LTRs activated by Epstein-Barr virus–induced transformation of B cells alter the transcriptome

Amy Leung,[1] Candi Trac,[1] Hiroyuki Kato,[1] Kevin R. Costello,[1,2] Zhaoxia Chen,[1] Rama Natarajan,[1,2] and Dustin E. Schones[1,2]

[1]Department of Diabetes Complications and Metabolism, Beckman Research Institute, City of Hope, Duarte, California 91010, USA; [2]Irell and Manella Graduate School of Biological Sciences, City of Hope, Duarte, California 91010, USA

Endogenous retroviruses (ERVs) are ancient viral elements that have accumulated in the genome through retrotransposition events. Although they have lost their ability to transpose, many of the long terminal repeats (LTRs) that originally flanked full-length ERVs maintain the ability to regulate transcription. While these elements are typically repressed in somatic cells, they can function as transcriptional enhancers and promoters when this repression is lost. Epstein-Barr virus (EBV), which transforms primary B cells into continuously proliferating cells, is a tumor virus associated with lymphomas. We report here that transformation of primary B cells by EBV leads to genome-wide activation of LTR enhancers and promoters. The activation of LTRs coincides with local DNA hypomethylation and binding by transcription factors such as RUNX3, EBF1, and EBNA2. The set of activated LTRs is unique to transformed B cells compared with other cell lines known to have activated LTRs. Furthermore, we found that LTR activation impacts the B cell transcriptome by up-regulating transcripts driven by cryptic LTR promoters. These transcripts include genes important to oncogenesis of Hodgkin lymphoma and other cancers, such as HUWE1/HECTH9. These data suggest that the activation of LTRs by EBV-induced transformation is important to the pathology of EBV-associated cancers. Altogether, our results indicate that EBV-induced transformation of B cells alters endogenous retroviral element activity, thereby impacting host gene regulatory networks and oncogenic potential.

[Supplemental material is available for this article.]

Endogenous retroviruses (ERVs) make up ~8% of the human genome (Lander et al. 2001). A complete ERV element contains two long terminal repeats (LTRs), containing the regulatory elements responsible for transcription, flanking the Gag-, Pol-, and Env-coding sequences needed for transposition activity (Mager and Stoye 2015). In the human genome, the vast majority of ERVs are solitary LTRs, and many of these still maintain the ability to regulate transcription under certain conditions. A subset of these LTRs has been co-opted as regulatory elements for other genes (Maksakova et al. 2008; Rowe and Trono 2011; Leung and Lorincz 2012; Thompson et al. 2016).

There are many examples in which LTR-driven transcripts play a role in pluripotent cells. In the early embryo of both humans and mice, specific ERV groups are active and drive expression of many transcripts (Evsikov et al. 2004; Peaston et al. 2004; Macfarlan et al. 2012; Maksakova et al. 2013; Göke et al. 2015). Likewise, both embryonic stem cells and induced pluripotent stem cells express LTR-driven transcripts important for pluripotency (Leung and Lorincz 2012; Fort et al. 2014; Lock et al. 2014; Lu et al. 2014; Ohnuki et al. 2014; Hashimoto et al. 2015; Babaian et al. 2016; Klawitter et al. 2016). In pathways other than development, LTR-driven transcripts have also been found to be expressed in autoimmune diseases and several types of cancer, including lymphoma, hepatocellular carcinoma, and prostate cancer (Lamprecht et al. 2010; Prensner et al. 2013; Lock et al. 2014; Hashimoto et al. 2015; Babaian et al. 2016; Chuong et al. 2016). While this work has led to an appreciation of the degree of aberrant LTR activation in normal and disease cells, the mechanisms that lead to LTR activation have remained unclear.

Epstein-Barr virus (EBV) was originally discovered in cells derived from a Burkitt's lymphoma biopsy. EBV can readily infect primary B lymphocytes (B cells) in vitro and transform them into continuously proliferating lymphoblastoid cell lines (LCLs), leading to the classification of EBV as a tumor virus (Epstein et al. 1964; Pope et al. 1973). In addition to lymphomas, EBV has been associated with many other cancer types, including nasopharyngeal carcinoma, gastric adenocarcinoma, and lymphoepithelioma-like carcinomas (Hsu and Glaser 2000). In memory B cells, EBV can remain quiescent in a latent state using one of four latency programs, all of which are characterized by distinct expression of EBV-encoded genes, including transcription factor genes (Thorley-Lawson 2001). Transformation of primary B cells leads to genome-wide transcriptional changes, including decreased transcription of apoptotic genes and increased transcription of proliferative genes (Allday 2013; Price and Luftig 2014; Price et al. 2017). EBV-induced transformation of primary B cells also leads to extensive DNA hypomethylation across the genome, similar to what is observed in cancer cells (Hernando et al. 2013; Hansen et al. 2014). It has furthermore been found that EBV infection in primary B cells transactivates a human ERV locus, ERVK-18 (also known as HERV-K18), leading to the expression of a superantigen important for T-cell response (Sutkowski et al. 1996). Based up on this, we set out to determine if EBV-induced transformation of B cells leads to activation of LTRs.

To evaluate the impact of EBV-induced transformation of primary B cells on LTRs across the genome, we examined H3K4me3,

a chromatin modification associated with regulatory activity, in donor-matched primary B cells and LCLs. Furthermore, we examined the dynamics of DNA methylation, the binding of transcription factors (TFs), and the transcriptional changes of LTRs in B cells and LCLs.

Results

H3K4me3 profiling reveals widespread enrichment at LTRs in LCLs

To explore the potential impact of EBV-induced transformation on the chromatin environment of B cells, we generated H3K4me3 profiles from three donor primary B cells and matched LCLs derived from the same donor B cells (Fig. 1A; Supplemental Fig. S1A). We identified a union set of 105,967 peaks (sites) of H3K4me3, a chromatin modification associated with enhancers and promoters, from B cells and LCLs. To identify genes that are differentially enriched for H3K4me3, we performed DESeq2 analysis (Methods; Supplemental Methods) on read counts found in the 2-kb region surrounding the promoter (i.e., ±1 kb) of genes (hg19 RefSeq annotation) that overlap a H3K4me3 site, and found 9654 sites as variable across the two groups (adjusted $P < 0.01$) with 5853 increased and 3801 decreased (Fig. 1B, top). Up-regulated sites were enriched

at promoters of genes with the top KEGG term of "pathways in cancer" (Supplemental Fig. S1B). Genes found in this KEGG term included many genes known to be up-regulated by EBV, including *FAS*, *IKBKB*, and *TRAF1* (Supplemental Fig. S1C; Hernando et al. 2013). These data demonstrate that the H3K4me3 profiles of primary B cells and LCLs capture known transcriptional changes at genes related to the transformation of B cells to LCLs.

