



# SCIENTIFIC REPORTS



OPEN

## Spontaneous reactivation of latent HIV-1 promoters is linked to the cell cycle as revealed by a genetic-insulators-containing dual-fluorescence HIV-1-based vector

Yik Lim Kok<sup>1,2</sup>, Stefan Schmutz<sup>2</sup>, Anne Inderbitzin<sup>1,2,3</sup>, Kathrin Neumann<sup>1,2</sup>, Audrey Kelley<sup>1,2,3</sup>, Lisa Jörmann<sup>1,2</sup>, Mohaned Shilaih<sup>1,2</sup>, Valentina Vongrad<sup>1,2</sup>, Roger D. Kouyos<sup>1,2</sup>, Huldrych F. Günthard<sup>1,2</sup> , Christian Berens<sup>1,2</sup>  & Karin J. Metzner<sup>1,2</sup>

Long-lived latently HIV-1-infected cells represent a barrier to cure. We developed a dual-fluorescence HIV-1-based vector containing a pair of genetic insulators flanking a constitutive fluorescent reporter gene to study HIV-1 latency. The protective effects of these genetic insulators are demonstrated through long-term (up to 394 days) stable fluorescence profiles in transduced SUP-T1 cells. Analysis of 1,941 vector integration sites confirmed reproduction of HIV-1 integration patterns. We sorted monoclonal cells representing latent HIV-1 infections and found that both vector integration sites and integrity of the vector genomes influence the reactivation potentials of latent HIV-1 promoters. Interestingly, some latent monoclonal cells exhibited a small cell subpopulation with a spontaneously reactivated HIV-1 promoter. Higher expression levels of genes involved in cell cycle progression are observed in these cell subpopulations compared to their counterparts with HIV-1 promoters that remained latent. Consistently, larger fractions of spontaneously reactivated cells are in the S and G2 phases of the cell cycle. Furthermore, genistein and nocodazole treatments of these cell clones, which halted cells in the G2 phase, resulted in a 1.4–2.9-fold increase in spontaneous reactivation. Taken together, our HIV-1 latency model reveals that the spontaneous reactivation of latent HIV-1 promoters is linked to the cell cycle.

Upon entry into a CD4<sup>+</sup> T cell, the human immunodeficiency virus type 1 (HIV-1) integrates its reverse-transcribed viral DNA into the host's genome<sup>1</sup>. The integrated provirus has two fates: it either continues its replication cycle to produce progeny virions or remains latent in the host cell<sup>1</sup>. The latent HIV-1 reservoir is unsusceptible to both the host individual's immune system and antiretroviral therapy (ART), which is currently only effective against active infections<sup>2</sup>. More importantly, ART cessation leads to rebound of HIV-1, thus necessitating lifelong therapy<sup>3</sup>.

Studies examining features driving the establishment and maintenance of HIV-1 latency have been limited by the low frequencies of cells latently infected with replication-competent HIV-1 in patients (~1–10<sup>2</sup> per 10<sup>6</sup> CD4<sup>+</sup> T cells)<sup>4,5</sup> and the lack of phenotypic markers to identify these cells<sup>6</sup>. To circumvent these obstacles, *in vitro* models were developed to recapitulate HIV-1 infection and latency. Earlier models used HIV-1-based vectors encoding one fluorescent reporter gene to transduce and subsequently identify cells harbouring an active or latent HIV-1 promoter, *i.e.*, those which expressed the fluorescent reporter gene constitutively or only when activated externally, respectively<sup>7–9</sup>. Dual-fluorescence vectors were later developed to identify cells harbouring a latent HIV-1 promoter without the need for external activation<sup>10–12</sup>. These second generation vectors have an additional

<sup>1</sup>Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland. <sup>2</sup>Institute of Medical Virology, University of Zurich, Zurich, Switzerland. <sup>3</sup>Life Science Zurich Graduate School, University of Zurich, Zurich, Switzerland. <sup>4</sup>Institute of Molecular Pathogenesis, Friedrich-Loeffler-Institut, Jena, Germany. Correspondence and requests for materials should be addressed to K.J.M. (email: [karin.metzner@usz.ch](mailto:karin.metzner@usz.ch))

fluorescent reporter gene whose expression is driven by a constitutive promoter to identify transduced cells irrespective of their HIV-1 promoter activity.

These models have proved to be valuable in understanding molecular features of HIV-1 latency. For instance, it has been shown that vector integration sites have an influence on the HIV-1 promoter activity<sup>7,13,14</sup>, and Dahabieh *et al.* showed that the HIV-1 promoter activity correlates with the degree of activation of cells during transduction<sup>11</sup>. Sherrill-Mix *et al.* recently showed that the association between HIV-1 latency and chromosomal position is model dependent<sup>15</sup>. This is likely a consequence of the heterogeneous nature of the latent HIV-1 reservoir<sup>16</sup>, thus arguing for novel models to complement existing ones in order to unravel key determinants of HIV-1 latency-associated events.

We have developed a novel HIV-1-based vector (LTatC[M]) consisting of two fluorescent reporter gene cassettes: Cerulean and mCherry (Fig. 1a). The expression of Cerulean is driven by the HIV-1 promoter in the viral 5' long terminal repeat (5' LTR) and supported by the HIV-1 transactivator of transcription (Tat) via a positive feedback loop whereas the expression of mCherry is driven by the constitutive human elongation initiation factor 4A1 (*heIF4A1*) promoter (Fig. 1a). As such, Cerulean reports the activity of the HIV-1 promoter (active or latent) and mCherry labels transduced cells, enabling the identification of cells harbouring an active HIV-1 promoter and simultaneously distinguishing cells harbouring a latent HIV-1 promoter from untransduced cells. Additionally, the mCherry cassette is flanked by a pair of genetic insulators, the chicken hypersensitive site 4 core (cHS4) and synthetic matrix attachment region (sMAR) (Fig. 1a), to prevent (i) transcriptional interference between Cerulean and mCherry<sup>17–19</sup> that might lead to artefactual HIV-1 latency and (ii) silencing of mCherry due to position-effect variegation<sup>20–25</sup>, thus ensuring its long-term constitutive expression.

In the present study, we first characterized our vector LTatC[M] and verified its capacity to reproduce features of active and latent HIV-1 infections. We then transduced SUP-T1 cells with LTatC[M] pseudotyped with vesicular stomatitis virus glycoprotein G (VSV-G) and subsequently sorted out bulk cell populations, as well as monoclonal cells, representing active and latent HIV-1 infections for further analysis. With our model, we observed that some monoclonal cells constantly exhibited a spontaneously reactivated HIV-1 promoter and found that this phenomenon is linked to the cell cycle.

## Results

### Transduction of SUP-T1 cells with LTatC[M] yields active and latent HIV-1 phenotypes.

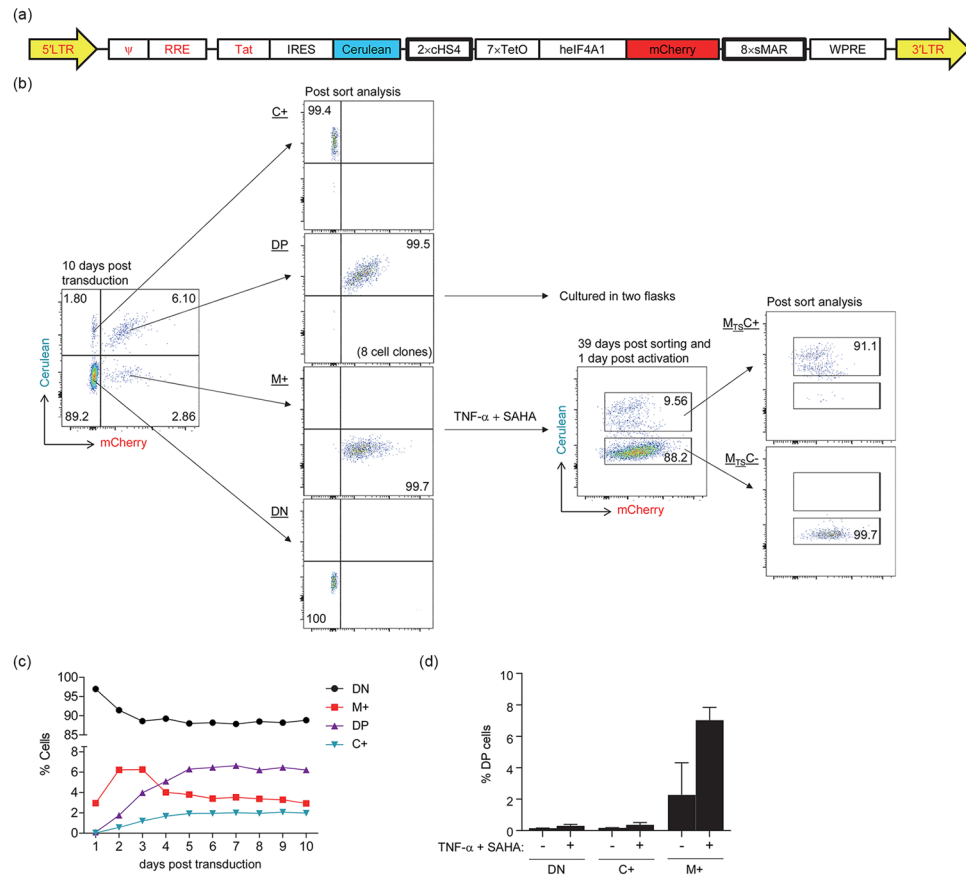
Transduction of SUP-T1 cells with VSV-G-pseudotyped LTatC[M] (Fig. 1a) yielded four cell populations: double positive (DP), single Cerulean positive (C+), single mCherry positive (M+), and double negative (DN) (Fig. 1b). The HIV-1 promoter in DP and C+ cells was active since Cerulean was expressed, and thus they represented active HIV-1 infections. Correspondingly, latent HIV-1 infections were represented by M+ cells due to their lack of Cerulean expression (Fig. 1b). Henceforth, DP and C+ will be called active and M+ will be called latent to denote the state of the HIV-1 promoter harboured in these cells and not of the host cells.

As we were interested in cell populations with stable phenotypes representing active and latent HIV-1 infections, we examined the kinetics of the emergence of the four cell populations: DP, C+, M+, and DN, for 10 days post transduction (Fig. 1c; Supplementary Fig. S1). The majority of transduced cells appeared to be latent up to 3 days post transduction, although this phenomenon was transient, and a stable fluorescence profile for all transduced cells was established only at 5 days post transduction. The initial fluctuations in fluorescence could be a result of differential temporal regulation of gene expression post integration at different vector integration sites and spontaneous reactivation of latent HIV-1 promoters in M+ cells. Similar kinetics were observed between transduction efficiencies of 5% and 15% (Fig. 1c; Supplementary Fig. S1). Therefore, we sorted with flow cytometry all four cell populations at 10 days post transduction with a purity of >99% for subsequent analyses (Fig. 1b).

