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

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

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

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

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

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

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

[#] Current Address: Department of Microbiology, The University of Hong Kong, Hong Kong

*Corresponding author

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

Abstract

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

Author summary

Human retroviruses such as HIV-1 and HTLV-1 (human T cell leukemia virus) can establish a latent infection in the host cell. However, these viruses need to be able to produce viral genome to propagate in a new host. HTLV-1-infected cells are transmitted through breastfeeding, blood transfusion and sexual contact, and HTLV-1 restores transcription once the infected cells are drawn from infected individuals. We measured the kinetics of the HTLV-1 transcriptional reactivation in blood cells isolated from HTLV-1^{+} individuals by single-molecule RNA FISH. Viral transcripts were visualised as diffraction-limited spots and their abundance was quantified at one-hour intervals. The onset of the virus transcription peaked after one to three hours of incubation. In each cell, a short period of slow HTLV-1 transcription was followed by a phase of rapid transcription. Computer simulation, based on experimental data on PBMCs from a leukemic patient with a single dominant HTLV-1-infected T cell clone, indicated that this rapid transcription from the HTLV-1 sense-strand promoter activated by Tax was terminated in less than an hour. The HTLV-1 antisense transcript *HBZ* was constantly produced at a low level, and 50% ± 20% of HTLV-1^{+} cells were negative for *HBZ* at a given time. These results demonstrate how rapidly HTLV-1 is reactivated and potentially becomes infectious, once HTLV-1^{+} cells are transmitted into a new host.

57 ## Introduction

58 Retroviral latency constitutes the main barrier to eradicating the infection from the host. The provirus is
59 not completely silent, but is intermittently transcribed to produce viral particles that infect new clones
60 of cells [1]. This intermittent transcription is partly due to the apparently stochastic nature of gene
61 regulation [2]. However, it is poorly understood what regulates the balance between retroviral latency
62 and transcription. Attempts to reactivate HIV-1 to expose it to drug therapy and the immune response -
63 the shock-and-kill strategy [3] - have not been successful so far because it has not been possible to
64 reactivate all the HIV-1 provirus in the reservoir [4]. Therefore, it is essential to understand the
65 mechanisms and the extent of heterogeneity in the viral reactivation. Human T cell leukemia virus type 1
66 (HTLV-1) is a pathogenic retrovirus, which causes fatal and disabling diseases [1]. HTLV-1 remains largely
67 latent *in vivo*, but undergoes transcriptional reactivation once the HTLV-1-infected lymphocytes are
68 drawn from the circulation [5]. Here, we exploit this pattern of predominant HTLV-1 latency *in vivo* and
69 spontaneous transcriptional reactivation *ex vivo* to study the kinetics of retroviral reactivation.
70
71 It is estimated that 10 million people across the globe are living with HTLV-1 [6]. HTLV-1 resides mainly
72 in $CD4^+$ T lymphocytes, and causes adult T cell leukemia/lymphoma (ATL), a $CD4^+$ T cell malignancy, in
73 some 5% of HTLV-1-infected individuals. HTLV-1 also causes, in up to another 5%, HTLV-1-associated
74 myelopathy/tropical spastic paraparesis (HAM/TSP; hereafter abbreviated to HAM) and other
75 inflammatory disorders [1]. ATL typically presents as a monoclonal expansion of HTLV-1-infected T
76 lymphocytes that is characterized by a unique integration site of the HTLV-1 provirus in the host genome
77 [7], with an underlying polyclonal population of lower abundance. By contrast, patients with HAM and
78 asymptomatic HTLV-1 carriers have a highly polyclonal population of ~10^4 HTLV-1^+ T cell clones [7, 8].
79

80 HTLV-1 transcribes both the sense and antisense strand of its 9 kb provirus. The long terminal repeats
81 (LTRs) located at the ends of the provirus serve as the promoters for sense and antisense transcription
82 respectively. HTLV-1 Tax, produced from the sense strand, is a transcriptional transactivator protein that
83 triggers NF-κB activation and other transcriptional cascades as well as its own promoter at the LTR
84 (reviewed in ref [9]), exerting a strong positive feedback on the sense-strand transcription of the
85 provirus. The HTLV-1 transcript *tax* is undetectable *in vitro* in freshly isolated cells from most HTLV-1-
86 infected individuals [10]. However, the presence of chronically activated cytotoxic T lymphocytes (CTLs)
87 that recognize sense-strand-encoded products, particularly the highly immunodominant Tax protein
88 [11], indicates that the sense strand is intermittently expressed *in vivo*. By contrast, transcripts of the
89 HTLV-1 bZIP factor (*HBZ*) from the antisense strand are usually detectable *in vitro* at the cell population
90 level in peripheral blood mononuclear cells (PBMCs) obtained from HTLV-1-infected individuals [10].
91
92 In a previous study, we applied single-molecule RNA FISH (smFISH) [12, 13] to quantify HTLV-1 sense and
93 antisense transcription in HTLV-1^{+} T cell clones isolated from HTLV-1-infected individuals by limiting
94 dilution and maintained *in vitro* [14]. We observed that the HTLV-1 sense strand is transcribed in an
95 occasional yet intense transcriptional burst. The antisense transcript *HBZ*, however, was much more
96 constantly expressed, albeit at a low level.
97
98 To propagate between cells and between individuals, HTLV-1 must restore the sense-strand
99 transcription from latency and produce viral components to form a virological synapse at the site of cell-
100 to-cell contact [15]. The infection spreads to a new host when HTLV-1-infected T lymphocytes are
101 transmitted by breastfeeding, sexual contact and blood transfusion. HTLV-1 sense transcripts rapidly
102 become detectable when the HTLV-1-infected T lymphocytes are drawn from the circulation: the
103 transcriptional reactivation is mediated by p38 MAP kinase activation [5] in response to diverse forms of

104 cell stress. In the present study, we quantified this transient HTLV-1 transcriptional reactivation using
105 smFISH [14] and used stochastic simulation to derive numerical estimates of the parameters of HTLV-1
106 reactivation from latency.
107

108 ## Results

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

110 HTLV-1-infected PBMCs carry a single copy of the provirus in the genome. HTLV-1 rapidly initiates
111 transcriptional reactivation of the sense strand once the PBMCs are introduced into culture. We used
112 smFISH to quantify the frequency and intensity of HTLV-1 reactivation at the single-cell level (Fig 1A).
113 We removed $CD8^+$ cells to avoid CTL-mediated killing of HTLV-1-infected PBMCs that express sense-
114 strand antigens (particularly Tax) during culture [16]. The PBMCs were sampled every hour and
115 subjected to smFISH.
116

117 The sequence of representative images of cells (from a patient with HAM, coded TDZ) in Fig 1B shows
118 the increase in the abundance of HTLV-1 sense transcripts over time. Sense-strand transcripts were
119 virtually absent at 0 h (i.e. without incubation), and became detectable in a fraction of cells after 2 h of
120 incubation. Subsequently, a bright spot appeared in the nucleus in those cells in which the sense-strand
121 transcripts were present. The bright spot is an indication of rapid transcription [17], where the
122 transcripts are produced faster than they diffuse away from the site of transcription, resulting in the
123 accumulation of mRNAs at the position of the HTLV-1 provirus in the nucleus. This site contains nascent,
124 unspliced transcripts, as revealed by an additional set of probes that targets *gag* (Fig 1C) which is
125 otherwise spliced out of the sense transcripts (Fig 1A), indicating that the rate of transcription is even
126 faster than that of the splicing reaction [12]. In this paper, we define a transcriptional burst in a single
127 cell as the presence of viral transcripts in the nucleus [18, 19], as opposed to the rapid increase in the

128 abundance of HTLV-1 transcripts over time at the population level. The occurrence of the intense spot in
129 the nucleus was typically associated with the presence of ~10 to 20 sense transcripts in the nucleus. This
130 observation is consistent with the idea that the positive feedback exerted by Tax protein on the sense-
131 strand promoter occurs after a short interval following the first appearance of sense-strand transcripts,
132 during which *tax* mRNA enters the cytoplasm, is translated, and Tax protein re-enters the nucleus.
133 Thereafter, we observed cells carrying a large number of sense-strand transcripts - up to 300 molecules
134 (Fig 1B).
135

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

137 The population of PBMCs imaged by smFISH contained both HTLV-1-infected and uninfected cells. We
138 wished to analyse the distribution of the HTLV-1 mRNA count within the HTLV-1-infected population.
139 We used the proviral load to calculate the number of HTLV-1-infected cells in the total population (Fig
140 2A).
141

142 The distribution of the number of sense-strand transcripts per HTLV-1$^+$ cell over time is shown in Fig 2B.
143 In each HTLV-1$^+$ subject examined (TBC, TDZ, LGZ and LGB), the total number of transcripts increased
144 continuously during the 8 hours of incubation. The distribution in the sample from one patient with
145 HAM (TBC) presented a long tail of cells that carried a larger number of sense transcripts (up to 250
146 molecules), with a continuous distribution of cells that carried fewer than 50 transcripts. By contrast, a
147 single mode was identified at around 100 molecules in an ATL patient (LGZ) carrying a single dominant
148 clone.
149

