

Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses

Hitoshi Ohtani,^{1,3} Minmin Liu,^{1,3} Wanding Zhou,¹ Gangning Liang,² and Peter A. Jones¹

¹Van Andel Research Institute, Grand Rapids, Michigan 49503, USA; ²Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA

We provide a comprehensive genomic and epigenomic map of the more than 500,000 endogenous retroviruses (ERVs) and fragments that populate the intergenic regions of the human genome. The repressive epigenetic marks associated with the ERVs, particularly long terminal repeats (LTRs), show a remarkable switch in silencing mechanisms, depending on the evolutionary age of the LTRs. Young LTRs tend to be CpG rich and are mainly suppressed by DNA methylation, whereas intermediate age LTRs are associated predominantly with histone modifications, particularly histone H3 lysine 9 (H3K9) methylation. Young LTRs can be reactivated by treatment with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (5-aza-CdR) alone, but their level of expression is much increased by 5-aza-CdR treatment plus knockdown of one of several H3K9 methyltransferases or of the H3K27 methyltransferase EZH2. The removal of cytosine methylation led to rapid, widespread increases in H3K9me3 in the LTRs. Intermediate age LTRs had lower CpG densities and were not up-regulated by 5-aza-CdR treatment, but they were sensitive to knockdown of H3K9 methyltransferases. Unlike the situation in embryonic stem cells, the polycomb repressive complex (PRC2) has a minor role in LTR suppression by itself and is only a player after removal of cytosine methylation in the analyzed cancer cell line. Up-regulation of LTRs and induction of "viral mimicry" is rapidly becoming of interest for predicting cancer patient response to epigenetic therapies. Understanding the mechanism for LTR suppression is of major importance in order to improve patient treatment strategies.

[Supplemental material is available for this article.]

Transposable elements including *Alus*, long interspersed nuclear elements (LINEs), and endogenous retroviruses (ERVs) make up >40% of the human genome, yet they have been refractory to mapping and epigenomic analysis due to their highly repetitive nature (Gregory 2005). Epigenomic analysis of these elements has not been reported, and they are often specifically excluded from mapping projects because of the difficulty in precisely determining their locations. New developments in sequencing technology, coupled with software that can map specific repetitive elements, have simplified this task (Treangen and Salzberg 2011; Kim et al. 2013), as reported in recent genome-wide studies of repetitive elements (Göke et al. 2015; Liu et al. 2016; Brocks et al. 2017). Here, we focused on the location of epigenomic marks (including DNA methylation and histone modifications) on the approximately 450,000 long terminal repeats (LTRs) that are the putative promoters of ERVs. We concentrated on mapping their distributions in the intergenic regions, excluding those located between the transcription start sites and transcription end sites of protein coding genes and long noncoding RNAs (lincRNAs). This was done to simplify the expression profiles, because it was more likely that the transcripts were being driven by the LTRs rather than being included in the transcription unit driven by a host gene. The majority of human LTRs are functionally inactive, but a minority are expressed in somatic cells; some LTRs can be expressed following the removal

of repressive epigenetic marks (Niwa and Sugahara 1981; Seifarth et al. 2005; Rowe et al. 2010; Anwar et al. 2017; Brocks et al. 2017). Recent studies on mouse embryonic cells have reported that LTR silencing mechanisms depend on DNA methylation, heterochromatin conformation, and the PRC2 complex (Leung and Lorincz 2012; Walter et al. 2016), but less is known about the epigenetic silencing mechanisms of LTRs in human cancer cells.

Because the expression of ERVs is potentially dangerous to the host cell (Tang et al. 2015), understanding the repressive mechanisms is important. Earlier studies have implicated the aberrant expression of ERVs in autoimmune disease pathogenesis (Kassiotis and Stoye 2016). However, this "enemy within" may also play a beneficial role in cancer therapy. For example, inducing their expression may improve the responses of patients who are treated with DNA demethylating drugs, because a state of "viral mimicry" can be induced that encourages destruction of the cancer cells by the innate and adaptive immune systems (Chiappinelli et al. 2015; Roulois et al. 2015; Liu et al. 2016).

We present a comprehensive map of the distribution of LTRs within the intergenic regions of human somatic cell DNA, and we show that the silencing mechanisms differ depending upon the evolutionary age of the particular LTR family. DNA methylation is a major mechanism responsible for repression of LTRs that have more recently transposed into the human germline. The

³These authors contributed equally to this work.

Corresponding author: peter.jones@vai.org

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.234229.118>.

© 2018 Ohtani et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

majority of LTR sequences that are of intermediate and old evolutionary ages are associated with H3K9me2/3 marks. However, H3K27me3, the mark of the PRC2 repressive complex (Mozzetta et al. 2015), does not play a direct role in LTR silencing, although the complex is key to the silencing of developmentally important genes (Azuara et al. 2006; Bernstein et al. 2006; Sharma et al. 2010).

Our data provide extensive maps of intergenic LTR distribution and the epigenetic marks with respect to evolutionary age, and it demonstrates how these repressed sequences can be reactivated following a lessening of histone and DNA methylation. This information will be important in the design of clinical trials to increase the sensitivity of a patient's tumor cells to treatment.

Results

Genome-wide mapping of intergenic ERVs and their epigenetic marks

The ancestral primate germline has been colonized by ERVs for at least the last 100 million years (Lowe et al. 2007; Vargiu et al. 2016). Transcriptionally competent ERVs tend to be CpG rich and to acquire cytosine methylation in the germline, which is probably essential for their silencing. Because of the well-known tendency of 5-methylcytosine (5mC) to deaminate to thymine (Duncan and Miller 1980), the ERVs and other transposable elements (such as *Alus* and *LINES*) show a progressive depletion of CpG content as a function of their evolutionary ages after colonization of the germline (Yang et al. 1996; Egger et al. 2004). It has been difficult to precisely map the position of ERVs because of their repetitive nature, but this has changed with the advent of long-read sequencing technologies and computational approaches that allow for the mapping of distinct ERVs to particular genomic regions.

We first determined the location of ERVs based on RepeatMasker (Smit et al. 2013–2015) and their associated epigenetic marks by whole-genome bisulfite sequencing (WGBS) and H3K9me3 ChIP-seq in the human colon cancer cell line HCT116 (Fig. 1A). To simplify the epigenetic profiles, we focused most of our analyses on ERVs located in intergenic regions (Supplemental Fig. S1). The intergenic ERVs are widely distributed throughout the genome and are found almost everywhere except at cen-

tromeric regions and transcription start sites (Fig. 1A,B; Takai and Jones 2004). The densities of ERVs were positively correlated with the H3K9me3 mark (Fig. 1C), and they tend to be concentrated in regions of heterochromatin, which are marked by the presence of H3K9me3.

Role of DNA methylation in ERV silencing

We examined the potential role of DNA methylation in the silencing of ERV transcription in intergenic regions. We first compared the cytosine methylation levels in HCT116 cells to that in

