

RESEARCH

Open Access



# HIV integration sites in latently infected cell lines: evidence of ongoing replication

Jori Symons<sup>1†</sup>, Abha Chopra<sup>2†</sup>, Eva Malatinkova<sup>3</sup>, Ward De Spiegelaere<sup>3</sup>, Shay Leary<sup>2</sup>, Don Cooper<sup>2</sup>, Chike O. Abana<sup>4</sup>, Ajantha Rhodes<sup>1</sup>, Simin D. Rezaei<sup>1</sup>, Linos Vandekerckhove<sup>3</sup>, Simon Mallal<sup>2,4</sup>, Sharon R. Lewin<sup>1,5‡</sup> and Paul U. Cameron<sup>1,5\*‡</sup>

## Abstract

**Background:** Assessing the location and frequency of HIV integration sites in latently infected cells can potentially inform our understanding of how HIV persists during combination antiretroviral therapy. We developed a novel high throughput sequencing method to evaluate HIV integration sites in latently infected cell lines to determine whether there was virus replication or clonal expansion in these cell lines observed as multiple integration events at the same position.

**Results:** We modified a previously reported method using random DNA shearing and PCR to allow for high throughput robotic processing to identify the site and frequency of HIV integration in latently infected cell lines. Latently infected cell lines infected with intact virus demonstrated multiple distinct HIV integration sites (28 different sites in U1, 110 in ACH-2 and 117 in J1.1 per 150,000 cells). In contrast, cell lines infected with replication-incompetent viruses (J-Lat cells) demonstrated single integration sites. Following in vitro passaging of the ACH-2 cell line, we observed a significant increase in the frequency of unique HIV integration sites and there were multiple mutations and large deletions in the proviral DNA. When the ACH-2 cell line was cultured with the integrase inhibitor raltegravir, there was a significant decrease in the number of unique HIV integration sites and a transient increase in the frequency of 2-LTR circles consistent with virus replication in these cells.

**Conclusion:** Cell lines latently infected with intact HIV demonstrated multiple unique HIV integration sites indicating that these cell lines are not clonal and in the ACH-2 cell line there was evidence of low level virus replication. These findings have implications for the use of latently infected cell lines as models of HIV latency and for the use of these cells as standards.

## Background

Despite the success of suppressive combination antiretroviral therapy (cART), HIV persists as integrated provirus in long lived latently infected cells, typically resting memory CD4<sup>+</sup> T-cells [1, 2]. Latently infected memory CD4<sup>+</sup> T-cells are rare in individuals on cART, occurring at a frequency of 10–100 per million cells [3], and therefore, are difficult to study ex vivo. Multiple in vitro models

of HIV latency have been developed including latently infected cell lines and primary T-cells [4]. Understanding the location and frequency of HIV integration in the host genome in models of HIV latency as well as resting CD4<sup>+</sup> T-cells from HIV-infected individuals on cART can potentially provide insights into the origin of infection, clonal expansion and potentially the response to latency reversing agents [5].

Latently infected cell lines are established following infection with either intact, replication-competent virus or mutated, replication-defective viruses. Examples of cell lines infected with replication competent virus include U1, ACH-2 and J1.1 cells [6–9] and with replication defective virus include J-Lat, where the cell lines

\*Correspondence: paul.cameron@unimelb.edu.au

<sup>†</sup>Jori Symons and Abha Chopra are joint first authors

<sup>‡</sup>Paul U. Cameron and Sharon R. Lewin are joint senior authors

<sup>1</sup>The Peter Doherty Institute for Infection and Immunity, University of Melbourne and Royal Melbourne Hospital, 792 Elizabeth St, Melbourne, VIC 3000, Australia

Full list of author information is available at the end of the article

are monoclonal and harbour a single integration site [10, 11]. In CD4<sup>+</sup> T-cells from HIV-infected individuals on cART, several groups have recently shown a significant expansion of latently infected cells with a distinct site of integration, consistent with clonal expansion in vivo [5, 12–14]. Understanding whether similar patterns of integration occur in in vitro models of HIV latency and in patient derived cells is important, if these models are to be used to study the establishment, maintenance and reversal of latency.

Strategies to determine sites of HIV integration include sequencing and cloning [15, 16] or bulk sequencing [5, 12, 13, 17]. Most bulk sequencing approaches use restriction enzymes or random shearing of genomic DNA followed by PCR, using primers in the long terminal repeat (LTR) and a linker [5, 12, 13, 17]. Random shearing leads to different sized PCR products. Therefore, if an identical HIV integration site is detected but the length of the PCR product is different, it is most likely that this HIV integration sites was derived from a clonally expanded cell. Another method of determining the frequency of HIV integration sites is by limiting dilution of genomic DNA based on the estimated copies of HIV integrated DNA followed by loop amplification, and sequencing using primers located in the LTR [14].

Here, we describe a method to significantly streamline the assessment of HIV integration sites using robotic processing. Using this method, we evaluated HIV integration sites in commonly used latently infected cell lines and demonstrated that multiple cell lines that are traditionally used to study latency have evidence of productive infection.

## Methods

### Latently infected cell lines

Cells were obtained from NIH AIDS reagent program (Table 1) and were maintained in culture medium (CM) (RPMI 1640 medium (Life Technologies) supplemented with 10% (v/v) heat inactivated FCS, 100 µg/ml penicillin, 100 µg/ml streptomycin (Life Technologies) at 37 °C and 5% CO<sub>2</sub>. Cells were divided in a ratio of 1:6 or 1:10 twice weekly.

