Seminars in Cell & Developmental Biology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

A piece of the pi(e): The diverse roles of animal piRNAs and their PIWI partners

Robyn S.M. Lim[a,b,**], Toshie Kai[a,b,*]

[a] Temasek Life Sciences Laboratory, National University of Singapore, Singapore 117604, Singapore
[b] Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore

ARTICLE INFO

Article history:
Available online xxx

Keywords:
PIWI-interacting RNA (piRNA)
PIWI protein
piRNA biogenesis
Transposable elements
Gene regulation
Evolutionary conservation

ABSTRACT

Small non-coding RNAs are indispensable to many biological processes. A class of endogenous small RNAs, termed PIWI-interacting RNAs (piRNAs) because of their association with PIWI proteins, has known roles in safeguarding the genome against inordinate transposon mobilization, embryonic development, and stem cell regulation, among others. This review discusses the biogenesis of animal piRNAs and their diverse functions together with their PIWI protein partners, both in the germline and in somatic cells, and highlights the evolutionarily conserved aspects of these molecular players in animal biology.

1. Introduction

Three broad classes of endogenous small RNAs, microRNAs (miRNAs), endogenous small interfering RNAs (esiRNAs), and PIWI-interacting RNAs (piRNAs), are well known for their roles in cellular processes. These small RNAs act in concert with Argonaute proteins, which are found in many organisms [1,2]. miRNAs and esiRNAs associate with a subset of Argonaute proteins that is collectively referred to as the AGO clade [3–6]. Mature miRNAs and esiRNAs are approximately 21–24 nucleotides (nt) in length and are canonically generated from double-stranded regions (dsRNA) of RNA precursors [4–19]. The dsRNA-slicing enzyme Dicer is required for the maturation of AGO-bound small RNAs [4,10,15,20–23]. These small RNAs operate by guiding their AGO effector molecule partners to targets through base pairing, and they predominantly function in gene silencing [4–6,12–19,24–32], although there are several exceptions [33–38]. Other details of the miRNA and esiRNA pathways and their functions have been reviewed elsewhere [39–48].

The objective of this review is to provide a broad overview of the third class of small RNAs that associate with Argonaute proteins, the PIWI-interacting RNAs (piRNAs), and their protein partners. We cover the biogenesis of piRNAs and highlight their diverse features and functions in both the germline and in somatic cells of various animals. Unlike miRNAs and esiRNAs, piRNAs are typically longer, with a length of 24–31 nt, originate from single-stranded RNA (ssRNA) precursors, and do not require Dicer for maturation [49–52]. First identified in the fruit fly, *Drosophila melanogaster*, piRNAs were initially designated as repeat-associated small interfering RNAs (rasiRNAs) because they were found to map to repetitive elements and transposons and to participate in their suppression [49,50,53,54]. Subsequently, because rasiRNAs were found to bind to effector Argonaute proteins of the PIWI clade but not to those

Abbreviations: AGO, Argonaute clade protein; Ago3, Argonaute 3; Armi, Armitage; Arx, Asterix; Aub, Aubergine; CBC, cap-binding complex; Dnmt, DNA methyltransferase; dsRNA, double-stranded RNA; ESC, embryonic stem cell; esiRNA, endogenous small interfering RNA; GSC, germline stem cell; H3K4me2, histone 3 lysine 4 dimethylation; H3K9me3, histone 3 lysine 9 trimethylation; Hop, Hsp70/Hsp90 organizing protein; HP1a, heterochromatin protein 1a; Hsp90, heat shock protein 90; iPS, induced pluripotent stem; lncRNA, long non-coding RNA; LRSC, lineage-restricted somatic stem cell; m^7G, 5′ terminal 7-methylguanylate cap; *masc*, *masculinizer*; MB, mushroom body; miRNA, microRNA; mitoPLD, mitochondrial phospholipase D; MOV10L1, Moloney leukemia virus 10-like 1; mRNA, messenger RNA; MSC, multipotent stem cell; nt, nucleotides; Piwi, P-element-induced wimpy testis; PIWI, PIWI clade protein; Pol II, RNA polymerase II; PSC, pluripotent stem cell; PTGS, post-transcriptional gene silencing; rasiRNA, repeat-associated small interfering RNA; RdRP, RNA-dependent RNA polymerase, Rhi-Del-Cuff; RDC, Rhino-Deadlock-Cutoff; scnRNA, scan RNA; sDMA, symmetrically dimethylated arginine; ssRNA, single-stranded RNA; Su(var)3–9, suppressor of variegation 3–9; TE, transposable element; TGS, transcriptional gene silencing; TSS, transcription start site; TUD, Tudor domain-containing protein; UAP56, 56-kDa U2AF-associated protein; UTR, untranslated region; Vas, Vasa; Zuc, Zucchini.

