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

Amy Leung[1,*], Candi Trac[1], Hiroyuki Kato[1], Rama Natarajan[1,2], Dustin E Schones[1,2,*]

[1]Department of Diabetes Complications and Metabolism, Beckman Research Institute, City of Hope, Duarte, CA, USA

[2]Irell & Manella Graduate School of Biological Sciences, City of Hope, Duarte, CA, USA

***Corresponding Authors:**

Amy Leung (aleung@coh.org), Dustin E. Schones (dschones@coh.org)

Amy Leung, Ph.D.
Department of Diabetes Complications and Metabolism
Beckman Research Institute
City of Hope
Gonda South 2015
1500 E. Duarte Rd
Duarte, CA, 91016
Email: aleung@coh.org
Phone: 626-218-2278

Dustin Schones, Ph.D.
Department of Diabetes Complications and Metabolism
Beckman Research Institute
City of Hope
Gonda South 2015
1500 E. Duarte Rd
Duarte, CA, 91016
Email: dschones@gmail.com
Phone: 626-218-9319

Running title: EBV-activated LTRs alter the transcriptome of B cells

Keywords: Retrotransposons, LTRs, H3K4me3, EBV, cryptic promoters

ABSTRACT

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 regulate proximal transcripts when this repression is diminished. It has become clear in recent years that a subset of LTR promoters are activated during pathological conditions, including cancer. However, the events driving LTR activation are not well understood. Epstein-Barr virus (EBV), which transforms primary B cells into continuously proliferating cells, is a tumor virus first associated with lymphomas. We report here that transformation of primary B cells by EBV leads to genome-wide activation of LTR promoters. The activation of LTRs coincides with local DNA hypomethylation and binding by transcription factors such as RUNX3, EBF1, and EBNA2. The set of EBV-activated LTRs is unique to transformed B cells when compared to cell lines known to upregulate LTRs. Furthermore, we found that EBV-induced LTR activation impacts the B cell transcriptome by upregulating >100 coding and non-coding transcripts. These transcripts include isoforms of *CSF1R and IRF5*, which are both important to oncogenesis of Hodgkin lymphoma. These data suggest that the activation of LTRs by EBV may be 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.

.

INTRODUCTION

Endogenous retroviruses (ERVs) make up ~8% of the human genome (Lander et al. 2001). A complete ERV locus contains two long terminal repeats (LTRs) flanking the Gag-, Pol- and Env-coding sequences needed for transposition activity (Mager and Stoye 2015). Active ERVs can transpose to new locations in the genome through RNA intermediates transcribed by RNA polymerase II (Stoye 2012). The transcription of ERVs is driven by LTRs which contain both enhancer and promoter elements (Stoye 2012). These loci are generally marked by DNA methylation and repressive histone modifications to inhibit their activity in the host genome (Wolf and Goff 2008; Rowe and Trono 2011; Friedli et al. 2014; Liu et al. 2014). In the human genome, the vast majority of ERVs are solitary LTRs, and thus incapable of transposition, but still maintain the ability to enhance and promote transcription. With repression mechanisms intact, a subset of LTRs has been co-opted in the genome as “cryptic promoters” of transcription (Maksakova et al. 2008; Rowe and Trono 2011; Leung and Lorincz 2012; Thompson et al. 2016). This LTR-driven transcription occurs in a cell- and tissue-specific manner (Faulkner et al. 2009)

There are many examples in which LTR-driven transcripts, both protein-coding and non-coding, play a role in pluripotent cells. In the early embryo of both humans and mice, specific ERV groups are transcriptionally active and drive expression of many transcripts in the developing cells (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, suggesting that they are 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). 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 (Prensner et al. 2013; Lock et al. 2014; Hashimoto et al. 2015; Babaian 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.

Originally discovered in cells derived from a Burkitt’s lymphoma biopsy, Epstein-Barr virus (EBV) can readily infect primary B lymphocytes (B cells) in vitro and in vivo 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). EBV can infect B cells and epithelial cells (Sixbey et al. 1984; Walling et al. 2001; Pegtel et al. 2004). In memory B cells, EBV can remain quiescent in a latent state using one of four latency programs, all which are characterized by distinct expression of EBV-encoded genes, including transcription factors (Thorley-Lawson 2001). Transformation of primary B cells leads to genome-wide transcriptional changes including a decrease of apoptotic genes and an increase of proliferative genes (Allday 2013; Price and Luftig 2014; Price et al. 2017). EBV-induced transformation of primary B cells also leads to DNA hypomethylation at about two-thirds of the genome (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, HERV-K18, leading to the expression of a superantigen important for T cell response (Sutkowski et al. 1996).

To evaluate the impact of EBV infection of primary B cells on LTRs across the genome, we profiled H3K4me3, a chromatin modification associated with active transcription, in primary B cells and LCLs. We report here that LCLs have widespread activation of LTRs compared to primary B cells. LTR activation is coincident with loss of DNA methylation in LCLs at LTR loci and binding by B cell transcription factors (TFs), non-specific TFs (e.g. MYC), and EBNA2, a transcription factor encoded by the viral genome. A consequence of LTR activation is the expression of a number of chimeric mRNAs and non-coding RNAs originating from LTR promoters. Finally, we found that EBV transformation is sufficient to promote the expression of LTR-driven chimeric transcripts for *CSF1R* and *IRF5*, which are expressed in lymphomas but not in primary B cells. This suggests that LTR activation induced by EBV may support expression of key genes related to cancer development.

RESULTS

H3K4me3 profiling reveals widespread enrichment of the modification at LTRs in LCLs

To explore the potential impact of EBV-transformation on the chromatin environment of B cells, we generated H3K4me3 profiles from three donor primary B cells and compared them to published H3K4me3 profiles from three LCLs (Kasowski et al. 2013), including the well-studied GM12878 cell line (**Fig. 1A,B; Supplemental Fig. S1A**). We identified 26,018 H3K4me3 sites in the primary B cells and 30,319 sites in the LCLs with a union set of 35,245 sites. To identify

differentially enriched H3K4me3 sites genome-wide, we performed DESeq2 analysis on the union set of sites and found 13,031 sites as variable across the two groups (adjusted *P <0.01*), with 5,582 increased and 7,449 decreased (**Fig. 1C**). Upregulated sites were enriched at promoters of genes with the top KEGG term of "Pathways in Cancer" (**Fig. S1B**). Genes found in this KEGG term included many genes known to be upregulated by EBV infection, including *FAS, NFKBIA, and TRAF1* (**Fig. S1C**)(Hernando et al. 2013). These data demonstrate that the