150 The average count of the HTLV-1 sense-strand transcripts in the top 25% of all the infected cells was 100
151 to 300 molecules (S1 Fig). A steeper increase in the count was observed between 1 and 2 hours than in
152 the first hour, which accords with the notion that the HTLV-1 sense-strand expression is biphasic.
153
154 We postulated that the polyclonal structure of HTLV-1-infected T cell clones in subjects with non-
155 malignant HTLV-1 infection [7] explains the observed diversity in the distribution of the number of
156 sense-strand transcripts per cell, whereas ATL patients with a monoclonal expansion of an HTLV-1$^+$ T cell
157 clone would show more uniform distribution of the number of sense transcripts. We plotted the
158 variance vs. the mean of the observed transcript count in the four HTLV-1$^+$ subjects with observations at
159 nine time points (Fig 2C). The mean and variance from all the four samples showed a uniform power-law
160 relationship with the exponent 1.7, suggesting that there is no fundamental qualitative difference in the
161 kinetics of transcription between the samples examined.
162
163 **HTLV-1 reactivation peaks after a few hours of *ex vivo* culture**
164 To further characterize the transcriptional reactivation of the HTLV-1 sense strand, we quantified the
165 fraction of cells that had become positive for sense transcripts at a given time during short-term culture
166 of PBMCs (Fig 3A). Cells were defined as positive for sense transcripts when they carried 3 or more
167 spots, to discount spurious spots due to noise (S2 Fig; see Materials and Methods). At 0 h, none of the
168 HTLV-1$^+$ cells from each of the four HTLV-1$^+$ subjects examined (TBC, TDZ, LGZ and LGB) were positive for
169 sense-strand transcripts. After 8 h of incubation, the fraction of positive cells rose to between 0.2 and
170 0.8, depending on the HTLV-1$^+$ subject examined. We assumed that the waiting time until we observed 3
171 or more molecules in a cell is an accumulation of single stochastic processes, and therefore fitted the
172 plots presented in Fig 3A with a cumulative gamma distribution [20]. The probability distribution
173 function derived from the fitted curve (Fig 3B) shows that the onset of sense-strand transcriptional

174 reactivation peaked within 1 to 3 hours after the start of *ex vivo* culture. The timing of onset was more
175 broadly distributed in samples from patients with HAM (TBC and TDZ) than from those with ATL (LGZ
176 and LGB): the 75% quantiles of the probability distribution were 3.9 h and 4.9 h for TBC and TDZ
177 respectively, whereas those for LGZ and LGB were less than 3 h. However, in each case the onset of the
178 transcriptional reactivation was completed by 8 h *ex vivo* incubation in those cells that expressed
179 detectable HTLV-1 mRNAs (20% to 80% of the total population, again depending on the HTLV-1$^+$ subject
180 examined); the remaining cells showed no sign of proviral transcription during the period of
181 observation.
182

183 **Stochastic simulation predicts a burst termination following the promoter activation**

184 From the results presented above, we established the following characteristics of the transient and
185 heterogeneous HTLV-1 transcription in *ex vivo* PBMCs during the short-term culture. (1) The timing of
186 the onset of transcriptional reactivation peaks in the first few hours of *ex vivo* incubation. (2) The sense-
187 strand transcription is restored with an initially low transcription rate. (3) Subsequently the rate of
188 transcription of the sense strand increases rapidly: we postulate that this rapid sense-strand
189 transcription is triggered by the strong positive feedback loop exerted by Tax protein. (4) Only a subset
190 of HTLV-1$^+$ PBMCs (20% to 80%, depending on the subject) restore the sense transcription.
191

192 We wished to identify the smallest number of parameters needed to explain the observed kinetics of
193 proviral transcription. We applied a three-state promoter model [21] with some modifications (Fig 4A):
194 the model allows the HTLV-1 sense promoter to transit from the "Off" to "On" state, and further
195 progress to the "On-Tax" state by binding Tax protein; transcription occurs at the On state and the On-
196 Tax state, with distinct transcription rates k_5 and k_6, respectively ($k_5 < k_6$). We used the smFISH data from
197 the ATL patient LGZ for the parameter optimisation, because the single dominant clone of HTLV-1$^+$

198 PBMCs in this leukemic subject requires only a single kinetic parameter for each reaction rate in the
199 model, as opposed to a polyclonal population in which each clone would require a unique reaction rate
200 to be assigned. The simulations of this three-state model are shown in Fig 4B. The three-state model
201 failed to capture the observed distribution of the transcript count in LGZ (Fig 4B, upper panel): it led to
202 an overestimation of the number of sense-strand transcripts at the later time points. Therefore we
203 assumed an additional reaction path in the three-state model (Fig 4B, lower panel), which allows the
204 promoter to transit directly from the On-Tax state to the Off state with the parameter k_{10}. This model
205 fitted the smFISH data well (Fig 4B, lower panel; see also S1 Text) and accurately interpolated missing
206 data (S3 Fig); the resulting parameter estimates (posterior distributions) are presented in Fig 4C. The
207 parameter k_{10}, which is a strong determinant of the duration of the sense-strand transcription,
208 converged within the range of 1 to 3 (h^{-1}) (Fig 4C and S4 Fig), indicating that the HTLV-1 sense-strand
209 promoter terminates the Tax-activated rapid transcription in less than an hour (mean duration ~ $1/k_{10}$)
210 in this putatively malignant clone. The duration of the rapid transcription phase was measured in each
211 cell in the simulation by recording the interval between the initiation of the On-Tax promoter state and
212 the direct transition from the On-Tax to the Off state (Fig 4D). These results indicate that the HTLV-1
213 sense-strand transcription is rapidly terminated.
214

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

216 Lastly, we investigated the expression of the HTLV-1 antisense transcript *HBZ*. PBMCs isolated from one
217 asymptomatic HTLV-1^+ carrier (HBL), three patients with HAM (TBC, TDK and TDZ) and four with ATL
218 (LGZ, LFV, LFK and LGB) were examined by smFISH (Fig 5A). The distribution of the *HBZ* mRNA count per
219 HTLV-1^+ PBMCs (without incubation) is shown in Fig 5B; the mean (<m>) is indicated in the inset. These
220 results lead to two conclusions. First, the number of *HBZ* mRNA transcripts carried per cell was very low:
221 frequently only one or two molecules of *HBZ* mRNA were present per cell; and unexpectedly 20% to 70%

222 of HTLV-1$^+$ PBMCs in the subjects (except LFK) were negative for *HBZ* at a given time. In the subject LFK,
223 almost all the HTLV-1$^+$ PBMCs were negative for *HBZ* (also see S2 Fig). Second, the distribution of the
224 number of *HBZ* molecules conformed to the Poisson distribution of the parameter <m>, the observed
225 mean mRNA count (Fig 5B, black lines; the p-value of the Kolmogorov-Smirnov test is indicated in the
226 inset). This observation indicates that the *HBZ* transcripts are produced and degraded at a constant
227 average rate (Fig 5C). Our previous study showed that the half-life of *HBZ* mRNA is 4.4 h [14], from which
228 we estimate the degradation rate δ to be ln2 / 4.4 = 0.1575 h^{-1}. Since the mean number of *HBZ* mRNA
229 molecules carried per cell <m> ranged from 0.380 to 1.31 (Fig 5B, except LFK), we estimate that the
230 mean production rate of *HBZ* mRNA (k = <m>·δ) falls within the range 0.0599 to 0.206 h^{-1} (Fig 5C).
231

232 ## Discussion

233 It is assumed that the HTLV-1 sense strand is expressed *in vivo* [1], because a strong, chronically
234 activated CTL response against Tax and other sense-strand-encoded proteins is readily detectable in
235 PBMCs from HTLV-1$^+$ subjects *ex vivo* [11]. However, in the present study HTLV-1 sense-strand
236 transcripts were rarely detected in fresh PBMCs without incubation, using the sensitive and specific
237 method of smFISH, in any of the HTLV-1$^+$ subjects examined (Figs 1B and 2B). The results show that
238 HTLV-1 sense-strand transcription is rare in the circulation. We postulate that HTLV-1 expresses the
239 sense strand in different anatomical compartments of the body other than the blood, such as bone
240 marrow and lymphatic organs. This notion is supported by the observation that *tax* is highly expressed in
241 bone marrow in Japanese macaques infected with simian leukemia virus STLV-1, the simian counterpart
242 of HTLV-1 [22]. The smFISH technique could be applied to analyse specimens from those organs, to
243 identify and localize the subset of cells that express *tax*, and to reveal the structural and spatial
244 distribution of HTLV-1 expression *in situ*.
245

246 Whereas the sense-strand transcript *tax* is usually undetectable by reverse-transcription-PCR in PBMCs
247 isolated from HTLV-1$^+$ subjects, the HTLV-1 antisense transcript *HBZ* is invariably detected in the same
248 subjects [10] at the cell population level. This observation raised the hypothesis that HBZ protein and
249 *HBZ* mRNA play an essential part in HTLV-1 persistence (reviewed in ref [23]).
250
251 In our previous study, we examined *HBZ* expression by smFISH in five HTLV-1-infected T cell clones
252 isolated by limiting dilution from HTLV-1$^+$ patients and maintained *in vitro* [14]. While the distribution of
253 the *HBZ* mRNA count in three of the clones showed a small but significant departure from the
254 expectation derived from the Poisson distribution, the other two clones conformed to the Poisson
255 distribution (S5 Fig). The slight deviation from the Poisson distribution observed in certain clones may
256 result from a small burst of *HBZ* transcription [14] (S2 Text). The average *HBZ* mRNA count per cell in
257 those five clones was in a small range (1.67 to 2.92), suggesting that the *HBZ* production rate varies little
258 with the genomic insertion site of the HTLV-1 provirus.
259
260 In PBMCs circulating in HTLV-1-infected individuals, we assume that the great majority of HTLV-1$^+$ clones
261 are capable of expressing *HBZ*, because its sequence is highly conserved and because *HBZ* appears to be
262 necessary for the persistence of HTLV-1$^+$ T lymphocytes *in vivo* [23]. In the *ex vivo* PBMCs, we did not
263 observe strong transcriptional bursts of the antisense strand. The distribution of the *HBZ* count in the
264 patients with ATL conformed well with the Poisson expectation (Fig 5B), except in the sample from the
265 patient coded LGB. This leukemic patient was anomalous in that only 20% of the total population
266 expressed the proviral sense strand, even though a single dominant clone was present in the peripheral
267 blood. The malignant clone present in another ATL patient (LFK), which essentially expressed no *HBZ*,
268 had presumably acquired mutations that allowed the cells to proliferate independently of the HTLV-1
269 provirus. The *HBZ* count in the patients with HAM, with a polyclonal HTLV-1$^+$ T cell population, also

