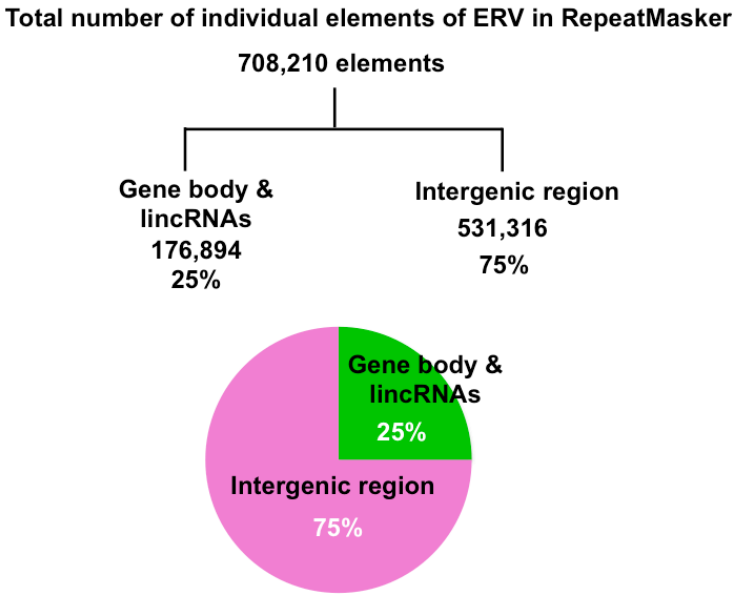
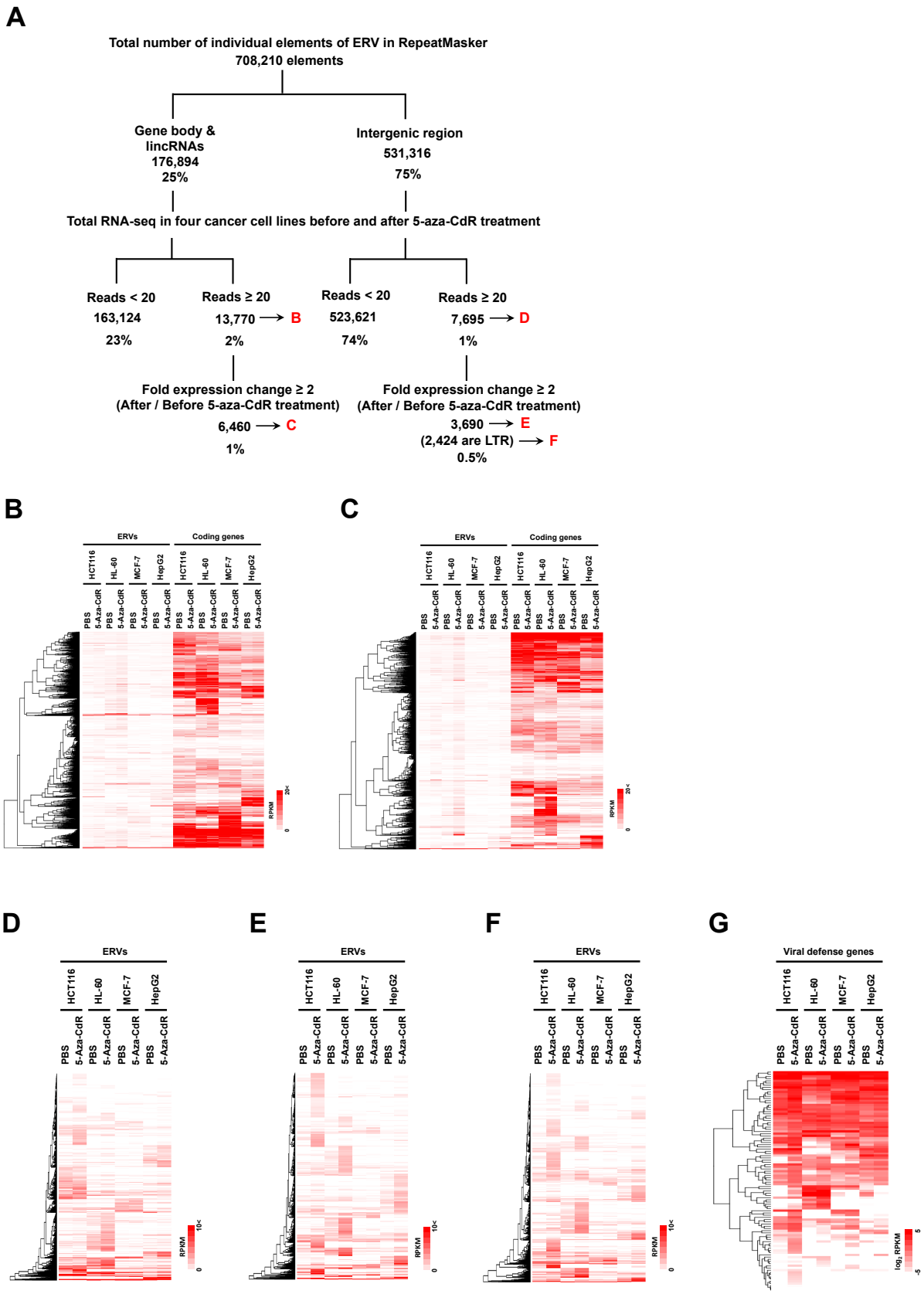


