## Author Summary

HTLV-1 is a human retrovirus, estimated to infect over 10 million individuals worldwide, which causes the inflammatory disease HTLV-1-associated Myelopathy/Tropical Spastic Paraparesis and an aggressive malignancy known as Adult T-cell Leukemia/Lymphoma. The mechanisms that allow the virus to maintain a life-long infection are not fully understood. Here we identified attributes of the host genome flanking the integrated HTLV-1 provirus associated with integration targeting and spontaneous expression of the provirus in vitro, and clonal expansion in vivo. Spontaneous expression (after short-term culture) of the viral protein Tax, which is known to drive proliferation of the infected cell, was significantly more frequent among less expanded clones, suggesting that Tax-expressing clones are more efficiently controlled by the immune response. Certain transcription start sites immediately upstream of the viral integration site were associated with virus latency, which in turn was associated with clonal expansion in vivo.


We hypothesize that the genomic integration site of HTLV-1 determines the pattern and intensity of expression of the plus and minus proviral strands, which in turn determine the equilibrium abundance and the pathogenic potential of an infected T cell clone in vivo. To test this hypothesis, we used our recently described protocol [11] of high-throughput mapping and quantification of proviral integration sites in fresh primary PBMCs from HTLV-1-infected individuals.

## Results

### HTLV-1 preferentially integrates within 1 kb of a host transcription start site and is strongly biased to specific transcription factor binding sites

To identify genomic factors associated with the targeting of HTLV-1 integration, we infected Jurkat T cells by short co-culture with the HTLV-1-producing cell line MT2. The integration sites were then analysed using our high-throughput protocol and compared to a control list of random sites in the human genome. Figure 1A illustrates the possible orientations (same or opposite) of the nearby genomic features, such as transcription start sites, either upstream or downstream of the integrated provirus.

We previously showed [11] that 47% of de novo HTLV-1 proviral integration events lie within a RefSeq gene. This frequency is slightly higher than expected by chance, but is much lower than that observed for HIV (~70%), which uses the host protein LEDGF to target proviral integration to genes [21]. As expected by chance, ~50% of proviruses integrated within host genes were in the same transcriptional orientation as the host gene (Figure 1D, in vitro).

Gillet et al [11] reported a significantly higher than expected proportion of in vitro integration sites within 10 kb of a RefSeq gene. We extended this analysis to identify the optimal (most frequent) distance between the integration site and the nearest host transcription start site (TSS). The results (Figure 1B) show a peak preference (measured by the odds ratio, OR, observed/expected) towards integration in proximity to TSS at ~1 kb of the integrated provirus (upstream or downstream); the OR gradually diminished until it reached 1 (same as random expectation) at ~1 Mb from the integration site (Figure 1B). There was a small bias (non-significant for in vitro integration) towards integration with a TSS downstream of the integration site (Figure 1C, in vitro).

Similarly, we observed a bias (up to 2-fold greater than random) towards integration in proximity to CpG islands; again, the bias reached a peak at 1 kb from the nearest CpG island (supplementary Figure S4).

We showed previously [11] that HTLV-1 provirus preferentially integrates in transcriptionally active regions of the host genome. To test the hypothesis that specific transcription factor binding sites (TFBS) influence HTLV-1 proviral targeting, expression and clonal abundance, we used data on genome-wide TFBS ChIP-seq: where available, from primary $CD4^+$ T cells; otherwise, from T cells or other human cell types; see Table S3 for complete listing of the datasets used.

In vitro integration sites showed a remarkably strong bias (compared with random sites) towards integration in proximity to specific TFBS, in particular STAT1, p53, HDACs (e.g. HDAC3, HDAC6) and HATs (e.g. p300, CBP) (Table S3). In most cases the effect was localized to within 100–1000 bases of the integration site (Figure 2A) and declined sharply at greater distances. Two patterns were observed in this biased integration. First, the preference towards integration in proximity to TFBS was typically symmetrical (e.g. p300), i.e. equally strong upstream and downstream of the integration site but in some cases was asymmetrical (e.g. STAT1), with a bias towards one side (often downstream). Second, in many cases we observed a sharp decrease in the preferential integration at 10 bases from the TFBS, such as STAT1 Figure 2A). This pattern was consistently observed across several in vitro and in vivo datasets (supplementary Figures S1, S2).

Because certain TFBS are frequently co-located in the human genome [22], we wished to test which TFBS were independently associated with targeting of the integration site. First, a likelihood ratio test was used to test whether the TFBS was selectively associated with integration either upstream or downstream of the integration site, and each TFBS was then tested individually using a univariate model. We then combined all significant factors using a step-down multivariate logistic regression analysis until only independently significant ($p<0.05$) factors remained. Most factors that were independently associated with integration site targeting occurred with equal frequency upstream or downstream of the integration site (Figure 2B, see also supplementary Table S7). The factors with the highest odds ratios were the transcription factor p53 and the histone deacetylase HDAC6.

### Effect of HTLV-1 integration sites on clonal expansion

We previously reported [11] a significant association between certain features in the flanking genome and in vivo expansion of the infected T-cell clone. Here, we found that proviruses integrated within a gene were more frequent in larger (more abundant) clones than in smaller clones in vivo, but only when the provirus was integrated in the same transcriptional orientation as the host gene (Figure 1D); the frequency of integration in the opposite orientation was not positively correlated with clonal abundance.

High clone abundance (Figure 1C, top two bins) was associated with the presence of a host TSS within 1 kb downstream of the provirus; here, the transcriptional orientation of the provirus had less effect on abundance than in the case of proviruses integrated within a host gene. The excess frequency of TSS downstream (but not upstream) was much higher in integration sites in vivo than in vitro, in particular when the provirus was integrated in the same orientation as the nearby host gene ($p(\text{same}) < 10^{-5}$; $p(\text{opposite}) < 0.05$, $\chi^2$ test). The presence of a host CpG island within 1 kb downstream was also selectively associated with clone high abundance (Figure 1E).

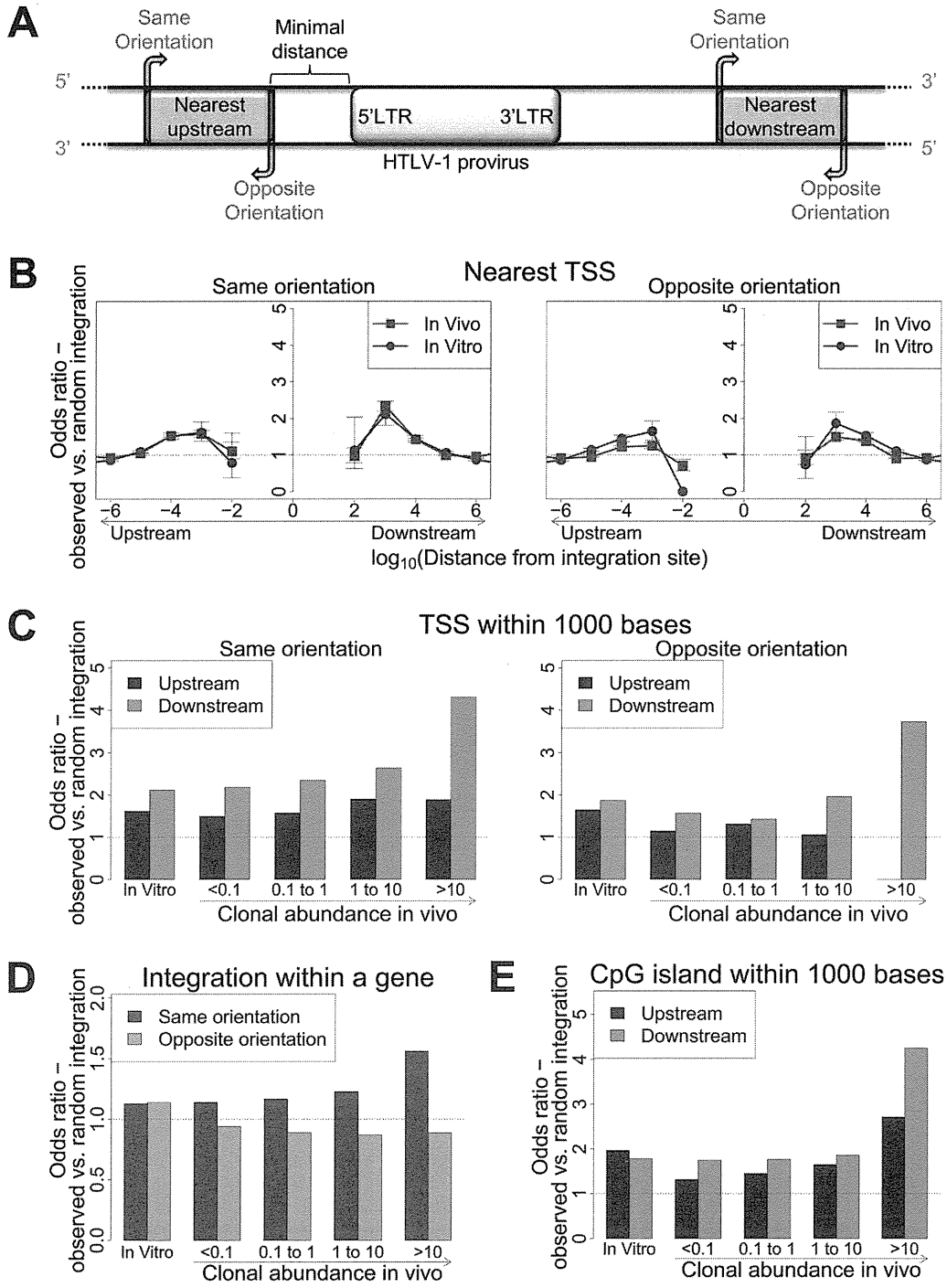
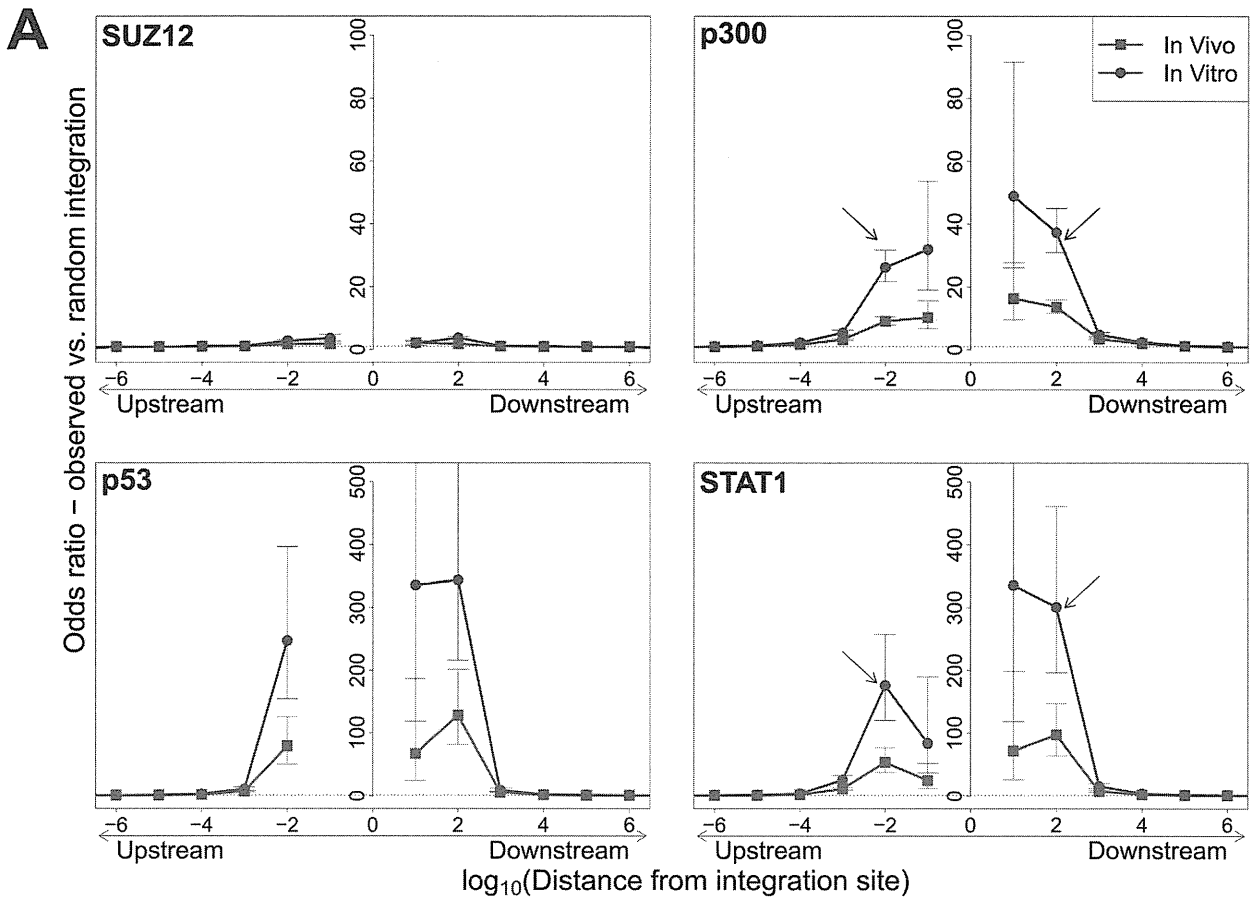



**Figure 1. Genomic environment at HTLV-1 proviral integration site determines integration in vitro and abundance in vivo.** (A) Blue blocks denote a genomic feature such as a transcription start site. The distance to the nearest genomic feature is calculated (unless otherwise stated) separately for features upstream (closer to 5′ LTR) and downstream of the provirus. Unless otherwise stated, distance is calculated to the nearest end of the genomic feature. Where the genomic feature has an orientation (i.e. transcription units) its orientation relative to the transcriptional orientation of the provirus is indicated as "same" or "opposite". (B) to (E): proportion of observed integration sites compared to random expectation. (B) Frequency of integration in proximity to transcriptional units (RefSeq). In vitro denotes a combined dataset from two independent experiments (see Table 1). (C) Frequency of integration within 1 kb of a TSS according to clonal abundance (cells in a given clone per 10 000 PBMCs). (D) The excess frequency (compared with random) of observing a provirus within a transcription unit was greater among abundant clones in vivo integrated in the same transcriptional orientation (blue) but not in opposite orientation (orange). (E) The excess frequency (compared with random) of observing a provirus within 1 kb of a host CpG island increased with increasing clonal abundance, in particular where the CpG island lay downstream of the integration site.

