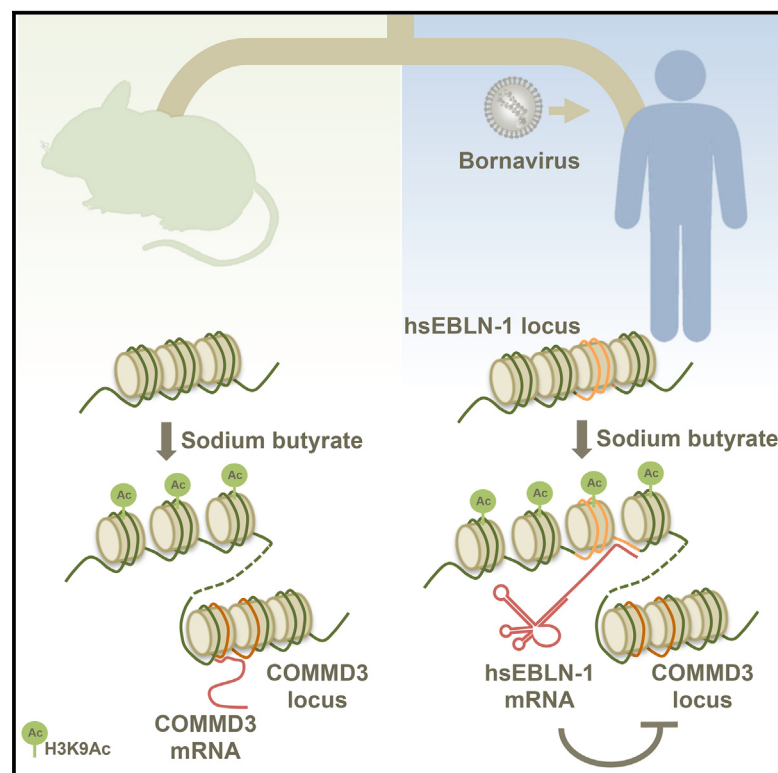


Cell Reports

Transcription Profiling Demonstrates Epigenetic Control of Non-retroviral RNA Virus-Derived Elements in the Human Genome

Graphical Abstract



Authors

Kozue Sofuku, Nicholas F. Parrish, Tomoyuki Honda, Keizo Tomonaga

Correspondence

thonda@virus.kyoto-u.ac.jp (T.H.),
tomonaga@virus.kyoto-u.ac.jp (K.T.)

In Brief

DNA sequences derived from ancient bornaviruses (endogenous bornavirus-like nucleoprotein elements, EBLNs) are present in vertebrate genomes. Sofuku et al. demonstrate that *Homo sapiens* EBLN-1 (hseBLN-1) is repressed epigenetically in many tissues. Expression of hseBLN-1 in a limited, tissue-specific manner suggests a role in regulating neighboring gene expression and genome function.

Highlights

- Seven *H. sapiens* endogenous bornavirus-like nucleoprotein elements are expressed as RNA in at least one tissue
- Expression of some hseBLNs, for example, hseBLN-1, is tissue-specific
- Expression from the hseBLN-1 locus is epigenetically silenced in most tissues
- Integration of hseBLN-1 in the genome affects the expression of a neighboring gene

Accession Numbers

AB934382.1
AB934381.1



Sofuku et al., 2015, Cell Reports 12, 1548–1554
September 8, 2015 ©2015 The Authors
<http://dx.doi.org/10.1016/j.celrep.2015.08.007>

CellPress

Transcription Profiling Demonstrates Epigenetic Control of Non-retroviral RNA Virus-Derived Elements in the Human Genome

Kozue Sofuku,¹ Nicholas F. Parrish,¹ Tomoyuki Honda,^{1,*} and Keizo Tomonaga^{1,*}

¹Department of Viral Oncology, Institute for Virus Research, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

*Correspondence: thonda@virus.kyoto-u.ac.jp (T.H.), tomonaga@virus.kyoto-u.ac.jp (K.T.)

<http://dx.doi.org/10.1016/j.celrep.2015.08.007>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

Endogenous bornavirus-like nucleoprotein elements (EBLNs) are DNA sequences in vertebrate genomes formed by the retrotransposon-mediated integration of ancient bornavirus sequence. Thus, EBLNs evidence a mechanism of retrotransposon-mediated RNA-to-DNA information flow from environment to animals. Although EBLNs are non-transposable, they share some features with retrotransposons. Here, to test whether hosts control the expression of EBLNs similarly to retrotransposons, we profiled the transcription of all *Homo sapiens* EBLNs (hsEBLN-1 to hsEBLN-7). We could detect transcription of all hsEBLNs in at least one tissue. Among them, hsEBLN-1 is transcribed almost exclusively in the testis. In most tissues, expression from the hsEBLN-1 locus is silenced epigenetically. Finally, we showed the possibility that hsEBLN-1 integration at this locus affects the expression of a neighboring gene. Our results suggest that hosts regulate the expression of endogenous non-retroviral virus elements similarly to how they regulate the expression of retrotransposons, possibly contributing to new transcripts and regulatory complexity to the human genome.

INTRODUCTION

Recently, we and others discovered that numerous vertebrate genomes contain endogenous sequences related to non-reverse-transcribing RNA and DNA viruses, not only retroviruses. Bornavirus, a nonsegmented negative-strand RNA virus, is unique because it is the only non-retroviral RNA virus with endogenous elements in the human genome (Horie et al., 2010). The majority of endogenous fragments of bornavirus in human genomes appear to have originated from reverse transcription and integration of nucleoprotein (N) mRNA of ancient bornavirus through long interspersed nuclear element-1 (LINE-1 or L1) activity. They are thus a unique form of processed pseudogenes and evidence a mechanism of retrotransposon-

mediated RNA-to-DNA information flow from virus to host. We call these sequences endogenous bornavirus-like nucleoproteins (EBLNs) (Horie et al., 2010). A number of EBLNs are identified in the genomes of various mammalian species, including primates, rodents, and afrotherians (Horie et al., 2010). Whereas we have recently shown that an EBLN in the ground squirrel genome can inhibit Borna disease virus (BDV) replication (Fujino et al., 2014), it is unknown whether EBLNs have any biological significance in humans. At least seven independently integrated EBLNs are present in the human genome (Table S1). The motifs resembling the start signal motifs (SSMs) of BDV transcription were well conserved in flanking sequences of *Homo sapiens* EBLNs (hsEBLNs) except for hsEBLN-7, and hsEBLNs contain the target site duplications (TSDs) and poly A tracts, hallmarks of L1 retrotransposition process (Figure 1A). In one case the N mRNA appears to have been integrated as a chimera with a short interspersed nuclear element (SINE) (Horie et al., 2010). While EBLNs are not thought to be transposable themselves, they may share features with transposing elements on account of their origin via retrotransposition.

Transposable elements, such as L1 and endogenous retroviruses (ERVs), induce genomic changes through several mechanisms. Retrotransposition of LINEs themselves occurs in as many as one in eight human births and has been implicated in human disease (Kazazian, 1999). In addition, retrotransposons provide the molecular machinery for reduplication of host genes, generating processed pseudogenes and new fusion genes (Esnault et al., 2000; Sayah et al., 2004). ERVs create new transcriptional landscapes by promoting transcription of protein-coding genes proximal to their insertion site (Whitelaw and Martin, 2001).

Retrotransposition mediated by these elements can have deleterious effects on the hosts, which have evolved redundant strategies that constrain retrotransposition. For example, in response to retrotransposon overexpression, host cells activate programmed cell death pathways (Chuma, 2014; Noutsopoulos et al., 2010). Should retrotransposon-encoded proteins become translated, many proteins can interact with them to inhibit their function (Goodier et al., 2013). However, at the front line of defense, epigenetic modification of transposon loci limits their transcription.