We next examined the possibility that EBV-induced transformation influences transcriptional activity, as measured by increase in H3K4me3 enrichment, of LTR promoters and enhancers across the genome. Taking all LTR sites (RepeatMasker) (Smit et al. 2013–2015) found in the human genome, we performed DESeq2 analysis with read counts mapping to LTR sites that overlap a H3K4me3 site. Of the 2138 differentially enriched LTR sites (adjusted $P < 0.05$) discovered by this analysis, the vast majority (2039) display an increase in H3K4me3 enrichment in LCLs compared with B cells (Fig. 1B, bottom). These LTRs sites can be found at the transcriptional start sites (TSSs) of both long noncoding RNAs (e.g., *LINC02363*) (Fig. 1A) and protein-coding genes (Supplemental Fig. S2). In comparison to TSSs of genes, which show roughly equivalent degrees of increase and decrease in H3K4me3 enrichment in LCLs compared with B cells (Fig. 1B, top), LTRs have a marked increase in H3K4me3 in the LCLs (Fig. 1B, bottom). We next distinguished the LTRs and genes that are Group I, "activated" by complete gain of H3K4me3 (i.e., not enriched for H3K4me3 in B cells); Group II, "up-regulated" by increase in H3K4me3 (i.e., enriched with H3K4me3 in B cells); Group III, "down-regulated" (i.e., decrease in H3K4me3, but still enriched in LCLs; and Group IV, "repressed" (i.e., no H3K4me3 enrichment in LCLs) (Fig. 1C). With this classification, 1134 LTRs are found in Group I and from hereon will be referred to as "activated LTRs." These data indicate a widespread activation of LTR promoters after EBV-induced transformation. ERVK-18, an ERV1 element known to be induced by EBV infection, is found in the activated group as expected (Supplemental Fig. S3). Examining the specific subfamilies of LTR that are activated by EBV transformation revealed that the majority of the activated LTRs belong to either the ERV1 or ERVL-MaLR/ERVL families of LTRs (Fig. 1D). While there is great variety in the subfamilies of LTRs (241 subfamilies) (Supplemental Table S1), we found a dramatic activation of MLT subfamily members than expected by chance (e.g., MLT1D [$P = 0.013$], MLT1K [$P = 0.008$], and MLT1F1 [$P = 0.002$]), and MER41 subfamily members (e.g., MER41A [$P < 0.001$], and MER41B [$P = 0.011$]) (Supplemental Table S1). These data indicate that specific subfamilies of LTRs may be more susceptible to activation. To assess if there are any DNA rearrangements at LTRs that may lead to activation of these LTRs, we performed PCR analysis of genomic DNA and confirmed that there is no
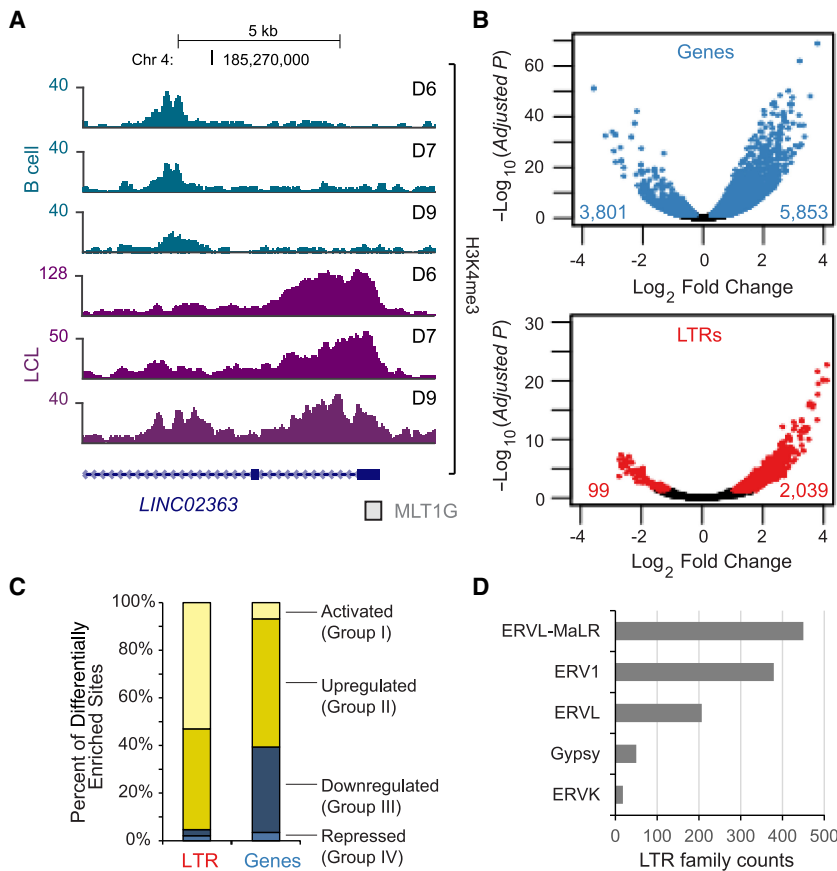

Figure 1. EBV-induced transformation of B cells activates LTRs. (*A*) Genome browser tracks of H3K4me3 profiles at *LINC02363*, a long noncoding gene, with the MLT1G LTR at the transcriptional start site (TSS). (*B*) Volcano plots of normalized H3K4me3 read counts in LCLs relative to B cells at TSSs of genes (*top*) and at LTRs (*bottom*). (*C*) Percentage of differentially enriched H3K4me3 LTRs and genes classified as "activated," "up-regulated," "down-regulated," and "repressed." (*D*) Enumeration of the LTR families that are activated in LCLs (Group I).

obvious rearrangements in the two assayed LTRs (Supplemental Fig. S4).

Cryptic LTR promoter and enhancer activation coincides with loss of DNA methylation and binding of transcription factors

LTRs can be silenced through repressive histone modifications (Rowe and Trono 2011; Leung and Lorincz 2012; Liu et al. 2014), corepressors (Jacobs et al. 2014; Wolf et al. 2015), and DNA methylation (Wolf and Goff 2008). EBV transformation of B cells has been shown to induce large-scale loss of DNA methylation, akin to the loss of DNA methylation in cancer cells (Hernando et al. 2013; Hansen et al. 2014). We therefore examined the DNA methylation profiles of activated LTRs after EBV transformation. Utilizing whole-genome bisulfite sequencing data from primary B cells and matched LCLs (GSE49627) (Hansen et al. 2014), we examined DNA methylation and H3K4me3 (this paper) at the activated LTRs. Comparing the average CpG methylation level change at LTRs in three donor B cells with the three LCLs revealed that activated LTRs are generally hypomethylated upon EBV treatment and have corresponding increases in H3K4me3 (Fig. 2A, representative example Fig. 2B). In contrast to the large-scale hypomethylated blocks previously reported in EBV-induced B cell immortalization (Hansen et al. 2014), the DNA methylation changes we observed were more local changes, specific to the LTR elements (Fig. 2B; Supplemental Fig. S5). These data support the idea that hypomethylation is concurrent with LTR activation. We next addressed whether loss of DNA methylation itself is sufficient to activate LTRs. Examining DNA methylation at all LTRs, regardless of activation status, indicated that hypomethylation does not necessarily result in gain of H3K4me3 modifications after EBV-induced transformation (Fig. 2C), suggesting that loss of DNA methylation is not sufficient to activate LTRs in B cells and that additional regulatory factors are required.