### Sample collection for integration site analysis

To assess HIV integration sites in a panel of latently infected cell lines, all cells were passaged ten times in a 1:6 dilution twice weekly. High molecular weight genomic DNA (gDNA) was isolated from every other passage (i.e. passages 0, 2, 4, 6, 8 and 10) according to manufacturer's protocol (Blood & Cell Culture DNA Mini Kit, QIAGEN). Passage zero was before the start of culture and sample was obtained from cryovials received

**Table 1 Cell lines analysed in this study from NIH AIDS reagent program**

| Cell line         | Replication competent virus | Site of mutation | Proviral copies reported per cell | Reference |
|-------------------|-----------------------------|------------------|-----------------------------------|-----------|
| U1                | Yes                         | Mutation in tat  | 2                                 | [6]       |
| ACH-2             | Yes                         | Mutation in tar  | 1                                 | [7, 8]    |
| J1.1              | Yes                         | Wild type        | 1                                 | [9]       |
| J-Lat 8.4         | No                          | Frame shift env  | 1                                 | [10]      |
| J-Lat 9.2         | No                          | Frame shift env  | 1                                 | [10]      |
| J-Lat 10.6        | No                          | Frame shift env  | 1                                 | [10]      |
| J-Lat 15.4        | No                          | Frame shift env  | 1                                 | [10]      |
| J-Lat tat-GFP 8.2 | No                          | LTR-tat-GFP      | 1                                 | [10, 11]  |
| J-Lat tat-GFP A1  | No                          | LTR-tat-GFP      | 1                                 | [10, 11]  |
| J-Lat tat-GFP H2  | No                          | LTR-tat-GFP      | 1                                 | [10, 11]  |
| J-Lat tat-GFP A72 | No                          | LTR-tat-GFP      | 1                                 | [10, 11]  |

*ENV* envelope, *LTR* long terminal repeat, *tat* trans-activator of transcription, *GFP* green fluorescent protein

from the NIH AIDS reagent program. To assess the efficiency of the HIV integration site analysis method, we spiked 1.5 µg high molecular weight gDNA from (peripheral blood mononuclear cells (PBMC) corresponding to approximately  $2 \times 10^5$  cells) derived from an HIV negative donor with high molecular weight gDNA from J-Lat 8.4, J-Lat 15.4 and J-Lat tat-GFP 82 cell line corresponding to a total of 125 cells. To test specificity, we analysed 1.5 µg gDNA PBMC from an HIV negative donor.

In some experiments ACH-2 cells were cultured in CM with and without 1 µM raltegravir (NIH AIDS reagent program). Cells were passaged nine times in a 1:10 dilution twice weekly and between passage 4 and 12, high molecular weight gDNA was isolated according to manufacturer's protocol (Blood and Cell Culture DNA Mini Kit, QIAGEN). HIV integration sites and frequency of 2-LTR circles were then quantified.

### High throughput HIV integration analysis method

High molecular weight DNA corresponding to  $1.5 \times 10^5$  cells was randomly digested for 20 min at 22 °C with NEBNext dsDNA fragmentase according to manufacturer's protocol (New England Biolabs, Ipswich, MA) to obtain short 200–1200 bp fragments for processing. Digested DNA was purified with AMPure XP (Agencourt Beckman Coulter, Nyon, Switzerland) in a 0.7:1 ratio, to remove small DNA fragments (<200 bp), buffer and fragmentase. The product was end-repaired, A-tailed and 10 pmol of linker (Additional file 1: Table S1) was ligated to the product using End-repair, ligation module (New England Biolabs) according to the manufacturer's instructions. After purification, restriction digest was performed with BglIII (New England Biolabs) to remove

the upstream sequence of HIV and prevent sequencing of proviral HIV DNA. Purified product was amplified using a nested approach. The first round PCR was two staged with a first step of LTR dependent linear amplification with the biotinylated primer LTR1\_F (Additional file 1: Table S1) for 98 °C for 3 min, followed by 12 cycles of 98 °C for 1 min, 60 °C for 30 s and 72 °C for 1 min, and then 10 min at 72 °C. In the second step, the PCR reaction was spiked with 1 µL 10 µM linker primer, 1stLink\_Primer (Additional file 1: Table S1) that could only anneal when there was LTR dependent elongation for increased specificity, and then the same cycling conditions for 27 cycles. After purification of the PCR product with the AMPure XP (Agencourt Beckman Coulter) to remove unbound biotinylated primers, biotinylated PCR products were selected by streptavidin-coupled Dynabead selection according to the manufacturer's protocol (Thermo Fisher, Malaga, Australia). Subsequently, a second round PCR was performed with the same cycling conditions for 35 cycles with the mid tagged primers LTR2\_F and 2ndLink\_Primer (Additional file 1: Table S1). Nested PCR was performed with Gotaq hotstart polymerase M5006 chemistry according to the manufacturer's protocol (Promega, Madison, WI).

The purified products were quantitated using Nanodrop (Thermo Fisher, Malaga, Australia). Post quantification, up to 12 samples were pooled in equimolar amounts and a sequence library was prepared using the Kapa Hyper Library prep kit (Kapa Biosystems, Wilmington, MS) using 500 ng of pooled product. The library was further quantified using qPCR and sequenced by 300 bp paired-end sequencing on Illumina MiSeq (Additional file 1: Fig. S1a) at appropriate concentrations. All the steps in the HIV integration site analysis methods were optimized for robotic processing to minimize contamination risk and increase throughput.

#### Data analysis

To determine an HIV integration site, the MiSeq reads were paired and needed to contain a multiplex-identifier (MID) sequence on either end of LTR and linker based sequence products. These sample read pairs were subsequently checked for LTR primer sequence and the remaining LTR nucleotide sequence ending on the nucleotide sequence CA. No mismatches were allowed in this sequence excluding CA (2 in primer sequence, 1 in LTR end). Furthermore, the linker primer had to be present. The sequence of the matched read pairs after the LTR or linker were trimmed to 50 base pairs and subsequently the trimmed read pairs were grouped according to 100% sequence match. Grouped sequences had to have  $\geq 3$  reads to proceed. Well represented sequences were used for chromosomal alignment which was determined

using the Blat-UCSC Genome Browser (GRCH38/hg38). An HIV integration site was called if the results had  $\geq 10$  reads, and the frequency was determined by a length difference of  $\geq 2$  nucleotides (Additional file 1: Fig. S2).