270 conformed to the Poisson distribution. We conclude that *HBZ* is produced constantly in each infected
271 cell, at ~0.1 molecules per hour on average; due to the low average rate of transcription, around half of
272 all HTLV-1-infected cells are *HBZ*-negative at any one time.
273
274 It has been reported that *HBZ* mRNA promotes cell proliferation *in vitro* [10, 24]. If this observation
275 holds true in naturally-infected cells, with the low abundance of *HBZ* mRNA molecules observed in this
276 study, it becomes increasingly important to understand how *HBZ* mRNA exerts such functions.
277
278 The HBZ protein level is also of great interest because the $CD8^+$ T cell response against HBZ is a
279 significant determinant of the HTLV-1 proviral load [25], the main correlate of the risk of HAM and ATL
280 [1]. A previous study from others estimated, by western blotting, that the abundance of endogenous
281 HBZ protein is 17 molecules per cell in an ATL case [26]. In addition to the abundance of *HBZ* mRNA
282 quantified in the present study, measuring the half-life of HBZ protein will permit the estimation of the
283 variance in HBZ protein expression as well as the rate of HBZ translation.
284
285 During the *ex vivo* incubation, the distribution of *HBZ* mRNA molecules per cell was stable compared to
286 the changes in the sense-strand expression (S6 Fig) in three of the four subjects examined (TBC, TDZ and
287 LGB). This observation indicates that the antisense strand is not reactivated in concert with the sense
288 strand. In the fourth subject examined (LGZ), the abundance of *HBZ* mRNA in the population decayed
289 over time (S6 Fig, LGZ). Since the majority of HTLV-1^+ cells in LGZ (80%) restored the sense-strand
290 transcription, it is possible that transcriptional interference - by the collision of two converging
291 polymerase complexes - diminished the antisense transcription in this subject. However, in each of the
292 remaining three subjects studied, there was no correlation between the sense-strand transcription and
293 *HBZ* transcription. These results suggest that transcriptional interference between the sense- and

294 antisense-strands does not play an important part in regulating HTLV-1 proviral transcription in most
295 infected T cell clones.
296
297 By applying smFISH to the transient HTLV-1 transcriptional reactivation of the sense strand, we found
298 that reactivation was highly heterogeneous in *ex vivo* PBMCs, as shown by three observations. First, only
299 a subset of the HTLV-1^{+} PBMCs reactivated proviral transcription; second, the time of onset of
300 transcription varied between cells; and third, the intensity of reactivation (i.e. the number of transcripts
301 produced in a given period of time) varied widely in the population. Our results indicate two distinct
302 phases of sense-strand transcription: the initial low-rate transcription produces ~10 to 20 mRNA
303 molecules; thereafter a phase of rapid transcription occurs, probably attributable to the positive
304 feedback exerted by the viral transactivator protein Tax. Transcriptional heterogeneity is likely to arise
305 from the difference between clones in the HTLV-1 proviral integration site in the host genome.
306 However, even in the ATL subjects examined in this study (LGZ and LGB), who carried one dominant
307 clone in which (by definition) each cell shares the same integration site, a significant variance between
308 cells in reactivation intensity was observed, which was equivalent to that of HTLV-1^{+} patients with
309 polyclonal insertions (Fig 2C). Such heterogeneity in transcription is often ascribed to the stochasticity
310 that inevitably arises in biological processes.
311
312 In this study, we demonstrated that HTLV-1 sense-strand reactivation in PBMCs occurs in a burst of rapid
313 transcription. But how long does such a burst last in the nucleus? The precise duration of a burst in
314 individual cells cannot be determined by smFISH, which captures only a single timepoint. By contrast,
315 the stochastic simulation of the three-state promoter model (Fig 4B and S1 Text) estimated that the
316 average duration of the intense transcription in PBMCs from a leukemic patient (LGZ) was less than an
317 hour: the model required the HTLV-1 sense-strand promoter to revert from the On-Tax state to the Off.

318 Without invoking this extra parameter, cells did not terminate the transcription during the 8 h of the
319 simulation, so the number of transcripts overshot the observed smFISH data (Fig 4B and S1 Text). These
320 parameters were estimated from a population of PBMCs containing a single dominant clone: other
321 HTLV-1-infected clones may differ in the magnitude of the kinetic parameters, although the
322 fundamental molecular mechanisms are unlikely to vary. Indeed, the putative malignant clone in the
323 patient LGZ might have been selected *in vivo* by the host immune response for rapid termination of
324 sense-strand transcription: *tax* expression is lost in ~50% of cases of ATL, and it is believed that the
325 strong cytotoxic T lymphocyte response to Tax is responsible for this silencing.
326
327 Mahgoub et al. reported that the average duration of Tax expression was ~19 hours [27] in an HTLV-1-
328 infected cell line (MT-1) that was established from a patient with ATL, using a GFP reporter system to
329 indicate the production of Tax protein. By contrast, we used uncultured *ex vivo* PBMCs, which carry a
330 single copy of the HTLV-1 provirus in each cell [28], and quantified the HTLV-1 transcripts by smFISH,
331 which allows direct estimation of the gene transcription kinetics. Based on our model, we postulate that
332 the HTLV-1 sense strand, which encodes *tax* and is activated by the strong positive feedback of Tax
333 protein, terminates the transcription after a short period, even if Tax protein is still present in the cell.
334 However, it is possible that repeated sense-strand bursts occur in rapid succession, due to repeated
335 reattachment of transcription complexes to the HTLV-1 promoter, resulting in the longer duration of Tax
336 presence in the cell observed by Mahgoub et al.
337
338 Our data, together with inferences from the mathematical model, suggest that active termination of a
339 transcriptional burst is critical in shaping the dynamics of HTLV-1 sense-strand transcription. We
340 examined the transient (non-equilibrium) dynamics of HTLV-1 transcriptional bursting in primary,
341 naturally-infected lymphocytes, whereas others studied Tat transcription from synthetic Tat circuits to

342 analyse the mechanisms that produce bimodality in Tat expression over an extended parameter regime
343 in a steady-state system [29-31].
344
345 The mechanism of this rapid termination of the sense-strand burst remains unknown, as in mammalian
346 gene expression in general. The HTLV-1 accessory protein p30 retains *tax* mRNA in the nucleus, which
347 results in a reduction of Tax protein production, leading in turn to reduced viral transcription [32].
348 However, p30 is unlikely to account for the rapid termination of the HTLV-1 transcriptional burst in
349 PBMCs observed in the present study, because the abundance of Tax protein continues to rise in PBMCs
350 during 8 h *ex vivo* culture [33]. At least three factors can be identified that may contribute to
351 transcriptional silencing. First, epigenetic modifications at the 5 promoter including re-ubiquitination of
352 histone 2A lysine 119 by polycomb repressive complex 1 (PRC1) [5] and methylation (or
353 hydroxymethylation) of cytosine [34]. Second, sequestration of the transcription co-factors CBP/p300 by
354 HBZ protein [35]. Third, a progressive reduction in Tax translation caused by the accumulation of Rex
355 protein [36]. Recent evidence suggests that transcription factors form phase-separated droplets where
356 transcriptional bursts occur, and intrinsically disordered domains of transcription factors are particularly
357 responsible for forming droplets [37, 38]. HTLV-1 Tax protein has a number of binding partners in the
358 cell [39] and forms speckles in the nucleus that colocalize with active transcription spots [40]. These
359 observations raise the possibility that Tax contributes to the formation of phase-separated droplets in
360 the nucleus that serve as the locus for HTLV-1 transcription. External perturbation can disperse phase-
361 separated droplets [38] and might therefore terminate the burst, leading to a refractory period of
362 complete transcriptional silence.
363
364 In HIV-1 infection, the potential use of shock-and-kill treatment is still under debate because all
365 approaches used to date have activated only a small fraction of the proviruses in a host. The results

366 presented here demonstrate that the extent of heterogeneity in retrovirus reactivation is diverse in a
367 population of naturally-infected cells, and suggest a quantitative approach for examining the efficacy of
368 the treatment.
369

370 ## Materials and Methods

371 ### Ethics statement

372 EDTA-anticoagulated blood from HTLV-1-infected patients was collected at St Mary's Hospital (Imperial
373 College London, UK). All donors gave written informed consent in accordance with the Declaration of
374 Helsinki to donate blood samples to the Communicable Diseases Research Tissue Bank, approved by the
375 UK National Research Ethics Service (15/SC/0089).
376

377 ### Patient-derived PBMCs

378 Information on the HTLV-1^{+} subjects examined in this study (four ATL patients, three HAM patients and
379 one asymptomatic carrier) is available in Table 1. PBMCs were isolated using Histopaque (Sigma, H8889)
380 and washed with PBS. Unless mentioned otherwise, they were stored in liquid nitrogen with FBS/10%
381 DMSO until use.
382
383
384
385
386
387
388

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

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

[a]measured in the sample collected on 04 Jul 2016
[b]Samples were used immediately after PBMC isolation, without freeze-thawing.

