A MILI-independent piRNA biogenesis pathway empowers partial germline reprogramming

Lina Vasiliauskaité1,2, Dimitrios Vitsios3, Rebecca V Berrens4, Claudia Carrieri1,2, Wolf Reik4,5, Anton J Enright3 & Dónal O’Carroll1,2

In mice, the pathway involving PIWI and PIWI-interacting RNA (PIWI–piRNA) is essential to re-establish transposon silencing during male-germline reprogramming. The cytoplasmic PIWI protein MILI mediates piRNA-guided transposon RNA cleavage as well as piRNA amplification. MIWI2’s binding to piRNA and its nuclear localization are proposed to be dependent upon MILI function. Here, we demonstrate the existence of a piRNA biogenesis pathway that sustains partial MIWI2 function and reprogramming activity in the absence of MILI.

The erasure of DNA methylation during mammalian male germline reprogramming and the concomitant derepression of transposons initiate the re-silencing process through the production of primary piRNAs1. In mice, these piRNAs are bound to the cytoplasmic PIWI protein MILI, which is a piRNA-directed RNA endonuclease that mediates piRNA amplification of secondary piRNAs, ensuring transposon repression2 (Supplementary Fig. 1). These secondary piRNAs, through sequence complementarity, guide the PIWI protein MIWII to direct de novo DNA methylation of LINE1 and IAP elements1–5 (Supplementary Fig. 1). The loss of MILI or MIWII results in defective de novo DNA methylation and depression of LINE1s and IAPs, causing meiotic arrest3,5,6. In addition to undergoing meiotic arrest, Miwi2−/− mice progressively lose germ cells and are fully aspermatogenic by 9 months of age2,3. We found that Miwi2−/− mice present a milder germ-cell-loss phenotype with ~50% of tubules being spermatogenic at 1 year of age (Fig. 1a,b). This discrepancy in phenotype indicates that MIWII may have a MILI-independent role during fetal germ cell reprogramming.

We therefore FACS-purified undifferentiated spermatogonia and performed whole-genome bisulfite sequencing. While the loss of MILI and MIWII predominantly affects LINE1 and IAP methylation4,5,7 (Supplementary Fig. 2), we found, in accordance with our hypothesis, that Miwi2 deficiency has a more severe impact on genome remethylation than MilI disruption (Fig. 1c and Supplementary Fig. 2). This discrepancy contrasts with a previous report3, and probably arises from the fact that in the previous study, MiI−/− and Miwi2−/− spermatogonia were differentially isolated with varying degrees of somatic cell contamination8. These somatic cells have normal genomic methylation levels, and thus a differential degree of contamination confounds a quantitative comparative analysis. Here we observed that the methylation of 1,704 loci were dependent upon MIWII function, whereas 258 loci showed selective dependence upon MILI (Fig. 1c). To determine whether the methylation of these loci could be attributed to the piRNA pathway, we analyzed whether piRNAs bound to MILI or MIWII fetal gonad ribonucleoproteins (RNPs) map to these loci. Indeed, piRNAs from both RNPs were complimentary to these loci (Fig. 1d). Next, we analyzed uniquely mapping piRNAs, which excluded transposon-associated piRNAs, and found that unique piRNAs from the respective RNPs map to both MIWII- and MILI-dependent loci (Fig. 1e). In summary, Miwi2 deficiency has a greater impact on spermatogenesis and fetal germline reprogramming in comparison to the loss of MilI.