Next, DN, C+, and M+ cell populations sorted from LTatC[M]-transduced SUP-T1 cells were treated with TNF- $\alpha$  and SAHA for 24 hours and the induction of DP cells from these cell populations was measured with flow cytometry. Approximately 7% of M+ cells expressed Cerulean and mCherry after induction with TNF- $\alpha$  and SAHA (Fig. 1d), showing that latent M+ cells could be activated to become double positive. The expression of either Cerulean or mCherry in DN cells and mCherry in C+ cells could not be induced with TNF- $\alpha$  and SAHA (Fig. 1d).

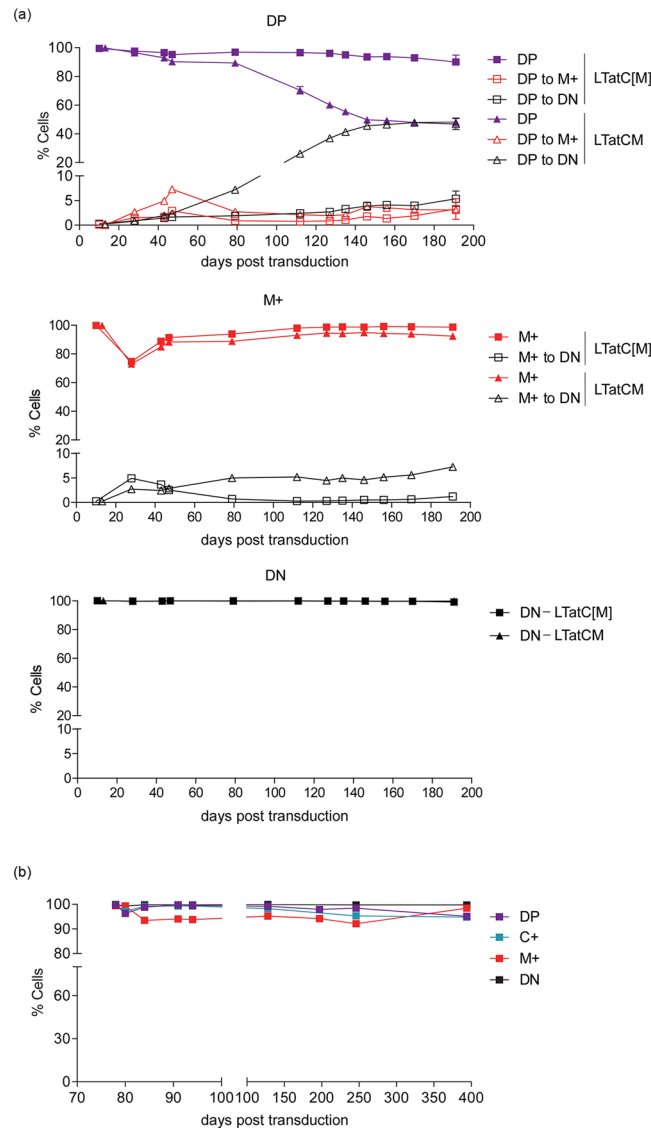
### The genetic insulators, cHS4 and sMAR, confer long-term stability of fluorescence profiles in LTatC[M]-transduced SUP-T1 cells.

The short-term kinetics and ratio of active (DP and C+) to latent cells (M+) were comparable between LTatC[M] and a vector variant with no genetic insulators flanking the mCherry cassette (LTatCM) (Supplementary Fig. S1), demonstrating that the genetic insulators did not influence the emergence of latent cells. As the present study focused on the long-term maintenance and persistence of HIV-1 latency, we examined whether the genetic insulators were capable of protecting the mCherry cassette from being silenced over time by monitoring the fluorescence profiles of the sorted DP, M+, and DN cell populations transduced with LTatC[M] or LTatCM up to 191 days post transduction. Over 90% of cell populations transduced with LTatC[M] retained their initial fluorescence profiles for the entire period analysed (Fig. 2). In contrast, mCherry expression in DP cells transduced with LTatCM decreased gradually down to 48% at 191 days post transduction (Fig. 2a). The decrease in mCherry expression in LTatCM-transduced M+ cells was less drastic, although it remained consistently lower by ~5% compared to the LTatC[M]-transduced M+ cells (Fig. 2a). Our results were reproducible in a second transduction of SUP-T1 cells with LTatC[M]. Consistently, over 90% of all sorted cell populations in the second transduction retained their initial fluorescence profiles analysed for 394 days post transduction (Fig. 2b). Therefore, the pair of genetic insulators, cHS4 and sMAR, flanking the mCherry cassette confer long-term protection to mCherry expression from silencing.



**Figure 1.** Transduction of SUP-T1 cells with LTatC[M], a dual-fluorescence HIV-1-based vector. **(a)** Schematic diagram of LTatC[M]. The expression of Cerulean is driven by HIV-1 5' long terminal repeat (LTR) and supported by HIV-1 transactivator of transcription (Tat) via a positive feedback loop. The expression of mCherry is driven by the constitutive human elongation initiation factor 4A1 (*helf4A1*) promoter and the entire cassette is protected from position-effect variegation by a pair of genetic insulators: two copies of the chicken hypersensitive site 4 core (cHS4) and eight copies of synthetic matrix attachment region (sMAR). ψ: HIV-1 packaging signal; RRE; HIV-1 Rev response element; IRES: Internal ribosomal entry site; TetO: Tet operator; WPRE: Woodchuck hepatitis virus post-transcriptional regulatory element. **(b)** Flow cytometric sorting strategy and purity of transduced SUP-T1 cell populations. SUP-T1 cells transduced with LTatC[M] were sorted 10 days post transduction into four cell populations: single Cerulean positive (C+), double positive (DP), single mCherry positive (M+), and double negative (DN). DP cells were divided equally and cultured in two flasks. A portion of M+ cells was expanded in culture for 39 days and then activated with TNF- $\alpha$  and SAHA for one day prior to sorting into cell populations that became Cerulean<sup>+</sup> (M<sub>T<sub>S</sub></sub>C+) and that remained Cerulean<sup>-</sup> (M<sub>T<sub>S</sub></sub>C-). **(c)** Fluorescence profiles of the four cell populations that arose from transduction of SUP-T1 cells with LTatC[M]: DN, M+, DP, and C+, were measured with flow cytometry daily for 10 days. Each datapoint represents the mean of two independent transductions (n = 2) and error bars depict standard error means. Some error bars are within datapoints. **(d)** DN, C+, and M+ cell populations sorted from LTatC[M]-transduced SUP-T1 cells were treated with TNF- $\alpha$  and SAHA for one day and the induction of DP cells from these cell populations was measured with flow cytometry. In the case of DN cell population, the induction of C+ and M+ cells was also measured but not distinguished from DP cell induction. The means of two independent measurements (n = 2) are indicated by the heights of the bars and standard error means are indicated by the error bars.

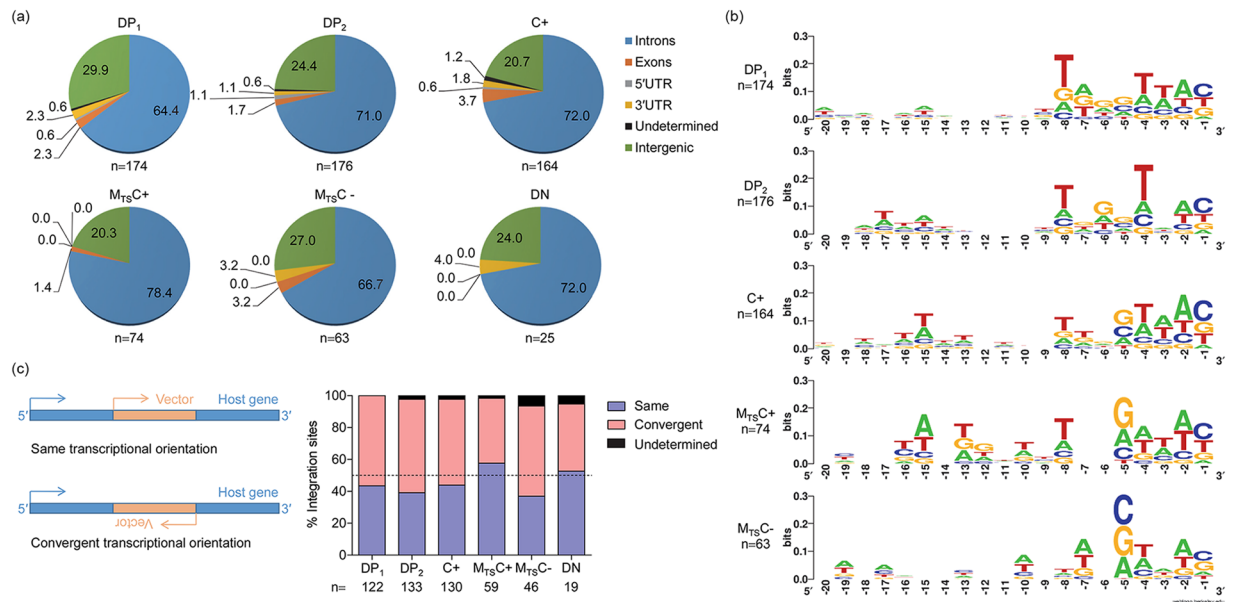
**LTatC[M] reproduces HIV-1 integration site patterns.** To ensure that the genetic-insulators-containing LTatC[M] reproduced integration site patterns of HIV-1, vector integration sites from the sorted cell populations, DP, C+, M<sub>T<sub>S</sub></sub>C+, M<sub>T<sub>S</sub></sub>C-, and DN, were sequenced and mapped with the Integration Site Analysis Pipeline (InStAP)<sup>26</sup>. M<sub>T<sub>S</sub></sub>C+ and M<sub>T<sub>S</sub></sub>C- cells were sorted from latent M+ cells in which Cerulean expression was inducible and non-inducible with TNF- $\alpha$  and SAHA, respectively (Fig. 1b). A total of 676 unique LTatC[M] integration sites were analysed. Across all cell populations, the majority of vector integration sites were found in transcription units, although slightly more in the inducible M<sub>T<sub>S</sub></sub>C+ cells (Fig. 3a), and in a convergent transcriptional orientation relative to vector-hosting genes, except for M<sub>T<sub>S</sub></sub>C+ cells (Fig. 3c). The HIV-1 signature weakly conserved palindromic sequence was also observed at vector integration sites for all cell populations (Fig. 3b). These features were reproducible in a second transduction experiment in which a total of 1,265 unique vector integration sites were analysed (Supplementary Fig. S2).



