Discovery and Characterization of a Hidden Retroviral Enhancer by Viral DNA-capture-seq Approach

Misaki Matsuo
Kumamoto University
Takaharu Ueno
Kansai Medical University
Kazuaki Monde
Kumamoto University https://orcid.org/0000-0002-8314-5306
Benjy Jek Yang Tan
Kumamoto University https://orcid.org/0000-0002-8275-7378
Paola Miyazato
Kumamoto University
Kyosuke Uchiyama
Kumamoto University
Saiful Islam
Kumamoto University
Hiroo Katsuya
Kumamoto University
Shinsuke Nakajima
Kansai Medical University
Masahito Tokunaga
Imamura General Hospital
Kisato Nosaka
Kumamoto University
Hiroyuki Hata
Kumamoto University
Atae Utsunomiya
Imamura General Hospital
Jun-ichi Fujisawa
Kansai Medical University
Yorifumi Satou (y-satou@kumamoto-u.ac.jp)
Kumamoto University https://orcid.org/0000-0002-1495-7810

Article

Keywords: retrovirus, enhancer, viral DNA-capture-seq, HTLV-1, SRF, ELK-1

Posted Date: February 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-141856/v1

Discovery and Characterization of a Hidden Retroviral Enhancer by Viral DNA-capture-seq Approach

Misaki Matsuo[1,2], Takaharu Ueno[3], Kazuaki Monde[4], Benjy Jek Yang Tan[1,2], Paola Miyazato[1,2], Kyosuke Uchiyama[1,2], Saiful Islam[1,2], Hiroo Katsuya[1,5], Shinsuke Nakajima[3], Masahito Tokunaga[6], Kisato Nosaka[7], Hiroyuki Hata[8], Atae Utsunomiya[6], Jun-ichi Fujisawa[3], Yorifumi Satou[1,2]*

[1] Division of Genomics and Transcriptomics, Joint Research Center for Human Retrovirus Infection, Kumamoto University, Kumamoto, 860-8556, Japan,

[2] International Research Center for Medical Sciences (IRCMS), Kumamoto University, Kumamoto, 860-0811Japan,

[3] Department of Microbiology, Kansai Medical University, Osaka, 573-1010, Japan,

[4] Department of Microbiology, Faculty of Life Sciences, Kumamoto University, Kumamoto, 860-8556, Japan,

[5] Division of Hematology, Respiratory Medicine and Oncology, Saga University, Saga, 849-8501, Japan,

[6] Department of Hematology, Imamura General Hospital, Kagoshima, 890-0064, Japan,

[7] Department of Hematology, Rheumatology and Infectious Disease, Kumamoto University Hospital, Kumamoto, 860-8556, Japan,

[8] Division of Informative Clinical Sciences, Faculty of Life Sciences, Kumamoto University, Kumamoto, 862-0972, Japan.

*Correspondence and requests for materials should be addressed to Y.S.

1-1-1 Honjo, Chuo-ku, Kumamoto

860-8556 Japan

Tel: +81-96-373-6830

E-mail: y-satou@kumamoto-u.ac.jp

Abstract

Human T-cell leukemia virus type 1 (HTLV-1) is a retrovirus that causes a cancer of infected cells called adult T-cell leukemia (ATL). There is both sense and antisense transcription from the integrated provirus. Sense transcription tends to be suppressed, but antisense transcription is constitutively active *in vivo* even in proviruses lacking the 5' long terminal repeat (LTR), a known viral enhancer and promoter. Various efforts have been made to elucidate the regulatory mechanism of HTLV-1 provirus for several decades; however, it remains unknown how HTLV-1 antisense transcription is maintained. Here, using proviral DNA-capture followed by high-throughput sequencing, we found a previously unidentified viral enhancer not in the LTR but in the middle of the HTLV-1 provirus. The host transcription factors, SRF and ELK-1, bind to this enhancer region both in cell lines and in freshly isolated ATL cells. HTLV-1 containing mutations in the SRF- and ELK-1-binding sites markedly decreased chromatin openness at the viral enhancer, viral gene transcription, and enhancing effects on host gene transcription near the viral integration site. Aberrant host genome transcription was observed at nearby integration sites in defective proviruses containing the enhancer in ATL cells. This finding reveals how the exogenous retrovirus achieves persistent infection in the host via the internal viral enhancer and resolves certain long-standing questions concerning HTLV-1 infection. We anticipate that the DNA-capture-seq approach can be applied to analyze regulatory mechanisms of other oncogenic viruses integrated into the host cellular genome.

Keywords: retrovirus, enhancer, viral DNA-capture-seq, HTLV-1, SRF, ELK-1

Human T-cell leukemia virus type 1 (HTLV-1) is an exogenous retrovirus endemic to some tropical regions. HTLV-1 infection is associated with human diseases such as adult T-

cell leukemia/lymphoma (ATL) [1-3]. Because HTLV-1 is a retrovirus, its viral RNA genome is reverse transcribed into double-stranded DNA that is then integrated into the host genomic DNA, forming a provirus, which serves as a template for generating new viral particles. A characteristic of HTLV-1 infection is that the virus maintains its copy number during chronic infection not via production of free viral particles but via clonal expansion and persistence of infected T-cell clones [4,5]. Viral genes encoded in the provirus are known to play roles in the persistence and expansion of infected cells. Tax is a viral protein encoded in the plus strand of HTLV-1 and possesses oncogenic functions, such as anti-apoptosis and cell proliferation [6,7]. HBZ is encoded in antisense transcripts from the 3′ LTR and also plays a pivotal role in viral persistence and pathogenesis [4,8]. Antisense transcription from the 3'LTR is constitutively activated at the population level even in ATL cells, whereas sense transcription from the 5′LTR is frequently silenced *in vivo* [9,10], suggesting that this proviral expression pattern is beneficial for the virus to persist in the host and predisposes infected cells to malignant cellular transformation [5].

We previously reported that there is an insulator region in the HTLV-1 provirus [11]. While the viral insulator may explain the distinct transcriptional pattern between the 5'LTR and the 3'LTR of the HTLV-1 provirus, it cannot explain the large difference in transcriptional activity between the 3′LTR and the 5′LTR. Mobile DNA elements in the genome can be dangerous to host cells because they act as genome mutagens and may induce genomic instability. Therefore, the host cell has evolved defense systems to transcriptionally and post-transcriptionally silence such mobile elements. For example, the KRAB ZnF-Trim28-Setdb1-ZFP809 complex induces transcriptional silencing of murine leukemia virus in embryonic stem cells [12,13]. Constitutive activation of antisense transcription from HTLV-1 raises the possibility that there may exist a regulatory mechanism that actively maintains transcription from the 3′LTR.

In this study, we screened transcriptional regulatory regions within the HTLV-1 provirus to identify nucleosome-free regions (NFRs), using a highly sensitive micrococcal nuclease sequencing (MNase-seq) approach, following our recently developed HTLV-1 DNA-capture-seq protocol [14,15]. The results reveal an internal HTLV-1 enhancer, which has not been identified for 40 years since Poiesz et al identified HTLV-1 in 1980 [2].

Results

MNase-seq with HTLV-1 DNA-capture identified a significant nucleosome-free region in the HTLV-1 provirus

Transcriptional regulatory regions in the genome, such as promoters and enhancers, are generally nucleosome-free because they need to be accessed by transcription factors, epigenetic modifiers or chromatin remodelers, to exert regulatory function. We utilized our recently developed HTLV-1 DNA-capture-seq approach, which enables a several thousand-fold increase in the detection sensitivity of HTLV-1 sequence [14,15]. We analyzed two HTLV-1-infected T-cell lines, ED and TBX-4B. ED is an ATL cell line derived from an ATL patient, in which sense transcription of the provirus is silenced by DNA methylation and nonsense mutation of the *tax* gene, while the antisense transcription remains active [16,17] (Figure 1a). TBX-4B is also a T-cell clone derived from an ATL patient; however, it is not an ATL clone [18]. Sense transcription was more abundant than antisense transcription from the HTLV-1 provirus in TBX-4B cells (Fig. 1a) possibly due to cultivation *ex vivo* [19]. To identify previously uncharacterized transcriptional regulatory regions, we screened for NFRs in the HTLV-1 provirus in an unbiased manner by performing MNase-seq analysis, where MNase preferentially digests genomic DNA lacking nucleosomes (Figure 1b). MNase-seq demonstrated a sharp NFR signal at the ~7,100 nt position of HTLV-1 in ED cells, close to the insulator region we recently reported [11] (Figure 1c). Because insulator regions are known to have regulatory function, they generally possess open (nucleosome-free) chromatin, and so can frequently be identified using MNase-seq. The most nucleosome-depleted region was present between the insulator region and the 3′LTR (Figure 1c). This region is part of exon 3 of the *tax* gene; however, there have been no previous reports regarding its possible function as a DNA element. We further asked if the NFR is also observed in *in vivo* samples in addition to the *in vitro* cell lines by analyzing peripheral mononuclear cells (PBMCs) freshly

isolated from ATL patients and an asymptomatic carrier. We found that the NFR was also present in the same region as in the HTLV-1 infected cell lines (Figure 1d), indicating that the NFR is present *in vivo* in naturally virus-infected individuals as well as *in vitro* cell lines.

Figure 1

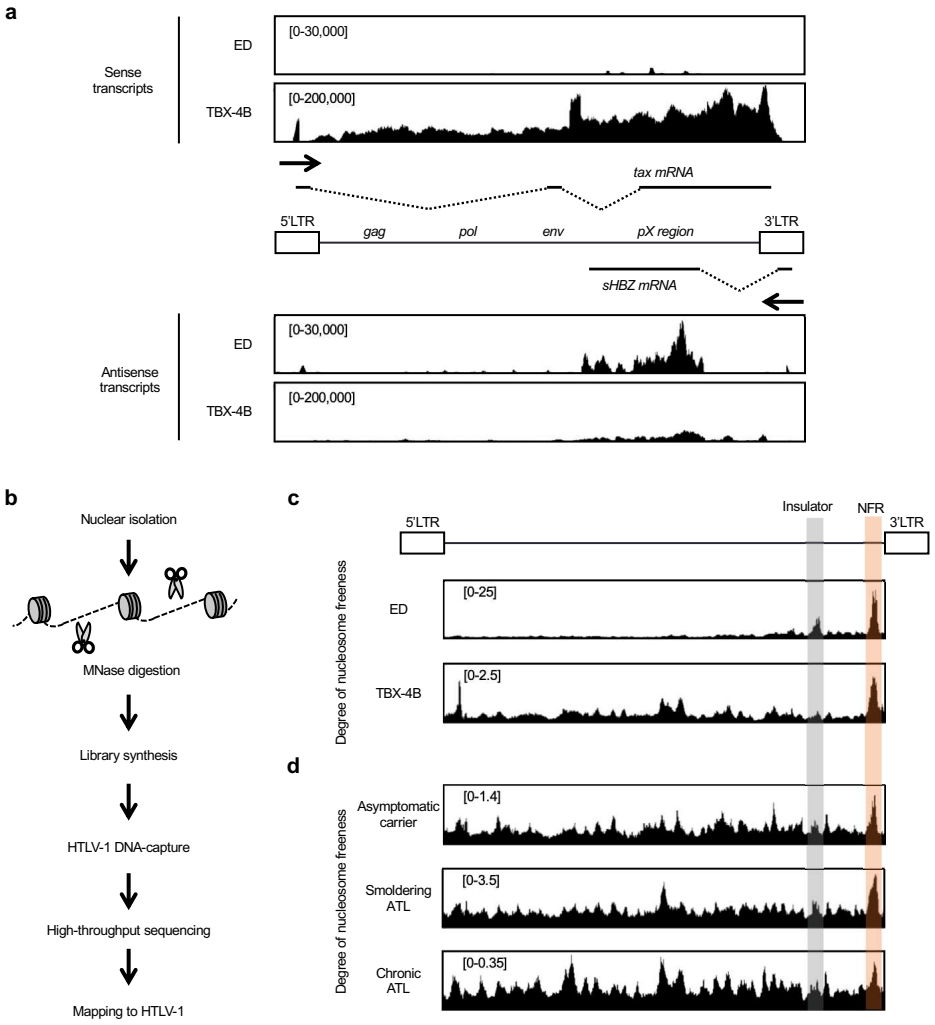


Fig. 1 Transcriptome pattern and nucleosome positions in the HTLV-1 provirus. **a** Stranded proviral transcriptome are visualized at the sense (above) or antisense (below) orientation in ED cells and TBX-4B cells by Integrative Genomics Viewer (IGV; https://software.broadinstitute.org/software/igv/). **b** Experimental workflow of MNase-seq with HTLV-1 DNA-capture-seq. **c**, **d** MNase-seq of ED cells, TBX4B cells, (**c**) and PBMCs of asymptomatic HTLV-1 carrier and ATL patients (**d**). Degree of nucleosome freeness is shown as the MNase-seq value normalized to the input DNA-seq value. Orange-shaded region indicates the NFR location and gray-shaded region indicates the insulator region. ATL, adult T-cell leukemia/lymphoma; NFR, nucleosome-free region.

The nucleosome free region harbors enhancer-related histone modifications and produces enhancer RNAs

To investigate the functional role of the most significant NFR, we performed promoter assays with the promoter of the *HBZ* gene [20] (Figure 2a). Promoter activity was enhanced by insertion of the NFR either upstream or downstream of the promoter and in a sense or antisense orientation, indicating that the NFR has an enhancer function (Figure 2b). We also evaluated the effect of the NFR on the 5′LTR, which is the promoter of the sense transcription in the HTLV-1 provirus (Figure 2c). The promoter activity of the 5′LTR was enhanced but by a much smaller factor than that observed for the 3′LTR (Figure 2c). T-cell stimulation with TNF-α or Phorbol 12- Myristate 13-Acetate (PMA)/Ionomycin did not enhance promoter activity but marginally increased promoter/enhancer activity (Figure 2d).

We next analyzed enhancer-related histone modifications within the HTLV-1 proviral region. Chromatin immunoprecipitation sequencing (ChIP-seq) signals of enhancer-related histone modifications [21], including H3K27Ac, H3K4me1, and H3K4me2, were high around the NFR in ED cells (Figure 2e, upper panel). Consistent with the high level of transcriptional activity from the 5′LTR in clone TBX-4B, in which both the 5′ and 3′ LTRs are transcriptionally active (Figure 1a), there was a wide distribution of enhancer-related histone modifications in this clone (Figure 2e, lower panel). It has been reported that enhancer regions express enhancer RNAs (eRNAs) - non-coding RNAs with divergent orientation from the center of the enhancer [22]. Thus, we performed native elongating transcript-cap analysis of gene expression (NET-CAGE) to detect eRNAs [23] . NET-CAGE identifies the sequence of the 5′ region of mRNAs or non-coding RNA adjacent to the cap structure using nascent RNA, which is useful in identifying transcriptional start sites and eRNAs with high resolution. eRNAs from the intragenic HTLV-1 enhancer region were detected in ED cells

(Figure 2f). These findings demonstrate that the NFR in the HTLV-1 pX region harbors several fundamental features of an enhancer region.

