

piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals

NICHOLAS F. PARRISH,^{1,8,9} KAN FUJINO,^{1,8,11} YUSUKE SHIROMOTO,^{2,10} YUKA W. IWASAKI,³ HONGSEOK HA,⁴ JINCHUAN XING,⁴ AKIKO MAKINO,^{1,5} SATOMI KURAMOCHI-MIYAGAWA,² TORU NAKANO,² HARUHIKO SIOMI,³ TOMOYUKI HONDA,^{1,6} and KEIZO TOMONAGA^{1,6,7}

¹Department of Viral Oncology, Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan

²Department of Pathology, Medical School and Graduate School of Frontier Biosciences, Osaka University, Osaka 565-0871, Japan

³Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan

⁴Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA

⁵Center for Emerging Virus Research, Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan

⁶Department of Tumor Viruses, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan

⁷Department of Mammalian Regulatory Network, Graduate School of Biostudies, Kyoto University, Kyoto 606-8507, Japan

ABSTRACT

Endogenous bornavirus-like nucleoprotein elements (EBLNs) are sequences within vertebrate genomes derived from reverse transcription and integration of ancient bornaviral nucleoprotein mRNA via the host retrotransposon machinery. While species with EBLNs appear relatively resistant to bornaviral disease, the nature of this association is unclear. We hypothesized that EBLNs could give rise to antiviral interfering RNA in the form of PIWI-interacting RNAs (piRNAs), a class of small RNA known to silence transposons but not exogenous viruses. We found that in both rodents and primates, which acquired their EBLNs independently some 25–40 million years ago, EBLNs are present within piRNA-generating regions of the genome far more often than expected by chance alone ($P = 8 \times 10^{-3} - 6 \times 10^{-8}$). Three of the seven human EBLNs fall within annotated piRNA clusters and two marmoset EBLNs give rise to bona fide piRNAs. In both rats and mice, at least two of the five EBLNs give rise to abundant piRNAs in the male gonad. While no EBLNs are syntenic between rodent and primate, some of the piRNA clusters containing EBLNs are; thus we deduce that EBLNs were integrated into existing piRNA clusters. All true piRNAs derived from EBLNs are antisense relative to the proposed ancient bornaviral nucleoprotein mRNA. These observations are consistent with a role for EBLN-derived piRNA-like RNAs in interfering with ancient bornaviral infection. They raise the hypothesis that retrotransposon-dependent virus-to-host gene flow could engender RNA-mediated, sequence-specific antiviral immune memory in metazoans analogous to the CRISPR/Cas system in prokaryotes.

Keywords: endogenous viral elements; retrotransposon; piRNA; CRISPR/Cas; paleovirology

INTRODUCTION

Retroviruses integrate DNA copies of their RNA genome into that of their host, transferring genetic information in a direction not predicted by the central dogma (Crick 1958). Understanding how these gene transfers influence the immunologic distinction between host “self” and viral

“nonself” is of practical importance. Interestingly, several host genes derived from retroviruses limit infection by related viruses (Ikeda and Odaka 1983; Hainaut et al. 1990; Mura et al. 2004; Aswad and Katzourakis 2012; Fujino et al. 2014; Yap et al. 2014), demonstrating that lateral gene flow from viruses to their vertebrate hosts can mediate transgenerational immune memory; all known examples are mediated by production of dominant negative proteins. RNA-mediated transgenerational antiviral immunity, while common via the CRISPR/Cas system in prokaryotes

⁸These authors contributed equally to this work.

⁹Present address: Section of Surgical Sciences, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

¹⁰Present address: Laboratory of Gene Expression and Regulation, The Wistar Institute, Philadelphia, PA 19104, USA

¹¹Present address: Department of Microbiology II, School of Veterinary Medicine, Azabu University, Kanagawa 252-5201, Japan

Corresponding authors: nicholas.parrish@vanderbilt.edu, tomonaga@virus.kyoto-u.ac.jp

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.052092.115>.

© 2015 Parrish et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

(Sorek et al. 2008), has thus far been reported in only one metazoan, *Caenorhabditis elegans* (Rechavi et al. 2011). Immunity through the CRISPR/Cas system requires lateral transfer of genetic information: short nonself nucleic acid “spacers” are integrated into specialized genomic arrays of repeated sequences. Transcripts from these specialized self loci are processed into ribonucleoprotein complexes capable of degrading nonself targets on the basis of Watson–Crick base-pairing with spacer RNA (Barrangou et al. 2007). A conceptually similar system mediated by PIWI-interacting RNAs (piRNAs) appears broadly conserved in metazoans (Grimson et al. 2008), yet the confirmed targets are largely limited to endogenous transposable elements.

In addition to retroviruses, RNA-only viruses have also given rise to sequences in many metazoan genomes, although the specific mechanisms involved in this horizontal RNA-to-DNA information flow are less clear (Belyi et al. 2010; Horie et al. 2010; Katzourakis and Gifford 2010). We have studied one class of these sequences, called endogenous bornavirus-like nucleoprotein elements (EBLNs), because they are the only riboviral endogenous elements known in humans. *Homo sapiens* EBLNs (hsEBLNs) contain poly(dA:dT) tracts, recognizable transcription start sites, and are flanked by target-site duplications, strongly suggesting that they represent viral mRNA integrated by a retrotransposon (Esnault et al. 2000; Belyi et al. 2010). These EBLNs were integrated ~40 million years ago, coincident with the peak of host processed pseudogene formation by a similar mechanism (Zhang et al. 2003). Because species with EBLNs appear relatively protected against modern day bornaviruses (Belyi et al. 2010), negative-strand RNA viruses that can cause neurological disease (Tomonaga et al. 2002), we questioned if they could influence antiviral immunity like some endogenous retroviral elements.

HsEBLN-1 and *-2* contain long open reading frames (ORFs) with the potential to code for proteins of 366 and 225 amino acids, respectively. In cell culture experiments, overexpression of bornaviral nucleoprotein prevents infection, presumably because a specific stoichiometry of replication complex components is critical (Geib et al. 2003). Thus overexpression of EBLN-encoded proteins could potentially have prevented ancient bornaviral replication. Indeed, we recently showed that the *Ictidomys tridecemlineatus* genome contains an EBLN that was integrated more recently than human EBLNs, shares over 75% of amino acids in common with some extant avian bornaviruses, and can block bornaviral replication when overexpressed in human cells (Fujino et al. 2014). However, most EBLNs in primates and rodents have disrupted ORFs (Horie et al. 2010), and there is no evidence of selection to maintain the ORFs of EBLNs in primates (Kobayashi et al. 2011). Thus most EBLNs either had no function, had a protein-coding function that has been lost, or perhaps had a function not related to encoding a protein.

We recently observed that all seven human EBLNs are expressed as RNAs, some exclusively in the adult testis (K Sofuku, N Parrish, T Honda, and K Tomonaga, in prep.). Because the native promoter sequence is not mobilized during pseudogene formation, the probability of seven out of seven randomly chosen host processed pseudogenes being expressed is low; by the highest estimates, only about one-third of host processed pseudogenes are transcribed (Zheng et al. 2007; Guo et al. 2014). This led us to investigate the possibility that mammalian EBLNs encode antiviral RNAs. The previously noted similarities between the piRNA pathway and the CRISPR/Cas immune system (Karginov and Hannon 2010) made piRNAs an attractive candidate. Analogous to CRISPR guide RNA, piRNA, in complex with a PIWI-clade argonaute protein partner, target transposons for transcriptional and post-transcriptional silencing (Siomi and Kuramochi-Miyagawa 2009; Ishizu et al. 2012). Similar to CRISPR arrays, piRNA precursor molecules are transcribed from discrete loci (“piRNA clusters”) that cover a small percentage of the total genome (Lau et al. 2006; Aravin et al. 2007). It is thought that these loci act as “traps,” in the sense that nucleic acid elements transposing into them will subsequently be silenced (Malone and Hannon 2009). PiRNAs are most abundant in the germline, consistent with a role in genome defense, yet some piRNA pathway components are detectable in somatic cells where their functional relevance is unclear. Here we show that multiple EBLNs, integrated independently in two mammalian lineages at least 20 million years ago, give rise to piRNAs, and present evidence consistent with selection for EBLNs that integrated into piRNA-generating loci.