doi:10.1371/journal.ppat.1003271.g001



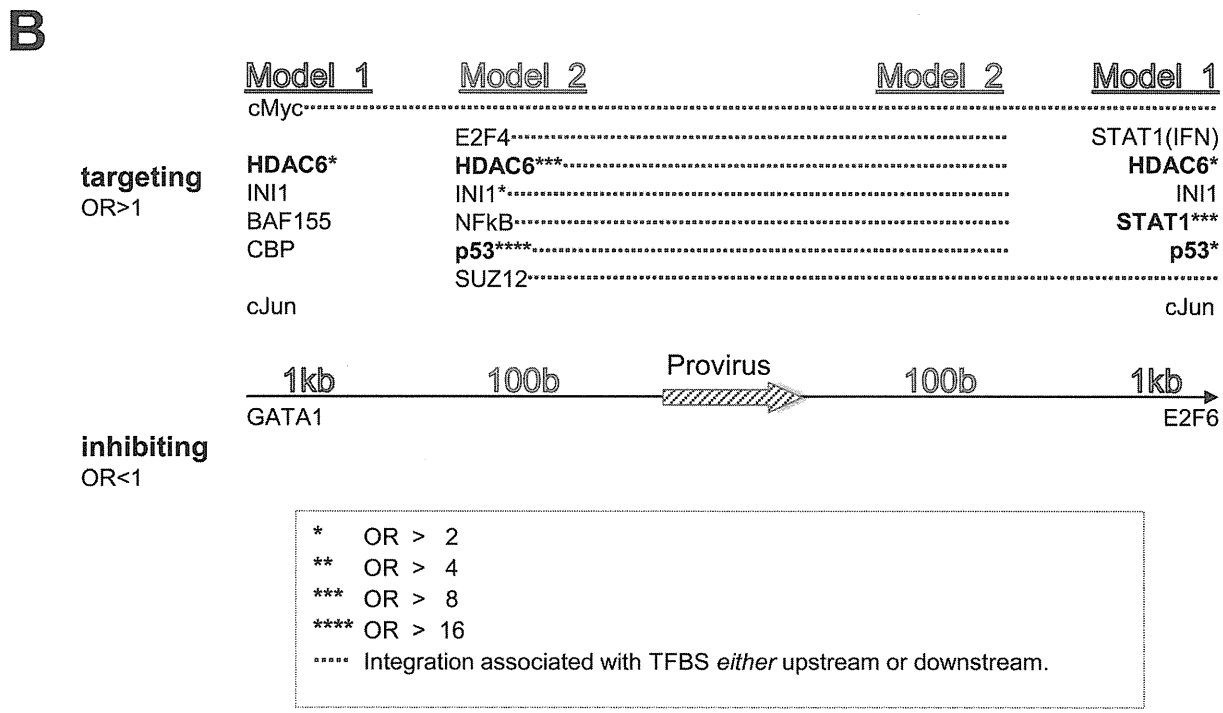




**Figure 2. Influence of host TFBS on integration site targeting.** (A) Bias in integration in proximity to TFBS (based on ChIP-seq experiments), measured by the odds ratio compared to random expectation. Four representative plots are shown; see also supplementary information. The excess frequency of integration in proximity to TFBS was frequently greater in in vitro infection than in clones isolated from PBMCs in vivo, and greater in low abundance clones in vivo than high abundance clones in vivo (see bottom right panel and supplementary information). Arrows indicate a

symmetrical (p300) or asymmetrical (STAT1) bias towards integration in proximity to TFBS, as well as a lower bias in close proximity to IS (STAT1). See also supplementary Table S4 for underlying data. (B) TFBS independently associated with integration frequency in vitro were identified by multivariate analysis. OR – odds ratio. TFBS shown above the line were associated with an excess frequency of integration compared with random (OR>1); TFBS below the line were significantly less likely to lie near the provirus (OR<1). Model 1 and Model 2 (carried out independently) test for TFBS within 1 kb and 100 bp of IS, respectively.
doi:10.1371/journal.ppat.1003271.g002

Integration sites observed in vivo showed a similar bias towards proximity to TFBS, with two important differences. First, the OR was in each case lower than that observed in in vitro integration. Second, the magnitude of the bias (OR) declined as clonal abundance increased (Figure 3; supplementary Figure S3).

## Effect of HTLV-1 integration site on Tax expression

We wished to identify features of the genomic integration site that favour expression of the HTLV-1 provirus. We hypothesized that the genomic environment flanking the proviral integration site determines the rate of spontaneous expression of the HTLV-1 transactivator protein Tax by a given infected T-cell clone: that is, the proportion of cells in that clone that express Tax within a given time interval. $CD8^+$ T-cells were depleted from fresh unstimulated PBMCs of 10 infected HAM/TSP patients (to preclude CTL-mediated lysis), and the $CD8^-$ population was incubated in vitro overnight to allow spontaneous expression of the Tax protein [15]. We then sorted the cells by flow cytometry to isolate $Tax^+$ and $Tax^-$ cells and analysed the integration sites in the two cell fractions.

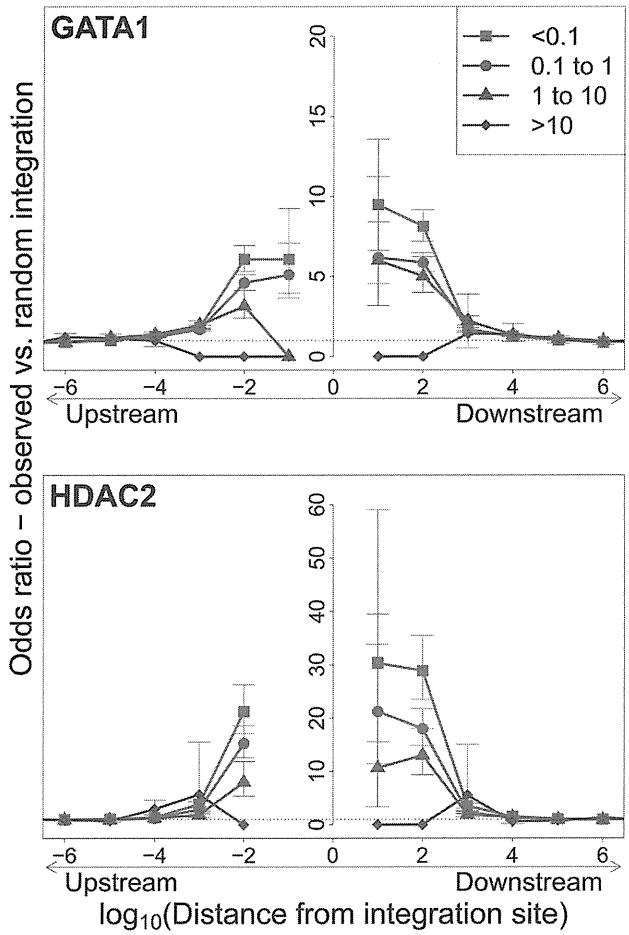



**Figure 3. Influence of host TFBS on clonal abundance.** Bias in integration in proximity to TFBS (based on ChIP-seq experiments), measured by the odds ratio compared to random expectation. Two representative plots are shown; see also supplementary Figure S3. The excess frequency of integration in proximity to TFBS was greater in low abundance clones in vivo than high abundance clones in vivo. See also supplementary Table S5 for underlying data.
doi:10.1371/journal.ppat.1003271.g003

We measured the proportion of each clone that spontaneously expressed Tax by quantifying individual integration sites in the $Tax^+$ and $Tax^-$ cells, (Figure 4E, and supplementary Figure S7). The observed proportion of $Tax^+$ cells per clone varied between 0% and 100%. The majority of clones, regardless of clonal abundance, were either >90% $Tax^+$ or >90% $Tax^-$. This observation is consistent with the hypothesis that the rate of spontaneous expression of Tax is an intrinsic property of each clone and is determined by the proviral integration site.

When the provirus was integrated within a host gene, we observed a slight but significant excess frequency of $Tax^+$ cells compared with $Tax^-$ cells (46% vs 43% respectively, $p<10^{-3}$, $\chi^2$ test). However, while the proviruses in the $Tax^+$ cells were found with equal frequency in the same or the opposite transcriptional orientation to the host gene in which they were integrated, the $Tax^-$ cells were significantly more frequently present in the same orientation as the host gene (52% of $Tax^+$ vs 59% of $Tax^-$ cells, $p<10^{-15}$, $\chi^2$ test). Thus, T cell clones that were 100% $Tax^-$ were significantly more likely to carry a provirus in the same orientation as the host gene (Figure 4B).

The relative position (upstream or downstream of the integration site) and the transcriptional orientation of the nearest host gene influenced not only the clonal abundance (Figure 1) but also spontaneous Tax expression. Where the nearest host gene lay in the same transcriptional orientation as the HTLV-1 provirus, the presence of a host TSS (Figure 4A) or CpG island (Figure 4C) within 1 kb upstream of the provirus was associated with silencing of Tax, whereas a TSS or CpG island within 1 kb downstream was associated with Tax expression. The closer the upstream gene was to the integration site, the lower was the proportion of $Tax^+$ cells if the gene was in the same orientation (Figure 4D). In contrast, where the nearest host gene was in opposite transcriptional orientation, this asymmetrical effect of the nearby host gene was not observed (Figure 4A, right-hand panel; Figure 4D).

The mean proportion of $Tax^+$ cells in one clone (across all clone abundance classes) was 60%. We wished to test whether proximity to TFBS would alter this proportion. We found that the presence of certain TFBS (including STAT1, cJun, NRSF) within 1 kb upstream of the integration site was associated with a higher proportion of $Tax^+$ cells in the respective T-cell clone (Figure 5A). A notable exception was BRG-1, which showed a strong opposite asymmetric effect: cells containing a BRG-1 site just upstream of the provirus were more likely to be $Tax^-$, whereas cells with a BRG-1 site just downstream of the provirus were more likely to be $Tax^+$ (Figure 5A, top left panel).

To identify the TFBS that were independently and significantly associated with spontaneous Tax expression, a logistic regression analysis was carried out as described above (Figure 2B) for integration site targeting. The results (Figure 5B, see also supplementary Table S7) confirmed the asymmetric effects of the BRG-1 binding site, and in addition revealed significant

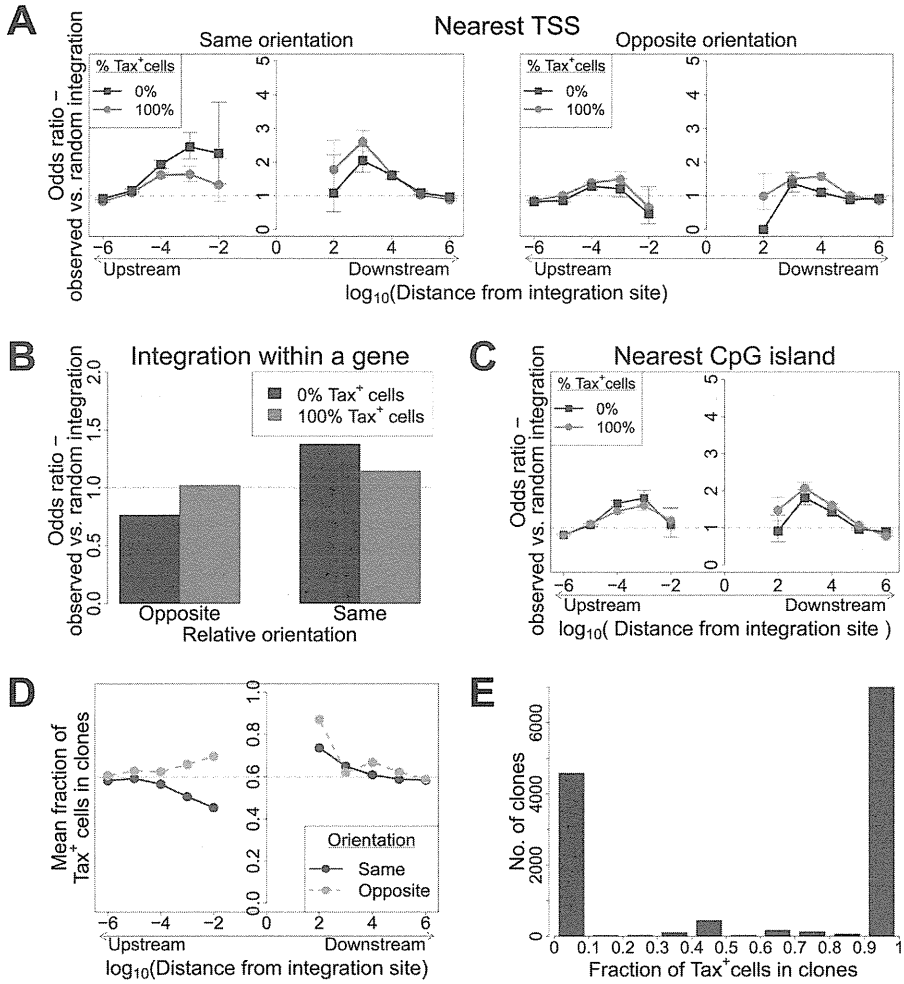
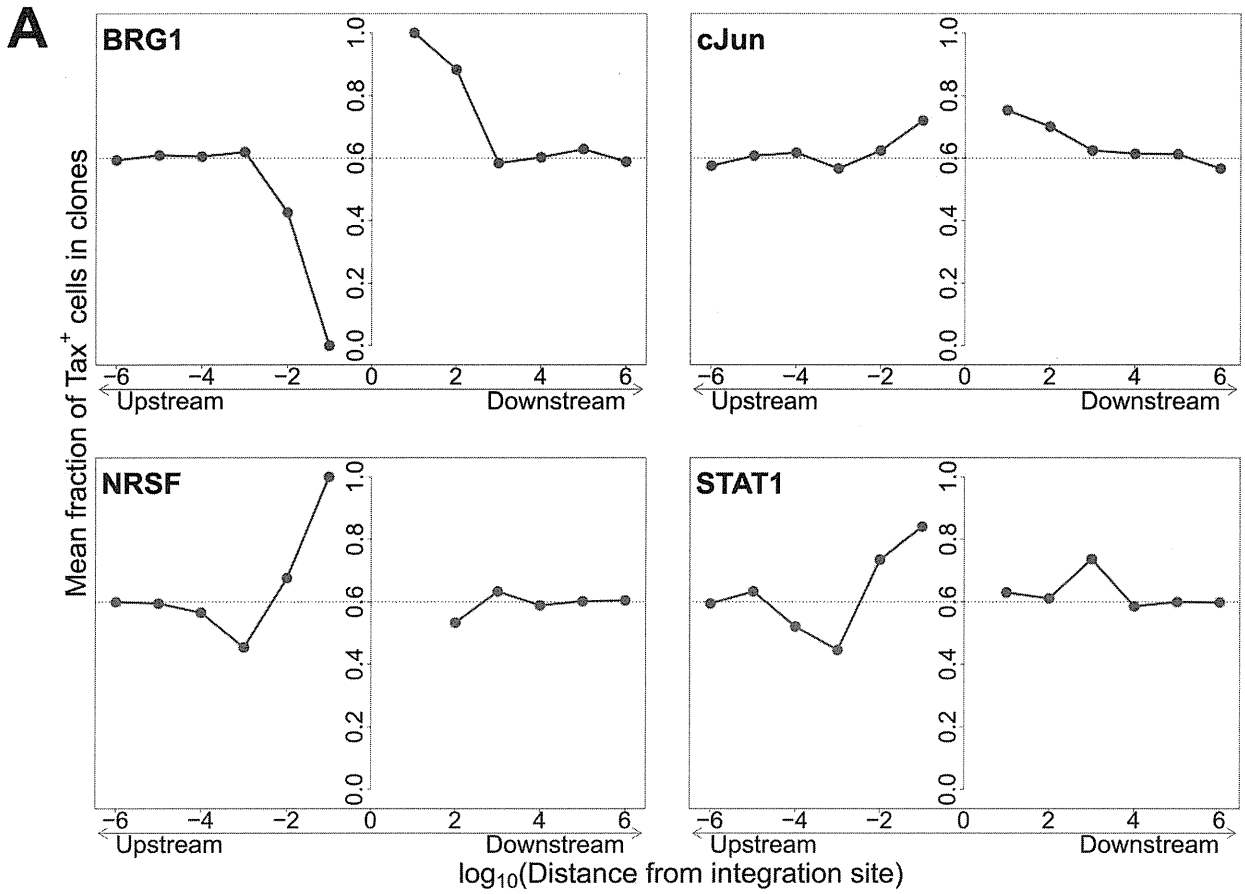