Figure 2

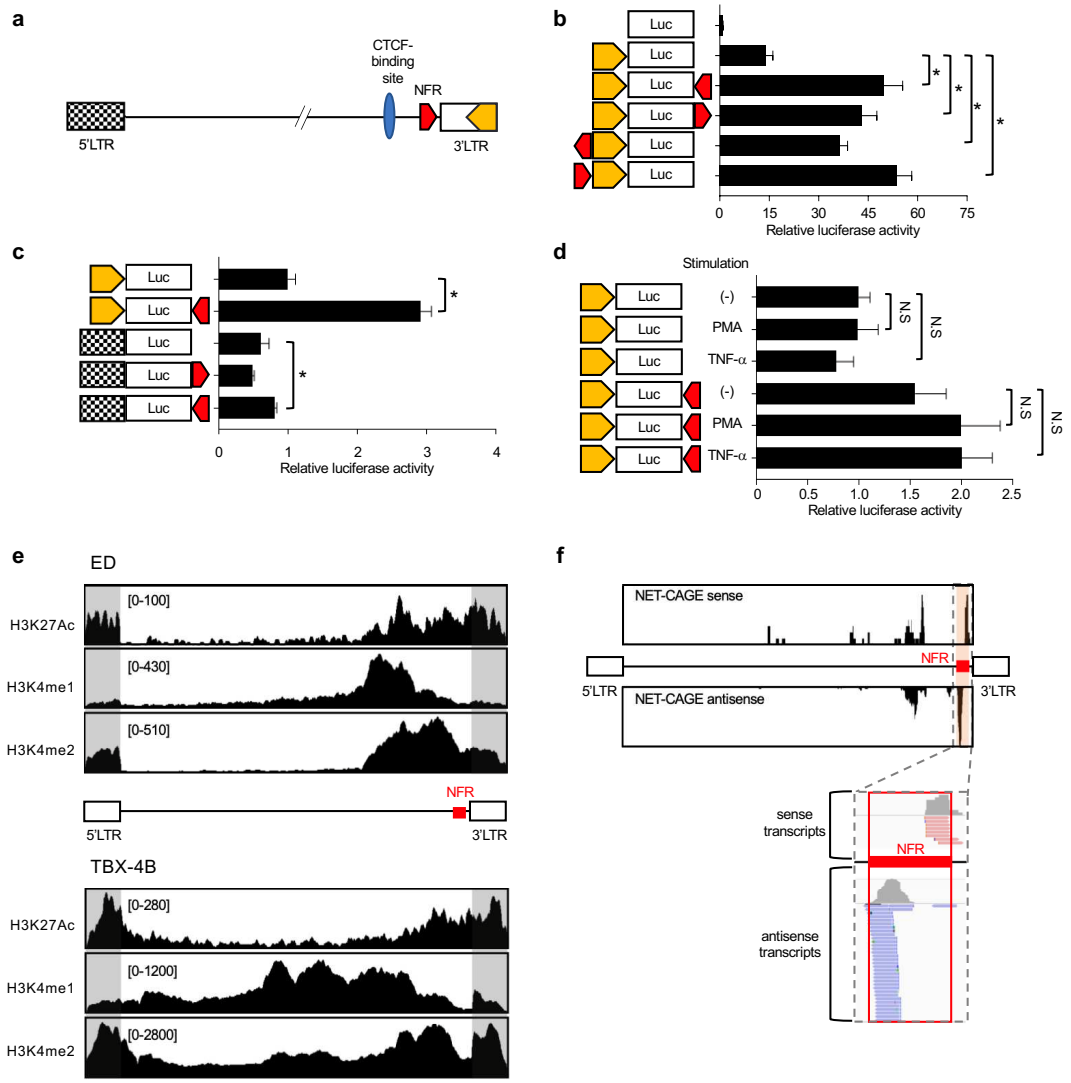


Fig. 2 The nucleosome region harbors enhancer-related histone modifications and produces enhancer RNAs.
a Schematic of the HTLV-1 provirus structure. The 5′LTR (black plaid), CTCF-binding site (blue) [11], NFR (red), and the HBZ promoter (yellow) [20] are shown.
b-d Transcriptional regulatory function of the NFR was analyzed by luciferase reporter assays in Jurkat cells. The HBZ promoter [20] (**b**, **c**) and 5′LTR (**c**) were used as a promoter. PMA and TNF-α were used for cell stimulation (**d**). Luciferase activity was normalized to Renilla activity. Representative data of three independent experiments is shown as fold change to pGL4-basic (**b**), pGL4-basic-HBZ promoter (**c**, **d**) (Student's t-test, *P < 0.05).
e H3K27Ac (top), H3K4me1 (middle), and H3K4me2 (bottom) occupancy within the provirus in ED (upper) and TBX-4B (lower) cells. ChIP-seq signals were visualized by IGV. Gray-shaded areas indicate the ChIP signal mapped to LTRs.
f NET-CAGE results using nuclear lysates of ED cells in the sense (above) or antisense (below) orientations, demonstrating eRNAs at the NFR. The bottom panel is an enlarged

image of the signals around the NFR. NET-CAGE signals were visualized by IGV. Luc, luciferase; NFR, nucleosome-free region; N.S., not significant.

The host transcription factors SRF and ELK-1 bind to the intragenic HTLV-1 enhancer

The NFR region we identified in this study is ~160 bp in length. We performed transcription factor binding prediction with the NFR sequence based on the consensus binding motif of various transcription factors and found several candidates (Figure 3a). We analyzed their binding to the NFR using highly sensitive ChIP-seq analysis with an HTLV-1 DNA-capture approach [14]. The results demonstrated that SRF and ELK-1 co-localized to the NFR of the HTLV-1 proviral DNA (Figure 3b). Since SRF is involved in the regulation of the 5′LTR [24], we also observed the SRF signal in the 5′LTR region in TBX-4B cells, in which *tax* expression is active. Most importantly, the binding of SRF and ELK-1 to the NFR was observed in PBMCs freshly isolated from HTLV-1-infected individuals, indicating that this molecular mechanism is actually ongoing *in vivo* in infected individuals.

Next, we performed electrophoretic mobility shift assays (EMSA) to investigate whether SRF and ELK-1 binding to the NFR depends on DNA sequence. We generated oligonucleotide probes for the NFR with a wild-type (WT) sequence (NFR-wt) and negative control probes targeting viral regions other than the NFR (Figure 3c). We observed a band shift when combining the NFR-wt probe and nuclear extract of 293T cells transfected with SRF and ELK-1 expression vectors (Figure 3c). Addition of either anti-SRF or anti-ELK-1 antibodies induced a band supershift, demonstrating the involvement of SRF and ELK-1 in the detected band (Figure 3c). We further generated oligonucleotide probes with mutations in the SRF and/or ELK-1 consensus binding sequence. Mutant 1 (mut1), mutant 2 (mut2), and mutant 3 (mut3) contain mutations in the SRF, ELK-1, or both SRF and ELK-1 binding sites, respectively (Figure 3d). To investigate whether the mutations alter transcription factor

binding to the NFR, we performed competition EMSA and found marked reduction in the binding activity of mutant probes to SRF and ELK-1 compared with that of the WT probe (Figure 3e). Because all mutants markedly decreased the formation of a ternary complex of SRF/ELK-1 on the NFR DNA, we used mut3 for subsequent experiments and found a remarkable reduction in the enhancer activity of the NFR after introducing the mutation (Figure 3f). These results demonstrate that SRF and ELK-1 binding to the NFR plays an indispensable role in enhancer activity.

Figure 3

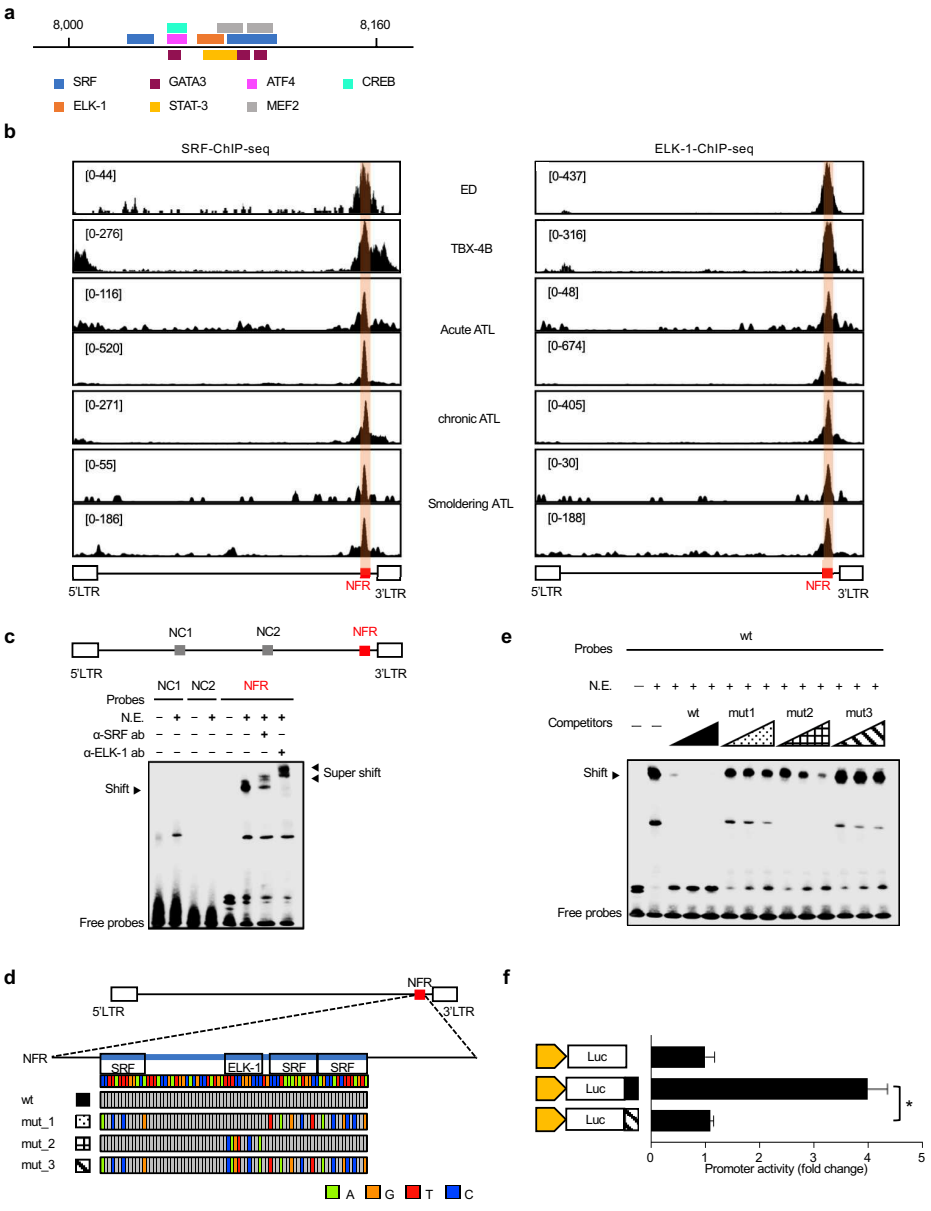


Fig. 3 SRF and ELK-1 bind to the NFR in a DNA sequence-dependent manner.
a The prediction of transcription factor binding to the NFR was performed by using TFBIND (http://tfbind.hgc.jp/) [25]. Candidate transcription factor binding sites are shown.
b The localization of SRF (left) and ELK-1 (right) to the NFR in cell lines and PBMCs of ATL patients. ChIP-seq signals were visualized by IGV. Orange-shaded region indicates the NFR.
c The binding ability of SRF and ELK-1 to the NFR oligonucleotides was analyzed by EMSA. Biotinylated DNA probes of 120 bp for the NFR (red) and negative control regions (gray) were incubated with nuclear extract of 293T cells transfected with SRF and ELK-1 expression vectors (N.E.). NFR-SRF/ELK-1 complexes and super-shifted complexes, which were detected with the anti-SRF and the anti-ELK-1 antibody, are indicated by arrowheads.
d The position of introduced mutations used are shown as green (A), orange (G), red (T) and blue (C).
e EMSA competition analysis with wt- and mutated-NFR oligonucleotides. Biotin-labeled wt-NFR probe and 100, 200, or 300 times more non-labeled competitor oligonucleotides were used in this assay. Each non-labeled competitor sequence is shown in (**d**) and is indicated wt as black, mut1 as dots pattern, mut2 as lattice pattern and mut3 as diagonal stripe pattern.
f Transcriptional regulatory function of the wt (black) or mut (pattern) NFRs was analyzed using the HBZ promoter (yellow) in Jurkat cells by luciferase assay. Luciferase activity was normalized to Renilla activity. Representative data of three independent experiments is shown as fold change to pGL4-basic-HBZ promoter (Student's t-test, *P < 0.05). ATL, adult T-cell leukemia/lymphoma; NFR, nucleosome-free region; wt, wild-type; mut, mutant; N.E., nuclear extract.

The SRF and ELK-1 plays a critical role in HTLV-1 enhancer function

Next, we investigated the functional role of SRF/ELK-1 binding to the NFR in the context of the whole viral sequence. As the NFR is located in the coding region of the *tax* gene, we generated mutations of the SRF/ELK-1 binding site without altering the amino acid sequence of the Tax protein. The nucleotides substitutions could change stability of mRNA and translational efficiency, but we confirmed that introduction of mut1, mut2, or mut3 did not change Tax protein levels (Figure S1a). We constructed HTLV-1 mutant molecular clones (HTLV-1-mut) containing the same mutations as mut3 (Figure 3d) and then transfected HTLV-1-wt or mut plasmids into 293T cells. After quantifying viral gene expression in the transiently transfected cells and viral production in the culture supernatant (Figure 4a), we found a marginal decrease of p19 production in the supernatant of mut plasmid-transfected cells; however, there was no statistically significant difference (Figure 4b). Nevertheless, there was a significant reduction in *tax* and *HBZ* expression at the mRNA level (Figure 4c). Next, we generated Jurkat T cells infected with HTLV-1-wt or mut by co-

culturing with the transfected 293T cells (Figure 4d). We used JET cells - Jurkat T cells stably transfected with a reporter plasmid to monitor Tax expression - as host cells. We sorted Tax-expressing cells 3 days after infection and then analyzed provirus sequences, proviral load, and the distribution of HTLV-1 integration sites (ISs) in the sorted bulk cell populations. We performed DNA sequencing of whole integrated provirus by DNA-capture-seq and confirmed that the proviral sequences of JET cells infected with HTLV-1-wt and mut were the same as the plasmid sequences used for transfection (Figure 4e). The proviral load of HTLV-1-mut-transfected JET cells was lower than that of HTLV-1-wt-transfected ones (Figure 4f). We next analyzed whether mutations in the SRF/ELK-1 binding site actually reduced SRF/ELK-1 binding to the NFR in the infected cells *in vivo*. We performed ChIP-seq analysis for SRF and ELK-1 and observed SRF/ELK-1 binding in wt-HTLV-1- infected JET cells but not in mutant virus-infected JET cells (Figure 4g). Viral IS analysis demonstrated that there were hundreds of different ISs in each JET cell infected with HTLV-1-wt or mut (Figure 4h). Distribution of viral IS was not so different between the WT and mutant HTLV-1-infected JET cells in terms of the relationship with the host gene and epigenetic environment (Figure. S2a and S2b). We then evaluated expression levels of *tax* and *HBZ* in JET cells infected with wt- or mut-HTLV-1 and found that infected cells with mut-HTLV-1 showed a significant reduction in *tax* and *HBZ* expression ($p < 0.05$; Figure 4i). Taking into consideration that there was a similar distribution of ISs between wt-HTLV-1 and mut-HTLV-1 infected cells, their different proviral expression was thus due to the mutation introduced in the NFR of the HTLV-1 provirus and not due to a different distribution of HTLV-1 ISs. HBZ was previously reported to confer antiapoptotic phenotypes to Jurkat T cells [26]; therefore, we analyzed susceptibility to apoptosis induced by T-cell activation and found that JET cells infected with HTLV-1-mut were more susceptible to activation-induced T-cell death than those infected with HTLV-1-wt (Figure 4j). We further analyzed the effect

of mutations in the NFR on chromatin status and found that the mutations induced a decrease in the chromatin openness of the NFR (Figure 4k). These findings demonstrate that SRF and ELK-1 binding to the enhancer plays a critical role in the enhancer function.

