bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The transcriptome of HTLV-1-infected primary cells following reactivation reveals changes to host gene expression central to the proviral life cycle

Aris E. N. Aristodemou[1], David S. Rueda[1, 2], Graham P. Taylor[1], Charles R. M. Bangham[1]*

1 Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK

2 Single Molecule Imaging Group, MRC-London Institute of Medical Sciences, London, UK

* Corresponding author: c.bangham@imperial.ac.uk

Abstract

Infections by Human T cell Leukaemia Virus type 1 (HTLV-1) persist for the lifetime of the host by integrating into the genome of $CD4^+$ T cells. Proviral gene expression is core to proviral survival and the maintenance of the proviral load, through the pro-proliferative changes it induces in infected cells. Despite their role in HTLV-1 infection and a persistent cytotoxic T lymphocyte response raised against them, proviral transcripts from the sense-strand are rarely detected in fresh cells extracted from the peripheral blood, and have recently been found to be expressed intermittently by a small subset of cells at a given time. *Ex vivo* culture of infected cells prompts synchronised proviral expression in infected cells from peripheral blood, allowing the study of factors involved in reactivation in primary cells. Here, we used bulk RNA-seq to examine the host transcriptome over six days *in vitro*, following proviral reactivation in primary peripheral $CD4^+$ T cells isolated from subjects with non-malignant HTLV-1 infection. Infected cells displayed a conserved response to reactivation, characterised by discrete stages of gene expression, cell division and subsequently horizontal transmission of the virus. We observed widespread changes in Polycomb gene expression following reactivation, including an increase

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in PRC2 transcript levels and diverse changes in the expression of PRC1 components. We hypothesize that these transcriptional changes constitute a negative feedback loop that maintains proviral latency by re-deposition of H2AK119ub1 following the end of proviral expression. Using RNAi, we found that certain deubiquitinases, *BAP1*, *USP14* and *OTUD5* each promote proviral transcription. These data demonstrate the detailed trajectory of HTLV-1 proviral reactivation in primary HTLV-1-carrier lymphocytes and the impact on the host cell.

Author summary

Human T cell Leukemia Virus type I (HTLV-1) is a retrovirus which causes an aggressive leukemia or lymphoma, or a chronic inflammatory disease of the central nervous system, in a subset ($\approx 10\%$) of affected carriers. Whilst the virus is only intermittently expressed by infected cells, the virus persists in its host by increasing the proliferation rate and survival of infected cells. It is therefore imperative to understand the mechanisms that control the activation and deactivation of the virus. We examined the expression of host and viral genes during HTLV-1 reactivation in cells freshly isolated from patients' blood. The infected cells displayed consistent changes in gene expression over six days. We observed changes in the expression of the Polycomb group of epigenetic modifiers, known to impact HTLV-1 transcription, which may form a negative-feedback mechanism allowing the virus to return to a latent (quiescent) state following activation. We additionally identified a set of three deubiquitinases that increase expression of the virus. Our data detail the changes in gene expression underlying core aspects of the HTLV-1 life cycle in primary cells, and provide a resource for further investigation.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

As obligate genomic parasites, retroviruses depend extensively upon host cell machinery to progress through their life cycle. Human T cell leukaemia virus type 1 (HTLV-1) is the first retrovirus discovered infecting humans (1) and is the causative agent of a debilitating neuropathy, HTLV Associated Myelopathy/Tropical Spastic Paraparesis, and an aggressive malignancy, Adult T cell Leukaemia, each affecting $\approx 5\%$ of carriers (2; 3; 4). The risk of disease is correlated with the maintenance of a high proviral load (PVL) in carriers (5; 6; 7). A persistently active cytotoxic T lymphocyte response to antigen produced from the proviral sense-strand accompanies infection, irrespective of disease status (8), and despite the low frequency of expression of viral proteins in freshly isolated cells*in vivo* (9). The proviral antisense product, *HBZ*, is more readily detected at the RNA level (10; 11; 12); however, it is poorly immunogenic (13; 14).

Ex vivo culture of HTLV-1-carrier primary PBMCs stimulates proviral sense-strand expression (15; 16), dominated by *tax/rex* transcription and Tax protein expression (17). The gradual accumulation of Rex results in increased nuclear export of unspliced and singly-spliced mRNA species, with consequent depletion of *tax/rex* mRNA and production of structural and enzymatic viral components necessary for virion formation and horizontal transmission. Viral expression is accompanied by changes in host gene expression driven mainly by the pluripotent effects of Tax and HBZ upon host signaling pathways (18), the cell cycle (19) and apoptosis (20; 21).

Investigations of HTLV-1-infected samples at the single-cell level have uncovered high levels of heterogeneity in proviral gene expression, suggesting transcriptional bursting of the proviral promoter and intermittent expression (11; 12). Live imaging of Tax expression using a short-lived GFP reporter revealed these transient expression episodes in MT-1 cells (21), a finding later confirmed by the use of a fluorescent timer reporter for Tax expression which allows identification of cells which have recently terminated Tax expression (22). Intermittent expression offers a

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

plausible mechanism for limiting the exposure of antigen to adaptive immunity whilst reaping the benefits of proviral gene expression, particularly those conferred by Tax and HBZ, which together contribute to infected cell survival and proliferation (18; 23).

Integrated into the host genome, the provirus is subject to the same molecular mechanisms which regulate host gene expression. Chromatin immunoprecipitation (ChIP) experiments have identified the distribution of multiple epigenetic modifications along the provirus (24; 25; 26; 27; 28; 29). Histone tail post-translational modifications commonly associated with increased transcriptional activity - H3K27ac, H3K9ac and H3K4me3 - are found on the provirus, and differ between Tax^{+} and Tax^{-} cells (28). Repressive histone modifications are also associated with the provirus, in particular H2AK119ub1 and H3K27me3 (26; 27), respectively deposited by the Polycomb repressive complexes 1 and 2 (PRC1/2) (30). Proviral reactivation has been demonstrated to depend upon the removal of H2AK119ub1 from the proviral promoter in the $5'$ LTR (27), but the roles of epigenetic chromatin modifications in active silencing of the provirus are unknown.

Whilst it is established that HTLV-1 gene expression perturbs host transcription, and many such changes have been invoked in the oncogenesis of ATL, few of these results have been obtained or confirmed in PBMCs from HTLV-1 carriers, and even fewer have examined expression over a longer time frame allowing the observation of changes which occur over both the early and late phases of proviral reactivation. We analysed the transcriptomes of primary HTLV-1-infected $CD4^{+}$ T cells cultured *ex vivo* over six days in six subjects with non-malignant HTLV-1 infection and three uninfected control subjects. The resulting RNA-seq timecourse captured the switch from early to late sense-strand gene expression. We observed a marked difference in host gene expression between infected and uninfected samples, with the former showing conserved changes in gene expression over time as cells enter the cell cycle and commence the horizontal transmission of virions. We investigated the pathways stimulated during *ex vivo* culture

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in uninfected cells, finding that these are already active in non-malignant infected cells at the point of entry into culture. Finally, we observed widespread dysregulation of Polycomb gene expression in infected cells during proviral reactivation.

Results

Isolation of HTLV-1-infected population using CADM1

To obtain a profile of HTLV-1 and host gene expression over the reactivation cycle, we used frozen PBMC samples collected from consenting HTLV-1 positive and uninfected volunteers. Our sample consisted of two asymptomatic HTLV-1 carriers (HES and HQL) and four individuals diagnosed with HAM/TSP (TCL, TCQ, TDK and TDX). The cells were thawed, enriched for $CD4^+$ lymphocytes by negative selection, sorted for viability and high CADM1 expression and samples were withdrawn for analysis at 0, 24, 48, 96, 120 and 144 hours after the start of *in vitro* culture. RNA was extracted at each timepoint and used for qRT-PCR quantification of HTLV-1 transcripts and quality control prior to submission for sequencing (Fig. 1A).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

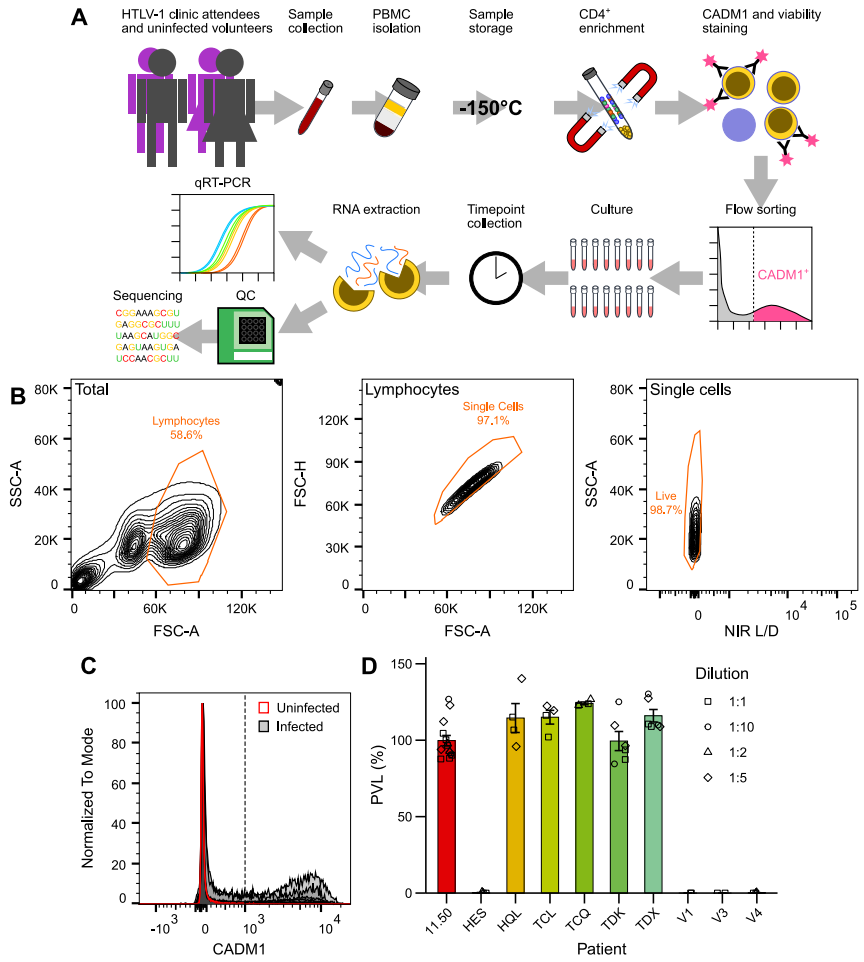

Figure 1: Sampling of HTLV-1 infected cells along expression cycle *ex vivo*.

(A) Overview of sample preparation and collection. (B) Gating scheme used to obtain single live lymphocytes from peripheral blood lymphocyte samples. (C) Gating of CADM1^{high} cells was performed based on CADM1 levels in uninfected controls. Dotted line represents CADM1 intensity threshold for infected samples. (D) ddPCR measurement of PVL of CADM1^{high}-sorted cells used for RNA-seq at t_0. Infected T cell clone 11.50, which contains a single copy HTLV-1 provirus per cell (31; 32), was used as a positive control. Columns and error bars represent mean ±SEM.

CADM1^{high} cell sorting was performed before culture to ensure that infected cell gene expression signatures were not diluted by the presence of uninfected PBMCs. High levels of CADM1 are expressed exclusively by HTLV-1-infected CD4$^+$ T cells upon their extraction from the circulation (Fig. 1C), and represent ≈ 65% of the total PVL (33). Samples from un-

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

infected individuals were processed identically, to control for the effects of thawing, cell-sorting and subsequent *ex vivo* culture. DNA was extracted from a subset of cells following sorting to measure PVL by ddPCR (34). PVLs in sorted HTLV-1 patient samples were found to be $\approx 100\%$ (Fig. 1D). Given that each infected cell contains on average a single proviral copy (31), we inferred that our sorted samples consisted entirely of infected cells. No viral copies were detected in the uninfected control samples (V1, V3 and V4). No viral copies were detected by PCR in the sample for one infected subject, HES, however, this is likely due to polymorphism in the primer — or probe — binding sites in the provirus; proviral transcripts were later detected in this sample both by qRT-PCR and RNA-seq (Fig. 2).

Polyclonal infected cell populations respond homogeneously to *ex vivo* culture

Prior to sequencing, RNA from patient samples was subjected to qRT-PCR to examine the trajectories of HTLV-1 transcripts (Fig. S1). A decrease in *tax/rex* transcript abundance was observed following the 24-hour timepoint, together with a decrease in total sense-strand transcripts, detected using a probe complementary to the pX region that is present in all sense-strand transcripts. Signal from the *gag-pro-pol* polytranscript first sense-strand intron, detected using probes complementary to *gag*, remained relatively constant following the first 24 hours (Fig. S1). The overshoot in spliced transcript abundance combined with the monotonic increase of unspliced product is consistent with the precursor-depletion-mediated negative feedback, mediated by Rex in HTLV-1 (17) and by Rev in HIV-1 (35). A similar overshooting trajectory can be observed in sequenced reads assigned to proviral sense-strand transcript (Fig. 2A). The abundance of reads aligning to the proviral sense strand correlated well with the qRT-PCR results obtained using the pX primer pair (Fig. 2B). Reads aligning to the HTLV-1 provirus were absent from the uninfected samples (Fig. 2A), except in sample V4 t_{24}, where a small ($\approx$0.5% of infected sample mean) number of reads aligned to the sense and antisense proviral strands (Fig. S2). These

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reads most likely represent contamination from infected samples, and this sample was therefore excluded from further analyses. No reads mapping to HTLV-1 were detected in other timepoints for the same patient, suggesting that t_{24} was the only timepoint affected by contamination.

