# Two modes of targeting transposable elements by piRNA pathway in human testis

ILDAR GAINETDINOV,[1] YULIA SKVORTSOVA,[1] SOFIA KONDRATIEVA,[1] SERGEY FUNIKOV,[2] and TATYANA AZHIKINA[1]

[1]Department of Genomics and Postgenomic Technologies, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997, Russia
[2]Department of Structural, Functional and Evolutionary Genomics, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991, Russia

## ABSTRACT

**PIWI proteins and their partner small RNAs, termed piRNAs, are known to control transposable elements (TEs) in the germline. Here, we provide evidence that in humans this control is exerted in two different modes. On the one hand, production of piRNAs specifically targeting evolutionarily youngest TEs (L1HS, L1PA2-L1PA6, LTR12C, SVA) is present both at prenatal and postnatal stages of spermatogenesis and is performed without involvement of piRNA clusters. On the other hand, at postnatal stages, piRNAs deriving from pachytene clusters target "older" TEs and thus complement cluster-independent piRNA production to achieve relevant targeting of virtually all TEs expressed in postnatal testis. We also find that converging transcription of antisense-oriented genes contributes to the origin of genic postnatal prepachytene clusters. Finally, while a fraction of pachytene piRNAs was previously shown to arise from long intergenic noncoding RNAs (lincRNAs, i.e., pachytene piRNA cluster primary transcripts), we ascertain that these are a specific set of lincRNAs that both possess distinguishing epigenetic features and are expressed exclusively in testis.**

**Keywords: PIWI proteins; piRNA cluster; piRNA pathway; transposable elements**

## INTRODUCTION

A special class of small RNAs called piRNAs (PIWI-interacting RNAs) has been extensively studied since its initial discovery more than 10 years ago (Aravin et al. 2001, 2003, 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Vagin et al. 2006). One of the primary functions of the piRNA pathway is control of transposable elements (TEs) (Vagin et al. 2004; Aravin et al. 2007, 2008; Brennecke et al. 2007; Carmell et al. 2007; Kuramochi-Miyagawa et al. 2008; Gebert and Rosenkranz 2015; Rosenkranz et al. 2015b). Coincidentally, expression of piRNA pathway proteins and piRNAs themselves is mostly restricted to germ cells, where TEs pose a threat to genomic stability and faithful transmission of hereditary information to the next generation.

The role of the piRNA pathway in germ cell development has been codified by a host of studies in a range of model organisms (for reviews, see Han and Zamore 2014; Le Thomas et al. 2014; Czech and Hannon 2016; Hirakata and Siomi 2016). In the mammalian testis, the current picture of piRNA biogenesis was obtained from experiments in mice. There are three stages of piRNA production during animal development: fetal/prenatal, prepachytene/postnatal, and pachytene/postnatal. At the fetal/prenatal stage, TE-derived piRNAs bound to PIWI proteins have been shown to participate in transcriptional silencing of these TEs through DNA methylation (Aravin et al. 2007, 2008; Kuramochi-Miyagawa et al. 2008; Manakov et al. 2015; Nagamori et al. 2015; Kojima-Kita et al. 2016). Later in development, most postnatal piRNAs are produced from specific genomic regions termed piRNA clusters and divided into two types, depending on the stage of their expression: prepachytene clusters arising from genic sequences, and pachytene clusters of intergenic origin (Li et al. 2013a). The details of both biogenesis and function of these two types of postnatal piRNAs in the developing testis remain unclear; however, several reports have proposed various hypotheses relating pachytene piRNAs to TE silencing or the control of gene expression in either an indiscriminate or targeted manner (Gou et al. 2014; Hirano et al. 2014; Gebert et al. 2015; Goh et al. 2015; Zhang et al. 2015; Vourekas et al. 2016).

**Corresponding author: ildargv@gmail.com**
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.060939.117.

Though earlier work established that piRNA pathway features are conserved between humans and other mammals (Aravin et al. 2006; Girard et al. 2006; Yang et al. 2013; Ha et al. 2014; Roovers et al. 2015; Rosenkranz et al. 2015a; Rounge et al. 2015; Williams et al. 2015), due to ethical and technical limitations, research progress in humans has been slower. Specifically, while all postnatal human piRNAs were previously mapped to piRNA clusters (Ha et al. 2014; Roovers et al. 2015; Rosenkranz et al. 2015a; Williams et al. 2015), these clusters were ill-defined in the genome. Also, no prior study has performed a thorough analysis of TE targeting by the piRNA pathway at different stages of spermatogenesis (i.e., prenatal and postnatal). In this work, we conducted a comprehensive analysis of both newly generated small RNA libraries and dozens of publicly available small RNA and RNA-seq libraries. These concordant analyses allowed us to reveal new features of TE targeting in the human testis, as well as propose possible determinants governing commitment of a genomic locus to piRNA production.

## RESULTS

### Definition of human postnatal prepachytene and pachytene piRNA clusters

In order to study piRNAs targeting transposable elements (TEs) in human testis, we generated small RNA libraries from postnatal testis and also used publicly available data sets from both prenatal and postnatal testis (Supplemental Table S1A; Ha et al. 2014; Williams et al. 2015). Notably, as in previous studies on human, marmoset, and mouse small RNAs (Hirano et al. 2014; Rounge et al. 2015; Kumar et al. 2016; Sharma et al. 2016), we observed the presence of 31- to 34-nt-long 5′-tiRs (long 5′ fragments of tRNAs, Kumar et al. 2016). Moreover, using modified bioinformatics pipeline (see Materials and Methods), we were able to detect non-templated tailing of miRNAs, piRNAs, and 5′-tiRs with different biases for tailed nucleotides and nucleotide combinations (Supplemental Tables S1B, S1B1 and Supplemental Fig. S1).

Since most piRNAs in postnatal testis are known to be produced from a limited number of genomic loci named piRNA clusters (Rosenkranz 2016), we defined these clusters in humans using an algorithm empirically devised in our group and based on the approach used in seminal studies (see Materials and Methods) (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006). We also compared this method with the probabilistic algorithm proposed by Rosenkranz and Zischler (2012) and the meta-analysis performed by Chirn et al. (2015). As a matter of fact, overlap among the three sets of clusters (empirical, probabilistic, and from the meta-analysis) was as high as 80%–100% (median: 94%, Supplemental Table S1C). Consequently, to gain complementary perspectives during further analyses, we used all three piRNA cluster sets and confirmed that the results obtained with each set did not diverge significantly. All three sets were defined with a 10 RPKM threshold, which is relevant to the values used in the original studies (1 RPKM threshold was also used and yielded similar results; data not shown).

Importantly, postnatal testis in mammals is reported to possess two major piRNA cluster populations: prepachytene and pachytene—named after temporally separated stages of spermatogenesis at which they are found. The locations of these clusters are different: Prepachytene piRNAs originate mostly from clusters overlapping 3′UTRs and exons of coding genes, whereas pachytene piRNAs arise from intergenic loci (Li et al. 2013a,b). Another distinguishing feature of these two populations is the piRNA length bias: Prepachytene piRNAs peak at 26 nt since PIWIL2 is the major PIWI protein expressed at this stage, unlike pachytene piRNAs that are almost exclusively 30 nt long because PIWIL1 is more abundant at the pachytene stage (Li et al. 2013a,b). This length bias was recently confirmed by IP experiments in humans (Williams et al. 2015) and we also could observe two primary peaks of piRNAs in the postnatal testis samples analyzed in this study (Supplemental Fig. S1). Therefore, because there are no available data sets of small RNAs from different stages of spermatogenesis for humans, we used the length of reads to computationally separate prepachytene from pachytene piRNA clusters. Clusters were considered prepachytene if the fraction of 26-nt reads was bigger than the fraction of the 30-nt reads fraction, and vice versa for pachytene clusters. Using just this parameter appears to be sufficient: 26%–59% (median: 53%) of the length of presumably prepachytene (26 nt) piRNA clusters overlaps coding exons/3′UTRs, while 77%–93% (median: 83%) of the length of presumably pachytene (30 nt) piRNA clusters maps to intergenic regions (Supplemental Fig. S2). Hence, we will further refer to these computationally staged sets as prepachytene and pachytene clusters, respectively. Additionally, since pachytene clusters account for about 10 times more piRNA reads than prepachytene ones (~40% versus ~4%, Supplemental Fig. S2), we will mostly focus on pachytene clusters as a source of piRNAs targeting TEs.