Figure 4

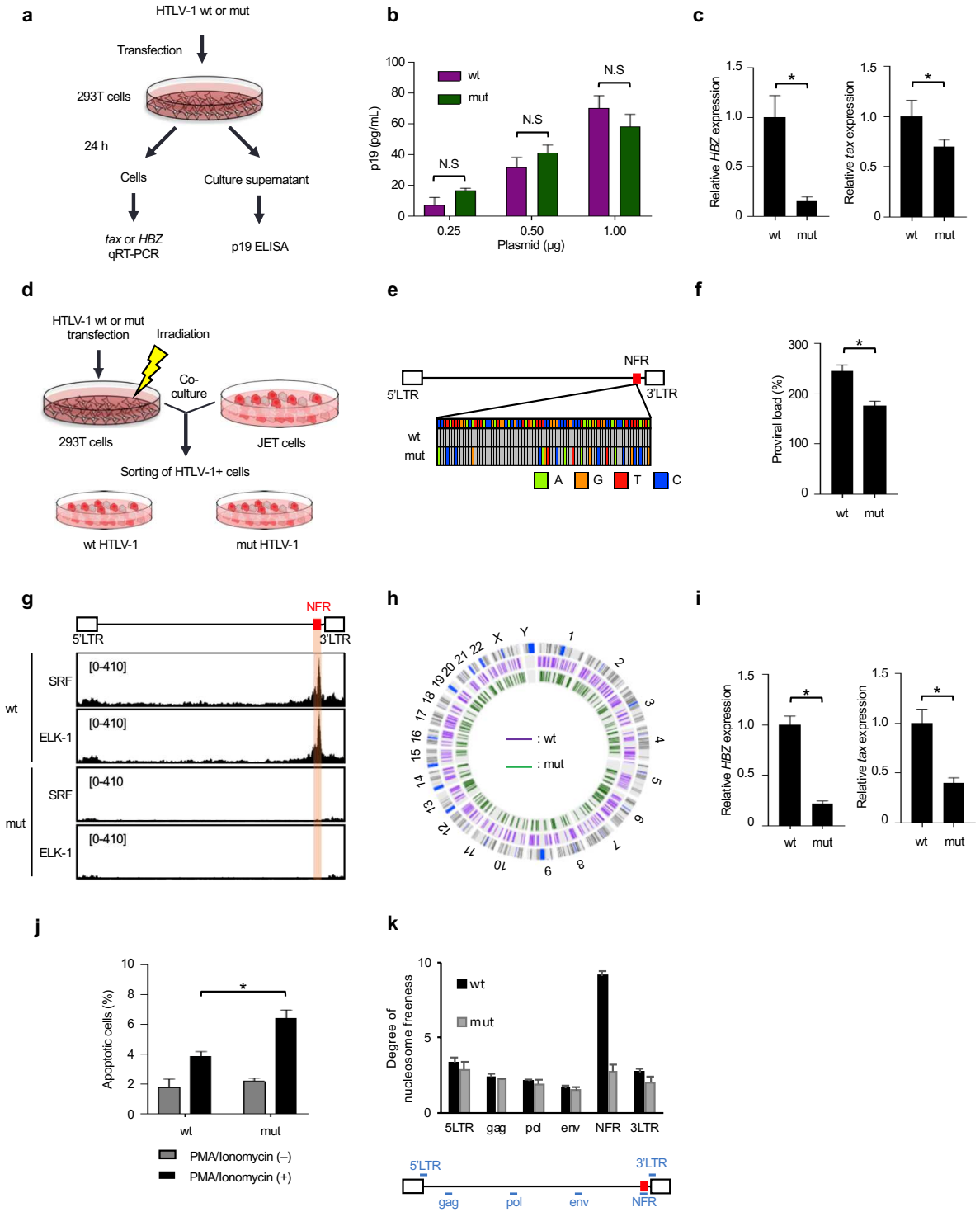


Fig. 4 Generation and characterization of the HTLV-1 infectious clone with mutations in the SRF/ELK-1 binding site.
a Diagram illustrating the experimental workflow of transient transfection evaluation using HTLV-1-wt or mut molecular clones.
b p19 was quantified for analyzing infectivity by ELISA using the supernatant of 293T cells transfected with HTLV-1-wt or mut (0.25, 0.5, 1 μg). Results are expressed as the mean ± SEM of three experiments performed in duplicate.
c qRT-PCR results of HBZ (left) and tax (right) levels after transient transfection using HTLV-1-wt or mut. 18S rRNA was quantified as an internal control. Data shown are representative of two independent experiments (Student's t-test, *P < 0.05).
d Experimental workflow of establishing stable cells infected with wt or mut HTLV-1.
e The HTLV-1 NFR sequences of JET cells infected with HTLV-1-wt or mut was analyzed by HTLV-1 DNA-capture-seq. Each detected mutation is indicated green as A, orange as G, red as T and blue as C.
f Proviral load in JET cells infected with HTLV-1-wt or mut was measured by digital droplet PCR. Values reported were derived from three biological replicates (Student's t-test, *P < 0.05).
g Localization of SRF and ELK-1 in cells infected with HTLV-1-wt (above) or mut (below). SRF and ELK-1 ChIP-seq signals were visualized by IGV. Orange-shaded area represents the NFR location.
h The distribution of integration sites (ISs) in bulk JET cells infected with HTLV-1_wt (purple line) or mut (green line) is shown in a circus plot. Outer gray ring represents human chromosomes, cytogenetic bands shown as black lines and the centromere shown as blue line. Numbers and letters surrounding the circos represent human chromosomes.
i Representative qRT-PCR results of HBZ (left) and tax (right) levels in JET cells infected with HTLV-1-wt or mut. 18S rRNA was quantified as an internal control. Data shown are representative of two independent experiments (Student's t-test, *P < 0.05).
j Cell apoptosis was detected by Annexin V staining after stimulation with PMA and ionomycin. Values reported were derived from two biological replicates (Student's t-test, *P < 0.05).
k MNase assay of in JET cells infected with HTLV-1-wt (above) or mut (below). Degree of open chromatin was evaluated by MNase treatment and ddPCR. Values of indicated proviral regions are shown after normalization to non-MNase digestion sample value. NFR, nucleosome-free region; wt, wild-type; mut, mutant.

The intragenic viral enhancer induces upregulation of host genome transcription near the viral IS

The presence of an intragenic viral enhancer in the HTLV-1 provirus raises the possibility that it acts as an ectopic enhancer to activate transcription in host cellular genomic DNA, resulting in changes in host gene expression near the viral IS. To investigate the effect of HTLV-1 integration on host gene expression near the ISs, we cloned JET cells infected with wt- or mut-HTLV-1 by limiting dilution from bulk cell populations (Figure 4d) and established five clones infected with HTLV-1-wt with one to four proviruses per clone (Figure 5a). We also established four clones infected with mut-HTLV-1 containing one to two proviruses per clone. The characteristics of each individual clone are listed in Table S1.

We then performed RNA-seq analysis using these clones and found read-through transcripts around the IS of the JET wt-HTLV-1-infected clone (Figure 5b) but not in the mutant infected clones (Figure 5c). We further tested whether an ectopic enhancer inserted by the HTLV-1 would alter host gene expression near ISs. The proportion of upregulated genes in JET clones infected with HTLV-1-wt was significantly higher than those in mutant HTLV-1 clones ($P < 0.01$; Figure 5d). It has been reported that viral CTCF plays a role in chromatin looping with the host CTCF-binding site and induces changes in host gene transcription [27]. Thus, we also analyzed CTCF binding to the host gene near ISs and found a high frequency of CTCF-binding sites in upregulated host genes (Figure 5d). We then used CRISPR/Cas9 to introduce the mutation that abrogated SRF-ELK-1 binding to the enhancer region (Figure 3d and 3e, Figure 4g) of a clone infected with wt-HTLV-1. SRF/ELK-1 ChIP-seq peaks in wt-HTLV-1-infected cells were abolished in the CRISPR-mutated cells, thereby reducing proviral transcription both at sense and antisense direction (Figure 5e and 5f) and chromatin openness in the enhancer region (Figure 5g). We analyzed the whole proviral sequence of the wt- and CRISPR-mutated clones by HTLV-1 DNA-capture and observed the expected mutations in the enhancer in mutant clones, while other regions were identical between the wt and mut clones, demonstrating that SRF and ELK-1 play an important role in the enhancer function.

Figure 5

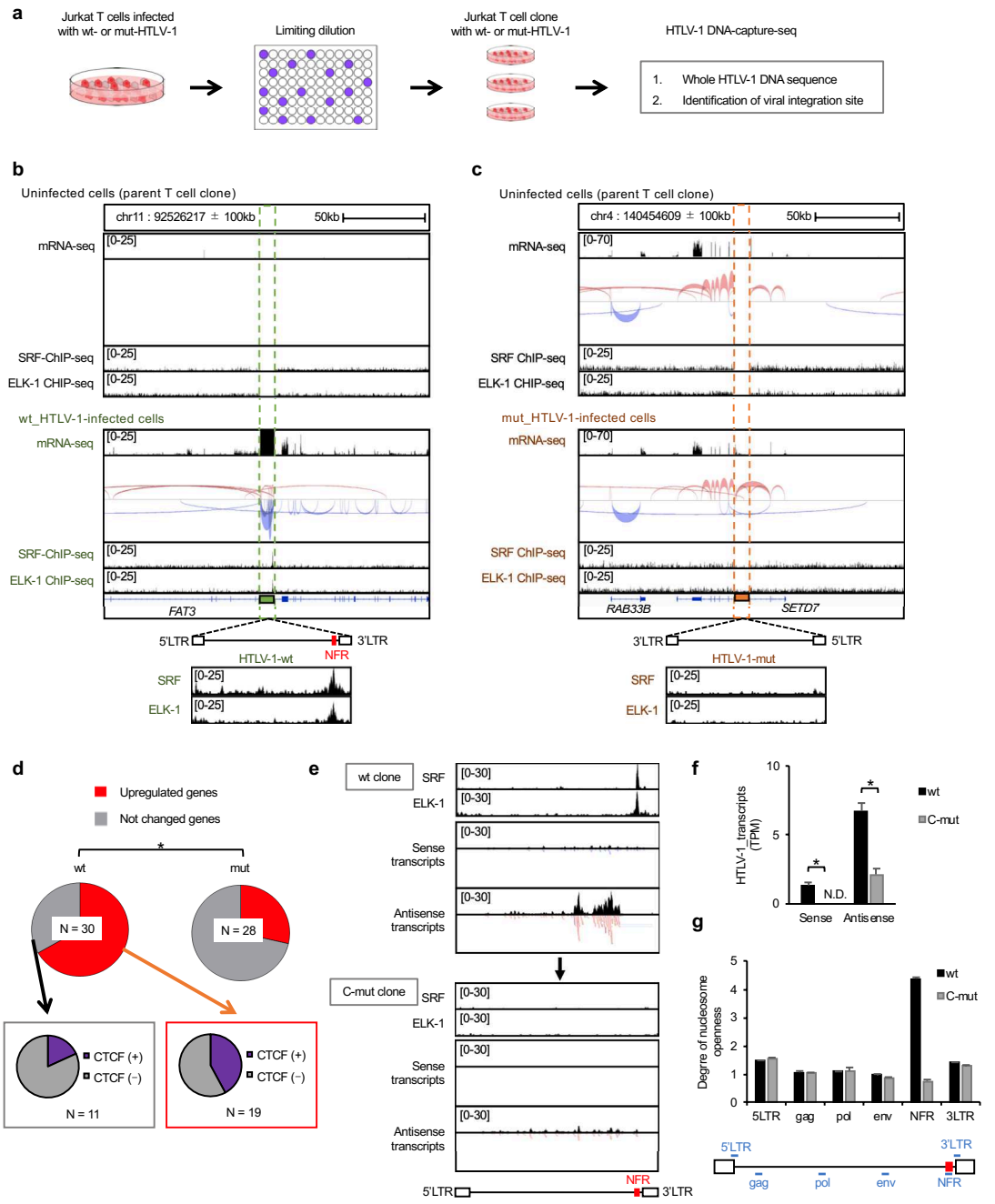


Fig. 5 Establishment and characterization of Jurkat T cell clones infected with wt- and mut-HTLV-1.

a Experimental workflow of establishing infected clones with wt or mut HTLV-1 by limiting dilution.

b, **c** Local transcriptome and splice junction near viral integration site are visualized in a JET clone infected with HTLV-1-wt (**b**) and mut (**c**) by IGV. The splice junctions are shown in red as sense direction and blue as antisense direction. The thickness of red and blue line is indicated the frequency of detection of specific splices. Host genes near the IS and direction of HTLV-1 provirus is indicated below the graph. SRF/ELK-1 ChIP-seq results are also shown in each ATL clone.

d The fraction of upregulated genes in JET clones infected with HTLV-1-wt (above left) and mut (above right). Presence or absence of CTCF ChIP-seq signals in the upregulated group (below right) or 'no change' group (below left) of JET clones infected with HTLV-1-wt. Chi-square test, $P < 0.05$.
e mRNA-seq and SRF/ELK-1 ChIP-seq results of JET cells infected with HTLV-1-wt (above) and CRISPR-mutated HTLV-1 (below). Representative results from two independent experiment are visualized by IGV.
f Level of proviral expression of JET cells infected with HTLV-1-wt or CRISPR-mutated (C-mut) clone. Data are generated as transcripts per million reads (TPM) from two independent mRNA-seq analyses.
g MNase assay of JET cells infected with wild type or CRISPR-mutated HTLV-1. Degree of open chromatin was evaluated by MNase treatment and ddPCR. Values of indicated proviral regions are shown after normalization to non-MNase digestion sample value. NFR, nucleosome-free region; wt, wild-type; mut, mutant; C-mut, CRISPR-mutant.

SRF and ELK-1 localization to the enhancer and aberrant host genome transcription near the proviral integration site in fresh PBMCs

We further investigated the effect of HTLV-1 integration on viral and host genomes by performing mRNA-seq analysis using freshly isolated PBMCs from five ATL cases. All five cases had a high proviral load (Figure S3a) and had a clonally expanded ATL clone (Figure S3b). Consistent with previous reports [9,28], proviral expression in the sense orientation was lower than that in the antisense orientation (Figure 6a). There was read-through proviral transcription in the sample with HTLV-1 ISs in the host genomic region but not in other samples without HTLV-1 ISs (Figure 6b and 6c), as previously reported [29]. Interestingly, an ATL case with a defective provirus lacking the 5'LTR, also exhibited readthrough transcription from the virus to the flanking host genome (Figure 6d). More importantly, there were clear peaks of SRF and ELK-1 ChIP-seq signals in integrated proviruses, indicating SRF and ELK-1 play a role in the transcriptional regulation (Figure 6b–6d). PBMCs contain not only ATL cells but also non-ATL infected T cells, uninfected T cells and various non-T cells; thus, the mRNA-seq data shown in Figure 6b–6d are derived from the sum of all PBMC subsets. To see the effect of HTLV-1 ISs on the host genome with high accuracy at single-cell resolution, we performed single-cell RNA-seq analysis using PBMCs from five ATL cases including the same ATL case as in Figure 6b and 6c, and in ATL cases containing

375 defective proviruses. Based on the T-cell receptor (TCR) clonotype and transcriptome data,
376 we performed clustering analysis and found that the ATL clones, which were identified by
377 the T-cell receptor (TCR) clonotype, clustered differently from the other $CD4^+$ T cell clones
378 (Figure 6e and 6f). We then compared the transcriptome near viral IS of $CD4^+$ T cells among
379 five ATL cases. There was remarkable upregulation of the local transcriptome only in the
380 sample with viral integration (Figure 6g and 6h, Figure S4a-4c, left panels). Furthermore,
381 there was a significant increase of the local transcriptome in the ATL clone but not in non-
382 ATL $CD4^+$ T cell clones (Figure 6g and 6h, Figure S4a-4c, right panels). These data support
383 the idea that the intragenic viral enhancer we identified in this study plays a role in persistent
384 proviral expression and aberrant transcription of the integrated host genome by recruiting
385 SRF and ELK-1.

