Vertebrate lineages exhibit diverse patterns of transposable element regulation and expression across tissues

Giulia I.M. Pasquesi¹,², Blair W. Perry¹, Mike W. Vandewege³, Robert P. Ruggiero⁴, Drew R. Schield¹,⁵, and Todd A. Castoe¹‡

¹Department of Biology, 501 S. Nedderman Dr, University of Texas at Arlington, Arlington, TX 76019 USA
²Department of Molecular, Cellular, and Developmental Biology, 3415 Colorado Ave, University of Colorado, Boulder, CO 80303
³Department of Biology, ENMU Station 33, 1500 S Ave K, Eastern New Mexico University, Portales, NM 88130 USA
⁴Department of Biology, One University Plaza, MS 6200, Southeast Missouri State University, Cape Girardeau, MO 63701 USA
⁵Department of Ecology and Evolutionary Biology, 1900 Pleasant St, University of Colorado, Boulder, CO 80309

‡Corresponding Author: Todd A. Castoe, Department of Biology, University of Texas at Arlington, Arlington, TX 76010 USA
Email: todd.castoe@uta.edu phone: 817-272-9084 fax: 817-272-9615

Running Head: transposable element regulation and expression in vertebrate tissues

Author Contributions: GIMP and TAC conceived the project; GIMP, BWP, DRS, and TAC collected the data; GIMP, MWV, BWP, and DRS analysed the data; GIMP, MWV, RPR and TAC wrote the manuscript; all authors reviewed the manuscript and approved its submission.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Data deposition: New unassembled RNA sequencing data has been deposited at NCBI under the following accessions: PRJNA608231. The authors declare that all data and scripts used in this study are available via public databases or available from the corresponding author upon request.

Abstract

Transposable elements (TEs) comprise a major fraction of vertebrate genomes, yet little is known about their expression and regulation across tissues, and how this varies across major vertebrate lineages. We present the first comparative analysis integrating TE expression and TE regulatory pathway activity in somatic and gametic tissues for a diverse set of 12 vertebrates. We conduct simultaneous gene and TE expression analyses to characterize patterns of TE expression and TE regulation across vertebrates, and examine relationships between these features. We find remarkable variation in the expression of genes involved in TE negative regulation across tissues and species, yet consistently high expression in germline tissues, particularly in testes. Most vertebrates show comparably high levels of TE regulatory pathway activity across gonadal tissues except for mammals, where reduced activity of TE regulatory pathways in ovarian tissues may be the result of lower relative germ cell densities. We also find that all vertebrate lineages examined exhibit remarkably high levels of TE-derived transcripts in somatic and gametic tissues, with recently-active TE families showing higher expression in gametic tissues.

http://mc.manuscriptcentral.com/gbe
Although most TE-derived transcripts originate from inactive ancient TE families (and are likely incapable of transposition), such high levels of TE-derived RNA in the cytoplasm may have secondary, unappreciated biological relevance.

**Keywords**: somatic transposable element expression, transposable element cellular derived transcripts, transposable element silencing, PIWI, ovary, testis
Introduction

Transposable elements (TEs) represent the largest identifiable fraction of vertebrate genomes (Chalopin, et al. 2015; Smit, et al. 2013-2015) despite the fact that they are fundamentally mutagens that propagate through the insertion of new copies. Though ubiquitous, the composition and abundance of TEs is highly variable across vertebrate genomes (Chalopin, et al. 2015; Kapusta, et al. 2017; Pasquesi, et al. 2018; Platt, et al. 2018). This variability is the result of complex processes acting at both the levels of TEs and the host genome, including population demography (Lynch and Conery 2003; Neafsey, et al. 2004; Xue, et al. 2018), the evolutionary history of TEs that have infected host genomes (Gilbert, et al. 2012; Kordis and Gubensek 1998; Pasquesi, et al. 2018), and the ability of the host to repress TE mobilization (Aravin, et al. 2008; Ozata, et al. 2019). TE insertions may negatively impact the fitness of their host (Boissinot, et al. 2006; Lynch and Walsh 2007) and have been shown to disrupt open reading frames and regulatory regions, alter chromosome structure, and promote genomic rearrangements (Beck, et al. 2011; Burns 2017; Callinan and Batzer 2006; Gasior, et al. 2006; Sen, et al. 2006; Vogt, et al. 2014). Yet, increasing evidence for the roles of TEs in rewiring regulatory networks and driving evolutionary innovation (Agrawal, et al. 1998; Bourque, et al. 2008; Chuong, et al. 2016; Lynch, et al. 2015; Makałowski, et al. 2017; Zeng, et al. 2018).
2018) counterbalances a simplistic view that TEs are exclusively associated with deleterious impacts on host genomes.

Host genomes have evolved multiple mechanisms to negatively regulate TE activity (i.e., reviewed in Goodier 2016), with the primary mechanism being epigenetic modification to silence TE-containing chromatin (Jacobs, et al. 2014; Reik 2007; Sanchez-Luque, et al. 2019; Slotkin and Martienssen 2007). Gonadal germ cell development, however, requires genome-wide erasure of methylation patterns in primordial germ cells to establish cell potency (Surani, et al. 2007). This leaves transposons temporarily unsuppressed by chromatin silencing and thus capable of generating heritable insertions until chromatin structure is reestablished (Hajkova, et al. 2002; Kato, et al. 2007; Molaro, et al. 2014). Safeguarding of the genome against this TE propagation in the germline is primarily accomplished by the PIWI:piRNA (PIWI interacting RNAs) pathway (Aravin and Tuschl 2005; Lim and Kai 2015), a specific small RNA interference mechanism that limits TE proliferation at both the transcriptional level through de novo methylation of TE loci and the post-transcriptional level by targeting and degrading TE transcripts (Aravin, et al. 2007; Siomi, et al. 2011; Weick and Miska 2014).

Previous studies of TE expression and regulation have primarily focused on analyses of critical tissues or temporal windows for novel TE insertions, including testes (i.e., Handel and Schimenti 2010;
Shi, et al. 2007) and early embryonic-stage tissues (Feusier, et al. 2019; He, et al. 2019; Richardson, et al. 2017). Fewer studies have examined the extent of somatic TE activity (Faulkner and Billon 2018; Faulkner, et al. 2009; Garcia-Perez, et al. 2016; Loreto and Pereira 2017; Soumillon, et al. 2013), although there is evidence for biologically-relevant levels of TE transposition in certain somatic tissues, such as the brain, and for elevated levels of TE activation in somatic tissues associated with ageing or disease (Anwar, et al. 2017; Bedrosian, et al. 2016; Callinan and Batzer 2006; De Cecco, et al. 2013; Faulkner and Garcia-Perez 2017; Kreiling, et al. 2017).

Currently, our understanding of variation in TE expression and TE regulation across somatic and gametic tissues is based primarily on studies of mammal and bird species (Soumillon, et al. 2013), and remarkably little is known about how TE expression and TE regulation may vary across the vertebrate tree of life.

Here, we examine patterns of TE expression and regulation in somatic and gametic tissues from 12 species that represent a sampling of all major vertebrate lineages (Supplementary file S1). We leverage this sampling to (i) quantify the effects of conserved TE regulatory mechanisms on TE expression levels within and across vertebrate lineages; and (ii) evaluate whether non-mammalian vertebrate species follow mammalian patterns of TE regulation and expression. Our integrated analyses provide new evidence for the uniqueness of
mammalian germline biology compared to that of other vertebrates, highlight many features of TE regulation shared across vertebrate lineages, and raise new questions about the biological significance of broad expression of TE-derived transcripts in somatic and gametic tissues that appears to be ubiquitous across vertebrates.