DNA methylation plays a major role in ERV silencing (Bourc'his and Bestor, 2004; Walsh et al., 1998; Yoder et al., 1997). Methylation of histone H3 lysine 9 (H3K9) is also required for ERV

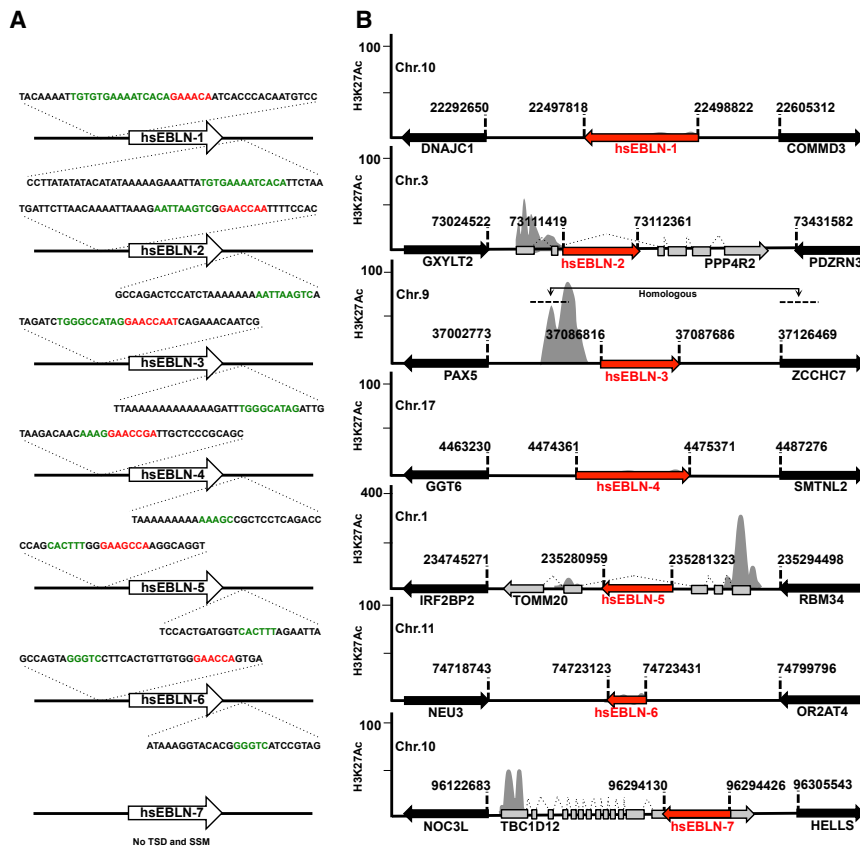


Figure 1. hsEBLN Loci in the Human Genome

(A) The SSM-like motif (red) and the TSD (green) around hsEBLN loci.

(B) The numbering corresponds to nucleotide positions in human genome (UCSC as of November 15, 2013). Red arrows indicate the region of homology with BDV N protein as annotated in Table S1. Black and gray arrows indicate neighboring genes around hsEBLNs. The enrichment of H3K27ac histone marks around hsEBLNs is plotted on the y axis in a.u.

silencing when DNA methylation levels are insufficient for ERV repression (Dong et al., 2008; Leung and Lorincz, 2012; Matsui et al., 2010). As for ERVs, DNA methylation and H3K9 methylation are critical for L1 silencing (Castro-Diaz et al., 2014; Di Giacomo et al., 2013; Ishizu et al., 2012; Tachibana et al., 2007). Furthermore, L1 retrotransposition is controlled by P-element-induced wimpy testis (PIWI)-interacting RNAs (piRNAs) in concert with the DNA methylation machinery (Bourc'his and Bestor, 2004; Kuramochi-Miyagawa et al., 2008) and endogenous RNAi in mouse (Tam et al., 2008; Watanabe et al., 2008) and human cells (Yang and Kazazian, 2006).

We hypothesized that expression of hsEBLNs might be epigenetically controlled by the host similarly to retrotransposons, even though they are not autonomously transposable. Here, to test this possibility, we profiled the transcription pattern of EBLNs in human tissues. We detected transcription of all seven hsEBLNs in at least one tissue examined. hsEBLN-1 was expressed in a testis-specific manner, and its expression was restricted by epigenetic modifications, reminiscent of retrotransposons.

RESULTS

Expression Profiles of hsEBLN RNAs

Figure 1B shows each hsEBLN within its genomic context, including nearest protein-coding genes. Although transcription from some hsEBLN loci has been predicted bioinformatically

do not encode promoter sequences. We therefore searched putative nearby promoters that might drive hsEBLN transcription. UE loci were accompanied by upstream enrichment of histone H3 lysine 27 acetylation (H3K27ac), an active histone mark, and transcription factor binding in various cells at University of California, Santa Cruz (UCSC) Genome Browser (Figure 1B). The hsEBLN-2 and hsEBLN-5 loci are located in an intron of the *PPP4R2* and *TOMM20* genes, respectively. The hsEBLN-7 locus is located in an exon of the *TBC1D12* gene. The upstream sequence (chr9: 37079001–37086816) of the hsEBLN-3 locus, where H3K27ac was identified within various cell types, was highly homologous (92.2% identity) to the upstream region (chr9: 37119617–37129512) of the *ZCCHC7* gene, a neighboring gene of hsEBLN-3, where H3K27ac enrichment was also identified (Figure 1B). On the other hand, we detected putative promoter activity upstream of the TS hsEBLN-1 locus, where transcription factor binding, but not H3K27ac enrichment, was detected in UCSC Genome Browser, suggesting hsEBLN-1 may contain a putative promoter sequence with a suboptimal activity (Figure S1). Transcripts from all seven hsEBLN loci were detected in human testis, consistent with the report of elevated expression of processed pseudogenes in testis (Figure 2H) (Marques et al., 2005). Taken together, all hsEBLN loci have the potential to be transcribed because of a “borrowed” promoter sequence, as is the case with some other processed pseudogenes (Bradley et al., 2004; McCarrey and Thomas, 1987; Mighell et al., 2000).

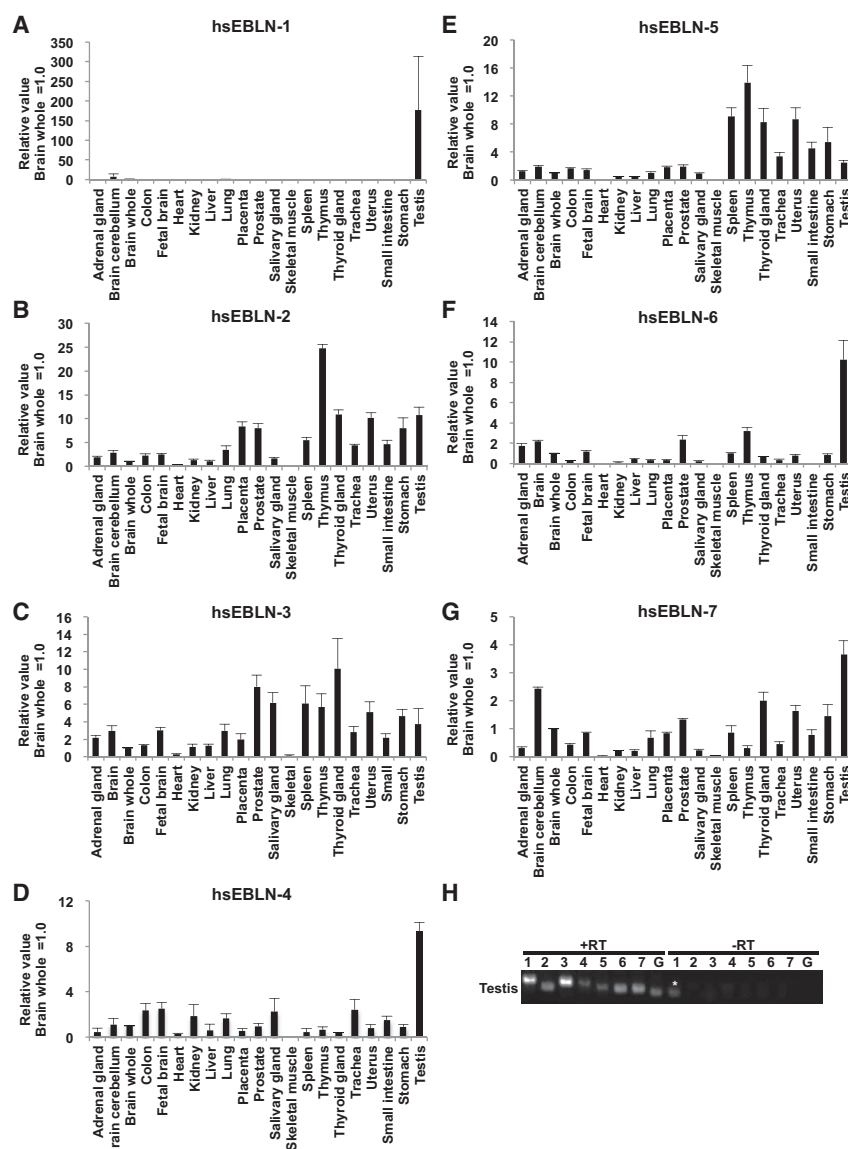


Figure 2. Expression Profiles of hseBLN RNAs

(A–G) Quantitative analysis of hseBLN RNAs in various human tissues. (A) hseBLN-1. (B) hseBLN-2. (C) hseBLN-3. (D) hseBLN-4. (E) hseBLN-5. (F) hseBLN-6. (G) hseBLN-7. The expression level of hseBLN RNA was normalized to that of GAPDH mRNA. The relative levels of hseBLN RNAs in various tissues were calculated with that in brain whole. Values are expressed as the mean + SD.