* Corresponding author. Present address: Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka Suita, Osaka 565-0871, Japan. Tel.: +81 6 6879 7972; fax: +81 6 6879 7972.
** Corresponding author. Present address: Department of Pharmacy, The Academia, 20 College Road, Singapore General Hospital, Singapore 169856, Singapore. Tel.: +65 6576 7813; fax: +65 6221 5142.
E-mail addresses: robyn.lim.s.m@sgh.com.sg (R.S.M. Lim), toshie.kai@fbs.osaka-u.ac.jp (T. Kai).

http://dx.doi.org/10.1016/j.semcdb.2015.10.025
1084-9521/

of the AGO clade, they were renamed as piRNAs [50–52,54–57]. The PIWI clade of Argonaute proteins was named after its founding member, the *Drosophila piwi* gene, abbreviated from *P-element-induced wimpy testis* (for clarity, PIWI refers to a protein of the PIWI clade, whereas Piwi refers to the individual *Drosophila* protein) [58]. The members of this clade, the PIWI proteins (PIWIs), are phylogenetically distinct from AGO proteins and, apart from their presence in some protists, have not been observed beyond the animal kingdom, unlike AGO proteins [1,59].

Because many of the initial functional and mechanistic insights into these molecules were derived from work in fly ovaries, in which the majority of piRNAs map to transposable elements (TEs), the most well-studied role of the piRNA pathway is in TE silencing [53,54,56,60,61]. The piRNA pathway primarily targets class I TEs, the retrotransposons, which propagate through an RNA intermediate (reviewed in [62,63]). piRNAs and PIWIs are mainly expressed in the gonads of many of the animals that have been studied, and they predominantly appear to negatively regulate their targets; however, exceptions do exist (discussed later) (Fig. 1). Although PIWIs are also known to participate in genome rearrangement in ciliates [64–67] (reviewed in [68–70]), these non-metazoan PIWIs associate with small RNAs, termed scan RNAs (scnRNAs), which are generated from dsRNA precursors in a Dicer-dependent manner [71,72]. Thus, ciliate PIWIs and scnRNAs might have evolved separately from metazoan PIWIs and piRNAs. In this review, we will focus on piRNAs and PIWIs in animals and will not cover them in protists.

Throughout the review, we also present a recurrent perspective: the evolutionary conservation of piRNAs and PIWIs, highlighting the known conserved features and functions of these molecules. Although there are differences between the species, many attributes appear to be conserved across metazoans, emphasizing the utility of piRNAs and PIWIs in animal biology.

2. The germline piRNA pathway

The piRNA pathway is required for protecting the germline genome against damaging levels of TE activity and for normal gametogenesis. Its importance in the germline is highlighted by the observation that mutations in piRNA pathway components often result in DNA damage caused by excessive TE mobility and sterility [60,61,73–85]. Within germline cells, many evolutionarily conserved components of the piRNA pathway, including cytoplasmic PIWIs, are enriched at the nuage—a membraneless, perinuclear organelle that is common to all examined metazoan germline cells [86–88]—conceiving the theory that the nuage is a site of piRNA production and TE silencing [52,56,60,61,89]. The nuage is alternatively known as P granules in *Caenorhabditis elegans* and is believed to contribute to cytoplasmic bodies such as sponge bodies, chromatoid bodies, balbiani bodies, and mitochondrial clouds because it often colocalizes with mitochondria (reviewed in [90–92]). Hence it may be no surprise that certain components involved in piRNA biogenesis are also associated with mitochondria [93–95] (reviewed in [96]).