Short-term culture of patient PBMCs

Patient-derived PBMCs were thawed and washed with PBS (Gibco, 14190) containing 2 mM EDTA and 0.1% FBS. CD8$^+$ cells were removed with Dynabeads (Invitrogen, 11147D) to avoid CTL killing against HTLV-1 Tax$^+$ cells. Cells were suspended (1×10^6 cells/mL) in RPMI1640 (Sigma, R0883) supplemented with 2 mM L-glutamine (Gibco, 25030-81), penicillin/streptomycin (Gibco, 15070-063) and 10% fetal bovine serum (Gibco, 10500-064). One mL of cell suspension (1×10^6 cells) was aliquoted and dispensed into 5 mL polystyrene round-bottom tubes (Falcon, 352054) to keep the cell density consistent between assays. Cells were incubated in 5% CO_2 at 37°C until sampled. One aliquot was sampled at each time point (every hour) and subjected to single-molecule RNA FISH.

402 **Single-molecule RNA FISH**

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

407

408 Hybridization was performed as previously described [14]. Namely, permeabilized cells were hybridized
409 overnight at 37°C with probes that target either *gag* (HTLV-1 sense strand; labeled with FAM), *tax*
410 (HTLV-1 sense strand; labeled with Quasar 670) or *HBZ* (HTLV-1 antisense strand; labeled with Quasar
411 570) in hybridization buffer (Stellaris, SMF-HB1-10). Cells were washed on the coverslip twice with wash
412 buffer A (Stellaris, SMF-WA1-60), and then wash buffer B (Stellaris, SMF-WB1-20). DAPI staining was
413 included during the second wash with the buffer A. The cells were embedded in VECTASHIELD mounting
414 medium (Vector Laboratories, H-1000) for imaging.

415

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

419

420 **Image analysis**

421 Stacked TIFF images containing 54 z-slices with a 300 nm interval were obtained using Micro Manager
422 [41]. Each slice is composed of four channels (DAPI, FAM, Quasar 570 and Quasar 670). The stacked TIFF
423 images were split into the four channels. Cells were segmented using the DAPI channel with an in-house
424 algorithm which identifies the edge of the nucleus by signal intensity (the algorithm is available at DOI
425 10.17605/OSF.IO/R6Q8T). Dead cells were identified by the abnormal morphology of nucleus and

426 omitted from the analysis. Spots in each of the three channels (FAM, Quasar 570 and Quasar 670) were
427 detected with FISH-QUANT [42]. A threshold for the signal intensity was set to distinguish the mRNA
428 spots from spurious signals: an example is shown in S2 Fig containing the samples that were hybridized
429 and imaged in the same batch - LGZ fixed at 7 h; LFK and LFV fixed without incubation. The image data
430 are available at DOI 10.17605/OSF.IO/R6Q8T (open under CC0 1.0 Universal Licence).
431

432 **Proviral load measurement**

433 DNA was extracted from unfixed cells using DNeasy Blood & Tissue Kit (Qiagen, 69504). The reaction of
434 20 µL volume was set up using ddPCR Supermix (Bio-Rad, 186-3023) containing 900 nM each of forward
435 and reverse primers and 250 nM FAM-labeled probe to detect the HTLV-1 *pX* locus (forward primer, 5'-
436 CTC CTT CCG TTC CAC TCA AC-3'; reverse primer, 5'-GTG GTA GGC CTT GGT TTG AA-3'; probe, 5'-CGC
437 CTA TGA TTT CCG GGC CCT G-3'). The copy number of a host gene *RPP30* was determined using 1 µL of
438 HEX-labeled PrimePCR RPP30 (Bio-Rad, 10031243) in the reaction to calibrate the amount of DNA input.
439 Droplets were generated with QX200 Droplet Generator (Bio-Rad). PCR was performed (95°C for 10 min;
440 40 cycles of 94°C for 30 sec and 60°C for 1 min; 98°C for 10 min) with C1000 Touch thermal cycler (Bio-
441 Rad). Fluorescent signal was detected using QX200 Droplet Reader (Bio-Rad).
442

443 **Half-life estimation of the HTLV-1 sense transcripts**

444 Two HTLV-1 infected T-cell clones [14] were treated with 10 µg/ml Actinomycin D (Sigma, SBR00013).
445 Total RNA was extracted at 0, 3, 6 and 9 hours of treatment using RNeasy Plus Mini Kit (Qiagen, 74134),
446 and was reverse-transcribed to cDNA using Transcriptor First Strand cDNA Synthesis Kit with random
447 primers (Roche, 04897030001). The relative abundance of HTLV-1 sense-strand transcripts at each time
448 point was measured by qPCR with Fast SYBR Green Master Mix (Thermo Fisher Scientific, 4385612)
449 (forward primer, 5'-CCG GCG CTG CTC TCA TCC CGG T-3'; reverse primer, 5'-GGC CGA ACA TAG TCC CCC

450 AGA G-3′ [5]). The half-life of the HTLV-1 sense transcripts was estimated by non-linear regression curve
451 fitting (one phase decay) using GraphPad Prism version 8.2.1 for Windows (GraphPad Software).
452

453 **Stochastic simulation of the transient HTLV-1 sense-strand transcription**

454 A Gillespie stochastic simulation algorithm was employed [43]. The molecular species and the schematic
455 diagram of reactions are shown in Fig 4A. We assumed three possible states for the HTLV-1 sense-strand
456 promoter: the "Off" state where essentially no transcription takes place; the "On" state from which
457 sense transcripts are produced at a low transcription rate; and the "On-Tax" state where the rapid
458 transcription takes place due to the strong positive feedback in the presence of Tax protein. The HTLV-1
459 sense promoter was set at the Off state in each cell with no sense transcripts and Tax protein in the
460 initial condition. In modelling the HTLV-1 transient transcription in the patient LGZ, 20 % of the total
461 cells in the simulation were confined to the Off promoter state (Fig 3A, LGZ). A random parameter
462 search was performed to identify the sets of parameters that fit to the smFISH data (S1 Text). The
463 parameter search space is shown in Table 2. The limits of the parameter search space were set using the
464 data from smFISH experiments on the timing of the onset of transcription (for k_1) and the number of
465 sense transcripts observed prior to the rapid transcription (for k_5 and k_3). The sense-strand mRNA
466 degradation rate (k_7, ln2 / 4.5 = 0.154 h^{-1}) was determined experimentally in the present study (S7 Fig).
467 The degradation rate of Tax protein (k_9) was obtained from Rende et al. [44]. One million parameter sets
468 were tested for each respective model. The goodness of fit for each parameter set was assessed by
469 squared deviation summed across the bins in every time point. See also the supporting information (S1
470 Text); the code for model iii is available at DOI 10.17605/OSF.IO/R6Q8T.
471
472
473

Table 2. Parameters used in the stochastic simulation.

Reaction	Parameter	Description	Range for parameter search (h^{-1})
Transition	k_1	Off to On	0.25 to 2.5
	k_2	On to Off	0.01 to 10
	k_3	On to On-Tax	0.1 to 10
	k_4	On-Tax to On	0.01 to 10
	k_{10}	On-Tax to Off	0.01 to 10
Transcription	k_5	From On state	10 to 50
	k_6	From On-Tax state	k_5 to 100
Translation	k_8	Tax translation	0.01 to 10
Degradation	k_7	Sense mRNA	0.154 (fixed)
	k_9	Tax protein	0.105 (fixed)

Acknowledgements

We thank Graham Taylor, Lucy Cook, the donors and research nurses in the National Centre for Human Retrovirology, Imperial College; Dr Aileen Rowan for discussion on the ATL subjects; Dr Martin Billman for help with the single-molecule RNA FISH; and the Imperial College Research Computing Service (DOI: 10.14469/hpc/2232) for the use of the high-performance computer cluster; and the Gen Foundation for their support.

482 ## References

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

486 2. Singh A, Weinberger LS. Stochastic gene expression as a molecular switch for viral latency. Curr
487 Opin Microbiol. 2009;12(4):460-6. Epub 2009/07/15. doi: 10.1016/j.mib.2009.06.016. PubMed PMID:
488 19595626; PubMed Central PMCID: PMCPMC2760832.

489 3. Darcis G, Van Driessche B, Van Lint C. HIV Latency: Should We Shock or Lock? Trends Immunol.
490 2017;38(3):217-28. Epub 2017/01/12. doi: 10.1016/j.it.2016.12.003. PubMed PMID: 28073694.

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

495 5. Kulkarni A, Taylor GP, Klose RJ, Schofield CJ, Bangham CR. Histone H2A monoubiquitylation and
496 p38-MAPKs regulate immediate-early gene-like reactivation of latent retrovirus HTLV-1. JCI Insight.
497 2018;3(20). Epub 2018/10/20. doi: 10.1172/jci.insight.123196. PubMed PMID: 30333309; PubMed
498 Central PMCID: PMCPMC6237452.

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

502 7. Gillet NA, Malani N, Melamed A, GormLey N, Carter R, Bentley D, et al. The host genomic
503 environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. Blood.
504 2011;117(11):3113-22. Epub 2011/01/14. doi: 10.1182/blood-2010-10-312926. PubMed PMID:
505 21228324; PubMed Central PMCID: PMCPMC3062313.

