The proteins CMT3, UHRF1 and ESP1 repress select transposable elements in the model plant *Arabidopsis thaliana*

Brittany Cox[1], Majeda Al Rawahneh*, Jacob Anderson*, Corey Berta*, Evan Boucher*, Ryan Bridenstine*, Kimberlee Eakins*, Rachel Embly*, Candace Flynn*, Andrew Francis*, Jenna Gallagher*, Brian Jones*, Clint Miller*, Forrest Payne*, Sylvia Phimsouay*, Nandita Rajshekhar*, Robert Rosia*, Scott Spencer*, Kaitlyn Stevelak*, David Sun*, Valerie Childress[2], Saivageethi Nuthikattu[3], Kaushik Panda[3], R. Keith Slotkin [4,5]*

* Undergraduate Student enrolled in the Autumn 2013 Department of Molecular Genetics Laboratory course MG5601
[1] Primary Author of the Report
[2] Non-degree Graduate Student enrolled in the Autumn 2013 Department of Molecular Genetics Laboratory course MG5601
[3] Graduate Teaching Assistant in the Autumn 2013 Department of Molecular Genetics Laboratory course MG5601
[4] Professor of the Autumn 2013 Department of Molecular Genetics Laboratory course MG5601
[5] Department of Molecular Genetics, 500 Aronoff Laboratory, 318 West 12th Ave., Columbus, OH, 43210.

ABSTRACT Transposable elements (TEs) are mobile fragments of DNA that occupy large portions of eukaryotic genomes and can harmfully mutate the genome when they are active. To suppress TE activity, eukaryotic cells repress TE transcription using multiple elaborate mechanisms that are not well understood. The Autumn 2013 Molecular Genetics 5601 Laboratory course aimed to discover specific proteins in the plant *Arabidopsis thaliana* which act to maintain the epigenetic silencing of multiple TEs. In half of a semester we successfully genotyped 17 homozygous mutants, measured the expression of seven different TEs by qRT-PCR in these confirmed mutant lines, and additionally investigated the DNA methylation levels of specific mutant-TE combinations by McrBC-PCR. From our analysis we identified three proteins that function to suppress TE transcription. We unexpectedly found that the putative mRNA-binding protein ESP1 functions to maintain the silencing of particular select types of TEs, and the DNA methylation mutants cmt3 and uhrf1 have only minor reactivation of specific TEs, not a genome-wide global TE activation as previously suggested. Our data demonstrates that an undergraduate laboratory course can successfully identify new proteins involved in the repression of TEs while learning the fundamentals of molecular genetics and experimental biological research.

INTRODUCTION

Transposable elements (TEs) are “jumping” DNA fragments that can excise and reinsert themselves, or insert copies of themselves, into regions of the genome where they had not been previously located. TEs occupy half of the human genome (de Koning, Gu, Castoe, Batzer, & Pollock, 2011), and have been particularly active in the primate lineage (Cordaux & Batzer, 2009). When expressed and actively duplicating, TEs are highly mutagenic, as TEs can insert into genes (disrupting their protein-coding function) or generate breaks in chromosomes. TE duplication is closely associated with the progression of cancer (Lee et al., 2012) and thus repression of TE activity is critically important to the stability of the genome.

Although much of the genome of any given eukaryote is composed of TEs, in a healthy cell TEs are not mutagenic because they are successfully repressed. Cells spend considerable energy to repress TEs, and once initially repressed the cell has mechanisms to keep the TE heritably or “epigenetically” silenced indefinitely. Epigenetics is the study of heritable gene expression not involving alteration of the DNA sequence, but rather the repression of transcription, and the inheritance of that suppressed or silenced state. Epigenetic regulation can be passed from one cell to another, or from parent to offspring generations. Through epigenetic regulation the cell maintains TEs in a silenced state over long

evolutionary timespans. DNA methylation is a large focal point of epigenetic studies, as DNA methylation patterns can be inherited from parent to daughter cells and highly methylated regions such as TEs are targeted for transcriptional silencing. TEs are the major targets of DNA methylation, and it has been hypothesized that DNA methylation and epigenetic regulation evolved to suppress TEs (Cerutti & Casas-Mollano, 2006).