To gain a better understanding of the regulatory factors involved in activating LTRs, we next examined transcription factor binding at activated LTRs in GM12878 cells using ChIP-seq data for TFs generated from ENCODE (Dunham et al. 2012), along with EBNA2 (GSE29498), RPBJ (GSE29498), and NF-κB subunits (GSE55105) (Zhao et al. 2011, 2014). Overall, we found that there are 36 transcription factors bound to at least 25% of the activated LTR sites (Fig. 3A). A number of these transcription factors are up-regulated in LCLs, including RUNX3, EBF1, and EBNA2 (Supplemental Fig. S6), as expected based upon existing literature (Spender et al. 2002; Tempera et al. 2016). RUNX3, a transcription factor expressed throughout the hematopoietic system that is specifically up-regulated by EBV during LCL development and promotes cell proliferation (Spender et al. 2002; Brady et al. 2009), is bound at ∼67% of LTRs activated by EBV-induced transformation. EBF1, which is important for B cell lineage development (Boller and Grosschedl 2014), is also among the most enriched transcription factors at activated LTRs. IKAROS/IKZF1, which binds to ∼25% of activated LTRs, plays key roles in the latent-lytic switch of EBV in B cells, indicating that perhaps LTR activation is influenced by EBV-induced pathways (Iempridee et al. 2014). EBNA2, a TF encoded in the EBV genome, is bound to ∼30% of the EBV-activated LTRs (Fig. 3A), Given that many transcription factors bind to common regulatory regions (Wang et al. 2013), we examined the co-occupancy of transcription factors bound at activated LTRs by performing hierarchical clustering of pairwise correlations of binding patterns (Methods). This analysis revealed clusters that appear to co-occupy LTRs together (Fig. 3B). EBNA2 cobinds with RBPJ, as has been previously demonstrated (Zhao et al. 2011). Other clusters include a RUNX3 cluster with many of the factors binding the largest fraction of activated LTRs, a general TF/chromatin cluster with EP300/CHD2, and an NF-κB cluster with RELA, RELB, REL (also known as c-Rel), and NFKB2 (also known as p52) (Fig. 3B). We next asked whether a particular subfamily is more likely to be bound by the top 10 transcription factors. Overall, the LTR subfamily bound by these transcription factors is diverse (Supplemental Fig. S7; Supplemental Table S2). We next examined the number of EBV-activated LTRs that give rise to transcripts using cap-analysis of gene expression (CAGE) data from GM12878 cells (ENCODE) (Hoffman et al. 2013). In total, ∼31% of EBV-activated LTRs overlap CAGE-enriched sites, indicating this subset of LTRs can initiate transcription (Fig. 3C). Those without CAGE evidence are likely either transcribing weakly or serving as regulatory regions for other genes.

To determine if RUNX3 binding at LTRs is necessary to drive transcription of LTR-driven transcripts, we performed
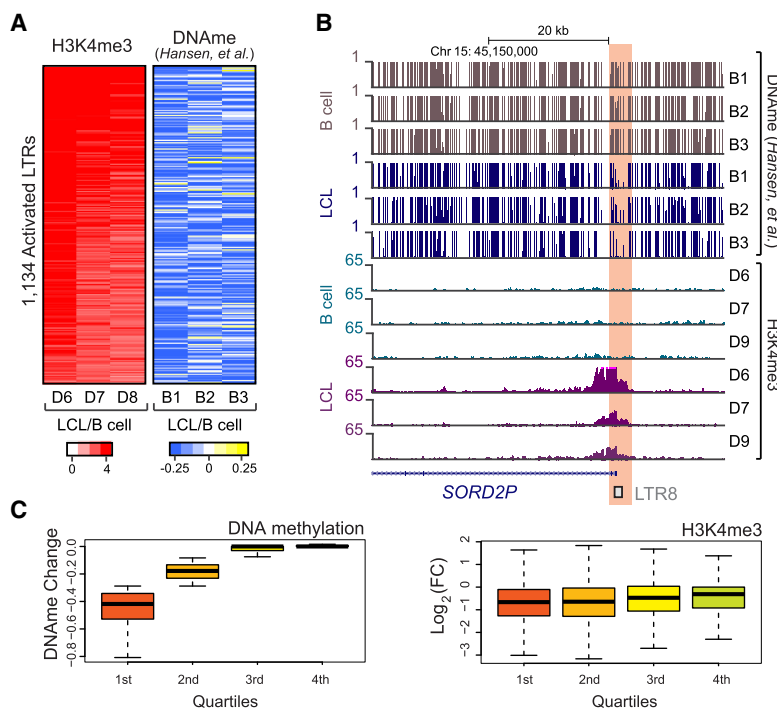
Figure 2. Activated LTRs are hypomethylated in LCLs. (*A, left*) Heatmap of H3K4me3 in three B cell donors and three matched LCLs at activated LTRs; (*right*) heatmap of DNA methylation changes in matched B cells and LCLs at activated LTRs. (*B*) Genome browser tracks of DNA methylation and H3K4me3 profiles in B cells and LCLs at a *SORD2P* locus driven by an LTR8. (*C*) Boxplot of DNA methylation changes (*left*) and H3K4me3 change (*right*) between B cells and LCLs stratified into quartiles based on DNA methylation change.