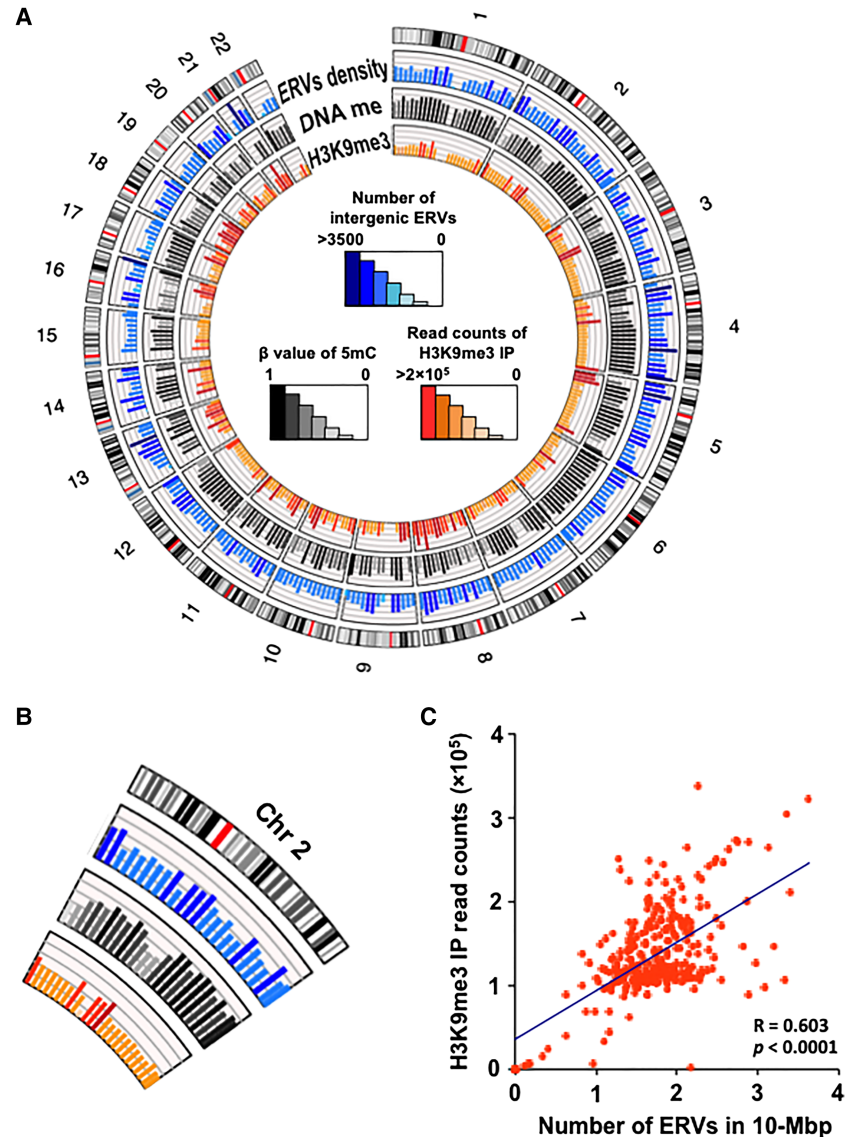


Figure 1. Genome-wide distribution of intergenic ERVs. (A) Circular genome mapping of ERV density, DNA methylation, and H3K9me3 (heterochromatin mark). The first track (outermost) represents the chromosome ideogram. Dark bands indicate heterochromatin regions, and red bands indicate centromeres. The second track represents the number of ERVs in each 10-Mb bin shown as histograms. The third track indicates beta values of DNA methylation of ERVs averaged in the bins, and the fourth track indicates the total read counts of H3K9me3 IP in the bins. In general, regions of high ERV density correlated with high levels of DNA methylation or heterochromatin marks. (B) Enlargement of Chromosome 2 in A. (C) Scatterplot represents the correlation between ERV density and enrichment of H3K9me3. The x-axis and y-axis correspond to the second and fourth tracks in A, respectively.

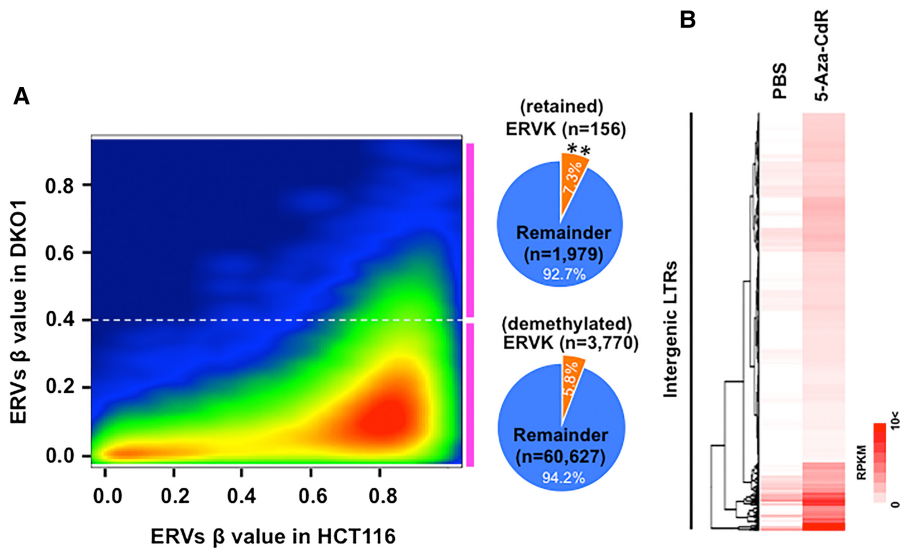


Figure 2. Expression of silenced intergenic LTRs can be induced by 5-aza-CdR in HCT116 cells. (A) A smoothing scatterplot showing the correlation between DNA methylation level of intergenic ERVs in HCT116 (x-axis) and DKO1 (y-axis) cell lines. The cold-to-warm color scale indicates a low-to-high density of intergenic ERVs. The dotted line represents the beta value of DNA methylation at 0.4 in DKO1. Each pie chart represents the percentage of ERVK in the two groups (retained: beta value ≥ 0.4 , and demethylated: beta value < 0.4). P -values were calculated using the χ^2 test for comparisons with the two groups: (**) $P < 0.01$. (B) Heatmap showing the expression of up-regulated 988 intergenic LTRs after 5-aza-CdR treatment. Red indicates expression levels of LTRs from total RNA-seq in PBS or 5-aza-CdR-treated HCT116: (RPKM) reads per kb per million mapped reads.

derivative DKO1 cells, in which $\sim 95\%$ of the methylation has been removed by genetic knockout of *DNMT3B* and partial removal of *DNMT1* (Rhee et al. 2002; Egger et al. 2006). Most of the ERV sequences in the wild-type cells lost methylation in the DKO1 cells, although a small subset retained partial methylation (Fig. 2A). This implies that the methylation is functionally important in ERV suppression, an argument made by the Surani group to explain the continued methylation of a subset of ERVs during the extensive genome-wide demethylation of the human germ line after fertilization (Tang et al. 2015). In particular, the evolutionarily young ERVK family is more resistant to extensive genome-wide demethylation during human germline development (Tang et al. 2015). The evolutionarily young ERVK family was significantly ($P < 0.01$) more methylated than the remaining ERVs, suggesting that they required methylation for their quiescence and cancer cell survival (Fig. 2A). However, virtually no ERV maintained the 90%–100% methylation found in the parent HCT116 cells. This is unlike the situation we observed for a subset of host gene promoters in the same cell line (De Carvalho et al. 2012).

By performing total RNA-seq on four cell lines (HCT116, HL-60, MCF-7, and HepG2), we determined the approximate number of intergenic ERVs that could be up-regulated by treatment with 5-aza-CdR to evaluate how many ERVs were inducible and, by inference, had functional promoters (Supplemental Fig. S2A; Supplemental Table S1). We identified a mapping rate of $\sim 90\%$ after unique mapping of 75-bp RNA-seq; therefore, we believe that the majority of LTR elements can be detected as shown in Methods. We again focused on the intergenic ERVs, because it is difficult to distinguish whether some ERVs are up-regulated due to changes in expression of host genes following drug treatment (Supplemental Fig. S2B,C). About 8000 ERVs were expressed in the four cell lines (read counts more than 20), and about 4000 ERVs were up-regulated by more than twofold after a transient 24-h

exposure to 5-aza-CdR (Supplemental Fig. S2D,E). We then concentrated our attention on LTRs, which are putative promoters capable of initiating transcription. Approximately 2500 LTRs were involved in the up-regulated ERVs in the four cell lines (Supplemental Fig. S2F). About 1000 of the up-regulated LTRs were identified in HCT116 (Fig. 2B). The expression patterns of ERVs were replicated in two independent experiments (Supplemental Fig. S3). The vast majority of LTRs were not inducible by pharmacological inhibition of DNA methylation in any of the four cell lines tested, suggesting that most of LTRs in intergenic regions are not transcriptionally competent. However, our study was limited to only four cell lines, which showed small overlaps between the up-regulated LTRs after treatment with 5-aza-CdR (Supplemental Fig. S2F). Clearly, the number and type of LTRs expressed after DNA demethylation was cell-type-specific, suggesting that other factors (such as specific transcription factors) are required for expression. The total number of potential transcriptionally competent LTRs will require analysis of many more cell lines.

In addition, induction of LTRs leads to activation of the viral defense pathways (Chiappinelli et al. 2015; Roulois et al. 2015). We confirmed this activation in all four cell lines (Supplemental Fig. S2G), and the patterns of viral defense gene up-regulation by 5-aza-CdR were also cell-line-specific. Further studies will be necessary to determine the tumor type specificity for treatment response.