*Arabidopsis thalian*a is a small flowering plant that is used as a model organism in a variety of biomedical and agricultural research studies. Although not as complex as animals, this plant provides insight into the biological and genetic functions of vertebrates. Its benefits as a model organism include its short lifecycle, prolific seed production, easy transgene transformation, and available mutants in nearly all genes. For our purposes Arabidopsis can provide specific information about the genes and mechanisms responsible for the epigenetic suppression of TEs. Unraveling the genes and mechanisms responsible for the epigenetic suppression of TEs will proceed much faster in the model plant Arabidopsis, and this information can then be applied to studies of TEs and epigenetic regulation in vertebrate and cancer systems.

The goal of our project is to identify the genes involved in maintaining the silenced state of TEs in Arabidopsis. Many genes are already known to maintain the silencing of TEs in both Arabidopsis and vertebrates (Slotkin & Martienssen, 2007), and recent data suggests that RNA processing factors may play a role in this silencing (Dou et al., 2013). To investigate if RNA processing factors as well as other pathways and mechanisms play a role in the silencing of Arabidopsis TEs, we will identify and examine confirmed homozygous mutants in several different pathways to determine if they have a loss of TE epigenetic transcriptional silencing and express TEs. Quantitative Reverse-Transcription PCR (qRT-PCR) will be used to amplify and quantify TE expression in each mutant. This data will be compared to control reference lines with known active or silenced TEs. Finally, DNA from those mutants with significant TE expression will be assayed to determine if the reactivated TEs display loss of DNA methylation. By identifying mutants that reactivate TE expression, and determining if TE methylation is lost in these mutants, we will determine what molecular mechanisms are responsible for the epigenetic silencing of TEs.

METHODS

Seeds of 21 putative *Arabidopsis thaliana* transgene (T-DNA) insertion mutants were ordered from the OSU Arabidopsis Biological Resource Center and grown on soil. These mutants each have a T-DNA insertion into a gene, interrupting and mutating that gene. Control plants of wild-type Columbia (wt Col) and ddm1 mutants were also grown. Mutants in the DDM1 gene have known global activation of TEs (Lippman et al., 2004), while wt Col plants have transcriptionally silenced TEs. Leaves were collected from juvenile Arabidopsis plants. To confirm homozygous T-DNA mutations, DNA from 4 individuals of each mutant line was extracted and genotyped by PCR. DNA extractions were performed as fractional precipitations of DNA from ground leaves, and the genotyping was performed with two PCR reactions: the first to identify the presence or absence of the wild-type (wt) allele, and the second to confirm the presence or absence of the T-DNA insertion. PCR products were verified using gel electrophoresis and individual homozygous mutant plants that contain only T-DNA insertion alleles (and not wt alleles) for each mutant line were investigated further.

Confirmed homozygous mutant plants advanced to the RNA extraction phase, in which inflorescence (flower bud) tissue from each plant was collected for analysis. RNA was extracted using a fractional precipitation from ground tissue and all RNA samples were quantified via Nanodrop and normalized to 1ug of total RNA. Each RNA sample was treated with DNase enzyme via the "DNase Free Kit" (Life Technologies) and cDNA was generated from the RNA using an oligo-dT primer and Tetro Reverse Transcriptase (Bioline, Inc.). To assess cDNA quality, we amplified an intron-spanning PCR product followed by gel electrophoresis to monitor the splicing of a control gene At2g20610. Only cDNA that was free of contaminating genomic DNA and 100% spliced was used for further analysis.

Each sample of confirmed cDNA proceeded to qRT-PCR analysis. Using Sensi-Mix qPCR Mix (Bioline), we measured the expression level of the Athila6 TE gag/pol protein coding region. In addition, for some mutant plants we measured the expression of six additional TEs from the same cDNA. For each qRT-PCR reaction, we used three technical replicates of 20uL reactions and 1-3 biological replicates, depending on how many plants genotyped as homozygous mutants (Table 1). Relative expression values of TEs were determined

