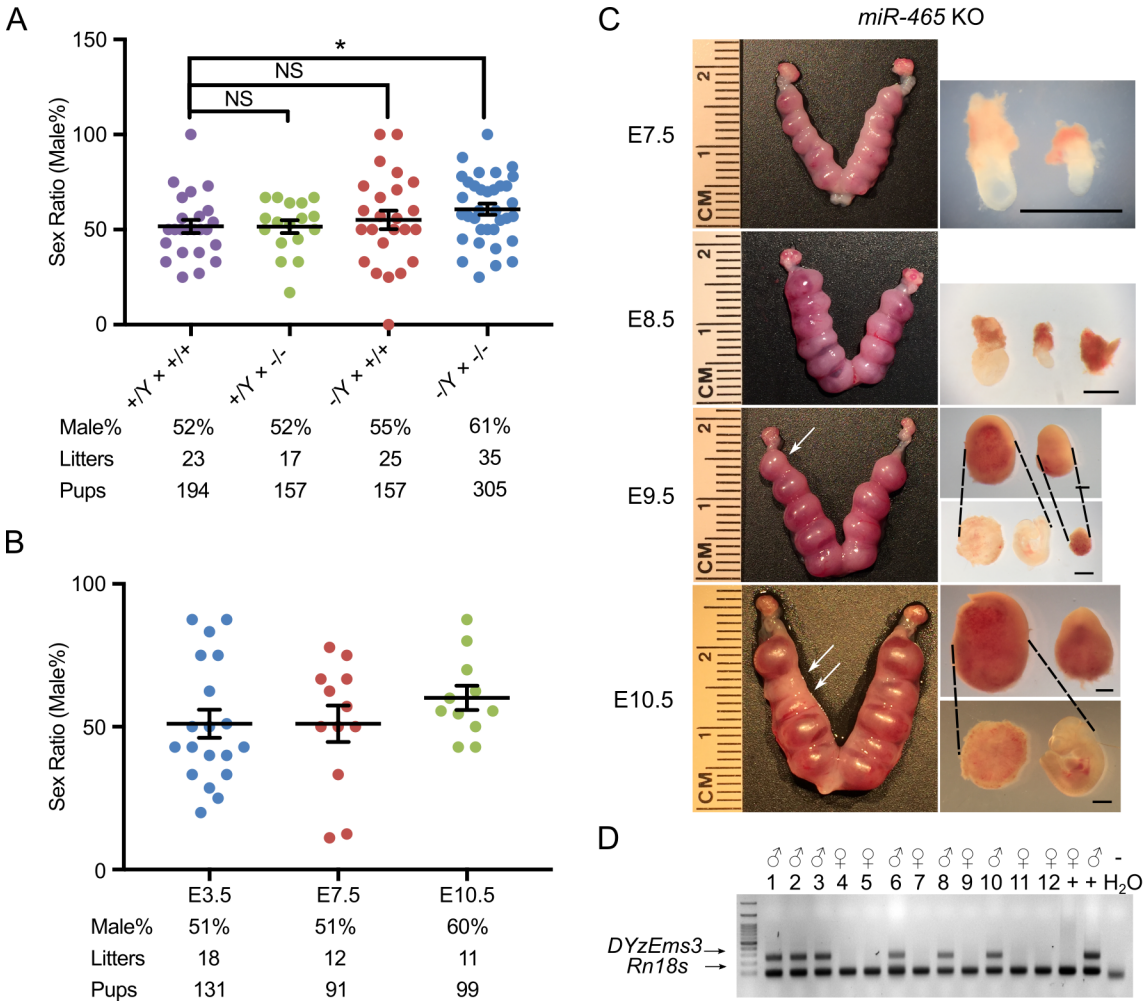

Figure 3-3 *MiR-465*$^{-/-}$ homozygous inbreeding showed highly skewed sex ratio.

A. Sex ratio is highly skewed in homozygous inbreeding (-/Y × -/-), rather than in homozygous outcrossing with WT ((-/Y× +/+) or (+/Y × -/-)) or WT inbreeding (+/Y × +/+). * indicates $p<0.05$, NS indicates not significant. Two sample T-test was used for statistical analysis. B. Skewed sex ratio happens at E10.5, rather than at E3.5 or E7.5, n stands for numbers of breeding pairs. C. Representative figures showed some female embryos were reabsorbed during E8.5-E10.5. D. Representative figure of genotyping data, *DYzEms3* is a Y chromosome-specific

repetitive sequence, *Rn18s* is the internal control. Male shows two bands (*DYzEms3* & *Rn18s*), while female only shows one band (*Rn18s*).

