RESEARCH PAPER

Mobile Genetic Elements 2:1, 1–7; January/February 2012; © 2012 Landes Bioscience

Do human transposable element small RNAs serve primarily as genome defenders or genome regulators?

Kevin J. Lee,[1] Andrew B. Conley,[1] Victoria V. Lunyak[2] and I. King Jordan[1,3,*]

[1]School of Biology, Georgia Institute of Technology; Atlanta, GA, USA; [2]Buck Institute for Age Research; Novato, CA, USA; [3]PanAmerican Bioinformatics Institute; Santa Marta, Magdalena, Colombia

Keywords: gene expression, genome regulation, RNA processing, small RNA, RNA interference

Abbreviations: mRNA, messenger RNA; miRNA, microRNA; piRNA, PIWI interacting RNA; sRNA, small RNA; TE, transposable element; UTR, untranslated region

It is currently thought that small RNA (sRNA) based repression mechanisms are primarily employed to mitigate the mutagenic threat posed by the activity of transposable elements (TEs). This can be achieved by the sRNA guided processing of TE transcripts via Dicer-dependent (e.g., siRNA) or Dicer-independent (e.g., piRNA) mechanisms. For example, potentially active human L1 elements are silenced by mRNA cleavage induced by element encoded siRNAs, leading to a negative correlation between element mRNA and siRNA levels. On the other hand, there is emerging evidence that TE derived sRNAs can also be used to regulate the host genome. Here, we evaluated these two hypotheses for human TEs by comparing the levels of TE derived mRNA and TE sRNA across six tissues. The genome defense hypothesis predicts a negative correlation between TE mRNA and TE sRNA levels, whereas the genome regulatory hypothesis predicts a positive correlation. On average, TE mRNA and TE sRNA levels are positively correlated across human tissues. These correlations are higher than seen for human genes or for randomly permuted control data sets. Overall, Alu subfamilies show the highest positive correlations of element mRNA and sRNA levels across tissues, although a few of the youngest, and potentially most active, Alu subfamilies do show negative correlations. Thus, Alu derived sRNAs may be related to both genome regulation and genome defense. These results are inconsistent with a simple model whereby TE derived sRNAs reduce levels of standing TE mRNA via transcript cleavage, and suggest that human cells efficiently process TE transcripts into sRNA based on the available message levels. This may point to a widespread role for processed TE transcripts in genome regulation or to alternative roles of TE-to-sRNA processing including the mitigation of TE transcript cytotoxicity.

Introduction

Eukaryotic genomes harbor numerous transposable element (TE) sequences that are capable of moving from one location in the genome to another. This transpositional activity entails the genomic insertion of relatively large sequences and often leads to highly deleterious mutations. TE insertions can cause protein coding sequence mutations or premature termination of transcription in gene regions, can disrupt normal patterns of gene expression by targeting regulatory sequences and can lead to chromosomal breakage and re-arrangements.[1,2] Thus, TEs can be extremely mutagenic, and so genomes must have some way to control their activity.

A variety of transposition repression mechanisms have evolved to mitigate the threat that TEs pose to genome integrity.[3,4] These include DNA methylation,[1,5,6] repressive histone modifications,[7-13] the activity of cytosine deaminases and DNA repair proteins[14-16] and even the physical elimination of TE sequences from the genome.[17] In addition, results from recent studies are taken to point to a number of small RNA (sRNA) based mechanisms that may be are employed for the repression of TEs.[18] sRNAs refer to a number of different short RNA species processed from longer transcripts such as Dicer-dependent short interfering RNAs (siRNAs) or Dicer-independent PIWI-interacting RNAs (piRNAs). For example, the RNA interference (RNAi) pathway in *Caenorhabditis elegans* uses TE-derived sRNAs generated from double-stranded RNA (dsRNA) by Dicer to represses the transposition of DNA-type elements.[19] In Drosophila, piRNAs processed from TEs via a distinct 'ping-pong' amplification method are used to repress transposition in the germline thereby

*Correspondence to: I. King Jordan; Email: king.jordan@biology.gatech.edu
Submitted: 10/13/11; Revised: 12/12/11; Accepted: 12/13/11
http://dx.doi.org/10.4161/mge.2.1.19031

blocking the inheritance of TE-induced mutations and safeguarding development.[20-22] TE-derived sRNAs in mouse are used to repress the transcription of retrotransposons in oocytes.[23]

Close to 50% of the human genome sequence is derived from TEs.[24] While the vast majority of these elements are no longer capable of transposition, there remain a handful of active elements, LINE-1 (L1) and Alu sequences for the most part,[25] that pose a substantial mutagenic threat.[26] Work done on L1s provides the best characterized example of sRNA regulation for a human TE.[27] Full-length, potentially active L1 elements encode an antisense promoter in their 5' UTR.[28,29] Bi-directional transcriptional activity from both the canonical L1 sense promoter and the anti-sense promoter leads to the production of dsRNA, which is processed into L1-specific sRNAs.[27] These L1 sRNAs were shown to repress transposition by degrading full-length L1 mRNA transcripts. Thus, for human L1s an inverse correlation has been observed between the levels of L1 mRNA and element sRNA.