#### HIV DNA sequencing

DNA was extracted from pelleted ACH2 cells using QIAamp DNA Blood Mini kit (QIAGEN Inc., Hilden, Germany). Two overlapping amplicons spanning the HIV near full length genomes were PCR amplified using nested PCR with Platinum Taq DNA High Fidelity polymerase (Invitrogen, Carlsbad, CA). Each 25 µl reaction mixture contained 2.5 µl reaction mix (10 $\times$ ), 0.5 µl of each primer (0.5 mM), 0.25 µl Platinum TaqHigh Fidelity mix, 1.5 µl MgSO<sub>4</sub> (3 mM), 1.25 µl dNTP mix (2 mM) and 2.5 µl of template DNA. The two step cycling conditions were 94 °C for 2 min; 15 cycles of 94 °C for 15 s, 60 °C for 20 s, and 68 °C for 6 min, 20 additional cycles of 94 °C for 15 s, 65 °C for 20 s, 68 °C for 6 min and finally 68 °C for 10 min. Amplicons were verified by agarose gel electrophoresis and quantified using Nanodrop. Amplicons were pooled in equimolar amounts. 500 ng of pooled product was used for preparing the indexed libraries for 454FLX and Illumina MiSeq platforms. For 454FLX, the library was prepared using the GS FLX Titanium Rapid Library Preparation Kit as per manufacturer's protocol (Roche 454 Life Sciences, Branford, CT). Library was quantified using the Kapa library quantification kit with 454FLX primer premix (Kapa Biosystems) and sequenced on the 454 FLX Titanium XL + instrument (454 Life Sciences. Roche Applied Sciences, Branford, CT). For Illumina sequencing, indexed libraries was prepared using the PCR free Kapa Hyper Prep kit as per the manufacturer's protocol (Kapa Biosystems). The libraries were quantified using the KAPA Library Quantitation qPCR kit with Illumina Primer Premix (Kapa biosystems) and sequenced on Illumina MiSeq using 600 cycle MiSeq reagent kits (version 3). We analysed transition and transversion mutation additionally to insertions and deletion in the envelope gene with these two platforms. We called a mutation if the minority read count exceeded 1% of the total read count.

#### 2-LTR circles and CCR5

2-LTR circles were measured as previously described [18]. Briefly, 2-LTR circle sequences were amplified by nested PCR with the first round PCR conditions including a denaturation step of 8 min at 95 °C and 16 cycles of amplification (95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min), followed by an elongation step of 5 min at 72 °C and the second round denaturation step of 10 min at 95 °C and 40 cycles of amplification (95 °C for 10 s, 60 °C for 30 s) followed by an elongation step of 5 min at 72 °C.

The number of 2-LTR circles were normalized to cellular input by performing a separate real-time PCR assay to quantify the CCR5 coding sequence as previously described [19].

### Statistical analysis

The numbers of unique integration sites of all passages were scored independent of frequency. Differences during passaging were assessed using linear regression models. Differences in 2-LTR circles when the ACH-2 cell line was cultured with and without raltegravir was analyzed by a one-tailed paired *t* test. Differences in the decay of the number of ACH-2 integration sites with and without raltegravir were determined by regression analysis. Differences were considered statistically significant at  $p < 0.05$ . Gene ontology analysis was performed using the GENE SeT Analysis Toolkit [20]. Differences in ontology were analysed by Chi square test. All statistical analysis was performed using GraphPad Prism 6 and STATA.

## Results

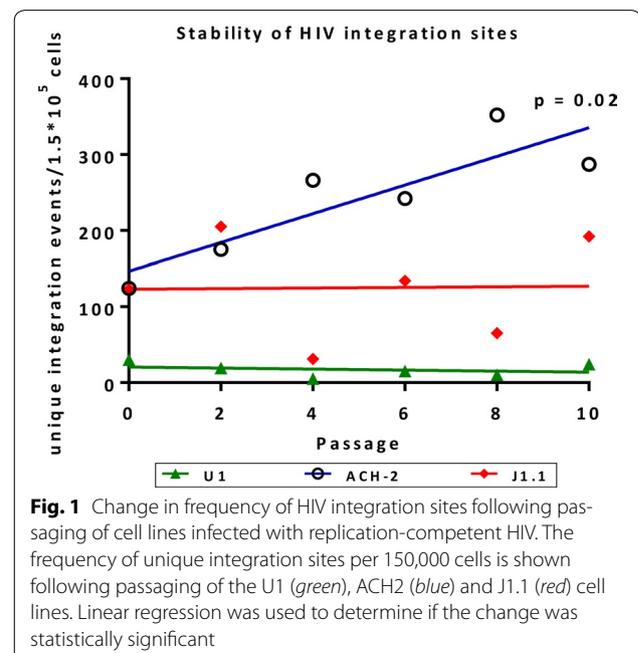
### Development of a novel high throughput assay to study HIV integration

We assessed the efficiency of a novel high throughput method to study HIV integration using a mix of uninfected (Jurkat and PBMC from HIV-uninfected donors) and infected cell lines (J-Lat 8.4, J-Lat 15.4 and J-Lat tat-GFP 82). We detected 24–30 of 125 (19–24%) integration sites in a background of 1.5  $\mu$ g gDNA from PBMCs from an HIV negative donor ( $n = 3$ ). To test the specificity of our assay, we assessed HIV integration sites in 1.5  $\mu$ g of gDNA derived from PBMCs from an HIV negative donor and no HIV integration sites were detected (data not shown,  $n = 3$ ). We included several modifications to a recently published method [12] (Additional file 1: Fig. S1b), but the main adaptation that improved throughput was enzymatic random digestion with NEBNext dsDNA fragmentase, based on previous reports showing similar efficacy in DNA fragmentation to sonication [21]. After fragmentation, the use of nd-repair ligation module (New England Biolabs) which is a single step allowed ligation without product clean-up and therefore reduced the number of steps considerably. Furthermore, in our protocol we omitted linker ligation. By not digesting the linker, the second round nested PCR could also be performed with barcoded linker primers. Therefore the reverse reads were barcoded and sample specific. These changes enabled the whole procedure to be performed in a 96 well format allowing for up to 96 samples to be analysed in 16 h, which was run entirely on a robotic platform, making sample processing consistent. Only sample preparation was performed manually which meant that on average integration site analysis of 6–8 samples could be completed in 24 h.