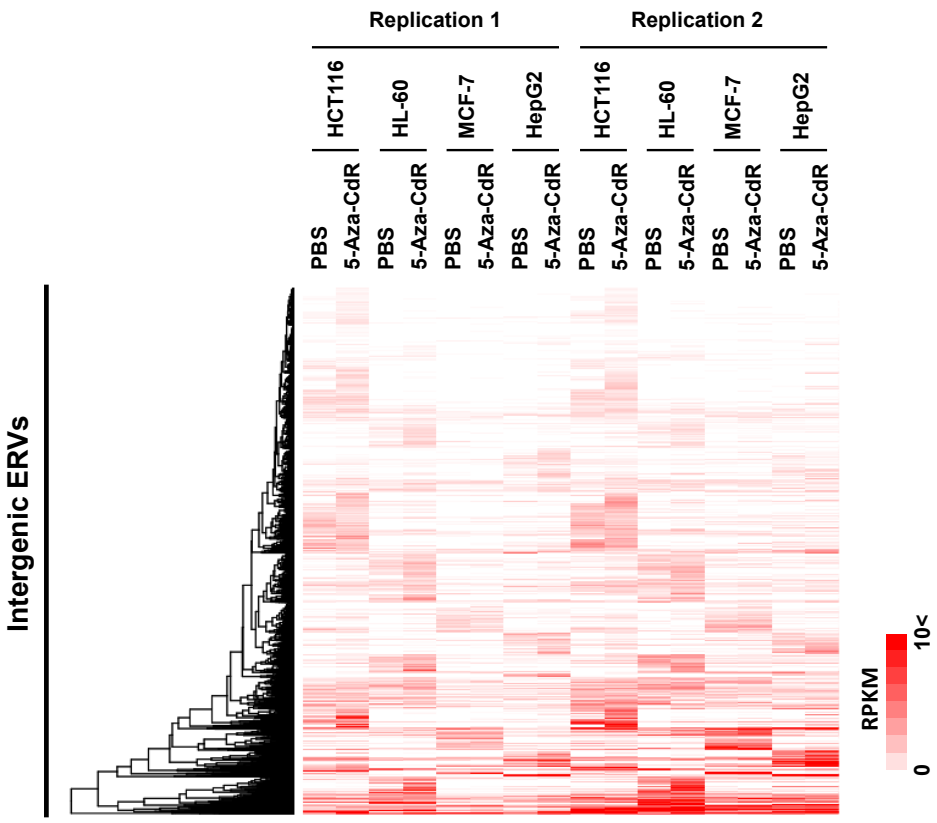
Supplemental Figure S1. Genomic distribution of individual ERV elements in the human genome (for Figure 1A).

Removed from this analysis were the 172,505 ERV elements that overlapped coding gene bodies (beyond the region from transcription start site to end site) and the 4,389 ERVs that overlapped 561 lincRNAs in the NCBI RefSeq database.


A
Total number of individual elements of ERV in RepeatMasker
708,210 elements
Gene body & lincRNAs
176,894
25%
Intergenic region
531,316
75%
Total RNA-seq in four cancer cell lines before and after 5-aza-CdR treatment
Reads < 20
163,124
23%
Reads ≥ 20
13,770 → B
2%
Reads < 20
523,621
74%
Reads ≥ 20
7,695 → D
1%
Fold expression change ≥ 2
(After / Before 5-aza-CdR treatment)
6,460 → C
1%
Fold expression change ≥ 2
(After / Before 5-aza-CdR treatment)
3,690 → E
(2,424 are LTR) → F
0.5%
B
C
D
E
F
G
ERVs
Coding genes
Viral defense genes
HCT116
HL-60
MCF-7
HepG2
PBS
5-Aza-CdR
RPKM
log2 RPKM

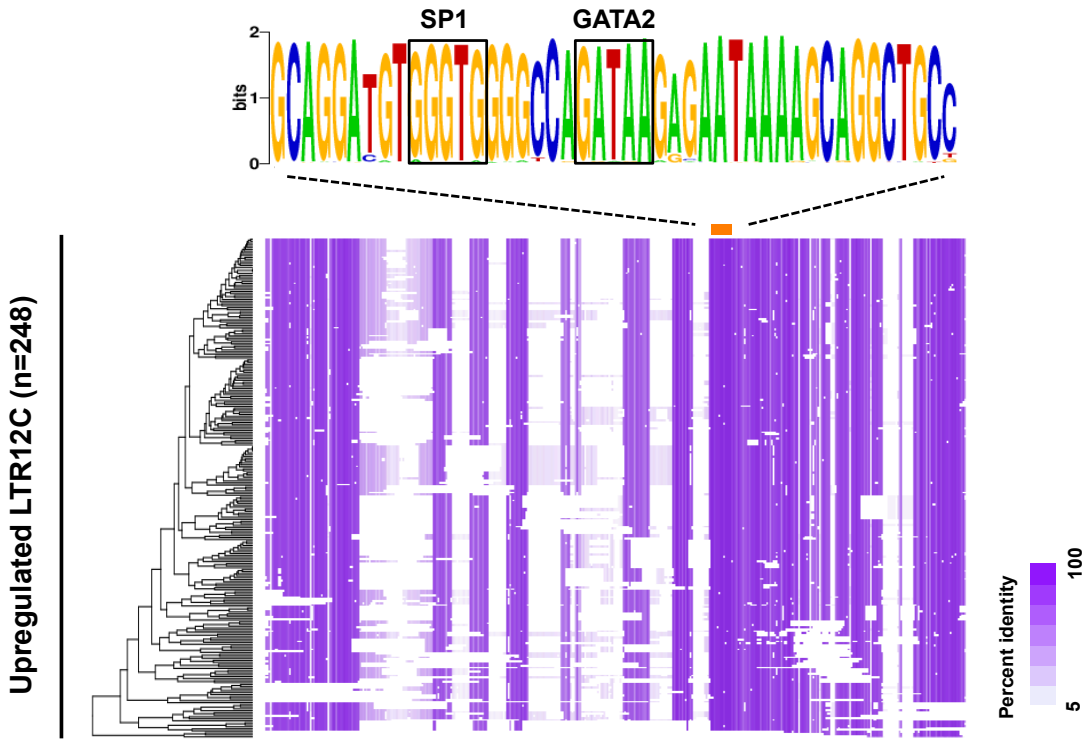
Supplemental Figure S2. Intergenic ERVs are upregulated in four 5-aza-CdR-treated cancer cell lines (for Figure 2B).