Figure 3. Activated LTRs are binding sites for many transcription factors in GM12878 cells and can promote transcription. (*A*) Bar graphs of the percentage of activated LTRs bound by each transcription factor. (*B*) Correlation plots (Pearson's) of transcription factor co-occupancy at EBV-activated LTRs. Hierarchical clustering was performed to identify transcription factors that co-occur. (*C*) Percentage of activated LTRs with CAGE sequences in GM12878 cells, indicating promoter acti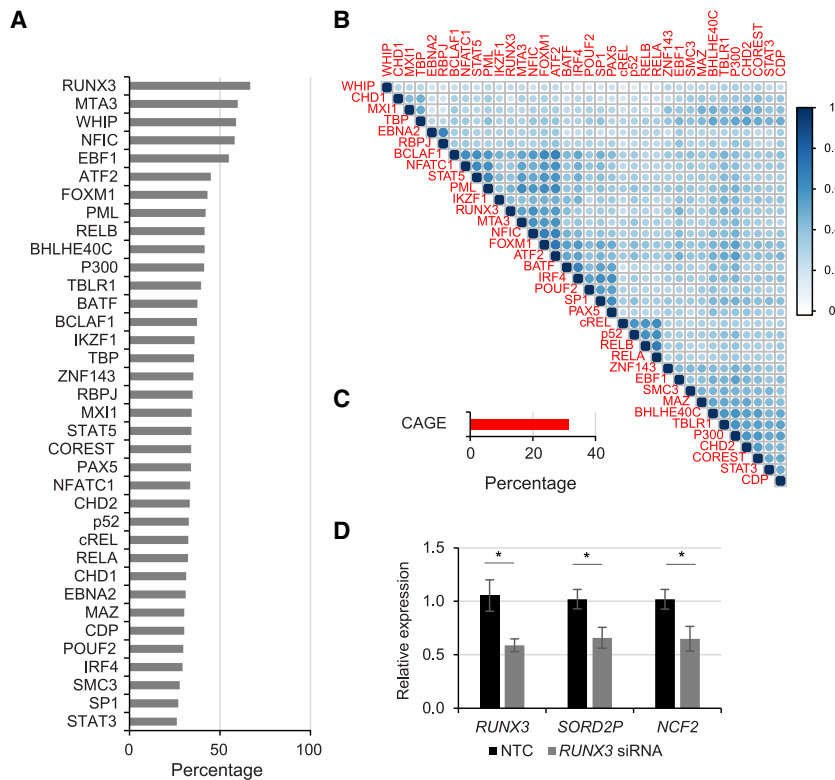vity. (*D*) RT-qPCR of *RUNX3*, *SORD2P*, and *NCF2* transcripts after siRNA knockdown of *RUNX3* transcripts. (*) $P < 0.05$.

siRNA knockdown of *RUNX3* and examined the expression of several genes driven by cryptic LTR promoters (Fig. 3D). Upon *RUNX3* knockdown, the genes we examined were reduced in expression, consistent with RUNX3 binding at cryptic LTR promoters contributing to the expression of these genes.

LTRs activated by EBV are mostly unique to LCLs

Given that LTRs were previously shown to be active in other cells lines, including H1 (embryonic stem cell) and K562 (immortalized myelogenous leukemia) cells (Djebali et al. 2012; Fort et al. 2014), we next addressed whether the LTRs activated by EBV-induced transformation are unique to LCLs or are generally active in cell lines. By using H3K4me3 profiles generated by ENCODE from GM12878, H1, Hela S3, HUVEC, HepG2, and K562 cells (Dunham et al. 2012), we assessed the enrichment of H3K4me3 across these cell lines at activated LTRs. By using hierarchical clustering of the H3K4me3 profiles generated from two different sequencing centers within the ENCODE project, we found that the profiles from the same cell lines cluster together and that the vast majority of the activated LTRs are not enriched for H3K4me3 in other cells (Fig. 4A). These data suggest that the activation of LTRs by EBV-induced transformation occurs at a subset of loci that are unique from those activated in other cells. Indeed, analysis of H3K4me3 enrichment at all LTRs across the six cells revealed that each cell line has a distinct profile of active LTRs (Supplemental Fig. S8). We further profiled H3K4me3 profiles in

different normal tissue types from RoadMap data (Roadmap Epigenomics Consortium et al. 2015) and found that the enrichment at the activated LTRs is low compared with LCLs (Supplemental Fig. S9). These data agree with previous reports that LTR expression and activity is cell-type specific (Faulkner et al. 2009; Jacques et al. 2013; Xie et al. 2013).

We furthermore examined whether these activated LTRs in GM12878 cells are classified as transcriptionally active regulatory regions. At specific loci, such as at *LINC01055* (Fig. 4B), the LTR marked by H3K4me3 is classified by combined analysis from Segway and ChromHMM (Ernst and Kellis 2010, 2012; Hoffman et al. 2012, 2013) as a TSS. By examining the functional classification from Segway and ChromHMM of activated LTRs across the genome, we found that the majority of the activated LTRs are classified as promoters or enhancers (>80%). In contrast, these same regions are mostly classified as repressed regions in other cell lines (Fig. 4C).

EBV-induced activation of LTRs is sufficient to drive lymphoma- and cancer-related transcripts

Activation of cryptic LTR promoters has been demonstrated to promote expression of TE-driven transcripts (Babaian and Mager 2016; Elbarbary et al. 2016). To assess the potential impact of LTR activation, we examined whether activated LTRs can lead to expression of cryptic transcripts. A THE1B LTR drives the expression of a noncanonical isoform of *CSF1R*, leading to lineage inappropriate expression of the gene in lymphoma cells (Lamprecht et al. 2010), and a LOR1a LTR drives the expression of *IRF5*, which is a key regulator in the development of lymphomas (Babaian et al. 2016). By RT-qPCR analysis of RNA isolated from B cells and LCLs, we found that the noncanonical transcript is indeed expressed in LCLs but not in B cells (Supplemental Fig. S10). The canonical *CSF1R* transcript, however, is expressed in both LCLs and primary B cells, which was previously observed (Baker et al. 1993). Similarly, in addition to the canonical transcript, LCLs express the noncanonical *IRF5* isoform driven by the LOR1a LTR (Supplemental Fig. S10). These data indicate that EBV-induced transformation of B cells is sufficient to activate LTRs driving transcripts known to be important in the development of lymphomas. We further examined the presence of additional LTR-driven transcripts. By using StringTie reference guided transcript assembly with RefSeq gene annotations, we found that there are 31 LTR-driven transcripts with five previously unannotated that are up-regulated in the LCLs (Supplemental Table S3; Supplemental Fig. S11).