**Materials and methods**

We used previously published poly-A-selected RNAseq datasets to compare expression levels of TE-derived transcripts and genes involved in the negative regulation of TEs in testes, ovaries, and up to seven somatic tissues (brain, heart, kidney, liver, muscle, spleen and small intestine) across 12 vertebrate species that included representatives of fish, amphibians, reptiles, and mammals (Supplementary file S1). Additionally, we included available purified oocyte cell datasets for five species (Supplementary file S1) for comparison to ovary and testis whole-tissue RNAseq. Raw RNAseq data were first filtered for prokaryote and eukaryote rRNA transcripts using SortMeRNA v2.1 (Kopylova, et al. 2012), and then quality and adapter trimmed in Trimmomatic 0.36 (Bolger, et al. 2014). Detailed information for each analysis is provided in the supplemental methods. For each species, reads were mapped using STAR v2.7.0f (Dobin, et al. 2013) to the latest genome version and annotation .gff files available on the NCBI Genome database (Sayers, et al. 2019). All genomes used in this study are associated with high-quality
repeat element annotations that incorporated intensive species-specific repeat identification (see Supplementary Methods). STAR was run using default parameters, discarding chimeric transcripts, and allowing for a maximum of 100 mapped reads per locus (as suggested in Jin, et al. 2015).

Gene and TE-derived transcript expression levels were simultaneously estimated using TEtranscript (Jin, et al. 2015). To assign mapped reads to a genomic locus, TEtranscript requires two annotation files that specify gene and repeat element coordinates, respectively. TE index structures were built from RepeatMasker .out files (Smit, et al. 2013-2015), and gene index structures were built from the same gene annotation files used when running STAR (detailed information on the protocol used to build the input .gtf files are provided in the supplemental methods). TEtranscript was run using default parameters, the -multi multi-mapper mode flag, and specifying whether transcriptome data was stranded or not. Expression levels of TE-derived reads that originated from recently active TE copies were estimated in a second, separate analysis (we refer to this as the “recent-TE” dataset). This second analysis was required to effectively survey recent-TEs because TEtranscript analyses do not retain locus coordinates, which prevented us from being able to subsample recent-TEs directly from the primary inclusive analysis. For this analysis, we provided TEtranscript with a filtered .gtf
annotation file that contains only TE loci with less than 2% Kimura 2-parameter distance from the consensus. For each species, normalization of TE-derived and gene-derived raw read counts across tissues was performed using the \texttt{estimateSizeFactors-estimateDispersions-counts(normalized=TRUE)} functions in \texttt{DESeq2} v1.20 (Love, et al. 2014) after removing elements with less than 10 mapped reads across all samples. We normalized and performed statistical analyses using both the total-TE and the recent-TE dataset, and compared normalized gene expression values and results between the two (Supplementary file S2). Normalized expression values displayed only minimal variation between total-TE and recent-TE datasets, and statistical analysis results were unaffected, thus we only report results based on normalized expression values associated with the recent-TE dataset.

To assess the relationships between TE expression levels and TE regulatory pathway gene levels, we compared recent-TE expression levels to 5 sets of TE regulatory genes: (i) genes participating in the PIWI:piRNA pathway (Carbon, et al. 2009; PIWI pathway hereafter); (ii) genes involved in the small RNA interference pathway (Carbon, et al. 2009; siRNA pathway); (iii) genes involved in transcriptional regulation of TEs (e.g., responsible for \textit{de novo} DNA or histone methylation; Hutchins and Pei 2015; Wylie, et al. 2016); (iv) other genes previously identified to negatively impact TE mobilization.
and/or insertion at the post-transcriptional level (e.g., Apobec; Goodier 2016); and (v) the combined magnitude all TE regulatory genes (which corresponds to all 79 conserved genes belonging the four previous sets; Supplementary file S3).

Patterns of within-species variation in expression levels across tissues were assessed by performing principal component analyses (PCAs) on blind variance stabilizing transformed data (Anders and Huber 2010). Because of the heterogeneous nature of our data, between species comparisons were performed using percentages of the transcriptome following normalization of read counts to limit biases due to different methods of tissue processing, library preparation, sequencing technology and dataset quality (Dunn, et al. 2018; Sudmant, et al. 2015). To calculate percentages of TE regulatory gene expression, only normalized expression counts derived from annotated genes were used to calculate the total (i.e., TE-derived transcripts were excluded). To investigate relationships in expression patterns across vertebrates, we performed phylogenetic independent contrast (PIC) linear regressions, Spearman rank correlation analyses and PCAs using the phytools package in R (Revell 2012). Additional methodological details for statistical analyses performed in this study are provided in the supplemental methods.
Results

TE regulatory mechanisms are active in somatic and gametic tissues across vertebrate lineages

Our analysis of gene expression for a combined set of 79 genes involved in TE negative regulatory mechanisms (Supplementary file S3) demonstrates substantial variation in expression across tissues and species. We find that all categories of negative regulators (i.e., repressors of TE activity) are expressed in both somatic and germline tissues at widely varying levels, with roughly 2.5 times higher average expression in the germline (Supplementary file S2, Figure 1A and Supplementary fig. S1). Of all regulatory pathways, the PIWI:piRNA pathway shows higher expression in the germline compared to both somatic tissues (average 16.51-fold higher) and other regulatory gene sets in the germline (1.67-fold higher; Supplementary file S2). In contrast, genes involved in the siRNA pathway have consistently low expression in somatic and germline tissues, while genes involved in transcriptional and post-transcriptional regulation of TE activity show wide variation in expression across species and tissues (Figure 1A; Supplementary fig. S1). We also find that negative transcriptional regulators of TE expression on average are expressed at levels similar to the PIWI pathway in the germline, with two-fold higher expression than in somatic tissues (Supplementary file S2). This finding is consistent with elevated levels of
chromatin modification and the deposition of histone and DNA methylation markers in germline tissues (Greenberg and Bourc'his 2019; Stewart, et al. 2016).

Patterns of TE regulatory mechanism activation across tissues and vertebrate lineages

To assess variation in expression patterns of TE regulatory pathways among tissues and across lineages, we used multivariate clustering methods to summarize and differentiate trends of expression. Within-species principal component analyses (PCAs) on gene expression of PIWI pathway genes show distinct, individual clustering of germline tissues in non-mammal species, such that expression patterns in testes and ovaries are distinct from each other and from somatic tissues. In contrast, only testes show a distinct profile in mammals, while PIWI pathway levels in mammalian ovarian tissues fall within the variance of somatic tissues (Figure 1B left panel; Supplementary fig. S2A). No clear tissue clustering patterns are observed in pathway-specific analyses of the siRNA, transcriptional, and post-transcriptional regulatory pathways (Supplementary fig. S2B-D), except for a consistent trend of tissue separation driven by the ovary among non-mammal species. Broadly, these other regulatory pathways show cross-tissue profiles similar to those of the PIWI pathway, but with greater variance among somatic tissues (Figure 1B right panel). We further measured the contribution of each gene to
the principal component determination, and find that the five genes with the highest contribution scores all belong to the PIWI pathway for the majority of species (Supplementary fig. S3).

To understand how vertebrate lineages differ on the basis of how they regulate TEs in the germline, we directly compared variation in expression levels of TE regulatory pathways between species in germline tissues, specifically. Phylogenetically-correct PCAs for the set of PIWI pathway genes, genes from the three other regulatory mechanisms (i.e., ‘other pathways’), and all mechanisms combined demonstrate distinct TE regulatory mechanism expression patterns in the mammal species analyzed compared to non-mammalian species, largely driven by variation in TE regulatory activity in the ovaries (Supplementary fig. S4). Comparisons of the first principal component between the PIWI pathway and ‘other pathways’ distinguish testes gene expression patterns in the alligator and snake species from all other vertebrates (Figure 1C above). In contrast, we find that ovary expression patterns in human and mouse cluster independently from other vertebrate species, with the distinction being driven mostly by variation in expression of genes in the PIWI pathway (Figure 1C below).

