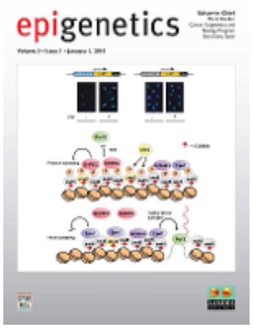

Epigenetics

ISSN: 1559-2294 (Print) 1559-2308 (Online) Journal homepage: http://www.tandfonline.com/loi/kepi20

Variable DNA methylation of transposable elements: The case study of mouse Early Transposons

Daphne Reiss, Ying Zhang, Arefeh Rouhi, Max Reuter & Dixie L. Mager

To cite this article: Daphne Reiss, Ying Zhang, Arefeh Rouhi, Max Reuter & Dixie L. Mager (2010) Variable DNA methylation of transposable elements: The case study of mouse Early Transposons, Epigenetics, 5:1, 68-79, DOI: 10.4161/epi.5.1.10631

To link to this article: http://dx.doi.org/10.4161/epi.5.1.10631

Epigenetics 5:1, 68-79; January 1, 2010; © 2010 Landes Bioscience

Variable DNA methylation of transposable elements

The case study of mouse Early Transposons

Daphne Reiss,[1,†] Ying Zhang,[1] Arefeh Rouhi,[1] Max Reuter[2,*] and Dixie L. Mager[1,*]

[1]Terry Fox Laboratory; BC Cancer Agency; Vancouver, BC CA; and Deptartment of Medical Genetics; University of British Columbia; [2]Research Department for Genetics, Evolution and Environment; Faculty of Life Sciences; University College London; London, UK

[†]Current address: CNRS (UMR 5558); University of Lyon Lab. Biométrie & Biologie Evolutive; Villeurbanne, France

Key words: inter-individual, variability, DNA methylation, retrotransposons, mouse, long terminal repeats

Abbreviations: TE, transposable element; (H)ERV, (human) endogenous retrovirus; ETn, early transposon; IAP, intracisternal-A-particle; LTR, long terminal repeat

Phenotypic variation stems from both genetic and epigenetic differences between individuals. In order to elucidate how phenotypes are determined, it is necessary to understand the forces that generate variation in genome sequence as well as its epigenetic state. In both contexts, transposable elements (TEs) may play an important role. It is well established that TE activity is a major generator of genetic variation, but recent research also suggests that TEs contribute to epigenetic variation. Stochastic epigenetic silencing of some TE insertions in mice has been shown to cause phenotypic variability between individuals. However, the prevalence of this phenomenon has never been evaluated. Here, we use 18 insertions of a mouse Endogenous Retrovirus (ERV) family, the Early Transposons (ETns), to detect insertion-dependent determinants of DNA methylation levels and variability between both cells and individuals. We show that the structure and age of insertions influence methylation levels and variability, resulting in a subgroup of loci that displays unexpectedly high variability in methylation and suggesting stochastic events during methylation establishment. Despite variation in methylation according to the age and structure of each locus, homologous CpG sites show similar tendencies in methylation levels across loci, emphasizing the role of the insertion's sequence in methylation determination. Our results show that differences in methylation of ETns between individuals is not a sporadic phenomenon and support the hypothesis that ERVs contribute to phenotypic variability through their stochastic silencing.

Introduction

An organism's phenotype can vary due to its genotype as well as the environment in which it develops. However, the relationship between the phenotype and its genetic and environmental basis is far from deterministic; even genetically identical individuals reared under carefully replicated environmental conditions vary phenotypically, and sometimes to a large degree. It is becoming increasingly clear that at least a part of this hitherto unexplained variation could be accounted for by epigenetic differences between individuals. Epigenetics relates to phenotypic variation in two ways. First, it seems to be, at least partly, the interface for the interaction between environment and genotype.[1,2] Second, although frequently associated with genetic variation,[3,4] epigenetic variation occurs even when variation of genetic and environmental factors is absent or too small to be measurable.[5-7]

Intriguingly, among the few examples of natural epigenetic variation reported so far in animals,[8] almost all involve transposable elements (TEs). Some of these studies suggest that phenotypic variation can arise from stochastic methylation of TE insertions inducing the so-called "metastable epialleles."[9] For example, genomic insertions of Intracisternal A-type Particle elements (IAPs), a family of mouse Endogenous Retroviruses (ERVs), generate new phenotypes by driving ectopic or aberrant expression of mouse genes.[10-12] The phenotypes arising from the IAP-driven mis-expression display hyper-variable penetrance between isogenic mice.[6,9,13] Detailed molecular analyses have shown that within an isogenic background, IAP methylation can vary not only between individuals but also between cells of the same individual, even within the same tissue.[6,9,13,14] However, other studies suggest that TE methylation is not generally stochastic. For example, Human Endogenous Retrovirus (HERV) insertions show more deterministic patterns of methylation with similar methylation levels across individuals and consistent differences between cells.[15,16] Similarly, the majority of TE loci show consistent methylation levels in different plant ecotypes despite

*Correspondence to: Max Reuter and Dixie L. Mager; Email: m.reuter@ucl.ac.uk and dmager@bccrc.ca
Submitted: 10/15/09; Accepted: 11/13/09
Previously published online: www.landesbioscience.com/journals/epigenetics/article/10631

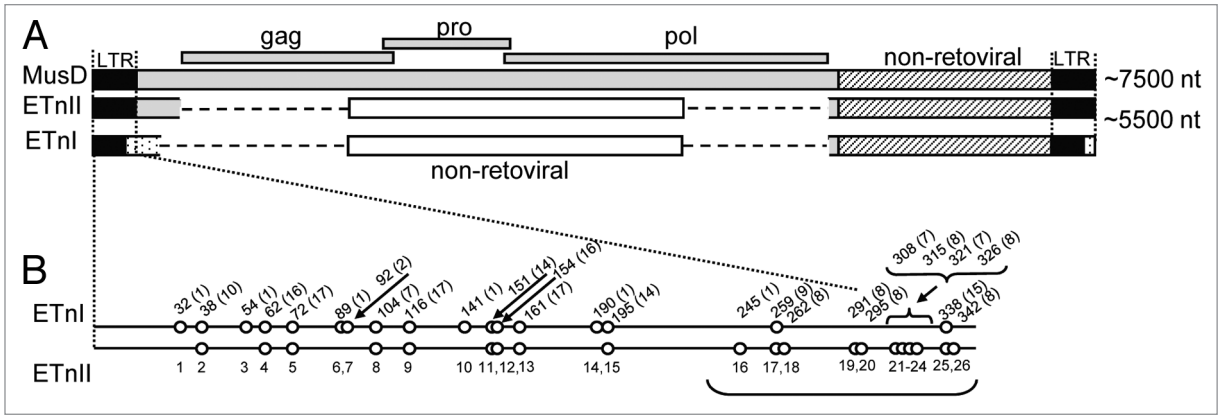

Figure 1. Structure of the MusD and ETn transposable elements. (A) Sequence features of MusD and ETn elements. The coding competent MusD contains three open reading frames encoding the proteins necessary for retrotransposition *gag, pro, pol*. In the non-coding competent ETnII and ETnI elements, these ORFs have been replaced by a shorter non-coding sequence of unknown origin. ETnI elements further differ from the other two subfamilies downstream of the LTR, as indicated by a dotted pattern. (B) A close up of the LTRs of ETnI and ETnII showing the location of CpG dinucleotides (circles numbered 1–26) present in at least one of the 18 ETn insertions chosen for this analysis. Numbers on the top of the line indicate the CpG positions on the consensus sequence and numbers in parentheses indicate how many of the 18 LTRs bear this CpG site. The region towards the 3' end in which ETnI and ETnII sequences differ is indicated by a horizontal parenthesis.

differences in the genomic background.[3] The apparent contradiction between these studies indicates the need for investigations into the factors that underlie the level of methylation and its variation. Elucidating the stochastic and deterministic elements of methylation is important because of the aforementioned involvement of epigenetic effects in determining the phenotype. Moreover, such an investigation may also help to gain insight into the mechanism of TE methylation, details of which are only starting to emerge at least in mammals.[17,18]

The methylation status of a TE locus will likely depend both on its genomic context, that is the insertion site, and the sequence of the TE itself. Although insertion site effects on methylation are well documented[19,20] and reviewed elsewhere,[21] an effort to identify TE-dependent determinants is lacking. Indeed, differences in methylation establishment attributed to the TE sequence itself concern TEs from different families, too dissimilar to provide any insight into the precise factors underlying these differences.[18,22] Here, we focus on properties of the element itself and determine factors that underlie the degree of DNA methylation as well as the variability of methylation between cells of the same tissue and between individuals. As a model system we chose the MusD/Early Transposon (ETn) family of ERVs in mice, and examined methylation patterns of 18 ETn long terminal repeats (LTRs) inserted in similar genomic contexts but differing in their age and structure. We found significant effects of both ETn age and structure on levels and variability of methylation. Moreover, an analysis of paralogous CpG sites between LTR loci showed that such CpGs present similar methylation patterns across loci, confirming the role of the TE sequence in methylation establishment.

