The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome

Yorifumi Satou[a,b,c,d,e,1], Paola Miyazato[b,c,d], Ko Ishihara[c,e], Hiroko Yaguchi[a], Anat Melamed[a], Michi Miura[a], Asami Fukuda[b,c,d], Kisato Nosaka[f], Takehisa Watanabe[g,2], Aileen G. Rowan[a], Mitsuyoshi Nakao[e,g], and Charles R. M. Bangham[a,1]

[a]Section of Virology, Division of Infectious Diseases, Imperial College, London W2 1PG, United Kingdom; [b]Centre for AIDS Research, Kumamoto University, Kumamoto 860-0811, Japan; [c]Priority Organization for Innovation and Excellence, Kumamoto University, Kumamoto 860-0811, Japan; [d]International Research Centre for Medical Science, Kumamoto University, Kumamoto 860-0811, Japan; [e]Core Research for Evolutionary Science and Technology, Japan Science and Technology Agency, Saitama 332-0012, Japan; [f]Cancer Centre, Kumamoto University Hospital, Kumamoto University, Kumamoto 860-0811, Japan; and [g]Department of Medical Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto 860-0811, Japan

Edited by Robert C. Gallo, Institute of Human Virology, University of Maryland School of Medicine, Baltimore, MD, and approved February 1, 2016 (received for review December 4, 2014)

Human T-lymphotropic virus type 1 (HTLV-1) is a retrovirus that causes malignant and inflammatory diseases in ∼10% of infected people. A typical host has between 10^4 and 10^5 clones of HTLV-1–infected T lymphocytes, each clone distinguished by the genomic integration site of the single-copy HTLV-1 provirus. The HTLV-1 bZIP (*HBZ*) factor gene is constitutively expressed from the minus strand of the provirus, whereas plus-strand expression, required for viral propagation to uninfected cells, is suppressed or intermittent in vivo, allowing escape from host immune surveillance. It remains unknown what regulates this pattern of proviral transcription and latency. Here, we show that CTCF, a key regulator of chromatin structure and function, binds to the provirus at a sharp border in epigenetic modifications in the pX region of the HTLV-1 provirus in T cells naturally infected with HTLV-1. CTCF is a zinc-finger protein that binds to an insulator region in genomic DNA and plays a fundamental role in controlling higher order chromatin structure and gene expression in vertebrate cells. We show that CTCF bound to HTLV-1 acts as an enhancer blocker, regulates HTLV-1 mRNA splicing, and forms long-distance interactions with flanking host chromatin. CTCF-binding sites (CTCF-BSs) have been propagated throughout the genome by transposons in certain primate lineages, but CTCF binding has not previously been described in present-day exogenous retroviruses. The presence of an ectopic CTCF-BS introduced by the retrovirus in tens of thousands of genomic locations has the potential to cause widespread abnormalities in host cell chromatin structure and gene expression.

retrovirus | latency | epigenetics | HTLV-1 | CTCF

Retroviruses integrate a dsDNA copy of their genome, the provirus, into the genome of the cell they infect. Human T-lymphotropic virus type 1 (HTLV-1) is an exogenous retrovirus, widespread in the tropics. Most infected people are asymptomatic carriers, but ∼10% develop a malignant or inflammatory disease. Adult T-cell leukemia (ATL) is a leukemia of HTLV-1–infected $CD4^+$ T cells. ATL cells frequently contain chromosomal abnormalities, and the disease is refractory to conventional chemotherapy. HTLV-1–associated myelopathy/tropical spastic paraparesis (HAM/TSP) is a chronic inflammatory disease of the spinal cord. HTLV-1 spreads within the individual both by cell-to-cell transmission and by clonal proliferation of infected cells: HTLV-1 gene products induce proliferation and enhance survival of infected cells (1, 2). In addition to the viral genes that encode enzymes and structural proteins, HTLV-1 encodes several regulatory and accessory genes in the pX region, between the *env* gene and the 3′ long terminal repeat (LTR). The *HBZ* gene is constitutively expressed from the minus strand of the integrated provirus (3), whereas plus-strand expression, required for viral propagation to uninfected cells, is suppressed or intermittently expressed in vivo, allowing escape from host immune surveillance (2, 4). It is unknown how HTLV-1 maintains this chromatin state and strand-selective transcription.

One mechanism used by HTLV-1 to suppress transcription of the plus strand is methylation of the 5′ LTR, whereas there is little DNA methylation in the 3′ LTR (5). The DNA methylation is sharply reduced at the middle of the provirus (6). This observation raised the question of whether there is a regulatory mechanism that divides the methylated 5′ part from the unmethylated 3′ part of the provirus, perhaps to allow the constitutive expression of the *HBZ* gene that appears to be required for clonal persistence of HTLV-1 (3, 7).

A chromatin insulator is a DNA region that separates transcriptionally active and inactive regions by binding to certain proteins. The best-characterized insulator-binding protein in higher eukaryotes is CTCF, an 11 zinc-finger protein highly conserved from flies to humans (8), which binds to tens of thousands of sites in the human genome and regulates chromatin structure, transcriptional activation, repression, silencing, imprinting, and alternative splicing (9). We therefore set out to test the hypothesis that CTCF binds to the HTLV-1 provirus at an epigenetic border and regulates proviral transcription.

Significance

The retrovirus human T-lymphotropic virus type 1 (HTLV-1) causes inflammatory and malignant diseases in humans. To maintain latency and avoid immune detection in vivo, HTLV-1 minimizes expression of genes on the plus-strand of the integrated provirus but allows constitutive expression of the minus-strand gene, which maintains clonal persistence. It is not understood how this gene expression is regulated. We show that CTCF, a master regulator of chromatin structure and gene expression, binds to HTLV-1, forms loops between the provirus and host genome, and alters expression of proviral and host genes. Because a typical HTLV-1–infected host carries >10^4 infected T-cell clones, each containing a provirus integrated in a different genomic site, CTCF binding gives HTLV-1 the potential to cause widespread abnormalities in the human genome.

