RESEARCH ARTICLE

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

Martin R Billman¹, David Rueda², Charles R M Bangham ³¹

¹Section of Virology, Division of Infectious Diseases, Department of Medicine, Imperial College London, Norfolk Place, London, UK
²Section 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 G2/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.
Corresponding authors: David Rueda (d.rueda@imperial.ac.uk), Charles R M Bangham (c.bangham@imperial.ac.uk)

Author roles: Billman MR: Preparation, Writing – Review & Editing; Rueda D: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft; Bangham CRM: Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing

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 1; referees: 2 approved, 1 approved with reservations] Wellcome Open Research 2017, 2:87 (doi: 10.12688/wellcomeopenres.12469.1)

Copyright: © 2017 Billman MR et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was supported by the Wellcome Trust [100291 and 099858] and a core grant of the MRC London Institute of Medical Sciences (RCUK MC-A658-5TY10).

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. 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. 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, 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. Several actions of Tax and HBZ are mutually antagonistic, but both Tax and HBZ play crucial roles in viral persistence, gene expression and leukaemogenesis. 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 (e.g. by flow cytometry) or nucleic acid (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. In addition, assays of viral expression in a cell population mask 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 RNA fluorescent in situ hybridisation (smFISH) to quantify the transcripts of tax 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 tax gene. 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.

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.

The clones used, their integration sites and the patients they were derived from are summarised in Table 1:

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

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% CO2, 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.
RNA-FISH probes were designed using Stellaris’s online probe design tool (version 4.1), with the reference genome AB513134 (ref. 14) used as input for the tax probe (post-splice coding sequence) and hbz (post-splice coding sequence and 3'-UTR as reported in ref. 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 tax with a Quasar-670 fluorophore.

Fluorescence microscopy

All coverslips were imaged using an Olympus IX70 inverted wide-field 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 (tax-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 v2.2.015, with a pipeline adapted from ref. 16.

Images were then analysed with an adapted version of FISH-Quant, v31718, 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 ref. 16. Cell-cycle gates were determined by visual inspection of the histograms, following the rule that cells in G1/M have double the DNA content (and therefore integrated intensity) of cells in G1/G0 (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 tax and hbz bursts shown in Figure 4 were all analysed using chi-squared tests, while the relationship between hbz and tax expression and cell cycle stage was examined using binary and multinomial logistic regression with the “glm” and “multinom” functions respectively in R, v3.3.319.

The expected frequency of hbz-negative cells in each clone was calculated from the observed mean hbz spot count in the respective clone, using the Poisson distribution (Supplementary Figure 1). 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 (containing the tax sequence) and the minus-strand (hbz) (Figure 1a). 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. 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\(^2\); the provirus is complete in four clones, while in the fifth clone (Clone E) the 5’-LTR is deleted and the tax gene thereby silenced, a phenomenon common in leukaemic clones\(^9\). The diffraction-limited signal of a single spot in smFISH corresponds to a single mRNA molecule\(^\)\(^1\). In total, 19,477 individual cells were analysed, quantifying tax 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 tax 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, tax presented a bimodal expression profile. In most clones, only a small fraction of cells were tax+, but these cells expressed tax at a much greater level than hbz. Across all clones, cells with ‘high’ tax (≥ 100 spots, Figure 1b.v) made up 4.9% of all cells, but contributed over 80% of all tax mRNA detected. The variation in total tax 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 tax expression. The maximum level of tax observed in clone D was much lower; however, all four clones had a bimodal distribution of tax 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 tax expression was that many tax+ cells also had a large nuclear spot which was typically much brighter than the average spot (Figure 2a). We surmised that these bright spots correspond to transcriptional bursts, which are increasingly recognized in mammalian genes\(^22\). 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 tax 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.

tax and hbz bursts differ in intensity

The size of tax and hbz bursts was 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, tax bursts were much larger: the largest bursts were estimated to contain hundreds of transcripts. tax bursts were also more variable than hbz bursts in size (spot count). Bursts, whether tax or hbz, were not uniformly distributed throughout the nucleus, and their spatial localization differed between clones (Supplementary Figure 4).

tax is rapidly exported, hbz remains largely intranuclear

Tax protein generates a strong, persistently activated T-lymphocyte response in infected people\(^4\). 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\(^3\).\(^\)\(^6\). We therefore investigated the proportion of both transcripts that could be found in the nucleus or cytosol of cells.