Results

Choice of insertions. To test whether age and structure of TE insertions affect methylation, we identified individual ERV insertions in the mouse genome and characterized their methylation in several mice from different litters. To identify candidate insertions for this study, the assembled C57BL/6 (B6) genomic sequence was screened for insertions satisfying the following three criteria: First, the insertions had to be full length, to allow an estimation of their age by comparison of their 5' and 3' LTR. Upon insertion of an LTR-retroelement, both LTRs are identical and then diverge from each other following the insertion event.[23] Second, insertions had to be in similar genomic contexts to avoid effects of different epigenetic states of the flanking sequences. Thus, the insertions were chosen in introns of genes actively transcribed in the tissues examined, this context offering the best-defined environment, since genomic regions outside genes may bear transcribed sequences with unknown tissue specificity. Finally, insertions had to be selectively "neutral," meaning that they should have the least possible effect on expression of the surrounding gene, to avoid any selective forces that may bias the insertion's methylation. Therefore, copies belonging to the MusD/ETn family of retrotransposons inserted in the anti-sense direction to the surrounding gene were chosen for the analysis. Deleterious effects of intronic MusD/ETns on gene expression have been well defined[24] and nearly always occur when the insertions are in the same direction as the gene, suggesting that anti-sense insertions have a limited, if any, impact on fitness. The MusD/ETn family is the most suitable for our analysis since, after examination of two other active mouse ERV families, IAPs and MTs, we found that the latter have only one CpG site in their LTRs while the former comprise mostly very young copies of 99.5–100% identity between LTRs.

This screen identified 26 MusD/ETn insertions satisfying the above criteria. Among these, only 1 belonged to the MusD subfamily with the remaining 25 being ETns, which are the non-autonomous counterparts of the coding competent MusDs[25,26] (**Fig. 1**). The fact that only one MusD copy was detected by this screen is surprising since there are about 90 MusD copies

Table 1. Insertions analyzed in this study

Name	Gene[a]	Tissue[b]	Strain[c]	Indiv[d]	Type[e]	CpG[f]	Id (%)[g]
LTR-cdh	*Cdh23*	heart	A/J	4	II	16	317/317 (100)
LTR-dym	*Dym*	muscle	A/J	2	II	19	319/319 (100)
LTR-atp	*Atp9a*	brain	A/J	4	II	17	319/319 (100)
LTR-gem	*Gem*	lung	A/J	4	II	13	317/317(100)
LTR-snt	*Sntb1*	liver	A/J	4	II	17	317/317 (100)
LTR-tex	*Tex21*	testis	A/J	2	II	18	316/317 (99.7)
LTR-shb	*Sh3bp4*	lung	B6	2	II	15	314/317 (99)
LTR-rik	*A830018L16Rik*	brain	B6	2	II	18	309/317 (97.5)
LTR-tbc	*Tbc1d12*	pancreas	A/J	2	I	9	322/322 (100)
LTR-mtm	*Mtmr7*	brain	B6	2	I	9	320/322 (99.4)
LTR-cde	*Cd84*	heart	B6	2	I	11	319/321 (99.8)
LTR-vnn	*Vnn3*	liver	B6	2	I	8	321/322 (99.7)
LTR-nat	*Nmnat2*	brain	A/J	4	I	10	307/310 (99)
LTR-gvi	*Gvin1*	intestine	A/J	2	I	9	307/314 (97.8)
LTR-lhf	*Lhfp*	lung	A/J	4	I	9	321/321 (100)
LTR-gph	*Gphn*	brain	A/J	4	I	7	315/320 (98.4)
LTR-abc	*Abcc4*	prostate	A/J	2	I	10	320/332 (96.4)
LTR-mct	*Mctp1*	brain	A/J	4	I	5	324/339 (95.6)

[a]gene in the intron of which the ETn is inserted. [b]tissue used to determine ETn methylation and in which the gene is expressed. [c]mouse strain used for the analysis. B6 = C57BL/6. [d]the number of individuals analyzed. [e]type I refers to ETnI and type II to ETnII insertions. [f]number of CpGs present in the 5' LTR studied here. [g]ratio of identical nucleotides between 5' and 3' LTRs and % of identity in parentheses.

in the B6 strain (excluding solo LTRs[27]) and this number is only 2.7 times smaller than the number of ETn copies. The lack of full length MusDs in introns could reflect a higher impact of these particular elements on fitness as has been suggested previously for this and other TE families.[19,20,28] Indeed, longer TEs are more subject to elimination by selection than shorter ones likely because of increased ectopic recombination occurrence.[29] Therefore, we decided to focus our study on ETns. This subfamily is further divided in two types, ETnI and ETnII (**Fig. 1**). Among the 25 insertions, 19 were ETnI and 6 ETnII in accordance with their relative abundance in the genome (200 versus 40 respectively excluding solo LTRs[27]). The consensus sequences of these two types are 94–98% identical. In fact, ETnI and ETnII were defined based on a short fragment located in their 5' region with no apparent homology between ETnI and ETnII.[30] This fragment is 270 bp long for ETnI and slightly longer—410 bp—for ETnII (**Fig. 1**). It includes the last 90 nucleotides of the LTR and a small section of 5' internal region.[30] Moreover, this region is CpG-rich in ETnIIs but CpG poor in ETnIs: ETnII-LTRs carry about two times more CpG dinucleotides than ETnIs (**Table 1, Fig. 1**) and the additional CpGs are concentrated in the 3' region of the ETnII-LTR. Even though the structural difference between the two types is small (5–7% of the total length), the ETnII type includes the most active elements since most new mutations induced by MusD/ETn transposition involve ETnII copies.[24] However, it should be noted that the MusD/ETn subfamilies correspond to a classification based on structure and not on phylogenetic relationships between copies.[27] Indeed, this family seems to be subject to numerous between-copy recombination events resulting in different evolutionary histories for different parts of the elements.[26]

Out of the 25 suitable ETn insertions identified in our screen, we selected 14 for methylation analysis. With the view of testing for an eventual role of an insertion's age on its methylation pattern, our selection aimed at maximizing the age range across selected insertions. The 14 ETns chosen vary between 95.6% and 100% identity between their 5' and 3' LTRs. Assuming a mutation rate of 4.5×10^{-9} per base per year in mouse,[31] this corresponds roughly to an age range from 0 to 5 MY. Insertions with identical LTRs only imply that insertions are younger than 350,000 years. We therefore also included three polymorphic ETnII insertions present only in the A/J strain and one ETnII insertion present in A/J and the related A/WySn strain but absent in B6. These insertions were chosen from a recently published study that identified ETn elements variably present among mouse strains.[32] In fact, the A/J strain is known to display a relatively high rate of ETnII transposition.[24] Thus, specific A/J insertions are likely to have occurred after establishment of the strain no more than 80 to 100 years ago.[33] To confirm that these polymorphic insertions had 100% identical 5' and 3' LTRs and belonged to the ETn subfamily (and not MusD), the first 1.6 kb as well as their LTRs were sequenced. The 5' and 3' LTRs of the four insertions were 100% identical (**Table 1**). In total 18 insertions were studied, 10 ETnI and 8 ETnII (**Tables 1 and S1**).