506 8. Laydon DJ, Melamed A, Sim A, Gillet NA, Sim K, Darko S, et al. Quantification of HTLV-1 clonality
507 and TCR diversity. PLoS Comput Biol. 2014;10(6):e1003646. Epub 2014/06/20. doi:
508 10.1371/journal.pcbi.1003646. PubMed PMID: 24945836; PubMed Central PMCID: PMCPMC4063693.
509 9. Kulkarni A, Bangham CRM. HTLV-1: Regulating the Balance Between Proviral Latency and
510 Reactivation. Front Microbiol. 2018;9:449. Epub 2018/04/05. doi: 10.3389/fmicb.2018.00449. PubMed
511 PMID: 29615991; PubMed Central PMCID: PMCPMC5867303.
512 10. Satou Y, Yasunaga J, Yoshida M, Matsuoka M. HTLV-I basic leucine zipper factor gene mRNA
513 supports proliferation of adult T cell leukemia cells. Proc Natl Acad Sci U S A. 2006;103(3):720-5. Epub
514 2006/01/13. doi: 10.1073/pnas.0507631103. PubMed PMID: 16407133; PubMed Central PMCID:
515 PMCPMC1334651.
516 11. Jacobson S, Shida H, McFarlin DE, Fauci AS, Koenig S. Circulating CD8+ cytotoxic T lymphocytes
517 specific for HTLV-I pX in patients with HTLV-I associated neurological disease. Nature.
518 1990;348(6298):245-8. Epub 1990/11/15. doi: 10.1038/348245a0. PubMed PMID: 2146511.
519 12. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. Science.
520 1998;280(5363):585-90. Epub 1998/05/09. PubMed PMID: 9554849.
521 13. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA
522 molecules using multiple singly labeled probes. Nat Methods. 2008;5(10):877-9. Epub 2008/09/23. doi:
523 10.1038/nmeth.1253. PubMed PMID: 18806792; PubMed Central PMCID: PMCPMC3126653.
524 14. Billman MR, Rueda D, Bangham CRM. Single-cell heterogeneity and cell-cycle-related viral gene
525 bursts in the human leukaemia virus HTLV-1. Wellcome Open Res. 2017;2:87. Epub 2017/10/25. doi:
526 10.12688/wellcomeopenres.12469.2. PubMed PMID: 29062917; PubMed Central PMCID:
527 PMCPMC5645716.

528 15. Igakura T, Stinchcombe JC, Goon PK, Taylor GP, Weber JN, Griffiths GM, et al. Spread of HTLV-I
529 between lymphocytes by virus-induced polarization of the cytoskeleton. Science. 2003;299(5613):1713-
530 6. Epub 2003/02/18. doi: 10.1126/science.1080115. PubMed PMID: 12589003.
531 16. Hanon E, Hall S, Taylor GP, Saito M, Davis R, Tanaka Y, et al. Abundant tax protein expression in
532 CD4+ T cells infected with human T-cell lymphotropic virus type I (HTLV-I) is prevented by cytotoxic T
533 lymphocytes. Blood. 2000;95(4):1386-92. Epub 2000/02/09. PubMed PMID: 10666215.
534 17. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells.
535 PLoS Biol. 2006;4(10):e309. Epub 2006/10/20. doi: 10.1371/journal.pbio.0040309. PubMed PMID:
536 17048983; PubMed Central PMCID: PMCPMC1563489.
537 18. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, et al. Transcriptional burst
538 frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A.
539 2012;109(43):17454-9. Epub 2012/10/16. doi: 10.1073/pnas.1213530109. PubMed PMID: 23064634;
540 PubMed Central PMCID: PMCPMC3491463.
541 19. Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, et al.
542 Genomic encoding of transcriptional burst kinetics. Nature. 2019;565(7738):251-4. Epub 2019/01/04.
543 doi: 10.1038/s41586-018-0836-1. PubMed PMID: 30602787.
544 20. Schwabe A, Bruggeman FJ. Single yeast cells vary in transcription activity not in delay time after
545 a metabolic shift. Nat Commun. 2014;5:4798. Epub 2014/09/03. doi: 10.1038/ncomms5798. PubMed
546 PMID: 25178355.
547 21. Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, et al. Transcription factors modulate c-
548 Fos transcriptional bursts. Cell Rep. 2014;8(1):75-83. Epub 2014/07/02. doi:
549 10.1016/j.celrep.2014.05.053. PubMed PMID: 24981864; PubMed Central PMCID: PMCPMC5555219.
550 22. Furuta R, Yasunaga JI, Miura M, Sugata K, Saito A, Akari H, et al. Human T-cell leukemia virus
551 type 1 infects multiple lineage hematopoietic cells in vivo. PLoS Pathog. 2017;13(11):e1006722. Epub

552 2017/12/01. doi: 10.1371/journal.ppat.1006722. PubMed PMID: 29186194; PubMed Central PMCID:
553 PMCPMC5724899.
554 23. Ma G, Yasunaga J, Matsuoka M. Multifaceted functions and roles of HBZ in HTLV-1 pathogenesis.
555 Retrovirology. 2016;13:16. Epub 2016/03/17. doi: 10.1186/s12977-016-0249-x. PubMed PMID:
556 26979059; PubMed Central PMCID: PMCPMC4793531.
557 24. Mitobe Y, Yasunaga J, Furuta R, Matsuoka M. HTLV-1 bZIP Factor RNA and Protein Impart
558 Distinct Functions on T-cell Proliferation and Survival. Cancer Res. 2015;75(19):4143-52. Epub
559 2015/09/19. doi: 10.1158/0008-5472.CAN-15-0942. PubMed PMID: 26383166.
560 25. Macnamara A, Rowan A, Hilburn S, Kadolsky U, Fujiwara H, Suemori K, et al. HLA class I binding
561 of HBZ determines outcome in HTLV-1 infection. PLoS Pathog. 2010;6(9):e1001117. Epub 2010/10/05.
562 doi: 10.1371/journal.ppat.1001117. PubMed PMID: 20886101; PubMed Central PMCID:
563 PMCPMC2944806.
564 26. Raval GU, Bidoia C, Forlani G, Tosi G, Gessain A, Accolla RS. Localization, quantification and
565 interaction with host factors of endogenous HTLV-1 HBZ protein in infected cells and ATL. Retrovirology.
566 2015;12:59. Epub 2015/07/05. doi: 10.1186/s12977-015-0186-0. PubMed PMID: 26140924; PubMed
567 Central PMCID: PMCPMC4491271.
568 27. Mahgoub M, Yasunaga JI, Iwami S, Nakaoka S, Koizumi Y, Shimura K, et al. Sporadic on/off
569 switching of HTLV-1 Tax expression is crucial to maintain the whole population of virus-induced leukemic
570 cells. Proc Natl Acad Sci U S A. 2018;115(6):E1269-E78. Epub 2018/01/24. doi:
571 10.1073/pnas.1715724115. PubMed PMID: 29358408; PubMed Central PMCID: PMCPMC5819419.
572 28. Cook LB, Rowan AG, Melamed A, Taylor GP, Bangham CR. HTLV-1-infected T cells contain a
573 single integrated provirus in natural infection. Blood. 2012;120(17):3488-90. Epub 2012/09/08. doi:
574 10.1182/blood-2012-07-445593. PubMed PMID: 22955925; PubMed Central PMCID: PMCPMC3482858.

575 29. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV. Stochastic gene expression in a
576 lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell.
577 2005;122(2):169-82. Epub 2005/07/30. doi: 10.1016/j.cell.2005.06.006. PubMed PMID: 16051143.
578 30. Razooky BS, Cao Y, Hansen MMK, Perelson AS, Simpson ML, Weinberger LS. Nonlatching positive
579 feedback enables robust bimodality by decoupling expression noise from the mean. PLoS Biol.
580 2017;15(10):e2000841. Epub 2017/10/19. doi: 10.1371/journal.pbio.2000841. PubMed PMID:
581 29045398; PubMed Central PMCID: PMCPMC5646755.
582 31. Aull KH, Tanner EJ, Thomson M, Weinberger LS. Transient Thresholding: A Mechanism Enabling
583 Noncooperative Transcriptional Circuitry to Form a Switch. Biophys J. 2017;112(11):2428-38. Epub
584 2017/06/08. doi: 10.1016/j.bpj.2017.05.002. PubMed PMID: 28591615; PubMed Central PMCID:
585 PMCPMC5474885.
586 32. Nicot C, Dundr M, Johnson JM, Fullen JR, Alonzo N, Fukumoto R, et al. HTLV-1-encoded p30II is a
587 post-transcriptional negative regulator of viral replication. Nat Med. 2004;10(2):197-201. Epub
588 2004/01/20. doi: 10.1038/nm984. PubMed PMID: 14730358.
589 33. Rowan AG, Suemori K, Fujiwara H, Yasukawa M, Tanaka Y, Taylor GP, et al. Cytotoxic T
590 lymphocyte lysis of HTLV-1 infected cells is limited by weak HBZ protein expression, but non-specifically
591 enhanced on induction of Tax expression. Retrovirology. 2014;11:116. Epub 2014/12/17. doi:
592 10.1186/s12977-014-0116-6. PubMed PMID: 25499803; PubMed Central PMCID: PMCPMC4282740.
593 34 Miura M, Miyazato P, Satou Y, Tanaka Y, Bangham CRM. Epigenetic changes around the pX
594 region and spontaneous HTLV-1 transcription are CTCF-independent. Wellcome Open Res. 2018;3:105.
595 Epub 20187/12/11. doi: 10.12688/wellcomeopenres.14741.2. PubMed PMID: 30607369; PubMed
596 Central PMCID: PMCPMC6305241.
597 35. Clerc I, Polakowski N, Andre-Arpin C, Cook P, Barbeau B, Mesnard JM, et al. An interaction
598 between the human T cell leukemia virus type 1 basic leucine zipper factor (HBZ) and the KIX domain of