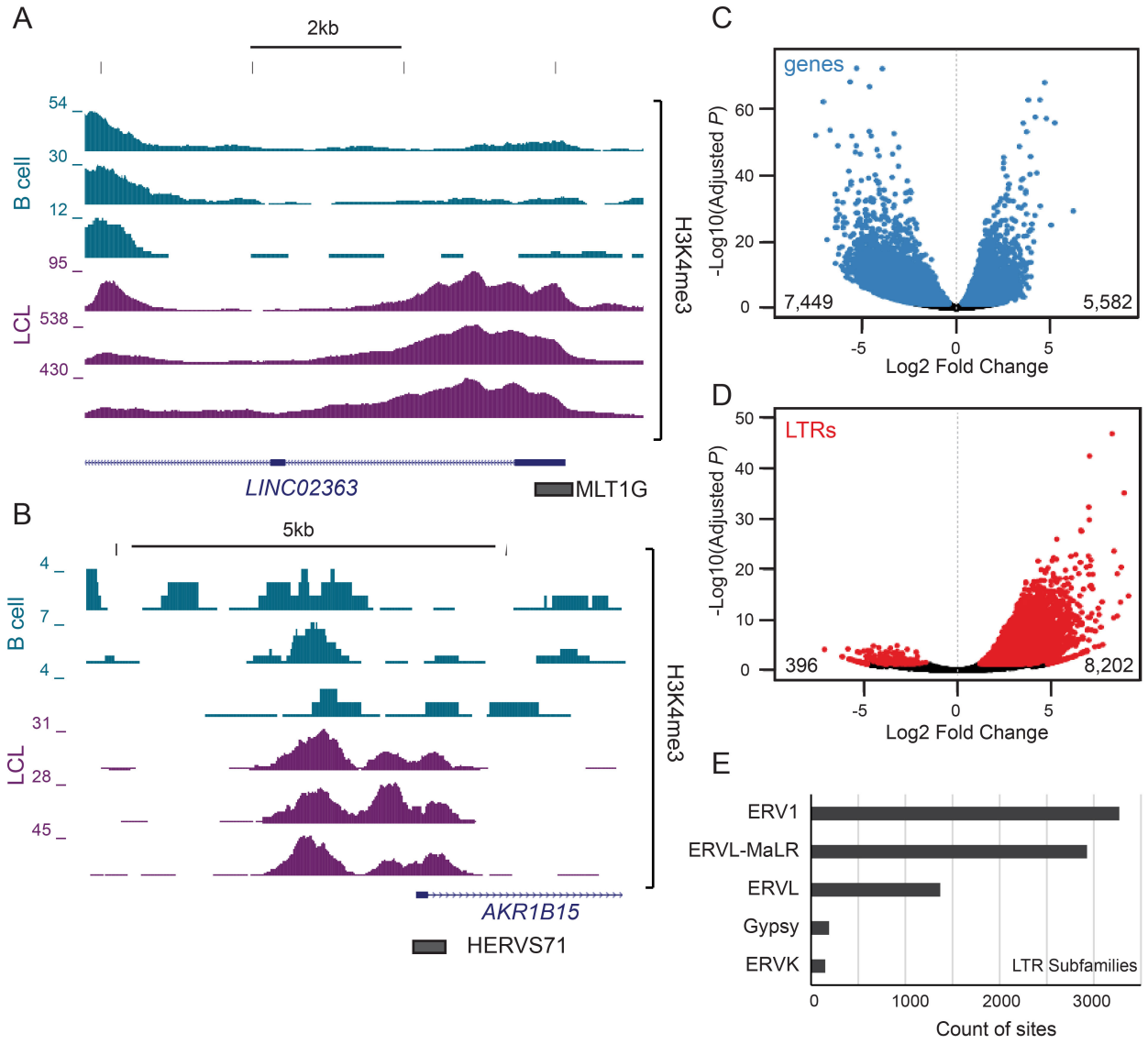

Figure 1: EBV-induced transformation of B cells activates LTRs. Genome browser tracks of H3K4me3 profiles at (*A*) *LINC02363*, a long non-coding gene, and (*B*) *AKR1B15*, a protein-coding gene. Three donor B cells are shown in teal and three LCLs are shown in purple. Volcano plots of normalized H3K4me3 read counts in LCLs relative to B cells at (*C*) transcriptional start sites of genes showing 7,499 gene TSSs with decreased H3K4me3 and 5,582 gene TSSs with increased H3K4me3, and at (*D*) LTR loci showing 396 LTRs with decreased H3K4me3 and 8,202 LTRs with increased H3K4me3.. Blue and red, respectively, are locus with adjusted *P* <0.01. (E) Enumeration of the LTRs subfamilies that show increase in H3K4me3 enrichment in LCLs.

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 the transcriptional activity of LTR promoters across the genome. Taking all LTR sites found in the human genome (RepeatMasker (Smit et al. 2013-2015), we performed DESeq2 analysis with read counts mapping to LTR sites in the B cells and LCLs. Of the 8,596 differentially enriched LTR sites (adjusted $P < 0.01$), the vast majority (8,202) display increase in H3K4me3 enrichment in LCLs compared to B cells (**Fig. 1D**). These LTRs sites can be found at the transcriptional start sites (TSSs) of both long non-coding RNAs (e.g. *LINC02363*, **Fig. 1A**) and protein-coding genes (e.g *AKR1B15*, **Fig. 1B**). In comparison to transcriptional start sites of genes, which show roughly equivalent degrees of increase and decrease in H3K4me3 enrichment in LCLs compared to B cells (**Fig. 1C**), LTRs have a marked increase in H3K4me3 in the LCLs (**Fig. 1D**). These data suggest that widespread upregulation of LTR promoters occurs upon EBV transformation. HERV-K18, an ERV1 element previously reported to be induced by EBV infection, demonstrated increased H3K4me3 in LCLs compared to B cells, as expected (Supplemental **Fig. S2**). Examining the specific subfamilies of LTR promoters that are activated by EBV transformation revealed that the majority of the activated LTRs belong to either the ERV1 or ERVL-MaLR/ERVL subfamilies of LTRs (**Fig. 1E**). These results show that EBV transformation pervasively impacts the transcriptional activity of LTRs across the genome (**Supplemental Table S1, Supplemental Fig S3**).