In light of this work on the sRNA regulation of human L1s, we hypothesized that if the predominant role of TE-derived sRNAs is to repress transposition by means of transcript cleavage, as the levels of TE-specific sRNA go up, there should be a concomitant decrease in TE mRNA levels genome-wide. If this is the case, we expect to observe a negative correlation between TE mRNA and TE sRNA levels. On the other hand, if TE generated sRNAs are primarily being utilized by the genomes in which they reside to facilitate the regulation of host genes, one may expect to see a positive correlation between levels of TE-derived mRNA and sRNA. This would suggest that TE-derived transcripts are efficiently processed by the host cellular machinery, based on available levels of RNA messages, in a way that does not reduce the overall efficacy of TE expression. Under this scenario, TEs would be dynamically regulated to express transcripts that are destined to be processed and function in sRNA based cellular regulatory pathways as opposed to simply serving as transposition intermediates.

Consistent with a potential role for TE transcripts in genome regulation, it has recently been shown that human TEs initiate transcription on a massive scale and are also dynamically regulated among different cell types; this includes the expression of numerous relatively ancient TEs that are no longer capable of transposing.[30] Furthermore, there are several recent examples illustrating that TE-derived sRNAs can in fact regulate host genes. In *Drosophila melanogaster*, TE-derived piRNAs play a critical role in embryonic patterning by targeting a specific host gene message.[31] piRNAs derived from the *roo* and *412* retrotransposons facilitate cleavage of the *nos* mRNA via interactions with its 3' UTR thereby establishing a posterior-to-anterior gradient that is critical for proper head and thorax segmentation. In the human genome, TE-derived miRNAs[32] have been shown to play diverse roles in cancer by regulating both tumor suppressor[33] and ocogenes.[34]

In attempt to distinguish between these two roles for TE-derived sRNAs in the human genome, namely whether TE sRNAs serve primarily as genome defenders or as genome regulators, we explored the relationship between levels of TE mRNA and TE sRNA across six tissues. We found that levels of TE-derived mRNA and sRNA are positively correlated across different tissues, with gene-rich Alu elements showing the strongest correlations. Despite previous work showing an inverse relationship between L1 element expression and the generation of sRNAs,[27] L1 mRNA levels were also positively correlated with levels of sRNA. These data are not consistent with the widespread cleavage of TE mRNA by TE sRNA, and raise the possibility that numerous TE-derived transcripts are processed to yield sRNAs that function to regulate the host genome.

Results

Mapping of human mRNA and sRNA sequence data. Levels of mRNA and sRNA were compared across human tissues for individual genes and TE subfamilies. To do this, we used publicly available paired sets of mRNA and sRNA data generated with high-throughput sequencing techniques from six human tissues: brain, heart, kidney, liver, lung and skeletal muscle (Supplementary **Table S1**). Sequence tags were mapped to the human genome reference sequence and co-located with genes and TEs as described in the Materials and Methods section. A recently developed algorithm for mapping ambiguous tags was used to ensure maximal coverage of repetitive TE sequences for the short sequence tags used.[35] This algorithm ensures that the best single genomic location for each multi-mapping tag is chosen, thus ensuring deeper coverage of TE sequences than would be achieved if multi-mapping tags were discarded. In addition, a series of quality controls designed for high-throughput sequence data were implemented to ensure the reliability of the sequences used (**Figs. S1–3**).

Results of the tag-to-genome mapping for the six human tissues analyzed here are shown in **Table 1**. There were ~26–134 million reads for the mRNA libraries and ~3–7 million reads for the sRNA libraries. After processing reads to eliminate adaptor sequences, sRNA sequences mapped to the human genome with extremely high fidelity. The majority of sRNA reads mapped to known miRNA loci, and ~1–2% mapped to TE sequences. mRNA reads mapped to the genome with lower fidelity, but a greater percentage mapped to TEs. The vast majority (90%) of sRNA sequence tags analyzed here were 19–24nt in length suggesting that they are miRNAs or endogenous siRNAs, as opposed to longer piRNAs, as can be expected since they were isolated from somatic tissue (**Fig. S4**). In mammalian genomes, small RNA based regulation of TEs is primarily attributed siRNAs as opposed to piRNAs, which appear to function in TE control exclusively in the male germline.[36]

Correlation of mRNA and sRNA levels for genes and TEs. For individual genes and individual TE subfamilies, mRNA vs. sRNA levels were regressed and the resulting correlation coefficients and slopes were determined (**Fig. 1B**). Regressing mRNA and sRNA levels across tissues in this way controls for any differences in the library preparations used prior to high-throughput sequencing since relative levels of expression are compared. The distributions of the correlation coefficients and slopes were then evaluated to determine the overall relationships between mRNA and sRNA levels across tissues for genes and TEs