As reported previously, certain transcription factors might also be responsible for the regulation of evolutionarily young LTRs (Du et al. 2016; Brocks et al. 2017; Sundaram et al. 2017). Transcription factor binding motifs, including SP1 and GATA2, are conserved in LTR12C elements up-regulated by DNA methyltransferase and histone deacetylase inhibitors (Brocks et al. 2017). We confirmed the conservation of transcription factor binding motifs of up-regulated LTR12C elements in this study (Supplemental Fig. S4): The transcription factor binding DNA motifs of SP1 and GATA2 were conserved in LTR12C elements up-regulated by 5-aza-CdR treatment.

Different histone lysine 9 methyltransferases silence distinct LTRs

Because we were mainly interested in the abilities of LTRs to initiate transcription, we used siRNAs to knock down a subset of potential repressors including histone lysine methyltransferases targeting H3K9me2/3 (EHMT2; also known as G9a, SETDB1, SETDB2, SUV39H1, and SUV39H2), H3K27me3 (EZH2), and the scaffold protein TRIM28 to determine the roles of these proteins in silencing LTRs within HCT116 cells. RNA-seq analysis of the LTR regions following knockdown (KD) of individual factors showed increased expression of different LTRs in all targeted KD approaches (Fig. 3A; Supplemental Table S2).

KDs of *SETDB2*, *SUV39H2*, and *EZH2* were only marginally effective at inducing LTR transcripts, so they were not widely involved in silencing of these LTRs. However, the EHMT2, SETDB1,

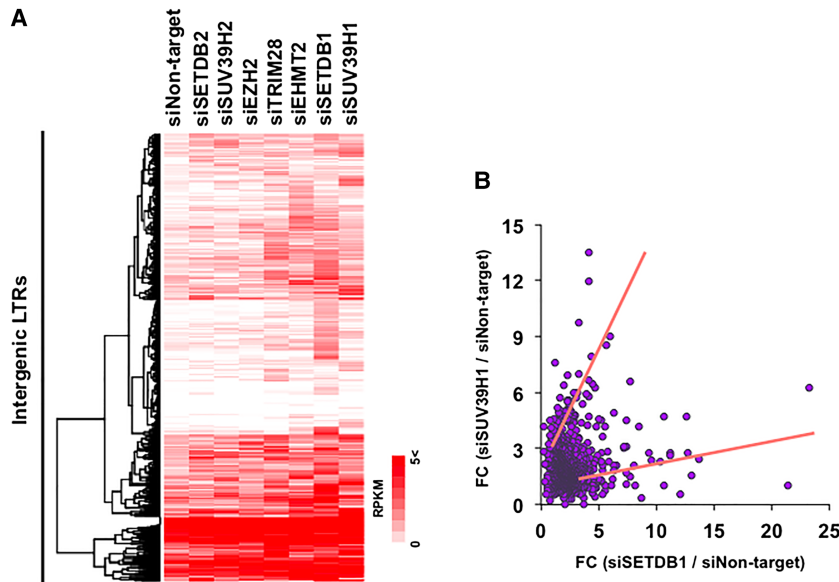


Figure 3. EHMT2, SETDB1, and SUV39H1 are required to maintain silencing of some LTRs in HCT116 cells. (A) The expression of up-regulated 791 intergenic LTRs after knockdown (KD) of represented genes. Red indicates the expression level of LTRs based on total RNA-seq in Nontarget (control), *SETDB2*, *SUV39H2*, *EZH2*, *TRIM28*, *EHMT2*, *SETDB1*, and *SUV39H1* KD samples. (B) The fold expression changes in *SETDB1* and *SUV39H1* KD samples.

and SUV39H1 histone methyltransferases apparently directly suppressed subsets of LTRs. Although they targeted the modification of the same lysine residue, the abilities of these histone methyltransferases to silence the LTRs was most likely dependent on their different roles in modifying specific chromatin regions. For example, EHMT2 and SETDB1 are considered to be methyltransferases active in euchromatic and facultative heterochromatic regions, whereas SUV39H1 is more active in constitutive heterochromatic regions (Mozzetta et al. 2015). The gene knockdown study also demonstrated that KD of *TRIM28* elicited the expression of a unique subset of LTRs, consistent with the idea that the chromatin regulators examined are specialized with respect to their abilities to induce LTR expression (Rowe et al. 2010).

The different abilities of *SETDB1* and *SUV39H1* KDs to cause LTR expression are explored in the scatterplot in Figure 3B. Although both enzymes are H3K9 trimethyltransferases, they silenced distinct LTR elements (with some overlap), and the results were confirmed by quantitative RT-PCR (Supplemental Fig. S5). These data also showed the specificity of histone methyltransferases in silencing LTRs in different epigenomic regions.

Distinct LTRs are silenced by DNA or histone methylation

We next compared the 988 LTRs up-regulated by inhibition of DNA methylation to the 791 up-regulated by knockdown of the subset of histone methyltransferases examined earlier and found that there was little overlap (Fig. 4A; Supplemental Fig. S6). The 929 LTRs uniquely up-regulated by 5-aza-CdR treatment had significantly higher densities of CpGs (defined as the number of CpGs per 100 bp) than the 732 up-regulated by histone methyltransferase KDs or the remaining 448,387 (Fig. 4B). Most LTRs (the 448,387) were not affected by treatment with either histone methyltransferase KDs or DNA methylation inhibition and had lower CpG densities than those responding to treatment.

We also explored the potential differences between those LTRs silenced by DNA methylation or H3K9 lysine methylation, focusing on the evolutionary ages of the LTRs. We first examined the CpG densities of different LTR classes relative to their established evolutionary ages (Supplemental Fig. S7). It was necessary to compare these parameters in LTR classes rather than in individual LTRs because multiple members of LTR classes are present in the genome (Göke et al. 2015), and only subsets of these can be activated by altering methylation marks (Supplemental Fig. S2A). The youngest LTR class, ERVK (Vargiu et al. 2016), showed a statistically higher CpG density relative to the older ERV1 and ERVL (Supplemental Fig. S7). Gypsy, which is the oldest of the LTRs, showed the lowest CpG density. These data are consistent with many observations that the methylation of CpG sequences in the germline results in their subsequent mutation from CpG to TpG due to the enhanced mutational frequency at methylated CpGs (Duncan and Miller 1980; Yang et al. 1996). We also estimated evolution-

ary ages of 368 LTR families based on CpG density and divergence level of intergenic LTRs to their consensus sequence using RepeatMasker. The two indexes to estimate evolutionary age were correlated with each other (Supplemental Fig. S8A,B). The 929 LTRs up-regulated by 5-aza-CdR treatment shown in Figure 4B had significantly lower divergence levels than the 732 specifically up-regulated by HMT KDs or the remainder (Supplemental Fig. S9).

Most of the LTRs up-regulated by 5-aza-CdR were considerably younger than those up-regulated by histone methyltransferase KD, and this correlated with the age of the LTR families (Fig. 4C; Supplemental Table S3). We normalized the number of up-regulated LTRs in each family by using the number of LTRs in the family and by accounting for the fact that evolutionarily old LTR families are more numerous than young LTR families (Supplemental Fig. S8B). The up-regulation of younger LTRs by 5-aza-CdR was also confirmed in the HL-60, MCF-7, and HepG2 cell lines (Supplemental Fig. S10). These data suggest that the primary mechanism for suppression of evolutionarily young LTRs (which have high CpG densities) is DNA methylation, whereas intermediate age LTRs (lower CpG density) are regulated by histone modifications. A majority of the evolutionarily oldest members of the families, such as the ERVL and Gypsy, are not capable of initiating transcription following either DNA methylation inhibition or knockdown of histone methyltransferase enzymes (Fig. 4C). Consistent with these results, CpG sites in evolutionarily young LTRs were methylated at higher levels than in evolutionarily old LTRs (Fig. 4D). An inverse correlation was found between evolutionary age and the H3K9me3 and H3K9me2 marks on LTRs (Fig. 4E; Supplemental Fig. S11). Evolutionarily older LTRs were primarily associated with H3K9me2/3 marks; younger LTRs were less frequently associated with such marks. These results are consistent with the HMT KD study (Fig. 4C) in that evolutionarily young LTRs were not highly up-regulated after HMT knockdown. The results were confirmed by use of ENCODE CHIP-seq data