The physiological and molecular discrepancies between Miwi2 and MilI deficiency could indicate that a non-canonical piRNA biogenesis and partial piRNA-dependent MIWII function exists in the absence of MILI. To test this hypothesis, we generated total small RNA libraries from wild-type and MilI−/− embryonic day 16.5 (E16.5) fetal testes. Although at low levels, piRNAs are clearly present in the MilI−/− libraries (Fig. 2a and Supplementary Fig. 3a). Importantly, these are bona fide piRNAs, as all the piRNAs found in the absence of MILI are present in wild-type fetal gonads (Fig. 2b). LINE1 and IAP piRNAs are generated in greatly reduced quantities and, as expected, the hallmarks of piRNA amplification are absent in the MilI−/− libraries (Supplementary Fig. 3b–e). In the MilI mutant, piRNAs cognate to the 1,704 MIWII-dependent loci were still present but at lower levels (Fig. 2c and Supplementary Fig. 3f), with a two- and three-fold reduction in multi- or uniquely mapping piRNAs observed, respectively (Supplementary Fig. 3g). In contrast, the MilI−/− libraries showed an 18-fold reduction in the piRNAs uniquely mapping to the 256 loci whose methylation is dependent upon MILL function (Supplementary Fig. 3g).

Because piRNAs exist in MilI−/− fetal gonocytes, albeit in reduced quantities, MIWII should retain partial nuclear localization. We next analyzed the localization of MIWII in E16.5 MilI−/− fetal gonocytes by means of confocal immunofluorescence, using a rabbit polyclonal antibody against MIWII (Fig. 2d and Supplementary Fig. 4a). This analysis revealed that a portion of MIWII indeed retains nuclear localization (Fig. 2d). Quantification of the signal in wild-type and Miwi2−/− fetal testes revealed that the anti-MIWI2 antibody gave a non-negligible level of nonspecific background staining (Supplementary Fig. 4b). To independently confirm this finding, we generated an N-terminal HA-epitope-tagged MIWII

1 European Molecular Biology Laboratory (EMBL), Monterotondo, Italy. 2 MRC Centre for Regenerative Medicine, School of Biological Sciences, University of Edinburgh, Edinburgh, UK. 3 European Bioinformatics Institute, Hinxton, Cambridge, UK. 4 Epigenetics Programme, Babraham Institute, Cambridge, UK. 5 Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. Correspondence should be addressed to D.O’C. (donal.ocarroll@ed.ac.uk).

Received 7 March 2017; accepted 26 April 2017; published online 22 May 2017; doi:10.1038/nsmb.3413
Loss of MIWI2 results in more severe physiological and molecular phenotype than loss of MILI. (a) Percentage of aspermatogenic tubules in Mili<sup>−/−</sup> and Mili<sup>+/−</sup> 1-year-old mice. Graph shows mean and s.d. (n = 5 animals). Source data are available online. (b) Hematoxylin- and eosin-stained testis cross sections of Mili<sup>−/−</sup> and Mili<sup>+/−</sup> 1-year-old mice. (c) Comparisons of the percentage of CpG methylation in wild-type (WT) and Mili<sup>−/−</sup>, WT and Miwi<sup>−/−</sup> as well as Mili<sup>+/−</sup> and Miwi<sup>−/−</sup> undifferentiated spermatogonia are shown. Blue dots represent significantly differentially methylated regions (DMR). Number (#) of DMRs in each comparison is shown. (d,e) piRNAs (d) and uniquely mapping piRNAs (e) from MIWI2 and MILI RNPs mapped to loci whose methylation is dependent upon MIWI2 and MILI, respectively. Positive (red) and negative (blue) values indicate sense and antisense piRNAs, respectively. Graphs show averages of values obtained from two animals.

(8A–MIWI2) mouse allele (Miwi<sub>2</sub>)<sup>HA</sup> (Supplementary Fig. 5), permitting the use of high quality monoclonal HA antibodies. Confocal immunofluorescence of HA–MIWI2 in Miwi<sup>−/−</sup> fetal gonocytes confirmed partial nuclear localization of MIWI2 with dramatically reduced levels of nonspecific background signal (Fig. 2e and Supplementary Fig. 4c,d).