Table 1. Results of the tag-to-genome mapping for mRNA and sRNA sequence libraries for six human tissues

	Reads per tissue	Reads after clipping	Reads that map to hg18	% reads mapped	Reads that map to TEs	% of mapping reads that map to TEs	Reads that map to genes	% of mapping reads that map to genes
mRNA								
brain	34,493,914	n/a	28,389,338	82.3	1,001,006	3.5	24,194,582	85.2
heart	40,338,602	n/a	32,751,816	81.2	571,069	1.7	26,665,851	81.4
kidney	83,696,940	n/a	42,051,713	50.2	3,828,411	9.1	33,587,016	79.9
liver	125,090,140	n/a	73,281,292	58.6	6,769,796	9.2	64,212,056	87.6
lung	25,862,057	n/a	19,808,655	76.6	3,138,208	15.8	16,434,340	83.0
muscle	45,280,908	n/a	36,984,450	81.7	919,399	2.5	32,413,952	87.6
sRNA								
brain	5,021,339	2,977,817	2,939,957	98.7	33,102	1.1	2,452,355	83.4
heart	5,901,910	4,937,144	4,921,992	99.7	42,284	0.9	4,701,738	95.5
kidney	2,869,903	2,135,001	2,108,413	98.8	23,959	1.1	1,720,229	81.6
liver	6,312,578	3,448,077	3,422,122	99.2	74,695	2.2	860,191	25.1
lung	7,294,106	4,808,564	4,709,583	97.9	62,764	1.3	3,652,715	77.6
muscle	3,793,410	3,537,750	3,532,680	99.9	38,019	1.1	3,458,249	97.9

(**Fig. 1B**). In particular, we sought to evaluate whether there was an overall negative or positive relationship between mRNA and sRNA levels for TE subfamilies in order to distinguish between the genome defense vs. genome regulator hypotheses for the primary role of human TE sRNAs.

The distribution of correlation coefficients for 760 human TE subfamilies is highly skewed toward the positive end with the peak value closest to a perfect correlation of 1 (**Fig. 2A**). The distribution is substantially different from a control distribution generated by randomly shuffling mRNA and sRNA vectors for TE subfamilies, which is far more bell shaped with a peak just below 0 (**Fig. 2A**). The distribution of correlation coefficients for genes is also skewed toward the positive end of the scale but the effect is far less pronounced than seen for TEs (**Fig. 2B**). TE subfamilies show a median mRNA vs. sRNA correlation coefficient of 0.62, which is significantly greater than seen for human genes or for the random control (TEs × genes $W = 2.6 \times 10^6$, $p < 10^{-10}$; TEs × control $W = 4.7 \times 10^5$, $p < 10^{-10}$). In other words, human TE mRNA and sRNA levels show a more consistently positive relationship than seen for genes or than can be expected by chance given the underlying data values being analyzed.

A similar set of patterns are observed when the distributions of the slopes of the linear regression lines are considered (**Fig. S5**). Although the shapes of the observed vs. random control distributions are more similar, the observed TE sloped distribution is shifted to the right indicating that mRNA vs. sRNA slopes are greater than would be expected by chance alone. The median TE slope value is also significantly higher than seen for genes or for the random control (TEs vs. genes $W = 3.3 \times 10^6$, $p = 9.1 \times 10^{-9}$; TEs vs. control $W = 4.5 \times 10^5$, $p < 10^{-10}$). Thus for human TEs, as mRNA levels increase, sRNA levels increase more precipitously than seen for human genes.

We also compared the correlation coefficient and slope distributions for the most abundant individual TE families or classes: LTR elements, DNA-type elements (i.e. cut-and-paste transposons), L1 and Alu. LTR, DNA and L1 groups all show similar median positive correlation coefficient values, whereas Alu has a significantly higher median value than the rest (**Fig. 3A**; Alu verus LTR $W = 8808$ $p = 0.01$). The pattern seen for the comparison of slopes is similar with Alus having an even more pronounced difference from the other TE families (**Fig. 3B**; Alu vs. L1 $W = 3369$ $p = 6.7 \times 10^{-11}$).

Discussion

Genome defense vs. genome regulation. sRNA regulatory pathways are thought to be critical for the control of TEs,[3,18] and accordingly TE-derived sRNAs have mainly been considered in light of this paradigm. In this report, we evaluated the relationship between levels of human TE mRNA and TE sRNA in attempt to try and discriminate between this classic view on the role of TE sRNAs and the alternative possibility that TE sRNAs play functional roles for the host, i.e. the genome defense vs. genome regulation hypotheses. To do this, we built upon the logic of previous studies of human TE silencing based on TE sRNAs. In the human genome, sRNAs were previously shown to defend the genome against transposition by repressing the expression of L1 TEs.[27] In this case, an increase in L1 generated sRNA levels led to a decrease in element mRNA levels via transcript cleavage. We sought to evaluate whether a similar inverse relationship between TE mRNA vs. sRNA levels could be seen across TE subfamilies genome-wide. On the contrary, we found that TE mRNA and sRNA levels are positively related (**Figs. 2 and S5**), consistent with a possible role for TE-derived sRNAs in genome regulation.