**Figure 4. Genomic environment at HTLV-1 proviral integration site associated with proviral expression after 18 h in culture.** CD8-depleted PBMCs were placed in culture overnight and sorted by flow cytometry to isolate $Tax^+$ and $Tax^-$ cells, followed by integration site analysis of sorted cells. (A)–(C): proportion of observed integration sites compared to random expectation. (A) Frequency of integration in proximity to transcriptional units (RefSeq) in clones that were 100% $Tax^+$ or 100% $Tax^-$, according to the relative transcriptional orientation of the provirus and the host gene. The peak of integration at 1 kb mirrors that observed in vivo in unsorted cells (Figure 1B). However, the integration site in $Tax^-$ clones was more likely than in $Tax^+$ clones to possess a nearby upstream TSS in the same orientation, and less likely to lie nearby a downstream TSS in the same orientation (or any relative position in the opposite orientation). (B) The provirus in $Tax^-$ clones (blue) was oriented in the same transcriptional sense as the host gene in which it was integrated more frequently than random. The orientation of $Tax^+$ clones (pink) did not differ from random. (C) Frequency of integration in proximity to CpG islands in clones that were 100% $Tax^+$ or 100% $Tax^-$. The peak of integration at 1 kb mirrors that observed in vivo in unsorted cells and in vitro (Figure S4). (D) Mean fraction of $Tax^+$ cells in clones with a TSS at a given distance (log scale) from the integration site, according to the relative transcriptional orientation of the provirus and the host TSS. The dotted line denotes the mean fraction of $Tax^+$ cells across all clones. (E) Frequency distribution of clones according to the frequency of $Tax^+$ cells in the respective clones. See supplementary Figure S7 for detailed frequency distribution separated according to clone abundance.
doi:10.1371/journal.ppat.1003271.g004

asymmetric associations between Tax expression and several other TFBS, notably STAT1, NRSF, and HDAC1. Thus, a STAT1 binding site 100 bp upstream of the provirus strongly favoured Tax expression, but the presence of a downstream STAT1 binding site was not an independent predictor of Tax expression after multivariate analysis. Conversely, an NRSF binding site 100 bp downstream was a significant predictor of Tax negativity, but the closest upstream NRSF binding site was not independently associated with Tax expression. The asymmetry of these associations contrasts with the predominantly symmetrical associations observed between TFBS and integration site targeting (Figure 2B), and suggests a mechanistic interaction between transcription of the provirus and transcription of the flanking host genome.

## $Tax^+$ cells are more frequent in low-abundance clones

To test the hypothesis that the level of Tax expression is correlated with the in vivo abundance of the infected T cell clone, we divided all detected clones into four abundance bins based on the total number of cells observed in each clone. There was a significant negative correlation between clone abundance and the proportion of $Tax^+$ cells in the respective abundance bin (Figure 5). That is, small clones were more likely to be $Tax^+$, and this likelihood decreased as clone abundance increased. We conclude that, at least in cells from HAM/TSP patients, the majority of spontaneous Tax expression observed is due to the large number of low-abundance clones, rather than a small number of high-abundance clones.



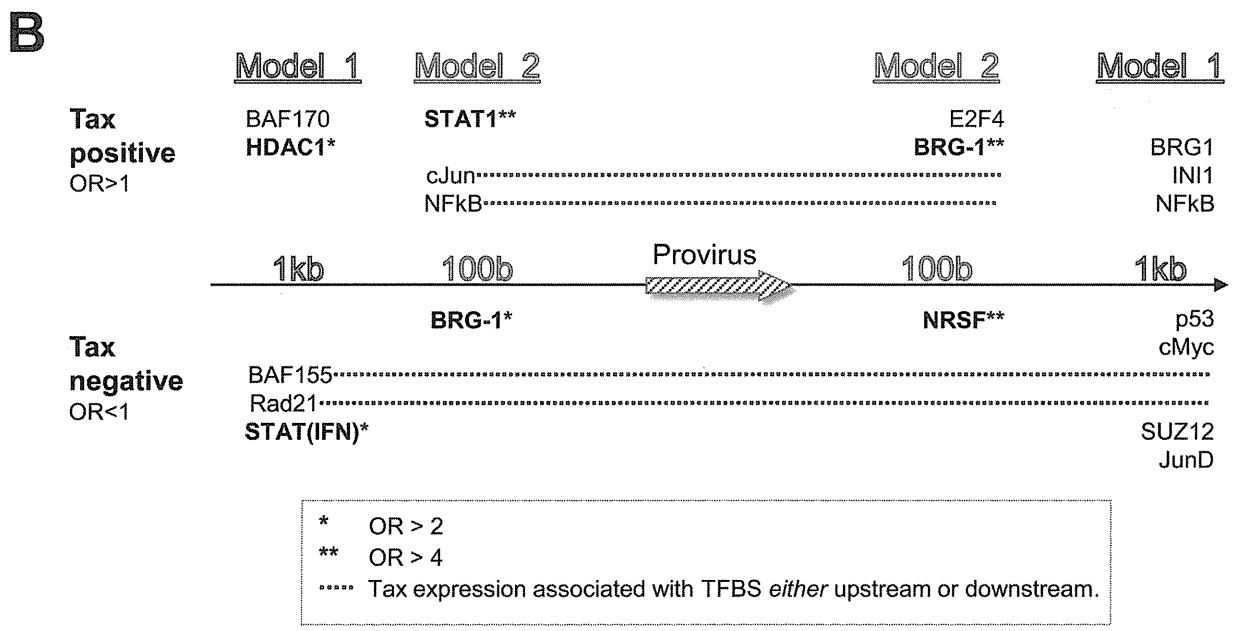




**Figure 5. Influence of proximity to TFBS on Tax expression.** (A) Mean fraction of Tax$^+$ cells in clones with a TFBS (based on ChIP-seq experiments) at a given distance from the IS. Four representative plots are shown. (B) TFBS that were independently associated with Tax expression were identified by multivariate analysis, outcome measure . TFBS shown above the line were associated with Tax expression (OR>1); TFBS below the line were associated with Tax silencing (OR<1). Model 1 and Model 2 (carried out independently) test for TFBS within 1 kb and 100 bp of IS, respectively.

doi:10.1371/journal.ppat.1003271.g005

## Discussion

An understanding of the regulation of proviral latency is required for attempts to eradicate latent retroviruses and to optimize retroviral vectors for in vitro and in vivo use. In HIV-1 infection, a reservoir of latently infected cells persists indefinitely in the face of antiretroviral drug therapy and precludes eradication of the infection (reviewed in [23]). In HTLV-1 infection, proviral expression is difficult to detect in fresh PBMCs: however, the strong, chronically activated host immune response and the selective oligoclonal proliferation of HTLV-1-infected T cells argue that the virus is continuously or intermittently expressed in vivo [4,24].

The abundance of an HTLV-1-infected T cell clone in vivo will be determined by the net effect of two main selection forces: its ability to proliferate and its susceptibility to killing by the strong CTL response [4]. If these forces acted upon all clones equally, the clones would have the same relative abundance in the host. However, Gillet et al [11] showed a wide variation in clone abundance both within and between infected individuals and over time. We hypothesized that this variation between clones is caused by the genomic environment of the integrated provirus, by determining the frequency and intensity of expression of proviral genes, in particular Tax and HBZ, which in turn promote cell proliferation and thereby confer a selective advantage on the infected T cell clone.

To identify the host genomic factors that determine integration site targeting, we mapped and quantified proviral integration sites isolated from two independent in vitro infection experiments. We assume that the pattern of integration observed in short-term in vitro infection reflects the initial pattern of integration in vivo, before the selection exerted during chronic infection. The results confirmed our previous observations [10,11] that the virus is targeted to transcriptionally active regions of the genome, within or near to a host gene. There was no bias in the orientation of the provirus in the initial infection, indicating that the bias observed in integration sites isolated from PBMCs is a result of the long-term selection forces acting on the infected clones in vivo.

We observed a bias towards integration in proximity to particular transcription factor binding sites. This bias was remarkably strong in certain cases (STAT1, NRSF) in single-factor analysis. Because clusters of different TFBS are frequent in the genome [22], we carried out a multivariate (logistic regression) analysis to identify the TFBS that were independently and significantly associated with an excess frequency of integration. The results (Figure 2B) confirmed the identification of p53, HDAC6 and STAT1 as significant independent correlates of integration. Further independent predictors of integration included Ini1 (see below), cMyc, cJun and NF-kB (Figure 2B). p53 and STAT1 both play important roles in HTLV-1 infection. HTLV-1 dysregulates p53 signalling pathways in vivo [25]; it is not known whether insertional mutagenesis contributes to this dysregulation. HTLV-1 also causes widespread activation of interferon-stimulated genes in vivo, including the key transcriptional regulator STAT1 [25]. A strong association was reported between STAT1 and MLV integration [26]; the authors attributed this to an association between MLV integration and particular epigenetic marks (H3K4me3, H3K4me1 and H3K9ac) at the integration site.

The proportion of all integration sites near any one TFBS was in the minority. This observation indicates that proximity to the transcription factor binding site itself is not sufficient for integration, but suggests that these transcription factors (or an associated host factor) increase the efficiency of proviral integration. Host factors associated with HIV integrase have been thoroughly studied [27]; the most important is the lens epithelium-derived growth factor (LEDGF/p75, [28]), which determines integrase localization [29] and targeting of HIV integrase to transcription units [21]. A study of host factors associated with HTLV-1 integrase is currently underway.

The observed bias towards integration near certain TFBS was predominantly symmetrical and short-range, reaching a maximum at 100b from the integration site and falling to random expectation at ~10 kb (Figure 2A). In many instances the bias dropped sharply at less than 100b from the integration site: we suggest that this drop is due to steric hindrance between the pre-integration complex and the DNA-bound transcription factor.

In contrast to the symmetry observed in the association between genomic features (such as TFBS) and the frequency of initial integration, we found significant asymmetric interactions between the flanking host genome and the integrated provirus in determining clonal abundance and spontaneous proviral expression. Both the relative position of the nearest host gene (upstream or downstream of the provirus) and its relative transcriptional orientation showed significant associations with clone abundance and expression. Previous studies [1,2] reported contradictory evidence on the role of an upstream same-sense host promoter in either promoting or suppressing proviral transcription. More recently, Shan et al [30] have shown in Bcl-2-transduced $CD4^+$ T cells, infected in vitro with GFP expressing modified HIV, that persistent expression of GFP was associated with opposite sense orientation, while inducible expression was associated with same sense orientation. The evidence obtained here demonstrates that, in natural HTLV-1 infection, the presence of a same-sense host gene promoter upstream of the integrated provirus is associated with inhibition of spontaneous proviral expression, suggesting the operation of transcriptional interference. We conclude that the transcriptional interaction between host and HTLV-1 operates at two levels. First, at a regional level – within 10 kb of the provirus – transcriptional activity of the flanking host genome favours proviral gene expression [10,11], presumably because of accessibility of the euchromatin to transcription complexes. Second, at a local level – within 100b to 1000b – transcriptional interference by a same-sense host promoter within 1 kb upstream can override the regional effect and inhibit proviral transcription.

Two observations reported here demonstrate that proviral integration and expression are not determined simply by the accessibility of chromatin. First, whereas some TFBS were associated with HTLV-1 proviral integration more frequently than expected, the frequency of other TFBS showed no such bias. Second, the asymmetric associations observed between proviral orientation and position with respect to flanking host genes and the abundance and expression of the HTLV-1 provirus argue for a mechanistic interaction between transcription of the HTLV-1 provirus and transcription of the flanking host genome.

An observation of particular interest is the opposing effect of a BRG-1 binding site upstream and downstream of the provirus (Figure 5A). BRG-1, one of the two ATPase components required for the activity of the SWI/SNF complex [31], controls gene expression by remodelling chromatin, i.e. by repositioning nucleosomes to control the access of transcriptional complexes to the DNA. BRG-1 can cause both gene repression [32] and gene activation [33]; the balance appears to depend on which other subunits are recruited to the SWI/SNF complex [34]. Easley et al [35] found that BRG-1 is required for Tax expression and HTLV-1 replication in vitro, and Rafati et al [36] found that the BAF subclass of the SWI/SNF complex repressed HIV-1 transcription whereas the PBAF subclass promoted transcription. Our observation (Figure 5A) that a BRG-1 site upstream of the provirus is

associated with silencing of Tax, while a BRG-1 site downstream is associated with Tax expression, is consistent with our conclusion (above) that transcriptional interference dominates the transcriptional interaction between the provirus and the flanking host genome.