Between-lineage variation in gametic tissue expression of TE regulatory pathway genes

To further characterize variation in TE regulatory activity across
lineages, we calculated Z-scores of expression relative to the mean expression of all genes for TE regulatory genes with orthologs identified in at least 8 of 12 species (Figure 2). Hierarchical clustering of Z-scores across tissues identified five distinct clusters: vertebrate testes, non-mammal ovaries, vertebrate brain, mammalian ovaries, and a mixed cluster of somatic tissues from diverse lineages (Figure 2). In contrast to the single testes germline cluster, we find two clusters of TE regulatory expression profiles from vertebrate ovaries. The first cluster includes all non-mammal species, in which expression profiles resemble TE regulation profiles in the testes. The other cluster is comprised solely of mammals, in which expression levels in the ovary are similar to those observed in somatic tissues. The only exception to this pattern is that the human ovary profile clustered with brain. Differences in relative gene expression levels in vertebrate ovaries are further supported by comparative analyses of differential gene expression (DE) between germline and somatic tissues. Multiple genes are significantly differentially expressed in the ovaries of non-mammal species while none are differentially expressed in the mouse or human, and few genes show significant differential expression in the platypus and opossum (Supplementary fig. S5).
TE-derived transcript abundance across tissues and vertebrate lineages.

To characterize TE transcription levels and composition across vertebrate tissues, we compared expression levels of total TE-derived transcripts (total-TE dataset; Supplementary fig. S6), as well as transcripts derived only from recently inserted TEs (recent-TE dataset; Supplementary fig. S7). Total-TE expression is substantial in both germline and somatic tissues across all species analyzed, although at variable levels within and between species (Supplementary figs. S6 and S8). For example, while total-TE derived transcripts comprise on average 6.55% of the transcriptome, values range from 0.26% in the chicken muscle to 23.44% in the opossum spleen (Supplementary fig. S8; Supplementary file S4). Among sampled species, the chicken and human exhibit the lowest total-TE average expression levels (2.66% and 2.92% of the total transcriptome, respectively), due mainly to very low TE transcription levels in somatic tissues (1.52% and 2.35% of the transcriptome on average, respectively). The highest average levels of total TE expression are found in the two snake species, the prairie rattlesnake and boa constrictor (13.75% and 12.16% of the transcriptome, respectively).

Our analyses also show that germline tissues do not always exhibit higher average total-TE expression levels than somatic tissues in vertebrates. For example, the clawed frog, prairie rattlesnake,
platypus, and opossum all exhibit higher average total-TE expression in somatic tissues compared to germline tissues. In the prairie rattlesnake, platypus, and opossum, this is driven by expression levels that are generally elevated in all or several somatic tissues. In the case of the clawed frog, this pattern is driven by comparatively low expression levels of total TE transcripts in the germline (which are the lowest across all vertebrate species analyzed). Despite high variance in TE expression levels across tissues, several tissues have relatively consistent trends across species. For example, the testes exhibit significantly greater expression on average compared to the ovary (pairwise Wilcoxon test p-value = 0.02; Supplementary fig. S8); this trend is consistent across all species except the opossum, where expression in ovaries is higher than in testes. Additionally, the brain has consistently high total-TE transcription levels across species, which is also higher than expression in testes (average of 9.61% versus 9.07% of the transcriptomes made up by TE-derived transcripts in the brain and testes, respectively). Conversely, muscle and ovary exhibit consistently low total-TE expression levels (Supplementary fig. S8; Supplementary file S4). Average expression in the muscle is significantly lower than that of testes (Wilcoxon test p-value = 0.01), brain (Wilcoxon test p-value = 0.04) and spleen (Wilcoxon test p-value = 0.03), and average expression in ovaries is significantly lower than in testes and brain (Wilcoxon test p-values = 0.02 and
0.04, respectively).

Recent-TEs are expressed in both germline and somatic tissues across vertebrates, although at lower levels (0.14% of the transcriptome on average across tissues and species) compared to all TE-derived transcripts (Supplementary figs. S8-10; Supplementary file S4). While lower overall, proportional expression levels of recent-TEs are variable across species and tissues (e.g., from 0.003% in boa muscle to 1.94% in zebrafish testes), similar to trends in total-TE transcript levels. However, pairwise comparisons testing for differences in average expression levels across species per tissue were not significant (Kruskal-Wallis rank sum test p-value = 0.39; pairwise Wilcoxon test p-values > 0.5). In contrast to the total-TE transcript dataset, average recent-TE expression is highest in the testes (0.27%, although though this is driven primarily by high testis expression in the zebrafish), followed by the small intestine and the brain (0.22% and 0.19%, respectively). We found multiple examples of divergent levels of recent TE transcript expression among species within major vertebrate lineages. For example, while mouse exhibits among the highest average recent TE expression levels, human has low average recent-TE expression levels (Supplementary file S4; Supplementary figs. S7C and S9).

Overall, our analyses demonstrate that recent and total TE expression levels in somatic tissues are poor predictors of one another. For
example, the small intestine has a higher relative fraction of the transcriptome derived from recent-TEs, while the brain and the spleen have higher fractions of the transcriptome made by TE-derived transcripts that originated from more ancient (and presumably non-mobilizing) TE families (Supplementary fig. S9; Supplementary file S4). Differences between recent and total TE expression among germline tissues tend to be clade-specific. In the testes, mammal and non-mammal species have similar average total-TE expression levels (8.43% vs. 9.4%, respectively), but remarkably different recent-TE expression levels (0.14% and 0.33% respectively). With the exception of the zebrafish, however, recent-TE expression levels are very similar (0.14% and 0.10%), in agreement with findings for total-TE expression. In contrast, mammalian ovaries exhibit more than two-fold greater TE expression than non-mammal species (2.56-fold for recent-TEs and 2.16-fold for total-TEs; Supplementary figs. S9 and S10, Supplementary file S4). There is also a positive relationship between the fold-change in TE expression levels (total-TE/recent-TE) between testes and ovaries at the phylogenetic scale (Supplementary fig. S11), and TE-family composition in testis and ovary is very similar for total-TE transcripts. Yet, analyses of recent-TE transcriptional levels highlight sexually dimorphic TE expression, with some specific TE families being expressed exclusively in either ovaries or testes. For example, CR1-LINEs are expressed in the python ovary but not in the testis, and the opposite pattern is observed in the platypus...
(Supplementary fig. S10). Despite tissue-specific expression of some TE families in the recent-TE transcriptome of testes and ovaries, significant associations (linear regression \( p \)-values < 0.005) are still found between relative TE composition of germline tissues for both total and recent TE expression for each species (Supplementary fig. S12).

Relationships between genome and transcriptome TE composition in germline tissues

A stochastic model of genome-wide transcription predicts that a vast majority of the genome is transcribed at some level (Encode Project Consortium 2012; Hangauer, et al. 2013). To test whether such a model applies to TEs across vertebrate lineages, we compared relative expression levels of 16 major TE families in the germline to the relative TE composition of the genome for each species analyzed (Supplementary file S5). Each vertebrate species is characterized by a strong significant linear relationship between gametic tissue total-TE expression and the relative genomic abundance of TEs for each respective genome (linear regression \( p \)-values < 0.04; Supplementary file S6, Figure 3, Supplementary fig. S13). We observe similar trends in relative recent-TE transcriptome composition and relative abundance of recently inserted TE-copies in the genome (\( p \)-values for all linear regressions are reported in Supplementary file S6). However, coefficients of determination are generally lower for
recent-TEs than for total-TEs, and in some species we find a lack of support for the relationship between genome TE content and TE transcriptional levels in the recent-TE matched comparisons (e.g., chicken, anole, and mouse ovary). This pattern likely stems from multiple instances of TE subfamilies being entirely absent in germline transcriptomes but detectable in the genomes of these species, a trend that is observed in mammals and birds in particular (Figure 3). Finally, comparisons of the relative total genomic TE composition to the relative abundance of recent-TEs in germline transcriptomes reveal a lack of associations in testes and ovaries for most species. Mammal species represent an exception to this general trend, however, as they do exhibit significant linear correlations between genomic TE composition and recent-TE expression in both tissues, although with low coefficients of determination (linear regression p-values < 0.04, Supplementary file S6).