The higher average correlation coefficient and slope values seen for the relatively young Alu family of TEs (**Fig. 3**) was an

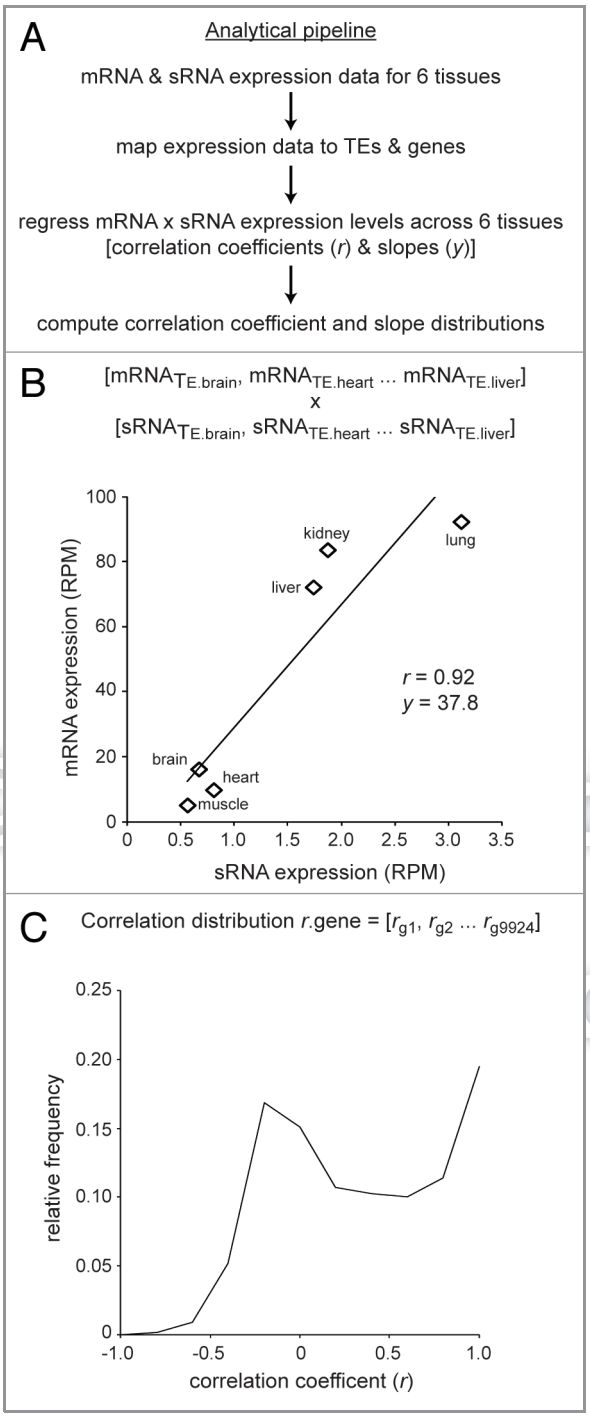

Figure 1. Scheme of the analytical pipeline and tools presented herein. (A) Analytical pipeline overview. (B) Example of the linear regression and correlation analysis used to compare mRNA vs. sRNA levels for individual TE subfamilies and genes across six human tissues. (C) Example of the distribution of the resulting correlation coefficients for all genes.