**Figure 2.** Longitudinal fluorescence profiles of transduced SUP-T1 cell populations. **(a)** SUP-T1 cells were transduced with LTatC[M] (squares) or a vector variant with no genetic insulators in the mCherry cassette (LTatCM) (triangles). Double positive (DP), single mCherry positive (M+), and double negative (DN) cell populations were sorted 10 days post transduction and their fluorescence profiles were measured with flow cytometry for 191 days post transduction. Shaded symbols depict percentages of cells with the fluorescence initially sorted for whereas open symbols depict percentages of cells that became DN and M+. Each datapoint of the DP data set represents the mean of duplicate cultures ( $n = 2$ ) and error bars depict standard error means. Some error bars are within datapoints. **(b)** Longitudinal fluorescence profiles of sorted SUP-T1 cell populations up to 394 days post transduction in a second transduction experiment.

**LTatC[M] integration site selection and genome integrity influence reactivation potentials of HIV-1 promoters in TNF- $\alpha$  and SAHA-responsive ( $M_{TS}C+$ ) and -non-responsive ( $M_{TS}C-$ ) single mCherry positive cell clones.** The influence of HIV-1 integration sites on the reactivation potentials of latent HIV-1 promoters has been demonstrated previously<sup>7,13,14</sup>. To further explore this in our model, we sorted from latent cell populations (M+) from two transductions: 22 latent cell clones in which Cerulean expression was inducible with TNF- $\alpha$  and SAHA ( $M_{TS}C+$ ) and 18 latent cell clones in which Cerulean expression was non-inducible with TNF- $\alpha$  and SAHA ( $M_{TS}C-$ ), as well as 8 active cell clones (DP) (Fig. 1b). Vector integration sites were sequenced for all cell clones, and cell clones with identical vector integration sites were treated as biological replicates, resulting in 7  $M_{TS}C+$ , 13  $M_{TS}C-$ , and 6 DP independent cell clones.

We subsequently evaluated the reactivation potentials of latent HIV-1 promoters after expansion of  $M_{TS}C+$  and  $M_{TS}C-$  cell clones in culture without TNF- $\alpha$  and SAHA, during which  $M_{TS}C+$  cell clones reverted to their initial phenotype of single mCherry positive. Surprisingly, highly variable reactivation potentials (26–89%) were observed among the seven  $M_{TS}C+$  cell clones with distinct vector integration sites (Fig. 4a; Table 1). To examine the effects of vector-hosting genes' activity upon HIV-1 infection on the reactivation potentials of latent HIV-1



**Figure 3.** Integration site patterns of LTatC[M] in various SUP-T1 cell populations. **(a)** Genomic distributions of integrated LTatC[M]. **(b)** Consensus sequences of 20 nucleotides upstream of 5'LTR. Double negative (DN) is not shown due to low number of integration sites. **(c)** Transcription orientations of intragenic LTatC[M] relative to its host genes. DP<sub>1</sub>: double positive replicate 1; DP<sub>2</sub>: double positive replicate 2; C+: single Cerulean; M<sub>TS</sub>C+: TNF- $\alpha$  and SAHA-responsive single mCherry positive; M<sub>TS</sub>C-: TNF- $\alpha$  and SAHA-non-responsive single mCherry positive.

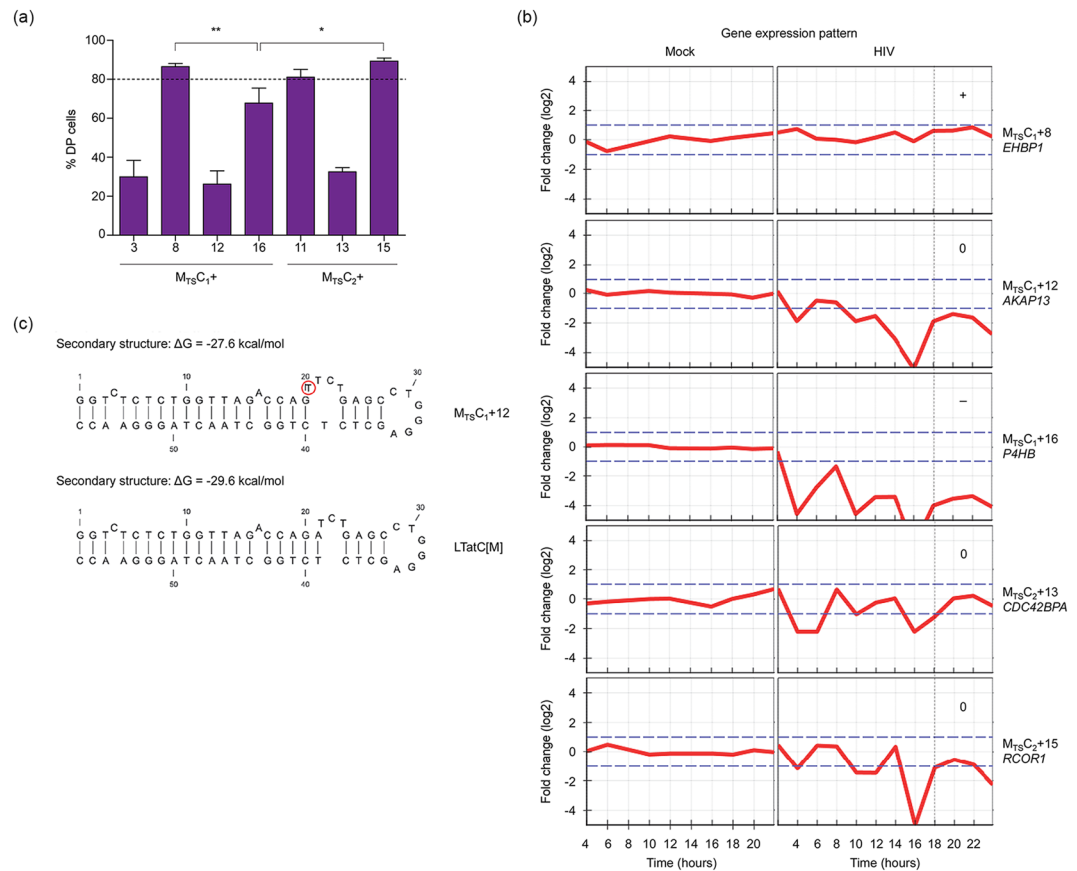
promoters, we used data from the Patterns of Expression and Analysis of Clusters of HIV/Host interactions (PEACHi) database<sup>27</sup> (Fig. 4b). Data for M<sub>TS</sub>C<sub>1</sub>+3 (intergenic) and M<sub>TS</sub>C<sub>2</sub>+11 (*CTNND1/TMX2-CTNND1*) were not available on PEACHi. Of the five gene expression patterns for which data were available, *EHBPI* (M<sub>TS</sub>C<sub>1</sub>+8), *AKAP13* (M<sub>TS</sub>C<sub>1</sub>+12), *P4HB* (M<sub>TS</sub>C<sub>1</sub>+16), *CDC42BPA* (M<sub>TS</sub>C<sub>2</sub>+13), and *RCOR1* (M<sub>TS</sub>C<sub>2</sub>+15), only *P4HB* was significantly downregulated (~8-fold;  $P < 0.01$ ) post HIV-1 integration, *i.e.* > 18 hours post infection<sup>27</sup> (Fig. 4b). While this could explain why M<sub>TS</sub>C<sub>1</sub>+16 had a significantly lower reactivation potential than M<sub>TS</sub>C<sub>1</sub>+8 and M<sub>TS</sub>C<sub>1</sub>+15, it did not explain the much lower reactivation potentials of M<sub>TS</sub>C<sub>1</sub>+12 and M<sub>TS</sub>C<sub>2</sub>+13 since the vector-hosting genes in these cell clones were not significantly downregulated post HIV-1 integration (Fig. 4a,b).

Therefore, we examined the Cerulean cassettes of all cell clones to determine whether mutations contributed to low reactivation potentials of latent HIV-1 promoters. None (0/6) of the DP cell clones analysed had any mutations in their Cerulean cassettes whereas mutations were found in 5/7 M<sub>TS</sub>C+ cell clones (Fig. 5). Notably, M<sub>TS</sub>C<sub>1</sub>+12 had a mutation in the HIV-1 transactivation response (TAR) element, which was predicted to disrupt the 3-nucleotide bulge essential for HIV-1 Tat binding and subsequent transcription elongation from the HIV-1 promoter<sup>28,29</sup> (Fig. 4c), and M<sub>TS</sub>C<sub>2</sub>+13 had numerous mutations throughout its HIV-1 5' LTR (Supplementary Table S1). Mutations in these cell clones could account for their low reactivation potentials. The mutations in the HIV-1 Tat region found in M<sub>TS</sub>C<sub>1</sub>+8 and M<sub>TS</sub>C<sub>1</sub>+16 (Fig. 5; Supplementary Table S1) have been reported to have wild-type transactivation activities<sup>30,31</sup>. Interestingly, no mutation was found in M<sub>TS</sub>C<sub>1</sub>+3 and M<sub>TS</sub>C<sub>2</sub>+15 while the reactivation potentials of latent HIV-1 promoters in these clones differed by ~60% (Fig. 4a), further showing the influence of vector integration sites on the reactivation potentials of latent HIV-1 promoters. Taken together, our data provide evidence that the reactivation potentials of latent HIV-1 promoters are influenced by both vector integration sites and integrity of the Cerulean cassettes.

We then examined the Cerulean cassettes of the 18 M<sub>TS</sub>C- cell clones, the expression of which was not inducible with TNF- $\alpha$  and SAHA. The Cerulean cassettes of 17/18 cell clones contained large internal deletions in the HIV-1 Tat and/or Cerulean region (Fig. 5). Although cell clone MT<sub>2</sub>-5 did not have internal deletions, multiple mutations were found throughout its Cerulean cassette (Fig. 5), *e.g.* one in the stem of HIV-1 TAR and another resulting in a C37Y amino acid substitution in HIV-1 Tat, which likely abolished transactivation by Tat<sup>28,29,31</sup>. Thus, defective vector genome is the main cause of non-inducible HIV-1 promoters in M<sub>TS</sub>C- cell clones. This phenomenon has also been observed in cells from HIV-1-infected individuals<sup>4</sup>.