---

**Figure 1** piRNAs are produced and MIWI2 is partially localized to the nucleus in the absence of MILI. (a) Length distribution of piRNAs in wild-type (WT) and Mili<sup>+/−</sup>, E16.5 fetal testes. (b) piRNA expression analysis in WT and Mili<sup>−/−</sup>, E16.5 fetal testes. Blue dots represent read counts of individual piRNAs in log<sub>2</sub> scale. (c) Unique piRNAs from WT and Mili<sup>−/−</sup>, E16.5 fetal testes mapped to loci whose methylation is dependent upon MIWI2 and MILI, respectively. Positive (red) and negative (blue) values indicate sense and antisense piRNAs, respectively. Graphs in a–c show averages of values obtained from two animals. (d) Immunofluorescence with anti-MIWI2 antibody (d) and anti-HA antibody (e) on fetal testes of the indicated genotypes with a single fetal gonocyte shown. Scanline profile plots (bottom) represent relative fluorescence intensity of MIWI2 (d) and HA (e) (green) staining in the nucleus (blue) along the analyzed trajectory (represented as arrow in d and e).
The selective loss of MILI’s endonuclease activity attenuates piRNA amplification but leaves primary piRNA biogenesis intact, resulting in an MIWI2 RNP that can partially function to silence IAP\(^2\), whereas the complete loss of MILI function was thought to abrogate all piRNA biogenesis and MIWI2 reprogramming activity\(^1\). Here, we show that non-canonical primary piRNA biogenesis exists in the absence of MILI that results in MIWI2 partial nuclear localization and reprogramming activity. This residual reprogramming activity is supported by the fact that more severe methylation and spermatogenesis defects are observed in $\text{Miwi}^{2/−}$ than in $\text{Mili}^{2/−}$ mice. The piRNA biogenesis in $\text{Mili}^{2/−}$ mice may resemble the phased primary piRNA pathway identified in *Drosophila*, *Bombyx mori* and mouse, in which PIWI binding to cleaved transcripts from the amplification cycle coupled with Zucchini-mediated endonucleolytic cleavage results in additional piRNA biogenesis\(^9\)–\(^12\). While phased piRNA biogenesis in *Drosophila* can occur in the absence of piRNA amplification\(^13\), we could not detect phasing in the $\text{Mili}^{2/−}$ piRNA population (data not shown), likely because of the extensive 3' trimming that occurs during mouse piRNA biogenesis\(^14\). In summary, we show that mouse fetal primary piRNA biogenesis can exist in the absence of MILI, and this residual piRNA pathway contributes to germline reprogramming.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. GA 310206 awarded to D.O'C.

**AUTHOR CONTRIBUTIONS**

L.V. contributed to the design, execution and analysis of the majority of experiments. D.V. performed the bioinformatic piRNA analysis. R.V.B. generated the whole-genome bisulfite libraries and performed the bioinformatic analysis of genomic methylation. C.C. designed, generated and validated the $\text{Miwi}^{2HA}$ allele. A.J.E. and W.R. oversaw all bioinformatic analyses performed. D.O'C. conceived and supervised this study. L.V. and D.O'C. wrote the final version of the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Aravin, A.A. *et al.* Mol. Cell **31**, 785–799 (2008).
2. De Fazio, S. *et al.* Nature **480**, 259–263 (2011).
3. Carmell, M.A. *et al.* Dev. Cell **12**, 503–514 (2007).
4. Kuramochi-Miyagawa, S. *et al.* Genes Dev. **22**, 908–917 (2008).
5. Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G.J. Science **316**, 744–747 (2007).
6. Kuramochi-Miyagawa, S. *et al.* Development **131**, 839–849 (2004).
7. Molaro, A. *et al.* Genes Dev. **28**, 1544–1549 (2014).
8. Manakov, S.A. *et al.* Cell Rep. **12**, 1234–1243 (2015).
9. Han, B.W., Wang, W., Li, C., Weng, Z. & Zamore, P.D. Science **348**, 817–821 (2015).
10. Mohn, F., Handler, D. & Brennecke, J. Science **348**, 812–817 (2015).
11. Homolka, D. *et al.* Cell Rep. **12**, 418–428 (2015).
12. Yang, Z. *et al.* Mol. Cell **61**, 138–152 (2016).
13. Wang, W. *et al.* Mol. Cell **59**, 819–830 (2015).
14. Saxe, J.P., Chen, M., Zhao, H. & Lin, H. *EMBO J.* **32**, 1869–1885 (2013).
ONLINE METHODS

