Transposons and the PIWI pathway: genome defense in gametes and embryos

Stewart J Russell1 and Jonathan LaMarre2

1CReATE Fertility Centre, Toronto, Ontario, Canada and 2Department of Biomedical Sciences, University of Guelph, Guelph, Ontario, Canada

Correspondence should be addressed to J LaMarre; Email: jlamarre@uoguelph.ca

Abstract

Hiding in plain sight within the genome of virtually every eukaryotic organism are large numbers of sequences known as transposable elements (TEs). These sequences often comprise 50% or more of the DNA in many mammals and are transcriptionally constrained by DNA methylation and repressive chromatin marks. Individual TEs, when relieved of these epigenetic constraints, can readily move from one genomic location to another, either directly or through RNA intermediates. Demethylation and removal of repressive histone marks during epigenetic reprogramming stages of gametogenesis and embryogenesis render the genome particularly susceptible to increased TE mobilization, which has significant implications for the fidelity of genome replication and subsequent viability of the progeny. Importantly, however, TEs have functionally integrated themselves into developmental events to the extent that complete suppression precludes normal gamete and embryo development. Consequently, multiple mechanisms have evolved to limit the extent of TE expression and mobilization during reprogramming without completely suppressing it. One of the most important TE repression mechanisms is the PIWI/piRNA pathway, in which 25–32 nucleotide RNA molecules known as piRNAs associate with Argonaute proteins from the PIWI clade to form piRISC complexes. These complexes target and silence TEs post-transcriptionally and through the induction of epigenetic changes at the loci from which they are expressed. This review will briefly discuss the intricate molecular détente between TE expression and its suppression by the PIWI pathway, with particular emphasis on mammalian species including human, bovine and murine.

Genomic assailants from within: transposable elements

The trans-generational propagation of genes from parent to offspring is an indispensable component of the process of reproduction. Faithful genome replication is threatened in eukaryotic organisms by viral elements that have the potential to cause mutations and compromise the fitness of subsequent generations or damage the genome such that reproduction is no longer possible. Surprisingly perhaps, the richest source of these viral elements lies within an organism’s own DNA in the form of integrated sequences known as transposable elements (TEs) that are often virus derived.

TEs are mobile DNA sequences present within the eukaryotic genome (Fedoroff 2012). They were discovered in plants by Barbara McClintock where they were observed to move within and between chromosomes and subsequently named ‘jumping genes’ (McClintock 1950). In higher eukaryotes, these sequences constitute a significant fraction of the genome; approximately 50% of human and bovine genomes are composed of TEs and their derivatives (Lander et al. 2001). While a large number of these elements are evolutionarily conserved, more than one quarter appear to be species specific (Adelson et al. 2009). Most TEs are fragmented and cannot undergo independent transposition, but some remain intact and competent to mobilize within the genome. The frequency of activation is reflected in the observation that approximately 10% of all spontaneous genome mutations in rodents appear to result from TE mobilization (Kazazian 1998). Surprisingly perhaps, mechanisms to completely eliminate or permanently suppress transposons within the genome have not evolved, raising important questions with respect to the high potential fitness costs of continually replicating large amounts of DNA with no obvious coding or regulatory functions. In particular, the possibility must be considered that TEs provide unrecognized advantages to host genomes.

Transposable elements are categorized in two classes and a number of subclasses based on the mechanism of transposition. Class I elements, also known as retrotransposons, (Wicker et al. 2007, 2008) move through RNA intermediates that are transcribed in a manner similar to coding genes. Intact retrotransposons can mobilize through transcription then re-integrate into the genome through the activities of reverse transcriptase.
(RT) and endonuclease enzymes that are encoded by the TE (Wicker et al. 2007). Common examples of this class are long terminal repeat (LTR)/endogenous retroviruses (ERV) and non-LTR retrotransposons, including the long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (Fig. 1A).

Within the LINE retrotransposons, LINE1 (L1) is an autonomous element (encoding the proteins required for retrotransposition) normally consisting of a few thousand base pairs (bp) (Swergold 1990). It is transcribed similarly to genic mRNA, utilizing RNA Polymerase II (PoI II) followed by 5’ capping and polyadenylation. Two open reading frames (ORFs) are present in a complete L1 transcript; ORF1 encodes an RNA-binding protein and ORF2 encodes a protein with RT and endonuclease activities (Ostertag & Kazazian 2001, Martin 2006). Together, ORF1 and ORF2 facilitate L1 RNA translocation to the nucleus, reverse transcription and reintegration into DNA (Ostertag & Kazazian 2001) (Fig. 1B). The L1 element has been described as the only autonomous non-LTR retrotransposon in mammals, with the exception of marsupials and ruminants which have LINE RTE repeats in addition to LINE1 (Deininger et al. 2003). The bovine genome encodes nine complete copies of the LINE RTE BovB, which includes a reverse transcriptase similar to L1 (Adelson et al. 2009) suggesting that the L1 class of retrotransposons may be active in this species.

In contrast to LINEs, SINEs are non-autonomous and transcribed by PoI III (Krouter et al. 2009). SINEs (and some fragmented LINEs) capitalize on the L1 retrotransposition machinery, specifically requiring ORF2 from L1 for successful retrotransposition (Krouter et al. 2009). There is greater variation in SINE classes between species, with bovine SINEs dominated by BOV-A2, Bov-tA and tRNA insertions (Adelson et al. 2009) and human by Alu and SVA type SINEs (Ostertag et al. 2003, An et al. 2004). SINE elements range from 85 to 500 base pairs and contain an internal PoI III promoter (Deininger 2002). However, the majority of both LINEs and SINEs in mammalian genomes have lost their functional promoters (Deininger 2002).

