Kinetics of HTLV-1 reactivation from latency quantified by single-molecule RNA FISH and stochastic modelling

Short title: Single-cell quantification of HTLV-1 reactivation kinetics

Michi Miura[1], Supravat Dey[2], Abhyudai Singh[2], David Rueda[3,4] and Charles R. M. Bangham[1]*

[1] Section of Virology, Division of Infectious Diseases, Department of Medicine, Imperial College London, Norfolk Place, London W2 1PG, UK

[2] Department of Electrical and Computer Engineering, University of Delaware, Newark DE 19716, USA

[3] Section of Virology, Division of Infectious Diseases, Department of Medicine, Imperial College London

[4] Single Molecule Imaging Group, MRC London Institute of Medical Sciences, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK

*Corresponding author

c.bangham@imperial.ac.uk (CRMB)

Abstract

The human T cell leukemia virus HTLV-1 establishes a persistent infection *in vivo* in which the viral sense-strand transcription is usually silent at a given time in each cell. However, cellular stress responses trigger the reactivation of HTLV-1, enabling the virus to transmit to a new host cell. Using single-molecule RNA FISH, we measured the kinetics of the HTLV-1 transcriptional reactivation in peripheral blood mononuclear cells (PBMCs) isolated from HTLV-1$^+$ individuals. The abundance of the HTLV-1 sense and antisense transcripts was quantified hourly during incubation of the HTLV-1-infected PBMCs *ex vivo*. We found that, in each cell, the sense-strand transcription occurs in two distinct phases: the initial low-rate transcription is followed by a rapid transcriptional burst. The onset of transcription peaked between 1 and 3 hours after the start of *in vitro* incubation. The variance in the transcription intensity was similar in polyclonal HTLV-1$^+$ PBMCs (with tens of thousands of distinct provirus insertion sites), and in samples with a single dominant HTLV-1$^+$ clone. A stochastic simulation model was developed to estimate the parameters of HTLV-1 proviral transcription kinetics. The model indicated that the average duration of an HTLV-1 sense-strand transcriptional burst is less than one hour. HTLV-1 antisense transcription was stable during reactivation of the sense-strand. The antisense transcript *HBZ* was constantly produced at ~0.1 molecules per hour per HTLV-1$^+$ cell; however, 20% to 70% of HTLV-1-infected cells were *HBZ*-negative at a given time. HTLV-1-infected cells are exposed to a range of stresses when they are drawn from the host, which initiate the viral reactivation. We conclude that whereas antisense-strand transcription is stable throughout the stress response, the HTLV-1 sense-strand reactivation is highly heterogeneous and occurs in short, self-terminating bursts.

Author summary

Human retroviruses such as HIV-1 and HTLV-1 (human T cell leukemia virus) can establish a latent infection in the host cell. However, these viruses need to be able to produce viral genome to propagate in a new host. HTLV-1-infected cells are transmitted through breastfeeding, blood transfusion and sexual contact, and HTLV-1 restores transcription once the infected cells are drawn from infected individuals. We measured the kinetics of the HTLV-1 transcriptional reactivation in blood cells isolated from HTLV-1^{+} individuals by single-molecule RNA FISH. Viral transcripts were visualised as diffraction-limited spots and their abundance was quantified at one-hour intervals. The onset of the virus transcription peaked after one to three hours of incubation. In each cell, a short period of slow HTLV-1 transcription was followed by a burst of rapid transcription. Computer simulation, based on the experimental results, indicated that HTLV-1 terminates the transcriptional bursting in less than an hour. The HTLV-1 antisense transcript *HBZ* was constantly produced at a low level, and 20% to 70% of HTLV-1^{+} cells were negative for *HBZ* at a given time. These results demonstrate how rapidly HTLV-1 is reactivated and potentially becomes infectious, once HTLV-1^{+} cells are transmitted into a new host.

Introduction

Retroviral latency constitutes the main barrier to eradicating the infection from the host. The provirus is not completely silent, but is intermittently transcribed to produce viral particles that infect new clones of cells [1]. This intermittent transcription is partly due to the apparently stochastic nature of gene regulation [2]. However, it is poorly understood what regulates the balance between retroviral latency and transcription. Attempts to reactivate HIV-1 to expose it to drug therapy and the immune response - the shock-and-kill strategy [3] - have not been successful so far because it has not been possible to reactivate all the HIV-1 provirus in the reservoir [4]. Therefore, it is essential to understand the mechanisms and the extent of heterogeneity in the viral reactivation. Human T cell leukemia virus type 1 (HTLV-1) is a pathogenic retrovirus, which causes fatal and disabling diseases [1]. HTLV-1 remains largely latent *in vivo*, but undergoes transcriptional reactivation once the HTLV-1-infected lymphocytes are drawn from the circulation [5]. Here, we exploit this pattern of predominant HTLV-1 latency *in vivo* and spontaneous transcriptional reactivation *ex vivo* to study the kinetics of retroviral reactivation.

It is estimated that 10 million people across the globe are living with HTLV-1 [6]. HTLV-1 resides mainly in $CD4^+$ T lymphocytes, and causes adult T cell leukemia/lymphoma (ATL), a $CD4^+$ T cell malignancy, in some 5% of HTLV-1-infected individuals. HTLV-1 also causes, in up to another 5%, HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP; hereafter abbreviated to HAM) and other inflammatory disorders [1]. ATL typically presents as a monoclonal expansion of HTLV-1-infected T lymphocytes that is characterised by a unique integration site of the HTLV-1 provirus in the host genome [7]. By contrast, patients with HAM and asymptomatic HTLV-1 carriers have a polyclonal population of ~10^4 HTLV-1^+ T cell clones [7, 8].

HTLV-1 transcribes both the sense and antisense strand of its 9 kb provirus. The long terminal repeats (LTRs) located at the ends of the provirus serve as the promoters for sense and antisense transcription respectively. HTLV-1 Tax, produced from the sense strand, is a transcriptional transactivator protein that triggers NF-κB activation and other transcriptional cascades as well as its own promoter at the LTR (reviewed in ref [9]), exerting a strong positive feedback on the sense-strand transcription of the provirus. The HTLV-1 transcript *tax* is undetectable *in vitro* in freshly isolated cells from most HTLV-1-infected individuals [10]. However, the presence of chronically activated cytotoxic T lymphocytes (CTLs) that recognise sense-strand-encoded products, particularly Tax [11], indicates that the sense strand is intermittently expressed *in vivo*. By contrast, HTLV-1 transcribes HTLV-1 bZIP factor (*HBZ*) from the antisense strand. The *HBZ* transcript is usually detectable *in vitro* in peripheral blood mononuclear cells (PBMCs) obtained from HTLV-1-infected individuals [10].

In a previous study, we applied single-molecule RNA FISH (smFISH) [12, 13] to quantify HTLV-1 sense and antisense transcription in HTLV-1^{+} T cell clones isolated from HTLV-1-infected individuals by limiting dilution and maintained *in vitro* [14]. We observed that the HTLV-1 sense strand is transcribed in an occasional yet intense transcriptional burst. The antisense transcript *HBZ*, however, was much more constantly expressed, albeit at a low level.

To propagate between cells and between individuals, HTLV-1 must restore the sense-strand transcription from latency and produce viral components to form a virological synapse at the site of cell-to-cell contact [15]. HTLV-1-infected T lymphocytes are transmitted to a new host by breastfeeding, sexual contact and blood transfusion. HTLV-1 sense transcripts rapidly become detectable when the HTLV-1-infected T lymphocytes are drawn from the circulation: the transcriptional reactivation is mediated by p38 MAP kinase activation [5] in response to diverse forms of cell stress. In the present study, we quantified this

transient HTLV-1 transcriptional reactivation using smFISH [14] and used stochastic simulation to derive numerical estimates of the parameters of HTLV-1 reactivation from latency.

Results

HTLV-1 sense-strand shows two distinct phases of transcription in *ex vivo* culture

HTLV-1-infected PBMCs carry a single copy of the provirus in the genome. HTLV-1 rapidly initiates transcriptional reactivation of the sense-strand once the PBMCs are introduced into culture. We used smFISH to quantify the frequency and intensity of HTLV-1 reactivation at the single-cell level (Fig 1A). We removed $CD8^+$ cells to avoid CTL-mediated killing of HTLV-1-infected PBMCs that express sense-strand antigens (particularly Tax) during culture [16]. The PBMCs were sampled every hour and subjected to smFISH.

The sequence of representative images of cells (from a patient with HAM, coded TDZ) in Fig 1B shows the increase in the abundance of HTLV-1 sense transcripts over time. Sense-strand transcripts were virtually absent at 0 h (i.e. without incubation), and became detectable in a fraction of cells after 2 h of incubation. Subsequently, a bright spot appeared in the nucleus in those cells in which the sense-strand transcripts were present. The bright spot is an indication of rapid transcription [17], in which the transcripts are produced faster than they diffuse away, resulting in the accumulation of mRNAs at the site of transcription. This site contains nascent, unspliced transcripts, as revealed by an additional set of probes that targets *gag* (Fig 1C) which is otherwise spliced out of the sense transcripts (Fig 1A), indicating that the rate of transcription is even faster than that of the splicing reaction [12]. In this paper, we denote the occurrence of the intense spot in the nucleus (i.e. a rapid transcription) as a transcriptional burst [18, 19], as opposed to the rapid increase in the abundance of HTLV-1 transcripts over time at the population level. The occurrence of a transcriptional burst was typically associated with the presence of ~10 to 20 sense

transcripts in the nucleus. This observation is consistent with the idea that the positive feedback exerted by Tax on the sense-strand promoter occurs after a short interval following the first appearance of sense-strand transcripts, during which *tax* mRNA enters the cytoplasm, is translated, and Tax protein re-enters the nucleus. Thereafter, we observed cells carrying a large number of sense-strand transcripts - up to 300 molecules (Fig 1B).

Cell-to-cell variation in the intensity of HTLV-1 sense-strand reactivation in HTLV-1$^+$ PBMCs

The population of PBMCs imaged by smFISH contained both HTLV-1-infected and uninfected cells. We wished to analyse the HTLV-1 mRNA count within the HTLV-1-infected population. We used the proviral load to calculate the number of HTLV-1-infected cells in the total population (Fig 2A).