(A) Bioinformatic workflow to identify DNA methylation–regulated intergenic LTRs in the four cell lines. (B) The expression level of ERVs and overlapping coding genes. The expression patterns of ERVs and coding genes are overlapped. (C) The upregulation patterns of ERVs and overlapping coding genes are closely correlated. (D) The expression level of intergenic ERVs before and after 5-aza-CdR treatment. (E) Upregulated intergenic ERVs after 5-aza-CdR treatment. (F) Expression level of upregulated ERVs with LTRs after 5-aza-CdR treatment. (G) Expression level of defense response to virus (GO:0051607) and type I interferon signaling pathway (GO:0060337) genes in Gene Ontology. The 100 genes were upregulated at least in one cell line more than 2-fold after 5-aza-CdR treatment, relative to PBS treatment.



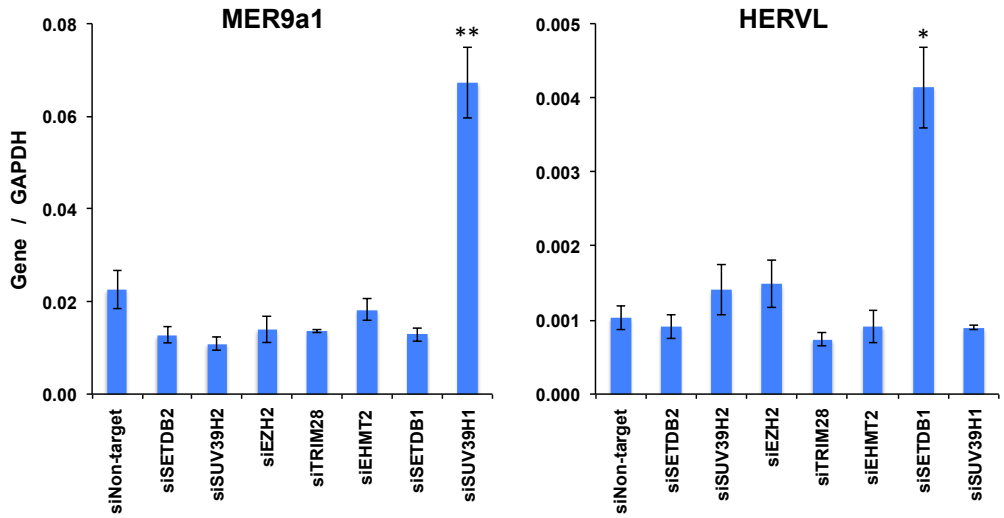
Supplemental Figure S3. The expression patterns of intergenic ERVs are replicated at high level in two independent experiments by RNA-seq before and after 5-aza-CdR treatment in four cancer cell lines (for Figure 2B).

Comparison of two biological replicates of RNA-seq data for intergenic ERVs.



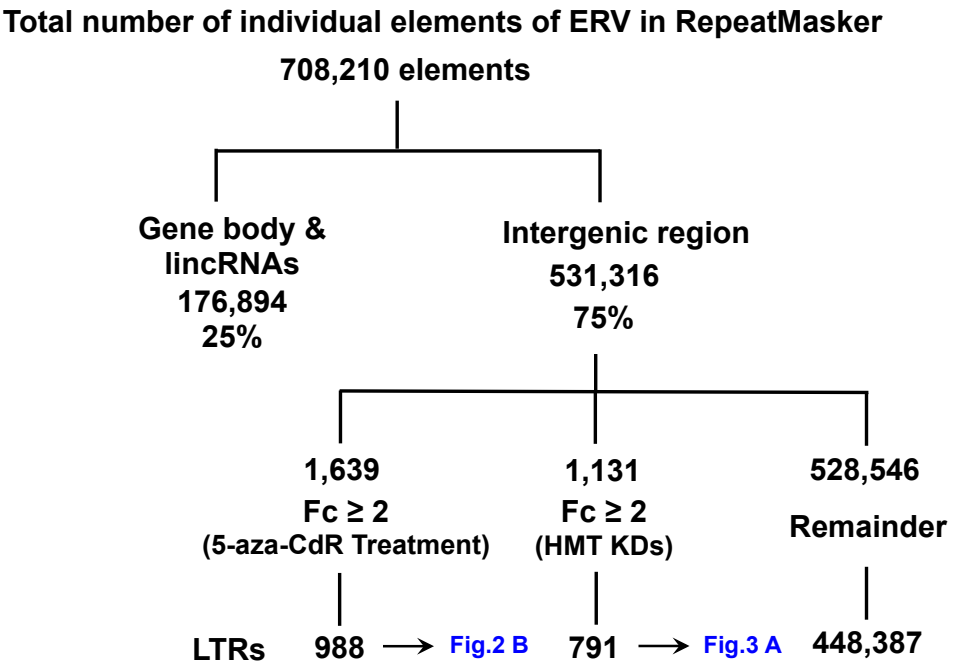
Supplemental Figure S4. Transcription factor binding DNA motifs are conserved in upregulated LTR12C elements by 5-aza-CdR treatment.