Endogenous retroviruses (ERVs) are less abundant in the bovine genome, at between 0.4 and 4.3% of total DNA and approximately 13,000 copies, depending on the detection algorithm applied (Garcia-Etxebarria & Jugo 2010). Structurally, ERVs encode the canonical retroviral GAG, POL and ENV proteins over 5–10 kilobases (kb), flanked by LTRs of between 300 and 1200 bps. The gag gene encodes the core viral proteins that bind the retroviral RNA genome, the pol region encodes the RT and integrase responsible for genomic reintegration and the env region encodes surface and transmembrane proteins that facilitate receptor binding and membrane fusion (Ryan 2004). Importantly, ERVs show considerable variability in sequence due to the accumulation of mutations, insertions and deletions.
Transposons and the PIWI pathway

Stoye 2012). Moreover, recombination events involving LTRs of ERVs can lead to excision of the coding regions, leaving behind ‘Solo LTRs’ that are much more abundant than full or partial ERVs (Benachenhou et al. 2009). Recent studies by Svoboda and colleagues (Franke et al. 2017) have demonstrated that LTRs have strong potential to play important roles in gene evolution and expression control in oocytes and early embryos of many different species including bovine.

The second major class of TEs consists of Class II elements, also called DNA-transposons, which replicate through a DNA intermediate, by rolling circle replication or through other unknown mechanisms (Wicker et al. 2007). Class II TEs comprise only about 2% of the bovine genome and will not be discussed further here. For a recent discussion of DNA TEs in mammals (see Hickman & Dyda 2015).

Transposable element expression

The expression of TEs is normally suppressed epigenetically by methylation of cytosine and guanine residues in DNA and repressive marks on histones (Huda et al. 2010, Zamudio et al. 2015). However, as described in more detail below, TEs become transcribed at higher frequency during periods of genome demethylation, increasing the probability of transposition events. In principle, transposition leads to one of three outcomes: positive, negative or neutral effect on host viability. The net effect depends on the integration site and downstream impacts of the new TE on the expression or function of nearby genes. TEs have contributed substantially to the evolution of the bovine genome in which an average of 18 TE insertions per gene (in regulatory and coding regions) have been identified (Almeida et al. 2007). While no formal associations between TEs and heritable diseases in cattle are apparent in the literature, many examples of diseases resulting from retrotransposition are present in humans (Kazazian et al. 1988). TE mobilization is also observed in different human cancers, but whether TE expression is the cause, or a consequence, of genomic instability is not clear (reviewed in Belancio et al. 2010). Mechanistically, disease-causing TE insertions typically disrupt the coding sequence or processing (i.e. splicing) of mRNAs (Maksakova et al. 2006). As with any mutation, TE insertion would be predicted to cause more negative than positive effects on the host cell, and therefore, most cells or progeny in which retrotransposition occur are likely subject to negative selection.

TE expression in embryos

Despite the deleterious potential of TE expression and reintegration, specific TEs and their truncated transcripts appear to be beneficial, or at least required, during embryogenesis where they are frequently expressed (Beraldi et al. 2006). In support of a functional requirement for specific TE expression, suppression of the L1 ORF1 in murine zygotes increases the rate of embryo arrest (Beraldi et al. 2006). A recent investigation by Jachowicz et al. demonstrated that not only is LINE1 expression required for early mouse embryogenesis, but that it must be subsequently silenced for successful progression to the blastocyst stage (Jachowicz et al. 2017). Furthermore knockdown of MuERV-L, one of the first transcripts generated from the quiescent mouse genome after fertilization, leads to a developmental block at the four-cell stage (Kigami et al. 2003). Moreover, ample evidence suggests that specific TE transcripts are actively regulated throughout development in many species (Ge 2017).

Evolutionary selection for TE expression may have resulted from the tendency of these elements to contain regulatory domains that are frequently co-opted by genes near the insertion sites (Töhönen et al. 2015, Franke et al. 2017). The regulatory mechanisms in which they participate can be particularly complex, making them difficult to fully elucidate (Sokol et al. 2015). The contribution of TEs to gene regulation in mouse and human genomes has been investigated at the genome level (Faulkner et al. 2009). TEs situated upstream of 5’ untranslated regions (UTRs) tend to function as enhancers or alternate promoters (Bejerano et al. 2006, Franke et al. 2017). The promoters for L1 and LINE-like elements are internal to the transcription start site, which prevents promoter loss and provides potential alternative start sites for genes into which they transpose (Swergold 1990). In addition, a large fraction of Refseq genes (~25%) contain one or more TE insertions in their 3’ UTRs (Faulkner et al. 2009). Globally, these transcripts show decreased expression as TE coverage in their 3’ UTRs increases, suggesting that negative regulatory functions can often be ascribed to repeat sequences in this region. In support of this, SINE Alu elements contain a poly(T) sequence that contributes approximately 40% of the sequence present in known 3’ UTR AU-rich elements (ARE) (An et al. 2004). AREs are critical for mRNA stability and regulation through binding of proteins that facilitate or impair mRNA degradation (reviewed in Barreau et al. 2005). In addition to altering the regulatory regions of existing genes, TEs participate in the generation of retrogenes – host mRNAs that have undergone reverse transcription, with the resultant cDNAs incorporated into the genome as intron-less genes (Buzdin et al. 2003). Recent work by Hendrickson et al. (2017) demonstrates that the retrogene paralogs DUX/DUX4 are indispensable for embryo cleavage in mice and humans, highlighting a particularly relevant example of this process in embryos. In a fascinating twist on this observation, transcripts induced by the DUX retrogene product (a transcription factor) include the endogenous retroviruses M/HERVL, directly suggesting
an explanation for the persistence of this TE in the genome. Finally, expression of the maternally inherited factor Stella in oocytes is required for embryo progression past the 4-cell stage (Payer et al. 2003). Recently, Stella has been shown to induce the expression of ERVs during the maternal-to-zygotic transition (MZT), which appears to be critical for mouse embryo development (Huang et al. 2017), reinforcing the importance of regulated TE expression during development.