Figure 6

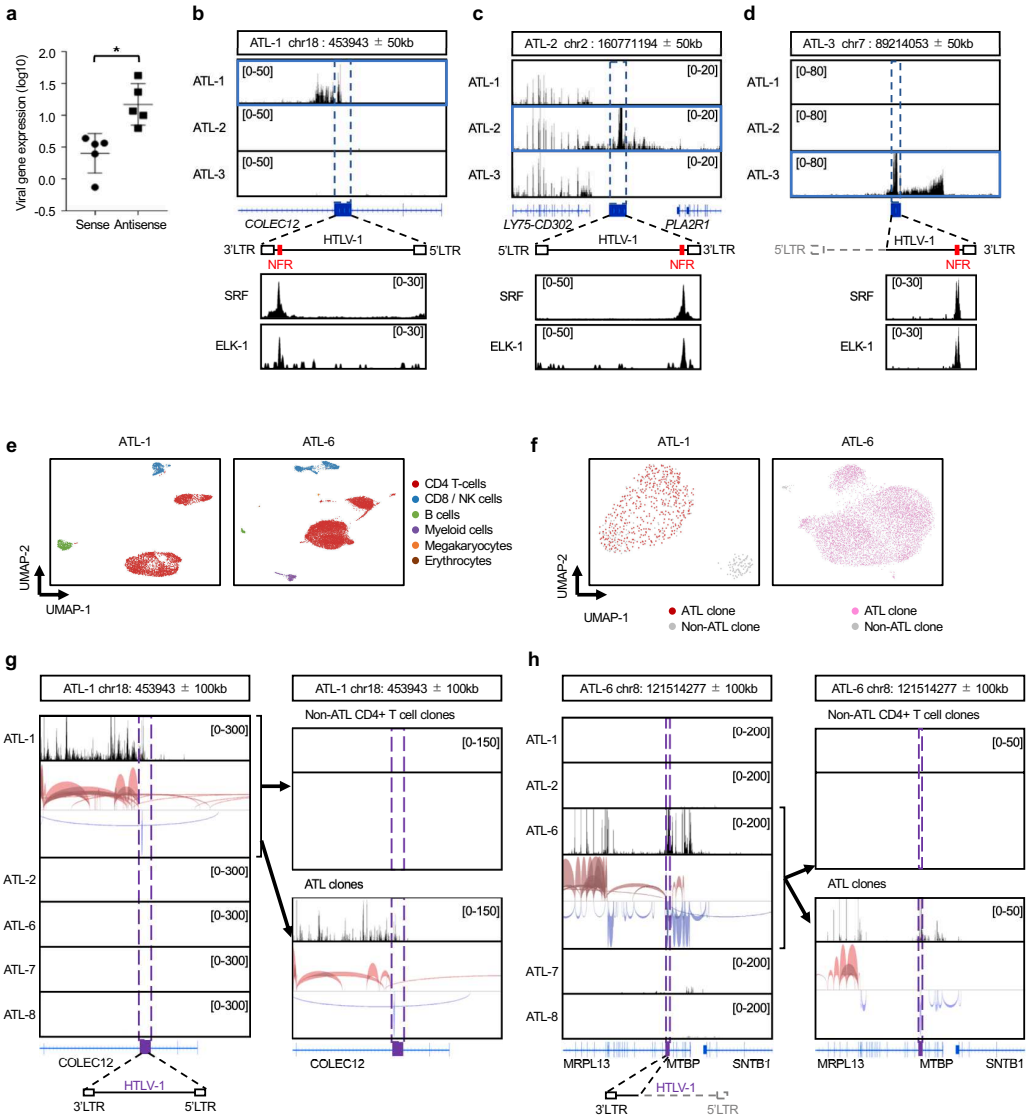


Fig. 6 Transcriptional characterization of the provirus and the flanking host genomes in freshly isolated PBMCs from infected individuals.

a The level of sense or antisense proviral expression in fresh PBMCs from five ATL patient samples. Data shown are transcript per million for each case.

b-d Visualization of mRNA-seq data of three ATL cases at around each viral IS. Host genes near the IS and direction of HTLV-1 provirus is indicated below the graph. SRF/ELK-1 ChIP-seq results are also shown in each ATL clone. The ATL sample with HTLV-1 IS in the region is highlighted with a blue square.

e scRNA-data of PBMCs from the indicated ATL cases. Cell clustering analysis was performed with a nonlinear dimensionality reduction method, uniform manifold approximation and projection (UMAP). Each cell cluster was annotated by expression pattern of marker gene for PBMC subsets.

f We defined ATL cells as T cells containing the most abundant TCR. ATL clones are shown in red (ATL-1) and pink (ATL-6).

g, **h** Local transcriptome including viral integration site are visualized by IGV. We obtained scRNA-seq data from five ATL cases. The data shown were region with viral IS of ATL-1(**g**) and ATL-6 (**h**), respectively. Data from CD4+ T cells are shown in the left panel. CD4+ T cells are further divided into non-ATL cells (right, upper panel) and ATL cells (right, lower panel).

Discussion

The size of the HTLV-1 genome is just over 9,000 bp. To achieve persistent infection in the host, HTLV-1 encodes several viral genes by alternative splicing in its small genome. In addition, the provirus is transcribed from both the 3'LTR and the 5'LTR [20,30,31]. It has been reported that antisense transcription is frequently expressed *in vivo*, whereas sense transcription is typically silenced or expressed only intermittently [9,17,32,33]. It has not been understood how HTLV-1 antisense transcription remains selectively active. In the present study, we demonstrated the presence of a previously uncharacterized viral enhancer in the HTLV-1 pX region, exploiting the high efficiency and resolution of the viral DNA-capture-seq approach. The enhancer we identified here is located at the 3′ side of the insulator region in the provirus (Figure 2a). We suggest that the internal enhancer region near the 3′LTR may have two distinct functions: first, to drive the frequent antisense transcription from the 3'LTR (Figure 5e and 5f), and second, to cooperate with the viral insulator to inhibit the spread of heterochromatin from the 5′LTR towards the 3′LTR. The antisense transcript HBZ plays an indispensable role in viral persistence [9,34] and therefore the intragenic viral enhancer would also contribute to viral persistence via HBZ upregulation. Consistent with this notion, the intragenic viral enhancer and insulator are maintained even in defective type proviruses that is observed 20-30% of ATL cells [15,35,36].

There are several thousand different HTLV-1-infected T cell clones in an infected individual. After long-term clinical latency, a specific clone may undergo malignant transformation, causing the syndrome of ATL. A key question that remains is how a certain clone is selected as an ATL clone from various infected clones. Previous reports demonstrated that the HTLV-1 provirus tends to integrate near cancer-related genes in ATL cells [29,37], indicating that aberrant host genome transcription by viral integration may

contribute to the multistep oncogenic process induced by HTLV-1 infection. As previously reported, HTLV-1 contains CTCF-binding sites and therefore viral integration generates an ectopic CTCF-binding site in the host genome, which induces deregulation of host gene transcription via chromatin looping [27]. We demonstrate here that HTLV-1 generates an ectopic enhancer region together with CTCF-binding site in the host genome. These findings indicate that HTLV-1 induces a distinct type of alteration of the host transcriptome via chromatin looping, and thereby upregulates cancer-related genes near ISs and might contributes to the selection of a specific infected cell for clonal expansion during the early phase of leukemogenesis.

Mobile DNA elements, including endogenous retroviruses or foreign DNA elements introduced by exogenous retroviruses, can be dangerous for the host cell because they disturb cellular genomic homeostasis. Mammalian cells have an evolutionally acquired host defense system that silences such elements in genomic DNA. For example, murine leukemia virus (MLV) is silenced by Trim28—a well-characterized transcriptional co-repressor [38]—and ZFP809 to prevent further viral spread in embryonic stem cells [13]. Although little is known regarding the precise molecular mechanisms behind silencing of the HTLV-1 provirus in the host genome, the HTLV-1 5′LTR is frequently silenced by DNA methylation or histone modifications [39] or transcribed only intermittently [32,33]. This suggests that a host defense mechanism plays a role in selecting viral infected clones with silenced HTLV-1 proviral DNA. As a result, there is no detectable viremia in the serum of HTLV-1-infected individuals. However, the virus maintains the ability to re-activate viral transcription when the virus needs to induce *de novo* infection from an infected host to an uninfected host. We showed here that HTLV-1 recruits the host transcription factors SRF and ELK-1 to an NFR in proviral DNA to sustain chromatin openness and proviral transcription in host cells. The

molecular mechanism should enable the virus to be latent but maintain an ability to reactivate viral expression when infected cells need to induce *de novo* infection from the infected to uninfected host.

HTLV-1 has co-existed with humans for the past 20,000–30,000 years [40]. The virus may have evolved this strategy—presence of an internal insulator and enhancer region in the provirus—to achieve persistent infection under pressure from the host system to silence foreign DNA elements as well as from the host immune response. Usage of lentiviral/retroviral vectors for gene therapy or for generation of induced pluripotent stem cells (iPS) has been under intense research and development [41]. Lentiviral and retroviral vectors integrate into host genomic DNA and form a provirus in the target cells; however, the provirus tends to be silenced by the host defense mechanism as described above. Various efforts have been made to optimize the lentiviral and retroviral vector to prevent silencing of integrated provirus, such as the introduction of insulator or enhancer elements [41,42]. Retrovirus vector insertion can trigger deregulated cell proliferation, most likely driven by retrovirus enhancer activity on cancer-related genes [43]. It is surprising that an exogeneous virus HTLV-1 has by itself evolved a similar system, obtaining an insulator, an enhancer, and a chromatin-opening element in the retroviral genome. This experiment of nature may provide insights into how an exogenous retrovirus achieves persistent infection in humans and also how to tackle the silencing of foreign DNA elements to maintain chromatin openness and transgene transcription without causing transformation of host cells.

In conclusion, we have analyzed HTLV-1 provirus integrated in the host genome with high resolution and efficiency using HTLV-1-DNA-capture sequencing approach and discovered internal viral enhancer in HTLV-1 genome. This finding provides clues to help solve several

long-lasting questions related with HTLV-1 persistence and pathogenesis. Viral DNA-capture-seq approaches can be applied to studies aiming to understand transcriptional regulatory mechanism of other oncogenic viruses integrated into the host cellular genomic DNA.

Methods

Ethics statement.

All protocols involving human subjects were reviewed and approved by the Kumamoto University Institutional Review Board (approval number 263). The study was carried out in accordance with the guidelines proposed in the Declaration of Helsinki. Informed written consent was obtained from all subjects in this study.

Cell culture.

ED, 293T, Jurkat, and JET cells infected WT or mutant HTLV-1 molecular clones were cultured in RPMI-1640 medium (Thermo Fisher Scientific, Waltham, MA) supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin, and 100 μg/mL streptomycin. TBX-4B cells were cultured in RPMI-1640 supplemented with 20% FBS, interleukin-2 (200 U/mL; Cosmo Bio Co., Ltd., Tokyo, Japan), 100 U/mL penicillin, and 100 μg/mL streptomycin.

Generation of reporter constructs.

The HBZ promoter, 3′LTR300[20], and 5′LTR were amplified from ED cells. The NFR was amplified from ED cells and the NFR mutant was generated by gBlocks ® Gene Fragments (Integrated DNA Technologies, Coralville, IA). Using XhoI and HindIII restriction sites, each promoter construct was inserted into pGL4-basic (Promega, Madison, WI) which

includes the luciferase reporter gene. The NFR was inserted into pGL4-3′LTR300 or pGL4-5′LTR using BamHI or KpnI restriction sites while the NFR mutant was inserted into pGL4-3′LTR300 using the BamHI restriction site. Primers associated with each construct and the NFR mutant are listed in Table S2.

mRNA-seq and qRT-PCR.

RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and treated with DNase. For mRNA-seq, mRNA libraries were prepared using NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® Multiplex Oligos for Illumina (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions. Libraries were run as 75-cycle-single end reads on a NextSeq 550 (Illumina, San Diego, CA) using a high-output flow cell. cDNA was synthesized using ReverTra Ace® qPCR RT Master Mix (Toyobo, Osaka, Japan) according to the manufacturer's instructions. qPCR was performed using Thunderbird SYBR qPCR mix (Toyobo) and run on an Applied Biosystems® StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific); primers used are listed in Table S3.

Preparation and cultivation of HTLV-1-infected cells *in vitro.*

293T cells were transfected with a wt or enhancer-mutated HTLV-1 molecular clone[44] by polyethylenimine (PEI) and then irradiated with 30 Gy. The irradiated 293T cells were co-cultured with JET cells for 3 days[45], after which tdTomato-positive cells were sorted by FACS Aria™ (Becton, Dickinson and Company, Franklin Lakes, NJ), and cultured in RPMI supplemented with 10% FBS, 100 U/mL penicillin, and 100 μg/mL streptomycin for 2 weeks.

Proviral load (PVL) measurement.

We estimated the number of infected cells by quantifying the copy number of the *tax* gene normalized to the copy number of the *ALB* gene by using digital droplet PCR as previously described but with minor modifications [15]. PVL was calculated as follows, PVL (%) = [(copy number of *tax*)/(copy number of albumin)/2] × 100. Primer sequences are listed in Table S3

HTLV-1 DNA-capture seq.

HTLV-1 DNA-capture-seq was performed as previously described [15] with minor modifications. Briefly, 1 μg genomic DNA was fragmented by sonication using a Picoruptor (Diagenode s.a., Liège, Belgium) to produce 300–500-bp fragments. The DNA library was generated using an NEBNext Ultra II DNA Library Prep Kit for Illumina and Multiplex Oligos for Illumina (New England Biolabs). DNA-seq libraries were used for HTLV-1 sequence enrichment with HTLV-1 specific probes, after which enriched libraries were amplified by additional PCR. Enriched libraries were quantified using P5P7 primers and then sequenced via Illumina MiSeq or NextSeq.

MNase assay and MNase-seq.

Cells (1.0×10^6 for cell lines or 2.0×10^6 for patient PBMCs) were lysed using cell lysis buffer (0.05% Triton X-100, 2 mM PMSF, 5 mM sodium butyrate, 100× protease inhibitor cocktail) or PBMC lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Nonidet-P40). Extracted nuclei were digested by MNase (TaKaRa Bio, Kusatsu, Japan) for 5–20 minutes at 37 °C after which the reaction was stopped by the addition of 20 mM ethylenediaminetetraacetic acid (EDTA). After deproteination with proteinase K solution (Nacalai Tesque, Kyoto, Japan), MNase digestion samples were purified using a PCR Purification Kit (Qiagen). MNase-seq libraries were prepared by the NEBNext Ultra II DNA

Library Prep Kit for Illumina and Multiplex Oligos for Illumina (New England Biolabs), after which the efficiency was quantified using P5P7 primers and then sequenced using Illumina MiSeq. Also, MNase digestion sample and input sample were measured by QX200 droplet digital PCR system (BIO-RAD, Hercules, CA). Primer sequences are listed in Table S5.

ChIP-seq.

ChIP assays were performed using the SimpleChIP® Enzymatic Chromatin IP Kit (Cell Signaling Technology, Danvers, MA) according to manufacturer's instructions. Briefly, cells (4×10^6) were fixed in 1% formaldehyde for 10 min at room temperature, quenched in glycine solution, and washed in ice-cold PBS. Nuclei were extracted by lysis buffer (buffer A) and then samples were digested by MNase for 20 min at 37 °C and sonicated for 30 s on and 30 s off for 5–8 minutes using Bioruptor UCD-300 (Cosmo Bio Co., Ltd.) to break the nuclear membrane. Extracted chromatin was immunoprecipitated using anti-H3K27Ac (#07-360; Millipore, Burlington, MA), H3K4me1 (#ab8895; Abcam, UK, England), H3K4me2 (#ab7766, Abcam), SRF (#5147; Cell Signaling Technology), and ELK-1 (#ab32106; Abcam) antibodies. ChIP sample libraries were prepared by NEBNext Ultra II DNA Library Prep Kit for Illumina and Multiplex Oligos for Illumina (New England Biolabs), after which the efficiency was quantified using P5P7 primers and sequenced using Illumina MiSeq or NextSeq.

Luciferase reporter assays.

Jurkat cells were harvested 24 h after transfection with each reporter construct, using Turbofect Transfection Reagent (Thermo Fisher Scientific). Luciferase assays were then performed using the Dual-Glo Luciferase Assay System (Promega) according to the

manufacturer's instructions, and luminescence was detected using GloMax ® 20/20 Luminometer (Promega).

NET-CAGE.

Nascent RNAs were extracted from the nuclei of ED cells following previously described [23]. NET-CAGE libraries were generated using CAGE library preparation kit (K.K. DNAFORM, Yokohama, Japan) according to manufacturer's instructions. Briefly, cDNA was synthesized from 5μg nascent RNAs. The 5'cap-structures of nascent RNAs were labeled by biotin for the cap-trapping step. After removing Remaining RNA fragments without 5'cap structure by RNaseONE enzyme, enriched cDNA by cap-trapping was used for linker ligation and library generation. NET-CAGE Libraries were quantified by qPCR and sequenced using Illumina NextSeq.

