

SCIENTIFIC REPORTS



OPEN

DNA methylation and hydroxymethylation analyses of the active LINE-1 subfamilies in mice

Yui Murata¹, Miki Bundo^{1,2}, Junko Ueda³, Mie Kubota-Sakashita³, Kiyoto Kasai⁴, Tadafumi Kato³ & Kazuya Iwamoto¹ 

Retrotransposon long interspersed nuclear element-1 (LINE-1) occupies a large proportion of the mammalian genome, comprising approximately 100,000 genomic copies in mice. Epigenetic status of the 5' untranslated region (5'-UTR) of LINE-1 is critical for its promoter activity. DNA methylation levels in the 5'-UTR of human active LINE-1 subfamily can be measured by well-established methods, such as a pyrosequencing-based assay. However, because of the considerable sequence and structural diversity in LINE-1 among species, methods for such assays should be adapted for the species of interest. Here we developed pyrosequencing-based assays to examine methylcytosine (mC) and hydroxymethylcytosine (hmC) levels of the three active LINE-1 subfamilies in mice (Tf1, A, and GfII). Using these assays, we quantified mC and hmC levels in four brain regions and four nonbrain tissues including tail, heart, testis, and ovary. We observed tissue- and subfamily-specific mC and hmC differences. We also found that mC levels were strongly correlated among different brain regions, but mC levels of the testis showed a poor correlation with those of other tissues. Interestingly, mC levels in the A and GfII subfamilies were highly correlated, possibly reflecting their close evolutionary relationship. Our assays will be useful for exploring the epigenetic regulation of the active LINE-1 subfamilies in mice.

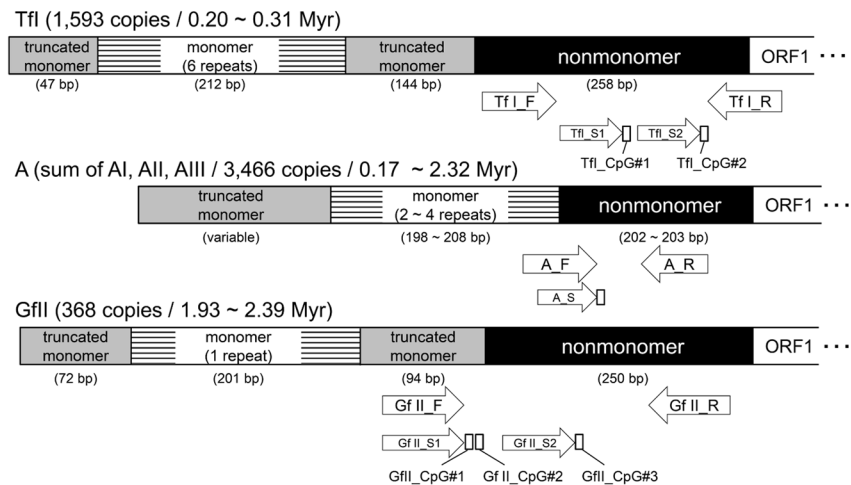
A large proportion of the mammalian genome is occupied by transposons and their related sequences, including a retrotransposon called long interspersed nuclear element-1 (LINE-1). LINE-1 is an approximately 6-kb genomic element, composing approximately 20% of the mammalian genome^{1–3}. The total amount of LINE-1 in mouse is estimated to approximately 100,000 copies in mice⁴. A full-length LINE-1 contains 5' untranslated region (UTR), open reading frame (ORF) 1, ORF2, and 3'-UTR, and it can amplify its copy number in the genome via transcription and reverse transcription through a process called retrotransposition^{5,6}.

In humans, only the youngest LINE-1 subfamily, Hs, retains retrotransposition activity⁷. The DNA methylation level of the 5'-UTR of LINE-1 Hs is widely used to estimate the global DNA methylation level in the human genome⁸, and it can be measured in cells or tissues affected by various disorders, such as cancer^{9–14}, tumors¹⁵, heart disease¹⁶, and neuropsychiatric disorders^{17–20}.

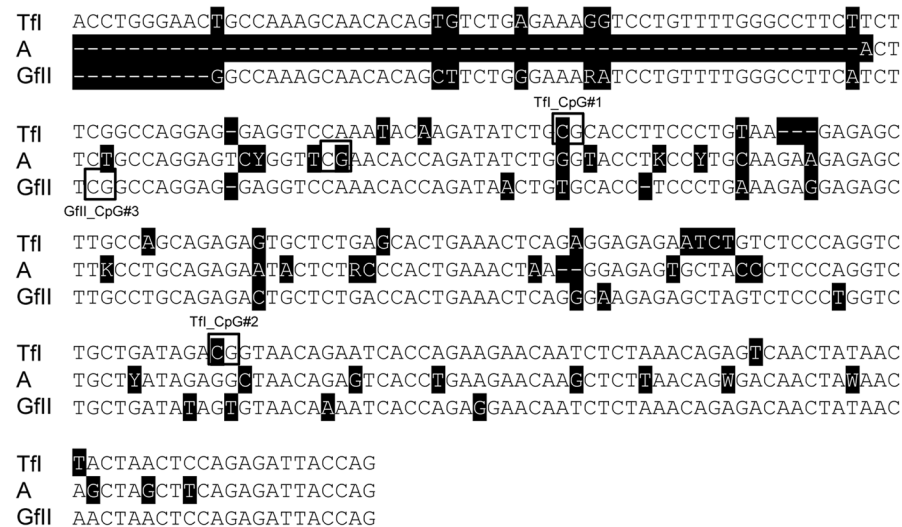
The structure of LINE-1 varies considerably across species. In mice, at least three LINE-1 subfamilies (Tf, A, and Gf) retain retrotransposition activities, representing more than 9,000 full-length copies in the mouse genome^{4,21–23}. The total number of active LINE-1 elements in mice is more than 10-fold higher compared to humans²⁴. In addition, repeat tandems, called monomers, occur within the 5'-UTR, and are not found in the human LINE-1²⁵. Therefore, DNA methylation assays developed for the 5'-UTR of human LINE-1 cannot be applied to other species.