### Diverse integration sites in latently infected cell lines infected with intact replication-competent viruses

We next analysed HIV integration sites in established latently infected cell lines generated using intact replication-competent viruses or defective, replication-incompetent viruses. Cell lines reported to be infected with a replication-competent virus demonstrated multiple unique integration sites (number of unique integrations sites per 150,000 cells was 24–30, 89–124 and 123 in U1 ( $n = 2$ ), ACH-2 ( $n = 2$ ) and J1.1 cells ( $n = 1$ ) respectively). Whereas the number of total integration sites detected in 150,000 cells was 667–754, 589–649 and 577 in U1 ( $n = 2$ ), ACH-2 ( $n = 2$ ) and J1.1 cells ( $n = 1$ ) respectively). In contrast, all cell lines infected with a replication-incompetent virus (i.e. J-Lat cell lines) had a single unique integration site. This observation raised the possibility that there was ongoing infection of uninfected cells during cell passaging. In the ACH-2 cell line, the number of unique integration sites significantly increased over time in culture ( $p = 0.02$ ), whereas it remained stable in J1.1 and U1 (Fig. 1). In contrast, all HIV integration sites in the J-Lat cell lines remained stable over time and were restricted to a single unique site (Table 2) (with a median of 388 (range 265–413) of total integration sites detected per 150,000 cells). The HIV integration sites in J-Lat cells 8.4, 9.2 and 15.4 were detected as previously reported [22].

We also analysed the orientation of integrated HIV and if there were differences in integration within genes (intron, exon) or in nongenic regions in the U1, ACH-2 and J1.1 cell lines from passage zero and passage ten. We



**Table 2 HIV integration sites in latently infected cell lines**

| Cell line         | Replication competent virus | Chromosome <sup>a</sup> | Position <sup>b</sup> | Gene <sup>c</sup> | Gene orientation <sup>d</sup> | HIV orientation <sup>e</sup> | Percentage of events <sup>f</sup> |
|-------------------|-----------------------------|-------------------------|-----------------------|-------------------|-------------------------------|------------------------------|-----------------------------------|
| U1                | Yes                         | 2                       | 48,177,527            | AC079807.4        | –                             | +                            | ≈38%                              |
|                   |                             | X                       | 38,811,467            |                   |                               | –                            | ≈55%                              |
|                   |                             | 19                      | 34,452,847            | UBA2              | +                             | +                            | ≈3%                               |
| ACH-2             | Yes                         | 7                       | 33,019,791            | NT5C3A            | –                             | –                            | ≈37%                              |
|                   |                             | 9                       | 128,111,651           | SLC25A25-AS       | –                             | –                            | ≈13%                              |
|                   |                             | 2                       | 26,386,668            | EPT1              | +                             | +                            | ≈2%                               |
|                   |                             | 17                      | 81,558,479            | NPLOC4            | +                             | –                            | ≈1%                               |
| J1.1              | Yes                         | 11                      | 685,243               | DEAF1             | –                             | –                            | ≈41%                              |
|                   |                             | 12                      | 54,257,973            | CBX5              | –                             | –                            | ≈38%                              |
| J-Lat 8.4         | No                          | 1                       | 7,946,384             | FUBP1             | –                             | –                            | 100%                              |
| J-Lat 9.2         | No                          | 19                      | 4,381,104             | PPP5C             | +                             | +                            | 100%                              |
| J-Lat 10.6        | No                          | 9                       | 136,468,579           | SEC16A            | –                             | +                            | 100%                              |
| J-Lat 15.4        | No                          | 19                      | 34,441,293            | UBA2              | +                             | +                            | 100%                              |
| J-Lat tat-GFP 8.2 | No                          | 10                      | 39,936,068            |                   |                               | +                            | 100%                              |
| J-Lat tat-GFP A1  | No                          | X                       | 34,073,326            |                   |                               | –                            | 100%                              |
| J-Lat tat-GFP H2  | No                          | 2                       | 171,821,429           | SLC25A12          | –                             | +                            | 100%                              |
| J-Lat tat-GFP A72 | No                          | X                       | 45,038,538            | KDM6A             | +                             | –                            | 100%                              |

<sup>a</sup> The particular chromosome where the HIV integration site is located

<sup>b</sup> The first human genome nucleotide following 3'-end LTR

<sup>c</sup> The gene where HIV is integrated

<sup>d</sup> Transcription direction of the gene

<sup>e</sup> Transcription direction of HIV

<sup>f</sup> Percentage of events where the particular HIV integration site was detected in all passages

found that most integration sites were within genes and specifically introns, and there was no preference in orientation of HIV integration and this did not change following passaging (Fig. 2).

