Biochimica et Biophysica Acta 1839 (2014) 463–475

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Global characterization and target identification of piRNAs and endo-siRNAs in mouse gametes and zygotes

Jesús García-López [a], Juan de Dios Hourcade [a], Lola Alonso [b], David B. Cárdenas [a], Jesús del Mazo [a,*]

[a] Department of Cellular and Molecular Biology, Centro de Investigaciones Biológicas (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain
[b] Bioinformatics Service, Centro de Investigaciones Biológicas (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 24 March 2014
Accepted 7 April 2014
Available online 24 April 2014

Keywords:
piRNAs
Endo-siRNA
Deep sequencing
Transposable element
Germ cell
Zygote

ABSTRACT

A set of small RNAs known as rasRNAs (repeat-associated small RNAs) have been related to the down-regulation of Transposable Elements (TEs) to safeguard genome integrity. Two key members of the rasRNAs group are piRNAs and endo-siRNAs. We have performed a comparative analysis of piRNAs and endo-siRNAs present in mouse oocytes, spermatozoa and zygotes, identified by deep sequencing and bioinformatic analysis. The detection of piRNAs and endo-siRNAs in the spermatozoa and revealed also in zygotes, hints to their potential delivery to oocytes during fertilization. However, a comparative assessment of the three cell types indicates that both piRNAs and endo-siRNAs are mainly maternally inherited. Finally, we have assessed the role of the different rasRNA molecules in connection with amplification processes by way of the "ping-pong cycle". Our results suggest that the ping-pong cycle can act on other rasRNAs, such as tRNA- and rRNA-derived fragments, thus not only being restricted to TEs during gametogenesis.

1. Introduction

Mammalian genomes contain an abundant number of repeated DNA sequences, more than 50% in humans [1]. Highly repetitive DNA is normally localized in the genome in tandem clusters of multiple copies of untranslated regions, as for example satellite DNA, as well as in clusters of transcribed genes such as histones, ribosomal RNA (rRNAs) and transfer RNAs (tRNAs) [2,3]. However, a large number of DNA sequence repeats correspond to transposable elements (TEs) categorized in three main groups as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) or long terminal repeats (LTRs). Since repetitive DNA regions can be transcribed into RNA molecules, the transcripts from LINEs, SINEs or LTRs can be self-propagated via RNA mediated transposition. TEs have played decisive roles in the evolution of genomes having accumulated in large portions of eukaryotic genomes. However, their transposition can induce, among other alterations, insertion mutations in protein-coding genes, enhancer or promoter modifications, abnormal chromatin rearrangements and also chromosome breakage [4].

Maintaining genomic integrity is essential during germ cell formation and subsequently in the early morphogenetic development of multicellular organisms. Mechanisms involved in genome protection evolved to guard from potentially deleterious transcription effects. Some groups of small RNAs, acting through RNA interference, have been related to the posttranscriptional regulation of genes encoded by highly repeated DNA [5,6]. These repeat-associated small RNAs (rasRNAs) include piRNAs and endo-siRNAs [7–9]. piRNAs and endo-siRNAs are generated by different pathways, yet both can act through RNAi mechanisms that control transposon repression and TE mobility [6,10–12]. The interaction between piRNAs/endo-siRNAs and repeat elements of genomes is dual, on the one hand repeat transcripts can be the source of piRNA and endo-siRNA biogenesis, and on the other hand transcripts of repeat elements are targeted by rasRNAs.