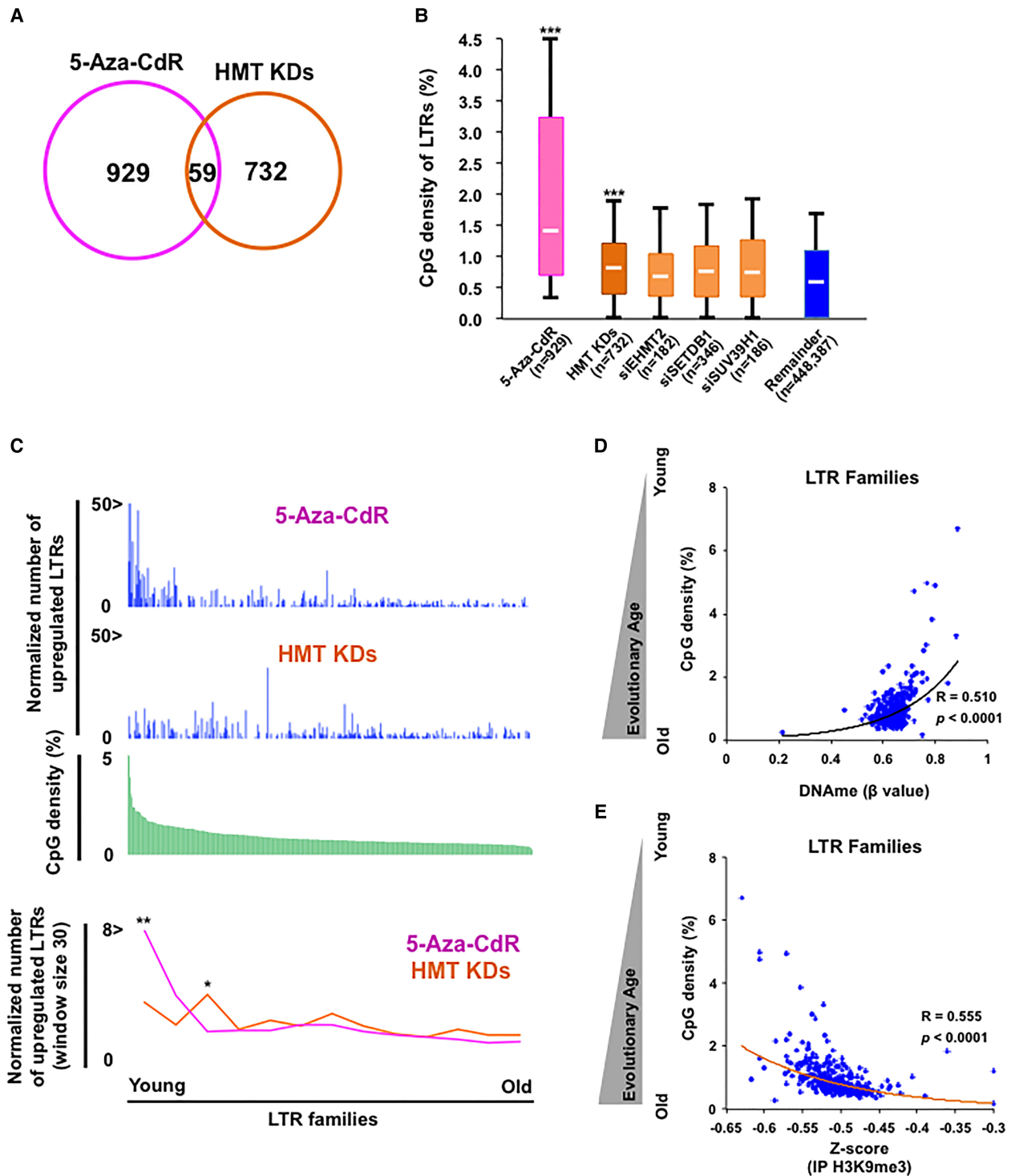


Figure 4. Silencing mechanisms of LTRs are switched in an age-dependent manner. (A) LTRs overlapping between up-regulated LTRs by 5-aza-CdR treatment and knockdown of histone methyltransferases (HMTs). (B) The distribution of CpG densities in up-regulated LTRs by 5-aza-CdR or HMT KDs. The boxes on 5-aza-CdR, HMT KDs, *siEHMT2*, *siSETDB1*, and *siSUV39H1* indicate uniquely up-regulated LTRs in each group. “Remainder” refers to LTRs that were not up-regulated by any treatment. Each box represents the data between the 25th and 75th quartiles. The whiskers are drawn down to the 10th percentile and up to the 90th percentile. The differences between the 5-aza-CdR or HMT KDs groups and remainder are significant. *P*-values were calculated using Mann-Whitney *U* test: (***) $P < 0.001$. (C) The distribution of normalized number of up-regulated LTRs in each family. Blue bars indicate the number of up-regulated LTRs per 1000 LTRs in each LTR family (top). Green bars indicate average CpG density in each member in 368 LTR families (middle). Sliding window analysis represents the normalized number of up-regulated LTRs along CpG density (bottom); window size was 30 LTR families. *P*-values were calculated using the Wilcoxon signed-rank test for comparisons with the two groups: (*) $P < 0.05$; (**) $P < 0.01$. (D) The correlation between CpG density (evolutionary age) and DNA methylation of LTR families. (E) The correlation between CpG density (evolutionary age) and H3K9me3 of LTR families. Z-score values were calculated from RPKM based on ChIP-seq data in PBS and 5-aza-CdR-treated HCT116 (Fig. 5A).

(Supplemental Fig. S12; The ENCODE Project Consortium 2012). DNA methylation is therefore dominant in the control of LTRs especially for evolutionarily young LTRs, whereas histone modification plays a more important role in the regulation of intermediate age LTRs.

Synergy between inhibition of DNA and histone methylation in induction of LTRs and viral defense genes

We observed increases in H3K9me2/3 in the LTRs after transient treatment with 5-aza-CdR (Fig. 5A; Supplemental Fig. S13), which demonstrated potential epigenetic switching after DNA methylation inhibition, possibly to maintain LTR silencing, and which suggested that synergistic effects could be expected from inhibiting both DNA methylation and histone modifications. Thus, we next tested whether 5-aza-CdR treatment plus the knockdown of a histone methyltransferase or *TRIM28* (“dual depletion”) would show synergy. In general, we observed strongly increased expression of LTRs up-regulated by 5-aza-CdR alone (Fig. 5B; Supplemental Table S4).

There is a concern regarding the knockdown efficiencies of HMTs, which were between 50% and 80% (Supplemental Fig. S14A,B), so that our experiments may have identified fewer transcriptionally competent ERVs than the total number of potentially transcribed LTRs. Nevertheless, it seems unlikely that further knockdown would result in up-regulation of evolutionarily young LTRs, because the H3K9me2/3 was largely absent from them (Fig. 4E; Supplemental Fig. S11). The knockdown of *SETDB2*, *SUV39H2*, and *EZH2*, which showed limited up-regulation of LTRs by themselves (Fig. 3A), showed increases in up-regulation of multiple LTRs when combined with 5-aza-CdR treatment (Fig. 5B). Although knockdown of *EZH2* alone showed little LTR up-regulation, it produced an altered spectrum of up-regulated LTRs with 5-aza-CdR treatment. This is probably due to a process in which removal of DNA methylation leads to the subsequent recruitment of EZH2, resulting in suppression by the polycomb repressive complex PRC2 (Lay et al. 2015). Therefore, the same kinds of chromatin dynamics appear to be used both by LTRs and genes. LTRs up-regulated in dual depletion experiments were more often evolutionarily young LTRs (Fig. 5C). Thus, dual depletion of epigenetic silencing marks overcame histone modification-induced silencing and led to synergistic up-regulation of LTRs.

In summary, evolutionarily young LTRs are predominantly silenced by DNA methylation, whereas intermediate age LTRs are silenced by histone methylations (H3K9me2/3). Therefore, the silencing mechanism switched from DNA methylation to histone methylation during host genome evolution (Fig. 5D).