The advantages of TE expression

Increased awareness that TEs may confer fitness advantages to their hosts has contributed to a significant shift in the perception of their roles and importance. While previously considered parasitic ‘passengers’ in the genome that were occasionally responsible for disease, TEs are now more commonly considered ‘evolutionary drivers’ and ‘exaptation mediators’ that exist symbiotically in the genome (Elbarbary et al. 2016).

In addition to their roles in altering the landscape of regulatory regions in coding genes, a number of studies have demonstrated that TE expression is increased in response to stress. This association is most firmly established in plants, where tissue damage, pathogen exposure or heat stress increases TE expression (reviewed in Wessler 1996, Capy et al. 2000). In yeast, Ty retrotransposon expression is increased after ethanol exposure (Stanley et al. 2010). TE expression changes in response to stress are not as widely documented in mammals. Interestingly, higher numbers of Alu elements are present in the regulatory elements of genes involved in stress and immune responses (van de Lagemaat et al. 2003), suggesting that common regulatory elements may control their expression. Furthermore, SINE-mediated mRNA regulation has been observed in mammals under stressful conditions (reviewed in Elbarbary et al. 2016). In addition to stress, hormonal stimuli can drive TE activation, HERV expression in human breast cancer cells and LTR VL30 expression in mouse reproductive organs are increased in response to steroid hormones (Ono et al. 1987, Schiff et al. 1991). Based on these observations, it is reasonable to speculate that TE expression also depends on environmental stimuli in the context of early mammalian development. Regardless of where TEs transpose within the genome or the stimuli that cause them to do so, selection for the accumulated beneficial changes they induce has clearly contributed to their persistence.

TE mobilization during reprogramming

While emerging evidence suggests that regulated TE expression is associated with normal embryogenesis, unregulated, widespread expression and reintegration of mobile elements during reprogramming poses significant risks to the genome. Reprogramming refers to the erasure and re-writing of specific epigenetic marks, namely DNA methylation and histone modifications. In the context of reproduction, reprogramming occurs twice: once during gametogenesis and again during early embryogenesis (Morgan et al. 2005). The resultant chromatin state leads to TE de-repression (reviewed in Leung & Lorincz 2012). Human embryos express TEs from various classes including L1, Alu, SVA and HERV-K (Adjaye et al. 1997, Guo et al. 2014, Grow et al. 2015). HERV-K transcription is driven by OCT4 and produces functional proteins that are present in the blastocyst and capable of RNA binding, although their developmental roles are unknown (Grow et al. 2015).

Two published studies have investigated the expression of TEs during bovine reprogramming. In both in vitro-produced (IVP) and -cloned (SCNT) embryos, marked induction of three ERVs was observed in one study (Bui et al. 2009). A 1000-fold increase in ERV1_1_BT was seen between the four-cell and morula stages for both the IVP and SCNT groups. Independently, seven TE families were examined between the four-cell and blastocyst stages (Li et al. 2016). Low or highly variable expression was observed in four of the seven families, and L1, BovB and ERV1_1_BT showed increased expression from the four-cell to blastocyst stages, with ERV1_1_BT showing a similar increase to that seen in the Bui study. These findings highlight marked changes in TE expression in the bovine embryo during the MZT.

Some inference regarding TE activity can be obtained from evaluating the overall abundance of specific TEs in the genome; L1 BT and BovB comprise 11 and 10% of the bovine genome respectively (Adelson et al. 2009). Seventy-three potentially active copies of L1 BT and nine of BovB are present. The SINE element Bov-tA is substantially more abundant, with approximately 1.5 million copies. Based on these data, the L1 BT family of TEs is likely the most active, along with SINE elements that exploit the L1 ORF genes in their replication/reintegration cycles. In the endogenous retrovirus (ERV) category, 24 families and ~13,622 total ERV elements are present (Garcia-Etxebarria & Jugo 2010) suggesting substantial recent activity in evolutionary terms.