Both age and structure affect ETn methylation levels and variability. We determined the methylation of all 18 ETn insertions by bisulfite-sequencing of their 5' LTR. For each insertion, we sequenced between 15 and 52 clones of PCR products

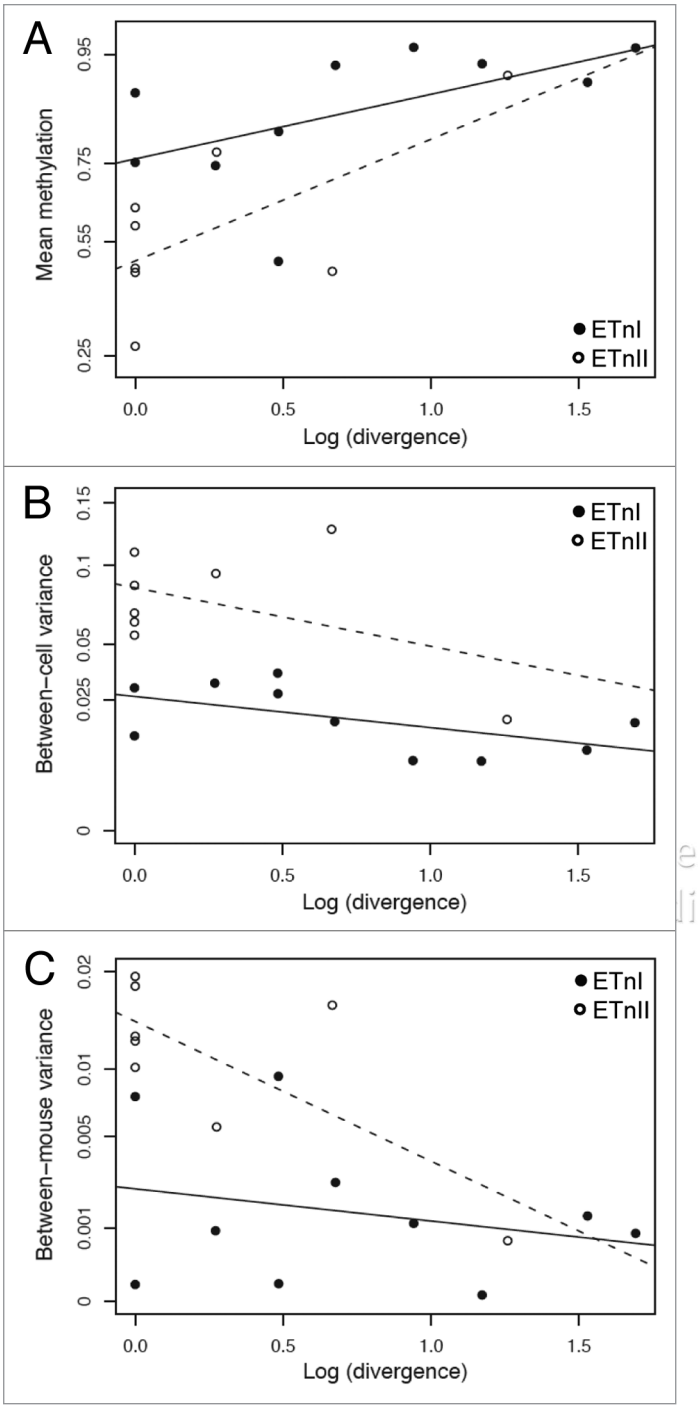

Figure 2. Effect of insertion age and structure on methylation. Plots showing the effect of insertion age (log (percentage divergence +1)) on mean methylation (A), between-cell variance in methylation (B) and between-mouse variance in methylation (C). Data are shown separately for ETnI (filled circles, solid lines) and ETnII (open circles, dashed lines). On all three panels, means and variances are plotted as arcsine-transformed data (as used in the analyses), but axes show raw (un-transformed) values to provide a more informative representation.

amplified from extractions of genomic DNA from 2 or 4 individual mice (**Table 1**). In order to ensure a homogenous epigenetic context for all insertions, methylation was determined in a tissue where the surrounding gene is expressed (**Table 1**). All clones retained for the analysis were independent (see Methods). The methylation patterns of all 18 LTRs are shown in **Figure S1**.

We used Analysis of Covariance (ANCOVA) to establish how patterns of methylation varied with an insertion's age (divergence between 5' and 3' LTRs) and structure (ETnI versus ETnII). We first analyzed mean methylation levels, calculated for each insertion by averaging the proportion of methylated CpGs across clones and averaging these values across the replicate mice. Statistical analysis showed that mean methylation was significantly affected by both the age of insertions (mean methylation arcsine-transformed; age = log (percentage divergence +1); $F_{1,14}$ = 17.3, $p < 0.001$) and the structure of the insertions ($F_{1,14}$ = 5.8, $p < 0.05$). However, the age-effect did not significantly differ between the two types. (interaction term, $F_{1,14}$ = 0.8, NS). Thus, as shown in **Figure 2A**, the methylation level of young ETnI is higher than that of young ETnII insertions, but methylation of both types then increases over time.

We analyzed variation in methylation at two levels, variation between cells of the same individual and variation in mean methylation levels between mice. To assess between-cell variance, we relied on the fact that all clones retained for the analysis are independent (i.e., generated by different DNA templates during PCR) and considered, for subsequent analyses, that methylation of a clone corresponds to the methylation of a cell for a given LTR-locus. Of course, clone methylation actually measures the state of an allele, and there could be variation in methylation between alleles within a cell, which is not being assessed here. For each insertion, the between-cell variance in methylation was calculated and averaged over replicate mice. Between-mouse variation was calculated as the variance in an insertion's average methylation level across mice. Statistical analyses of both types of variation revealed similar results to those obtained for average methylation. The age of insertions had a significant effect on both variances (between-cells: $F_{1,14}$ = 16.9, p = 0.001; between-mouse: $F_{1,14}$ = 13.2, p = 0.003) as did the type of insertion (between-cells: $F_{1,14}$ = 19.22, $p < 0.001$; between-mouse: $F_{1,14}$ = 10.9, p = 0.005), but no significant interaction was found (between-cells: $F_{1,14}$ = 0.4, NS; between-mouse: $F_{1,14}$ = 2.3, NS). While the significance of the effects are similar between mean methylation and its variances, the direction of the effects is not. Thus, both measures of methylation variance are higher for ETnII than for ETnI and decrease with age at similar rates for both types (**Fig. 2B and C**).

High between-cell variability of methylation for ETnII insertions. Our previous analyses treated mean and variance as two independent aspects of methylation. However, both measures are linked if we consider methylation as a binomial probabilistic process with two possible states, methylated and unmethylated, for a given CpG site. If we consider the chances of individual CpGs within an LTR being methylated to be independent and to occur with the same probability across all cells of an individual, we can calculate the predicted variance in methylation between cells for a given average methylation level *p*. Based on the variance of the binomial distribution, this variance should be *np (1-p)*, where *n* is the number of CpGs in an LTR. Comparing the observed

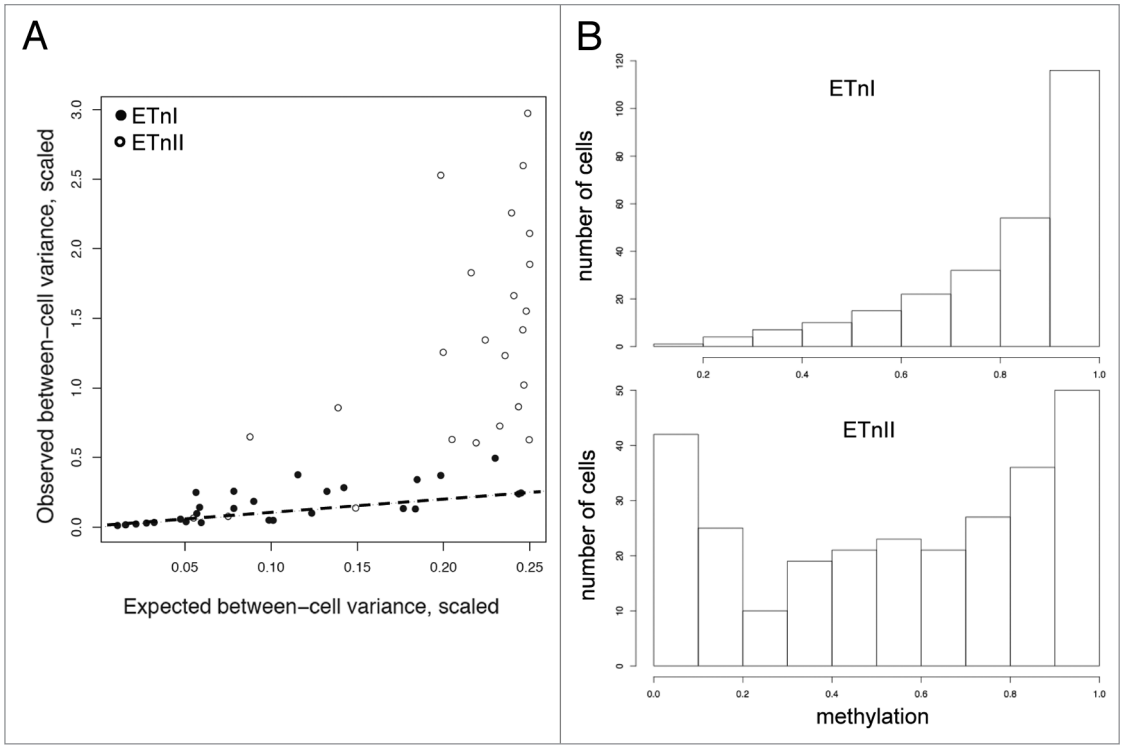

Figure 3. Excess of between-cell variance in methylation for ETnII. (A) Plot showing the observed versus the expected between-cell variance of each LTR in each replicate mouse tested. Black dots correspond to ETnI insertions and white to ETnII. The dashed line indicates the distribution for which observed equals expected between-cell variance according to the binomial model. Both observed and expected variances are scaled for the number of cells assayed. (B) Distribution of methylation levels across individual cells for ETnI (upper) and ETnII insertions (lower).