The distribution of the number of sense-strand transcripts per HTLV-1$^+$ cell over time is shown in Fig 2B. In each HTLV-1$^+$ subject examined (TBC, TDZ, LGZ and LGB), the total number of transcripts increased continuously during the 8 hours of incubation. The distribution in the sample from one patient with HAM (TBC) presented a long tail of cells that carried a larger number of sense transcripts (up to 250 molecules), with a continuous distribution of cells that carried fewer than 50 transcripts. By contrast, a single mode was identified at around 100 molecules in an ATL patient (LGZ) carrying a single dominant clone.

We postulated that the polyclonal structure of HTLV-1-infected T cell clones in subjects with non-malignant HTLV-1 infection [7] would explain the observed diversity in the distribution of the number of sense-strand transcripts per cell; by contrast, ATL patients with a monoclonal expansion of an HTLV-1$^+$ T cell clone would show more uniform distribution of the number of sense transcripts. We plotted the variance vs. the mean of the observed transcript count for the four HTLV-1$^+$ subjects with nine time points (Fig 2C). The mean and variance from all the four samples showed a uniform power-law relationship with

the exponent 1.7, suggesting that there is no fundamental qualitative difference in the kinetics of transcription between the samples examined.

HTLV-1 reactivation peaks after a few hours of *ex vivo* culture

To further characterise the transcriptional reactivation of the HTLV-1 sense strand, we quantified the fraction of cells that had become positive for sense transcripts at a given time during short-term culture of PBMCs (Fig 3A). Cells were defined as positive for sense transcripts when they carried 3 or more spots, to discount spurious spots due to noise (S1 Fig; see Methods). At 0 h, none of the HTLV-1$^+$ cells from each of the four HTLV-1$^+$ subjects examined (TBC, TDZ, LGZ and LGB) were positive for sense-strand transcripts. After 8 h of incubation, the fraction of positive cells rose to between 0.2 and 0.8, depending on the HTLV-1$^+$ subject examined. We assumed that the waiting time until we observed 3 or more molecules in a cell is an accumulation of single stochastic processes, and therefore fitted the plots presented in Fig 3A with a cumulative gamma distribution [20]. The probability distribution function derived from the fitted curve (Fig 3B) shows that the onset of sense-strand transcriptional reactivation peaked within 1 to 3 hours after the start of *ex vivo* culture. The timing of onset was more broadly distributed in samples from patients with HAM (TBC and TDZ) than from those with ATL (LGZ and LGB): the 75% quantiles of the probability distribution were 3.9 h and 4.9 h for TBC and TDZ respectively, whereas those for LGZ and LGB were less than 3 h. However, in each case the onset of the transcriptional reactivation was completed by 8 h *ex vivo* incubation in those cells that expressed detectable HTLV-1 mRNAs (20% to 80% of the total population, again depending on the HTLV-1$^+$ subject examined); the remaining cells showed no sign of proviral transcription during the period of observation.

Stochastic simulation predicts a burst termination following the promoter activation

From the results presented above, we established the following characteristics of the transient and heterogeneous HTLV-1 transcription in *ex vivo* PBMCs during the short-term culture. (1) The timing of the onset of transcriptional reactivation peaks in the first few hours of *ex vivo* incubation. (2) The sense-strand transcription is restored with an initially low transcription rate. (3) Subsequently the sense-strand transcription occurs in a rapid burst. We postulate that the sense-strand burst is triggered by the strong positive feedback loop on the sense-strand transcription exerted by Tax protein. (4) Only a subset of HTLV-1^+ PBMCs (20% to 80%, depending on the subject) restore the sense transcription.

We wished to identify the smallest number of parameters needed to explain the observed kinetics of proviral transcription. We applied a three-state promoter model [21] with some modifications (Fig 4A): the model allows the HTLV-1 sense promoter to transit from the "Off" to "On" state, and further progress to the "On-Tax" state by binding Tax protein; transcription occurs at the On state and the On-Tax state, with distinct transcription rates k_5 or k_6, respectively ($k_5 < k_6$). We used the smFISH data from the ATL patient LGZ for the parameter optimisation, because the single dominant clone of HTLV-1^+ PBMCs in this leukemic subject requires only a single kinetic parameter for each reaction rate in the model, as opposed to a polyclonal population in which each clone would require a unique reaction rate to be assigned. The simulations of this three-state model are shown in Fig 4B. The three-state model failed to capture the observed distribution of the transcript count in LGZ: it led to an overestimation of the number of sense-strand transcripts. Therefore we assumed an additional reaction path in the three-state model (Fig 4C), which allows the promoter to transit directly from the On-Tax state to the Off state with the parameter k_{10}. This model produced fitted the smFISH data well (Fig 4C); the resulting parameter estimates are listed in S1 Table. The parameter k_{10} was estimated to lie within the range of 1 to 5 (h^{-1}), indicating that the HTLV-1 sense-strand promoter terminates the transcriptional burst in less than an hour. The sense-strand

burst duration was measured for each cell in the simulation by recording the interval between the initiation of the On-Tax promoter state and the direct transition from the On-Tax to the Off state (Fig 4D). These results indicate that the HTLV-1 sense-strand burst is rapidly terminated.

Between 20% and 70% of HTLV-1-infected PBMCs are *HBZ*-negative at a given time

Lastly, we investigated the expression of the HTLV-1 antisense transcript *HBZ*. PBMCs isolated from one asymptomatic HTLV-1^+ carrier (HBL), three patients with HAM (TBC, TDK and TDZ) and four with ATL (LGZ, LFV, LFK and LGB) were examined by smFISH (Fig 5A). The distribution of the *HBZ* mRNA count per HTLV-1^+ PBMCs (without incubation) is shown in Fig 5B; the mean (<m>) is indicated in the inset. These results lead to two conclusions. First, the number of *HBZ* mRNA transcripts carried per cell was very low: frequently only one or two molecules of *HBZ* mRNA were present per cell; and unexpectedly 20% to 70% of HTLV-1^+ PBMCs in the subjects (except LFK) were negative for *HBZ* at a given time. In the subject LFK, almost all the HTLV-1^+ PBMCs were negative for *HBZ* (also see S1 Fig). Second, the distribution of the number of *HBZ* molecules conformed to the Poisson distribution of the parameter <m>, the observed mean mRNA count (Fig 5B, black lines; the p-value of the Kolmogorov-Smirnov test is indicated in the inset). This observation indicates that the *HBZ* transcripts are produced and degraded at a constant average rate (Fig 5C). Our previous study showed that the half-life of *HBZ* mRNA is 4.4 h [14], from which we estimate the degradation rate δ to be ln2 / 4.4 = 0.1575 h^{-1}. Since the mean number of *HBZ* mRNA molecules carried per cell <m> ranged from 0.380 to 1.31 (Fig 5B, except LFK), we estimate that the mean production rate of *HBZ* mRNA ($k = \langle m \rangle \cdot \delta$) falls within the range 0.0599 to 0.206 h^{-1} (Fig 5C).

Discussion

It is assumed that HTLV-1 *tax* is expressed *in vivo* [1], because a strong, chronically activated CTL response against Tax is readily detectable in PBMCs from HTLV-1^+ subjects *ex vivo* [11]. However, in the present

study HTLV-1 sense-strand transcripts were rarely detected in fresh PBMCs without incubation, using the sensitive and specific method of smFISH, in any of the HTLV-1^{+} subjects examined (Figs 1B and 2B). The results show that HTLV-1 sense-strand transcription is rare in the circulation. We postulate that HTLV-1 expresses *tax* in different anatomical compartments of the body other than the blood, such as bone marrow and lymphatic organs. This notion is supported by the observation that *tax* is highly expressed in bone marrow in Japanese macaques infected with simian leukemia virus STLV-1, the simian counterpart of HTLV-1 [22]. The smFISH technique could be applied to analyse specimens from those organs, to identify and localise the subset of cells that express *tax*, and to reveal the structural and spatial distribution of HTLV-1 expression *in situ*.

Whereas the sense-strand transcript *tax* is usually undetectable by reverse-transcription-PCR in PBMCs isolated from HTLV-1^{+} subjects, the HTLV-1 antisense transcript *HBZ* is invariably detected in the same subjects [10] at the cell population level. This observation raised the hypothesis that HBZ protein and *HBZ* mRNA play an essential part in HTLV-1 persistence (reviewed in ref [23]).

In our previous study, we examined the *HBZ* expression by smFISH in five HTLV-1-infected T cell clones isolated by limiting dilution from HTLV-1^{+} patients and maintained *in vitro* [14]. While the distribution of the *HBZ* mRNA count in three of the clones showed a small but significant departure from the expectation derived from the Poisson distribution, the other two clones conformed to the Poisson distribution (S2 Fig). The slight deviation from the Poisson distribution observed in those clones may result from a small burst of *HBZ* transcription [14] (S1 Text). The average *HBZ* mRNA count per cell in those five clones was in a small range (1.67 to 2.92), suggesting that the *HBZ* production rate varies little with the genomic insertion site of the HTLV-1 provirus.

In PBMCs circulating in HTLV-1-infected individuals, we assume that every HTLV-1$^+$ clone is capable of expressing *HBZ*, because *HBZ* appears to be necessary for the persistence of HTLV-1$^+$ T lymphocytes *in vivo* [23]. In the *ex vivo* PBMCs, we did not observe strong transcriptional bursts of the antisense strand. The distribution of the *HBZ* count in the patients with ATL conformed well to the Poisson expectation (Fig 5B), except the patient coded LGB. This leukemic patient was anomalous in that only 20% of the total population expressed the proviral sense strand, even though a single dominant clone was present in the peripheral blood. The malignant clone present in another ATL patient (LFK), which essentially expressed no *HBZ*, had presumably acquired a set of mutations that allowed the cells to proliferate independently of the HTLV-1 provirus. The *HBZ* count in the patients with HAM, with a polyclonal HTLV-1$^+$ T cell population, also conformed to the Poisson distribution. We conclude that *HBZ* is produced constantly - around 0.1 molecules per hour on average - in each infected cell; due to the low average rate of transcription, around half of the total infected population are *HBZ*-negative at any one time.

It has been reported that *HBZ* mRNA promotes cell proliferation *in vitro* [10, 24]. If this observation holds true in naturally-infected cells, with the low abundance of *HBZ* mRNA molecules observed in this study, it becomes increasingly important to understand how *HBZ* mRNA exerts such functions.