Studies on TEs during reprogramming in gametes have focused primarily on the mechanisms responsible for TE suppression, particularly in the mouse, where knockout of several different repressive pathways arrests meiosis through genomic instability in both sexes (Aravin et al. 2007a, Carmell et al. 2007, Kuramochi-Miyagawa et al. 2008, Flemr et al. 2013). In particular, widespread activation of the L1 and ERV families during spermatogenesis is observed when repressive mechanisms are inhibited (Aravin et al. 2007a, Flemr et al. 2013). Activation of TE transcription and its consequences during oogenesis is less extensively studied, although recent studies employing oocyte-
specific deletion of the histone methyltransferase Setd1b highlight the need for some level of epigenetic suppression of TE expression (Eymery et al. 2016). In one example of the requirement for TE activation, LINE1 ORF1 protein is required for exit from meiotic arrest in mouse oocytes, while overexpression inhibits progression to metaphase II (Luo et al. 2016). ORF1 is also expressed at variable levels in fetal oocytes, with overexpression leading to meiotic prophase I defects (Malki et al. 2014). Studies on the murine LTR TE mouse transcript (MT) expressed in neonatal to mature oocytes (Park et al. 2004, Holt et al. 2006) suggest that it accounts for 10% of total oocyte transcripts (Peaston et al. 2004) although no function is recognized. In support of its importance, RNAi-mediated inhibition of MT expression in GV oocytes inhibited germinal vesicle breakdown (GVBD), implicating MT in oocyte maturation (Peaston et al. 2004). Overall, TE regulation appears important to ensure that some level of TE expression occurs during the reprogramming stage of gametogenesis without excessive activation. The functions of TE expression in this context and the evolutionary forces that drive their selection are not known.

Control of TE expression during reprogramming

Given the delicate balance between the positive effects of limited TE expression during reprogramming, and the severe genomic damage that results from TE overexpression, it is perhaps not surprising that a number of distinct pathways have evolved to control their expression. Endogenous siRNAs appear to be particularly important in mouse oocytes, where suppression of Dicer or Ago2 increases the levels of MT transcripts and abnormalities in spindle formation during meiosis I (Watanabe et al. 2008, Stein et al. 2015). This pathway may be most important in the mouse and rat, which have acquired a unique oocyte-specific variant of Dicer (Dicer<sub>o</sub>) (Flemr et al. 2013). However, recent studies on L1 targeting endo-siRNAs in the pig suggest that Dicer and Ago2 are also important in other species (Zhang et al. 2017). Small RNA pathways in which targeting RNAs derived from tRNAs (known as tRFs) have recently been shown to suppress LTR TEs in mice (Schorn et al. 2017). In this pathway, 18 nt tRFs selectively antagonize RT action (an obligate step in replication for this TE class) while 22nt tRFs act as endo-siRNAs, participating in RISC-dependent destruction of TE transcripts. Additional pathways involving the proteins MARF1 (Su et al. 2012), Stella (Huang et al. 2017), RIF-1 (Li et al. 2017) and TRIM28 (Tao et al. 2018) are also important in TE repression and act through epigenetic and transcriptional modulation, but lie beyond the scope of this review. The most widely studied pathway involved in the control of TE expression is the PIWI/piRNA pathway.

PIWI proteins and piRNAs – permissive, targeted TE defense

A clade of Argonaute family members known as PIWI proteins and their associated small (25–32 nt) RNAs form the central components of a genome protection mechanism known as the PIWI/piRNA pathway (Aravin et al. 2006, Saito et al. 2006, Vagin et al. 2006, Brennecke et al. 2007). The pathway was originally described in Drosophila gametes where its role in limiting TE expression was first recognized (Lin & Spradling 1997, Aravin et al. 2001, Deng & Lin 2002, Vagin et al. 2006). The name is derived from the phenotype ‘P-element Induced Wtmpy testis’ (PIWI), where a mutation in the PIWI gene of Drosophila was induced by a Class I transposon (P-element) to which some strains are particularly susceptible (Lin & Spradling 1997). The importance in reproduction was immediately evident as both male and female progeny are sterile (Lin & Spradling 1997). The first identified mechanism of action involved the interaction of PIWI proteins with piRNAs of specific sequences complementary to RNA targets (i.e. expressed TEs) to form a piRNA-induced silencing complex (piRISC) that specifically cleaves the RNA target using RNase activity of the complex (Aravin et al. 2006, Girard et al. 2006, Gunawardane et al. 2007). Additional potential roles in targeting mRNA transcripts have recently been reported in germline and somatic cells (Rajasethupathy et al. 2012, Chen et al. 2013, Kwon et al. 2014, Nandi et al. 2016, Russell et al. 2017), although the extent and importance of this targeting is not yet clear. This review will focus primarily on the established roles of PIWI proteins and their associated piRNAs in the control of TE expression.

A systematic examination of mutations that impacted germline stem cell viability led to the identification of the first PIWI gene in Drosophila (Lin & Spradling 1997) where PIWI gene mutations result in cellular loss (Cox et al. 1998). The Drosophila PIWI proteins, Argonaute3 (Ago3) and Aubergine (Aub), were subsequently characterized based on their structural similarity and co-localization with PIWI in the pole plasm (Wilson et al. 1996, Harris & Macdonald 2001, Williams & Rubin 2002). Structurally all proteins of the PIWI family contain a PIWI domain, which has RNase activity, in addition to R-terminal, PIWI/Agoonaute/Zwille and MID domains (Cora et al. 2014). The importance of the PIWI pathway in mammalian reproduction was first demonstrated in the mouse, where the PIWI homolog Mlwi (PIWI-like 1 or PIWIL1) was shown to be essential for spermatogenesis (Deng & Lin 2002). Mouse PIWIL1 is expressed during spermatogenesis between the pachyten stage and the round spermatid stage where it silences TEs and specific mRNAs (Reuter et al. 2011, Gou et al. 2014). A similar expression pattern is observed in the bull (Russell et al. 2016) and dog (Stalker et al. 2016). The key mechanism of PIWI action was identified when it was shown to
bind a class of small RNAs that guide complementary interactions with target RNA transcripts including TEs (Aravin et al. 2006, Girard et al. 2006, Lau et al. 2006, Brennecke et al. 2007, Carmell et al. 2007).