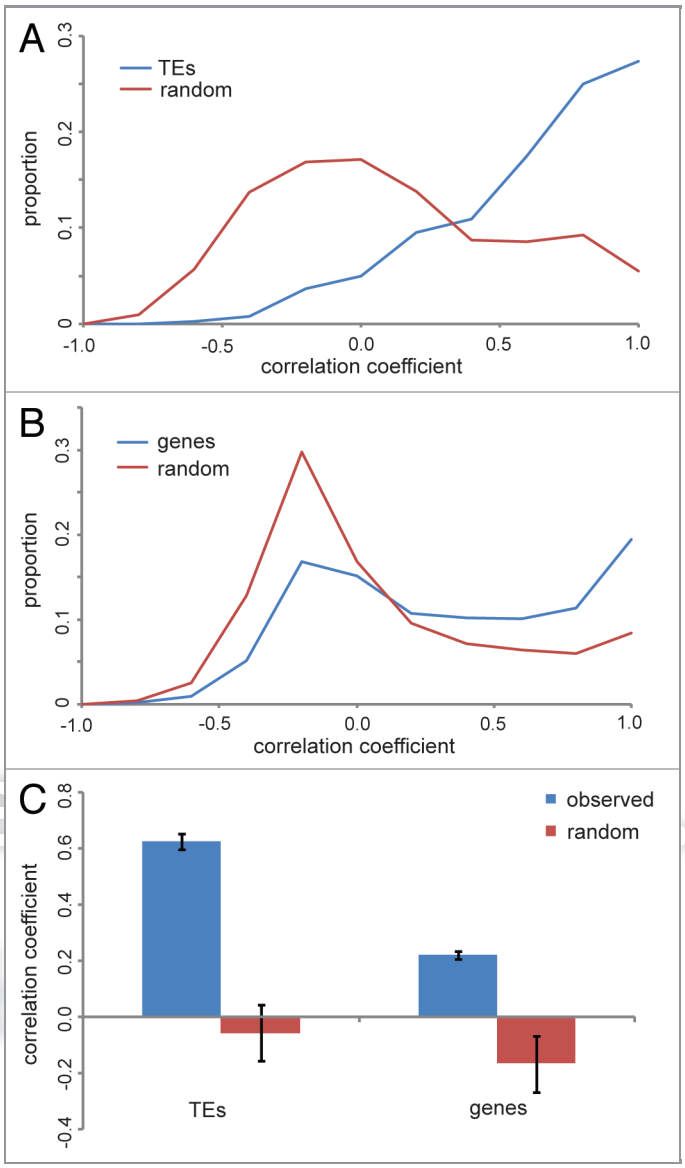

Figure 2. mRNA vs. sRNA correlation coefficient distributions for human TE subfamilies and genes across six tissues. (A) Observed (blue) and randomized (red) correlation coefficient distributions for TE subfamilies. (B) Observed (blue) and randomized (red) correlation coefficient distributions for genes. (C) Correlation coefficient median ± standard error values for TE subfamily and gene observed (blue) vs. random (red) distributions.

unexpected observation. If TE-derived sRNAs are being used primarily to degrade mRNA transcripts in order to defend the genome against transposition, one may expect that the youngest and most potentially active TE subfamilies would show the most pronounced negative correlation between mRNA and sRNA levels. Similarly, if older elements that are no longer capable of transposing have been domesticated to transcribe RNAs with functional utility for the host, then those element families should show higher mRNA-to-sRNA positive correlations. This was clearly not the case here. However, when individual Alu element subfamilies were considered separately younger AluY subfamilies did show some evidence for genome defense by virtue of having negative TE mRNA-to-sRNA correlations; in fact, AluY subfamilies were the only ones to show such negative correlations. For example, the youngest AluY subfamily, AluYb with an estimated age of 1.9 my, has a TE mRNA-to-siRNA correlation of

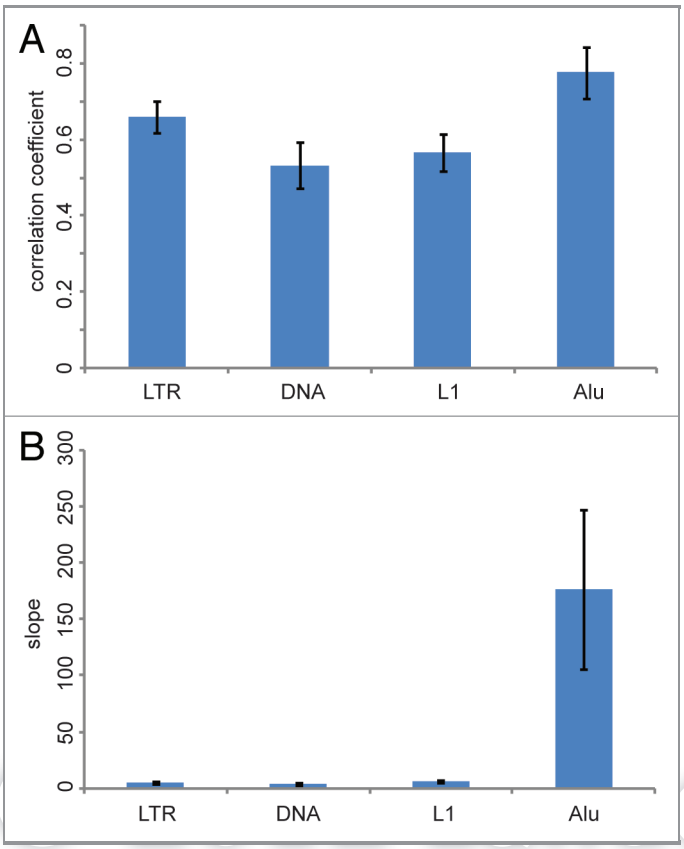

Figure 3. Median ± standard error values for the (A) correlation coefficient and (B) slope distributions for individual TE family (classes).

r = -0.30. Furthermore, when the relative ages of all Alu subfamilies are considered with respect to their TE mRNA-to-sRNA correlations, younger families overall show lower correlation values (Alu subfamily age vs. TE mRNA-to-siRNA correlation r = 0.43, t = 2.7, p = 0.01). Thus, for Alus there is evidence in favor of both genome defense and genome regulation hypotheses with respect to the roles of TE sRNA. These results are consistent with a variety of roles in genome regulation and organization that have been ascribed to Alu element sequences and transcripts.[37-39] L1 subfamilies, on the other hand, do not show any evidence for genome defense when analyzed in a similar way.