### Postnatal piRNAs relevantly target most expressed TEs through a combination of cluster-dependent and cluster-independent biogenesis

To study TE targeting by piRNAs, we used the following metrics of TE activity and TE-derived piRNA production: genomic, transcriptomic, piRNA cluster fractions as well as piRNAs produced from inside and outside piRNA clusters, ping-pong *Z*-score, 1U/10A bias. Since not all copies of TEs are still active, the genomic fraction of a TE is expected to reflect its evolutionary history of expansion. Therefore, we built our analyses around the TE transcriptomic fraction because transcription is the closest step to a retrotransposition event of a TE. We also invoked TE age rank determined by defragmentation approach and recent studies on evolutionarily young TEs to assess the relationship between TE

age and other parameters used in the study (Boissinot et al. 2000; Ostertag et al. 2003; Khan et al. 2006; Giordano et al. 2007; Bennett et al. 2008). Specifically, we will further refer to L1PA subfamilies of LINEs, AluY subfamilies of SINEs, LTR12C, and SVA elements as young TEs.

Initially, to gain a perspective of TE transcriptional activity across human somatic and germline tissues, we compared TE transcription in 26 publicly available RNA-seq data sets (Supplemental Table S2A) and found that TE transcriptional landscapes do not vary significantly between testis and somatic tissues both at prenatal and postnatal stages (Pearson's $r = 0.82–0.99$, median: 0.96, Supplemental Table S2B, S2B1). Active youngest TE subfamilies contribute the most variability between the tissues (e.g., relative standard deviation >0.8 for L1PA2-7, LTR12C, SVA_D, etc.; Supplemental Table S2C). Furthermore, the TE transcriptional landscape also appears to be virtually undisturbed throughout human male primordial germ cell development (PGCs, weeks 4–19, single-cell RNA-seq data, 149 samples, Pearson's $r = 0.48–0.99$, median: 0.94, Supplemental Table S2D; Guo et al. 2015). Similarly, TE subfamilies with the most variable RPM (reads per million) values in PGCs were the youngest ones (e.g., RSD > 0.7 for L1PA2-6, AluY subfamilies, Supplemental Table S2E). However, we noticed that though the level of expression of different TEs linearly correlated between various data sets, the total level of TE transcriptional activity (the sum of RPM values for all TEs) was different between fetal somatic and fetal testis tissues (total values in Supplemental Table S2B1; Supplemental Fig. S3A). This could be due to the fact that fetal PGCs undergo genome demethylation and concomitant chromatin resetting, which could lead to activation of TE transcription (Tang et al. 2015).

Interestingly, in all RNA-seq data sets, transcription of "older" TE subfamilies was enriched in the 3′UTRs of genic mRNAs, while their intergenic counterparts were depleted (Supplemental Fig. S3B). On the other hand, youngest TEs were transcribed from all genomic locations where they happened to have integrated (examples of L1PA3 and L1PA4 in Supplemental Fig. S3B).

When comparing piRNA repertoires across various TE subfamilies in different postnatal testis samples, we found them to be very similar (Pearson's $r = 0.77–0.99$, median: 0.95, Supplemental Table S2F). Crucially, these piRNA repertoires also resembled TE transcriptional landscapes from postnatal testis (Pearson's $r = 0.57–0.87$, median: 0.67, Supplemental Table S2F; Figure 1A). That is, transcriptional landscape in postnatal testis is relevantly represented by the corresponding amounts of piRNAs for each TE subfamily. Outliers were mostly comprised of LTR TEs, which are overrepresented in the piRNA repertoire compared to their fractions in transcriptome (Fig. 1A). It could be explained by the abundant production of LTR-deriving piRNAs from pachytene clusters. Indeed, if the genomic fraction of a TE in piRNA clusters is compared with random intergenic controls, it is LTRs that appear to be enriched in pachytene clusters (Fig. 1B). In contrast, some L1 elements seem to be depleted from pachytene clusters compared to the random intergenic controls (Fig. 1B), which could be due to their preference to integrate in gene-poor heterochromatic regions and the notion that chromatin structure of pachytene clusters resembles that of protein-coding genes in euchromatic areas of the genome.

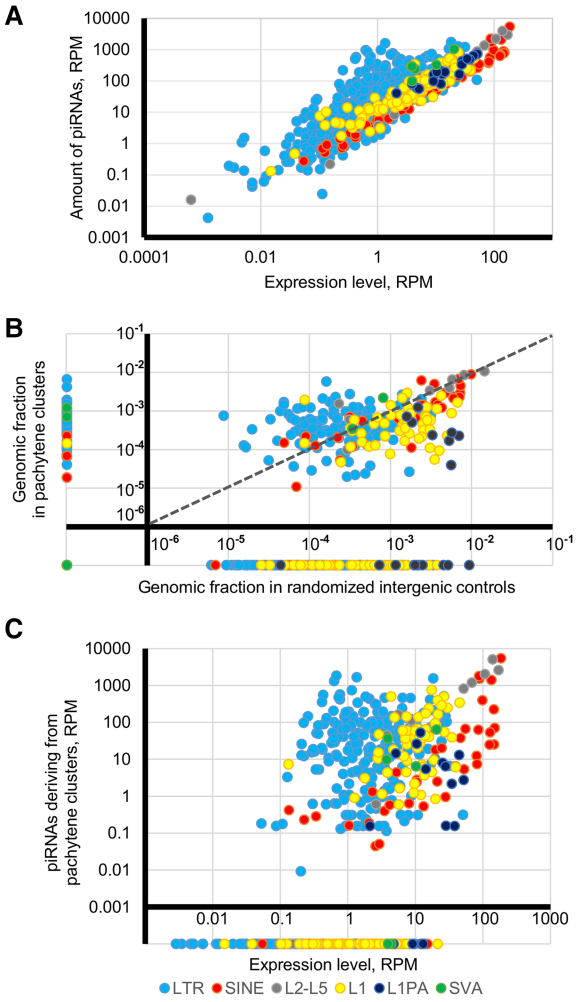


**FIGURE 1.** piRNAs deriving from transposable elements (TEs) in postnatal testis. (*A*) Relationship between expression level of a TE and amount of all piRNAs deriving from it. (*B*) Relationship between genomic fractions of TE subfamilies in pachytene clusters and in randomized intergenic controls demonstrates enrichment of pachytene piRNA clusters with LTRs and depletion of LINE elements from them. TE subfamilies not presented in randomized controls or pachytene clusters are shown to the *left* of the *y*-axis and *below* the *x*-axis, respectively. (*C*) Relationship between expression level of a TE and amount of piRNAs deriving from pachytene clusters only. TE subfamilies without reads deriving from pachytene clusters are shown *below* the *x*-axis. All results are presented for pachytene clusters constructed with empirical approach (averaged across all postnatal testis small RNA-seq data sets under study and three postnatal testis RNA-seq data sets).

Similar results were obtained when we looked at unique piRNA reads that could be unambiguously assigned either inside or outside pachytene piRNA clusters (Fig. 1C).

