Check for updates

RESEARCH ARTICLE

REVISED **Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1** [version 2; referees: 2 approved, 1 approved with reservations]

Martin R Billman[1], David Rueda[2,3], Charles R M Bangham iD [1]

[1]Section of Virology, Division of Infectious Diseases, Department of Medicine, Imperial College London, Norfolk Place, London, UK
[2]Single Molecule Imaging Group, MRC London Institute of Medical Sciences, Du Cane Road, London, UK
[3]Section of Virology, Division of Infectious Diseases, Department of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London , UK

Abstract

Background: The human leukaemia virus HTLV-1 expresses essential accessory genes that manipulate the expression, splicing and transport of viral mRNAs. Two of these genes, *tax* and *hbz*, also promote proliferation of the infected cell, and both genes are thought to contribute to oncogenesis in adult T-cell leukaemia/lymphoma. The regulation of HTLV-1 proviral latency is not understood. *tax,* on the proviral plus strand, is usually silent in freshly-isolated cells, whereas the minus-strand-encoded *hbz* gene is persistently expressed at a low level. However, the persistently activated host immune response to Tax indicates frequent expression of *tax in vivo*.
Methods: We used single-molecule RNA-FISH to quantify the expression of HTLV-1 transcripts at the single-cell level in a total of >19,000 cells from five T-cell clones, naturally infected with HTLV-1, isolated by limiting dilution from peripheral blood of HTLV-1-infected subjects.
Results: We found strong heterogeneity both within and between clones in the expression of the proviral plus-strand (detected by hybridization to the *tax* gene) and the minus-strand (*hbz* gene). Both genes are transcribed in bursts; *tax* expression is enhanced in the absence of *hbz*, while *hbz* expression increased in cells with high *tax* expression. Surprisingly, we found that *hbz* expression is strongly associated with the S and G_2/M phases of the cell cycle, independent of *tax* expression. Contrary to current belief, *hbz* is not expressed in all cells at all times, even within one clone. In *hbz*-positive cells, the abundance of *hbz* transcripts showed a very strong positive linear correlation with nuclear volume.
Conclusions: The occurrence of intense, intermittent plus-strand gene bursts in independent primary HTLV-1-infected T-cell clones from unrelated individuals strongly suggests that the HTLV-1 plus-strand is expressed in bursts *in vivo*. Our results offer an explanation for the paradoxical correlations observed between the host immune response and HTLV-1 transcription.

Open Peer Review

Referee Status: ✓ ✓ ?

	Invited Referees		
	1	**2**	**3**
REVISED **version 2** published 11 Dec 2017			
version 1 published 22 Sep 2017	✓ report	✓ report	? report

1 **Vincenzo Ciminale**, University of Padua, Italy

2 **Ali Bazarbachi**, American University of Beirut, Lebanon
Hiba El Hajj, American University of Beirut, Lebanon

3 **Marc Sitbon** iD , Institute of Molecular Genetics of Montpellier, France

Discuss this article

Comments (0)

Corresponding authors: David Rueda (d.rueda@imperial.ac.uk), Charles R M Bangham (c.bangham@imperial.ac.uk)

Competing interests: No competing interests were disclosed.

How to cite this article: Billman MR, Rueda D and Bangham CRM. **Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1 [version 2; referees: 2 approved, 1 approved with reservations]** Wellcome Open Research 2017, **2**:87 (doi: 10.12688/wellcomeopenres.12469.2)

First published: 22 Sep 2017, **2**:87 (doi: 10.12688/wellcomeopenres.12469.1)

Introduction

Human T-lymphotropic virus type 1 (HTLV-1) was the first human retrovirus to be discovered and infects approximately 10 million individuals around the world, causing an aggressive leukaemia or lymphoma or progressive lower limb weakness and paralysis in approximately 10% of infected individuals[1]. Once infection is established in CD4+ and CD8+ T-lymphocytes, it persists lifelong in the host. The virus appears to be latent in the blood; however, the continuously activated anti-HTLV-1 immune response indicates frequent or persistent expression of HTLV-1 in vivo[1–3]. The regulation of HTLV-1 expression in vivo is not well understood.

In addition to the gag, pol and env genes common to all exogenous retroviruses, HTLV-1 encodes a pX region[1], which undergoes alternative splicing to express six accessory proteins that regulate transcription, splicing and transport of viral mRNAs. The accessory proteins also manipulate several key functions in the host cell. The two most important pX products are Tax, on the plus strand of the genome, and HBZ, the only gene encoded on the minus strand[4,5]. Several actions of Tax and HBZ are mutually antagonistic, but both Tax and HBZ play crucial roles in viral persistence, gene expression and leukaemogenesis[5,6]. Understanding how their expression is controlled is a key step towards understanding latency and expression of HTLV-1 in the host.

Earlier studies of HTLV-1 proviral expression have focused, at the cell population level, on detection either of protein[2,7,8] (e.g. by flow cytometry) or nucleic acid[8,9] (e.g. by qPCR). Neither of these approaches is appropriate for HBZ, because it is expressed at a level near the limit of detection of current assays, including qPCR. However, the immune response to HBZ is an important correlate of the outcome of HTLV-1 infection[10]. In addition, assays of viral expression in a cell population masks any

heterogeneity of expression at the single-cell level. It is imperative to identify the extent and causes of such single-cell heterogeneity in order to understand the regulation of proviral latency.

We describe the use of single-molecule fluorescent in situ hybridisation (smFISH) to quantify the transcripts of plus-strand and hbz mRNA in individual cells of naturally-infected T-cell clones, isolated from patients' peripheral venous blood. We found that both the plus-strand and the minus-strand of the HTLV-1 provirus are expressed in intermittent bursts, with a surprising level of heterogeneity at the single-cell level in the expression of both the hbz gene and, especially, the plus-strand. The results reveal fundamental differences in the regulation of transcription of the provirus plus- and minus-strands, and suggest an explanation for the paradoxical differential effectiveness of the cytotoxic T-lymphocyte immune response to Tax and HBZ that is characteristic of HTLV-1 infection[11].

Methods

Derivation of T-lymphocyte clones from infected patients

Peripheral blood mononuclear cells (PBMCs) were isolated from the donated blood of HTLV-1+ patients, before individual clones were isolated and cultured as described in 12. Cells were distributed in 96-well plates at ~1 cell/well, using limiting dilution. The cells were then cultured with irradiated feeder cells, PHA, IL-2 and the retroviral integrase inhibitor raltegravir. Wells containing proliferating cells were tested for infection and proviral integrity using PCR. Linker-mediated PCR was then used as previously described to identify the proviral integration site and to verify that the population was indeed monoclonal[13].

The clones used, their integration sites and the patients they were derived from are summarised below:

Clone	Patient	HTLV-1 Integration site
A	TBJ	Chr 4: 70,567,285
B	TBX	Chr 22: 44,323,168
C	TBW	Chr 19: 33,829,548
D	TBJ	Chr 14: 46,204,950
E	TCX	Chr 16: 53,601,059
F	TBW	Uninfected

Cell culture, preparation and fixation

Patient-derived T-lymphocyte clones were cultured in RPMI-1640 medium (Sigma-Aldrich) with added L-glutamine (Invitrogen), penicillin and streptomycin (Invitrogen) and 10% AB human serum (Invitrogen) at 37°C, 5% CO_2. IL-2 (Promokine) was added to the culture every 3 days, and the concentration of raltegravir (Selleck) was maintained throughout cell culture. In addition, the cells were activated every 14 days by the addition of beads coated with antibodies against CD2, CD3 and CD28 (Miltenyi-Biotech). All experiments were carried out on cells harvested on Day 8 of this cycle, after addition of fresh media on Day 7. Each clone was analysed in triplicate; cells from each triplicate sample were cultured separately for at least 24 hours before fixation.

Cells were added to glass coverslips (SLS, 12mm, number 1) coated with poly-L-lysine (Sigma-Aldrich), before being fixed in 2% formaldehyde (Life Technologies, in PBS) at room temperature for 15 minutes. Cells were then transferred to 70% ethanol for permeabilization or long-term storage at -20°C.

Actinomycin D treatment

To quantify transcript half-life and intracellular distribution, and to investigate the link between bursts and transcription, cells were treated with actinomycin D (ActD) to block transcription. ActD (Sigma-Aldrich) was dissolved in DMSO (Sigma-Aldrich) at 1.5 mg/ml before being added to RPMI medium to a final concentration of 1.5 µg/ml. Two clones were tested, each in two biological replicates; DMSO alone was used as a loading control. After addition of ActD or DMSO, cells were fixed after 0, 2, 4, 6 and 22 hours of culture at 37°C before being stained, imaged and analysed.

RNA-fluorescence in-situ hybridization (RNA-FISH)

RNA-FISH was carried out in accordance with the manufacturer's protocols for the Stellaris probe system (Biosearch Technologies). Cells were permeabilised in 70% ethanol before being washed and incubated overnight at 37°C with the desired RNA-FISH probes (Stellaris). The cells were subsequently washed with DAPI (Sigma-Aldrich) and mounted in Vectashield mounting medium (Vectashield) on glass slides (Vector) and sealed with clear nail-varnish for imaging.

RNA-FISH probes were designed using Stellaris's online probe design tool (version 4.1), with the reference genome AB513134 (Fan *et al.,* 2010[14]) used as input for the plus-strand probe (post-splice coding sequence) and *hbz* (post-splice coding sequence and 3'-UTR as reported in Kamentsky *et al.,* 2011[15]); the oligonucleotides used in each probe set are shown in Supplementary Figure 13. Probes targeting *hbz* were labelled with a Quasar-570 fluorophore and those targeting the plus-strand with a Quasar-670 fluorophore.

Fluorescence microscopy

All coverslips were imaged using an Olympus IX70 inverted widefield microscope with a 100x-1.35NA UPlanApo oil objective and a Spectra Light Engine illumination source (Lumencor), with a resolution of 64 nm per pixel, and a spacing of 300 nm per optical slice. Z-offset was applied to correct for chromatic aberration, and exposures were adjusted for each respective wavelength to optimize the signal-noise ratio and minimize photobleaching. Excitation filters used were 390/18 (DAPI), 575/25 (*hbz*-Quasar 570) and 632/22 (plus-strand-Quasar 670), with corresponding emission filters of 457/20, 632/60 and 692/40. A Coolsnap HQ camera (Roper Scientific) was used for initial images; the camera was later replaced with an ORCA-Flash 4.0 V2 digital CMOS camera (Hamamatsu).

Image processing and analysis

After acquisition, all images passed through the following pipeline:

1. Visual inspection of image-stacks to ensure cells were captured in their entirety in the z-axis. Individual cells were excluded from the analysis. If more than a handful of cells were 'cut-off', the image was discarded.

2. Discarding of unneeded optical slices above and below cells in the image-stack, reducing image size to minimize subsequent processing and computing time.

3. Correction for uneven illumination. An illumination function was calculated using CellProfiler v2.2.0[15], with a pipeline adapted from Roukos *et al.,* 2015[16].

Images were then analysed with an adapted version of FISH-Quant, v3[17,18], run on MATLAB. The adaptation allowed step number 4, below:

1. Outlines marking cells and nuclei were automatically generated by thresholding and then manually double-checked.

2. All local maxima signals were characterised, to determine the signal intensity threshold which will best separate true and false positives, and calculate an optimal intensity threshold to define transcriptional 'bursts'.

3. Spots passing intensity and shape thresholds were counted in each individual cell.

4. The stage of the cell-cycle was identified from the integrated intensity of each cell's nuclear DAPI signal. While carried out in FISH-Quant, this process was adapted from the method described in Roukos *et al.,* 2015[16]. Cell-cycle gates were determined by visual inspection of the histograms, following the rule that cells in G_2/M have double the DNA content (and therefore integrated intensity) of cells in G_0/G_1 (Figure 5a; Supplementary Figure 15). Varying the width of the cell-cycle gates by 10% did not materially affect the conclusions.

5. The individual measurements – counts of putative single mRNA molecules, nascent 'burst' counts and cell cycle stage were collated and interpreted.

Finally, cells containing putative bursts were analysed individually using Imaris 3D-analysis software (Bitplane) to identify the 3D location of the burst relative to the 'centre-of-gravity' of the respective cell nucleus; the intranuclear location was then normalised by nuclear volume. Nuclear volume was estimated from the circle of best fit to the periphery of the DAPI staining, assuming a perfectly spherical nucleus.

Statistical analysis

For each clone, three biological replicates were used, unless stated otherwise in figure legends. With exception of the logistic regression, all statistical analyses were carried out on Graphpad Prism 6, which was also used to create all graphs. Significance was symbolised in the following manner: ns ($p > 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$) and **** ($p < 0.0001$).

In Figure 2, chi-squared tests were used to analyse burst frequency (Figure 2c) and Mann-Whitney tests to analyse burst size and location (Figure 2); all tests were two-tailed.