(H) RT-PCR analysis of hseBLN RNA expression in human testis. Lane 1, hseBLN-1; lane 2, hseBLN-2; lane 3, hseBLN-3; lane 4, hseBLN-4; lane 5, hseBLN-5; lane 6, hseBLN-6; lane 7, hseBLN-7; lane G, GAPDH. * Non-specific band.

previously (Horie et al., 2010). On the basis of these observations, we used OL cells to investigate the silencing mechanisms of the hseBLN-1 locus.

First, we evaluated the contribution of histone deacetylation to silencing of the hseBLN-1 locus. Treatment of OL cells with a histone deacetylase (HDAC) inhibitor, sodium butyrate (SB) or trichostatin A (TSA), led to dose-dependent induction of hseBLN-1 RNA and histone H3 acetylation (H3ac) (Figures 3B, 3C, S2A, and S2B). Chromatin immunoprecipitation (ChIP) revealed that SB treatment increased the level of histone H3 lysine 9 acetylation (H3K9ac) on the hseBLN-1 promoter (Figure S2C). These results demonstrate that histone deacetylation contributes to silencing of the hseBLN-1 locus in OL cells. We next evaluated the contribution of DNA methylation to silencing of the locus. A DNA methylation inhibitor, 5-aza-2'-deoxycytidine (5-Aza), enhanced transcription from the hseBLN-1 locus to a similar level as SB, suggesting DNA methylation also

contributes to silencing of the hseBLN-1 locus (Figures S2D and S2E).

Hierarchical Regulation of hseBLN-1 by Multiple Epigenetic Modifications

We then evaluated the contribution of histone methylation to hseBLN-1 silencing using siRNA against various histone lysine methyltransferases (HKMTs) (siG9a, siEZH2, or siESET). All three siRNAs efficiently downregulated the target mRNA (Figures S2F–S2H). However, there was no significant change in the expression of hseBLN-1 RNA by these siRNA treatments (Figure 3D). Since hseBLN-1 silencing is enforced by at least two other mechanisms, we considered that reductions in HKMTs may be insufficient to induce transcription from the hseBLN-1 locus. We therefore knocked down HKMTs in the presence of SB or 5-Aza. In the presence of SB, siESET led

Regulation of hseBLN-1 Expression by Histone Deacetylation and DNA Methylation

The testis-enriched expression pattern of TS loci raised the possibility of host silencing of these sequences in somatic tissues, similar to retrotransposons. Such silencing would be of potential interest, especially because these sequences are homologous to an exogenous virus. The TS hseBLN-1 locus has a 1,004-nt sequence of continuous homology to the N gene of BDV. Because of this homology and its striking testis-specific expression, we focused on the hseBLN-1 locus for additional research.

We examined whether transcription from the hseBLN-1 locus is also restricted in human cultured cell lines derived from two somatic tissues. As shown in Figure 3A, transcription from the hseBLN-1 locus was limited in OL (human oligodendrogloma cells) and 293T (human embryonic kidney cells) cells, although it was detected by increasing amplification cycles as reported

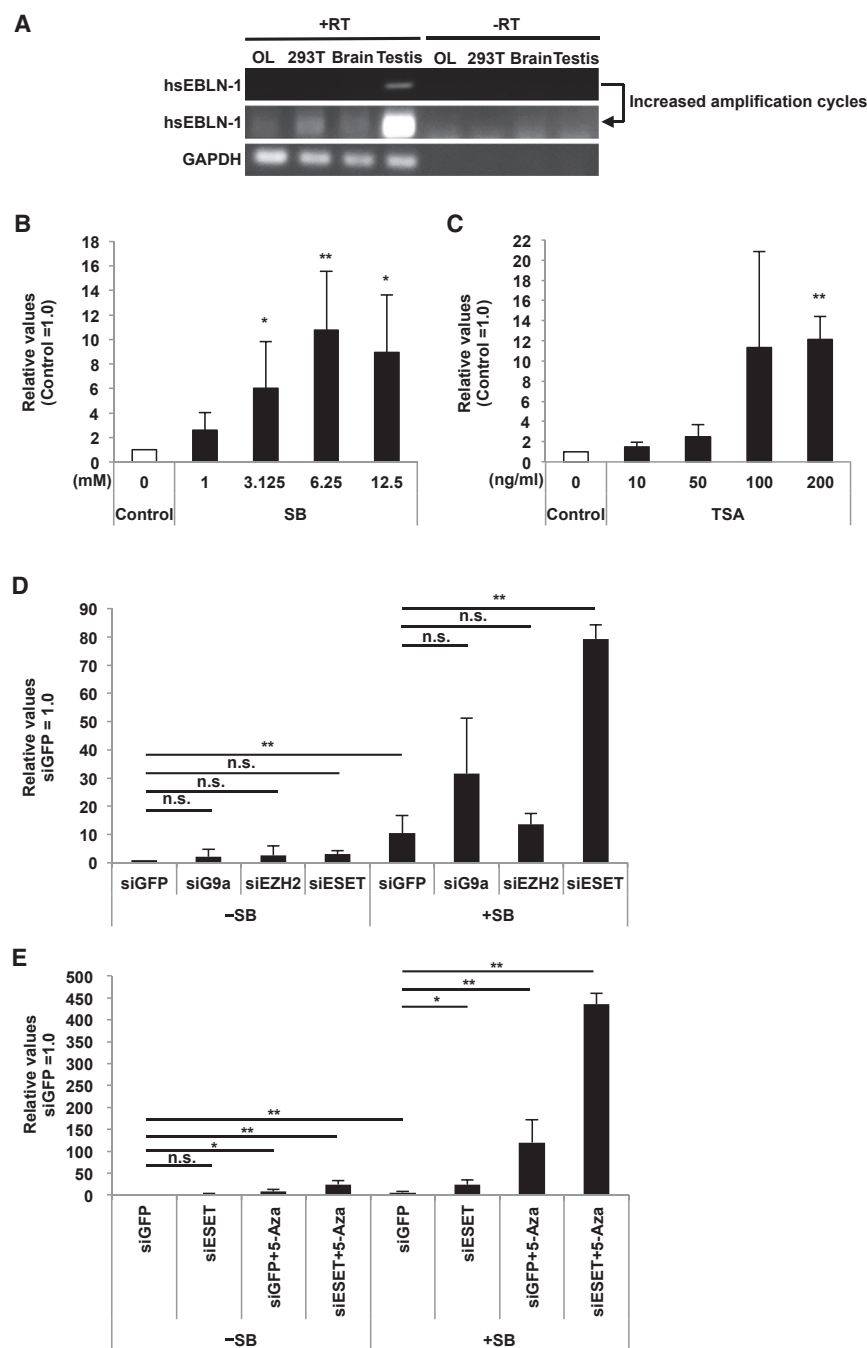


Figure 3. Epigenetic Regulation of hEBLN-1 RNA Expression

(A) RT-PCR analysis of hEBLN-1 RNA expression in human brain, testis, OL, and 293T cells. GAPDH mRNA expression is indicated as a loading control. (B and C) Quantitative analysis of hEBLN-1 RNA after treatment of HDAC inhibitors. OL cells were incubated for 18 hr after addition of 12.5 mM SB (B) or 200 ng/ml TSA (C). The expression level of hEBLN-1 RNA was normalized to that of GAPDH mRNA. The relative values were calculated using untreated group (Control). (D) Quantitative analysis of hEBLN-1 RNA after treatment with siG9a, siEZH2, or siESET with or without SB. (E) Quantitative analysis of hEBLN-1 RNA after treatment with siESET and 5-Aza with or without SB. OL cells were incubated for 3 days after treatment with siESET and 10 μ M 5-Aza with or without SB. The relative values were calculated using siGFP-treated group without SB (siGFP). Values are expressed as the mean $\pm$ SD. * $p < 0.05$, ** $p < 0.01$, and n.s., no significance (Student's t test). At least three independent experiments were performed.