Strikingly, young subfamilies of L1s (L1PA2-7 and L1HS) as well as LTR12C and SVA_D are either depleted from piRNA fractions in pachytene clusters or are completely absent from those (Fig. 1B,C; Supplemental Table S2G). It implies that targeting of these TEs by piRNAs is performed without involvement of cluster-driven piRNA production, since a relevant amount of piRNAs matching these TEs seems to derive from outside clusters (compare Fig. 1A,C).

At the same time, for approximately half of those TEs that had a considerable share of piRNAs arising from pachytene clusters, we were able to detect a moderate-to-significant trend for more cluster piRNAs to be antisense to TEs than piRNAs from outside clusters (Supplemental Table S2H). Another finding was the length bias of TE-deriving piRNAs from clusters that tend to be 30 nt long, whereas those from outside clusters are predominantly 26 nt (Supplemental Table S2I). This fact might reveal a possible separation of roles between different PIWI proteins.

Notably, we could also observe the essential signs of TE silencing by piRNAs. First, compared to transcriptome, there are considerable 1U and 10A biases characteristic for primary and secondary piRNA biogenesis, respectively (Supplemental Table S2J,K). Second, the *Z*-score for ping-pong signature is higher than the threshold of 2.58 for the majority of TE subfamilies (except most MIRs and some old LTRs, Supplemental Table S2L).

Because different classes of TEs use different strategies for expansion and exhibit various evolutionary histories, we further analyzed the relation between age and genomic/transcriptomic/piRNA clusters fractions for each of the three major classes separately. First, for SINEs, the most remarkable feature was a high degree of correlation between SINE subfamily fractions in the genome and the transcriptome (Pearson's $r = 0.95–0.99$, Supplemental Table S2M) and fractions in the transcriptome and piRNA repertoire (Pearson's $r = 0.59–0.97$, median: 0.74, Supplemental Table S2M). Furthermore, there is also a positive correlation between piRNA fraction and the age of a SINE subfamily (Pearson's $r = 0.73–0.83$, median: 0.81, Supplemental Table S2M). In fact, the degree of correlation between the age and genomic/transcriptomic/piRNA/piRNA cluster fractions of a SINE subfamily appears to be higher than for other classes of TEs (compare Supplemental Table S2M–O). One of the possible explanations could be the fact that Alu elements, which are the most abundant type of SINE class, hijack L1 reverse transcriptase for their retrotransposition (Dewannieux et al. 2003). Furthermore, less than half of the length of an Alu element is required for an actual retrotransposition (Ahl et al. 2015). Therefore, even an evolutionarily ancient and truncated insertion might still be capable of serving as substrate for a new retrotransposition event. From an evolutionary perspective, this setting and the higher positive correlations between genomic/transcriptomic/piRNAs/piRNA cluster fractions could support the following hypothesis: The older a SINE subfamily is, the longer it has been retrotransposing and the bigger the fraction of the genome it has taken and the more transcripts it produces. In the end, older SINE subfamilies take a larger fraction of pachytene piRNA clusters, which produce more piRNAs.

On the contrary, while LTR-containing TE subfamilies demonstrate a positive correlation between their genomic and transcriptomic fractions (Pearson's $r = 0.62–0.83$, Supplemental Table S2N) as well as transcriptomic and piRNA fractions (Pearson's $r = 0.38–0.62$, median: 0.50, Supplemental Table S2N), relation to age is almost uncorrelated (Pearson's $r = 0.06/0.10–0.17$ for genomic/transcriptomic fractions, Supplemental Table S2N). This might be due to the heterogeneity of LTR subfamilies having disparate evolutionary histories of expansion. In addition, the contrast between SINE and LTR piRNA fractions' relation to the age could be connected with different mechanisms for generation of transcripts. Indeed, transcription of SINEs is believed to be performed by PolIII, which does not need a load of specific transcription factors (TFs). Conversely, LTR promoters are recognized by PolII requiring certain combination of TFs, and mutations accumulating with time might easily disturb this intricate assembly, leading to a restricted evolutionary period of LTR active transposition.

Finally, when studying LINE elements, we were able to observe a pattern similar to LTRs: positive correlation between genomic and transcriptomic shares (Pearson's $r = 0.84–0.87$) and between transcriptomic and piRNA fractions (Pearson's $r = 0.47–0.90$, median: 0.81). However, no clear relation to age was detected. Only when we focused on the youngest primate-specific subgroup of LINEs (L1PA) did we note a negative correlation between age and genomic/transcriptomic fractions (Pearson's $r = -0.56/-0.67–0.70$, respectively, Supplemental Table S2O) as well as with piRNA shares (Pearson's $r = -0.28–0.77$, median: $-0.66$, Supplemental Table S2O). This might imply that younger L1PA TEs are more successful at acquiring genomic space and producing transcripts, which are targeted by a bigger piRNA fraction. Notably, although all piRNAs deriving from L1PA members possess a 1U bias (primary biogenesis feature, Supplemental Table S2P), only the youngest ones (L1PA1-L1PA5) display a prominent 10A bias (secondary biogenesis sign, Supplemental Table S2P). As noted earlier, most L1PA elements lack their copies within pachytene clusters and are presumably targeted by piRNAs with the help of a mechanism different from pachytene cluster-driven piRNA biogenesis.

## piRNA pathway targets only evolutionary young TE subfamilies in prenatal testis

When we analyzed small RNA data sets from prenatal testis, we found that piRNAs do not arise from clusters as in postnatal testis, which is consistent with previous studies on mice and humans (Aravin et al. 2008; Williams et al. 2015). Moreover, young subfamilies constitute the top TEs driving piRNA production at this stage (Fig. 2A). It is known that

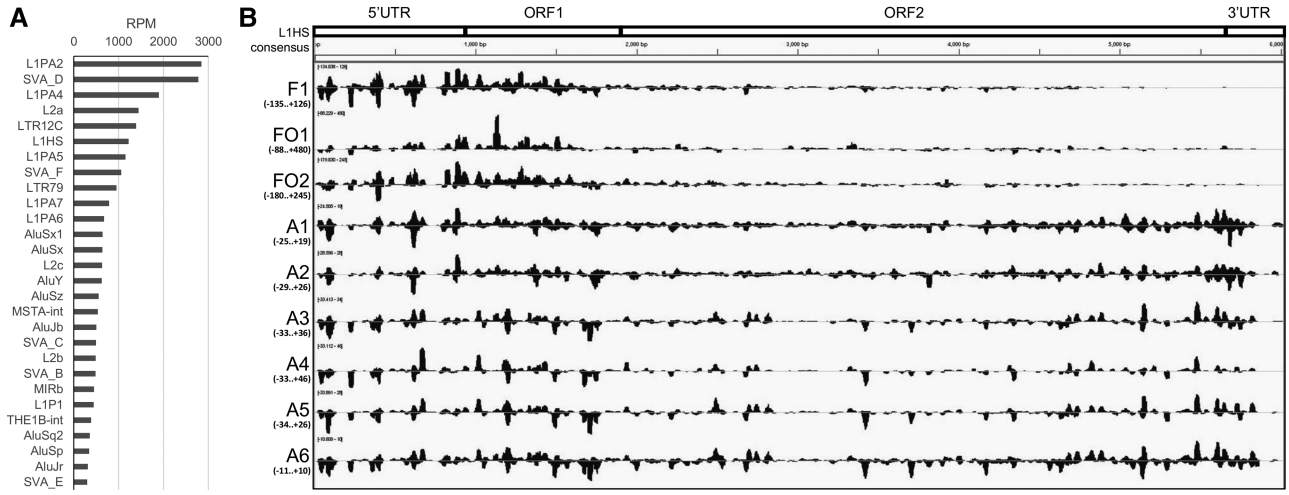


**FIGURE 2.** Production of piRNAs deriving from transposable elements (TEs) in prenatal testis. (*A*) Top 30 TE subfamilies producing most piRNAs sorted in descending order for library F1 (prenatal testis, nonoxidized). (*B*) Distribution of piRNA reads along L1HS consensus in prenatal (F1, nonoxidized; FO1–FO2, oxidized) and postnatal testis libraries (A1–A6, nonoxidized). Scale ranges in RPM (reads per million) are given in brackets.