The Ini1 subunit of SWI/SNF interacts directly with HIV-1 integrase [37]; a fraction of Ini1 moves transiently to the cytoplasm to associate with the HIV-1 preintegration complex [38]. However, the function of Ini1 in HIV-1 proviral integration and expression in vivo is not understood. We found that genomic sites for Ini1 binding were significantly associated with HTLV-1 proviral integration (Figure 2B). The presence of an Ini1 site 1 kb downstream of the provirus was associated with spontaneous Tax expression, similar to the effect of the downstream BRG-1 site. Finally, the SWI/SNF subunit BAF155 was overrepresented 1 kb upstream of the integration site (Figure 2B), and was associated with Tax silencing when present either upstream or downstream of the provirus (Figure 5B).

The HTLV-1 Tax protein acts in concert with host cell transcription factors (notably CBP/p300) on the promoter/enhancer in the viral 5′ LTR, driving plus-strand transcription in a strong positive feedback loop. Tax also acts on response elements for NF-kB, CREB and the serum response factor (SRE) to upregulate expression of a wide range of host genes [39]. Finally, Tax promotes cell cycle progression by accelerating passage through G1 and inhibiting the G1/S and G2/M checkpoints [40]. The net effect of Tax expression is therefore to drive activation and proliferation of the infected T cell. We previously reported that spontaneous Tax expression in fresh unstimulated PBMCs was associated with proliferation of the respective cell in vivo [41]. We therefore expected to observe a positive correlation between the frequency of spontaneous Tax expression by a given clone ex vivo and the abundance of that clone in vivo. However, the results obtained here (Figure 6) demonstrate the opposite, i.e. a highly significant negative correlation. This correlation is likely to be caused by the strong host immune response to the virus. The Tax protein is highly immunodominant in the class 1 MHC-restricted cytotoxic T lymphocyte (CTL) response to HTLV-1 [14,42], and the *tax* gene is frequently silenced in vivo by mutation or epigenetic changes such as DNA methylation in both untransformed and malignantly transformed (leukemic) cells [43]. Cells that express a high level of Tax are killed by CTLs faster than low Tax-expressing cells [44]. Therefore, suppression or loss of Tax expression may confer a survival advantage on the infected clone in vivo. We conclude that the small (low-abundance) HTLV-1-infected clones express Tax at a higher rate and turn over faster in vivo than the high-abundance clones. It is possible that the critical role of Tax in the HTLV-1 lifecycle is not to maintain clone abundance but rather to promote virion production and infection of new cells by cell contact via the virological synapse [45,46].

The negative correlation observed between clone abundance and the percentage of $Tax^+$ cells, although it was highly significant in all patients combined, was not uniform in every patient. In a small number of patients (in particular those with a high oligoclonality index, Supplementary Figure S8), the most abundant clones (bin 4, clones with greater than 10 cells) contained a high proportion of $Tax^+$ cells, suggesting that certain antigen-expressing clones escaped control by the immune response, for example by CTL escape mutations in the *tax* gene [47]. However, clone-specific sequencing of exon 3 of the *tax* gene of 38 clones (>10 cells) from 8 patients did not reveal significant differences in the occurrence of Tax mutations between clones with a high or low frequency of $Tax^+$ cells; and only in one patient was a

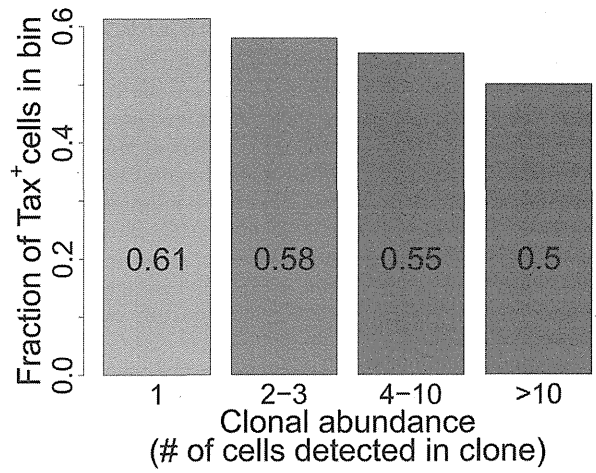



**Figure 6. $Tax^+$ cells are more frequent in smaller clones.** Mean fraction of $Tax^+$ cells in bins of increasing clonal abundance (total number of cells in each respective clone). The fraction of $Tax^+$ cells was negatively correlated with clonal abundance ($p<10^{-16}$, $\chi^2$ test for trend).
doi:10.1371/journal.ppat.1003271.g006

difference in amino acid sequence found between one clone and the others (data not shown).

Interestingly, while $Tax^+$ cells were more frequent in low-abundance clones, certain features favouring proviral expression (e.g. a downstream host TSS) also favoured clonal expansion. The association between clonal abundance and proviral integration within 1 kb of a downstream host TSS was maintained even within $Tax^-$ clones, consistent with the idea that the selective expansion of these clones is driven by other proviral genes.

These observations raise the possibility that the equilibrium abundance of an HTLV-1-infected T cell clone in vivo is determined not by Tax but by the HBZ gene, encoded on the negative strand of the provirus. Satou et al showed that *HBZ* mRNA promoted proliferation of the infected cell, and whereas Tax expression is frequently undetectable, HBZ appears to be persistently expressed in fresh cells isolated from both non-malignant cases of HTLV-1 infection and cases of adult T-cell leukemia/lymphoma [16]. Further, Macnamara et al recently showed that the CTL response to HBZ is a critical determinant of the equilibrium proviral load in vivo [17]

In this study we examined Tax expression only among $CD4^+$ T cells. A small percentage of infected cells are $CD8^+$ T cells [48,49]; it is possible that the genomic factors that determine targeting and expression of HTLV-1 differ in $CD8^+$ cells. Also, the propensity of a cell to express Tax was measured by quantifying the frequency of spontaneous Tax expression after 18 hours incubation in vitro. Two lines of evidence suggest that this measure is relevant to HTLV-1 infection and pathogenesis in vivo. First, cells which express Tax ex vivo turn over faster in vivo [41]. Second, the proportion of $CD4^+$ cells that express Tax after overnight culture is significantly associated with the HTLV-1 inflammatory disease HAM/TSP [50]. The individuals studied here were all patients with HAM/TSP: the mean level of spontaneous Tax expression is lower in asymptomatic HTLV-1 carriers, but it is unlikely that the molecular mechanisms that govern proviral latency differ qualitatively between asymptomatic carriers and patients with HAM/TSP.

It will be important to compare the present results with the genomic factors associated with HBZ expression or silencing. At

present this cannot be done by flow-sorting because existing HBZ-specific antibodies are insufficiently sensitive to detect the low expression levels of HBZ protein in primary cells. We are currently testing the hypothesis that Tax-specific and HBZ-specific CTL clones selectively lyse different clonal populations in vitro.

We have identified host genomic factors that determine the integration site, the proviral expression and selective clonal expansion of HTLV-1 in natural infection in vivo: these factors are summarized in Figure 7. These results open the way to test the molecular mechanisms involved.

## Materials and Methods

### Ethics statement

Blood samples were donated by HTLV-1-infected individuals attending the HTLV-1 clinic at the National Centre for Human Retrovirology (Imperial College Healthcare NHS trust) at St Mary's Hospital, London UK, with fully informed written consent. This study was approved by the UK National Research Ethics Service (NRES reference 09/H0606/106).

### DNA samples (Table 1)

PBMCs were isolated using Histopaque-1077 (Sigma-Aldrich) and cryopreserved in FBS (Gibco) containing 10% DMSO (Sigma-Aldrich). DNA extraction was carried out using the DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's protocol.

**Table 1.** IS datasets used.

| Dataset | total IS | total infected individuals | reference |
|---|---|---|---|
| in vitro (1) | 4521 | N/A | [11] |
| in vitro (2) | 1805 | N/A | This publication |
| In vivo (1) | 78563[1] | 63 | [11] |
| In vivo (2) | 20202 | 10 | This publication |
| Tax Negative | 6700 | 10 (pooled) | This publication |
| Tax Positive | 13054 | 10 (pooled) | This publication |
| Random UIS | 176505 | N/A | This publication |

[1]For the purpose of this work, only one time point was used for each patient (most recent available if multiple time points were originally analysed).
doi:10.1371/journal.ppat.1003271.t001

### In vitro infection

In vitro infection was carried out in two independent assays as previously described [11]. Jurkat (JKT) cells were co-cultured for 3 h with γ-irradiated ($^{137}$Cs, 40,000 cGy) MT2 cells [51], labelled with anti-CD4 MicroBeads (Miltenyi). MT2 cells were then depleted from the co-culture using magnetic separation (Miltenyi), and infected JKT cells were maintained in culture for 14 days in RPMI (supplemented by L-glutamine, penicillin, streptomycin) containing 10% FBS for 18 hours at 37C with 5% $CO_2$. Genomic

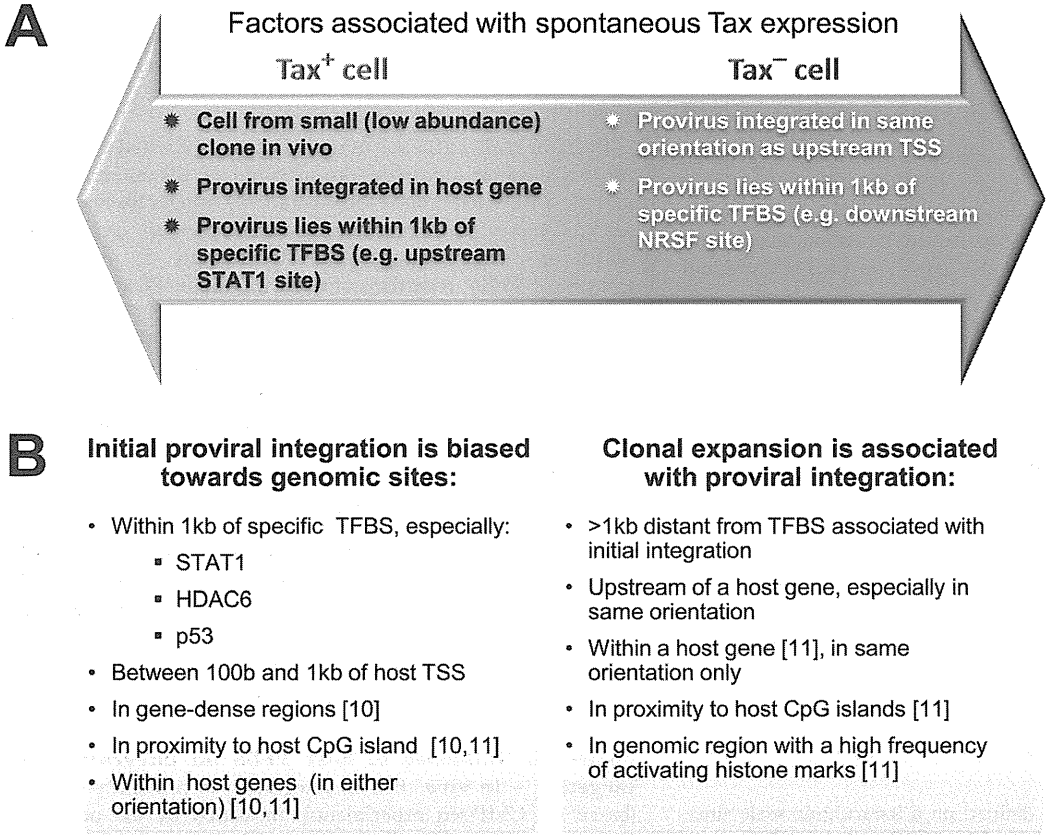



**Figure 7. Genomic correlates of HTLV-1 proviral targeting, clonal expansion and proviral expression.** (A) Factors associated with the presence or absence of spontaneous Tax expression by a given cell after short-term (18 h) in vitro incubation. (B) Features of the genomic environment of the provirus associated either with initial integration (left panel), or clonal expansion in vivo (right panel). Findings were made in the present study unless otherwise stated. TSS – transcription start site. TFBS – transcription factor binding site.
doi:10.1371/journal.ppat.1003271.g007

DNA was extracted and the proviral integration sites (IS) analysed as previously described [11]. IS from MT2 were also analysed to exclude possible contamination of the JKT IS. No contaminating MT2 IS were found after 14 days.

### Tax sorting

See also supplementary Figure S5, supplementary Table S6. PBMCs from 10 patients with HAM/TSP with a high proviral load (range 12.2–50.6 copies per 100 PBMC) were depleted of $CD8^+$ cells using magnetic depletion (Miltenyi) and incubated in RPMI (supplemented by L-glutamine, penicillin, streptomycin) containing 10% FBS for 18 hours at 37C with 5% $CO_2$. After 18 h culture, the cells were stained for intracellular expression of Tax (anti-Tax mAb LT4) and sorted using FACS (FACSAria IIIU, BD Biosciences) to isolate two populations of live $CD4^+$ cells based on Tax expression. Gates were set (FACSDiva, BD Biosciences) to ensure a clear demarcation between the $Tax^+$ and $Tax^-$ populations (Figure S6). DNA was extracted from whole unsorted PBMCs from each patient and analysed separately to identify the patient of origin of each clone; 46% of the clones were attributed in this way. To calculate the fraction of $Tax^+$ cells in a given clone, the frequency of $Tax^+$ and $Tax^-$ cells were normalized to the mass of genomic DNA per cell from each respective cell population, to correct for experimental variation in efficiency of genomic DNA isolation (Table S6),

### Analysis of IS

Identification and quantification of proviral integration sites was done as previously described [11]. HTLV-1 infected DNA was randomly sheared by sonication (Covaris S2) and then blunt-ended (Klenow polymerase) and ligated to a partly double-stranded DNA linker. Following a nested PCR step, the resulting DNA libraries were deep sequenced using the Illumina GA-II platform. DNA sequence was aligned to the human genome reference (UCSC hg18, excluding haplotype and "random" sequences) using the ELAND algorithm. Distinct IS were grouped based on integration site and quantified based on number of distinct shear sites isolated and the respective patient's proviral load.

DNA sequences from ~190000 random sites in the human genome (hg18) were generated using Galaxy [52,53,54] and back-aligned to the human genome using the same pipeline to eliminate any potential bias due to alignment limitations.

### Calculation of clonal abundance

The absolute abundance of a given clone was defined as the number of proviral copies of that clone per $10^4$ PBMCs. Given $n_i$ - the number of proviral copies for the $i^{th}$ clone, and S – the total number of clones identified in the sample, the absolute abundance was calculated for PBMC samples according to the following formula:

$$\text{absolute abundance} = \text{PVL} \times \frac{n_i}{\sum_{i=1}^{S} n_i}$$

Clone abundance bins were defined on a logarithmic scale since proviral load (used in calculation of abundance) follows a logarithmic distribution [55]. The number of clones in each clone abundance bin is given in Table S1. For samples sorted for Tax protein expression, where proviral load data were not available, the clonal abundance bins were set according to proviral copy count.