Author contributions: Y.S., K.N., M.N., and C.R.M.B. designed research; Y.S., P.M., K.I., H.Y., A.M., M.M., A.F., T.W., and A.G.R. performed research; Y.S., P.M., K.I., H.Y., A.M., M.M., M.N., and C.R.M.B. analyzed data; K.N. made clinical diagnoses; and Y.S. and C.R.M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: ChIP-sequencing data reported in this paper have been deposited in the DNA Data Bank of Japan (accession no. DRA004162).

[1]To whom correspondence may be addressed. Email: c.bangham@imperial.ac.uk or y-satou@kumamoto-u.ac.jp.

[2]Present address: Department of Gastroenterology and Hepatology, Graduate School of Medical Sciences, Kumamoto University, Kumamoto 860-0811, Japan.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1423199113/-/DCSupplemental.

Results

Epigenetic Border in the HTLV-1 Provirus. We used chromatin immunoprecipitation (ChIP) assays to identify epigenetic marks over the HTLV-1 provirus (Fig. 1*A*), including H3K36me3 and H3K9Ac, each associated with actively transcribed genes; H3K4me3, characteristic of enhancer and promoter regions; and H2A.Z, a histone variant frequently observed in promoter and enhancer regions (10). The results showed a sharp border in each of these epigenetic modifications in the pX region of the provirus in both a nonmalignant primary T-cell clone (Fig. 1*B*) and an ATL cell line (Fig. 1*C*). The epigenetic border was more sharply defined in the ATL cell line, in which the spliced HBZ transcript (sHBZ) is transcribed but plus-strand transcription is significantly suppressed (Fig. S1), than in nonmalignant T-cell clones, in which both tax, a plus-strand transcript, and sHBZ, a minus-strand transcript, are actively transcribed. The same pattern was also observed in three other HTLV-1–infected T-cell lines (Fig. S2 *A–C*) and in peripheral blood mononuclear cells (PBMCs) freshly isolated from patients with ATL (Fig. S3 *A–F*). In addition, a border was observed in DNA methylation, consistent with previous observations (6) (Fig. 1 *D* and *E*). The results are summarized schematically in Fig. 1*F*. The observed pattern of epigenetic modifications was consistent with the observation that transcription of the proviral minus-strand exceeds transcription of the plus-strand both ex vivo (4) and frequently in HTLV-1–infected clones in vitro (Fig. S1). In contrast, in the provirus of another human retrovirus, HIV-1, the mark H3K36me3 showed a progressive rise, with no border (Fig. 1*G*); this result was expected, because H3K36me3 is characteristic of actively transcribed gene bodies, increasing toward the 3′ end of the gene (10). These observations, together with previous reports (5, 6), suggested the presence of a regulatory mechanism that establishes the distinctive pattern of respective epigenetic modifications in the 5′ and 3′ portions of the HTLV-1 provirus.

The Host Insulator-Binding Protein CTCF Binds to the HTLV-1 Provirus in Vivo. Insulator elements demarcate the boundary between transcriptionally active (euchromatic) and inactive (heterochromatic) regions of the genome (11) and prevent enhancers from activating inappropriate promoters. We therefore tested the hypothesis that CTCF binds to HTLV-1 at the observed boundary in epigenetic modifications (Fig. 1). First, we searched for the consensus CTCF DNA-binding motif (9) (Fig. S4 *A–C*) in the HTLV-1 genome and identified several candidate binding sites (Fig. 2*A*); the site in pX showed the highest similarity to the consensus sequence. CTCF ChIP assays revealed CTCF binding in pX [viral CTCF-binding site (vCTCF-BS)], but not in the other proviral regions tested (Fig. 2 *B* and *C*); similar results were obtained in fresh PBMCs from patients with either HAM/TSP (Fig. 2*D*) or ATL (Fig. 2*E*), demonstrating that CTCF binds to the pX region of HTLV-1 in natural infection in vivo. In contrast, CTCF did not bind to the provirus of HIV-1 (Fig. 2*F*), consistent with the absence of a border in the epigenetic mark H3K36me3 within the HIV-1 provirus (Fig. 1*G*).

ChIP-Sequencing Analysis Showed a Sharp Peak of CTCF Binding in the Region of the Epigenetic Border. To define CTCF binding and the epigenetic border in the HTLV-1 provirus with greater precision,