Mutant name	Mutant gene number	Mutant allele	Number verified homozygous mutant plants	Number of successful RNA isolations	Number of successful cDNA synthesis reactions
esp1	AT1G73840	*SALK_078793C*	2	2	2
swinger	AT4G02020	*SALK_109121C*	4	3	2
rrp6l3	AT2G32415	*SALK_122492C*	3	1	1
uhrf1/vim1	AT1G57820	*SALK_050903C*	4	3	3
mos4	AT3G18165	*SALK_019535C*	1	1	1
drd1	AT2G16390	*CS9948*	0	N/A	N/A
rrp6l2	AT5G35910	*SALK_113786C*	3	2	2
pol4/pol5	AT1G63020 / AT2G40030	*SALK_128428 & SALK_017795*	0	N/A	N/A
ago6	AT2G32940	*SALK_031553*	3	1	1
shh1	AT1G15215	*SALK_074540C*	0	N/A	N/A
wrky6	AT1G62300	*SALK_012997C*	4	2	2
cmt2	AT4G19020	*SALK_012874*	3	3	3
cmt3	AT1G69770	*SALK_148381*	2	2	2
h1.1	AT1G06760	*SALK_128430*	3	2	2
sde3	AT1G05460	*SALK_092019*	3	2	2
at1g77270	AT1G77270	*SALK_070487C*	4	3	3
at1g21560	AT1G21560	*SALK_141662C*	3	2	2
nerd	AT2G16485	*SALK_093814*	4	3	3
ago2	AT1G31280	*SALK_003380*	4	3	3
hog1	AT4G13940	*SALK_023915C*	0	N/A	N/A
hst1	AT3G05040	*SALK_006481C*	4	1	1
Total			54	36	35

Table 1. Mutants analyzed.

using the control gene At1g08200 as a reference housekeeping gene.

For the analysis of DNA methylation levels, 500ng of purified DNA from each individual was digested with the McrBC enzyme (New England Biolabs) at 37C for 2 hours, and the enzyme was inactivated at 65C for 20 minutes. The McrBC enzyme digests methylated DNA, reducing the amount of PCR template if the input DNA is methylated. The same procedure was performed for mock reactions of each sample, using water in place of the McrBC enzyme. This was followed by PCR amplification of +McrBC and mock (-McrBC) reactions for the TE promoter regions of interest, as well as an unmethylated control gene (At2g20610, TyrAT) that acts as a loading control and to ensure enzyme specificity by excluding the possibility of random DNA degradation. Samples were PCR amplified and subjected to gel electrophoresis to allow for semi-quantitative measurement of methylation levels for each mutant individual at each TE promoter analyzed.

RESULTS

Confirmation of mutant lines and successful cDNA synthesis

To investigate the genes and mechanisms responsible for the epigenetic silencing of Arabidopsis TEs, we grew and genotyped 84 plants representing 21 different mutant lines, which were chosen because they were either mutants for pathways that had not previously been investigated for TE activity, or had conflicting previous published data on TE activity. We genotyped these 21 lines and identified 54 plants from 17 mutant lines that were verified as homozygous for the mutation (Table 1). From the verified homozygous mutant individuals, we successfully isolated RNA from 36 plants representing 1-3 biological replicates of the 17 mutant lines. We successfully generated cDNA that was free of genomic DNA contamination for 35 individual plants representing all 17 homozygous lines (Table 1). In addition to these mutant lines, we also produced cDNA from wt Col and ddm1 control plants.

Analysis of TE expression levels

To determine which mutants have TE expression, we used all 35 successful cDNA synthesis reactions in a qRT-PCR survey of the expression of the Athila6 TE, using PCR primers to the gag/pol region of this TE. The Athila6 TE has two protein-coding regions that are differentially regulated: gag/pol and env. Figure 1 shows the relative expression of all 17 mutant homozygous lines compared to wt Col and ddm1 controls. Our qRT-PCR surveys showed that most mutants had Athila6 gag/pol relative expression at similar levels (or below) the levels of wt Col (Figure 1), demonstrating that these mutants do not have TE activation. The positive control ddm1 had very high expression (Figure 1). Importantly, the mutants esp1, cmt3 and uhrf1 displayed expression levels that were more than twice as high as the reference wt Col. We chose these three mutants to continue further analysis.