#### Identity of genes where HIV integrates in latently infected cell lines

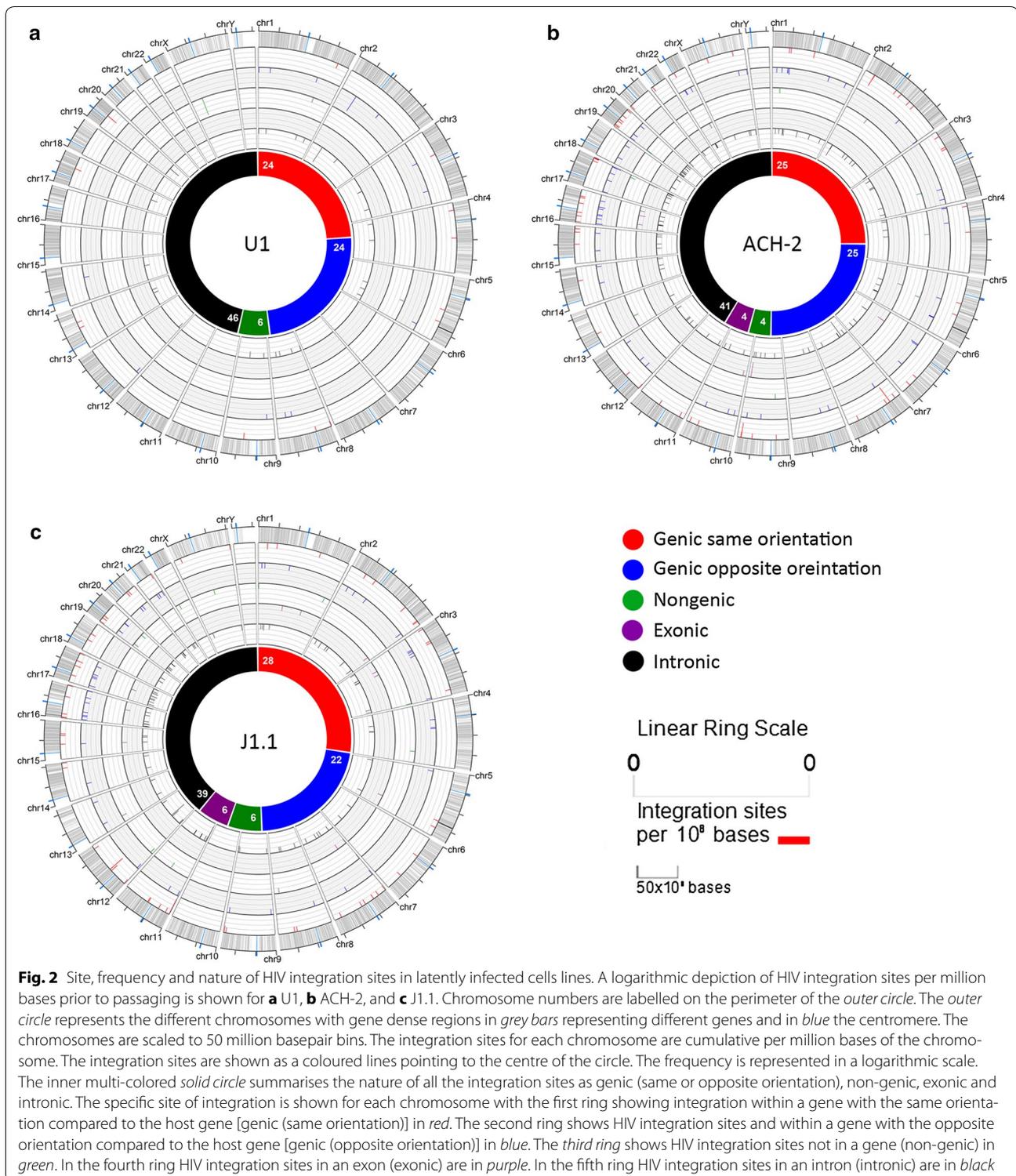
We next analysed the specific gene locations of HIV integration sites that were frequently detected during in vitro passaging. In the U1 cell line, we detected two major sites with high frequency. One site was in chromosome 2 in the *AC079807.4* gene with the 3' LTR junction at nucleotide position 48,177,527 whereas the other site was found in chromosome X in a functionally unknown region. Interestingly, an additional site was continuously detected in chromosome 19, in the ubiquitin-like modifier activating enzyme 2 (*UBA2*) gene, in about 4 percent of events that were detected (Table 2).

In the ACH-2 cell line, the major HIV integration site was detected in chromosome 7 in the cytosolic 5'-nucleotidase 3A (*NT5C3A*) gene as previously reported [17, 23]. We also detected an additional integration site in chromosome 9 which was always detected in high frequency in the solute carrier family 25 antisense RNA (*SLC25A25-AS*) noncoding RNA gene as recently

reported [17]. Additionally, we detected two sites with multiple integration events intermittently, one in chromosome 2 in the ethanolaminephosphotransferase 1 (*EPT1*) gene, the other in chromosome 17 in the nuclear protein localization 4 (*NPLOC4*) gene (Table 2).