Relationships between recent TE expression and TE regulatory activity

Considering multiple lines of evidence for differential regulation of TE activity in germline tissues, we tested the relationships between the magnitude of host response against TEs (particularly the relative activation of the PIWI pathway) and recent-TE expression in germline and somatic tissues (Figure 4; Supplementary fig. S14A). Relationships were tested using PIC Spearman’s rank-order correlation analyses (Supplementary fig. S14, Supplementary file S8) and, for
germline tissues, PIC linear regression analyses (Figure 5, Supplementary files S7). We found evidence for significant associations between expression levels of recent-TEs and both PIWI pathway genes and the entire set of genes involved in TE regulation only in the testes (PIC linear regression p-values = 0.02 and 0.004, respectively; Figure 5, Supplementary file S7). All other correlation analyses lacked significant associations between recent-TE and regulatory mechanism expression levels, including analyses where somatic and germline tissues were combined and analyses of combined germline tissues (Supplementary fig. S14, Supplementary file S8). Relationships between recent-TE expression and regulatory activity in ovaries were not significant in analyses of all species as well as analyses in which mammals, which exhibited particularly low expression of PIWI pathway genes and above average recent-TE expression, were excluded (p-values > 0.05; Figure 5, Supplementary file S4, Supplementary fig. S14D).

Comparison of expression patterns between oocytes, ovaries, and testes

Our analyses demonstrate broad differences in expression profiles of TE regulatory mechanisms and TE-derived transcripts between testes and ovaries, as well as between mammal and non-mammal ovarian tissues. To evaluate the possibility that our findings are linked to lower fractions of germ cells in ovaries relative to testes, we
analyzed available data from purified oocyte cell populations of two non-amniote (zebrafish and clawed frog) and three amniote (chicken, human and mouse) species. In the zebrafish and clawed frog, oocyte TE regulation and expression profiles recapitulate observations derived from ovarian tissues (Supplementary figs. S15 and S16). Expression of genes involved in TE regulation are similar to that observed in the ovary, although oocyte cells have noticeably higher expression of genes belonging to the siRNA pathway compared to somatic tissues (Supplementary figs. S15A and S16). Similarly, estimates of both total and recent TE-derived transcript expression appear similar to those of the ovary (Supplementary figs. S6A and S7A). Oocytes also exhibited a complement of high relative activation of TE regulatory mechanisms with comparatively lower recent-TE expression, consistent with observations in ovaries (Supplementary fig. S15A). In the chicken, oocyte cells share features of TE regulation with the ovary (Supplementary fig. S15), yet expression levels of recent-TEs are higher than in both ovary and testis. TE regulatory genes in human and mouse oocytes show remarkably distinct clustering patterns (Supplementary fig. S16), but in both species we find support for active regulation of TEs (Supplementary fig. 15A); this is in contrast with profiles of TE regulation in the ovaries. Specifically, expression levels of genes involved in negative TE regulations are more similar to those of the testis than those of the ovary, particularly for PIWI pathway genes. We also find that the oocyte
cell populations show unique profiles of total and recent-TE relative expression compared to ovaries and testes in both species (Figure 3, Supplementary fig. S15C, Supplementary file 5). Our results suggest that human oocytes predominantly express Alu SINEs, whereas the testis and ovary show additional relevant contributions of ERV LTRs and L1 LINEs to the transcriptome. In mouse oocytes, most of the recent-TE derived transcriptome appears to originate from ERV LTRs, but in the ovary and testis L1 LINEs are highly represented as well. Collectively, our analyses suggest that ovaries and oocytes are similar in the zebrafish and the clawed frog, but show distinct characteristics of regulatory mechanism and TE expression in human and mouse.

Discussion

A vertebrate-wide perspective on TE expression and TE regulatory pathway activity

To date, studies of TE expression have primarily focused on analysis of male germline and embryonic tissues in mammals (e.g., human and mouse) with the goal of understanding mechanisms that regulate TE activity during developmental stages associated with genome-wide DNA demethylation, and therefore critical for the vertical propagation of TEs (Ernst, et al. 2017; Hajkova, et al. 2002; Richardson et al., 2017; Surani, et al. 2007). Our integrated analyses across germline and somatic tissues shed new light on the variation that exists in
both TE expression and regulatory mechanisms among vertebrates, and highlight major differences between germline patterns in mammals compared to other vertebrate lineages. Our results also raise new questions about the relatively high, yet variable, levels of TE-derived transcripts across somatic and gametic tissues in vertebrates, and underscore the poorly understood relationships between TE regulation and TE transcript expression.

Overlooked complexity of TE negative regulation in the vertebrate germline

Despite major differences in evolutionary history and genomic composition of vertebrate TE landscapes, evidence of active TE repression via multiple conserved regulatory pathways appears to be a shared feature of vertebrates somatic and gametic tissues. Expression levels of TE repression mechanisms are particularly variable in ovaries across vertebrate lineages, yet appear to be relatively conserved in the testes. Mammals in particular appear to express genes involved in regulating TE expression in the ovary at a low level similar to expression in somatic tissues, which directly contrasts the active regulatory signature observed in the ovary of other vertebrates. Reduced expression levels of TE regulatory genes in mammalian ovaries may explain why polymorphic TE insertions that have developmental origins in the female early embryo and late germline exhibit the highest transmission rates in mice (Richardson
These findings raise questions regarding the biological basis and selective drivers that underlie reduced ovarian TE regulation in mammals compared to other vertebrate lineages. One potential explanation may involve differences in mitotic rates in mammals. Previous studies of TE activity and repression have focused specifically on the male germline over the female germline due to higher mitotic and meiotic rates during spermatogenesis (Handel and Schimenti 2010), and other previous studies have indicated that TE activity positively correlates with tissue-specific cell mitotic rates (Navarro, et al. 2019). To further explore these relationships, we analyzed expression data derived from oocyte cell populations for five vertebrate species to evaluate if differences in the proportion of germ cells in vertebrate ovarian tissues explain the distinct profiles recovered for mammal and non-mammal species. Although the limited taxonomic sampling prevents us from drawing broad conclusions, our results suggest a relationship between expression patterns of TEs and TE regulatory mechanisms in the ovary and the ratio of somatic-to-germ cells in female gametic tissues. In species characterized by the deposition of numerous eggs (e.g., zebrafish and clawed frog), we found profiles of TE regulatory mechanisms and TE expression in oocytes to be remarkably similar to those of the ovary. These findings agree with the existence of an ovarian germline stem
cell (OGSCs) population to replenish the oocyte pool (Hanna and Hennebold 2014). In contrast, human and mouse (where the presence of OGSCs is still debated; Hanna and Hennebold 2014) show profiles of TE expression and regulation that more closely resemble testis profiles than those of ovary, suggesting a lower fraction of germ cells and their precursors in the ovary. While chicken oocyte cells share features of TE regulation with the chicken ovary, our analyses seem to agree with previous findings that do not support an OGSC population (Motono, et al. 2008; Nakamura, et al. 2013).

Across animals, ovaries are characterized by a cell population in meiotic arrest (Sagata 1996). Our analyses provide indirect support that differences likely exist not only in the frequency and magnitude of oocyte activation across lineages (Abrieu, et al. 2001), but also between germ cells at the same maturation stage of closely related species (i.e., MII oocytes of human and mouse). Future study of the variation in key features of ovarian biology across vertebrates, including mitotic and meiotic rates, stage of oocyte maturation at the onset of meiotic arrest, as well as the presence of OGSCs, may prove valuable for examining links between variation in characteristics of ovarian biology and the activity of TE regulatory mechanisms across vertebrate lineages.

Few previous studies have focused on TE regulatory mechanisms outside of the mammalian germline (Lim, et al. 2013; Malki, et al. 2014;
Watanabe, et al. 2008), which limits the context for comparison of our results across tissues in vertebrates. Our conclusion that PIWI pathway genes are expressed at similar levels in testes and ovaries is broadly consistent with previous studies in the zebrafish, clawed frog, and anole (Houwing, et al. 2007; Kirino, et al. 2009; Zhang, et al. 2017), while expression of PIWI mRNAs or piRNAs have not been detected in previous studies of chicken ovaries (Sun, et al. 2017).