¹Department of Molecular Brain Science, Graduate School of Medical Sciences, Kumamoto University, 1-1-1 Honjo, Chuo-ku, Kumamoto City, Kumamoto, 860-8556, Japan. ²PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi City, Saitama, 332-0012, Japan. ³Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-city, Saitama, 351-0198, Japan. ⁴Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan. Correspondence and requests for materials should be addressed to K.I. (email: iwamotok@kumamoto-u.ac.jp)

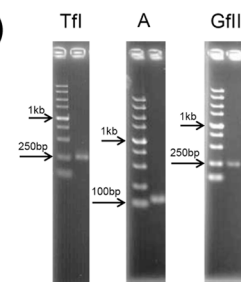
a)



b)



c)



d)

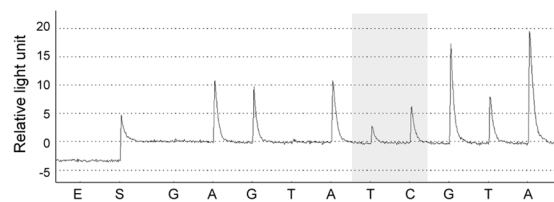


Figure 1. Pyrosequencing-based DNA methylation assay of the active LINE-1 subfamilies in mice. **(a)** Schematic representation of 5'-UTR of the active LINE-1 (Tfl, A, and GfII) in mice, drawn from previous studies^{23,35}. Designed primers and examined CpG sites are indicated. 5'-UTR of the active LINE-1 subfamilies generally consists of a variable number of monomers and truncated monomers as well as a nonmonomer region upstream of ORF1. Definitions of monomers and nonmonomers were determined previously^{23,35}. Numbers in brackets indicate the total copy number of full length elements, and rough age of the subfamily⁴. Myr, million years; ORF, open reading frame; F, forward primer; R, reverse primer; S, sequence primer. **(b)** Alignment of the nonmonomer sequences of Tfl, A, and GfII. Nonconserved sequences among these subfamilies are highlighted in black. The analyzed CpGs within the nonmonomer sequences (Tfl_CpG#1, Tfl_CpG#2, A, and GfII_CpG#3) are boxed. Note that GfII_CpG#1 and GfII_CpG#2 are located in the truncated monomer sequence. **(c)** Agarose gel electrophoresis analysis of the PCR amplicons. The expected amplicon sizes of Tfl, A, and GfII are 261 bp, 129 bp, and 244 bp, respectively. A 100-bp DNA ladder (TakaraBio) was used. Arrows indicate respective sizes

of bands. **(d)** A representative pyrogram in pyrosequencing analysis (TfI_CpG#2). The pyrosequencing reaction starts with an input of enzyme (E) followed by substrates (S). The sequence after S is as listed as dispensation order in Table 1. The shaded site is the cytosine analyzed in this study.

We developed pyrosequencing-based assays to examine the levels of DNA methylation and hydroxymethylation in the 5'-UTR of the active LINE-1 subfamilies in mice. Hydroxymethylcytosine (hmC), which is widely distributed across tissues and especially enriched in the brain, is modified from methylcytosine (mC) by the ten-eleven translocation (TET) enzymes^{26–28}. Because the widely used sodium bisulfite treatment cannot distinguish between mC and hmC, we used oxidation treatment prior to bisulfite modification²⁹. Using the developed assay, we examined cytosine modification status in several adult brain regions (frontal cortex, hippocampus, cerebellum, and basal ganglia) and in nonbrain tissues including tail, heart, testis, and ovary. We observed tissue- and subfamily-specific mC and hmC differences, which suggest complex epigenetic regulation of the active LINE-1 subfamilies in adult tissues in mice.

Results

Design of pyrosequencing-based assay. Consensus sequences of active LINE-1 subfamilies in mice (Tf, A, and Gf), which were characterized in a previous study⁴, were retrieved from Repbase^{30,31}. Based on these sequences, we designed PCR primers to the truncated monomer or nonmonomer region of each subfamily for a pyrosequencing assay. We successfully developed the pyrosequencing assay for TfI, A (including AI, AII, and AIII), and GfII (Fig. 1a,b and Table 1). In these assays, we obtained single-banded expected size of PCR amplicons (Fig. 1c), and expected peaks were deduced from the consensus sequence in the pyrogram (Fig. 1d). We confirmed the specificity of each assay by TA-cloning the bisulfite PCR products followed by Sanger sequencing. In the amplification step, these assays showed 95%, 97% and 100% specificity for types TfI, A and GfII, respectively (see Supplementary Figs S1 to S3). Linearity and sensitivity of the assay were assessed using the synthetic DNA samples with different methylation levels. We obtained high correlations ($R > 0.998$, see Supplementary Fig. S4).