In Figure 3, Unpaired t-tests with Welch's correction were used to analyse the changes in proportion of spots found within the nucleus (Figure 3b). The half-life of *hbz* after ActD treatment was estimated using both a one-phase decay exponential fit of the raw total *hbz* per cell, and by linear regression of log (total cellular *hbz* per cell) against time (Figure 3c).

The changes in frequency of plus-strand and *hbz* bursts shown in Figure 4 were all analysed using chi-squared tests, while the relationship between *hbz* and plus-strand expression and cell cycle stage was examined using binary and multinomial logistic regression with the "glm" and "multinom" functions respectively in R, v3.3.3[19].

The expected frequency of *hbz*-negative cells in each clone was calculated from the observed mean *hbz* spot count in the respective clone (Supplementary Table S1), using the Poisson distribution. The observed and expected numbers of *hbz*-positive and –negative cells were compared using chi-squared tests.

Results

smFISH reveals single-cell heterogeneity in HTLV-1 expression

To quantify HTLV-1 expression at the single-cell level, and to identify any cell-to-cell heterogeneity that is lost in population-averaged approaches such as qPCR, we designed fluorescent-labelled probes to image mRNA from the plus-strand and the minus-strand (*hbz*) of the provirus (Figure 1a). The probe sequence we used to detect the plus-strand, in the pX region of the provirus, is present in all plus-strand transcripts. These probes were used to stain cells from five HTLV-1-infected T-cell clones, derived from patients' PBMCs by limiting dilution; biological triplicates of each clone were studied in independent experiments. Each clone has a single, unique viral integration site[12]; the provirus is complete in four clones, while in the fifth clone (Clone E) the 5'-LTR is deleted and the plus-strand thereby silenced, a phenomenon common in leukaemic clones[20]. The diffraction-limited signal of a single spot in smFISH corresponds to a single mRNA molecule[21]. In total, 19,477 individual cells were analysed, quantifying plus-strand and *hbz* transcripts and the proportion of transcripts within the nucleus, and identifying the cell cycle stage in each cell.

The results show that *hbz* and the plus-strand have extremely different expression profiles (Figure 1b). *hbz* levels were very low, as expected from previous observations in naturally-infected cells[8]. A majority of cells were *hbz*+, but expression was not universal in any clone, ranging from 90% positive down to 65% (Supplementary Figure 1). The *hbz*+ cells formed a unimodal population of low-expressing cells, most containing between 0 and 5 transcripts (Figure 1b.ii). The frequency of cells with successively higher expression levels of *hbz* diminished rapidly: in ~20,000 cells, the highest observed *hbz* spot count was 25, although the range of expression varied between clones. The variation in total *hbz* mRNA between clones was explained by the variation in the proportion of *hbz*-expressing cells, not by differences in the mean level of *hbz* expression (Figure 1c).

In contrast, the plus-strand presented a bimodal expression profile. In most clones, only a small fraction of cells were plus-strand+, but these cells expressed the plus-strand at a much greater level than *hbz*. Across all clones, cells with 'high' plus-strand expression (≥ 100 spots, Figure 1b.v) made up 4.9% of all cells, but contributed over 80% of all plus-strand mRNA detected. The variation in total plus-strand RNA between clones was due to the variation in the proportion of cells with this high level of expression (Figure 1d).

Three of the clones studied (A-C) each had a similar range of intensity of expression. The maximum level of plus-strand RNA observed in clone D was much lower; however, all four clones had a bimodal distribution of plus-strand expression (Supplementary Figure 2).

HTLV-1 transcription occurs in bursts

In addition to the small proportion of cells with very high levels of expression, a characteristic feature of plus-strand expression was that many plus-strand+ cells also had a large nuclear spot which was typically much brighter than the average spot (Figure 2a). We surmised that these bright spots corresponded to transcriptional bursts, which are increasingly recognized in mammalian genes[22,23]. In a transcriptional burst, transcripts are created faster than they diffuse away, and consequently they cannot be individually resolved. The resulting spot is larger and brighter than the average spot (Figure 2b). The same characteristic spots were also identified for *hbz*, but with a smaller difference from average spots in intensity or size.

To test the hypothesis that the bright spots represent transcriptional bursts, we treated two of the clones (A and B) with ActD, to inhibit transcription and allow the putative bursts to disperse and disappear. Bursts were defined as intranuclear spots whose intensity significantly exceeded the main distribution of spot intensities (Supplementary Figure 3). In 19,437 analysed cells, ActD treatment strongly and progressively reduced the frequency of these bright spots of both plus-strand and *hbz* (Figure 2c). These observations are consistent with the view that the bright spots represent transcriptional bursts. A small proportion of the bright spots may be chance superpositions of multiple transcripts, but we conclude that a majority represent ongoing or recent transcription. Hereafter, we refer to these bright spots as bursts.

Plus-strand bursts and *hbz* bursts differ in intensity

The burst sizes of the plus-strand and *hbz* were estimated by averaging thousands of individual diffraction-limited spots, which putatively represent single transcripts, and superposing their point-spread function (PSF) on that of the burst to estimate the number of transcripts present, given the 3D shape and intensity of that burst. Using this technique, we observed that *hbz* bursts were uniformly small in size, typically containing 3 to 4 spots, with very low cell-to-cell variation (Figure 2d). Conversely, plus-strand bursts were much larger: the largest bursts were estimated to contain hundreds of transcripts. Plus-strand bursts were also more variable than *hbz* bursts in size (spot count). Bursts, whether plus-strand or *hbz*, were not uniformly distributed throughout the
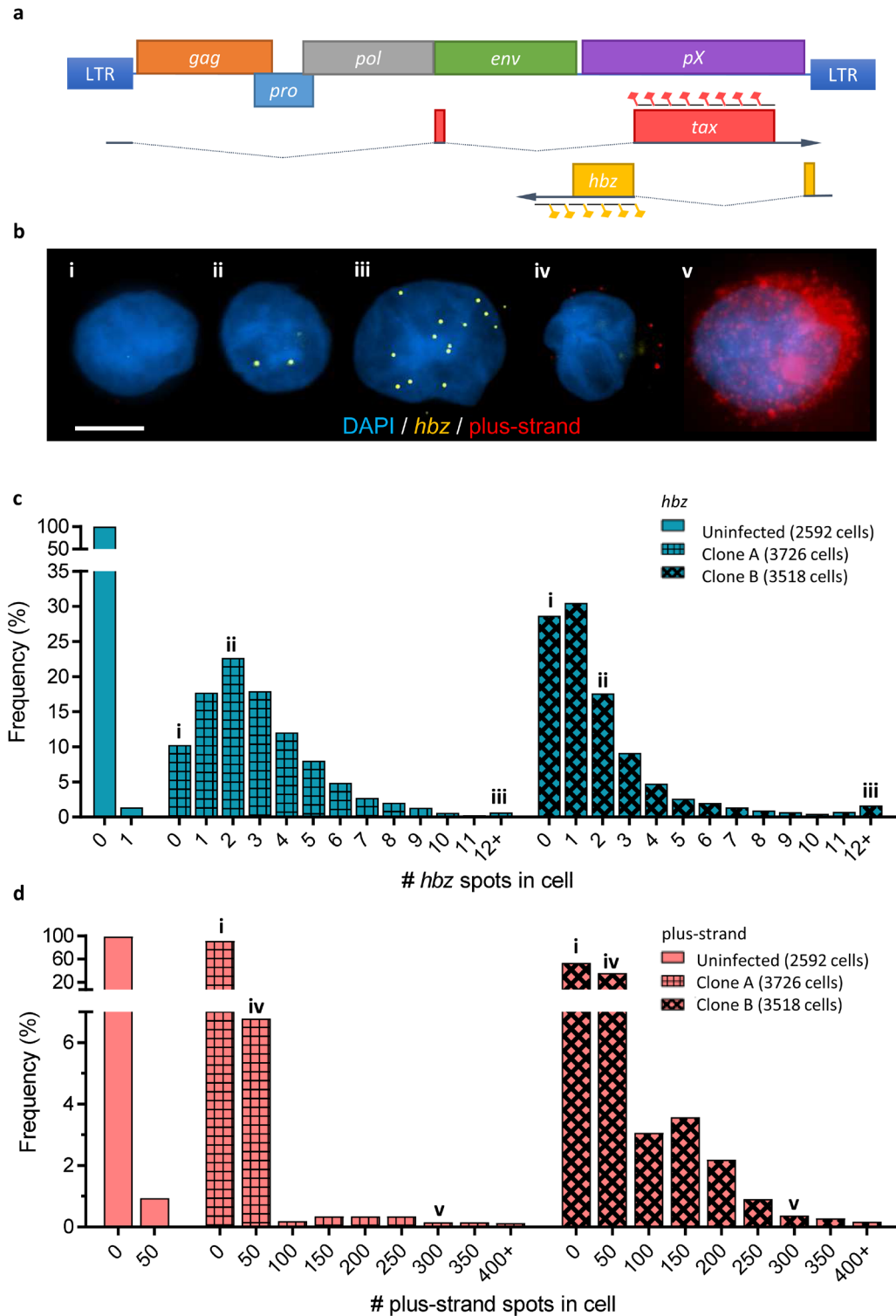
Figure 1. Single-molecule RNA-FISH reveals heterogeneity of plus-strand and *hbz* expression within and between clonal cell populations. a. Schematic of the HTLV-1 genome, showing the spliced transcripts of the plus-strand and *hbz*; the sequences targeted by probes on the plus-strand and the minus are marked by red and yellow flags respectively. **b**. Composite fluorescent micrograph with representative examples of cells expressing different levels of *hbz*: i) silent, ii) low (<5 spots) and iii) high (≥5 spots); and plus-strand RNA: iv) low (<100 spots), v) high (≥100 spots). The Roman numerals show the respective cells' positions in panels **c** and **d**. Scale bar is 5 μm. **c**. Frequency distribution of *hbz* expression (number of *hbz* spots/cell) in individual cells, from one uninfected and two HTLV-1-infected clones. **d**. Frequency distribution of plus-strand expression in individual cells (same clones as in **c**); bins show the number of cells with 0-1 plus-strand spots ("0"), 2-50 spots ("50"), 51-100 spots ("100"), 101-150 spots ("150") etc.

Figure 2. Both plus-strand and *hbz* are expressed in bursts; plus-strand bursts are much larger than *hbz* bursts. a. Representative plus-strand and *hbz* bursts. Dashed lines show the positions of the intensity profiles in panel b. Scale bar is 5 μm. **b**. Bursts are defined as specifically stained intranuclear spots that are significantly brighter than the main population of spots. **c**. Frequency of 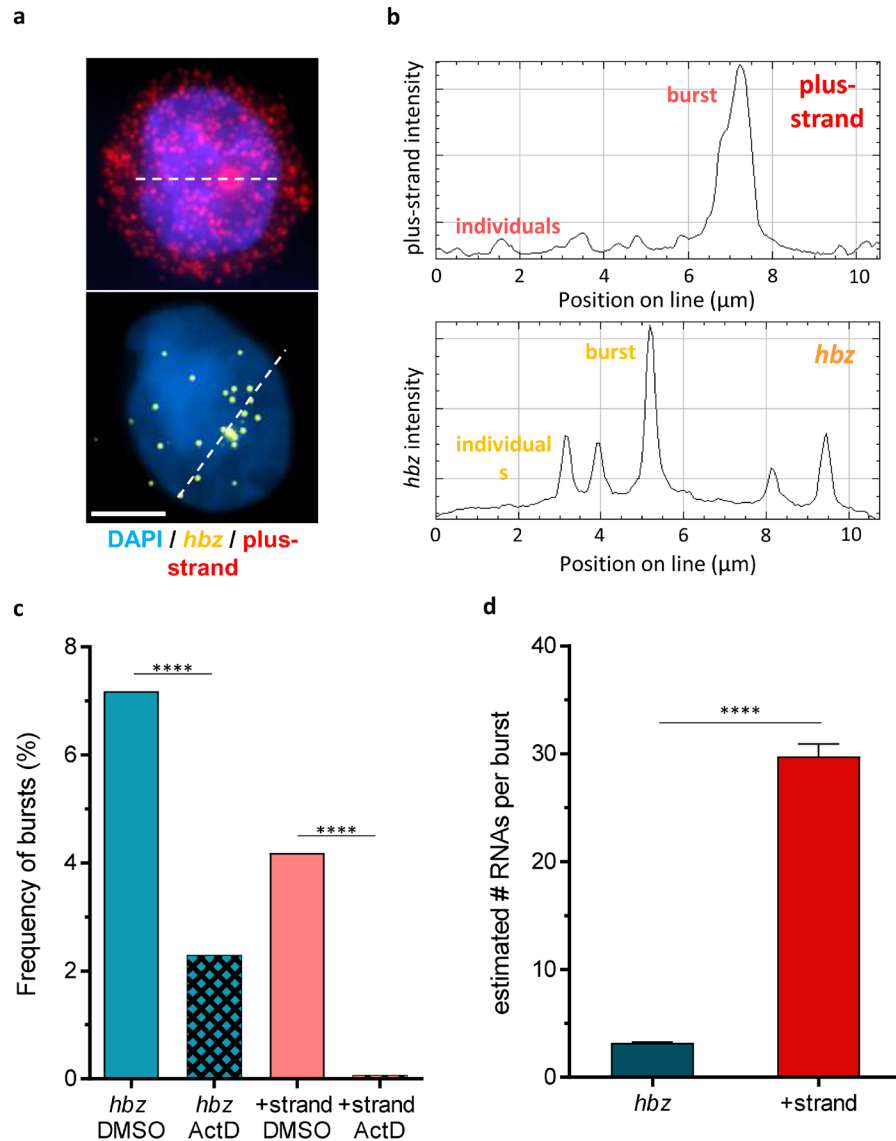bursts is significantly reduced by blocking transcription with actinomycin D (n = 12,317 DMSO-treated cells, 7,120 actinomycin-D-treated cells from two replicates each of clones A and B). **d**. *hbz* bursts are small and uniform in size, whereas plus-strand bursts are significantly larger and more variable. SEM is shown, with 514 plus-strand and 800 *hbz* bursts from the pooled plus-strand-competent clones.

nucleus, and their spatial localization differed between clones (Supplementary Figure 4).