Interestingly, the zebrafish is also known to produce sex-specific piRNAs from distinct genomic TE loci (Zhou, et al. 2010); if this mechanism exists in other vertebrates, it may provide an explanation for sexually dimorphic expression of recent-TEs in the germline.

**TE regulatory pathways do not clearly demarcate somatic and gametic tissues**

Our comparative analyses illustrate that expression of genes involved in the negative regulation of TEs is not limited to the germline. Among the four categories of TE regulatory mechanisms analyzed, only expression levels of the PIWI:piRNA pathway consistently discriminated at least one germline tissue from somatic tissues based on variation in across-tissue gene expression. In contrast, endogenous small interfering RNA (siRNA), transcriptional, and post-transcriptional pathways are all characterized by relatively consistent expression levels across germline and somatic tissues. Our analyses therefore support the canonical view of PIWI pathway genes
and associated piRNAs are a hallmark of gametic tissues, and the 
vanguard of germline genome integrity.

TE expression and TE repression mechanisms have been extensively 
studied in somatic tissues, but mostly in association with cancer, 
aging, and other diseases (Burns 2017; Jang 2019; Kazazian 1998; 
Kreiling, et al. 2017). Those studies have led to the collective view 
that, because of the threat that TE mobilization poses to genome 
integrity and structure, their expression is severely restricted at 
both transcriptional and post-transcriptional levels. Subsequent 
studies found exceptions to this pattern in the central nervous 
system and in specific developmental stages, where expression of 
specific elements promotes cellular mosaicism and the correct 
execution of cell specification pathways, respectively (Baillie, et 
al. 2011; Hackett, et al. 2017; Weissman and Gage 2016). Broadly, our 
findings indicate that genes traditionally associated with the 
germline (e.g., genes in the PIWI:piRNA pathway; Ponnusamy, et al. 
2017) also exhibit detectable expression in somatic tissues, although 
often at low levels, and vice-versa (e.g., genes in the siRNA 
pathway; Stein, et al. 2015).

The brain is the only somatic tissue where de novo TE insertions have 
been identified outside of the germline in non-pathological 
conditions (e.g., Baillie, et al. 2011; Weissman and Gage 2016) 
Understanding how TE activity is regulated in the central nervous

http://mc.manuscriptcentral.com/gbe
system is therefore a topic of primary interest. In the context of our study, we find a single, distinct profile of TE regulation common to all vertebrate brain tissues characterized by higher relative expression of transcriptional regulators (e.g., TRIM28 and methyltransferases). This finding suggests that a conserved landscape of TE activity may exist in the central nervous system across vertebrates. PIWI genes and most members of the PIWI:piRNA pathway show little to no expression in the brain, suggesting that this regulatory mechanism does not play a role in TE regulation in the brain across vertebrates (Supplementary figs. S1 and S17), and that other repressive mechanisms may have evolved to regulate TE mobilization in the central nervous system (Grassi, et al. 2019). These results further support the roles of TE regulation in somatic tissues, possibly through the evolution of compensatory or reinforcing mechanisms, or the cooption of existing mechanisms for TE regulation (Levine, et al. 2016).

**Interpretations of TE-derived transcript abundance**

Our analyses demonstrate that TE-derived transcripts on average comprise a notably large fraction of the transcriptomes of germline and somatic tissues across vertebrate lineages. We expected a priori that a majority of TE-derived transcripts would originate from recent active TE families, yet this pattern was not observed in any of the species analyzed. Instead, TE-derived transcripts originate from a
variety of recent and ancient TEs families among the species studied. These findings, corroborated by the identification of similar relative composition of genomes and TE transcriptomes across species, support hypotheses from studies in mammals evoking a stochastic transcription model, in which the majority of the genome is pervasively transcribed (Encode Project Consortium 2012; Hangauer, et al. 2013). Although the majority of TE-derived transcripts may not have biological activity related to insertional mutagenesis or replication, it remains an open question whether the abundant pool of TE-derived cellular RNAs have other biologically relevant impacts in gene regulation (e.g., lncRNAs and microRNAs), unappreciated roles due to their sheer abundance (e.g., mass-effect competition for RNA catabolic processes, RNA metabolism, and interference with translation) or potential cooption as regulatory elements (Chuong, et al. 2016; Cordaux and Batzer 2009; Lippman, et al. 2004; van de Lagemaat, et al. 2003). It is notable that particular somatic tissues in some species exhibit distinctly high estimates of TE transcripts (e.g., 23.44% in opossum spleen). Because our approach cannot differentiate between legitimate TE-mRNAs and transcriptional read-through, it is unclear whether one or both of these may explain these high expression levels. Specifically in the case of spleen tissues, it is plausible that our inferences of TE-derived transcripts may be confounded by TE-related gene expression (e.g., recombinase activating genes) in tissues with high levels of immune cell
activity.

To focus on TE-derived transcripts that are likely to be relevant sources of mutation and transpositional activity, we restricted our analyses to transcripts that originated only from recently-active TEs (i.e., recent-TEs). Such recent TEs are likely to be more strongly targeted by negative regulatory mechanisms (Sun, et al. 2017; Vandewege, et al. 2016). We find that recent-TEs are expressed in both germline and somatic tissues across vertebrates, although at far lower levels (mean = 0.14% of the transcriptome) compared to total TE-derived transcripts (mean = 6.55%). Recent-TE expression tends to be highest in the testes, followed by the small intestine and the brain. Our results also indicate that patterns of recent-TE expression in mammals are unique among vertebrates analyzed. Mammals have relatively higher levels of expression in the ovaries, such that mammalian ovaries and testes show similar recent-TE expression levels. We also identified multiple examples of highly divergent levels of recent-TE transcript expression among species within major lineages, suggesting that substantial variation likely exists across species within major vertebrate lineages.

Variation in genome assembly quality may impact reconstructions of genomic TEs, especially recent TEs. While this is expected to impact estimates of genomic TE content for recent TEs, this should have a minor impact on inferences of TE activity as long as at least some
copies of recent-TE sequences are reconstructed correctly for each repeat type per genome. Accordingly, our genome-wide mapping of RNAseq reads to annotated genomic repeat sequences to infer transcript levels for each type of TE should be robust across genome assemblies of reasonable quality, since our inferences of transcriptional activity of TE types are not dependent on the number of recent TE copies, but rather on having at least some closely-related examples of each TE type present in the genome annotation.

Our analyses of the relative composition of TE-derived transcripts in vertebrate gametic tissue transcriptomes highlight extreme variability in evolutionary TE dynamics across lineages. Some species (e.g., squamate reptiles), show fairly equal representation of all major TE subfamilies in their recent TE-derived transcriptomes. Other species, including human, mouse, and chicken, have been characterized by the extinction of most TE families, such that few elements are thought to remain capable of generating novel insertions (Gagnier, et al. 2019). In human and mouse we inferred traces of DNA transposon-derived transcripts in the recent-TE dataset, which we expect likely represent results derived from transcriptional read-through. On a broad level, however, our analyses agree with recent studies (e.g., Feusier et al., 2019) suggesting high retrotransposition rates in the human germline of L1 LINEs and Alu/SVA SINEs. We also found that human testes, ovaries, and oocytes have different relative recent-TE
derived transcripts composition, with L1 and Alu/SVA transcripts being present in the testis (30% and 41% of the recent-TE transcriptome, respectively), yet in oocytes we estimated Alu/SVA transcripts as comprising up 92% of the recent-TE transcriptome. Similar sexual dimorphism in relative recent-TE-derived transcript composition was also observed in mouse and chicken, but not in zebrafish or clawed frog. Substantial differences therefore exist in TE expression and TE regulation between male and female germline tissues, and relationships between sex-specific germline tissue expression patterns further vary across vertebrate lineages.