The Expression of Neighboring Genes around the hEBLN-1 Locus

Our results above suggest that the host may significantly regulate the expression of the hEBLN-1 locus by the epigenetic machineries similarly to the regulation of retrotransposon expression. These observations raised the possibility that hEBLN loci could influence the expression of neighboring genes similarly to retrotransposons (McClintock, 1956). We therefore examined whether transcription from the hEBLN-1 locus affects the expression level of neighboring protein-coding genes, human *DNAJC1* and *COMMD3* genes. Treatment with SB or TSA led to transcription from the hEBLN-1 locus (Figures 3B and 3C) as well as up-regulation of *DNAJC1* expression (Figure 4A). On the other hand, the expression of human *COMMD3* gene was downregulated (Figure 4B). Given

that treatment of SB or TSA either enhances or does not affect gene transcription in general, this downregulation might be due to the influence of hEBLN-1 expression. To evaluate this, we suppressed hEBLN-1 RNA induction after SB treatment using siRNA against hEBLN-1 RNA (si-hEBLN-1) (Figure 4C). Suppression of the hEBLN-1 RNA induction eliminated the SB-induced downregulation of the *COMMD3* gene expression (Figure 4D). Taken together, hEBLN-1 RNA is likely to affect expression of the neighboring *COMMD3* gene.

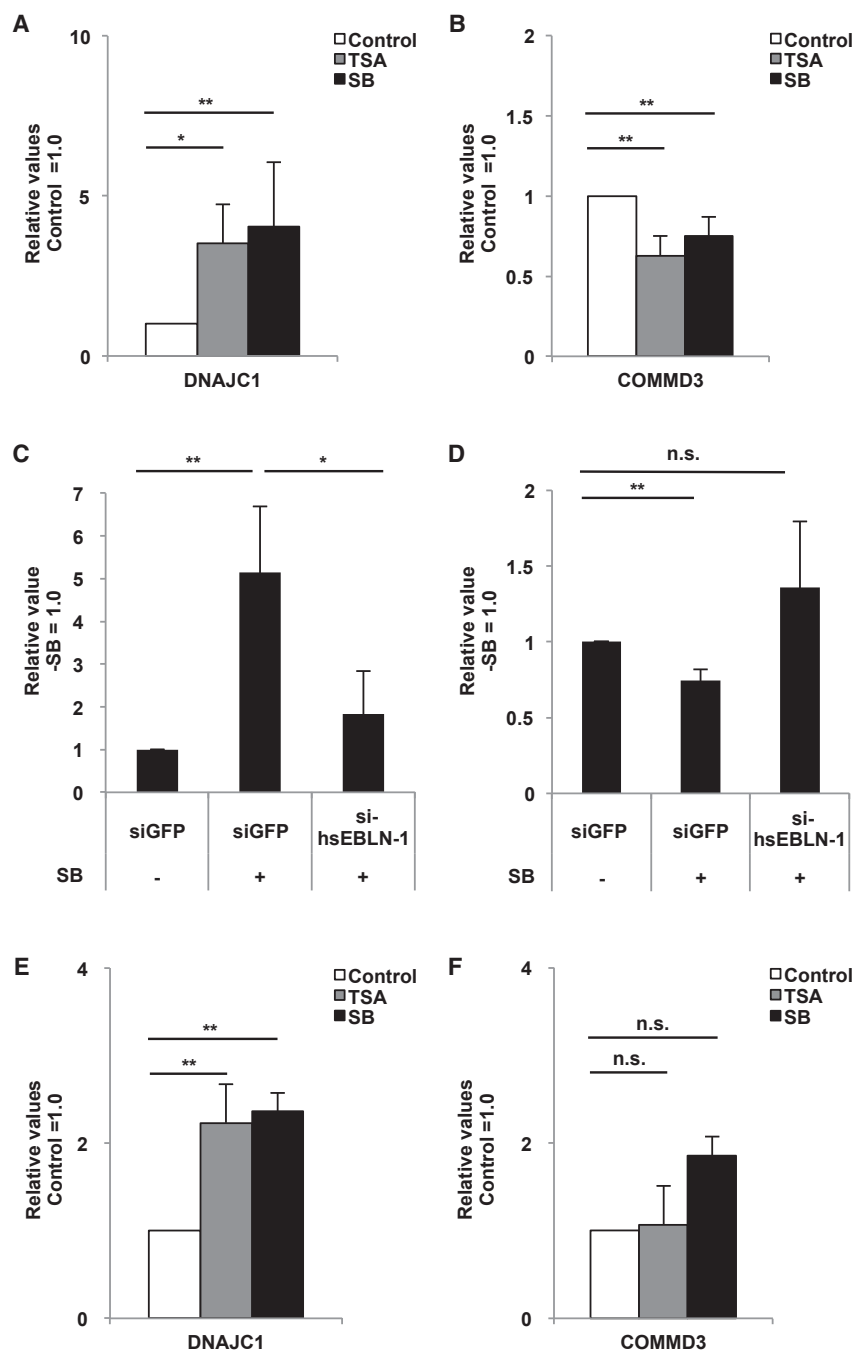


Figure 4. The Expression of Neighboring Genes around the hseBLN-1 Locus

(A and E) Quantitative analysis of the expression of the *DNAJC1* gene after treatment of SB or TSA. Human OL (A) or mouse B7 (E) cells were incubated for 18 hr after treatment of 12.5 mM SB or 200 ng/ml TSA.

(B and F) Quantitative analysis of the expression of the *COMMD3* gene after treatment of SB or TSA. OL (B) or B7 (F) cells were incubated for 18 hr after treatment of SB or TSA.

(C and D) Quantitative analysis of the expression of hseBLN-1 RNA (C) and the *COMMD3* gene (D) after treatment of the hseBLN-1-specific siRNA and SB. OL cells were treated with the hseBLN-1-specific siRNA and were incubated for 18 hr after treatment of SB. The expression levels of the indicated genes were normalized to that of GAPDH mRNA. The relative values were calculated using untreated group (Control). Values are expressed as the mean $\pm$ SD. * $p < 0.05$, ** $p < 0.01$, and n.s., no significance (Student's *t* test). At least three independent experiments were performed.

which do not (Figure S3A). As in human cells, the expression of monkey, mouse, and rat *DNAJC1* orthologs was upregulated by treatment with HDAC inhibitors (Figures 4E, S3B, and S3D). In contrast, HDAC inhibitor treatment led to decreased transcription of *COMMD3* orthologs in human and monkey cells, whereas transcription of mouse and rat *COMMD3* orthologs was not affected (Figures 4F, S3C, and S3E). These results suggest that EBLN integration may affect the epigenetic status of its neighboring gene, further supporting the idea that the expression of human *COMMD3* gene is influenced by the hseBLN-1 locus.

DISCUSSION

This is systematic evaluation of endogenous non-retroviral virus elements in the human genome. In this study, we demonstrated that all hseBLNs have the potential to be transcribed in some tissues. hseBLN loci can be categorized into two

We then investigated whether EBLN integration indeed influenced epigenetic regulation around the hseBLN-1 locus. EBLN-1 through EBLN-4 integrated into the genome of an ancestor of all haplorhini primates (Horie et al., 2010). Except for the EBLN, this region is syntenic between rodents and primates. Therefore we tested the effect of HDAC inhibitors on expression of *DNAJC1* and *COMMD3* orthologs in cells derived from the African green monkey (Vero cells), which has an hseBLN-1 ortholog in this locus, as well as cells from mouse (B7 pulmonary epithelial cells) and rat (C6 neuroblastoma cells),

groups based on expression profile: UE and TS loci. hseBLN loci were located in an existing gene or downstream of a potent promoter region. This is consistent with the proposed nature of EBLNs as pseudogenes generated from mRNA in an L1-dependent manner; processed pseudogenes are generally expressed only when they acquire new promoter and regulatory sequences or happen to integrate within an existing transcription unit (Bradley et al., 2004; McCarrey and Thomas, 1987; Mighell et al., 2000). Our observation that four of seven hseBLNs are currently transcribed fairly ubiquitously is unexpected based on

transcription of pseudogenes processed from host mRNA, of which only 4%–6% are transcribed (Harrison et al., 2005). The unexpected extent of transcription from hseBLN1s might suggest a possible function of some of these elements.