### Bioinformatic analysis of genomic environment

Transcription units and CpG island data were retrieved from the NCBI (ftp.ncbi.nih.gov/gene/) and UCSC tables [56], respectively. Annotations to the human genome were obtained from published datasets (Table S3) including ChIP-seq experiments on primary $CD4^+$ T cells where available; otherwise, data on human $CD4^+$ T cell lines or other human cell lines were used. We used the SISSRs algorithm [57] to identify the position of a putative transcription factor binding site in published ChIP-seq data where raw ChIP-seq data were available.

Annotations positions were compared to the IS using the R package hiAnnotator (http://malnirav.github.com/hiAnnotator), kindly provided by N. Malani and F. Bushman (University of Pennsylvania, USA).

### Statistical analysis

Statistical analysis was carried out using R version 2.13.0 (http://www.R-project.org/). Two separate logistic regression analyses were carried out, respectively, to identify independent predictors of HTLV-1 integration targeting and independent predictors of Tax positivity. Genomic annotations used to derive input variable were published ChIP-seq datasets (see Bioinformatic analysis above; Table S3). For integration targeting, the binary outcome measure was a "true" integration site (from 4521 identified in vitro integration sites) or a "false" integration site (45210 random genomic locations). For spontaneous Tax expression, the binary outcome was Tax positivity (20813 $Tax^+$ cells) or Tax negativity (10326 $Tax^-$ cells). Each TFBS was tested (presence or absence of the TFBS within a given distance of the integration site) as an independent predictor in each analysis. For each outcome variable, two separate analyses were carried out, respectively at two distances of the integration site - 100 bases and 1 kb.

First, for each TFBS and at each distance, we tested whether the relative position (upstream/downstream) of the integration site and the TFBS determined the outcome by using a likelihood ratio test to compare two competing models: 1) presence or absence of TFBS upstream or (separately) downstream; 2) presence or absence of TFBS, regardless of relative position. Next, we carried out univariate analysis of each individual TFBS, based on the model chosen by the likelihood ratio test. Only TFBS that were significant (p-value$<$0.05) after correction for multiple comparisons (Benjamini-Hochberg) were used in the multivariate analysis. Multivariate analysis was carried out using a step-down logistic regression method.

## Supporting Information

**Figure S1 Influence of host TFBS on integration site targeting – in vitro.** Bias in integration in proximity to TFBS (based on ChIP-seq experiments), measured by the odds ratio compared to random expectation. The bias was maintained across separate datasets, generated by independent in vitro experiments. Dotted line denotes random expectation (OR = 1).
(TIF)

**Figure S2 Influence of host TFBS on integration site targeting – in vivo.** Bias in integration in proximity to TFBS (based on ChIP-seq experiments), measured by the odds ratio compared to random expectation. The pattern of bias was maintained between different patient clinical groups. Dotted line denotes random expectation (OR = 1). ATLL = Adult T-cell leukaemia/lymphoma . HAM/TSP = HTLV-1 associated myelopathy/Tropical spastic paraparesis. AC = Asymptomatic carrier.
(TIF)

**Figure S3 Influence of host TFBS on clonal abundance in vivo.** Bias in frequency of integration in proximity to TFBS (based on ChIP-seq experiments), measured by the odds ratio compared to random expectation. TFBS that were associated with integration targeting showed a stronger bias (higher OR) in the clones least expanded in vivo. Clonal abundance is expressed as the number of cells in given clone per $10^4$ PBMCs. Dotted line denotes random expectation (OR = 1).
(TIF)

**Figure S4 The genomic environment at the HTLV-1 proviral integration site determines integration targeting in vitro and clonal abundance in vivo.** Frequency of integration in proximity to CpG islands in clones for in vitro (in blue) and in vivo (purple) integration.
(TIF)

**Figure S5 Protocol for flow-sorting of Tax-expressing cells.** (A) $CD8^+$ cell-depleted PBMCs were studied from 10 patients with HAM/TSP with a high proviral load. The cells were incubated overnight, fixed and stained for Tax and surface CD4 expression, and sorted on a high-speed flow cytometer (see Figure S6 for details). (B) Recovered cells from all 10 patients were combined in two pools, respectively $CD4^+Tax^+$ cells and $CD4^+Tax^-$ cells. (C) Genomic DNA was extracted from each pool of cells and integration site analysis carried out as described.
(TIF)

**Figure S6 Flow cytometry sorting by Tax expression.** (A) Representative FACS plots of the gating procedure used (from 1 of 10 samples studied). Lower middle panel shows gating of $CD4^+Tax^+$ ('Tax pos') and $CD4^+Tax^-$ ('Tax neg') populations; these gates were set to distinguish unequivocally between $Tax^+$ and $Tax^-$ populations. (B) Purity testing of $Tax^-$ sorted cells: $Tax^+$ cells not detected. C) Purity testing of $Tax^+$ sorted cells: 0.2% were $Tax^-$.
(TIF)

**Figure S7 Majority of HTLV-1-infected clones were either 100% $Tax^+$ or 0% $Tax^+$.** Frequency distribution of clones according to the frequency of $Tax^+$ cells in each respective clone, binned according to number of sister cells detected in sample: bin 1: 1 cell detected; bin 2: 2 or 3 cells detected; bin 3: 4 to 10 cells detected; bin 4: over 10 cells detected.
(TIF)

**Figure S8 $Tax^+$ cells were more frequent in smaller clones.** Mean fraction of $Tax^+$ cells within each bin, in bins of increasing clonal abundance (total number of cells in each respective clone). In the majority of patients there was an inverse correlation between clone abundance bin and fraction of $Tax^+$ cells: this correlation was highly significant in all patients combined ($P<10^{-32}$). However in certain patients (particularly those with a high oligoclonality Index) the most abundant clones contained a high proportion of $Tax^+$ cells. Clone abundance bins were defined as in Figure S7.
(TIF)

**Table S1 In vivo integration sites – sample data by clone abundance.**
(DOC)

**Table S2 In vivo integration sites – sample data by patient code.**
(DOC)

**Table S3 List of annotations datasets used.**
(DOC)

**Table S4 Odds ratio and clone counts data for a selection of TFBS – in vitro, in vivo vs. random sites.**
(XLS)

**Table S5 Odds ratio and clone counts data for a selection of TFBS – clonal abundance bins vs. random sites.**
(XLS)

**Table S6 Tax sorting experiment – sample data.**
(DOC)

**Table S7 Multivariate analysis results – detailed odds ratios and confidence intervals.**
(XLS)

## Acknowledgments

We thank Nirav Malani and Frederic D. Bushman at the department of Microbiology, University of Pennsylvania, Philadelphia, PA, USA for the list of random integration sites and for developing software packages; and Laurence Game, Nathalie Lambie and Adam Giess of the Genomics Laboratory at the MRC Clinical Sciences Centre, Hammersmith Hospital, London UK and Robert Sampson at the flow cytometry facility at St Mary's Campus, Imperial College, London UK. Finally we thank Lucy Cook, Heather Niederer and Becca Asquith and her team at the Section of Immunology (Imperial College) for discussions and comments, and the staff and donors at the National Centre for Human Retrovirology (Imperial College Healthcare NHS trust) at St Mary's Hospital, London UK.

## Author Contributions

Conceived and designed the experiments: AM GPT CRMB. Performed the experiments: AM NAG. Analyzed the data: AM. Contributed reagents/materials/analysis tools: YT GPT NAG DJL. Wrote the paper: AM CRMB. Clinical diagnosis and assessment: GPT. Designed the statistical analysis: AM DJL CRMB.

## References

1. Han Y, Lin YB, An W, Xu J, Yang HC, et al. (2008) Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe 4: 134–146.
2. Lenasi T, Contreras X, Peterlin BM (2008) Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe 4: 123–133.
3. Igakura T, Stinchcombe JC, Goon PK, Taylor GP, Weber JN, et al. (2003) Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. Science 299: 1713–1716.
4. Bangham CR (2009) CTL quality and the control of human retroviral infections. Eur J Immunol 39: 1700–1712.
5. Iwanaga M, Watanabe T, Utsunomiya A, Okayama A, Uchimaru K, et al. (2010) Human T-cell leukemia virus type I (HTLV-1) proviral load and disease progression in asymptomatic HTLV-1 carriers: a nationwide prospective study in Japan. Blood 116: 1211–1219.
6. Matsuzaki T, Nakagawa M, Nagai M, Usuku K, Higuchi I, et al. (2001) HTLV-I proviral load correlates with progression of motor disability in HAM/TSP: analysis of 239 HAM/TSP patients including 64 patients followed up for 10 years. J Neurovirol 7: 228–234.
7. Wattel E, Vartanian JP, Pannetier C, Wain-Hobson S (1995) Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. J Virol 69: 2863–2868.
8. Cook LB, Rowan AG, Melamed A, Taylor GP, Bangham CR (2012) HTLV-1-infected T cells contain a single integrated provirus in natural infection. Blood 120: 3488–3490.
9. Derse D, Crise B, Li Y, Princler G, Lum N, et al. (2007) Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. J Virol 81: 6731–6741.
10. Meekings KN, Leipzig J, Bushman FD, Taylor GP, Bangham CR (2008) HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. PLoS Pathog 4: e1000027.
11. Gillet NA, Malani N, Melamed A, Gormley N, Carter R, et al. (2011) The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. Blood 117: 3113–3122.

12. Jacobson S, Shida H, McFarlin DE, Fauci AS, Koenig S (1990) Circulating CD8+ cytotoxic T lymphocytes specific for HTLV-I pX in patients with HTLV-I associated neurological disease. Nature 348: 245–248.
13. Parker CE, Daenke S, Nightingale S, Bangham CR (1992) Activated, HTLV-1-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. Virology 188: 628–636.
14. Goon PK, Biancardi A, Fast N, Igakura T, Hanon E, et al. (2004) Human T cell lymphotropic virus (HTLV) type-1-specific CD8+ T cells: frequency and immunodominance hierarchy. J Infect Dis 189: 2294–2298.
15. Hanon E, Hall S, Taylor GP, Saito M, Davis R, et al. (2000) 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.
16. 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 U S A 103: 720–725.
17. Macnamara A, Rowan A, Hilburn S, Kadolsky U, Fujiwara H, et al. (2010) HLA class I binding of HBZ determines outcome in HTLV-1 infection. PLoS Pathog 6: e1001117.
18. Hilburn S, Rowan A, Demontis MA, MacNamara A, Asquith B, et al. (2011) In vivo expression of human T-lymphotropic virus type 1 basic leucine-zipper protein generates specific CD8+ and CD4+ T-lymphocyte responses that correlate with clinical outcome. J Infect Dis 203: 529–536.
19. Landry S, Halin M, Vargas A, Lemasson I, Mesnard JM, et al. (2009) Upregulation of human T-cell leukemia virus type 1 antisense transcription by the viral tax protein. J Virol 83: 2048–2054.
20. Gaudray G, Gachon F, Basbous J, Biard-Piechaczyk M, Devaux C, et al. (2002) The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. J Virol 76: 12813–12822.
21. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11: 1287–1289.
22. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.
23. Durand CM, Blankson JN, Siliciano RF (2012) Developing strategies for HIV-1 eradication. Trends Immunol 33: 554–562.
24. Bangham CR, Meekings K, Toulza F, Nejmeddine M, Majorovits E, et al. (2009) The immune control of HTLV-1 infection: selection forces and dynamics. Front Biosci 14: 2889–2903.
25. Tattermusch S, Skinner JA, Chaussabel D, Banchereau J, Berry MP, et al. (2012) Systems biology approaches reveal a specific interferon-inducible signature in HTLV-1 associated myelopathy. PLoS Pathog 8: e1002480.
26. Santoni FA, Hartley O, Luban J (2010) Deciphering the code for retroviral integration target site selection. PLoS Comput Biol 6: e1001008.
27. Rain JC, Cribier A, Gerard A, Emiliani S, Benarous R (2009) Yeast two-hybrid detection of integrase-host factor interactions. Methods 47: 291–297.
28. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem 278: 372–381.
29. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, et al. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem 278: 33528–33539.
30. Shan L, Yang HC, Rabi SA, Bravo HC, Shroff NS, et al. (2011) Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. J Virol 85: 5384–5393.
31. Euskirchen G, Auerbach RK, Snyder M (2012) SWI/SNF Chromatin-remodeling Factors: Multiscale Analyses and Diverse Functions. J Biol Chem 287: 30897–30905.
32. Ho L, Jothi R, Ronan JL, Cui K, Zhao K, et al. (2009) An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci U S A 106: 5187–5191.
33. De S, Wurster AL, Precht P, Wood WH, 3rd, Becker KG, et al. (2011) Dynamic BRG1 recruitment during T helper differentiation and activation reveals distal regulatory elements. Mol Cell Biol 31: 1512–1527.
34. Euskirchen GM, Auerbach RK, Davidov E, Gianoulis TA, Zhong G, et al. (2011) Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. PLoS Genet 7: e1002008.
35. Easley R, Carpio L, Guendel I, Klase Z, Choi S, et al. (2010) Human T-lymphotropic virus type 1 transcription and chromatin-remodeling complexes. J Virol 84: 4755–4768.
36. Rafati H, Parra M, Hakre S, Moshkin Y, Verdin E, et al. (2011) Repressive LTR nucleosome positioning by the BAF complex is required for HIV latency. PLoS Biol 9: e1001206.
37. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science 266: 2002–2006.
38. Turelli P, Doucas V, Craig E, Mangeat B, Klages N, et al. (2001) Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. Mol Cell 7: 1245–1254.
39. Matsuoka M, Jeang KT (2011) Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy. Oncogene 30: 1379–1389.
40. Marriott SJ, Semmes OJ (2005) Impact of HTLV-I Tax on cell cycle progression and the cellular DNA damage repair response. Oncogene 24: 5986–5995.
41. Asquith B, Zhang Y, Mosley AJ, de Lara CM, Wallace DL, et al. (2007) In vivo T lymphocyte dynamics in humans and the impact of human T-lymphotropic virus 1 infection. Proc Natl Acad Sci U S A 104: 8035–8040.
42. Kannagi M, Harada S, Maruyama I, Inoko H, Igarashi H, et al. (1991) Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8+ cytotoxic T cells directed against HTLV-I-infected cells. Int Immunol 3: 761–767.
43. Taniguchi Y, Nosaka K, Yasunaga J, Maeda M, Mueller N, et al. (2005) Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. Retrovirology 2: 64.
44. Kattan T, MacNamara A, Rowan AG, Nose H, Mosley AJ, et al. (2009) The avidity and lytic efficiency of the CTL response to HTLV-1. J Immunol 182: 5723–5729.
45. Nejmeddine M, Barnard AL, Tanaka Y, Taylor GP, Bangham CR (2005) Human T-lymphotropic virus, type 1, tax protein triggers microtubule reorientation in the virological synapse. J Biol Chem 280: 29653–29660.
46. Nejmeddine M, Negi VS, Mukherjee S, Tanaka Y, Orth K, et al. (2009) HTLV-1-Tax and ICAM-1 act on T-cell signal pathways to polarize the microtubule-organizing center at the virological synapse. Blood 114: 1016–1025.
47. Niewiesk S, Daenke S, Parker CE, Taylor G, Weber J, et al. (1995) Naturally occurring variants of human T-cell leukemia virus type I Tax protein impair its recognition by cytotoxic T lymphocytes and the transactivation function of Tax. J Virol 69: 2649–2653.
48. Richardson JH, Edwards AJ, Cruickshank JK, Rudge P, Dalgleish AG (1990) In vivo cellular tropism of human T-cell leukemia virus type 1. J Virol 64: 5682–5687.
49. Hanon E, Stinchcombe JC, Saito M, Asquith BE, Taylor GP, et al. (2000) Fratricide among CD8(+) T lymphocytes naturally infected with human T cell lymphotropic virus type I. Immunity 13: 657–664.
50. Asquith B, Mosley AJ, Heaps A, Tanaka Y, Taylor GP, et al. (2005) Quantification of the virus-host interaction in human T lymphotropic virus I infection. Retrovirology 2: 75.
51. Miyoshi I, Kubonishi I, Sumida M, Hiraki S, Tsubota T, et al. (1980) A novel T-cell line derived from adult T-cell leukemia. Gann 71: 155–156.
52. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15: 1451–1455.
53. Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11: R86.
54. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, et al. (2010) Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol Chapter 19: Unit 19 10 11–21.
55. Nagai M, Usuku K, Matsumoto W, Kodama D, Takenouchi N, et al. (1998) Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP. J Neurovirol 4: 586–593.
56. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, et al. (2004) The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32: D493–496.
57. Jothi R, Cuddapah S, Barski A, Cui K, Zhao K (2008) Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res 36: 5221–5231.

