Cell Cycle 12:11, 1657–1658; June 1, 2013; © 2013 Landes Bioscience

Defining piRNA primary transcripts

Xin Zhiguo Li, Christian K. Roy, Melissa J. Moore and Phillip D. Zamore*

Howard Hughes Medical Institute; RNA Therapeutics Institute and Department of Biochemistry and Molecular Pharmacology; University of Massachusetts Medical School; Worcester, MA USA

PIWI proteins and their associated small RNAs (PIWI-interacting RNAs, piRNAs) are essential for fertility in mammals.[1,2] piRNAs (24–35 nt) are longer than miRNAs (21–23 nt) and have 2'-*O*-methyl-modified 3' termini.[3-6] Like miRNAs, piRNAs bind members of the Argonaute family of proteins, but piRNAs are unique in that they guide PIWI proteins, a specialized subfamily of Argonaute proteins that are expressed mainly in germ cells.[1] The sequences of piRNAs are more diverse than any other known class of cellular RNAs. For example, our 8.8 million piRNA reads in a deep sequencing library from adult mouse testis comprised 2.7 million different piRNAs; > 90% of piRNA species were sequenced just once.

piRNAs map to large blocks of genomic sequence called clusters.[4,5] The architecture of piRNA clusters suggests that mature piRNAs derive from precursor transcripts via multiple RNA processing steps. While > 80% of piRNA reads in flies map to transposons, ~93% of adult mouse piRNA reads map to a single site in the genome. That so many mouse piRNAs map uniquely to the genome facilitates the study of their precursor transcripts. Using total RNA sequencing, we detected long RNAs (> 100 nt) from piRNA clusters. Chromatin immunoprecipitation of RNA polymerase II and III suggests that these transcripts are transcribed by RNA polymerase II. Consistent with mouse piRNA precursors being RNA polymerase II transcripts, the long RNAs bear a standard cap structure (detected by cap analysis of gene expression; CAGE), and their 3' ends possess a poly(A) tail (detected by polyadenylation site sequencing; PAS-seq).[7]

Historically, piRNA clusters have been defined computationally using mature piRNA sequences. Thus, clusters are not transcriptional units. When we began our studies, it was unknown whether piRNAs originated from single continuous transcripts or multiple short transcripts, as they do in nematodes. The location of piRNA precursor transcription start sites (TSSs) and promoters or whether they are spliced was also unknown. We discovered the first transcription factor regulating piRNA biogenesis by analyzing a subset of 15 clusters that appeared to be transcribed bidirectionally. In these clusters, a short piRNA region lacking piRNA separates piRNAs mapping to the genomic minus strand from those mapping to the plus strand. These putative bidirectional promoter regions contain binding motifs for the MYB family of transcription factors ($E = 8.3 \times 10^{-12}$). We used the subset of clusters for motif finding, because the computationally defined boundary of most clusters is more than 3 kbp away from experimentally determined TSSs. Accurately annotating the TSSs of piRNA cluster transcripts allowed us to discover that MYB motifs are significantly enriched ($E = 9.1 \times 10^{-28}$) in 100 piRNA promoters, including both bidirectional and unidirectional transcribed clusters.[7] This allowed us to define the transcription units of piRNA-producing loci and replace the previous computationally defined cluster annotations.

We identified primary piRNA transcripts by a combination of de novo and reference-based transcript assembly using paired-end total RNA sequencing. We further corrected the ends of these transcripts using CAGE and PAS. Our data allowed us to define 467 piRNA precursor transcripts derived from 214 piRNA-producing loci.[7] The 214 loci constitute just 0.33% of the mouse genome yet account for 95% of piRNAs in the adult mouse testis.

The 214 loci can be divided into two different types of piRNA-producing genes. Genic piRNA loci are known protein-coding genes that also produce piRNAs and are mostly expressed during early spermatogenesis before the pachytene stage of meiosis. Historically, piRNAs mapping to these genes have been referred to as pre-pachytene piRNAs.[8] Intergenic piRNA loci lie far in the genome from other annotated genes. Most piRNAs from these loci are detected after 12.5 days postpartum and have been called pachytene piRNA clusters.

The transcription factor A-MYB binds at the promoters of the pachytene piRNA-producing loci, and the loss of *A-Myb* depleted both piRNA precursor transcripts as well as mature piRNAs across the length of the loci.[7] Thus, our data represent the first formal demonstration that long RNA polymerase II transcripts are the precursors of mature piRNAs in mammals. That is, piRNAs are derived from long, continuous RNAs that are subsequently fragmented and processed into piRNAs (**Fig. 1**). The set of piRNA loci and well-annotated piRNA precursor transcripts defined by our work provide an invaluable resource for further study of piRNA biogenesis and function.