The HBZ protein level is also of great interest because the CD8$^+$ T cell response against HBZ is a significant determinant of the HTLV-1 proviral load [25], the main correlate of the risk of HAM and ATL [1]. A previous study from others estimated, by western blotting, that the abundance of endogenous HBZ protein is 17 molecules per cell in an ATL case [26]. However, this study did not report HBZ protein expression in non-malignant cells, or its variance in a polyclonal population. In addition to the abundance of *HBZ* mRNA quantified in the present study, measuring the half-life of HBZ protein will permit the estimation of the variance in HBZ protein expression as well as the rate of HBZ translation.

During the *ex vivo* incubation, the distribution of *HBZ* molecules per cell was stable compared to the changes in the sense-strand expression (S3 Fig) in three of the four subjects examined (TBC, TDZ and LGB). This observation suggests that the antisense strand does not undergo as strong a transcriptional burst as does the sense strand. In the fourth subject examined (LGZ), the abundance of *HBZ* mRNA in the population decayed over time (S3 Fig, LGZ). Since the majority of HTLV-1$^+$ cells in LGZ (80%) restored the sense-strand transcription, it is possible that transcriptional interference - by the collision of two converging polymerase complexes - diminished the antisense transcription in this subject. However, in each of the remaining three subjects studied, there was no correlation between the sense-strand transcription and *HBZ* transcription. We conclude that transcriptional interference between the sense- and antisense-strands does not play an important part in regulating HTLV-1 proviral transcription.

By applying smFISH to the transient HTLV-1 transcriptional reactivation of the sense strand, we found that reactivation was highly heterogeneous in *ex vivo* PBMCs, as shown by three observations. First, only a subset of the HTLV-1$^+$ PBMCs reactivated proviral transcription; second, the time of onset of transcription varied between cells; and third, the intensity of reactivation (i.e. the number of transcripts produced in a given period of time) varied widely in the population. Our results indicate two distinct phases of sense-strand transcription: the initial low-rate transcription produces ~10 to 20 mRNA molecules; thereafter a rapid transcriptional burst occurs, probably attributable to the positive feedback exerted by the viral transactivator protein Tax. This heterogeneity is likely to arise from the difference between clones in the HTLV-1 proviral integration site in the host genome. However, even in the ATL subjects examined in this study (LGZ and LGB), who carried one dominant clone in which (by definition) each cell shares the same integration site, a significant variance in reactivation intensity equivalent to that of HTLV-1$^+$ patients with polyclonal insertions was observed (Fig 2C). Such heterogeneity in transcription is often ascribed to the stochasticity that inevitably arises in biological processes.

In this study, we demonstrated that HTLV-1 sense-strand transcriptional reactivation in PBMCs occurs in an intense and rapid burst. But how long does such a burst last in the nucleus? By smFISH, we could not estimate the precise duration of a burst, because the cells rapidly accumulated hundreds of transcripts, making identification of a single bright spot difficult. By contrast, the stochastic simulation of the three-state promoter model (Fig 4C) estimated that the average burst duration was less than an hour: the model required the HTLV-1 sense-strand promoter to revert from the On-Tax state to the Off. Without invoking this extra parameter, cells did not terminate the transcription during the 8 h of the simulation, so the number of transcripts overshot the observed smFISH data (Fig 4B).

The mechanism of this rapid termination of the sense-strand burst remains unknown, as in mammalian gene expression in general. The HTLV-1 accessory protein p30 retains *tax* mRNA in the nucleus, which results in the reduction of Tax protein production, leading to the reduction of viral transcription [27]. However, p30 is unlikely to account for the rapid termination of the HTLV-1 transcriptional burst in PBMCs observed in the present study, because the abundance of Tax protein continues to rise in PBMCs during 8 h *ex vivo* culture [28]. Recent evidence suggests that transcription factors form phase-separated droplets where transcriptional bursts occur, and intrinsically disordered domains of transcription factors are particularly responsible for forming droplets [29, 30]. HTLV-1 Tax protein has a number of binding partners in the cell [31] and forms speckles in the nucleus that colocalise with active transcription spots [32]. These observations raise the possibility that Tax contributes to the formation of phase-separated droplets in the nucleus that serve as the locus for HTLV-1 transcription. External perturbation can disperse phase-separated droplets [30] and might therefore terminate the burst, leading to a refractory period of complete transcriptional silence.

In HIV-1 infection, the potential use of shock-and-kill treatment is still under debate because all approaches used to date have activated only a small fraction of the proviruses in a host. The results presented here demonstrate that the extent of heterogeneity in retrovirus reactivation is diverse in a population of naturally-infected cells, and suggest a quantitative approach for examining the efficacy of the treatment.

Materials and Methods

Patient-derived PBMCs

EDTA-anticoagulated blood from HTLV-1-infected patients was collected at St Mary's hospital (Imperial College London, UK). All donors gave written informed consent in accordance with the Declaration of Helsinki to donate blood samples to the Communicable Diseases Research Tissue Bank, approved by the UK National Research Ethics Service (15/SC/0089). Information on the HTLV-1$^+$ subjects examined in this study (four ATL patients, three HAM patients and one asymptomatic carrier) is available in Table 1. PBMCs were isolated using Histopaque (Sigma, H8889) and washed with PBS. Unless mentioned otherwise, they were stored in liquid nitrogen with FBS/10% DMSO until use.

Short-term culture of patient PBMCs

Patient-derived PBMCs were thawed and washed with PBS (Gibco, 14190) containing 2 mM EDTA/0.1% FBS. CD8$^+$ cells were removed with Dynabeads (Invitrogen, 11147D) to avoid CTL killing against HTLV-1 Tax$^+$ cells. Cells were suspended (1×10^6 cells/ml) in RPMI1640 (Sigma, R0883) supplemented with 2 mM L-glutamine (Gibco, 25030-81), penicillin/streptomycin (Gibco, 15070-063) and 10% fetal bovine serum (Gibco, 10500-064). One ml of cell suspension (1×10^6 cells) was aliquoted and dispensed into 5 ml polystyrene round-bottom tubes (Falcon, 352054) to keep the cell density consistent between assays.

Cells were incubated in 5% CO_2 at 37°C until sampled. One aliquot was sampled at each time point (every hour) and subjected to single-molecule RNA FISH.

Single-molecule RNA FISH

Cells were washed and suspended in PBS at a density of 5-10 × 10^6 cells/ml. An aliquot (20 µl) of the suspension was spread on to a coverslip that had been coated with poly-L-lysine (Sigma, P8920). Cells were fixed with 2% formaldehyde (Thermo Scientific, 28908) on the coverslip for 15 min. Coverslips were immersed in 70% ethanol and stored at -20°C until hybridisation.

Hybridisation was performed as previously described [14]. Namely, permeabilised cells were hybridized overnight at 37°C with probes that target either *gag* (HTLV-1 sense strand), *tax* (HTLV-1 sense strand) or *HBZ* (HTLV-1 antisense strand) in hybridisation buffer (Stellaris, SMF-HB1-10). Cells were washed on the coverslip twice with wash buffer A (Stellaris, SMF-WA1-60), and then wash buffer B (Stellaris, SMF-WB1-20). DAPI staining was included during the second wash with the buffer A. The cells were embedded in VECTASHIELD mounting medium (Vector Laboratories, H-1000) for imaging.

The coverslips were imaged with an Olympus IX70 inverted widefield microscope with a 100× 1.35NA UPlanApo oil objective lens (Olympus), a Spectra Light Engine illumination source (Lumencor) and an ORCA-Flash 4.0 V2 digital CMOS camera (Hamamatsu).

Image analysis

Stacked TIFF images containing 54 z-slices with a 300 nm interval were obtained using Micro Manager [33]. Each slice is composed of four channels (DAPI, GFP, DsRed and Cy5). The stacked TIFF images were split into the four channels. Cells were segmented using the DAPI channel with an in-house algorithm.

Dead cells were identified by the abnormal morphology of nucleus and omitted from the analysis. Spots in each of the three channels (GFP, DsRed and Cy5) were detected with FISH-QUANT [34]. A threshold for the signal intensity was set to distinguish the mRNA spots from spurious signals: an example is shown in S1 Fig containing the samples that were hybridised and imaged in the same batch - LGZ fixed at 7 h; LFK and LFV fixed without incubation.

Proviral load measurement

DNA was extracted from unfixed cells using DNeasy Blood & Tissue Kit (Qiagen, 69504). The reaction of 20 µl volume was set up using ddPCR Supermix (Bio-Rad, 186-3023) containing 900 nM each of forward and reverse primers and 250 nM FAM-labeled probe to detect the HTLV-1 *pX* locus (forward primer, 5'-CTC CTT CCG TTC CAC TCA AC-3'; reverse primer, 5'-GTG GTA GGC CTT GGT TTG AA-3'; probe, 5'-CGC CTA TGA TTT CCG GGC CCT G-3'). The copy number of a host gene *RPP30* was determined using 1 µl of HEX-labeled PrimePCR RPP30 (Bio-Rad, 10031243) in the reaction to calibrate the amount of DNA input. Droplets were generated with QX200 Droplet Generator (Bio-Rad). PCR was performed (95°C for 10 min; 40 cycles of 94°C for 30 sec and 60°C for 1 min; 98°C for 10 min) with C1000 Touch thermal cycler (Bio-Rad). Fluorescent signal was detected using QX200 Droplet Reader (Bio-Rad).

Stochastic simulation for the transient HTLV-1 sense-strand transcription

A Gillespie stochastic simulation algorithm was employed [35]. The molecular species and the schematic diagram of reactions are shown in Fig 4A. We assumed three possible states for the HTLV-1 sense-strand promoter: the "Off" state where essentially no transcription takes place; the "On" state from which sense transcripts are produced at a low transcription rate; and the "On-Tax" state where a transcriptional burst takes place due to the strong positive feedback in the presence of Tax protein. The HTLV-1 sense promoter was set at the Off state in each cell with no sense transcripts and Tax protein in the initial condition. In

modelling the HTLV-1 transient transcription in the patient LGZ, 20 % of the total cells in the simulation were confined to the Off promoter state (Fig 3A, LGZ). A random parameter search was performed to identify the sets of parameters that fit to the smFISH data. The parameter search space is shown in Table 2. The limits of the parameter search space were set using the data from smFISH experiments on the timing of the onset of transcription (for k_1) and the number of sense transcripts observed prior to a transcriptional burst (for k_5 and k_3). The limits of the parameter search for the mRNA and protein degradation rates were set according to the previous publication [36]. A number of parameter sets were tested (2,330 for the simulation set out in Fig 4B, and 9,725 in Fig 4C). The goodness of fit for each parameter set was assessed by squared deviation summed across the bins in every time point.