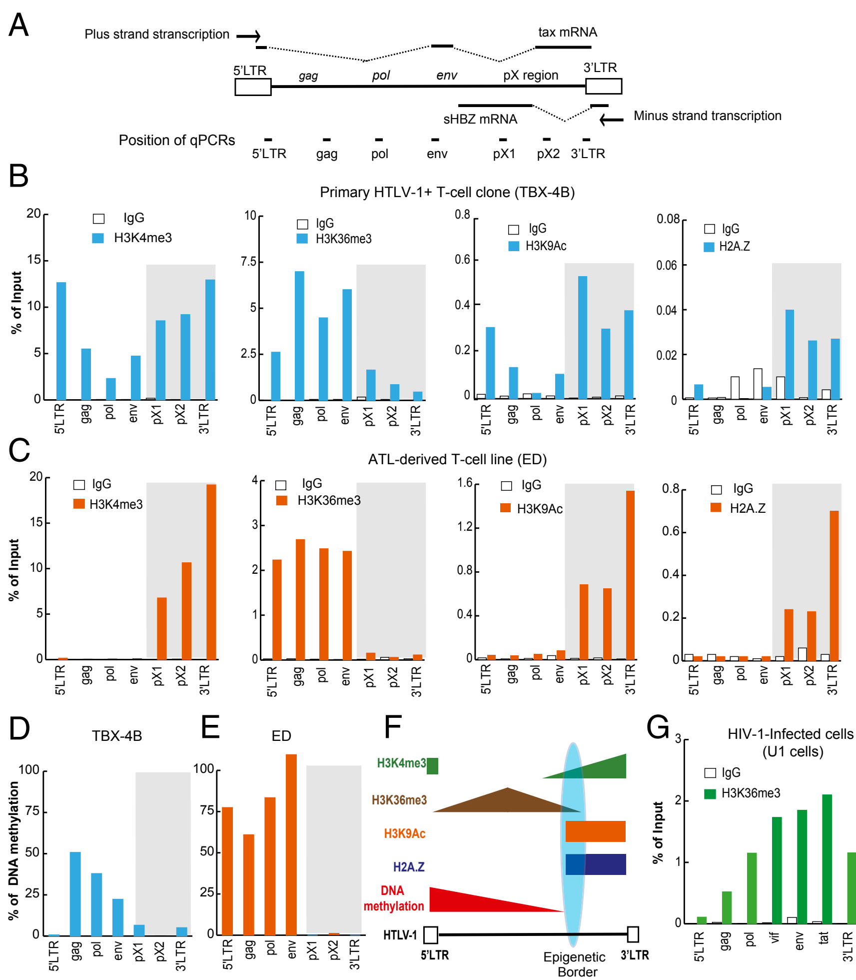

Fig. 1. Epigenetic border in the pX region of the HTLV-1 provirus. (*A*) HTLV-1 provirus showing the four main viral coding regions (*gag*, *pol*, *env*, and pX; above provirus), the regions analyzed by real-time quantitative PCR (qPCR) (primers shown in Table S1) in the ChIP assay (below provirus), and the viral transcripts tax and sHBZ. Distribution of H3K4me3, H3K36me3, H3K9Ac, and H2A.Z (*B* and *C*) and DNA methylation (*D* and *E*) over the HTLV-1 provirus in a primary nonmalignant clone (TBX-4B) (*B* and *D*) and an ATL cell line (ED) (*C* and *E*). The ChIP signal is shown as the percentage of input DNA. An IgG of unrelated specificity was used as a negative control. Representative results from three independent experiments are shown. (*F*) Schematic distribution of epigenetic modifications in the HTLV-1 provirus in relation to the CTCF-BS. (*G*) Distribution of H3K36me3 over the HIV-1 provirus in a latently infected cell line (U1).

MICROBIOLOGY

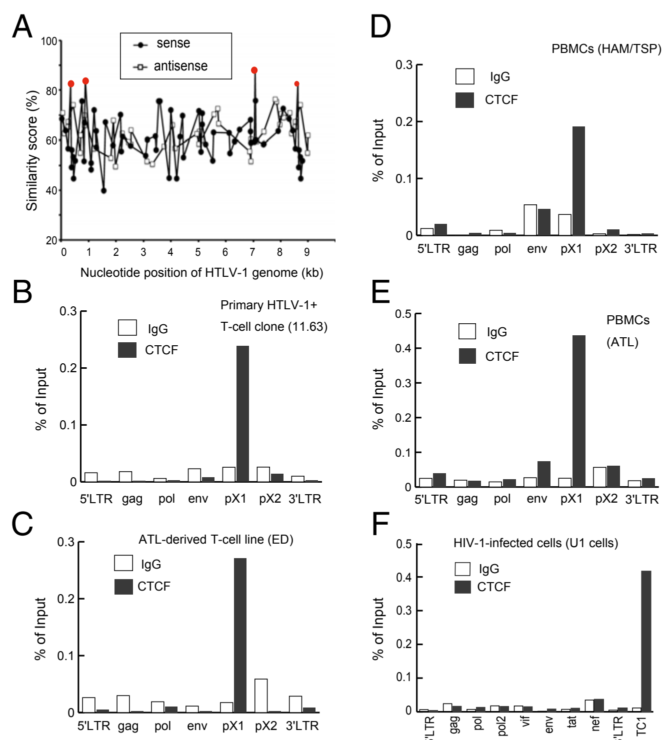

Fig. 2. Identification of CTCF (vCTCF)-BS in the HTLV-1 genome. (*A*) Similarity scores were calculated using the method shown in Fig. S4 to identify potential CTCF-BSs in the HTLV-1 genome (National Center for Biotechnology Information GenBank accession no. AB513134). The four positions with a score >80 (shown in red) were all in the plus-strand (sense). (*B*–*E*) CTCF specifically localized at the putative CTCF-BS of the pX region. Results of CTCF-ChIP assay in a nonmalignant HTLV-1–infected T-cell clone, 11.63 (*B*); an ATL cell line, ED (*C*); PBMCs from a patient with HAM/TSP (*D*); PBMCs from a patient with ATL (*E*); and an HIV-1–infected cell line (U1) (*F*). Representative results are shown from three (*B* and *C*) or two (*D* and *E*) independent experiments, respectively. TC-1, positive control region for CTCF binding in the human genome.

we performed ChIP-sequencing (ChIP-seq) analysis of a cell line (ED) derived from a patient with ATL, which contains a single-copy HTLV-1 provirus. Because the 5′ LTR and 3′ LTR are identical in sequence, sequence reads that lie within the LTR cannot be specifically mapped. LTR reads were therefore randomly mapped to either the 5′ or 3′ LTR. The results showed a single peak of CTCF binding in the pX region, consistent with the ChIP-quantitative PCR result (Fig. 3 and Fig. S5). The pattern of histone modifications and the histone variant H2A.Z also changed at the CTCF-binding region, extending a short distance beyond the binding site in each case, as observed by others (12).