PGCs undergo global genome demethylation that could lead to reactivation of these young TE subfamilies during development of fetal testis (Tang et al. 2015). However, this finding is still peculiar given the fact that the transcriptional landscapes of TEs in prenatal and postnatal testes are almost identical (Pearson's $r = 0.93–0.97$, median: 0.97, Supplemental Table S2B). In other words, at prenatal stage the piRNA pathway appears to be operating only in a cluster-independent manner targeting evolutionarily young subfamilies, while in postnatal testis, this mode is combined with cluster-dependent piRNA repertoire and together they cover the whole of TE transcriptional landscape.

Importantly, prenatal piRNAs targeting young subfamilies possess 1U (Supplemental Table S2Q) and 10A biases (Supplemental Table S2R) and display the *Z*-score for ping-pong signature higher than the threshold of 2.58 (Supplemental Table S2L). However, there is also a remarkable difference in distribution of L1PA-deriving piRNAs along TE consensus in prenatal and postnatal testis: Most piRNAs are concentrated around the 5′ ends of targeted L1PAs at prenatal stage, whereas in postnatal testis they evenly cover all the length of the TEs (Fig. 2B). This fact might imply involvement of transcriptional silencing in addition to posttranscriptional mechanisms for the control of L1PA expression. Together, these two routes might form a feed-forward loop in human prenatal testis similar to what was recently shown for flies (Han et al. 2015; Mohn et al. 2015; Senti et al. 2015; Wang et al. 2015).

## Prepachytene and pachytene piRNA clusters originate from converging transcription and testis-specific lincRNAs, respectively

Previous studies on prepachytene piRNA clusters probed for different factors defining genes whose 3′UTRs and exons will most contribute to production of piRNAs. Among possible determinants, level of gene expression was considered because PIWI machinery could simply recruit most expressed mRNAs into piRNA processing (Robine et al. 2009; Saito et al. 2009; Ha et al. 2014). During analysis of our data sets, we came to the same conclusion as earlier reports: There is no significant correlation between a gene's expression and its piRNA production levels. Also, many highly expressed genes do not produce piRNAs at all (Supplemental Fig. S4).

Nevertheless, a recurring setting came up with the top piRNA producing 3′UTRs: Each of them is almost invariably located adjacent to another piRNA-producing 3′UTR in antisense orientation so that both are transcribed in a converging manner (Fig. 3A). This layout might prompt piRNA production by slicing and further "phasing/inchworming" with PIWI proteins in the presence of two complementary strands (Han et al. 2015; Homolka et al. 2015; Mohn et al. 2015). Since ping-pong *Z*-scores for 3′UTR deriving piRNAs are above the threshold of 2.58 for almost all samples of postnatal testis (Supplemental Table S3A), this mechanism may indeed participate in the biogenesis of a prepachytene fraction of piRNAs. Notably, no signs of endo-siRNAs were detected for these cases (Fig. 3A, "21–23-nt reads" track, and low Z19-Z22 scores, data not shown).

To assess on the whole-genome scale, whether converging transcription is a common underlying feature of piRNA producing 3′UTRs, we looked at two parameters. First, how frequently is the adjacent 3′UTR located in antisense to the piRNA producing 3′UTR? For a random sample, the frequency is expected to be around 50%. Second, how close is a piRNA producing 3′UTR to another antisense oriented 3′UTR? For a randomly sampled set, the value will range from zero in the case of overlapping 3′UTRs to millions of bases. In fact, we observed a significant deviation of these

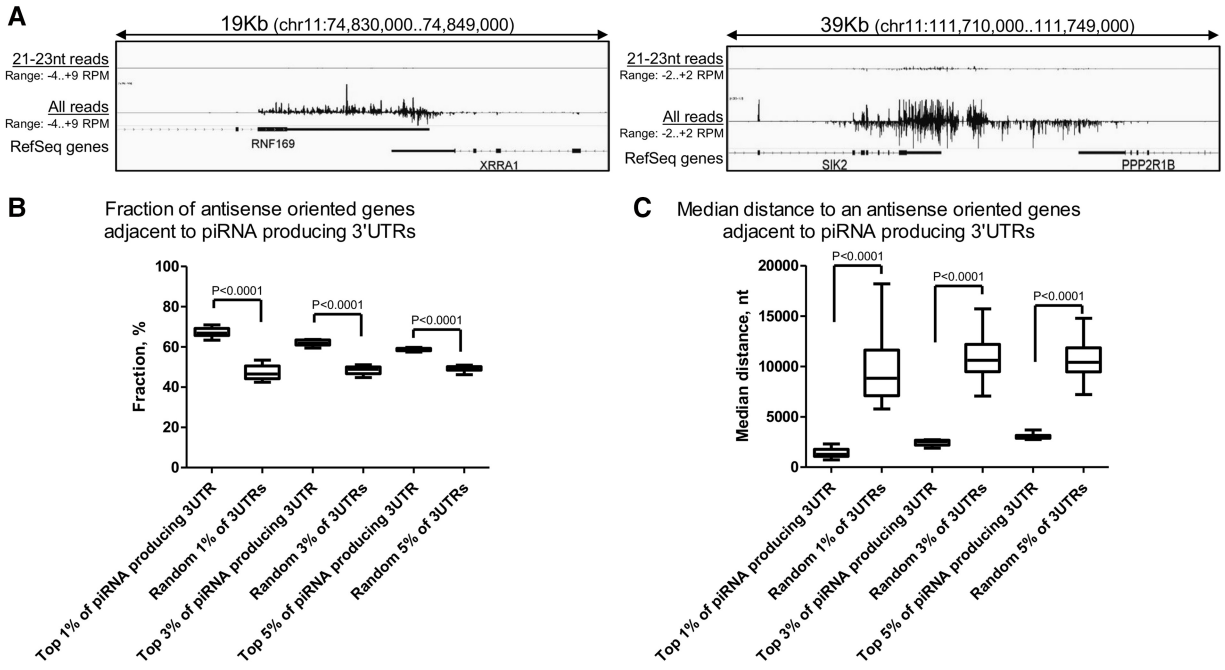


**FIGURE 3.** piRNA production from 3′UTRs is associated with converging transcription of genes (transcription of antisense-oriented 3′UTRs). (*A*) Examples of converging transcription of 3′UTRs accompanied by significant piRNA production. Genome coordinates (assembly hg38) and a separate track for 21- to 23-nt reads (putative endo-siRNAs) are shown. (*B*) Fractions of antisense oriented genes adjacent to 1/3/5% of top piRNA-producing 3′UTRs. Top piRNA-producing 3′UTRs were defined based on the density of piRNAs per kilobase. Controls are random size-matched sets of 3′UTRs. (*C*) Median distance to antisense-oriented genes adjacent to 1/3/5% of top piRNA-producing 3′UTRs. Top piRNA-producing 3′UTRs and controls were defined as in panel *B*. The graphs in panels *B* and *C* show median, interquartile range (box), and min/max range (whiskers) across 11 postnatal libraries. *P*-value for two-tailed Mann–Whitney test is presented.

two parameters for the top 1/3/5% piRNA-producing 3′UTRs (accounting for 28%–70% of all 3′UTR piRNA reads) from relevantly sized random sets of 3′UTRs (Fig. 3B,C).