2.1. piRNA biogenesis: the birth and propagation of piRNAs

2.1.1. piRNA precursor transcription

Much of our present mechanistic understanding of piRNA biogenesis has been shaped by work in *Drosophila* ovaries (Fig. 2A–C). Long, ssRNA precursors are transcribed from genomic regions known as piRNA clusters or loci that contain fragments of TEs and other repetitive elements [56]. In the female *D. melanogaster* germline, piRNA loci are either unidirectionally transcribed (uni-strand clusters) (Fig. 2A), giving rise to piRNAs that largely map to one genomic strand, or bidirectionally transcribed in a convergent manner (dual-strand clusters) (Fig. 2B), producing piRNAs that map to both strands [56]. Uni-strand cluster transcription displays all of the key attributes of canonical Pol II transcription: the transcription-associated dimethylation mark on histone 3 lysine 4, H3K4me2, and RNA Polymerase II (Pol II) are present at the promoters and transcription start sites (TSS) of these clusters, respectively, and transcripts are 5′ methyl-guanosine-capped and terminated [97] (Fig. 2A). How these uni-strand piRNA precursors are distinguished from other canonical Pol II transcripts remains to be determined. Canonical Pol II transcription of piRNA loci also occurs in the mouse testis, where most piRNAs originate from a single DNA strand [98–100]. In fact, unidirectional piRNA cluster transcription appears to be evolutionarily conserved, from non-bilaterians to bilaterians [52,57,98–108] (Fig. 2A). By contrast, the convergent transcription of dual-strand clusters in the *Drosophila* germline is non-canonical (Fig. 2B). Both H3K4me2 marks and Pol II peaks are absent from the promoters and TSS of dual-strand clusters, respectively [97]. In addition, splicing and transcription termination are inhibited, which is hypothesized to differentiate these piRNA precursors from other transcripts [97,109] (Fig. 2B). This non-canonical transcription process in the *Drosophila* germline is governed by the Rhino-Deadlock-Cutoff (Rhi-Del-Cuff, RDC) complex [97,109]. Of the three components of the RDC complex, which licenses bidirectional transcription, *rhi* is not conserved throughout the *Drosophila* genus, and orthologs of *del* are not present beyond the order Diptera [110], suggesting that the bias toward bidirectional transcription might have evolved with the RDC complex in the germline of *D. melanogaster* and, possibly, other drosophilids. Hence, unidirectional transcription might be a general, unifying feature of piRNA clusters in many metazoans, including non-bilaterians. Interestingly, piRNA precursor transcription may be controlled by distinct transcription factors: precursors of the mouse testicular pachytene piRNAs (the most abundant class of adult murine piRNAs) are regulated by the A-MYB transcription factor, whereas the transcription of *C. elegans* piRNAs, which originate from independent loci instead of piRNA clusters, is governed by Forkhead family transcription factors [100,111]. Although the bulk of piRNAs is supplied by precursors that are transcribed from dedicated genomic loci, it is worthwhile to note that up to a third of piRNAs may originate from the 3′ untranslated regions (UTRs) of many messenger RNAs (mRNAs) (discussed later) [112].

piRNA precursors are then transported out of the nucleus, although our current knowledge of the potential mechanisms is limited to the transport of precursors from dual-strand clusters in the fly germline [113] (Fig. 2B). The nuclear DEAD box RNA helicase, 56-kDa U2AF-associated protein (UAP56), which also participates in suppressing the splicing of dual-strand cluster precursors [109], is proposed to couple bidirectional cluster transcription to piRNA processing in the nuage [113] (Fig. 2B). UAP56 delivers precursor transcripts to the cytoplasmic DEAD box helicase, Vasa (Vas), a piRNA pathway nuage component [113] (Fig. 2B).

2.1.2. Primary piRNA biogenesis

Precursor transcripts are processed into mature primary piRNAs within the cytoplasm. First, piRNA precursors are resolved of secondary structures by the RNA helicase Armitage (Armi), also known as Moloney leukemia virus 10-like 1 (MOV10L1) in mouse [114,115] (Fig. 2C). This unwinding facilitates cleavage by the mitochondria-associated endonuclease, Zucchini (Zuc), the murine ortholog of which is known as mitochondrial phospholipase D (mitoPLD), to generate pre-piRNAs that bear a 5′ monophosphate, a feature of mature piRNAs [93,94,116–119] (Fig. 2C). Pre-piRNAs are then loaded onto PIWIs and the 3′ ends of the piRNAs are believed to be formed after trimming by an unidentified exonuclease [120] (Fig. 2C). A potential candidate is the 3′-to-5′ exonuclease,

Please cite this article in press as: Lim RSM, Kai T. A piece of the pi(e): The diverse roles of animal piRNAs and their PIWI partners. Semin Cell Dev Biol (2015), http://dx.doi.org/10.1016/j.semcdb.2015.10.025