RESULTS

Rodent EBLNs give rise to small RNA with characteristics of piRNAs

We noted EBLN-derived piRNAs while examining rodent EBLN sequences using a web-based genome browser (Karolchik et al. 2014). Several piRNAs identified in reports initially describing piRNAs (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006) overlap with *Mus musculus* and *Rattus norvegicus* EBLNs (Table 1). Most piRNAs are generated from genomic loci that are unannotated (Girard et al. 2006), as are most EBLNs. Thus to better ascertain the abundance and diversity of EBLN-derived small RNAs, we sequenced small RNAs from the testis of a 6-wk-old mouse. Small RNAs with sequence characteristics of primary piRNAs, namely a length over 26 nucleotides and enrichment of 5' uridine, mapped to three of the five mouse EBLNs (*mmEBLN-3* through *-5*) (Fig. 1A). The abundance of small RNAs mapped to *mmEBLN-5* was highest, with thousands of reads mapped to some regions of this sequence. The abundance of small RNAs mapped to *mmEBLN-3* and *-4* was lower, yet above

TABLE 1. GenBank annotated piRNAs derived from rodent EBLNs

	Accession	Name	Coordinates	Length	5' Nucleotide	
<i>Mus musculus</i>						
<i>EBLN-5</i>	DQ549815	mmu_piR_003423	chr9:54093758–54093787	30	U	
	DQ548337	mmu_piR_003006	chr9:54093761–54093790	30	U	
	DQ689528	mmu_piR_012394	chr9:54093806–54093834	29	U	
	DQ700150	mmu_piR_019537	chr9:54093814–54093842	29	U	
	DQ703437	mmu_piR_021794	chr9:54093845–54093874	30	U	
<i>EBLN-3</i>	DQ724667	mmu_piR_036414	chr9:54093868–54093898	31	U	
	DQ706487	mmu_piR_023855	chr18:67195283–67195313	31	U	
	DQ711296	mmu_piR_027183	chr18:67195514–67195543	30	U	
	DQ725925	mmu_piR_037293	chr18:67195670–67195699	30	U	
	DQ698401	mmu_piR_018347	chr18:67195905–67195932	28	A	
	DQ551691	mmu_piR_004100	chr18:67195934–67195965	32	U	
	NA	mmu_piR_038731	chr18:67196113–67196142	30	U	
<i>Rattus norvegicus</i>						
<i>EBLN-5</i>	DQ622271	rno_piR_009248	chr8:57747085–57747114	30	U	
	DQ621647	rno_piR_008949	chr8:57747093–57747122	30	U	
	DQ609231	rno_piR_003427	chr8:57747095–57747125	31	U	
	DQ617717	rno_piR_007299	chr8:57747130–57747158	29	U	
	DQ767477	rno_piR_038475	chr8:57747226–57747254	29	U	
	DQ750254	rno_piR_026877	chr8:57747268–57747296	29	U	
	DQ732898	rno_piR_015474	chr8:57747276–57747304	29	U	
	DQ605336	rno_piR_001654	chr8:57747283–57747311	29	U	
	DQ754944	rno_piR_029961	chr8:57747287–57747316	30	U	
	DQ757774	rno_piR_031873	chr8:57747336–57747365	30	U	
	DQ751527	rno_piR_027737	chr8:57747349–57747379	31	U	
	<i>EBLN-3</i>	DQ739763	rno_piR_019953	chr18:63553931–63553956	26	U

(U) Uracil, (A) adenosine.

the level of most genomic loci. In contrast, no small RNAs mapped to *mmEBLN-1* and *-2*. Examining the small RNAs mapped to the 100 kb flanking these EBLNs revealed that *mmEBLN-3* and *-5* were within loci with small RNAs abundantly mapped to a single genomic strand, while *mmEBLN-4* was near the end of such a locus (Fig. 1B). Notably, all piRNA-like small RNAs mapped to the antisense strand relative to the hypothetical ancient bornaviral N mRNA whose integration was responsible for EBLN formation. This experiment suggested that small RNAs with characteristics of piRNAs were generated from three of five *mmEBLN* loci, and that these EBLNs were in or near piRNA clusters. Thus we examined publically available sequence data sets and piRNA cluster annotations to further evaluate this possibility.

Rodent EBLNs are enriched in pachytene piRNA clusters

A piRNA is defined as a small RNA in complex with a PIWI protein, thus immunoprecipitation of intact ribonucleoproteins is required to unambiguously confirm the identity of any RNA molecule as a piRNA. Thus we first analyzed sequences from a comprehensive analysis of murine piRNA biogenesis (Li et al. 2013a). Consistent with our initial experiment, small RNAs mapped to three of five *mmEBLN*s

(*mmEBLN-3*, *-4*, and *-5*) (Table 2). These RNAs ranged in length from 25 to 31 nucleotides (nt) and >95% contained uridine as the 5' nucleotide (Fig. 1C). All were antisense relative to the proposed ancient bornaviral mRNA, consistent with the potential to post-transcriptionally silence such mRNA (Reuter et al. 2011). As expected of piRNAs derived from a primary piRNA transcript, and as opposed to secondary piRNAs generated during the so-called “ping-pong” amplification cycle that occurs prenatally in mice (Beyret et al. 2012), there was no enrichment of adenosine at the 10th nucleotide position (Wang et al. 2014). Li et al. (2013b) defined 417 piRNA precursor transcripts from which over 95% of mature murine piRNA are derived. Together these sequences cover only 0.28% of the mouse genome. Pachytene piRNA transcripts, which give rise to the predominant class of piRNAs found in adult testis, cover only 0.08% of the genome. Thus a short genomic range chosen at random has an exceedingly low likelihood of being within such a transcript. However, two of the five *mmEBLN*s (*mmEBLN-3* and *-5*) are within pachytene piRNA transcripts. Moreover, the piRNA precursors containing these two EBLNs give rise to the third and 13th highest density of mapped piRNAs from the adult mouse testis (Lau et al. 2006). *mmEBLN-4* is within 40 kb of the 3' termini of an annotated pachytene piRNA transcript. To estimate the probability of this apparent enrichment of EBLNs within piRNA-generating genomic loci, we calculated