Two additional murine PIWI homologs were identified in subsequent studies: MILI (PIWIL2) and MIWI2 (PIWIL4) (Aravin et al. 2006, Carmell et al. 2007). PIWIL2 is thought to be the oldest PIWI protein in evolutionary terms (Kuramochi-Miyagawa et al. 2008) and has the most extensive period of expression throughout spermatogenesis, starting at the primordial germ cell stage in utero and persisting to the pachytene stage (Kuramochi-Miyagawa et al. 2004, Aravin et al. 2006). Mice lacking PIWIL2 develop genomic instability and elevated TE expression after which spermatogenesis becomes blocked in prophase of meiosis I (Kuramochi-Miyagawa et al. 2004, Reuter et al. 2009). PIWIL2 is localized to the cytoplasm, primarily in perinuclear granules, a known site for RNA processing (Aravin et al. 2008). PIWIL4 is found in the pre-natal testis and is expressed from 15.5 days post-coitum until 3 days after birth in prospermatogonia (Aravin et al. 2008) coinciding with an important period of de novo DNA methylation (Lees-Murdock et al. 2003, Maatouk et al. 2006). PIWIL4 distribution is primarily nuclear, with limited expression in cytoplasmic granules (Aravin et al. 2008). One potentially important but less thoroughly studied member of the PIWI clade is PIWIL3, which is encoded in human, bovine (Roovers et al. 2015) and other genomes, but absent in the mouse and rat where most functional studies have been performed. Recent evidence suggests that PIWIL3 expression is normally limited to oocytes and embryos (Roovers et al. 2015, Virant-Klun et al. 2016, S J Russell and J LaMarre, unpublished observations). The roles and importance of this protein are largely unknown.

**Regulation of piRNA precursor expression from genomic ‘clusters’**

Most mature piRNAs are generated from longer RNA transcripts originating in specific piRNA-rich genomic regions called piRNA clusters (Brennecke et al. 2007). piRNA clusters consist mostly of repetitive element and transposon fragments that have integrated into the genome throughout the course of evolution and which reflect the history of transposition and virus integration into the genome of a species and its evolutionary ancestors (Aravin et al. 2007a). Several hundred to a few thousand clusters are present in the mammalian genomes examined (Lau et al. 2006, Aravin et al. 2007a, Brennecke et al. 2007, Gebert et al. 2015, Russell et al. 2017). The factors that regulate transcription from these clusters (enhancers, epigenetic status) remain poorly characterized. Transcription of piRNA precursor RNAs can occur from a single, or from both strands of DNA, and are respectively classified as uni- or dual-strand clusters (Aravin et al. 2006, Lau et al. 2006, Brennecke et al. 2007, Carmell et al. 2007). Uni- or bi-directional transcription from uni-strand clusters is mediated by RNA polymerase II (Pol II) from one transcription start site, precursors are then polyadenylated, 5’ capped and sometimes subjected to alternative splicing (Goriaux et al. 2014). Expression from dual-strand clusters may be due in part to overlapping, run-through transcription from nearby genes (Mohn et al. 2014). However, recent studies in flies have identified a novel heterochromatin-dependent transcriptional process through which the basal transcription machinery containing a variant of transcription factor II A named ‘Moonshiner’ associates with another protein known as Rhino to license transcription from heterochromatin at piRNA clusters (Andersen et al. 2017).

Two distinct ‘waves’ of piRNA expression occur in murine testes (Vourekas et al. 2012) and the piRNAs from these waves are named based on the timing of their expression. Pre-pachytene piRNAs are present during early spermatogenesis and are primarily targeted at TEs and associate with PIWIL2 and PIWIL4 (Aravin et al. 2008, Li et al. 2013). Pachytene piRNAs are abundantly expressed in pachytene spermatocytes and round spermatids in adult mouse testes and physically associate with PIWIL1 and PIWIL2 (Aravin et al. 2006, Girard et al. 2006). Transcription of pachytene piRNA precursors is driven by the transcription factor A-MYB, which also facilitates PIWIL1 transcription, increasing processing ‘capacity’ for the cluster-derived precursor piRNAs induced at that time (Li et al. 2013). Interestingly, in addition to TEs, pachytene piRNAs may also target mRNAs and have been strongly implicated in the widespread active decay of mRNAs that occurs late in spermatogenesis (Gou et al. 2014). PIWI pathway targeting of mRNAs may not be limited to testes, as piRNAs that appear to target mRNAs have also been recognized in bovine oocytes (Russell et al. 2017).