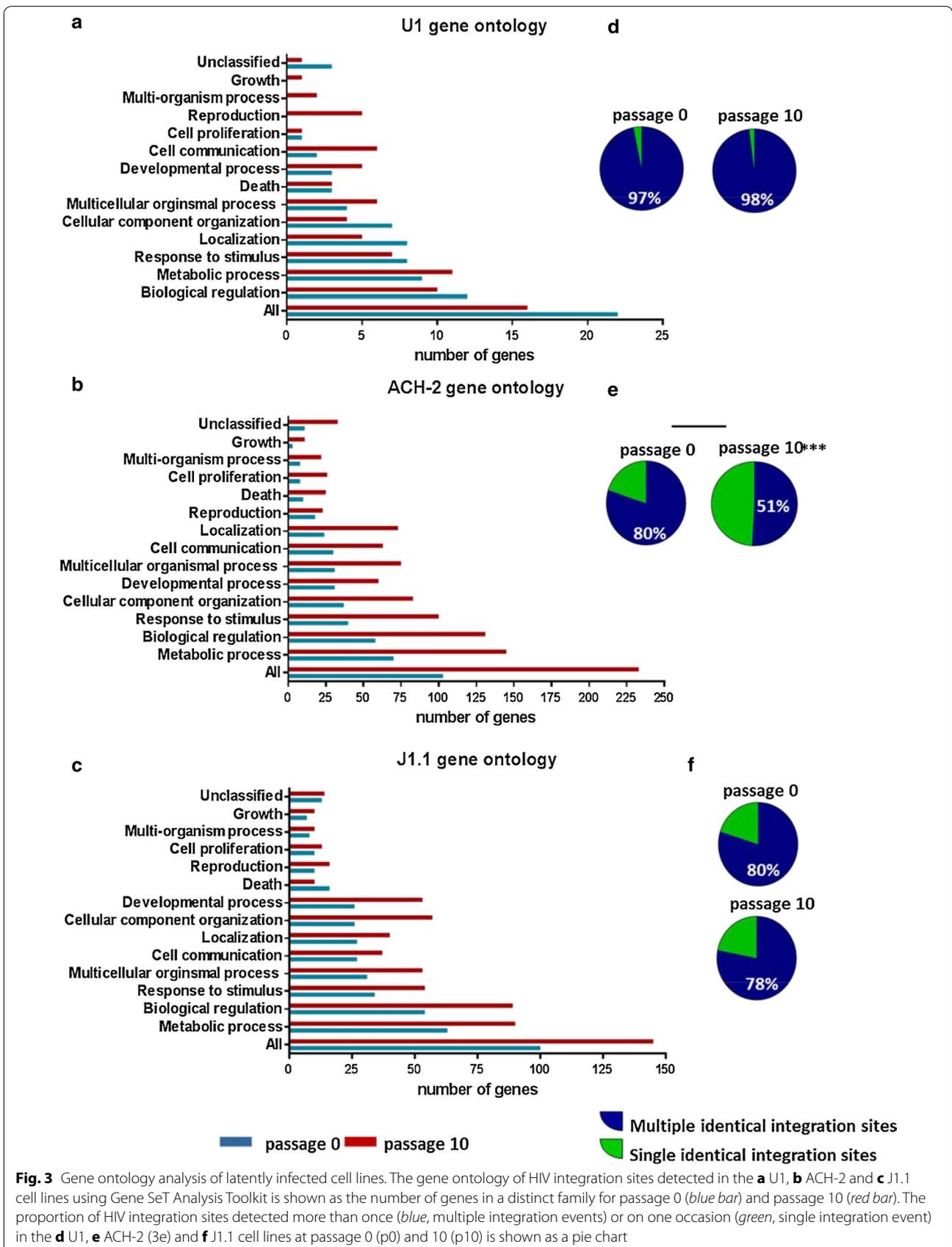
Although previous reports have described a single copy of integrated HIV per cell in the J1.1 cells [9], we detected two major integration sites in 150,000 cells with equal frequency in passage zero before start of cell culture. One was detected in chromosome 11 in the deformed epidermal autoregulatory factor 1 (*DEAF1*) gene and the other in chromosome 12 in the chromobox homolog 5 (*CBX5*) gene (Table 2). During the ten passages, the HIV integration site in all the J-Lat cell lines remained single and stable and the identified integration sites represented 100% of the total events detected (Table 2).

We analysed the genes where HIV integrated following passage zero and passage ten for the U1, ACH-2 and J1.1 cell lines. We observed that with passaging, the number of HIV integration sites in genes decreased in the U1 cell line (Fig. 3a), but increased in ACH-2 (Fig. 3b) and J1.1 (Fig. 3c) cell lines. The proportion of integration sites in genes associated with cellular proliferation was approximately 10% and did not significantly differ between the passages in all cells.



We then compared the integration sites in the cell lines with those found in CD4<sup>+</sup> T-cells from HIV-infected individuals on cART, which have been previously reported in the literature [5, 12–14, 24]. The number

of integration sites we detected in the U1, ACH-2 and J1.1 cell lines overlapped 22, 33 and 35% respectively when compared to HIV integration sites found in cells from HIV-infected individuals on ART. This suggests



that many sites of HIV integration identified in latently infected cell lines were not identified in cells from HIV-infected individuals on ART.

#### Multiple identical integration sites in latently infected cell lines

We analysed the percentage of identical integration events compared to total integration sites per cell line. Identical integration sites were defined as two or more integration events. In the U1 cell line, which contains an intact virus, most integration sites were identical and detected multiple times (97% of all integration sites detected were identical at passage 0) and did not change during passage (98% at passage 10) (Fig. 3d). In the ACH-2 cell line, we observed a significant decrease in the percentage identical integration events (from 80% at passage 0–51% at passage 10,  $p < 0.001$ ) (Fig. 3e). In the J1.1 cell line, the percentage of identical integration events did not significantly differ during passage (80% at passage 0, 78% at passage 10) (Fig. 3f).

#### Change in integration sites over time in culture is consistent with residual virus replication in latently infected ACH-2 cell line

To determine if the observed effects in the ACH-2 cell line were due to ongoing virus replication, we sequenced the integrated provirus from the ACH-2 cell line from passage 1 with two deep sequencing methods and analysed the envelope gene and demonstrated multiple transition and transversion mutations, insertions and deletions with 454 sequencing. Moreover with illumina Miseq sequencing we observed multiple transition (50) and transversion [36] mutations, insertions [8] and deletions [5] With 70–80% of the mutations, insertions and deletions detected in the Miseq data overlapping with the mutations found by 454 sequencing.