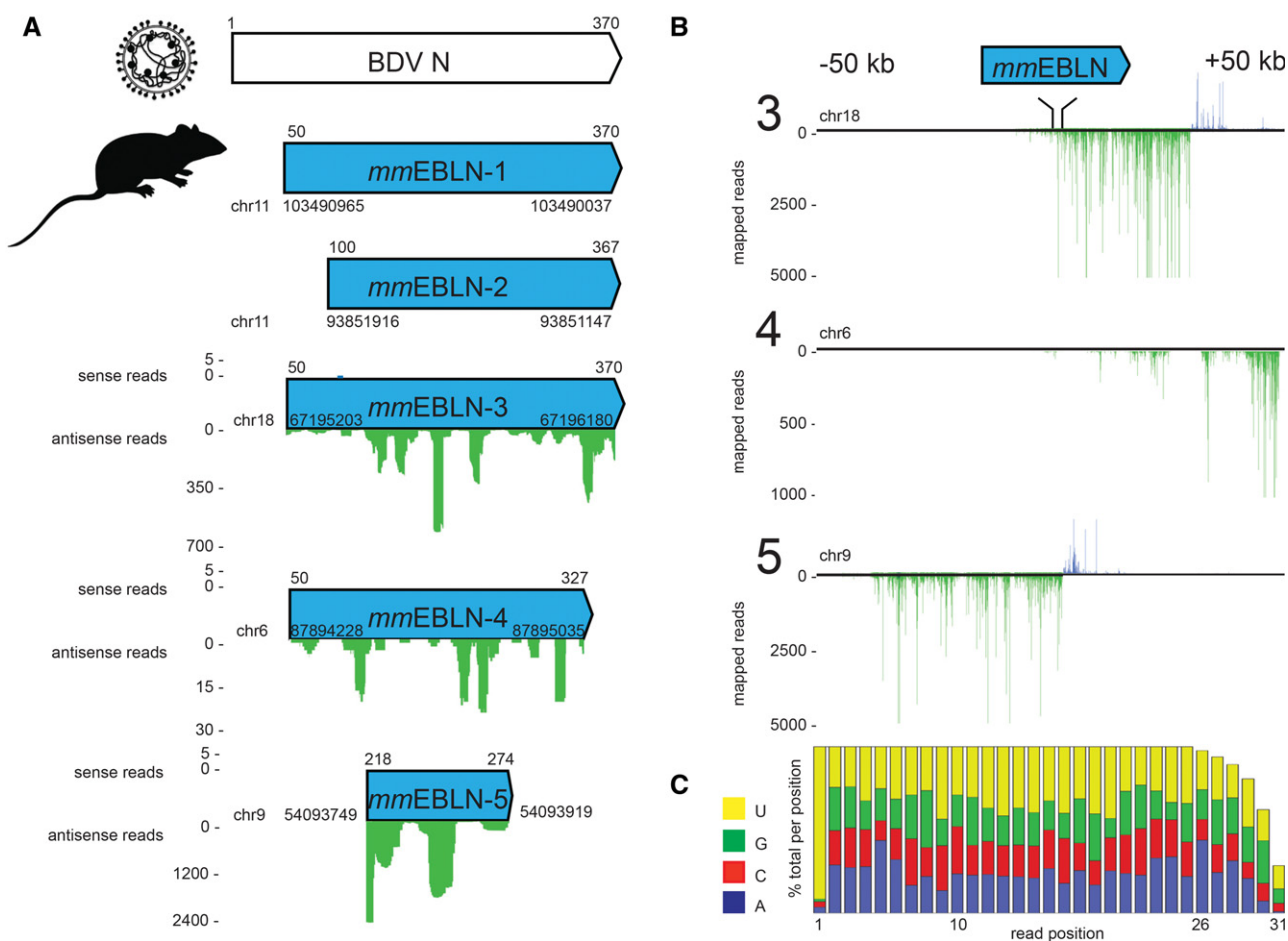


FIGURE 1. Murine EBLNs give rise to piRNA. (A) Testis small RNAs map to murine EBLNs. Borna disease virus (BDV), nucleoprotein (N) gene, and homologous murine EBLNs (1–5 as numbered) are depicted. Numbers *atop* each marker represent the EBLN length as defined by translated amino acid homolog to BDV determined in Arensburger et al. (2011). Numbers *below* each marker indicate the genomic coordinates of the murine EBLNs. Uniquely mapped small RNAs are shown for *EBLN-3* through *-5*; reads antisense to the predicted ancient bornaviral mRNA are plotted in green *below* each gene marker, sense reads in blue *above*. The *y*-axis indicates the number of mapped reads. (B) *mmEBLN-3* through *-5* are in or near clusters of mapped small RNAs. Small RNA reads matching the 50 kb upstream and downstream from *mmEBLN-3* through *-5* are shown. Reads mapping to the 5' to 3' strand are shown in blue *above* the line and those mapped to the opposite strand are shown in green *below* the line. The *y*-axis indicates the number of uniquely mapped reads and is arbitrarily truncated at 1000 or 5000 reads. (C) *mmEBLN*-mapped reads have characteristics of piRNAs. The nucleotide composition of small RNAs mapped to murine *EBLN-3* through *-5* is shown, with each base colored as indicated. Bases 1, 10, 26, and 31 are numbered. The *y*-axis indicates the percentage nucleotides at each position of all mapped reads corresponding to each base.

the cumulative binomial distribution of an event with a probability 0.0028 occurring twice or more in, as there are five recognized EBLNs, five trials ($\mathbb{P} = 8 \times 10^{-5}$).

Considering the possibility that this enrichment was perhaps limited to a single species, we also examined the genome of the laboratory rat (*Rattus norvegicus*). All EBLNs are in syntenic loci in mice and rats (Horie et al. 2013), suggesting that these EBLNs were integrated between 20 and 30 million years ago into a common ancestor shared by these species (Horie et al. 2013). PiRNA clusters are generally well conserved between mice and rats in terms of their genomic context (Assis and Kondrashov 2009), but not at the level of piRNA sequence (Aravin et al. 2006). As primary piRNA precursor transcripts have not been defined in the rat, we exam-

ined piRNA cluster annotations to determine if EBLN-derived piRNAs would be expected to be made in this species. Indeed, both orthologous EBLNs within piRNA clusters in mice were annotated in rat piRNA clusters (Girard et al. 2006; Lau et al. 2006), and *rnEBLN-4* was also annotated within a cluster in one study (Girard et al. 2006). As in the mouse, a low percentage of the overall rat genome gives rise to piRNAs and is annotated as a piRNA clusters (0.18% in the more inclusive annotation). The probability of this degree of enrichment of *rnEBLN*s in piRNA-generating loci occurring by chance, calculated as above, ranges from 6×10^{-8} to 1×10^{-6} based on the annotation used (Fig. 2A). In summary, three out of the five EBLNs in mice and rats give rise to piRNAs. As we are limited to detecting EBLNs that

TABLE 2. piRNA abundance from rodent and primate EBLNs

EBLN	Coordinates (strand)	piRNA precursor transcript ^a	piRNA cluster evolution ^b	RPKM ^c	
				Complementary to N mRNA ^d	Complementary to viral genomic RNA ^e
<i>mmeEBLN-1</i>	chr11:103490037–103490965 (–)	–		0	0
<i>mmeEBLN-2</i>	chr11:93851147–93851916 (–)	–		0	0
<i>mmeEBLN-3</i>	chr18:67195203–67196180 (+)	18-qE1-36451.1; chr18:67189100–67226114	Ancestral: human ID 5251; chr18:11631026–11717899 (marmoset) chr13:45297101–45343400	144.11	0
<i>mmEBLN-4</i>	chr6:87894228–87895035 (+)	–		5.84	0
<i>mmEBLN-5</i>	chr9:54093749–54093919 (+)	9-qA5.3-24188.1; chr9:54054980–54097630	Ancestral: human ID 4384; chr15:51535076–51605926 (marmoset) chr10:28388101–28440800	603.48	0
	Coordinates (strand)	piRNA cluster		RPKM	Complementary to viral genomic RNA
<i>rnEBLN-1</i>	chr10:92784712–92785688 (–)	–		0	0
<i>rnEBLN-2</i>	chr10:82624298–82625334 (+)	–		0	0
<i>rnEBLN-3</i>	chr18:63553420–63554192 (+)	cluster 14 ⁱ ; chr18:63622000–63658000	Ancestral: human ID 5251; chr13:45297101–45343400	ND	0
<i>rnEBLN-4</i>	chr4:122161250–122161697 (+)	cluster 85 ^g ; chr4:122158319–122199619	Ancestral: human ID 1264; chr3:128540550–128603591 (marmoset chr15:68183501–68217900)	ND	0
<i>rnEBLN-5</i>	chr8:57747079–57747381 (+)	cluster 6 ^f ; chr8:57725000–57777000	Ancestral: human ID 4384; chr15:51535076–51605926 (marmoset) chr10:28388101–28440800	ND	0
	Coordinates (strand)	piRNA cluster ^h		RPKM ^h	Complementary to viral genomic RNA
<i>hsEBLN-1</i>	chr10:22208838–22209999 (–)	–		0.12	0
<i>hsEBLN-2</i>	chr3:73062206–73063317 (+)	ID 1201; chr3:73061025–73070092		0	0.68
<i>hsEBLN-3</i>	chr9:37086679–37087837 (+)	ID 2841; chr9:37086668–37092983		0	1.71
<i>hsEBLN-4</i>	chr17:4570963–4572180 (+)	–		0.07	0
<i>hsEBLN-5</i>	chr1:235117490–235118542 (–)	–		0	0
<i>hsEBLN-6</i>	chr11:75011855–75012971 (–)	ID 3562; chr11:75001149–75023707		3.35	0
<i>hsEBLN-7</i>	chr10:94534165–94535277 (–)	–		5.28	0
	Coordinates (strand)	piRNA cluster ⁱ		RPKM ⁱ	Complementary to viral genomic RNA
<i>cjEBLN-1</i>	chr7:19830726–19831493 (–)	–		Complementary to N mRNA	0