The induction of LTRs results in a state of viral mimicry, in which viral defense genes are up-regulated in response to the presence of double- and single-stranded RNA in the cytoplasm (Chiappinelli et al. 2015; Roulois et al. 2015). We therefore examined the spectrum of viral defense genes induced by the knockdowns alone or by combination with 5-aza-CdR. Few viral defense genes were up-regulated by histone lysine methyltransferases KD alone, but many were up-regulated by 5-aza-CdR alone or by dual depletion (Fig. 6A; Supplemental Table S5). Knockdown of *EZH2*, *EHMT2*, or *SUV39H1* produced enhanced expression of the viral defense genes after 5-aza-CdR treatment. The synergy for expression changes of ERV and viral defense gene were also confirmed by quantitative RT-PCR showing up-regulation of an intact evolutionarily young ERV (ERV-Fc2) and a key pathogen recognition receptor involved in type I interferon signaling pathway

(DDX58; also known as RIG-I) (Fig. 6B). ERV-Fc2 is a member of the ERV1 family, but some ERV1 families are younger than ERVK families (Supplemental Fig. S8B). A detailed analysis of immune response pathways also showed that some pathways are further up-regulated by the dual depletion with 5-aza-CdR plus knockdown of *EZH2*, *EHMT2*, or *SUV39H1* (Fig. 6C). The data again showed a synergy between DNA methylation inhibition and a subset of histone-modifying enzymes.

To summarize, knockdown of histone modifiers did not cause a robust up-regulation of viral defense genes, so induction of intermediate age LTRs might not activate the viral mimicry pathway. Inhibiting DNA methylation led to the activation of viral defense genes, whose up-regulation was associated with the expression of evolutionarily young LTRs (Supplemental Fig. S15). However, further studies are needed to elucidate the detailed mechanisms by which evolutionarily young LTRs activate the innate immune system. Dual depletion with 5-aza-CdR plus histone methyltransferase KD produced strong increases in the expression of viral defense genes, probably in response to the increased up-regulation of the evolutionarily young LTRs (Figs. 5C, 6A). Those young LTRs might be the trigger for activation of viral defense genes. However, not all of them may play key roles, because further activation of viral defense genes was not seen in dual depletion by *SETDB1* KD plus 5-aza-CdR (Figs. 5B, 6A).

Overall, we have shown that expression of the LTRs is regulated by both the DNA and histone methylation, and we have demonstrated a synergy between the two for the up-regulation of the viral defense signature (Supplemental Fig. S15). Therefore, pharmacological inhibition of DNA methylation concurrent with targeting histone methyltransferases might substantially increase the state of viral mimicry. Our data also show a synergy between the two main silencing mechanisms, in which the DNA methylation signal is dominant with evolutionarily younger LTRs, and intermediate age LTRs rely on histone modification.

Discussion

Our results provide genome-wide information on the distribution of intergenic LTRs with respect to evolutionary age and the chromatin features associated with them. We have found important and unexpected changes in the epigenetic marks presumably used to suppress the activities of these potentially dangerous elements. Subsets of evolutionarily younger LTRs, which have high CpG density and high levels of CpG methylation, can be reactivated by 5-aza-CdR treatment alone, suggesting that DNA methylation plays a dominant role in their suppression. The potential deleterious effects of their expression are underscored by the fact that many ERVs retain partial methylation even in cells that have a 95% reduction in genomic 5-methylcytosine. It is also interesting that similar ERV families escape demethylation in the human germline during germ cell specification, once again suggesting that their activities are not compatible with cellular viability (Tang et al. 2015). In both of these cases, the ERV CpG methylation is not maintained near 100%, which is different from what we have reported for the promoters of coding genes in the HCT116 and DKO1 cells (De Carvalho et al. 2012).

We also provide an approximate number of the LTRs that have the capability of initiating transcription following demethylation of either DNA or histones, thus extending the data of Brocks et al. (2017), who focused on DNA methylation and histone acetylation. Approximately 5000 LTRs were found as transcripts after the treatment of four cell lines with 5-aza-CdR. Only about 800

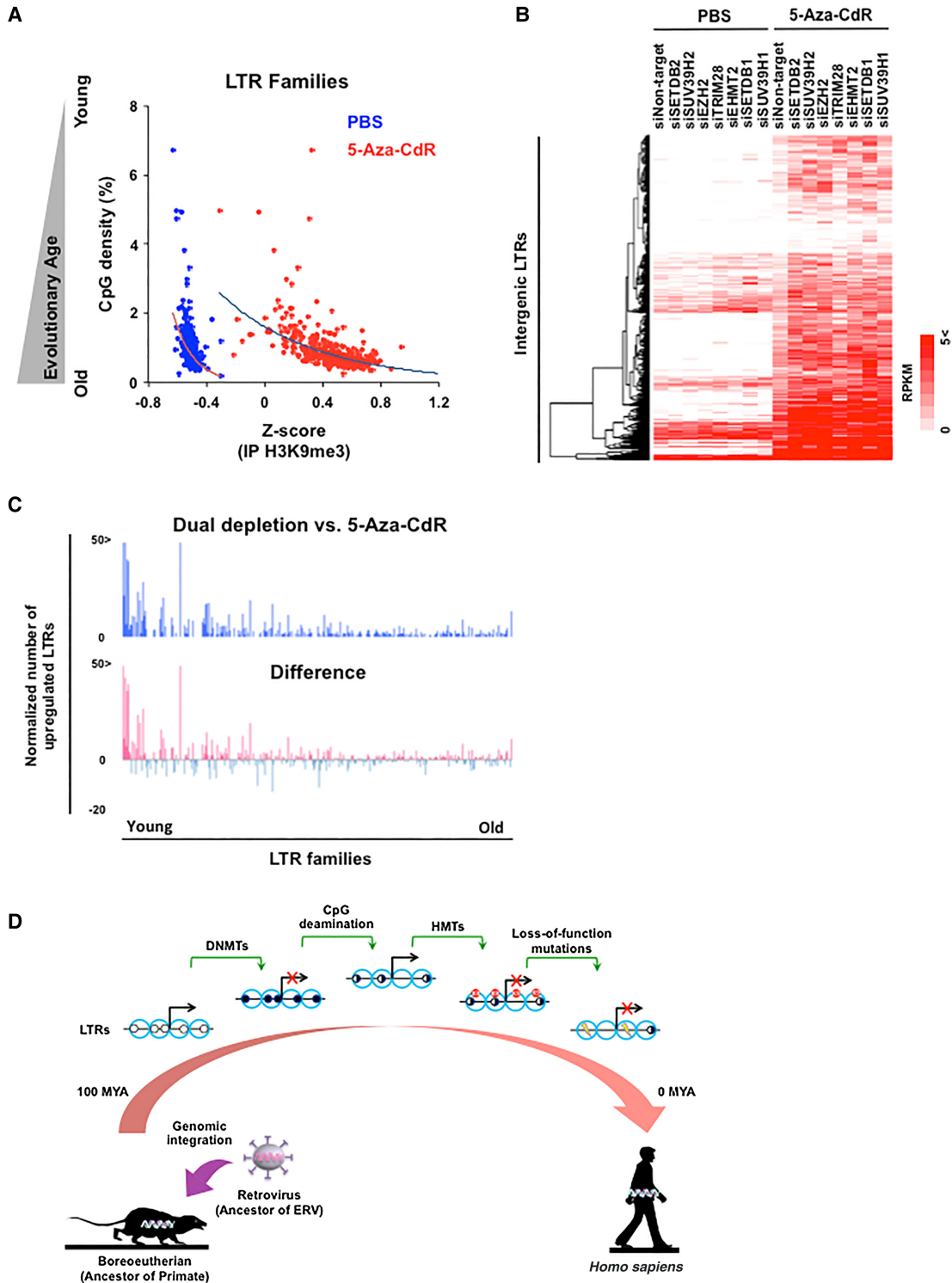


Figure 5. Dual depletion of DNA and histone methylation reveals an epigenetic switch in evolutionarily young LTRs. (A) The correlation between CpG density (evolutionary age) and the H3K9me3 mark in LTR families before and after 5-aza-CdR treatment. (B) The expression level of up-regulated 1033 intergenic LTRs upon dual depletion. Red indicates the expression of LTRs based on total RNA-seq. (C) The distribution of normalized number of up-regulated LTRs in each family. The bars indicate the number of up-regulated LTRs per 1000 in each family (*top*: dual depletion vs. 5-aza-CdR treatment; *bottom*: the difference between dual depletion vs. 5-aza-CdR treatment shown in the *top* panel and HMT KDs shown in Fig. 4C). (D) Evolutionary history of retrovirus silencing after integration into the host genome. Black dots represent DNA methylation in each CpG site, red dots represent H3K9me2/3, yellow represents DNA mutations, and blue circles represent nucleosomes. About 100 million years ago (MYA), retroviruses (ancestors of ERVs) were integrated into the boreoeutherian ancestor genome. LTR elements were silenced by DNA methylation at the beginning, but CpG sites might be lost due to CpG deamination several million years later. Therefore, the silencing mechanism switched from DNA methylation to histone methylation. Eventually, the LTR elements become silenced by the accumulation of loss-of-function genetic mutations.