ELSEVIER

Institut Pasteur

Microbes and Infection 15 (2013) 491–505

www.elsevier.com/locate/micinf

Original article

# Viral interference with host mRNA surveillance, the nonsense-mediated mRNA decay (NMD) pathway, through a new function of HTLV-1 Rex: implications for retroviral replication

Kazumi Nakano [a], Tomomi Ando [a,b], Makoto Yamagishi [a], Koichi Yokoyama [a], Takaomi Ishida [c], Takeo Ohsugi [d], Yuetsu Tanaka [e], David W. Brighty [f], Toshiki Watanabe [a,*]

[a] *Laboratory of Tumor Cell Biology, Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan*

[b] *Department of Virology II, National Institute of Infectious Diseases, 1-23-1, Toyama, Shinjukuku, Tokyo 162-8640, Japan*

[c] *Research Center for Asian Infectious Diseases, The Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan*

[d] *Center for Animal Resources and Development, The University of Kumamoto, 2-2-1, Honsho, Kumamoto 860-0811, Japan*

[e] *Department of Immunology, Graduate School of Medicine, University of the Ryukyus, 207 Uehara, Nishihara-cho, Nakagusuku, Okinawa 903-0215, Japan*

[f] *Division of Cancer Research, Medical Research Institute, University of Dundee, Scotland DD1 9SY, UK*

Received 9 January 2013; accepted 18 March 2013
Available online 27 March 2013

**Abstract**

Nonsense-mediated mRNA decay (NMD) is an essential and conserved cellular mRNA quality control mechanism. RNA signals to express viral genes from overlapping open reading frames potentially initiate NMD, nevertheless it is not clear whether viral RNAs are sensitive to NMD or if viruses have evolved mechanisms to evade NMD. Here we demonstrate that the genomic and full-length mRNAs of Human-T-cell Leukemia Virus type-I (HTLV-1), a retrovirus responsible for Adult T-cell Leukemia (ATL), are sensitive to NMD. They exhibit accelerated turnover in NMD-activated cells, while siRNA-mediated knockdown of NMD-master-regulator, UPF1, promotes enhanced stability of them. These effects on RNA stability were recapitulated by a reporter construct encoding the HTLV-1 translational frameshift signal of *gag-pol*. In agreement with the RNA stability, viral protein expression from the integrated provirus was inversely correlated with cellular NMD activity. We further demonstrated that the viral RNA-binding protein, Rex, approves the stability of viral RNA by inhibiting NMD. Significantly, Rex establishes a general block to NMD, as both NMD-responsive reporter transcripts and natural host-encoded NMD substrates were stabilized in the presence of Rex. Thus, we suggest that Rex not only stabilizes viral transcripts, but also perturbs cellular mRNA metabolism and host cell homeostasis via inhibition of NMD.
© 2013 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

*Keywords:* HTLV-1; HTLV-1 Rex; NMD; Retroviral genomic RNA; Host–pathogen interaction

## 1. Introduction

Nonsense-mediated mRNA decay (NMD) is an mRNA surveillance mechanism that is conserved in eukaryotic cells. The degradation of aberrant mRNAs containing premature termination codons (PTCs) is the most studied NMD function. Positioned upstream of the natural end of open reading frames (ORFs), PTCs are stop codons that arise from frameshifts due to mutations or aberrant mRNA processing events. Truncated proteins encoded by such abnormal mRNAs are often deleterious to cells because they may be structurally unstable and result in translation product aggregation or may function as dominant negative inhibitors of wild-type (WT) protein function [1]. Recently, it has been recognized that NMD function is important for eliminating aberrant mRNAs and

* Corresponding author. Tel.: +81 3 5449 5298; fax: +81 3 5449 5418.
*E-mail addresses:* tnabe@ims.u-tokyo.ac.jp, tnabe@k.u-tokyo.ac.jp (T. Watanabe).

1286-4579/$ - see front matter © 2013 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.micinf.2013.03.006

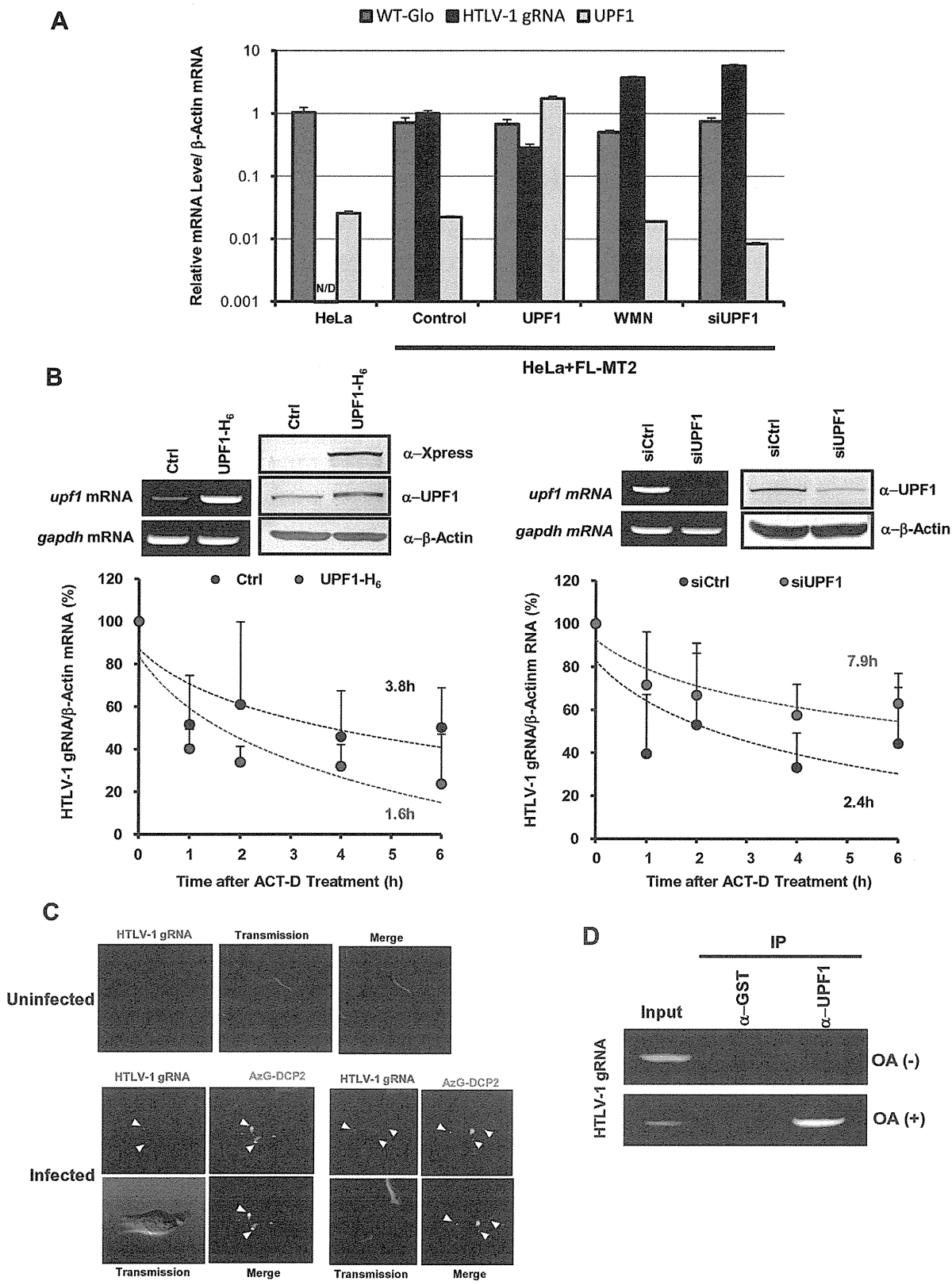



Fig. 1. HTLV-1 genomic unspliced RNA is a NMD target. (A) Changes in the cellular level of HTLV-1 unspliced mRNA were dependent on cellular NMD activity. The steady-state level of HTLV-1 genomic primary transcripts (HTLV-1 gRNA) accumulating from the pFL-MT2 infectious clone in HeLa cells was decreased by UPF1 overexpression and increased by NMD inhibition through siUPF1 transfection or wortmannin treatment. By contrast, the level of WT-β-globin mRNA (WT-Glo), which is not a NMD target, was not influenced by the cellular level of UPF1 or wortmannin treatment, indicating that the level of HTLV-1 genomic mRNA is selectively influenced by cellular NMD activity. (B) UPF1-dependent HTLV-1 genomic mRNA instability in HTLV-1 infected HeLa cells by co-cultivation with

controlling the expression levels of a considerable number of normal cellular mRNAs that possess NMD-activating structures, such as uORFs and introns within the 3′ untranslated region (UTR) [2]. The NMD machinery achieves these functions by coupling with the splicing and translational machinery [3,4]. Thus, NMD is an essential mechanism that governs mRNA quality and quantity in eukaryotic cells.

RNA viruses have compact genomes, but they have evolved elegant mechanisms to maximize coding potential and precisely regulate the expression of encoded genes. Overlapping reading frames, internal ribosome entry sites, alternative splicing, suboptimal Kozak sequences, and ribosomal frameshifting are among the varied mechanisms used to maximize genomic coding potential and regulate viral gene expression [5]. The presence of such signals in cellular mRNA is unusual, but wherever they occur, there is significant potential to activate the NMD pathway, which destabilizes RNA and increases mRNA turnover [6–8]. Programmed Ribosomal Frameshift (PRF) is a mechanism frequently used by viruses to alter the translational reading frame by shifting the ribosome at a frameshifting sequence often referred to as a "slippery site" [9]. Especially, *pgk1* mRNA stability assays with or without L-A viral −PRF signal showed that −PRF itself functioned as *cic*-acting destabilization element through NMD [7]. In addition, as many as one-third of all cellular mRNA variants produced by alternative splicing are reported to be potential targets of NMD [6,8]. As reviewed by Dickson and Wilusz [10], virus–host mRNA surveillance interface is attracting growing interest as a new aspect of host–pathogen interactions. Nevertheless, viral mechanisms to evade host mRNA surveillance to protect its own RNA has been mostly left uninvestigated. Accumulated knowledge indicates that viruses have various strategies to avoid or incapacitate host mRNA decay. Therefore, for RNA viruses, there are important unresolved questions. First, are RNA viruses that possess overlapping reading frames, alternative stop codons, and translational frameshift sequences sensitive to NMD? If so, do RNA viruses actively avoid the NMD surveillance pathway? Finally, are viral factors required for evasion of host-encoded NMD?

The human T-cell leukemia virus type I (HTLV-1) is a delta retrovirus that causes aggressive adult T-cell leukemia (ATL) in some infected individuals. The genomic RNA of this retrovirus encodes more than ten ORFs with associated stop codons within a full-length genomic RNA of 8685 nucleotides. HTLV-1 employs a number of mechanisms to achieve appropriate and ordered expression of these genes, including alternative splicing and PRF. In particular, Gag, Pro, and Pol, is translationally regulated by in-frame read-through and two −1 PRF signals at 1718 and 2245 nucleotides, respectively. In addition, the HTLV-1 genomic RNA contains two major splice sites. Unspliced HTLV-1 RNA yields Gag, Pro, and Pol proteins, singly spliced RNA produces Env, while the functional proteins derived from the pX region can be translated only from double-spliced mRNA. Therefore, we hypothesized that because of the unusual structure and processing signals, which include multiple stop codons, translational frameshifts, and a downstream splice acceptor site, the full-length HTLV-1 genomic RNA and *gag/pol* mRNA appear to be prime candidates for interference with viral mRNA accumulation via NMD.

Here we explored whether the presence of multiple stop codons and −1 PRF signals in HTLV-1 viral transcripts activates the host NMD. Also, we investigated how such viral RNA survives in the face of a highly efficient cellular mRNA quality control system.

## 2. Materials and methods

### 2.1. NMD reporter constructs and assays

To measure overall cellular NMD activity, we constructed NMD reporter plasmids based on widely used β-globin as well as on a HTLV-1 *gag* mRNA fragment. The β-globin-based NMD reporter plasmid was created based on Boelz et al. [11] with modification dependent on the type of experiment. Detailed methods of reporter construction and luciferase assays are available in Supplementary material.