Collectively, our analyses of the relationships between recent-TE expression and TE regulatory pathway activity provide evidence for divergent patterns between gametic tissues across vertebrates. In the male germline, there is a positive relationship across vertebrates between expression levels of recent-TEs and TE regulatory pathway activity. Given how the PIWI:piRNA pathway acts to repress TE (i.e., Lim and Kai 2015), these findings may suggest that activation of TE repressive mechanisms may be proportional to the magnitude of threat posed by TE expression and activity. This is consistent with previous findings that higher TE activity is associated with higher TE repressive mechanism activation in the host (Reznik et al., 2019). In contrast, no significant relationship was found between the expression of recent-TEs and TE repressive mechanisms in vertebrate
ovarian tissues, suggesting that the unique biology of the ovary may confer or necessitate unique mechanisms to prevent the potentially deleterious effects of TE activity.

Conclusions and future directions

Our comparative analyses of TE regulation and expression across vertebrate lineages suggest that active repression of TEs is accomplished by multiple conserved mechanisms, and represents a shared feature among germline and somatic vertebrate tissues. Our results also highlight highly unique sexually dimorphic TE-associated biology specific to gametic tissues. We find that patterns of TE regulation are remarkably distinct in mammalian ovarian tissues compared to other vertebrates, and that a shift towards decreased TE regulatory activity in ovaries occurred early in the evolution of the eutherian mammal lineage. Yet, analyses of oocyte cells for two mammal species suggest the possibility that lower expression levels of TE regulatory mechanisms may be due to a lower proportion of germ cells in the eutherian mammal ovary compared to other vertebrate species. These findings, together with other differences in TE regulation and expression identified among vertebrate lineages underscore the importance of studies of diverse vertebrate lineages and tissues for understanding the uniqueness of mammalian biology, and demonstrate the potential shortcomings of broad assumptions that diverse vertebrate model systems share common biological features and
regulatory mechanisms. Our findings also underscore challenges in understanding the relevance of TE-derived transcript abundance from analysis of RNA-seq data alone, and argue for future integration of approaches that quantify transpositionally competent TE-derived transcripts (Deininger, et al. 2017), allow for a better discrimination of TE-loci transcription and gene read-through, and leverage other functional data (Goubert, et al. 2019; Faulkner, et al. 2009; He, et al. 2019; Sun, et al. 2017).

Although our analyses focused on TE negative regulatory pathways and how they relate to expression levels of recent TE-derived transcripts, instances of genes and transcription factors that can promote TE activity (e.g., Runx3; Yang et al. 2003), or both negatively and positively regulate TEs in a context-specific fashion (e.g., YY1; Sanchez-Luque, et al. 2019; Athanikar, et al. 2004) have been reported. While beyond the scope of the our current study, we cannot dismiss that different expression levels of genes that promote TE transcription may result in difference in expression levels of specific TE subfamilies across tissues (e.g., Supplementary fig. S18), and investigation of such positive TE regulation mechanisms will be an important complement to our analyses of negative TE regulation.

Acknowledgements

Support was provided from startup funds from the University of Texas
at Arlington to TAC and the Society for the Study of Evolution (to GIMP). We acknowledge the Texas Advanced Computing Center (TACC) for providing access to computational resources.

**Competing interests**

The authors declare no competing interests.
References

Abrieu A, Doree M, Fisher D 2001. The interplay between cyclin-B-Cdc2 kinase (MPF) and MAP kinase during maturation of oocytes. Journal of Cell Science 114: 257-267.

Agrawal A, Eastman QM, Schatz DG 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394: 744-751.

Anders S, Huber W 2010. Differential expression analysis for sequence count data. Genome Biology 11. doi:10.1186/gb-2010-11-10-r106

Aravin AA, Tuschi T 2005. Identification and characterization of small RNAs involved in RNA silencing. FEBS Lett 579: 5830-5840. doi: 10.1016/j.febslet.2005.08.009

Aravin AA, Hannon GJ, Brennecke J 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318: 761-764. doi: 10.1126/science.1146484

Baillie JK, et al. 2011. Somatic retrotransposition alters the genetic landscape of the human brain. Nature 479: 534-537. doi: 10.1038/nature10531

Beck CR, Garcia-Perez JL, Badge RM, Moran JV 2011. LINE-1 Elements in Structural Variation and Disease. Annual Review of Genomics and Human Genetics, Vol 12: 187-215.

Bedrosian TA, Linker S, Gage FH 2016. Environment-driven somatic mosaicism in brain disorders. Genome Medicine 8.

Boissinot S, Davis J, Entezam A, Petrov D, Furano AV 2006. Fitness cost of LINE-1 (L1) activity in humans. PNAS 103: 9590-9594.

Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120. doi: 10.1093/bioinformatics/btu170

Bourque G, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Research 18: 1752-1762.

Burns KH 2017. Transposable elements in cancer. Nat Rev Cancer 17: 415-424. doi: 10.1038/nrc.2017.35

Callinan PA, Batzer MA 2006. Retrotransposable elements and human disease. Genome Dyn 1: 104-115. doi: 10.1159/000092503

Carbon S, et al. 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25: 288-289. doi: 10.1093/bioinformatics/btn615

Chalopin D, Naville M, Plard F, Galiana D, Volff JN 2015. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. Genome Biology and Evolution 7: 567-580.

Chuong EB, Elde NC, Feschotte C 2016. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. doi: 10.1038/nrg.2016.139

Encode Project Consortium 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74. doi: 10.1038/nature11247
Cordaux R, Batzer MA 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet 10: 691-703. doi: 10.1038/nrg2640

De Cecco M, et al. 2013. Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. Aging (Albany NY) 5: 867-883.

Deininger P, et al. 2017. A comprehensive approach to expression of L1 loci. Nucleic Acids Res 45: e31. doi: 10.1093/nar/gkw1067

Dobin A, et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15-21. doi: 10.1093/bioinformatics/bts635

Dunn CW, Zapata F, Munro C, Siebert S, Hejnol A 2018. Pairwise comparisons across species are problematic when analyzing functional genomic data. PNAS 115: E409-E417. doi: 10.1073/pnas.1707515115

Ernst C, Odom DT, Kutter C 2017. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. Nat Commun 8: 1411. doi: 10.1038/s41467-017-01049-7

Faulkner GJ, Billon V 2018. L1 retrotransposition in the soma: a field jumping ahead. Mob DNA 9: 22. doi: 10.1186/s13100-018-0128-1

Faulkner GJ, Garcia-Perez JL 2017. L1 Mosaicism in Mammals: Extent, Effects, and Evolution. TRENDS in Genetics 33: 802-816. doi: 10.1016/j.tig.2017.07.004

Faulkner GJ, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. Nature Genetics 41: 563-571. doi: 10.1038/ng.368

Feusier J, et al. 2019. Pedigree-based estimation of human mobile element retrotransposition rates. Genome Res 29: 1567-1577. doi: 10.1101/gr.247965.118

Gagnier L, Belancio VP, Mager DL 2019. Mouse germ line mutations due to retrotransposon insertions. Mobile DNA 10. doi: 10.1186/s13100-019-0157-4

Garcia-Perez JL, Widmann TJ, Adams IR 2016. The impact of transposable elements on mammalian development. Development 143: 4101-4114. doi: 10.1242/dev.132639

Gasior SL, Wakeman TP, Xu B, Deininger PL 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. Journal of Molecular Biology 357: 1383-1393. doi: 10.1016/j.jmb.2006.01.089

Gilbert C, Hernandez SS, Flores-Benabib J, Smith EN, Feschotte C 2012. Rampant Horizontal Transfer of SPIN Transposons in Squamate Reptiles. Molecular Biology and Evolution 29: 503-515.

Goodier JL 2016. Restricting retrotransposons: a review. Mob DNA 7: 16. doi: 10.1186/s13100-016-0070-z

Grassi DA, Jonsson ME, Brattas PL, Jakobsson J 2019. TRIM28 and the control of transposable elements in the brain. Brain Res 1705: 43-47. doi: 10.1016/j.brainres.2018.02.043

Greenberg MVC, Bourc'his D 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 20: 590-607. doi: 10.1038/s41580-019-0159-6

Hackett JA, Kobayashi T, Dietmann S, Surani MA 2017. Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. Stem Cell Reports 8: 1645-1658.