CTCF Binds Directly to HTLV-1 pX DNA in a Sequence-Dependent Manner. CTCF exerts its pleiotropic functions by interacting with various cofactors (8, 13), suggesting that CTCF might bind indirectly to HTLV-1. To test this possibility, we performed an EMSA using recombinant CTCF and three synthetic oligonucleotides, each corresponding to a candidate CTCF-BS in the HTLV-1 genome, and one from the well-characterized CTCF-binding DNA region H19/DMR, the differentially methylated region of the *H19* gene (14, 15) (Fig. S6*A*); each oligonucleotide contained a different variant of the consensus sequence for CTCF binding (Fig. S6*B*). The results showed specific binding of CTCF to vCTCF in the HTLV-1 pX region, but not to either the LTR or *env* (Fig. S6*C*). Competition assays confirmed these results (Fig. S6*D*). The introduction of six nucleotide substitutions in the consensus CTCF-binding sequence in vCTCF abrogated the ability of the oligonucleotide to compete with the WT pX sequence in binding to CTCF, showing that CTCF binds directly to HTLV-1 DNA in the pX region in a sequence-dependent manner (Fig. S6*E*). The consensus sequence for the CTCF-BS in pX is conserved both in the HTLV-1 sequences in the GenBank (Fig. S7*A*) and in all cases of ATL and asymptomatic carriers examined (Fig. S7*B*).

HTLV-1 pX Has Enhancer-Blocking Activity. CTCF bound to DNA between an enhancer and a promoter can block the activation of the promoter by the enhancer. A 1-kb DNA fragment of pX containing the CTCF-BS showed strong enhancer-blocking (EB) activity, similar to the positive control, H19/DMR (Fig. 4*A*). EB function was observed with pX inserted in either orientation between the enhancer and the promoter, and it was reduced, but not eliminated, when the pX fragment was inserted outside the promoter-enhancer region, suggesting that pX contains a bi-directional insulator function (Fig. 4*A*). The EB activity was detected using 1-kb, 300-bp, or 100-bp fragments (Table S2) of the pX sequence, with each containing the CTCF-binding region (Fig. 4*B*); the progressively stronger EB effect observed with longer fragments suggests that additional host factors bind the provirus near the CTCF site. In contrast, there was no EB activity with the HTLV-1 *env* sequence, which does not bind CTCF. A degree of promoter suppression was observed even when the pX sequence was located beyond the promoter-enhancer (Fig. 4*A*). In these

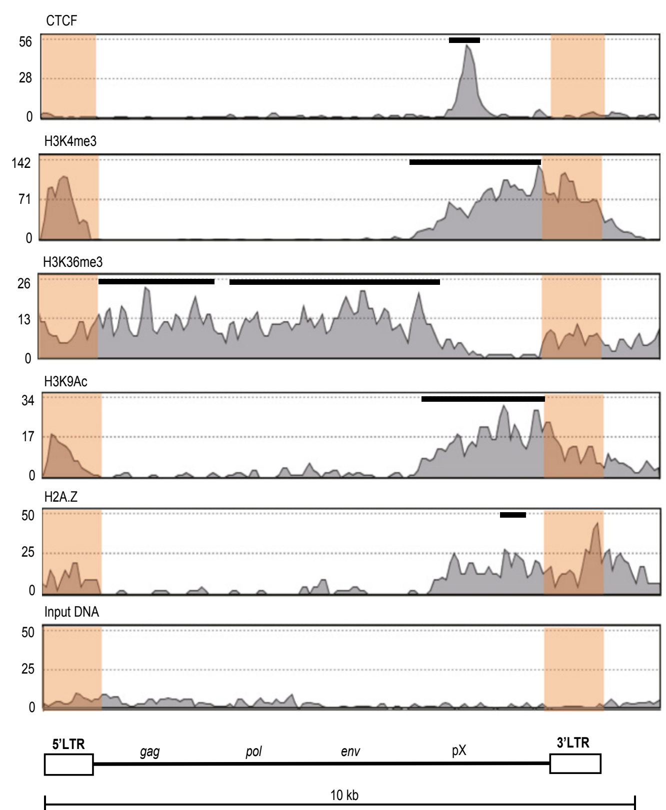

Fig. 3. Epigenetic landscape of HTLV-1 provirus analyzed by ChIP-seq. Enrichment of CTCF, H3K4me3, H3K36me3, H3K9Ac, and H2A.Z in ED cells is shown. These data are aggregated from two biological replicate experiments; each replicate showed a similar epigenetic profile. LTR sequences are shaded in orange. LTR reads were randomly mapped to the 5′ LTR or 3′ LTR, because their sequences are identical. Peaks of each signal were analyzed by the model-based analysis of ChIP-seq algorithm and are shown as black bars above the histogram. The significance of enrichment of each signal over input-DNA peaks was calculated with a cutoff *P* value of 10^{-5}. The ChIP-seq profile of the flanking human genome is shown at a larger scale (~100 kb) in Fig. S5.