Table 1. HTLV-1$^+$ subjects and PBMCs examined in this study.

Assay #	Patient code	Disease Status	Blood collected	Storage condition	Sampling time (h)	CD8 removal	Proviral load
1	TDZ	HAM	06 Jan 2015	Frozen	0,1,2,3,4,5,6,7,8	CD8-depleted	16.4%*
2	TBC	HAM	18 Sep 2017	Frozen	0,1,2,3,4,5,6,7,8	CD8-depleted	81.1%
3	LGZ	ATL	08 Jul 2014	Frozen	0,1,2,3,4,5,6,7,8	CD8-depleted	89.7%
4	HBL	Asymptomatic	01 Oct 2018	Fresh	0	Total PBMCs	1.95%
5	LFK	ATL	01 Oct 2018	Fresh	0	Total PBMCs	46.2%
6	LFV	ATL	01 Oct 2018	Fresh	0	Total PBMCs	94.6%
7	TDK	HAM	02 Oct 2018	Fresh	0	Total PBMCs	11.2%
8	LGB	ATL	10 Jun 2014	Frozen	0,1,2,3,4,5,6,7,8	CD8-depleted	88.8%

*measured in the sample collected on 04 Jul 2016

Table 2. Parameters used in the stochastic simulation.

Reaction	Parameter	Description	Range for parameter search (h^{-1})
Transition	k_1	Off to On	0.25 to 2.5
	k_2	On to Off	0.01 to 10
	k_3	On to On-Tax	0.1 to 10
	k_4	On-Tax to On	0.01 to 10
	k_{10}	On-Tax to Off	0.01 to 10
Transcription	k_5	From On state	10 to 50
	k_6	From On-Tax state	k5 to 100
Translation	k_8	Tax translation	0.01 to 10
Degradation	k_7	Sense mRNA	0.008 to 0.2
	k_9	Tax protein	0.002 to 0.2

Acknowledgements

This work was supported by the Wellcome Trust UK (Investigator Award to CRMB), the Medical Research Council UK (Project Grant to CRMB), the Gen Foundation (research grant to MM) and the Imperial College Biomedical Research Centre (BRC). We thank Graham Taylor, Lucy Cook, the donors and research nurses in the National Centre for Human Retrovirology, Imperial College; Dr Aileen Rowan for discussion on the ATL subjects; Dr Martin Billman for help with the single-molecule RNA FISH; and the Imperial College Research Computing Service (DOI: 10.14469/hpc/2232) for the use of the high-performance computer cluster.

References

1. Bangham CRM, Matsuoka M. Human T-cell leukaemia virus type 1: parasitism and pathogenesis. Philos Trans R Soc Lond B Biol Sci. 2017;372(1732). Epub 2017/09/13. doi: 10.1098/rstb.2016.0272. PubMed PMID: 28893939; PubMed Central PMCID: PMCPMC5597739.

2. Singh A, Weinberger LS. Stochastic gene expression as a molecular switch for viral latency. Curr Opin Microbiol. 2009;12(4):460-6. Epub 2009/07/15. doi: 10.1016/j.mib.2009.06.016. PubMed PMID: 19595626; PubMed Central PMCID: PMCPMC2760832.

3. Darcis G, Van Driessche B, Van Lint C. HIV Latency: Should We Shock or Lock? Trends Immunol. 2017;38(3):217-28. Epub 2017/01/12. doi: 10.1016/j.it.2016.12.003. PubMed PMID: 28073694.

4. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell. 2013;155(3):540-51. Epub 2013/11/19. doi: 10.1016/j.cell.2013.09.020. PubMed PMID: 24243014; PubMed Central PMCID: PMCPMC3896327.

5. Kulkarni A, Taylor GP, Klose RJ, Schofield CJ, Bangham CR. Histone H2A monoubiquitylation and p38-MAPKs regulate immediate-early gene-like reactivation of latent retrovirus HTLV-1. JCI Insight. 2018;3(20). Epub 2018/10/20. doi: 10.1172/jci.insight.123196. PubMed PMID: 30333309; PubMed Central PMCID: PMCPMC6237452.

6. Gessain A, Cassar O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. Front Microbiol. 2012;3:388. Epub 2012/11/20. doi: 10.3389/fmicb.2012.00388. PubMed PMID: 23162541; PubMed Central PMCID: PMCPMC3498738.

7. Gillet NA, Malani N, Melamed A, Gormley N, Carter R, Bentley D, et al. The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. Blood. 2011;117(11):3113-22. Epub 2011/01/14. doi: 10.1182/blood-2010-10-312926. PubMed PMID: 21228324; PubMed Central PMCID: PMCPMC3062313.

8. Laydon DJ, Melamed A, Sim A, Gillet NA, Sim K, Darko S, et al. Quantification of HTLV-1 clonality and TCR diversity. PLoS Comput Biol. 2014;10(6):e1003646. Epub 2014/06/20. doi: 10.1371/journal.pcbi.1003646. PubMed PMID: 24945836; PubMed Central PMCID: PMCPMC4063693.

9. Kulkarni A, Bangham CRM. HTLV-1: Regulating the Balance Between Proviral Latency and Reactivation. Front Microbiol. 2018;9:449. Epub 2018/04/05. doi: 10.3389/fmicb.2018.00449. PubMed PMID: 29615991; PubMed Central PMCID: PMCPMC5867303.

10. 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. 2006;103(3):720-5. Epub 2006/01/13. doi: 10.1073/pnas.0507631103. PubMed PMID: 16407133; PubMed Central PMCID: PMCPMC1334651.

11. Jacobson S, Shida H, McFarlin DE, Fauci AS, Koenig S. Circulating CD8+ cytotoxic T lymphocytes specific for HTLV-I pX in patients with HTLV-I associated neurological disease. Nature. 1990;348(6298):245-8. Epub 1990/11/15. doi: 10.1038/348245a0. PubMed PMID: 2146511.

12. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. Science. 1998;280(5363):585-90. Epub 1998/05/09. PubMed PMID: 9554849.

13. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008;5(10):877-9. Epub 2008/09/23. doi: 10.1038/nmeth.1253. PubMed PMID: 18806792; PubMed Central PMCID: PMCPMC3126653.

14. Billman MR, Rueda D, Bangham CRM. Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. Wellcome Open Res. 2017;2:87. Epub 2017/10/25. doi: 10.12688/wellcomeopenres.12469.2. PubMed PMID: 29062917; PubMed Central PMCID: PMCPMC5645716.

15. Igakura T, Stinchcombe JC, Goon PK, Taylor GP, Weber JN, Griffiths GM, et al. Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. Science. 2003;299(5613):1713-6. Epub 2003/02/18. doi: 10.1126/science.1080115. PubMed PMID: 12589003.

16. Hanon E, Hall S, Taylor GP, Saito M, Davis R, Tanaka Y, 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. 2000;95(4):1386-92. Epub 2000/02/09. PubMed PMID: 10666215.

17. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006;4(10):e309. Epub 2006/10/20. doi: 10.1371/journal.pbio.0040309. PubMed PMID: 17048983; PubMed Central PMCID: PMCPMC1563489.

18. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, et al. Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A. 2012;109(43):17454-9. Epub 2012/10/16. doi: 10.1073/pnas.1213530109. PubMed PMID: 23064634; PubMed Central PMCID: PMCPMC3491463.

19. Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, et al. Genomic encoding of transcriptional burst kinetics. Nature. 2019;565(7738):251-4. Epub 2019/01/04. doi: 10.1038/s41586-018-0836-1. PubMed PMID: 30602787.

20. Schwabe A, Bruggeman FJ. Single yeast cells vary in transcription activity not in delay time after a metabolic shift. Nat Commun. 2014;5:4798. Epub 2014/09/03. doi: 10.1038/ncomms5798. PubMed PMID: 25178355.

21. Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, et al. Transcription factors modulate c-Fos transcriptional bursts. Cell Rep. 2014;8(1):75-83. Epub 2014/07/02. doi: 10.1016/j.celrep.2014.05.053. PubMed PMID: 24981864; PubMed Central PMCID: PMCPMC5555219.

22. Furuta R, Yasunaga JI, Miura M, Sugata K, Saito A, Akari H, et al. Human T-cell leukemia virus type 1 infects multiple lineage hematopoietic cells in vivo. PLoS Pathog. 2017;13(11):e1006722. Epub 2017/12/01. doi: 10.1371/journal.ppat.1006722. PubMed PMID: 29186194; PubMed Central PMCID: PMCPMC5724899.

23. Ma G, Yasunaga J, Matsuoka M. Multifaceted functions and roles of HBZ in HTLV-1 pathogenesis. Retrovirology. 2016;13:16. Epub 2016/03/17. doi: 10.1186/s12977-016-0249-x. PubMed PMID: 26979059; PubMed Central PMCID: PMCPMC4793531.

24. Mitobe Y, Yasunaga J, Furuta R, Matsuoka M. HTLV-1 bZIP Factor RNA and Protein Impart Distinct Functions on T-cell Proliferation and Survival. Cancer Res. 2015;75(19):4143-52. Epub 2015/09/19. doi: 10.1158/0008-5472.CAN-15-0942. PubMed PMID: 26383166.

25. Macnamara A, Rowan A, Hilburn S, Kadolsky U, Fujiwara H, Suemori K, et al. HLA class I binding of HBZ determines outcome in HTLV-1 infection. PLoS Pathog. 2010;6(9):e1001117. Epub 2010/10/05. doi: 10.1371/journal.ppat.1001117. PubMed PMID: 20886101; PubMed Central PMCID: PMCPMC2944806.

26. Raval GU, Bidoia C, Forlani G, Tosi G, Gessain A, Accolla RS. Localization, quantification and interaction with host factors of endogenous HTLV-1 HBZ protein in infected cells and ATL. Retrovirology. 2015;12:59. Epub 2015/07/05. doi: 10.1186/s12977-015-0186-0. PubMed PMID: 26140924; PubMed Central PMCID: PMCPMC4491271.