Hajkova P, et al. 2002. Epigenetic reprogramming in mouse primordial germ cells. Mechanisms of Development 117: 15-23. doi: 10.1016/S0925-4773(02)00181-8
Handel MA, Schimenti JC 2010. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat Rev Genet 11: 124-136. doi: 10.1038/nrg2723

Hangauer MJ, Vaughan IW, McManus MT 2013. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. Plos Genetics 9: e1003569. doi: 10.1371/journal.pgen.1003569

Hanna CB, Hennebold JD 2014. Ovarian germline stem cells: an unlimited source of oocytes? Fertil Steril 101: 20-30. doi: 10.1016/j.fertnstert.2013.11.009

He J, et al. 2019. Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. Nat Commun 10: 34. doi: 10.1038/s41467-018-08006-y

Houwing S, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129: 69-82. doi: 10.1016/j.cell.2007.03.026

Hutchins AP, Pei D 2015. Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. Sci Bull (Beijing) 60: 1722-1733. doi: 10.1007/s11434-015-0905-x

Jacobs FMJ, et al. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature 516: 242.

Jang HS, et al. 2019. Transposable elements drive widespread expression of oncogenes in human cancers. Nature Genetics 51: 611-617. doi: 10.1038/s41588-019-0373-3

Jansz N 2019. DNA methylation dynamics at transposable elements in mammals. Essays in Biochemistry 63: 677-689. doi: 10.1042/ebc20190039

Jin Y, Tam OH, Paniagua E, Hammell M 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics 31: 3593-3599. doi: 10.1093/bioinformatics/btv422

Kapusta A, Suh A, Feschotte C 2017. Dynamics of genome size evolution in birds and mammals. PNAS 114: E1460-E1469.

Kato Y, et al. 2007. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Human Molecular Genetics 16: 2272-2280. doi: 10.1093/hmg/ddm179

Kazazian HH 1998. Mobile elements and disease. Current Opinion in Genetics & Development 8: 343-350.

Kirino Y, et al. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nature Cell Biology 11: 652-U478. doi: 10.1038/nceb1872

Kopylova E, Noe L, Touzet H 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28: 3211-3217. doi: 10.1093/bioinformatics/bts611

Kordis D, Gubensek F 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. PNAS 95: 10704-10709. doi: DOI 10.1073/pnas.95.18.10704

Kreiling JA, et al. 2017. Contribution of retrotransposable elements to aging. In. Human retrotransposons in health and disease: Springer. p. 297-321.
Levine AJ, Ting DT, Greenbaum BD 2016. P53 and the defenses against genome instability caused by transposons and repetitive elements. Bioessays 38: 508-513. doi: 10.1002/bies.201600031

Lim RS, Kai T 2015. A piece of the pi(e): The diverse roles of animal piRNAs and their PIWI partners. Semin Cell Dev Biol 47-48: 17-31. doi: 10.1016/j.semcdb.2015.10.025

Lim SL, et al. 2013. Conservation and expression of PIWI-interacting RNA pathway genes in male and female adult gonad of amniotes. Biol Reprod 86: 136. doi: 10.1095/biolreprod.113.111211

Lippman Z, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471-476. doi: 10.1038/nature02651

Loreto ELS, Pereira CM 2017. Somatizing the transposons action. Mob Genet Elements 7: 1-9. doi: 10.1080/2159256X.2017.1314236

Love MI, Huber W, Anders S 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550. doi: 10.1186/s13059-014-0550-8

Lynch M, Conery JS 2003. The origins of genome complexity. Science 302: 1401-1404.

Lynch M, Walsh B. 2007. The origins of genome architecture: Sinauer Associates Sunderland.

Lynch VJ, et al. 2015. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep 10: 551-561. doi: 10.1016/j.celrep.2014.12.052

Makałowski W, Kischka T, Makałowska I 2017. Contribution of transposable elements to human proteins. eLS.

Malki S, van der Heijden GW, O'Donnell KA, Martin SL, Bortvin A 2014. A Role for Retrotransposon LINE-1 in Fetal Oocyte Attrition in Mice. Developmental Cell 29: 521-533. doi: 10.1016/j.devcel.2014.04.027

Molaro A, et al. 2014. Two waves of de novo methylation during mouse germ cell development. Genes & Development 28: 1544-1549. doi: 10.1101/gad.244350.114

Motono M, Ohashi T, Nishijima K, Iijima S 2008. Analysis of chicken primordial germ cells. Cytotechnology 57: 199-205. doi: 10.1007/s10616-008-9156-x

Nakamura Y, Kagami H, Tagami T 2013. Development, differentiation and manipulation of chicken germ cells. Dev Growth Differ 55: 20-40. doi: 10.1111/dgd.12026

Navarro F, et al. 2019. TeXP: Deconvolving the effects of pervasive and autonomous transcription of transposable elements. bioRxiv: 648667.

Neafsey DE, Blumenstiel JP, Hartl DL 2004. Different regulatory mechanisms underlie similar transposable element profiles in pufferfish and fruitflies. Molecular Biology and Evolution 21: 2310-2318.

Pasquesi GIM, et al. 2018. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat Commun 9. doi: 10.1038/s41467-018-05279-1

Platt RN, Vandewege MW, Ray DA 2018. Mammalian transposable elements and their impacts on genome evolution. Chromosome Research 26: 25-43.
Ponnusamy M, Yan KW, Liu CY, Li PF, Wang K 2017. PIWI family emerging as a decisive factor of cell fate: An overview. European Journal of Cell Biology 96: 746-757. doi: 10.1016/j.ejcb.2017.09.004

Reik W 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447: 425-432.

Revell LJ 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3: 217-223. doi: 10.1111/j.2041-210X.2011.00169.x

Reznik B, Cincotta SA, Jaszczak RG, Mateo LJ, Shen J, Cao M, Baskin L, Ye P, An W, Laird DJ 2019. Heterogeneity of transposon expression and activation of the repressive network in human fetal germ cells. Development 146(12). doi: 10.1242/dev.171157

Richardson SR, et al. 2017. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. Genome Res 27: 1395-1405. doi: 10.1101/gr.219022.116

Sagata N 1996. Meiotic metaphase arrest in animal oocytes: Its mechanisms and biological significance. Trends in Cell Biology 6: 22-28. doi: 10.1016/0962-8924(96)81034-8

Sanchez-Luque FJ, et al. 2019. LINE-1 Evasion of Epigenetic Repression in Humans. Molecular Cell 75(3): 590-604

Sayers EW, et al. 2019. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 47: D23-D28. doi: 10.1093/nar/gky1069

Sen SK, et al. 2006. Human genomic deletions mediated by recombination between Alu elements. American Journal of Human Genetics 79: 41-53.

Shi X, Seluanov A, Gorbunova V 2007. Cell divisions are required for L1 retrotransposition. Mol Cell Biol 27: 1264-1270. doi: 10.1128/MCB.01888-06

Siomi MC, Sato K, Pezic D, Aravin AA 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol 12: 246-258. doi: 10.1038/nrm3089

Slotkin RK, Martienssen R 2007. Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics 8: 272-285. doi: 10.1038/nrg2072

Smit AFA, Hubley R, Green P. 2013-2015. RepeatMasker Genomic Datasets. In.

Soumillon M, et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep 3: 2179-2190. doi: 10.1016/j.celrep.2013.05.031

Stein P, et al. 2015. Essential Role for endogenous siRNAs during meiosis in mouse oocytes. Plos Genetics 11: e1005013. doi: 10.1371/journal.pgen.1005013

Stewart KR, Veselovska L, Kelsey G 2016. Establishment and functions of DNA methylation in the germline. Epigenomics 8: 1399-1413. doi: 10.2217/epi-2016-0056

Sudmant PH, Alexis MS, Burge CB 2015. Meta-analysis of RNA-seq expression data across species, tissues and studies. Genome Biol 16: 287. doi: 10.1186/s13059-015-0853-4

Sun YH, et al. 2017. Domestic chickens activate a piRNA defense against avian leukemia virus. Elife 6.