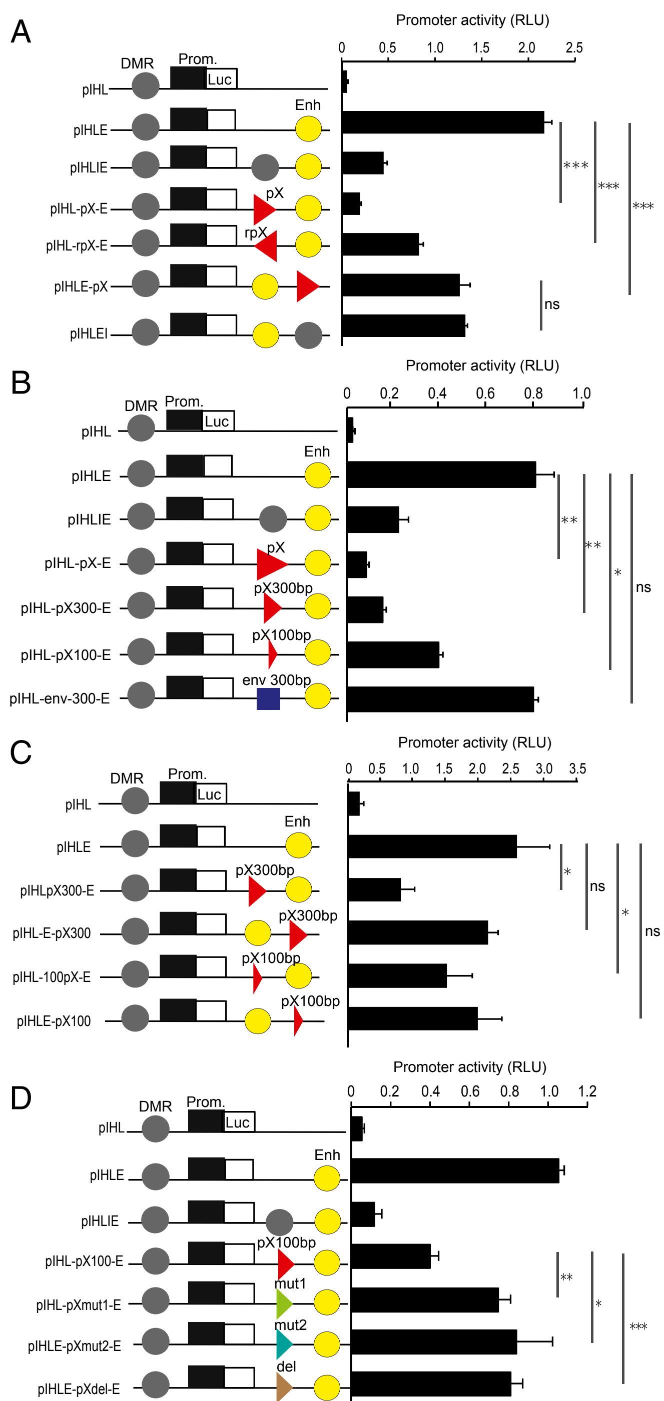

Fig. 4. HTLV-1 pX region exerts CTCF-dependent EB activity. (*A*) pIHL-pX-E and pIHL-rpX-E plasmids were constructed by inserting a fragment of ~1,000 bp, containing the CTCF-BS from the pX region, inserted in either a sense or antisense orientation between the promoter and the enhancer in pIHLE reporter vector. The *H19* DMR insulator was used as a positive control for insulator function. For pIHLE-pX, the *pX* fragment was inserted downstream of the enhancer in pIHLE reporter vector. Firefly luciferase activities were normalized to Renilla luciferase and are shown as relative light units (RLU). (*B*) Plasmids pIHL-pX300-E and pIHL-pX100-E were constructed by inserting fragments, containing the pX CTCF-BS, of ~300 bp and ~100 bp, respectively. The pIHL-env-300-E was constructed by inserting a 300-bp fragment from the HTLV-1 env region. (*C*) For pIHLE-pX300 and pIHLE-pX100, the *pX* fragment was inserted downstream of the enhancer. Plasmids linearized with the restriction enzyme MluI were used in the assay. (*D*) Plasmids pIHLpXmut1-E and pIHLpXmut2-E were generated by inserting a DNA fragment of 100 bp containing the CTCF-BS with the same nucleotide substitutions as shown in Fig. S6*B*. In pIHL-pXdel-E, the core CTCF-BS was deleted. DMR, *H19* DMR insulator; Enh, SV40 enhancer; Luc, *luciferase* gene; Prom., promoter of *H19* gene. Data shown are representative of three independent experiments (*t* test: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). ns, not significant.

experiments, we used circular plasmids, in which the enhancer might affect the promoter either upstream or downstream. We therefore repeated the EB assay after linearizing the plasmids. The results showed that 300-bp or 100-bp pX fragments exerted EB activity only when inserted between the promoter and the enhancer (Fig. 4*C*). The EB activity of the pX region was significantly reduced by mutation of the CTCF-BS (Fig. S6*B*), indicating that CTCF plays an important role in the EB function (Fig. 4*D*).

CTCF Influences Proviral Transcription and RNA Splicing. To investigate the effect of CTCF on proviral transcription, we transfected an ATL cell line (MT-1) with shRNA to knock down CTCF expression (Fig. S8 *A* and *B*). There was a slight suppressive effect of CTCF knockdown (KD) on the proviral transcriptome (Fig. S8*C*). Further analysis by quantitative RT-PCR (qRT-PCR) assay with transcript-specific primers (Fig. S8*D*) revealed significant repression of the HTLV-1 p30 gene (Fig. S8*E*). We further analyzed the effect of mutations in the viral CTCF-binding region by using the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 system in another ATL cell line, ED. We confirmed the loss of CTCF binding by mutating six key nucleotides (Fig. S6*B*; vCTCF-mut1) in the proviral core CTCF-BS (Fig. 5*A*). Initially, there was no significant change in expression of *tax* or *HBZ* (Fig. 5*B*), but a slight change was evident in the border of H3K36me3 (Fig. 5*C*). Because *HBZ* expression plays a critical role in the proliferation of ATL cells (3), ED cells with low *HBZ* expression may be counterselected during in vitro growth. To reduce the dependency of ED cell proliferation on *HBZ*, we cultured the cells in the presence of IL-2. After 4 wk, the loss of the vCTCF-BS (ΔvCTCF-BS) resulted in a significant decrease in sHBZ transcription (Fig. 5*D*). The *tax* gene was not reactivated; the provirus of ED cells appears to be epigenetically stabilized by DNA hypermethylation in the 5′ LTR (Fig. 1*E*). Together with the results shown in Figs. 1–5 and Figs. S1–S3 and S8, these data indicate that CTCF plays a role in the regulation of HTLV-1 proviral transcription and RNA splicing.