### 2.2. HTLV-1 expression in HeLa cells

To express HTLV-1 in HeLa cells, transfection of an infectious clone (pFL-MT2) [12] (for Fig. 1A) or co-cultivation with MT-2 (for Fig. 1B–D, 3C–E, 6B–D) was employed. Please see Supplementary material for more detail.

### 2.3. Inhibition of UPF1 and UPF2

The siRNA sequences against *upf1* mRNA and *upf2* mRNA are described elsewhere [13,14]. The sequence of the control siRNA was 5′-AGGUCGAACUACGGGUCAA(TT)-3′. Wortmannin is a PI3K inhibitor and known to inhibit UPF1 activity by inhibiting SMG1, the kinase of UPF1. For complete inhibition of UPF1 activity, cells were treated with 100 μM wortmannin for 19–20 h before sampling. For construction of

MT-2 cells. The stability of HTLV-1 genomic mRNA decreased in the cells overexpressing His-tagged UPF1 (UPF1-$H_6$) (half-life = 1.6 h) compared with control (Ctrl) cells transfected with the empty vector (half-life = 3.8 h). In contrast, the stability of HTLV-1 genomic mRNA in cells transfected with siUPF1 (half-life = 7.9 h) increased compared with that in cells transfected with the control siRNA (siCtrl) (half-life = 2.4 h). The half-life was calculated based on the regression curve (dashed line). Above each graph, semi-quantitative RT-PCR analysis of relative RNA levels (left panel) and Western blotting analysis of protein levels (right panel) are shown ($n = 4$, mean ± SE). (C) HTLV-1 genomic RNA, detected by in situ hybridization with the *gag/pol* cDNA probe in HTLV-1 infected HeLa cells, co-localized with AzamiGreen-DCP2, a P-body marker, indicating that a notable fraction of HTLV-1 genomic unspliced RNA (gRNA) accumulates in P-bodies. No clear signals in negative control HeLa cells without infection confirmed that the *gag/pol* cDNA probe specifically detected gRNA in infected HeLa cells. (D) Interaction between UPF1 and HTLV-1 unspliced mRNA. HTLV-1 genomic unspliced mRNA was found in the UPF1 complex co-immunoprecipitated with anti-UPF1 antibody from whole-cell lysate of MT-2 cells (top panel). Moreover, sustained phosphorylation of UPF1 through okadaic acid (OA) treatment increased the level of bound HTLV-1 genomic mRNA recovered (bottom panel).

an antisense (As)-*upf2* mRNA-expressing plasmid and suppression of UPF2, please see Supplementary material.

### 2.4. Measurement of mRNA stability

First, the cells ($3 \times 10^5$/mL) were resuspended in culture medium, and $1.5 \times 10^5$/500 μL of the cells were sampled as the 0 h sample. Second, the culture medium was replaced by that containing actinomycin D (5 μg/mL) and seeded at $1.5 \times 10^5$/500 μL in 5 wells of a 12-well plate. The cells were sampled 1, 2, 3, 4, and/or 6 h after the addition of actinomycin D to each well, and total RNA was extracted using Isogen (Nippon Gene Co., Ltd.) for the measurement of mRNA levels by real-time PCR. The levels of NMD target mRNAs (MAP3K14, IL6, DUSP10, Fyn, PTPRF, ARHGEF18, ASNS, and DEXI) were measured at each time point and standardized by the corresponding β-actin mRNA level, which is not a NMD target. Methods in establishment of CEM cells, which stably overexpress Rex protein are available in Supplementary material.

### 2.5. Protein expression plasmids

The UPF1 expression plasmid was constructed by inserting the full-length cDNA fragment of human *upf1* into the *Eco*RI site of pCDNA3.1/His C (Invitrogen) for His/Xpress-UPF1 expression. The expression plasmids of HTLV-1 regulatory proteins were constructed by inserting the PCR-amplified cDNA fragments of HTLV-1 *rex*, *tax*, *p30ii*, *p12*, *p13*, and *hbz* from a cDNA pool derived from MT-2 cells at the *Xho*I/*Spe*I sites of pME-FLAG for FLAG-tagged proteins. The *dcp2* cDNA fragment was inserted at the *Bam*HI site of pmAG1-MN1 (Amalgaam Co., Ltd.) for AzamiGreen-tagged DCP2 expression.

### 2.6. Quantitative and semi-quantitative RT-PCR and genomic DNA PCR

The levels of viral and host cell transcripts were measured by quantitative or semi-quantitative RT-PCR. Total RNA was extracted using Isogen (Nippon Gene Co., Ltd.) following the manufacturer's protocol. DNase I treatment was performed to eliminate genomic DNA contamination. Extracted total RNA samples were subjected to reverse transcription using SuperScript II (Invitrogen), followed by quantitative real-time PCR using SYBR Premix Ex Taq (Takara Bio Inc.) and thermal cycler dice (Takara) or by semi-quantitative PCR. The sequences of primers used for PCR are available in Supplementary material.

### 2.7. Western blotting and antibodies

For the Western blotting of HTLV-1 Tax, a monoclonal Tax antibody (LT-4) was used [15]. Rex antibody, used in Western blotting for the HTLV-1 Rex protein, was monoclonal rat antiserum (Tanaka, unpublished data), while HTLV-1 Gag-p19 and Gag-p24 antibodies were the monoclonal mouse antibodies, GIN-7 and NOR-1, respectively [16]. The following primary antibodies were purchased from the indicated companies: hUPF1 (#9435; Cell Signaling Technology Inc.), hUPF2 (ab28712-200; Abcam), FLAG (F3165; Sigma–Aldrich Corporation), GST (#27-4577-01; GE Healthcare Bioscience), Xpress (46-0528; Invitrogen), β-actin (sc-69879; Santa Cruz Biotechnology, Inc.). For the secondary antibodies conjugated with alkaline phosphatase, anti-mouse IgG (S372B; Promega), anti-rabbit IgG (S373B; Promega), or anti-goat IgG (V115A; Promega) was used depending on the host species of the primary antibody.

### 2.8. Indirect immunofluorescence in situ RNA hybridization assays

HeLa cells transiently expressing AzamiGreen-DCP2 was co-cultured with MT-2 at 37 °C for 24 h to express HTLV-1. MT-2 cells were removed by washing 3 times with PBS, then the infected HeLa cells were further incubated at 37 °C for 24 h. In situ hybridization of HTLV-1 unspliced mRNA was performed by incubating the fixed and permeabilized cells with DIG-labeled HTLV-1 *gag/pol* cDNA probes at 37 °C for 16 h. The hybridized probes were detected by immunocytochemistry using rhodamine-conjugated anti-DIG antibody (Roche). The subcellular localization of HTLV-1 unspliced mRNA and AzamiGreen-DCP2 was observed using a confocal laser scanning microscope (LSM510; Carl Zeiss AG). As negative control, HeLa cells without HTLV-1 infection were also subjected to the same hybridization procedure.

### 2.9. RNA immunoprecipitation (RIP) assay

RIP assay between UPF1 and HTLV-1 genomic unspliced mRNA was performed following a method described elsewhere [17]. UFP1 was co-immunoprecipitated from the whole cell lysate of MT-2 using goat polyclonal antibody against hUPF1 (Rent-1 (p-14), sc-18260; Santa Cruz Biotechnology, Inc.), and total RNA was extracted from the immunoprecipitant for detection of HTLV-1 genomic unspliced RNA (gRNA) by RT-PCR using primers for the HTLV-1 *gag* region. Immunoprecipitation by a goat polyclonal antibody against GST (GE Healthcare Bioscience) was performed as a negative control. Total RNA from the whole cell lysate (22% vol. of input to immunoprecipitation) was also extracted for RT-PCR of gRNA.

### 2.10. Statistical analyses

Throughout the present study, two-tailed paired Student's *t*-test was performed to test the statistical difference between the experimental groups. Asterisks in the figures indicate a significant difference between the tested groups ($*p < 0.05$; $**p < 0.01$; and $***p < 0.001$, $n > 3$).

## 3. Results

### 3.1. The cellular UPF1 level influences the turnover rate of HTLV-1 unspliced mRNA

To test the hypothesis that NMD inhibits or antagonizes HTLV-1 replication, the accumulation of genomic primary

mRNA was examined in human HeLa cells transfected with an infectious HTLV-1 molecular clone, pFL-MT2 [12], in the presence of the overexpressed key NMD-positive effector, UPF1, or following siRNA-mediated knockdown of UPF1 (Fig. 1A). In the presence of ectopic UPF1, the steady-state level of HTLV-1 genomic unspliced mRNA (gRNA) accumulating from the pFL-MT2 infectious clone significantly decreased, whereas it increased following NMD inhibition by wortmannin treatment or following knockdown with UPF1-specific siRNAs (Fig. 1A). The level of WT-β-globin mRNA, which is not a NMD target, was not significantly influenced by the cellular level of UPF1 or by wortmannin treatment, indicating that the level of HTLV-1 genomic RNA was selectively influenced by cellular NMD activity (Fig. 1A). Reverse-transcriptase (−) PCR with *gag* primers was conducted in all above cDNA samples and it was confirmed that no detectable level of genomic DNA was contaminated in these samples (data not shown). In addition, the viral genomic unspliced RNA was significantly destabilized following UPF1 overexpression in infected cells (half-life = 1.6 h) compared with control cells with endogenous levels of UPF1 (half-life = 3.8 h) (Fig. 1B, left panel). In stark contrast, HTLV-1 unspliced RNA was stabilized after UPF1 knockdown by siUPF1 in infected cells (half-life = 7.9 h) compared to the control cells transfected with the control siRNA (half-life = 2.4 h) (Fig. 1B right panel). In cells, the factors required for NMD activity are often associated with cytoplasmic foci known as processing bodies (P-bodies). Therefore, the subcellular localization of HTLV-1 genomic unspliced RNA was examined in HTLV-1 infected HeLa cells by in situ hybridization and compared with the distribution of Azami-Green labeled de-capping protein 2 (DCP2) protein, which is a known P-body marker. HTLV-1 unspliced RNA (gRNA) was observed throughout the cytoplasm, but a substantial fraction of the RNA co-localized with DCP2 in brightly stained granular centers, suggesting that a significant level of the viral RNA localized to the P-bodies (Fig. 1C). Moreover, immunoprecipitation assays demonstrated that UPF1 is in a complex with HTLV-1 unspliced RNA and the amount of viral RNA interacting with UPF1 depends on the phosphorylation status of UPF1, i.e., phosphorylated UPF1 with okadaic acid treatment interacted with a higher amount of HTLV-1 genomic unspliced RNA (Fig. 1D). These data support the notion that HTLV-1 unspliced RNA is specifically targeted by UPF1 for processing via the NMD pathway.

### 3.2. HTLV-1 derived reporter activity is NMD sensitive

To further examine the impact of NMD on HTLV-1 RNA stability, a luciferase-based reporter plasmid was constructed (Fig. 2A). The reporter comprises the renilla luciferase ORF fused in-frame with the 5′ end of a HTLV-1 fragment spanning the 3′ end of *gag* through *pro* and into the 5′ end of *pol* (1677–2594 nt). The HTLV-1 sequences are followed at the 3′ end by a β-globin splicing sequence, which is positioned for recruitment of the exon junction complex (EJC) downstream of the HTLV-1 frameshift fragment. Splicing at the correct site, i.e., between exon 2 and exon 3 of β-globin mRNA, has been confirmed by sequencing PCR-amplified transcripts produced from the reporter (data not shown). This reporter provides four possible translation patterns, of which three generate a PTC and are expected to trigger NMD (Fig. 2A). The impact of siRNA knockdown of UPF1 or UPF2, which led to the suppression of NMD activity, on HTLV-1 derived reporter activity was examined in HeLa cells. Of note, the reporter activities increased in HeLa cells treated with siUPF1 and siUPF2 (Fig. 2B). Thus, reporter transcripts spanning 1677–2594 nucleotides of HTLV-1 genomic RNA are stabilized by NMD inhibition and are consequently targeted by NMD.

### 3.3. NMD activity in HTLV1 transformed cell lines

To determine cellular NMD activity, we employed a luciferase-based β-globin reporter system, which encodes renilla luciferase that is linked to the WT β-globin gene sequence as a control (WT-Glo) or an essentially identical construct but with a PTC-harboring β-globin gene sequence downstream of renilla ORF (PTC-Glo) (Fig. 3A). This β-globin-based reporter provides a robust and reliable readout of cellular NMD activity in HeLa cells (Fig. 3A) and in the T-cell line, Jurkat (Fig. S1).

To determine if cell transformation by HTLV-1 specifically perturbs NMD activity, we tested NMD activity in control HTLV-1-unrelated cell lines (fibroblasts and T cell lines) and in HTLV-1 infected T cell lines. In these assays, the control HTLV-1-unrelated cell lines showed relative NMD activities of >80% of that in control HeLa cells, which have intact NMD activity. In contrast, the NMD values for HTLV-1 infected (i.e., HTLV-1-transformed and ATL-derived) T cell lines were <50% of that observed in control HeLa cells, indicating lower NMD activities (Fig. 3B). These results raised the possibility that NMD is partially suppressed or inhibited in HTLV-1 infected cells and in ATL cells.