Continued

TABLE 2. Continued

EBLN	Coordinates (strand)	piRNA precursor transcript ^a	piRNA cluster evolution ^b	RPKM ^c	
				Complementary to N mRNA ^d	Complementary to viral genomic RNA ^e
cjEBLN-2	chr15:49481896-49483299 (+)	-		0	0
cjEBLN-2	chr1:138998195-138998975 (+)	-		0	0
cjEBLN-4	GL25135.1:14259-15765 (+)	-		0	0
cjEBLN-5	chr19:45134664-45135692 (-)	-		0	0
cjEBLN-6	chr11:63682272-63683372 (-)	chr11:63677801-63689900		15.81	0
cjEBLN-7	chr12:81451040-81452148 (-)	-		1.23	0

(mmEBLN) *Mus musculus* EBLN; annotated based on mm9 (strand equivalent to hypothetical ancient mRNA). (mEBLN) *Rattus norvegicus* EBLN; annotated based on RGSC3.4. (hsEBLN) *Homo sapiens* EBLN; annotated based on GRCh38/hg38. (cjEBLN) *Callithrix jacchus* EBLN; annotated based on cj3.2. (ND) Not determined.

^aFrom Li et al. (2013a), Supplemental Table 2 (0.28% of the genome annotated as a piRNA cluster).

^bFrom Assis and Kondrashov (2009).

^cMapped from Li et al. (2013b); GEO accession number GSM1096587.

^dAccording to hypothetical ancient bornaviral N mRNA.

^eAs a negative-strand RNA virus, the bornavirus genomic strand is the reverse complement of the mRNA strand.

^fFrom Lau et al. (2006) (0.106% annotated as a piRNA cluster).

^gFrom Girard et al. (2006) (0.18% annotated as a piRNA cluster).

^hFrom Ha et al. (2014) (2.4% annotated as a piRNA cluster at RPKM cutoff of >1) liftover converted to GRCh38/hg38.

ⁱFrom Hirano et al. (2014) (0.12% annotated as a piRNA clus).

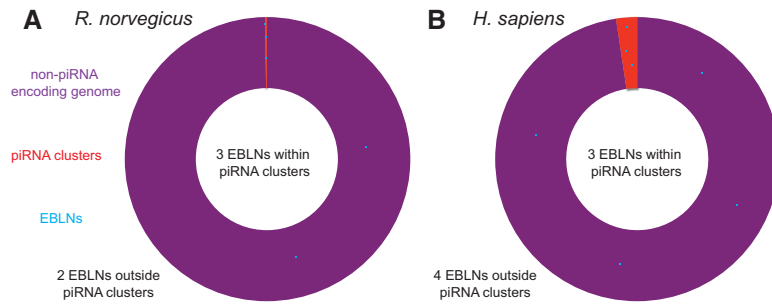


FIGURE 2. EBLNs are enriched within piRNA clusters. (A) The genome of *Rattus norvegicus* is depicted. Non-piRNA-generating sequences in the genome (purple) and piRNA clusters (red, 0.18%, as annotated in Girard et al. (2006) are drawn to scale. EBLNs are depicted as blue dots and are drawn larger than scale to allow visualization. The probability of the observed enrichment, estimated as described in the text, is 6×10^{-8} . (B) The genome of *Homo sapiens* is depicted as above, with 2.4% of the genome annotated as piRNA cluster. The probability of the observed enrichment is 4.5×10^{-4} .

have maintained recognizable homology with modern bornaviruses, we do not assume that the existing five rodent EBLNs are the only sequences derived from ancient bornaviruses that ever entered the genomes of these species. Nonetheless, given the relative sparsity of piRNA-generating genomic loci, we can reasonably exclude the hypothesis that rodent EBLNs were integrated at random and maintained neutrally thereafter.

piRNAs are generated from primate EBLNs

EBLNs are found in many vertebrate genomes (Belyi et al. 2010), and similar to retrotransposon insertions, are predicted to be essentially free of homoplasy. About 40 million years ago, EBLNs were integrated into a common primate ancestor prior to the divergence of the strepsirrhine and haplorrhine primates (Horie et al. 2010). Accordingly, these primate EBLNs were each integrated into different loci than rodent EBLNs. We hypothesized that if EBLN-derived piRNAs were functional, independent EBLN integrations with piRNA-generating capacity may have evolved convergently in these two lineages. Thus we determined if piRNA-like RNAs or piRNAs were made from the seven EBLNs in human and marmoset genomes, respectively, as deep sequencing of small RNAs from adult testes from both species has recently been performed (Ha et al. 2014; Hirano et al. 2014). Similar to pachytene piRNAs derived from adult mouse testes, the piRNA populations in the testis of adult marmosets are processed from primary piRNA precursor transcripts rather than ping-pong amplification. In humans, three EBLNs (*hsEBLN-2*, -3, and -6) were found to be within annotated piRNA clusters (Fig. 2B; total 2.4% annotated, $\mathbb{P} = 4.5 \times 10^{-4}$). *hsEBLN-7* was also found to give rise to piRNA-like RNA, yet was shorter than the arbitrary length cutoff used to define clusters (Table 2). In the marmoset, one EBLN (*cjEBLN-6*) was within an annotated cluster (total 0.12% annotated, $\mathbb{P} = 0.008$) and *cjEBLN-7* also gave rise to piRNAs associated with the marmoset PIWI-like 1 protein MARWI

(Hirano et al. 2014). As in rodents, the piRNAs produced by these EBLNs are antisense to the proposed ancient bornaviral mRNA. Experimental differences could explain why *hsEBLN-2* and -3 give rise to piRNA-like small RNAs in the testis while the syntenic marmoset EBLNs in the marmoset do not: the marmoset sequences represent bona fide piRNAs immunoprecipitated with MARWI, while the human testis small RNA was bulk-isolated and enriched for 2'-O-methylated small RNAs (Kirino and Mourelatos 2007). Alternatively, this could reflect loss of piRNA production from *cjEBLN-2* and -3 loci in the ~35 million years since the marmoset and human lineages diverged.

In any case, these observations confirm that independent EBLN integrations are enriched within piRNA- or piRNA-like RNA-generating loci in two mammalian lineages. The probability of this occurring due to chance, approximated as the union of the most likely probabilities when each lineage is considered independently, is very low ($\mathbb{P} = 6.4 \times 10^{-7}$).