Again, the differences between tax 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 tax transcripts were found in the nucleus in 3 of 4 tax-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 z-axis are erroneously marked as intranuclear. This explains why the proportion of tax within the nucleus
Figure 1. Single-molecule RNA-FISH reveals heterogeneity of tax and hbz expression within and between clonal cell populations.

a. Schematic of the HTLV-1 genome, showing the spliced transcripts of tax and hbz. 
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 tax: 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 tax expression in individual cells (same clones as in c).
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 tax 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 tax spots in many tax-expressing cells precluded a precise spot count
Figure 3. **tax** 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 **tax** and **hbz**, ii) a typical **tax**-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 **tax** 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.

of **tax** (Supplementary Figure 14); as a result, it was not possible to make a reliable estimate of **tax** mRNA 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 tax-expressing clones to test for a correlation between tax and hbz expression.

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

Many of the cells with high tax expression were hbz-negative (Figure 4a.ii). However, bursts of hbz expression were almost three times more frequent in tax-high cells than cells with no tax or low levels of tax (Figure 4a.iii, 4d). In fact, all of the 689 observed hbz bursts in low-tax or tax-silent cells lacked an accompanying tax burst (Figure 4a.iv). This effect was observed in all three clones (A, B and C) that contained cells with high tax expression. As further evidence of the reciprocal relationship between tax and hbz expression, cells containing both a tax burst and an hbz burst were significantly less frequent (2.9% of tax-high cells) than expected by chance (7.2%); of the tax-high cells, 14.6% had

**Figure 4.** Relationship between expression of tax and hbz in individual cells. a. Representative images of a i) tax-silent cell, ii) tax-high cell (≥100 spots) with a tax burst, iii) tax-high cell without a tax burst (note the presence of an hbz burst), iv) tax-low cell (<100 spots) and v) tax-silent cell, with low-level hbz. b. tax bursts occur more frequently in hbz-negative cells. c. tax bursts are indicative of very high tax expression, with the proportion of cells which have a burst increasing as the level of tax increases. d. hbz bursts occur more frequently in high-tax cells than in tax-negative or tax-low cells. Numbers above columns denote the number of cells in the corresponding population, from four pooled tax-competent clones.
an \textit{hbz} burst and 49.6\% had a \textit{tax} burst (p < 0.0001; chi-squared). Finally, among high-\textit{tax} cells (N = 757), \textit{hbz} bursts were significantly more frequent in cells without a \textit{tax} burst than in those with a \textit{tax} burst (23.2\% vs. 5.8\%; p < 10^{-5}, chi-squared), consistent with the notion that the \textit{hbz} burst inhibits \textit{tax} transcription and thereby terminates the \textit{tax} burst.

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

\textit{hbz} is not expressed in all cells at a given time
Since \textit{hbz} is expressed at a very low level, the unexpectedly high observed frequency of \textit{hbz}-negative cells might be due to a failure to detect \textit{hbz} mRNA in some cells. If the \textit{hbz} spots were distributed randomly among the cells in a clone, then the observed average frequency of \text{~}2 to 3 \textit{hbz} spots/cell would result in zero spots in \text{~}5\% to 18\% of cells respectively (Poisson distribution; Supplementary Figure 1). The observed frequency of \textit{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 < 0.01 in each case; Supplementary Figure 1). We note that tax 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 G2/M phase have elevated tax 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 (Figure 5a). The cell-cycle profiles were reproducible between replicate samples of each respective clone (Supplementary Figure 7), but there were marked differences between clones. Compared with cells in G0/G1, cells in G2/M showed a greater mean intensity (spot count) of expression of both hbz (Figure 5b, Supplementary Figure 8) and tax (Figure 5d, Supplementary Figure 9), and a higher frequency of hbz bursts. These effects were less pronounced in tax: only a small proportion of cells in G2/M were tax-high (Figure 5d). However, since tax-high cells contribute over 80% of total tax, an increase in tax-high cells from 2.6% in G0/G1 to 5.6% in G2/M resulted in a significant increase in total tax expression. tax-silent and tax-low cells were more likely to be in G0/G1, whereas tax-high cells were more likely to be in S or G2/M (Figure 5c), regardless of whether they contained tax 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 G2/M, as were one third of cells with low hbz and a burst.