EMSA.

293T cells (2×10^6) were harvested 24 h after transfection with pcDNA3-myc-SRF[46] and pCGN-ELK-1 (Addgene, Watertown, MA) using Turbofect Transfection Reagent (Thermo Fisher Scientific). After cell lysis in cell lysis buffer (10 mM Tris-HCl pH 8.0, 60 mM KCl, 1 mM EDTA, 1 mM DTT, 100 μM PMSF, 0.1% NP-40), nuclear lysates were extracted in nuclear extraction buffer (20 mM Tris-HCl pH 8.0, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25 mM glycerol). EMSA was then performed with the extracted nuclear lysates, biotin-labeled NFR-wt probe, and NFR-wt or mutant cold probes using Perfect NT Gel which is a 3-12% gradient polyacrylamide gel (#NTH-5X5HP; DRC, Tokyo, Japan) and the LightShift Chemiluminescent EMSA Kit (#20148; Thermo Fisher Scientific) according to the manufacturer's instructions. Nuclear lysates were mixed with 50 fmol biotin-labeled probes and 1 μg each of the SRF (#2185; Cell Signaling Technology) and ELK-1 (#ab32106;

Abcam) antibodies. For the competition assay, NFR-wt or mutant cold probes were added in the mixture of nuclear lysates and biotin-labeled NFR-wt probes. Probe sequences are listed in Table S4.

p19 ELISA.

293T cells (2×10^5) were transfected with HTLV-1-wt or mut molecular clone using HilyMax (Dojindo Laboratories, Kumamoto, Japan). After 24 h, the supernatants were collected and measured p19 presence by RETROtek HTLV p19 Antigen ELISA (ZeptoMetrix Corporation, Baffalo, NY) following the manufacturer's instruction.

Apoptosis analysis.

JET cells infected with HTLV-1-wt or mut molecular clone were stimulated with 100 ng/ml PMA and 2 μM Ionomycin and incubated for 24 h. After incubation, apoptotic cells were stained with annexin V by MEBCYTO® Apoptosis Kit (MBL, Nagoya, Japan) and detected by flow cytometry using BD FACSVerse™ (Becton, Dickinson and Company). Flow cytometry data was analyzed using FlowJo ™ (Becton, Dickinson and Company).

CRISPR/Cas9 mutagenesis.

Guide sequences were designed with the both edge of NFR in target and cloned into the pX330-U6-Chimeric BB-CBh-hSpCas9 plasmid (pX330; Addgene, 42230) as previously described [47]. The oligonucleotides for constructing guide sequence were listed in Table S6. Wt-HTLV-1 infected JET clone was co-transfected with two pX330 plasmid for each NFR edge, an expression vector with puromycin resistance gene and mut-enhancer cassette plasmid for HDR by electroporation using NEPA21 (NEPAGENE, Ichikawa, Japan). After puromycin selection, limiting dilution was performed to get single clone. CRISPR/Cas9

mediated mutant clone was confirmed the sequence which converted wt to mut by Sanger sequencing.

Single cell TCR (scTCR) analysis.

scTCR libraries were prepared on the 10x Genomics platform using Chromium™ Single Cell V (D) J Enrichment Kit, Human T cell and Chromium instrument (10x Genomics, Pleasanton, CA) according to the manufacture's protocol. Libraries were sequenced by Illumina HiSeq to obtain paired end reads using the following read length: read1_150 bp; read2_150 bp. The scTCR dataset was analyzed using Cell Ranger (10x Genomics).

Single cell RNA (scRNA)-seq.

scRNA libraries were prepared on the 10X Genomics platform using Chromium™ Single Cell 5' v2 Reagent Kit and Chromium instrument (10x Genomics) according to the manufacture's protocol. Libraries were sequenced by Illumina HiSeq to obtain paired end reads using the following read length: read1_26 bp; read2_91 bp. The scTCR dataset was mapped against the human hg19 reference genome analyzed using Cell Ranger (10x Genomics).

Bioinformatic analysis.

The peak detection in ChIP-seq analysis was performed as described previously[11]. Viral integration site and clonal abundance analysis was performed with HTLV-1-DNA-seq data as we reported [15]. RefSeq gene data was obtained from UCSC tables (https://genome.ucsc.edu/). Relationship between viral integration site and host genes or epigenetic microenvironment were analyzed using the R package hiAnnotator (http://github.com/malnirav/hiAnnotator) as described previously [48].

Statistical analysis.

Data were analyzed using a chi-squared test with GraphPad Prism 7 software (GraphPad Software Inc., La Jolla, CA) unless otherwise stated. Statistical significance was defined as $P < 0.05$.

Acknowledgements

We would like to thank M. Miura for providing R program to perform quality checks of the index reads using R program, M. Nakao for providing SRF expression vector, N. Misawa, S. Nagaoka and K. Sato for technical support and valuable discussion. We are also grateful to CRM. Bangham and S. Hino for their critical reading of the manuscript. This study was supported by grants from the Japan Society for the Promotion of Science (JSPS) KAKENHI (JP20H03724, and JP18KK0230 to Y.S., 16KK0206 and JP18K16122 to H.K., JP18K08437 and JP18KK0452 to PM) and Japan Agency for Medical Research and Development (AMED) (JP20jm0210074, JP20wm0325015, JP19fm0208012 and JP20fk0410023 to Y.S.) the Grant for Joint Research Project of the Institute of Medical Science, the University of Tokyo to Y.S., the grant from Kumamoto University Excellent Research Projects to YS, JST MIRAI to Y.S., and Kumamoto University Fellowship for Excellent Graduate Students to M.M.. The funders had no role in study design, data collection, data interpretation, or the discussion regarding submission for publication.

Additional information

Author contributions:

Study conception, YS; Methodology and Formal analysis, MM, TU, KM, BJYT, PM, JF and YS; Investigation; MM, TU, KM, BJYT, PM, KU, SI, HK, and SN; Data curation, MM, TU,

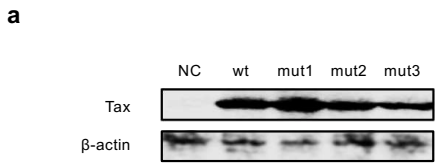
JF and YS; Resources, MT, KN, and AU, Writing/original draft, MM and YS; Writing/review and editing, MM, TU, KM, BJYT, PM, KU, SI, HK, SN, MT, KN, HH, AU, JF and YS; Supervision, HH, AU, JF and YS; Project administration and Funding acquisition PM, HK, and YS.

Data availability: Data on the findings reported here are available from the corresponding author upon request. We are preparing data deposition in the DNA Data Bank of Japan (DDBJ; accession no. DRA) regarding FASTQ files generated during this study will be open when the manuscript is published.

Competing financial interests: The authors declare no competing financial or nonfinancial interests.

Supplementary Figures

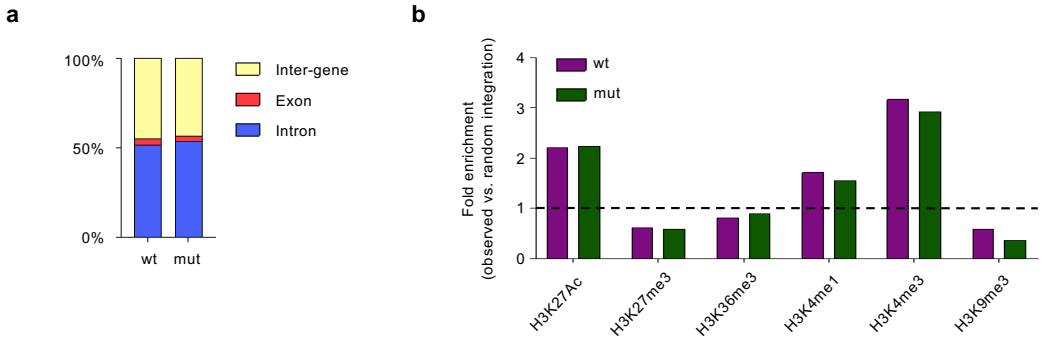
Figure S1



Supplementary Figure 1. Tax expression does not change between wt and mutants.

a Tax protein levels in nuclear lysates of cells transfected with wt or enhancer-region-mutated Tax-expression vectors. wt, wild-type; mut, mutant. NC, negative control

Figure S2



Supplementary Figure 2. Relationship between HTLV-1 ISs and genetic/epigenetic environment in JET cells infected with HTLV-1-wt or mut.

a The frequency of ISs within genes or inter-genes in JET cells infected with HTLV-1-wt or mut.

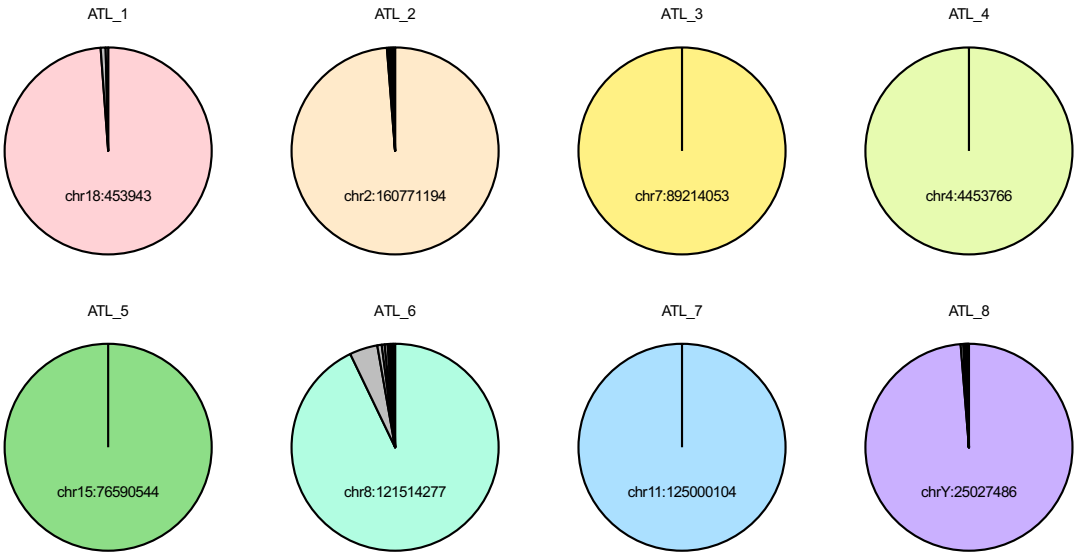
b Fold enrichment of IS distribution in each histone modification compared to random distributions in cells infected with each molecular clone.

Figure S3

a

		ATL_1	ATL_2	ATL_3	ATL_4	ATL_5	ATL_6	ATL_7	ATL_8
PVL (%)		43.64	58.75	121.6	102.0	62.7	23.75	33.4	66.56
Provirus type		full	full	5' defective	full	full	5' defective	5' defective	5' defective
Experiment	Bulk_ mRNA-seq	done	done	done	done	done	-	-	-
	Single cell_ RNA-seq	done	done	-	-	-	done	done	done

b

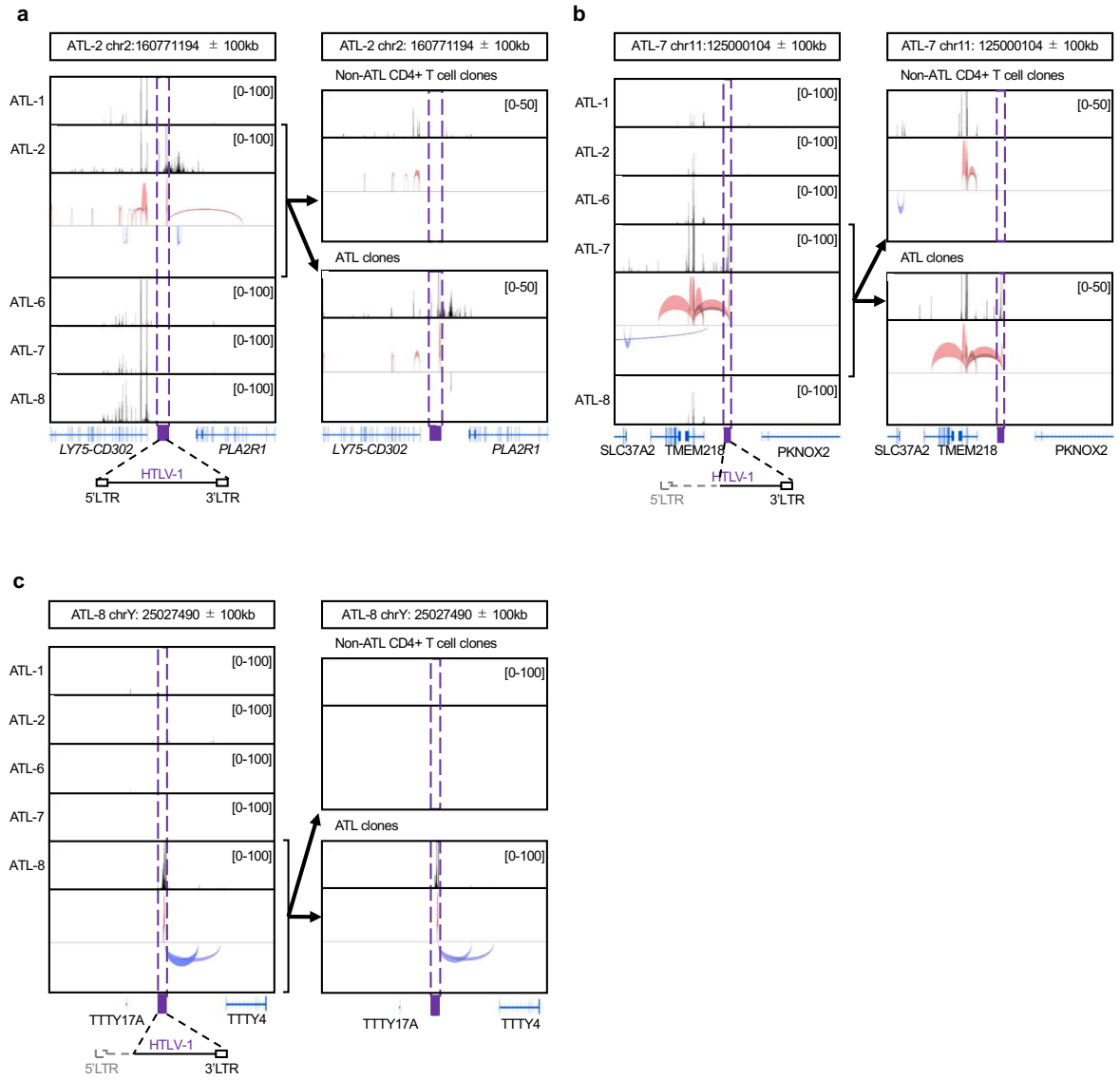


Supplementary Figure 3. The characterization of patient cells.

a Proviral load (PVL) and proviral structure, full-length type or defective type, of eight ATL cases are shown. These PVL were determined by ddPCR. These provirus types were evaluated by HTLV-1-capture-seq (Katsuya et al., 2019).

b Clonality of HTLV-1 infected cells in eight ATL cases shown in figure 6 and supplementary figure 4 were evaluated by DNA-capture-seq or LM-PCR as described previously [15].

Figure S4



Supplementary Figure 4. Transcriptional characterization of the provirus and the flanking host genomes in freshly isolated PBMCs from infected individuals. **a-c** Local transcriptome including viral integration site are visualized by IGV. We obtained scRNA-seq data from five ATL cases. The data shown were region with viral IS of ATL-2 (**a**), ATL-7 (**b**) and ATL-8 (**c**), respectively. Data from CD4+ T cells are shown in the left panel. CD4+ T cells are further divided into non-ATL cells (right, upper panel) and ATL cells (right, lower panel).

Supplementary Tables

Table S1. Integration site and strand direction of wt- and mut-infected clones

Clone name	Integration site		Proviral direction in the host genome
wt_1	chr11	92526218	+
	chrX	2718939	−
wt_2	chr2	36362768	−
	chr3	48487187	+
	chr6	158528725	−
	chr8	21638241	−
wt_3	chr4	80736011	−
	chr16	61845376	+
wt_4	chr9	2233717	−
	chr16	60365280	+
	chr16	67905552	+
wt_5	chr5	64256335	+
mut_1	chr4	140454610	−
	chr11	120298441	−
mut_2	chr2	60025417	+
mut_3	chr7	138733748	+
	chr9	139894554	+
mut_4	chr9	33403846	+
	chr11	65253493	−

The profile of the cells infected with each molecular clone. The integration site and the strand direction of wt-infected clones and mut-infected clones.