**Higher expression of a distinct set of cell cycle regulators in TNF- $\alpha$  and SAHA-responsive single mCherry positive (M<sub>TS</sub>C+) cell subpopulations with spontaneously reactivated HIV-1 promoters.** Some M<sub>TS</sub>C+ cell clones when reverted to their initial M+ phenotype upon TNF- $\alpha$  and SAHA withdrawal, constantly exhibited a small subpopulation of cells with a spontaneously reactivated HIV-1 promoter, *i.e.*, cells that became double positive without further external stimulation. These spontaneous double positive cells are termed M<sub>sp</sub>C+ while the majority of cells that remained single mCherry positive are termed M<sub>r</sub>C- here (Fig. 6a). To examine whether M<sub>sp</sub>C+ and M<sub>r</sub>C- phenotypes were reversible, we sorted out the





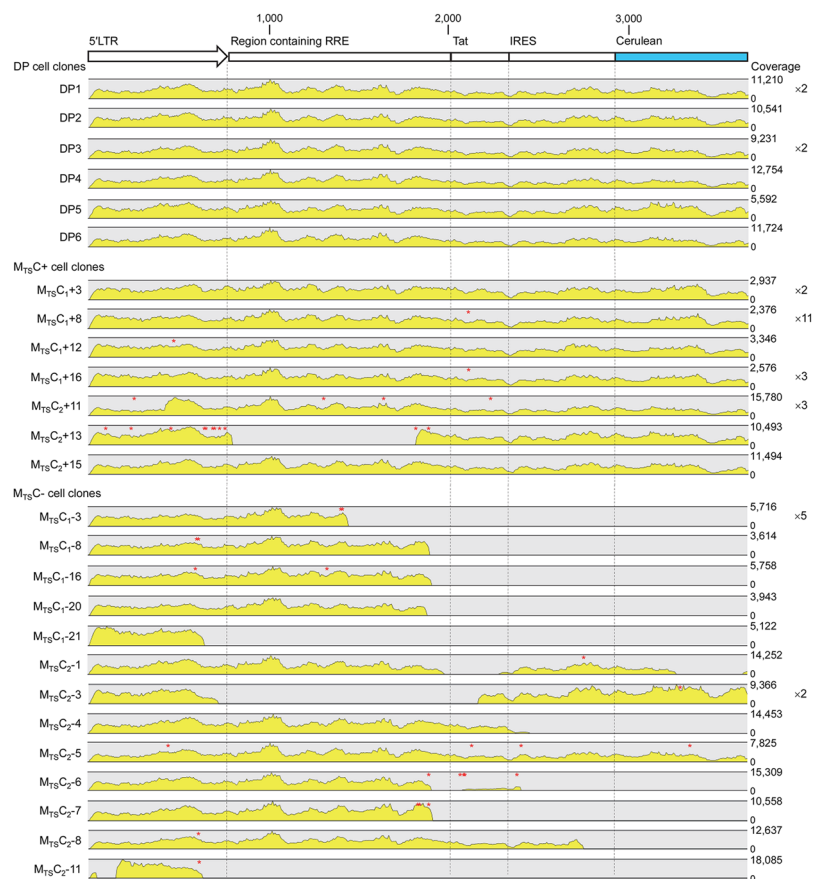
**Figure 4.** Factors influencing reactivation potentials of latent HIV-1 promoters in TNF- $\alpha$  and SAHA-responsive single mCherry positive ( $M_{TS}C+$ ) cell clones. **(a)** Reactivation potentials of latent HIV-1 promoters in TNF- $\alpha$  and SAHA-responsive single mCherry positive ( $M_{TS}C+$ ) cell clones with different LTatC[M] integration sites were reactivated with TNF- $\alpha$  and SAHA 54–106 days post sorting. The means of at least two independent measurements ( $n = 2–22$ ) are indicated by the heights of the bars and standard error means are indicated by the error bars. Two-tailed Mann-Whitney  $U$  test with 95% confidence level was used to test for statistical significance; \* $P < 0.05$ ; \*\* $P < 0.01$ . Subscripts 1 and 2 indicate two independent transduction and sorting experiments from which the cell clones were derived. DP: double positive. **(b)** Gene expression patterns in SUP-T1 cells during the first 24 hours upon infection with an HIV-1-based vector. Median log<sub>2</sub> fold changes in expression of genes into which LTatC[M] was found integrated in different  $M_{TS}C+$  cell clones are shown. The majority of the HIV-1 integration process are defined as complete at 18 hours post infection. + and – indicate significant ( $P < 0.01$ ) upregulation and downregulation in gene expression, respectively, whereas 0 indicates no significant changes in gene expression compared to data at time point 0 hour. Data were obtained from the Patterns of Expression and Analysis of Clusters of HIV/Host interactions (PEACHi) database<sup>27</sup>. Subscripts 1 and 2 indicate two independent transduction and sorting experiments from which the cell clones were derived. **(c)** Predicted RNA secondary structure of HIV-1 transactivation response element. The RNA secondary structures of HIV-1 transactivation response (TAR) element found in cell clone  $M_{TS}C_1+12$  (top) and in the original vector LTatC[M] (bottom) were predicted using CLC Main Workbench 7. Red circle indicates the mutation in the TAR element of  $M_{TS}C_1+12$ .

two cell subpopulations respectively from five  $M_{TS}C+$  cell clones representing two distinct vector integration sites:  $M_{TS}C_1+8$ ,  $M_{TS}C_1+28$ , and  $M_{TS}C_1+34$  which had LTatC[M] integrated into *EHBP1*, and  $M_{TS}C_2+11$  and  $M_{TS}C_2+20$  which had LTatC[M] integrated into *CTNND1* (Fig. 6a; Table 1). The percentages of spontaneous double positive cells that remained in the sorted  $M_{sp}C+$  cell subpopulations and those that emerged from the sorted  $M_rC-$  cell subpopulations over time were monitored with flow cytometry. For all  $M_{TS}C+$  cell clones, the majority of the sorted  $M_{sp}C+$  cells became single mCherry positive while a small fraction of  $M_rC-$  cells became double positive to the extents of their respective parental cell clones (Fig. 6c).

Given that the  $M_{sp}C+$  and  $M_rC-$  phenotypes were reversible, we hypothesized that the spontaneous reactivation of HIV-1 promoters was linked to the cell cycle. The transcriptomes of  $M_{sp}C+$  and  $M_rC-$  cells at 0 day post sorting were sequenced to identify differentially expressed genes between the two cell subpopulations (Supplementary Fig. S3). Cells of the same fluorescence phenotype and vector integration site were treated as biological replicates. Using the cut-offs of fold change  $> 1.5 \times (|\log_2 \text{ratio}| > 0.585)$  and  $P < 0.02$ , we identified 18 genes that had a higher expression level in  $M_{sp}C+$  cell subpopulations compared to  $M_rC-$  and were common between the two vector integration sites (Fig. 6b; Supplementary Table S2a). Of these 18 genes, three (16.7%) are

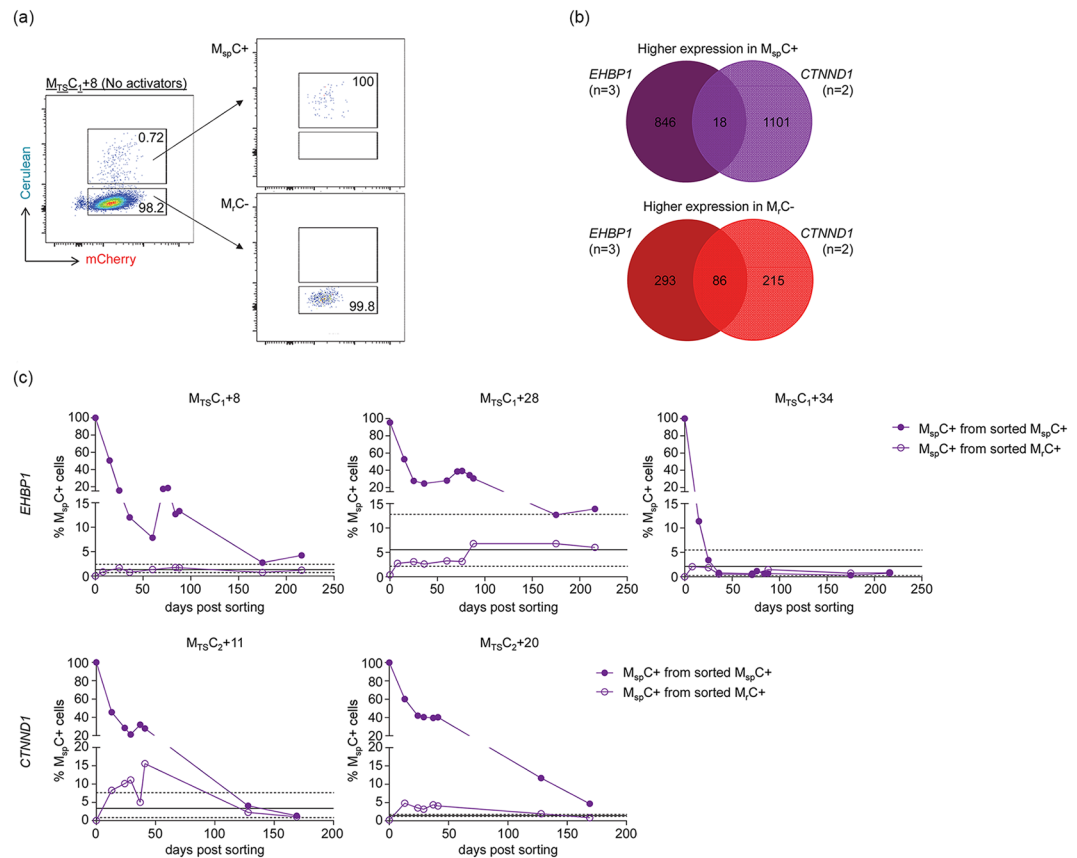
Cell clone	Integration locus	Gene	Intragenic region	Transcriptional orientation
M <sub>TSC</sub> C <sub>1</sub> +3	Chr3: 48710573	Intergenic	N/A	N/A
M <sub>TSC</sub> C <sub>1</sub> +8	Chr2: 63216606	<i>EHBP1</i>	Intron 15/17*	Convergent
M <sub>TSC</sub> C <sub>1</sub> +12	Chr15: 86127496	<i>AKAP13</i>	Intron 7	Same
M <sub>TSC</sub> C <sub>1</sub> +16	Chr17:79808805	<i>P4HB</i>	Intron 4	Convergent
M <sub>TSC</sub> C <sub>2</sub> +11	Chr11: 57536158	<i>CTNND1/TMX2-CTNND1**</i>	Intron 1/2**	Same
M <sub>TSC</sub> C <sub>2</sub> +13	Chr1: 227375035	<i>CDC42BPA</i>	Intron 5	Same
M <sub>TSC</sub> C <sub>2</sub> +15	Chr14: 103178189	<i>RCOR1</i>	Intron 7	Same

**Table 1.** Integration sites of LTatC[M] in TNF- $\alpha$  and SAHA-responsive single mCherry positive (M<sub>TSC</sub>C+) cell clones. Two other cell clones, M<sub>TSC</sub>C<sub>1</sub>+28 and M<sub>TSC</sub>C<sub>1</sub>+34, and another cell clone, M<sub>TSC</sub>C<sub>2</sub>+20, which were included in subsequent analyses, had the same LTatC[M] integration site as M<sub>TSC</sub>C<sub>1</sub>+8 and M<sub>TSC</sub>C<sub>2</sub>+11, respectively. \*Transcript variant; \*\**TMX2-CTNND1* is read-through long non-coding RNA. Subscripts 1 and 2 indicate two independent transduction and sorting experiments from which the clones were derived.