The sequence alignment of 248 LTR12C elements upregulated by 5-aza-CdR treatment. Each row represents one LTR12C element. Purple indicates percent identity (5 to 100%); white indicates gaps. A 2-kb flanking region is shown.



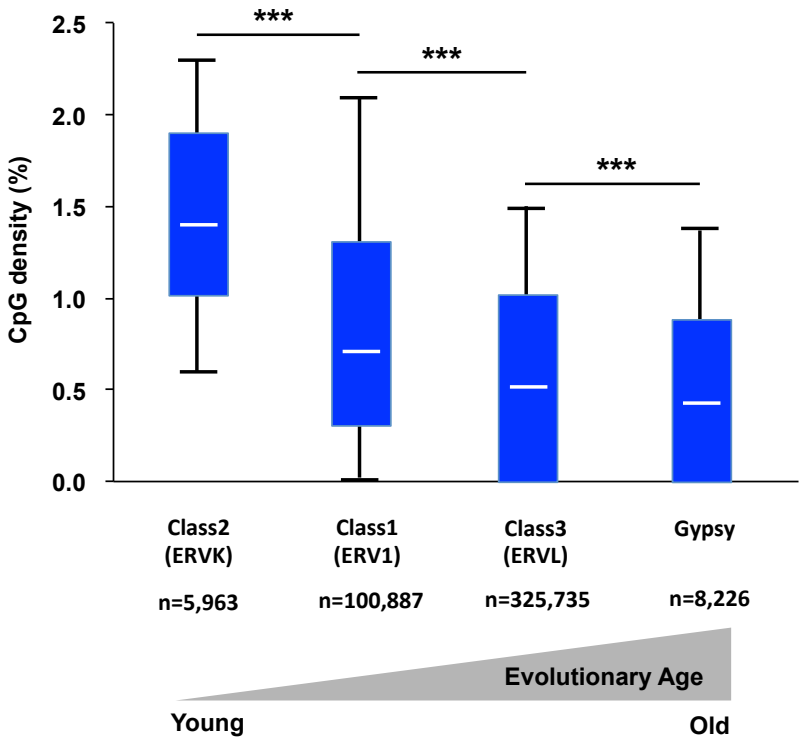
Supplemental Figure S5. Upregulation of intergenic LTRs by HMT KDs (for Figure 3).

The unique upregulation of LTRs in Figure 3B was confirmed by RT-qPCR. MER9a1 and HERVL were selected as examples showing specific upregulation by the SUV39H1 and SETDB1 knockdowns, respectively. Primers of MER9a1 were designed at specific loci; primers of HERVL were designed at multiple loci. Error bars represent SEM from three independent biological replicates. *P*-values were calculated using the two-tailed Student's *t*-test for comparison with the siNon-target sample. (*) $P < 0.05$, (**) $P < 0.01$.