Subsequently, we analysed the frequency of 2-LTR circles in the ACH-2 cell line cultured with and without 1  $\mu$ M raltegravir. Surprisingly, in the ACH-2 cell line, we detected 2-LTR circles in cells collected from each passage. We detected significantly more 2-LTR circles per million cells in the first three passages with raltegravir compared to without raltegravir ( $p = 0.04$ ,  $p = 0.02$  and  $p < 0.01$  for the first three passages compared to baseline). As expected, 2-LTR circles were detected in the positive control of the CEM cell line spiked with the 2-LTR plasmid while 2-LTR circles were not detected in the uninfected CEM cell line, in the CEM cell line spiked with pNL4-3 or the J-Lat 15.4 and 10.6 cell lines (Fig. 4a). We observed a decrease in the number of unique integration sites over time following twice weekly passage and a faster decay in the number of HIV integration sites in the ACH-2 cell line cultured with raltegravir compared

to without raltegravir (half-life of number of integration sites per passage was 1.81 versus 2.96 passage respectively) although the difference was not statistically significant ( $p = 0.57$ ) (Fig. 4b). Moreover, we observed a consolidation around the reported dominant integration site in chromosome seven from 32% in passage four to 74% in passage 12.

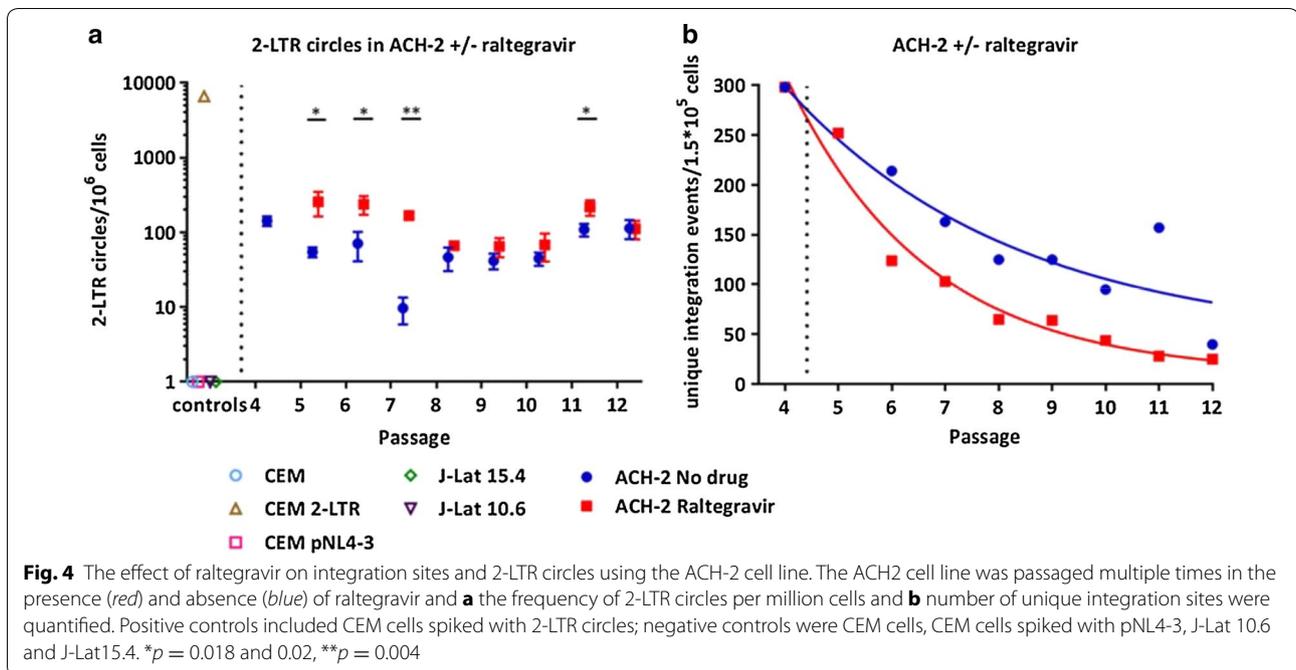
#### Discussion

In this study, we developed a high-throughput assay to analyse HIV integration sites using enzymatic random fragmentation. Our assay was specific and was able to detect 25% of the HIV integration sites using infected cell lines (J-Lats) in a background of human gDNA. In cell lines infected with replication-incompetent viruses, we observed single integration sites. In contrast, in cell lines infected with replication-competent viruses, we demonstrated multiple unique HIV integration sites and in the ACH-2 cell line, the diversity of integration sites increased over time in culture and reduced in the presence of the antiretroviral drug raltegravir. Therefore, these latently infected cell lines were not clonal and the findings were consistent with residual virus replication.

The U1 monocytic cell line was cloned from a population of chronically HIV-1 infected U937 cells and is one of the most-studied cell lines for HIV latency [7]. It has been reported that U1 cells contain two copies of integrated HIV DNA [7] and latency is maintained by mutations in Tat with one Tat containing a point mutation, H13L, and the other lacking the ATG initiation codon [25]. Our analysis demonstrated that the U1 cell line contains multiple distinct HIV integration sites that declined with passaging.

J1.1 cells are latently HIV infected cell line cloned by limiting dilution from HIV-infected Jurkat cells. Although, J1.1 cells are infected with wild type HIV<sub>LAV</sub>, the HIV-infected T-cell clone is defective in IL-2 production and Ca<sup>2+</sup> mobilization after CD3 stimulation [9]. Furthermore, J1.1 cells are reported to be infected with a single integrated copy per cell [9]. Interestingly, we detected multiple HIV integration sites and identified two major integration sites in equal frequency. Why there is preferential expansion or persistence of these specific integration sites is unclear.

The ACH-2 cell line was established by infection of the acute lymphoblastic leukemic T-cell line A3.01 with HIV<sub>LAI</sub>. Cells that survived were cloned resulting in a cell line that has been reported to contain a single integrated copy of HIV [6, 8]. The major HIV insertion site has been identified in chromosome 7 [23]. In addition, HIV latency is established by a point mutation in TAR (C37T), whereby this mutation impairs the response of the LTR to Tat [26]. Low level transcription of multiply spliced



transcripts persist in ACH-2 cells in the absence of stimulation [27], and here, we show that ACH-2 cells harbour multiple integration sites in addition to the major site in chromosome 7, as recently demonstrated by others [17].