599 p300/CBP contributes to the down-regulation of tax-dependent viral transcription by HBZ. J Biol Chem.
600 2008;283(35):23903-13. Epub 2008/07/05. doi: 10.1074/jbc.M803116200. PubMed PMID: 18599479;
601 PubMed Central PMCID: PMCPMC3259792.
602 36 D'Agostino DM, Cavallari I, Romanelli MG, Ciminale V. Post-transcriptional Regulation of HTLV
603 Gene Expression: Rex to the Rescue. Front Microbiol. 2019;10:1958. Epub 2019/08/22. doi:
604 10.3389/fmicb. 2019.01958. PubMed PMID: 31507567; PubMed Central PMCID: PMCPMC 6714889.
605 37. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for
606 Transcriptional Control. Cell. 2017;169(1):13-23. Epub 2017/03/25. doi: 10.1016/j.cell.2017.02.007.
607 PubMed PMID: 28340338; PubMed Central PMCID: PMCPMC5432200.
608 38. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription Factors
609 Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell.
610 2018;175(7):1842-55 e16. Epub 2018/11/20. doi: 10.1016/j.cell.2018.10.042. PubMed PMID: 30449618;
611 PubMed Central PMCID: PMCPMC6295254.
612 39. Boxus M, Twizere JC, Legros S, Dewulf JF, Kettmann R, Willems L. The HTLV-1 Tax interactome.
613 Retrovirology. 2008;5:76. Epub 2008/08/16. doi: 10.1186/1742-4690-5-76. PubMed PMID: 18702816;
614 PubMed Central PMCID: PMCPMC2533353.
615 40. Semmes OJ, Jeang KT. Localization of human T-cell leukemia virus type 1 tax to subnuclear
616 compartments that overlap with interchromatin speckles. J Virol. 1996;70(9):6347-57. Epub 1996/09/01.
617 PubMed PMID: 8709263; PubMed Central PMCID: PMCPMC190661.
618 41. Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N. Computer control of microscopes using
619 microManager. Curr Protoc Mol Biol. 2010;Chapter 14:Unit14 20. Epub 2010/10/05. doi:
620 10.1002/0471142727.mb1420s92. PubMed PMID: 20890901; PubMed Central PMCID:
621 PMCPMC3065365.

622 42. Mueller F, Senecal A, Tantale K, Marie-Nelly H, Ly N, Collin O, et al. FISH-quant: automatic
623 counting of transcripts in 3D FISH images. Nat Methods. 2013;10(4):277-8. Epub 2013/03/30. doi:
624 10.1038/nmeth.2406. PubMed PMID: 23538861.
625 43. Gillespie DT. General Method for Numerically Simulating Stochastic Time Evolution of Coupled
626 Chemical-Reactions. J Comput Phys. 1976;22(4):403-34. doi: Doi 10.1016/0021-9991(76)90041-3.
627 PubMed PMID: WOS:A1976CQ87900001.
628 44 Rende F, Cavallari I, Corradin A, Silic-Benussi M, Toulza F, Toffolo GM, et al. Kinetics and
629 intracellular compartmentalization of HTLV-1 gene expression: nuclear retention of HBZ mRNAs. Blood.
630 2011;117(18):4855-9. Epub 2011/03/11. doi: 10.1182/ blood-2010-11-316463. PubMed PMID:
631 21398577; PubMed Central PMCID: PMCPMC 5292588.

632 ## Supporting information

633 **S1 Fig. Average HTLV-1 sense-strand transcripts in PBMCs in *ex vivo* culture over time.**

634 **S2 Fig. Determination of the threshold for mRNA spot detection.**

635 **S3 Fig. Accuracy of data prediction by dynamical models.**

636 **S4 Fig Parameter estimation.**

637 **S5 Fig. Antisense *HBZ* expression in *in vitro* maintained HTLV-1^{+} T cell clones.**

638 **S6 Fig. HTLV-1 antisense expression in PBMCs *ex vivo* culture over time.**

639 **S7 Fig. Decay-rate of the HTLV-1 sense-strand transcripts.**

640 **S1 Text. Model Comparison.**

641 **S2 Text. An alternative model accounting for the departure from the Poisson distribution.**

642 ## Figure legends

643 **Fig 1 Intense transcription of the HTLV-1 sense strand in PBMCs identified by smFISH.**

644 (A) HTLV-1 transcripts, smFISH probes and sample preparation for smFISH. Three sets of probes for
645 smFISH are indicated that hybridize to HTLV-1 transcripts: sense-strand in the *pX* region (Q670, red
646 stars), *HBZ* (Q570, yellow stars), and unspliced sense transcripts containing *gag* region (FAM, green
647 stars). (B) Representative images of HTLV-1$^+$ PBMCs with the sense-strand transcripts at the indicated
648 time points. Blue area indicates the nucleus stained with DAPI, and red spots indicate the HTLV-1 sense
649 transcripts. Scale bar (white) = 5 µm. (C) A site of intense transcription identified by the probes for *gag*.
650 The image on the left shows the *gag* staining (green). The image overlaid with sense transcript spots
651 (red, shown in the middle) is presented on the right. The cell presented is from a HAM patient coded
652 TDZ sampled at 4 hours of incubation.

653

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

655 (A) Schematic diagram of the identification of the HTLV-1$^+$ population for analysis. The entire population
656 (in the black and grey rectangles) are the cells identified and analysed by microscopy. The size of the
657 HTLV-1-infected population (black rectangle) was estimated by multiplying the total cell count by the
658 proviral load as measured by ddPCR. Also see Materials and Methods. (B) Sense-transcript count per
659 HTLV-1$^+$ cell (bin size = 20, note the units) at each time point for the two patients with HAM (TBC and
660 TDZ) and two with ATL (LGZ and LGB). Note that the first few bins in each case are out of scale on the y-
661 axis. In the inset for each histogram is indicated the number of cells analysed. (C) The mean and variance
662 of the sense transcript count per HTLV-1$^+$ cell obtained from four patients, each containing nine time
663 points.

664

665

666 **Fig 3 Rate of onset of HTLV-1 reactivation.**

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

671

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

673 (A) Kinetic model of HTLV-1 proviral transcription, with three promoter states: Off, On, and On-Tax. (B)
674 Stochastic simulation of the model presented in panel A. The best 50 fits (mean and standard deviation
675 shown in black) are superimposed on the smFISH data from the ATL patient LGZ (red bars, reproduced
676 from Fig 2B). (C) The marginal distributions of the best 50 parameters (posterior; green bars); the
677 respective prior distributions are shown in blue. The mean and standard deviation are indicated in the
678 inset. (D) The interval between the initiation of the On-Tax promoter state and the direct transition from
679 the On-Tax to the Off state in the simulation. The frequency of the interval was fitted with an
680 exponential decay, and its probability density function is presented. The mean intervals (τ) ± standard
681 deviation for the prior set I (Table 2) and the prior set II (S4 Fig) are indicated in the inset.

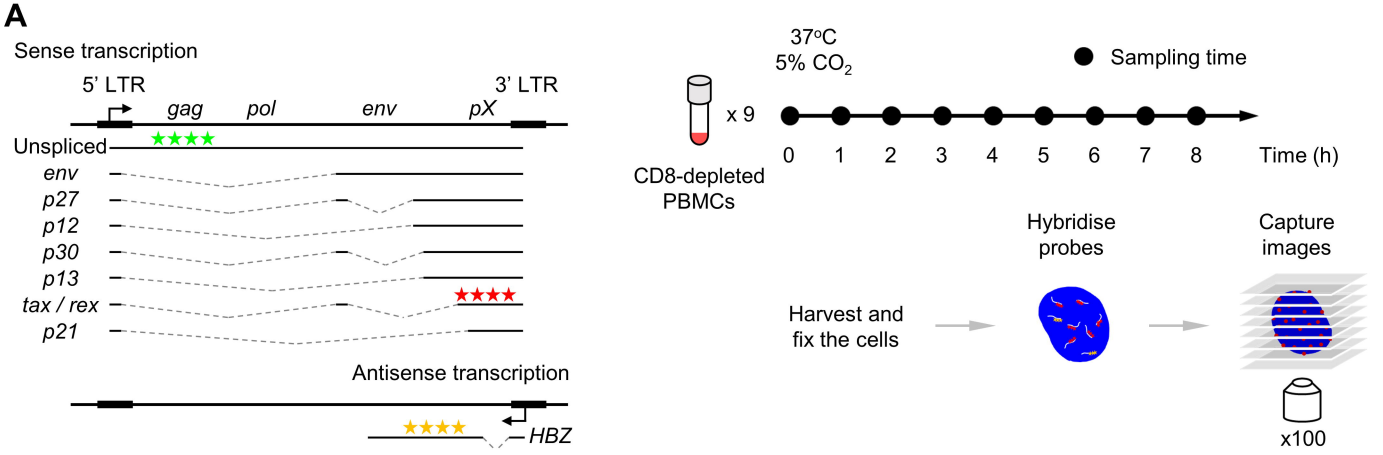
682

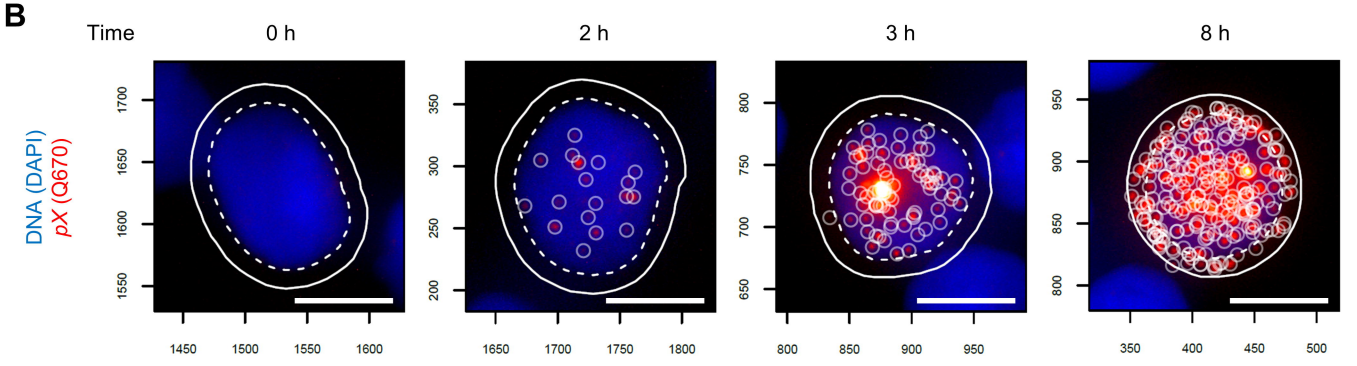
683 **Fig 5 Identification of HTLV-1 antisense transcripts by smFISH.**