**Figure 5.** Mutational analysis of Cerulean cassettes of double positive (DP), TNF- $\alpha$  and SAHA-responsive single mCherry positive (M<sub>TSC</sub>C+), and TNF- $\alpha$  and SAHA-non-responsive single mCherry positive (M<sub>TSC</sub>C-) cell clones. Cerulean cassettes of double positive (DP), TNF- $\alpha$  and SAHA-responsive single mCherry positive (M<sub>TSC</sub>C+), and TNF- $\alpha$  and SAHA-non-responsive single mCherry positive (M<sub>TSC</sub>C-) cell clones were amplified and sequenced with the Illumina MiSeq next-generation sequencing technology. The schematic diagram of the LTatC[M] Cerulean cassette is shown on top and sequence coverages are depicted as yellow peaks with the range for each cell clone shown on the right. Point mutations are denoted by red asterisks. The numbers of cell clones with the same integration sites and mutation patterns analysed are shown next to the sequence coverage ranges. Subscripts 1 and 2 indicate two independent transduction and sorting experiments from which the cell clones were derived.

known to promote cell cycle progression: *FOSB*, *NEAT1*, and *EGR1* (Supplementary Table S2a). For instance, the two top hits, *FOSB* and *NEAT1*, have been reported to induce cell cycle entry by activating cyclin D1 in mouse fibroblasts<sup>32</sup> and promote cell proliferation in cancer cells<sup>33</sup>, respectively. Furthermore, *NEAT1* and *EGR1* have also been reported to be involved in the HIV-1 replication cycle<sup>34,35</sup>. On the other hand, 86 genes were found

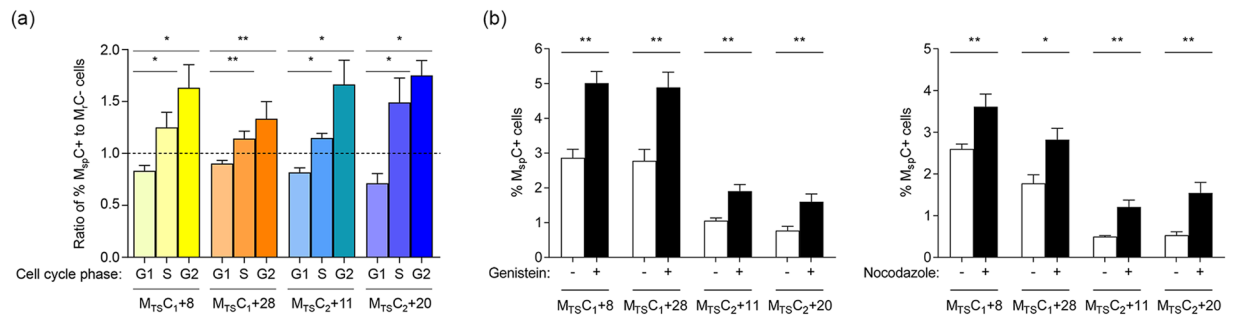


**Figure 6.** Longitudinal fluorescence profiles of spontaneous double positive ( $M_{sp}C+$ ) cell subpopulations and their counterparts that remained single mCherry positive ( $M_{rc}-$ ) sorted from TNF- $\alpha$  and SAHA-responsive single mCherry positive ( $M_{TS}C+$ ) cell clones. TNF- $\alpha$  and SAHA-responsive single mCherry positive ( $M_{TS}C+$ ) cell clones that exhibited a small subpopulation of spontaneous double positive ( $M_{sp}C+$ ) cells without treatment with activators were sorted by flow cytometry for  $M_{sp}C+$  cells and single mCherry cells that did not spontaneously become DP, *i.e.* non-spontaneous  $M+$  ( $M_{rc}-$ ). **(a)** Representative flow cytometric sorting strategy and purity of  $M_{sp}C+$  and  $M_{rc}-$  cell subpopulations from  $M_{TS}C_1+8$  cell clone. **(b)** Number of genes that had a higher expression level in  $M_{sp}C+$  (top) and  $M_{rc}-$  (bottom) cell subpopulations and were common between three  $M_{TS}C+$  cell clones with LTatC[M] integrated into *EHBP1* and two  $M_{TS}C+$  cell clones with LTatC[M] integrated into *CTNND1*. **(c)** Longitudinal percentages of  $M_{sp}C+$  cell subpopulations remaining and emerging from sorted  $M_{sp}C+$  (shaded circles) and  $M_{rc}-$  (open circles), respectively. Solid and dashed horizontal lines indicate the means and ranges of at least five independent measurements ( $n \geq 5$ ) of percentages of  $M_{sp}C+$  cells in the unsorted parental cell clones. The top three cell clones had the same LTatC[M] integration site in *EHBP1* whereas the bottom two cell clones had the same LTatC[M] integration site in *CTNND1*. Subscripts 1 and 2 indicate two independent transduction and sorting experiments from which the cell clones were derived.

to have a higher expression level in  $M_{rc}-$  cell subpopulations (Fig. 6b; Supplementary Table S2b). Of these 86 genes, 81 encode non-coding RNAs (Supplementary Table S2b), such as *RN7SK*, the RNA product of which sequesters CDK9 and CycT1 from acting as positive transcription elongation factors at the HIV-1 promoter<sup>36</sup>, and *RN7SL1*, *RN7SL2*, *RNY1*, and *RNY3*, which are cellular RNAs typically co-packaged in the HIV-1 viral particle<sup>37</sup>. Six (7.0%) and three (3.5%) of these 86 genes are known to promote and suppress cell cycle progression, respectively (Supplementary Table S2b).

**Larger fractions of spontaneous double positive ( $M_{sp}C+$ ) cells are in the S and G2 phases of the cell cycle compared to cells remaining as single mCherry positive ( $M_{rc}-$ ) and treatment with cell cycle regulators increased spontaneous reactivation.** To further explore the linkage between spontaneous activation of latent HIV-1 promoters and the cell cycle, four  $M_{TS}C+$  cell clones representing two vector integration sites were resolved into G1, S, and G2 phases of the cell cycle by flow cytometric analysis of cells with their genomic DNA content fluorescently labelled. A significant increase in percentages of cells in the S ( $P < 0.05$ ) and G2 ( $P < 0.001$ ) phases of the cell cycle was observed in  $M_{sp}C+$  cells in comparison to  $M_{rc}-$  cells (Supplementary Fig. S4). The ratio of percentages of  $M_{sp}C+$  cells to  $M_{rc}-$  cells in each of the three cell cycle phases are shown in Fig. 7a. Consistently, treatment of these cell clones with genistein and nocodazole, which halted SUP-T1 cells in the G2 phase of the cell cycle (Supplementary Fig. S5), increased the percentages of





**Figure 7.** Spontaneous reactivation of latent HIV-1 promoters is linked to the cell cycle. **(a)** Ratio of percentages of cells in the various phases of the cell cycle in spontaneous double positive ( $M_{sp}C+$ ) to non-spontaneous  $M+$  ( $M_{sp}C-$ ) cell subpopulations. The means of at least four independent measurements ( $n \geq 4$ ) are indicated by the heights of the bars and standard error means are indicated by the error bars. **(b)** Percentages of  $M_{sp}C+$  cells upon treatment with genistein (left) and nocodazole (right) in comparison to equally diluted DMSO control (-). The means of five independent measurements ( $n = 5$ ) are indicated by the heights of the bars and standard error means are indicated by the error bars. Two-tailed Mann-Whitney  $U$  test with 95% confidence level was used to test for statistical significance; \* $P < 0.05$ ; \*\* $P < 0.01$ . Subscripts 1 and 2 indicate two independent transduction and sorting experiments from which the cell clones were derived. M<sub>TSC</sub>C+: TNF- $\alpha$  and SAHA-responsive single mCherry positive.

spontaneous reactivation by 1.8–2.1- and 1.4–2.9-fold, respectively (Fig. 7b). Taken together with the transcriptomic data above, our findings suggest that spontaneous activation of latent HIV-1 promoters is linked to the cell cycle.

## Discussion

HIV-1-based vectors have previously been utilized to dissect various aspects of HIV-1 infection<sup>7,10,11</sup>. In the present study, we have developed a novel vector, LTatC[M], and characterized it in great detail with regard to its representation of HIV-1 infection and latency in SUP-T1 cells. In contrast to previous vectors, LTatC[M] contains a pair of genetic insulators, cHS4 and sMAR, that flank the constitutive mCherry cassette to protect it from position-effect variegation that might be exerted at some integration sites<sup>7,38</sup>. Additionally, mCherry expression is driven by the human *eIF4A1* promoter, which has been shown to yield high levels of gene expression in macrophage cell lines<sup>39</sup>, potentially enabling the use of LTatC[M] to study HIV-1 infection in this other target cell type of HIV-1. Furthermore, cHS4 is capable of alleviating transcriptional interference between tandem gene cassettes<sup>17–19</sup>. The functionality of the genetic insulators was evident in two ways: (i) stable expression of Cerulean and/or mCherry in long-term bulk and monoclonal cell cultures for up to 394 days and (ii) absence of reactivatable HIV-1 and *heIF4A1* promoters in the double negative cell population. In contrast, in a previous study using a dual-fluorescence vector without genetic insulators, a substantial fraction of double negative cells was found to harbour a reactivatable vector<sup>40</sup>. More importantly, these genetic insulators did not alter the kinetics and ratio of emergence of active and latent cells and vector integration site patterns compared to the control vector lacking the genetic insulators. With our model, we were able to examine HIV-1 latency in monoclonal cells with distinct vector integration sites. We focused on factors that contribute towards two aspects of HIV-1 latency reversal: (i) spontaneous reactivation and (ii) reactivation potentials upon external induction.

Spontaneous reactivation of latent HIV-1 promoters *in vitro* has also been observed by others<sup>7</sup>, although a mechanistic explanation is lacking. We found that spontaneous reactivation of latent HIV-1 promoters is linked to the cell cycle, as evidenced by transcriptomic profiling to identify differentially expressed genes between spontaneously reactivated and the remaining non-activated cell subpopulations within cell clones. This linkage was further corroborated by two observations: (i) larger fractions of spontaneously reactivated cells were in the S and G2 phases of the cell cycle compared to the non-activated cells and (ii) halting the cell cycle at the G2 phase with genistein and nocodazole increased the fraction of spontaneous reactivation. This is analogous to the partial reactivation of latent HIV-1 observed in central memory CD4<sup>+</sup> T cells undergoing homeostatic proliferation<sup>41</sup>. Furthermore, viral production is known to increase in the G2 phase<sup>42,43</sup>. Interestingly, two cell cycle regulators having a higher expression level in spontaneously reactivated cells, *NEAT1* and *EGR1*, have also been reported to be involved in the HIV-1 replication cycle. *NEAT1* is essential for the formation of nuclear paraspeckles and has been found to be upregulated in peripheral blood mononuclear cells of untreated HIV-1-infected individuals (productive infection) compared to those of virally suppressed (latent infection)<sup>44</sup>. Additionally, knockdown of *NEAT1* enhances HIV-1 production<sup>35</sup>. These studies suggest that *NEAT1* might be upregulated as an antiviral response to HIV-1 expression, *e.g.* HIV-1 Tat in spontaneously reactivated cells. *EGR1* encodes a transcriptional factor that regulates genes related to cell growth and differentiation. The expression of *EGR1* is upregulated by HIV-1 Tat<sup>45,46</sup> and its activation leads to the reversal of HIV-1 latency<sup>34</sup>, consistent with our observation that *EGR1* was expressed more in spontaneously reactivated cells. The vast majority of genes identified to have a higher expression level in non-activated cells were non-coding small nuclear RNAs (snRNAs) with unknown relevance to HIV-1 infection, except for *RN7SK*, which forms an RNA scaffold that binds CDK9 and CycT1, thus sequestering these molecules from acting as a positive transcription elongation factor at the HIV-1 promoter<sup>36</sup>. The role of non-coding RNAs, especially long non-coding RNAs and microRNAs, in the modulation of HIV-1

infection is of great interest<sup>47,48</sup> and snRNAs identified in our screen might represent an additional layer of potential candidates for further investigation in the context of HIV-1 latency establishment and maintenance.