Needless to say, this mechanism might not be the only one playing a role in the production of 3′UTR-derived piRNAs, and further experimental evidence is necessary to make the case for it and to assess its relative contribution. Incidentally, we compared different characteristics of piRNAs originating from the top 1/3/5% and the randomly chosen piRNA-producing 3′UTRs and found indirect evidence to support this mechanism: The top 1/3/5% piRNA-producing 3′UTRs exhibited a higher 10A bias and more unique piRNAs able to prime slicing through imperfect base-pairing (Supplemental Fig. S5).

The other type of piRNA genomic loci is referred to as pachytene clusters. As in marmoset, we were able to find pachytene clusters that are transcribed divergently from a promoter region shared with a protein-coding gene (Hirano et al. 2014): e.g., *BUB1*, *WDR1* as in marmoset and a new instance with *TDRD5*, which was described neither in mouse nor in marmoset.

However, most pachytene clusters tend to be located in intergenic regions and frequently overlap long intergenic noncoding RNAs (lincRNAs), unlike prepachytene clusters (Supplemental Fig. S6). An appreciable amount of human piRNAs was earlier reported to originate from lincRNAs; yet the original study did not divide them into prepachytene and pachytene fractions (Ha et al. 2014). Therefore, we looked into the nature of the relationship between these "piRNA-producing lincRNAs" and pachytene/30-nt piRNA clusters. While analyzing piRNAs deriving from "piRNA-producing lincRNAs," we observed a significant 10A bias and high ping-pong *Z*-score values for these piRNAs (Supplemental Table S3B). Another common feature of "piRNA-producing lincRNAs" was their testis-specific expression (Supplemental Fig. S7). We conducted a whole-genome survey with the data available from the GTEx project (RNA-seq assessed expression in 53 tissues from 544 human donors) (GTEx Consortium 2013; Melé et al. 2015). Importantly, a relevant "baseline" control is crucial in this analysis, since a considerable number of lincRNAs in annotation databases are found only in testis. In other words, when randomly sampling lincRNAs from any given annotation list, the chance of finding a testis-specific one is as high as 33% (Fig. 4), which is larger than for any other tissue. Furthermore, if this test is constrained to only testis-expressed lincRNAs, the number of testis-specific ones will increase to 50% (Fig. 4). Nevertheless, for "piRNA-producing lincRNAs," we observed that ∼72%–96% of them were testis-specific (Fig. 4). This finding, 10A bias, and ping-pong signature together imply that "piRNA-producing lincRNAs" should essentially be equated with pachytene piRNA primary transcripts.

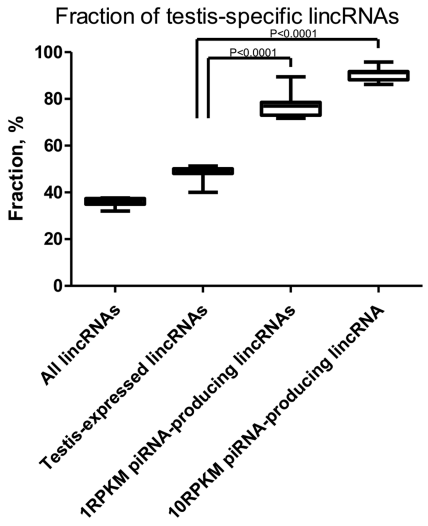


**FIGURE 4.** Expression of "piRNA-producing lincRNA" is mostly restricted to testis. Fraction of testis-specific lincRNAs among all lincRNAs, testis-expressed lincRNAs, and "piRNA-producing lincRNAs" selected with a threshold of 1 RPKM or 10 RPKM. The graph shows median, interquartile range (box), and min/max range (whiskers) across either 100 randomly sampled sets of lincRNAs (for all/testis-expressed lincRNAs from GENOCODE v.23) or 11 postnatal libraries (for 1RPKM and 10RPKM piRNA-producing lincRNAs). *P*-value for two-tailed Mann–Whitney test is presented.

We also observed only a very weak nonlinear correlation between the expression level of testis-specific lincRNAs and the amount of piRNAs produced from them (both normalized to length, Spearmans $\rho = 0.42$). This notion suggests that the level of lincRNA expression is not a strong predictor of piRNA production. Furthermore, not all testis-specific lincRNAs are "piRNA-producing lincRNAs" (Supplemental Fig. S8). Therefore, we probed for various epigenetic aspects as defining determinants of lincRNA commitment to piRNA production. First, using ChIP-seq data sets published for human pachytene spermatocytes (Lesch et al. 2016), we found that "piRNA-producing lincRNAs" are enriched for H3K4me3 and depleted for H3K27me3 chromatin modifications unlike all testis-expressed and all testis-specific lincRNAs (Supplemental Fig. S6). Importantly, pachytene clusters follow a similar pattern, which allows us to consider "piRNA-producing lincRNAs" and pachytene piRNA primary transcripts as the same entities. It is also in agreement with previous reports claiming that the epigenetic architecture of pachytene piRNA primary transcripts is similar to protein-coding genes: the presence of H3K4me3 enriched promoter region and lack of H3K27me3-marked facultative heterochromatin. Second, we looked at positions of "piRNA-producing lincRNAs" relative to such constitutive heterochromatic features as centromeres/telomeres and lamina-associated domains (LADs) (Guelen et al. 2008; Meuleman et al. 2013). Depletion of centromere/telomere fractions was detected for "piRNA-producing lincRNAs" compared to all testis-expressed/testis-specific lincRNAs with a similar pattern for pachytene clusters (Supplemental Fig. S6). On the contrary, a fraction of LADs was slightly but statistically significantly enriched in "piRNA-producing lincRNAs" compared to all testis-expressed/testis-specific lincRNAs (Supplemental Fig. S6). Taken together, these findings suggest that a testis-specific lincRNA is likely to be recruited for production of piRNAs (i.e., become a "piRNA-producing lincRNA"/pachytene piRNA primary transcript) given two features: protein-coding gene-like chromatin environment and proximity to lamina.

## DISCUSSION

A multitude of studies on the role of piRNAs and PIWI proteins in spermatogenesis in mammals have been published to date, but human studies have been cursory thus far. Here, we performed a comprehensive analysis of both newly generated small RNA libraries and publicly available data sets, enabling us to shed more light on piRNA biogenesis and TE targeting in prenatal and postnatal human testis.

Specifically, though previous reports noted some degree of proportionality between expression level and amount of piRNAs deriving from a TE (Mourier 2011; Vandewege et al. 2016), we studied this aspect in detail and found that the TE transcriptional landscape is matched by a relevant piRNA repertoire in the postnatal testis. This relationship is achieved through a combination of cluster-independent targeting of evolutionarily young TEs (L1HS, L1PA2-L1PA7, LTR12C, SVA_D), and pachytene cluster-driven targeting of "older" ones. Notably, only the first component controlling young TEs is functioning in prenatal testes. Among the possible candidates contributing to tethering piRNA machinery to young TEs are KRAB-ZNF proteins, since they have been shown to participate in silencing some L1PA and SVA elements in embryonic stem cells (Jacobs et al. 2014). Importantly, our findings in humans are in line with previous studies on silencing TEs in mammalian model organisms (De Fazio et al. 2011; Reuter et al. 2011; Di Giacomo et al. 2013; Hirano et al. 2014). The two routes to target TEs may reflect the need for both "piRNA cluster-dependent" long-term memory of TE invasion to control "older" repeats and an immediate ad hoc response, which is cluster-independent. Possibly, with time, as soon as a young TE subfamily member happens to integrate inside a pachytene piRNA cluster, its repression is mediated by cluster-deriving piRNAs, and evolutionary pressure to keep cluster-independent targeting might wear away.