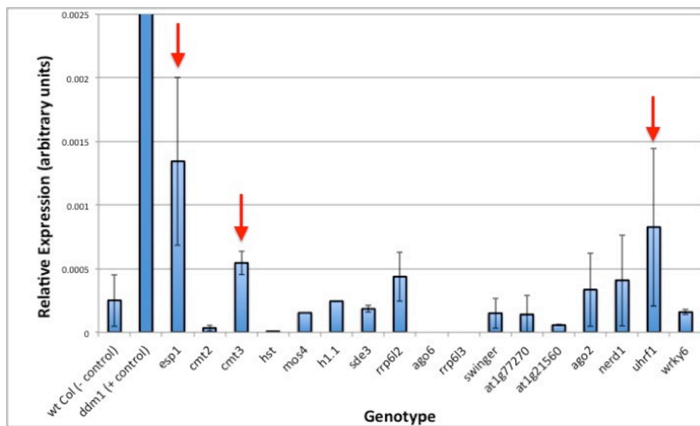

Figure 1. Athila6 gag/pol TE relative expression in 17 confirmed homozygous mutant lines as determined by qRT-PCR. The qRT-PCR mRNA expression analysis of the gag/pol protein-coding region of the Athila6 TE in 17 different mutants and two control lines. The ddm1 mutant is a positive TE expression control and has an expression level that is higher than the top of the Y-axis (0.032). The wt Col control is the wt baseline level for the epigenetically silenced TE. The red dotted line represents a 2-fold increase in expression compared to wt Col. Three mutant lines have expression above this 2-fold increase line: esp1, cmt3 and uhrf1. These three mutant values are shown with red arrows.

For the three mutants with elevated levels of Athila6 gag/pol expression, we aimed to determine if these mutants have a limited and selected reactivation of just some TEs, or have a global TE activation such as the control mutant ddm1 (Lippman et al., 2004). We performed additional qRT-PCR on the same cDNA as in Figure 1 with PCR primers for six different TEs. Figure 2 shows the relative expression of the mutant lines along with the wt Col and ddm1 controls for the Athila6 TE env region (Figure 2A), the AtMu1 TE (Figure 2B), the AtENSPM6 TE (Figure 2C), the AtGP1 TE (Figure 2D), the AtLINE1 TE (Figure 2E) and the AtCopia52 TE (Figure 2F). We found that the cmt3 mutant has reactivation of many different types of TEs, including Athila6 env region, AtMu1, AtLINE1, and AtCopia52 (Figure 2). In cmt3 mutants the AtCopia52 TE is activated to the same expression level as the positive control ddm1 mutant, while the other TEs showed a lower level of expression intermediate to wt Col and ddm1 (Figure 2). However, not all TEs are activated in cmt3 mutants, as the AtENSPM6 and AtGP1 TEs showed no evidence of reactivation, demonstrating that cmt3 has widespread activation of TEs, but not complete global TE reactivation like the ddm1 mutant.

The uhrf1 mutant displayed reactivation of select TEs (Figure 2). The AtENSPM6 and AtLINE1 TEs are more highly expressed in uhrf1 mutants than the ref-

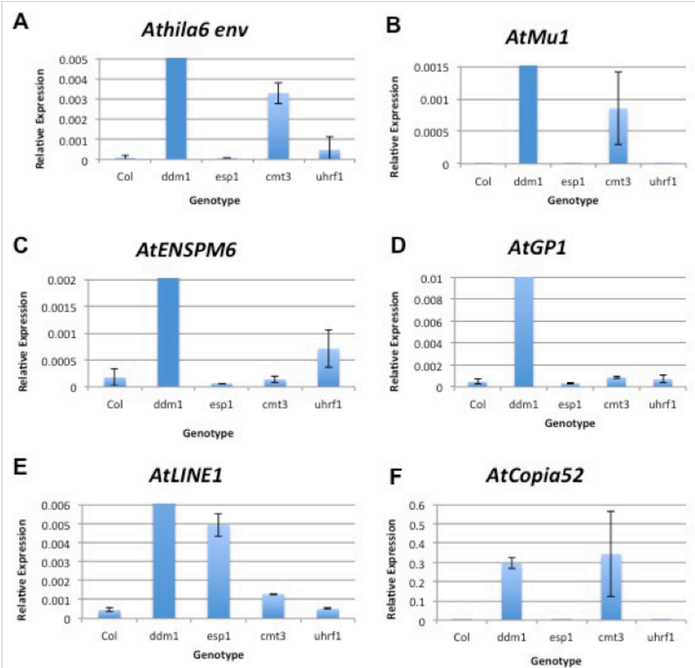

Figure 2. Relative expression of various TEs as determined by qRT-PCR. The three mutant lines that had more than a 2-fold increase in Athila6 gag/pol TE expression (from Figure 1) were tested for expression of six additional TEs (parts A-F). The ddm1 mutant is a positive TE expression control and has an expression level that is higher than the top of the Y-axis. The wt Col control is the wt baseline level for the epigenetically silenced TE.