The observations that larger fractions of spontaneously reactivated cells are in the later phases of the cell cycle and that spontaneous reactivation increased upon treatment with cell cycle regulators suggest that cell cycling might lead to spontaneous reactivation. This would provide a mechanistic explanation to the spontaneous reactivation of latent HIV-1 proviruses observed in HIV-1-infected individuals during treatment interruption as a result of homeostatic proliferation of latently infected cells. Our model, however, does not exclude the alternative; cell cycling is a result of spontaneous reactivation. Spontaneous reactivation of latent HIV-1 proviruses has been described as an intrinsic characteristic of the virus<sup>49,50</sup>. We detected a higher expression of *EGR1*, which is a positive cell cycle regulator, in spontaneously reactivated cells and upregulation of *EGR1* by HIV-1 Tat has been shown<sup>45,46</sup>. *Tat* is contained in our vector and it is an early-expressed gene in the HIV-1 replication cycle<sup>36</sup>. Thus, the intrinsic spontaneous reactivation of latent HIV-1 proviruses may drive cell cycling, potentially contributing towards clonal expansion of latently infected cells<sup>51–53</sup>. Of note, our minimalistic HIV-1-based vector does not encode, for example, HIV-1 *Vpr*, which is also known to modulate the host cell's cell cycling programme<sup>54</sup>. Nonetheless, individual components of the viral genome could be cloned into our vector, thus enabling the possibility to systematically examine individual or combined effects of these components on cell cycling as well as other aspects of the host cell.

HIV-1 integration site selection has been implicated to have an influence on whether the integrated provirus would be active or latent, as well as the reactivation potentials of latent proviruses<sup>13,14</sup>. Vector integration sites identified in our model conformed to HIV-1 integration site patterns<sup>8,26,55–57</sup>. Furthermore, cells representing latent HIV-1 infections consistently exhibited different vector integration site patterns, albeit modest, than their active counterparts in two independent transductions; higher frequencies in transcription units and in the same transcriptional orientation relative to vector-hosting genes. These differences between active and latent cells were also reported in a primary-cell model transduced with HIV-1<sub>NL4-3</sub> whose *env* was partially replaced with a fluorescent reporter gene<sup>58</sup>. More HIV-1 integrations in a convergent transcriptional orientation have also been observed in viraemic (active infections) than in treated (latent infections) individuals<sup>55</sup>. We further dissected cells with a latent HIV-1 phenotype into those that were inducible with TNF- $\alpha$  and SAHA to become active and those that were non-inducible. Inducible cell clones with distinct vector integration sites and expression profiles of vector-hosting genes displayed variable reactivation potentials. The reactivation potentials were dramatically attenuated in inducible cell clones with mutations in elements important for the expression of Cerulean, demonstrating that the integrity of the viral genome, in addition to features of the viral integration site, has an impact on HIV-1 latency reversal. On the other hand, we did not find any non-inducible cell clone with an intact Cerulean cassette, although the sample size of cells analysed was relatively small. Major internal deletions and mutations, possibly a result of the error-prone HIV-1 reverse transcriptase<sup>59,60</sup> and copy-choice recombination<sup>61,62</sup>, were the cause of these clones not being reactivatable. Such a phenomenon is also observed in cells derived from HIV-1-infected individuals<sup>4</sup>, although the rate and site of mutation and recombination are likely to differ in our system utilizing a minimalistic HIV-1-based vector. Nonetheless, our observation accentuates the need to carefully examine singly fluorescent cell populations, e.g. latency-representing cells, in HIV-1 latency models using HIV-1-based vector systems.

Besnard *et al.* recently, in a genome-wide shRNA screen, identified candidate genes that either promote or inhibit HIV-1 latency, and subsequently showed that the mechanistic target of rapamycin (mTOR) complex is a positive regulator of HIV-1 latency reversal<sup>63</sup>. In their screen, actin cytoskeleton reorganization pathways were also found to be potentially involved in the regulation of HIV-1 latency<sup>63</sup>. Interestingly, 4/7 inducible cell clones analysed had LTatC[M] integrated into genes related to actin cytoskeleton reorganization: *EHBPI*<sup>64</sup>, *AKAP13*<sup>65</sup>, *CTNND1*<sup>66</sup>, and *CDC42BPA*<sup>67</sup>. For example, EHBPI links actin to the cell membrane for endosomal tubulation<sup>64</sup> and CDC42BPA (Cell division control 42 binding protein kinase  $\alpha$ ) (CDC42 effector protein 5 is the second highest latency-promoting gene in the study of Besnard *et al.*)<sup>63</sup> promotes actin and myosin reorganization<sup>67</sup>. These four genes are also listed as latency promoting in the study of Besnard *et al.*<sup>63</sup>. The linkage between integration into these genes and HIV-1 latency establishment is currently unclear. Nevertheless, our observation complements current knowledge on HIV-1 latency and presents an interesting finding for further investigation.

## Conclusions

In summary, our well-characterized dual-fluorescence, genetic-insulators-containing HIV-1-based vector is capable of recapitulating various facets of HIV-1 infection and latency and invariably maintains active and latent HIV-1 phenotypes over a long observation period spanning at least one year. Application of this vector might provide insights into the pathomechanisms of HIV-1 latency. We applied this vector to study spontaneous reactivation of latent HIV-1 promoters in SUP-T1 cells and found a linkage between this phenomenon and the cell cycle. This is an important basis for further investigations on the mechanisms of maintenance of the HIV-1 reservoir and by which spontaneous viral rebound occurs in HIV-1-infected individuals off ART.

## Methods

**Plasmid vector construction.** LTatC[M] was synthesized by GeneArt (Thermo Fisher Scientific), except that the HIV-1 Rev response element (RRE) was amplified and cloned from pEV731<sup>7</sup> (kindly provided by Eric Verdin) with primer pair 5'NNNNNGTCGACCTCGAGATGGGTGCGAGAGCGTCAG3' and 5'NNNNNGTCGACGGTGGCATCGATAACCGTCGAG3', and the Woodchuck hepatitis virus post-transcriptional regulatory element (WPRES) was amplified from pCLX-UBI-GFP (kindly provided by Patrick Salmon; Addgene plasmid # 27245) with primer pair 5'NNNNNNGGATCCCGATAATCAACCTCTGGATTAC3' and 5'NNNNNACCGGTAATCCCGAGCGGGGAGG3'. Genetic elements in LTatC[M] are shown in Fig. 1a and tabulated in Supplementary Table S3.

**Virus pseudotyping and titration.** VSV-G-pseudotyped LTatC[M] were generated by transfecting 293 T cells (American Type Culture Collection) of at least 90% confluency cultured in DMEM (Sigma-Aldrich) supplemented with 10% (v/v) fetal bovine serum (FBS) (Thermo Fisher Scientific), 100 U/mL Penicillin, and 100 µg/mL Streptomycin (Thermo Fisher Scientific) in a T-150 flask with the following components: pLTatC[M] (20.8 µg), psPAX2 (20.8 µg) (kindly provided by Didier Trono; Addgene plasmid # 12260), pCAG-VSVG (10.4 µg) (kindly provided by Arthur Nienhuis and Patrick Salmon; Addgene plasmid # 35616), polyethylenimine (208 µg), and serum-free DMEM in a total volume of 3.4 mL. Medium was replaced 18 hours post transfection. Supernatant containing viral particles was harvested twice at 24 and 48 hours post medium change, and concentrated with PEG-it (System Biosciences) to a final volume of 450 µL.

VSV-G-pseudotyped virus stocks were titrated by transducing 0.2 million SUP-T1 cells<sup>68</sup> (kindly provided by Dharam Ablashi through the NIH AIDS Reagent Programme, Division of AIDS, NIAID, NIH) cultured in 200 µL of RPMI-1640 (Sigma-Aldrich) supplemented with 10% (v/v) FBS, 100 U/mL Penicillin, and 100 µg/mL Streptomycin in 48-well plates with virus stocks. Medium was replaced daily and cells were analysed with flow cytometry (BD Biosciences LSR Fortessa) for total positive cells, *i.e.*, cells expressing mCherry and/or Cerulean, at 5 days post transduction. The mean viral titre was calculated from transductions that yielded a total positive cells of <20%, thus ensuring a high probability of obtaining single integrants per cell<sup>69</sup>, and expressed as SUP-T1 transduction units/mL.

**SUP-T1 transduction and flow cytometric sorting.** SUP-T1 cells were transduced with VSV-pseudotyped LTatC[M] to yield 5–56% of positive cells. All four cell populations were sorted with flow cytometry (BD Biosciences FACS ARIA III) 4–13 days post transduction. For sorting of tumour necrosis factor alpha (TNF-α) (Bio-Techne) and suberanilohydroxamic acid (SAHA) (Bio-Techne)-inducible (M<sub>TSC</sub>+ ) and -non-inducible (M<sub>TSC</sub>- ) cell populations, M+ cells were treated with 0.01 µg/mL TNFα and 1 µM SAHA for 24 hours prior to sorting.

**Amplification, sequencing, and mapping of 5' vector integration junctions.** The 5' vector integration junctions were amplified, sequenced, and mapped as described previously<sup>26</sup>. Briefly, 5' vector integration junctions were amplified with the non-restrictive linear amplification-mediated PCR<sup>70</sup> and 12–14 pM purified PCR products were sequenced with the Illumina MiSeq platform using the MiSeq Reagent Kit v2 (300 cycles) (Illumina) with 8% PhiX. Mapping of sequencing reads to the human genome assembly GRCh37.p13 was performed using an in-house bioinformatic pipeline, Integration Site Analysis Pipeline (InStAP). All vector integration junctions are listed in Supplementary Table S4.