HTLV-1 Alters Local Higher Order Chromatin Structure and Gene Expression in the Host Genome. To test the hypothesis that the HTLV-1 provirus forms CTCF-mediated chromatin interactions with the flanking host genome, we used the chromosome conformation capture (3C) assay to analyze chromatin from ED cells containing either WT HTLV-1 or the mutant HTLV-1 ΔvCTCF-BS, in which CTCF binding was abrogated (Fig. 5*A*). A strong CTCF ChIP-seq signal [cellular CTCF (cCTCF)] was observed in the host genome, 48 kb upstream of the proviral integration site (Fig. 6*A*). Quantitative 3C analysis demonstrated that the cCTCF sites made long-range interactions with the WT provirus significantly more frequently than with the ΔvCTCF-BS mutant (Fig. 6*B*). These CTCF sites are oriented toward the provirus, which is integrated in the genome in the negative orientation (Fig. 6*A*). The vCTCF-BS and host CTCF-BS are therefore oriented toward each other; convergent CTCF-BSs appear to be required for chromatin loop formation (16–18). We also examined the effect of the ΔvCTCF-BS on local host gene expression: *TCHP*, *GIT2*, and *C12orf76* (Fig. 6*C*). Transcription of two host genes near an upstream CTCF-BS was affected by vCTCF-BS mutation (Fig. 6*D*). These results show that the integrated HTLV-1 provirus can form CTCF-dependent chromatin loops with the flanking host genome and may cause CTCF-dependent changes in host gene expression.

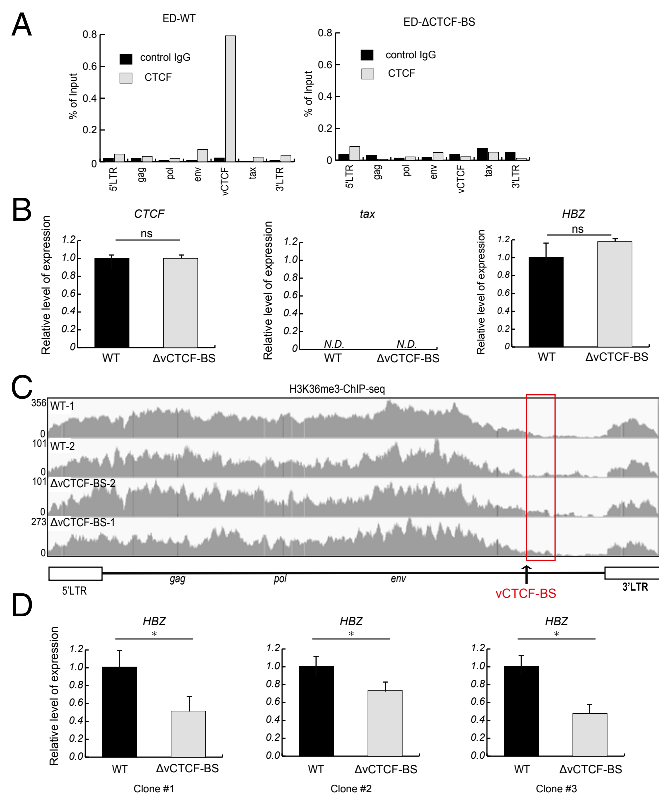

Fig. 5. Effect of mutations in the vCTCF sequence on proviral transcription and a histone modification, H3K36me3. (*A*) Introduction of mutations in the vCTCF sequence (using CRISPR-Cas9) abolished the CTCF binding to the pX region. Representative CTCF-ChIP-quantitative PCR results of ED-WT and ED ΔvCTCF-BS cells are shown. (*B*) Representative qRT-PCR result of viral transcripts from HTLV-1 provirus. 18S rRNA was also quantified as an internal control. Data shown are representative of two independent experiments. Tax expression is inhibited in ED cells by dense DNA methylation in the 5′ LTR and by a premature stop codon in the *tax* gene. *N.D.*, not detectable. (*C*) H3K36me3–ChIP-seq results from WT and ED ΔvCTCF-BS cells (two independent experiments in each case). Because the 5′ LTR and 3′ LTR are identical in sequence, the LTR reads were randomly mapped to either the 5′ LTR or the 3′ LTR. The black arrow indicates a vCTCF-BS. The red rectangle indicates the region immediately downstream of the CTCF-BS. (*D*) Representative qRT-PCR result of viral transcripts from HTLV-1 provirus. 18S rRNA was quantified as an internal control. Data shown are representative of two independent experiments on each of three independent sets of clones after long-term cultivation in the presence of IL-2 (100 IU/mL). Statistical significance was assessed by Student's *t* test (* $P < 0.05$).

Discussion

We show that the chromatin insulator protein CTCF binds directly to the HTLV-1 provirus. CTCF-BSs have been propagated throughout the genome by transposons in certain nonhuman primate lineages (19), and CTCF has been shown to regulate latency programs in the herpesviruses Epstein-Barr virus (EBV) (20) and Kaposi's sarcoma herpesvirus (KSHV) (21). However, CTCF binding has not been described in present-day exogenous retroviruses. The diverse functions ascribed to CTCF include enhancement or repression of transcription, promoting or blocking communication between enhancers and promoters, acting as a barrier to the propagation of epigenetic changes, and regulation of mRNA splicing (22, 23). We show here that KD of CTCF changed the pattern of splicing of the p30 transcript, in which the CTCF-BS is present near the splicing junction (Fig. S8 *D* and *E*). The actions of CTCF at a given location depend on the local context, involving selective interactions between CTCF and both transcriptional cofactors and the primary DNA sequence (24). Homodimerization of CTCF bound to two genomic sites can cause chromatin looping (8), forming structural chromatin loops (∼1 Mb in size), known as topologically associating domains (TADs), which are largely shared between cell types, and smaller functional loops (sub-TADs). Sub-TADs, typically contained within TADs, appear to regulate transcription by bringing a distant enhancer near the respective promoter (8, 16, 18, 25). However, although CTCF plays a central role in establishing chromatin structure, epigenetic borders, and transcriptional activity, removal of CTCF does not invariably lead to changes in chromatin looping (26) and epigenetic marks flanking the border (12). Further, gene expression can be changed *in cis* at some distance from the abrogated CTCF-BS (12, 27). The epigenetic pattern and chromatin structure may be maintained in the absence of CTCF by cohesin or by other DNA-binding factors, such as transcription factors, other zinc finger proteins (e.g., ZNF143), or RNAs (8, 13). We show here that CTCF binding to HTLV-1 plays