Of the LTR-driven cryptic transcripts identified in our data set, an additional transcript is related to cancer. The *HUWE1/HECTH9* locus has a cryptic/noncanonical isoform originating from an MLT1L LTR upstream of the canonical isoform (Fig. 5A; Supplemental Fig. S12). *HUWE1* is involved in the transcriptional
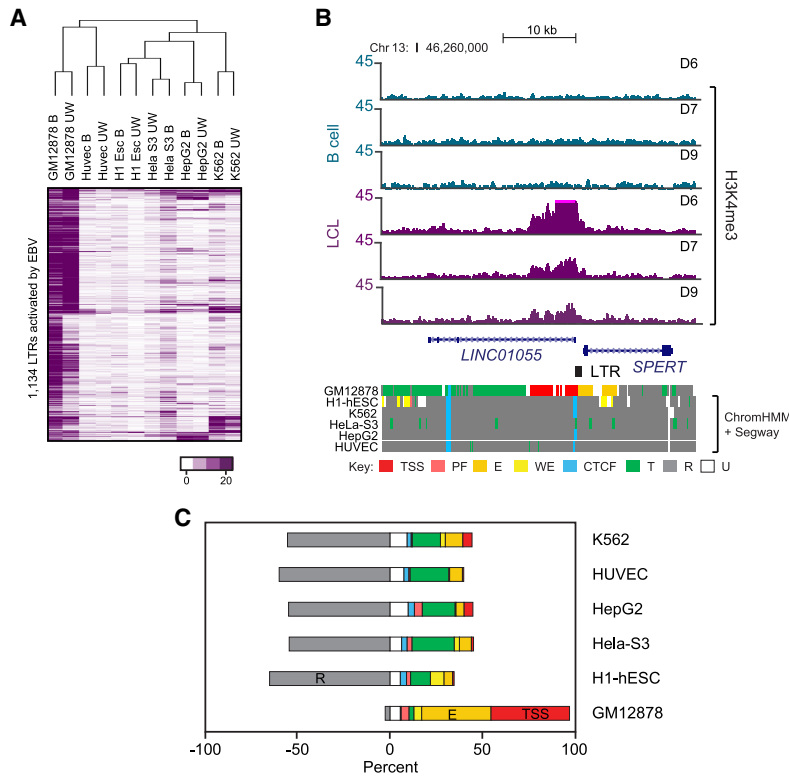
Figure 4. Transcriptional activity of activated LTRs is mostly specific to transformed B cells. (*A*) Hierarchical clustering of H3K4me3 read counts across cell lines from ENCODE data. (B) Broad; (UW) University of Washington; replicates from different sequencing centers. (*B*) Genome browser track of H3K4me3 profile in B cells and LCLs, and of ChromHMM + Segway combined data at the *LINC01055-SPERT1* locus. (TSS) Predicted promoter region including TSS; (PF) predicted promoter flanking region; (WE) predicted weak enhancer or open chromatin *cis* regulatory element; (CTCF) CTCF enriched element; (T) predicted transcribed region; (R) predicted repressed or low activity region; and (U) unclassified. (C) ChromHMM + Segway combined classification of all activated LTRs shown as a frequency of total for each ENCODE cell line (color schemes as in *B*).

activity of MYC and has been found to be essential to cell proliferation in cancer cells (Adhikary et al. 2005). From our RNA-seq data set, expression of this cryptic *HUWE1* transcript is clearly up-regulated in LCLs (Fig. 5B). To confirm the presence of this transcript, we performed RT-PCR specific to the isoform using primers unique to MLT1L and downstream exon 2 of *HUWE1* transcript. This analysis confirmed that the cryptic *HUWE1* transcript is expressed in the LCLs but not the B cells (Fig. 5C,D). Altogether, these data suggest that EBV-transformed LCLs specifically activate MLT1L at the *HUWE1/HECTH9* locus to express a cryptic isoform.

Discussion

Our results demonstrate that activation of LTRs by EBV-induced transformation occurs pervasively across the genome from various subfamilies of ERVs. The activation of LTRs involves both DNA hypomethylation and transcription factor binding. The transcriptional activity of activated LTRs is unique to transformed B cells compared with other cell types that are known to have active LTRs. Furthermore, we have shown that EBV-induced transformation of B cells leads to expression of many cryptic transcripts, including several that are relevant to lymphoma biology, as well as other cancers types.

Viruses have previously been implicated in the widespread activation of LTRs and long interspersed nuclear elements (LINEs) in hepatocarcinomas, for which tumors with viral etiology have a higher level of LTR and LINE activation (Shukla et al. 2013; Hashimoto et al. 2015; Schauer et al. 2018). Taken together with our data, this suggests that LTR activation is related to the host's response to viruses. The most obvious pathways responsible for LTR activation are those related to immune response to viruses. ERVs have previously been demonstrated to be evolutionarily co-opted by innate immune pathways (Chuong et al. 2016). Our current analysis of transcription factors bound to LTRs in GM12878 cells does find that NF-kB subunits can bind to ~40% of LTRs, suggesting these LTRs have an important role in regulating immune pathways. It has furthermore been demonstrated that the noncanonical isoform of *CSF1R* is regulated by an NF-κB binding site at the THE1B element (Lamprecht et al. 2010). Further analysis of the dynamics of transcription factor occupancy and signaling at LTRs across EBV infection time points will be necessary to fully elucidate this issue.

The most common transcription factor found at EBV-activated LTRs is RUNX3, which was bound to ~67% of activated LTRs. This transcription factor is highly up-regulated by EBV-encoded transcription factors (Spender et al. 2002; Brady et al. 2009). Previously, RUNX3 has been reported to be involved in LINE-1 transcription (Yang et al. 2003). To our knowledge, this is the first report of RUNX3 involved in LTR activation. A recent report indicated that RUNX3 binding can induce site-specific DNA hypomethylation (Suzuki et al. 2017). Taken together with our finding that RUNX3 binds to >50% of all activated LTRs, this suggests EBV-induced RUNX3 binding is important in the hypomethylation of activated LTRs.