**Export and processing – primary piRNA biogenesis**

The generation of piRNAs from primary cluster transcripts is termed primary piRNA biogenesis (Aravin et al. 2007b). After transcription, piRNA transcripts move from the nucleus to the nuage, a peri-nuclear ‘cloud’ containing structures also known as germ granules, Yb bodies, chromatoid bodies, pi-bodies and piP-bodies depending on the cell type and content (Buchan & Parker 2009, Ishizu et al. 2012, Pillai & Chuma 2012, Meikar et al. 2014). The nuage contains mRNA processing machinery and is a known site of decay, storage and quality assessment for RNA (Buchan & Parker 2009, Kulkarni et al. 2010, Meikar et al. 2014). Transcription of piRNA precursors is...
coupled to the processing machinery in the nuage by the DEAD box helicase UAP56 (Shen 2009, Zhang et al. 2012). Precursors to piRNAs remain single stranded (in contrast to siRNAs and miRNAs) or assume unrecognized structures prior to processing (Houwing et al. 2007). Nucleo-cytoplasmic shuttling is facilitated by the protein Maelstrom (Mael) in Drosophila and mice, which interestingly plays different roles in each species (Aravin et al. 2009, Sienski et al. 2012). In Drosophila, Mael participates in nuclear silencing of transposons in conjunction with PIWI but is not required for piRNA biogenesis (Sienski et al. 2012). In contrast, Mael is required for piRNA biogenesis during pachytene spermatogenesis in mice and binds piRNA precursors that associate with the nuage. Inactivation of Mael results in meiotic arrest and the dissociation of perinuclear granules containing PIWIL4, leading to failure of LINE1 repression (Aravin et al. 2009). These findings implicate Maelstrom in both pre-pachytene PIWIL4 functions and pachytene piRNA biogenesis.

Exported piRNA precursors are processed in several subsequent steps in the nuage, all of which appear closely associated with the mitochondria. An endonuclease known as Zucchini (Zuc; mZuc or mitoPLD in the mouse) processes precursors into intermediate fragments (Pane et al. 2007, Watanabe et al. 2011a, Ipsaro et al. 2012, Nishimasu et al. 2012) after the RNA helicase Armitage (Armi; MOV10L1 in the mouse), unwinds RNA secondary structure (Frost et al. 2010, Vourekas et al. 2015). Physical loading of piRNA intermediates onto PIWI proteins in the nuage requires the co-chaperones shutdown (Shu) and heat shock protein 90 (HSP90) (Olivieri et al. 2012, Izumi et al. 2013).

The next step in piRNA biogenesis is 3′ trimming, which occurs after intermediates are loaded onto PIWI proteins. Mature piRNA length (extent of trimming) appears to depend on the specific PIWI protein to which it is bound – the modal size of mature piRNAs bound to PIWIL1, PIWIL2 and PIWIL4 are 29–30, 26–27, 27–28 nt respectively (Kuramochi-Miyagawa et al. 2008, Vourekas et al. 2012). An enzyme known as Nibbler is a mut-7 homolog in Drosophila which likely mediates the 3′–5′ exonuclease activity (Han et al. 2011, Kawaoka et al. 2011, Wang et al. 2016) for piRNAs bound to PIWI proteins, however the mammalian homolog is unidentified. Nibbler activity depends on TDRKH (also called Papi), which is present at the mitochondrial membrane (Chen et al. 2009, Honda et al. 2013, Saxe et al. 2013). In the final processing step, fully cleaved piRNAs, when bound to PIWI proteins, become 3′-methylated through the activity of Hen1/Pimet, which produces a 2′-O-methyl modification (Horwich et al. 2007, Saito et al. 2007). At the completion of biogenesis, the mature piRNA–PIWI protein complex is considered a ‘piRNA-induced silencing complex’ (piRISC).

The ping-pong amplification loop – generation of secondary piRNAs

One of the most unique elements of the piRNA pathway is a secondary ‘feed-forward’ biogenesis process known as the ‘Ping-Pong’ amplification loop. This loop is considered the secondary piRNA biogenesis pathway (Aravin et al. 2008). In Drosophila, piRNAs generated through primary biogenesis and bound to Aubergine target TEs in the cytoplasm leading to cleavage of the target RNA through RNase slicer activity (Brennecke et al. 2007, Gunawardane et al. 2007). Rather than becoming degraded with the target, the resulting cleaved RNA products load directly onto another PIWI protein, Ago3, as intermediate piRNAs that then become trimmed and 2′-O-Me modified as described above (Han et al. 2015, Wang et al. 2016). These secondary Ago3 piRISC complexes resulting from this ‘ping’ step may then reciprocally target primary piRNA precursors (‘pong’ step) and amplify the number of targetting piRNAs available for silencing through the PIWI pathway (Brennecke et al. 2007). The reciprocal cleavage events, combined with the positional nature of cleavage through this pathway result in a ten nucleotide overlap in piRNA populations that have undergone ping-pong amplification (Aravin et al. 2008). This ping-pong ‘signature’ of overlapping complementary nucleotides is visible as a defined peak when plotted (for examples see Russell et al. 2017).

One important consequence of secondary piRNA generation through the ping-pong cycle is higher sequence diversity in piRISC complexes. Cleavage products from Ago3 in Drosophila when loaded onto PIWI, can direct transcriptional gene silencing (TGS) (Fig. 2; Senti et al. 2015). Similarly, mouse PIWIL2-cleaved targets can associate with PIWIL4, which also direct TGS in the nucleus after translocation (De Fazio et al. 2011). Primary and secondary piRNA biogenesis require a number of cofactors. Anchoring PIWI proteins to the scaffold in the nuage are a family of Tudor domain-containing proteins (TDRDs), TDRD9, TDRD1 and TDRDKH, that bind to symmetric dimethylarginines (sDMAs) on the PIWI proteins (Reuter et al. 2009, Shoji et al. 2009, Saxe et al. 2013). In the Drosophila nuage, the proteins Spindle-E, Krimper, Tejas, Tapas and Qin (Lim & Kai 2007, Malone et al. 2009, Patel & Kai 2010, Zhang et al. 2011) are involved as reviewed extensively elsewhere (Czech & Hannon 2016). In the mouse, a gamete-specific protein known as GTSF1 (Yoshimura et al. 2018) that was previously implicated in other aspects of PIWI/piRNA function (Dönert et al. 2013, Ohtani et al. 2013) has very recently been shown to be required for secondary piRNA generation. The protein Vasa appears to play an important part in piRNA biogenesis in the silkworm model of the PIWI pathway. Here, piRNA intermediates generated by PIWIL1 bind to Vasa and are protected from degradation, facilitating
their transfer to Ago3 (Xiol et al. 2014). Similarly, in mice, the Mouse Vasa Homolog (MVH) participates in retrotransposon suppression during fetal male and female gametogenesis; targeted MVH knockout generates a phenotype remarkably similar to PIWIL2 or PIWIL4 (Kuramochi-Miyagawa et al. 2010, Lim et al. 2013) supporting a conserved role across species.