variance to this expectation allows us to make inferences about the nature of the methylation process, in particular about the independence of methylation state between CpGs of the same LTR. If CpGs are methylated independently and with a constant probability (given by the average methylation *p*), then observed and predicted variance should be similar. An excess of observed variance, in contrast, suggests that cells differ in their methylation rates over and above the differences expected due to the stochasticity of a random process. We performed a comparison of this kind for all insertions and all mice included in our dataset. As already indicated by the ANOVA results above, within-mouse (between-cell) variance in methylation is far higher for ETnII insertions than for ETnI (**Fig. 3A**). However, the data also suggests differences in the process of methylation between the two types. For EtnI, the observed variance (corrected for the differences in the number of CpGs between different loci) is significantly higher than the expected one (Wilcoxon test on observed - expected values, V = 283, $p < 0.05$). However, the distribution of proportional excess [(observed - expected)/expected] is highly skewed and centered close to zero (median = 0.03, mean = 0.51), indicating that most LTRs show methylation patterns close to the probabilistic expectation. The situation is very different for ETnII. Here, the observed variance is on average about five-fold in excess of the expected value, a difference that is highly significant ($t_{23} = 8.4$, $p < 0.0001$). Furthermore, the distribution of proportional excess of variance is nearly symmetrical (median = 5.1, mean = 4.9), reflecting the fact that almost all LTRs in all mice show excess between-cell variance in methylation. This analysis indicates that between-cell variability in methylation of ETnI LTRs arises mostly as a result of random processes. The methylation of individual ETnII LTRs, in contrast, varies widely and significantly more than expected from the probabilistic process. This suggests that for ETnII additional factors are at work that result in either almost complete presence or complete absence of methylation of LTRs in individual cells. Indeed, the distribution of ETnII methylation levels when all LTR-loci in all cells and replicate mice are pooled together follows a bimodal distribution with many completely unmethylated and many heavily methylated cells as shown in **Figure 3B**.

No evidence for small RNA effects on methylation. Reports in Drosophila and mouse strongly suggest an involvement of small RNAs, in particular piRNAs and siRNAs, in the silencing of TE transcription through targeted chromatin modifications.[34-36] Although the mechanism of directed TE methylation in mouse remains elusive, a role of small RNAs in this process is increasingly supported.[17,18] Therefore, we could expect the relative abundance of small RNAs matching the sequence of an element's LTR to correlate with the degree or variation of its methylation. Genome-wide studies in Arabidopsis show that abundance of small RNAs that target a sequence and its methylation levels positively correlate.[37,38] Moreover, small RNAs could be the effectors through which the age and/or the structure influence methylation. For instance, ETnIs, which are more abundant in the genome than ETnIIs, could be targeted by a higher

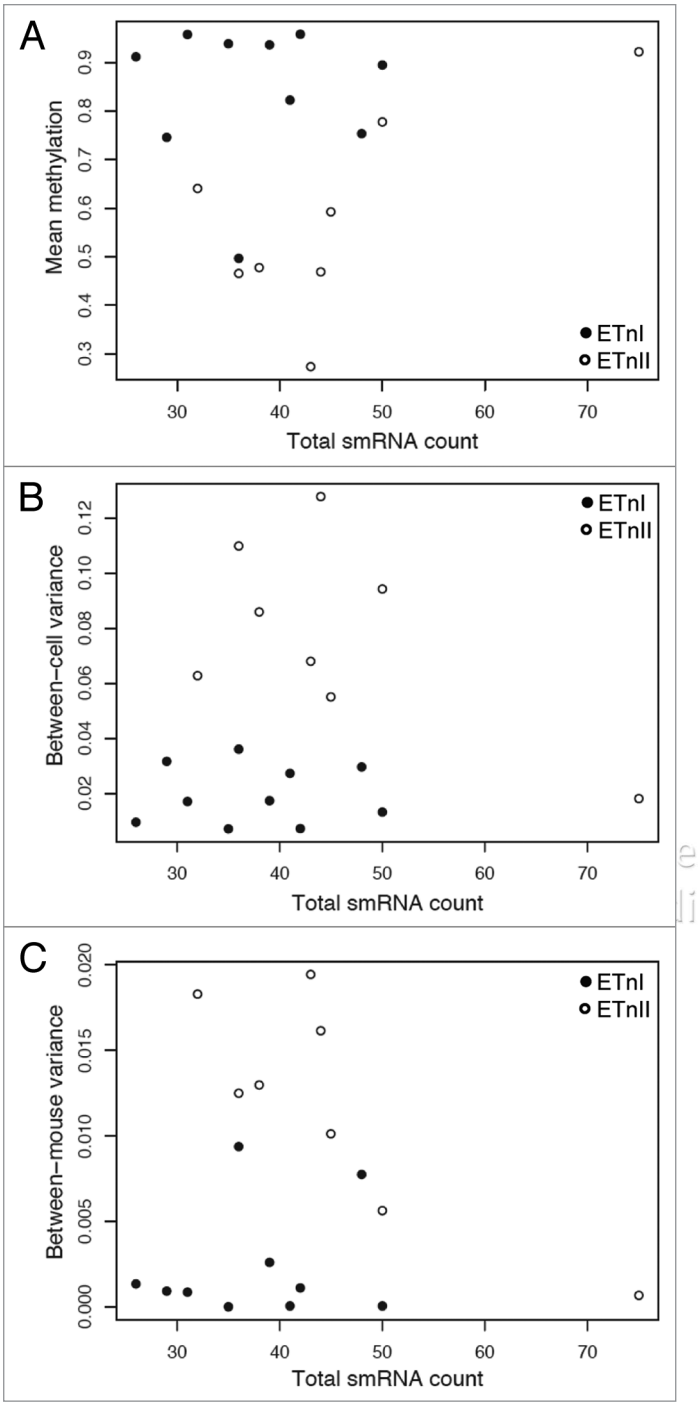

Figure 4. No correlation between methylation levels and small RNA targeting. Plot showing the relationship between the number of total small RNAs generating a perfect hit with each LTR and its upstream and downstream sequence (X-axes) and (A) the mean methylation (B) the between-cell variance in methylation and (C) the between-mouse variance in methylation of ETnI (solid circles) and ETnII (open circles).

number of small RNAs and thus be methylated more efficiently. To test this, we used publicly available databases of total small RNAs (24–33 bp long) from embryonic testis (day 16.5).[17] Here we assume that the relative abundance of ETn small RNAs in embryonic germline is not significantly different from that present in the somatic cells where methylation of ETn LTRs is established during development. For each of the 18 ETns, we queried the small RNA database with the ~322 bp LTR sequence, the 1,000 nucleotides upstream flanking sequence, and the first 1,000 nucleotides of ETn sequence downstream of the 5' LTR and recorded the number of perfect matches. **Figure 4A** shows the number of hits for each LTR-locus in relation to the methylation mean and variance. Analyzing general patterns of targeting, we find that the three sequence segments (flanking, LTR, downstream) are targeted differently by smRNAs while ETnI and ETnII are targeted similarly (ANOVA on hits per nucleotide as a function of sequence segment, ETn type and their interaction; segment: $F_{2,48} = 61.0$, $p < 0.001$; type: $F_{1,48} = 1.2$, NS; interaction: $F_{2,48} = 0.7$, NS). The segment effect in this analysis is a result of lower targeting of the upstream flanking region (**Fig. S2**), which received significantly fewer hits (on average 0.3 per 100 bp) than the LTR (2.5 hits) or the first 1,000 bp of the TE core sequence (3.0 hits). An exception is LTR-rik, which is inserted into an L1 element and therefore shows extremely high values of smRNA targeting of its upstream flanking sequence. This shows that smRNAs are indeed directed against ETn insertions, but target different parts of the elements indiscriminately. While we found evidence for general smRNA targeting of the elements, we could not detect any effect of targeting intensity on the level or the variation of ETn methylation (ANOVA of methylation level, average between-cell variance and between mouse variance as a function of total smRNA hits, element type and their interaction, all smRNA terms NS). This result is not affected by removing the insertion LTR-rik. Thus, this analysis provides no evidence for an effect of small RNAs from day 16.5 embryonic testis on ETn methylation levels or variability in tissues from adult mice.