EBV-induced transformation of B cells leads to global loss of DNA methylation, similar to what is observed in cancer cells (Hansen et al. 2014). Our results indicate that LTR loci become hypomethylated upon EBV-induced transformation and that a subset of these hypomethylated LTRs can then be transcriptionally activated. We found that DNA hypomethylation is insufficient to transcriptionally activate LTR loci. We furthermore found that many of these activated LTRs are bound by specific transcription factors, indicating that there may be both the need to derepress as well as activate by transcription factor binding. This derepression/activation model has been proposed previously by others (Lamprecht et al. 2010; Babaian and Mager 2016). This model suggests that while DNA hypomethylation can be achieved in different cells, the LTRs that are activated will be unique based upon the transcription factors that are present within the cell. In line with this model and previous data describing cell-type–specific binding of transcription factors at transposons (Xie et al. 2013), our data show that LTR activation is a cell-type–specific behavior.
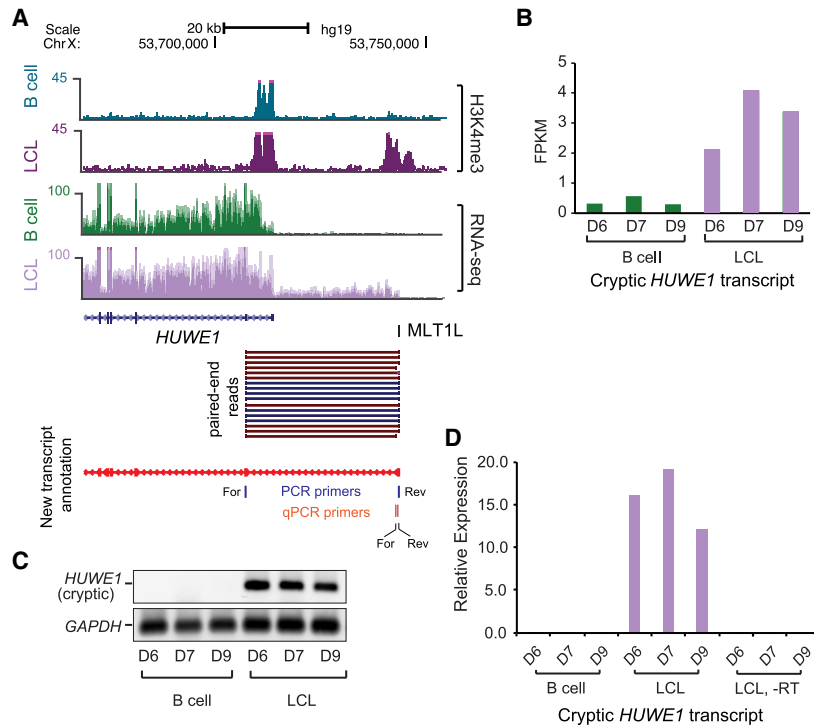
Figure 5. Activated LTRs can drive transcription of cryptic transcripts relevant to cancer. (*A*) Genome browser tracks for H3K4me3 and RNA-seq profile for B cells and LCLs at *HUWE1/HECTH9* locus. Shown in red is the cryptic transcript assembled with the supporting paired-end reads from RNA-seq data set. TSS of alternative cryptic transcript overlaps *MLT1L*. *Below* transcript annotation are the PCR primers used for RT-PCR (blue) and qPCR (orange). (For) Forward primer; (Rev) reverse primer. (*B*) Fragments per kilobase per million reads (FPKM) in each RNA data set. (*C*) RT-PCR analysis of cryptic *HUWE1* transcripts in B cells and matched LCLs. (*D*) RT-qPCR analysis of cryptic *HUWE1* expression in B cells and LCLs. (-RT) no reverse transcriptase.

With the previous observation that a *ERVK* locus is expressed during EBV infection to promote superantigen expression (Sutkowski et al. 1996, 2001), our finding that there is extensive LTR activation by EBV-induced transformation suggests that LTR activation may be an intrinsic process that the host cell has co-opted. The additional finding that these LTRs can lead to activation of *CSF1R* and *IRF5*, two genes important for lymphoma cell survival and proliferation, suggests that activation of LTRs can also have a benefit for the virus in promoting survival of the infected cell. The activation of these two genes important for oncogenesis, however, may lead to negative consequences for the host in the long run.

EBV has long been associated with increased risk of lymphoma (Young et al. 2016). Transcriptomes of lymphoma cells include many LTR-driven cryptic transcripts (Lock et al. 2014; Babaian et al. 2016). Analysis of our transcriptome and chromatin profiles reveals many additional LTR-driven transcripts, including a cryptic isoform of *HUWE1/HECTH9*. This E3 ubiquitin ligase has previously been shown to be important in regulating the transcriptional activity of MYC and, as such, may have an important role in LCL development as well as in the development of lymphomas related to EBV (Adhikary et al. 2005).

Based on our results, it is important to examine whether the subset of lymphomas that express LTR-driven transcripts, including Hodgkin lymphoma and diffuse large B cell lymphoma (DLBCL) (Lamprecht et al. 2010; Lock et al. 2014), have all been infected by EBV. We preliminarily addressed this by examining cell lines that express the noncanonical isoform of *IRF5* (Babaian

et al. 2016). We found that this isoform is expressed in both EBV-positive and EBV-negative B cell lymphoma cell lines tested by Babaian and colleagues (Babaian et al. 2016), indicating it is not unique to EBV transformation. We furthermore examined our RNA-sequencing profiles to determine if the cryptic transcripts identified in DLBCL are also expressed in the LCLs. *CSF1R* and *IRF5* were the only common transcripts, indicating that while LTR activation leading to cryptic transcripts occurs in both DLBCL and LCLs, the subset of cryptic transcripts is distinct in each cell type. This suggests that LTR activation does not solely rely upon viral-induced transformation, and that perhaps in the context of lymphoma development, activation of specific cryptic transcripts can occur through other means (other than viral-induced transformation).

Methods

Detailed methods are provided in the Supplemental Material.

Collection of B cells and growth of LCLs

Blood samples from de-identified healthy donors were obtained following institutional guidelines at the City of Hope (IRB no. 17022). B cell population was isolated using Dynabeads CD19 pan B magnetic beads (Invitrogen). Isolated B cells were collected for chromatin immunoprecipitation, DNA isolation, and RNA isolation. LCLs were derived from B cells isolated from primary B cells (from de-identified donors labeled D6, D7, and D9) using EBV collected from a B95-8 strain-containing marmoset cell line (Hui-Yuen et al. 2011). Briefly, B cells were incubated with EBV-containing supernatant, grown for >6 wk, frozen, and regrown to establish continual proliferating cell lines.

ChIP-seq, RNA-seq

Standard ChIP-seq and RNA-seq methods were used, as we have previously reported (Leung et al. 2014; Du et al. 2016). Details of sample preparation, sequencing, data processing, and downstream analysis are included in the Supplemental Materials including custom code (Supplemental_Code.tar.gz). All data were aligned to hg19 for comparison with existing data (Wang et al. 2013; Hansen et al. 2014). A comparison of LTR content in hg19 and hg38 is included in the Supplemental Material.

Transfections and knockdown of *RUNX3*

siRNA knockdown of *RUNX3* transcripts was performed with 1 million lymphoblastoid cells (D7 LCL) according to the manufacturer's protocol (Amaxa) using 50 nM pool of siRNA targeting *RUNX3* (Thermo Fisher Scientific, catalog no. AM16708; ID 107390 and 115509) or 50 nM of control siRNA (nontargeting control, Thermo Fisher Scientific). Cells were recovered after transfections and grown for 2 d before collection of RNA analysis. Total

RNA was extracted; cDNA was prepared as described above and used for qPCR analysis.

Data access

All new sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession no. GSE106851.

Acknowledgments

References

Adhikary S, Marinoni F, Hock A, Hulleman E, Popov N, Beier R, Bernard S, Quarto M, Capra M, Goettig S, et al. 2005. The ubiquitin ligase HectH9 regulates transcriptional activation by myc and is essential for tumor cell proliferation. *Cell* **123:** 409–421. doi: 10.1016/j.cell.2005.08.016

Allday MJ. 2013. EBV finds a polycomb-mediated, epigenetic solution to the problem of oncogenic stress responses triggered by infection. *Front Genet* **4:** 212. doi: 10.3389/fgene.2013.00212

Babaian A, Mager DL. 2016. Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* **7:** 24. doi: 10.1186/s13100-016-0080-x

Babaian A, Romanish MT, Gagnier L, Kuo LY, Karimi MM, Steidl C, Mager DL. 2016. Onco-exaptation of an endogenous retroviral LTR drives *IRF5* expression in Hodgkin lymphoma. *Oncogene* **35:** 2542–2546. doi: 10.1038/onc.2015.308

Baker AH, Ridge SA, Hoy T, Cachia PG, Culligan D, Baines P, Whittaker JA, Jacobs A, Padua RA. 1993. Expression of the colony-stimulating factor 1 receptor in B lymphocytes. *Oncogene* **8:** 371–378.