Table S2. Primer and reporter construct sequences

Primer	Sequence
XhoI-3'LTR300-F	5'-ctcgagTGTGTACTAAATTTCTCTCCTGGA-3'
HindIII-3'LTR300-R	5'-aagcttGCGTCCGCCGTCTAGGTAAGTT-3'
XhoI-5'LTR-F	5'-ctcgagTGACAATGACCATGAGCCCCAA-3'
HindIII-5'LTR-R	5'-aagcttTGTGTACTAAATTTCTCTCCTGGA-3'
BamHI-NFR-F	5'-ggatccTCCTTCCGTTCCACTCAAC-3'
BamHI-NFR-R	5'-ggatccTGGTAGGCCTTGGTTTGA-3'
NFR_mutant_construct	5'-TCCTTCCGTTCCACTCAACCCTCACCACTCCAGGACTCATCTGGACGTTTACCGATGGCACGCCTATGATCAGTGGCCCATGTCCAAAGGACGGTCAACCCTCGCTCGTGCTACAGTCCTCCTCCTTTATATTTCACAAATTTCAAACCAAGGCCTACCA-3'

Primers used for reporter constructs. The sequence of primers used for each reporter construct and the NFR mutant construct.

Table S3. Primer sequences for qRT-PCR and PVL measurement

Primers and probes	Sequence
qRT-PCR_*HBZ*-F	5'-GGACGCAGTTCAGGAGGCAC-3'
qRT-PCR_*HBZ*-R	5'-CCTCCAAGGATAATAGCCCG-3'
qRT-PCR_*tax*-F	5'-CCGGCGCTGCTCTCATCCCGGT-3'
qRT-PCR_*tax*-R	5'-GGCCGAACATAGTCCCCCAGAG-3'
ddPCR_*tax*-F	5'-CGGATACCCAGTCTACGTGTT-3'
ddPCR_*tax*-R	5'-CAGTAGGGCGTGACGATGTA-3'
ddPCR_*alb*-F	5'-TGCATGAGAAAACGCCAGTAA-3'
ddPCR_*alb*-R	5'-ATGGTCGCCTGTTCACCAA-3'
ddPCR_*tax*-probe	5'-/ 56-FAM/CTGTGTACA/ZEN/AGGCGACTGCC/3IABkFQ/ -3'
ddPCR_*alb*-probe	5'- /5HEX/TGACAGAGT/ZEN/CACCAAATGCTGCACAGAA/3IABkFQ/ -3'

Oligonucleotides for qRT-PCR and PVL measurement.
The sequence of primers and probes for qRT-PCR and PVL measurement.

Table S4. Labeled and non-labeled probe sequences for EMSA

Probes	Sequence
Biotin_wt-probe	5'- bio_ACTCAACCCTCACCACTCCAGGCCTTATTTGGACATTTACCGATGGCACGC CTATGATTTCCGGGCCCTGCCCTAAAGATGGCCAGCCATCTTTAGTACTACAGT CCTCCTCCTTTATATT -3'
Biotin_NC3100-probe	5'- bio_CTACTACTACTCTCAGAGGCCACAATGGCTTCCCTAATCTCCCATGGGTTG CCTGTGTCCGAAAACAAAACCCAGCAAACCCCTGGAACAATTAAGTTCCTAG GGCAGATAATTTCACCCA -3'
Biotin_NC5000-probe	5'- bio_CTGGTCTTAATAGCCGCCAGTGGAAAGGACCACAGGAGGCTCTCCAAGA AGCTGCCGGCGCTGCTCTCATCCCGGTAAGCGCTAGTTCTGCCCAGTGGATCC CGTGGAGACTCCTCAAGCG -3'
wt_cold-probe	5'- ACTCAACCCTCACCACTCCAGGCCTTATTTGGACATTTACCGATGGCACGCCT ATGATTTCCGGGCCCTGCCCTAAAGATGGCCAGCCATCTTTAGTACTACAGTCC TCCTCCTTTATATT -3'
mut1_cold-probe	5'- ACTCAACCCTCACCACTCCAGGACTCATCTGGACGTTTACCGATGGCACGCCT ATGATTTCCGGGCCCTGTCCAAAGGACGGTCAACCCTCGCTCGTGCTACAGTC CTCCTCCTTTATATT -3'
mut2_cold-probe	5'- ACTCAACCCTCACCACTCCAGGCCTTATTTGGACATTTACCGATGGCACGCCT ATGATCAGTGGCCCATGCCCTAAAGATGGCCAGCCATCTTTAGTACTACAGTCC TCCTCCTTTATATT -3'
mut3_cold-probe	5'- ACTCAACCCTCACCACTCCAGGACTCATCTGGACGTTTACCGATGGCACGCCT ATGATCAGTGGCCCATGTCCAAAGGACGGTCAACCCTCGCTCGTGCTACAGTC CTCCTCCTTTATATT -3'

The sequence of labeled and non-labeled Probes for EMSA are described.

Table S5. Primer sequences for MNase assay

Primers	Sequence
MNase-5'LTR_j-F	5'-GACAGCCCATCCTATAGCACTC-3'
MNase-5'LTR_j-R	5'-CTAGCGCTACGGGAAAAGATT-3'
MNase-gag-F	5'-CAGAGGAAGATGCCCTCCTATT-3'
MNase-gag-R	5'-GTCAACCTGGGCTTTAATTACG-3'
MNase-pol-F	5'-TTCCGCCACCGCACAAGTCG-3'
MNase-pol-R	5'-TGCCTTGGAAGGTGCCCAGG-3'
MNase-env-F	5'-CTGTTCCCACCCTAGGATCCCG-3'
MNase-env-R	5'-GAGGCTCTTTCCTGAGGCGAGG-3'
MNase-NFR-F	5'-CTCCTTCCGTTCCACTCAAC-3'
MNase-NFR-R	5'-GTGGTAGGCCTTGGTTTGAA-3'
MNase-3'LTR_j-F	5'-AATACACCAACATCCCCATTTC-3'
MNase-3'LTR_j-R	5'-GTTTTTCACTGGGAGGCTCTAA-3'

Oligonucleotides for MNase assay.

The sequence of primers for MNase assay are described.

Table S6. Oligonucleotide for guide sequence used in CRISPR/Cas9 system

Primers	Sequence
PX330-KOwt-1-F	5'-CACCGTCACCACTCCAGGCCTTATT-3'
PX330-KOwt-1-R	5'-AAACAATAAGGCCTGGAGTGGTGAC-3'
PX330-KOwt-2-F	5'-CACCGAGGACTGTAGTACTAAAGA-3'
PX330-KOwt-2-R	5'-AAACTCTTTAGTACTACAGTCCTC-3'

Oligonucleotide for guide sequence used in CRISPR/Cas9 system.

The sequence of oligonucleotide for constructing guide sequence cloned into pX330 plasmid are described.

References

1 Uchiyama, T., Yodoi, J., Sagawa, K., Takatsuki, K. & Uchino, H. Adult T-cell leukemia: clinical and hematologic features of 16 cases. *Blood* **50**, 481-492 (1977).

2 Poiesz, B. J. *et al.* Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A* **77**, 7415-7419 (1980).

3 Hinuma, Y. *et al.* Adult T-cell leukemia: antigen in an ATL cell line and detection of antibodies to the antigen in human sera. *Proc Natl Acad Sci U S A* **78**, 6476-6480, doi:10.1073/pnas.78.10.6476 (1981).

4 Matsuoka, M. & Jeang, K. T. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer* **7**, 270-280, doi:10.1038/nrc2111 (2007).

5 Bangham, C. R. M. Human T Cell Leukemia Virus Type 1: Persistence and Pathogenesis. *Annu Rev Immunol* **36**, 43-71, doi:10.1146/annurev-immunol-042617-053222 (2018).

6 Yoshida, M. Multiple viral strategies of HTLV-1 for dysregulation of cell growth control. *Annu Rev Immunol* **19**, 475-496, doi:10.1146/annurev.immunol.19.1.475 (2001).

7 Giam, C. Z. & Semmes, O. J. HTLV-1 Infection and Adult T-Cell Leukemia/Lymphoma-A Tale of Two Proteins: Tax and HBZ. *Viruses* **8**, doi:10.3390/v8060161 (2016).

8 Satou, Y. *et al.* HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo. *PLoS Pathog* **7**, e1001274, doi:10.1371/journal.ppat.1001274 (2011).

9 Satou, Y., Yasunaga, J., Yoshida, M. & Matsuoka, M. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. *Proc Natl Acad Sci U S A* **103**, 720-725, doi:10.1073/pnas.0507631103 (2006).

10 Usui, T. *et al.* Characteristic expression of HTLV-1 basic zipper factor (HBZ) transcripts in HTLV-1 provirus-positive cells. *Retrovirology* **5**, 34, doi:10.1186/1742-4690-5-34 (2008).

11 Satou, Y. *et al.* The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome. *Proc Natl Acad Sci U S A* **113**, 3054-3059, doi:10.1073/pnas.1423199113 (2016).

12 Wolf, D. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**, 46-57, doi:10.1016/j.cell.2007.07.026 (2007).

13 Wolf, D. & Goff, S. P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**, 1201-1204, doi:10.1038/nature07844 (2009).

14 Miyazato, P. *et al.* Application of targeted enrichment to next-generation sequencing of retroviruses integrated into the host human genome. *Sci Rep* **6**, 28324, doi:10.1038/srep28324 (2016).

15 Katsuya, H. *et al.* The Nature of the HTLV-1 Provirus in Naturally Infected Individuals Analyzed by the Viral DNA-Capture-Seq Approach. *Cell Rep* **29**, 724-735 e724, doi:10.1016/j.celrep.2019.09.016 (2019).

16 Maeda, M. *et al.* Origin of human T-lymphotrophic virus I-positive T cell lines in adult T cell leukemia. Analysis of T cell receptor gene rearrangement. *J Exp Med* **162**, 2169-2174, doi:10.1084/jem.162.6.2169 (1985).

17 Takeda, S. *et al.* Genetic and epigenetic inactivation of tax gene in adult T-cell leukemia cells. *Int J Cancer* **109**, 559-567, doi:10.1002/ijc.20007 (2004).

18 Cook, L. B., Rowan, A. G., Melamed, A., Taylor, G. P. & Bangham, C. R. HTLV-1-infected T cells contain a single integrated provirus in natural infection. *Blood* **120**, 3488-3490, doi:10.1182/blood-2012-07-445593 (2012).

19 Hanon, E. *et al.* Abundant tax protein expression in CD4+ T cells infected with human T-cell lymphotropic virus type I (HTLV-I) is prevented by cytotoxic T lymphocytes. *Blood* **95**, 1386-1392 (2000).

20 Yoshida, M., Satou, Y., Yasunaga, J., Fujisawa, J. & Matsuoka, M. Transcriptional control of spliced and unspliced human T-cell leukemia virus type 1 bZIP factor (HBZ) gene. *J Virol* **82**, 9359-9368, doi:10.1128/JVI.00242-08 (2008).

21 Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**, 7-18, doi:10.1038/nrg2905 (2011).

22 Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187, doi:10.1038/nature09033 (2010).

23 Hirabayashi, S. *et al.* NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat Genet* **51**, 1369-1379, doi:10.1038/s41588-019-0485-9 (2019).

24 Suzuki, T., Hirai, H., Fujisawa, J., Fujita, T. & Yoshida, M. A trans-activator Tax of human T-cell leukemia virus type 1 binds to NF-kappa B p50 and serum response factor (SRF) and associates with enhancer DNAs of the NF-kappa B site and CArG box. *Oncogene* **8**, 2391-2397 (1993).

25 Tsunoda, T. & Takagi, T. Estimating transcription factor bindability on DNA. *Bioinformatics* **15**, 622-630, doi:10.1093/bioinformatics/15.7.622 (1999).

26 Tanaka-Nakanishi, A., Yasunaga, J., Takai, K. & Matsuoka, M. HTLV-1 bZIP factor suppresses apoptosis by attenuating the function of FoxO3a and altering its localization. *Cancer Res* **74**, 188-200, doi:10.1158/0008-5472.CAN-13-0436 (2014).

27 Melamed, A. *et al.* The human leukemia virus HTLV-1 alters the structure and transcription of host chromatin in cis. *Elife* **7**, doi:10.7554/eLife.36245 (2018).

28 Kataoka, K. *et al.* Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet* **47**, 1304-1315, doi:10.1038/ng.3415 (2015).

29 Rosewick, N. *et al.* Cis-perturbation of cancer drivers by the HTLV-1/BLV proviruses is an early determinant of leukemogenesis. *Nat Commun* **8**, 15264, doi:10.1038/ncomms15264 (2017).

30 Fujisawa, J., Seiki, M., Kiyokawa, T. & Yoshida, M. Functional activation of the long terminal repeat of human T-cell leukemia virus type I by a trans-acting factor. *Proc Natl Acad Sci U S A* **82**, 2277-2281, doi:10.1073/pnas.82.8.2277 (1985).

31 Fujisawa, J., Seiki, M., Sato, M. & Yoshida, M. A transcriptional enhancer sequence of HTLV-I is responsible for trans-activation mediated by p40 chi HTLV-I. *EMBO J* **5**, 713-718 (1986).

32 Billman, M. R., Rueda, D. & Bangham, C. R. M. Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. *Wellcome Open Res* **2**, 87, doi:10.12688/wellcomeopenres.12469.2 (2017).

33 Mahgoub, M. *et al.* Sporadic on/off switching of HTLV-1 Tax expression is crucial to maintain the whole population of virus-induced leukemic cells. *Proc Natl Acad Sci U S A* **115**, E1269-E1278, doi:10.1073/pnas.1715724115 (2018).

34 Arnold, J. *et al.* Enhancement of infectivity and persistence in vivo by HBZ, a natural antisense coded protein of HTLV-1. *Blood* **107**, 3976-3982, doi:10.1182/blood-2005-11-4551 (2006).

35 Tamiya, S. *et al.* Two types of defective human T-lymphotropic virus type I provirus in adult T-cell leukemia. *Blood* **88**, 3065-3073 (1996).

36 Miyazaki, M. *et al.* Preferential selection of human T-cell leukemia virus type 1 provirus lacking the 5' long terminal repeat during oncogenesis. *J Virol* **81**, 5714-5723, doi:10.1128/JVI.02511-06 (2007).

37 Cook, L. B. *et al.* The role of HTLV-1 clonality, proviral structure, and genomic integration site in adult T-cell leukemia/lymphoma. *Blood* **123**, 3925-3931, doi:10.1182/blood-2014-02-553602 (2014).

38 O'Geen, H. *et al.* Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet* **3**, e89, doi:10.1371/journal.pgen.0030089 (2007).

39 Taniguchi, Y. *et al.* Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. *Retrovirology* **2**, 64, doi:10.1186/1742-4690-2-64 (2005).

40 Verdonck, K. *et al.* Human T-lymphotropic virus 1: recent knowledge about an ancient infection. *Lancet Infect Dis* **7**, 266-281, doi:10.1016/S1473-3099(07)70081-6 (2007).

41 David, R. M. & Doherty, A. T. Viral Vectors: The Road to Reducing Genotoxicity. *Toxicol Sci* **155**, 315-325, doi:10.1093/toxsci/kfw220 (2017).

42 Liu, M. *et al.* Genomic discovery of potent chromatin insulators for human gene therapy. *Nat Biotechnol* **33**, 198-203, doi:10.1038/nbt.3062 (2015).

43 Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415-419, doi:10.1126/science.1088547 (2003).

44 Mitchell, M. S. *et al.* Phenotypic and genotypic comparisons of human T-cell leukemia virus type 1 reverse transcriptases from infected T-cell lines and patient samples. *J Virol* **81**, 4422-4428, doi:10.1128/JVI.02660-06 (2007).