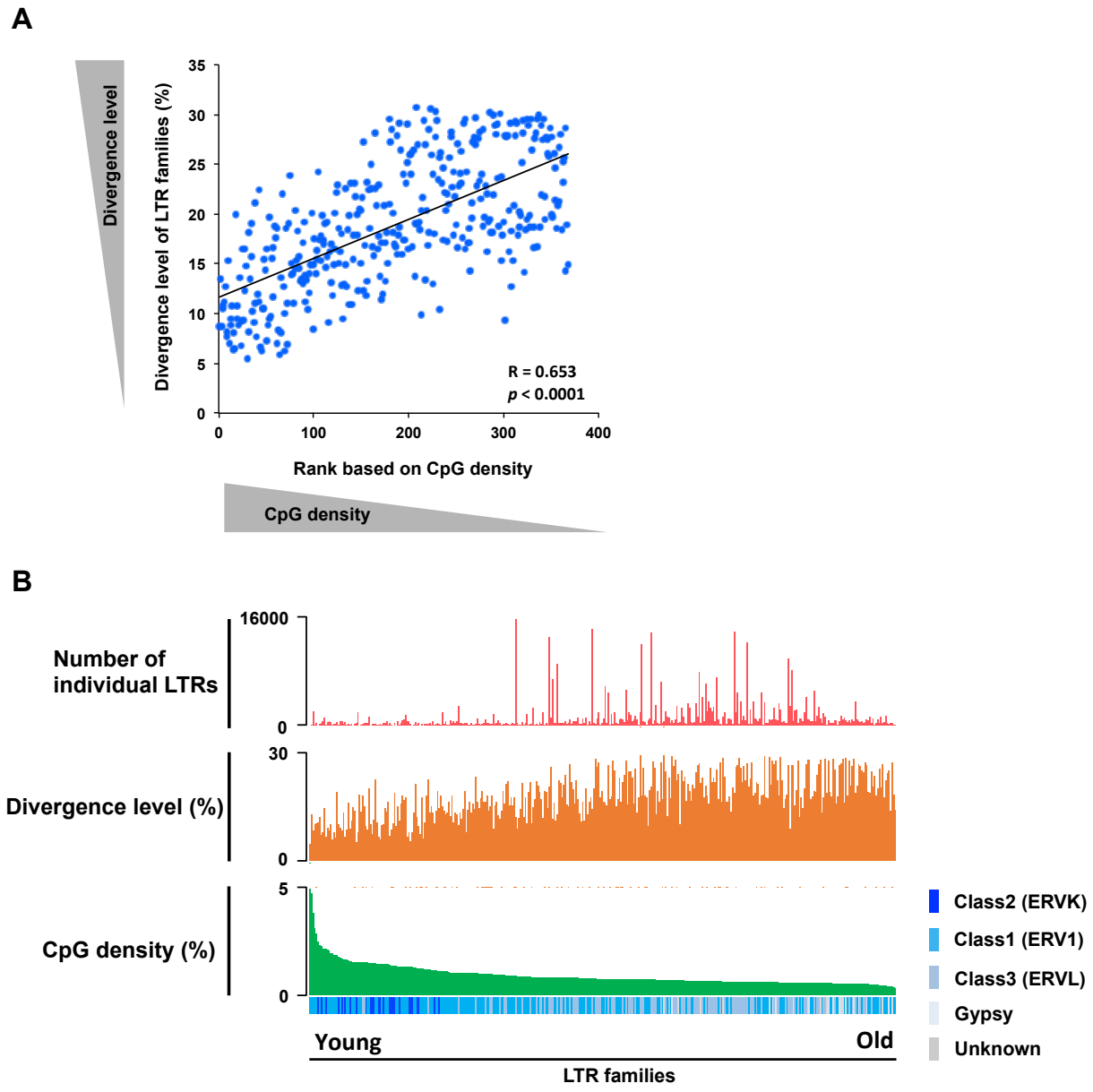
Supplemental Figure S6. Bioinformatic workflow to identify DNA methylation or histone methylation regulated intergenic ERVs in HCT116.

The number of upregulated intergenic ERVs after 5-aza-CdR treatment was 1,639, involving 988 LTRs for Figure 2B (read counts ≥ 20 and fold change ≥ 2). The number of upregulated intergenic ERVs in at least one KD sample (SETDB2, SUV39H2, EZH2, TRIM28, EHMT2, SETDB1, or SUV39H1) was 791 LTRs for Figure 3A.



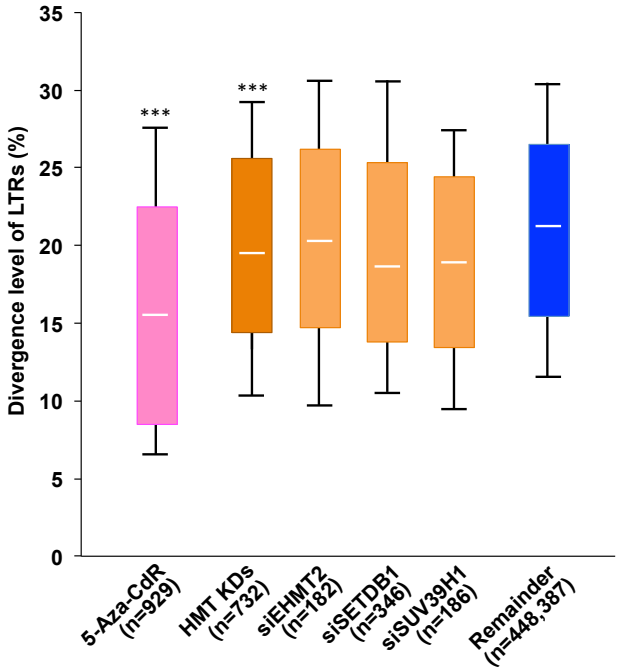
Supplemental Figure S7. Negative correlation between evolutionary age and CpG density of ERVs (for Figure 4B and C).

The distribution of CpG density in each ERV class (LTR only) and its evolutionary age. ERVK is known to be an evolutionarily young ERV; ERVL and Gypsy are known to be evolutionarily old. *P*-values were calculated using the Mann-Whitney *U* test. (***) $P < 0.001$.