Insertions with different structure differ in their promoter activity. Another factor possibly affecting methylation of these LTRs is their affinity for transcription factors (TFs). Binding of TFs may protect promoters from methylation. This phenomenon has been repeatedly reported for Sp1/Sp3 transcription factor,[39-41] and for a Moloney Leukemia Virus enhancer preventing maintenance-coupled de novo methylation.[42] Sp1/Sp3 is a transcription factor that binds the LTRs of ETns in three sites located in their 5' extremity[43] (**Table S2**). Since the structural difference between ETnI and ETnII overlaps the LTR downstream of the promoter sequence,[43] it could affect the transcription factor binding properties. As a consequence, differential TF affinity could account for structure effects on methylation. Similarly, age could modify TF binding sites: insertions accumulate mutations in their promoters and these may increasingly impair their binding properties over time. In order to investigate these effects, we analyzed the promoter activities of 11 out of our 18 insertions. Assuming that variance in TF binding is reflected in variance of promoter strength, we undertook luciferase assays in the mouse teratocarcinoma p19 cell line where these elements are known to be transcriptionally active[43] in order to determine and compare their LTR promoter activity. It should be noted that two pairs of ETnIIs (LTR-cdh/LTR-gem and LTR-snt/LTR-atp) are identical in sequence. Thus, the promoter activity of LTR-cdh and LTR-

snt determined by the luciferase assay are equivalent to the promoter activities of LTR-gem and LTR-atp, respectively. **Figure 5** shows the promoter activities of the 11 luciferase reporter gene constructs. Clearly, ETnII promoters show higher activity in p19 cells than ETnI promoters with an average activity of ETnII five times higher than ETnI (Wilcoxon test, U = 0, $p < 0.004$). This result is in accordance with the hypothesis that ETnIIs are less methylated than ETnI due to protection from methylation by TF binding. In contrast, we did not detect an effect of age on promoter activity, as neither within ETnIs nor ETnIIs did promoter activity correlate with the divergence between LTRs (Spearman rank correlations, both NS). The evidence for a link between promoter activity and methylation patterns is thus tentative. While a comparison between the two insertion types supports a connection, within-type analyses do not. It should be noted, however, that both tests suffer from small sample sizes. The comparison between types is basically unreplicated, whereas the number of data points for the age effect within subfamilies is very small for both samples (ETnI: N = 6, ETnII: N = 7). Thus, well-replicated data on more groups of ETns with different promoter strengths and methylation levels is needed to establish a correlation between these two features.

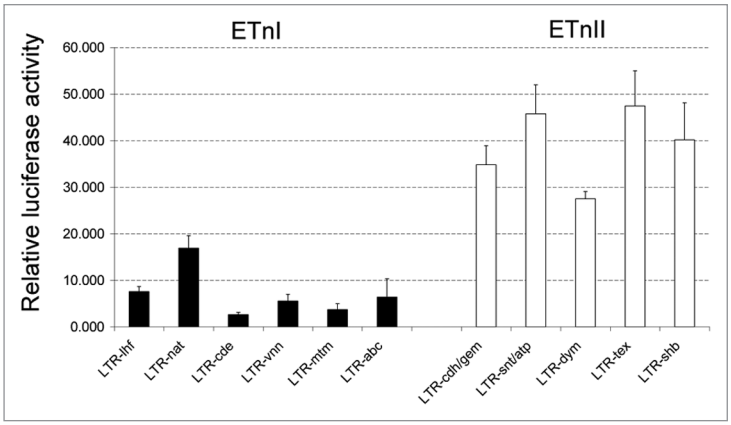

Figure 5. Promoter activities of 11 ETn LTRs. Promoter activity of 6 ETnI LTRs (in black) and 5 ETnII LTRs (in white) in p19 cells. The activity of each LTR relative to the promoterless pGL3B vector is shown.

Analysis of methylation levels of individual CpG sites. The clear difference between ETnI and ETnII methylation patterns suggests that the sequence of an insertion can have a profound influence on the dynamics of methylation. One characteristic that could underlie the effect of structure is the density of CpG dinucleotides present in the LTRs. Indeed, CpG density seems to affect levels of methylation since CpG-rich regions in promoters of housekeeping genes called CpG-islands normally remain unmethylated throughout development.[44,45] The region that distinguishes ETnI from ETnII is CpG-rich in ETnIIs but not in ETnIs (**Fig. 1B**). Even though this excess in CpG density is not statistically significant (the EMBOSS "cpgplot" CpG-island finder tool does not detect any CpG-island in this region) we cannot exclude that a higher CpG density is the structural factor that prevents ETnII-LTRs from methylation in some cells. In order to investigate potential effects of sequence structure, we determined whether methylation varied systematically along the LTR sequences. More specifically, we analyzed the average methylation level of individual CpG sites across cells and replicate mice for each LTR locus. This analysis showed clear and systematic patterns of site-specific methylation, both within and across ETn types. Firstly, we found that overall methylation increases from 5' to 3'. Accordingly, linear regressions of average methylation level (across cells and replicate mice, calculated separately for each locus and CpG) on position of CpGs along the LTR are positive for all insertions analyzed and similar between the two types of insertions (**Fig. 6**; ETnI: slopes ranging from 0.002 to 0.035, mean slope 0.012; ETnII: range 0.003 to 0.017, mean 0.012). The increase of methylation along the LTR contradicts the hypothesis that regions of denser CpG sites are less methylated, since it is the CpG-rich 3' end of the LTRs that tends to be more methylated for ETnII insertions (**Fig. 6B**). Second, we observe that individual CpG sites differ from the overall trend in a systematic manner. Thus, individual CpG sites deviate from the overall 5'-3' trend in a highly parallel manner across different insertions, meaning that across all loci some sites within the LTR are consistently over- or under-methylated, compared to their position in the LTR (**Fig. 7A**). Finally, we find that the pattern of over-and under-methylation is similar between ETnI and ETnII. Specifically, residual methylation of the 11 CpG sites that are common between ETnIs and ETnIIs is correlated between insertion types (**Fig. 7B**; rho = 0.71, $t_{10} = 3.2$, $p = 0.01$). This shows that even though ETnI and ETnII differ in their global average methylation, individual paralogous CpG loci show similar methylation tendencies, persisting even in the different contexts of individual insertions and/or the occurrence of stochastic events that might affect the overall methylation level of the LTR within an individual cell.

Discussion

Effect of structure on methylation. This analysis of 18 LTRs belonging to the ETn family of ERVs showed that methylation varies consistently with structure, age of the insertion and between individual CpG sites within the LTRs. The structural difference between ETnIs and ETnIIs consists of a short sequence of dissimilarity overlapping the 3' end of the LTR and the 5' internal sequence. Although this difference represents only 7% of their sequence, the effect on methylation is clear, with ETnIIs being less methylated than ETnIs and showing a high variance in methylation between cells as well as between individuals. This variability is due to completely unmethylated ETnII-loci in a high proportion of cells, resulting in a bimodal distribution of ETnII methylation.

One of the possible mechanisms that could underlie the structure's effect on methylation is a more efficient protection of ETnII LTRs from methylation by transcription factors. We have shown that ETnII LTRs on average have five times higher promoter activity in reporter gene assays compared with ETnI LTRs in embryonic carcinoma cells. The higher activity of ETnII promoters could be the result of an increased affinity for

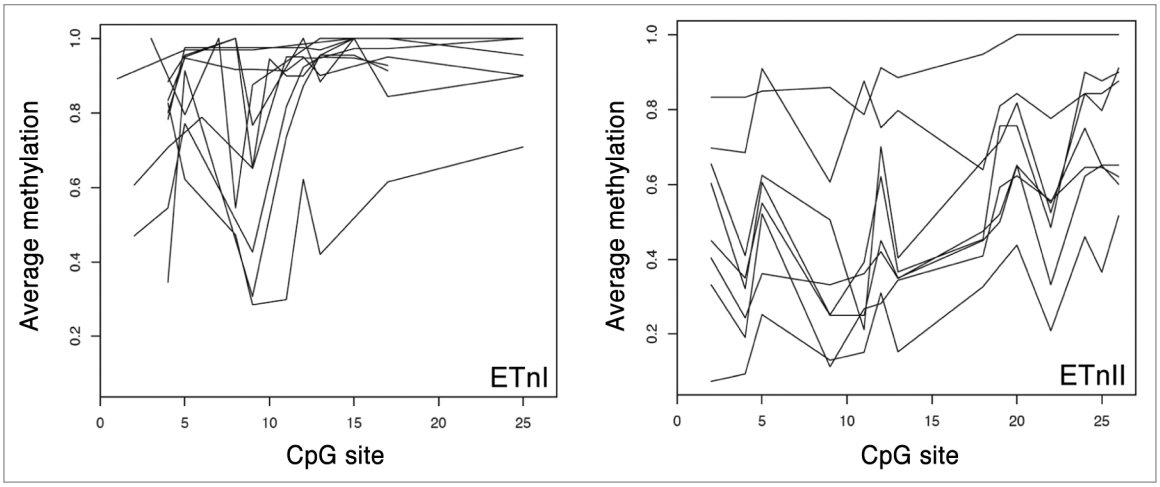

Figure 6. Methylation levels of individual CpG sites along the LTR. Methylation levels of individual CpG sites (numbers corresponding to the schematic representation in Fig. 1B), averaged over cells and replicate mice, for ETnI loci (left) and ETnII loci (right). Each line corresponds to an individual locus.