LTR promoter 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), co-repressors (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 if LTR promoters activated after EBV transformation become hypomethylated. Utilizing whole genome bisulfite sequencing data on primary B cells and LCLs (GSE49627) (Hansen et al. 2014), we examined DNA methylation changes at LTR sites enriched with H3K4me3 in LCLs compared to B cells. Comparing the average CpG methylation level change at LTRs in three donor B cells with the

three LCLs revealed that LTR promoters are generally hypomethylated upon EBV treatment and have corresponding increases in H3K4me3 (**Fig. 2A,** representative example **Fig. 2B**). This data supports the idea that EBV-induced hypomethylation allows LTRs to be activated. We next addressed whether loss of DNA methylation itself is sufficient to activate LTRs. Analysis of DNA methylation at all LTRs, regardless of EBV-activation status, shows that hypomethylation does

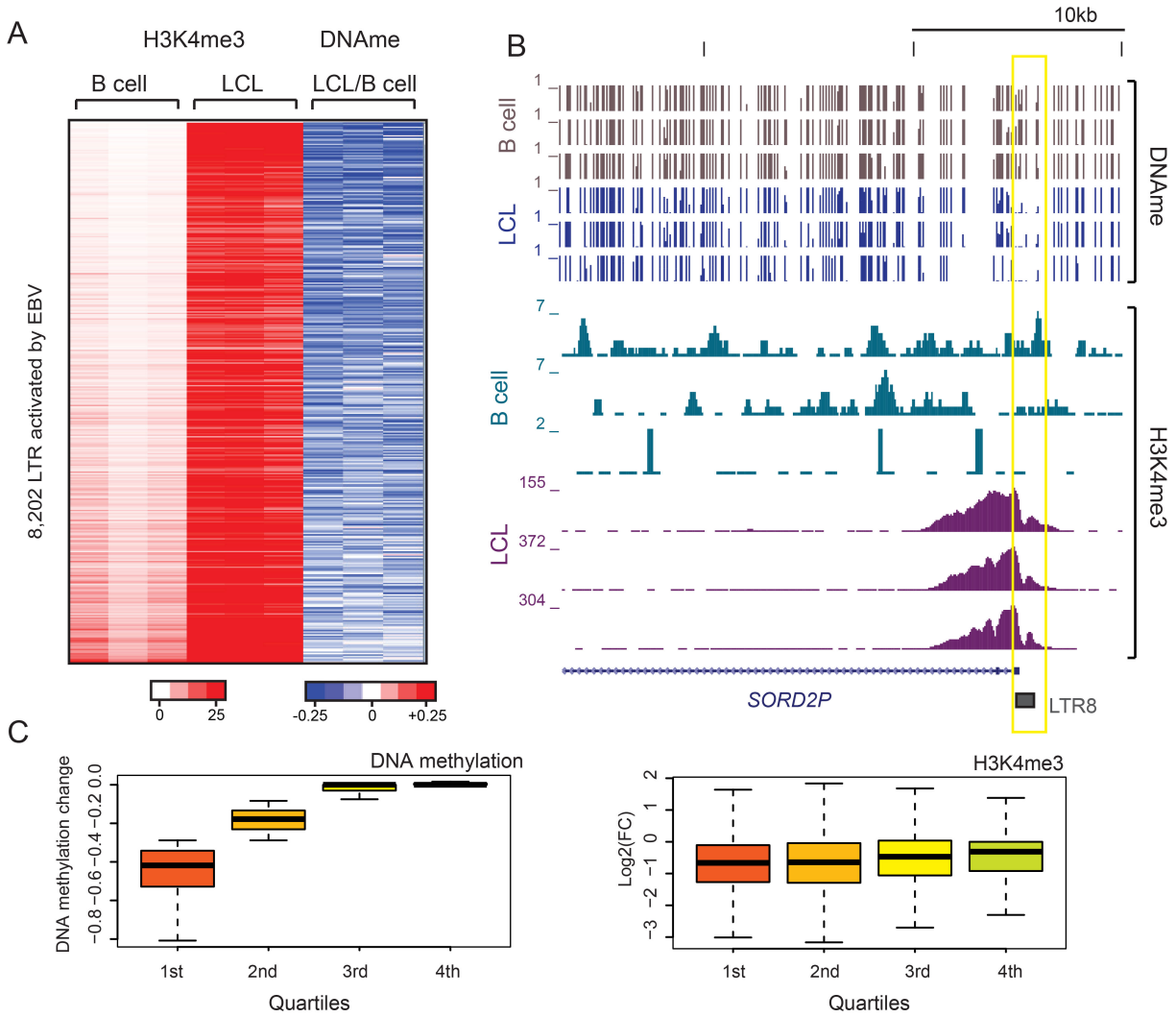

Figure 2: EBV-activated LTRs are hypomethylated in LCLs. (*A*) Heatmap of H3K4me3in three B cell donors and three LCLs at EBV-activated LTRs (red) and heatmap of DNA methylation changes in matched B cells and LCLs at EBV-activated LTRs. (*B*) Genomebrowser tracks of DNA methylation and H3K4me3 profile in B cells and LCLs at *SORD2P* locus driven by an LTR8. (C) (left) Boxplot of DNA methylation changes and (right) H3K4me3change between B cell and LCL stratified into quartiles based on DNA methylation change.

not result in gain of H3K4me3 modifications after EBV transformation (**Fig. 2C**). These data suggest 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 LTR promoters, we next examined transcription factor binding in GM12878 using data generated from ENCODE (Dunham et al. 2012). Overall, we found that there are 24 transcription factors bound to at least 20% of the activated LTR sites (**Fig. 3A**). The transcription factor bound to the greatest number of LTRs activated by EBV is RUNX3 (~50% EBV-activated LTRs), a transcription factor expressed throughout the hematopoietic system. Interestingly, RUNX3 is specifically activated by EBV during LCL development and promotes cell proliferation (Spender et al. 2002; Brady et al. 2009). EBF1, which is important for B cell lineage development (Boller and Grosschedl 2014), is also among the most enriched at