Mouse strains and experimentation. The Mili<sup>+</sup> and Miwi2<sup>+</sup> mouse alleles were previously produced in the O’Carroll laboratory<sup>12,13</sup> and maintained on a mixed 129/C57BL/6 genetic background. The Miwi2<sup>-tdTomato</sup> transcripational allele was used to isolate spermatogonia<sup>14</sup>. The Miwi2<sup>ΔHA</sup> allele was created using CRISPR-Cas9 gene-editing technology<sup>15</sup>. The sequence encoding an HA tag was inserted after the starting ATG codon of Miwi2 to generate the Miwi2<sup>ΔHA</sup> allele that results in an N-terminal HA–Miwi2 fusion protein. An sgRNA (5′ ACCCGACACGTCGTCTTT; reverse: 5′ CAGGATACGAAAAGGAAGA) flanking Miwi2 ATG with BamHI. The Miwi2<sup>ΔHA</sup> allele was also confirmed by sequencing. For the analysis of progressive germ cell loss, Mili<sup>−/−</sup> and Miwi2<sup>−/−</sup> mice were aged to one year and sacrificed, and then the testes were collected. All mice used in this study were male and bred and maintained in EMBL Mouse Biology Unit, Monterotondo, and subsequently in the Centre for Regenerative Medicine, Edinburgh. All procedures were done in accordance to the current Italian legislation (Art. 9, 27 Jan. 1992, nu 161) under license from the Italian Health Ministry or the UK Home Office regulations, respectively.

Spermatogonia isolation. Isolated postnatal-day-14 testes from mice of the respective genotypes that carried the Miwi2<sup>-tdTomato</sup> transcripational allele in heterozygosity were dealbulginated and digested with collagenase (0.5 mg/ml, Sigma) at 32 °C for 10 min, shaking in 1 ml of DMEM media (Life Technologies). Whole-genome bisulfite-sequencing libraries were generated from these samples by using a postbisulfite-adaptor tagging (PBAT) method as previously described<sup>17</sup>, using 10 cycles of PCR amplification. Libraries were sequenced using Illumina HiSeq 2000. Three biological replicates (cells independently isolated from three animals) were generated per genotype to ensure adequate power to detect statistically significant differences between the respective genotypes.

Histology. Testes were fixed in Bouin’s fixative (Sigma) overnight at 4 °C and paraffin embedded. 8-µm-thick sections were stained with hematoxylin and eosin using routine methods.

Immunofluorescence. Freshly dissected testes were embedded into OCT compound (Sakura), 8-µm sections were cut and fixed in 4% paraformaldehyde (Sigma) for 10 min at room temperature, then permeabilized for 10 min at room temperature in 0.1% Triton-X100 (Sigma). Subsequently, sections were blocked for 30 min at room temperature in 10% normal donkey serum (Sigma), 1% BSA (Sigma) and 0.1M glycine (Sigma). Primary antibody incubation was done overnight at 4 °C in the blocking buffer, using rabbit polyclonal anti-Miwi2 (1:200) and rabbit monoclonal anti-HA (1:500, Cell Signalling, 3724). Appropriate donkey anti-rabbit AlexaFluor488 secondary antibody (Life Technologies) was used in a dilution of 1:1000. DAPI (5 µg/µl, Life Technologies) was used to stain DNA. Samples were mounted with ProLong Gold Antifade Reagent (Life Technologies). A Leica TCS SP5 confocal microscope was used to acquire images. Five images were acquired per each biological replicate for each genotype. After acquisition, images were processed with ImageJ and Adobe Photoshop CS5. All images correspond to the same experiment were acquired and processed applying the same settings. To determine Miwi2 and HA immunofluorescence intensities in the nuclei of gonocytes, fetal testes cross sections were stained with Miwi2 and HA antibodies as described above. Fetal testes from two embryos were used per genotype and analyzed as biological duplicates, except the Miwi2<sup>−/−</sup> fetal testes, for which a biological triplicate was used. All data were collected in 2–3 independent immunofluorescence experiments. Acquired stacks of images were merged using ImageJ. The ROI Manager Tool in ImageJ was used to quantify relative fluorescence intensity in the nuclei where ROIs were manually defined, avoiding cytoplasmic signal.