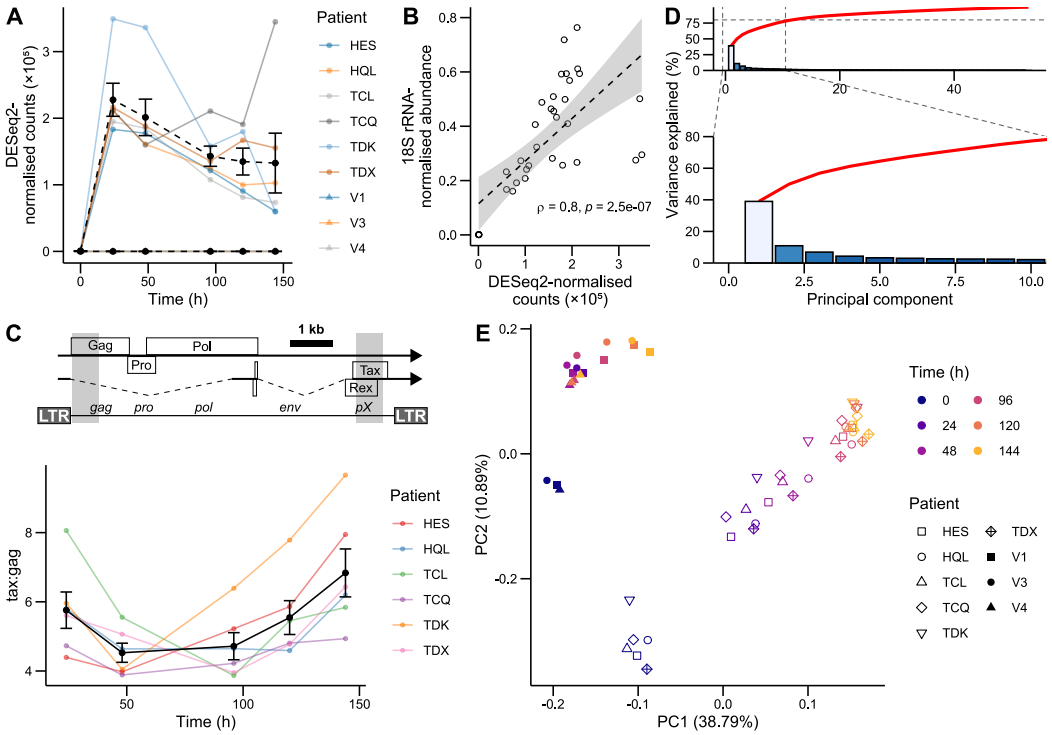

Figure 2: RNA-seq dataset captures progression of infected cells along HTLV-1 expression stages.

(A) Trajectory of sequenced reads mapped to the HTLV-1 sense-strand, excluding the LTRs. (B) Correlation between RNA-seq and qRT-PCR HTLV-1 transcript trajectories. Dotted line represents linear model fit. Shaded area represents linear model 95% confidence interval. Statistics shown were calculated using a paired Spearman rank correlation test. (C) Above: Schematic of proviral sequence with *tax/rex* and *gag-pro-pol* polytranscript shown. Dotted lines represent introns, boxes represent ORFs, and lines represent untranslated regions. Shaded regions correspond to to AB513134 regions 804–1,433 and 7,400–8,029, used to quantify reads aligning to *gag* and *tax* regions, respectively. Below: Ratio of reads aligning to proviral *tax* and *gag* regions, corresponding to the ratio of spliced to unspliced transcripts. Summary line and error bars represent mean ±SEM. (D) Scree plot showing contribution of principal components to total variance observed, with red line representing cumulative sum. Zoom inset focuses on principal components accounting for ≈ 80% of variance in the data. (E) PCA bi-plot of infected and uninfected cell gene expression.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We next examined the ratio of early to late viral transcripts over time, by quantifying the number of reads aligned by STAR (36) to regions corresponding to the Gag CDS or the final Tax exon. We observed a "U"-shaped trajectory in the ratio of double spliced to unspliced transcripts (Fig. 2C), corresponding to an initial switch towards the expression of genes for virion production ($\leq$ 96 hrs.), followed by cells switching back to the expression of spliced transcripts during a second wave of reactivation, either in the same cells or newly reactivating cells. We also examined the levels of antisense transcripts over time, which correlated positively and significantly with sense-strand transcripts (Fig. S3), corroborating previous observations of regulation of *hbz* by sense-strand reactivation (37). This correlation complicates the assignment of the observed effects on host cell transcription to products of either proviral strand.

Principal component analysis (PCA) was performed on variance-stabilising transformation (VST) normalised transcript count matrices obtained following Kallisto pseudoalignment, to visualise relationships in overall gene expression between samples. The first two PCs explain just under 50% of the total variance, with the contribution of subsequent individual PCs tailing off rapidly (Fig. 2D). Examination of sample clustering along these two PCs suggested that the main sources of variance in the data were HTLV-1 infection status and time in culture. The substantial difference in mean age between the infected and uninfected subjects might contribute to this variance (Fig. S4). Infected samples showed strong within-timepoint clustering up to and including 96 hours, suggesting a common gene expression progression profile. In contrast, uninfected samples changed relatively little following the initial stimulus of their introduction into culture (Fig. 2E).

An additional comparison was performed with the dataset obtained by (22), which utilised a fluorescent timer protein to distinguish between the successive phases of Tax expression

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Fig. S5). Clonal infected T cells clustered along with infected samples in the second principal component. Blue, double-positive and red timer protein populations, representing respectively early, mid- and post-Tax expression cell states, clustered together with the 24-hour samples. The PBMC dataset therefore captures a greater degree of diversity over proviral expression. The 96–144-hour timepoints did not coincide with the red population. This observation suggests that Tax expression might not have terminated by 144 hours. Double-negative populations were positioned between the 0 and 24-hour timepoints, potentially reflecting changes related to culture adaptation, as they fall closer to the late-timepoint uninfected samples in the first principal component.

These results suggest that the observed trajectory captures at least part of the switch between early and late proviral transcripts. Within the period of observation of 144 h, infected cells progress through distinct stages of gene expression caused by viral gene expression. This progression is conserved between individual patients, irrespective of disease status.

Discrete changes to cell state occur following HTLV-1 reactivation

To obtain an overview of the changes in gene expression underlying differences between measurements at different timepoints, we performed differential expression analysis of Kallisto transcript pseudocounts (38) using `DESeq2` (39). An interaction model accounting for differential expression over time between infected and uninfected cells was used, yielding 7,753/28,693 significantly differentially expressed genes (Likelihood-ratio test; Benjamini-Hochberg FDR-adjusted $p < 0.05$). A heatmap of these differentially expressed genes (Fig. 3A) showed clustering between samples consistent with that seen in the first two principal components (Fig. 2E): uninfected samples clustered separately from infected samples and showed similar expression patterns after 24 hours, with little clustering within individual timepoints. In contrast, infected samples clustered well by timepoint up to 120 hours, at which point patient-specific differences became

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

dominant (Fig. 3A).

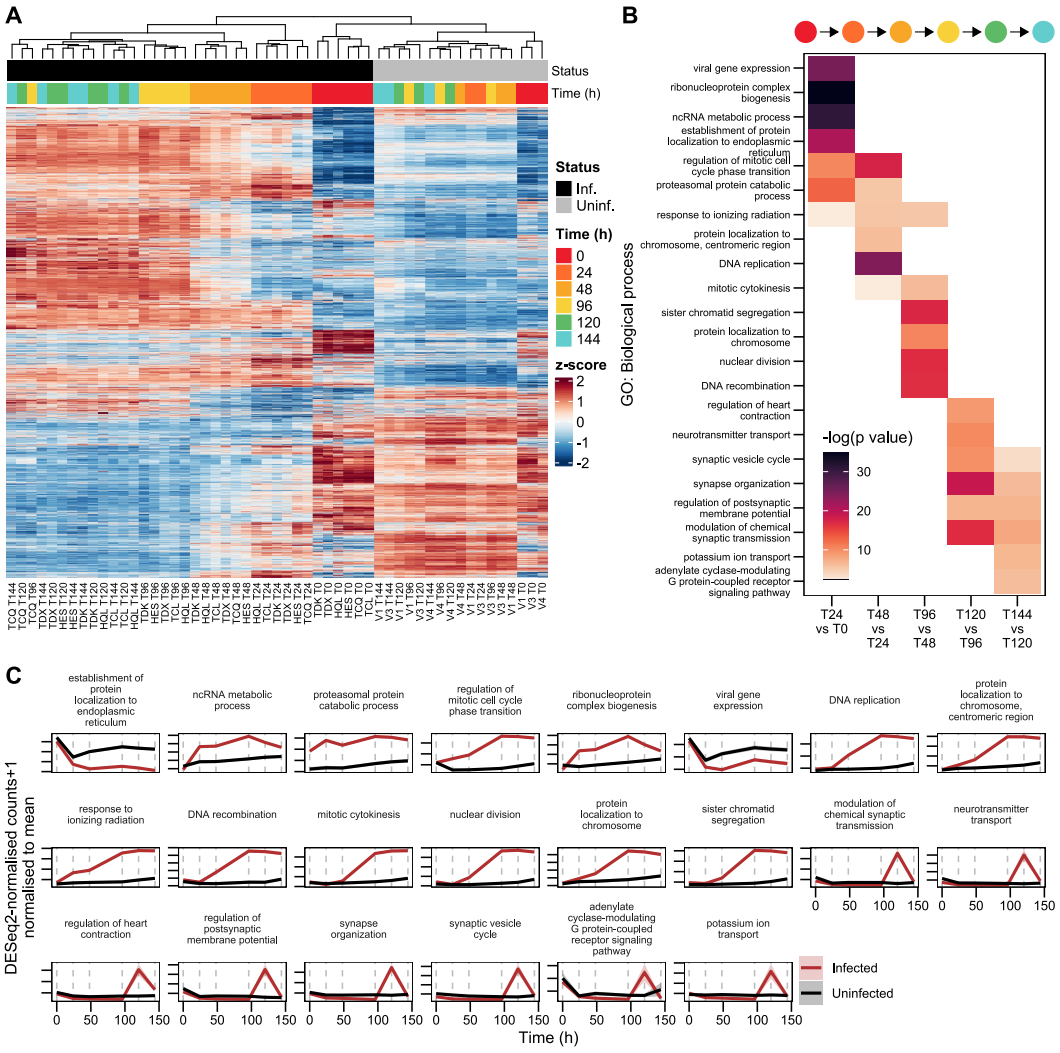

Figure 3: Timepoints separated by distinct changes in gene expression specific to infected cells.

(A) Heatmap of genes differentially expressed over time between infected and uninfected samples. (B) Top five (based on FDR-corrected p value) non-synonymous GO biological process terms associated with each interval. (C) Summary of mean-normalised trajectories of genes from for GO terms presented in panel B. Discontinuous lines indicate sampling timepoints. Shaded areas represent ±SEM.

We next examined further the differences between infected timepoints, to identify processes altered following HTLV-1 reactivation in infected cells. We performed over-representation analysis (ORA), using `clusterProfiler` (40) on gene sets identified as differentially expressed

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

between pairs of successive timepoints in infected cells. Over-represented Gene Ontology (GO) biological process terms (41; 42) within levels 5–10 of the GO term hierarchy were kept and further filtered to reduce redundancy, retaining only the lowest FDR-corrected p-value of terms with $\leq 50\%$ overlap (Fig. 3B).

The 0–24-hour interval was characterised by differential expression of genes annotated as being involved in the response to viral gene expression. However, these genes likely reflect a general response to stress, as their trajectory is shared with the uninfected control samples (Fig. 3C). Also over-represented in this interval were genes involved in ribosome biogenesis, ribonucleoprotein complex organisation and RNA catabolism, likely reflecting the initial adaptation to the novel *in vitro* nutritional environment. Viral gene expression is more likely to exert an additional impact on genes associated with these nutrient stress-related GO terms, as their trajectories can be seen to vary from those in uninfected controls responding to *ex vivo* culture alone (Fig. 3C).

DNA replication was triggered early in infected samples, and multiple related terms were enriched in the subsequent 24–48 and 48–96 hour intervals (Fig. 3B). We examined the trajectories of transcripts known to peak at distinct cell cycle stages, obtained from Cyclebase 3.0 (43). An increase in transcripts associated with all cell cycle stages was observed following t_{24}, indicating asynchronous entry into the cell cycle, given the simultaneous upregulation of all stage markers (Fig. S6). It is clear that this inferred entry into the cell cycle is due to the expression of proviral genes, as the trajectories of cell cycle associated terms in uninfected samples are distinct from those in infected samples (Fig. 3C, S6).

Terms associated with transition into the final two timepoints were more heterogeneous, comprising genes involved in synapse organisation and vesicle formation, as well as ion and metabolite transport. Terms related to vesicle formation or transport, and synapse-related terms, may represent horizontal infection, which occurs across a cell-cell junction and may

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

involve vesicle-based transfer of virions (44; 45). Horizontal infection is a key outcome of proviral plus-strand expression. At least some HTLV-1-infected cells must preserve the potential to reactivate, in order to retain the ability to infect a new host. Horizontal transmission can occur during *ex vivo* culture of $CD4^+$ patient cells. The transfer of new virions between cells was evidenced by proviral genome copy numbers, measured by ddPCR of extracted DNA, which surpassed one copy per cell (Fig. S7A–B). The number of reverse-transcribed proviral genomes grew exponentially, the increase first becoming apparent in the 24–48 hour interval (Fig. S7A). In contrast, synapse- and vesicle formation-related terms were enriched exclusively the in the 96–120 and 120–144 hour intervals (Fig. 3B–C). The observed rate of these increases in reverse-transcribed proviral genomes indicates that these terms are not involved in horizontal transmission. Instead, they may be associated with the action of accessory proteins thought to affect cellular ion concentrations, for example p12 (46) or p13 (47; 48).