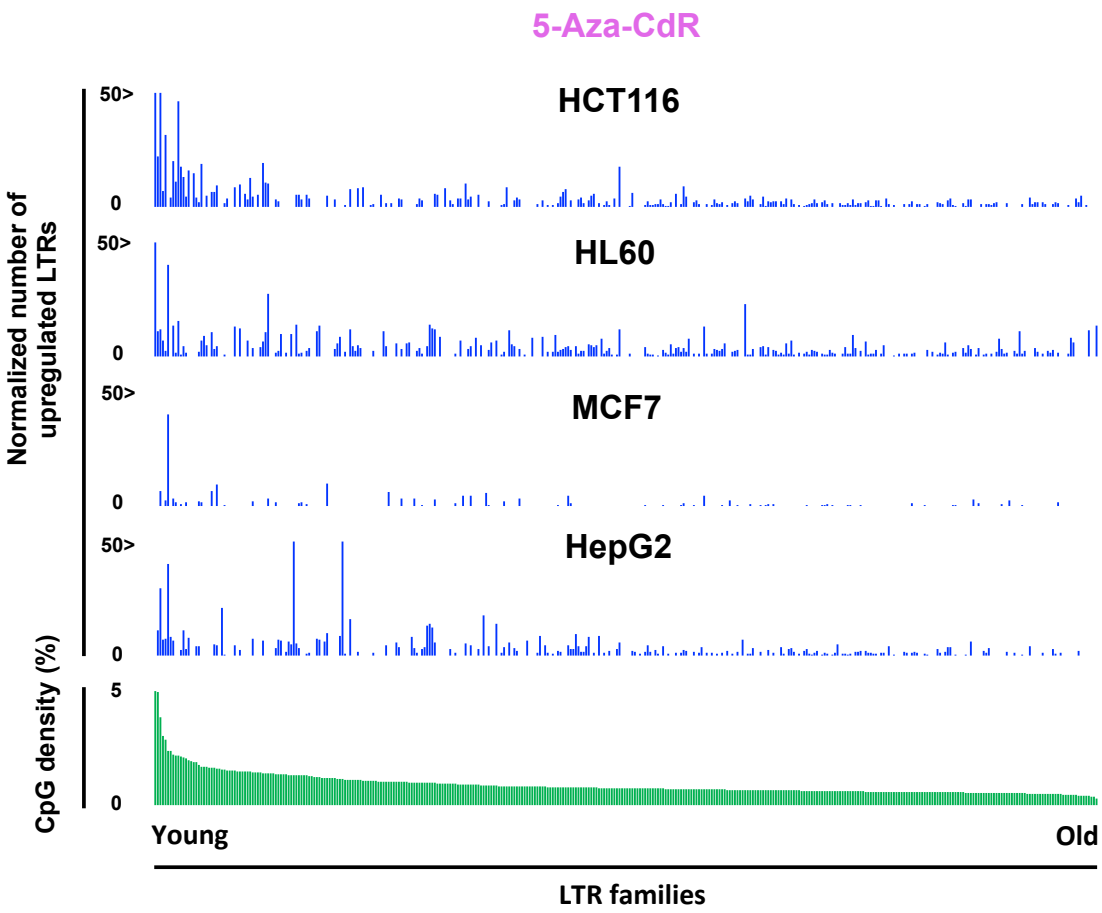
Supplemental Figure S8. Divergence levels of 368 LTR families are positively correlated with their CpG density

(A) The correlation between CpG density of LTR families and their divergence level. X-axis indicates rank of LTR families based on their CpG density. Y-axis indicates average divergence level of LTR families to their consensus sequences. The divergence levels are obtained from the RepeatMasker database. (B) Red bars indicate the number of LTRs in each family; brown bars indicate the average divergence level of LTR families; green bars indicate the average CpG density in each family.



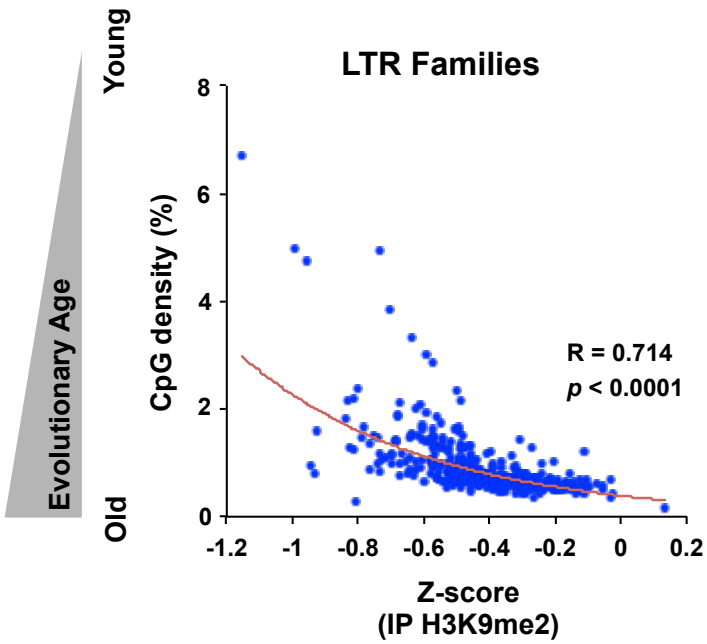
Supplemental Figure S9. Divergence level of LTRs that are reactivated by 5-aza-CdR treatment or HMT KDs (for Figure 4B).

The distribution of the divergence level of LTRs. The boxes for 5-aza-CdR, HMT KDs, siEHMT2, siSETDB1, and siSUV39H1 indicate uniquely upregulated LTRs in each group; the remainder is the control. Each box represents the data between the 25th and 75th quartiles. The whiskers are drawn down to 10th percentile and up to 90th percentile. The differences between the 5-aza-CdR and HMT KDs groups are significant relative to the remainder group. *P*-values were calculated using the Mann-Whitney *U*-test. (***) $P < 0.001$.