transcription factors thus preventing methylation. Indeed, the much lower endogenous levels of ETnI transcripts compared to ETnII transcripts in ES cells[28] could be due to a combination of heavier ETnI methylation and lower affinity for TFs. TF binding has been reported to occasionally prevent DNA methylation of promoters, as in the case of Sp1.[39-41] It has been shown that three Sp1-binding sites located in the promoter of the *Aprt* gene prevent methylation in the presence of Sp1 and all three sites are necessary for this effect.[41] Recent work also showed that a polymorphic Sp1/Sp3 binding site in the promoter of a tumor suppressor gene confers protection from aberrant methylation and silencing of the gene.[39] These data are directly relevant to our study, since ETns have three Sp1 sites in their LTRs that act synergistically to promote transcription.[43] However, our results suggest that Sp1/Sp3 sites alone are not sufficient to protect LTRs from methylation; most of the loci analyzed here show at least some methylation while having all three Sp1/Sp3 sites in their sequence (**Suppl. Table S2**).

Another mechanism that could explain the effect of the ETn's structure on methylation is CpG density. The ETnII-specific sequence is CpG-rich compared to that of ETnI. High density of CpG sites combined with transcription initiation signals present in the LTR may form a CpG island-like sequence and remain unmethylated.[44,45] Even though this hypothesis is not supported by our individual CpG site analysis, showing that methylation increases in this region of ETnII LTRs, we cannot rule out that CpG density protection acts in an "all-or-nothing" way, generating completely unmethylated sequences when efficient, but not explaining the methylation patterns within a sequence.

Effect of age on methylation. Age has a positive effect on levels of methylation and a negative effect on its between-cell and between-individual variance. This result seemingly contradicts a homology-dependent mechanism of silencing such as small RNA-directed methylation suggested by comparative and functional studies.[17,46] Under this hypothesis, the specific targeting of TEs is based on their repetitive nature implying that the less divergent an insertion is from the consensus (i.e., the youngest), the more efficiently it will be silenced by directed methylation. If small RNAs mediate this phenomenon, their abundance is expected to be negatively correlated with the age of the insertion that they match. Indeed, a negative correlation between the above variables is significant within ETnI loci but not ETnII (data not shown). Despite the expected effect of the age in ETnI small RNAs, methylation levels do not follow predictions of this scenario. The absence of correlation between small RNA abundance and methylation could be due to the different origins of the two datasets: the small RNA dataset was compiled from 16.5 day embryonic testis whereas we determined DNA methylation in adult tissues. Since generation of small RNAs depends on transcription of the elements,[17] it is possible that assessing small RNA populations at the embryonic developmental stages when ETns are most highly transcribed[47] would give a different result. An alternative hypothesis is that, since ETn insertions studied here all have LTRs with greater than 95% identity, they may not be divergent enough to observe significant differences in small RNA targeting.

Another possible mechanism accounting for the effect of age on methylation could be related to promoter strength of the LTRs. A weakening of the affinity for TFs as inactivating mutations occur within LTRs could result in a less efficient protection from methylation. However, as mentioned above, no correlation was found between promoter strength and age, suggesting that our insertions do not cover a time span large enough for such effects to be observed and ruling out the above hypothesis for this dataset.

The positive correlation between insertion age and methylation levels could also result from selection on insertions that differ inherently and permanently in their methylation, favoring those insertions that are more stably and heavily methylated. Thus, young ('unselected') insertions would show any pattern of methylation while among old ('selected') insertions only heavily methylated ones would remain. Despite our efforts to combine selective neutrality and well-defined genomic contexts by choosing insertions in introns in the anti-sense direction,[24] an effect of their presence on expression of the surrounding genes cannot be excluded. We put forward two arguments against the selection

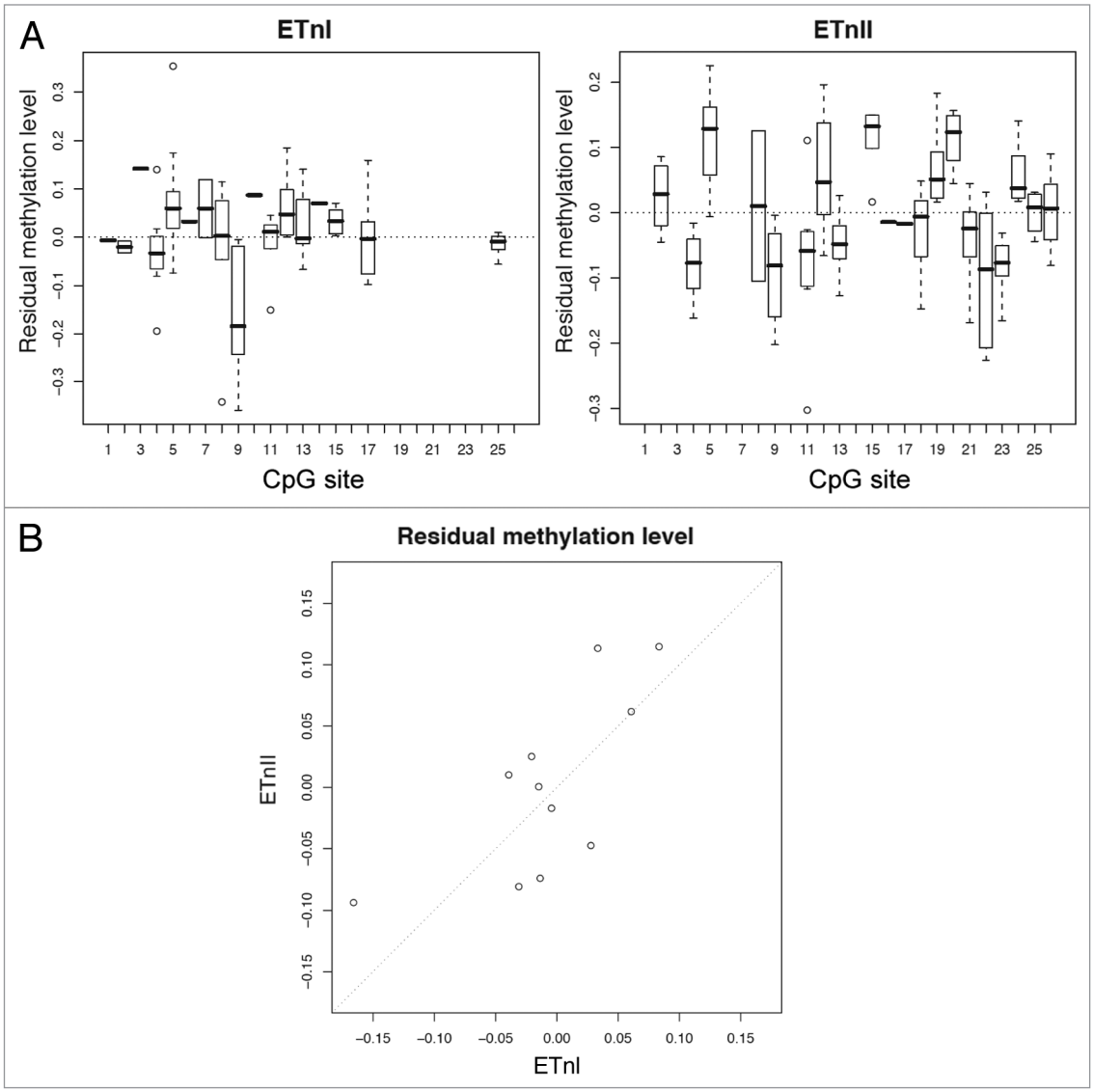

Figure 7. Residual methylation of homologous CpG sites is correlated between ETnI and ETnII loci. (A) Box plot showing the distribution of residual methylation levels of individual CpG sites across insertions of ETnI (left) and ETnII (right). (B) Correlation between mean residual methylation of homologous CpG sites in ETnI loci (x-axis) and ETnII loci (y-axis).

hypothesis. First, recent work in Arabidopsis suggests that selection on insertions near genes acts against methylated TEs rather than unmethylated ones.[37] Thus, the deleterious effects of TEs in these regions appear to derive from the changes in chromatin structure and not by transcriptional interference. Assuming that these results can be extrapolated to ETn in mice, selection should generate the opposite pattern of that observed in our study, a negative correlation between age and methylation level. Second, the selection hypothesis would predict that the correlation patterns described here should be driven by the very recent insertions that are A/J strain-specific (LTR-atp, LTR-cdh, LTR-dym and LTR-gem). Removing these from the dataset does not affect the effect of insertion's age on mean methylation ($p < 0.05$) or between-cell variance ($p < 0.01$) but renders the between-mouse variance non-significant ($p = 0.13$). The persistence of the effect of age on methylation levels and between-cell variance even in a dataset of reduced size means that our data cannot be explained by selection against inefficiently silenced loci.