27. Nicot C, Dundr M, Johnson JM, Fullen JR, Alonzo N, Fukumoto R, et al. HTLV-1-encoded p30II is a post-transcriptional negative regulator of viral replication. Nat Med. 2004;10(2):197-201. Epub 2004/01/20. doi: 10.1038/nm984. PubMed PMID: 14730358.

28. Rowan AG, Suemori K, Fujiwara H, Yasukawa M, Tanaka Y, Taylor GP, et al. Cytotoxic T lymphocyte lysis of HTLV-1 infected cells is limited by weak HBZ protein expression, but non-specifically enhanced on induction of Tax expression. Retrovirology. 2014;11:116. Epub 2014/12/17. doi: 10.1186/s12977-014-0116-6. PubMed PMID: 25499803; PubMed Central PMCID: PMCPMC4282740.

29. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. Cell. 2017;169(1):13-23. Epub 2017/03/25. doi: 10.1016/j.cell.2017.02.007. PubMed PMID: 28340338; PubMed Central PMCID: PMCPMC5432200.

30. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell. 2018;175(7):1842-55 e16. Epub 2018/11/20. doi: 10.1016/j.cell.2018.10.042. PubMed PMID: 30449618; PubMed Central PMCID: PMCPMC6295254.

31. Boxus M, Twizere JC, Legros S, Dewulf JF, Kettmann R, Willems L. The HTLV-1 Tax interactome. Retrovirology. 2008;5:76. Epub 2008/08/16. doi: 10.1186/1742-4690-5-76. PubMed PMID: 18702816; PubMed Central PMCID: PMCPMC2533353.

32. Semmes OJ, Jeang KT. Localization of human T-cell leukemia virus type 1 tax to subnuclear compartments that overlap with interchromatin speckles. J Virol. 1996;70(9):6347-57. Epub 1996/09/01. PubMed PMID: 8709263; PubMed Central PMCID: PMCPMC190661.

33. Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N. Computer control of microscopes using microManager. Curr Protoc Mol Biol. 2010;Chapter 14:Unit14 20. Epub 2010/10/05. doi: 10.1002/0471142727.mb1420s92. PubMed PMID: 20890901; PubMed Central PMCID: PMCPMC3065365.

34. Mueller F, Senecal A, Tantale K, Marie-Nelly H, Ly N, Collin O, et al. FISH-quant: automatic counting of transcripts in 3D FISH images. Nat Methods. 2013;10(4):277-8. Epub 2013/03/30. doi: 10.1038/nmeth.2406. PubMed PMID: 23538861.

35. Gillespie DT. General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions. J Comput Phys. 1976;22(4):403-34. doi: Doi 10.1016/0021-9991(76)90041-3. PubMed PMID: WOS:A1976CQ87900001.

36. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011;473(7347):337-42. Epub 2011/05/20. doi: 10.1038/nature10098. PubMed PMID: 21593866.

Supporting information

S1 Fig **Determination of the threshold for mRNA spot detection.**

S2 Fig **Antisense HBZ expression in in vitro maintained HTLV-1+ T cell clones.**

S3 Fig **HTLV-1 antisense expression in PBMCs *ex vivo* culture over time.**

S1 Table **Top 20 parameter sets that best fitted the smFISH data (LGZ).**

S1 File **An alternative model accounting for the departure from the Poisson distribution.**

Figure legends

Fig 1 The HTLV-1 sense-strand transcriptional burst identified by smFISH.

(A) HTLV-1 transcripts, smFISH probes and sample preparation for smFISH. Three sets of probes for smFISH are indicated that hybridise to HTLV-1 transcripts: sense-strand in the *pX* region (Q670, red stars), *HBZ* (Q570, yellow stars), and unspliced sense transcripts containing *gag* region (FAM, green stars). (B) Representative images of HTLV-1$^+$ PBMCs with the sense-strand transcripts at the indicated time points. Blue area indicates the nucleus stained with DAPI, and red spots indicate the HTLV-1 sense transcripts. Scale bar (white) = 5 µm. (C) A transcriptional bursting identified by the probes for *gag*. The image on the left shows the *gag*-staining (green). The image overlaid with sense transcript spots (red, shown in the middle) is presented on the right.

Fig 2 Sense transcript count per HTLV-1$^+$ cell over time.

(A) Schematic diagram of the identification of the HTLV-1$^+$ population for analysis. (B) Sense-transcript count per HTLV-1$^+$ cell (bin size = 10, note the units) at each time point for the two patients with HAM (TBC and TDZ) and two with ATL (LGZ and LGB). Note that the first few bins in each case are out of scale on the y-axis. In the inset for each histogram is indicated the number of cells analysed. (C) The mean and variance of the sense transcript count per HTLV-1$^+$ cell obtained from four patients, each containing nine time points.

Fig 3 Rate of onset of HTLV-1 reactivation.

(A) Increase in the fraction of cells that carry ≥3 sense transcripts over time (red dots). The black line for each HTLV-1$^+$ subject indicates a fitted curve with a cumulative gamma distribution. (B) Probability density of the gamma distribution obtained from panel A for each HTLV-1$^+$ subject. The inset numbers indicate the mean time of the onset of reactivation after the start of *ex vivo* culture.

Fig 4 Stochastic simulation of the transient HTLV-1 expression in PBMCs *ex vivo*.

(A) Kinetic model of HTLV-1 proviral transcription, with three promoter states: Off, On, and On-Tax. (B) Stochastic simulation of the model presented in panel A. The 10 best fits (grey lines) are superimposed on the smFISH data from the ATL patient LGZ (red bars, reproduced from Fig 2B). (C) Stochastic simulation of the model with the additional reaction channel. The 10 best fits (grey lines) are overlaid on the same smFISH data (red bars) as in panel B. (D) Frequency distribution of the simulated burst duration. Ten such distributions (blue lines) were obtained with the 10 parameter sets used to produce the data presented in panel C. A representative histogram from one of the distributions is shown in grey to guide the eye.

Fig 5 Identification of HTLV-1 antisense transcripts by smFISH.

(A) Typical image captured by smFISH of *HBZ* mRNA (yellow spot) in a cell (from a patient with HAM, coded TDZ) without *ex vivo* incubation. Blue indicates the nucleus stained with DAPI. Scale bar (white) = 5 µm. (B) Frequency distribution of the number of *HBZ* mRNA molecules carried per HTLV-1$^+$ PBMC. Yellow bars show the frequencies observed by smFISH; the number of HTLV-1$^+$ cells analysed (N) and the average number of *HBZ* molecules per HTLV-1$^+$ cell (<m>) are indicated in the inset. Black lines indicate the expected frequencies from a Poisson distribution with the parameter <m>. (C) The kinetic model for the antisense transcription. *HBZ* mRNA is transcribed at the rate k, and degraded at the rate δ.

Fig 1

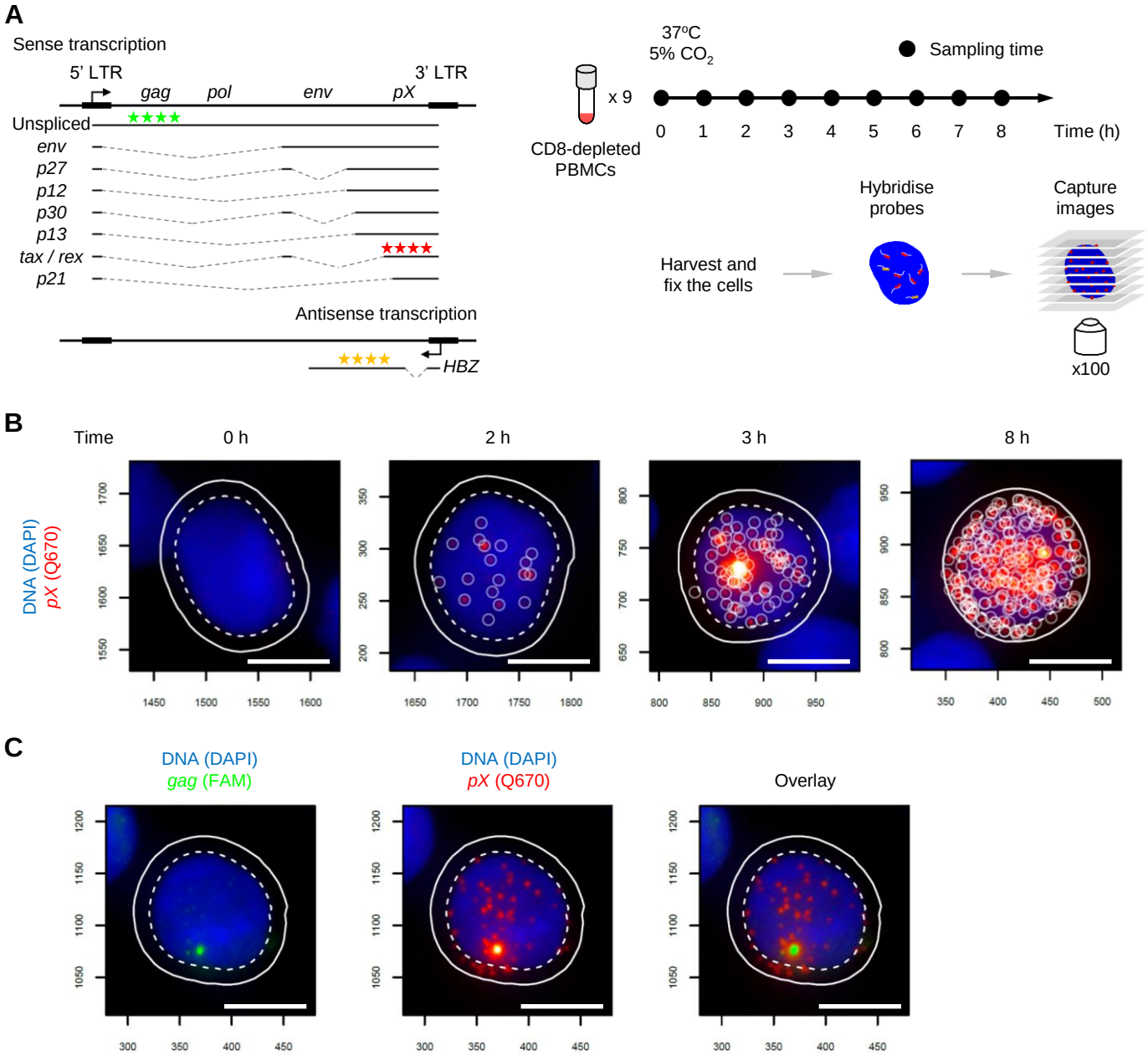

A
Sense transcription
5' LTR
3' LTR
gag
pol
env
pX
Unspliced
env
p27
p12
p30
p13
tax / rex
p21
Antisense transcription
HBZ
37ºC
5% CO2
Sampling time
x 9
CD8-depleted PBMCs
0 1 2 3 4 5 6 7 8
Time (h)
Hybridise probes
Capture images
Harvest and fix the cells
x100
B
Time
0 h
2 h
3 h
8 h
DNA (DAPI)
pX (Q670)
C
DNA (DAPI)
gag (FAM)
DNA (DAPI)
pX (Q670)
Overlay