mC and hmC levels of active LINE-1 subfamilies in mouse tissues. Using the developed assay, we measured mC and hmC levels in various brain and nonbrain tissues of adult mice. By using oxidation reaction before bisulfite modification²⁹, the actual mC level at each CpG site was determined. The hmC level could be estimated by subtracting the actual mC level from the bisulfite modification data. We did not find significant differences in mC or hmC levels among the four brain regions, frontal cortex (FC), hippocampus (Hp), cerebellum (Cb), and basal ganglia (BG), using analysis of variance (ANOVA) (Fig. 2). At the subfamily level, we found a relatively high hmC level in the TfI subfamily across all brain regions, while others showed lower or negligible levels, except in the FC and BG in A subfamily.

In contrast, we observed significant tissue- and subfamily-specific differences in mC and hmC levels in non-brain tissues (Fig. 3). For TfI (TfI_CpG#2), mC levels of the tail and testis were higher, while hmC levels were lower, compared with the brain and heart. In the nonmonomer regions of A and GfII (GfII_CpG#3), the mC level of the heart was lower and that of the testis was higher compared with the other tissues. In the truncated monomer region of GfII (GfII_CpG#1 and #2), nonbrain tissues showed lower mC and higher hmC levels to a similar extent compared with the brain.

Correlation of mC levels between tissues or subfamilies. We calculated pair-wise correlations of mC levels across all tissues (Fig. 4). As expected, each brain region showed strong correlations with the other three brain regions ($R > 0.8$). Interestingly, among the tissues tested, levels in the testis were poorly correlated with those in other tissues. Similarly, we calculated pair-wise correlations of mC levels across all CpG sites (Fig. 5). We found significant strong correlations between CpG sites for GfII_CpG#1 and GfII_CpG#2, as well as between GfII_CpG#3 and A (Fig. 5).

Discussion

We developed pyrosequencing-based DNA methylation assays of the 5'-UTR in the active LINE-1 subfamilies TfI, A, and GfII in mice, and we evaluated both mC and hmC levels of the brain regions and several other tissues. Given that there are about 1,600, 3,500, and 370 full-length copies in TfI, A, and GfII subfamilies, respectively⁴, the mC and hmC levels obtained with these assays indicate the average values for each subfamily.

Measuring the mC level at the monomer region is difficult, due to the variable number of monomer unit repeats. Therefore, we focused on the CpGs in the nonmonomer or truncated monomer regions. Compared to the monomer regions, the number of CpGs in nonmonomer regions seemed to be depleted. Consequently, in TfI, two out of three CpGs in the nonmonomer region were covered by our assay. Although there was only one conserved CpG in the nonmonomer region in both A and GfII, we successfully included these CpGs in the pyrosequencing assays (Fig. 1a,b).

Subfamily- and tissue-dependent epigenetic differences detected in this study suggest that each of the active LINE-1 subfamilies may have distinct roles in the adult brain and other tissues, which have not been well addressed to date. The four brain regions examined showed less variation in mC levels, although subfamily-dependent variations in hmC levels were observed. Among the nonbrain tissues examined, the testis had distinctive epigenetic profiles as exemplified by the low correlation coefficients in comparisons with the other tissues (Fig. 4). This finding might be related to the unique epigenetic regulation of transposons in spermatogenesis³². However, the tissues used in this analysis, except for the testis, were derived from female mice. Therefore, our correlation analysis may be affected by the sex difference. Overall, the hmC levels were relatively low (Figs 2

Subfamily	Primer type	Primer name	Primer sequence	Sequence to analyze	Dispensation order	CpG#
TfII	Forward	TfI_F	TTTGGGAATTGTTAAAGTAATATAG			
	Reverse	TfI_R	bio-CCATCTAATAATCTCTAAAATTAATA			
	Sequence 1	TfI_S1	GGAGGAGGTTTAAAT	ATAAGATATTTG <u>Y</u> GTA	TATAGATATAGTCGT	TfI_CpG#1
	Sequence 2	TfI_S2	TGTTTTTAGGTTTGTGAT	AGA <u>Y</u> GGTAATAGAATTAT	GAGTATCGTA	TfI_CpG#2
A	Forward	A_F	GTGAGTGGAATATAATTTTGTAGGA			
	Reverse	A_R	bio-AAAAATAACACTCTCCTTAATTTCAATAA			
	Sequence	A_S	GGAATATAATTTTGTAGG	AGTTYGGTT <u>Y</u> GAAATATTAGATATTTGGGTATT	GATGTCAGTCGAT	
GfII	Forward, Sequence 1	GfII_F, GfII_S1	GGGGTTATTTGATTTTG	GGATTT <u>Y</u> GTAG <u>Y</u> GGGTA	AGTATTCGTATGTCGT	GfII_CpG#1, CpG#2
	Reverse	GfII_R	bio-TTCCTCTAATAATTTTATTACACTA			
	Sequence 2	GfII_S2	TTGTTTTGGGTTTTTATTTT	<u>Y</u> GGTTAGGAGG	ATCGTAGA	GfII_CpG#3

Table 1. List of the primers designed, the analyzed sequences, and the dispensation orders. The 5'-biotinylated primers are indicated by the "bio-". Note that one amplicon is analyzed by two different sequencing primers in TfII and GfII. In GfII, the forward primer is also used as a sequencing primer (GfII_S1).