Given that high Tax can lead to expression of hbz25, the question arose whether hbz expression was independently correlated with the cell cycle, or correlated with tax. Using logistic regression analysis, we found that both high tax 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 tax than if they were tax-low or silent, and a cell with a given level of tax was on average 50% more likely to contain an hbz burst if it was in G2/M rather than G0/G1 (Supplementary Figure 12).

Both high hbz expression and an hbz burst were independently significantly associated with G2/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 tax was significantly correlated with G2/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 tax bursts in this clone. Although the level of tax expression in Clone D was low as defined in this study, a higher tax spot count in Clone D cells was still associated with S or G2/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 (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.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.18,19. 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.20,21

The results reported here offer an explanation for each of these paradoxes. By quantifying the frequency and intensity of hbz and tax 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). 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 (as indicated by the tax sequence) is not uniform, but rather exhibits a bimodal expression profile (Figure 1d). The minority of cells expressing tax at any one time express the transcripts at a very high level (Figure 4c). The tax transcripts are then rapidly exported to the cytosol (Figure 3b). The resulting intermittent but intense expression of the highly immunogenic Tax protein25 is sufficient to account for the observed preponderance of anti-Tax CTLs. 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 (tax), consistent with previous observations25,26, 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 suggested26 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.20,27 Our estimate of hbz half-life is higher than that made in previous studies25,28, 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\textsuperscript{24,25}. These previous observations, in combination with our observations of the correlation between levels of *tax* and *hbz* within individual cells (Figure 4), suggest the following model of the interaction between *tax* and *hbz*. *tax* is more likely to be expressed in an *hbz*-silent cell (Figure 4a.i), and once expressed *tax* typically reaches very high levels (Figure 4a.ii). A high level of *tax*, however, increases the probability of *hbz* expression\textsuperscript{28} (Figure 4a.iii), which in turn leads to cessation of *tax* expression. Subsequently, both *tax* 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 *tax* expression include glucose metabolism and the available molecular oxygen\textsuperscript{30}; 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 *tax* burst to develop, before *tax* 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\textsuperscript{9,26,30,34,35}. 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\textsuperscript{36,37}. 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\textsuperscript{38}; 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/M. This burst timing contrasts with the transcription rate of many cellular genes, which is reduced during S/G/M, probably to compensate for the doubling of the gene copy number\textsuperscript{39}. The occurrence of an *hbz* burst in S/G/M might confer two possible advantages on the virus. First, *hbz* promotes progression through the cell cycle\textsuperscript{40,41}. 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\textsuperscript{36}.

The observed longevity of HTLV-1-infected T-cell clones in vivo indicates that sustained cell proliferation plays a significant part in viral persistence\textsuperscript{31}. The relationship between *tax* and *hbz* expression and the cell cycle is therefore of critical importance. We found that cells in G/M had on average a significantly higher amount of *tax*, and the minority of cells with high *tax* were significantly more frequent in S or G/M than in G/G (Figure 5). In contrast, *tax* bursts were not significantly associated with S or G/M phase (Supplementary Figure 10), perhaps because of the low frequency and transient nature of *tax* bursts. The strength of this relationship between *tax* 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/M (rather than G/G) by an average of 1.45-fold (7.3-fold in *hbz*-high cells), whereas each additional *tax* transcript increased the odds by 1.004-fold (1.58-fold in *tax*-high cells). This correlation with *hbz* was maintained in the clone incapable of expressing *tax*, showing that the relationship between *hbz* and the cell cycle is independent of *tax*. Although several mechanisms have been described by which Tax protein and both HBZ protein and RNA promote cell proliferation\textsuperscript{34,35,40}, 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)\textsuperscript{34,42}.