**Amplification and sequencing of Cerulean cassettes.** The Cerulean cassettes, spanning from the HIV-1 5' LTR to cHS4, in transduced cells were amplified using Platinum Taq DNA polymerase high fidelity (Thermo Fisher Scientific) with the forward primer 5'GACAAGAGATCCTTGATCTGTGGATC3' and reverse primer 5'CACTGATAGGGAGTAAACATATGC3' or 5'GAAGGACAGCTTCAAGTAGTCG3'. The cycling conditions were 94 °C for 2 min, 40 cycles of 94 °C for 30 s, 55 °C for 30 s and 68 °C for 5 min 30 s, and a final extension of 68 °C for 10 min.

The amplicons were processed with the Nextera XT DNA Library Preparation Kit (Illumina) and subsequently sequenced with the MiSeq reagent Kit v2 (50 cycles) with 1% PhiX.

**Transcriptomic analysis.** Total RNA was extracted using AllPrep DNA/RNA Kit (Qiagen) and reverse-transcribed library was prepared and normalized using SMARTer Stranded Total RNA-Seq Kit – Pico Input Mammalian (Takara). The TruSeq SR Cluster Kit v4-cBot-HS or TruSeq PE Cluster Kit v4-cBot-HS (Illumina) was used for cluster generation with 8 pM pooled normalized libraries on the Illumina cBOT system. Sequencing was performed with the Illumina HiSeq. 2500; paired end at 2 × 126 bp or single end 126 bp using the TruSeq SBS Kit v4-HS (Illumina).

Bioinformatic analysis was performed using the R package ezRun (<https://github.com/uzh/ezRun>) within the data analysis framework SUSHI<sup>71</sup>. Raw reads were quality checked using Fastqc (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and FastQ Screen ([http://www.bioinformatics.babraham.ac.uk/projects/fastq\\_screen/](http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/)). Quality-controlled reads were aligned to the reference genome GRCh38 using the STAR aligner<sup>72</sup>. Expression counts were computed using featureCounts in the Bioconductor package Subread<sup>73</sup>. Differential expression analysis was performed using the edgeR package<sup>74</sup>, where raw read counts were normalized using the Trimmed Mean of M values (TMM) method<sup>75</sup> and differential expression was computed using the Generalized Linear Mode (GLM) likelihood ratio test. Quality checkpoints<sup>76</sup>, such as quality control of the alignment and count results were implemented in ezRun and applied throughout the analysis workflow to ensure correct data interpretation.

**Cell cycle analysis.** A total of 1 million cells were fixed in 1 mL of 2% (w/v) paraformaldehyde (Fluka) in Dulbecco's phosphate-buffered saline (DPBS) on ice for 15 min, washed with DPBS twice by centrifugation at 800 g for 5 min, resuspended in 1 mL of 0.1% (v/v) Triton X-100 (Sigma-Aldrich) in DPBS supplemented with 1% (v/v) FBS, treated with 0.1 mg/mL RNase A (Thermo Fisher Scientific), and stained with 1 µL of FxCycle Far Red Stain (Thermo Fisher Scientific) at room temperature for 30 min. Cell cycle profiles were modeled with the Cell Cycle function in FlowJo v10.0.8 [Watson (Pragmatic); CV of G2 peak = CV of G1 peak].

**Cell cycle regulators treatment.** A total of 0.1 million M<sub>TSC</sub>+ cells cultured in 200 µL of RPMI-1640 (10% (v/v) FBS, 100 U/mL Penicillin, and 100 µg/mL Streptomycin) or 1 million SUP-T1 cells cultured in 2 mL of RPMI-1640 (10% (v/v) FBS, 100 U/mL Penicillin, and 100 µg/mL Streptomycin) were treated with 30 µM

genistein (Selleckchem), 40 ng/mL nocodazole (Sigma-Aldrich), or equally diluted dimethyl sulfoxide (DMSO) (Sigma-Aldrich) for 24 hours prior to flow cytometric analysis.

**Data availability.** Data generated and analysed during this study are included in this published article and its supplementary information files unless stated otherwise, in which case are available from the corresponding author on reasonable request.

## References

- Greene, W. C. & Peterlin, B. M. Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nat. Med.* **8**, 673–680, <https://doi.org/10.1038/nm0702-673> (2002).
- Archin, N. M., Sung, J. M., Garrido, C., Soriano-Sarabia, N. & Margolis, D. M. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. *Nat. Rev. Microbiol.* **12**, 750–764, <https://doi.org/10.1038/nrmicro3352> (2014).
- Laskey, S. B. & Siliciano, R. F. A mechanistic theory to explain the efficacy of antiretroviral therapy. *Nat. Rev. Microbiol.* **12**, 772–780, <https://doi.org/10.1038/nrmicro3351> (2014).
- Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551, <https://doi.org/10.1016/j.cell.2013.09.020> (2013).
- Eriksson, S. *et al.* Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS Pathog.* **9**, e1003174, <https://doi.org/10.1371/journal.ppat.1003174> (2013).
- Brooks, D. G. & Zack, J. A. Effect of latent human immunodeficiency virus infection on cell surface phenotype. *J. Virol.* **76**, 1673–1681 (2002).
- Jordan, A., Defechereux, P. & Verdin, E. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J.* **20**, 1726–1738, <https://doi.org/10.1093/emboj/20.7.1726> (2001).
- Lewinski, M. K. *et al.* Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* **79**, 6610–6619, <https://doi.org/10.1128/jvi.79.11.6610-6619.2005> (2005).
- Tyagi, M., Pearson, R. J. & Karn, J. Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. *J. Virol.* **84**, 6425–6437, <https://doi.org/10.1128/jvi.01519-09> (2010).
- Calvanese, V., Chavez, L., Laurent, T., Ding, S. & Verdin, E. Dual-color HIV reporters trace a population of latently infected cells and enable their purification. *Virology* **446**, 283–292, <https://doi.org/10.1016/j.virol.2013.07.037> (2013).
- Dahabieh, M. S. *et al.* Direct non-productive HIV-1 infection in a T-cell line is driven by cellular activation state and NFkappaB. *Retrovirology* **11**, 17, <https://doi.org/10.1186/1742-4690-11-17> (2014).
- Matsuda, Y. *et al.* Epigenetic heterogeneity in HIV-1 latency establishment. *Sci. Rep.* **5**, 7701, <https://doi.org/10.1038/srep07701> (2015).
- Vranckx, L. S. *et al.* LEDGIN-mediated Inhibition of Integrase-LEDGF/p75 Interaction Reduces Reactivation of Residual Latent HIV. *EBioMedicine* **8**, 248–264, <https://doi.org/10.1016/j.ebiom.2016.04.039> (2016).
- Chen, H. C., Martinez, J. P., Zorita, E., Meyerhans, A. & Filion, G. J. Position effects influence HIV latency reversal. *Nat Struct Mol Biol* **24**, 47–54, <https://doi.org/10.1038/nsmb.3328> (2017).
- Sherrill-Mix, S. *et al.* HIV latency and integration site placement in five cell-based models. *Retrovirology* **10**, 90, <https://doi.org/10.1186/1742-4690-10-90> (2013).
- Katlama, C. *et al.* Barriers to a cure for HIV: new ways to target and eradicate HIV-1 reservoirs. *Lancet* **381**, 2109–2117, [https://doi.org/10.1016/s0140-6736\(13\)60104-x](https://doi.org/10.1016/s0140-6736(13)60104-x) (2013).
- Villemure, J. F., Savard, N. & Belmaaza, A. Promoter suppression in cultured mammalian cells can be blocked by the chicken beta-globin chromatin insulator 5'HS4 and matrix/scaffold attachment regions. *J. Mol. Biol.* **312**, 963–974, <https://doi.org/10.1006/jmbi.2001.5015> (2001).
- Yahata, K. *et al.* cHS4 insulator-mediated alleviation of promoter interference during cell-based expression of tandemly associated transgenes. *J. Mol. Biol.* **374**, 580–590, <https://doi.org/10.1016/j.jmb.2007.09.054> (2007).
- Uchida, N., Hanawa, H., Yamamoto, M. & Shimada, T. The chicken hypersensitivity site 4 core insulator blocks promoter interference in lentiviral vectors. *Hum Gene Ther Methods* **24**, 117–124, <https://doi.org/10.1089/hgtb.2012.152> (2013).
- Aker, M. *et al.* Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Hum. Gene Ther.* **18**, 333–343, <https://doi.org/10.1089/hum.2007.021> (2007).
- Li, C. L. & Emery, D. W. The cHS4 chromatin insulator reduces gammaretroviral vector silencing by epigenetic modifications of integrated provirus. *Gene Ther.* **15**, 49–53, <https://doi.org/10.1038/sj.gt.3303009> (2008).
- Tian, J. & Andreadis, S. T. Independent and high-level dual-gene expression in adult stem-progenitor cells from a single lentiviral vector. *Gene Ther.* **16**, 874–884, <https://doi.org/10.1038/gt.2009.46> (2009).
- Suttiprapa, S., Rinaldi, G. & Brindley, P. J. Prototypic chromatin insulator cHS4 protects retroviral transgene from silencing in *Schistosoma mansoni*. *Transgenic Res.* **21**, 555–566, <https://doi.org/10.1007/s11248-011-9556-0> (2012).
- Chung, J. H., Bell, A. C. & Felsenfeld, G. Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci USA* **94**, 575–580 (1997).
- Recillas-Targa, F. *et al.* Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci USA* **99**, 6883–6888, <https://doi.org/10.1073/pnas.102179399> (2002).
- Kok, Y. L. *et al.* Monocyte-derived macrophages exhibit distinct and more restricted HIV-1 integration site repertoire than CD4(+) T cells. *Sci. Rep.* **6**, 24157, <https://doi.org/10.1038/srep24157> (2016).
- Mohammadi, P. *et al.* 24 hours in the life of HIV-1 in a T cell line. *PLoS Pathog.* **9**, e1003161, <https://doi.org/10.1371/journal.ppat.1003161> (2013).
- Roy, S., Dellling, U., Chen, C. H., Rosen, C. A. & Sonenberg, N. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* **4**, 1365–1373 (1990).
- Berkhout, B. & Jeang, K. T. Detailed mutational analysis of TAR RNA: critical spacing between the bulge and loop recognition domains. *Nucleic Acids Res.* **19**, 6169–6176 (1991).
- Neuveut, C. & Jeang, K. T. Recombinant human immunodeficiency virus type 1 genomes with tat unconstrained by overlapping reading frames reveal residues in Tat important for replication in tissue culture. *J. Virol.* **70**, 5572–5581 (1996).
- Jeang, K. T., Xiao, H. & Rich, E. A. Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J. Biol. Chem.* **274**, 28837–28840 (1999).
- Brown, J. R. *et al.* Fos family members induce cell cycle entry by activating cyclin D1. *Mol. Cell. Biol.* **18**, 5609–5619 (1998).
- Li, Z. *et al.* Overexpression of long noncoding RNA, NEAT1 promotes cell proliferation, invasion and migration in endometrial endometrioid adenocarcinoma. *Biomed. Pharmacother.* **84**, 244–251, <https://doi.org/10.1016/j.biopha.2016.09.008> (2016).
- Krishnan, V. & Zeichner, S. L. Host cell gene expression during human immunodeficiency virus type 1 latency and reactivation and effects of targeting genes that are differentially expressed in viral latency. *J. Virol.* **78**, 9458–9473, <https://doi.org/10.1128/jvi.78.17.9458-9473.2004> (2004).
- Zhang, Q., Chen, C. Y., Yedavalli, V. S. & Jeang, K. T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio* **4**, e00596–00512, <https://doi.org/10.1128/mBio.00596-12> (2013).