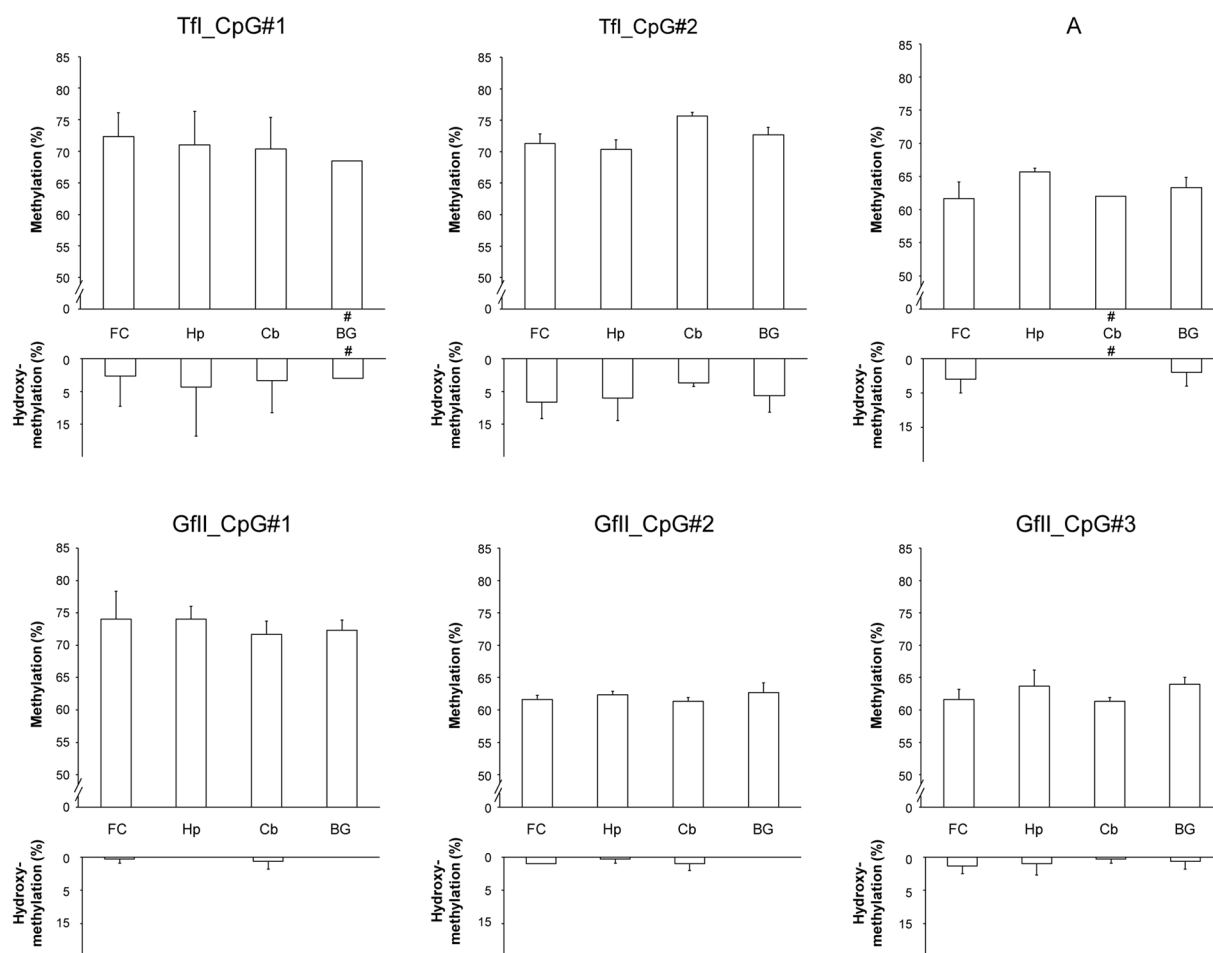


Figure 2. mC and hmC levels of the active LINE-1 subfamilies in the adult mouse brain regions. For TfII, two CpG sites located within the nonmonomer region; for A, one CpG site located within the nonmonomer region; and for GfII, three CpG sites (two located within the truncated monomer, and one within the nonmonomer region) were analyzed. Three samples were available for analysis for each CpG site, except for the samples indicated by # symbol, for which only two were available. Values are given as mean \pm standard deviation. mC, DNA methylation; hmC, hydroxymethylation; FC, frontal cortex; Hp, hippocampus; Cb, cerebellum; BG, basal ganglia.