45 Furuta, R. *et al.* Human T-cell leukemia virus type 1 infects multiple lineage hematopoietic cells in vivo. *PLoS Pathog* **13**, e1006722, doi:10.1371/journal.ppat.1006722 (2017).

46 Matsuzaki, K. *et al.* PML-nuclear bodies are involved in cellular serum response. *Genes Cells* **8**, 275-286 (2003).

47 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

48 Satou, Y. *et al.* Dynamics and mechanisms of clonal expansion of HIV-1-infected cells in a humanized mouse model. *Sci Rep* **7**, 6913, doi:10.1038/s41598-017-07307-4 (2017).

Figures

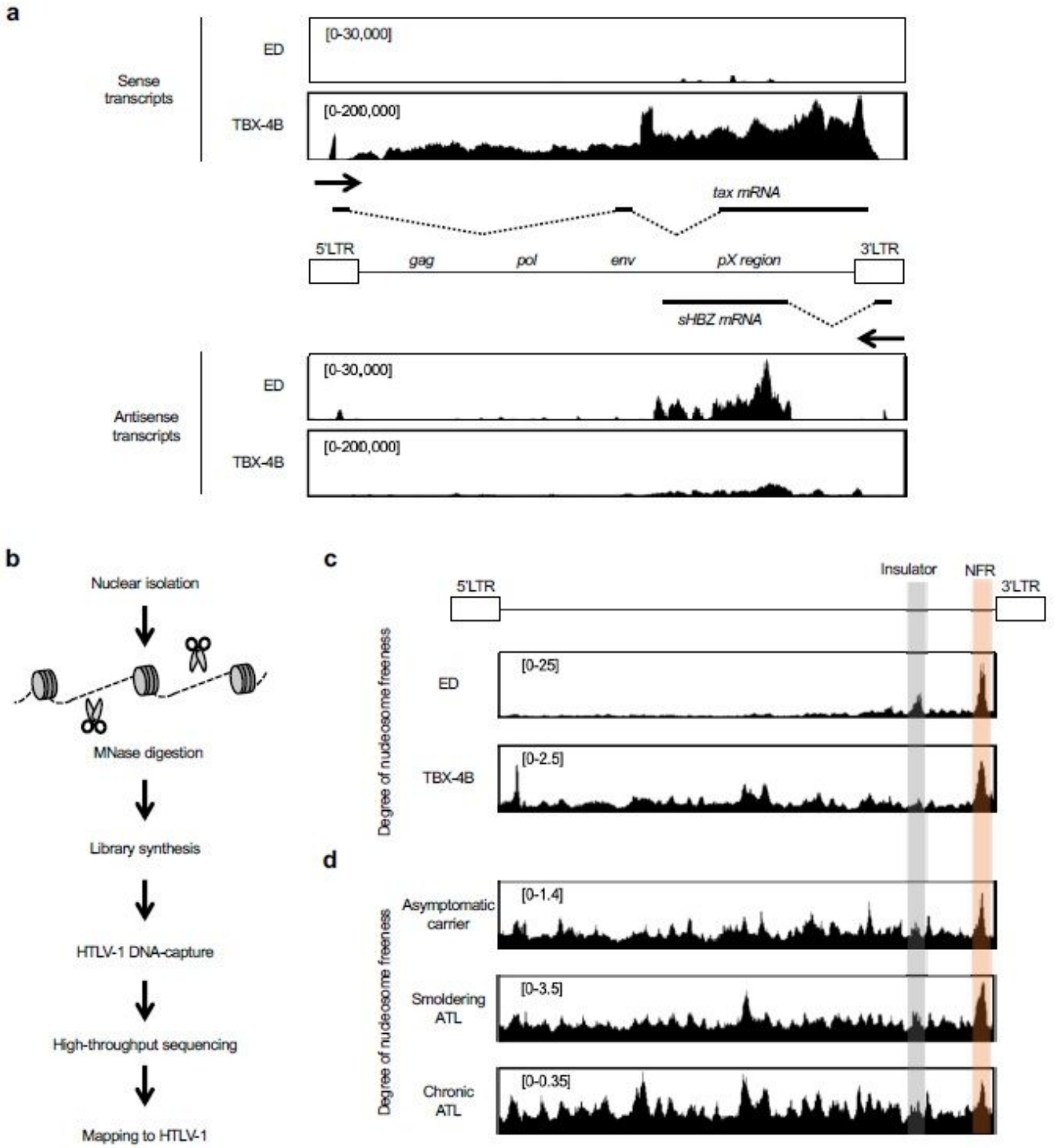


Figure 1

Transcriptome pattern and nucleosome positions in the HTLV-1 provirus. a Stranded proviral transcriptome are visualized at the sense (above) or antisense (below) orientation in ED cells and TBX-4B cells by Integrative Genomics Viewer (IGV; https://software.broadinstitute.org/software/igv/). b

Experimental workflow of MNase-seq with HTLV-1 DNA-capture-seq. c, d MNase-seq of ED cells, TBX4B cells, (c) and PBMCs of asymptomatic HTLV-1 carrier and ATL patients (d). Degree of nucleosome freeness is shown as the MNase-seq value normalized to the input DNA-seq value. Orange-shaded region indicates the NFR location and gray-shaded region indicates the insulator region. ATL, adult T-cell leukemia/lymphoma; NFR, nucleosome-free region.

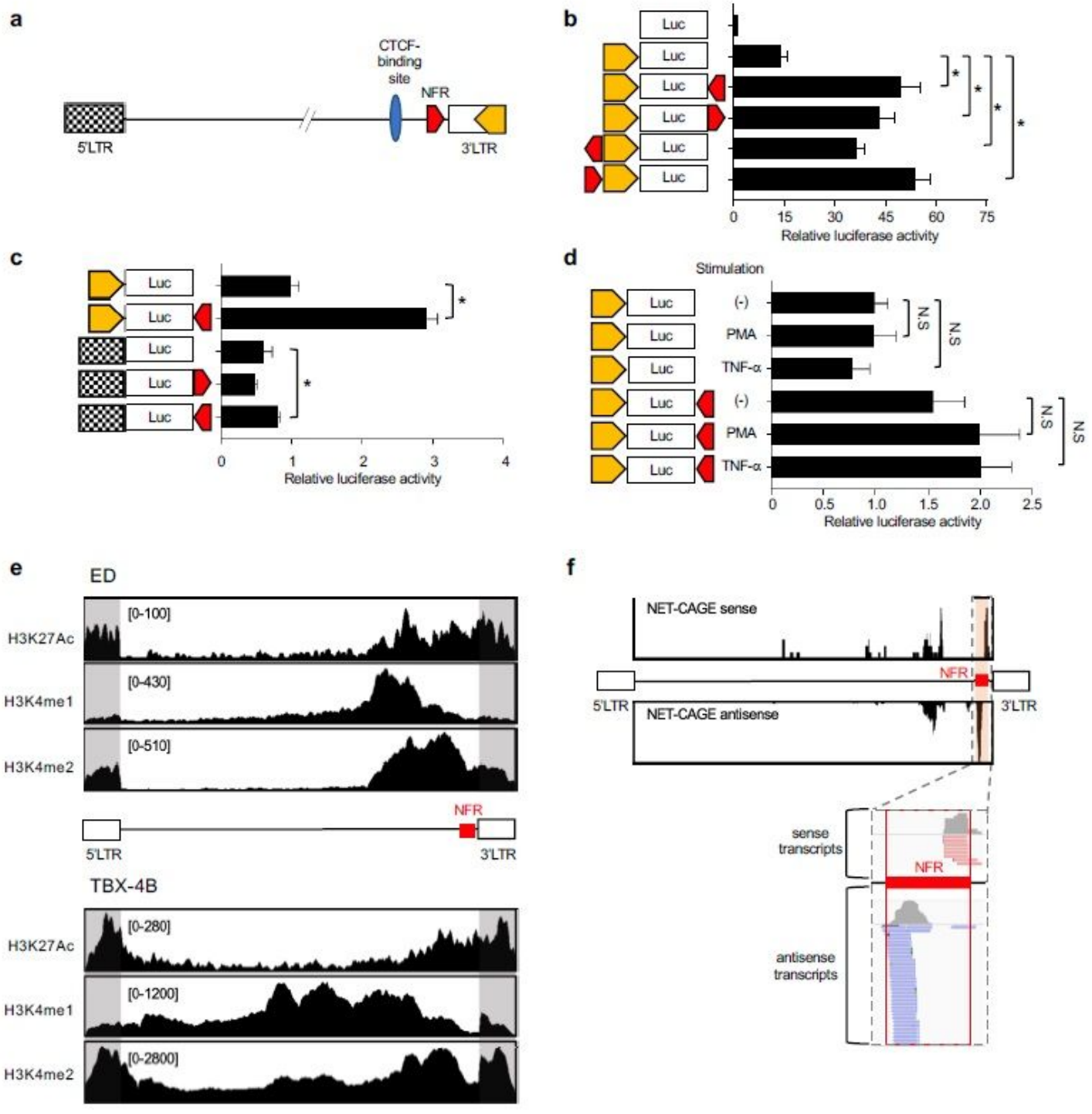


Figure 2

The nucleosome region harbors enhancer-related histone modifications and produces enhancer RNAs. a Schematic of the HTLV-1 provirus structure. The 5′LTR (black plaid), CTCF-binding site (blue) 11, NFR (red), and the HBZ promoter (yellow) 20 are shown. b-d Transcriptional regulatory function of the NFR was analyzed by luciferase reporter assays in Jurkat cells. The HBZ promoter 20 (b, c) and 5′LTR (c) were used as a promoter. PMA and TNF-α were used for cell stimulation (d). Luciferase activity was normalized to Renilla activity. Representative data of three independent experiments is shown as fold change to pGL4-basic (b), pGL4-basic-HBZ promoter (c, d) (Student's t-test, *P < 0.05). e H3K27Ac (top), H3K4me1 (middle), and H3K4me2 (bottom) occupancy within the provirus in ED (upper) and TBX-4B (lower) cells. ChIP-seq signals were visualized by IGV. Gray-shaded areas indicate the ChIP signal mapped to LTRs. f NET-CAGE results using nuclear lysates of ED cells in the sense (above) or antisense (below) orientations, demonstrating eRNAs at the NFR. The bottom panel is an enlarge image of the signals around the NFR. NET-CAGE signals were visualized by IGV. Luc, luciferase; NFR, nucleosome-free region; N.S., not significant.

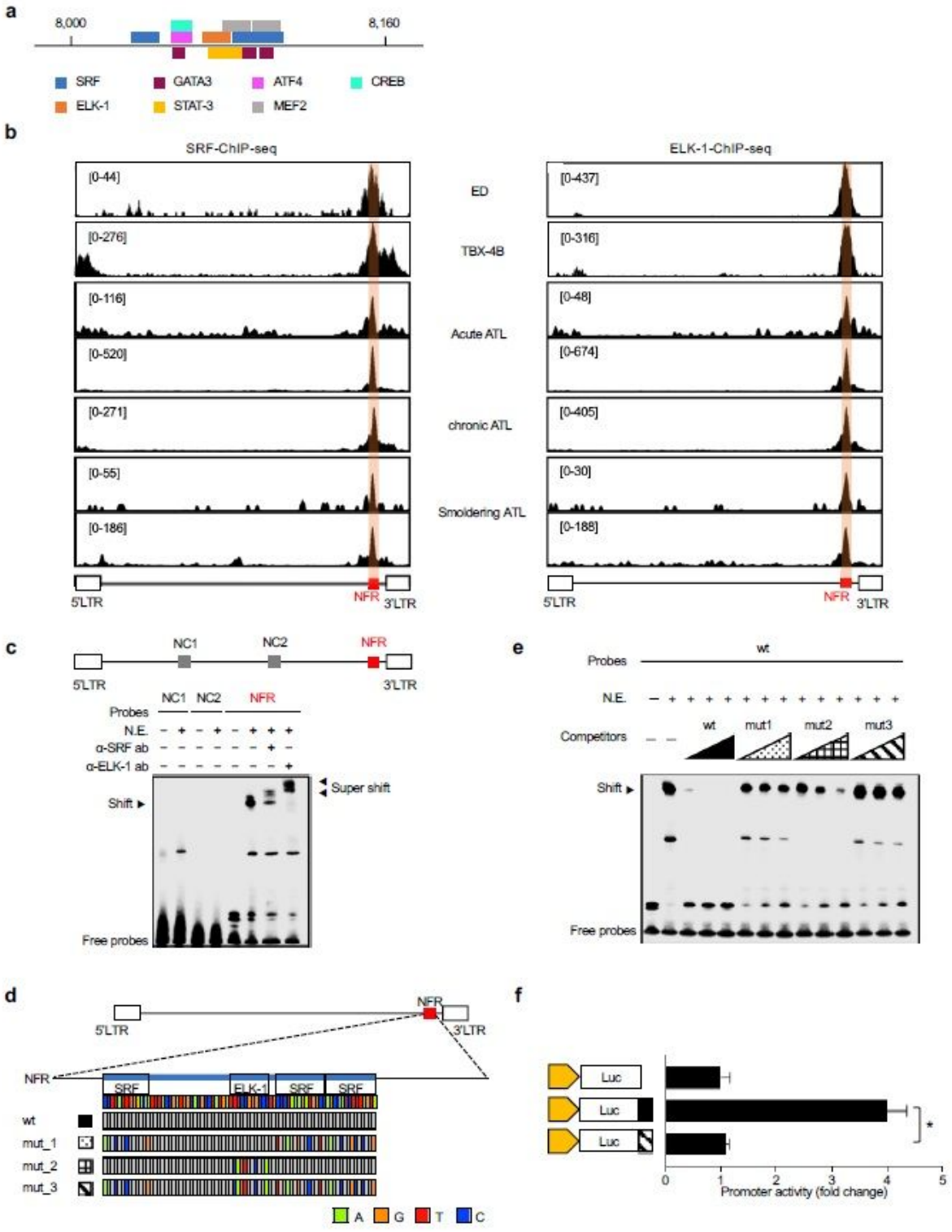


Figure 3

SRF and ELK-1 bind to the NFR in a DNA sequence-200 dependent manner. a The prediction of transcription factor binding to the NFR was performed by using TFBIND (http://tfbind.hgc.jp/) 25. Candidate transcription factor binding sites are shown. b The localization of SRF (left) and ELK-1 (right) to the NFR in cell lines and PBMCs of ATL patients. ChIP-seq signals were visualized by IGV. Orange-shaded region indicates the NFR. c The binding ability of SRF and ELK-1 to the NFR oligonucleotides was

analyzed by EMSA. Biotinylated DNA probes of 120 bp for the NFR (red) and negative control regions (gray) were incubated with nuclear extract of 293T cells transfected with SRF and ELK- expression vectors (N.E.). NFR-SRF/ELK-1 complexes and super-shifted complexes, which were detected with the anti-SRF and the anti-ELK-1 antibody, are indicated by arrowheads. d The position of introduced mutations used are shown as green (A), orange (G), red (T) and blue (C). e EMSA competition analysis with wt- and mutated-NFR oligonucleotides. Biotin-labeled wt- NFR probe and 100, 200, or 300 times more non-labeled competitor oligonucleotides were used in this assay. Each non-labeled competitor sequence is shown in (d) and is indicated wt as black, mut1 as dots pattern, mut2 as lattice pattern and mut3 as diagonal stripe pattern. f Transcriptional regulatory function of the wt (black) or mut (pattern) NFRs was analyzed using the HBZ promoter (yellow) in Jurkat cells by luciferase assay. Luciferase activity was normalized to Renilla activity. Representative data of three independent experiments is shown as fold change to pGL4-basic-HBZ promoter (Student's t-test, *P < 0.05). ATL, adult T-cell leukemia/lymphoma; NFR, nucleosome-free region; wt, wild-type; mut, mutant; N.E., nuclear extract.