Interestingly, only the youngest subfamilies of LINEs (L1Hs and L1PAs) exhibit negative correlation between their evolutionary age and abundance of piRNAs. This is probably because L1HS and several L1PA TEs are the only currently active autonomous retrotransposons in the human genome. At the same time, other LINEs and LTR TEs are mostly transpositionally incompetent and the selective pressure for the piRNA pathway to silence them is possibly decreased. On the contrary, abundance of SINE-derived piRNAs is positively

correlated with their age, genomic and transcriptomic fractions. This observation could be explained by the fact that even evolutionarily old but transcriptionally active Alu insertions are capable of retrotransposition and, therefore, need to be posttranscriptionally repressed. Conversely, since the fraction of actively retrotransposing TEs in *Drosophila melanogaster* is significantly higher than in humans and piRNA clusters in flies are specifically enriched in TE sequences, the relation between TE age and piRNA production is probably influenced by both transcriptional and transpositional activity of a TE in combination with limited selective advantage for host repression of retrotransposition (Lu and Clark 2010; Kelleher and Barbash 2013).

We also found that production of piRNAs from postnatal prepachytene clusters can be governed by converging transcription of 3′UTRs, which offers an elegant explanation for biogenesis of genic piRNAs. Moreover, we showed that a fraction of lincRNAs, here referred to as "piRNA-producing lincRNAs," are essentially pachytene piRNA primary transcripts, since their expression is restricted to the testis and piRNAs derived from them exhibit a 10A bias and a ping-pong signature. This observation prompts us to put forward a new framework for speculating on the origin of pachytene piRNAs and their contribution to TE silencing. Specifically, despite the fact that lincRNAs are generally depleted of TEs (Kapusta et al. 2013; Washietl et al. 2014), many of those lincRNAs that were recently acquired in the primate lineage possess a significantly higher fraction of TEs (Washietl et al. 2014). Moreover, expression of most of these hominid-specific lincRNAs is predominantly testis-specific (Washietl et al. 2014). Additionally, a sharp increase in the number of testis-specific lincRNAs is also seen in amniotes (Hezroni et al. 2015). Therefore, constant evolutionary turnover of sets of testis-specific lincRNAs might be, at least in part, responsible for the definition of pachytene piRNA repertoires, reminiscent of piRNA cluster evolution in rodents (Assis and Kondrashov 2009). It is also noteworthy that transcription of many of these evolutionary young lincRNAs is driven by LTR elements both in humans and rodents (Kapusta et al. 2013; Chen et al. 2016). Interestingly, there is a slight negative correlation between age rank of LTR subfamilies with their piRNA cluster fractions (Pearson's $r = -0.23/-0.19/-0.22$ for empirical/probabilistic/meta-analysis approaches, Supplemental Table S2N). In other words, the piRNA clusters are slightly enriched for young LTR subfamilies. Therefore, some testis-specific lincRNAs may provide a scaffold for pachytene piRNA production, serving as long single-stranded precursors and becoming "piRNA-producing lincRNAs"/pachytene piRNA primary transcripts.

Another peculiar finding is the seemingly contradicting epigenetic features of "piRNA-producing lincRNAs." First, they possess a chromatin environment similar to protein-coding genes (H3K4me3 enrichment and H3K27me3 depletion) and, second, they tend to be closer to lamina than testis-specific lincRNAs that do not produce piRNAs. However, both aspects are expected to support pachytene piRNA production: A gene-like chromatin environment is supposed to stimulate transcription of pachytene piRNA precursors, while proximity to the lamina can facilitate transport of these precursors outside the nucleus for further processing.

To conclude, the machinery involving the piRNA pathway in silencing evolutionarily recent TE subfamilies appears to operate both on prenatal and postnatal stages of spermatogenesis, in order to prevent the presence of RT activity at both stages of germ cell development. At the postnatal phase, pachytene cluster-driven piRNA production targets almost all TE subfamilies in late spermatogenesis. Despite this intricate mechanism designed for the purpose of tackling TE expansion, repeats have been consistently overcoming barriers to their transposition and spreading throughout genomes and populations (Haig 2016). This host–parasite coevolution is reminiscent of an arms race (Jacobs et al. 2014); it is also conceivable, however, that repeat expansion is successful partly due to the co-option of TEs within host regulatory networks (Chuong et al. 2016, 2017; Shapiro 2016). Perhaps there is simply insufficient evolutionary pressure to completely suppress TE expansion, or host organisms might have managed to strike the delicate balance between overspending host energy on absolute restriction of TEs from the genome and underspending such that TEs become detrimental to the host's existence.

## MATERIALS AND METHODS

### Tissue collection

Three samples of normal testis tissues adjacent to nonseminoma were obtained from three orchiectomy specimens. One sample of normal testis tissue was obtained from prostate cancer patients undergoing surgical castration. The samples were immediately frozen in liquid nitrogen. Histological evaluation was used to confirm the absence of tumor contamination. All patients provided written informed consent according to federal law, and the study was approved by the ethical committees of the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences and the Blokhin Russian Cancer Research Center, after reviewing patients' consent and information forms.

### RNA extraction, gene expression assays, and small RNA libraries generation

Total RNA extraction and purification was performed with TRIzol reagent (Thermo Fisher Scientific). Small RNA libraries were generated from 2 mkg of total RNA with the TruSeq Small RNA Library Preparation Kit (Illumina) according to manufacturer's instructions and sequenced on the HiSeq 2000 Sequencing System to generate 50-nt-long reads. Raw sequencing reads were deposited in SRA as bioproject PRJNA352412 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA352412): runs SRR4896698, SRR4896699, SRR4896700, SRR4896694.

## Small RNA data sets

Raw reads were quality filtered and trimmed with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Bowtie v1.0.0 (Langmead et al. 2009) was used to align trimmed reads to hg38 Human Genome Assembly with zero mismatches. Alignment procedure was based on the Tailor aligner (Chou et al. 2015). Briefly, unaligned reads were trimmed by 1 nt from the 3′ end and realigned again with zero mismatches allowed. This step was repeated two more times and only aligned reads (four alignments in total: after 0/1/2/3 nucleotides trimmed from 3′ end) were kept for further analyses.

The Integrative Genomics Viewer v2.3.74 (Robinson et al. 2011) was used to visualize alignment data. Statistical calculations were performed with R custom scripts or GraphPad Prism v5.00. Miscellaneous data processing was conducted with custom Bash/Awk/Perl scripts available on request.

Assessment of small RNA coverage of transposable elements was done with the approach described earlier in Pezic et al. (2014). Briefly, up to 10,000 genomic hits mapped with zero mismatches were reported for each read and $1/N$ score was assigned for each genomic hit found, where $N$ is the total number of all possible genomic hits for a read. Further, scores for all genomic hits were summed by a transposable element subfamily.

Ping-pong $Z$-score was calculated as described earlier (Zhang et al. 2011). Briefly, the $Z$-score is the difference of the number of piRNAs complementary to each other at a 5′-to-5′ distance of 10 nt and the mean number of piRNAs at background 5′-to-5′ distances (i.e., 1–9/11–20 nt), divided by the standard deviation of the numbers of piRNAs at background distances. $Z$-scores of Z19, Z20, Z21, and Z22 were calculated similarly with background distances at 1–18/20–22, 1–19/21–22, 1–20/22, 1–21, respectively.

Alignments to transposable element consensi were performed with two mismatches allowed and were used to interrogate TE-specific ping-pong $Z$-scores and distribution of reads along the consensi.

## RNA-seq data sets

Publicly available RNA-seq libraries used in the analyses are listed in Supplemental Table S2A.