In mouse, piRNAs are ≈23–32 nt long RNAs that act in association with PIWI proteins: MILI, MIWI and MIWI2 [13–15]. The biogenesis of piRNAs is not completely understood, nonetheless, two different generation pathways seem to be involved. The first way comprises the expression of specific areas of the genome encoding large RNA precursor molecules, which subsequently must be cleaved into smaller fragments [16,17]. The second however is more related to their function. It has been established that piRNAs bind to TEs [18,19] and that the binding between each piRNA and their TE target RNA takes place at the 5′ end of the piRNA sequence. Upon binding, poorly characterized proteins participate in the cleavage of TE transcripts after nucleotide position 10 at the 5′ end of the piRNA binding site. Cleavages generate new piRNAs derived from the RNA of TEs [18,19] in a process known as the "ping-pong model" or "ping-pong cycle" [14,20,21]. The ping-pong cycle generates characteristic cleavage specific piRNA sequence signatures [14,20] in this manner, piRNAs whose targets are the TE transcripts show a bias for uridine at the 5′ end (*5U*), whereas the piRNA sequences generated from TE cleavage show a bias for nucleotide

* Corresponding author. Tel.: +34 91 837 3112; fax: +34 91 536 0432.
E-mail address: jdelmazo@cib.csic.es (J. del Mazo).

http://dx.doi.org/10.1016/j.bbagrm.2014.04.006
1874-9399/

adenosine at position 10 (*10A*) [14,20]. These characteristics have been conserved in such diverse organisms as the mouse, *Drosophila*, zebrafish and silkworm [9,10,22,23].

Mutated PIWI family genes have been related to severe defects in male fertility [24,25], on the contrary, female fertility was noticed to be unaffected by PIWI protein inactivation [24,25]. Female fertility and early development progression seem to be associated with the presence and activity of DICER [26–30] and AGO2 [31–34]. These observations have been related to the functions of endogenous small interference RNAs (endo-siRNAs) present in the female germline.

Endo-siRNAs are ≈21–24 nt long sequences derived from convergent sense/antisense transcripts yet mainly proceed from transposon transcripts [6,18,35]. Contrary to piRNAs, endo-siRNAs have been determined as specific of the female germline and preimplantation embryos [18,35,36]. DICER and AGO2 proteins participate in endo-siRNA biogenesis and silencing activity respectively, while *Dicer* and *Ago2* null mice displayed lethality during early embryogenesis [37–39]. The regulatory role of endo-siRNA during mammalian gametogenesis is still poorly understood and endo-siRNA inheritance from gametes to embryo has not been clarified yet.

The piRNAs and endo-siRNAs are also known to hold crucial roles in transposon silencing in order to avoid epigenetic disorders that could compromise embryo viability at early developmental stages. In the present study, we have used high throughput sequencing technology to perform a global characterization of piRNA and endo-siRNA sequence populations present in mouse spermatozoa, oocytes and zygotes. Based on the RNA targets of piRNAs and endo-siRNAs we have evaluated their potential roles with regard to gamete differentiation and fertilization. Finally, we have tracked ping-pong cycle events from gametes to early zygote formation.

Our findings, suggest a vital role of maternally inherited piRNAs and endo-siRNAs in connection with the repression of transposons. However, transposable element regulation appears not to be the exclusive role performed by rasRNAs. We have also detected an elevated ratio of piRNAs and endo-siRNAs derived from rRNA or tRNA molecules which could consequently be involved in translation regulatory pathways.

2. Materials and methods

2.1. Animals

All procedures relating to the care and handling of the animals used in the present study, were carried out in the CIB-CSIC bioterium under specific pathogen-free (SPF), temperature (22 ± 1 °C) and humidity-controlled (50–55%) conditions. All animals were housed on a 12 h light/dark cycles with ad libitum access to food and water. Animal care and handling was carried out in accordance with the regulations of the Bioethics Committee of the *Consejo Superior de Investigaciones Científicas* (*CSIC*) that approved the study, and adhering to the European Commission guidelines.