Imaging of proviral transcripts using single-molecule RNA FISH revealed clustering of signal from both introns and exons of the provirus, consistent with accumulation of whole virus genomes (Fig. S7C). We quantified the fraction of imaged cells showing evidence of unspliced RNA clusters near the cytoplasmic periphery or between cells (Fig. S7C) over time in both patient samples. Since the cells were mixed before mounting on coverslips, it is likely that many cells showing polarisation of viral RNA had lost contact with other cells. Therefore cells with clustered RNA not in contact with other cells were also considered. The fraction of cells showing these RNA clusters increased over time (Fig. S7D), the trajectory resembling the observed increase in reverse-transcribed proviral genome counts (Fig. S7A), further suggesting the horizontal transmission of virions in culture.

We then obtained a list of genes associated with horizontal transmission, compiled by (49), and examined their trajectories in infected and uninfected samples (Fig. S7E). In light of the timing of the PVL increase observed in reactivating cells (Fig. S7A), it is likely that terms

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

which showed increased expression late during reactivation, such as *FSCN-1*, *VCAM-1*, *SDC-1*, *CCL22* and collagen had little impact upon the infection process in *ex vivo* patient samples; their upregulation was not necessary for the transfer of virions into target cells. Conversely, genes that showed substantial changes in expression before the 48-hour timepoint might play a more central role. Additional interesting patterns were observed in the trajectories of individual genes. First, the majority (21/32) of the infection-related gene set tested was significantly differentially expressed in infected cells, suggesting most related processes are triggered by proviral gene expression, rather than by the stress which reactivates the provirus upon *ex vivo* culture. The interaction between ICAM-1 and LFA-1 is important for establishing contact between an infecting cell and its target (50). *ICAM-1* and *LFA-1* were regulated in the opposite sense in infected cells. Opposing expression of these genes might limit horizontal transmission between infected cells, so limiting superinfection (31).

HTLV-1-infected cells and uninfected cells respond similarly to *ex vivo* culture

Multiple pathways influence the degree of proviral reactivation following the stimulation of infected cells; however, without appropriate uninfected controls it is difficult to distinguish between pathways and factors which trigger reactivation and those which merely modulate its intensity. We examined genes significantly differentially expressed in uninfected cells in the 0–24 hour interval (Wald test; BH FDR-corrected $p < 0.05$) to identify pathways affected by the combined stimuli of thawing from -150°C storage and *ex vivo* culture responsible for proviral reactivation. ORA was performed using the Hallmark gene sets from the Molecular Signatures Database (51). Several general stress and nutrient response pathways were highlighted (Fig. 4A), reflecting the introduction into the culture environment and thawing from -150°C storage. c-Myc and mTORC1 signaling are known to coordinate the response to nutrient stress (52; 53), with the former being influenced by the latter (54). The unfolded protein response is similarly

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

triggered by diverse extracellular stimuli or internal changes in metabolite levels and impacts multiple aspects of protein production and secretion (55). NF-κB expression can be activated by a wide variety of stressors (56).

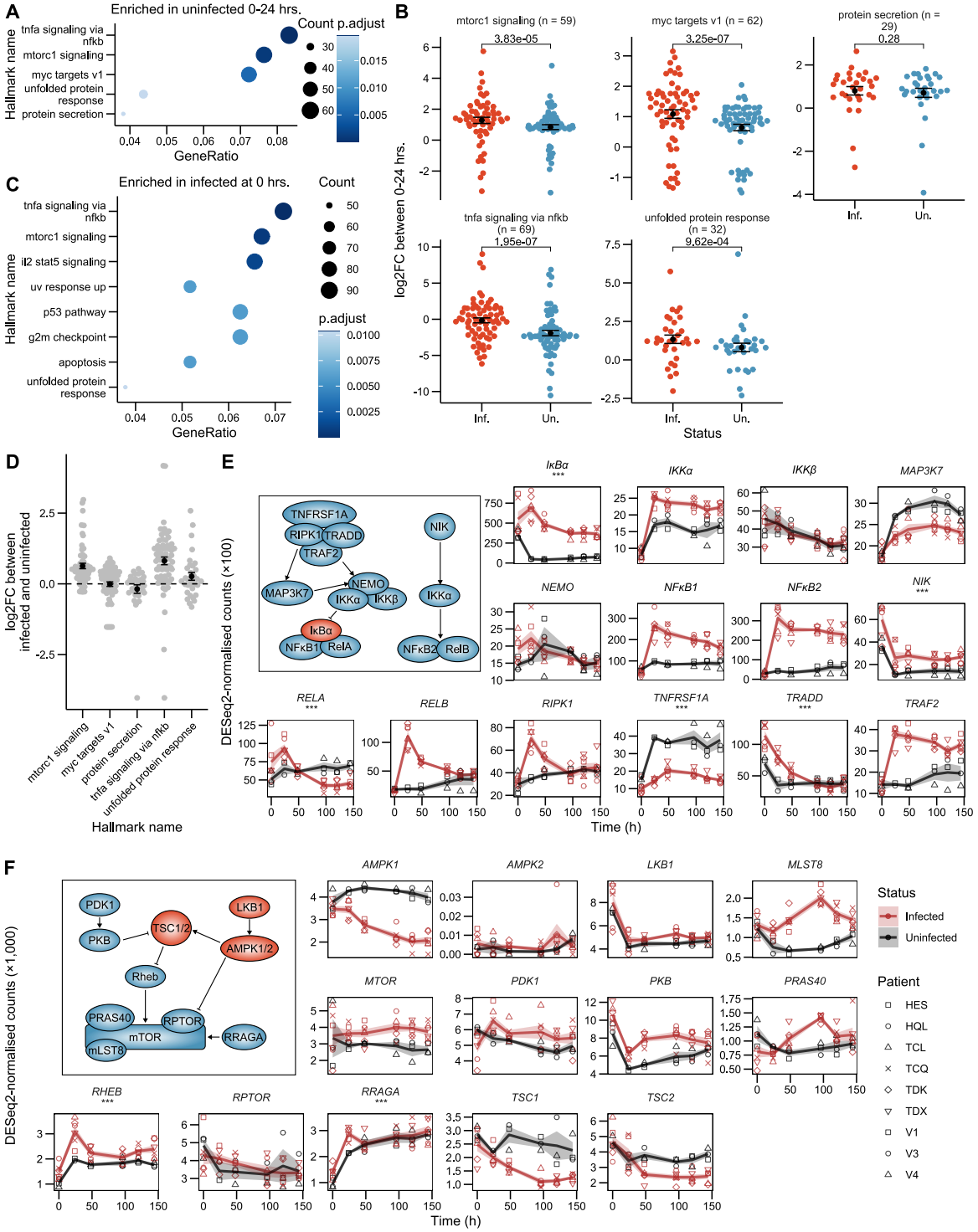

Figure 4: HTLV-1 infection does not alter responsiveness to reactivating factors.

(A) ORA using Hallmark gene sets of genes differentially expressed in uninfected samples during their

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

introduction to *ex vivo* culture (0–24 hrs.). (B) log_2-fold changes during reactivation (0–24 hrs.) in infected and uninfected samples. BH FDR-corrected paired Wilcoxon test p values shown. (C) ORA of genes differentially expressed between infected and uninfected samples at 0 hours. (D) Differential expression of genes within Hallmark sets between infected and uninfected samples at 0 hours. Summary bars represent mean ±SEM. (E–F) Trajectories of key TNFα signaling via NF-κB, and mTORC1 pathway components. Asterisks ('***') indicate genes which are significantly differentially expressed between infected and uninfected samples at 0 hours (Wald test; BH FDR-corrected $p < 0.05$). Summary lines and shaded areas represent mean ±SEM. Insets highlight factors with positive (blue) or inhibitory (red) effect on pathway.

We hypothesised that HTLV-1-infected cells are altered in their ability to react to these stimuli, in order to either regulate proviral reactivation. We compared the log_2-fold changes in genes associated with these Hallmark terms between 0–24 hours in infected and uninfected samples. Infected cells displayed greater differences in the expression of these genes for all terms examined, aside from protein secretion (Fig. 4B). This result is consistent with the known impact of HTLV-1 products on signalling in each of these pathways (57; 58; 59; 60). To distinguish between potential Tax transactivation and increased sensitivity to these stimuli, We examined the differences between infected and uninfected cells at 0 hours, prior to Tax-mediated transactivation. ORA of genes differentially expressed between infected and uninfected samples at t_0 revealed that genes associated with either NF-κB, mTORC1 signalling or the unfolded protein response were already differentially expressed before *ex vivo* culture (Fig. 4C). Indeed, genes in these pathways which are differentially expressed by uninfected samples in response to stress (Fig. 4A), were already upregulated in infected cells at t_0 (Fig. 4D), relative to uninfected samples.

We then examined the levels of key mediators of the NF-κB and mTORC1 pathways to identify potential differences in expression levels which might suggest differential sensitivity

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to reactivation stimuli, as opposed to the pre-activation of these pathways at t_0 (Fig. 4E–F). Multiple genes were differentially expressed in infected relative to uninfected samples, again highlighting the impact of HTLV-1 expression on these pathways. However, in the NF-κB pathway, only 4/11 of these factors were significantly differentially expressed at t_0. Moreover, we observed both positive effects on the pathway (upregulation of *TRADD* and *RELA*) and suppressive effects (downregulation of *TNFRSF1A*, upregulation of $I\kappa B\alpha$), suggesting that infected cells are not differentially sensitive to NF-κB stimuli prior to reactivation (Fig. 4E). Similarly, factors facilitating mTORC1 signaling did not differ between infected and uninfected cells at t_0, with the exception of *RRAGA* and *RHEB*, where increased expression was apparent in the infected cells (Fig. 4F). Therefore, it is unlikely that these pathways are more responsive in infected cells. Instead, it is more likely that there is some degree of pre-activation of the NF-κB and mTORC1 pathways in infected cells responsible for their differential expression between infected and uninfected cells at t_0.

These results indicate that *ex vivo* culture triggers a set of stress-related responses in all samples tested, and HTLV-1-infected cells showed an amplified response to these stimuli. This amplified response is more likely to be the result of synergism with HTLV-1 reactivation than a greater sensitivity to these stimuli, because key mediators of affected pathways were not differentially expressed prior to stimulation.

Histone ubiquitination factors most closely correlate with sense-strand transcription

We postulated that proviral transcription may trigger the expression of host factors that limit the duration of proviral expression. (22) observed that the expression of several components of the non-canonical PRC1 complex, which catalyses ubiquination of H2AK119, correlated with sense-strand expression in clonal HTLV-1-infected T cells. We obtained a set of genes corre-

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

lated with sense-strand transcription by performing Spearman rank correlation tests between the proviral sense-strand and host transcripts, followed by correction for multiple hypothesis testing (Benjamini-Hochberg false discovery rate) (61) (Fig. 5A). We then performed an ORA for enriched GO:BP terms containing the term "histone" within their description. No term was significantly over-represented (ORA hypergeometric test; BH FDR-corrected $p < 0.05$), however, the three top ranked terms related to histone ubiquitination (Fig. 5B, Table S1). The genes within these groups are components of PRC1 complexes (*RYBP*, *RING1B/RNF2*, *BCOR*, *KDM2B*) and other and ubiquitin ligases (Table S2). Over 90% these histone ubiquitination-associated genes were positively correlated with sense-strand expression (Fig. 5C). These results are consistent with the notion that PRC1 contributes to the active termination of the HTLV-1 plus-strand burst. In addition, the majority of these genes were differentially expressed between infected and uninfected cells (Fig. 5D), indicating an association with proviral expression rather than with the response to stress.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

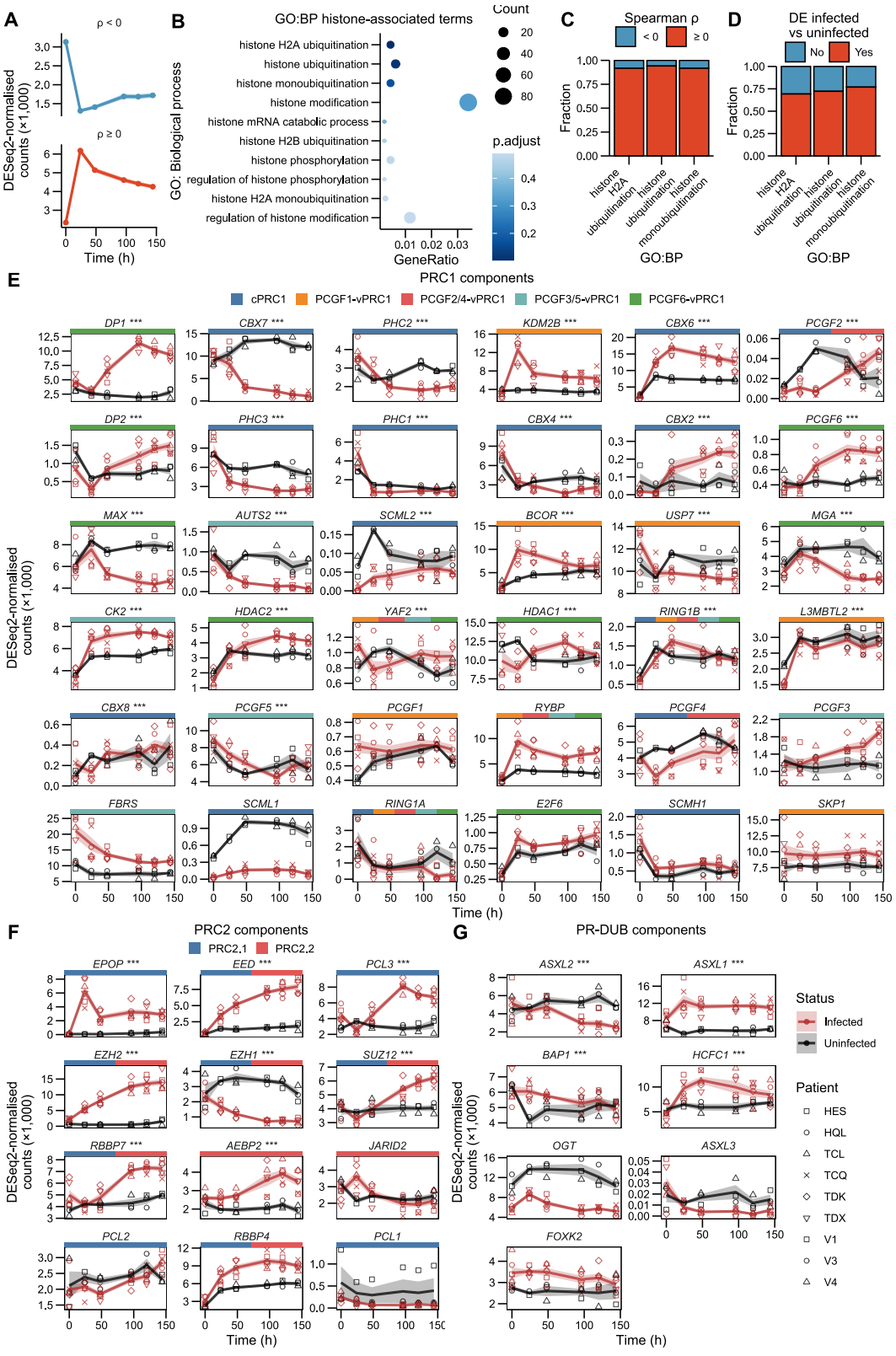

Figure 5: Polycomb effectors are differentially expressed during reactivation.