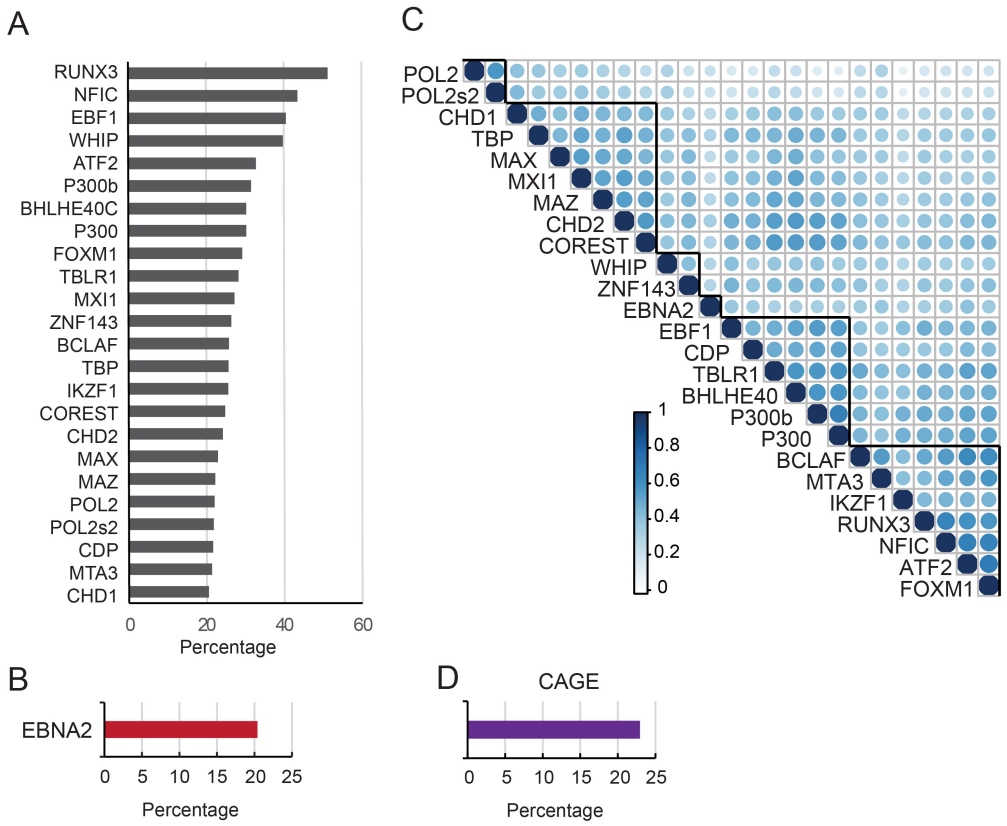

Figure 3: EBV-activated LTRs are binding sites for many transcription factors in GM12878 and can promote transcription.(*A-B*) Bargraphs of the percentage of EBV-activated LTRs bound by each transcription factor. (*C*) Correlation plots (Pearson) of transcription factor co-occupancy at EBV-activated LTRs. Hierarchical clustering was performed to identify transcription factors that co-occur frequently. (*D*) Percentage of EBV-activated LTRs with CAGE sequences.

activated LTRs. Another transcription factor enriched at LTR promoters is IKAROS/IKZF1transcription factor, which binds to ~25% of LTR promoters. This transcription factor plays key roles in the latent-lytic switch of EBV in B cells, indicating that perhaps LTR

promoter activation is influenced by EBV-induced pathways (Iempridee et al. 2014). Though not the most enriched, MAX and MAZ, which interact with c-MYC, are also found at activated LTRs.

We furthermore examined the binding of EBNA2, a transcription factor encoded by the EBV genome that can regulate host cell genes (Kempkes and Ling 2015). Using EBNA2 ChIP-seq data generated from LCLs (GSE29498) (Zhao et al. 2011), we found that ~20% of the EBV-activated LTR promoters are bound by EBNA2 (**Fig. 3B**), indicating these are targeted specifically by this viral-encoded transcription factor. EBNA2 has also been found to bind to super-enhancers during LCL development (Zhou et al. 2015). Intriguingly, we found that ~60% of the super-enhancers in LCLs have an EBV-activated LTR (63 of 187 sites). To examine the co-occupancy of transcription factors bound at activated LTRs we utilized hierarchical clustering of their binding patterns. This analysis revealed five clusters that appear to co-occupy LTRs together (**Fig. 3C**). Interestingly, EBNA2 does not belong to any of these five clusters, suggesting that EBNA2 does not specifically interact with any of the tested transcription factors to activate LTRs. We next examined the number of EBV-activated LTRs that give rise to transcripts using CAGE (Cap-analysis of gene expression) data from GM12878 (ENCODE)(Hoffman et al. 2013). In total, ~23% of EBV-activated LTRs overlap CAGE enriched sites, indicating these LTRs are initiating transcription (**Fig. 3D**).

LTRs activated by EBV are mostly unique to LCLs

LTRs have previously been shown to be active in other cells lines, including H1 (embryonic stem cell line) and K562 (immortalized myelogenous leukemia line) cells (Djebali et al. 2012; Fort et al. 2014). We next addressed whether the LTRs activated by EBV are unique to LCLs. 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 cells lines at the LTRs activated by EBV. Using hierarchical clustering of the H3K4me3 profiles from these cells, we found that the majority of the EBV-activated LTRs are not enriched for H3K4me3 in other cells (**Fig. 4A**). These data suggest that the activation of LTRs by EBV 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. S4**). This data coincides with previous reports that LTR expression and activity is cell type-specific (Faulkner et al. 2009; Jacques et al. 2013; Xie et al. 2013).