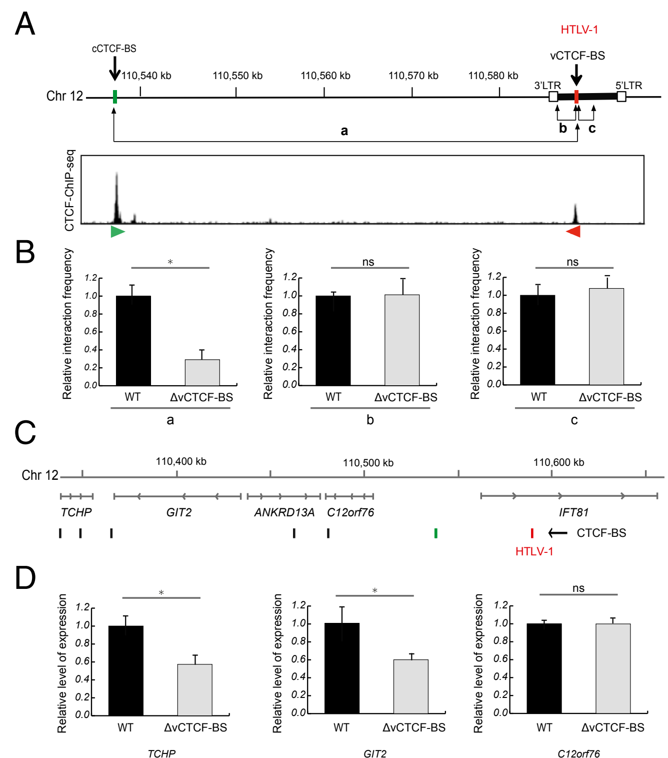

Fig. 6. HTLV-1 alters local higher order chromatin structure and gene expression in the host genome. (*A*) CTCF ChIP-seq data reveal binding of CTCF to the HTLV-1 provirus and the flanking host genome across a 60-kb region in ED cells. The green and red arrowheads denote the orientation of the respective CTCF-BS (inferred from the DNA sequence) in the cellular genome (cCTCF-BS) and the provirus (vCTCF-BS). (*B*) Frequency of long-range chromatin interactions was assessed by quantitative 3C analyses on six independent libraries (three ED-WT and three ED-ΔvCTCF-BS), using primers within the NlaIII fragments containing the cCTCF-BS (green bar in *A*) and vCTCF-BS (red bar in *A*) and the TaqMan probe encompassing the ligation junction of two fragments. Results are mean ± SE. The mean interaction frequency in the mutant was significantly lower than in the WT (three biological replicates of each) (Welch two-sample *t* test: * $P = 0.0186$). (*C*) Schematic figure of host genes near the integration site of HTLV-1 in ED cells. CTCF-binding regions in ED cells are also shown as vertical bars below the figure. The red bar shows a CTCF-BS within the HTLV-1 provirus, and the green bar shows the interacting host CTCF site shown in *A*. (*D*) Representative result of qRT-PCR analysis. Expression levels of *GIT2*, *TCHP*, and *C12orf76* genes were analyzed in ED-WT and ED-ΔvCTCF-BS cells after long-term cultivation in the presence of IL-2. Representative results from two independent experiments are shown. The level of expression of *GIT2* and *TCHP* was significantly lower in the ΔvCTCF-BS line than in the WT (Student's *t* test: ** $P < 0.01$; * $P < 0.05$).

a role in EB (Fig. 4) and proviral gene expression (Fig. 5); however, abolition of CTCF binding did not cancel the effect completely (Figs. 4*D* and 5*C*), suggesting that other cofactors also contribute to these mechanisms.

Binding of CTCF to the HTLV-1 provirus may serve several functions that benefit the virus, among which we identify three possibilities that are not mutually exclusive. First, by forming a barrier to epigenetic modifications, CTCF may cause or perpetuate differential transcription of the 5′ and 3′ parts of the provirus, resulting in persistent expression of *HBZ*, which is required for clonal persistence (2, 7), while promoting escape from the strong host immune response to the plus-strand gene products Tax, Pol, and Gag (4) by temporary silencing of the plus-strand genes (3, 5, 6). The results reported here and the previous observation of partial DNA methylation in the provirus (3, 5, 6) suggest that CTCF establishes an epigenetic border in the provirus, which may be stably maintained thereafter by DNA methylation. Second, CTCF may block aberrant activation of the 5′ and 3′ viral promoters by the enhancer present in the 3′ LTR and 5′ LTR, respectively. Third, CTCF might tether the provirus to the nucleolar periphery (28), allowing cell cycle-dependent proviral silencing, and thereby promoting immune escape. This possibility is consistent with our observation (29) of selective survival in vivo of HTLV-1 proviruses integrated in acrocentric chromosomes, which are associated with the nucleolar periphery. The impact of the HTLV-1 provirus on higher order host chromatin structure and the consequent effects on gene expression are adventitious, because they depend on the proviral integration site: They are therefore unlikely to benefit the virus systematically. The consequences of CTCF binding will differ between clones of HTLV-1–infected cells, because the proviral integration site is unique to each clone. A chromatin loop that brings either the 5′ LTR or the 3′ LTR adjacent to a host gene may dysregulate cell growth and cause or contribute to malignant transformation of the cell (30, 31). HTLV-1–infected individuals typically possess $>10^4$ infected T-cell clones, each containing one provirus in a unique genomic location (32). HTLV-1 therefore has the potential to cause widespread abnormalities in chromatin structure and gene expression of the human genome by inserting ectopic CTCF-BSs. These abnormalities may, in turn, contribute to the HTLV-1–associated diseases ATL and HAM/TSP (33).