(A) Mean trajectories of genes significantly (paired Spearman rank correlation test; FDR-corrected $p < 0.05$) correlated with sense-strand transcript levels. Shading represents ±SEM. (B) Top 10 terms

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ranked by p-value. (C) Fraction of genes annotated as histone ubiquitination-associated correlated either positively or negatively with sense-strand expression. (D) Fraction of genes annotated as histone ubiquitination-associated differentially expressed over time between infected and uninfected samples (Likelihood ratio test; BH FDR-corrected $p < 0.05$). (E) Expression trajectories of PRC1 complex components. Asterisks (***) denote significant differential expression between infected and uninfected samples (Likelihood ratio test; BH FDR-corrected $p < 0.05$). Coloured boxes represent PRC complexes comprising each gene (30). Shaded areas represent ±SEM. (F) Trajectories of PRC2 components. (G) Trajectories of PR-DUB complex components and associated factors.

We next examined changes in Polycomb group gene expression induced by proviral reactivation, to identify potential coordination between components of individual complexes which might distinguish a silencing mechanism. Multiple PRC1 complex components were differentially expressed over time between infected samples and uninfected controls (Fig. 5A). Most PCGF6-vPCR1 complex components were upregulated over the culture period, again suggesting that PRC1 contributes to silencing following reactivation, although the HTLV-1 provirus possesses few high-quality matches to E- and T-box motifs (Fig. S8) recognised by the MAX or MGA components of PCGF6-vPRC1 at a subset of its genomic target sites (62). Similarly, PCGF1-vPRC1 components were largely upregulated, potentially guiding the complex to the GC-rich unmethylated regions in the LTRs which are associated with proviral sense-strand expression. Components unique to cPRC1 were mostly downregulated, with the exception of *CBX6* and *CBX2*.

Core PRC2 components (*EED*, *EZH2*, *SUZ12* and *RBBP4/7*) increased over time in culture (Fig. 5B), potentially affecting the maintenance of the H3K27me3 histone mark observed during reactivation (26). Contrary to the other core components, *EZH1* was downregulated. Of the PRC2-variant subcomponents, *EPOP* showed the largest change, with other subcomponents changing relatively little, remaining expressed at some level throughout. PR-DUB deubiquiti-

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

nation complex components were not greatly altered during reactivation: with the exception of *HCFC1*, all were expressed in cells at the point of culture (Fig. 5C). As expected, *ASXL3* levels were expressed at a very low level throughout (Fig. 5C), reflecting their specific expression in neural tissues (63).

Diverse deubiquitinases are involved in HTLV-1 reactivation

H2AK119ub1 levels at the promoter were found influence proviral reactivation through the use of a broad-specificity deubiquitinase (DUB) inhibitor, PR-619 (27). PR-619 has multiple targets, hence the precise DUBs involved in HTLV-1 reactivation are not known. We examined our dataset for evidence of known PR-619 target DUBs being significantly upregulated during reactivation, *i.e.* the t_0–t_{24} interval. A PR-619 target list consisting of 27 genes was obtained from the manufacturer's website (`https://www.selleckchem.com/products/pr-619.html`).Of these PR-619 target genes, 13 were found to be significantly upregulated in infected cells between t_0 and t_{24} (Fig. 6A).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

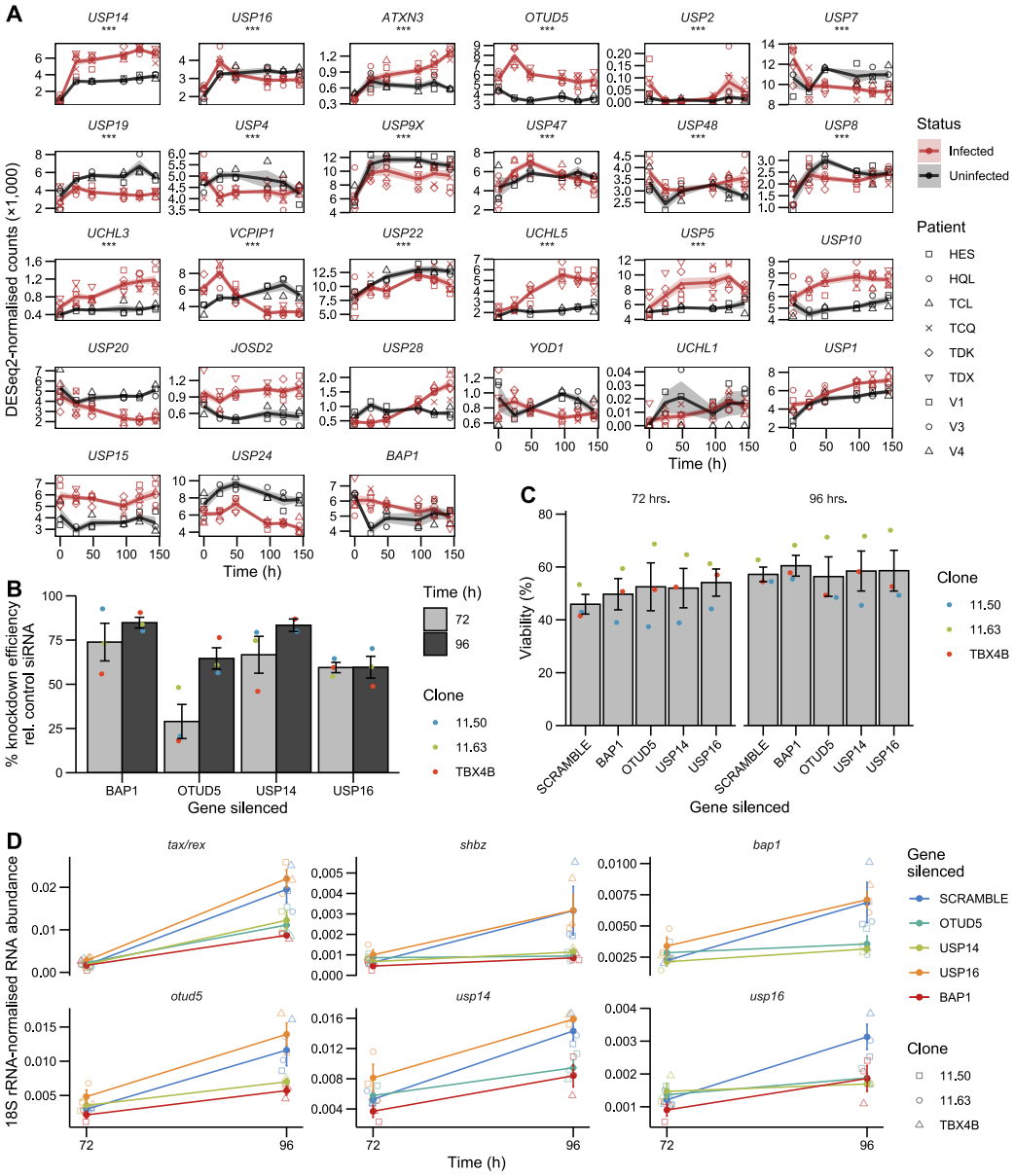

Figure 6: Knockdown of *BAP1*, *OTUD5* or *USP14* reduces proviral transcript levels in infected cells.

(A) Trajectories of PR-619 targets in infected and uninfected cells. Asterisks ('***') denote genes significantly differentially expressed in infected samples between 0–24 hours (Wald test; BH FDR-adjusted $p < 0.05$). Summary lines and shaded area represent mean ±SEM. (B) Knockdown efficiency following incubation with siRNA. Dot colours represent individual infected T cell clones carrying distinct proviral integration sites. Error bars ±SEM. (C) Viability measurements at 72 and 96 hours following the addition of siRNA targeting DUBs. Error bars ±SEM. (D) Changes in HTLV and host cell transcripts following DUB knockdown for 96 hours. Error bars ±SEM. (E) Changes in transcript levels upon DUB

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

inhibition after 72 and 96 hours. Error bars ±SEM.

We selected three DUBs for knockdown using siRNA-induced RNA interference, *USP16*, *USP14* and *OTUD5*. These DUBs were selected due to their high fold change during reactivation (*USP14*), or a stark difference in expression trajectory between infected and uninfected cells (*OTUD5*) or published involvement in H2A deubiquitination (*USP16*) (64; 65). We also targeted *BAP1*, which was not significantly differentially regulated during reactivation (Fig. 6A), but is targeted by PR-619 and is the deubiquitinase present in the PR-DUB complex. RNAi of genes involved in reactivation is impracticable using the PBMC reactivation system, because the time needed to deplete transcript and protein levels exceeds the reactivation timeframe. We therefore incubated three clonal infected T cell lines established from naturally infected cells (31) with siRNA targeting *USP14*, *USP16*, *OTUD5* or *BAP1*. These cells are maintained continuously in culture, and are therefore asynchronous with respect to proviral expression (11).

RNAi reduced target transcript levels, as measured at 72 and 96 hours post siRNA introduction, with mean expression levels falling as low as 15.2% relative to the non-targeting siRNA control (Fig. 6B). The siRNA delivery protocol reduced cell viability; however, individual deubiquitinase knockdown did not appear to reduce viability further. The impact of the siRNA protocol on cell viability is likely due to the omission of serum during incubation with siRNA, since the viability increased within 24 hours of the reintroduction of serum to the culture medium (Fig. 6C). Knockdown of *BAP1*, *OTUD5* or *USP14* resulted in the reduction of *tax/rex* transcript levels, suggesting a role for these genes in reactivation of sense-strand transcription (Fig. 6D). This reduction in plus-strand mRNA abundance was accompanied by similar changes in *shbz* transcript levels.

Knockdown of any one of these three DUBs resulted in a reduction in the levels of the other two, although the siRNA sequences used were specific to each respective protein (Fig. S9).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Knockdown of *USP16*, which is known to be involved in H2AK119ub1 deubiquitination, appeared to have no effect on any of the transcripts examined, although it itself is influenced by both *BAP1*, *OTUD5* and *USP14*. This lack of effect might be due to incomplete depletion, as knockdown efficiency only reached 60%. Although the knockdown efficiency was similar between 72 and 96 hours of siRNA incubation, a reduction in proviral transcripts or non-target DUBs only became apparent at 96 h (Fig. 6D). Serum was reintroduced at 72 hours, following t_{72} RNA sample collection, according to the manufacturer's instructions. It is likely that serum addition triggered synchronous reactivation of some cells, which then made the effects of DUB depletion evident, although protein levels might have been effectively depleted only at 96 hours. Regardless, these results confirm previous findings regarding the importance of deubiquitination for HTLV-1 reactivation and implicate specific deubiquitinases in the process.

Discussion

HTLV-1 is persistently expressed at the cell population level *in vivo*, as evidenced by the persistent activation of α-HTLV-1 CTLs in infected individuals (8). However, only a small fraction of infected cells in peripheral blood express sense-strand transcripts at a given time (9; 12). Proviral expression can be induced by *ex vivo* culture (66; 15; 16). Here, we have examined host transcriptome changes over six days in $CD4^+$ T cells from individuals with polyclonal, non-malignant HTLV-1 infections, comparing it to the behaviour of uninfected $CD4^+$ cells presented with identical stimuli.

Infected cells displayed a distinct, temporally conserved response to *ex vivo* culture. The processes distinguishing these timepoints are key to proviral survival, and include the modification of host cell signalling pathways, the initiation of horizontal infection, and entry into the cell cycle. Infected cells showed changes in gene expression consistent with cell cycle entry following

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

their introduction to *ex vivo* culture. In contrast, uninfected samples did not undergo the same changes, as evidenced by their trajectories of cell-cycle related terms (Fig. S6). The differential response to the culture and thawing stimuli provides further support for the maintenance of PVL through increased proliferation, driven by proviral gene expression (67; 68; 69; 70). The timing and ultimate outcome of cell cycle entry might vary substantially between infected cells, given the observed simultaneous upregulation of markers of distinct cell cycle stages. Indeed, it has been suggested that the intensity of sense-strand expression determines the fate of the HTLV-1-infected cell, with low expression levels resulting in proliferation and intense expression resulting in senescence and death (71).