A scenario which could explain the effect of age on methylation involves "gametic epigenetic inheritance,"[48] namely the meiotic transmission of epigenetic states. Although well documented in plants, heritable epigenetic modifications in mammals have so far only been shown to occur in the cases of the IAP-induced metastable epialleles mentioned in the Introduction and reviewed previously.[48] In fact, these IAPs are not only variably methylated but their epigenetic states are meiotically transmitted because the proportion of methylated IAP-alleles in the offspring depends on their methylation levels in the parents.[6] Moreover, it seems that the inherited mark is not methylation itself but another unknown epigenetic factor.[49] Thus, if a CpG site in a TE locus has a very low probability per generation to become permanently methylated by binding of such an epigenetic factor, methylation would

slowly accumulate and variance in methylation would decrease. A prediction of this hypothesis is that as an insertion gets older its "epigenetic heritability" should increase; meaning that the proportion of variation in methylation due to between mice differences, as opposed to the variation in methylation within mice (between-cell), should increase with age. However, our data show the opposite trend with epigenetic heritability decreasing with age (data not shown). Thus, variability in methylation of old insertions is essentially due to within mouse variation in methylation, which does not support the hypothesis of gametic epigenetic inheritance as underlying the age effect on methylation.

Effect of CpG position within an LTR on methylation. Our analysis of individual CpGs shows that methylation is not homogenous along a 5' LTR. Firstly, CpGs located in the 3' region of the LTR are more methylated than the CpGs located in its 5' region. This trend is observable for both types of ETns and is not attributable to the 5'-3' CpG density variation since CpG density increases from 5' to 3' for ETnIIs but decreases for ETnIs. This finding is in disagreement with the methylation patterns observed in CpG-islands where CpG density is negatively correlated with methylation levels[50,51] suggesting that methylation determinants for gene promoters and TEs are not the same.

In addition to the 5'-3' trend in methylation, there are individual CpG sites that are systematically over or under methylated given the LTR in which they reside and their position in this LTR. This tendency is the same even between ETn types when homologous CpG sites are considered. A similar phenomenon of a "relaxed site-specific" pattern is described in human CpG islands.[51] Unlike CpG-islands, neither an obvious pattern of periodicity in the distances between CpG sites nor preferential nucleotide composition of the flanking sequence was observed for ETn loci. However, the site-specific pattern observed for ETns demonstrates that methylation depends in part on DNA sequence, and that repeated sequences scattered in different genomic loci are treated by the methylation machinery in similar ways. Thus, TEs that belong to highly similar groups such as those of the loci studied here (low percent divergence, small structural differences) are methylated similarly at the same CpG sites. Nonetheless, the mean methylation of these loci can greatly differ as demonstrated by clear differences in methylation levels and variability according to the type and age of the insertion, showing that, in addition to deterministic factors, stochastic events are involved in methylation establishment.

Inter-individual variability in methylation. This study demonstrates that stochastic events in the establishment of methylation resulting in mosaicism within the same tissue occur more or less frequently according to certain characteristics of the insertion. A remaining question is whether the mechanisms underlying the between-cell variance are also responsible for the observed inter-individual variability in methylation. Whereas there is a significant correlation between the two types of variance ($r = 0.76$, $p = 0.0002$) when all the insertions are analyzed together, it becomes non significant when insertions are analyzed by type, most likely because of the reduced dataset. Even though high between-cell variance and high between-mouse variance are frequently observed among ETnIIs but not ETnIs, no clear conclusions can be made since they may not systematically coincide outside of these ETn types.

Our data suggest that inter-individual differences in methylation of TE insertions are a rather frequent phenomenon in mouse and are not confined to the IAPs. In the examples of inter-individual variability previously reported,[6,9,14] the variable methylation was detected only because variable IAP LTR activity caused observable differences in phenotype, raising the question of the real frequency of this phenomenon when it is not linked to an easily detectable phenotype. Here we examined random and a priori neutral copies using only structural and location-relative criteria. Of the 18 insertions, five (LTR-cdh, LTR-gem, LTR-atp LTR-snt and LTR-lhf) show different methylation levels between at least two replicate mice and these differences range between 20 and 32%. This result suggests that, indeed, inter-individual variability in methylation of TEs is common and might be a frequent underlying cause for observed metastable epialleles.

Generality of the conclusions and evolutionary implications. Our detailed study of TE methylation necessarily had to focus on a specific type of element and a limited number of individual insertions. It is therefore interesting to speculate about the generality of our findings. One important question is whether structural and age-related factors play a role in methylation establishment of other TEs. If the molecular factor underlying the structure's effect is promoter strength, this should be observable for other TEs with strong promoters. IAPs for example should also be sensitive to the same factors as they display highly variable methylation. However, ETns and IAPs are both ERVs, so it is unknown if the same phenomenon would occur for other TE types. Another aspect of generality is whether loci inserted in other genomic regions (far from genes, close to genes but outside introns) are also sensitive to the same factors (structure, age, CpG position in the LTR) described here. We expect that to be the case since these factors identified here are characteristic of the insertion itself. However, additional factors could be involved as genomic loci vary and result to different patterns of methylation. Third, it would be interesting to determine if these findings extend to germline cells. This is an important point for the evolution of TEs. In terms of transposition, the state of silencing is important only in germ cells and early embryogenesis, before the formation of the germ line. If the strong variability in methylation originates in early development, it implies that the most highly active copies (young, recombinants with high promoter activity) are also those that are less efficiently silenced, increasing their chances to transpose in germline cells. Although the molecular mechanisms that target DNA methylation in the soma are not characterized in mice, the higher transpositional activity of ETnII compared to ETnI strongly suggests that silencing of ETnIs in the germline is more efficient than silencing of ETnIIs. Finally, it is important to determine whether these phenomena would be encountered in other species. Inter-individual variability in TE methylation has also been observed in plants[52] and human[53] but more work is needed to define factors underlying variability in these taxa.

Materials and Methods

Choice of insertions. Insertions were selected based on a bioinformatics screen of full-length ETns in the assembled C57BL/6 genome (Build 36, Feb. 2006) with WU-BLAST 2.0 (Gish W, 1996–2004 http://blast.wustl.edu/). LTR identity was calculated without considering gaps. The genomic coordinates of all locations are given in **Table S1**. For the four insertions absent in B6, we selected intronic ETns present only in the A/J strain or in both A/J and the related strain A/WySn. These insertions were obtained from a recently published dataset of polymorphic ETns in different mouse strains,[32] and only insertions with an antisense orientation with respect to the enclosing gene were chosen.

Strains. Genomic DNA from tissues from 4 male A/J mice and two male C57BL/6 mouse was used. The A/J strain was used by default and C57BL/6 only for insertions absent in A/J. The tissue to be tested for each insertion was defined by using the microarray expression data from GNF Expression Atlas.[54] The expression of the gene in the selected tissue was verified by RT-PCR, using primers corresponding to the two exons encompassing the intron bearing the insertion (data not shown, primers available upon request).

Bisulfite sequencing. Bisulfite conversion, PCR of individual LTRs, cloning and sequencing were carried out as described previously[16] with the exception that only one round of PCR was necessary to amplify the ETns studied here. All primers used are shown in **Table S3**. All the sequences included in the analyses either displayed unique methylation patterns or unique C to T non-conversion errors (remaining C's not belonging to a CpG dinucleotide) after bisulfite treatment of the genomic DNA. This avoids considering several PCR-amplified sequences resulting from the same template molecule (provided by a single cell). All sequences had a conversion rate >96%.

Small RNA hit counting. Small RNA sequences were compared with the first 1.32 kb of each target ETn sequence and the 1 kb flanking region upstream the beginning of its 5' LTR. Both strands were screened and each perfect match (100% identical) is counted as one hit. Moreover, the abundance of small RNAs was also taken into account when calculating the total number of hits by multiplying the number of hits with the number of times that each small RNA had been sequenced. This part of work was accomplished with a self-developed Perl script.