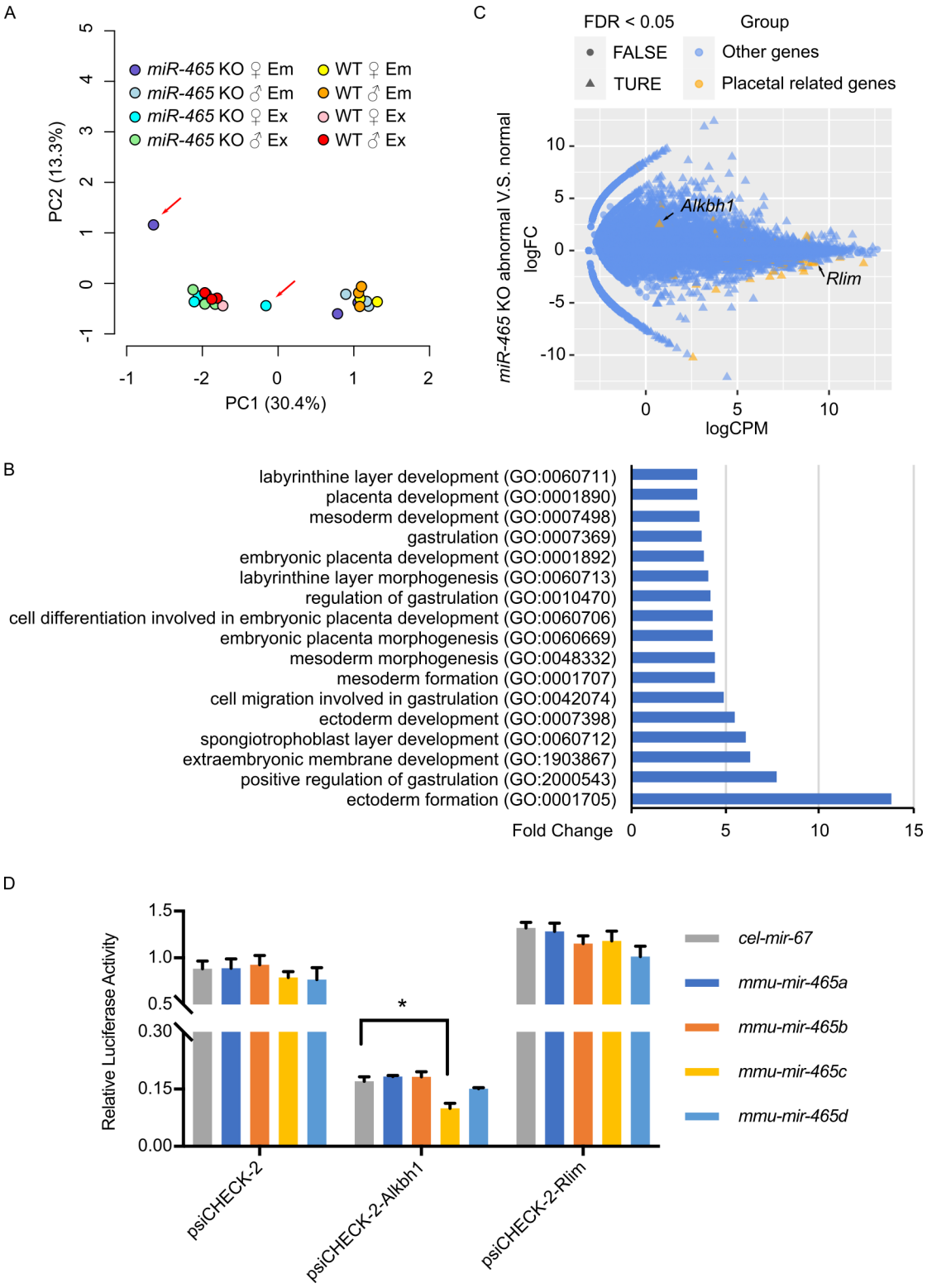

Figure 3-4 RNA-seq analysis of *miR-465* KO and WT mice.