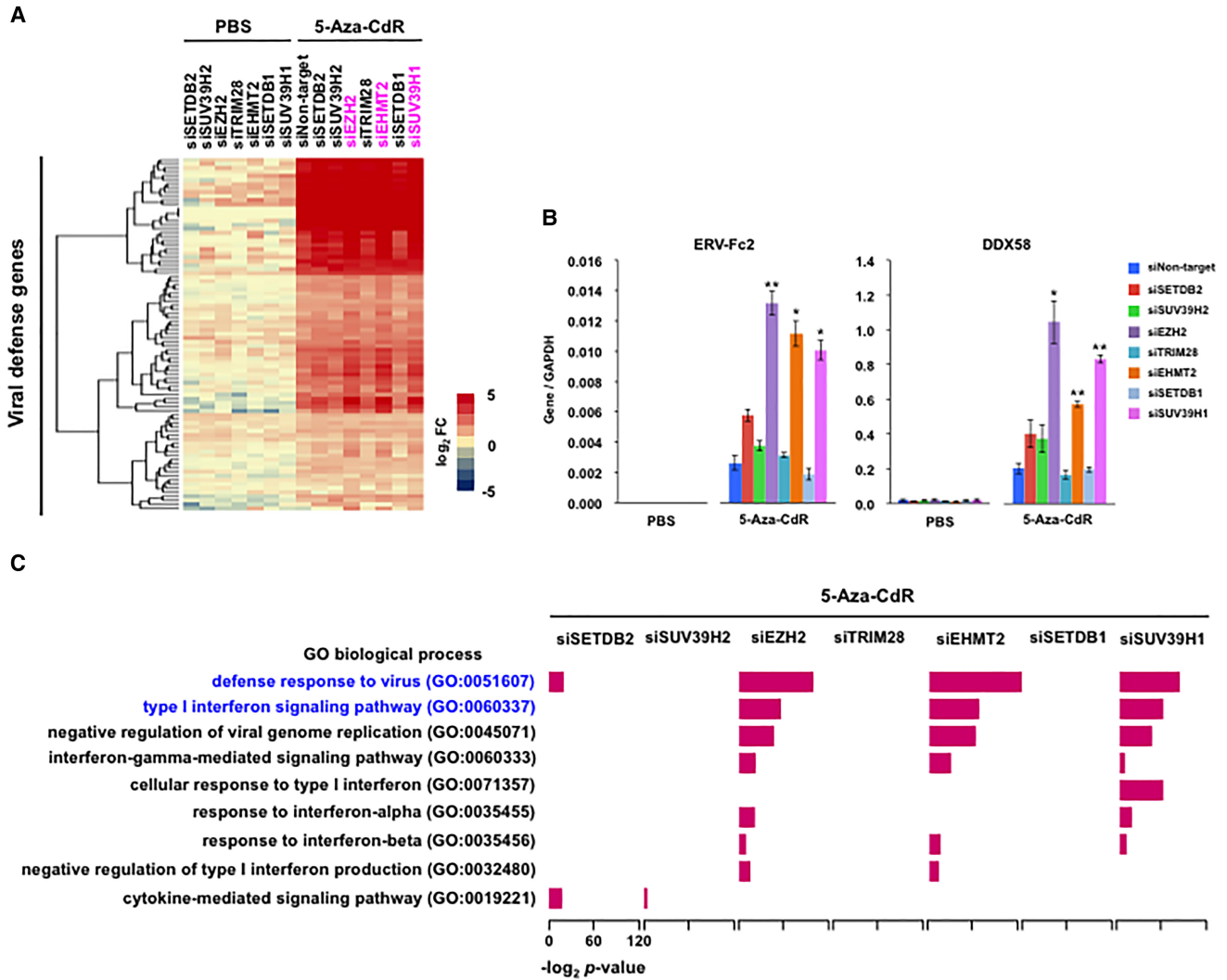


Figure 6. Synergistic effect of activation of viral defense pathway by dual depletion. (A) The expression levels of “defense response to virus” (GO:0051607) and “type I interferon signaling pathway” (GO:0060337) genes in Gene Ontology. The 87 genes were up-regulated at least in one sample more than twofold relative to nontarget siRNA. (B) Expression of *ERV-Fc2* and *DDX58* by RT-qPCR. Error bars represent SEM from three independent biological replicates. *P*-values were calculated using the two-tailed Student’s *t*-test for comparisons with 5-aza-CdR plus siNontarget samples. (C) Gene Ontology analysis of biological processes enriched for up-regulated genes by dual depletion. The x-axis indicates Bonferroni corrected $-\log_2 P$ -values in each dual depletion sample.

of them could be reactivated by the knockdown of H3K9me2/3 methyltransferases in HCT116 cells. The results suggest that only a minority of the total intergenic 450,000 LTRs are transcriptionally competent. However, because our study was limited to only four cell lines, the total number of transcriptionally competent LTRs will require analysis of many more cell lines and primary tissues.

We found a change in silencing mechanisms that was related to the evolutionary ages of the LTRs. Many of the intermediate age LTRs were silenced by histone lysine 9 methyltransferase activity, and although the methyltransferases modify the same lysine residue, different LTRs were up-regulated following knockdown of particular classes of methyltransferases. It has been reported that EHMT2, SETDB1, and EZH2 are required for ERV silencing (Leeb et al. 2010; Karimi et al. 2011; Leung and Lorincz 2012; Maksakova et al. 2013; Kato et al. 2018), but SUV39H1/H2 mainly targets major satellite and intact LINES (Bulut-Karslioglu et al. 2014). LTRs only show a minor up-regulation in SUV39H1/H2

double-knockout mouse embryonic stem cells (Elsässer et al. 2015). Our studies suggest that SUV39H1/H2 are involved in the silencing of LTRs in human cancer cells. In addition, although the function of SETDB2 remains poorly characterized, our finding suggests SETDB2 is involved in LTR silencing, which is interesting because it is induced during infection by influenza virus (Schliehe et al. 2015).

Our data showed that EHMT2, SETDB1, and SUV39H1 were involved in silencing LTRs in human somatic cells. Another three lysine methyltransferases—SETDB2, SUV39H2 and EZH2—had minimal effects on silencing: Few LTRs were up-regulated following their knockdown. On the other hand, these enzymes showed synergistic up-regulation of LTRs when DNA methylation was inhibited. *SETDB2* had low levels of RNA expression, which might be a reason why it did not play a key role in LTR silencing. The absolute expression level of *SUV39H1* RNA was lower than that of *SUV39H2* or *EZH2*. Since the latter two HMTs were not

involved in LTR silencing, the absolute expression levels may not be an essential factor in identifying HMTs required for LTR silencing.

Our findings demonstrate that some demethylated LTRs remained silenced by an epigenetic switch, which recruits other repressive marks such as H3K9me2/3 or H3K27me3 (Schones et al. 2014; Walter et al. 2016). This may explain why knockdown of these lysine methyltransferases cannot up-regulate LTRs but can synergize up-regulation in combination with DNA methylation inhibitors. We found that knockdown of *EZH2* had a significant influence on the levels and types of LTRs up-regulated following 5-aza-CdR treatment, although knocking down *EZH2* by itself had little effect. In addition, our results showed that, unlike *SUV39H1*, *SUV39H2* was necessary for LTR silencing only after loss of DNA methylation, whereas *SUV39H2* encodes a histone H3 lysine methyltransferase that shares 59% identity with *SUV39H1* (O'Carroll et al. 2000). The "epigenetic switch" in silencing mechanisms also is known to occur during development or disease; for example, primordial germ cells undergo a reciprocal loss of DNA methylation and gain of H3K9 methylation at mitotic/meiotic arrest (Cantone and Fisher 2013), and loss of DNA methylation leads to accumulation of H3K27me3 in gene promoters in colon cancer cells (Lay et al. 2015). We show that an epigenetic switch at LTRs occurs during host genome evolution (Supplemental Fig. S16), and that 5-aza-CdR treatment was capable of eliciting the switch at evolutionarily young LTRs. Basal DNA methylation status might therefore be a key effector in determining the silencing mechanism.