Fig 2

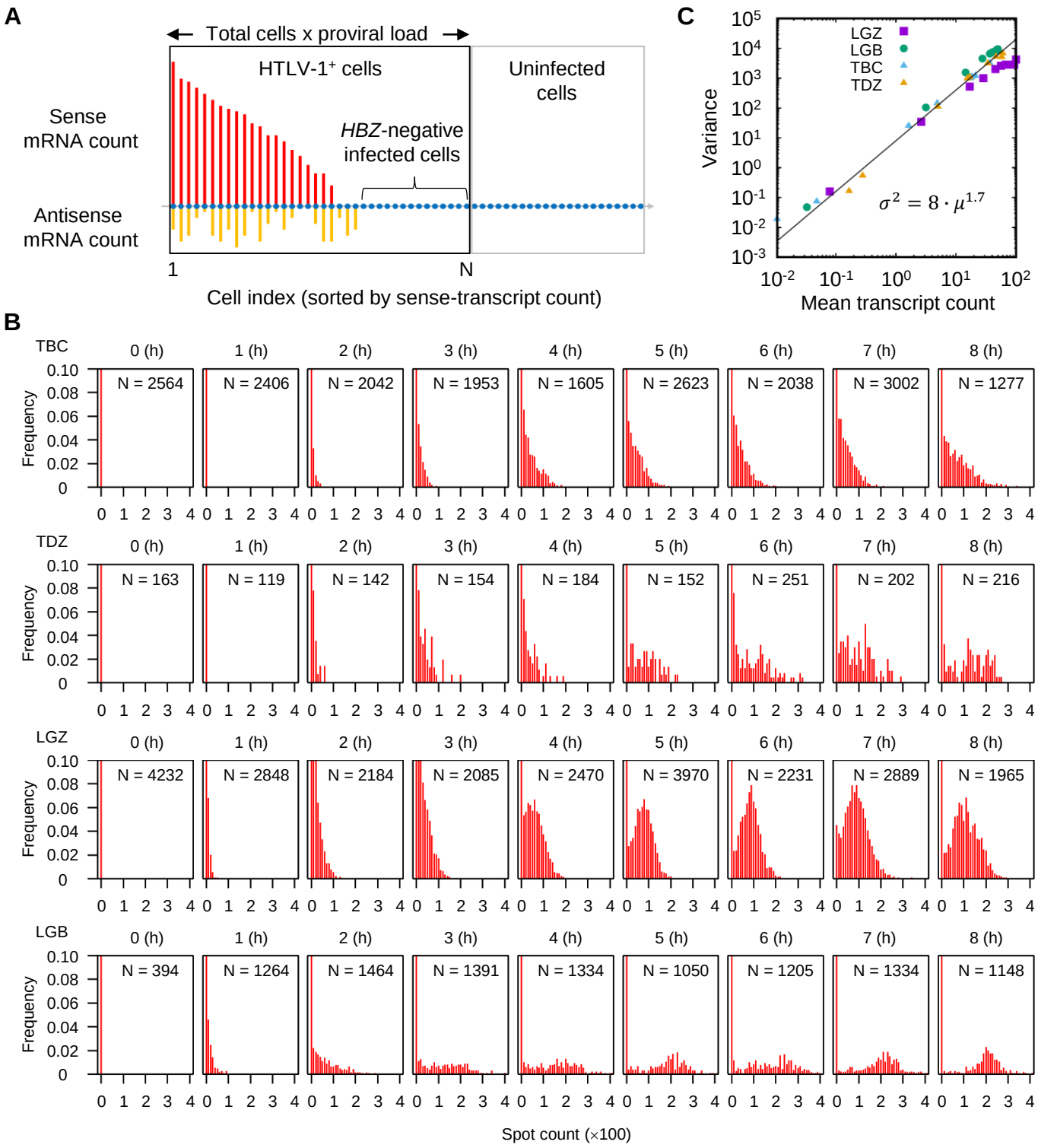

A
Total cells x proviral load
HTLV-1+ cells
Uninfected cells
Sense mRNA count
HBZ-negative infected cells
Antisense mRNA count
1
N
Cell index (sorted by sense-transcript count)
C
LGZ
LGB
TBC
TDZ
Variance
Mean transcript count
σ² = 8 · μ^1.7
B
TBC
TDZ
LGZ
LGB
0 (h)
1 (h)
2 (h)
3 (h)
4 (h)
5 (h)
6 (h)
7 (h)
8 (h)
N = 2564
N = 2406
N = 2042
N = 1953
N = 1605
N = 2623
N = 2038
N = 3002
N = 1277
N = 163
N = 119
N = 142
N = 154
N = 184
N = 152
N = 251
N = 202
N = 216
N = 4232
N = 2848
N = 2184
N = 2085
N = 2470
N = 3970
N = 2231
N = 2889
N = 1965
N = 394
N = 1264
N = 1464
N = 1391
N = 1334
N = 1050
N = 1205
N = 1334
N = 1148
Frequency
Spot count (×100)

Fig 3

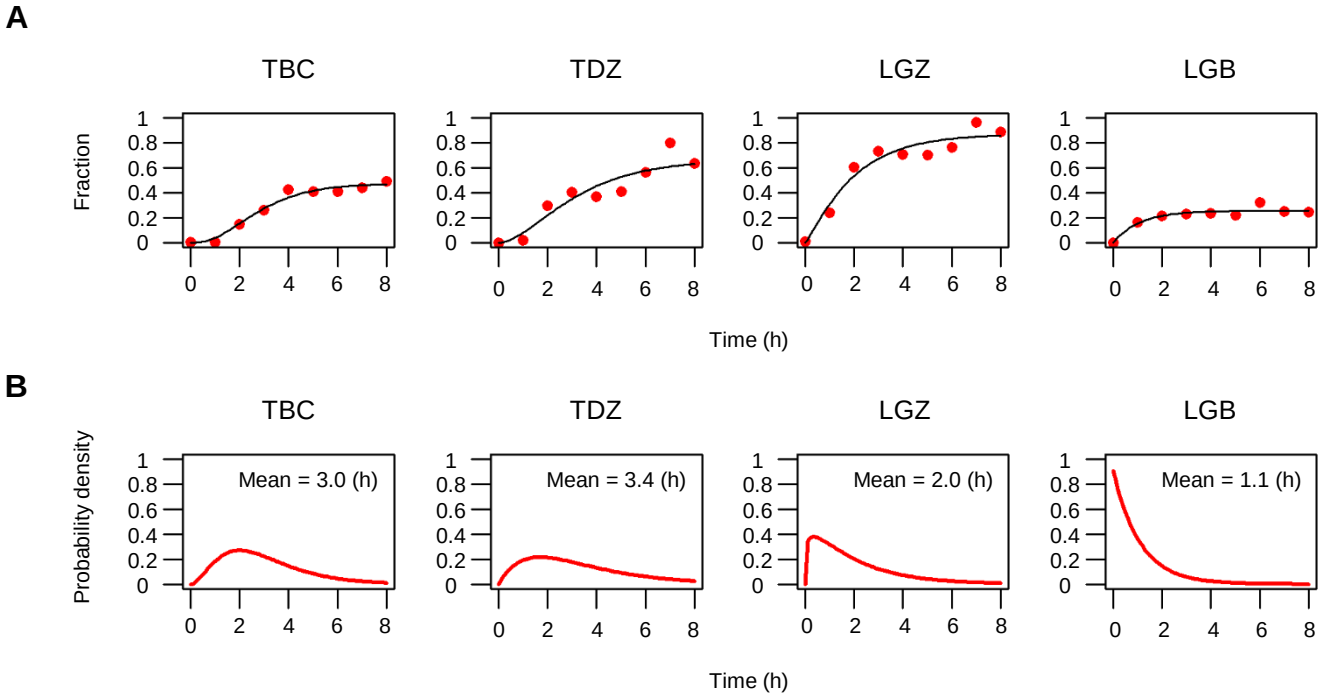

A
TBC
TDZ
LGZ
LGB
Fraction
Time (h)
B
TBC
TDZ
LGZ
LGB
Probability density
Mean = 3.0 (h)
Mean = 3.4 (h)
Mean = 2.0 (h)
Mean = 1.1 (h)
Time (h)