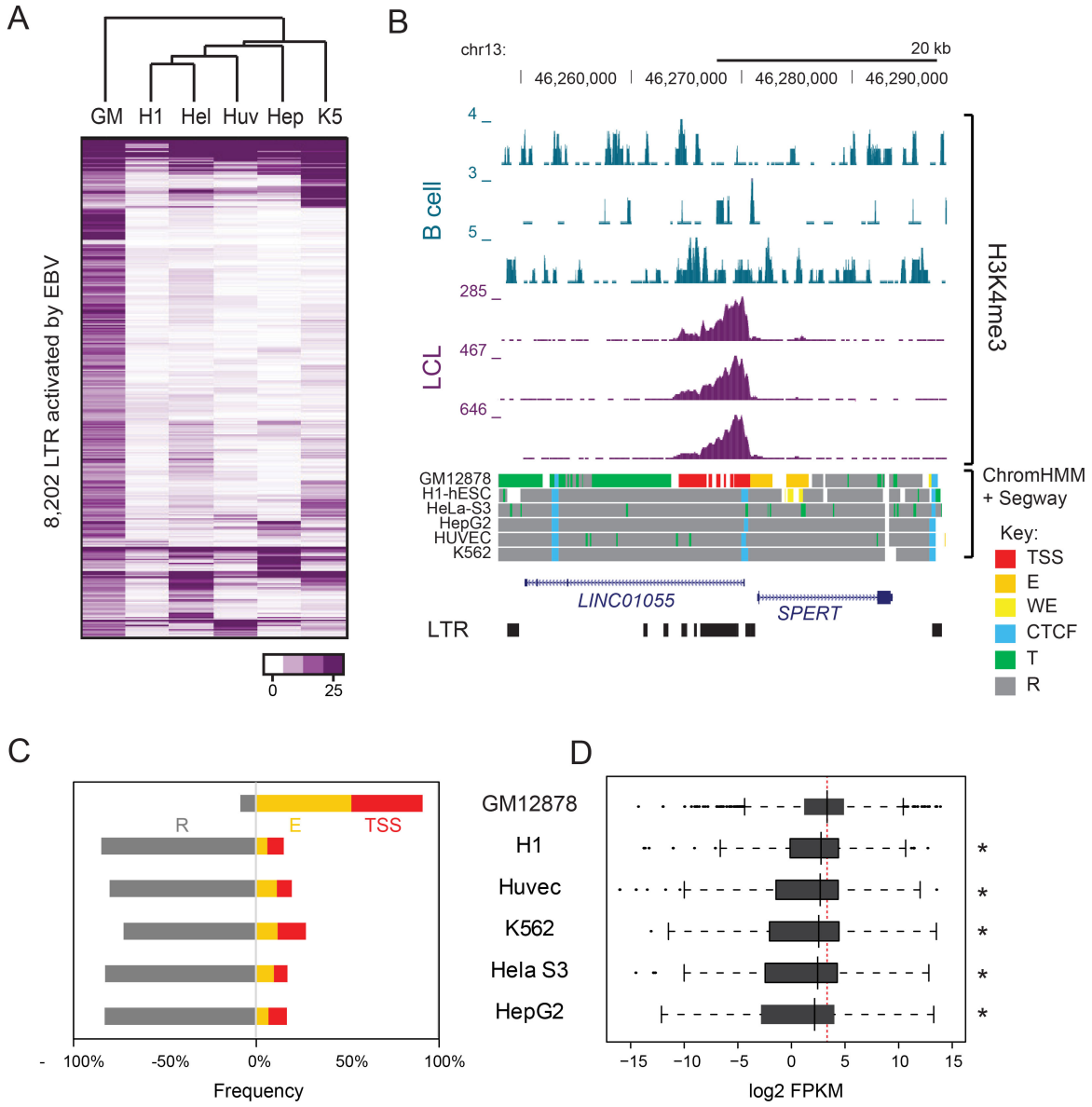

Figure 4: Transcriptional activity of EBV-activated LTRs is mostly specific to transformed B cells. (*A*) Hierarchical clustering H3K4me3 read counts across cell lines. (GM, GM12878 (LCL); H1, H1 embryonic stem cells; Hel, Hela-S3; Huv, Huvecs; Hep, HepG2; K5, K562).(*B*) Genome browser track of H3K4me3 profile in B cells and LCLs,and of ChromHmm+Segway combined data at the *LINC01066-SPERT* locus. (Key: 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. (*C*) ChromHmm and Segway combined classification of all EBV- activated LTRs shown as a frequency of total for each ENCODE cell line. Color scheme as in (B). (*D*) Boxplot of read counts for genes adjacent to EBV-activated LTR for ENCODE cell lines. (*, *P* val <0.01, compared to GM12878)

We furthermore examined whether these EBV-activated LTRs in GM12878 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; Ernst and Kellis 2012; Hoffman et al. 2012; Hoffman et al. 2013) as a transcriptional start site (TSS). Examining the functional classification of activated LTRs across the genome (Methods), we found that the majority of the EBV-activated LTRs are classified as promoters or enhancers (>85%) In contrast, these same regions are classified as repressed regions in other cell lines (**Fig. 4C**). Since these EBV-activated LTRs are classified as promoters and enhancers, we reasoned it is likely that the genes surrounding these are upregulated in LCLs. To test this, we obtained RNA-sequencing data from ENCODE (Dunham et al. 2012) and assessed the FPKMs of genes adjacent to EBV-activated LTR. As expected, this set of adjacent genes have a higher FPKM in GM12878 than in the other five cell lines indicating that these LTRs impact adjacent gene expression (**Fig. 4D**).

EBV-induced activation of LTRs is sufficient to drive lymphoma-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 on gene regulatory networks, we examined whether EBV-activated LTRs can lead to expression of annotated transcripts. We found that 103 of the EBV-activated LTRs are annotated as TSSs for protein-coding and long non-coding RNAs (lncRNAs) in LCLs (**Supplemental Table S1**). We found that EBV-activated LTRs can serve as the sole TSS for the gene, an alternative upstream TSS, or an alternative downstream TSS for protein-coding mRNAs (**Fig. 5A**). In contrast, for lncRNAs, EBV-activated LTRs are only found as an alternative upstream TSS or the sole TSS. To assess the expression differences induced by EBV activated LTRs, we examined the expression of these 102 transcripts using GTEx data (Carithers and Moore 2015; Guan et al. 2016), for EBV-induced B cells and blood. We found these transcripts strongly trend towards increased expression in the LCLs (**Fig. 5B**). Furthermore, expression of these TE-driven transcripts is highest in EBV-transformed B cells compared to all other tissues examined by GTEx data (**Supplemental Fig. S5**). Notably, the tissue with the next highest expression of these TE-driven transcripts is testis, which is known to have elevated expression of TE-driven transcripts compared to other tissues (Faulkner et al. 2009). Within our dataset, we found two of the EBV-activated LTRs are relevant to lymphomas. A THE1B LTR drives the expression of a non-canonical isoform of *CSF1R*, leading to lineage