An elegant series of recent studies (Han et al. 2015, Mohn et al. 2015) have identified mechanisms by which a subset of piRNAs are ‘phased’ during the biogenesis process, further increasing the variability of targeting piRNAs generated by the pathway. Phased piRNA generation proceeds when PIWI proteins (PIWI in the fly and PIWIL2 in the mouse) are recruited to the 5’ end of precursor transcripts and deﬁne a cleavage position 3’ to the complex. The endonuclease Zucchini (MitoPLD in the mouse) is then recruited and cleaves the precursor transcript at this 3’ position, deﬁning the 3’ end of the piRNA and leaving a new 5’ terminus on the precursor, upon which the process repeats. In the mouse the piRNAs generated are further processed by Nibbler in association with TDRKH (Hayashi et al. 2016). The multiple pathways described highlight the complexity of biogenesis and suggest that additional factors, which participate in piRNA biogenesis and TE silencing are likely to be identiﬁed as the pathway continues to be characterized.

Mechanisms of piRNA suppression of TEs

Through the generation of piRISC complexes, the PIWI pathway suppresses TE and gene expression through several distinct mechanisms. Complementary RNA targets such as retrotransposon-encoded RNAs are cleaved by RNase slicer activity (Fig. 1A) (reviewed in Hirakata & Siomi 2016). Although less extensively studied, piRISC complexes also help induce epigenetic changes in chromatin conformation and DNA methylation patterns at speciﬁc loci in ﬂies (Fig. 1B; Sienski et al. 2012) and mice (Kuramochi-Miyagawa et al. 2008). Cleavage of TE RNA transcripts by piRNA-guided slicer activity of piRISC complexes is the best-studied and ﬁrst-recognized function of the PIWI pathway (Reuter et al. 2011, Zhang et al. 2015). This post-transcriptional silencing pathway mediated by PIWI proteins participates in both TEs and mRNA cleavage. Targeting and cleavage by piRISC complexes appears more stringent than for some other RISC complexes, requiring a minimum of 16–22 continuous bases for efﬁcient target decay (Zhang et al. 2015, Yuan et al. 2016). In spite of this, the canonical post-transcriptional silencing by PIWI proteins has many similarities with other Argonaute-dependent RISC pathways that have been reviewed extensively elsewhere (Wilczynska & Bushell 2015). In addition to mediating classical RNA cleavage events, some studies have suggested that murine PIWIL1 can also direct target decay through recruitment of the deadenylase CAF1 during spermiogenesis (Fig. 2; Gou et al. 2014). For targeting through this mechanism, the requirement for piRISC-target complementarity is not strict and mismatches similar to those observed in miRNA-target interactions are possible (Gou et al. 2014). Follow-up studies will be required to determine the importance and scope of this pathway in cells where PIWIL1 is expressed.

Although it has been most widely studied in the context of post-transcriptional destruction of RNAs originating from genomic TE sequences, the PIWI pathway has also been implicated in TGS through directed DNA methylation and histone modiﬁcations (Fig. 1B; Carmell et al. 2007, Klenov et al. 2014). In studies with Drosophila, Piwi localization in both the nucleus and cytoplasm was demonstrated and Piwi depletion changed the pattern of chromatin silencing of speciﬁc TEs (Klenov et al. 2011). Subsequently, Piwi binding to nascent RNAs transcribed from target loci was shown to result in the deposition of repressive H3K9me3 marks on nearby histones (Le-Thomas et al. 2013, Post et al. 2014). Further epigenetic inﬂuence was observed with a mutated, slicer-incompetent Piwi (incapable of endonucleolytic cleavage) which retained the ability to transcriptionally silence TE expression through the prevention of RNA Pol II occupancy at speciﬁc loci in conjunction with the cofactor Mael (Klenov et al. 2011, Sienski et al. 2012). Mammalian PIWIL4 has similar roles in transcriptional repression, but mediates changes in DNA methylation rather than altering histones (Aravin et al. 2008). One early hypothesis in which PIWIL2 was thought to generate piRNAs that subsequently
bind to PIWIL4 and direct targeted methylation has recently been challenged by reports showing substantial differences in TE repression in mouse testes by PIWIL2 and PIWIL4 (Manakov et al. 2015). While abundant evidence suggests that the PIWI pathway participates in both post-transcriptional and transcriptional TE silencing, it is important to note that the relative importance of the two processes is likely to vary with the stage of gamete differentiation or embryo development during which it is expressed, as demonstrated very recently during spermatogenesis (Inoue et al. 2017). The mechanisms and consequences of epigenetic changes directed by the PIWI pathway remain very active areas of investigation and are likely to yield fascinating insight into reprogramming and development.