Fig 4

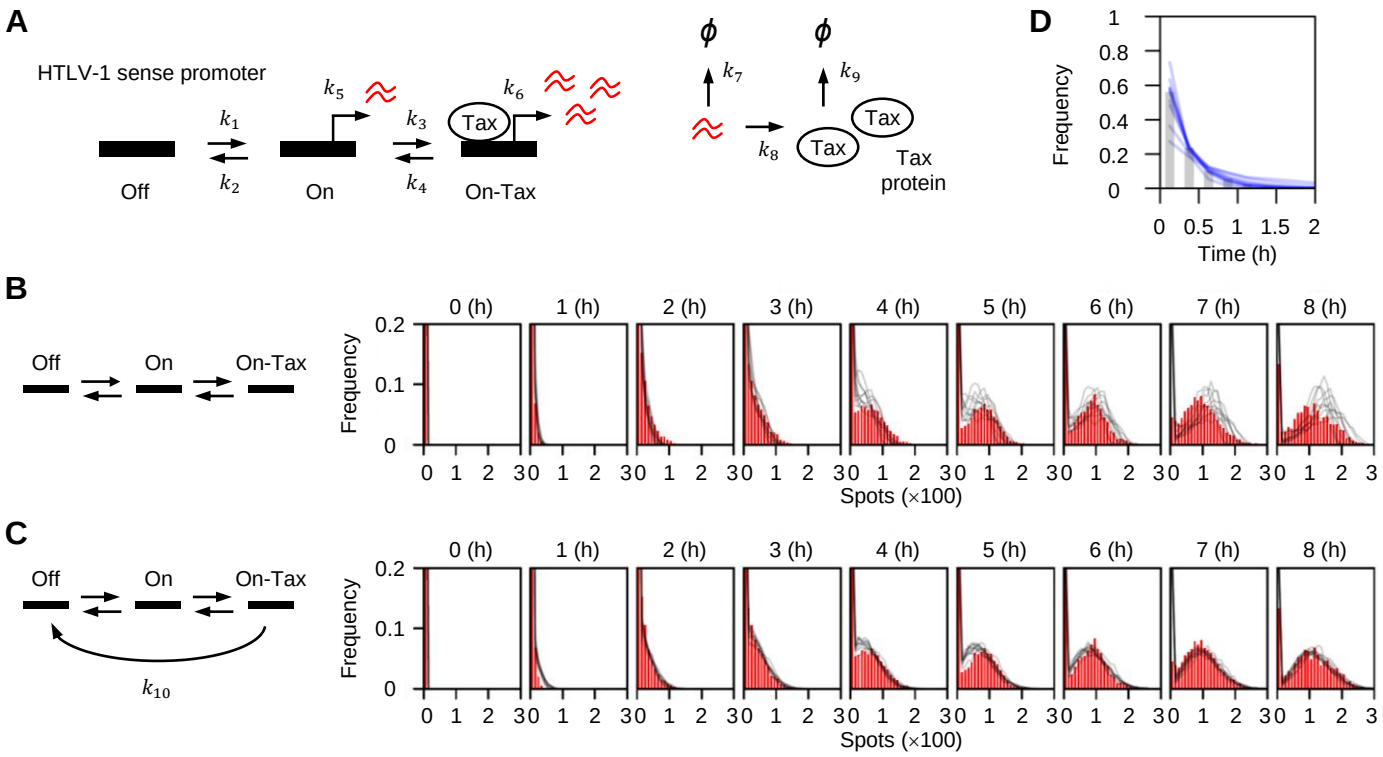
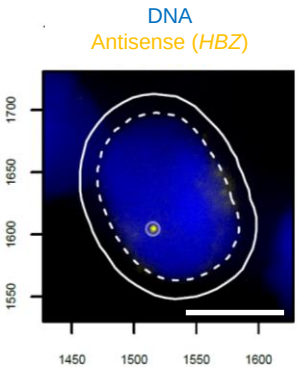
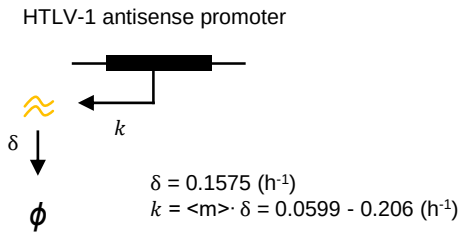

A
HTLV-1 sense promoter
Off
On
On-Tax
Tax
Tax protein
B
Off
On
On-Tax
Frequency
Spots (×100)
C
Off
On
On-Tax
Frequency
Spots (×100)
D
Frequency
Time (h)

Fig 5

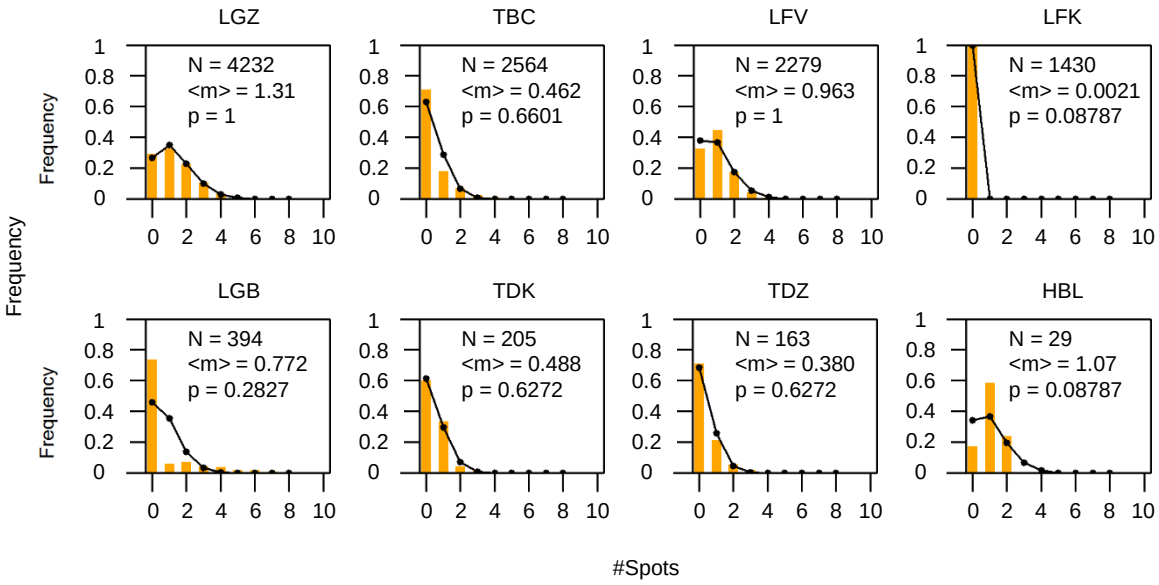
A


DNA
Antisense (*HBZ*)
1700
1650
1600
1550
1450
1500
1550
1600

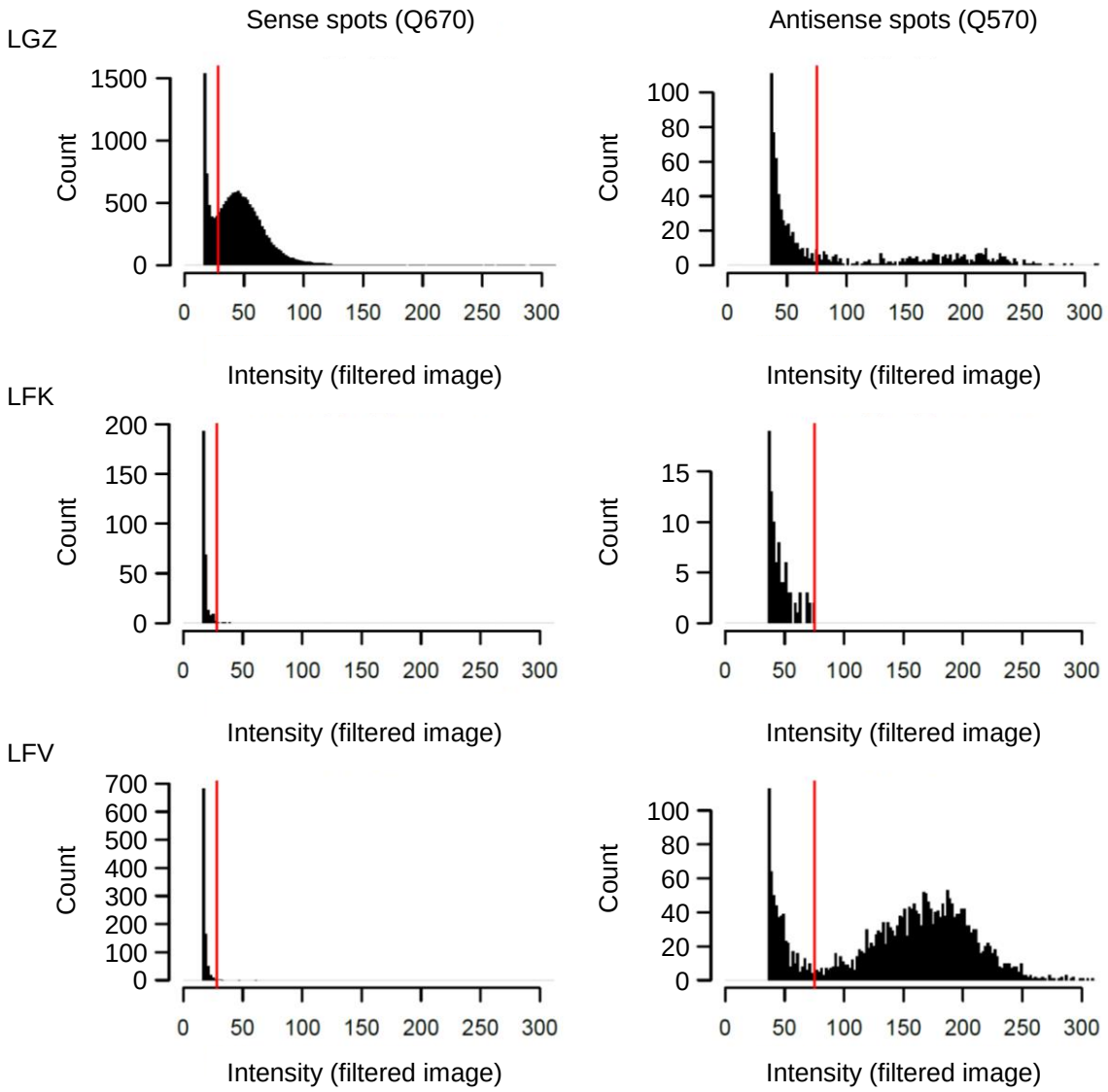
C


HTLV-1 antisense promoter
k
δ
ϕ
δ = 0.1575 (h-1)
k = <m>· δ = 0.0599 - 0.206 (h-1)

B


LGZ
N = 4232
<m> = 1.31
p = 1
TBC
N = 2564
<m> = 0.462
p = 0.6601
LFV
N = 2279
<m> = 0.963
p = 1
LFK
N = 1430
<m> = 0.0021
p = 0.08787
LGB
N = 394
<m> = 0.772
p = 0.2827
TDK
N = 205
<m> = 0.488
p = 0.6272
TDZ
N = 163
<m> = 0.380
p = 0.6272
HBL
N = 29
<m> = 1.07
p = 0.08787
Frequency
#Spots