For protein-coding genes, after filtering with FASTX-Toolkit, a conventional TopHat v2.0.9/Cufflinks v2.1.1 pipeline with default parameters and reference annotation from Ensembl release 83 was used (Trapnell et al. 2009, 2010, 2012; Kim et al. 2013).

For quantification of transposable elements' transcription level, we applied the same approach as for small RNA data sets with an additional preliminary step: All FASTX-Toolkit quality-filtered reads were trimmed to 50 nt to standardize the pipeline for RNA-seq data sets from different studies.

## ChIP-seq data sets

H3K4me3, H3K27me3 ChIP-seq data sets as well as corresponding Input controls for pachytene spermatocytes from three human individuals (Lesch et al. 2016) were filtered with the FASTX-Toolkit and aligned to hg38 Human Genome Assembly with Bowtie2 v2.1.0 with default parameters (Langmead and Salzberg 2012). H3K4me3 and H3K27me3 peaks were called with MACS v2.0.9 (Zhang et al. 2008) using default parameters for "narrow" and "broad" peaks, respectively. Only peaks detected for all three individuals were further considered for analysis.

## piRNA clusters definition

For empirical identification of piRNA cluster boundaries, we divided the genome into 100-bp bins and calculated piRNA coverage in them to further merge bins with density above 10RPKM and located within 1 kb of each other. Probabilistic piRNA cluster calling was done with TBr2_duster.pl/sRNAmapper.pl/ proTRAC_2.1.2.pl pipeline by Rosenkranz and Zischler (Rosenkranz 2016; Rosenkranz and Zischler 2012) with proTRAC parameters as follows: -pdens 0.05 -1Tor10A 0.1 -1Tand10A 0.1 -clstrand 0.5 -clsplit 0.05. Finally, a list of 1299 of human piRNA clusters from meta-analysis by Chirn et al. (2015) was also used. At the last filtering step, we kept only those empirical/probabilistic/meta-analysis clusters, whose piRNA coverage was higher than either 1RPKM or 10RPKM threshold, the fraction of 25- to 31-nt reads was at least 70%, and the length was more than 1 kb. These were further divided into prepachytene and pachytene clusters depending on whether the 26-nt or 30-nt fraction was bigger, respectively. For pachytene clusters, an additional filter of minimal length (3 kb) was used. Randomized controls were generated with "bedtools shuffle" (excluding protein-coding gene regions for pachytene clusters) (Quinlan and Hall 2010).

## GTEx data

Expression data quantified in RPKM were downloaded from http://gtexportal.org/ for all tissues available, and both means and standard deviations were separately calculated for "testis" and "non-testis" samples. A transcript was considered testis-specific if its "testis" mean expression was at least two "non-testis" standard deviations above the "non-testis" mean expression.

## lincRNAs

GENCODE v23 annotation (Harrow et al. 2012) filtered to remove small noncoding RNAs was used to create a list of 8455 testis-expressed and 3837 testis-specific lincRNA transcripts: All noncoding transcripts that do not overlap a protein-coding gene on either strand were considered. If lincRNA transcripts overlapped on the same strand they were merged and expression quantification for the most highly expressed one was used for the merged entity.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

We thank Professor Vladimir Gvozdev and members of his department at the Institute of Molecular Genetics, (Moscow, Russia) for useful comments on our work; Dr. Mikhail Klenov and Dr. Sergey Ryazansky at the Institute of Molecular Genetics (Moscow, Russia) for critical reading of the manuscript; Dr. Alexey Tryakin at the Blokhin Russian Cancer Research Center (Moscow, Russia) for assistance with surgical sample collection and histological

evaluation; Dr. Ilgar Mamedov at the Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry (Moscow, Russia) for help with high-throughput sequencing; and Dr. Olga Zatsepina at the Engelhardt Institute of Molecular Biology (Moscow, Russia) for guidance on small RNA libraries construction. This study was supported by the program of the Presidium of the Russian Academy of Sciences "Molecular and Cellular Biology" and the Russian Foundation for Basic Research grant no. 16-34-01193 mol_a to S.K.

*Received January 28, 2017; accepted August 23, 2017.*

## REFERENCES

Ahl V, Keller H, Schmidt S, Weichenrieder O. 2015. Retrotransposition and crystal structure of an Alu RNP in the ribosome-stalling conformation. *Mol Cell* **60:** 715–727.

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol* **11:** 1017–1027.

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T. 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* **5:** 337–350.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442:** 203–207.

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316:** 744–747.

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31:** 785–799.

Assis R, Kondrashov AS. 2009. Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. *Proc Natl Acad Sci* **106:** 7079–7082.

Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE. 2008. Active *Alu* retrotransposons in the human genome. *Genome Res* **18:** 1875–1883.

Boissinot S, Chevret P, Furano AV. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol Biol Evol* **17:** 915–928.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128:** 1089–1103.

Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12:** 503–514.

Chen J, Shishkin AA, Zhu X, Kadri S, Maza I, Guttman M, Hanna JH, Regev A, Garber M. 2016. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol* **17:** 19.

Chirn GW, Rahman R, Sytnikova YA, Matts JA, Zeng M, Gerlach D, Yu M, Berger B, Naramura M, Kile BT, et al. 2015. Conserved piRNA expression from a distinct set of piRNA cluster loci in eutherian mammals. *PLoS Genet* **11:** e1005652.

Chou MT, Han BW, Hsiao CP, Zamore PD, Weng Z, Hung JH. 2015. Tailor: a computational framework for detecting non-templated tailing of small silencing RNAs. *Nucleic Acids Res* **43:** e109.

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351:** 1083–1087.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18:** 71–86.

Czech B, Hannon GJ. 2016. One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. *Trends Biochem Sci* **41:** 324–337.

De Fazio S, Bartonicek N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, Antony C, Moreira PN, Enright AJ, O'Carroll D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480:** 259–263.

Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35:** 41–48.

Di Giacomo M, Comazzetto S, Saini H, De Fazio S, Carrieri C, Morgan M, Vasiliauskaite L, Benes V, Enright AJ, O'Carroll D. 2013. Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol Cell* **50:** 601–608.

Gebert D, Rosenkranz D. 2015. RNA-based regulation of transposon expression. *Wiley Interdiscip Rev RNA* **6:** 687–708.

Gebert D, Ketting RF, Zischler H, Rosenkranz D. 2015. piRNAs from pig testis provide evidence for a conserved role of the piwi pathway in post-transcriptional gene regulation in mammals. *PLoS One* **10:** e0124860.

Giordano J, Ge Y, Gelfand Y, Abrusan G, Benson G, Warburton PE. 2007. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. *PLoS Comput Biol* **3:** e137.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442:** 199–202.

Goh WS, Falciatori I, Tam OH, Burgess R, Meikar O, Kotaja N, Hammell M, Hannon GJ. 2015. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev* **29:** 1032–1044.

Gou LT, Dai P, Yang JH, Xue Y, Hu YP, Zhou Y, Kang JY, Wang X, Li H, Hua MM, et al. 2014. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res* **24:** 680–700.

Grivna ST, Beyret E, Wang Z, Lin H. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20:** 1709–1714.

GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45:** 580–585.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453:** 948–951.

Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, Yong J, Hu Y, Wang X, Wei Y, et al. 2015. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161:** 1437–1452.

Ha H, Song J, Wang S, Kapusta A, Feschotte C, Chen KC, Xing J. 2014. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics* **15:** 545.

Haig D. 2016. Transposable elements: self-seekers of the germline, team-players of the soma. *Bioessays* **38:** 1158–1166.

Han BW, Zamore PD. 2014. piRNAs. *Curr Biol* **24:** R730–R733.