A. PCA analysis of embryo and extraembryonic tissues from *miR-465* KO and WT mice. The red arrows indicate the outlier *miR-465* KO female. B. GO term of dysregulated genes between the outlier and normal-look *miR-465* KO female. C. Scatter plot of dysregulated genes between the outlier and normal-look *miR-465* KO female. FDR > 0.05 and FDR < 0.05 were indicated as circle and triangle, respectively. The genes related to placental development were labeled in orange, and the rest genes were labeled in blue. D. Luciferase assay for *Alkbh1* and *Rlim*. psiCHECK-2 is the empty vector used, the *cel-mir-67* is the negative control miRNA, firefly luciferase activity is used as the internal control. * indicates the $p < 0.05$.

CHAPTER IV: Conclusion and future direction.

Since the discovery of miRNA in C. elegans in 1993 [1], the research focus on miRNAs have increased dramatically. miRNAs appear to influence almost all cellular events[2]. Ablation of miRNA processing enzymes led to early embryonic lethality [3-6], conditional KO of Dicer or Drosha also led to impaired spermatogenesis and male infertility[7], indicating important roles of miRNA in early embryonic development as well as spermatogenesis.

In our studies, we showed that the X linked *miR-506* family fine-tunes spermatogenesis (Chapter II). Previous studies showed that the *miR-506* family only exists in placental mammals and one marsupial [8], using Multiz Alignment & Conservation, we found that the *miR-506* family exists as early as in green sea turtle. Besides, we found that the *miR-506* family has undergone divergent expansion that may be driven by L1 retrotransposition. The *miR-506* family is highly expressed in testis, germ cells, and sperm, and play an important role in fine-tuning spermatogenesis. The targets of the *miR-506* family seem conserved through humans, mice, and rats despite their highly divergent sequence similarities, and the miRNAs within this family compensate with each other. ~7nt of the seed region made it hard to identify miRNA targets. Our results indicate that the ablation of *miR-506* family induced dysregulation of target genes, and further analysis suggests that these targets are shared among humans, mice, and rats. Further work may need to be done by performing CLIP using the available Halo-Ago2 mice[9], although some targets may be missing due to either the reduced activity

of *Ago2* or that the X linked *miR-506* family may be bound by *Ago4* rather than *Ago2*[10].

Among the *miR-506* family, we found that the *miR-465* cluster is involved in sex dimorphism regulation through modulating early extraembryonic development (Chapter III). The *miR-465* cluster is preferentially expressed in extraembryonic tissues at E7.5. Ablation of this cluster led to female-biased lethality, whereas breeding of -/Y or -/- with wild-type didn't induce significant sex bias, suggesting that the sex bias was not due to biased X/Y sperm counts. Further dissection showed that the sex bias happened at E10.5, whereas not at E3.5 or E7.5. The embryos were absorbed during E8.5~E10.5. Further RNA-seq identified that abnormal embryos have upregulated *Alkbh1*, which is a tRNA demethylase and was shown to impair translation elongating upon upregulation[11]. Further work may need to be done by determining if overexpression of Alkbh1 could lead to female-biased lethality and if mutations in the *miR-465* binding site of *Alkbh1* 3'UTR could restore the biased sex ratio.

References

1. Lee, R.C., R.L. Feinbaum, and V. Ambros, The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell, 1993. 75(5): p. 843-54.
2. Bartel, D.P., Metazoan MicroRNAs. Cell, 2018. 173(1): p. 20-51.
3. Bernstein, E., et al., Dicer is essential for mouse development. Nat Genet, 2003. 35(3): p. 215-7.
4. Wang, Y., et al., DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet, 2007. 39(3): p. 380-5.
5. Alisch, R.S., et al., Argonaute2 is essential for mammalian gastrulation and proper mesoderm formation. PLoS Genet, 2007. 3(12): p. e227.
6. Morita, S., et al., One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. Genomics, 2007. 89(6): p. 687-96.
7. Wu, Q., et al., The RNase III enzyme DROSHA is essential for microRNA production and spermatogenesis. J Biol Chem, 2012. 287(30): p. 25173-90.
8. Zhang, F., et al., Evolution of an X-Linked miRNA Family Predominantly Expressed in Mammalian Male Germ Cells. Mol Biol Evol, 2019. 36(4): p. 663-678.
9. Li, X., et al., High-Resolution In Vivo Identification of miRNA Targets by Halo-Enhanced Ago2 Pull-Down. Mol Cell, 2020.

10. Modzelewski, A.J., et al., AGO4 regulates entry into meiosis and influences silencing of sex chromosomes in the male mouse germline. Dev Cell, 2012. 23(2): p. 251-64.
11. Liu, F., et al., ALKBH1-Mediated tRNA Demethylation Regulates Translation. Cell, 2016. 167(3): p. 816-828 e16.