We observed an increase in the copy number of viral DNA per cell in DNA extracted from patient PBMCs cultured over time, corroborating prior observations of horizontal transmission in patient PBMCs cultured *ex vivo* (44). This increase likely reflects the reverse transcription of vDNA following horizontal transmission of virions into target cells. Examination of several genes thought to be involved in the horizontal infection process confirmed their differential expression in infected samples. The expression trajectories of these genes varied substantially, with some becoming expressed only following the observed increase in reverse transcribed proviral copies. One notable example is *VCAM-1*, which is nearly undetectable at 0 hours and increases appreciably only by ≈ 48 hours. Interestingly, antibody-mediated blocking of VCAM-1 reportedly had little effect on HTLV-1 transmission through EVs, relative to ICAM-1 (45). A similar pattern was observed for *FSCN-1*, which has been shown to influence the efficacy of infection *in vitro* between the ATL cell line MT-2 and Jurkat cells (49). These observations highlight important differences between primary cells and the models used in earlier work, supporting increased use of primary cells wherever possible. We also observed opposite effects of HTLV-1 expression on *ICAM-1* and *LFA-1*, which interact between infected and uninfected cells to promote virological synapse formation (50; 72). Despite recent reports of possible recombination

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

events in the evolutionary history of some HTLV-1 strains (73; 74), the presence of a single integrated provirus in the majority of infected cells (31) suggests HTLV-1 has mechanisms to restrict superinfection (75). The downregulation of LFA-1 during reactivation might reduce the likelihood of superinfection following an encounter between infected cells, although LFA-1 is expressed in infected cells before reactivation at a higher level than in uninfected cells.

Infected cells showed a greater response to stimuli associated with thawing and *ex vivo* culture than did uninfected controls. Since the key regulators of the pathways activated by these stressors were not differentially expressed at t_0, we concluded that this greater response was due to the effects of proviral gene expression, which is known to induce both NF-κB (76; 77) and mTOR signaling (78; 59; 79), as opposed to an altered responsiveness to external stimuli. Nevertheless, it was evident that infected cells displayed some activation of these pathways at t_0, relative to uninfected cells, suggesting chronic activation of these pathways in nonmalignant infected cells.

The proposed mechanism for mTOR activation by HBZ (79) is compatible with activation prior to reactivation, as HBZ is detectable immediately upon isolation of infected cells from patient blood (10; 12). How NF-κB activation is maintained in the absence of constant sense-strand expression is less clear. HBZ expression is thought to antagonise NF-κB activity, through its suppressive effects on RelA (80). In addition, the effects of Tax expression on NF-κB have been reported as transient and correlated with the expression of Tax (22). Indeed, we observed trajectories similar to sense-strand expression for multiple mediators of NF-κB signaling, so it is unclear how transient proviral expression could account for constitutive activation. In ATL cells, NF-κB is chronically activated (81) and mutations affecting the NF-κB pathway are frequently detected (82). Oncogenic driver mutations can be detected in clonally expanded premalignant infected cells up to 10 years before the onset of ATL (83). However, the infections examined here are nonmalignant and polyclonal: somatic mutations are therefore unlikely to

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cause chronic activation of NF-κB signaling in these cell populations.

Before the initiation of *ex vivo* culture and the experimental start point (t_0), cells were extracted from donors, isolated using gradient density centrifugation, frozen, thawed, stained, and sorted: this procedure imposes stress on the cell, which might therefore already start to activate proviral transcription. Timecourse analyses of the dynamics of reactivation of immediate-early stress response genes show that *c-FOS* transcript levels reach a maximum at $\approx$ 15 minutes post-stimulation, followed by a return to basal levels by 60 minutes (84). *c-FOS* is expressed at a high level at the 0-hour timepoint in our dataset (Fig. S10), and is then near zero at all subsequent timepoints. This suggests that experimental t_0 lies at under 60 minutes following the initial stimulus. The difference in mean age between our infected and uninfected controls might also contribute to our result. Ageing has been observed to influence the steady-state expression of stress-pathways, including both mTORC1 (85) and NF-κB (86). However, it is possible that age-related upregulation of these pathways contributes to the increased pathogenesis of HTLV-1 in older individuals.

Proviral sense-strand expression displayed an overshoot, which was also apparent in qRT-PCR using primers against either *tax/rex* or the pX region. This pattern is apparent in several publications, at both the RNA and protein levels (24; 17; 29). The overshoot shape can arise through negative feedback, and does so through precursor transcript depletion in HIV, where Rev removes full-length proviral transcripts before splicing can occur (35). A similar mechanism likely exists in HTLV-1, where the RNA export factor Rex can exert negative feedback upon sense-strand expression (87; 17). This negative feedback has been proposed to lead to oscillations in Tax expression (87), consistent with live imaging of proviral expression using a d2EGFP reporter (21).

Host gene expression, potentially driven by the provirus, may also contribute to the induction or maintenance of a latent state following reactivation. Histone modification levels influence

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

proviral transcription (24; 27), are associated with Tax expression (28) and change during reactivation (26; 27). Genes associated with histone ubiquitination, the removal of which is necessary for proviral reactivation (27), were found to be better represented amongst genes which correlate with sense-strand expression than other histone modifications, indicating a more direct relationship with proviral expression. These findings are consistent with our previous results using naturally-infected T cell clones (22).

We found that HTLV-1 reactivation alters the expression of multiple Polycomb group genes, with several PRC1 components showing trajectories which mirror HTLV-1 expression, whilst core PRC2 components show a general increase over time. Microarray-derived gene expression patterns observed in ATL samples relative to healthy $CD4^+$ T cells have previously shown that multiple PRC2 and a subset of PRC1 components were upregulated in infected samples (88; 89). PRC2 upregulation may be related to the increased EZH2 occupancy detected at the proviral promoter post-reactivation (27). H3K27me3 (deposited by PRC2) levels are also maintained through reactivation (26). It is possible that the maintenance of H3K27me3 by PRC2 during reactivation allows the rapid re-establishment of H2AK119ub1 by cPRC1, in accordance with the canonical model of sequential Polycomb repression (30). Transient induction of immediate-early genes by vascular endothelial cell growth factor in human umbilical vein endothelial cells occurs through the activation of their bivalent (H3K4me3 and H3K27me3 positive [(90)]) promoters whilst H3K27me3 is maintained and EZH2 accumulates at the promoter. cPRC1 then deposits H2AK119ub1, coinciding with the reduction in gene expression (91). The sense-strand promoter shares several functional features with immediate-early genes, including bivalent promoter histone marks (26; 28; 29), independence from *de novo* protein synthesis for reactivation (27) and responsiveness to p38-MAPK signaling (27). Whilst it is tempting to draw parallels between these systems, the immediate-early gene response studied by (91) is far more short-lived ($<$ 4 hrs.), with transcript levels falling by 60 minutes post-stimulation. It is possible

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that Tax-mediated transactivation postpones this decrease, thereby maintaining sense-strand expression.

Transcript trajectories alone are insufficient to infer a mechanism of epigenetic regulation. It will be important to investigate the kinetics of H2AK119ub1 re-establishment at the provirus, and its potential role in active silencing of the provirus. Regardless, the findings of (27) concerning the impact of DUB function on proviral transcription have been confirmed by the DUB RNAi results presented. We found that siRNA knockdown of *USP14*, *OTUD5* or *BAP1* reduced *tax/rex* transcript counts in clonal infected T cells. The importance of H2AK119 deubiquitination to sense-strand transcription has not previously been assayed in infected T cell clones, and suggests that H2AK119ub1 is re-established in cells intermittently expressing the provirus.

Knockdown of *USP14*, *OTUD5* or *BAP1* also resulted in the reduction of *shbz* transcription, in contrast to previous results showing that *shbz* levels were unaffected by 2-hour DUB inhibition whilst *tax/rex* levels were reduced (27). We suggest that the longer period (24 hours) of reduced *tax/rex* levels in our data is responsible for this discrepancy and that the reduction in *tax/rex*, and hence Tax transactivation, is responsible for the observed reduction in *shbz* transcripts.

Whilst *BAP1* is a component of the PR-DUB complex, and its knockdown is known to induce the spread of H2AK119ub1 (92), *OTUD5* and *USP14* may influence proviral transcription through distinct mechanisms. USP14 is recognised as a proteasome-associated DUB, its inhibition increasing proteasome activity (93). Proteasomal function is key for the processing of both p105 and p100 during NF-κB activation (94) and USP14 inhibition or knockdown has specifically been implicated in the suppression of NF-κB signaling (95; 96). OTUD5 has been associated with stabilisation of chromatin modifiers, increasing chromatin accessibility during embryonic development (97). However, knockdown of either *USP14*, *OTUD5* or *BAP1* resulted in a reduction of the other DUBs tested, including *USP16*. All three enzymes were differentially expressed in infected cells, showing increased expression in the 0–24 hour interval relative to

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

uninfected controls. We therefore suggest that this joint regulation is caused by the disruption of proviral transcription following knockdown of any one of these proteins.

The data presented here demonstrate key changes in host gene expression following proviral reactivation in primary PBMCs from HTLV-1 carriers. These changes are contrasted to the transcriptomes of uninfected control samples faced with identical stimuli. The results demonstrate the induction of genes involved in the increased replication of infected cells in response to stimulation, the induction of genes involved in horizontal transmission, and chronic activation of core signalling pathways NF-κB and mTORC1 in infected cells. The relationship between PcG gene expression and HTLV reactivation was investigated, showing differential expression of multiple components of PRC1 and PRC2 in infected samples, which might contribute to proviral latency. Finally, DUBs *USP14*, *OTUD5* and *BAP1* were found to influence proviral reactivation, potentially through distinct mechanisms. Whilst a broad range of processes have been investigated here, this analysis represents only a small subset of the total analyses which could be performed using this dataset. To facilitate its further use and allow for rapid browsing of specific transcript trajectories, a browser-compatible application has been launched at `https://aenaristodemou.shinyapps.io/htlv_reactivation_trajectories/`.

Acknowledgements

We thank Parisa Amjadi from the CL3 Cell Sorting Facility at The Centre for Immunology and Vaccinology at Imperial College London. We thank Theodosios Kyriakou and the Oxford Genomics Centre for library preparation and RNA Sequencing. We thank Rob Klose (University of Oxford) for a fruitful discussion regarding Polycomb group proteins, Helen Kiik and Anat Melamed (Imperial College London) for bioinformatics advice. We thank Imperial College Research Computing Service for use of the high-performance computing cluster

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(https://doi.org/10.14469/hpc/2232). A.E.N.A. is funded by a Wellcome Trust UK 4-year PhD studentship. C.R.M.B. is funded by a Wellcome Trust UK investigator award. D.S.R. is supported by a core grant of the MRC-London Institute of Medical Sciences (UKRI MC-A658-5TY10), a Wellcome Trust Collaborative Grant (206292/Z/17/Z).

Author contributions

Conceptualization: A.E.N.A. and C.R.M.B. Data curation: A.E.N.A. Software: A.E.N.A. Formal analysis: A.E.N.A. Validation: A.E.N.A. Investigation: A.E.N.A. Visualization: A.E.N.A. Methodology: A.E.N.A. Writing—original draft: A.E.N.A. and C.R.M.B. Project administration: A.E.N.A. and C.R.M.B. Resources: A.E.N.A., D.S.R., G.P.T. and C.R.M.B. Supervision: C.R.M.B. Funding acquisition: A.E.N.A. and C.R.M.B. Writing—review and editing: A.E.N.A., D.S.R., G.P.T. and C.R.M.B.

Materials & methods

Patient samples

Blood samples were donated by HTLV-1 patients to the Communicable Diseases Research Tissue Bank, approved by the UK National Research Ethics Service (Communicable Diseases Research Tissue Bank ref.: 20/SC/0226). Written informed consent was obtained from all donors, in accordance with the Declaration of Helsinki. Healthy volunteers consented to sample collection and genomic analysis.

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples by density gradient centrifugation, using Histopaque-1077 Hybri-Max (Merck, Cat. No. H8889).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Cells were then frozen in fetal calf serum (FCS) (ThermoFisher Scientific, Cat. No. 10500064) 10% DMSO (v/v) and stored at -150°C until use.

Table 1: HTLV-1$^+$ subjects and healthy volunteers.

Sample ID	Sex	Age	Infection status	Disease status
HES	F	77	Infected	Asymptomatic carrier
HQL	F	63	Infected	Asymptomatic carrier
TCL	F	65	Infected	HAM/TSP
TCQ	M	79	Infected	HAM/TSP
TDK	M	74	Infected	HAM/TSP
TDX	F	65	Infected	HAM/TSP
V1	F	33	Uninfected	NA
V3	M	24	Uninfected	NA
V4	F	25	Uninfected	NA

Fluorescence-activated cell sorting of infected and uninfected samples

Frozen PBMC samples were thawed by the addition of room-temperature (RT) PBS, and cells were subsequently washed in PBS. Cell counts and viability were measured using the LUNA-FL dual fluorescence cell counter (Logos Biosystems, Cat. No. L20001). CD4$^+$ T cells were enriched using the EasySep Human CD4$^+$ T Cell Enrichment Kit (STEMCELL Technologies, Cat. No. 19052) according to the manufacturer's instructions. Cells were then stained using a viability dye (ThermoFisher Scientific, Cat. No. L10119) and subsequently with biotinylated α-CADM1 antibody (MBL, Cat. No. CM004-6) and streptavidin-conjugated secondary antibody (BioLegend, Cat. No. 405226). Cells were resuspended in 4°C PBS 2% FCS before cell sorting.