erence wt Col, but less than the positive control ddm1. The other TEs tested (AtMu1, AtGP1 and AtCopia52) were not activated in uhrf1 mutants, while the Athila6 env TE region showed an increase, however the experimental standard deviation was too high to accurately determine if this TE is reactivated in this mutant line. Thus, we conclude that the uhrf1 mutant results in the reactivation of only specific TEs.

Lastly, the esp1 mutant displayed TE activation of only two TEs, the Athila6 gag/pol region assayed in Figure 1 and the AtLINE1 element in Figure 2E. For both of these TEs, the expression level is substantially lower than the maximum reactivation potential level, which is shown in the ddm1 mutant. Therefore, we conclude that esp1 has reactivation of select TEs.

Analysis of TE DNA methylation levels

We next aimed to determine if the TE activation observed in Figures 1 and 2 were due to a loss in DNA methylation and thus loss of epigenetic silencing of the TE. We digested DNA with the McrBC enzyme, which degrades methylated DNA. We then PCR amplified the promoter regulatory region of the Athila6 gag/pol region (which is called the Athila6 LTR) from the digested (or mock digested) DNA. We again compared the

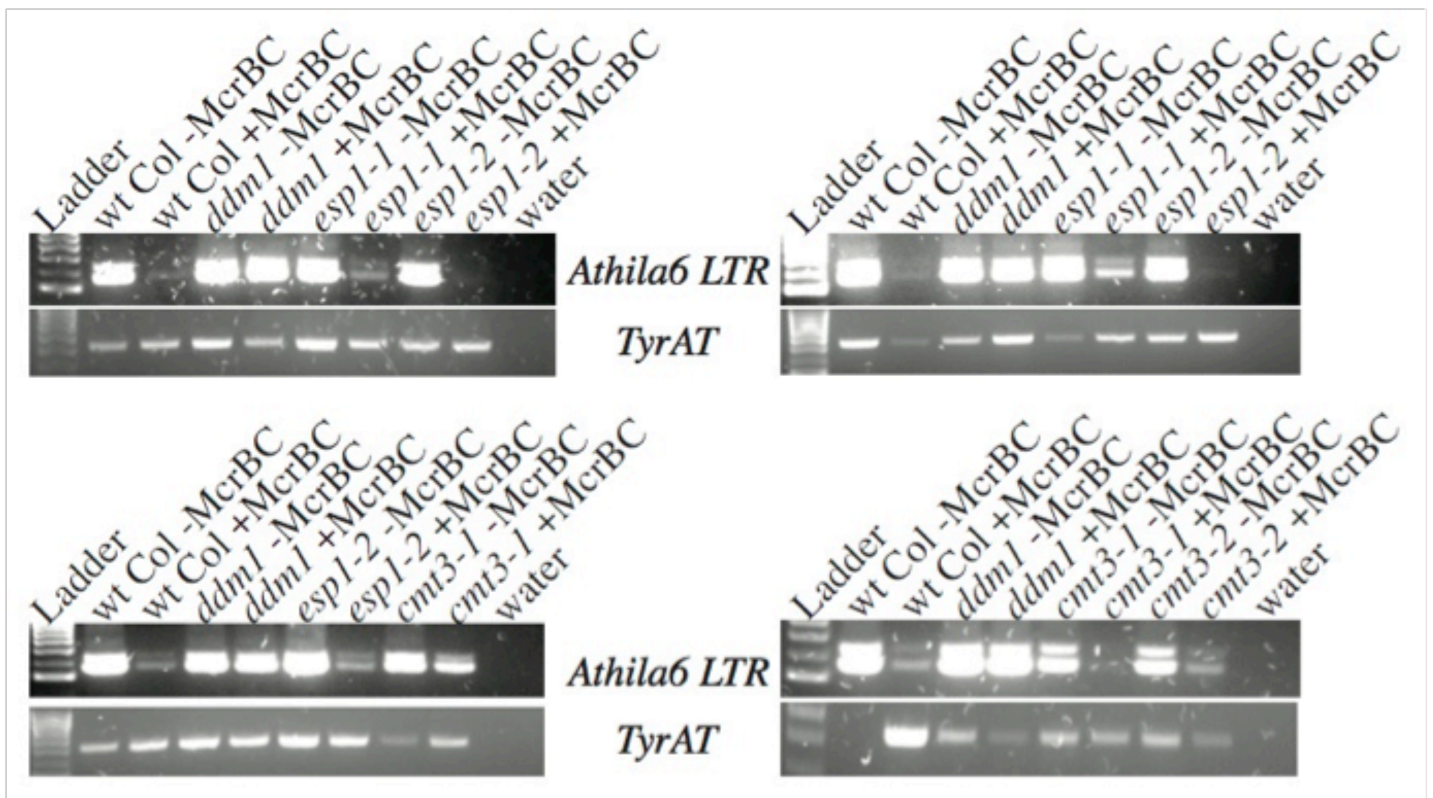

Figure 3. Analysis of DNA methylation of the Athila6 LTR by McrBC-PCR. Analysis of DNA methylation levels of the Athila6 gag/pol promoter and regulatory region Athila6 LTR. For each sample, the DNA was either treated with the McrBC enzyme (+McrBC) or subjected to a mock reaction without McrBC enzyme (-McrBC). The +McrBC or –McrBC samples were PCR amplified for the Athila6 LTR, as well as an un-methylated control region of the genome, TyrAT (At2g20610). The TyrAT control shows equal loading and that the McrBC enzyme does not degrade unmethylated DNA. When a PCR-targeted region of the genome is methylated there is a difference in the level of amplification between the +McrBC and –McrBC samples (such as in the wt Col Athila6 LTR reaction). When a PCR-targeted region of the genome is not methylated, there is no level of amplification difference between the +McrBC and –McrBC samples (such as in the ddm1 Athila6 LTR reaction). Two biological replicates (-1 and -2) were performed for each mutant genotype, and each replicate was subjected to 2-3 repeat digestions and PCR reactions.