Transcription from the hseBLN-1 locus was repressed in most tissues except for testis by multiple epigenetic mechanisms, dominantly by histone deacetylation and DNA methylation. These results indicate that hseBLN-1 is repressed by the host epigenetically, similar to retrotransposons, even though it has no transposition activity. However, there are some differences between mechanisms silencing hseBLN-1 and those silencing retrotransposons. One such difference is that transcription from the hseBLN-1 locus is rather reversible since an HDAC inhibitor or 5-Aza alone can reactivate the locus. This reversible control of transcription from the hseBLN-1 locus is similar to silencing of exogenous retroviruses (Blazkova et al., 2009; Nabel and Baltimore, 1987). On the other hand, silencing of endogenous retrotransposons is mediated by redundant mechanisms, and transcription from these elements is not easily induced (Di Giacomo et al., 2013; Leung and Lorincz, 2012; Matsui et al., 2010). Another difference is that histone deacetylation plays a major role in silencing the hseBLN-1 locus, again more similar to that of exogenous retroviruses (Williams et al., 2006) than to ERV or L1 where DNA methylation, but not histone deacetylation, plays a major role (Matsui et al., 2010). Collectively, silencing mechanisms of the hseBLN-1 locus might be more similar to those of exogenous retroviruses than to those of endogenous retrotransposons.

Retrotransposons were first characterized as elements capable of controlling the expression of neighboring genes (McClintock, 1956). Indeed, ERVs, L1s, and processed pseudogenes can influence the expression of neighboring genes by various mechanisms. For example, retrotransposons or transposed genes can regulate the expression of homologous genes in other loci by competing for transcription factor(s) (Kalmukova et al., 1998; Livak, 1990). Additionally, retrotransposons may provide alternative promoters, exons, terminators, and splice junctions to protein-coding loci, facilitating genome evolution (Babushok et al., 2007; Häslér et al., 2007; Kazazian, 2004; Speck, 2001). hseBLN1s were formed via horizontal gene transfer from the environment to mammals, which is an unexplored mechanism of genome innovation in mammals. Here, using hseBLN-1 as an example, we showed the possibility that endogenous non-retroviral virus elements could be a novel source of regulatory control of nearby protein-coding genes. In the rodent genome, lacking an orthologous EBLN at this locus, the expression of rodent *COMMD3* gene was not affected by SB or TSA. This suggests that the *COMMD3* locus itself is not inherently regulated by histone acetylation in the absence of an EBLN. When the expression from the EBLN locus in the human or monkey genome was induced, the expression of *COMMD3* orthologs was downregulated. si-hseBLN-1 abrogated this regulation of the *COMMD3* gene. Perhaps hseBLN-1 RNA may function as a long non-coding RNA that scaffolds transcriptional repressors for the *COMMD3* gene around the locus, downregulating its expression. EBLN-encoded proteins could potentially have biological functions as shown previously (Fujino et al., 2014). However, the involvement of the protein encoded by hseBLN-1 in the

COMMD3 gene expression is unlikely, because the protein is located in the cytoplasm (Fujino et al., 2014). As piRNAs generated from rodent EBLN sequences are found in the NCBI database, it is possible that piRNAs are generated from hseBLN-1 and are involved in this regulation. We cannot exclude the possibility that other SB- or TSA-responsive factors contribute to downregulation of the *COMMD3* gene in anthropoid cells, but our data highlight a previously underappreciated source of epigenetic regulation as a consequence of EBLN integration. Further studies are clearly required for the understanding of significance of EBLNs in this regard.

EXPERIMENTAL PROCEDURES

Cells

OL cells were cultured in DMEM (Gibco) containing 5% fetal bovine serum (FBS). B7 cells were cultured in DMEM containing 10% FBS. Cells were cultured at 37°C under 5% CO₂.

Chemicals and siRNAs

SB, TSA, and 5-Aza were purchased from Sigma-Aldrich. siESET was purchased from Santa Cruz Biotechnology. siGFP, siG9a, siEZH2, and si-hseBLN-1 were purchased from Sigma-Aldrich. The sequences of siRNAs are listed in Table S2.

RT-PCR Analysis

Total RNAs were extracted using TRIzol reagent (Invitrogen). Total RNAs of human tissues were purchased from Clontech. Reverse transcription was performed using Verso cDNA Synthetic Kit (Thermo Scientific). Real-time RT-PCR assays were carried out using Thunderbird SYBR qPCR Mix (Toyobo). The primers used are listed in Table S2.

Knockdown by siRNA

OL cells were transfected with siG9a, siEZH2, siESET, si-hseBLN-1, and siGFP using Lipofectamine 2000 (Invitrogen).

ACCESSION NUMBERS

Monkey *COMMD3* and *DNAJC1* partial mRNAs reported in this paper have been deposited to the GenBank and are available under accession numbers GenBank: AB934382.1 and GenBank: AB934381.1.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.08.007>.

AUTHOR CONTRIBUTIONS

K.S. and T.H. designed and performed experiments; K.S. and T.H. analyzed data; T.H. and K.T. conceived the project; and K.S., N.F.P., T.H., and K.T. wrote the manuscript.

ACKNOWLEDGMENTS

This study was supported in part by KAKENHI grant numbers 26253027, 26670225, 15H01259 (K.T.) and 25860336 (T.H.); the Core-to-Core Program A, Advanced Research Networks (K.T.) from Japan Society for the Promotion of Science (JSPS); a Basic Science and Platform Technology Program for Innovative Biological Medicine (K.T.) from Japan Agency for Medical Research and Development (AMED); a Grant-in-Aid for Scientific Research on Innovative Areas (24115709 and 25115508) (T.H.) from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan; and grants from the Takeda

Science Foundation, Senri Life Science Foundation, Suzuken Memorial Foundation, and The NOVARTIS Foundation (Japan) for the Promotion of Science (T.H.).

Received: April 7, 2015

Revised: July 3, 2015

Accepted: August 3, 2015

Published: August 27, 2015

REFERENCES

- Babushok, D.V., Ostertag, E.M., and Kazazian, H.H., Jr. (2007). Current topics in genome evolution: molecular mechanisms of new gene formation. *Cell. Mol. Life Sci.* **64**, 542–554.
- Belyi, V.A., Levine, A.J., and Skalka, A.M. (2010). Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog.* **6**, e1001030.
- Blazkova, J., Trejbalova, K., Gondo-Rey, F., Halfon, P., Philibert, P., Guiguen, A., Verdin, E., Olive, D., Van Lint, C., Hejnar, J., and Hirsch, I. (2009). CpG methylation controls reactivation of HIV from latency. *PLoS Pathog.* **5**, e1000554.
- Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99.
- Bradley, J., Baltus, A., Skaletsky, H., Royce-Tolland, M., Dewar, K., and Page, D.C. (2004). An X-to-autosome retrogene is required for spermatogenesis in mice. *Nat. Genet.* **36**, 872–876.
- Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P., and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409.
- Chuma, S. (2014). LINE-1 of evidence for fetal oocyte attrition by retrotransposon. *Dev. Cell* **29**, 501–502.
- Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliaskaite, L., Benes, V., Enright, A.J., and O'Carroll, D. (2013). Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Mol. Cell* **50**, 601–608.
- Dong, K.B., Maksakova, I.A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H.W., Lam, L.L., Mager, D.L., Schübeler, D., et al. (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* **27**, 2691–2701.
- Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**, 363–367.
- Fujino, K., Horie, M., Honda, T., Merriman, D.K., and Tomonaga, K. (2014). Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. *Proc. Natl. Acad. Sci. USA* **111**, 13175–13180.
- Goodier, J.L., Cheung, L.E., and Kazazian, H.H., Jr. (2013). Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* **41**, 7401–7419.
- Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N., and Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* **33**, 2374–2383.
- Häsler, J., Samuelsson, T., and Strub, K. (2007). Useful 'junk': Alu RNAs in the human transcriptome. *Cell. Mol. Life Sci.* **64**, 1793–1800.
- Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M., and Tomonaga, K. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463**, 84–87.
- Ishizu, H., Siomi, H., and Siomi, M.C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes Dev.* **26**, 2361–2373.
- Kalmykova, A.I., Dobritsa, A.A., and Gvozdev, V.A. (1998). Su(Ste) diverged tandem repeats in a Y chromosome of *Drosophila melanogaster* are transcribed and variously processed. *Genetics* **148**, 243–249.
- Kazazian, H.H., Jr. (1999). An estimated frequency of endogenous insertional mutations in humans. *Nat. Genet.* **22**, 130.
- Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632.
- Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22**, 908–917.
- Leung, D.C., and Lorincz, M.C. (2012). Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem. Sci.* **37**, 127–133.
- Livak, K.J. (1990). Detailed structure of the *Drosophila melanogaster* stellate genes and their transcripts. *Genetics* **124**, 303–316.
- Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A., and Kaessmann, H. (2005). Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol.* **3**, e357.
- Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931.
- McCarrey, J.R., and Thomas, K. (1987). Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature* **326**, 501–505.
- McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216.
- Mighell, A.J., Smith, N.R., Robinson, P.A., and Markham, A.F. (2000). Vertebrate pseudogenes. *FEBS Lett.* **468**, 109–114.
- Nabel, G., and Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature* **326**, 711–713.
- Noutsopoulos, D., Markopoulos, G., Vartholomatos, G., Kolettas, E., Kolaitis, N., and Tzavaras, T. (2010). VL30 retrotransposition signals activation of a caspase-independent and p53-dependent death pathway associated with mitochondrial and lysosomal damage. *Cell Res.* **20**, 553–562.
- Sayah, D.M., Sokolskaja, E., Berthou, L., and Luban, J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569–573.
- Speak, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* **21**, 1973–1985.
- Tachibana, M., Nozaki, M., Takeda, N., and Shinkai, Y. (2007). Functional dynamics of H3K9 methylation during meiotic prophase progression. *EMBO J.* **26**, 3346–3359.
- Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., and Hannon, G.J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534–538.
- Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–117.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539–543.
- Whitelaw, E., and Martin, D.I. (2001). Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* **27**, 361–365.
- Williams, S.A., Chen, L.-F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E., and Greene, W.C. (2006). NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO J.* **25**, 139–149.
- Yang, N., and Kazazian, H.H., Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* **13**, 763–771.
- Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340.