CADM1^{high} infected cells were sorted based on the CADM1 signal intensity of uninfected

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cells. Uninfected controls were sorted solely based on size, granularity, forward-scatter height to area ratio (singlet gating) and viability. Sorting was performed at 4°C into 5 ml polystyrene FACS tubes (Corning, Cat. No. 352054) pre-filled with 200 μl FCS.

ddPCR for PVL measurement

Droplet digital PCR (ddPCR) was used to quantify proviral load, as described in **(author?)** (34). DNA was extracted from cells at t_0, using a DNeasy Blood and Tissue kit (Qiagen, Cat. No. 69504) and subsequently quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). Where possible, genomic DNA dilutions of 5 and 2.5 ng/μl were prepared. Otherwise, gDNA was used at the extracted concentration, along with an additional dilution. Percentage proviral load was calculated as: (HTLV pX copy number) / (Albumin copy number / 2) $\cdot$ 100.

Table 2: ddPCR primer and probe sequences (34).

ID	**Sequence (5′ to 3′)**
tax FWD	CGGATACCCAGTCTACGTGTT
tax REV	CAGTAGGGCGTGACGATGTA
tax PROBE	/56-FAM/CTGTGTACA/ZEN/AGGCGACTGCC/3IABkFQ/
alb FWD	TGCATGAGAAAACGCCAGTAA
alb REV	ATGGTCGCCTGTTCACCAA
alb PROBE	/5HEX/TGACAGAGT/ZEN/CACCAAATGCTGCACAGAA/3lABkFQ/

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 3: Thermal cycler programme used for ddPCR.

PCR step	Temperature (°C)	Duration (min)	# cycles
Initial activation	95	10	1
Denaturation	94	0.5	40
Annealing/extension	58	1	40
Enzyme deactivation	98	10	1
Hold	4	∞	1

Cell culture

Sorted cells were cultured in 2 ml RPMI 5.5 mM glucose (1:1 mixture of Cat. No. 11875093 and Cat. No. 11879020 [Thermo-Fisher Scientific]) supplemented with 20% FCS (v/v), Penicillin/Streptomycin (100 U/ml and 100 μg/ml, Thermo Fisher Scientific, Cat. No. 15140122) and 10 μM raltegravir (Selleck Chemicals, MK-0518), in polystyrene round bottom 5 ml tubes (Corning, Cat. No. 352054). Cells were harvested at 0, 24, 48, 96, 120 and 144 hours, for viability and density measurements and RNA extraction.

RNA extraction

Cultured cells were washed once in 10 ml PBS before RNA extraction using the RNeasy Plus Micro kit (Qiagen, Cat. No. 74034). RNA was eluted in 30 μl RNase free H_2O and stored at -80°C until use.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

qRT-PCR

3 μl (10%) of extracted RNA was used for preliminary analysis of proviral transcription using qRT-PCR. cDNA was generated using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Cat. No. 04379012001), following the manufacturer's instructions. Random nucleotide hexamers were used to prime the reverse transcription reaction.

qPCR of cDNA was performed in 384-well plates (Thermo Fisher Scientific, Cat. No. 4309849), using Fast SYBR Green Master Mix reagent (Thermo Fisher Scientific, Cat. No. 4385612) for amplification detection. A ViiA 7 Real-Time PCR System (Thermo Fisher Scientific) was used to amplify and quantify cDNA. The relevant cycling protocol is shown in Table 4.

Table 4: Thermal cycler protocol for qPCR.

Step	**Temperature (°C)**	**Duration**
Hold 1	50	2'
Hold 2	95	10'
Melt (45×)	95	15"
Elongation (45×)	60	1'
Melt curve 1	95	15"
Melt curve 2	60	1'
Dissociation	95	15"

Prior to relative target quantification, background signal threshold and primer efficiency values (E) were calculated for each primer target using LinRegPCR (98). Target RNA quantity (R_0) was calculated as $Threshold/E^{C_q}$, where C_q is the quantification cycle where a reaction's fluorescent signal intensity crosses the background threshold. R_0 values for each target tran-

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

617 script were then normalised relative to 18S rRNA levels.

Table 5: Primers used for qRT-PCR (25).

ID	Sequence FWD (5′ to 3′)	Sequence REV (5′ to 3′)
tax/rex	CCGGCGCTGCTCTCATCCCGGT	GGCCGAACATAGTCCCCCAGAG
gag	CAGAGGAAGATGCCCTCCTATT	GTCAACCTGGGCTTTAATTACG
pX	CTCCTTCCGTTCCACTCAAC	GTGGTAGGCCTTGGTTTGAA
18S rRNA	GTAACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG

618 ## Sequencing

619 Prior to sequencing, extracted RNA quality and concentration were measured using the Bio-
620 analyser RNA Pico kit (Agilent, Cat. No. 5067-1513). Library preparation and sequencing
621 were performed by the Oxford Genomics Centre (`https://www.well.ox.ac.uk/ogc/`).
622 PolyA-enriched cDNA libraries were prepared. 150 bp paired-end fragments were sequenced
623 using the NovaSeq6000 sequencing platform (Illumina). Reads for each sample were collected
624 over two individual sequencing runs.

625 ## Alignment

626 ### STAR alignment

627 An alignment to visualise reads aligning to individual parts of the HTLV provirus was performed
628 using STAR v2.7.9a (36). Reads were aligned to the GRCh38 assembly of the human genome,
629 with HTLV-1 genome sequence AB513134 (99) appended as an additional chromosome. A
630 GTF file containing genomic feature information for the human genome (ref. GRCh38) was
631 obtained from the Ensembl database and the coordinates for HTLV-1, excluding the LTRs,

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

were appended. Reads cannot be unambiguously mapped to either LTR due to their near-identical sequence, hence these regions were excluded.

Kallisto pseudo-alignment

Transcript counts were estimated using Kallisto version 0.46.2 (38). The transcript index was generated from cDNA sequences for GRCh38 obtained from Ensembl. The entire length of the provirus (AB513134), excluding the LTRs, in the sense and antisense orientations were used as transcript sequences to quantify HTLV sense and antisense transcripts respectively.

Differential expression analysis

Transcript differential expression analysis was performed on Kallisto transcript counts. Linear models for comparison of sample transcript count matrices were generated using the `DESeq2` package (39) for R (100). Models were generated iteratively: first incorporating differences between time in culture, then adding the effect of HTLV-1 infection and, finally, adding the interaction term between infection status and time in culture. Likelihood Ratio Tests were used to quantify the number of genes identified as significant by more complex models.

To obtain sets genes of genes differentially expressed between sequential timepoints in infected cells, `DESeq2` models were generated using the same data, switching the intercept to allow for specific pairwise comparisons. Global p-value correction was then used to correct for the multiple pairwise comparisons performed using the Benjamini-Hochberg false discovery rate (BH-FDR) (61).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Identification of sense-strand correlated genes

Sense-strand expression correlates were obtained by performing pairwise Spearman rank correlation tests between sense-strand transcript counts and all other identified transcripts. p-value correction for multiple hypothesis testing was subsequently performed using BH-FDR.

Over-Representation Analyses

Over-representation analyses (ORA) were performed using the `clusterProfiler` package for R (40). Ensembl gene identifiers were converted to Entrez IDs using `biomaRt` (101; 102). The `clusterProfiler::enrichGO` function was used in conjunction with the org.Hs.eg.db database (103) to perform ORA against the biological process clade of the Gene Ontology (GO) (41; 42). A cutoff value of 0.05 for FDR-corrected p-values was set for significant results. A gene background of all the genes quantified in the Kallisto pseudoalignment was used. For the ORA performed in figure 3, GO terms in levels < 5 and > 10 were removed using the `clusterProfiler::dropGO` function. Remaining redundant terms were removed using the `clusterProfiler::simplify` function with a 50% overlap threshold, retaining the term with the lowest adjusted p-value.

Horizontal infection culture

Patients' samples were thawed from -150°C storage, as described above, and cultured at 37°C, 5% CO_2 in RPMI 5.5 mM glucose supplemented with 20% FCS (v/v), Penicillin/Streptomycin (100 U/ml and 100 μg/ml). DNA was extracted as for other ddPCR samples, and quantification was performed following the protocol described by **(author?)** (34).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 6: HTLV-1 patient samples used for timecourse PVL measurements.

Sample ID	Sex	Age	Infection status	Disease status
TCQ	M	77	Infected	HAM/TSP
TDT	F	47	Infected	HAM/TSP
TET	F	56	Infected	HAM/TSP
V3	M	24	Uninfected	NA

Clonal infected T cell culture

Clones 11.50, 11.63 and TBX4B were cultured in RPMI-1640 supplemented with Penicillin/Streptomycin (100 U/ml and 100 μg/ml), L-Glutamine and 20% FCS (v/v), 100 U/ml IL-2 (Miltenyi Biotec, Cat. No. 130-097-745) and 10 μM raltegravir (Selleck Chemicals, MK-0518) at 5% CO_2, 37°C. Transduced lines were maintained in 1 μg/ml puromycin (Thermo-Fisher Scientific, Cat. No. A1113803). Replacement of half the culture media with fresh media was performed 2–3 times per week.

RNA interference

Transcript knockdown was performed using Dharmacon Accell SMARTpool siRNA (Horizon Discovery) which incorporates four siRNA sequences for a target to knockdown gene expression. $2 \cdot 10^5$ cells were resuspended in 200 μl Accell delivery media supplemented with 100 U/ml IL-2, 10 μM raltegravir and 1 μM of the relevant SMARTpool siRNA in a V-bottom 96-well plate. Cells were incubated at 37°C, 5% CO_2 for 72 hours before lysis and RNA extraction using the RNeasy Micro Plus extraction kit (Qiagen, Cat. No. 74034). A further technical replicate was resuspended in RPMI supplemented with 20% FCS, 100 U/ml IL-2 and 10 μM raltegravir, cultured for a further 24 hours before lysis and RNA extraction.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

siRNA sequence alignment was performed using Clustal Omega (104; 105) and visualised using Jalview v2.11.2.2 (106).

Table 7: Primers used to examine candidate deubiquitinase expression levels.

Primers were designed against the "Ensembl Canonical" transcript variant for each gene. Annealing sites matching exon-exon splicing points were selected.

ID	FWD (5′ to 3′)	REV (5′ to 3′)
ATXN3	CCATCTTCCACGAGAAACAAGA	GTAACTCCTCCTTCTGCCATTC
OTUD5	GCCGACTACTTCTCCAACTATG	GATGGGTTCTGTGCTGTACT
BAP1	AAATACTCACCCAAGGAGCTG	CTGAGCCAGCATGGAGATAAA
USP14	GTGGATTGACAAACCTTGGTAAC	CAAATCTCTAAGGGCTGCAGTA
USP5	GGGAAACAGTATGTGGAGAGAC	TCCGCCTTCAACACCAATAG
UCHL5	GCAGTAAGGCCTGTCATAGAA	GATCTGTATCCATGGGTTCCTC
UCHL3	CCTGAAGAACGAGCCAGATAC	TTCCGCCCATCTAATTCATAGAG
USP22	GGCCACTACACCAGCTTTAT	TGATAGAACAGCAAGTACCCTTC
USP9X	GCTGAACGAATGGCAGAATG	CCTGTGCTGCCCAGATATTA
USP8	TACGATGGCAGGTGGAAAC	GCCCACCGTAGTGATTTGA
USP16	CCTCACTGTCTGGTTCTTAGTTT	CATTATCTTTCTCTGCTGGCTTTG
USP47	GCATTTGCTAGTGTGGAAGAAG	AGCTGTAAGGTCAGCAGATAAG
USP19	GAGGAGGAAGAGGAAGAGAAGA	CAAAGGAGCGGTCATGGAA
VCPIP1	AACCATGGGTATGGCTGATG	GAACCTGACTTACAGCCTCTTT

Motif search

To identify partial matches to E- and T-box motifs in the HTLV-1 sequence (AB513134), FIMO (107) was used to identify matches at a significance threshold of 0.001. E- and T-box motifs

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

692 used (MA0058.3 and MA0690.2 respectively) were obtained from JASPAR (108).

693 Statistical analyses and plotting

694 Statistical analyses were performed using R statistical computing software (100). Data were
695 plotted using `ggplot2` (109), with the exception of flow cytometry scatter and contour plots,
696 which were generated using FlowJo v10.8.1 (BD Life Sciences).

697 Sample collection, cell sorting and incubation

698 Please refer to sections describing sample collection, preparation for cell sorting and culture of
699 patient PBMCs in section : Materials & methods.

Table 8: HTLV-1 patient sample information.

Sample ID	Sex	Age	HTLV-1 infection status	Disease status
HHL	F	34	Infected	Asymptomatic carrier
P201890	F	64	Infected	Asymptomatic carrier

700 Coverslip preparation

701 Twelve millimeter diameter circular No. 1 coverslips (Scientific Laboratory Supplies, Cat. No.
702 MIC3304) were washed in 100% EtOH and then placed inside individual wells of 12-well plates
703 (Corning, Cat. No. 353043) to dry. After drying, coverslips were covered with 100 μl of 0.1%
704 (w/v) poly-L-lysine solution (Merck, Cat. No. P8920) for 10 minutes at RT. Coverslips were
705 then washed twice with 1 ml H_2O and left to dry for $\approx$ 30 minutes. Cells were washed 2$\times$ in
706 PBS and layered onto coverslips, in PBS ($\approx$ 1250 cells/μl), and left to sediment for 10 minutes,
707 RT. Two 1 ml PBS washes were then performed to remove unadhered cells. Cells were fixed

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for 1 minute in 1 ml PBS 2% formaldehyde (Thermo-Fisher Scientific, Cat. No. 28906) before being fixed for a further 14 minutes in fresh PBS 2% formaldehyde. Following two PBS washes, cells were permeabilised with 4°C 70% EtOH and stored at -20°C until imaging.