Han BW, Wang W, Li C, Weng Z, Zamore PD. 2015. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* **348:** 817–821.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22:** 1760–1774.

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* **11:** 1110–1122.

Hirakata S, Siomi MC. 2016. piRNA biogenesis in the germline: from transcription of piRNA genomic sources to piRNA maturation. *Biochim Biophys Acta* **1859:** 82–92.

Hirano T, Iwasaki YW, Lin ZY, Imamura M, Seki NM, Sasaki E, Saito K, Okano H, Siomi MC, Siomi H. 2014. Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate. *RNA* **20:** 1223–1237.

Homolka D, Pandey RR, Goriaux C, Brasset E, Vaury C, Sachidanandam R, Fauvarque MO, Pillai RS. 2015. PIWI slicing

and RNA elements in precursors instruct directional primary piRNA biogenesis. *Cell Rep* **12:** 418–428.

Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes *ZNF91/93* and SVA/L1 retrotransposons. *Nature* **516:** 242–245.

Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* **9:** e1003470.

Kelleher ES, Barbash DA. 2013. Analysis of piRNA-mediated silencing of active TEs in *Drosophila melanogaster* suggests limits on the evolution of host genome defense. *Mol Biol Evol* **30:** 1816–1829.

Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res* **16:** 78–87.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36.

Kojima-Kita K, Kuramochi-Miyagawa S, Nagamori I, Ogonuki N, Ogura A, Hasuwa H, Akazawa T, Inoue N, Nakano T. 2016. MIWI2 as an effector of DNA methylation and gene silencing in embryonic male germ cells. *Cell Rep* **16:** 2819–2828.

Kumar P, Kuscu C, Dutta A. 2016. Biogenesis and function of transfer RNA-related fragments (tRFs). *Trends Biochem Sci* **41:** 679–689.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22:** 908–917.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. *Science* **313:** 363–367.

Lesch BJ, Silber SJ, McCarrey JR, Page DC. 2016. Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nature Genet* **48:** 888–894.

Le Thomas A, Tóth KF, Aravin AA. 2014. To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biol* **15:** 204.

Li XZ, Roy CK, Dong X, Bolcun-Filas E, Wang J, Han BW, Xu J, Moore MJ, Schimenti JC, Weng Z, et al. 2013a. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol Cell* **50:** 67–81.

Li XZ, Roy CK, Moore MJ, Zamore PD. 2013b. Defining piRNA primary transcripts. *Cell Cycle* **12:** 1657–1658.

Lu J, Clark AG. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res* **20:** 212–227.

Manakov SA, Pezic D, Marinov GK, Pastor WA, Sachidanandam R, Aravin AA. 2015. MIWI2 and MILI have differential effects on piRNA biogenesis and DNA methylation. *Cell Rep* **12:** 1234–1243.

Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, et al. 2015. Human genomics. The human transcriptome across tissues and individuals. *Science* **348:** 660–665.

Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, Reinders M, Wessels L, van Steensel B. 2013. Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23:** 270–280.

Mohn F, Handler D, Brennecke J. 2015. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* **348:** 812–817.

Mourier T. 2011. Retrotransposon-centered analysis of piRNA targeting shows a shift from active to passive retrotransposon transcription in developing mouse testes. *BMC Genomics* **12:** 440.

Nagamori I, Kobayashi H, Shiromoto Y, Nishimura T, Kuramochi-Miyagawa S, Kono T, Nakano T. 2015. Comprehensive DNA methylation analysis of retrotransposons in male germ cells. *Cell Rep* **12:** 1541–1547.

Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* **73:** 1444–1451.

Pezic D, Manakov SA, Sachidanandam R, Aravin AA. 2014. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev* **28:** 1410–1428.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842.

Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, Antony C, Sachidanandam R, Pillai RS. 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480:** 264–267.

Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC. 2009. A broadly conserved pathway generates 3′UTR-directed primary piRNAs. *Curr Biol* **19:** 2066–2076.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29:** 24–26.

Roovers EF, Rosenkranz D, Mahdipour M, Han CT, He N, Chuva de Sousa Lopes SM, van der Westerlaken LA, Zischler H, Butter F, Roelen BA, et al. 2015. Piwi proteins and piRNAs in mammalian oocytes and early embryos. *Cell Rep* **10:** 2069–2082.

Rosenkranz D. 2016. piRNA cluster database: a web resource for piRNA producing loci. *Nucleic Acids Res* **44:** D223–D230.

Rosenkranz D, Zischler H. 2012. proTRAC—a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics* **13:** 5.

Rosenkranz D, Han CT, Roovers EF, Zischler H, Ketting RF. 2015a. Piwi proteins and piRNAs in mammalian oocytes and early embryos: from sample to sequence. *Genom Data* **5:** 309–313.

Rosenkranz D, Rudloff S, Bastuck K, Ketting RF, Zischler H. 2015b. Tupaia small RNAs provide insights into function and evolution of RNAi-based transposon defense in mammals. *RNA* **21:** 911–922.

Rounge TB, Furu K, Skotheim RI, Haugen TB, Grotmol T, Enerly E. 2015. Profiling of the small RNA populations in human testicular germ cell tumors shows global loss of piRNAs. *Mol Cancer* **14:** 153.

Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC. 2009. A regulatory circuit for *piwi* by the large Maf gene *traffic jam* in *Drosophila*. *Nature* **461:** 1296–1299.

Senti KA, Jurczak D, Sachidanandam R, Brennecke J. 2015. piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev* **29:** 1747–1762.

Shapiro JA. 2016. Exploring the read-write genome: mobile DNA and mammalian adaptation. *Crit Rev Biochem Mol Biol* **52:** 1–17.

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F, et al. 2016. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351:** 391–396.

Tang WW, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. 2015. A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161:** 1453–1467.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511–515.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protoc* **7:** 562–578.

Vagin VV, Klenov MS, Kalmykova AI, Stolyarenko AD, Kotelnikov RN, Gvozdev VA. 2004. The RNA interference proteins and vasa locus are

involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol* **1:** 54–58.

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313:** 320–324.

Vandewege MW, Platt RN II, Ray DA, Hoffmann FG. 2016. Transposable element targeting by piRNAs in Laurasiatherians with distinct transposable element histories. *Genome Biol Evol* **8:** 1327–1337.

Vourekas A, Alexiou P, Vrettos N, Maragkakis M, Mourelatos Z. 2016. Sequence-dependent but not sequence-specific piRNA adhesion traps mRNAs to the germ plasm. *Nature* **531:** 390–394.

Wang W, Han BW, Tipping C, Ge DT, Zhang Z, Weng Z, Zamore PD. 2015. Slicing and binding by Ago3 or Aub trigger Piwi-bound piRNA production by distinct mechanisms. *Mol Cell* **59:** 819–830.

Washietl S, Kellis M, Garber M. 2014. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res* **24:** 616–628.

Williams Z, Morozov P, Mihailovic A, Lin C, Puvvula PK, Juranek S, Rosenwaks Z, Tuschl T. 2015. Discovery and characterization of piRNAs in the human fetal ovary. *Cell Rep* **13:** 854–863.

Yang Q, Hua J, Wang L, Xu B, Zhang H, Ye N, Zhang Z, Yu D, Cooke HJ, Zhang Y, et al. 2013. MicroRNA and piRNA profiles in normal human testis detected by next generation sequencing. *PLoS One* **8:** e66809.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137.

Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, Weng Z, Theurkauf WE, Zamore PD. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol Cell* **44:** 572–584.

Zhang P, Kang JY, Gou LT, Wang J, Xue Y, Skogerboe G, Dai P, Huang DW, Chen R, Fu XD, et al. 2015. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res* **25:** 193–207.