Plus-strand RNA is rapidly exported, *hbz* remains largely intranuclear

Tax protein generates a strong, persistently activated T-lymphocyte response in infected people[10]. HBZ protein, in contrast, is expressed near the limit of detection of current methods, and T-cell receptor avidity for HBZ epitopes is correlated with both the proviral load and the disease outcome[3,10]. We therefore investigated the proportion of both transcripts that could be found in the nucleus or cytosol of cells.

Again, the differences between plus-strand RNA and *hbz* were stark. About 90% of all *hbz* spots were found in the nucleus, in each clone examined (Figure 3b). This proportion did not change significantly after transcription had been blocked for 22 hours with ActD (Figure 3a.iii).

In contrast, ~60% of plus-strand transcripts were found in the nucleus in 3 of 4 plus-strand-expressing clones. This percentage also fell after transcription was blocked (Figure 3a,b). These estimates of the proportion of intranuclear mRNAs are not exact because they are derived from 2D maximum projections of the 3D data: spots lying above or below the nucleus along the
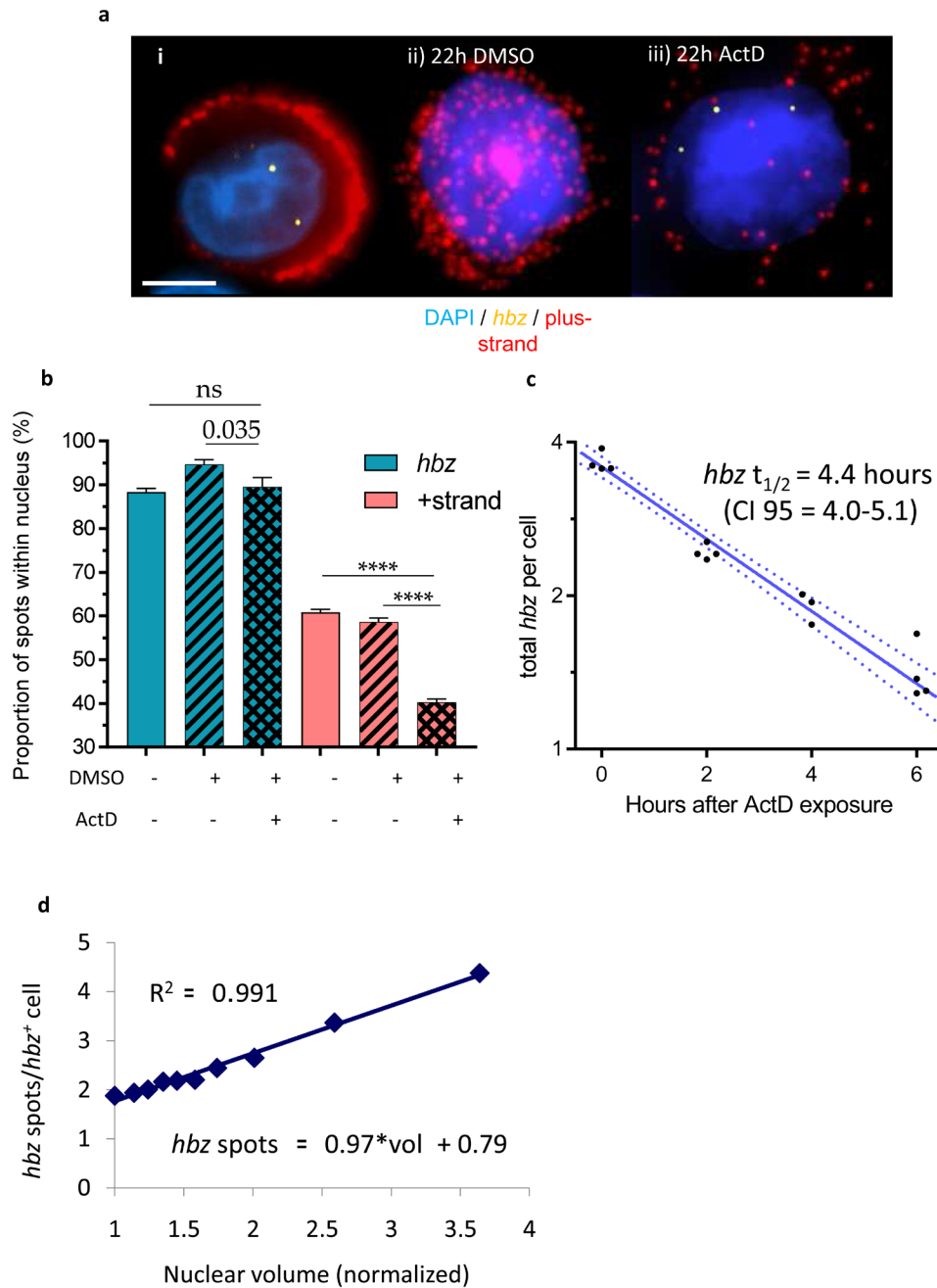
Figure 3. plus-strand RNA is exported from the nucleus, whereas the majority of *hbz* is retained. a. Representative images showing i) an optical slice illustrating the difference in cellular localisation of plus-strand RNA and *hbz*, ii) a typical plus-strand-high cell after 22 hours in DMSO and iii) a cell after 22 hours in actinomycin D. Image a.i is a single slice from a z-stack whereas ii and iii are maximum projections. Scale bar is 5 μm. **b**. *hbz* mRNA remains largely intranuclear, whereas plus-strand mRNA is mostly cytoplasmic; this difference is accentuated by blocking transcription with actinomycin D. n = 936 untreated, 390 DMSO-treated and 195 actinomycin D-treated cells, each from two pooled biological replicates of clones A and B. **c**. Rate of disappearance of total *hbz* spots following treatment with actinomycin D. The slope of the log-transformed data is the rate of decay. Individual measurements from replicates are shown, along with a linear regression and 95% confidence intervals (dashed lines); one of the four replicates failed at the 4 hour mark. n = 8126 total cells. **d**. Mean *hbz* spot count in *hbz*-positive cells was strongly correlated with nuclear volume.

z-axis are erroneously marked as intranuclear. This explains why the proportion of plus-strand RNA within the nucleus did not fall below 40% after 22 hr treatment with ActD, when 3D images indicate that the true proportion was lower. (Figure 3a.i, iii). The fourth clone studied (clone D) was again an exception: in this clone, the proportion of plus-strand RNA within the nucleus was comparable to that of *hbz* (Supplementary Figure 5).

hbz mRNA abundance correlates with nuclear volume

The nuclear volume in each cell was estimated from the diameter of the DAPI staining in the maximum-projection image. There was a strong linear correlation between the nuclear volume and the number of *hbz* spots per cell (Figure 3d). However, after taking *hbz* spot count into account, the frequency of *hbz* bursts was not correlated with nuclear volume (Table S1d).

hbz mRNA has a half-life of 4.4 hours

We estimated the half-life of *hbz* mRNA from the decline in *hbz* spot count during treatment with ActD. The high density of plus-strand spots in many plus-strand-expressing cells precluded a precise spot count of plus-strand RNA (Supplementary Figure 14); as a result, it was not possible to make a reliable estimate of plus-strand RNA half-life.

Two biological replicates each of two separate clones were exposed to ActD (or DMSO as control) for 0, 2, 4 and 6 hours (Supplementary Figure 6). A total of 15,524 cells were studied. The total counts of *hbz* spots, including the estimated sizes of bursts, were used to estimate the mean *hbz* content per cell; this mean value decreased over time (Figure 3c), with a half-life of 4.4 hours (95% confidence interval of 5.1 hours to 4.1 hours).

Transcription of the plus-strand is not independent of transcription of the minus-strand

Tax protein drives transcription of its own gene, in a positive feedback loop that is inhibited by HBZ protein[24]. We therefore examined four plus-strand-expressing clones to test for a correlation between plus-strand and *hbz* expression.

Plus-strand expression was strongly anti-correlated with the presence of *hbz*: cells were ~4 times more likely to have a plus-strand burst if they were *hbz*-negative (Figure 4b); *hbz*-negative cells also had a higher average amount of plus-strand RNA. The count of plus-strand spots was positively correlated with the fraction of cells containing a plus-strand burst (Figure 4c) in all clones studied, including clone D, although this clone differed from the other clones studied in having a lower peak plus-strand expression and a higher proportion of plus-strand mRNA retained in the nucleus.

Many of the cells with high plus-strand expression were *hbz*-negative (Figure 4a.ii). However, bursts of *hbz* expression were almost three times more frequent in plus-strand-high cells than cells with no plus-strand or low levels of plus-strand RNA (Figure 4a.iii, 4d). In fact, all of the 689 observed *hbz* bursts in low-plus-strand cells lacked a plus-strand burst (Figure 4a.iv).

This effect was observed in all three clones(A, B and C) that contained cells with high plus-strand expression. As further evidence of the reciprocal relationship between plus-strand and *hbz* expression, cells containing both a plus-strand burst and an *hbz* burst were significantly less frequent (2.9% of plus-strand-high cells) than expected by chance (7.2%); of the plus-strand-high cells, 14.6% had an *hbz* burst and 49.6% had a plus-strand burst (p < 0.0001; chi-squared). Finally, among high-plus-strand cells (N = 757), *hbz* bursts were significantly more frequent in cells without a plus-strand burst than in those with a plus-strand burst (23.2% vs. 5.8%; $p < 10^{-9}$, chi-squared), consistent with the notion that the *hbz* burst inhibits plus-strand transcription[24] and thereby terminates the plus-strand burst.

These data (Figure 4a) suggest a model for the temporal progression of HTLV-1 expression: see Discussion.

hbz is not expressed in all cells at a given time

Since *hbz* is expressed at a very low level, the unexpectedly high observed frequency of *hbz*-negative cells might be due to a failure to detect *hbz* mRNA in some cells. If the *hbz* spots are distributed randomly among the cells in a clone, then the observed average frequency of ~2 to 3 *hbz* spots/cell would result in zero spots in ~5% to 18% of cells respectively (Poisson distribution; Supplementary Table S1). The observed frequency of *hbz*-negative cells was close to this random expectation (observed/expected = 1.1 and 1.2 respectively) in Clones D and E, but *hbz*-negative cells occurred at twice the expected frequency in Clones A, B and C (observed/expected = 1.9, 2.0 and 1.9 respectively) ($p < 10^{-38}$ in each case; Supplementary Table S1). We note that the plus-strand is expressed at a high level in Clones A, B and C, but is low or absent in Clones D and E. These results show that the observed *hbz* spots were not randomly distributed among the cells, implying that some cells in each clone were not expressing *hbz* at a given instant.