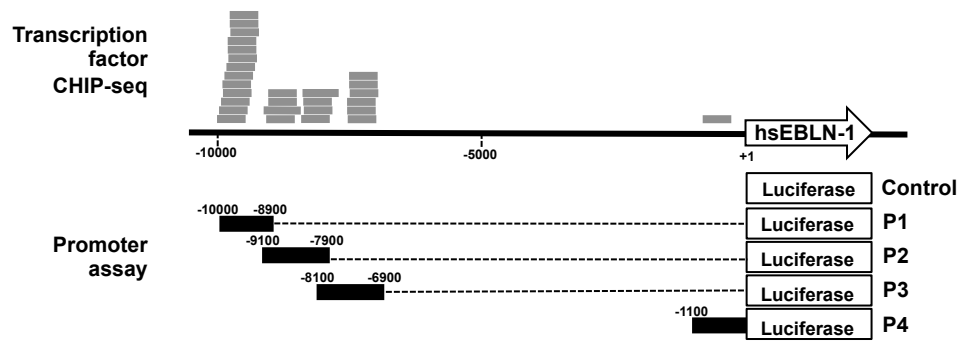
Cell Reports

Supplemental Information

**Transcription Profiling Demonstrates
Epigenetic Control of Non-retroviral RNA
Virus-Derived Elements in the Human Genome**

Kozue Sofuku, Nicholas F. Parrish, Tomoyuki Honda, and Keizo Tomonaga

A



B

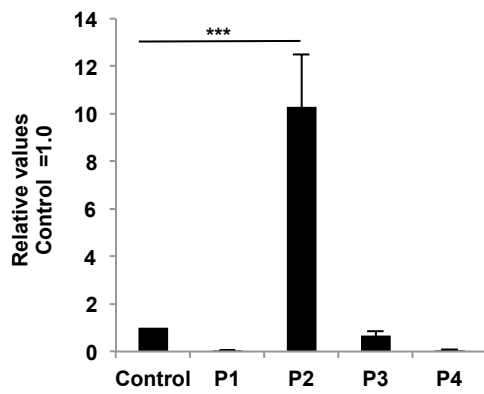


Figure S1

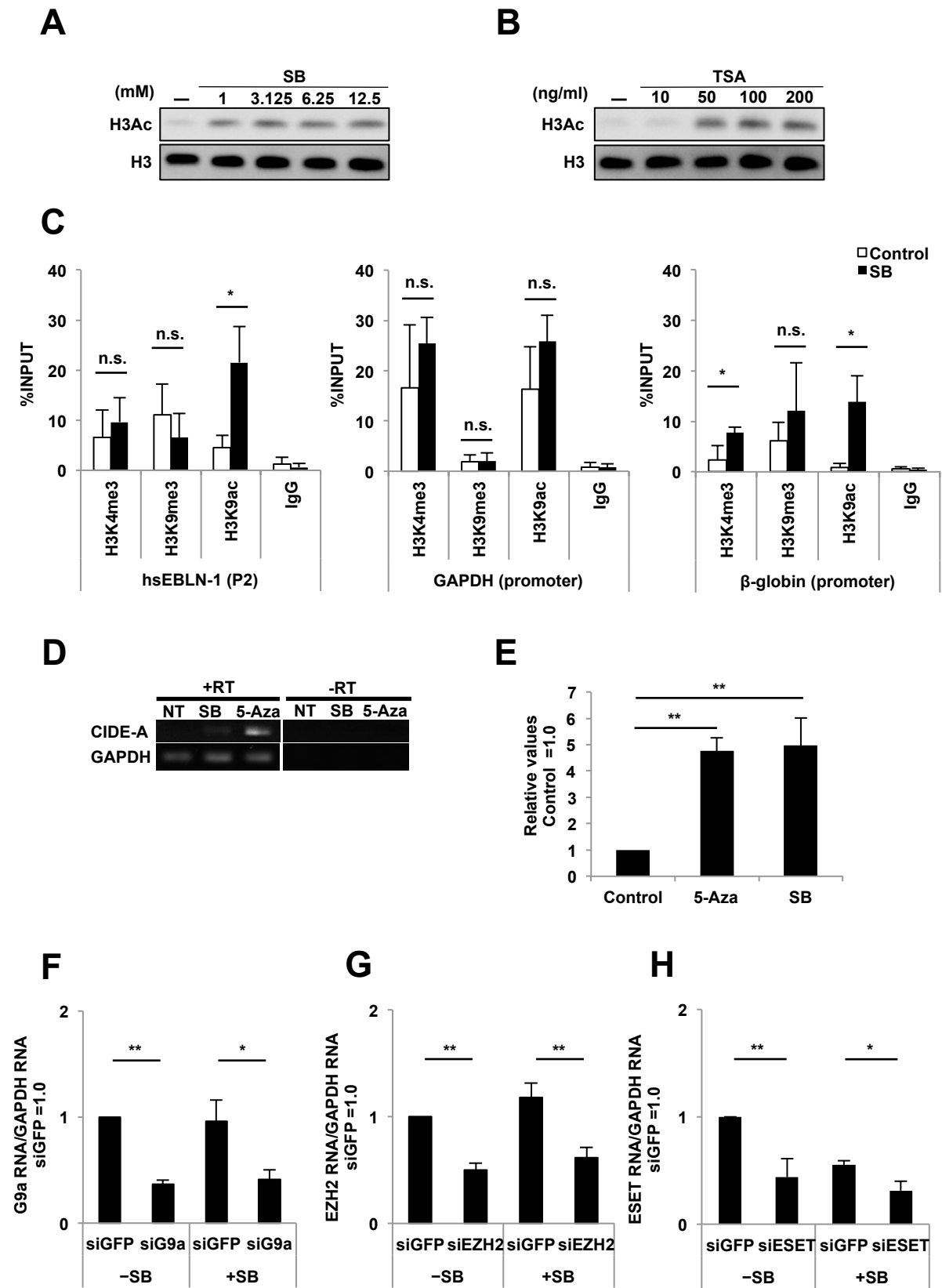
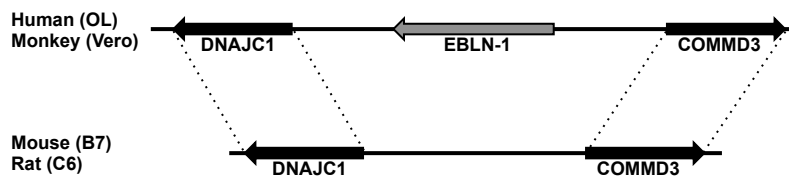
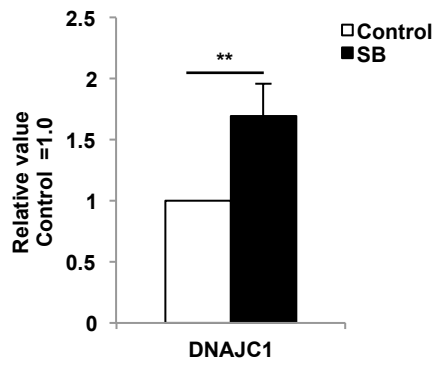
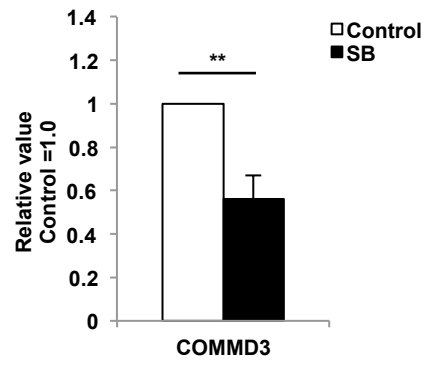
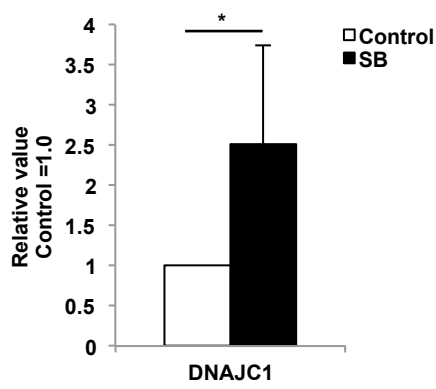
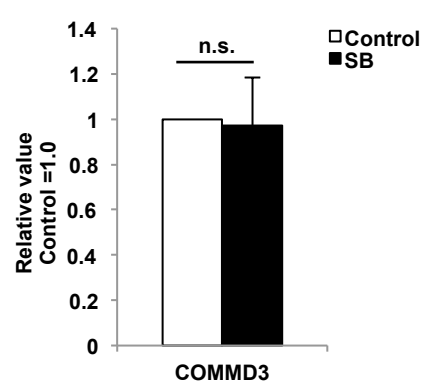