Bisulfite conversion. For bisulfite conversion, DNA was isolated from sorted cells as described above. Whole-genome bisulfite-sequencing libraries were generated using a postbisulfite-adaptor tagging (PBAT) method as previously described<sup>18</sup>, using 10 cycles of PCR amplification. Libraries were sequenced using Illumina HiSeq 2000. Three biological replicates (cells independently isolated from three animals) were generated per genotype to ensure adequate power to detect statistically significant differences between the respective genotypes.

Bisulfite sequencing analysis. Raw sequence reads were trimmed to remove both poor quality calls and adapters using Trim Galore (v0.4.1, www.bioinformatics.babraham.ac.uk/projects/trim_galore/, Cutadapt version 1.8.1, parameters:--paired). Trimmed reads were aligned to the mouse genome (GRCm38) in paired-end mode to be able to use overlapping parts of the reads only. Once alignments were carried out with Bismark v0.14.4<sup>19</sup> with the following set of parameters: paired-end mode:--phat. Reads were then deduplicated with deduplicate_bismark selecting a random alignment for position that was covered more than once. Cpg methylation calls were extracted from the deduplicated mapping output, ignoring the first 6 bp of each read to reduce the methylation bias typically observed in PBAT libraries using the Bismark methylation extractor (v0.14.4) with the following parameters (i) paired-end mode:--ignore 6--ignore_r2 6, (ii) single-end mode:--ignore 6. Cpg methylation calls were analyzed using R and SeqMonk software (http://www.bioinformatics. babraham.ac.uk/projects/seqmonk/). 50 adjacent Cpg running window probes were generated and percentage of methylation determined for probes containing at least five reads and three Cpgs on the pooled replicate data. The methylation level was expressed as the mean of individual Cpg site graphs. Graphing and statistics were performed using Seqmonk and RStudio.

Transposon analysis in bisulfite sequencing data. Repeat locations for a predefined set of repeat classes of interest were extracted from the premasked RepeatMasker libraries (mm10 - Dec 2011 - RepeatMasker open-4.0.5, http://www.repeatmasker.org/species/musMus.html). Repeat instances overlapping annotated genes in the Ensembl gene set were removed to avoid mixing signals from genic expression with specific expression of repetitive sequences. Methylation levels at the repeat instances were quantitated by summing up all methylation calls and nonmethylation calls for all instances of each class of repeat and calculating the percentage of methylated Cs over all Cs.

Small RNA libraries. Isolated E16.5 fetal testes were homogenized in 1 ml of Qiazol (QIAGEN). Extracted total-fetal-testes RNA was used to generate small RNA libraries using NEBNext Multiplex Small RNA Library Prep Set for Illumina (Set 1, NEB), following manufacturer’s instructions, with 20 cycles of PCR amplification. Small RNA libraries were generated from two biological replicates (RNA isolated from fetal testes from two animals) per genotype to ensure adequate confidence in observations.