Cells in S phase and G_2/M phase have elevated plus-strand and *hbz* expression

The cell-cycle stage of individual cells was identified by quantifying the integrated intensity of its DAPI-stained nucleus, which is linearly correlated with DNA content[16] (Figure 5a). The cell-cycle profiles were reproducible between replicate samples of each respective clone (Supplementary Figure 15), but there were marked differences between clones. Compared with cells in G_0/G_1, cells in G_2/M showed a greater mean intensity (spot count) of expression of both *hbz* (Figure 5b, Supplementary Figure 8) and the plus-strand (Figure 5d, Supplementary Figure 9), and a higher frequency of *hbz* bursts. These effects were less pronounced in the plus-strand: only a small proportion of cells in G_2/M were plus-strand-high (Figure 5d). However, since plus-strand-high cells contribute over 80% of total plus-strand RNA, an increase in plus-strand-high cells from 2.6% in G_0/G_1 to 5.6% in G_2/M resulted in a significant increase in total plus-strand expression. Plus-strand-silent and plus-strand-low cells were more likely to be in G_0/G_1, whereas plus-strand-high cells were more likely to be in S or G_2/M (Figure 5e), regardless of whether
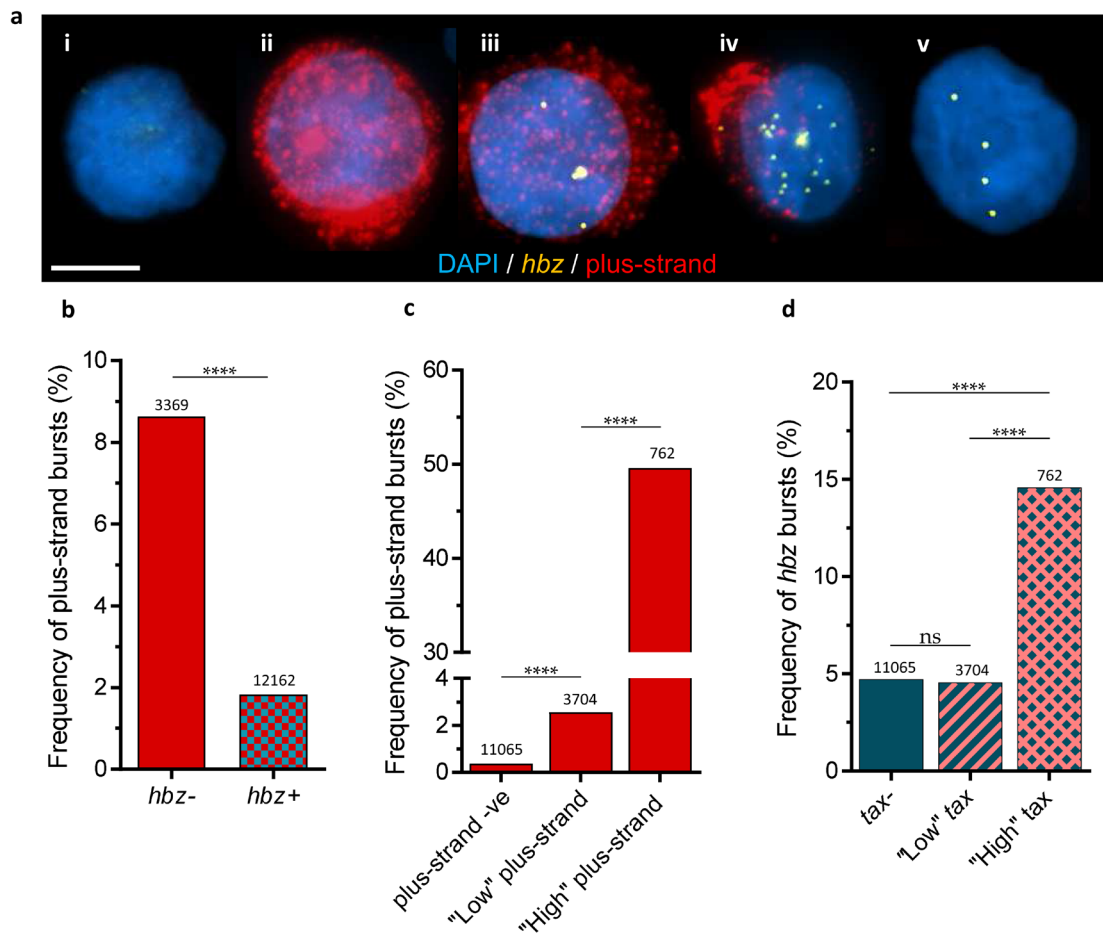
Figure 4. Relationship between expression of the plus-strand and *hbz* in individual cells. a. Representative images of a i) plus-strand-silent cell, ii) plus-strand-high cell (≥100 spots) with a plus-strand burst, iii) plus-strand-high cell without a plus-strand burst (note the presence of an *hbz* burst), iv) plus-strand-low cell (<100 spots) and v) plus-strand-silent cell, with low-level *hbz*. **b**. plus-strand bursts occur more frequently in *hbz*-negative cells. **c**. plus-strand bursts are indicative of very high plus-strand expression, with the proportion of cells which have a burst increasing as the level of plus-strand RNA increases. **d**. *hbz* bursts occur more frequently in high-plus-strand cells than in plus-strand-negative or plus-strand-low cells. Numbers above columns denote the number of cells in the corresponding population, from four pooled plus-strand-competent clones.

they contained plus-strand bursts. *hbz* expression showed a much stronger link with the cell cycle. Over half of all cells with high *hbz* (Figure 5c) were in S or G_2/M, as were one third of cells with low *hbz* and a burst.

Given that high Tax can lead to expression of *hbz*[25] the question arose whether *hbz* expression was independently correlated with the cell cycle, or correlated with plus-strand expression. Using logistic regression analysis, we found that both high plus-strand RNA and the cell cycle stage were independently correlated with *hbz* bursts (Supplementary Figure 10, Supplementary Figure 11; p < 0.0001). Cells at a given stage of the cell cycle were on average 5.8 times more likely to have an *hbz* burst if they had high plus-strand expression than if they were plus-strand-low or silent, and a cell with a given level of plus-strand RNA was on average 50% more likely to contain an *hbz* burst if it was in G_2/M rather than G_0/G_1 (Supplementary Figure 12).

Both high *hbz* expression and an *hbz* burst were independently significantly associated with G_2/M in all clones examined (p = 7.8×10^{-27} and p = 9.5×10^{-4}, respectively; logistic regression analysis, Supplementary Figure 10, Supplementary Figure 11). Similarly, high plus-strand RNA was significantly correlated with G_2/M in two of the three clones; the lack of significance in clone A may be attributable to sampling error on account of the rarity of plus-strand bursts in this clone. Although the level of plus-strand expression in Clone D was low as defined in this study, a higher plus-strand spot count in Clone D cells was still associated with S or G_2/M.

Discussion

In order for HTLV-1 to survive and propagate in the host, regulation of proviral latency is crucial. Three paradoxes regarding HTLV-1 proviral latency and the host immune response have remained unexplained. First, a strong cytotoxic T-lymphocyte
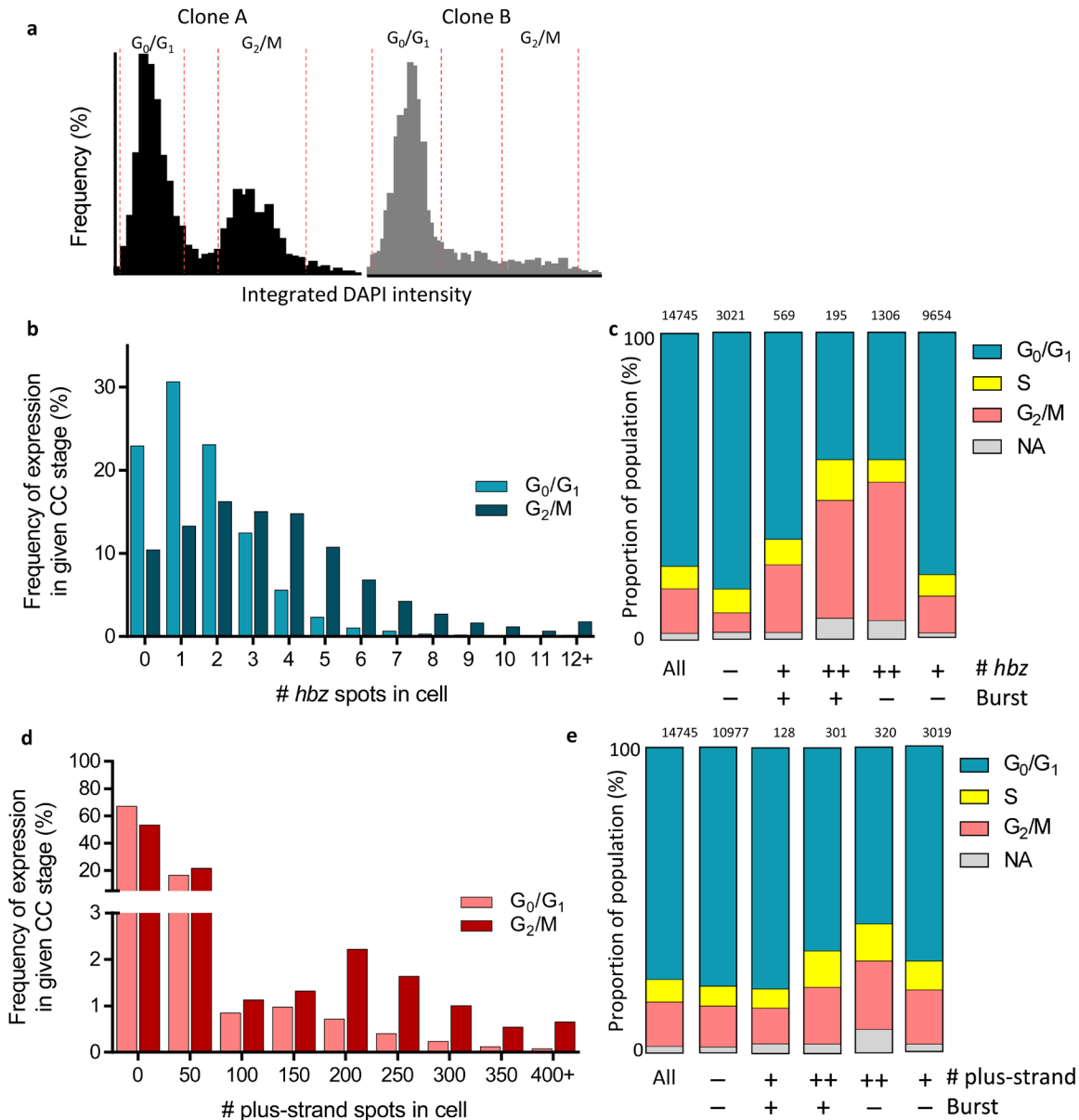
Figure 5. Expression of HTLV-1 plus-strand and *hbz* mRNAs varies with the stage of the cell cycle. a. Integrated intensity of DAPI in individual cells from replicates of two HTLV-1-infected T-cell clones. **b**. Cells in G_2/M are more frequently *hbz*+ and express higher levels of *hbz* than do cells in G_0/G_1. **c**. Consistent with this observation, cells with high levels of *hbz* mRNA and/or an *hbz* burst are more likely to be found in S or G_2/M ($p < 0.0001$, logistic regression analysis). **d**, **e**. Similarly, cells with high levels of plus-strand mRNA and/or a plus-strand burst are more likely to be found in S or G_2/M. Cells are categorized according to the level of expression of mRNAs: – (1 or 0 plus-strand spots, 0 *hbz* spots), + (2-99 plus-strand spots, 1-4 *hbz* spots) and ++ (>99 plus-strand spots, >4 *hbz* spots), as well as by the presence or absence of bursts. "NA" denotes cells whose integrated nuclear intensity was too dim or bright to fit in one of the three cell cycle bins. Number of cells with a given level of viral expression stated above each bar. The four plus-strand-competent clones were pooled for this analysis, of which three had three biological replicate samples each, and one had two biological replicates; total n = 14,745 cells.

(CTL) response to Tax protein is seen in all immunocompetent hosts, but the magnitude of this response does not explain the observed wide variation between individuals in the proviral load[1,11]. Second, the CTL response to HBZ is typically weak or undetectable, and *hbz* expression is low in polyclonal PBMCs, but the presence of a detectable anti-HBZ CTL response is associated with a lower PVL and a lower prevalence of inflammatory disease[3,10]. Third, the CTL response to Tax is chronically activated, implying frequent exposure to newly-synthesized Tax protein *in vivo*, but *tax* expression is usually undetectable in fresh PBMCs[26].

The results reported here offer an explanation for each of these paradoxes. By quantifying the frequency and intensity of *hbz* and plus-strand expression, we show that in clones of naturally-infected, CD4+ T-cells, both the plus-strand and minus-strand genes of HTLV-1 are expressed in bursts, which differ strongly in intensity and frequency between the two strands (Figure 2). The sequence used here to detect the plus-strand of the provirus is present in all the viral plus-strand mRNAs. However, it is established that *tax* mRNA is the first and most abundant species[26]: while some of the plus-strand signal detected in this study will represent the other plus-strand mRNA, species, the majority will be *tax* mRNA, especially in the early stages of the transcriptional burst.

If this pattern of HTLV-1 gene expression accurately represents the pattern of expression *in vivo*, the paradoxes can be explained as follows. At the single-cell level, expression of the plus strand is not uniform, but rather exhibits a bimodal expression profile (Figure 1d). The minority of cells expressing plus-strand RNA at any one time express the transcripts at a very high level (Figure 4c). The plus-strand transcripts are then rapidly exported to the cytosol (Figure 3b). The resulting intermittent but intense expression of the highly immunogenic Tax protein is sufficient to account for the observed preponderance of anti-Tax CTLs[27]. However, since the proportion of cells expressing Tax at any instant is low (of the order of 1% to 10%), the anti-Tax CTL response has a limited impact on the proviral load set-point.