Our understanding of the PIWI pathway has been most strongly and consistently informed by the many studies of its different roles in TE suppression. However, any review of this pathway would be incomplete without briefly mentioning several emerging roles for the pathway in the control of gene expression during gamete and embryo development. One of the first described examples of this phenomenon is presented in studies demonstrating PIWI-dependent, piRNA-directed DNA methylation of the paternally imprinted Rasgrf1 locus in mice (Watanabe et al. 2011b). Through targeting TE sequences within a differentially methylated region adjacent to this locus, piRNAs direct methylation and silencing. Importantly however, piRNAs do not direct all imprinting events, many of which are unaffected after PIWI pathway suppression (Watanabe et al. 2011b). While piRNA-directed imprinting processes appear to require TE sequences, several recent studies in gametocytes suggest that mRNAs can also act as direct PIWI pathway targets, particularly during spermatogenesis (Gou et al. 2014) and, more recently, in oocytes (Russell et al. 2017). The mechanisms through which such targeting has evolved and its overall importance in fertility and gamete development remain active fields of study.

Conclusions

The structure and function of non-coding regions of the genome, much of which consists of TE sequences, has only recently become the focus of intensive study. Normally, TEs are epigenetically silenced, but it has become apparent that their limited activation during reprogramming is an essential feature of gametogenesis and development, contributing directly to these processes through mechanisms that remain largely uncharacterized. The evolutionary advantages selecting for some level of TE expression may lie in the expanded regulatory potential they confer after reintegration within regulatory domains, a process that has contributed sequences to more than one quarter of human coding genes (Faulkner et al. 2009) and driven adaptive change. However, excessive TE activation is clearly detrimental and leads to infertility and developmental problems. These combined observations suggest that, rather than completely eliminating TE expression, control systems have evolved to promote a delicate balance between permission and suppression of TE activation during reprogramming. The characteristics of the PIWI pathway are consistent with this – TEs are suppressed only after the piRNAs that target them are transcribed and processed. A secondary amplification loop that depends on the presence of TE transcripts enhances suppression once initiated. Phasing and trimming of piRNAs during the biogenesis process expands the repertoire of targets and helps specify the nature of target suppression (post-transcriptional vs transcriptional). Evolutionary expansion of the PIWI pathway to target coding sequences, possibly as a consequence of TE integration into coding and transcriptional regulatory regions of the genome, have expanded functionality in ways we are only just beginning to understand. Future studies on the PIWI pathway will need to resolve a number of key issues: the identity of many cofactors in mammalian species, the factors that initiate piRNA biogenesis, particularly in somatic cells and the mechanisms by which they do so are particularly important questions. With respect to reproduction in particular, the overall importance of the PIWI pathway in mammalian oogenesis and embryogenesis remain unknown. However, regardless of the specific importance of the PIWI pathway itself in any one species, the ongoing tension between ancient virus sequences that hide within host DNA, and the defense pathways that keep them in check, continues to shape the genomic landscape and profoundly influence the control of gene expression that guides gametogenesis and the earliest phases of embryonic development.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Acknowledgements

The authors wish to thank Clifford Librach, Leanne Stalker, Allison Tscherner, Graham Gilchrist, Pavneesh Madan and Allan King for helpful discussions. Studies on the PIWI pathway in gametes and embryos in the LaMarre laboratory have been supported by NSERC (Canada), the Ontario Ministry of Agriculture Food and Rural Affairs and by a Michelson Grant in Reproductive Biology from the Michelson Found Animals Foundation.


Frost RJ, Hamra FK, Richardson JA, Qi X, Bassel-Duby R & Olson EN 2010 MOVTOL1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. PNAS 107 11847–11852. (https://doi.org/10.1073/pnas.1007158107)


Ge SS 2017 Exploratory bioinformatics investigation reveals importance of ‘junk’ DNA in early embryo development. BMC Genomics 18 200. (https://doi.org/10.1186/s12864-017-3566-0)

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. (https://doi.org/10.1371/journal.pone.0124860)


Ge SS 2017 Exploratory bioinformatics investigation reveals importance of ‘junk’ DNA in early embryo development. BMC Genomics 18 200. (https://doi.org/10.1186/s12864-017-3566-0)

Hirakata S & Siomi MC 2017 Switching of dominant retrotransposon silencing strategies from posttranscriptional to transcriptional mechanisms during male germ-cell development in mice. PLoS Genetics 13 e1006926. (https://doi.org/10.1371/journal.pgen.1006926)


Jachowicz JW, Bing X, Pontabry J, Boškovic A, Rando OJ & Torres-Padilla M-E 2017 LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nature Genetics 49 1502–1510. (https://doi.org/10.1038/ng.945)


Knoutter EN, Belancio VP, Wagstaff BJ & Roy-Engel AM 2009 The RNA polymerase dictates ORF requirement and timing of LINE and SINE retrotransposition. PLoS Genetics 5 e1000458. (https://doi.org/10.1371/journal.pgen.1000458)


Transposons and the PIWI pathway


Saxe JP, Chen M, Zhao H & Lin H 2013 Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. EMBO Journal 32 1869–1885. (https://doi.org/10.1038/emboj.2013.121)


Watanabe T, Tomizawa S-I, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A et al. 2011b Role for piRNAs and noncoding RNA in de novo DNA methylation of the


Received 27 April 2018
First decision 18 June 2018
Revised manuscript received 16 July 2018
Accepted 19 July 2018