The intermediate age LTRs up-regulated by KD of histone methyltransferases did not elicit a strong antiviral response. This suggests that the viral mimicry response, which may be crucial for patient responses to epigenetic drugs, depends on the expression of younger LTRs silenced by DNA methylation. However, it is important to remember that three histone methyltransferases (*EHMT2*, *SUV39H1*, and *EZH2*) that produced little LTR expression when knocked down alone had strong synergistic effects with DNA demethylation in up-regulating the viral defense system. Preliminary data from our laboratory shows that combination treatment with a DNA methylation and an *EHMT2* inhibitor may be useful in epigenetic therapy for cancer patients (M Liu, S Thomas, A DeWitt, W Zhou, Z Madaj, H Ohtani, SB Baylin, G Liang, PA Jones, in prep.). Although some studies have suggested that the LTR transcripts might be detected by host innate immune sensors, the type of LTR has not yet been characterized (Blomberg 2000; Colmegna and Garry 2006; Roulois et al. 2015). Our results therefore show previously unknown links between evolutionarily young LTRs and the host viral defense system.

In summary, evolutionarily young LTRs are silenced by DNA methylation, whereas intermediate age LTRs are silenced by histone methylation. The evolutionarily old LTRs are more likely inactivated by the accumulation of loss-of-function genetic mutations.

Methods

Cell lines and drug treatment

HCT116 (colorectal carcinoma), HL-60 (acute myeloid leukemia), MCF-7 (breast carcinoma), and HepG2 (hepatocellular carcinoma) cells were obtained from the American Type Culture Collection (ATCC). DKO1 cells (biallelic knockout of *DNMT3B* and hypomorphic *DNMT1^{ΔE2-5}* in HCT116 cells) were a gift from Drs. Bert Vogelstein and Steve Baylin. HCT116 and HL-60 cells were treated

with PBS (control) or 300 nM 5-aza-CdR for 24 h. For MCF-7 and HepG2 cells, the dose and exposure time were 100 nM and 72 h. All cells were harvested 5 d after treatment.

RNA-seq

Total RNA from cell lines was purified using a Direct-zol RNA MiniPrep Kit (Zymo Research). cDNA libraries were prepared using RNA HyperPrep Kits with RiboErase (KAPA Biosystems), according to the manufacturer's instructions, and were sequenced as single-end 75 bases on a NextSeq 500 instrument (Illumina). Sequencing reads were aligned against the human GRCh37/hg19 reference genome to report consistent results with our previous study (Liu et al. 2016) by using TopHat2 v.2.1.0 (Kim et al. 2013) or HISAT (Kim et al. 2015). The mapping rates were between 89% and 96%. Reads per kb per million mapped reads (RPKM) values were calculated using uniquely mapped reads. The number of overlapping mapped reads with ERV elements was obtained using BEDTools (Quinlan and Hall 2010). The definitions of ERV elements were based on RepeatMasker. According to the RepeatMasker, ~98% of the ERV elements were found in both of GRCh37 and GRCh38 reference genomes. Pathway enrichment analysis was performed by Gene Ontology (Ashburner et al. 2000; The Gene Ontology Consortium 2017).

siRNA transfection

siRNA-mediated gene knockdown was performed using DharmaFECT2 transfection reagents (Dharmacon) with siRNA at 25 nM final concentration. All cells were harvested 3 d after transfection. SMARTpool siRNAs targeting histone methyltransferase were obtained from Dharmacon (*siEHMT2*: M-006937-01-0005; *siSUV39H1*: M-009604-02-0005; *siSUV39H2*: M-008512-00-0005; *siSETDB1*: M-020070-00-0005; *siSETDB2*: M-014751-00-0005; *siEZH2*: M-004218-03-0005). siRNA targeting *TRIM28* was obtained from Sigma-Aldrich (NM_005762). Nontargeting siRNA (D-001210-01-20) was used as control.

RT-qPCR

cDNA synthesis was performed with an iScript cDNA Synthesis Kit (Bio-Rad) after DNA digestion by TURBO DNase (Thermo Fisher Scientific), according to the manufacturer's instructions. qPCR reactions were carried out in CFX96 Real-Time PCR Detection System (Bio-Rad) with a KAPA SYBR FASTqPCR KIT. The conditions of qPCR reaction were as follows: 3 min at 95°C followed by 40 cycles of 10 sec at 95°C, 20 sec at 60°C, and 20 sec at 72°C. Primers are listed in Supplemental Table S6.

Dual depletion of DNA and histone methylation

Dual depletion of DNA and histone methylation was carried out by combining 5-aza-CdR treatment and siRNA-mediated gene knockdown. HCT116 cells were treated with 300 nM 5-aza-CdR or PBS for 24 h and transfected with siRNAs 2 d after treatment. Cells were harvested 5 d after 5-aza-CdR treatment.

Whole-genome DNA methylation analysis

NOME-seq data on HCT116 and DKO1 were obtained from our previous study (Lay et al. 2015). Reads were aligned to the NCBI human reference genome assembly GRCh37 using BSMAP (Xi and Li 2009) with the following parameters: "-p 27 -s 16 -v 10 -q 2 -A AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACA CGACGCTCTTCCGATCT -A AGATCGGAAGAGCGTCGTGTAA GGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT."

We marked duplicated reads using Picard tools (<http://broadinstitute.github.io/picard>, version 1.38). Reads with non-unique mapping were filtered out. We also filtered out three bases from the 5' end of the first read in pair to avoid reported nonconversion due to reannealing close to adaptor sequences (Berman et al. 2011). The DNA methylation rate of cytosine in HCG sequence context was extracted using BISCUIT tools (<https://github.com/zwdzwd/biscuit>). Cytosines overlapping mutations, including C to T mutations identified from the complementary strands, were excluded. Methylation rates on the cytosine and its neighboring cytosine on the complementary strand in an HCGD sequence context were merged. Merged cytosine methylation rate with fewer than 10 reads of coverage were excluded from analysis.

ChIP-seq

Single nucleosome preparation was performed according to the Dilworth laboratory native ChIP protocol (Brand et al. 2008). Briefly, 5-aza-CdR-treated (or control) HCT116 cells (10^7 cells) were harvested, washed twice, and resuspended in ice-cold buffer N (15 mM Tris pH 7.5, 15 mM NaCl, 60 mM KCl, 8.5% [w/v] sucrose, 5 mM MgCl₂, 1 mM CaCl₂, 1 mM DTT, 200 μM PMSE, and 1× cOmplete Mini EDTA-free Protease Inhibitor Cocktail [Roche]). To prepare nuclei, cells were lysed in 1 mL lysis buffer (buffer N supplemented with 0.3% NP-40 substitute [Sigma]) for 10 min at 4°C, and nuclei were collected by centrifugation (500g for 5 min at 4°C), resuspended in 1 mL of buffer N, and then sedimented through 7.5 mL sucrose cushion (10 mM HEPES pH 7.9, 30% [w/v] sucrose, 1.5 mM MgCl₂) at 13,000g using a Sorvall swinging bucket for 12 min at 4°C. To isolate single nucleosomes, the nuclei were digested with MNase (1 unit Worthington MNase per 70 μg of chromatin for 10 min at 37°C). The nucleosomes were then purified by hydroxyapatite chromatography and adjusted to a concentration of 20 μg/mL with ChIP buffer 1 (25 mM Tris pH 7.5, 5 mM MgCl₂, 100 mM KCl, 10% [v/v] glycerol, and 0.1% [v/v] NP-40 substitute) and analyzed using 2% agarose gel. H3K9me2 and H3K9me3 ChIP were performed as previously described (Grzybowski et al. 2015), using 5 μg of nucleosomes pulled down with 10 μg of anti-H3K9me2 (Abcam ab1220) or anti-H3K9me3 (Active Motif AM39161) antibody on Dynabeads Protein G (Invitrogen) for 2 h at 4°C. Ten percent of the initial chromatin for each IP was set aside to serve as ChIP input. The beads were washed 3× with ChIP buffer 2 (10 mM Tris pH 7.5, 5 mM MgCl₂, 300 mM KCl, 10% [v/v] glycerol, and 0.1% [v/v] NP-40 substitute), twice with ChIP buffer 3 (10 mM Tris pH 7.5, 250 mM LiCl, 1 mM EDTA, 0.5% sodium deoxycholate, and 0.5% [v/v] NP-40 substitute), and twice with 1× TE buffer, followed with two elution steps in elution buffer (50 mM Tris pH 7.5, 1 mM EDTA, and 1% [w/v] SDS). After proteinase K (Roche) digestion (for 1 h at 65°C), sample DNA was purified using Agencourt AMPure XP beads (Beckman Coulter) prior to qPCR analysis. Sequencing libraries were prepared using TruSeq ChIP Library Prep Kit (Illumina) and sequenced as single-end 75 bases on a NextSeq 500 instrument (Illumina) at the Van Andel Research Institute Genomics Core. This procedure yielded between 90 and 120 million sequence reads per sample, >85% of which aligned to human genome version GRCh37 (UCSC Genome Bioinformatics Group).