Our results showing a positive correlation between TE mRNA and TE sRNA levels are consistent with two recent observations that also suggest that TE sRNAs should be considered with respect to possible roles that they may play in genome regulation. First of all, TEs were shown to be highly transcribed and dynamically regulated in the human and mouse genomes.[30] This includes numerous ancient TEs that are no longer capable of transposition and thus would not need to be repressed by their host genome. Second, it has recently been shown that TE-derived sRNAs can directly interact with host genes to regulate their expression. This has been seen for TE-derived piRNAs in Drosophila[31] and for TE-derived miRNAs in human.[32-34]

We would like to emphasize that the correlations observed here do not equal causation. Rather, the results we obtained point to the possibility that TE-derived sRNAs play some role in genome regulation. Nevertheless, we feel that the data reported here represent an important and worthwhile observation in light of the emphasis currently placed on sRNA based TE repression mechanisms.

Alternative roles for TE transcript processing. TE transcript processing by enzymes such as Dicer is typically thought to be related to the repression of transposition. However, it may also be possible that TE transcripts need to be efficiently processed to mitigate some other non-transposition related threats that they pose to the cells. In other words, accumulation of the TE transcripts themselves, or simply dysregulation of the TEs, may be toxic to the cellular environment and cells may efficiently process TE transcripts to mitigate this toxicity. For example, accumulation of unprocessed Alu transcripts based on Dicer deficiency has been linked to age-related macular degeneration in humans.[40] Dysregulated Alu transcription has also been related to the senescence of adult human stem cells, and sRNA based silencing of Alu transcription restores the self-renewing phenotype of these cells.[41] If organisms have evolved efficient mechanisms that process TE transcripts to mitigate their toxicity, one might also expect to see the kinds of positive correlations between TE mRNA and sRNA levels reported here across cellular phenotypes.

It may also be the case that sRNA based cleavage of TE transcripts for the purposes of repression of transposition does not necessarily lead to the predicted negative correlations between sRNA and mRNA levels. sRNA based silencing mechanisms are used to repress TE expression and transposition in *Arabidopsis thaliana* gametes. TEs are expressed in the vegetative nucleus cells of *A. thaliana* pollen but not in the sperm cells that pass on genetic material to successive generations.[42] Apparently, the TEs that are expressed in the vegetative nucleus are efficiently processed to yield sRNAs in accordance with the availability of full-length TE messages. In this case, it was proposed that TE activation in the vegetative nucleus may be used to provide sRNAs that are passed to the sperm cells to repress transposition therein. In other words, the repression mechanism is indirect in the sense that TEs from one nucleus are activate to provide sRNAs for TE silencing in another nucleus. This kind of mechanism could lead to positive correlations between TE mRNA and TE sRNA levels across cellular compartments with TE derived sRNAs exerting their repressive effects elsewhere in the organism.

Finally, it is worth noting that the two possible roles for TE-derived sRNAs are not mutually exclusive. It is clearly a fact that TE sRNAs are used to repress transposition, but it is becoming increasingly evident that TEs are widely expressed and dynamically regulated to yield non-coding RNAs, which in turn can be efficiently processed into sRNAs that interact with host genes to affect their regulation. The genome-scale results reported here suggest that the second view warrants serious consideration and raise the possibility that sRNA based mechanisms may have initially evolved to repress transposition but now serve primarily in genome regulation.

Materials and Methods

RNA sequence data and mapping. The levels of mRNA and sRNA for human TEs and genes analyzed in this study are based on a series of previous RNA-seq studies for full-length transcripts[43-45] and short RNAs[46] (**Table S1**), and the mRNA and sRNA sequence read data from these studies were obtained from the NCBI Sequence Read Archive (SRA - http://www.ncbi.nlm.nih.gov/sra). mRNA and sRNA data were analyzed from six human tissues: brain, hear, liver, lung, kidney and skeletal muscle. All RNA sequences analyzed here were characterized using the Illumina platform under the conditions described in **Table S1**. mRNA sequences were isolated from total RNA using oligo-T magnetic beads, and sRNA sequences were isolated from total RNA using 18–35 nt size fractionation.