Figure S2

A**B****C****D****E****Figure S3**

Supplemental Figure Legends

Figure S1. Identification of the hsEBLN-1 promoter (Related to Figure 2). (A) Schematic map of the hsEBLN-1 locus. Grey bar indicates a transcription factor binding region based on Transcription Factor ChIP-seq in UCSC Genome Browser. Constructs used in the promoter assay (Control, P1, P2, P3 and P4) are indicated. The region indicated in black box were cloned upstream of a *Guassia luciferase* gene in the pGLuc-Basic vector. (B) Promoter assay of regions around the putative promoter site. The luciferase activity of each construct was normalized with that of co-expressed *Cypridina* luciferase. The relative values were calculated using empty pGLuc-Basic-transfected 293T cells (Control). Values are expressed as the mean + S.D. *** $P < 0.005$ (Student's *t* test). At least three independent experiments were performed.

Figure S2. Epigenetic regulation of hsEBLN loci (Related to Figure 3). (A, B) Western blot analysis of the effect of HDAC inhibitors, SB (A) or TSA (B). The levels of histone H3 acetylation (H3ac) and H3 were detected using antibodies against panH3Ac and H3, respectively. The level of H3 is indicated as a loading control. (C) ChIP analysis of histone H3 lysine 9 acetylation (H3K9ac), histone H3 lysine 9 trimethylation (H3K9me3) and histone H3 lysine 4 trimethylation (H3K4me3) in OL cells treated with 12.5 mM SB was performed using antibodies against H3K9ac, H3K9me3 and H3K4me3 respectively. Enrichment of promoter regions of hsEBLN-1 (P2), *GAPDH*, and *β -globin* was quantified by real-time RT-PCR analysis. The enrichment of the indicated regions is shown as percentage of input. (D) Induction of cell death-inducing DFFA-like effector A (*CIDE-A*), a known 5-Aza-responsive gene, after treatment of 10 μ M 5-Aza for 72 h. The expression

of *CIDE-A* was analyzed by RT-PCR. GAPDH mRNA expression is indicated as a loading control. (E) Quantitative analysis of hsEBLN-1 RNA after treatment of 5-Aza. OL cells were incubated for 72 h after addition of 10 μ M 5-Aza. As a positive control, OL cells treated with 12.5 mM SB for 18 h were used. The expression level of hsEBLN-1 RNA was normalized to that of GAPDH mRNA. The relative values were calculated using untreated group (Control). (F-H) Quantitative analysis of the target mRNA after treatment with siG9a (F), siEZH2 (G), or siESET (H). OL cells were incubated for 2 days after treatment with siG9a, siEZH2, or siESET with or without 12.5 mM SB. The expression level of HKMT RNAs were analyzed by real-time RT-PCR and normalized to that of GAPDH mRNA. The relative values were calculated using siRNA against GFP (siGFP)-treated group without SB (siGFP). Values are expressed as the mean + S.D. * $P < 0.05$, ** $P < 0.01$, n.s. no significance (Student's t test). At least three independent experiments were performed.

Figure S3. The expression of neighboring genes around hsEBLN-1 locus (Related to Figure 4). (A) The genome arrangement around the EBLN-1 loci in the human and monkey genomes and the syntenic loci in the mouse and rat genomes. (B, D) Quantitative analysis of the expression of the *DNAJC1* gene after treatment of SB. Monkey Vero cells (B) or rat C6 cells (D) was incubated for 18 h after treatment of 12.5 mM SB. The expression of the *DNAJC1* gene was analyzed by real-time RT-PCR using *DNAJC1*-specific primers. (C, E) Quantitative analysis of the expression of the *COMMD3* gene after treatment of SB. Vero cells (C) or C6 cells (E) was incubated for 18 h after treatment of SB. The expression of the *COMMD3* gene was analyzed by real-time RT-PCR using *COMMD3*-specific primers. The expression levels of the *DNAJC1* and

COMMD3 genes were normalized to that of *GAPDH* gene. The relative values were calculated using untreated group (Control). Values are expressed as the mean + S.D. * $P < 0.05$, ** $P < 0.01$, n.s. no significance (Student's t test). At least three independent experiments were performed.

Supplemental Tables

Table S1. hsEBLN loci in the human genome (Related to Figure 1).

EBLN	Chr.	Location #		Base	Blat	Longest ORF region	ORF *	Remark
		Start	End	(nt)	E-value		(aa)	
hsEBLN-1	10	22498822	22497818	1004	4.00E-68	22498935-22497813	366	An intron of the PPP4R2 gene
hsEBLN-2	3	73111419	73112361	942	7.00E-40 /2.00E-05	73111225-73112050	272	
hsEBLN-3	9	37086816	37087686	870	3.00E-29 /3.00E-13	37086793-37087147	-	An intron of the TOMM20 gene
hsEBLN-4	17	4474361	4475371	1010	3.00E-32	4474831-4475170	-	
hsEBLN-5	1	235281323	235280959	364	2.00E-07	235281334-235281049	-	An exon of the TBC1D12 gene
hsEBLN-6	11	74723431	74723123	308	7.00E-05	74723197-74723092	-	
hsEBLN-7	10	96294426	96294130	296	6.00E-05	96294356-96294047	-	

* The amino acid length of putative protein coded in the longest ORF (>600 nt) located at the locus homologous to BDV N. “-” indicates that there is no ORF (>600nt) at the locus.

Table S2. Primers and siRNAs used in this paper (Related to Figures 2-4 and S1-3).