methylation levels to wt Col and ddm1 controls (Figure 3). As expected, in wt Col this Athila6 LTR region is fully methylated (+McrBC PCR does not amplify) while in ddm1 mutants this region loses DNA methylation resulting in transcriptional activation (Figure 3). We performed two biological replicates of the esp1 mutant (esp1-1 and esp1-2), as well as 2-3 technical repeats of each sample (Figure 3). We found that one biological replicate of the esp1 McrBC PCR reaction (esp1-1) repeatedly amplified to a higher level than wt Col, suggesting that the esp1-1 individual has less DNA methylation than wt Col. However, the other esp1 biological replicate (esp1-2) did not show this reduction in DNA methylation (Figure 3). From this data, coupled with the fact that the amplification level in esp1-1 is only slightly higher than wt Col, we cannot conclude if the esp1 mutant has a loss of DNA methylation. We saw even greater variation between repeats for the cmt3 mutant, and therefore we are unable to determine in our assay if cmt3 or esp1 have altered TE DNA methylation levels.

CONCLUSION

We determined that the cmt3 mutant has widespread activation of TEs. The CMT3 gene is a known DNA methyltransferase protein that acts to maintain CHG context DNA methylation (where H=A,C or T) and maintain TEs in an epigenetically silenced state (Kato, Miura, Bender, Jacobsen, & Kakutani, 2003; Tompa et al., 2002). We tested the cmt3 mutant because the previously published literature provided conflicting data and results on whether the loss of CMT3 protein function leads to a global activation of TEs. Our data demonstrates that in cmt3 mutants there is widespread reactivation of TEs, but not global (whole genome) activation like the ddm1 mutant. In addition, while one TE (AtCopia52) has the same expression level in cmt3 as in ddm1, the other six TEs assayed have lower expression levels in cmt3 compared to ddm1. This sug-

gests that there is only minor or partial reactivation of TEs in cmt3 mutants. We found that cmt3 mutants lose very little or no DNA methylation using our McrBC analysis (Figure 3) compared to the total loss of DNA methylation observed in ddm1 mutants, supporting our qRT-PCR results and conclusion that the CMT3 protein plays a minor but measurable role in the maintenance of TE silencing.