Quality control analysis of RNA sequence data was done using the FastQC program (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/), and only tags within the expected size range (19–24 nt) for miRNA or siRNA were considered for subsequent analysis. RNA sequence reads were mapped to the human genome reference sequence (NCBI36/hg18) using the program Bowtie[47] with a threshold of ≤ 2 mismatches allowed. The most likely mapping locations for reads that mapped to more than one location were rescued using the Gibbs sampling strategy for multi-mapping tags.[35] mRNA and sRNA sequence tags mapped and processed in this way were co-located with human gene and TE loci annotated in the UCSC Genome Browser.[48] The locations of human genes were taken from the Known Genes track[49] and the locations of human TEs, along with their class/family/subfamily designations, were taken from the RepeatMasker track.[50]

Statistical analysis. For each TE subfamily and each gene locus, tissue-specific reads per million (RPM) counts were computed for mRNA and sRNA. Then for each TE subfamily (n = 903) and each gene (n = 25,246), least squares linear regression was used to compare mRNA vs. sRNA levels across the six tissues, and the correlation coefficient and slope values were determined. A matched series of random correlation coefficients and slopes were calculated by randomly shuffling the underlying tissue-specific mRNA and sRNA RPM counts for each TE subfamily and each gene and performing the same linear regression analysis. Median values for the distributions of the correlation coefficient and slope values were compared using the Wilcoxon rank sum test.

Acknowledgments

K.J.L., A.B.C and I.K.J. were supported by the School of Biology, Georgia Institute of Technology. V.V.L. was supported by the National Institutes of Health pilot projects on UL1 DE019608 and the Buck Institute Trust Fund.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Note

Supplementary materials can be found at:
www.landesbioscience.com/journals/mge/article/19031/

References

1. Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 1997; 94:7704-11; PMID:9223252; http://dx.doi.org/10.1073/pnas.94.15.7704
2. McClintock B. Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol 1951; 16:13-47; PMID:14942727
3. Levin HL, Moran JV. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 2011; 12:615-27; PMID:21850042; http://dx.doi.org/10.1038/nrg3030
4. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 2007; 8:272-85; PMID:17363976; http://dx.doi.org/10.1038/nrg2072
5. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem 2005; 74:481-514; PMID:15952895; http://dx.doi.org/10.1146/annurev.biochem.74.010904.153721
6. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 1997; 13:335-40; PMID:9260521; http://dx.doi.org/10.1016/S0168-9525(97)01181-5
7. Gendrel AV, Lippman Z, Yordan C, Colot V, Martienssen RA. Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science 2002; 297:1871-3; PMID:12077425; http://dx.doi.org/10.1126/science.1074950
8. Kondo Y, Issa JP. Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. J Biol Chem 2003; 278:27658-62; PMID:12724318; http://dx.doi.org/10.1074/jbc.M304072200
9. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, et al. Role of transposable elements in heterochromatin and epigenetic control. Nature 2004; 430:471-6; PMID:15269773; http://dx.doi.org/10.1038/nature02651
10. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J 2005; 24:800-12; PMID:15678104; http://dx.doi.org/10.1038/sj.emboj.7600545
11. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007; 448:553-60; PMID:17603471; http://dx.doi.org/10.1038/nature06008
12. Pauler FM, Sloane MA, Huang R, Regha K, Koerner MV, Tamir I, et al. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res 2008; 19:221-33; PMID:19047520; http://dx.doi.org/10.1101/gr.080861.108
13. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 2002; 297:1833-7; PMID:12193640; http://dx.doi.org/10.1126/science.1074973
14. Chiu YL, Greene WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol 2008; 26:317-53; PMID:18304004; http://dx.doi.org/10.1146/annurev.immunol.26.021607.090350
15. Crow YJ, Hayward BE, Parmar R, Robins P, Leitch A, Ali M, et al. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. Nat Genet 2006; 38:917-20; PMID:16845398; http://dx.doi.org/10.1038/ng1845
16. Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell 2008; 134:587-98; PMID:18724932; http://dx.doi.org/10.1016/j.cell.2008.06.032
17. Yao MC, Chao JL. RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. Annu Rev Genet 2005; 39:537-59; PMID:16285871; http://dx.doi.org/10.1146/annurev.genet.39.073003.095906
18. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. Cell 2009; 136:656-68; PMID:19239887; http://dx.doi.org/10.1016/j.cell.2009.01.045
19. Vastenhouw NL, Plasterk RH. RNAi protects the Caenorhabditis elegans germline against transposition. Trends Genet 2004; 20:314-9; PMID:15219396; http://dx.doi.org/10.1016/j.tig.2004.04.011
20. Juliano C, Wang J, Lin H. Uniting Germline and Stem Cells: The Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms. Annu Rev Genet 2011; 45:447-69; PMID:21942366; http://dx.doi.org/10.1146/annurev-genet-110410-132541
21. Khurana JS, Theurkauf W. piRNAs, transposon silencing, and Drosophila germline development. J Cell Biol 2010; 191:905-13; PMID:21115802; http://dx.doi.org/10.1083/jcb.201006034