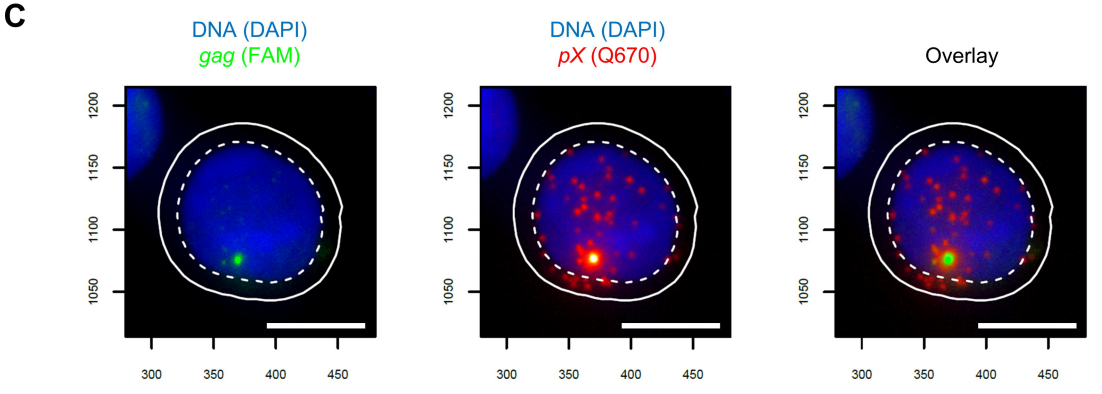
684 (A) Typical image captured by smFISH of *HBZ* mRNA (yellow spot) in a cell (from a patient with HAM,
685 coded TDZ) without *ex vivo* incubation. Blue indicates the nucleus stained with DAPI. Scale bar (white) =
686 5 µm. (B) Frequency distribution of the number of *HBZ* mRNA molecules carried per HTLV-1$^+$ PBMC.
687 Yellow bars show the frequencies observed by smFISH; the number of HTLV-1$^+$ cells analysed (N) and the
688 average number of *HBZ* molecules per HTLV-1$^+$ cell (<m>) are indicated in the inset. Black lines indicate

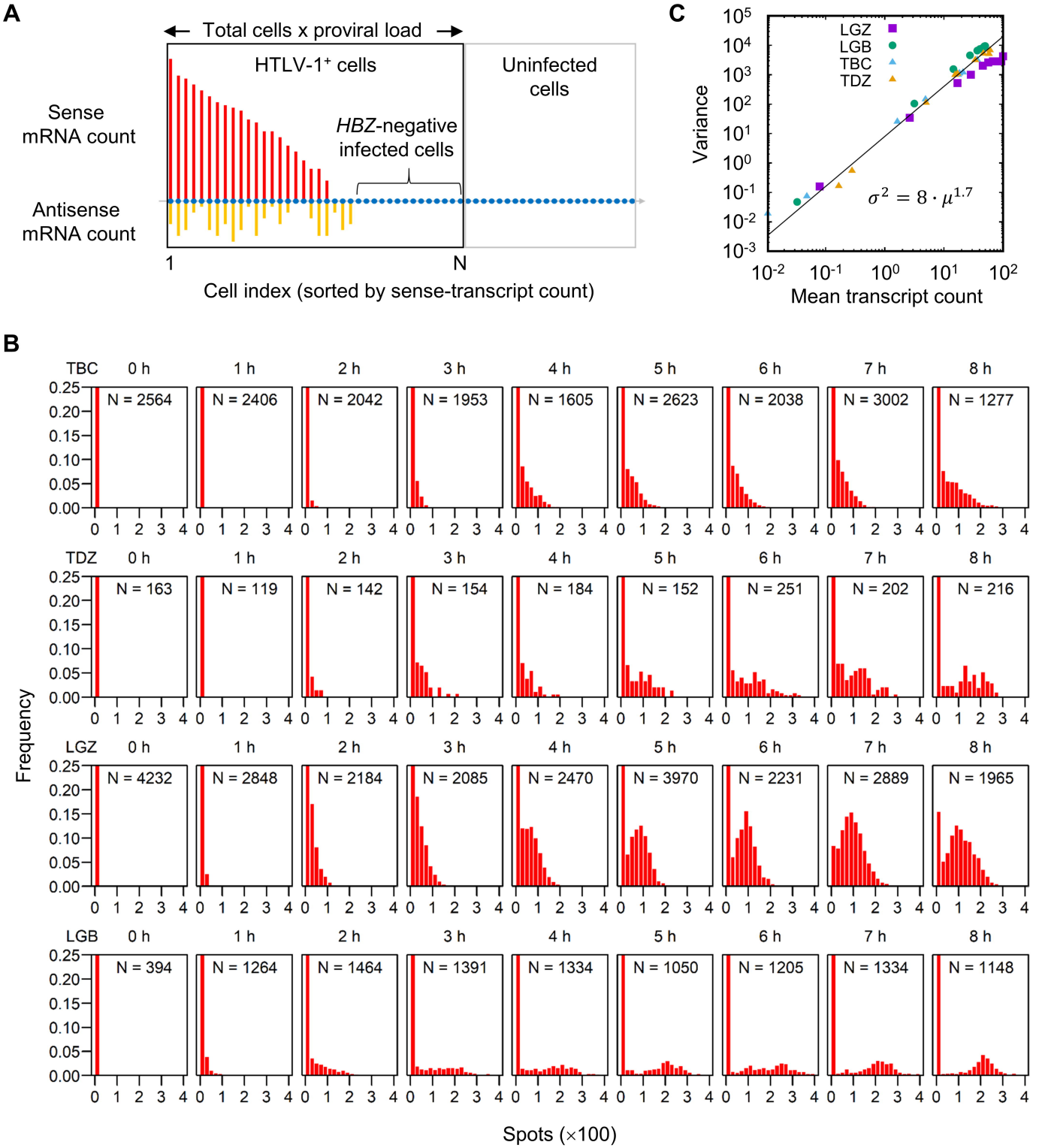
689 the expected frequencies from a Poisson distribution with the parameter <m>. (C) The kinetic model for

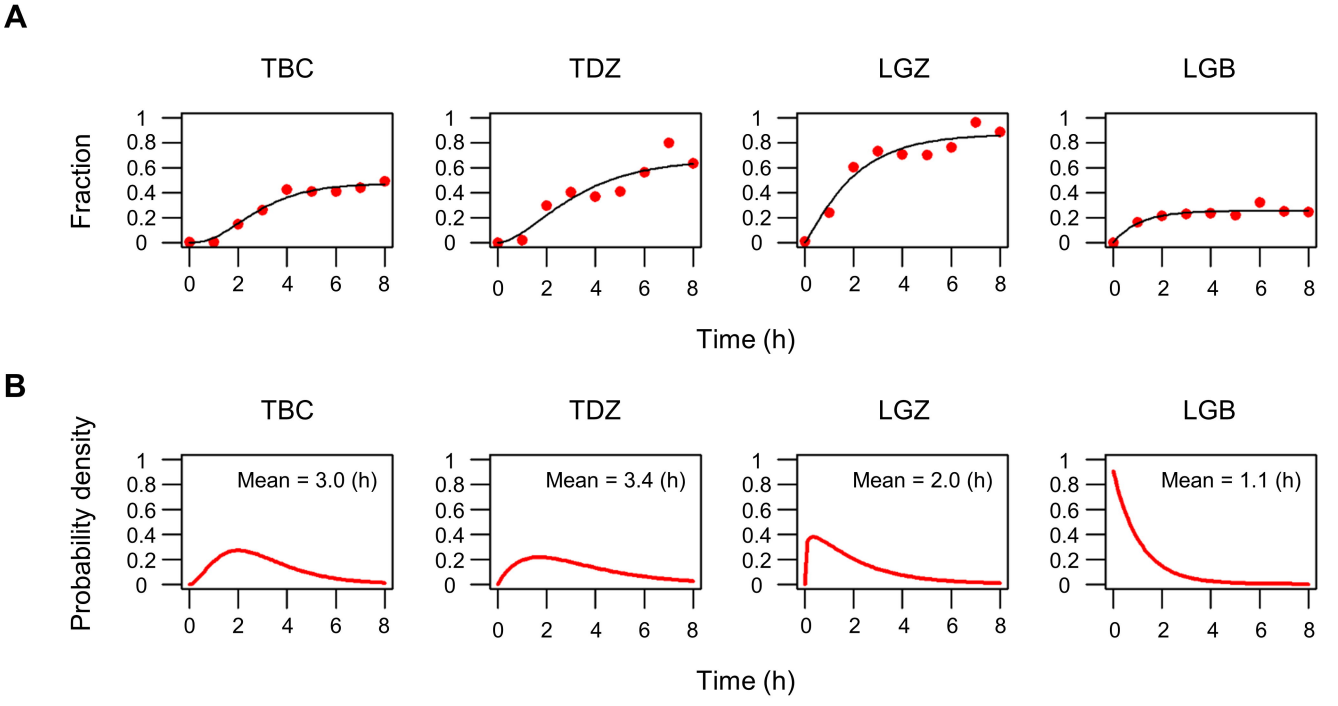
690 the antisense transcription. *HBZ* mRNA is transcribed at the rate *k*, and degraded at the rate δ.