Boller S, Grosschedl R. 2014. The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function. *Immunol Rev* **261:** 102–115. doi: 10.1111/imr.12206

Brady G, Whiteman HJ, Spender LC, Farrell PJ. 2009. Downregulation of RUNX1 by RUNX3 requires the RUNX3 VWRPY sequence and is essential for Epstein-Barr virus-driven B-cell proliferation. *J Virol* **83:** 6909–6916. doi: 10.1128/JVI.00216-09

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351:** 1083–1087. doi: 10.1126/science.aad5497

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. 2012. Landscape of transcription in human cells. *Nature* **489:** 101–108. doi: 10.1038/nature11233

Du J, Leung A, Trac C, Lee M, Parks BW, Lusis AJ, Natarajan R, Schones DE. 2016. Chromatin variation associated with liver metabolism is mediated by transposable elements. *Epigenetics Chromatin* **9:** 28. doi: 10.1186/s13072-016-0078-0

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74. doi: 10.1038/nature11247

Elbarbary RA, Lucas BA, Maquat LE. 2016. Retrotransposons as regulators of gene expression. *Science* **351:** aac7247. doi: 10.1126/science.aac7247

Epstein MA, Achong BG, Barr YM. 1964. Virus particles in cultured lymphoblasts from Burkitt's lymphoma. *Lancet* **1:** 702–703. doi: 10.1016/S0140-6736(64)91524-7

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28:** 817–825. doi: 10.1038/nbt.1662

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9:** 215–216. doi: 10.1038/nmeth.1906

Evsikov AV, de Vries WN, Peaston AE, Radford EE, Fancher KS, Chen FH, Blake JA, Bult CJ, Latham KE, Solter D, et al. 2004. Systems biology of the 2-cell mouse embryo. *Cytogenet Genome Res* **105:** 240–250. doi: 10.1159/000078195

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41:** 563–571. doi: 10.1038/ng.368

Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, Bonetti A, Voineagu I, Bertin N, Kratz A, et al. 2014. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* **46:** 558–566. doi: 10.1038/ng.2965

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. doi: 10.1016/j.stem.2015.01.005

Hansen KD, Sabunciyan S, Langmead B, Nagy N, Curley R, Klein G, Klein E, Salamon D, Feinberg AP. 2014. Large-scale hypomethylated blocks associated with Epstein-Barr virus–induced B-cell immortalization. *Genome Res* **24:** 177–184. doi: 10.1101/gr.157743.113

Hashimoto K, Suzuki AM, Dos Santos A, Desterke C, Collino A, Ghisletti S, Braun E, Bonetti A, Fort A, Qin XY, et al. 2015. CAGE profiling of ncRNAs in hepatocellular carcinoma reveals widespread activation of retroviral LTR promoters in virus-induced tumors. *Genome Res* **25:** 1812–1824. doi: 10.1101/gr.191031.115

Hernando H, Shannon-Lowe C, Islam AB, Al-Shahrour F, Rodríguez-Ubreva J, Rodríguez-Cortez VC, Javierre BM, Mangas C, Fernández AF, Parra M, et al. 2013. The B cell transcription program mediates hypomethylation and overexpression of key genes in Epstein-Barr virus-associated proliferative conversion. *Genome Biol* **14:** R3. doi: 10.1186/gb-2013-14-1-r3

Hoffman MM, Buske OJ, Wang J, Weng ZP, Bilmes WS. 2012. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9:** 473–476. doi: 10.1038/nmeth.1937

Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E, et al. 2013. Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* **41:** 827–841. doi: 10.1093/nar/gks1284

Hsu JL, Glaser SL. 2000. Epstein–Barr virus-associated malignancies: epidemiologic patterns and etiologic implications. *Crit Rev Oncol Hematol* **34:** 27–53. doi: 10.1016/S1040-8428(00)00046-9

Hui-Yuen J, McAllister S, Koganti S, Hill E, Bhaduri-McIntosh S. 2011. Establishment of Epstein-Barr virus growth-transformed lymphoblastoid cell lines. *J Vis Exp* **8:** pii: 3321. doi: 10.3791/3321.

Iempridee T, Reusch JA, Riching A, Johannsen EC, Dovat S, Kenney SC, Mertz JE. 2014. Epstein-Barr virus utilizes Ikaros in regulating its latent-lytic switch in B cells. *J Virol* **88:** 4811–4827. doi: 10.1128/JVI.03706-13

Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes *ZNF91/93* and SVA/L1 retrotransposons. *Nature* **516:** 242–245. doi: 10.1038/nature13760

Jacques PÉ, Jeyakani J, Bourque G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* **9:** e1003504. doi: 10.1371/journal.pgen.1003504

Klawitter S, Fuchs NV, Upton KR, Muñoz-Lopez M, Shukla R, Wang JC, Garcia-Cañadas M, Lopez-Ruiz C, Gerhardt DJ, Sebe A, et al. 2016. Reprogramming triggers endogenous L1 and *Alu* retrotransposition in human induced pluripotent stem cells. *Nat Commun* **7:** 10286. doi: 10.1038/ncomms10286

Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, Köchert K, Bouhlel MA, Richter J, Soler E, et al. 2010. Derepression of an endogenous long terminal repeat activates the *CSF1R* proto-oncogene in human lymphoma. *Nat Med* **16:** 571–579. doi: 10.1038/nm.2129

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle V, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921. doi: 10.1038/35057062

Leung DC, Lorincz MC. 2012. Silencing of endogenous retroviruses: When and why do histone marks predominate? *Trends Biochem Sci* **37:** 127–133. doi: 10.1016/j.tibs.2011.11.006

Leung A, Parks BW, Du J, Trac C, Setten R, Chen Y, Brown K, Lusis AJ, Natarajan R, Schones DE. 2014. Open chromatin profiling in mice livers reveals unique chromatin variations induced by high fat diet. *J Biol Chem* **289:** 23557–23567. doi: 10.1074/jbc.M114.581439

Liu S, Brind'Amour J, Karimi MM, Shirane K, Bogutz A, Lefebvre L, Sasaki H, Shinkai Y, Lorincz MC. 2014. *Setdb1* is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev* **28:** 2041–2055. doi: 10.1101/gad.244848.114