Small RNA sequencing analysis. All samples were initially aligned against Rfam<sup>20</sup> in order to filter out tRNA sequences. tRNA hits with an alignment identity score >90% were excluded from the rest of the analysis. Analysis of the filtered samples was then performed using Sequelcmp<sup>21</sup>. Input reads were first trimmed from the 3’ adapter with Reaper (using default configuration for read
geometry without barcode) and deduplicated with Tally, which are both part of the Kraken suite of tools. The length distribution of all cleaned reads between 18 and 32 nucleotides was recorded in order to check for depletion of piRNA sequences between the WT and Mili−/− conditions. Cleaned reads were later filtered by length (retaining only 24–32 nucleotide long sequences) and aligned against the Mouse genome (Ensembl release 66), allowing up to two mismatches and reporting up to 20 hits per sequence when analyzing for all mappers. In the case of uniquely mapped sequences, Bowtie call from the SequenceImp pipeline was tweaked using the parameter −m = 1 (parameter −k was set to 1 for both the unique and all mappers cases). BAM output files from the alignment step were intersected with BED files containing the coordinates of 1,704 and 258 loci, whose methylation is dependent upon MIWI2 and MILI, respectively. PiRNA counts within each locus were calculated as the average number of fragments aligning against the locus, divided by the size of the locus region in 1 kb units. Expression densities were limited within the interval (−100, 100) in order to filter out outliers from three overexpressed loci and thus increase density resolution for all loci. Regarding the piRNA differential expression analysis, a custom database of all 26–31 nucleotide long unique sequences found across all WT and Mili−/− replicates was initially built. Each sample replicate was then aligned against this database, and expression levels of all matching sequences were quantified between the two conditions. As for the quantification of LINE1 and IAP repeats, the analysis was performed using the ‘features’ step of SequenceImp for repeat elements, allowing up to three mismatches and correcting the read counts to the number of genome-mapping reads. The ping-pong signatures and 1U–10A content of the LINE1 and IAP elements were also calculated as part of this step. In all cases, normalization was performed based on the total number of reads of transcripts that remained unchanged between the two conditions. Specifically, the types of transcripts that were used for normalization (based on the official Ensembl Genebuild annotation) were: miRNAs, rRNAs, snRNAs, snoRNAs, processed transcripts, aa_tRNAs, Mt_tRNAs, other_tRNAs, miscRNAs and RNA_repeats. Significance was assessed using the BoostRatio algorithm.

**Code availability.** Computer code used to analyze small RNA sequencing and whole-genome bisulfite sequencing data is available upon request from A.J.E. or D.V. and R.V.B. or W.R., respectively.

**Data availability.** Small RNA sequencing data have been deposited in [http://www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena) with the accession code PRJEB19792. WGBS data have been deposited in [http://www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress) with the accession code E-MTAB-5561. MIWI2 and MILI RNP-associated piRNA libraries were used from a previously published study. Source data for Figure 1a and Supplementary Figures 3–5 are available online. The imaging data that support the findings of this study, and all other data, are available from the corresponding author upon reasonable request.

---

15. Di Giacomo, M. et al. Mol. Cell 50, 601–608 (2013).
16. Carrieri, C. et al. J. Exp. Med. [http://dx.doi.org/10.1084/jem.20161371](http://dx.doi.org/10.1084/jem.20161371) (2017).
17. Yang, H. et al. Cell 154, 1370–1379 (2013).
18. Miura, F. & Ito, T. DNA Res. 22, 13–18 (2015).
19. Krueger, F. & Andrews, S.R. Bioinformatics 27, 1571–1572 (2011).
20. Nawrocki, E.P. et al. Nucleic Acids Res. 43, D130–D137 (2015).
21. Davis, M.P., van Dongen, S., Abreu-Goorder, C., Bartonicek, N. & Enright, A.J. Methods 63, 41–49 (2013).
22. Clèries, R. et al. Comput. Biol. Med. 42, 438–445 (2012).