Rodent EBLNs were integrated into existing piRNA clusters

Precisely what defines a piRNA precursor transcript as such is currently unclear (Vourekas et al. 2015), as are the determinants of innate immune recognition of bornaviral nucleic acids (Martin et al. 2011). In invertebrates, piRNA-like RNAs can be generated from viral sequences (Wu et al. 2010; Morazzani et al. 2012; Léger et al. 2013), and knock-down of piRNA pathway components has a proviral effect in mosquito cells (Schnettler et al. 2013) and flies (Zamboni et al. 2006). Viral piRNA biogenesis in these invertebrates is assumed to be due to an interaction between viral genomic or transcript RNAs with PIWI proteins and piRNA pathway components in the absence of an integrated DNA intermediate. Moreover, in *Aedes aegypti* cells distinct PIWI proteins are required for piRNA biogenesis from viral precursors compared with endogenous transposon precursors. Thus presumably some feature of at least some viral nucleic acids is sufficient to determine their recognition for processing into piRNA-like molecules. The RNA-binding proteins involved in piRNA precursor processing (e.g., MOV10L1) may share common specificities with RNA-binding proteins involved in recognizing infectious viruses (e.g., MOV10 [Wang et al. 2010]), thus it is conceivable that integration of an EBLN into an existing transcriptional unit could have been responsible for that transcript's definition as a piRNA precursor. If this were the case, the convergent evolution described above could have arisen in a direct, mechanistic fashion. We can reject this possibility in several instances, as some piRNA clusters are syntenic between primates and rodents

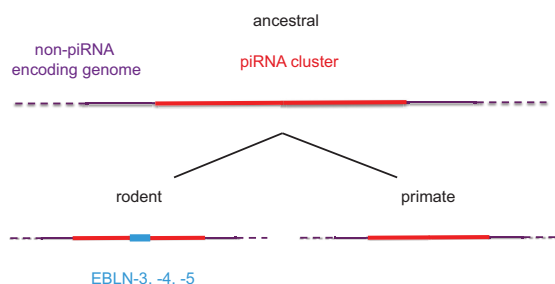


FIGURE 3. Rodent EBLNs were integrated into existing piRNA clusters. The genome of the common rodent/primate ancestor is depicted, with sequence predicted to generate piRNAs shown in red and the non-piRNA encoding genome in purple. After the divergence of rodents and primates, an EBLN (blue) was integrated into this pachytene piRNA precursor transcriptional unit. This pattern of evolution occurred for each EBLN currently found in rodent genomes, suggesting that EBLN sequences were not themselves required for piRNA-generating capacity to their surrounding piRNA cluster sequences.

(Hirano et al. 2014). Indeed, rodent *EBLN-3* and *-5* are within large intergenic piRNA clusters also present in humans, as is the rat piRNA cluster containing *EBLN-4* (Table 2; Fig. 3). The marmoset EBLNs generating bona fide piRNAs (*cjEBLN-6* and *-7*) are within the 3' untranslated regions of genes (*NEU3* and *TBC1D12*) that also generate piRNAs in the mouse, albeit much less abundantly (Robine et al. 2009). Thus at least some EBLNs were integrated into preexisting piRNA clusters. Beyond this, we cannot exclude an influence of EBLN integration on the life/death evolutionary dynamics of piRNA clusters (Assis and Kondrashov 2009).

DISCUSSION

Two models could account for the convergent evolution of piRNA-generating EBLNs in two mammalian lineages: one in which EBLNs were preferentially integrated within piRNA clusters, and another in which EBLNs integrated randomly and those within piRNA clusters were positively selected. We cannot exclude the first model, but consider it unlikely based on the distribution of other nucleic acids mobilized by retrotransposons: Among the over 10,000 and 5000 GENCODE-annotated (Harrow et al. 2012) processed pseudogenes in the human and mouse genomes, respectively, few are within piRNA clusters (Ha et al. 2014; Hirano et al. 2014; Watanabe et al. 2015). Furthermore, LINE-1-mobilized sequences in cultured cells (Berry et al. 2006), tumors (Cooke et al. 2014; Tubio et al. 2014), and mice bearing an engineered LINE-1 element (An et al. 2006) are not preferentially targeted to piRNA clusters. Finally, recent processed pseudogene (Abyzov et al. 2013) and retrotransposon insertions in human genomes are not enriched within piRNA-generating loci (Y Zhang and P Gerstein, pers. comm.). Nonetheless, experiments to determine if, under certain circumstances, nonself nucleic acids are preferentially trapped within piRNA clusters are warranted (Kawaoka et al. 2013).

Enrichment of EBLNs within piRNA-generating loci in two mammalian lineages is consistent with natural selection (Fig. 4). EBLN-derived piRNAs may have fortuitously silenced a host gene shared by these lineages, or multiple EBLNs could have hitchhiked with linked alleles. Another parsimonious hypothesis is that EBLN-derived piRNAs silenced bornavirus and thus protected against death or reduced fecundity due to bornaviral infection, reminiscent of transposon (Khurana et al. 2011) or errantiviral (Prud'homme et al. 1995) resistance after sequences from these elements are integrated into *Drosophila* piRNA clusters. Such an effect is feasible and consistent with current models of the mammalian piRNA system if ancient bornaviruses were transmitted vertically via the gamete. Indeed, the ancient bornaviruses giving rise to EBLNs infected germ cells; otherwise a noncanonical mechanism of gene transfer from soma to germline, in violation of Weismann's law, was responsible for EBLN formation (Pittoggi et al. 2006). Of note, modern bornaviruses can be transmitted vertically (Okamoto et al. 2003; Kerski et al. 2012). EBLN integration into piRNA clusters could thus have resulted in viral silencing in germ cells, similar to the transgene silencing observed after insertion of identical sequences into piRNA clusters (Yamamoto et al. 2013). As transcriptional silencing via repressive chromatin modification, rather than post-transcriptional silencing, appears the dominant mechanism of piRNA-mediated silencing, it is notable that modern bornaviruses, unlike most RNA viruses, replicate in the nucleus and interact directly with chromatin (Matsumoto et al. 2012). Further, while an antiviral effect of RNA interference has been detected in certain mammalian systems (Li et al. 2013b), it is considered to have been largely superseded in mammalian antiviral innate immunity by type I interferons (Cullen et al. 2013; Cullen 2014). However, overexpression of interferon α prevents germ-cell development in mice and its receptor is not expressed on pachytene spermatocytes (Satie et al. 2011), perhaps censoring this antiviral mechanism in the germline (Pare and Sullivan 2014).

PiRNA-like RNAs and/or PIWI proteins have been described in primate pluripotent cells (Marchetto et al. 2013), human hematopoietic cells (Sharma et al. 2001; Cichocki et al. 2010), and some other somatic cells including neurons (Lee et al. 2011; Yan et al. 2011; Rajasethupathy et al. 2012), a critical target cell of modern bornaviruses. Notably neurons, like germ cells, are permissive to retrotransposition (Upton et al. 2015) and relatively nonresponsive to type I interferons (Lin et al. 2013; Kreit et al. 2014). Transcripts of the murine PIWI-clade partner of most pachytene piRNAs (MIWI) can be detected in a distribution overlapping highly BDV-susceptible cells in the cerebellum, dentate gyrus, and olfactory bulb (Lein et al. 2007; Ackermann et al. 2010). Thus EBLN-derived piRNAs could potentially have protected stem or somatic cells like neurons from bornavirus-induced pathology, although more work is required to assess the feasibility of this model. A class of piRNA-like RNAs derived from the