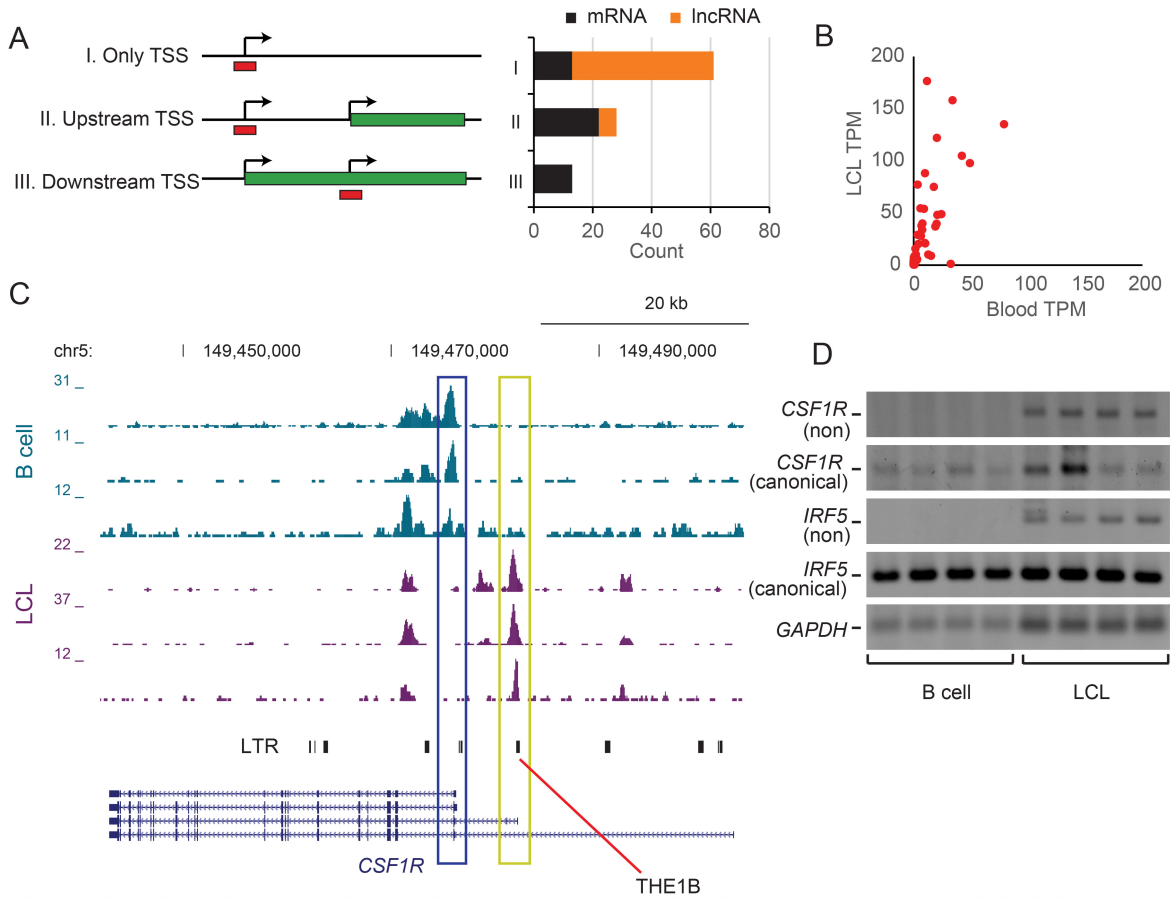

Figure 5: EBV-activated LTRs can drive transcription of mRNAs and lncRNAs. (*A*) (*left*) Schematic of relative position of LTR from transcriptional start site (TSS) of annotated gene. (red, LTR; green, annotated gene; arrow, TSS). (*right*) Count of mRNA and lncRNAs withLTR positioned as in (A). (*B*) Average transcripts per million (TPM) for each mRNA and lncRNAin (A) for blood and LCL analyzed byGTEx. (*C*) Genome browser tracks for H3K4me3 profile for B cells and LCLs at *CSF1R* locus. (Blue box; TSS with decreased H3K4me3 enrichment in LCL, yellow box; TSS with increased H3K4me3 enrichment in LCL.) (*D*) RT-PCR analysis of canonical and noncanonical (non) transcripts from *CSF1R* and *IRF5* locus and transcripts from *GAPDH* as control in B cell and LCLs.

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). At the *CSF1R* locus, we found that the H3K4me3 enrichment appears at the non-canonical promoter (THE1B site) in LCLs (**Fig. 5C**). By RT-QPCR analysis of RNA isolated from B cells and LCLs, we found that the non-canonical transcript is indeed expressed in LCLs, but not in B cells (**Fig. 5D**). The canonical *CSF1R* transcript, however, is expressed in both LCLs and primary B cells, which has previously been observed (Baker et al. 1993). Similarly, in addition to the canonical transcript, LCLs express the non-canonical *IRF5* isoform driven by the LOR1a LTR (**Fig. 5D**). 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.

DISCUSSION

It is well established that transposable elements can alter gene regulatory networks in both normal and diseased states (Babaian and Mager 2016; Elbarbary et al. 2016). Previously, it has been shown that EBV infection of B cells drives expression of a HERV-K locus found within the intron of CD48 (Sutkowski et al. 1996; Sutkowski et al. 2001). This leads to expression of a superantigen proposed to be important for immune response to the mounting infection and perhaps clearance of the viral load (Sutkowski et al. 2001). Here, we report that activation of LTRs by EBV occurs pervasively across the genome from varying subfamilies of ERVs. The activation of LTRs involves both DNA hypomethylation and transcription factor binding. The transcriptional activity of EBV-activated LTRs is unique to transformed B cells compared to 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 chimeric transcripts including two that are relevant to lymphoma biology.

Viruses have previously been implicated in the widespread activation of LTR in hepatocarcinomas, as tumors with viral etiology display a higher level of LTR activation (Hashimoto et al. 2015). Taken together, with our data, this suggests that LTR activation is related to the host's response to viruses. The most obvious pathway responsible for LTR activation are those related to immune response to viruses. Interestingly, ERVs have been evolutionarily co-opted by innate immune pathways (Chuong et al. 2016). Our current analysis of transcription factors bound to LTRs in GM12878 does not find any immune-related transcription factors that are bound to EBV-activated LTRs at high levels (>20%). However, it is possible that binding of immune-related transcription factors such as NF-κB to LTRs occur early during EBV infection/transformation resulting in an open chromatin state that allows additional factors to bind in the transformed cell. It has previously been demonstrated that the non-canonical isoform of *CSF1R* is regulated by an NF-κB binding site at the THE1B element (Lamprecht et al. 2010). Future 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.

While we did not see immune factors bound to LTRs, the most common transcription factor found at EBV-activated LTRs is RUNX3. This transcription factor is highly upregulated by EBV-encoded transcription factors (Spender et al. 2002; Brady et al. 2009). Previously, RUNX3 has been reported to be involved in L1 (Long interspersed nuclear element -1) transcription (Yang et

al. 2003). To our knowledge, this is the first report of RUNX3 involved in LTR activation. There has also been recent report that RUNX3 can induced site-specific DNA hypomethylation (Suzuki et al. 2017). Along with our finding that RUNX3 binds to >50% of all EBV-activated LTRs, EBV-induced RUNX3 activity may be an important factor in the hypomethylation of EBV-activated LTRs.

Based on our results, it is important to examine whether the subset of lymphomas, including Hodgkins lymphoma and diffuse large B-cell lymphoma (Lamprecht et al. 2010; Lock et al. 2014), that express LTR-driven transcripts, have been infected by EBV. We have preliminarily addressed this by examining the cell lines that express the non-canonical isoform of *IRF5* (Babaian et al. 2016) and found that it is expressed in both EBV-positive and EBV-negative B cell lymphoma cell lines. This suggest that LTR activation does not solely rely upon viral activation, and that perhaps in the context of lymphoma development, activation of specific TE-driven transcripts can occur through other means (other than viral activation).

Global loss of DNA methylation is a hallmark of cancer cells. EBV-induced transformation is one exogenous stimulus that can lead to global loss of DNA methylation (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. Interestingly, 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.

With the previous observation that a HERV-K locus is expressed to promote superantigen expression (Sutkowski et al. 1996; Sutkowski et al. 2001), our finding that there is extensive LTR activation suggests that LTR activation by EBV may be an intrinsic process that the host cell has co-opted to benefit itself. Interestingly, 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 EBV activation of LTRs can also have a benefit for the virus in promoting survival of the infected cell. The activation of these two genes, however, may have, in the long run, negative consequences for the host in the form of the oncogenesis. EBV has long been associated with increased risk of lymphoma (Young et al. 2016). Interestingly, analysis of lymphoma cell transcriptomes revealed many TE-driven chimeric transcripts (Lock et al. 2014; Babaian et al. 2016). However, transformed B cells only express a subset of those described in lymphoma, indicating that perhaps additional events that promote LTR activation occur in the development of lymphoma.