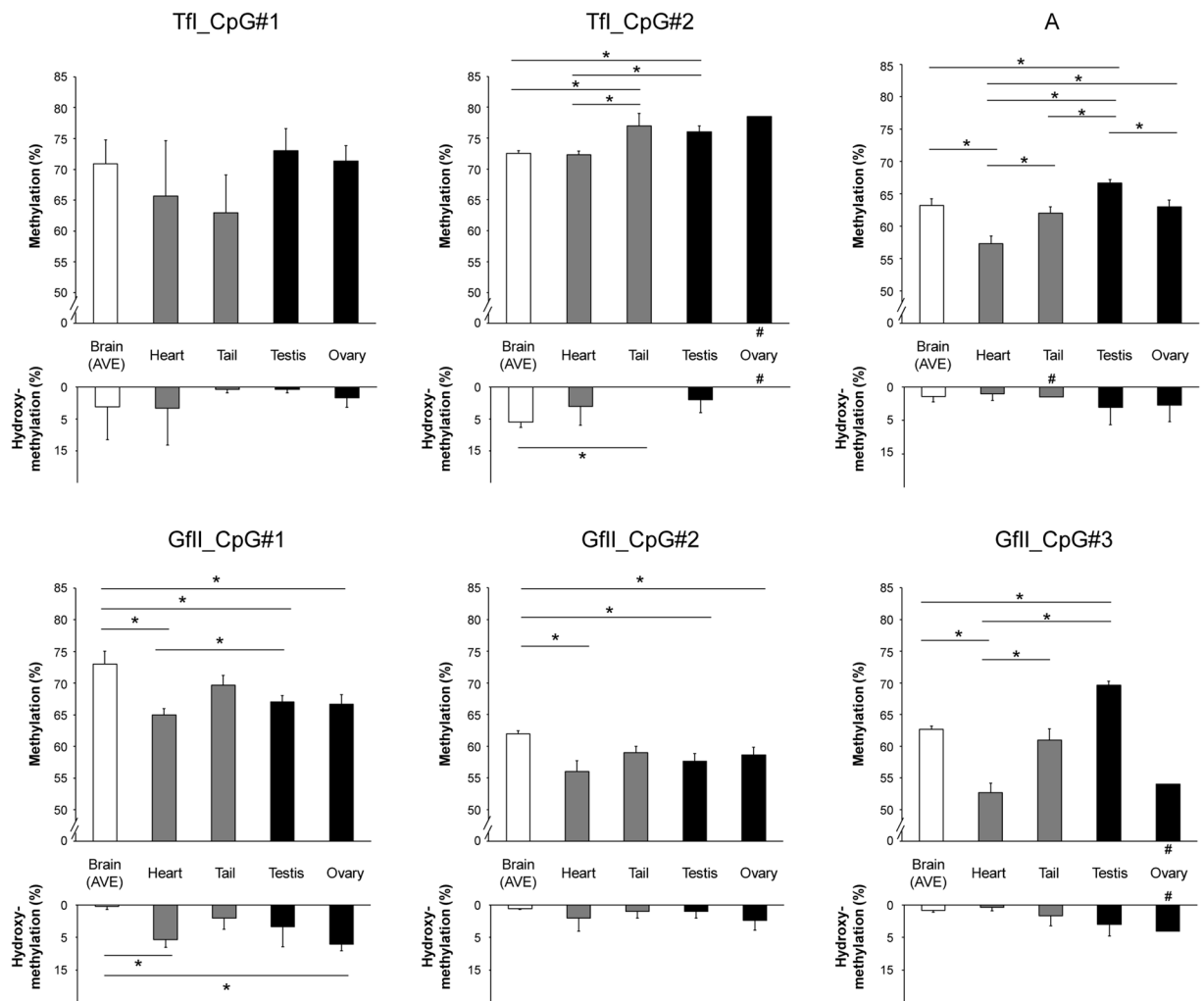


Figure 3. mC and hmC levels of the active LINE-1 subfamilies in the brain and nonbrain tissues. The data from four brain regions were averaged and treated as a single tissue, Brain (AVE). Three samples were available for analysis for each CpG site, except for the samples indicated by # symbol, for which only two were available. Values are given as mean \pm standard deviation. *Statistically significant (ANOVA followed by Tukey's test, $P < 0.05$). mC, DNA methylation; hmC, hydroxymethylation.

	Frontal cortex	Hippocampus	Cerebellum	Basal ganglia	Brain (AVE)	Heart	Tail	Testis	Ovary
Frontal cortex		0.884	0.901	0.861	0.963	0.787	0.664	0.494	0.742
Hippocampus	<0.001		0.838	0.863	0.939	0.707	0.627	0.445	0.664
Cerebellum	<0.001	<0.001		0.904	0.958	0.941	0.868	0.640	0.887
Basal ganglia	<0.001	<0.001	<0.001		0.947	0.821	0.854	0.602	0.746
Brain (AVE)	<0.001	<0.001	<0.001	<0.001		0.849	0.764	0.568	0.801
Heart	<0.001	0.001	<0.001	<0.001	<0.001		0.859	0.630	0.895
Tail	0.003	0.005	<0.001	<0.001	<0.001	<0.001		0.639	0.773
Testis	0.037	N.S.	0.006	0.011	0.014	0.005	0.004		0.677
Ovary	0.001	0.005	<0.001	0.001	<0.001	<0.001	<0.001	0.004	

Figure 4. Correlations of actual mC levels of the active LINE-1 subfamilies between two tissues. Pair-wise Pearson's correlation between two tissues was calculated using actual mC levels of two CpG sites from Tfl, one CpG site from A, and three CpG sites from Gfil. *R* and *P* values are indicated in upper right and lower left, respectively. Data from four brain regions were averaged and treated as a single brain tissue, Brain (AVE).

and 3), with individual levels ranging from 0% up to 13%. In general, the brain tissues showed higher hmC levels than the nonbrain tissues for Tfl (Fig. 3). In contrast, for Gfil, the brain tissues were poorly hydroxymethylated, compared with the other nonbrain tissues.