In contrast, the proviral minus strand (*hbz*) is expressed at a much lower level than the proviral plus strand, consistent with previous observations[28,29], and much more uniformly across cells (Figure 1c). However, *hbz* mRNA is less likely to be exported from the nucleus (Figure 3b), again in agreement with earlier studies[26]. This low expression combined with low translocation explains why HBZ protein levels are very low in physiological conditions. We have previously suggested[26] that HBZ protein expression is kept low in order to minimize exposure of the virus to the immune response; consistent with this hypothesis, HBZ protein is a weak T-cell immunogen[30]. HBZ protein is frequently undetectable in fresh PBMCs, but Baratella *et al*.[31] reported the detection of HBZ protein in the cytoplasm of PBMCs isolated from patients with the inflammatory disease HAM/TSP. Our estimate of *hbz* half-life is higher than that made in previous studies[32,33], but is set apart by using naturally-infected cells and absolute quantification.

Earlier investigations have pointed to an antagonistic relationship between Tax and HBZ: Tax can drive HBZ expression, which can then compete with Tax for the host factors which are necessary for expression driven from the promoter in the 5'-LTR[24,25]. These previous observations, in combination with our observations of the correlation between levels of plus-strand RNA and *hbz* within individual cells (Figure 4), suggest the following model of the interaction between plus-strand expression and *hbz*. The plus-strand is more likely to be expressed in an *hbz*-silent cell (Figure 4a.i), and once expressed the plus-strand typically reaches very high levels (Figure 4.ii). A high level of plus-strand RNA, however, increases the probability of *hbz* expression[25] (Figure 4iii), which in turn leads to cessation of plus-strand expression. Subsequently, both plus-strand and *hbz* transcripts decay, and the provirus remains latent until another factor upregulates expression, or the cell becomes silent, restarting the potential cycle. The factors that regulate the onset of plus-strand expression include glucose metabolism and the available molecular oxygen[34]; cellular activation and stress may also contribute.

It has long been held that *hbz* is constitutively expressed in infected cells. However, all previous studies quantified HBZ expression in cell populations, using bulk-averaged techniques such as qPCR or antibody titres. Our single-molecule and single-cell approach indicates that *hbz* is not expressed constantly in all cells, even within a single clone. If the pattern of proviral expression we observed is also found *in vivo*, a temporary lapse in *hbz* expression could allow a plus-strand burst to develop, before plus-strand expression declines either under the influence of *hbz*, or immune detection and destruction. The pattern of frequent, low-level expression of *hbz* is consistent with the notion that the primary function of HBZ – at both protein and mRNA levels – is to maintain clonal persistence[5,6,28,35,36]. Regardless of the cause, limiting expression of the highly immunogenic Tax protein to intermittent bursts allows the virus to optimise the protein's effects in manipulating the host cell, to drive viral replication and clonal proliferation, while minimising its exposure to the intense anti-Tax immune response.

The cause of the observed cell-to-cell heterogeneity in *hbz* expression is unknown. It is also unclear precisely how the relationship between Tax and HBZ is mediated. The previously described mechanisms of interaction between Tax and HBZ act at the protein level[37,38]. However, given the very low proportion of *hbz* transcripts exported from the nucleus in any given cell and the low abundance of HBZ protein found in naturally-infected cells, it is possible that the *hbz* RNA plays a major part in this relationship. There was a very strong linear correlation between *hbz* spot count and nuclear volume (Figure 3d). Certain cellular mRNAs show a similar dependence on nuclear volume or cell volume[39]; the mechanisms responsible are not yet known. However, the frequency of *hbz* bursts was independent of nuclear volume, but was strongly associated with $S/G_2/M$. This burst timing contrasts with the transcription rate of many cellular genes, which is reduced during $S/G_2/M$, probably to compensate for the doubling of the gene copy number[40].

The occurrence of an *hbz* burst in S/G$_2$/M might confer two possible advantages on the virus. First, *hbz* promotes progression through the cell cycle[41,42]. Second, the burst will increase the abundance of both HBZ mRNA and protein molecules, which are normally present at limiting frequency, and thereby ensure efficient partitioning of the molecules between the two daughter cells[43].

The observed longevity of HTLV-1-infected T-cell clones *in vivo* indicates that sustained cell proliferation plays a significant part in viral persistence[13]. The relationship between the cell cycle and expression of the plus-strand and *hbz* is therefore of critical importance. We found that cells in G$_2$/M had on average a significantly higher amount of plus-strand RNA, and the minority of cells with high plus-strand RNA were significantly more frequent in S or G$_2$/M than in G$_0$/G$_1$ (Figure 5). In contrast, plus-strand bursts were not significantly positively associated with S or G$_2$/M phase (Supplementary Figure 10), although the power of this test was limited by the low frequency and transient nature of plus-strand bursts. The strength of this relationship between plus-strand expression and the cell cycle varied between clones; in contrast, the relationship between *hbz* and the cell cycle was remarkably strong and consistent between clones. The logistic regression analysis indicated that each additional *hbz* transcript detected increased the odds of a cell being in G$_2$/M (rather than G$_0$/G$_1$) by an average of 1.45-fold (7.3-fold in *hbz*-high cells), whereas each additional plus-strand transcript increased the odds by 1.004-fold (1.58-fold in plus-strand-high cells). This correlation with *hbz* was maintained in the clone incapable of expressing the plus-strand, showing that the relationship between *hbz* and the cell cycle is independent of plus-strand expression. Although several mechanisms have been described by which Tax protein and both HBZ protein and RNA promote cell proliferation[35,36,41], it remains possible that cell division itself enhances proviral expression: that is, proviral expression may be both cause and consequence of cell division. Since HTLV-1 persists in the host by driving proliferation and avoiding the immune response, it is a logical strategy to align its proviral expression with the host cell cycle, particularly for such an immunogenic product as Tax. We have summarised our current interpretation of the regulation of HTLV-1 transcription and replication in a model (Figure 6)[44–46].

Previous evidence suggested that *hbz* is constitutively expressed in HTLV-1-infected T-cells, albeit at a low level[28,35]. However, at a given time, *hbz* was not expressed in all cells in any of the clones studied here. The observed frequency of hbz-negative cells is inconsistent with a random (Poisson) distribution of *hbz* spots (Supplementary Table S1); the departure from the Poisson distribution was consistently greater in the clones that expressed high levels of plus-strand RNA. We infer that *hbz* is not constitutively expressed in all cells. Either there is a fraction of cells within each clone that do not express *hbz* or, more likely, each cell passes through an *hbz*-negative phase, perhaps following a plus-strand burst. It is now important to identify the factors that trigger re-expression of *hbz*.

In summary, we show that plus-strand expression from the HTLV-1 provirus varies widely between individual cells, even within the same clone, and that minus-strand expression, while much more homogeneous than the plus-strand, is not present in all cells at all times. The contrast between the expression of the two strands is further demonstrated by difference between the frequency and intensity of their respective transcriptional bursts. Once expressed, the majority of plus-strand and *hbz* transcripts have different fates, with the majority of *hbz* retained in the
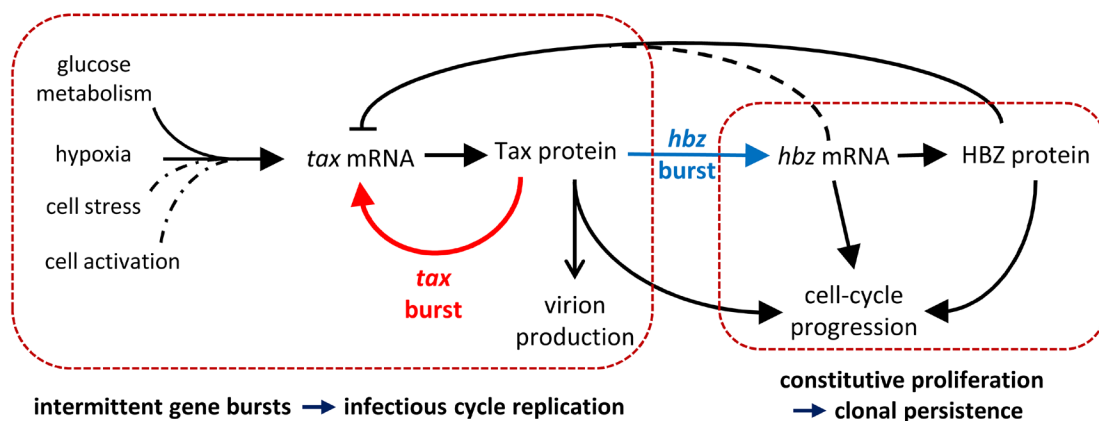
Figure 6. Model of regulation of HTLV-1 transcription and replication *in vivo*. Constitutive expression of *hbz* mRNA maintains clonal longevity by promoting cellular proliferation. Intermittent bursts of plus-strand expression, driven by bursts of *tax*, have three chief consequences: production of HTLV-1 virions and resulting infectious spread; accelerated cell-cycle progression; and stimulation of *hbz,* of which the RNA and protein both have proliferative effects. HBZ protein can also inhibit *tax* transcription, terminating the plus-strand burst. It is not known whether *hbz* mRNA also counteracts plus-strand transcription, but it is suggested by our results. The factors that regulate the frequency of plus-strand bursts include glucose metabolism and hypoxia; cell stress and cell activation may also trigger plus-strand expression. Aside from Tax, other triggers for *hbz* expression remain to be discovered. Dashed line represents a link indicated by results in the present study; dotted lines represent hypothetical links.

nucleus while plus-strand RNA is exported to the cytosol. Our results also show that plus- and minus-strand expression do not occur independently of each other, but rather that, at the single-cell level, plus-strand expression is more likely to occur in the absence of *hbz*, whose expression is in turn more likely in cells with high plus-strand expression. Finally, proviral expression is correlated with the phase of the cell cycle, with clear implications for driving proliferation and evasion of immunosurveillance. It remains to be tested how closely the transcriptional behaviour of HTLV-1 observed in these naturally-infected T-cell clones represents the transcriptional behaviour of the virus *in vivo*, both qualitatively and quantitatively. We are now applying the techniques described here to quantify HTLV-1 plus- and minus-strand transcription in PBMCs both directly *ex vivo* and after short-term *in vitro* incubation. The ability to quantify these relationships at the single-cell level opens exciting avenues to elucidate the phenomenon of HTLV-1 latency.

Ethical statement

HTLV-1-infected T-cell clones were derived from peripheral blood samples given by donors attending the National Centre for Human Retrovirology (NCHR) at Imperial College Healthcare NHS Trust, St Mary's Hospital, London. 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).

Data availability

Sample image data are available at: https://osf.io/b9mnd/. Further data are available on request from the authors; the large dataset (~2 Tb) is most reliably transferred on a hard disk.

Supplementary material

Supplementary File 1. Contains Supplementary Figures 1 to 15.

Click here to access the data.

Supplementary Table S1. Analysis of frequency of *hbz*-negative cells in each of five HTLV-1-infected clones, A to E.

Click here to access the data.

References

1. Bangham CRM: **Human T Cell Leukemia Virus Type 1: Persistence and Pathogenesis.** *Annu Rev Immunol.* 2018; **36**: 25–53.
   PubMed Abstract | **Publisher Full Text**

2. Hanon E, Hall S, Taylor GP, *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): 1386–1392.
   PubMed Abstract

3. Hilburn S, Rowan A, Demontis MA, *et al.*: **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.* 2011; **203**(4): 529–536.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Gaudray G, Gachon F, Basbous J, *et al.*: **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.* 2002; **76**(24): 12813–12822.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Boxus M, Willems L: **Mechanisms of HTLV-1 persistence and transformation.** *Br J Cancer.* 2009; **101**(9): 1497–1501.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Ma G, Yasunaga J, Matsuoka M: **Multifaceted functions and roles of HBZ in HTLV-1 pathogenesis.** *Retrovirology.* 2016; **13**: 16.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Satou Y, Yasunaga J, Zhao T, *et al.*: **HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo.** *PLoS Pathog.* 2011; **7**(2): e1001274.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Shiohama Y, Naito T, Matsuzaki T, *et al.*: **Absolute quantification of HTLV-1 basic leucine zipper factor (HBZ) protein and its plasma antibody in HTLV-1 infected individuals with different clinical status.** *Retrovirology.* 2016; **13**: 29.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Li M, Kesic M, Yin H, *et al.*: **Kinetic analysis of human T-cell leukemia virus type 1 gene expression in cell culture and infected animals.** *J Virol.* 2009; **83**(8): 3788–3797.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. MacNamara A, Rowan A, Hilburn S, et al.: **HLA class I binding of HBZ determines outcome in HTLV-1 infection.** *PLoS Pathog.* 2010; **6**(9): e1001117.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Kattan T, MacNamara A, Rowan AG, et al.: **The avidity and lytic efficiency of the CTL response to HTLV-1.** *J Immunol.* 2009; **182**(9): 5723–5729.
PubMed Abstract | **Publisher Full Text**

12. Cook LB, Rowan AG, Melamed A, et al.: **HTLV-1-infected T cells contain a single integrated provirus in natural infection.** *Blood.* 2012; **120**(17): 3488–3490.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Gillet NA, Malani N, Melamed A, et al.: **The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones.** *Blood.* 2011; **117**(11): 3113–3122.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Fan J, Ma G, Nosaka K, et al.: **APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes *in vivo*.** *J Virol.* 2010; **84**(14): 7278–7287.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Kamentsky L, Jones TR, Fraser A, et al.: **Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software.** *Bioinformatics.* 2011; **27**(8): 1179–1180.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Roukos V, Pegoraro G, Voss TC, et al.: **Cell cycle staging of individual cells by fluorescence microscopy.** *Nat Protoc.* 2015; **10**(2): 334–348.
PubMed Abstract | **Publisher Full Text**

17. Mueller F, Senecal A, Tantale K, et al.: **FISH-quant: automatic counting of transcripts in 3D FISH images.** *Nat Methods.* 2013; **10**(4): 277–278.
PubMed Abstract | **Publisher Full Text**

18. Tsanov N, Samacoits A, Chouaib R, et al.: **smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability.** *Nuc Acids Res.* 2016; **44**(22): e165.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. R Core Team: **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. 2017.