Lock FE, Rebollo R, Miceli-Royer K, Gagnier L, Kuah S, Babaian A, Sistiaga-Poveda M, Lai CB, Nemirovsky O, Serrano I, et al. 2014. Distinct isoform

of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. *Proc Natl Acad Sci* **111:** E3534–E3543. doi: 10.1073/pnas.1405507111

Lu X, Sachs F, Ramsay L, Jacques PÉ, Göke J, Bourque G, Ng HH. 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* **21:** 423–425. doi: 10.1038/nsmb.2799

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487:** 57–63. doi: 10.1038/nature11244

Mager DL, Stoye JP. 2015. Mammalian endogenous retroviruses. *Microbiol Spectr* **3:** MDNA3-0009-2014. doi: 10.1128/microbiolspec.MDNA3-0009-2014

Maksakova IA, Mager DL, Reiss D. 2008. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell Mol Life Sci* **65:** 3329–3347. doi: 10.1007/s00018-008-8494-3

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. doi: 10.1186/1756-8935-6-15

Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, Nakamura M, Tokunaga Y, Nakamura M, Watanabe A, et al. 2014. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci* **111:** 12426–12431. doi: 10.1073/pnas.1413299111

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7:** 597–606. doi: 10.1016/j.devcel.2004.09.004

Pope JH, Scott W, Moss DJ. 1973. Human lymphoid cell transformation by Epstein-Barr virus. *Nat New Biol* **246:** 140–141. doi: 10.1038/newbio246140a0

Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Erho N, Ghadessi M. 2013. The long noncoding RNA *SChLAP1* promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet* **45:** 1392–1398. doi: 10.1038/ng.2771

Price AM, Luftig MA. 2014. Dynamic Epstein–Barr virus gene expression on the path to B-cell transformation. *Adv Virus Res* **88:** 279–313. doi: 10.1016/B978-0-12-800098-4.00006-4

Price AM, Dai J, Bazot Q, Patel L, Nikitin PA, Djavadian R, Winter PS, Salinas CA, Barry AP, Wood KC, et al. 2017. Epstein-Barr virus ensures B cell survival by uniquely modulating apoptosis at early and late times after infection. *eLife* **6:** pii: e22509. doi: 10.7554/eLife.22509

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317–330. doi: 10.1038/nature14248

Rowe HM, Trono D. 2011. Dynamic control of endogenous retroviruses during development. *Virology* **411:** 273–287. doi: 10.1016/j.virol.2010.12.007

Schauer SN, Carreira PE, Shukla R, Gerhardt DJ, Gerdes P, Sanchez-Luque FJ, Nicoli P, Kindlova M, Ghisletti S, Santos AD, et al. 2018. L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res* **28:** 639–653. doi: 10.1101/gr.226993.117

Shukla R, Upton KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, et al. 2013. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153:** 101–111. doi: 10.1016/j.cell.2013.02.032

Smit AF, Hubley R, Green P. 2013–2015. RepeatMasker Open 4.0. http://www.repeatmasker.org.

Spender LC, Cornish GH, Sullivan A, Farrell PJ. 2002. Expression of transcription factor AML-2 (RUNX3, CBFα-3) is induced by Epstein-Barr virus EBNA-2 and correlates with the B-cell activation phenotype. *J Virol* **76:** 4919–4927. doi: 10.1128/JVI.76.10.4919-4927.2002

Sutkowski N, Palkama T, Ciurli C, Sekaly RP, Thorley-Lawson DA, Huber BT. 1996. An Epstein-Barr virus–associated superantigen. *J Exp Med* **184:** 971–980. doi: 10.1084/jem.184.3.971

Sutkowski N, Conrad B, Thorley-Lawson DA, Huber BT. 2001. Epstein-Barr virus transactivates the human endogenous retrovirus HERV-K18 that encodes a superantigen. *Immunity* **15:** 579–589. doi: 10.1016/S1074-7613(01)00210-2

Suzuki T, Maeda S, Furuhata E, Shimizu Y, Nishimura H, Kishima M, Suzuki H. 2017. A screening system to identify transcription factors that induce binding site-directed DNA demethylation. *Epigenetics Chromatin* **10:** 60. doi: 10.1186/s13072-017-0169-6

Tempera I, De Leo A, Kossenkov AV, Cesaroni M, Song H, Dawany N, Showe L, Lu F, Wikramasinghe P, Lieberman PM. 2016. Identification of *MEF2B*, *EBF1*, and *IL6R* as direct gene targets of Epstein-Barr virus (EBV) nuclear antigen 1 critical for EBV-infected B-lymphocyte survival. *J Virol* **90:** 345–355. doi: 10.1128/JVI.02318-15

Thompson PJ, Macfarlan TS, Lorincz MC. 2016. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol Cell* **62:** 766–776. doi: 10.1016/j.molcel.2016.03.029

Thorley-Lawson DA. 2001. Epstein-Barr virus: exploiting the immune system. *Nat Rev Immunol* **1:** 75–82. doi: 10.1038/35095584

Wang J, Zhuang JL, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong XJ, Kundaje A, Cheng Y, et al. 2013. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *J Biomol Struct Dyn* **31:** 49–50. doi: 10.1080/07391102.2013.786511

Wolf D, Goff SP. 2008. Host restriction factors blocking retroviral replication. *Annu Rev Genet* **42:** 143–163. doi: 10.1146/annurev.genet.42.110807.091704

Wolf G, Greenberg D, Macfarlan T. 2015. Spotting the enemy within: targeted silencing of foreign DNA in mammalian genomes by the Kruppel-associated box zinc finger protein family. *Mob DNA* **6:** 17. doi: 10.1186/s13100-015-0050-8

Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, et al. 2013. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45:** 836–841. doi: 10.1038/ng.2649

Yang N, Zhang L, Zhang Y, Kazazian HH Jr. 2003. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res* **31:** 4929–4940. doi: 10.1093/nar/gkg663

Young LS, Yap LF, Murray PG. 2016. Epstein–Barr virus: more than 50 years old and still providing surprises. *Nat Rev Cancer* **16:** 789–802. doi: 10.1038/nrc.2016.92

Zhao B, Zou J, Wang HF, Johannsen E, Peng CW, Quackenbush J, Mar JC, Morton CC, Freedman ML, Blacklow SC, et al. 2011. Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth. *Proc Natl Acad Sci* **108:** 14902–14907. doi: 10.1073/pnas.1108892108

Zhao B, Barrera LA, Ersing I, Willox B, Schmidt SCS, Greenfeld H, Zhou H, Mollo SB, Shi TT, Takasaki K, et al. 2014. The NF-κB genomic landscape in lymphoblastoid B cells. *Cell Rep* **8:** 1595–1606. doi: 10.1016/j.celrep.2014.07.037