Surani MA, Hayashi K, Hajkova P 2007. Genetic and epigenetic regulators of pluripotency. Cell 128: 747-762. doi: 10.1016/j.cell.2007.02.010

http://mc.manuscriptcentral.com/gbe
van de Lagemaat LN, Landry JR, Mager DL, Medstrand P 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. TRENDS in Genetics 19: 530-536. doi: 10.1016/j.tig.2003.08.004

Vandewege MW, Platt RN, Ray DA, Hoffmann FG 2016. Transposable Element Targeting by piRNAs in Laurasiatherians with Distinct Transposable Element Histories. Genome Biology and Evolution 8: 1327-1337.

Vogt J, et al. 2014. SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. Genome Biol 15: R80. doi: 10.1186/gb-2014-15-6-r80

Watanabe T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453: 539-U539. doi: 10.1038/nature06908

Weick EM, Miska EA 2014. piRNAs: from biogenesis to function. Development 141: 3458-3471. doi: 10.1242/dev.094037

Weissman IL, Gage FH 2016. A Mechanism for Somatic Brain Mosaicism. Cell 164: 593-595.

Xue AT, Ruggiero RP, Hickerson MJ, Boissinot S 2018. Differential Effect of Selection against LINE Retrotransposons Among Vertebrates Inferred From Whole-Genome Data and Demographic Modeling. Genome Biology and Evolution.

Yang N, Zhang L, Zhang Y, Kazazian HH Jr. 2003. An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res 31(16): 4929-4940

Yao Q, et al. 2018. Ribonuclease activity of MARF1 controls oocyte RNA homeostasis and genome integrity in mice. PNAS 115: 11250-11255. doi: 10.1073/pnas.1809744115

Zeng L, Pederson SM, Kortschak RD, Adelson DL 2018. Transposable elements and gene expression during the evolution of amniotes. Mob DNA 9: 17. doi: 10.1186/s13100-018-0124-5

Zhang B, et al. 2017. Identification and Characterization of a Class of MALAT1-like Genomic Loci. Cell Reports 19: 1723-1738.

Zhou X, et al. 2010. Profiling sex-specific piRNAs in zebrafish. Genetics 186: 1175-1185. doi: 10.1534/genetics.110.122234
Figure Legends

Figure 1. - Expression patterns of key genes involved in negative regulation of transposable element (TE) activity in germline and somatic vertebrate tissues. (A) For each species, heatmaps show log2-transformed within-species normalized expression levels of main pathways involved in TE silencing. Warm colors (yellow) represent higher total expression levels of genes in the pathway across tissues. (B) Principal component analyses (PCA) for the PIWI:piRNA pathway (left) and all other regulatory pathways (siRNA pathway, transcriptional and post-transcriptional TE silencing mechanisms; right) reflect variance in gene expression profiles across tissues for each species. While non-mammal species show discrimination of both germline tissues (testis in green and ovary in maroon) form somatic tissues (empty grey circles) and from each other in respect to PIWI pathway genes, gene expression in the mammalian ovary falls within the variability of somatic tissues. (C) PCA for the testis (above) and ovary (below) show species clustering based on the principal component of the PIWI pathway (x axis) and all other regulatory pathways (y axis). Per each species, coordinates were extracted from the corresponding phylogenetically independent contrast (PIC) PCAs. Cold colors represent non-amniote vertebrates, warm colors reptiles, and magenta mammal species.

Figure 2. - Hierarchical clustering Z-score heatmap of TE regulatory genes in germline and somatic vertebrate tissues. Analysis of differential expression of a subset of key conserved genes (present in at least 8 species) involved in TE silencing suggests the existence of 5 main expression profiles across vertebrate tissues: vertebrate testis, characterized by the highest activation status of the PIWI:piRNA pathway and transcriptional regulators; ovary of non-mammal species, with expression patterns similar to the testis; mammalian ovary (to the exclusion of humans), which shows a sharp decreased expression of PIWI genes; other somatic tissues (average Z-scores across heart, kidney, liver, muscle, spleen and small intestine after individual tissue heatmap supported the existence of a single cluster); and vertebrate brain.

Figure 3. - Relationship between genomic and transcriptomic TE relative composition in the male germline. Area of the circles in the balloon plot reflects the percentage of major TE subfamilies (blue = DNA transposons, green = LTRs, grey = PLE and DIRS, yellow = LINEs, violet = SINEs) relative to the total genomic TE content (top row, grey background) and to the total TE transcriptome (white background). In the box, the same relationship is displayed for recently inserted TE copies (with a Kimura distance < 2%) and recent-TEs in the transcriptome. Values to the left report the real proportion of TEs (TE content %) in the genomes and transcriptomes. We find support for high TE transcription in testis transcriptomes (up to 15%), which perfectly match the relative composition of the genome. In contrast, for recent-TEs some families are entirely missing in the transcriptome despite their presence in the genomic background. Balloon plot additionally highlights variability in TE landscapes across vertebrates.
**Figure 4.** – Expression levels of recent-TEs and their negative regulatory mechanisms in vertebrate somatic and germline tissues. Heatmap shows comparative expression levels of recent-TEs (top row), total regulatory pathways (PIWI:piRNA, siRNA, transcriptional and post-transcriptional), and details of the contribution of PIWI:piRNA pathway and all remaining silencing mechanism (bottom section) across vertebrate tissues. Comparative gene expression is reported as percentage of the transcriptome following within species normalization. Whereas human, xenopus and chicken show the lowest levels of recent-TE expression in both germline and somatic tissues, vertebrate tissues show moderate to high contribution of TEs to tissue transcriptomes, which is consistently highest in the testis, and reduced in non-mammal ovary.

**Figure 5.** – Relationship between expression levels of recent-TEs and their negative regulatory pathways. (A) Linear regressions and PICs support a significant positive relationship between recent-TE expression and host response (PIWI pathway and total response) in the testis, whereas in the ovary they suggest the opposite, although not significant, trend. (B) Patterns of species TE expression levels in the testis (x-axis) and ovary (y-axis). Recent-TE transcriptome percentages were corrected by the PIWI pathway to test for a correlation in expression levels. Mammal species show a consistent trend in the ovary where lower regulatory activity brings to increased TE transcription, matched by the testis although in favor of the PIWI pathway, compared to non-mammal species. In contrast, non-mammal species show a consistent host response proportional to TE activity in the ovary (constant TE:PIWI ratio), but higher variability in the testis, with some species that are more efficient at contrasting TEs.
| Species                  | DNA transposons | LTR       | LINEs       | SINEs       |
|--------------------------|-----------------|-----------|-------------|-------------|
| **Genome**               |                 |           |             |             |
| Danio rerio              | 49.76           | 9.10      | 3.08        | 1.94        |
| Xenopus laevis           | 26.69           | 2.85      | 0.13        | 0.02        |
| Alligator mississippiensis| 37.04           | 15.01     | 0.31        | 0.06        |
| Gallus gallus            | 8.31            | 11.33     | 0.30        | 0.05        |
| Anolis carolinensis      | 35.28           | 7.68      | 4.69        | 0.25        |
| Boa constrictor          | 28.80           | 14.53     | 0.43        | 0.03        |
| Python molurus           | 22.09           | 8.01      | 0.26        | 0.08        |
| Crotalus viridis         | 35.83           | 13.41     | 1.32        | 0.21        |
| Ornithorhynchus anatinus | 53.54           | 11.70     | 3.20        | 0.19        |
| Monodelphus domesticus   | 54.28           | 8.4       | 2.95        | 0.10        |
| Homo sapiens             | 49.90           | 5.71      | 1.68        | 0.03        |
| Mus musculus             | 41.89           | 7.85      | 5.53        | 0.24        |

**DNA transposons**
- **K2D < 2**
- **Testis**

**Total TE (%)**
- **Other**
- **ToMar**
- **hAT**
- **helitron**
- **Other**
- **ERV**
- **Gypsy**
- **Other**
- **CR1**
- **L2**
- **RTE**
- **L1**
- **Other**
- **Alu-SVA**
- **MIR**