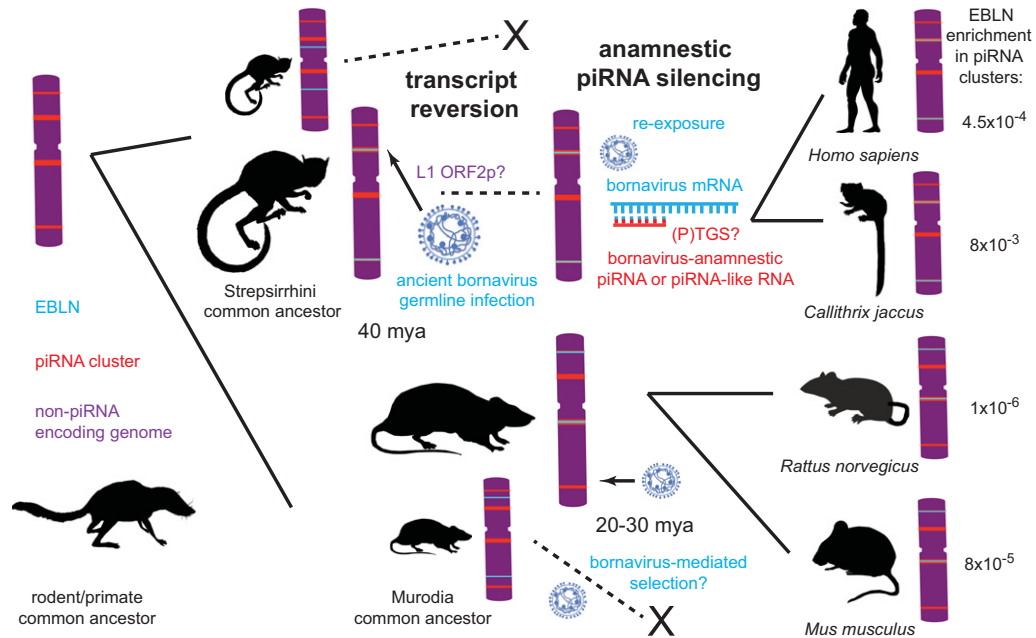


FIGURE 4. Conceptual model of mammalian EBLN formation and selection. A model chromosome of a common rodent/primate ancestor is shown at left. The non-piRNA encoding genome is shown in purple, piRNA clusters in red, and EBLN integrations are indicated in blue. Multiple sequences from ancient bornaviruses were integrated into the germline of individuals of the indicated lineages via transcript reversion ([mya] millions of years ago). The likely candidate for such activity in the primate lineage is the LINE-1 ORF2p, encoded by the host genome. We hypothesize that natural selection, perhaps mediated by differential survival or fecundity when animals with EBLNs faced subsequent bornavirus infection, could be responsible. One potential mechanism for this selective advantage, piRNA or piRNA-like RNA-guided viral gene silencing via post-transcriptional or transcriptional gene silencing ([P]TGS), is depicted. The observation that EBLNs are enriched within piRNA clusters in multiple species of each lineage is improbable in the absence of selection, and an approximation of this probability is listed for each species.

same genomic loci that give rise to pachytene piRNAs, yet with different genetic requirements for their biogenesis, have recently been described in a variety of somatic cell types (Ortoger et al. 2014). In addition, dicer-dependent endogenous small interfering RNAs are produced from some piRNA clusters (Watanabe et al. 2008). Thus production of piRNAs from EBLNs in the testis could correlate with production of other, more relevant, RNA molecules from these loci elsewhere.

It has previously been suggested that transposon silencing may not be the sole function of mammalian pachytene piRNAs, most of which are highly complementary only to the locus from which they are transcribed and are derived from unannotated intergenic regions depleted in transposon-derived sequences relative to the genome as a whole (Aravin et al. 2007). While their potential targets are thus unclear, piRNA clusters in rodents (Assis and Kondrashov 2009) and humans (Lukic and Chen 2011) evolve rapidly under positive selection. Recent reports suggest that piRNAs may target mRNAs to which they are only partially complementary, similar to miRNA:target interactions (Zhang et al. 2015). However, the biochemically confirmed functions of PIWI proteins require extensive piRNA:target complementarity (Reuter et al. 2011). Finding that pachytene piRNAs are made from sequences derived from an exogenous virus suggests an alternative explanation for these observations,

as well as a potential role for the transcription of piRNA precursors outside the germline: similar to CRISPR spacers (Mojica et al. 2005), they may serve an immunologic function by targeting sequences foreign to the genome from which they are derived (Sagy et al. 2014).

Genetic material laterally transferred from nonretroviral viruses to host genomes has been noted for a wide variety of pathogens and hosts (Zhdanov 1975; Belyi et al. 2010; Gilbert and Feschotte 2010; Kapoor et al. 2010; Katzourakis and Gifford 2010; Liu et al. 2010, 2011). Several of these gene transfer events can be clearly attributed to retrotransposon activity (Ballinger et al. 2012), and in at least one case the gene transfer is associated with viral resistance (Maori et al. 2007). We hypothesize that, lacking RNA-dependent RNA-polymerases used to amplify RNAs for viral silencing in organisms for which this function is well-established (Baulcombe 2004; Rechavi et al. 2011), some metazoans could use a multistep process: reverse transcription of viral transcripts followed by RNA polymerization templated on viral complementary DNA (cDNA). Sufficient reverse transcriptase activity acting on viral RNA is present in some mammalian cells to allow laboratory time-scale experiments (Pittoggi et al. 2006; Horie et al. 2010); LINE-1 ORF2p has recently been suggested to be the responsible enzyme in human cells (Shimizu et al. 2014). While no function has been ascribed to cDNA generated by this activity in mammals,

similar viral cDNA is a source of antiviral interfering RNA in *Drosophila* (Goic et al. 2013).

The described interactions between infectious viruses and the transposons that comprise the bulk of mammalian host genomes are consistent with the hypothesis that transposons are symbionts integral to genomic stress responses (McClintock 1984), including antiviral immune responses (Zeng et al. 2014; Yu et al. 2015). The piRNA system in mammals is known to silence quasi-nonsel self transposon nucleic acids; our observations raise the hypothesis that, as for CRISPR/Cas in prokaryotes, truly exogenous nonself nucleic acids from infecting viruses can be targeted by piRNA-like RNAs, but that this requires genetic information flow in an unexpected retrotransposon-dependent manner (Nuñez et al. 2015). We refer to this hypothetical mechanism as viral transcript reversion with anamnestic piRNA silencing (TRAPS) to indicate the host-directed nature of the reverse transcription involved, its potential role in heritable immune memory, and the specialized genomic loci involved in capturing the nonself information. Testing this hypothesis is of proximal relevance to human health, namely in arbovirus/vector interactions (Arensburger et al. 2011) and resistance to bornaviral disease.

MATERIALS AND METHODS

EBLN annotation

The genomic locations of rodent and primate EBLNs have been described elsewhere (Belyi et al. 2010; Horie et al. 2010, 2013). These studies defined EBLNs on the basis of potential ORFs or regions of amino acid-based homologies with BDV N. In order to evaluate the hypothesis that small RNA generated from these loci could have biological function, EBLN annotation in Table 2 is inclusive of the entire inserted BDV N mRNA-like sequence block (i.e., for primate EBLNs, the sequence flanked by TSDs).

Small RNA sequencing

One 6-wk-old BALB/c mouse was purchased from Charles River Laboratories, Japan. Testis total RNA including small RNA fraction was collected from the mouse using a miRNeasy mini Kit (QIAGEN). RNA quality was confirmed by 2100 Agilent Bioanalyzer (Agilent Technologies). After quality confirmation, cDNA libraries were constructed from the testis RNAs by TruSeq small RNA sample prep kit (Illumina). Small RNA sequencing was performed using an Illumina HiSeq (50SE) machine by Hokkaido Systems Science. The sequence data were mapped onto mm9 using BWA (Li and Durbin 2009) allowing up to two mismatches. Reads mapping multiply were assigned at random to a single map site. Of 29,959,596 total reads, 27,968,611 reads were mapped, of which 3,184,031 mapped repetitively.

Mapping published small RNA-seq data

To analyze murine piRNAs corresponding to EBLNs, we used GSM1096587 as small RNA data of 6-wk mouse testis (Li et al. 2013a). RNAs of 25 to 31 nt in length were mapped to mEBLNs

by Bowtie (Langmead et al. 2009) allowing up to two mismatches. This bioinformatic analysis was performed using Galaxy (<https://usegalaxy.org>).

Cumulative binomial probability

The likelihood of x EBLNs or more, out of a total of n EBLNs for a given species, being found within a piRNA cluster if piRNA clusters occupy p percent of the genome was approximated as the probability \mathbb{P} of x successes or more in n Bernoulli trials, each with a probability p :

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

piRNA cluster evolution

Ancestral piRNA clusters were as determined by Assis and Kondrashov (2009), who considered rodent clusters in detail. To identify the homologous primate clusters in human and marmoset, LASTZ alignments of murine piRNA clusters to each of these species were viewed using Ensembl and compared with annotations in Ha et al. (2014) and Hirano et al. (2014).

Murine genes with 3' UTRs giving rise to piRNA were determined by Robine et al. (2009).