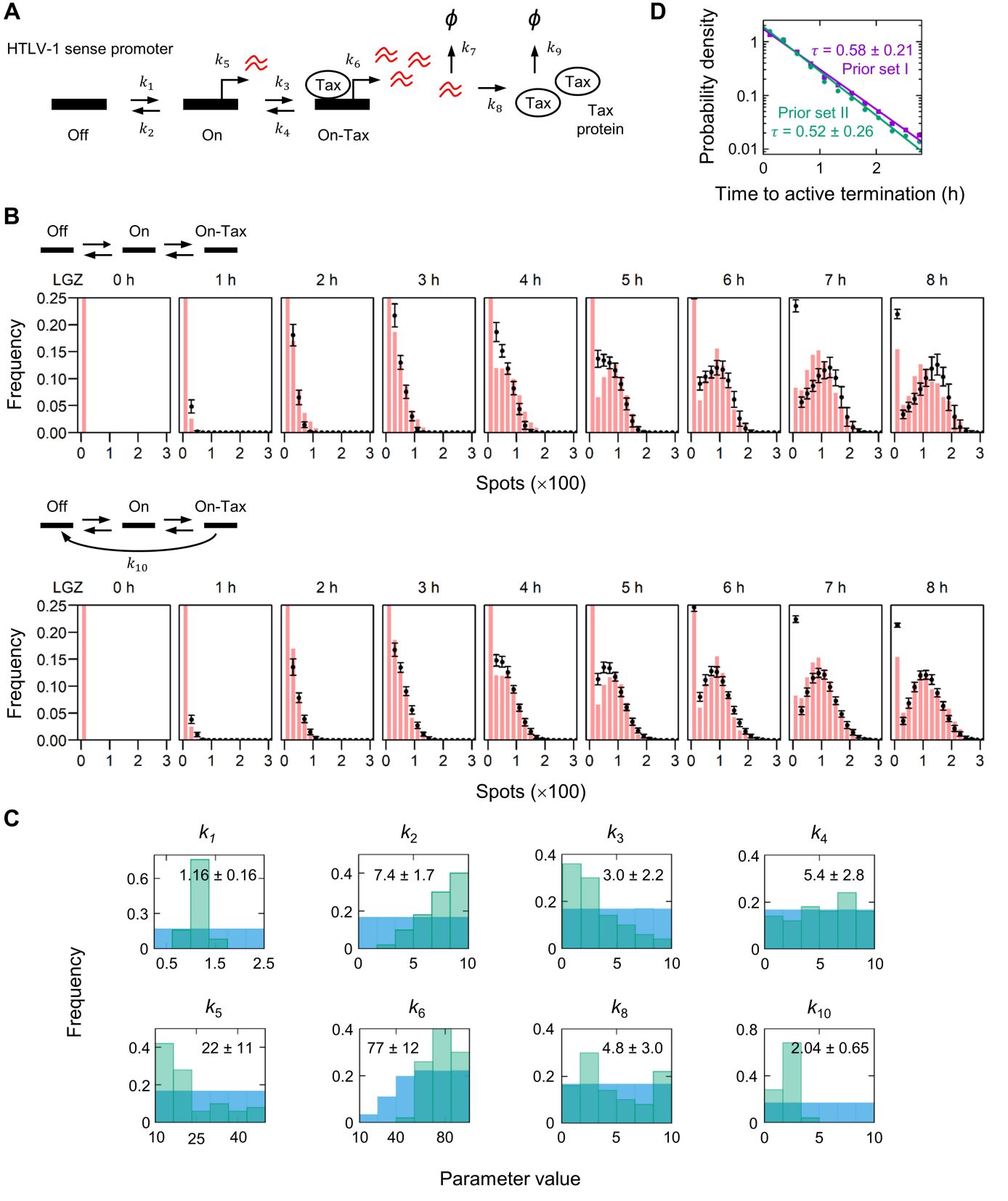

A
Sense transcription
5' LTR
3' LTR
gag
pol
env
pX
Unspliced
env
p27
p12
p30
p13
tax / rex
p21
Antisense transcription
HBZ
37°C
5% CO2
Sampling time
x 9
CD8-depleted PBMCs
0 1 2 3 4 5 6 7 8
Time (h)
Hybridise probes
Capture images
Harvest and fix the cells
x100

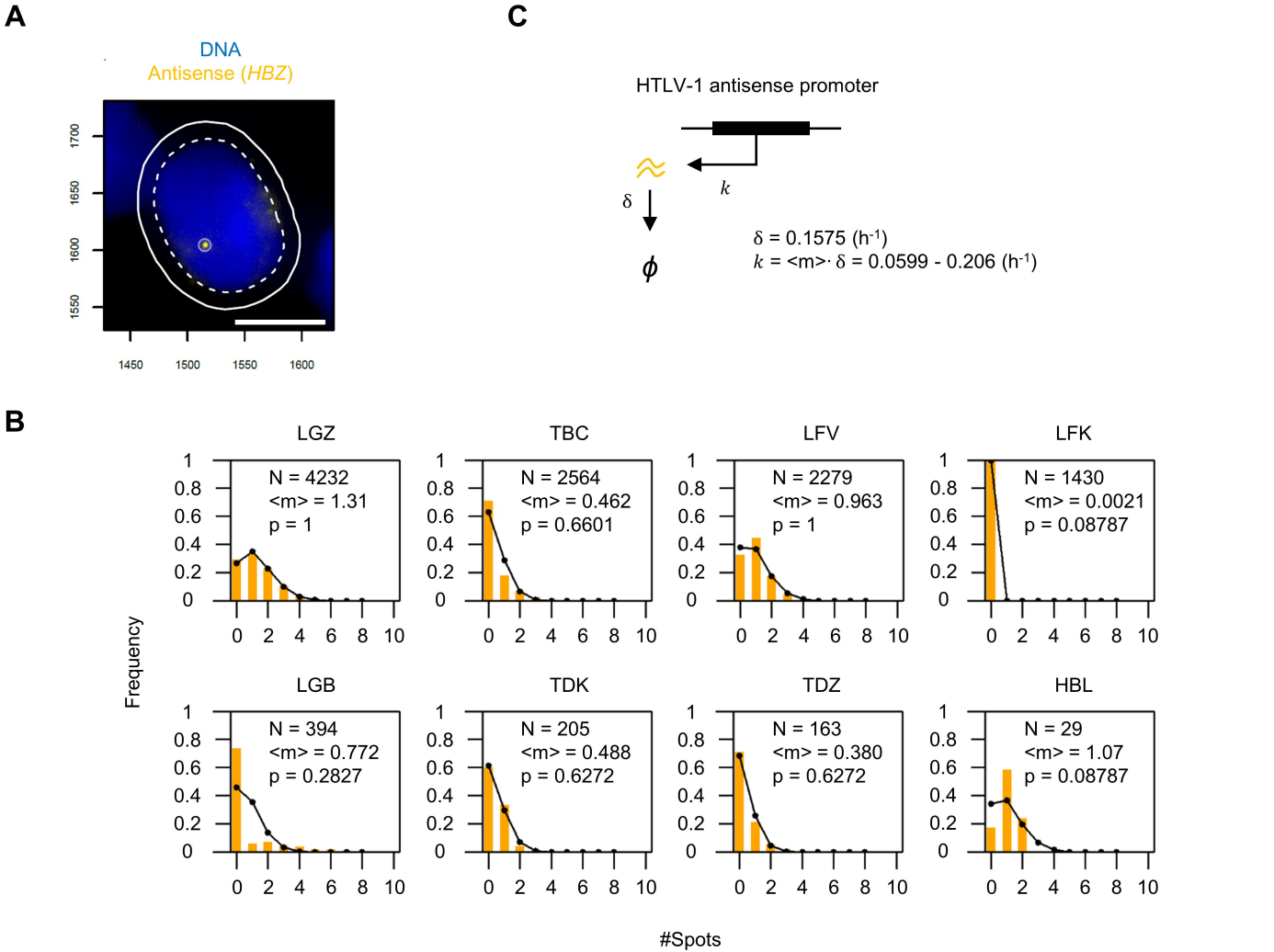

B
Time
0 h
2 h
3 h
8 h
DNA (DAPI)
pX (Q670)


C
DNA (DAPI)
gag (FAM)
DNA (DAPI)
pX (Q670)
Overlay


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


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


A
HTLV-1 sense promoter
Off
On
On-Tax
Tax
Tax protein
B
Off
On
On-Tax
LGZ
0 h
1 h
2 h
3 h
4 h
5 h
6 h
7 h
8 h
Frequency
Spots (×100)
C
Parameter value
1.16 ± 0.16
7.4 ± 1.7
3.0 ± 2.2
5.4 ± 2.8
22 ± 11
77 ± 12
4.8 ± 3.0
2.04 ± 0.65
D
Probability density
Time to active termination (h)
τ = 0.58 ± 0.21
Prior set I
Prior set II
τ = 0.52 ± 0.26


A
DNA
Antisense (HBZ)
1700
1650
1600
1550
1450
1500
1550
1600
C
HTLV-1 antisense promoter
δ
k
ϕ
δ = 0.1575 (h-1)
k = <m>· δ = 0.0599 - 0.206 (h-1)
B
LGZ
N = 4232
<m> = 1.31
p = 1
TBC
N = 2564
<m> = 0.462
p = 0.6601
LFV
N = 2279
<m> = 0.963
p = 1
LFK
N = 1430
<m> = 0.0021
p = 0.08787
LGB
N = 394
<m> = 0.772
p = 0.2827
TDK
N = 205
<m> = 0.488
p = 0.6272
TDZ
N = 163
<m> = 0.380
p = 0.6272
HBL
N = 29
<m> = 1.07
p = 0.08787
Frequency
#Spots