Materials and Methods

This study was conducted according to the principles expressed in the Declaration of Helsinki and was approved by the review boards of the respective institutions (National Research Ethics Service, UK; Ethics Committee of Kumamoto University Graduate School of Medicine, Japan). Written informed consent was obtained from each subject. Full details of the materials and methods used are provided in *SI Materials and Methods*, including cell culture, qRT-PCR, EMSA, ChIP assay, methylated DNA immunoprecipitation assay, EB assay, shRNA KD of CTCF, RNA-sequencing, CRISPR modification of the HTLV-1 genome, quantitative 3C, and identification of candidate CTCF-BSs in the HTLV-1 genome.

ACKNOWLEDGMENTS. We thank Graham Taylor for clinical samples and Michiyo Tokunaga for technical support. We thank Nicholas Proudfoot, Douglas Higgs, Jim Hughes, Matthias Merkenschlager, and members of the C.R.M.B. laboratory and Y.S. laboratory for discussion. This work was supported by a senior investigator award from the Wellcome Trust UK (to C.R.M.B.), Leuka (to C.R.M.B.), Japan Society for the Promotion of Sciences Grant-in-Aid for Scientific Research scheme (KAKENHI) (Grant 26461428 to Y.S.), Takeda Science Foundation (to Y.S.), the Japan Science of Technology Agency (to Y.S.), Core Research for Evolutionary Science and Technology (to M.N.), and The Naito Foundation (to M.M.).

1. Igakura T, et al. (2003) Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science* 299(5613):1713–1716.
2. Matsuoka M, Jeang KT (2007) Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer* 7(4):270–280.
3. Satou Y, Yasunaga J, Yoshida M, Matsuoka M (2006) HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. *Proc Natl Acad Sci USA* 103(3):720–725.
4. Bangham CR, et al. (2009) The immune control of HTLV-1 infection: Selection forces and dynamics. *Front Biosci (Landmark Ed)* 14:2889–2903.
5. Koiwa T, et al. (2002) 5′-long terminal repeat-selective CpG methylation of latent human T-cell leukemia virus type 1 provirus in vitro and in vivo. *J Virol* 76(18):9389–9397.
6. Taniguchi Y, et al. (2005) Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. *Retrovirology* 2:64.
7. Arnold J, et al. (2006) Enhancement of infectivity and persistence in vivo by HBZ, a natural antisense coded protein of HTLV-1. *Blood* 107(10):3976–3982.
8. Ong CT, Corces VG (2014) CTCF: An architectural protein bridging genome topology and function. *Nat Rev Genet* 15(4):234–246.
9. Kim TH, et al. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128(6):1231–1245.
10. Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823–837.
11. Cuddapah S, et al. (2009) Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 19(1):24–32.
12. Narendra V, et al. (2015) CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347(6225):1017–1021.
13. Bouwman BA, de Laat W (2015) Getting the genome in shape: The formation of loops, domains and compartments. *Genome Biol* 16:154.
14. Hark AT, et al. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405(6785):486–489.
15. Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405(6785):482–485.
16. Rao SS, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665–1680.
17. Guo Y, et al. (2015) CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162(4):900–910.
18. Tang Z, et al. (2015) CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* 163(7):1611–1627.
19. Schmidt D, et al. (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148(1-2):335–348.
20. Tempera I, Lieberman PM (2014) Epigenetic regulation of EBV persistence and oncogenesis. *Semin Cancer Biol* 26:22–29.
21. Kang H, Cho H, Sung GH, Lieberman PM (2013) CTCF regulates Kaposi's sarcoma-associated herpesvirus latency transcription by nucleosome displacement and RNA polymerase programming. *J Virol* 87(3):1789–1799.
22. Hadjur S, et al. (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460(7253):410–413.
23. Witcher M, Emerson BM (2009) Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol Cell* 34(3):271–284.
24. Nakahashi H, et al. (2013) A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Reports* 3(5):1678–1689.
25. Mishiro T, et al. (2009) Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J* 28(9):1234–1245.
26. Yang R, et al. (2015) Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus. *Nucleic Acids Res*, 10.1093/nar/gkv1358.
27. Ishihara K, Oshimura M, Nakao M (2006) CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* 23(5):733–742.
28. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G (2004) CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* 13(2):291–298.
29. Cook LB, et al. (2014) The role of HTLV-1 clonality, proviral structure, and genomic integration site in adult T-cell leukemia/lymphoma. *Blood* 123(25):3925–3931.
30. Babaei S, Akhtar W, de Jong J, Reinders M, de Ridder J (2015) 3D hotspots of recurrent retroviral insertions reveal long-range interactions with cancer genes. *Nat Commun* 6:6381.
31. Hacein-Bey-Abina S, et al. (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* 118(9):3132–3142.
32. Laydon DJ, et al. (2014) Quantification of HTLV-1 clonality and TCR diversity. *PLOS Comput Biol* 10(6):e1003646.
33. Zhang J, Markus J, Bies J, Paul T, Wolff L (2012) Three murine leukemia virus integration regions within 100 kilobases upstream of c-myb are proximal to the 5′ regulatory region of the gene through DNA looping. *J Virol* 86(19):10524–10532.
34. Folks TM, Justement J, Kinter A, Dinarello CA, Fauci AS (1987) Cytokine-induced expression of HIV-1 in a chronically infected promonocyte cell line. *Science* 238(4828):800–802.
35. Watanabe T, et al. (2012) Higher-order chromatin regulation and differential gene expression in the human tumor necrosis factor/lymphotoxin locus in hepatocellular carcinoma cells. *Mol Cell Biol* 32(8):1529–1541.
36. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
37. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
38. Kimata JT, Wong FH, Wang JJ, Ratner L (1994) Construction and characterization of infectious human T-cell leukemia virus type 1 molecular clones. *Virology* 204(2):656–664.