Fluorescent probe hybridisation

Coverslips were removed from EtOH and excess liquid removed by blotting on filter paper. Coverslips were subsequently submerged in wash buffer A (10% [v/v] de-ionised formamide, 20% [v/v] Stellaris wash buffer A [LGC Biosearch Technologies, Cat. No. SMF-WA1-60], 70% H_2O) for 10–15 minutes. The Stellaris RNA probe mixture (Table S3) was prepared by adding 1:100 of the relevant probes (final concentration 40 μM) to hybridisation buffer (10% [v/v] de-ionised formamide, 90% [v/v] Stellaris hybridisation buffer [Cat. No. SMF-HB1-10]). Probe mixture was pipetted onto a hybridisation cover (Thermo-Fisher Scientific, Cat. No. H18200) resting on a glass microscope slide (Thermo-Fisher Scientific, Cat. No. J2800AMNZ). Coverslips were removed from wash buffer A, blotted and placed onto the probe mix, cells facing down. Coverslips were then covered in rubber cement (Marabu, Cat. No. 29010017000), to prevent evaporation, and incubated overnight at 37°C, protected from light. Following hybridisation, coverslips were incubated in wash buffer A for 30 minutes at 37°C, 60 RPM. Coverslips were then transferred into fresh wash buffer A supplemented with 5 ng/ml DAPI and incubated for 30 minutes at 37°C, 60 RPM. After 5-minute incubation in Stellaris wash buffer B (Cat. No. SMF-WB1-20) at room temperature, cells were blotted and mounted onto glass slides in Vectashield antifade mounting medium (Vector Laboratories, Cat. No. H-1000-10). Coverslip edges were sealed with transparent nail polish and left to set for 10 minutes prior to imaging.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

smFISH probes

Novel Stellaris smFISH probes targeting proviral transcripts were designed using the manufacturer's browser-based design tools (`https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer`) (Table S3).

Microscopy

Imaging was performed on an Olympus IX70 inverted fluorescence microscope (Olympus Corporation) fitted with an ORCA-Flash 4.0 V2 digital CMOS camera (Hamamatsu Photonics), using the 100× objective. Micromanager 2.0 software (110) was used to coordinate the acquisition of z-stacks in four channels (Table 9). 54 z-stack slices for each field of view were obtained at 300 nm intervals.

Table 9: Channels used in acquisition of smFISH images.

Channel	Excitation λ (nm)	Emission λ (nm)	Exposure time (ms)
DAPI	390/18	457/20	100
Autofluorescence	475/28	514/30	500
Quasar 570	575/25	632/60	500
Quasar 670	632/22	692/40	500

Image processing

Images were first corrected for uneven illumination using channel-specific illumination functions as used by (11) and (12). Excess image slices ($\leq$4) at either end of the z-stack were removed in parallel, based on the focal centre of the stack, where the standard deviation of intensity is greatest.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

[1] Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proceedings of the National Academy of Sciences. 1980;77(12):7415. doi:10.1073/pnas.77.12.7415.

[2] Kaplan JE, Osame M, Kubota H, Igata A, Nishitani H, Maeda Y, et al. The risk of development of HTLV-I-associated myelopathy/tropical spastic paraparesis among persons infected with HTLV-I. Journal of acquired immune deficiency syndromes. 1990;3(11):1096–1101.

[3] Maloney EM, Cleghorn FR, Morgan O, Rodgers-Johnson P, Cranston B, Jack N, et al. Incidence of HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) in Jamaica and Trinidad. Journal of acquired immune deficiency syndromes and human retrovirology: official publication of the International Retrovirology Association. 1998;17(2):167–170.

[4] Arisawa K, Soda M, Endo S, Kurokawa K, Katamine S, Shimokawa I, et al. Evaluation of adult T-cell leukemia/lymphoma incidence and its impact on non-Hodgkin lymphoma incidence in southwestern Japan. International journal of cancer. 2000;85(3):319–324.

[5] Nagai M, Usuku K, Matsumoto W, Kodama D, Takenouchi N, Moritoyo T, et al. Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP. Journal of neurovirology. 1998;4(6):586–593.

[6] Jeffery KJM, Usuku K, Hall SE, Matsumoto W, Taylor GP, Procter J, et al. HLA alleles determine Human T-lymphotropic virus-I (HTLV-I) proviral load and the risk

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of HTLV-I-associated myelopathy. Proceedings of the National Academy of Sciences. 1999;96(7):3848–3853.

[7] Iwanaga M, Watanabe T, Utsunomiya A, Okayama A, Uchimaru K, Koh KR, et al. Human T-cell Leukemia Virus type I (HTLV-1) proviral load and disease progression in asymptomatic HTLV-1 carriers: a nationwide prospective study in Japan. Blood. 2010;116(8):1211–1219.

[8] Parker CE, Daenke S, Nightingale S, Bangham CRM. Activated, HTLV-1-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. Virology. 1992;188(2):628–636.

[9] Gessain A, Louie A, Gout O, Gallo RC, Franchini G. Human T-cell leukemia-lymphoma virus type I (HTLV-I) expression in fresh peripheral blood mononuclear cells from patients with tropical spastic paraparesis/HTLV-I-associated myelopathy. Journal of virology. 1991;65(3):1628–1633.

[10] Satou Y, Yasunaga JI, Yoshida M, Matsuoka M. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(3):720–725.

[11] Billman MR, Rueda D, Bangham CRM. Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1. Wellcome open research. 2017;2.

[12] Miura M, Dey S, Ramanayake S, Singh A, Rueda DS, Bangham CRM. Kinetics of HTLV-1 reactivation from latency quantified by single-molecule RNA FISH and stochastic modelling. PLOS Pathogens. 2019;15(11):e1008164. doi:10.1371/journal.ppat.1008164.

[13] MacNamara A, Rowan A, Hilburn S, Kadolsky U, Fujiwara H, Suemori K, et al. HLA class I binding of HBZ determines outcome in HTLV-1 infection. PLoS pathogens. 2010;6(9):e1001117.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[14] Sugata K, Yasunaga Ji, Mitobe Y, Miura M, Miyazato P, Kohara M, et al. Protective effect of cytotoxic T lymphocytes targeting HTLV-1 bZIP factor. Blood, The Journal of the American Society of Hematology. 2015;126(9):1095–1105.

[15] Minato Si, Itoyama Y, Goto I, Yamamoto N. Expression of HTLV-I antigen in cultured peripheral blood mononuclear cells from patients with HTLV-I associated myelopathy. Journal of the neurological sciences. 1988;87(2-3):233–244.

[16] Hanon E, Hall S, Taylor GP, Saito M, Davis R, Tanaka Y, et al. Abundant tax protein expression in CD4 T cells infected with Human T-cell Lymphotropic Virus type I (HTLV-I) is prevented by cytotoxic T lymphocytes. Blood. 2000;95(4):1386–1392.

[17] Rende F, Cavallari I, Corradin A, Silic-Benussi M, Toulza F, Toffolo GM, et al. Kinetics and intracellular compartmentalization of HTLV-1 gene expression: nuclear retention of HBZ mRNAs. Blood. 2011;117(18):4855–4859.

[18] Matsuoka M, Yasunaga Ji. Human T-cell Leukemia Virus type 1: replication, proliferation and propagation by Tax and HTLV-1 bZIP factor. Current opinion in virology. 2013;3(6):684–691.

[19] Marriott SJ, Semmes OJ. Impact of HTLV-I Tax on cell cycle progression and the cellular DNA damage repair response. Oncogene. 2005;24(39):5986–5995.

[20] Takahashi M, Higuchi M, Makokha GN, Matsuki H, Yoshita M, Tanaka Y, et al. HTLV-1 Tax oncoprotein stimulates ROS production and apoptosis in T cells by interacting with USP10. Blood, The Journal of the American Society of Hematology. 2013;122(5):715–725.

[21] Mahgoub M, Yasunaga Ji, Iwami S, Nakaoka S, Koizumi Y, Shimura K, et al. Sporadic on/off switching of HTLV-1 Tax expression is crucial to maintain the whole population of virus-induced leukemic cells. Proceedings of the National Academy of Sciences. 2018; p. 201715724.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[22] Kiik H, Ramanayake S, Miura M, Tanaka Y, Melamed A, Bangham CR. Time-course of host cell transcription during the HTLV-1 transcriptional burst. PLoS pathogens. 2022;18(5):e1010387.

[23] Matsuoka M, Mesnard JM. HTLV-1 bZIP factor: the key viral gene for pathogenesis. Retrovirology. 2020;17(1):1–8.

[24] Belrose G, Gross A, Olindo S, Lézin A, Dueymes M, Komla-Soukha I, et al. Effects of valproate on Tax and HBZ expression in HTLV-1 and HAM/TSP T lymphocytes. Blood. 2011;118(9):2483–2491.

[25] Satou Y, Miyazato P, Ishihara K, Yaguchi H, Melamed A, Miura M, et al. The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome. Proceedings of the National Academy of Sciences. 2016;113(11):3054–3059.

[26] Kulkarni A, Mateus M, Thinnes CC, McCullagh JS, Schofield CJ, Taylor GP, et al. Glucose Metabolism and Oxygen Availability Govern Reactivation of the Latent Human Retrovirus HTLV-1. Cell chemical biology. 2017;24(11):1387. e3.

[27] Kulkarni A, Taylor GP, Klose RJ, Schofield CJ, Bangham CRM. Histone H2A monoubiquitylation and p38-MAPKs regulate immediate-early gene-like reactivation of latent retrovirus HTLV-1. JCI Insight. 2018;3(20). doi:10.1172/jci.insight.123196.

[28] Miura M, Miyazato P, Satou Y, Tanaka Y, Bangham CRM. Spontaneous HTLV-1 transcription is accompanied by distinct epigenetic changes in the 5′ and 3′ long terminal repeats. Wellcome open research. 2018;3.

[29] Cheng X, Joseph A, Castro V, Chen-Liaw A, Skidmore Z, Ueno T, et al. Epigenomic regulation of human T-cell leukemia virus by chromatin-insulator CTCF. PLoS pathogens. 2021;17(5):e1009577.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

838 [30] Blackledge NP, Klose RJ. The molecular principles of gene regulation by Polycomb re-
839 pressive complexes. Nature Reviews Molecular Cell Biology. 2021;22(12):815–833.

840 [31] Cook LB, Rowan AG, Melamed A, Taylor GP, Bangham CRM. HTLV-1–infected T cells
841 contain a single integrated provirus in natural infection. Blood. 2012;120(17):3488–3490.

842 [32] Melamed A, Yaguchi H, Miura M, Witkover A, Fitzgerald TW, Birney E, et al. The
843 human leukemia virus HTLV-1 alters the structure and transcription of host chromatin
844 in cis. eLife. 2018;7:e36245. doi:10.7554/eLife.36245.

845 [33] Manivannan K, Rowan AG, Tanaka Y, Taylor GP, Bangham CRM. CADM1/TSLC1
846 identifies HTLV-1-infected cells and determines their susceptibility to CTL-mediated lysis.
847 PLoS pathogens. 2016;12(4):e1005560.

848 [34] Katsuya H, Islam S, Tan BJY, Ito J, Miyazato P, Matsuo M, et al. The
849 Nature of the HTLV-1 Provirus in Naturally Infected Individuals Analyzed
850 by the Viral DNA-Capture-Seq Approach. Cell Reports. 2019;29(3):724–735.e4.
851 doi:https://doi.org/10.1016/j.celrep.2019.09.016.

852 [35] Hansen MM, Wen WY, Ingerman E, Razooky BS, Thompson CE, Dar RD, et al. A post-
853 transcriptional feedback mechanism for noise suppression and fate stabilization. Cell.
854 2018;173(7):1609–1621.

855 [36] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
856 universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

857 [37] Landry S, Halin M, Vargas A, Lemasson I, Mesnard JM, Barbeau B. Upregulation of
858 Human T-cell Leukemia Virus type 1 antisense transcription by the viral tax protein.
859 Journal of virology. 2009;83(4):2048–2054.

860 [38] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quan-
861 tification. Nature biotechnology. 2016;34(5):525–527.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

862 [39] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
863 RNA-seq data with DESeq2. Genome biology. 2014;15(12):1–21.

864 [40] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological
865 themes among gene clusters. Omics: a journal of integrative biology. 2012;16(5):284–287.

866 [41] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology:
867 tool for the unification of biology. Nature genetics. 2000;25(1):25–29.

868 [42] The Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine.
869 Nucleic acids research. 2021;49(D1):D325–D334.

870 [43] Santos A, Wernersson R, Jensen LJ. Cyclebase 3.0: a multi-organism database on cell-
871 cycle regulation and phenotypes. Nucleic acids research. 2015;43(D1):D1140–D1144.

872 [44] Igakura T, Stinchcombe JC, Goon PKC, Taylor GP, Weber JN, Griffiths GM, et al.
873 Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton.
874 Science. 2003;299(5613):1713–1716.

875 [45] Pinto DO, DeMarino C, Pleet ML, Cowen M, Branscome H, Al Sharif S, et al. HTLV-1
876 extracellular vesicles promote cell-to-cell contact. Frontiers in microbiology. 2019; p. 2147.

877 [46] Ding W, Albrecht B, Luo R, Zhang W, Stanley JR, Newbound GC, et al. Endoplasmic
878 reticulum and cis-Golgi localization of human T-lymphotropic virus type 1 p12I: associa-
879 tion with calreticulin and calnexin. Journal of virology. 2001;75(16):7672–7682.

880 [47] D'Agostino DM, Ranzato L, Arrigoni G, Cavallari I, Belleudi F, Torrisi MR, et al. Mi-
881 tochondrial alterations induced by the p13II protein of human T-cell leukemia virus type
882 1: critical role of arginine residues. Journal of Biological Chemistry. 2002;277(37):34424–
883 34433.