ACKNOWLEDGMENTS

We thank Masayuki Horie, Cedric Feschotte, Craig B. Wilen, Bryan R. Cullen, Alexei A. Aravin, Andrew Z. Fire, Rafi Ahmed, Arnold J. Levine, Sara Cherry, and P. Jeremy Wang for helpful discussions and/or comments on the manuscript. N.F.P. thanks Bruce R. Levin for encouraging a hunt for a mammalian CRISPR-like immune system. N.F.P. was supported by the Japan Society for the Promotion of Science short-term postdoctoral fellowship award #PE13075. This study was supported in part by Funding Program for Next Generation World-Leading Researchers (NEXT program), KAKENHI grant numbers 26253027 and 26670225, and the Core-to-Core Program A, Advanced Research Networks from the Japan Society for the Promotion of Science (JSPS) (K.T.); grants from Takeda Science Foundation (K.T.); and KAKENHI grant numbers 25115508 and 25860336 from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan (T.H.).

Received April 16, 2015; accepted July 8, 2015.

REFERENCES

- Abyzov A, Iskow R, Gokcumen O, Radke DW, Balasubramanian S, Pei B, Habegger L; 1000 Genomes Project Consortium, Lee C, Gerstein M. 2013. Analysis of variable retroduplications in human populations suggests coupling of retrotransposition to cell division. *Genome Res* **23**: 2042–2052.
- Ackermann A, Guelzow T, Staeheli P, Schneider U, Heimrich B. 2010. Visualizing viral dissemination in the mouse nervous system, using a green fluorescent protein-expressing Borna disease virus vector. *J Virol* **84**: 5438–5442.
- An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, Triplett C, Boeke JD. 2006. Active retrotransposition by a synthetic L1 element in mice. *Proc Natl Acad Sci* **103**: 18662–18667.

- 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.
- Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. 2011. The mosquito *Aedes aegypti* has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. *BMC Genomics* **12**: 606.
- Assis R, Kondrashov AS. 2009. Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. *Proc Natl Acad Sci* **106**: 7079–7082.
- Aswad A, Katzourakis A. 2012. Paleovirology and virally derived immunity. *Trends Ecol Evol* **27**: 627–636.
- Ballinger MJ, Bruenn JA, Taylor DJ. 2012. Phylogeny, integration and expression of sigma virus-like genes in *Drosophila*. *Mol Phylogenet Evol* **65**: 251–258.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712.
- Baulcombe D. 2004. RNA silencing in plants. *Nature* **431**: 356–363.
- Belyi VA, Levine AJ, Skalka AM. 2010. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog* **6**: e1001030.
- Berry C, Hannehalli S, Leipzig J, Bushman FD. 2006. Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput Biol* **2**: e157.
- Beyret E, Liu N, Lin H. 2012. piRNA biogenesis during adult spermatogenesis in mice is independent of the ping-pong mechanism. *Cell Res* **22**: 1429–1439.
- Cichocki F, Lenvik T, Sharma N, Yun G, Anderson SK, Miller JS. 2010. Cutting edge: KIR antisense transcripts are processed into a 28-base PIWI-like RNA in human NK cells. *J Immunol* **185**: 2009–2012.
- Cooke SL, Shlien A, Marshall J, Pipinikas CP, Martincorena I, Tubio JM, Li Y, Menzies A, Mudie L, Ramakrishna M, et al. 2014. Processed pseudogenes acquired somatically during cancer development. *Nat Commun* **5**: 3644.
- Crick FH. 1958. On protein synthesis. *Symp Soc Exp Biol* **12**: 138–163.
- Cullen BR. 2014. Viruses and RNA interference: issues and controversies. *J Virol* **88**: 12934–12936.
- Cullen BR, Cherry S, tenOever BR. 2013. Is RNA interference a physiologically relevant innate antiviral immune response in mammals? *Cell Host Microbe* **14**: 374–378.
- Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**: 363–367.
- Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K. 2014. Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. *Proc Natl Acad Sci* **111**: 13175–13180.
- Geib T, Sauder C, Venturelli S, Hässler C, Staeheli P, Schwemmler M. 2003. Selective virus resistance conferred by expression of Borna disease virus nucleocapsid components. *J Virol* **77**: 4283–4290.
- Gilbert C, Feschotte C. 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. *PLoS Biol* **8**: pii: e1000495.
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199–202.
- Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, Blanc H, Gausson V, Vera-Otarola J, Cristofari G, Saleh MC. 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. *Nat Immunol* **14**: 396–403.
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**: 1193–1197.
- Guo X, Lin M, Rockowitz S, Lachman HM, Zheng D. 2014. Characterization of human pseudogene-derived non-coding RNAs for functional potential. *PLoS One* **9**: e93972.
- 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.
- Hainaut P, Castellazzi M, Gonzales D, Clause N, Hilgers J, Crepin M. 1990. A congenic line of the BALB/c mouse strain with the endogenous mouse mammary tumor virus proviral gene Mtv-3: tissue-specific expression and correlation with resistance to mouse mammary tumor virus infection and tumorigenesis. *Cancer Res* **50**: 3754–3760.
- 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.
- 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.
- Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Ikuta K, Jern P, Gojobori T, Coffin JM, et al. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463**: 84–87.
- Horie M, Kobayashi Y, Suzuki Y, Tomonaga K. 2013. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. *Philos Trans R Soc Lond B Biol Sci* **368**: 20120499.
- Ikeda H, Odaka T. 1983. Cellular expression of murine leukemia virus gp70-related antigen on thymocytes of uninfected mice correlates with Fv-4 gene-controlled resistance to Friend leukemia virus infection. *Virology* **128**: 127–139.
- Ishizu H, Siomi H, Siomi MC. 2012. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes Dev* **26**: 2361–2373.
- Kapoor A, Simmonds P, Lipkin WI. 2010. Discovery and characterization of mammalian endogenous parvoviruses. *J Virol* **84**: 12628–12635.
- Karginov FV, Hannon GJ. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* **37**: 7–19.
- Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, et al. 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* **42**: D764–D770.
- Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. *PLoS Genet* **6**: e1001191.
- Kawaoka S, Hara K, Shoji K, Kobayashi M, Shimada T, Sugano S, Tomari Y, Suzuki Y, Katsuma S. 2013. The comprehensive epigenome map of piRNA clusters. *Nucleic Acids Res* **41**: 1581–1590.
- Kerski A, de Kloet AH, de Kloet SR. 2012. Vertical transmission of avian bornavirus in Psittaciformes: avian bornavirus RNA and anti-avian bornavirus antibodies in eggs, embryos, and hatchlings obtained from infected sun conures (*Aratinga solstitialis*). *Avian Dis* **56**: 471–478.
- Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, Theurkauf WE. 2011. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **147**: 1551–1563.
- Kirino Y, Mourelatos Z. 2007. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat Struct Mol Biol* **14**: 347–348.
- Kobayashi Y, Horie M, Tomonaga K, Suzuki Y. 2011. No evidence for natural selection on endogenous borna-like nucleoprotein elements after the divergence of Old World and New World monkeys. *PLoS One* **6**: e24403.
- Kreit M, Paul S, Knoops L, De Cock A, Sorgeloos F, Michiels T. 2014. Inefficient type I interferon-mediated antiviral protection of primary mouse neurons is associated with the lack of apolipoprotein I9 expression. *J Virol* **88**: 3874–3884.
- 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.
- Lee EJ, Banerjee S, Zhou H, Jammalamadaka A, Arcila M, Manjunath BS, Kosik KS. 2011. Identification of piRNAs in the central nervous system. *RNA* **17**: 1090–1099.
- Léger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppée JY, Bonnefoy E, Bouloy M. 2013. Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells. *J Virol* **87**: 1631–1648.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**: 168–176.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- 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 Y, Lu J, Han Y, Fan X, Ding SW. 2013b. RNA interference functions as an antiviral immunity mechanism in mammals. *Science* **342**: 231–234.
- Lin CC, Wu YJ, Heimrich B, Schwemmler M. 2013. Absence of a robust innate immune response in rat neurons facilitates persistent infection of Borna disease virus in neuronal tissue. *Cell Mol Life Sci* **70**: 4399–4410.
- Liu H, Fu Y, Jiang D, Li G, Xie J, Cheng J, Peng Y, Ghabrial SA, Yi X. 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *J Virol* **84**: 11876–11887.
- Liu H, Fu Y, Li B, Yu X, Xie J, Cheng J, Ghabrial SA, Li G, Yi X, Jiang D. 2011. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. *BMC Evol Biol* **11**: 276.
- Lukic S, Chen K. 2011. Human piRNAs are under selection in Africans and repress transposable elements. *Mol Biol Evol* **28**: 3061–3067.
- Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136**: 656–668.
- Maori E, Tanne E, Sela I. 2007. Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology* **362**: 342–349.
- Marchetto MC, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL, Paquola AC, Desai KN, Herai RH, Weitzman MD, et al. 2013. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**: 525–529.
- Martin A, Hoefs N, Tadewaldt J, Staeheli P, Schneider U. 2011. Genomic RNAs of Borna disease virus are elongated on internal template motifs after realignment of the 3' termini. *Proc Natl Acad Sci* **108**: 7206–7211.
- Matsumoto Y, Hayashi Y, Omori H, Honda T, Daito T, Horie M, Ikuta K, Fujino K, Nakamura S, Schneider U, et al. 2012. Bornavirus closely associates and segregates with host chromosomes to ensure persistent intranuclear infection. *Cell Host Microbe* **11**: 492–503.
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792–801.
- Mojica FJ, Diez-Villaseñor C, García-Martínez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174–182.
- Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. 2012. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathog* **8**: e1002470.
- Mura M, Murcia P, Caporale M, Spencer TE, Nagashima K, Rein A, Palmarini M. 2004. Late viral interference induced by transdominant Gag of an endogenous retrovirus. *Proc Natl Acad Sci* **101**: 11117–11122.
- Núñez JK, Lee AS, Engelman A, Doudna JA. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**: 193–198.
- Okamoto M, Hagiwara K, Kamitani W, Sako T, Hirayama K, Kirisawa R, Tsuji M, Ishihara C, Iwai H, Kobayashi T, et al. 2003. Experimental vertical transmission of Borna disease virus in the mouse. *Arch Virol* **148**: 1557–1568.
- Ortogoero N, Schuster AS, Oliver DK, Riordan CR, Hong AS, Hennig GW, Luong D, Bao J, Bhetwal BP, Ro S, et al. 2014. A novel class of somatic small RNAs similar to germ cell pachytene PIWI-interacting small RNAs. *J Biol Chem* **289**: 32824–32834.
- Pare JM, Sullivan CS. 2014. Distinct antiviral responses in pluripotent versus differentiated cells. *PLoS Pathog* **10**: e1003865.
- Pittoggi C, Beraldi R, Sciamanna I, Barberi L, Giordano R, Magnano AR, Torosantucci L, Pescarmona E, Spadafora C. 2006. Generation of biologically active retro-genes upon interaction of mouse spermatozoa with exogenous DNA. *Mol Reprod Dev* **73**: 1239–1246.
- Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A. 1995. Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* **139**: 697–711.
- Rajasethupathy P, Antonov I, Sheridan R, Frey S, Sander C, Tuschl T, Kandel ER. 2012. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149**: 693–707.
- Rechavi O, Minevich G, Hobert O. 2011. Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**: 1248–1256.
- 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.
- Sagy O, Shamir R, Rechavi O. 2014. Examination of exhaustive cloning attempts reveals that *C. elegans* piRNAs, transposons, and repeat sequences are efficiently cloned in yeast, but not in bacteria. *Front Genet* **5**: 275.
- Satie AP, Mazaud-Guittot S, Seif I, Mahé D, He Z, Jouve G, Jégou B, Dejuq-Rainsford N. 2011. Excess type I interferon signaling in the mouse seminiferous tubules leads to germ cell loss and sterility. *J Biol Chem* **286**: 23280–23295.
- Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, Fazakerley JK, Kohl A, Fragkoudis R. 2013. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. *J Gen Virol* **94**: 1680–1689.
- Sharma AK, Nelson MC, Brandt JE, Wessman M, Mahmud N, Weller KP, Hoffman R. 2001. Human CD34⁺ stem cells express the *hiwi* gene, a human homologue of the *Drosophila* gene *piwi*. *Blood* **97**: 426–434.
- Shimizu A, Nakatani Y, Nakamura T, Jinno-Oue A, Ishikawa O, Boeke JD, Takeuchi Y, Hoshino H. 2014. Characterisation of cytoplasmic DNA complementary to non-retroviral RNA viruses in human cells. *Sci Rep* **4**: 5074.
- Siomi MC, Kuramochi-Miyagawa S. 2009. RNA silencing in germlines—exquisite collaboration of Argonaute proteins with small RNAs for germline survival. *Curr Opin Cell Biol* **21**: 426–434.
- Sorek R, Kunin V, Hugenholtz P. 2008. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181–186.
- Tomonaga K, Kobayashi T, Ikuta K. 2002. Molecular and cellular biology of Borna disease virus infection. *Microbes Infect* **4**: 491–500.
- Tubio JM, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, et al. 2014. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**: 1251343.
- Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sánchez-Luque FJ, Bodea GO, Ewing AD, Salvador-Palomeque C, van der Knaap MS, Brennan PM, et al. 2015. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161**: 228–239.
- Vourekas A, Zheng K, Fu Q, Maragkakis M, Alexiou P, Ma J, Pillai RS, Mourelatos Z, Wang PJ. 2015. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes Dev* **29**: 617–629.