A strong positive correlation in mC levels was found between Gfil_CpG#1 and Gfil_CpG#2 (Fig. 5). Since these two CpGs were located within the truncated monomer region (Fig. 1), they are likely to be under the same epigenetic regulation. Unexpectedly, the mC level of A showed a strong correlation with Gfil_CpG#3, but not with other CpG sites in Gfil (Gfil_CpG#1 or Gfil_CpG#2) or Tfl (Tfl_CpG#1 or Tfl_CpG#2). This observation

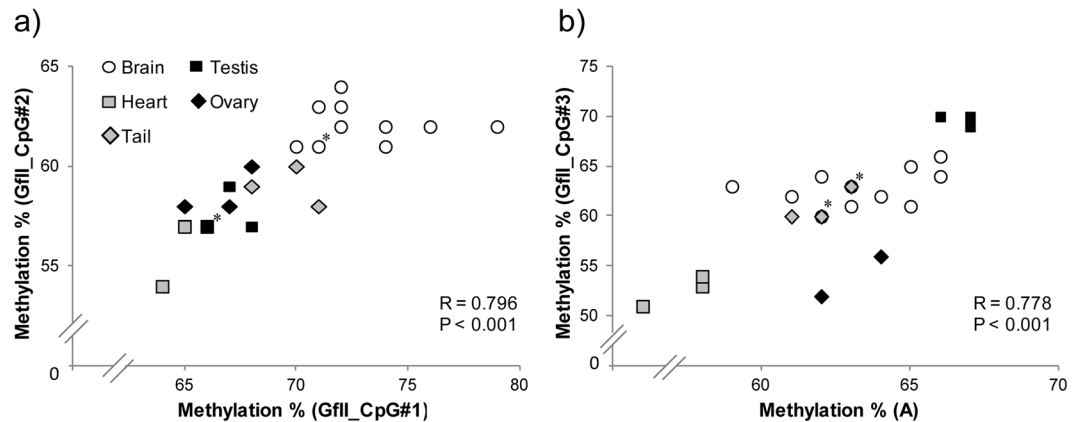


Figure 5. Correlation of actual mC levels between different two CpG sites. Pairs of (a) GfII_CpG#1 and GfII_CpG#2, and (b) A and GfII_CpG#3 were the only ones that showed significant correlations ($P < 0.05$ after Bonferroni correction). Each symbol represents data from one individual tissue ($N = 12$ for brain regions and $N = 3$ each for other tissues, except for the ovary tissue in GfII_CpG#3, for which only two were available). Symbols with * on the right corner indicate that two samples are overlapped on each other in the graph, because they showed the same DNA methylation levels.

may be attributable to their evolutionary relationship. Phylogenetic analysis based on the conserved ORF2 sequences previously revealed that A (AI, AII, and AIII) and GfII are clustered in the same phylogenetic branch⁴. Therefore, epigenetic status of each subfamily may reflect evolutionary aspect, despite that A family consists of A-type monomer and Gf family has F-type monomer, which is distinct from A-type in terms of sequence and structure of monomer region³³.

The pyrosequencing-based assays we developed would be useful for rapid and accurate high-throughput assessment of epigenetic regulation of the active LINE-1 subfamilies in mice.

Materials and Methods

Consensus sequences of mouse LINE-1 elements. We retrieved consensus sequences of the active LINE-1 subfamilies in mice (Tf, A, and Gf), which were characterized in a previous study⁴, from Repbase^{30,31}. Because they contained a variable number of monomer units, we tried to design primer pairs within the nonmonomer or truncated monomer region to avoid amplifying nonspecific sequences (Fig. 1 and Table 1). Because the consensus sequences were very similar among the A subfamilies (AI, AII, and AIII), the primers were designed to amplify all of these subfamilies.

Sample preparation. Adult wild-type mice (C57BL/6) were used for collection of tissues, including frontal cortex, hippocampus, cerebellum, basal ganglia, heart, tail, testis, and ovary. All tissues were collected from female mice, except for the testis, which were collected from males ($N = 3$ for each tissue, unless otherwise described). The tissues were stored at -80°C until use. The tissues were treated with 0.1 mg/ml of proteinase K (Roche) dissolved in DNA extraction buffer, and incubated at 55°C overnight. Genomic DNA was extracted using the standard phenol–chloroform methods. All animal experiments were approved by the local animal experiment committees of RIKEN (Wako, Saitama, Japan). Animal experiments were carried out in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals. All efforts were made to minimize the number of animals used and their suffering.

Synthetic DNA samples were prepared to test the linearity and sensitivity of the assay. Sequences of fully-methylated (i.e. -CG-), and unmethylated (i.e. -TG-) were purchased from a manufacturer (FASMAC) (see Supplementary Table S1). These samples were mixed together to prepare different DNA methylation levels (0~100%, 10% intervals), and subjected to the assays as described below. Between the methylation range where we detected in the actual mouse samples (i.e. 50~70% for A and 50~80% for TfI and GfII), we prepared samples with 5% interval levels. Ninety-picograms, or two-hundred and ninety picograms of the mixed DNA samples were used in PCR for A, TfI and GfII subfamilies, respectively.