2.2. Spermatozoa collection

Cauda epididymis and vasa deferentia from mature CD-1 males were collected in 500 μl of M2 medium after which adipose tissue and blood vessels were removed. Cleaned structures were placed in a new 200 μl drop of M2 medium covered with mineral oil. Epididymal fluid was squeezed out and sperm was suspended in M2 medium. Concentrations were determined with a Neubauer hemocytometer. The sperm suspension was layered in a 15 ml conical tube on a discontinuous 90%/45% Percoll gradient in a 1:1:1 ratio (cell sample: 45% Percoll: 90% Percoll) and centrifuged at 700 *g* during 20 min. The bottom pellet was washed in PBS and centrifuged at 350 *g* during 5 min. In order to entirely remove contaminant cells, a hiposmotic shock was induced by the resuspension of spermatozoa in H_2O-DEPC. Finally, the pellet was resuspended and homogenized in 100 μl of TRIzol® Reagent (Invitrogen) and frozen at −80 °C until use.

2.3. Oocyte and zygote collection

Fully-grown oocytes and zygotes were collected from the oviducts of mice as have been described previously [40]. Briefly, superovulation was induced in 4–5-week-old C57BL6 female mice by intraperitoneal injection of 5 IU pregnant mare serum gonadotropin (PMSG), followed 48 h later by 5 IU of human chorionic gonadotropin (HCG). After HCG administration, female mice were mated with DBA/6J males. Fully-grown oocytes and zygotes were treated with hyaluronidase (300 μg/ml, Sigma H3884) in order to remove cumulus cells by passage through several drops of M2 medium (Sigma M7167). To carry out the high-throughput sequencing a total of 15,210 fully-grown oocytes and 15,416 zygotes were collected. In all cases, the zona pellucida was removed by incubation in 30 μl of prewarmed acidic Tyrode solution under mineral oil during approximately 15 s, followed by careful washing in three drops of M2 medium under a stereomicroscope. Samples were stored in TRIzol® Reagent (Invitrogen) and frozen at −80 °C until use.

2.4. RNA purification and sequencing

Total RNA from 15,200 metaphase II oocytes, 500 million of spermatozoa and 15,400 zygotes was isolated using TRIzol® Reagent (Invitrogen) according to the instructions provided by the manufacturer. RNA concentrations were quantified measuring absorbance (A260/280 ratio) on a NanoDrop Spectrophotometer ND-1000 (NanoDrop). RNA integrity, considering the particular profile of spermatozoa, was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). Illumina protocols (available on the Illumina website at www.illumina.com/support) were followed to prepare the small RNA libraries from total RNA. Briefly, 3′ and 5′ adapters were ligated to each RNA molecule existing in total RNA previously isolated from spermatozoa, oocytes and zygotes. After adapter ligations an RT reaction was performed to cDNA synthesis. In order to avoid any bias, the cDNA was PCR amplified using common primers that were designed against adapter sequences. After cDNA amplification, the small RNA fraction was isolated by gel purification. Small RNA fraction was sequenced by high-throughput deep sequencing using an Illumina® Hiseq2000 sequencer.

2.5. Data analysis

Adaptor sequences were identified and removed (DNAVision SA, Belgium). Sequencing using the Illumina platform yielded sequences with a length range of 18–32 nucleotides. Using bioinformatics approaches we proceeded to classify all the small RNA sequences according to their identity with the annotated small RNA sequences found in public databases. The remaining sequences that had no identity with any small RNA sequence, ncRNA sequence or mRNA sequence was classified as unannotated (see Supplemental Fig. S5). The term "read" will strictly refer to the number of times that any RNA sequence has been detected, the term "sequence" will imply the specific succession of nucleotides sequenced, and the term "tag" will stand for the identifier of the defined element to which each sequence belongs. Therefore, the sum of all reads concerning a defined tag represents its expression value. For example, regarding the sequences: *TAAAACTGTATTTGAATTTGGGGC*; *TAAAACTGTATTTGAATTTGGGGCGA* (*read twice*) and *TAAAACTGTATTTG AATTTGGGGCGAGC*; this case represents 4 reads and 3 sequences although only 1 tag, the underlying reason being that all these sequences which differ in length were previously identified as the piRNA PIR222563 of the "RNAdb 2.0" [41–43]. Therefore, the sum of reads represents the expression value of the piRNA PIR222563. In order to compare the tag representation of different samples, we applied the DeSeq tool of the R/Bioconductor software package [44–46].