The gene UHRF1 is also known as VIM1. The protein product of this gene has known roles in the propagation of CG context DNA methylation (Feng et al., 2010; Woo, Dittmer, & Richards, 2008; Woo, Pontes, Pikaard, & Richards, 2007), and therefore we hypothesized that it may play a role in the maintenance of TE transcriptional silencing. For five of the TEs analyzed, we found no evidence of reactivation in uhrf1 mutants. For two TEs (Athila6 gag/pol and AtENSPM6) we observed very minor relative expression increases compared to the wt Col baseline level. Therefore, we conclude that the UHRF1 protein plays a minor role in the suppression of transcription from very select TEs. We did not measure the DNA methylation levels of the uhrf1 mutant, however from the literature we would expect a reduction in CG context DNA methylation from TEs (Woo et al., 2008). However, this reduction of TE methylation must not be enough to reactivate transcription from most TEs.

ESP1 is annotated as an mRNA binding protein, and has been found to regulate the post-transcriptional degradation of mRNAs by RNA interference (Herr, Molnar, Jones, & Baulcombe, 2006). We found that esp1 mutants have a minor and specific reactivation of only two TEs (Athila6 gag/pol and AtLINE1), while five TEs showed no evidence of TE activation. The reactivation of any TEs in the esp1 mutant line is somewhat unexpected and exciting, as the ESP1 protein is thought to function only after transcription and mRNA production (post-transcriptionally) (Herr et al., 2006). It is currently unknown if the accumulation of TE mRNA in esp1 mutants is due to the epigenetic reactivation of the TE promoters (transcriptional regulation), or if some TEs (such as Athila6 gag/pol and AtLINE1) have a low level of mRNA transcription that is normally degraded with the help of the ESP1 protein, but in esp1 mutants this post-transcriptional degradation does not occur and the TE mRNAs accumulate. To determine if there is loss of epigenetic silencing of the TE promoter, we attempted to assay the DNA methylation level in esp1 mutants. Although we obtained conflicting data, at least one biological replicate of esp1 DNA showed loss of TE DNA methylation in two repeated experiments. Further experimentation is required to determine how the ESP1 protein functions to suppress TE mRNA accumulation. Nevertheless, the identification of an RNA binding protein that regulates the mRNA accumulation of some TEs demonstrates that mRNA biding and processing factors do play a role in the suppression of TE expression in Arabidopsis.

Acknowledgments

The Authors thank the OSU Arabidopsis Biological Resource Center for supplying reagents and Eduardo Acosta for course logistics. This work was made possible due to teaching funds supplied by the Department of Molecular Genetics and the National Science Foundation Career Award MCB-1252370 to R.K.S.

REFERENCES

Cerutti, H., & Casas-Mollano, J. A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. Current Genetics, 50(2), 81–99.

Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. Nature Reviews Genetics, 10(10), 691.

de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genetics, 7(12), e1002384.

Dou, K., Huang, C.-F., Ma, Z.-Y., Zhang, C.-J., Zhou, J.-X., Huang, H.-W., et al. (2013). The PRP6-like splicing factor STA1 is involved in RNA-directed DNA methylation by facilitating the production of Pol V-dependent scaffold RNAs. Nucleic Acids Research, 41(18), 8489–8502.

Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences of the United States of America, 107(19), 8689–8694.

Herr, A. J., Molnar, A., Jones, A., & Baulcombe, D. C. (2006). Defective RNA processing enhances RNA silencing and influences flowering of Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America, 103(41), 14994–15001.

Kato, M., Miura, A., Bender, J., Jacobsen, S. E., & Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Cu rent Biology, 13(5), 421–426.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., et al. (2012). Landscape of Somatic Retrotransposition in Human Cancers. Science, 337 (6097), 967-971.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature, 430(6998), 471–476.

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics, 8(4), 272–285.

Tompa, R., McCallum, C. M., Delrow, J., Henikoff, J. G., van Steensel, B., & Henikoff, S. (2002). Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. Current Biology, 12(1), 65–68.

Woo, H. R., Dittmer, T. A., & Richards, E. J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genetics, 4(8), e1000156.

Woo, H. R., Pontes, O., Pikaard, C. S., & Richards, E. J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes and Development, 21(3), 267–277.