- Wang X, Han Y, Dang Y, Fu W, Zhou T, Ptak RG, Zheng YH. 2010. Moloney leukemia virus 10 (MOV10) protein inhibits retrovirus replication. *J Biol Chem* **285**: 14346–14355.
- Wang W, Yoshikawa M, Han BW, Izumi N, Tomari Y, Weng Z, Zamore PD. 2014. The initial uridine of primary piRNAs does not create the tenth adenine that is the hallmark of secondary piRNAs. *Mol Cell* **56**: 708–716.
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539–543.
- Watanabe T, Cheng EC, Zhong M, Lin H. 2015. Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Res* **25**: 368–380.
- Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, Ding SW. 2010. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci* **107**: 1606–1611.
- Yamamoto Y, Watanabe T, Hoki Y, Shirane K, Li Y, Ichiiyanagi K, Kuramochi-Miyagawa S, Toyoda A, Fujiyama A, Oginuma M, et al. 2013. Targeted gene silencing in mouse germ cells by insertion of a homologous DNA into a piRNA generating locus. *Genome Res* **23**: 292–299.
- Yan Z, Hu HY, Jiang X, Maierhofer V, Neb E, He L, Hu Y, Hu H, Li N, Chen W, et al. 2011. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res* **39**: 6596–6607.
- Yap MW, Colbeck E, Ellis SA, Stoye JP. 2014. Evolution of the retroviral restriction gene Fv1: inhibition of non-MLV retroviruses. *PLoS Pathog* **10**: e1003968.
- Yu Q, Carbone CJ, Katlinskaya YV, Zheng H, Zheng K, Luo M, Wang PJ, Greenberg RA, Fuchs SY. 2015. Type I interferon controls propagation of Long Interspersed Element-1. *J Biol Chem* **290**: 10191–10199.
- Zambon RA, Vakharia VN, Wu LP. 2006. RNAi is an antiviral immune response against a dsRNA virus in *Drosophila melanogaster*. *Cell Microbiol* **8**: 880–889.
- Zeng M, Hu Z, Shi X, Li X, Zhan X, Li XD, Wang J, Choi JH, Wang KW, Purrington T, et al. 2014. MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. *Science* **346**: 1486–1492.
- Zhang Z, Harrison PM, Liu Y, Gerstein M. 2003. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* **13**: 2541–2558.
- 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.
- Zhdanov VM. 1975. Integration of viral genomes. *Nature* **256**: 471–473.
- Zheng D, Frankish A, Baertsch R, Kapranov P, Reymond A, Choo SW, Lu Y, Denoeud F, Antonarakis SE, Snyder M, et al. 2007. Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res* **17**: 839–851.