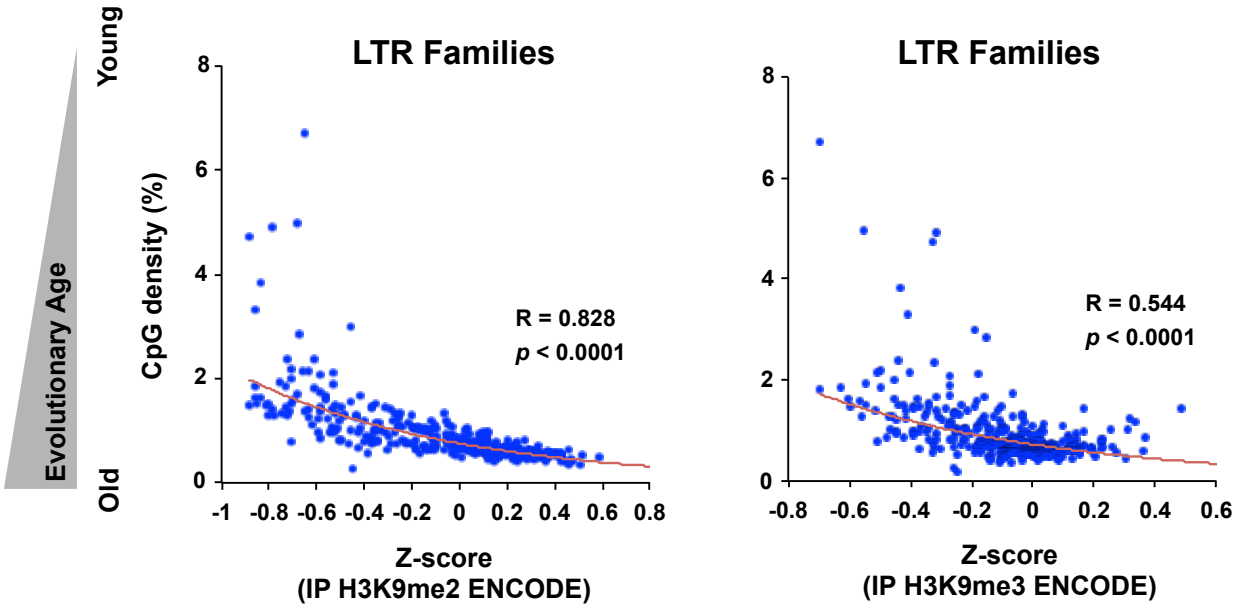
Supplemental Figure S10. The distribution of normalized number of upregulated LTRs in LTR families in four cancer cell lines by 5-aza-CdR (for Figure 4C).

Blue bars indicate the number of upregulated LTRs per 1,000 LTR elements in each LTR family. Green bars indicate the average CpG density in each LTR family (n = 368).



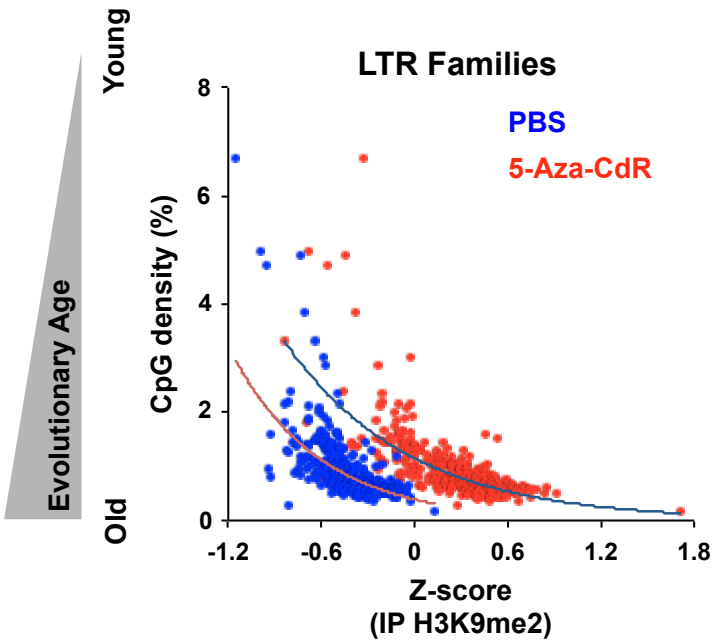
Supplemental Figure S11. Positive correlation between evolutionary age and H3K9me2 of LTR families (for Figure 4E).

The correlation between CpG density (evolutionary age) and H3K9me2 of LTR families. Z-score values were calculated from RPKM based on ChIP-seq data before and after 5-aza-CdR treatment of HCT116 cells (Supplemental Fig. S13). The exponential trend line illustrates a positive correlation between evolutionary age and H3K9me2 in LTR families.



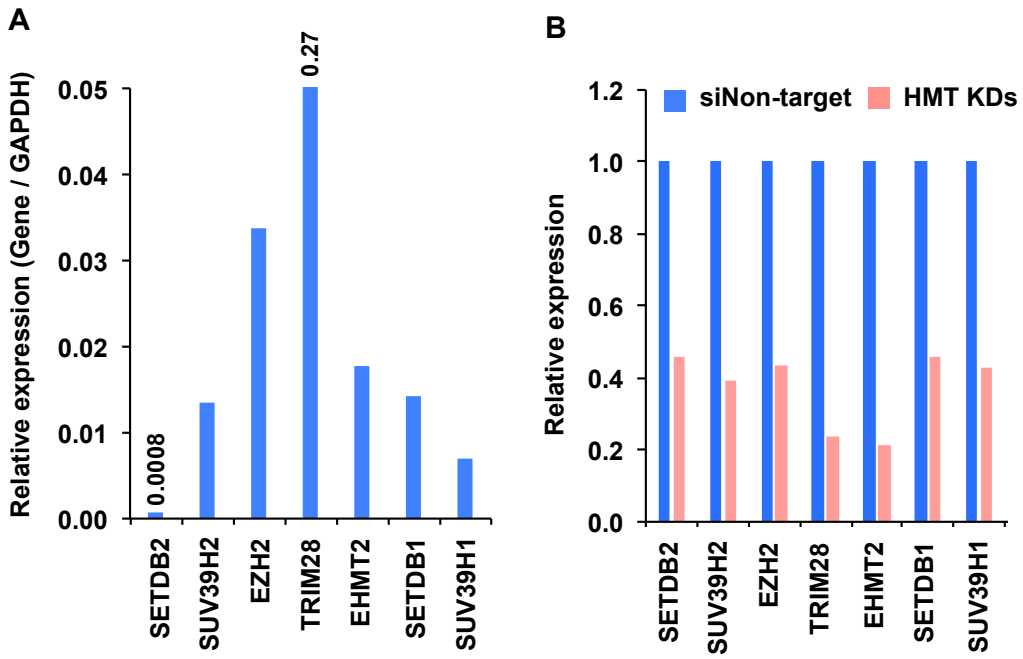
Supplemental Figure S12. Replication study based on ChIP-seq data from ENCODE project (for Figure 4E and Supplemental Figure S11).