22. Simonelig M. Developmental functions of piRNAs and transposable elements: A Drosophila point-of-view. RNA Biol 2011; 8:754-9; PMID:21712652; http://dx.doi.org/10.4161/rna.8.5.16042
23. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 2008; 453:539-43; PMID:18404146; http://dx.doi.org/10.1038/nature06908
24. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921; PMID:11237011; http://dx.doi.org/10.1038/35057062
25. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet 2007; 23:183-91; PMID:17331616; http://dx.doi.org/10.1016/j.tig.2007.02.006
26. Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. Hum Mutat 2005; 25:207-21; PMID:15643617; http://dx.doi.org/10.1002/humu.20133
27. Yang N, Kazazian HH, Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol 2006; 13:763-71; PMID:16936727; http://dx.doi.org/10.1038/nsmb1141
28. Nigumann P, Redik K, Matlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 2002; 79:628-34; PMID:11991712; http://dx.doi.org/10.1006/geno.2002.6758
29. Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 2001; 21:1973-85; PMID:11238933; http://dx.doi.org/10.1128/MCB.21.6.1973-1985.2001
30. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 2009; 41:563-71; PMID:19377475; http://dx.doi.org/10.1038/ng.368
31. Rouget C, Papin C, Boureux A, Meunier AC, Franco B, Robine N, et al. Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature 2010; 467:1128-32; PMID:20953170; http://dx.doi.org/10.1038/nature09465
32. Piriyapongsa J, Marino-Ramirez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. Genetics 2006; 176:1323-37; PMID:17435244; http://dx.doi.org/10.1534/genetics.107.072553
33. Lee DY, Deng Z, Wang CH, Yang BB. MicroRNA-378 promotes cell survival, tumor growth, and angiogenesis by targeting SuFu and Fus-1 expression. Proc Natl Acad Sci USA 2007; 104:20350-5; PMID:18077375; http://dx.doi.org/10.1073/pnas.0706901104
34. Meng F, Wehbe-Janek H, Henson R, Smith H, Patel T. Epigenetic regulation of microRNA-370 by interleukin-6 in malignant human cholangiocytes. Oncogene 2008; 27:378-86; PMID:17621267; http://dx.doi.org/10.1038/sj.onc.1210648
35. Wang J, Huda A, Lunyak VV, Jordan IK. A Gibbs sampling strategy applied to the mapping of ambiguous short-sequence tags. Bioinformatics 2010; 26:2501-8; PMID:20871106; http://dx.doi.org/10.1093/bioinformatics/btq460
36. Okamura K, Lai EC. Endogenous small interfering RNAs in animals. Nat Rev Mol Cell Biol 2008; 9:673-8; PMID:18719707; http://dx.doi.org/10.1038/nrm2479
37. Bettecken T, Frenkel ZM, Trifonov EN. Human nucleosomes: special role of CG dinucleotides and Alu-nucleosomes. BMC Genomics 2011; 12:273; PMID:21627783; http://dx.doi.org/10.1186/1471-2164-12-273
38. Pandey R, Mukerji M. From 'JUNK' to just unexplored noncoding knowledge: the case of transcribed Alus. Brief Funct Genomics 2011; 10:294-311; PMID:21987713; http://dx.doi.org/10.1093/bfgp/elr029
39. Ponicsan SL, Kugel JF, Goodrich JA. Genomic gems: SINE RNAs regulate mRNA production. Curr Opin Genet Dev 2010; 20:149-55; PMID:20176473; http://dx.doi.org/10.1016/j.gde.2010.01.004
40. Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, et al. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. Nature 2011; 471:325-30; PMID:21297615; http://dx.doi.org/10.1038/nature09830
41. Wang J, Geesman GJ, Hostikka SL, Atallah M, Blackwell B, Lee E, et al. Inhibition of activated pericentromeric SINE/Alu repeat transcription in senescent human adult stem cells reinstates self-renewal. Cell Cycle 2011; 10:3016-30; PMID:21862875; http://dx.doi.org/10.4161/cc.10.17.17543
42. Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 2009; 136:461-72; PMID:19203581; http://dx.doi.org/10.1016/j.cell.2008.12.038
43. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature 2008; 456:470-6; PMID:18978772; http://dx.doi.org/10.1038/nature07509
44. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 2008; 18:1509-17; PMID:18550803; http://dx.doi.org/10.1101/gr.079558.108
45. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 2008; 40:1413-5; PMID:18978789; http://dx.doi.org/10.1038/ng.259
46. Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol 2010; 11:R56; PMID:20507594; http://dx.doi.org/10.1186/gb-2010-11-5-r56
47. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10:R25; PMID:19261174; http://dx.doi.org/10.1186/gb-2009-10-3-r25
48. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002; 12:996-1006; PMID:12045153
49. Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC Known Genes. Bioinformatics 2006; 22:1036-46; PMID:16500937; http://dx.doi.org/10.1093/bioinformatics/btl048
50. Smit AF, Hubley R, Green P. RepeatMasker Open-3.0. 1996-2007.