Constructs, transfection and luciferase assays. The 11 LTRs were amplified by PCR using primers located in the flanking and the internal sequences (**Table S4**) followed by nested PCR in order to eliminate the LTR's surrounding sequence and introduce a KpnI restriction site (**Table S4**). The digested sequences were cloned in the KpnI cloning site of the pGL3B luciferase reporter vector (Promega). All constructs were verified by sequencing. Transfections of teratocarcinoma p19 cells were as described[43] with the following modifications: 24 hours prior to transfection 24 well plates were seeded at a density of 3×10^4. For transfections 500 ng of construct, 50 ng of the Renilla lucferase vector pRL-TK and 0.5 µl of Lipofectamine-2000 (Invitrogen) were used. The data were standardized to the internal Renilla luciferase control and expressed with regard to the residual luciferase activity of the promoterless pGL3-Basic (pGL3B) vector. The results are means and standard deviations from three separate experiments performed in duplicate.

Statistical analysis. Parametric and non-parametric tests were used as indicated in the text. Where appropriate, variables were transformed as indicated in the text in order to assure linearity of relationships as well as normally and homogeneously distributed residuals. All analyses were performed in R 2.6.2.[55]

Acknowledgements

We thank Carolyn Brown, Sylvain Charlat, Matt Lorincz and Irina Maksakova for comments on versions of this manuscript and Liane Gagnier for technical support. This work was supported by a grant from the Canadian Institutes of Health Research to D.L.M. and a fellowship from the Natural Environment Research Council UK (NE/D009189/1) to M.R. Y.Z. is supported by an Alexander Graham Bell Studentship from the Natural Sciences and Engineering Research Council of Canada.

Note

Supplementary materials can be found at:
www.landesbioscience.com/supplement/ReissEPI5-1-Sup.pdf

References

1. Wolff GL, Kodell RL, Moore SR, Cooney CA. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. FASEB J 1998; 12:949-57.
2. Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C. Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 2004; 427:164-7.
3. Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, et al. Epigenetic natural variation in *Arabidopsis thaliana*. PLoS Biol 2007; 5:174.
4. Zhai J, Liu J, Liu B, Li P, Meyers BC, Chen X, et al. Small RNA-directed epigenetic natural variation in *Arabidopsis thaliana*. PLoS Genet 2008; 4:1000056.
5. Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, Woychik RP. Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes Dev 1994; 8:1463-72.
6. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 1999; 23:314-8.
7. Raj A, van Oudenaarden A. Nature, nurture or chance: stochastic gene expression and its consequences. Cell 2008; 135:216-26.
8. Richards EJ. Inherited epigenetic variation—revisiting soft inheritance. Nat Rev Genet 2006; 7:395-401.
9. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends Genet 2002; 18:348-51.
10. Argeson AC, Nelson KK, Siracusa LD. Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. Genetics 1996; 142:557-67.
11. Perry WL, Copeland NG, Jenkins NA. The molecular basis for dominant yellow agouti coat color mutations. Bioessays 1994; 16:705-7.
12. Vasicek TJ, Zeng L, Guan XJ, Zhang T, Costantini F, Tilghman SM. Two dominant mutations in the mouse fused gene are the result of transposon insertions. Genetics 1997; 147:777-86.
13. Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, et al. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc Natl Acad Sci USA 2003; 100:2538-43.
14. Druker R, Bruxner TJ, Lehrbach NJ, Whitelaw E. Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. Nucleic Acids Res 2004; 32:5800-8.
15. Reiss D, Mager DL. Stochastic epigenetic silencing of retrotransposons: does stability come with age? Gene 2007; 390:130-5.
16. Reiss D, Zhang Y, Mager DL. Widely variable endogenous retroviral methylation levels in human placenta. Nucleic Acids Res 2007; 35:4743-54.
17. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 2008; 31:785-99.

18. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev 2008; 22:908-17.
19. Rosin FM, Watanabe N, Cacas JL, Kato N, Arroyo JM, Fang Y, et al. Genome-wide transposon tagging reveals location-dependent effects on transcription and chromatin organization in Arabidopsis. Plant J 2008; 55:514-25.
20. Xie H, Wang M, Bonaldo Mde F, Smith C, Rajaram V, Goldman S, et al. High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum. Nucleic Acids Res 2009; 37:4331-40.
21. Talbert PB, Henikoff S. Spreading of silent chromatin: inaction at a distance. Nat Rev Genet 2006; 7:793-803.
22. Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, et al. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 2003; 35:88-93.
23. Johnson WE, Coffin JM. Constructing primate phylogenies from ancient retrovirus sequences. Proc Natl Acad Sci USA 1999; 96:10254-60.
24. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet 2006; 2:2.
25. Mager DL, Freeman JD. Novel mouse type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences. J Virol 2000; 74:7221-9.
26. Ribet D, Harper F, Dewannieux M, Pierron G, Heidmann T. Murine MusD retrotransposon: structure and molecular evolution of an "intracellularized" retrovirus. J Virol 2007; 81:1888-98.
27. Baust C, Gagnier L, Baillie GJ, Harris MJ, Juriloff DM, Mager DL. Structure and expression of mobile ETnII retroelements and their coding-competent MusD relatives in the mouse. J Virol 2003; 77:11448-58.
28. Maksakova IA, Zhang Y, Mager DL. Preferential epigenetic suppression of the autonomous MusD over the nonautonomous ETn mouse retrotransposons. Mol Cell Biol 2009; 29:2456-68.
29. Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. Size matters: non-LTR retrotransposable elements and ectopic recombination in Drosophila. Mol Biol Evol 2003; 20:880-92.
30. Shell BE, Collins JT, Elenich LA, Szurek PF, Dunnick WA. Two subfamilies of murine retrotransposon ETn sequences. Gene 1990; 86:269-74.
31. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420:520-62.
32. Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, Mager DL. Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. PLoS Genet 2008; 4:1000007.
33. Festing MF. Properties of inbred strains and outbred stocks, with special reference to toxicity testing. J Toxicol Environ Health 1979; 5:53-68.
34. Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, et al. Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev 2007; 21:2300-11.
35. Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, et al. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell 2007; 12:503-14.
36. Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, Birchler JA, et al. Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science 2004; 303:669-72.
37. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 2009; 19:1419-28.
38. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 2008; 133:523-36.
39. Boumber YA, Kondo Y, Chen X, Shen L, Guo Y, Tellez C, et al. An Sp1/Sp3 binding polymorphism confers methylation protection. PLoS Genet 2008; 4:1000162.
40. Macleod D, Charlton J, Mullins J, Bird AP. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev 1994; 8:2282-92.
41. Mummaneni P, Yates P, Simpson J, Rose J, Turker MS. The primary function of a redundant Sp1 binding site in the mouse aprt gene promoter is to block epigenetic gene inactivation. Nucleic Acids Res 1998; 26:5163-9.
42. Lorincz MC, Schubeler D, Hutchinson SR, Dickerson DR, Groudine M. DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent de novo methylation. Mol Cell Biol 2002; 22:7572-80.
43. Maksakova IA, Mager DL. Transcriptional regulation of early transposon elements, an active family of mouse long terminal repeat retrotransposons. J Virol 2005; 79:13865-74.
44. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 1985; 40:91-9.
45. Cooper DN, Taggart MH, Bird AP. Unmethylated domains in vertebrate DNA. Nucleic Acids Res 1983; 11:647-58.
46. Meunier J, Khelifi A, Navratil V, Duret L. Homology-dependent methylation in primate repetitive DNA. Proc Natl Acad Sci USA 2005; 102:5471-6.
47. Loebel DA, Tsoi B, Wong N, O'Rourke MP, Tam PP. Restricted expression of ETn-related sequences during post-implantation mouse development. Gene Expr Patterns 2004; 4:467-71.
48. Youngson NA, Whitelaw E. Transgenerational epigenetic effects. Annu Rev Genomics Hum Genet 2008; 9:233-57.
49. Blewitt ME, Vickaryous NK, Paldi A, Koseki H, Whitelaw E. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS Genet 2006; 2:49.
50. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 2007; 39:457-66.
51. Zhang Y, Rohde C, Tierling S, Jurkowski TP, Bock C, Santacruz D, et al. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet 2009; 5:1000438.
52. Rangwala SH, Elumalai R, Vanier C, Ozkan H, Galbraith DW, Richards EJ. Meiotically stable natural epialleles of Sadhu, a novel Arabidopsis retroposon. PLoS Genet 2006; 2:36.
53. Sandovici I, Kassovska-Bratinova S, Loredo-Osti JC, Leppert M, Suarez A, Stewart R, et al. Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. Hum Mol Genet 2005; 14:2135-43.
54. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, et al. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA 2002; 99:4465-70.
55. R Development Core Team. R Foundation for Statistical Computing. Vienna, Austria 2008.