20. Takeda S, Maeda M, Morikawa S, et al.: **Genetic and epigenetic inactivation of *tax* gene in adult T-cell leukemia cells.** *Int J Cancer.* 2004; **109**(4): 559–567.
PubMed Abstract | **Publisher Full Text**

21. Raj A, van den Bogaard P, Rifkin SA, et al.: **Imaging individual mRNA molecules using multiple singly labeled probes.** *Nat Methods.* 2008; **5**(10): 877–879.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Raj A, Peskin CS, Tranchina D, et al.: **Stochastic mRNA synthesis in mammalian cells.** *PLoS Biol.* 2006; **4**(10): e309.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Tantale K, Mueller F, Kozulic-Pirher A, et al.: **A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting.** *Nat Commun.* 2016; **7**: 12248.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Lemasson I, Lewis MR, Polakowski N, et al.: **Human T-cell leukemia virus type 1 (HTLV-1) bZIP protein interacts with the cellular transcription factor CREB to inhibit HTLV-1 transcription.** *J Virol.* 2007; **81**(4): 1543–1553.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

25. Yoshida M, Satou Y, Yasunaga J, et al.: **Transcriptional control of spliced and unspliced human T-cell leukemia virus type 1 bZIP factor (HBZ) gene.** *J Virol.* 2008; **82**(19): 9359–9368.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Rende F, Cavallari I, Corradin A, et al.: **Kinetics and intracellular compartmentalization of HTLV-1 gene expression: nuclear retention of *HBZ* mRNAs.** *Blood.* 2011; **117**(18): 4855–4859.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Goon PK, Biancardi A, Fast N, et al.: **Human T cell lymphotropic virus (HTLV) type-1-specific CD8+ T cells: frequency and immunodominance hierarchy.** *J Infect Dis.* 2004; **189**(12): 2294–2298.
PubMed Abstract | **Publisher Full Text**

28. Saito M, Matsuzaki T, Satou Y, et al.: ***In vivo* expression of the HBZ gene of HTLV-1 correlates with proviral load, inflammatory markers and disease severity in HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM/TSP).** *Retrovirol.* 2009; **6**: 19.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Satou Y, Yasunaga J, Yoshida M, et al.: ***HTLV-1 basic leucine zipper factor* gene

mRNA supports proliferation of adult T cell leukemia cells.** *Proc Natl Acad Sci U S A.* 2006; **103**(3): 720–725.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. Rowan AG, Suemori K, Fujiwara H, 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.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

31. Baratella M, Forlani G, Raval GU, et al.: **Cytoplasmic Localization of HTLV-1 HBZ Protein: A Biomarker of HTLV-1-Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP).** *PLoS Negl Trop Dis.* 2017; **11**(1): e0005285.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

32. Cavallari I, Rende F, Bona MK, et al.: **Expression of Alternatively Spliced Human T-Cell Leukemia Virus Type 1 mRNAs Is Influenced by Mitosis and by a Novel *cis*-Acting Regulatory Sequence.** *J Virol.* 2015; **90**(3): 1486–1498.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

33. Mocquet V, Neusiedler J, Rende F, et al.: **The human T-lymphotropic virus type 1 tax protein inhibits nonsense-mediated mRNA decay by interacting with INT6/EIF3E and UPF1.** *J Virol.* 2012; **86**(14): 7530–7543.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

34. Kulkarni A, Mateus M, Thinnes CC, et al.: **Glucose Metabolism and Oxygen Availability Govern Reactivation of the Latent Human Retrovirus HTLV-1.** *Cell Chem Biol.* in press, 2017; **24**(11): 1377–1387.e3.
PubMed Abstract | **Publisher Full Text**

35. Satou Y, Miyazato P, Ishihara K, et al.: **The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome.** *Proc Natl Acad Sci U S A.* 2016; **113**(11): 3054–3059.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. Matsuoka M, Yasunaga JI: **Human T-cell leukemia virus type 1: replication, proliferation and propagation by Tax and HTLV-1 bZIP factor.** *Curr Opin Virol.* 2013; **3**(6): 684–691.
PubMed Abstract | **Publisher Full Text**

37. Basbous J, Arpin C, Gaudray G, et al.: **The HBZ factor of human T-cell leukemia virus type I dimerizes with transcription factors JunB and c-Jun and modulates their transcriptional activity.** *J Biol Chem.* 2003; **278**(44): 43620–43627.
PubMed Abstract | **Publisher Full Text**

38. Clerc I, Polakowski N, André-Arpin C, et al.: **An interaction between the human T cell leukemia virus type 1 basic leucine zipper factor (HBZ) and the KIX domain of p300/CBP contributes to the down-regulation of tax-dependent viral transcription by HBZ.** *J Biol Chem.* 2008; **283**(35): 23903–23913.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

39. Padovan-Merhar O, Nair GP, Biaesch AG, et al.: **Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms.** *Mol Cell.* 2015; **58**(2): 339–352.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Skinner SO, Xu H, Nagarkar-Jaiswal S, et al.: **Single-cell analysis of transcription kinetics across the cell cycle.** *eLife.* 2016; **5**: e12175.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Mitobe Y, Yasunaga J, Furuta R, et al.: **HTLV-1 bZIP Factor RNA and Protein Impart Distinct Functions on T-cell Proliferation and Survival.** *Cancer Res.* 2015; **75**(19): 4143–4152.
PubMed Abstract | **Publisher Full Text**

42. Satou Y, Yasunaga J, Yoshida M, et al.: ***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): 720–725.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

43. Soltani M, Singh A: **Effects of cell-cycle-dependent expression on random fluctuations in protein levels.** *R Soc Open Sci.* 2016; **3**(12): 160578.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

44. Matsuoka M, Jeang KT: **Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy.** *Oncogene.* 2011; **30**(12): 1379–1389.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

45. Overbaugh J, Bangham CR: **Selection forces and constraints on retroviral sequence variation.** *Science.* 2001; **292**(5519): 1106–1109.
PubMed Abstract | **Publisher Full Text**

46. Yoshida M: **Multiple viral strategies of HTLV-1 for dysregulation of cell growth control.** *Ann Rev Imm.* 2001; **19**: 475–496.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Referee Status: ✔ ✔ ?

? **Marc Sitbon** (iD)

UMR 5535, CNRS, Institute of Molecular Genetics of Montpellier, Montpellier, France

Billman *et al.* report here a detailed analysis of plus-strand (+strand) "*tax*" and anti-sense (*hbz*) RNA transcripts in HTLV-1 T cell clones that have been cultured *ex vivo* from CD4+CD25+ T cells isolated from PBMCs of 4 different patients. Each clone, harboring a single entire copy of the HTLV-1 genome, was cultured under periodic activation with IL-2, and anti CD2, CD3 and CD28 mAbs. All analyses were performed at day 8 of the 14 day activation cycle, when most cells are returning to a resting state. Single cell analyses of HTLV-1 RNA expression were based on integration of single molecule RNA-FISH and DAPI nuclear staining with occasional actinomycin D treatment.

Under these conditions, the authors observed that the vast majority of cells harbored detectable levels of *hbz* transcripts, while accumulation of +strand transcripts (which they refer to as "*tax*", see below) was more heterogeneous, varying from <10% in cells from 2 of the clones, to 50-60% of cells from the other 2 clones. An additional clone that harbored a deleted provirus, and thereby did not produce +strand transcripts, served as control. The authors further showed that both sense and anti-sense transcripts could be detected within larger nuclear spots in 4 to 7% of the cells analyzed, which the authors interpreted as transcriptional bursts; this interpretation was confirmed by the significant drop of bursts observed after treatment with actinomycin D. Act D further allowed the authors to show that +strand "*tax*" transcripts were rapidly accumulating to the cytoplasm while *hbz* transcripts remained in the nucleus. According to the authors, and in agreement with former observations by several teams, *hbz* and *tax* transcripts were in general inversely related. Also, the authors report that both types of transcripts tended to accumulate in cells in G2/M as compared to G0/G1 (monitored as a function of DAPI signal).

Elucidation of the *in vivo* dynamics of Hbz versus Tax production at the single cell level should indeed help elucidate the *in vivo* regulation of latent versus productive HTLV-1 infection, and provide precious clues on the pathogenic outcomes following infection. Despite an impressive amount of data cross-analyses based on +strand and anti-sense transcription of T cell clones derived from patients' PBMCs, the actual relevance of the model for *in vivo* situations is questionable. Thus, all analyses were performed on 4 clones isolated from a subpopulation of CD4+CD25+ T cells (as described in a reference and not mentioned here), which after being isolated from patients PBMCs, were cultured for a long period of time in conditions that do not reproduce *in vivo* conditions (hyperoxia, hyperglycemia, strong recurrent activations and absence of immune pressure). As such, several of the authors' conclusions (in the title, abstract and discussion) are not warranted (see below).

As presented, the study brings up other several additional points:

1. The authors state that all HTLV-1 +strand transcripts are detected with the so-called "*tax*" probe. Nevertheless, the constant designation of the transcripts as "tax" throughout the article can be very misleading. It would be more accurate to refer to the results obtained with this probe in the text and figures as "+strand", or another more inclusive designation, rather than "*tax*" solely.

2. Nearly 60 to 90% of cells in all 5 clones express *hbz* antisense transcripts, but depending on the clone studied, the presence of +strand transcripts, as detected with the "*tax"* probe, are much more heterogeneous (<10% in clones A and D, and 50-60% in clones B & C). Given this overwhelming heterogeneity, how can the authors exclude that it is not due to variation in Tax-encoding transcripts, but solely to other viral +strand transcripts. In other words, the authors cannot exclude that levels of Tax-encoding transcripts were homogeneously low in all clones.

3. Although the authors insist on the pitfalls of qPCR to analyze single cell transcription, there have been several reports on microfluidics technology being successfully used to monitor single cell RNA expression (i.e. Moussy *et al*. 2017. PLoS Biol 15(7):e2001867). The application of microfluidics on PBMCs directly isolated from patients would likely allow a more accurate assessment of the *in vivo* status of RNA transcripts.

4. As mentioned above, given the experimental model of subcloning and culture (at the non-physiological 20% oxygen tension, which artificially modulates HTLV-1 activation, as previously published by the authors), and in the absence of immune pressure, the claim that the results presented here are relevant for *in vivo* situations and "*offer an explanation for the paradoxical correlations observed between the host immune response and HTLV-1 transcription*" is not warranted.

5. Since the correlation between "transcription bursts" and the G2/M cell cycle stage is not established for +strand "tax" transcripts, and pertains to <7% of cells in which *hbz* is detected, the claim of a relationship between viral gene bursts and cell cycle (as stated in the general conclusion and the title) does not appear justified. Therefore, the intricate array of cross analyses and statistical evaluation performed here and addressed lengthily in the discussion is not likely to reflect *in vivo* situations.

6. Statistical significance of the correlation between high *hbz* expression, *hbz* bursts and the G2/M stages should be directly mentioned in Fig. 5 and the corresponding text. The clones used to obtain the data provided in panel 5b are not indicated (it appears that clones A and B were used here, whereas data in 5c are mentioned in the figure legend to result from the average of clones A, B, C and D).

7. Given the reservations mentioned above (an *in vitro* system that does not recapitulate *in vivo* conditions, the lack of distinction between *tax* versus +strand transcripts, and the lack of data on Tat and HBZ proteins), the discussion remains too speculative.