*Correspondence to: Plillip D. Zamore; Email: phillip.zamore@umassmed.edu
Submitted: 04/22/13; Accepted: 04/23/13
http://dx.doi.org/10.4161/cc.24989
Comment on: Li XZ, et al. Mol Cell 2013; 50:67-81; PMID:23523368; http://dx.doi.org/10.1016/j.molcel.2013.02.016

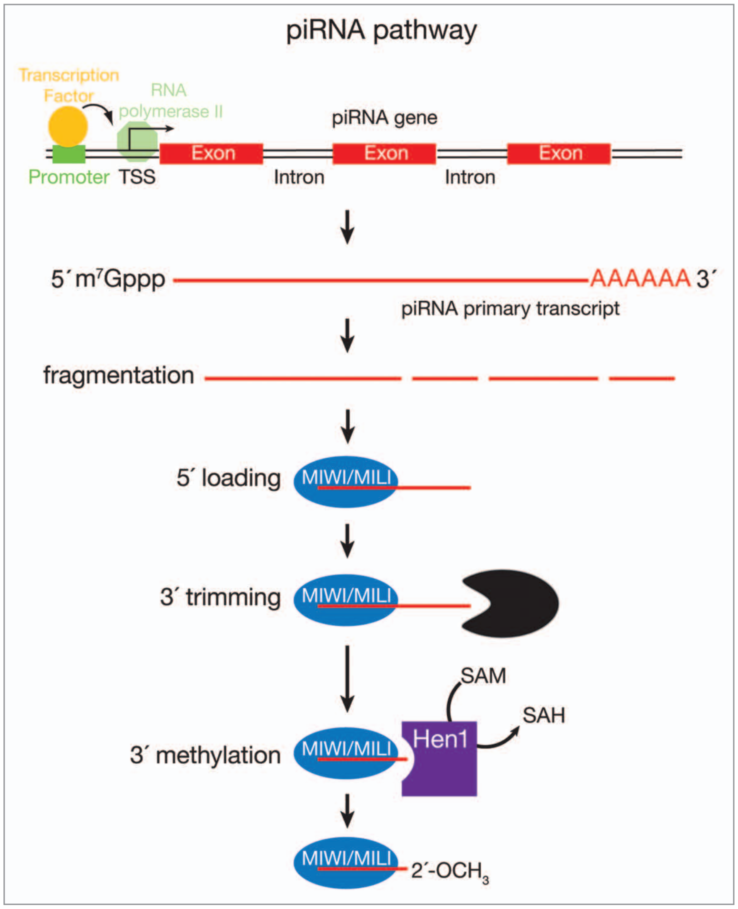

Figure 1. A model for piRNA biogenesis. Primary piRNA transcripts are transcribed by RNA polymerase II and contain 5' caps, exons and introns and poly(A) tails. The transcription of pachytene piRNA genes is controlled by A-MYB; transcription factor(s) (TF) controlling pre-pachytene piRNA genes remain to be discovered. Current models of piRNA biogenesis propose that PLD6 determines the 5' end of piRNA intermediates with lengths > 30 nt. These intermediates are proposed to then be loaded into PIWI proteins. After PIWI binding, a nuclease is thought to trim the 3' end of the piRNA to the length characteristic of the particular bound PIWI protein. Finally, further trimming is prevented by addition of a 2'-*O*-methyl group to the 3' end of the mature piRNA by the *S*-adenosylmethionine-dependent methyltransferase Hen1.

References

1. Deng W, et al. Dev Cell 2002; 2:819-30; PMID:12062093; http://dx.doi.org/10.1016/S1534-5807(02)00165-X
2. Kuramochi-Miyagawa S, et al. Development 2004; 131:839-49; PMID:14736746; http://dx.doi.org/10.1242/dev.00973
3. Aravin A, et al. Nature 2006; 442:203-7; PMID:16751777
4. Lau NC, et al. Science 2006; 313:363-7; PMID:16778019; http://dx.doi.org/10.1126/science.1130164
5. Girard A, et al. Nature 2006; 442:199-202; PMID:16751776
6. Grivna ST, et al. Genes Dev 2006; 20:1709-14; PMID:16766680; http://dx.doi.org/10.1101/gad.1434406
7. Li XZ, et al. Mol Cell 2013; 50:67-81; PMID:23523368; http://dx.doi.org/10.1016/j.molcel.2013.02.016
8. Aravin AA, et al. Science 2007; 316:744-7; PMID:17446352; http://dx.doi.org/10.1126/science.1142612