S1 Fig



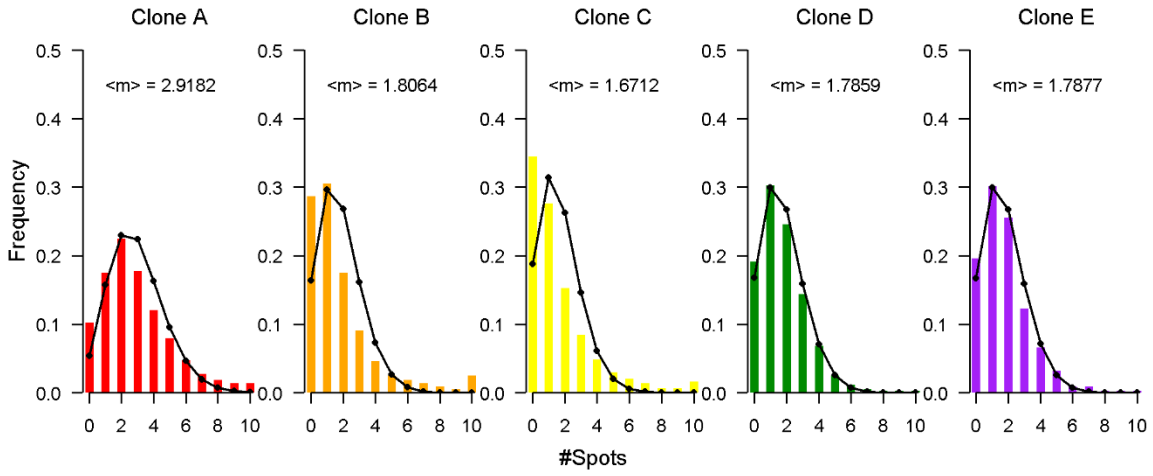
S1 Fig Determination of the threshold for mRNA spot detection.

Histograms of the intensity of spots pre-detected with FISH-QUANT (1) with a conservative threshold setting. The first column shows the spot intensity for Q670, and the second column for Q570. Subsequently, a tighter threshold was set for each of the channels (red vertical lines) to remove the spurious signals. The samples shown in the panel (LGZ incubated and fixed at 7h, LFK and LFV fixed without incubation) were hybridised and imaged in a single batch.

Reference

(1) Mueller F, Senecal A, Tantale K, Marie-Nelly H, Ly N, Collin O, et al. FISH-quant: automatic counting of transcripts in 3D FISH images. Nat Methods. 2013;10(4):277-8. Epub 2013/03/30. doi: 10.1038/nmeth.2406. PubMed PMID: 23538861.

S2 Fig



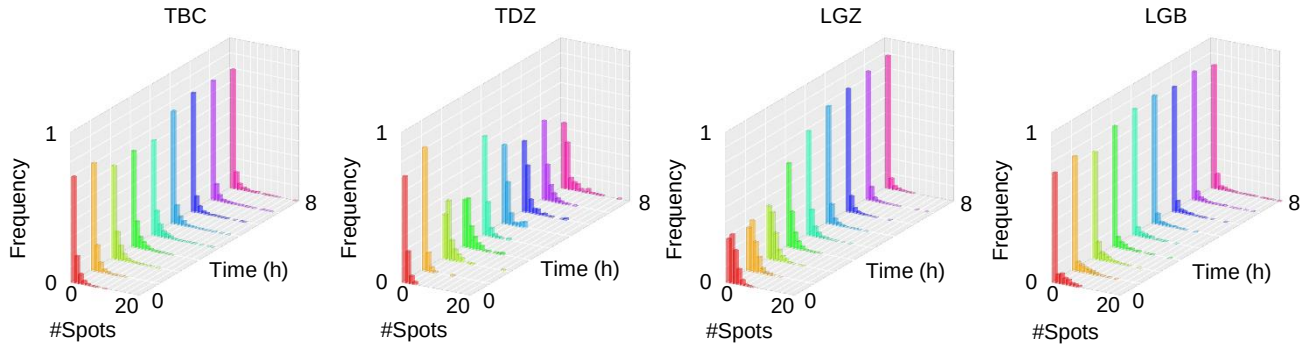
S2 Fig Antisense *HBZ* expression in *in vitro* maintained HTLV-1⁺ T cell clones.

The distribution of the number of *HBZ* molecules per cell in *in vitro* maintained HTLV-1⁺ T cell clones. Histograms were reproduced from the graphs presented in Billman *et al* (1). The mean number of *HBZ* molecules (<m>) is indicated in the inset. (The last bin is indicated as 10 or more spots in the original paper (1). However, it is treated as =10 spots when the mean <m> was reproduced in this figure.) The black line indicates the theoretical Poisson distribution with the parameter <m>, the observed mean number of *HBZ* molecules.

Reference

(1) Billman MR, Rueda D, Bangham CRM. Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. Wellcome Open Res. 2017;2:87. Epub 2017/10/25. doi: 10.12688/wellcomeopenres.12469.2. PubMed PMID: 29062917; PubMed Central PMCID: PMCPMC5645716.

S3 Fig



S3 Fig HTLV-1 antisense expression in PBMCs *ex vivo* culture over time.

HBZ mRNA count per HTLV-1^{+} cell at successive timepoints during *in vitro* incubation. The data from two patients with HAM (TBC and TDZ) and two with ATL (LGZ and LGB) are presented.

S1 Table

S1 Table Top 20 parameter sets that best fitted the smFISH data (LGZ).

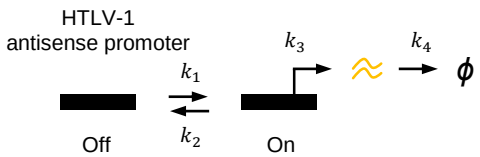
Transition					Transcription		Translation	Degradation	
k_1	k_2	k_3	k_4	k_{10}	k_5	k_6	k_8	k_7	k_9
1.595	8.224	9.837	7.644	4.789	15.4	91.9	2.544	0.0785	0.0132
1.272	8.985	1.216	3.053	2.872	42.6	83.0	5.790	0.1101	0.0544
1.095	6.269	8.059	3.075	2.437	20.3	90.4	3.208	0.1687	0.1703
1.190	9.770	6.431	6.358	2.768	46.6	81.6	3.163	0.0911	0.0553
1.292	9.419	6.147	5.521	3.151	21.7	84.6	6.098	0.1085	0.0109
1.311	9.306	5.842	2.461	2.470	36.0	60.8	1.553	0.0181	0.1460
1.019	8.213	5.118	2.707	3.283	31.5	81.6	9.186	0.0390	0.0740
1.023	7.859	2.436	4.458	1.283	33.8	47.3	8.328	0.0773	0.0147
1.243	9.578	4.568	3.580	3.768	26.4	83.1	9.271	0.0708	0.0472
1.200	4.058	0.924	3.833	0.867	18.6	59.2	7.120	0.1812	0.1934
1.146	2.699	7.969	2.387	3.830	11.8	87.0	1.638	0.0424	0.1212
1.536	9.049	9.062	3.598	3.435	17.3	63.2	7.486	0.0261	0.1078
1.255	9.010	6.955	5.407	1.786	28.7	71.9	1.786	0.1907	0.1059
1.893	9.823	5.064	2.686	5.917	12.4	98.1	4.789	0.0295	0.1118
1.343	9.362	9.700	3.129	2.423	25.5	73.4	3.691	0.1681	0.0320
1.605	8.837	6.730	7.361	3.212	19.7	86.2	2.312	0.1907	0.0388
1.745	7.933	6.367	6.245	3.445	10.1	83.8	8.238	0.1200	0.0456
1.311	4.703	6.011	6.344	2.316	20.8	76.0	1.873	0.1905	0.1180
1.651	9.306	5.720	4.733	3.311	13.0	73.9	5.087	0.1156	0.0459
1.337	7.397	6.281	9.775	3.896	28.9	99.0	2.825	0.1136	0.1366

The 10 parameters (k_1 to k_{10}) are classified into four categories (the promoter transition, transcription, translation and degradation). Each row indicates the iteration of the stochastic simulation with the 10 parameters. The 20 sets of parameters that best fitted the smFISH data (LGZ) are listed.

S1 Text

An alternative model accounting for the departure from the Poisson distribution

We observed the deviation of the *HBZ* count per cell in some of the *in vitro* maintained HTLV-1$^+$ T cell clones from the expectation with the Poisson distribution (1). Of note in this observation is the higher number of *HBZ*-negative cells in those clones at a given time. We suspected that this deviation is due to the antisense strand being transcribed in a small and occasional burst. Therefore, we applied the following two-state model to the clone B as an example.



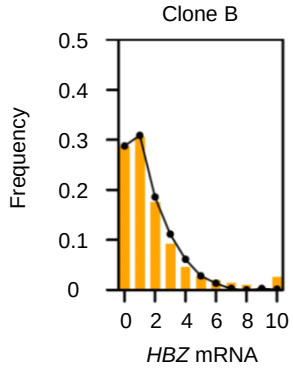
The number of *HBZ* molecules <m> at the steady state can be described as

$$< m > = \frac{k_3[On]}{k_4} = \frac{k_3}{k_4} \cdot \frac{k_1}{k_1 + k_2}$$

We estimate the degradation rate k_4 to be 0.1575 (h^{-1}) based on our previous publication (1). We set out the parameter search for k_1 and k_3 within the range of 0.001 to 10 and 0.05 to 50, respectively. The parameter k_2 was constrained by

$$k_2 = \frac{k_1 k_3}{< m > \cdot k_4} - k_1$$

We obtained a parameter set (k_1=0.658, k_2=94.524, k_3=41.14 and k_4=0.158) which yielded the distribution indicated below with the black line.



With this parameter set, we estimate the burst frequency and intensity as follows:

burst frequency = k_1[Off] = $k_1 \cdot k_2$ / (k_1+k_2) ~ 0.653 (h^{-1})

burst intensity = k_3 / k_2 ~ 0.435 (molecules / burst)

Reference

(1) Billman MR, Rueda D, Bangham CRM. Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. Wellcome Open Res. 2017;2:87. Epub 2017/10/25. doi: 10.12688/wellcomeopenres.12469.2. PubMed PMID: 29062917; PubMed Central PMCID: PMCPMC5645716.