The cell-type specificity of LTR activation may have important consequences. In transformed B cells, > 100 chimeric transcripts are activated, and these have the potential to impact cellular behavior. Further analysis is needed to understand the potential function of these EBV-activated TE-driven transcripts. The location and arrangement of LTRs can govern the perhaps unique set of chimeric transcripts that can be expressed in any given cell type. These LTR-driven transcripts may be uniquely expressed in a particular disease state and therefore used as biomarkers of disease or unique therapeutic targets.

METHODS

Collection of B cells and growth of lymphoblastoid cell lines

Blood samples from de-identified healthy donors were obtained following institutional guidelines at the City of Hope (IRB# 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.

Lymphoblastoid cell lines (GM12878, GM18505, GM19099) were obtained from Coriell Biorepository. LCLs were grown according to recommended procedures from the biorepository. LCLs, kept at low passages, were collected and used from RNA isolation. Additional LCLs were derived from B cells isolated from donor B cells using EBV collected from B95-9 strain containing marmoset cell line (Hui-Yuen et al. 2011). Briefly, B cells were incubated with EBV-containing supernatant, grown for >6 weeks, frozen, and regrown to establish continual proliferating cell lines.

Gene expression

RNA was isolated from donor B cells and LCLs using Trizol (Life Technologies). cDNA was generated from isolated RNAs using SuperScript IV Reverse Transcriptase Kit (Invitrogen) with random primers. PCR was performed for *GAPDH,* and isoforms of *CSF1R* and *IRF5*. Primers are available in Supplemental Table 2.

Chromatin immunoprecipitation (ChIP) sequencing

In B cells from donors, ChIPs were carried out with anti-H3K4me3 antibody (Abcam, ab8580) using standard ChIP protocols. Sequencing libraries were generated using the Kapa Hyper Prep kit (Kapa Biosystems) and sequenced using Illumina HiSeq 2500 technology. We obtained on average ~42M 100bp x 100bp paired-end reads per library (H3K4me3 and input, Supplemental Table S3). Sequences were aligned to the hg19 reference genome using Bowtie (Langmead 2010) with default options, except only uniquely mapped reads were retained (-m 1 option). BEDtools (Quinlan and Hall 2010; Quinlan 2014) was used to generate wiggle tracks which were visualized in the UCSC Genome Browser. Peaks of H3K4me3 were analyzed using MACS (Zhang et al. 2008). For LCL analysis, H3K4me3 and input FASTQ sequencing data for GM12878, GM19240, and GM2255 were downloaded from Gene Expression Omnibus (GEO) database (GSE50893) (Kasowski et al. 2013) and analyzed as with the B cell data. To identify H3K4me3 differentially enriched at genes, counts at peaks overlapping RefSeq transcriptional start sites were analyzed using DESeq2 comparing B cells to LCLs (Love et al. 2014). To identify H3K4me3 differentially enriched LTRs, all LTR loci (RepeatMasker Open-4.0) (Smit et al. 2013-2015) were obtained and counts of H3K4me3 reads were analyzed using DESeq2. For gene ontology analysis, differentially enriched H3K4me3 peaks at transcriptional start sites (RefSeq annotation, hg19) were analyzed. Genes with the transcriptional start sites were used as the list using DAVID Functional annotation Tool (Huang et al. 2009a; Huang et al. 2009b).

For EBNA2 binding sites, we downloaded ChIP-sequencing data for EBNA2 from GEO (GSE29498) and aligned the fastq data using Bowtie as with H3K4me3 data above. Peaks of enrichment were determined using MACS with default parameters.

Whole-genome bisulfite sequencing (WGBS) and bisulfite Sanger sequencing

To assess the changes in DNA methylation genome-wide, we obtained WGBS data of B cells and donor-matched LCLs from GEO database (GSE49629) (Hansen et al. 2014). BS-Seeker2 (Guo et al. 2013) was used to obtain DNA methylation levels genome-wide. To obtain changes in DNA methylation, individual CpG methylation levels overlapping LTRs were obtained and averaged across the LTR locus and the values were compared B cells to matched LCLs.

ENCODE transcription factor, chromatin, and gene expression analysis

For analysis of transcription factor binding at LTRs, ENCODE data for transcription factor ChIP-sequencing from GM12878 (Dunham et al. 2012) were downloaded from UCSC Genome Browser. We surveyed ~100 transcription factors. To determine occupancy of transcription factors at EBV-activated LTRs, presence or absence of transcription factor binding site from ENCODE at EBV-activated LTR was determine using BEDtools. For co-occupancy analysis, Jaccard Index was determined and hierarchical clustering by correlation was performed.

To compare H3K4me3 enrichment at EBV-activated LTRs in GM12878 and other cell lines (H1, Hela S3, HepG2, Huvec, and K562), H3K4me3 ChIP-seq fastq data was downloaded from UCSC Genome Browser. Reads were aligned to hg19 genome as described above using Bowtie. Number of reads (counts) was determined for each cell line and hierarchical clustering was performed across each LTR and across the cell lines. For the genome segmentation analysis, Combined Segway and ChromHMM data (Ernst and Kellis 2010; Ernst and Kellis 2012; Hoffman et al. 2012; Hoffman et al. 2013) were downloaded from UCSC Genome Browser. For each EBV-activated LTR, overlapping chromatin state was determined. For those with two or more overlapping chromatin states, the state with the greatest portion of the LTR was selected for frequency analysis.

To determine the impact of EBV-activated LTRs on neighboring genes in GM12878 and other ENCODE cell lines, we downloaded ENCODE gene expression data from UCSC Genome Browser (http://genome.ucsc.edu/encode/downloads.html). The immediate adjacent gene from EBV-activated LTR was determined and FPKM of that gene was analyzed across all cell lines.

DATA ACCESS

All new sequencing data generated from this study is deposited to NCBI GEO repository (GSE106851).

ACKNOWLEDGMENTS

This work was supported by K01DK104993 (AL); R01HL106089, R01HL087864 and RO1DK065073 (RN); R01DK112041, R01CA220693 and R56DK071122 (DES). Research reported in this publication included work performed in the Integrative Genomics Core supported by the National Cancer Institute of the National Institutes of Health under award number P30CA33572. We thank members of the Schones Laboratory and Javier Gordon Ogembo for their comments and suggestions. We also thank Sandrine Lacoste and members of the O'Connor laboratory for providing us with the EBV-containing cells and for their technical advice.

DISCLOSURE DECLARATION

Authors declare there are no conflicts of interest.

REFERENCES

Allday MJ. 2013. EBV finds a polycomb-mediated, epigenetic solution to the problem of oncogenic stress responses triggered by infection. *Frontiers in genetics* **4**: 212.

Babaian A, Mager DL. 2016. Endogenous retroviral promoter exaptation in human cancer. *Mobile DNA-Uk* **7**.

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**(19): 2542-2546.

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**(2): 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**(1): 102-115.

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**(13): 6909-6916.

Carithers LJ, Moore HM. 2015. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank* **13**(5): 307-308.