8. In this study, there is no assessment of the relationship between transcript and protein levels; moreover, since the half-lives of Tax and Hbz proteins in this model are not established, it is not possible to evaluate the actual timing of a potential cross-regulation between the 2 proteins. In this regard, the authors should discuss a recent paper by Baratella *et al.* (PLos Negl Trop Dis Jan 2017) that describes the specific accumulation of the HBZ protein in the cytoplasm of PBMCs directly isolated from TSP/HAM patients.

9. In suppl. fig 7 - there is an SEM indicated in the histogram of clone C while there was only 2 replicates, according to the figure legend. Also, in suppl. Figs 8 and 9 there appears to be a color and legend inversion for clones B, C, and D, and clone A does not seem to fit the G2/M correlation with high levels of +strand "tax" transcripts.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Retrovirology, hematology, cancer

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

✓ **Ali Bazarbachi** , **Hiba El Hajj**
Department of Internal Medicine, American University of Beirut, Beirut, Lebanon

This is a well-written report from leading investigators in the field of HTLV-I replication. In this study, the authors screened more than 19 000 cells derived from five naturally infected T cell clones isolated from PBMCs of HTLV-1 healthy donors. They used single-molecule RNA FISH to quantify, at the single cell level, the transcripts of two main products of the HTLV-I pX region, Tax encoded by the plus-strand of the provirus and HBZ encoded by its minus-strand. They observed a strong intra- and inter-clonal heterogeneity in the expression of *tax* and *hbz* genes with Tax being expressed at high levels in few cells whereas *hbz* is expressed at lower levels in most cells. They also report that both genes are transcribed in intermittent bursts and that tax expression is enhanced in the absence of HBZ but that *hbz* expression is enhanced by Tax. Finally, they show that HBZ expression is mostly associated with G2/M phases of the cell cycle, and that its abundance correlates with the nuclear volume. They conclude that the main function of this frequent low level of *hbz* transcript and protein is to maintain clonal persistence and that

conversely, intense transient intermittent bursts of *tax* drive viral replication and clonal proliferation while escaping the robust anti Tax CTL response. These results strongly suggest that the plus strand of HTLV-I is expressed in bursts *in vivo* explaining the hitherto unexplained paradoxical observations of a strong anti-Tax immune response in all immuno-competent HTLV-I infected individuals while Tax is undetectable in their freshly isolated PBMCs. This is an important study in the field that defeats multiple previous paradigms. The following experiments should strengthen the manuscript:

1. Authors used short-term cultures of activated cells in order to isolate individual cells. However, they need to show *tax* and *hbz* expression at the single cell level in freshly isolated PBMCs since cell culture and activation are likely to affect *tax* and *hbz* expression.

2. Figure 1: It is surprising that 1% of uninfected clones (negative control) express 50 *tax* transcripts and only 1 *hbz* transcript. More uninfected clones should be tested. Furthermore, in order to validate *tax* and *hbz* probes, they show the specificity/sensitivity of smFISH on one clone generated from HTLV-I infected cell lines like having different expression levels of Tax and HBZ.

3. It is important to show the viral load and the status of expression of Tax and HBZ in total PBMCs of the 5 HTLV-I positive donors, and compare it with that of the 5 T cell clones that were selected. How can the authors be sure that the selected clone from each donor is a representative one of all the HTLV-1 infected cells in that individual? Did they screen other clones from each of the five donors?

4. Figure 2: Images of *tax* bursts are more convincing that those of *hbz*. Figure 2D shows that the estimated number of *hbz* RNA per burst is around 3 to 4 with little intercellular variation, which is not classical for bursts. How can the authors rule out the possibility that these represent superposition of multiple transcripts. In 2C, they show a three-fold reduction in the frequency of *hbz* bursts in the presence of actinomycine D compared to a total disappearance of *tax* bursts. What is the percentage reduction of *hbz* bursts relative to the reduction of total *hbz* transcripts?

5. Figure 3: Estimation of intranuclear mRNA should be done in zStacks to rule out counting transcripts above and below the nucleus.

6. Figure 4: Authors show increased frequency of *hbz* bursts in high tax positive cells. Again, what about the total amount of *hbz* transcripts in these cells?

7. Elevated *tax* and *hbz* expression in S and G2/M phase of the cell cycle is solely based on the integrated intensity of the DAPI stained nucleus (2n versus 4n). It would be nice to perform synchronization of cells to confirm these findings.

8. Supplementary Figure 9: The negative control (uninfected cells) and clone E should be shown.

9. Supplementary Figure 10: The odds ratio for G2/M phase is much higher for cells with >4 *hbz* compared to cells with >99 tax whereas that for S phase is quite similar. Does this indicate that *hbz* expression results in G2/M arrest? Is there any correlation between G2/M phase and nuclear size in these cells?

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 12 October 2017

✔ **Vincenzo Ciminale**
Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy

In the present study Billman *et al.* used single-molecule RNA-FISH to quantify the expression of HTLV-1 transcripts at the single-cell level. This analysis was carried out in five T-cell clones isolated by limiting dilution from peripheral blood of HTLV-1-infected subjects. Results revealed strong heterogeneity in plus- and minus- strand expression of the viral genome both within and between clones, suggesting that viral genes are transcribed in transcriptional bursts. Interestingly, *hbz* expression was associated with the S and G /M phases of the cell cycle, and was correlated with nuclear volume.

In general, this is an excellent study that applies state-of the-art techniques to tackle some key unanswered questions on the regulation of HTLV-1 gene expression and their possible connection with the viral persistence strategies and pathogenic properties *in vivo*.

The study is very thorough and the data presented are solid and convincing. I have a few suggestions that the Authors might want to consider to improve the quality of this paper.

1. It would be useful to indicate of the location of the probes on the scheme of the genome shown in Fig. 1A.

2. Most of the Tax probes will hybridize to ALL plus-strand transcripts, including the highly expressed *gag*/*pol*, *env* and p21 mRNAs. Although the authors clearly state that "tax is used as a proxy for plus-strand proviral expression" referring to *Tax* may be misleading and is an oversimplification that does not fit well such a careful and thorough study. Indicating this signal as "plus-strand" throughout the paper would be more appropriate.

3. Immunofluorescence images should also show the individual channels, not just the overlay.

4. Although the association between a single mRNA molecule and a spot in the FISH analysis was shown in the original paper describing this methodology, it would have been helpful to see a comparison (in at least part of the samples) between the copy number measured by FISH and the copy number measured by absolute quantification in real-time RT-PCR.

5. The presented data provide very strong evidence of the heterogeneity of plus- and minus- strand transcription heterogeneity even within a clonal population with the same proviral integration site. This finding suggests (but does not prove) that viral expression, especially from the plus-strand might follow a pulsatile pattern, which would explain the apparent paradox of a strong immune response and low viral expression in vivo. Time-lapse experiments on live cells expressing a fluorescent reporter of Tax activity (e.g. LTR-unstableGFP) would provide formal proof for this key point.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 30 Nov 2017

Charles Bangham, Department of Immunology, Imperial College London, UK

Billman et al 2017: response to reviewers' comments
Reviewer 1 (Ciminale)
In the present study Billman et al. used single-molecule RNA-FISH to quantify the expression of HTLV-1 transcripts at the single-cell level. This analysis was carried out in five T-cell clones isolated by limiting dilution from peripheral blood of HTLV-1-infected subjects. Results revealed strong heterogeneity in plus- and minus- strand expression of the viral genome both within and between clones, suggesting that viral genes are transcribed in transcriptional bursts. Interestingly, hbz expression was associated with the S and G /M phases of the cell cycle, and was correlated

with nuclear volume.

In general, this is an excellent study that applies state-of the-art techniques to tackle some key unanswered questions on the regulation of HTLV-1 gene expression and their possible connection with the viral persistence strategies and pathogenic properties in vivo.

The study is very thorough and the data presented are solid and convincing. I have a few suggestions that the Authors might want to consider to improve the quality of this paper.

1. It would be useful to indicate of the location of the probes on the scheme of the genome shown in Fig. 1A.

Thank you for the suggestion - the probe locations have been added to Figure 1A.

2. Most of the Tax probes will hybridize to ALL plus-strand transcripts, including the highly expressed gag/pol, env and p21 mRNAs. Although the authors clearly state that "tax is used as a proxy for plus-strand proviral expression" referring to Tax may be misleading and is an oversimplification that does not fit well such a careful and thorough study. Indicating this signal as "plus-strand" throughout the paper would be more appropriate.

At the beginning of the Results section we made this point clear by stating 'In this paper, *tax* is used as a proxy for plus-strand proviral expression: the *tax* target sequence is present in all plus-strand transcripts.' However, we accept that the reader may overlook this important point when reading the rest of the paper, and accordingly we have systematically replaced reference '*tax'* expression by 'plus-strand' expression in the revised manuscript.

3.   Immunofluorescence images should also show the individual channels, not just the overlay.

We show only the overlay in the main paper for clarity and brevity; a sample of the data is available online, and the full dataset (~2 Tb) is available from the authors on request.

4.   Although the association between a single mRNA molecule and a spot in the FISH analysis was shown in the original paper describing this methodology, it would have been helpful to see a comparison (in at least part of the samples) between the copy number measured by FISH and the copy number measured by absolute quantification in real-time RT-PCR.

We agree that it might be helpful to examine formally the correlation between mRNA levels detected by the two techniques (smFISH and qRT-PCR). However, the results in the present study are corroborated by positive and negative controls, and the addition of qRT-PCR results is unlikely to extend or alter the conclusions reached here.

5. The presented data provide very strong evidence of the heterogeneity of plus- and minus-strand transcription heterogeneity even within a clonal population with the same proviral integration site. This finding suggests (but does not prove) that viral expression, especially from the plus-strand might follow a pulsatile pattern, which would explain the apparent paradox of a strong immune response and low viral expression in vivo. Time-lapse experiments on live cells expressing a fluorescent reporter of Tax activity (e.g. LTR-unstableGFP) would provide formal proof for this key point.

Thank you for this interesting suggestion. We are in fact already pursuing this approach.

Reviewer 2 (Bazarbachi/El Hajj)
This is a well-written report from leading investigators in the field of HTLV-I replication. In this study, the authors screened more than 19 000 cells derived from five naturally infected T cell clones isolated from PBMCs of HTLV-1 healthy donors. They used single-molecule RNA FISH to quantify, at the single cell level, the transcripts of two main products of the HTLV-I pX region, Tax encoded by the plus-strand of the provirus and HBZ encoded by its minus-strand. They observed a strong intra- and inter-clonal heterogeneity in the expression of tax and hbz genes with Tax being expressed at high levels in few cells whereas hbz is expressed at lower levels in most cells. They also report that both genes are transcribed in intermittent bursts and that tax expression is enhanced in the absence of HBZ but that hbz expression is enhanced by Tax. Finally, they show that HBZ expression is mostly associated with G2/M phases of the cell cycle, and that its abundance correlates with the nuclear volume. They conclude that the main function of this frequent low level of hbz transcript and protein is to maintain clonal persistence and that conversely, intense transient intermittent bursts of tax drive viral replication and clonal proliferation while escaping the robust anti Tax CTL response. These results strongly suggest that the plus strand of HTLV-I is expressed in bursts in vivo explaining the hitherto unexplained paradoxical observations of a strong anti-Tax immune response in all immuno-competent HTLV-I infected individuals while Tax is undetectable in their freshly isolated PBMCs. This is an important study in the field that defeats multiple previous paradigms. The following experiments should strengthen the manuscript:

1.   Authors used short-term cultures of activated cells in order to isolate individual cells. However, they need to show tax and hbz expression at the single cell level in freshly isolated PBMCs since cell culture and activation are likely to affect tax and hbz expression.

We agree that it is of great importance and interest to quantify HTLV-1 expression in freshly isolated PBMCs. To do this, we are indeed now applying the smFISH technique described here to fresh PBMCs.

2.   Figure 1: It is surprising that 1% of uninfected clones (negative control) express 50 tax transcripts and only 1 hbz transcript. More uninfected clones should be tested. Furthermore, in order to validate tax and hbz probes, they show the specificity/sensitivity of smFISH on one clone generated from HTLV-I infected cell lines like having different expression levels of Tax and HBZ.

We apologize for the confusion. In fact, the frequency of cells with > 1 candidate spot in the negative control cells was ~0.6%. The legend to Figure 5 has now been modified to clarify this point.

3.   It is important to show the viral load and the status of expression of Tax and HBZ in total PBMCs of the 5 HTLV-I positive donors, and compare it with that of the 5 T cell clones that were selected. How can the authors be sure that the selected clone from each donor is a representative one of all the HTLV-1 infected cells in that individual? Did they screen other clones from each of the five donors?
We have previously shown that a typical HTLV-1-infected host carries > 10^5 infected T-cell clones, it is therefore not practical to study a representative sample of individual clones. Rather, the present study was designed to establish points of principle. We aim to tackle the question of

representativeness by quantifying mRNA expression in fresh polyclonal PBMCs isolated from infected individuals: this work is now in progress.