Primers

Detected gene	Fw/Rv	5'→3'	Reference
hsEBLN-1	Fw	TCGAGTGACCCGATATGCAC	Self-designed
	Rv	ATCAAGTAGCTCGCACGCAG	
hsEBLN-2	Fw	TCTCATTTCCACCCTGTGACC	Self-designed
	Rv	CCTCGTTCTCATTCCCAACC	
hsEBLN-3	Fw	GGGTGCCCTGTATCACCCCG	Self-designed
	Rv	TCCAGTCCCACTGGATGCCT	
hsEBLN-4	Fw	CTGTTGTGACCCGGTGGGCA	Self-designed
	Rv	TGGTCCGACCCACGGTCTCG	
hsEBLN-5	Fw	TTGGCAGTGGAGATTCACCTG	Self-designed
	Rv	ATTGGCTGCTGTGGAAGAC	
hsEBLN-6	Fw	TGCAGCTTTGTTGACTGGAG	Self-designed
	Rv	TTAGCGGGTGGTAATTCCTCAC	
hsEBLN-7	Fw	TGCATTTCTTGGCGGTGAG	Self-designed
	Rv	TTCTGCTTCCTCACATTGGC	
Human GAPDH	Fw	AGCGAGATCCCTCCAAAATC	Self-designed
	Rv	AAATGAGCCCCAGCCTTCTC	
hsEBLN-1	Fw	CAGGGACTGGTAGGAGGTGA	Self-designed
(P2)	Rv	CATGGCTGAAGGAGGAGAAG	
Human GAPDH	Fw	TACTAGCGGTTTTACGGGCG	Kimura H. et al.
(Promoter)	Rv	TCGAACAGGAGGAGCAGAGAGCGA	Cell Struct Funct. 2008
Human β -globin	Fw	GGGCTGAGGGTTTGAAGTCC	Kimura H. et al.
(Promoter)	Rv	CATGGTGTCTGTTTGAGGTTGC	Cell Struct Funct. 2008
Human CIDE-A	Fw	AAGAGGTCGGGAATAGCGAGAGTC	Li D. et al.
	Rv	CTGCATCCCTATCCACACGTGAACC	Nucleic Acids Res. 2008
Human G9a	Fw	TGCTCATGCTGGAGAACTG	Self-designed
	Rv	TTGCCATAGTAACGGGCATC	
Human EZH2	Fw	CCCTGACCTCTGTCTTACTTGTGGA	Fujii S. et al.
	Rv	ACGTCAGATGGTGCCAGCAATA	J Biol Chem. 2008
Human ESET	Fw	GACTCTCTGAGACAACTCCAAGGA	Watanabe H. et al.
	Rv	CAGGGATTGAGGGAGGAACA	Cancer Cell Int. 2008
Human DNAJC1	Fw	TTGGTTGGAAAACAGGAACC	Self-designed
	Rv	TCCTTCAGTTGCTTGGCTTT	

Human COMMD3	Fw	AAAGCACCGAGCTGACAAGT	Self-designed
	Rv	CGCCAAGAAACATCCGTTAT	
Mouse DNAJC1	Fw	AGGAGCAAATGGATGACTGG	Self-designed
	Rv	TCACAGATCGACCCAATTCA	
Mouse COMMD3	Fw	AAAGCATTGTCATGCAGCAG	Self-designed
	Rv	AAGCGCCAAGAAACATCAGT	
Vero DNAJC1	Fw	TTGGTTGGAAAACAGGAACC	Self-designed
	Rv	TCCTTCAGTTGCTTGGCTTT	
Vero COMMD3	Fw	AAAGCACCGAGCTGACAAGT	Self-designed
	Rv	CGCCAAGAAACATCCGTTAT	
Rat DNAJC1	Fw	GATTGGCGACAGCCTGTATT	Self-designed
	Rv	TGCATCCACACTCTTGCTTC	
Rat COMMD3	Fw	GTTTCCTGGCGATTGGAATA	Self-designed
	Rv	TTCCCCACCAAATCCTGTAA	
Mouse GAPDH	Fw	GCACAGTCAAGGCCGAGAAT	Chen S. et al.
	Rv	GCCTTCTCCATGGTGGTGAA	Proc Natl Acad Sci U S A. 2006
Vero β -actin	Fw	GCGCGGCTACAGCTTACCAC	Self-designed
	Rv	GGGCGCCAGGGCAGTAATCTC	
Rat GAPDH	Fw	TCCCTCAAGATTGTCAGCAA	Silva WI. et al.
	Rv	AGATCCACAACGGATACATT	Int J Dev Neurosci. 2005

siRNAs

Target gene		5'→3'
G9a	Duplex	CCAUGCUGUCAACUACCAUGGTT
EZH2	Duplex	CUCAAGAGGUUCAGACGAGCUTT
GFP (control siRNA)	Duplex	GGCUACGUCCAGGAGCGCACCTT
hsEBLN-1	Duplex	UGUUAAGUGUAUAUGGAUCCATT

Supplemental Experimental Procedures

Cells

OL, 293T, Vero and C6 cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco) containing 5% fetal bovine serum (FBS).

Chemicals and antibodies

Sodium butyrate (SB), trichostatin A (TSA), and 5-aza-2'-deoxycytidine (5-Aza) were purchased from Sigma-Aldrich. Anti-pan histone H3 (cloneA3S, Millipore), anti-histone acetyl H3 (Upstate), anti-H3K4me3 (Active Motif), anti-H3K9me3 (Active Motif), and anti-H3K9ac (Active Motif) antibodies were purchased from the commercial sources listed. Normal rabbit IgG was purchased from Sigma-Aldrich. Horseradish peroxidase (HRP)-coupled goat anti-rabbit antibody was purchased from Jackson ImmunoResearch. siRNA against ESET was purchased from Santa Cruz Biotechnology. siRNAs against GFP, G9a, EZH2, and hEBLN-1 RNA were purchased from Sigma-Aldrich.

Promoter assay

The reporter constructs were generated by cloning a fragment of prospected promoter region that was approximately 1.2 kb into the pGLuc-Basic vector (New England BioLabs). Cells were co-transfected with the reporter construct or an empty vector and the pSV40-CLuc vector (New England BioLabs) using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, luciferase activity was measured in cell culture supernatants using BioLux Luciferase assay kit (New England BioLabs). The luciferase activity of each construct was normalized with that of co-expressed *Cypridina* luciferase. The relative values were calculated using empty pGLuc-Basic-transfected 293T cells (Control).

RT-PCR analysis

Total RNAs were extracted from OL, Vero, and C6 cells by using TRIzol reagent

(Invitrogen). Total RNAs of human tissues were purchased from Clontech. The RNAs except for brain whole, colon, kidney and liver were from pools of donors. First-strand cDNAs were synthesized from aliquots of 1 µg of total RNA by a Verso cDNA Synthesis Kit (Thermo Scientific) using random hexamers. PCR was performed in a total volume of 25 µl containing 1 µl of cDNA and 2 U of TaKaRa ExTaq polymerase (Takara). The PCR was performed at 94°C for 30 sec, at the annealing temperature for 30 sec, and at 72°C for 30 sec. The optimal number of amplification cycles and annealing temperature were changed for each primer. Real-time RT-PCR assays were carried out using Thunderbird SYBR qPCR Mix (Toyobo) in the 7900HT Fast Real-Time PCR System (Applied Biosystems), as described previously (Hayashi et al., 2009). The PCR was performed 40 cycles at 95°C for 15 sec and at 60°C for 33 sec. The primers used are listed in Table S2.

Western blotting analysis

Cell lysates were separated on 15% SDS polyacrylamide gel and transferred onto a PVDF membrane. The membrane was probed with anti-histone H3 or anti-H3ac antibody overnight at 4°C. After washing, the membrane was probed with HRP-coupled goat anti-rabbit antibodies. The blots were developed using ECL Plus Western Blot Detection Reagents (GE Healthcare) according to the manufacturer's instructions.

Chromatin immunoprecipitation (ChIP) analysis

ChIP analysis was performed using OneDay ChIP Kit (Diagenode). Briefly, OL cells were cross-linked with 1% formaldehyde for 8 min at room temperature. Cells were re-suspended in the provided ChIP buffer including proteinase inhibitors. Samples were sonicated with a Bioruptor (Diagenode) and incubated with 100 U/ml micrococcal

nuclease (Worthington Biochemical) for 40 min at 37°C to obtain fragments with a size range between 100 and 500 bp. After centrifugation, the supernatants were incubated with the appropriate antibodies overnight at 4°C. After the incubation, immunoprecipitation, DNA elution, and DNA purification were performed according to the manufacturer's instructions. The levels of DNA loci of interest in the immunoprecipitates were quantified by real-time RT-PCR analysis.

Knockdown by siRNA

OL cells were transfected with siRNAs against G9a, EZH2, ESET, GFP and hEBLN-1 RNA using Lipofectamine 2000. The cells were incubated for 48 or 72 h and collected for real-time RT-PCR analysis.

Supplemental References

Hayashi, Y., Horie, M., Daito, T., Honda, T., Ikuta, K., and Tomonaga, K. (2009). Heat shock cognate protein 70 controls Borna disease virus replication via interaction with the viral non-structural protein X. *Microbes Infect.* 11, 394–402.