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**(6277): 1083-1087.

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**(7414): 101-108.

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**(7414): 57-74.

Elbarbary RA, Lucas BA, Maquat LE. 2016. Retrotransposons as regulators of gene expression. *Science* **351**(6274).

Epstein MA, Achong BG, Barr YM. 1964. Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. *Lancet* **1**(7335): 702-703.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**(8): 817-U894.

-. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**(3): 215-216.

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**(2-4): 240-250.

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. *Nature genetics* **41**(5): 563-571.

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. *Nature genetics* **46**(6): 558-566.

Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Diaz N, Rowe HM, Ecco G, Unzu C, Planet E, Lombardo A. 2014. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res* **24**.

Göke J, Lu X, Chan Y-S, Ng H-H, Ly L-H, 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**.

Guan P, Rao A, Volpi S, Koester S, Moore HM, Consortium G. 2016. Normal tissue and data resources for cancer research from the GTEx program. *Cancer Res* **76**.

Guo WL, Fiziev P, Yan WH, Cokus S, Sun XG, Zhang MQ, Chen PY, Pellegrini M. 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *Bmc Genomics* **14**.

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**(2): 177-184.

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**(12): 1812-1824.

Hernando H, Shannon-Lowe C, Islam AB, Al-Shahrour F, Rodriguez-Ubreva J, Rodriguez-Cortez VC, Javierre BM, Mangas C, Fernandez 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**(1).

Hoffman MM, Buske OJ, Wang J, Weng ZP, Bilmes JA, Noble WS. 2012. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9**(5): 473-U488.

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**(2): 827-841.

Hsu JL, Glaser SL. 2000. Epstein-Barr virus-associated malignancies: epidemiologic patterns and etiologic implications. *Crit Rev Oncol Hemat* **34**(1): 27-53.

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**(1): 1-13.

-. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**(1): 44-57.

Hui-Yuen J, McAllister S, Koganti S, Hill E, Bhaduri-McIntosh S. 2011. Establishment of Epstein-Barr Virus Growth-transformed Lymphoblastoid Cell Lines. *Jove-J Vis Exp*(57).

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**(9): 4811-4827.

Jacobs FMJ, 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**.

Jacques P-E, Jeyakani J, Bourque G. 2013. The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements. *PLoS Genet* **9**.

Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, Boyle AP, Zhang QC, Zakharia F, Spacek DV et al. 2013. Extensive variation in chromatin states across humans. *Science* **342**(6159): 750-752.

Kempkes B, Ling PD. 2015. EBNA2 and Its Coactivator EBNA-LP. *Curr Top Microbiol* **391**: 35-59.

Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, Kochert 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**(5): 571-579.

Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Langmead B. 2010. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics* **Chapter 11**: Unit 11 17.

Leung DC, Lorincz MC. 2012. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem Sci* **37**.

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**.

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. *Proceedings of the National Academy of Sciences of the United States of America* **111**(34): E3534-3543.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**(12): 550.

Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH. 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nature structural & molecular biology* **21**(4): 423-425.

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**(7405): 57-+.

Mager DL, Stoye JP. 2015. Mammalian Endogenous Retroviruses. *Microbiology spectrum* **3**(1): MDNA3-0009-2014.

Maksakova IA, Mager DL, Reiss D. 2008. Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cellular and molecular life sciences : CMLS* **65**(21): 3329-3347.

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. *Epigenet Chromatin* **6**.

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. *Proceedings of the National Academy of Sciences of the United States of America* **111**(34): 12426-12431.

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**(4): 597-606.

Pegtel DM, Middeldorp J, Thorley-Lawson DA. 2004. Epstein-Barr virus infection in ex vivo tonsil epithelial cell cultures of asymptomatic carriers. *J Virol* **78**(22): 12613-12624.

Pope JH, Scott W, Moss DJ. 1973. Human lymphoid cell transformation by Epstein-Barr virus. *Nature: New biology* **246**(153): 140-141.

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. *Nature genetics* **45**.

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**.

Price AM, Luftig MA. 2014. Dynamic Epstein-Barr Virus Gene Expression on the Path to B-Cell Transformation. *Adv Virus Res* **88**: 279-313.

Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current protocols in bioinformatics* **47**: 11 12 11-34.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.

Rowe HM, Trono D. 2011. Dynamic control of endogenous retroviruses during development. *Virology* **411**(2): 273-287.

Sixbey JW, Nedrud JG, Raabtraub N, Hanes RA, Pagano JS. 1984. Epstein-Barr Virus-Replication in Oropharyngeal Epithelial-Cells. *New Engl J Med* **310**(19): 1225-1230.

Smit AF, Hubley R, Green P. 2013-2015. RepeatMasker Open-4.0.

Spender LC, Cornish GH, Sullivan A, Farrell PJ. 2002. Expression of transcription factor AML-2 (RUNX3, CBF alpha-3) is induced by Epstein-Barr virus EBNA-2 and correlates with the B-cell activation phenotype. *J Virol* **76**(10): 4919-4927.

Stoye JP. 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat Rev Microbiol* **10**(6): 395-406.

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**(4): 579-589.

Sutkowski N, Palkama T, Ciurli C, Sekaly RP, Thorley-Lawson DA, Huber BT. 1996. An Epstein-Barr virus-associated superantigen. *The Journal of experimental medicine* **184**(3): 971-980.

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**(1): 60.

Thompson PJ, Macfarlan TS, Lorincz MC. 2016. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Mol Cell* **62**(5): 766-776.

Thorley-Lawson DA. 2001. Epstein-Barr virus: exploiting the immune system. *Nat Rev Immunol* **1**(1): 75-82.

Walling DM, Flaitz CM, Nichols CM, Hudnall SD, Adler-Storthz K. 2001. Persistent productive Epstein-Barr virus replication in normal epithelial cells in vivo. *J Infect Dis* **184**(12): 1499-1507.

Wolf D, Goff SP. 2008. Host Restriction Factors Blocking Retroviral Replication. *Annu Rev Genet* **42**.

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**.

Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL. 2013. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nature genetics* **45**.

Yang N, Zhang L, Zhang Y, Kazazian HH. 2003. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res* **31**(16): 4929-4940.

Young LS, Yap LF, Murray PG. 2016. Epstein-Barr virus: more than 50 years old and still providing surprises. *Nat Rev Cancer* **16**(12): 789-802.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.

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. *Proceedings of the National Academy of Sciences of the United States of America* **108**(36): 14902-14907.

Zhou HF, Schmidt SCS, Jiang SZ, Willox B, Bernhardt K, Liang J, Johannsen EC, Kharchenko P, Gewurz BE, Kieff E et al. 2015. Epstein-Barr Virus Oncoprotein Super-enhancers Control B Cell Growth. *Cell Host Microbe* **17**(2): 205-216.