4.   *Figure 2: Images of tax bursts are more convincing that those of hbz. Figure 2D shows that the estimated number of hbz RNA per burst is around 3 to 4 with little intercellular variation, which is not classical for bursts. How can the authors rule out the possibility that these represent superposition of multiple transcripts. In 2C, they show a three-fold reduction in the frequency of hbz bursts in the presence of actinomycine D compared to a total disappearance of tax bursts. What is the percentage reduction of hbz bursts relative to the reduction of total hbz transcripts?*

Since the abundance of hbz mRNA is so low under all circumstances, there is indeed no strong distinction in the majority of cells between continuous hbz transcription and hbz bursts. The approach we have taken is described in detail in the Materials and Methods section.

5.   *Figure 3: Estimation of intranuclear mRNA should be done in zStacks to rule out counting transcripts above and below the nucleus.*

In developing the current protocol, we did indeed quantify the transcripts in z-stacks lying above or below the nucleus in individual cells. It was not practicable to inspect individual z-stacks on the large total number of cells studied (N > 19,000). However, the small error introduced by quantifying the distribution in 2D would not alter the major conclusions reached here on the distinction between the distribution of plus-strand transcripts and *hbz* transcripts.

6.   *Figure 4: Authors show increased frequency of hbz bursts in high tax positive cells. Again, what about the total amount of hbz transcripts in these cells?*

The total *hbz* spot count is indeed raised in the cells with a high plus-strand spot count. However, the logistic regression analysis (see Supplementary Figure 10), showed that the *hbz* spot count was correlated with the occurrence of *hbz* bursts, so it may not be meaningful to separate the two parameters here.

7.   *Elevated tax and hbz expression in S and G2/M phase of the cell cycle is solely based on the integrated intensity of the DAPI stained nucleus (2n versus 4n). It would be nice to perform synchronization of cells to confirm these findings.*

This is a good suggestion, and in fact these experiments are already underway.

8.   *Supplementary Figure 9: The negative control (uninfected cells) and clone E should be shown.*

Since the frequency of cells in the negative (uninfected) control (see point 2 above) and the tax-negative Clone E was <1%, the cell-cycle analysis of the frequency distribution of smFISH spots was not carried out with these clones.

9.   *Supplementary Figure 10: The odds ratio for G2/M phase is much higher for cells with >4 hbz compared to cells with >99 tax whereas that for S phase is quite similar. Does this indicate that hbz expression results in G2/M arrest? Is there any correlation between G2/M phase and nuclear size in these cells?*

It has previously been shown that hbz promotes progression through the cell cycle (Satou et al

2006; Mitobe et al 2015: refs. 39 and 40 in the paper). We therefore infer that this is the cause of the high odds ratio referred to, although our results do not exclude the possibility that a proportion of these cells are in G2/M arrest.

Reviewer 3 (Sitbon)

Billman et al. report here a detailed analysis of plus-strand (+strand) "tax" and anti-sense (hbz) RNA transcripts in HTLV-1 T cell clones that have been cultured ex vivo from CD4+CD25+ T cells isolated from PBMCs of 4 different patients. Each clone, harboring a single entire copy of the HTLV-1 genome, was cultured under periodic activation with IL-2, and anti CD2, CD3 and CD28 mAbs. All analyses were performed at day 8 of the 14 day activation cycle, when most cells are returning to a resting state. Single cell analyses of HTLV-1 RNA expression were based on integration of single molecule RNA-FISH and DAPI nuclear staining with occasional actinomycin D treatment.

Under these conditions, the authors observed that the vast majority of cells harbored detectable levels of hbz transcripts, while accumulation of +strand transcripts (which they refer to as "tax", see below) was more heterogeneous, varying from <10% in cells from 2 of the clones, to 50-60% of cells from the other 2 clones. An additional clone that harbored a deleted provirus, and thereby did not produce +strand transcripts, served as control. The authors further showed that both sense and anti-sense transcripts could be detected within larger nuclear spots in 4 to 7% of the cells analyzed, which the authors interpreted as transcriptional bursts; this interpretation was confirmed by the significant drop of bursts observed after treatment with actinomycin D. Act D further allowed the authors to show that +strand "tax" transcripts were rapidly accumulating to the cytoplasm while hbz transcripts remained in the nucleus. According to the authors, and in agreement with former observations by several teams, hbz and tax transcripts were in general inversely related. Also, the authors report that both types of transcripts tended to accumulate in cells in G2/M as compared to G0/G1 (monitored as a function of DAPI signal).

Elucidation of the in vivo dynamics of Hbz versus Tax production at the single cell level should indeed help elucidate the in vivo regulation of latent versus productive HTLV-1 infection, and provide precious clues on the pathogenic outcomes following infection. Despite an impressive amount of data cross-analyses based on +strand and anti-sense transcription of T cell clones derived from patients' PBMCs, the actual relevance of the model for in vivo situations is questionable. Thus, all analyses were performed on 4 clones isolated from a subpopulation of CD4+CD25+ T cells (as described in a reference and not mentioned here), which after being isolated from patients PBMCs, were cultured for a long period of time in conditions that do not reproduce in vivo conditions (hyperoxia, hyperglycemia, strong recurrent activations and absence of immune pressure). As such, several of the authors' conclusions (in the title, abstract and discussion) are not warranted (see below).

We agree that it remains to be proved how well the pattern of HTLV-1 transcription we report in these clones in vitro reflects the behaviour of the virus in vivo. In the paper we do not of course claim to have provided this proof: rather, we state 'If this pattern of HTLV-1 gene expression accurately represents the pattern of expression in vivo, the paradoxes [three identified] can be explained as follows:'; and elsewhere 'The results reported here offer an explanation for each of these paradoxes' and that our results 'strongly suggest[s] that the HTLV-1 plus-strand is expressed in bursts in vivo.' That is, we believe these results offer, but do not prove, a simple and logical

explanation of the observations in vivo. In the revised manuscript we have modified the conclusion section of the Abstract, and added a section to the Discussion (see point 7 below), to further clarify this important point:

As presented, the study brings up other several additional points:

1.   The authors state that all HTLV-1 +strand transcripts are detected with the so-called "tax" probe. Nevertheless, the constant designation of the transcripts as "tax" throughout the article can be very misleading. It would be more accurate to refer to the results obtained with this probe in the text and figures as "+strand", or another more inclusive designation, rather than "tax" solely.

At the beginning of the Results section we made this point clear by stating 'In this paper, *tax* is used as a proxy for plus-strand proviral expression: the *tax* target sequence is present in all plus-strand transcripts.' However, we accept that the reader may overlook this important point when reading the rest of the paper, and accordingly we have systematically replaced reference '*tax'* expression by 'plus-strand' expression in the revised manuscript.

2.   Nearly 60 to 90% of cells in all 5 clones express hbz antisense transcripts, but depending on the clone studied, the presence of +strand transcripts, as detected with the "tax" probe, are much more heterogeneous (<10% in clones A and D, and 50-60% in clones B & C). Given this overwhelming heterogeneity, how can the authors exclude that it is not due to variation in Tax-encoding transcripts, but solely to other viral +strand transcripts. In other words, the authors cannot exclude that levels of Tax-encoding transcripts were homogeneously low in all clones.

We have addressed this point in the revised manuscript by replacing '*tax'* with 'plus-strand' (see point 1 above). In fact, while there is a theoretical possibility that plus-strand transcripts other than *tax* dominate, it is well established that *tax* is the first plus-strand transcript, is most abundant, and crucially *tax* expression is required for efficient expression of other plus-strand RNAs. Also, even if *tax* were uniformly low in the plus-strand-expressing cells, the single-cell heterogeneity of *tax* would still be much more heterogeneous than that of *hbz.*

3.   Although the authors insist on the pitfalls of qPCR to analyze single cell transcription, there have been several reports on microfluidics technology being successfully used to monitor single cell RNA expression (i.e. Moussy et al. 2017. PLoS Biol 15(7):e2001867). The application of microfluidics on PBMCs directly isolated from patients would likely allow a more accurate assessment of the in vivo status of RNA transcripts.

We agree that it is of great importance and interest to quantify HTLV-1 expression in freshly isolated PBMCs. To do this, we are now applying the smFISH technique described here to fresh PBMCs.

4.   As mentioned above, given the experimental model of subcloning and culture (at the non-physiological 20% oxygen tension, which artificially modulates HTLV-1 activation, as previously published by the authors), and in the absence of immune pressure, the claim that the results presented here are relevant for in vivo situations and "offer an explanation for the paradoxical correlations observed between the host immune response and HTLV-1 transcription" is not warranted.

Our recently published evidence (Kulkarni et al 2017, Cell Chemical Biology **24**, 1–11) showed that

the oxygen tension modulated plus-strand HTLV-1 transcription, but this effect was approximately two orders of magnitude smaller than the variation in plus-strand abundance reported here, and there was no evidence that oxygen tension initiated plus-strand expression.

5.   Since the correlation between "transcription bursts" and the G2/M cell cycle stage is not established for +strand "tax" transcripts, and pertains to <7% of cells in which hbz is detected, the claim of a relationship between viral gene bursts and cell cycle (as stated in the general conclusion and the title) does not appear justified. Therefore, the intricate array of cross analyses and statistical evaluation performed here and addressed lengthily in the discussion is not likely to reflect in vivo situations.

The statistical analysis provides objective evidence of an association between transcription and cell cycle stage, and further shows that this association is independent of the other factors considered.

6.   Statistical significance of the correlation between high hbz expression, hbz bursts and the G2/M stages should be directly mentioned in Fig. 5 and the corresponding text. The clones used to obtain the data provided in panel 5b are not indicated (it appears that clones A and B were used here, whereas data in 5c are mentioned in the figure legend to result from the average of clones A, B, C and D).

Thank you for this useful suggestion. In the revised manuscript we have included a reference to the statistical test and significance of the association in the legend to Figure 5; it is also mentioned in the text of the Results section. The cells used in these experiments were the four *tax*-competent clones: this was described in the figure legend; the legend has been modified to make it clearer.

7.   Given the reservations mentioned above (an in vitro system that does not recapitulate in vivo conditions, the lack of distinction between tax versus +strand transcripts, and the lack of data on Tat and HBZ proteins), the discussion remains too speculative.

As noted above, we stated in the manuscript 'If this pattern of HTLV-1 gene expression accurately represents the pattern of expression *in vivo* [...]' and that our results 'offer an explanation' of certain observations *in vivo*. However, we agree that it is important to remind the reader of the actual and potential differences in HTLV-1 transcription between *in vivo* and *in vitro*, and accordingly we have added the following sentence to the end of the Discussion, to make clear what remains hypothetical: 'It remains to be tested how closely the transcriptional behaviour of HTLV-1 observed in these naturally-infected T-cell clones represents the transcriptional behaviour of the virus *in vivo*, both qualitatively and quantitatively. We are now applying the techniques described here to quantify HTLV-1 plus- and minus-strand transcription in PBMCs both directly *ex vivo* and after short-term *in vitro* incubation.' We have also modified the concluding section of the abstract, to say 'If so [if the *in vitro* behaviour accurately represents the *in vivo* behaviour], then our results offer an explanation...'

8.   In this study, there is no assessment of the relationship between transcript and protein levels; moreover, since the half-lives of Tax and Hbz proteins in this model are not established, it is not possible to evaluate the actual timing of a potential cross-regulation between the 2 proteins. In this regard, the authors should discuss a recent paper by Baratella et al. (PLos Negl Trop Dis Jan 2017) that describes the specific accumulation of the HBZ protein in the cytoplasm of PBMCs directly isolated from TSP/HAM patients.

Thank you for drawing our attention to this interesting paper by Baratella et al. In the revised manuscript we have included the following sentence in the Discussion section to refer to this work: 'HBZ protein is frequently undetectable in fresh PBMCs, but Baratella et al. [46] reported the detection of HBZ protein in the cytoplasm of PBMCs isolated from patients with the inflammatory disease HAM/TSP.' A detailed examination of the relationship between mRNA and protein levels is beyond the scope of the present paper.

9.   In suppl. fig 7 - there is an SEM indicated in the histogram of clone C while there was only 2 replicates, according to the figure legend. Also, in suppl. Figs 8 and 9 there appears to be a color and legend inversion for clones B, C, and D, and clone A does not seem to fit the G2/M correlation with high levels of +strand "tax" transcripts.

The standard error of the mean is defined for N > 1. Thank you for pointing out the inversions of colour and legend in Figures 8 and 9: these have now been corrected. Indeed, Clone A did not fit the G2/M correlation with a high frequency of plus-strand transcripts: this point is discussed in the legend to Supplementary Figure 9.

Competing Interests: No competing interests were disclosed.