Oxidative treatment and sodium bisulfite modification. Oxidative and sodium bisulfite treatments were performed using TrueMethyl Kit (Cambridge Epigenetix), according to the manufacturer's instructions. In brief, 1.5- μg DNA samples were fragmented to approximately 5 kb by sonication (Covaris). Half the volume of the sample (equivalent to 750 ng of DNA) was subjected to oxidation followed by bisulfite modification (oxidative bisulfite sequencing, oxBS), and the other half was used for bisulfite modification only (bisulfite sequencing, BS). After purification using bead columns, both oxBS and BS samples were denatured with 50 mM NaOH at 37°C for 30 min. One microliter of oxidation solution included in the kit was added to the oxBS sample, while 1 μl of water was added to BS sample. Samples were incubated at 40°C for 30 min in a thermal cycler and then centrifuged for 10 min at 14,000 G. The supernatants were removed and directly underwent sodium bisulfite treatment following the manufacturer's protocol. The samples were then desulfonated and washed with the provided buffers. All reagents and columns were included in the kit (Cambridge Epigenetix).

Bisulfite PCR for pyrosequencing assay. One microliter of oxBS or BS DNA sample was used for PCR amplification. The reaction mixture contained the following reagents; 1 × PCR amplification buffer (Invitrogen), 1 M betaine, 0.2 mM dNTP (TakaraBio), 3.0 mM MgCl₂, 0.4 mM primers including biotin-labeled primer, 2.0 ng single-stranded DNA binding protein (Promega), and 5 U of Platinum *Taq* DNA polymerase (Invitrogen). PCR conditions were as follows: 3 min at 95 °C followed by 40 cycles of 10 s at 98 °C, 30 s at 58 °C, and 30 s at 72 °C for Tfl and GflI; 3 min at 95 °C followed by 40 cycles of 10 s at 98 °C, 30 s at 55 °C, and 30 s at 72 °C for A. Primer sequences are listed in Table 1.

Pyrosequencing. Pyrosequencing was performed as described previously³⁴. Briefly, the PCR mixture was treated with 4 μl of streptavidin-sepharose beads (Amersham Biosciences) and 54 μl of 2 × binding buffer (Qiagen). The samples were denatured with 0.2 N NaOH and washed thoroughly with washing buffer (Qiagen). The beads were then suspended in 50 μl of annealing buffer (Qiagen) containing 0.2 μM of the sequencing primers (Table 1), and the suspension was heated to 96 °C for 2 min. Pyrosequencing was performed using a Pyromark Gold Q96 Reagents kit (Qiagen) with a PSQ 96MA instrument (Qiagen) following the manufacturer's protocols.

TA cloning and Sanger sequencing. PCR products prepared as described above were subjected to TA-cloning using TOPO TA Cloning kit (Thermo Fisher Scientific) and One Shot TOP10 Chemically Competent *E. coli* (Thermo Fisher Scientific), according to the manufacturer's instructions. After transformation, the single colonies were checked by colony PCR using M13 Forward (5'-GTAAACGACGGCCAG-3') and M13 Reverse primer (5'-CAGGAAACAGCTATGAC-3'). Sanger sequencing was then performed using T3 primer (5'-ATTAACCTCACTAAAGGGA-3') (Eurofins Genomics K.K.).

Data analysis. The mC level was estimated from the oxBS sample by pyrosequencing. The hmC level was estimated by subtracting the oxBS level from the BS level. Statistical analysis was performed using Student's t-test for the comparison of the two groups. For other tests, analysis of variance (ANOVA) followed by Tukey's test was conducted. For correlation analysis, Pearson's correlation test was used. For the purpose of comparison across different tissues, data from the four brain regions were averaged and treated as a single tissue. $P < 0.05$ was considered significant.

Sequence data of bisulfite PCR amplicon was processed using Sequencher (ver. 4.10.1., Gene Code Corporation) and aligned with the consensus sequences of the target subfamily using Genetyx (ver. 13, Genetyx Corporation). Phylogenetic trees were drawn using UPGMA method, implemented in Genetyx.