Data access

Raw and processed data sets from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE108177 and GSE108205.

Acknowledgments

We thank Van Andel Research Institute (VARI) Genomics Core, Bioinformatics Core, David Nadziejka and Tim Triche for technical editing of the article, and all members in the Jones group for helpful discussion. This work was supported by National Cancer Institute Grant R35CA209859.

Author contributions: H.O. performed the experiments and computational analyses. M.L. prepared ChIP-seq libraries. W.Z. assisted with the computational analyses. G.L. and P.A.J. supervised the study. All authors contributed significant discussion.

References

- Anwar SL, Wulaningsih W, Lehmann U. 2017. Transposable elements in human cancer: causes and consequences of deregulation. *Int J Mol Sci* **18**: E974.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.
- Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merckenschlager M, et al. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**: 532–538.
- Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA, et al. 2011. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* **44**: 40–46.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315–326.
- Blomberg J. 2000. [Newly discovered human retroviruses. Association with disease is still undetermined]. *Lakartidningen* **97**: 3597–3599, 3602–3603.
- Brand M, Rampalli S, Chaturvedi CP, Dilworth FJ. 2008. Analysis of epigenetic modifications of chromatin at specific gene loci by native chromatin immunoprecipitation of nucleosomes isolated using hydroxyapatite chromatography. *Nat Protoc* **3**: 398–409.
- Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, Li J, Zhang B, Hou Y, Laudato S, et al. 2017. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet* **49**: 1052–1060.
- Bulut-Karslioglu A, De La Rosa-Velázquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, Galán C, Winter GE, Engst B, Gerle B, et al. 2014. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* **55**: 277–290.
- Cantone I, Fisher AG. 2013. Epigenetic programming and reprogramming during development. *Nat Struct Mol Biol* **20**: 282–289.
- Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A, et al. 2015. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**: 974–986.
- Colmegna I, Garry RF. 2006. Role of endogenous retroviruses in autoimmune diseases. *Infect Dis Clin North Am* **20**: 913–929.
- De Carvalho DD, Sharma S, You JS, Su SF, Taberlay PC, Kelly TK, Yang X, Liang G, Jones PA. 2012. DNA methylation screening identifies driver epigenetic events of cancer cell survival. *Cancer Cell* **21**: 655–667.
- Du J, Leung A, Trac C, Lee M, Parks BW, Lusic AJ, Natarajan R, Schones DE. 2016. Chromatin variation associated with liver metabolism is mediated by transposable elements. *Epigenetics Chromatin* **9**: 28.
- Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**: 560–561.
- Egger G, Liang G, Aparicio A, Jones PA. 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**: 457–463.
- Egger G, Jeong S, Escobar SG, Cortez CC, Li TW, Saito Y, Yoo CB, Jones PA, Liang G. 2006. Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. *Proc Natl Acad Sci* **103**: 14080–14085.
- Elsässer SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. 2015. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522**: 240–244.
- The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.
- The Gene Ontology Consortium. 2017. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* **45**: D331–D338.

- Göke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, Szczerbinska I. 2015. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**: 135–141.
- Gregory TR. 2005. Synergy between sequence and size in large-scale genomics. *Nat Rev Genet* **6**: 699–708.
- Grzybowski AT, Chen Z, Ruthenburg AJ. 2015. Calibrating ChIP-Seq with nucleosomal internal standards to measure histone modification density genome wide. *Mol Cell* **58**: 886–899.
- Karimi MM, Goyal P, Maksakova IA, Bilensky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M, et al. 2011. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements and chimaeric transcripts in mouse ES cells. *Cell Stem Cell* **8**: 676–687.
- Kassiotis G, Stoye JP. 2016. Immune responses to endogenous retroelements: taking the bad with the good. *Nat Rev Immunol* **16**: 207–219.
- Kato M, Takemoto K, Shinkai Y. 2018. A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. *Nat Commun* **9**: 1683.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357–360.
- Lay FD, Liu Y, Kelly TK, Witt H, Farnham PJ, Jones PA, Berman BP. 2015. The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Res* **25**: 467–477.
- Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A. 2010. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* **24**: 265–276.
- Leung DC, Lorincz MC. 2012. Silencing of endogenous retroviruses: When and why do histone marks predominate? *Trends Biochem Sci* **37**: 127–133.
- Liu M, Ohtani H, Zhou W, Ørskov AD, Charlet J, Zhang YW, Shen H, Baylin SB, Liang G, Grønbaek K, et al. 2016. Vitamin C increases viral mimicry induced by 5-aza-2'-deoxycytidine. *Proc Natl Acad Sci* **113**: 10238–10244.
- Lowe CB, Bejerano G, Haussler D. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci* **104**: 8005–8010.
- Maksakova IA, Thompson PJ, Goyal P, Jones SJ, Singh PB, Karimi MM, Lorincz MC. 2013. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* **6**: 15.
- Mozzetta C, Boyarchuk E, Pontis J, Ait-Si-Ali S. 2015. Sound of silence: the properties and functions of repressive Lys methyltransferases. *Nat Rev Mol Cell Biol* **16**: 499–513.
- Niwa O, Sugahara T. 1981. 5-Azacytidine induction of mouse endogenous type C virus and suppression of DNA methylation. *Proc Natl Acad Sci* **78**: 6290–6294.
- O'Carroll D, Scherthan H, Peters AH, Opravil S, Haynes AR, Laible G, Rea S, Schmid M, Lebersorger A, Jerratsch M, et al. 2000. Isolation and characterization of *Suv39h2*, a second histone H3 methyltransferase gene that displays testis-specific expression. *Mol Cell Biol* **20**: 9423–9433.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW, et al. 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416**: 552–556.
- Roulois D, Yau HL, Singhanian R, Wang Y, Danesh A, Shen SY, Han H, Liang G, Pugh TJ, Jones PA, et al. 2015. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* **162**: 961–973.
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, et al. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**: 237–240.
- Schliehe C, Flynn EK, Vilagos B, Richson U, Swaminathan S, Bosnjak B, Bauer L, Kandasamy RK, Griesshammer IM, Kosack L, et al. 2015. The methyltransferase Setdb2 mediates virus-induced susceptibility to bacterial superinfection. *Nat Immunol* **16**: 67–74.
- Schones DE, Chen X, Trac C, Setten R, Paddison PJ. 2014. G9a/GLP-dependent H3K9me2 patterning alters chromatin structure at CpG islands in hematopoietic progenitors. *Epigenetics Chromatin* **7**: 23.
- Seifarth W, Frank O, Zeilfelder U, Spiess B, Greenwood AD, Hehlmann R, Leib-Mösch C. 2005. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *J Virol* **79**: 341–352.
- Sharma S, Kelly TK, Jones PA. 2010. Epigenetics in cancer. *Carcinogenesis* **31**: 27–36.
- Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Sundaram V, Choudhary MNK, Pehrsson E, Xing X, Fiore C, Pandey M, Maricque B, Udawatta M, Ngo D, Chen Y, et al. 2017. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat Commun* **8**: 14550.
- Takai D, Jones PA. 2004. Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. *Mol Biol Evol* **21**: 463–467.
- Tang WW, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. 2015. A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161**: 1453–1467.
- Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13**: 36–46.
- Vargiu L, Rodriguez-Tomé P, Sperber GO, Cadeddu M, Grandi N, Blikstad V, Tramontano E, Blomberg J. 2016. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* **13**: 7.
- Walter M, Teissandier A, Pérez-Palacios R, Bourc'his D. 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5**: e11418.
- Xi Y, Li W. 2009. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* **10**: 232.
- Yang AS, Gonzalzo ML, Zingg JM, Millar RP, Buckley JD, Jones PA. 1996. The rate of CpG mutation in *Alu* repetitive elements within the p53 tumor suppressor gene in the primate germline. *J Mol Biol* **258**: 240–250.

Received January 10, 2018; accepted in revised form June 29, 2018.