### 3.4. Impact of HTLV-1 infection on NMD activity

To examine whether HTLV-1 infection has any effect on the host cell NMD activity, sHeLa cells, which stably express both firefly-PTC-Glo and renilla-WT-Glo reporter genes, were infected with HTLV-1 by co-cultivation with MT-2 cells. Successful infection of HTLV-1 from MT-2 to sHeLa cells was confirmed by detection of HTLV-1 unspliced mRNAs by semi-quantitative RT-PCR (Fig. 3C). The expression of HTLV-1 unspliced mRNA was evaluated 1–4 days after infection. The highest levels of unspliced RNA were observed during the first 2 days of infection, corresponding to the peak levels of HTLV-1 *tax/rex* mRNAs and the Tax and Rex proteins (Fig. 3C). To confirm successful infection in HeLa cells co-cultured with MT-2, the expression pattern of the Tax protein was observed by immunocytochemistry (Fig. S2A). Genomic fingerprinting PCR of the human MCT118 locus in the infected sHeLa cells showed no MT-2-derived bands, confirming that viral proteins detected in the infected sHeLa cells were attributed to HTLV-1

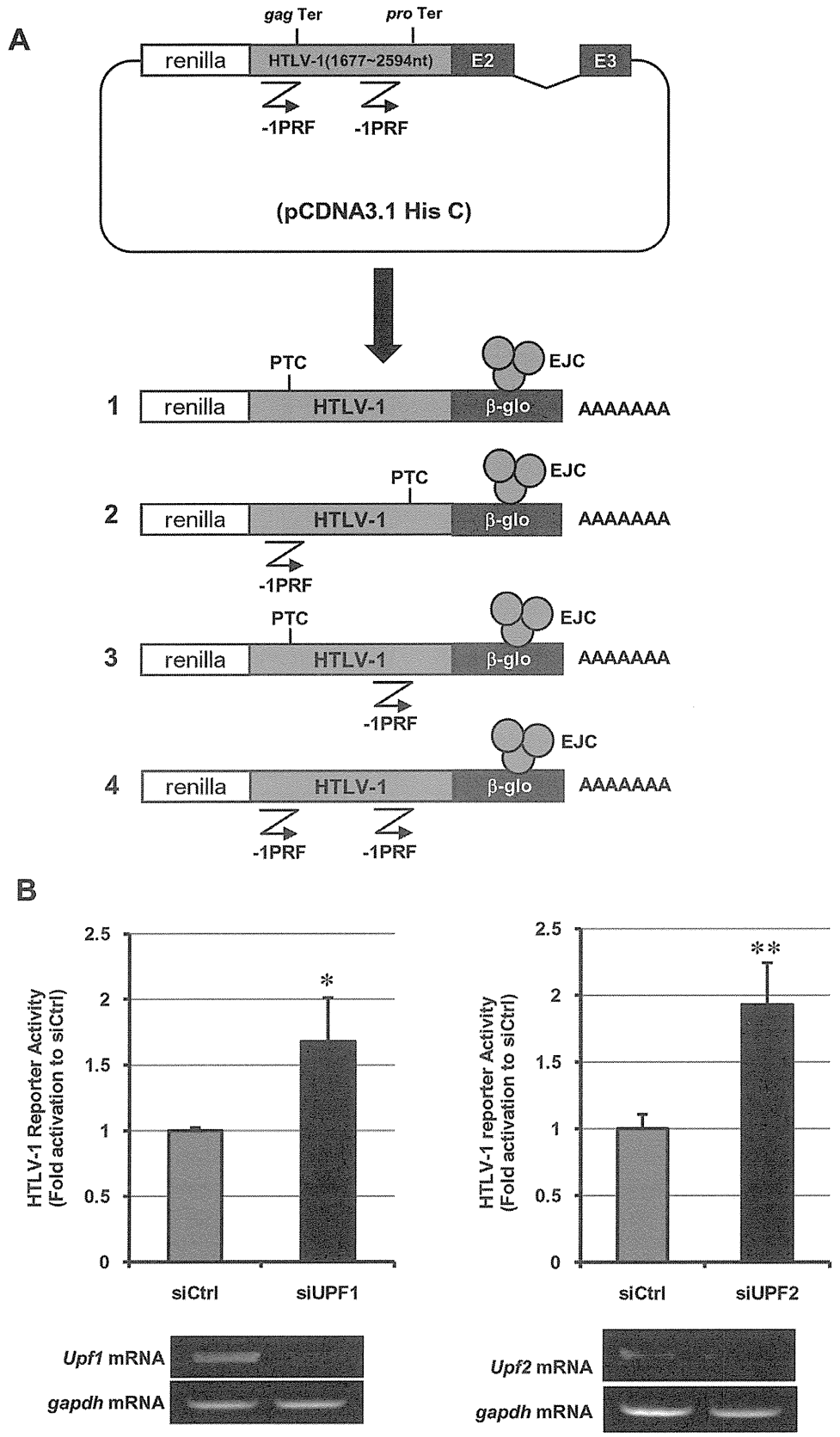



Fig. 2. HTLV-1 derived reporter activity is NMD sensitive. (A) A schematic structure of the HTLV-1 derived reporter plasmid and four possible translation patterns. As indicated, the HTLV-1 fragment in the reporter plasmid contains two −1 PRF signals (shown by ⇝) and 3/4 of these translation patterns result in PTC. (B) The HTLV-1 derived reporter activity was NMD dependent in HeLa cells. The relative HTLV-1 reporter activity was increased by NMD inhibition through siRNA-mediated suppression of UPF1 and UPF2 ($n = 3$–5, mean ± SD; $*p < 0.05$; $**p < 0.01$).

infection but not to contamination of co-cultured MT-2 cells (for genomic fingerprinting PCR, see Supplementary material and methods for more details) (Fig. S2B). NMD activity in these HTLV-I-infected sHeLa cells was significantly suppressed on day 1 ($p < 0.05$) and on day 2 ($p < 0.01$) after infection compared with that in control sHeLa cells co-cultured with uninfected cells (Fig. 3D). These results indicated that HTLV-1 caused a temporal suppression of NMD activity. Of note, the timing of maximum suppression of NMD activity corresponded to the peak expression of HTLV-1 *tax/rex* mRNAs and proteins and correlated with peak accumulation of HTLV-1 unspliced mRNA (Fig. 3C and D). Moreover, co-culture of sHeLa cells with MT-2 in the presence of the reverse transcriptase inhibitor azidothymidine (AZT) resulted in a dose-dependent rescue of NMD activity, supporting the idea that the suppression of NMD activity is indeed caused by HTLV-1 infection of target sHeLa cells and that this suppression requires reverse transcription and provirus integration (Fig. 3E).

### *3.5. HTLV-1 Rex is the principal viral factor inhibiting host NMD activity*

Next, we examined HTLV-1 gene products for the ability to suppress host NMD activity. The relative NMD activity in sHeLa cells that ectopically express p27Rex (Rex), p21Rex, p30II, and Tax was analyzed using the dual-luciferase NMD reporter assay at 24,36, and 48 h after transfection. As shown in Fig. 4A, NMD activity was significantly suppressed 24 h after transfection when p27Rex or p21Rex was expressed. In these experiments, the HTLV-1 transcriptional transactivator Tax was also found to inhibit NMD activity, but this effect was consistently less than that observed for p27Rex or p21Rex.

The most significant impact on host NMD activity was the dose-dependent suppression exerted by p27Rex and to a slightly lesser degree, p21Rex (Fig. 4B). Of note, NMD reporter constructs do not contain sequences corresponding to the highly structured RxRE. Consequently, we suggest that the suppression of NMD activity by Rex is not due to RxRE-dependent and Rex-mediated nuclear export of unspliced mRNAs. Indeed, nuclear export of unspliced β-globin mRNA, transcribed from the reporter plasmid, was not enhanced by Rex (Fig. S3A). We also confirmed that insertion of RxRE after the β-globin sequence of the NMD reporter plasmid did not exhibit any significant influence on Rex-induced NMD inhibition (Fig. S3B). Moreover, our study demonstrated that p21Rex, which lacks the N-terminal RxRE-binding domain, also exhibited NMD inhibitory activity, supporting the view that NMD inhibition is due to a genetically separable aspect of Rex function that is independent of the arginine-rich RxRE-binding motif. Other HTLV-1 accessory proteins, such as p12, p13, and antisense-encoded HBZ, did not show any significant influence on the cellular NMD activity (Fig. S4).

### *3.6. Rex inhibits global NMD activity of the cell*

On confirmation that Rex stabilizes chimeric NMD reporter mRNA, we tested if Rex stabilizes endogenous NMD target mRNA. Fig. 5 shows the decay rates of NMD target mRNAs in CEM cells stably expressing Rex or in control cells showing no Rex expression. The selected mRNAs contained uORFs (MAP3K14, IL6, DUSP10, Fyn, PTPRF, ARHGEF18, and ASNS) or 3′ UTR intron (DEXI) as NMD-inducing features and stabilization under UPF1 knockdown was confirmed (MAP3K14, PTPRF, ARHGEF18, and ASNS) or expected (DUSP10, Fyn, IL6, and DEXI) by Mendell et al. [2]. The graphs show that these mRNA substrates for NMD are significantly stabilized by Rex overexpression in a T cell line. These NMD target mRNAs were specifically stabilized in CEM-Rex, since the stability of β-actin mRNA, a non-NMD target, was almost the same between CEM-empty and CEM-Rex when the stabilities of β-actin mRNA and NMD target mRNA were separately illustrated (data not shown). These results provide strong evidence that Rex serves as a general inhibitor of NMD.

### *3.7. NMD inhibition by Rex enhances HTLV-1 expression*

Finally, we examined the effects of p27Rex, p21Rex, p30II, and Tax on HTLV-1 derived reporter activity. As shown in Fig. 6A, p27Rex significantly increased HTLV-1 reporter activity, which was very similar to that following NMD inhibition by siUPF1 or siUPF2 (Fig. 2B). p21Rex also significantly elevated HTLV-1 reporter activity, but to a lesser extent compared with p27Rex, whereas p30II and Tax did not significantly influence reporter activity (Fig. 6A). As anticipated from the accrued data and early reports by others [18–20], we also demonstrated that unspliced HTLV-1 mRNA was significantly stabilized in cells containing p27Rex (half-life = 11.3 h) compared with control cells containing empty vectors (half-life = 2.3 h) (Fig. 6B). On the other hand, p30II and Tax did not show significant effects on stabilization of unspliced HTLV-1 mRNA (half-lives = 3.4 h and 2.8 h, respectively). Interestingly, p21Rex, which does not have NLS and hence does not localize to the nucleus, also stabilized unspliced HTLV-1 unspliced mRNA (half-life = 11.0 h). These data further support that stabilization of viral genomic RNA by Rex is, at least partially, achieved by its function in the cytoplasm. Significantly, viral particle production, as determined by HTLV-1 Gag p53 (precursor), p24, and p19 protein expression levels, was increased under conditions of NMD inhibition by antisense-*upf2* mRNA overexpression that suppressed UPF2 protein expression (Fig. 6C). In addition, this increased level of Gag protein, especially in p53 and p19, was comparable to that observed following p27Rex and p21Rex overexpression (Fig. 6D). Given that p27Rex and p21Rex significantly stabilizes HTLV-1 *gag* mRNA (Fig. 6B), these data suggest that Rex enhances HTLV-1 replication by stabilizing unspliced viral transcripts via the suppression of host NMD activity.

## 4. Discussion

In this study, we demonstrate that HTLV-1 genomic and *gag/pol* RNA is recognized by UPF1, the principal regulator and initiator of NMD, and is thereby targeted for destruction

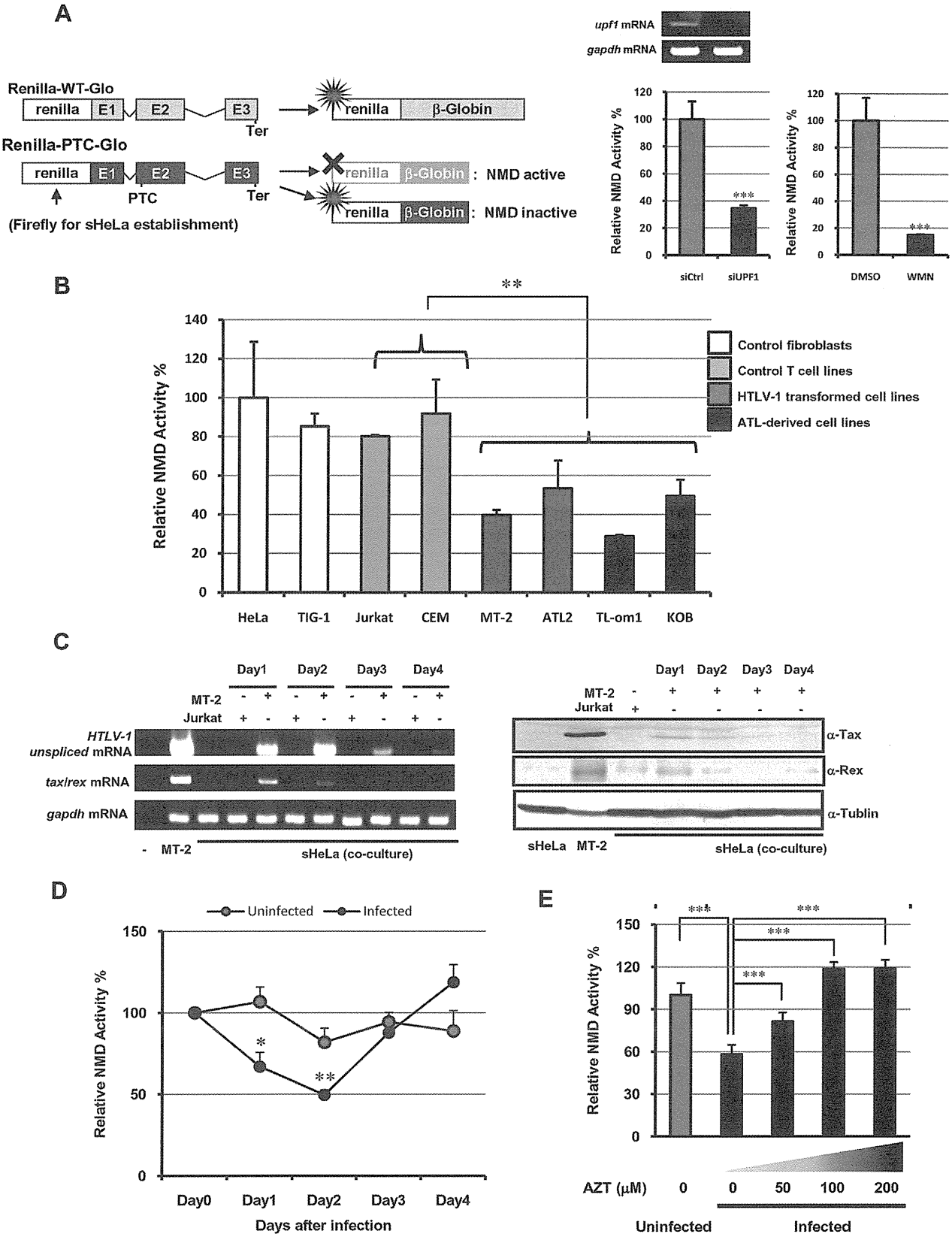



Fig. 3. HTLV-1 infection inhibits cellular NMD activity. (A) A schematic representation of the β-globin NMD reporter plasmids. PTC indicates a premature termination codon and Ter indicates a WT stop signal. The renilla fragment of Renilla-PTC-Glo was replaced with the firefly fragment for establishment of sHeLa cells, which stably expressed both Renilla-WT-Glo and Firefly-PTC-Glo, thus WT-Glo and PTC-Glo expression level was detected as renilla and firefly luciferase activity, respectively. ✹ indicates the luciferase activity. The renilla activity of Renilla-WT-Glo is always active and not influenced by the cellular NMD activity.