36. Karn, J. & Stoltzfus, C. M. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb. Perspect. Med.* **2**, a006916, <https://doi.org/10.1101/cshperspect.a006916> (2012).
37. Tian, C., Wang, T., Zhang, W. & Yu, X. F. Virion packaging determinants and reverse transcription of SRP RNA in HIV-1 particles. *Nucleic Acids Res.* **35**, 7288–7302, <https://doi.org/10.1093/nar/gkm816> (2007).
38. Akhtar, W. *et al.* Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154**, 914–927, <https://doi.org/10.1016/j.cell.2013.07.018> (2013).
39. Quinn, C. M. *et al.* The human eukaryotic initiation factor 4A1 gene (EIF4A1) contains multiple regulatory elements that direct high-level reporter gene expression in mammalian cell lines. *Genomics* **62**, 468–476, <https://doi.org/10.1006/geno.1999.6031> (1999).
40. Dahabieh, M. S., Ooms, M., Simon, V. & Sadowski, I. A doubly fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection. *J. Virol.* **87**, 4716–4727, <https://doi.org/10.1128/jvi.03478-12> (2013).
41. Bosque, A., Famiglietti, M., Weyrich, A. S., Goulston, C. & Pannelles, V. Homeostatic proliferation fails to efficiently reactivate HIV-1 latently infected central memory CD4+ T cells. *PLoS Pathog.* **7**, e1002288, <https://doi.org/10.1371/journal.ppat.1002288> (2011).
42. Goh, W. C. *et al.* HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr *in vivo*. *Nat. Med.* **4**, 65–71 (1998).
43. Thierry, S. *et al.* Cell cycle arrest in G2 induces human immunodeficiency virus type 1 transcriptional activation through histone acetylation and recruitment of CBP, NF- $\kappa$ B, and c-Jun to the long terminal repeat promoter. *J. Virol.* **78**, 12198–12206, <https://doi.org/10.1128/jvi.78.22.12198-12206.2004> (2004).
44. Jin, C. *et al.* Detection of the long noncoding RNAs nuclear-enriched autosomal transcript 1 (NEAT1) and metastasis associated lung adenocarcinoma transcript 1 in the peripheral blood of HIV-1-infected patients. *HIV Med.* **17**, 68–72, <https://doi.org/10.1111/hiv.12276> (2016).
45. Kim, N., Kukkonen, S., Gupta, S. & Aldovini, A. Association of Tat with promoters of PTEN and PP2A subunits is key to transcriptional activation of apoptotic pathways in HIV-infected CD4+ T cells. *PLoS Pathog.* **6**, e1001103, <https://doi.org/10.1371/journal.ppat.1001103> (2010).
46. Fan, Y., Zou, W., Green, L. A., Kim, B. O. & He, J. J. Activation of Egr-1 expression in astrocytes by HIV-1 Tat: new insights into astrocyte-mediated Tat neurotoxicity. *J. Neuroimmune Pharmacol.* **6**, 121–129, <https://doi.org/10.1007/s11481-010-9217-8> (2011).
47. Trypsteen, W. *et al.* Differential expression of lncRNAs during the HIV replication cycle: an underestimated layer in the HIV-host interplay. *Sci. Rep.* **6**, 36111, <https://doi.org/10.1038/srep36111> (2016).
48. Barichiev, S., Naidoo, J. & Mhlanga, M. M. Non-coding RNAs and HIV: viral manipulation of host dark matter to shape the cellular environment. *Front Genet* **6**, 108, <https://doi.org/10.3389/fgene.2015.00108> (2015).
49. Razooky, B. S., Pai, A., Aull, K., Rouzine, I. M. & Weinberger, L. S. A hardwired HIV latency program. *Cell* **160**, 990–1001, <https://doi.org/10.1016/j.cell.2015.02.009> (2015).
50. Singh, A., Razooky, B., Cox, C. D., Simpson, M. L. & Weinberger, L. S. Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys. J.* **98**, L32–34, <https://doi.org/10.1016/j.bpj.2010.03.001> (2010).
51. Ikeda, T., Shibata, J., Yoshimura, K., Koito, A. & Matsushita, S. Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. *J. Infect. Dis.* **195**, 716–725, <https://doi.org/10.1086/510915> (2007).
52. Wagner, T. A. *et al.* HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573, <https://doi.org/10.1126/science.1256304> (2014).
53. Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183, <https://doi.org/10.1126/science.1254194> (2014).
54. Faust, T. B., Binning, J. M., Gross, J. D. & Frankel, A. D. Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription. *Annu Rev Virol* **4**, 241–260, <https://doi.org/10.1146/annurev-virology-101416-041654> (2017).
55. Cohn, L. B. *et al.* HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432, <https://doi.org/10.1016/j.cell.2015.01.020> (2015).
56. Schroder, A. R. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529 (2002).
57. Brady, T. *et al.* HIV integration site distributions in resting and activated CD4+ T cells infected in culture. *AIDS* **23**, 1461–1471, <https://doi.org/10.1097/QAD.0b013e32832caf28> (2009).
58. Shan, L. *et al.* Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. *J. Virol.* **85**, 5384–5393, <https://doi.org/10.1128/jvi.02536-10> (2011).
59. Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H. & Kunkel, T. A. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J. Biol. Chem.* **264**, 16948–16956 (1989).
60. Patel, P. H. & Preston, B. D. Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. *Proc. Natl. Acad. Sci. USA* **91**, 549–553 (1994).
61. Sanchez, G., Xu, X., Chermann, J. C. & Hirsch, I. Accumulation of defective viral genomes in peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected individuals. *J. Virol.* **71**, 2233–2240 (1997).
62. Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9**, 617–626, <https://doi.org/10.1038/nrmicro2614> (2011).
63. Besnard, E. *et al.* The mTOR Complex Controls HIV Latency. *Cell Host Microbe* **20**, 785–797, <https://doi.org/10.1016/j.chom.2016.11.001> (2016).
64. Wang, P. *et al.* RAB-10 Promotes EHBP-1 Bridging of Filamentous Actin and Tubular Recycling Endosomes. *PLoS Genet* **12**, e1006093, <https://doi.org/10.1371/journal.pgen.1006093> (2016).
65. Mayers, C. M. *et al.* The Rho guanine nucleotide exchange factor AKAP13 (BRX) is essential for cardiac development in mice. *J. Biol. Chem.* **285**, 12344–12354, <https://doi.org/10.1074/jbc.M110.106856> (2010).
66. Boguslavsky, S. *et al.* p120 catenin regulates lamellipodial dynamics and cell adhesion in cooperation with cortactin. *Proc. Natl. Acad. Sci. USA* **104**, 10882–10887, <https://doi.org/10.1073/pnas.0702731104> (2007).
67. Leung, T., Chen, X. Q., Tan, I., Manser, E. & Lim, L. Myotonic dystrophy kinase-related Cdc42-binding kinase acts as a Cdc42 effector in promoting cytoskeletal reorganization. *Mol. Cell. Biol.* **18**, 130–140 (1998).
68. Ablashi, D. V. *et al.* Human herpesvirus-7 (HHV-7): current status. *Clin. Diagn. Virol.* **4**, 1–13 (1995).
69. Fehse, B., Kustikova, O. S., Bubenheim, M. & Baum, C. Poisson—it's a question of dose. *Gene Ther.* **11**, 879–881, <https://doi.org/10.1038/sj.gt.3302270> (2004).
70. Paruzynski, A. *et al.* Genome-wide high-throughput integrate analyses by nrLAM-PCR and next-generation sequencing. *Nat. Protoc.* **5**, 1379–1395, <https://doi.org/10.1038/nprot.2010.87> (2010).
71. Hatakeyama, M. *et al.* SUSHI: an exquisite recipe for fully documented, reproducible and reusable NGS data analysis. *BMC Bioinformatics* **17**, 228, <https://doi.org/10.1186/s12859-016-1104-8> (2016).
72. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, <https://doi.org/10.1093/bioinformatics/bts635> (2013).
73. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108, <https://doi.org/10.1093/nar/gkt214> (2013).



74. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).
75. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25, <https://doi.org/10.1186/gb-2010-11-3-r25> (2010).
76. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13, <https://doi.org/10.1186/s13059-016-0881-8> (2016).

## Acknowledgements

We would like to thank Eric Verdin (UCSF, California, USA) for pEV731, Patrick Salmon (UNIGE, Geneva, Switzerland) for pCLX-UBI-GFP, Didier Trono (EPFL, Lausanne, Switzerland) for psPAX2, Arthur Nienhuis (SICRH, Tennessee, USA) and Patrick Salmon for pCAG-VSVG, Christina Ewald, Philipp Schätzle, Cornelius Fischer, and Thomas Liechti (Flow Cytometry Facility, UZH, Zurich, Switzerland) for cell sorting services, and Catharine Aquino and Weihong Qi (Functional Genomics Centre, UZH, Zurich, Switzerland) for RNA-Seq services. SUP-T1 cells were obtained through the NIH AIDS Reagent Programme, Division of AIDS, NIAID, NIH from Dharam Ablashi. This study was funded by the Swiss National Science Foundation grant 310030\_141067 (to KJM and HFG), the HIV-X MRD project 51MRP0\_158328 (to HFG and KJM), and the University of Zurich Clinical Research Priority Project: Viral Infectious Diseases: Zurich Primary HIV Infection (to HFG). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## Author Contributions

Y.L.K. designed the study, performed the experiments, analysed the data, and wrote the manuscript. S.S. and A.I. supported the experiments and analysis of the data. K.N., A.K. and L.J. supported the experiments. M.S., V.V. and R.D.K. supported the analysis of the data. C.B. and H.F.G. participated in the design of the study. K.J.M. designed the study, supported the analysis of the data, and edited the manuscript. All authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-28161-y>.

**Competing Interests:** K.J.M. has received travel grants and honoraria from Gilead Sciences, Roche Diagnostics, Tibotec, Bristol-Myers Squibb, and Abbott; the University of Zurich has received research grants from Gilead, Roche, and Merck Sharp & Dohme for studies that K.J.M. serves as principal investigator, and advisory board honoraria from Gilead Sciences. All other authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018