References

- Gibbs, R. A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521, <https://doi.org/10.1038/nature02426> (2004).
- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921, <https://doi.org/10.1038/35057062> (2001).
- Mouse Genome Sequencing, C. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562, <https://doi.org/10.1038/nature01262> (2002).
- Sookdeo, A., Hepp, C. M., McClure, M. A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* **4**, 3, <https://doi.org/10.1186/1759-8753-4-3> (2013).
- Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691–703, <https://doi.org/10.1038/nrg2640> (2009).
- Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632, <https://doi.org/10.1126/science.1089670> (2004).
- Skowronski, J., Fanning, T. G. & Singer, M. F. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol* **8**, 1385–1397 (1988).
- Yang, A. S. *et al.* A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* **32**, e38, <https://doi.org/10.1093/nar/gnh032> (2004).
- Bollati, V. *et al.* Differential repetitive DNA methylation in multiple myeloma molecular subgroups. *Carcinogenesis* **30**, 1330–1335, <https://doi.org/10.1093/carcin/bgp149> (2009).
- Balassiano, K. *et al.* Aberrant DNA methylation of cancer-associated genes in gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Cancer Lett* **311**, 85–95, <https://doi.org/10.1016/j.canlet.2011.06.038> (2011).
- Ikeda, K. *et al.* Long interspersed nucleotide element 1 hypomethylation is associated with poor prognosis of lung adenocarcinoma. *Ann Thorac Surg* **96**, 1790–1794, <https://doi.org/10.1016/j.athoracsur.2013.06.035> (2013).
- Iwagami, S. *et al.* LINE-1 hypomethylation is associated with a poor prognosis among patients with curatively resected esophageal squamous cell carcinoma. *Ann Surg* **257**, 449–455, <https://doi.org/10.1097/SLA.0b013e31826d8602> (2013).
- Park, S. Y. *et al.* Alu and LINE-1 hypomethylation is associated with HER2 enriched subtype of breast cancer. *PLoS One* **9**, e100429, <https://doi.org/10.1371/journal.pone.0100429> (2014).
- Harada, K. *et al.* LINE-1 methylation level and patient prognosis in a database of 208 hepatocellular carcinomas. *Ann Surg Oncol* **22**, 1280–1287, <https://doi.org/10.1245/s10434-014-4134-3> (2015).
- Igarashi, S. *et al.* A novel correlation between LINE-1 hypomethylation and the malignancy of gastrointestinal stromal tumors. *Clin Cancer Res* **16**, 5114–5123, <https://doi.org/10.1158/1078-0432.CCR-10-0581> (2010).
- Baccarelli, A. *et al.* Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology* **21**, 819–828, <https://doi.org/10.1097/EDE.0b013e3181f20457> (2010).
- Bollati, V. *et al.* DNA methylation in repetitive elements and Alzheimer disease. *Brain Behav Immun* **25**, 1078–1083, <https://doi.org/10.1016/j.bbi.2011.01.017> (2011).
- Burghardt, K. J., Pilsner, J. R., Bly, M. J. & Ellingrod, V. L. DNA methylation in schizophrenia subjects: gender and MTHFR 677C/T genotype differences. *Epigenomics* **4**, 261–268, <https://doi.org/10.2217/epi.12.25> (2012).
- Misiak, B. *et al.* Lower LINE-1 methylation in first-episode schizophrenia patients with the history of childhood trauma. *Epigenomics*, 1–11, <https://doi.org/10.2217/epi.15.68> (2015).
- Viana, J. *et al.* Epigenomic and transcriptomic signatures of a Klinefelter syndrome (47,XXY) karyotype in the brain. *Epigenetics* **9**, 587–599, <https://doi.org/10.4161/epi.27806> (2014).

21. Martin, S. L. Characterization of a LINE-1 cDNA that originated from RNA present in ribonucleoprotein particles: implications for the structure of an active mouse LINE-1. *Gene* **153**, 261–266 (1995).
22. Saxton, J. A. & Martin, S. L. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *J Mol Biol* **280**, 611–622, <https://doi.org/10.1006/jmbi.1998.1899> (1998).
23. Severynse, D. M., Hutchison, C. A. 3rd & Edgell, M. H. Identification of transcriptional regulatory activity within the 5' A-type monomer sequence of the mouse LINE-1 retroposon. *Mamm Genome* **2**, 41–50 (1992).
24. Ostertag, E. M. & Kazazian, H. H. Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* **35**, 501–538, <https://doi.org/10.1146/annurev.genet.35.102401.091032> (2001).
25. Adey, N. B., Schichman, S. A., Hutchison, C. A. 3rd & Edgell, M. H. Composite of A and F-type 5' terminal sequences defines a subfamily of mouse LINE-1 elements. *J Mol Biol* **221**, 367–373 (1991).
26. Kinney, S. M. *et al.* Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *J Biol Chem* **286**, 24685–24693, <https://doi.org/10.1074/jbc.M110.217083> (2011).
27. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930, <https://doi.org/10.1126/science.1169786> (2009).
28. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935, <https://doi.org/10.1126/science.1170116> (2009).
29. Booth, M. J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937, <https://doi.org/10.1126/science.1220671> (2012).
30. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**, 11, <https://doi.org/10.1186/s13100-015-0041-9> (2015).
31. Jurka, J. Repeats in genomic DNA: mining and meaning. *Curr Opin Struct Biol* **8**, 333–337 (1998).
32. Frost, R. J. *et al.* MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proc Natl Acad Sci USA* **107**, 11847–11852, <https://doi.org/10.1073/pnas.1007158107> (2010).
33. Padgett, R. W., Hutchison, C. A. 3rd & Edgell, M. H. The F-type 5' motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. *Nucleic Acids Res* **16**, 739–749 (1988).
34. Ikegame, T. *et al.* DNA methylation analysis of BDNF gene promoters in peripheral blood cells of schizophrenia patients. *Neurosci Res* **77**, 208–214, <https://doi.org/10.1016/j.neures.2013.08.004> (2013).
35. Goodier, J. L., Ostertag, E. M., Du, K. & Kazazian, H. H. Jr. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res* **11**, 1677–1685, <https://doi.org/10.1101/gr.198301> (2001).

Acknowledgements

This work was supported by MEXT/JSPS KAKENHI Grant Numbers, 15J03188, 24116009, 15H04891, 15K09801, and 24116005. This research is partially supported by the Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS), the Strategic Research Program for Brain Sciences, and the Advanced Research and Development Programs for Medical Innovation from Japan Agency for Medical Research and development (AMED). We would like to thank Christine Clark, Jason Mellad, and Hanna Schutz from Cambridge Epigenetix for the advice and technical support for this study.

Author Contributions

Y.M. participated in designing the study, performing all experiments and statistical analyses, and drafting the manuscript. K.I. and M.B. participated in directing the study and finalizing the manuscript. J.U., M.K.S. and T.K. helped performing the experiments and preparing the manuscript. K.K. contributed for finalizing the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-14165-7>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017