The correlation between CpG density (evolutionary age) and histone methylation marks (ENCODE) of LTR families. H3K9me2 is shown at left and H3K9me3 is shown at right. The positive correlations in Figure 4E and Supplemental Figure S11 were confirmed.



Supplemental Figure S13. H3K9me2 marks are increased in LTR families after 5-aza-CdR treatment (for Figure 5A).

The correlation between CpG density (evolutionary age) and H3K9me2 of LTR families before and after 5-aza-CdR treatment. The Z-scores of LTR families were high after 5-aza-CdR treatment relative to PBS treatment.



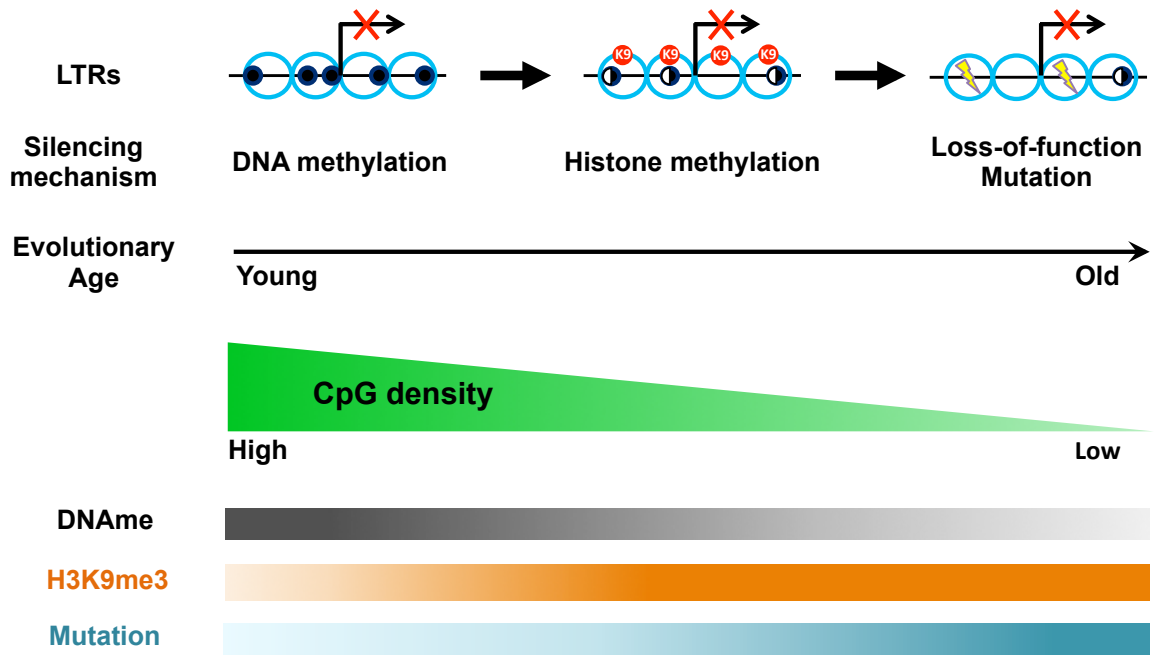
Supplemental Figure S14. Expression level and knockdown efficiency of HMTs.

(A) The bars represent expression of HMTs relative to GAPDH in HCT116. (B) Red bars represent expression relative to siNon-target samples (blue bars) in each HMT KD sample based on RNA-seq.

Treatments	Activation			
	Evolutionary age of LTRs			Viral defense genes
	Young	Intermediate	Old	
5-Aza-CdR	+	-	-	+
HMT KDs	-	+	-	-
Dual depletion	+ +	+ +	-	+ +

Supplemental Figure S15. Specifically targeted LTRs and viral defense genes by 5-aza-CdR, HMT KDs, and dual depletion.

Color represents expression level (+ +: high, +: medium, -: low). The highest 50 LTR families in CpG density were defined as evolutionarily young age (CpG density > 1.34%). The lowest 50 in CpG density were defined as evolutionarily old age (CpG density < 0.54%). The remainders were defined as intermediate age. Viral defense genes are from Figure 6A.



Supplemental Figure S16. A switch of silencing mechanisms in LTRs during host genome evolution.

Black dots represent DNA methylation in each CpG site; red dots represent H3K9me2/3; yellow represents DNA mutations; blue circles represent nucleosomes. LTR elements undergo different types of epigenetic silencing after integration into the host genome. The spontaneous loss-of-function mutations over evolutionary time have been ascribed a role in LTRs silencing.