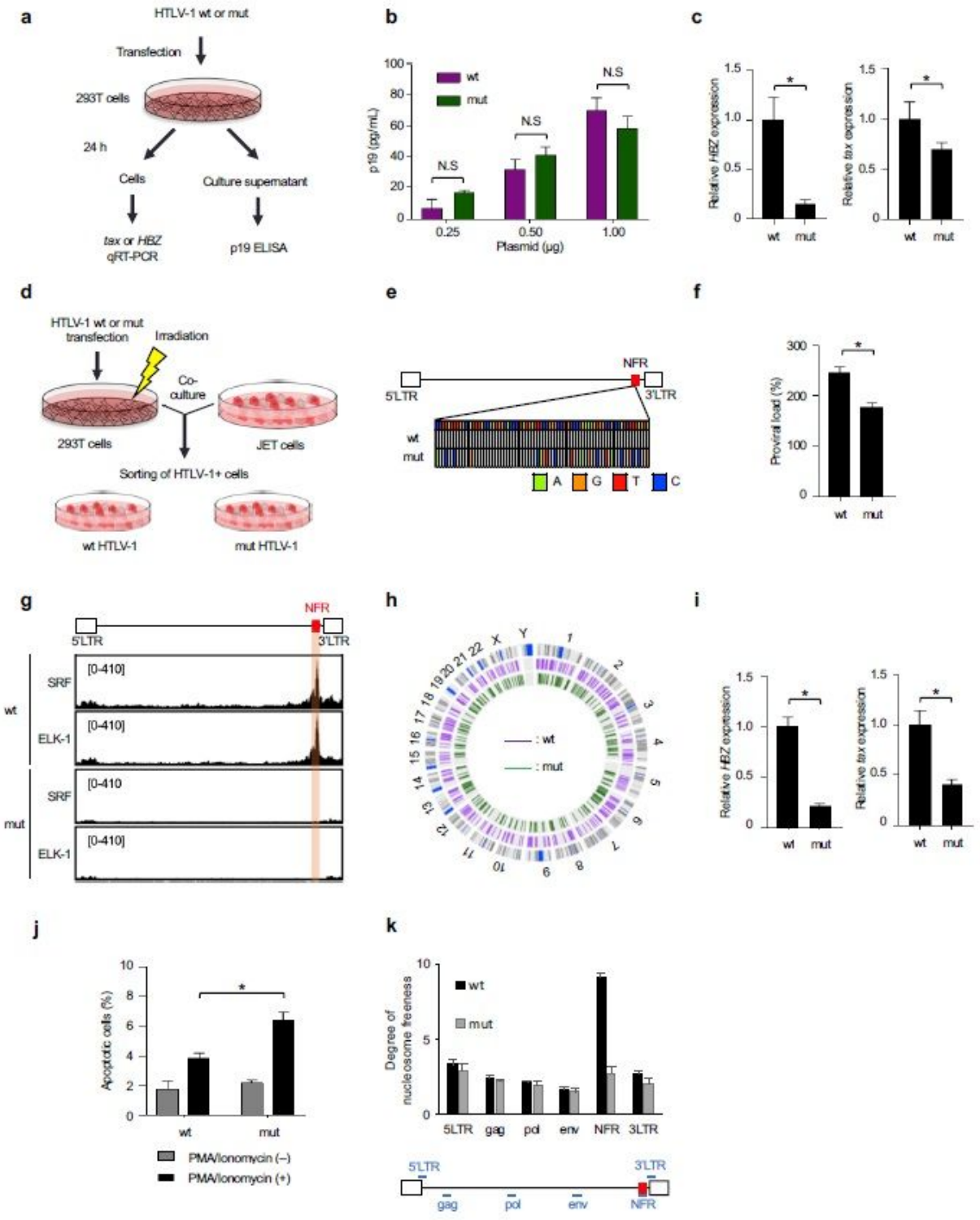


Figure 4

Generation and characterization of the HTLV-1 infectious clone with mutations in the SRF/ELK-1 binding site. a Diagram illustrating the experimental workflow of transient transfection evaluation using HTLV-1-wt or mut molecular clones. b p19 was quantified for analyzing infectivity by ELISA using the supernatant of 293T cells transfected with HTLV-1-wt or mut (0.25, 0.5, 1 µg). Results are expressed as the mean ± SEM of three experiments performed in duplicate. c qRT-PCR results of HBZ (left) and tax (right) levels

after transient transfection using HTLV- 1-wt or mut. 18S rRNA was quantified as an internal control. Data shown are representative of two independent experiments (Student's t-test, *P < 0.05). d Experimental workflow of establishing stable cells infected with wt or mut HTLV-1. e The HTLV-1 NFR sequences of JET cells infected with HTLV-1-wt or mut was analyzed by HTLV-1 DNA-capture-seq. Each detected mutation is indicated green as A, orange as G, red as T and blue as C. f Proviral load in JET cells infected with HTLV-1-wt or mut was measured by digital droplet PCR. Values reported were derived from three biological replicates (Student's t-test, *P < 0.05). g Localization of SRF and ELK-1 in cells infected with HTLV-1-wt (above) or mut (below). SRF and ELK-1 ChIP-seq signals were visualized by IGV. Orange-shaded area represents the NFR location. h The distribution of integration sites (ISs) in bulk JET cells infected with HTLV-1_wt (purple line) or mut (green line) is shown in a circus plot. Outer gray ring represents human chromosomes, cytogenetic bands shown as black lines and the centromere shown as blue line. Numbers and letters surrounding the circos represent human chromosomes. i Representative qRT-PCR results of HBZ (left) and tax (right) levels in JET cells infected with HTLV-1-wt or mut. 18S rRNA was quantified as an internal control. Data shown are representative of two independent experiments (Student's t-test, *P < 0.05). j Cell apoptosis was detected by Annexin V staining after stimulation with PMA and ionomycin. Values reported were derived from two biological replicates (Student's t-test, *P < 0.05). k MNase assay of in JET cells infected with HTLV-1-wt (above) or mut (below). Degree of open chromatin was evaluated by MNase treatment and ddPCR. Values of indicated proviral regions are shown after normalization to non-MNase digestion sample value. NFR, nucleosome-free region; wt, wild-type; mut, mutant.

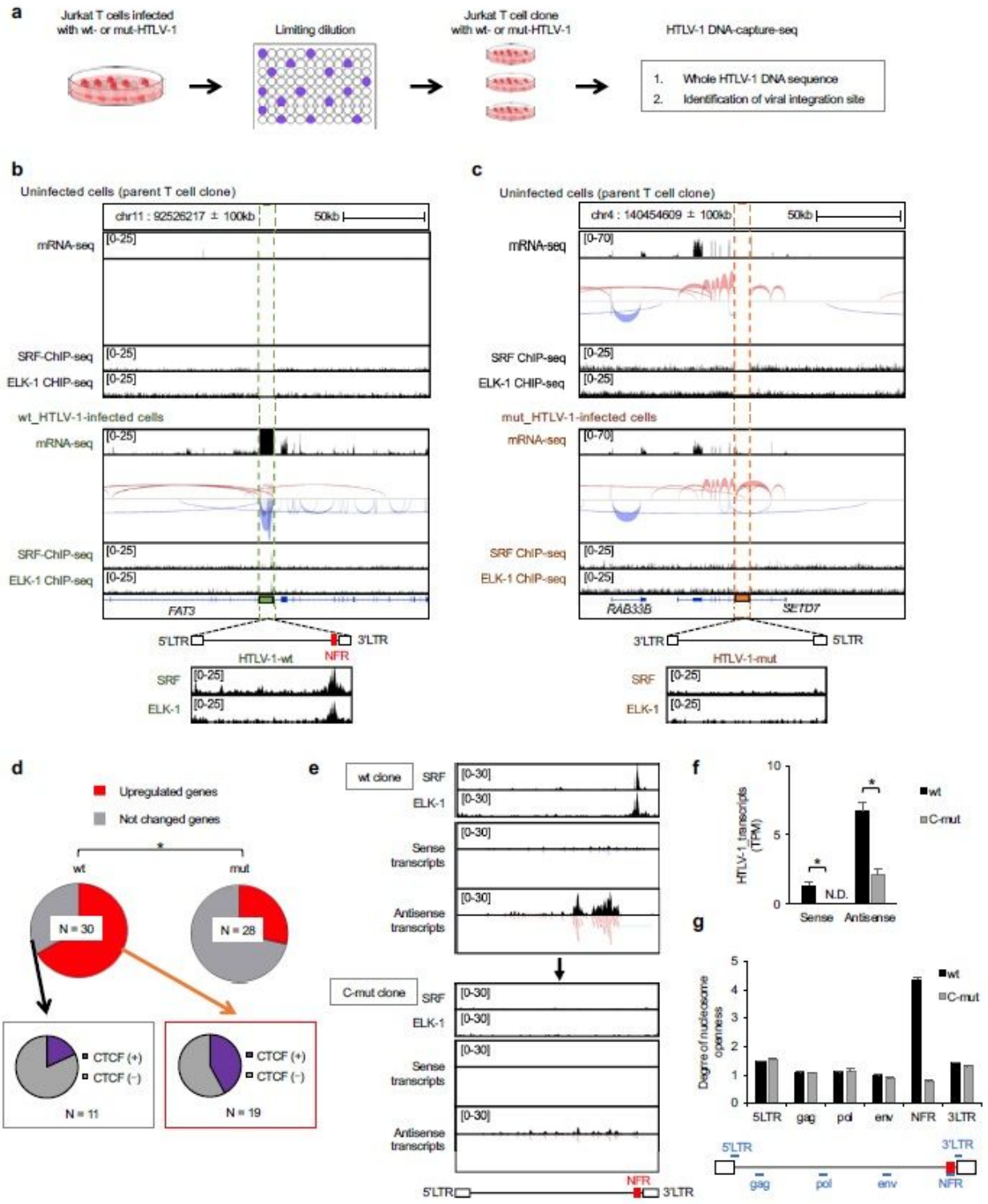


Figure 5

Establishment and characterization of Jurkat T cell clones infected with wt- and mut- HTLV-1. a Experimental workflow of establishing infected clones with wt or mut HTLV-1 by limiting dilution. b, c Local transcriptome and splice junction near viral integration site are visualized in a JET clone infected with HTLV-1-wt (b) and mut (c) by IGV. The splice junctions are shown in red as sense direction and blue as antisense direction. The thickness of red and blue line is indicated the frequency of detection of

specific splices. Host genes near the IS and direction of HTLV-1 provirus is indicated below the graph. SRF/ELK-1 ChIP-seq results are also shown in each ATL clone.d The fraction of upregulated genes in JET clones infected with HTLV- 1-wt (above left) and mut (above right). Presence or absence of CTCF ChIP-seq signals in the upregulated group (below right) or 'no change' group (below left) of JET clones infected with HTLV-1-wt. Chi-square test, $P < 0.05$. e mRNA-seq and SRF/ELK-1 ChIP-seq results of JET cells infected with HTLV-1-wt (above) and CRISPR-mutated HTLV-1 (below). Representative results from two independent experiment are visualized by IGV. f Level of proviral expression of JET cells infected with HTLV-1-wt or CRISPR-mutated (C350 mut) clone. Data are generated as transcripts per million reads (TPM) from two independent mRNA-seq analyses. g MNase assay of JET cells infected with wild type or CRISPR-mutated HTLV-1. Degree of open chromatin was evaluated by MNase treatment and ddPCR. Values of indicated proviral regions are shown after normalization to non-MNase digestion sample value. NFR, nucleosome-free region; wt, wild-type; mut, mutant; C-mut, CRISPR-mutant.

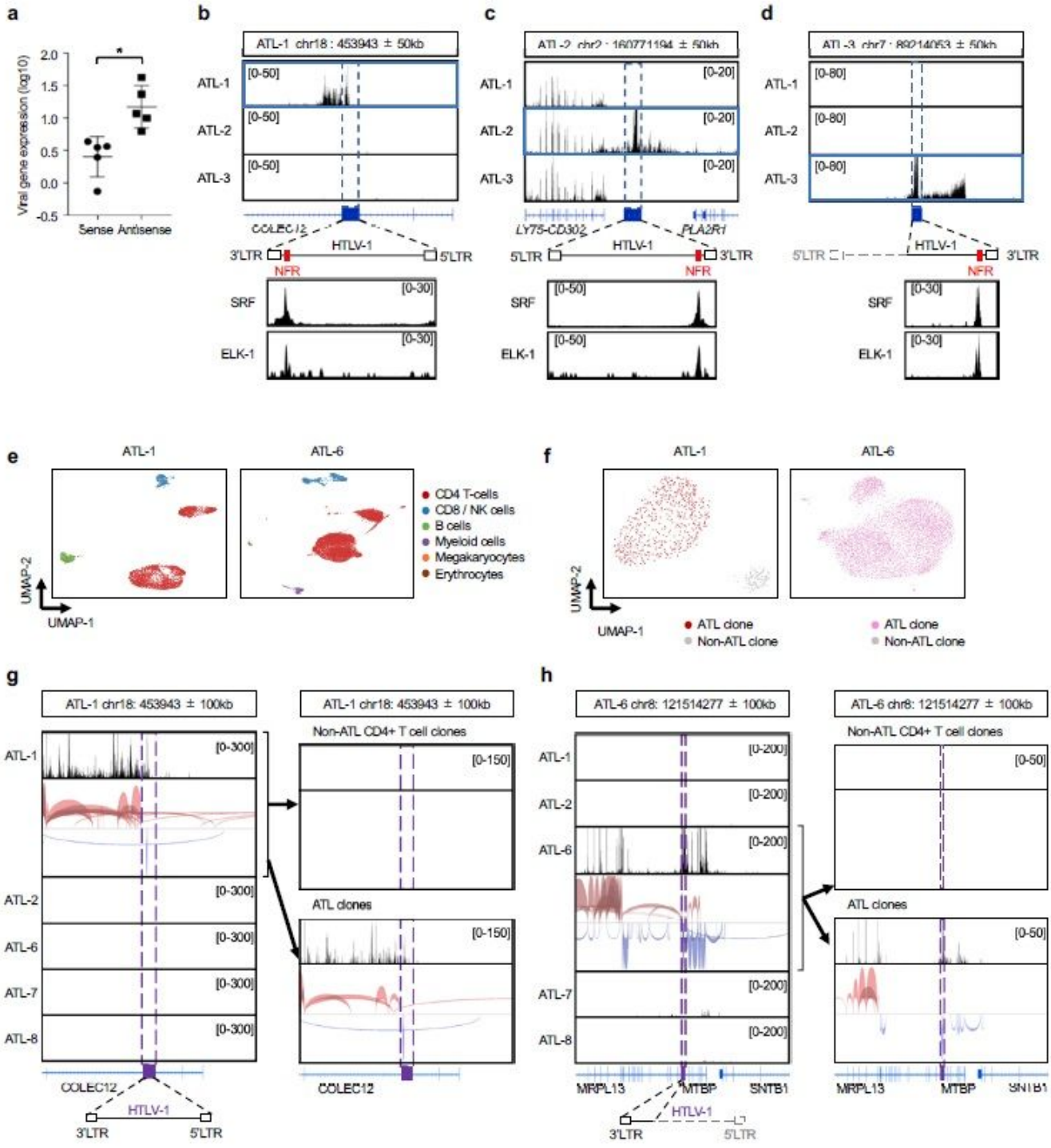


Figure 6

Transcriptional characterization of the provirus and the flanking host genomes in freshly isolated PBMCs from infected individuals. a The level of sense or antisense proviral expression in fresh PBMCs from five ATL patient samples. Data shown are transcript per million for each case. b-d Visualization of mRNA-seq data of three ATL cases at around each viral IS. Host genesnear the IS and direction of HTLV-1 provirus is indicated below the graph. SRF/ELK-1 ChIP seq results are also shown in each ATL clone. The ATL sample with HTLV-1 IS in the region is highlighted with a blue square. e scRNA-data of PBMCs from the

indicated ATL cases. Cell clustering analysis was performed with a nonlinear dimensionality reduction method, uniform manifold approximation and projection (UMAP). Each cell cluster was annotated by expression pattern of marker gene for PBMC subsets. f We defined ATL cells as T cells containing the most abundant TCR. ATL clones are shown in red (ATL-1) and pink (ATL-6). g, h Local transcriptome including viral integration site are visualized by IGV. We obtained scRNA-seq data from five ATL cases. The data shown were region with viral IS of ATL-1(g) and ATL-6 (h), respectively. Data from CD4+ T cells are shown in the left panel. CD4+ T cells are further divided into non-ATL cells (right, upper panel) and ATL cells (right, lower panel).