884 [48] Silic-Benussi M, Cannizzaro E, Venerando A, Cavallari I, Petronilli V, La Rocca N, et al.
885 Modulation of mitochondrial K+ permeability and reactive oxygen species production by

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the p13 protein of human T-cell leukemia virus type 1. Biochimica et Biophysica Acta (BBA)-Bioenergetics. 2009;1787(7):947–954.

[49] Gross C, Thoma-Kress AK. Molecular mechanisms of HTLV-1 cell-to-cell transmission. Viruses. 2016;8(3):74.

[50] Fukudome K, Furuse M, Fukuhara N, Orita S, Imai T, Takagi S, et al. Strong induction of ICAM-1 in human T cells transformed by human T-cell-leukemia virus type 1 and depression of ICAM-1 or LFA-1 in adult T-cell-leukemia-derived cell lines. International journal of cancer. 1992;52(3):418–427.

[51] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. Cell systems. 2015;1(6):417–425.

[52] Kim J, Guan KL. mTOR as a central hub of nutrient signalling and cell growth. Nature cell biology. 2019;21(1):63–71.

[53] Kress TR, Sabo A, Amati B. MYC: connecting selective transcriptional control to global RNA production. Nature Reviews Cancer. 2015;15(10):593–607.

[54] Csibi A, Lee G, Yoon SO, Tong H, Ilter D, Elia I, et al. The mTORC1/S6K1 pathway regulates glutamine metabolism through the eIF4B-dependent control of c-Myc translation. Current biology. 2014;24(19):2274–2280.

[55] Hetz C, Zhang K, Kaufman RJ. Mechanisms, regulation and functions of the unfolded protein response. Nature reviews Molecular cell biology. 2020;21(8):421–438.

[56] Lingappan K. NF-κB in oxidative stress. Current opinion in toxicology. 2018;7:81–86.

[57] Duyao M, Kessler D, Spicer D, Bartholomew C, Cleveland J, Siekevitz M, et al. Transactivation of the c-myc promoter by human T cell leukemia virus type 1 tax is mediated by NF kappa B. Journal of Biological Chemistry. 1992;267(23):16288–16291.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[58] Ku SC, Lee J, Lau J, Gurumurthy M, Ng R, Lwa SH, et al. XBP-1, a novel human T-lymphotropic virus type 1 (HTLV-1) tax binding protein, activates HTLV-1 basal and tax-activated transcription. Journal of virology. 2008;82(9):4343–4353.

[59] Yoshita M, Higuchi M, Takahashi M, Oie M, Tanaka Y, Fujii M. Activation of mTOR by human T-cell leukemia virus type 1 Tax is important for the transformation of mouse T cells to interleukin-2-independent growth. Cancer Science. 2012;103(2):369–374.

[60] Fochi S, Mutascio S, Bertazzoni U, Zipeto D, Romanelli MG. HTLV deregulation of the NF-κB pathway: An update on tax and antisense proteins role. Frontiers in microbiology. 2018;9:285.

[61] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 1995;57(1):289–300.

[62] Endoh M, Endo TA, Shinga J, Hayashi K, Farcas A, Ma KW, et al. PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. Elife. 2017;6:e21064.

[63] Shukla V, Rao M, Zhang H, Beers J, Wangsa D, Wangsa D, et al. ASXL3 is a novel pluripotency factor in human respiratory epithelial cells and a potential therapeutic target in small cell lung cancer. Cancer research. 2017;77(22):6267–6281.

[64] Yang W, Lee YH, Jones AE, Woolnough JL, Zhou D, Dai Q, et al. The histone H2A deubiquitinase Usp16 regulates embryonic stem cell gene expression and lineage commitment. Nature communications. 2014;5(1):1–15.

[65] Gu Y, Jones AE, Yang W, Liu S, Dai Q, Liu Y, et al. The histone H2A deubiquitinase Usp16 regulates hematopoiesis and hematopoietic stem cell function. Proceedings of the National Academy of Sciences. 2016;113(1):E51–E60.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

933 [66] Tochikura T, Iwahashi M, Matsumoto T, Koyanagi Y, Hinuma Y, Yamamoto N. Effect
934 of human serum anti-HTLV antibodies on viral antigen induction in in vitro cultured
935 peripheral lymphocytes from adult T-cell leukemia patients and healthy virus carriers.
936 International journal of cancer. 1985;36(1):1–7.

937 [67] Furukawa Y, Fujisawa J, Osame M, Toita M, Sonoda S, Kubota R, et al. Fre-
938 quent clonal proliferation of human T-cell leukemia virus type 1 (HTLV-1)-infected
939 T cells in HTLV-1-associated myelopathy (HAM-TSP). Blood. 1992;80(4):1012–1016.
940 doi:10.1182/blood.V80.4.1012.1012.

941 [68] Cavrois M, Gessain A, Wain-Hobson S, Wattel E. Proliferation of HTLV-1 infected circu-
942 lating cells in vivo in all asymptomatic carriers and patients with TSP/HAM. Oncogene.
943 1996;12(11):2419–2423.

944 [69] Asquith B, Zhang Y, Mosley AJ, de Lara CM, Wallace DL, Worth A, et al. In vivo
945 T lymphocyte dynamics in humans and the impact of human T-lymphotropic virus 1
946 infection. Proceedings of the National Academy of Sciences. 2007;104(19):8035–8040.

947 [70] Bangham CR, Miura M, Kulkarni A, Matsuoka M. Regulation of latency in the human
948 T cell leukemia virus, HTLV-1. Annual Review of Virology. 2019;6:365–385.

949 [71] Philip S, Zahoor MA, Zhi H, Ho YK, Giam CZ. Regulation of Human T-Lymphotropic
950 Virus Type I Latency and Reactivation by HBZ and Rex. PLoS Pathogens.
951 2014;10(4):e1004040. doi:10.1371/journal.ppat.1004040.

952 [72] Barnard AL, Igakura T, Tanaka Y, Taylor GP, Bangham CR. Engagement of specific T-
953 cell surface molecules regulates cytoskeletal polarization in HTLV-1–infected lymphocytes.
954 Blood. 2005;106(3):988–995.

955 [73] Desrames A, Cassar O, Gout O, Hermine O, Taylor GP, Afonso PV, et al. Northern

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

956 African strains of human T-lymphotropic virus type 1 arose from a recombination event.
957 Journal of virology. 2014;88(17):9782–9788.

958 [74] Cassar O, Desrames A, Marçais A, Gout O, Taylor GP, Hermine O, et al. Multiple recom-
959 binant events in human T-cell leukemia virus type 1: complete sequences of recombinant
960 African strains. Emerging microbes & infections. 2020;9(1):913–923.

961 [75] Nethe M, Berkhout B, Van der Kuyl AC. Retroviral superinfection resistance. Retrovi-
962 rology. 2005;2(1):1–13.

963 [76] Ballard DW, Bohnlein E, Lowenthal JW, Wano Y, Franza BR, Greene WC. HTLV-I tax
964 induces cellular proteins that activate the κB element in the IL-2 receptor α gene. Science.
965 1988;241(4873):1652–1655.

966 [77] Xiao G, Cvijic ME, Fong A, Harhaj EW, Uhlik MT, Waterfield M, et al. Retroviral onco-
967 protein Tax induces processing of NF-κB2/p100 in T cells: evidence for the involvement
968 of IKKα. The EMBO journal. 2001;20(23):6805–6815.

969 [78] Ikezoe T, Nishioka C, Bandobashi K, Yang Y, Kuwayama Y, Adachi Y, et al. Longitudinal
970 inhibition of PI3K/Akt/mTOR signaling by LY294002 and rapamycin induces growth
971 arrest of adult T-cell leukemia cells. Leukemia research. 2007;31(5):673–682.

972 [79] Mukai R, Ohshima T. HTLV-1 HBZ positively regulates the mTOR signaling pathway
973 via inhibition of GADD34 activity in the cytoplasm. Oncogene. 2014;33(18):2317–2328.

974 [80] Zhao T, Yasunaga JI, Satou Y, Nakao M, Takahashi M, Fujii M, et al. Human T-cell
975 Leukemia Virus type 1 bZIP factor selectively suppresses the classical pathway of NF-κB.
976 Blood. 2009;113(12):2755–2764. doi:10.1182/blood-2008-06-161729.

977 [81] Mori N, Fujii M, Ikeda S, Yamada Y, Tomonaga M, Ballard DW, et al. Constitutive
978 activation of NF-κB in primary adult T-cell leukemia cells. Blood, The Journal of the
979 American Society of Hematology. 1999;93(7):2360–2368.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[82] Kataoka K, Nagata Y, Kitanaka A, Shiraishi Y, Shimamura T, Yasunaga Ji, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. Nature genetics. 2015;47(11):1304.

[83] Rowan AG, Dillon R, Witkover A, Melamed A, Demontis MA, Gillet NA, et al. Evolution of retrovirus-infected premalignant T-cell clones prior to adult T-cell leukemia/lymphoma diagnosis. Blood. 2020;135(23):2023–2032.

[84] Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, et al. Transcription factors modulate c-Fos transcriptional bursts. Cell reports. 2014;8(1):75–83.

[85] Jin J, Kim C, Xia Q, Gould TM, Cao W, Zhang H, et al. Activation of mTORC1 at late endosomes misdirects T cell fate decision in older individuals. Science immunology. 2021;6(60):eabg0791.

[86] Salminen A, Kaarniranta K. NF-κB signaling in the aging process. Journal of clinical immunology. 2009;29(4):397–405.

[87] Corradin A, Camillo BD, Rende F, Ciminale V, Toffolo G, Cobelli C. Retrovirus HTLV-1 Gene Circuit: A Potential Oscillator for Eukaryotes. In: Altman RB, Dunker AK, Hunter L, Murray T, Klein TE, editors. Biocomputing 2010: Proceedings of the Pacific Symposium, Kamuela, Hawaii, USA, 4-8 January 2010. World Scientific Publishing; 2010. p. 421–432. Available from: `http://psb.stanford.edu/psb-online/proceedings/psb10/corradin.pdf`.

[88] Yamagishi M, Nakano K, Miyake A, Yamochi T, Kagami Y, Tsutsumi A, et al. Polycomb-mediated loss of miR-31 activates NIK-dependent NF-κB pathway in adult T cell leukemia and other cancers. Cancer cell. 2012;21(1):121–135.

[89] Fujikawa D, Nakagawa S, Hori M, Kurokawa N, Soejima A, Nakano K, et al. Polycomb-

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

dependent epigenetic landscape in adult T-cell leukemia. Blood, The Journal of the American Society of Hematology. 2016;127(14):1790–1802.

[90] Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125. doi:10.1016/j.cell.2006.02.041.

[91] Kanki Y, Muramatsu M, Miyamura Y, Kikuchi K, Higashijima Y, Nakaki R, et al. Bivalent-histone-marked immediate-early gene regulation is vital for VEGF-responsive angiogenesis. Cell Reports. 2022;38(6):110332.

[92] Conway E, Rossi F, Fernandez-Perez D, Ponzo E, Ferrari KJ, Zanotti M, et al. BAP1 enhances Polycomb repression by counteracting widespread H2AK119ub1 deposition and chromatin condensation. Molecular cell. 2021;81(17):3526–3541.

[93] Lee BH, Lee MJ, Park S, Oh DC, Elsasser S, Chen PC, et al. Enhancement of proteasome activity by a small-molecule inhibitor of USP14. Nature. 2010;467(7312):179–184.

[94] Siebenlist U, Franzoso G, Brown K. Structure, regulation and function of NF-kappaB. Annual review of cell biology. 1994;10(1):405–455.

[95] Min Y, Lee S, Kim MJ, Chun E, Lee KY. Ubiquitin-specific protease 14 negatively regulates toll-like receptor 4-mediated signaling and autophagy induction by inhibiting ubiquitination of TAK1-binding protein 2 and Beclin 1. Frontiers in immunology. 2017;8:1827.

[96] Zhao Y, Zong F. Inhibiting USP14 ameliorates inflammatory responses in trophoblast cells by suppressing MAPK/NF-κB signaling. Immunity, Inflammation and Disease. 2021;9(3):1016–1024.

[97] Beck DB, Basar MA, Asmar AJ, Thompson JJ, Oda H, Uehara DT, et al. Linkage-specific deubiquitylation by OTUD5 defines an embryonic pathway intolerant to genomic variation. Science advances. 2021;7(4):eabe2116.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[98] Čikoš Š, Bukovská A, Koppel J. Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis. BMC molecular biology. 2007;8(1):1–14.

[99] Fan J, Ma G, Nosaka K, Tanabe J, Satou Y, Koito A, et al. APOBEC3G Generates Nonsense Mutations in Human T-Cell Leukemia Virus Type 1 Proviral Genomes ¡i¿In Vivo¡/i¿. Journal of Virology. 2010;84(14):7278–7287. doi:10.1128/JVI.02239-09.

[100] R Core Team. R: A Language and Environment for Statistical Computing; 2021. Available from: `https://www.R-project.org/`.

[101] Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21(16):3439–3440.

[102] Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature protocols. 2009;4(8):1184–1191.

[103] Carlson M. org.Hs.eg.db: Genome wide annotation for Human; 2021.

[104] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011;7(1):539.

[105] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research. 2019;47(W1):W636–W641.

[106] Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–1191.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523159; this version posted January 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[107] Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017–1018.

[108] Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic acids research. 2022;50(D1):D165–D173.

[109] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY, USA: Springer-Verlag New York; 2009. Available from: `http://ggplot2.org`.

[110] Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N. Advanced methods of microscope control using μManager software. Journal of biological methods. 2014;1(2).