Previous evidence suggested that *hbz* is constitutively expressed in HTLV-1-infected T-cells, albeit at a low level\textsuperscript{24,34}. 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 Figure 1); the departure from the Poisson distribution was consistently greater in the clones that expressed high levels of *tax*. 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 *tax* 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 *tax* and *hbz* transcripts have different fates, with the majority of *hbz* retained in the nucleus while *tax* 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, *tax* expression is more
Figure 6. Model of regulation of HTLV-1 transcription and replication in vivo. Constitutive expression of \textit{hbz} mRNA maintains clonal longevity by promoting cellular proliferation. Intermittent bursts of \textit{tax} expression have three chief consequences: production of HTLV-1 virions and resulting infectious spread; accelerated cell-cycle progression; and stimulation of \textit{hbz}, of which the RNA and protein both have proliferative effects. HBZ protein can also inhibit \textit{tax} transcription, terminating the \textit{tax} burst. It is not known whether \textit{hbz} mRNA also counteracts \textit{tax} transcription, but it is suggested by our results. The factors that regulate the frequency of \textit{tax} bursts include glucose metabolism and hypoxia; cell stress and cell activation may also trigger \textit{tax} expression. Aside from Tax, other triggers for \textit{hbz} expression remain to be discovered. Dashed line represents a link indicated by results in the present study; dotted lines represent hypothetical links.

Figure 6. Model of regulation of HTLV-1 transcription and replication in vivo. Constitutive expression of \textit{hbz} mRNA maintains clonal longevity by promoting cellular proliferation. Intermittent bursts of \textit{tax} expression have three chief consequences: production of HTLV-1 virions and resulting infectious spread; accelerated cell-cycle progression; and stimulation of \textit{hbz}, of which the RNA and protein both have proliferative effects. HBZ protein can also inhibit \textit{tax} transcription, terminating the \textit{tax} burst. It is not known whether \textit{hbz} mRNA also counteracts \textit{tax} transcription, but it is suggested by our results. The factors that regulate the frequency of \textit{tax} bursts include glucose metabolism and hypoxia; cell stress and cell activation may also trigger \textit{tax} expression. Aside from Tax, other triggers for \textit{hbz} expression remain to be discovered. Dashed line represents a link indicated by results in the present study; dotted lines represent hypothetical links.

likely to occur in the absence of \textit{hbz}, whose expression is in turn more likely in cells with high \textit{tax}. Finally, proviral expression is correlated with the phase of the cell cycle, with clear implications for driving proliferation and evasion of immunosurveillance. 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**

The full raw dataset that supports the study findings is too large to be feasibly hosted by a repository. Sample image data are available at: https://osf.io/b9mnd/. Further data are available on request from the corresponding author (c.bangham@imperial.ac.uk);

the large dataset (~2 Tb) is most reliably transferred on a hard disk.

**Competing interests**

No competing interests were disclosed.

**Grant information**

This work was supported by the Wellcome Trust [100291 and 099858] and a core grant of the MRC London Institute of Medical Sciences (RCUK MC-A658-5TY10).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Acknowledgements**

The authors thank the staff and donors at the NIHR Biomedical Research Centre at Imperial College Healthcare NHS Trust, Dirk Dormon, Chad Whilding and Sheila Xie (MRC London Institute of Medical Sciences) for help with the image analysis and FISH techniques, Florian Mueller for his adaptation of FISH-Quant to allow cell cycle staging and helpful discussions, and members of the Bangham laboratory for helpful discussion.

**Supplementary material**

Supplementary File 1 contains Supplementary Figures 1 to 15.

Click here to access the data.

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

Click here to access the data.
References


Data Source
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 to 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.
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.

Referee Report 16 October 2017

doi:10.21956/wellcomeopenres.13502.r26294

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.

---

**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 \textit{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...
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 than 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.