Here, we present multiple lines of evidence consistent with residual virus replication in, ACH-2 cell line, and evidence suggestive for multiple rounds of replication occurring in the U1 and J1.1 cells. The evidence included diversity of integration sites in all three cell lines that was stable or increasing study directly links several lines of evidence demonstrating residual viral over time, and in the ACH-2 cell line, an increase in diversity during passage, numerous mutations in the proviral DNA and an increase in 2-LTR circles in the presence of raltegravir. Other groups have demonstrated viral production in the latently infected cell lines, U1, ACH-2 and J1.1 cells demonstrating continuous expression of viral RNA [28] and detection of 2-LTR circles in the ACH-2 cell line [29]. Our results are also consistent with the observed increase in integration sites in the HIV infected cell line H61-D during culture [30]. Olivares et al. demonstrated that during prolonged culture of these cells up to 100 passages, integration sites were detected in addition to the two sites found at passage zero [30].

Recently, Sunshine et al. [17] hypothesised that the presence of infrequent integration sites in the ACH-2 cell line indicated low-level viral replication. Our study presents several lines of evidence demonstrating residual viral replication in the latently infected cell line ACH-2. Additionally, our study is the first to suggest ongoing

low level replication in other cell lines infected with replication-competent viruses commonly used to study HIV latency. It is important to note that there is no perfect model to study HIV latency in vitro. Latently infected cell lines remain valuable in the assessment of latency or in activation studies using latency reversing agents. The advantages of latently infected cell lines include a high frequency of infected cells, the capacity to repeat experiments reproducibly and the ability to answer specific mechanistic questions that often isn't possible in primary cells.

It is possible that in our study, we increased the chance of cell-to-cell transmission of HIV by passaging cells at low dilution, which would allow for cell clumping. It has been previously demonstrated that cell-to-cell transmission increases with higher cell density [31, 32]. The degree of cell clumping could therefore have an impact on the level of residual virus replication in this cell line. Interestingly, the ACH-2 cell line is reported to not express CD4 on the cell surface, although transcription of CD4 can be induced [33]. Similarly, the H61-D cell line that demonstrated an increase in integration sites during culture did not express CD4 at the cell surface [30]. Cell-to-cell transmission can occur through a non-CD4 dependent process [34, 35] or CD4 could be expressed at low levels in the virological synapse.

Similar to other studies, we did not identify a high frequency of the same genes harbouring HIV integration in the latently infected cell lines that have been described in CD4<sup>+</sup> T-cells from HIV-infected individuals on cART

[5, 12–14, 24]. HIV integration in BACH2 and MKL2 [5, 14], although commonly found in patient derived cells were not detected in all the different cell lines. It has been suggested that the selective pressure for expansion of HIV integration sites into genes associated with cell proliferation is absent, since these cell lines have been previously immortalized. However, the proportion of HIV integration sites in genes associated with cell proliferation was similar in the U1, ACH-2 and J1.1 cell lines and in the CD4<sup>+</sup> T-cells from HIV-infected individuals on cART [5, 14]. Similar findings were recently demonstrated also using the ACH-2 cell line as well as a primary T-cell latency model [17]. Given that the same orientation of HIV integration relative to the host gene increases transcriptional efficiency and latency [36], this might potentially explain the higher observed basal HIV reverse transcriptase activity in ACH-2 compared to the U1 cell line [37]. There were several limitations in our study. First, we only report HIV integration sites in latently infected cell lines. However, others have compared cell lines to primary cell models, and observed similar patterns of HIV integration with respect to increased integration in intronic regions and no specificity for HIV integration orientation [15, 38, 39]. In addition, there was some overlap in the sites of integration in the cell lines and published work on cells from HIV-infected individuals on cART [5, 12, 13, 24]. Second, we were not able to definitely demonstrate if these cells harboured multiple HIV integration sites per cell or whether these cell lines contained multiple cells with different HIV integration sites. We suspect that multiple integration sites per cell, in cell lines and in vivo is rare but can occur [30, 40–42]. Moreover, it seems unlikely that a new integration event in a cell line is associated with the loss of the dominant integration site.

## Conclusions

We present a specific and efficient new method to analyse the location and frequency of HIV integration sites suitable for high throughput processing. Using HIV latently infected cell lines we demonstrated that latently infected cell lines that are infected with a replication-competent HIV had multiple HIV integration sites, and that this was likely due to low level replication. These findings have implications for the use of some latently infected cell lines as models for HIV latency whereas they are not truly latent and one must take into account that some cells might have a different number of integrated proviruses per cell than previously thought when one uses these cells for standardization of assays.

## Additional file

**Additional file 1.** Schematic overview of methods used for HIV integration site analysis.

## Authors' contributions

JS, AC, EM, COA, AR, and SDR performed the experiments. SL and DC created sequence analysis program and formatted data. WDS, LV, SM, SRL and PUC designed the study and provided intellectual input and with JS and AC analysed data. JS, SRL and PUC wrote the manuscript. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> The Peter Doherty Institute for Infection and Immunity, University of Melbourne and Royal Melbourne Hospital, 792 Elizabeth St, Melbourne, VIC 3000, Australia. <sup>2</sup> Institute for Immunology and Infectious Diseases (IIID), Murdoch University, Murdoch, WA, Australia. <sup>3</sup> HIV Translational Research Unit, Department of Internal Medicine, Faculty of Medicine and Health Sciences, Ghent University Hospital, Ghent University, Ghent, Belgium. <sup>4</sup> Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, TN 37232, USA. <sup>5</sup> Department of Infectious Diseases, Alfred Hospital and Monash University, Melbourne, Australia.

## Acknowledgements

The following reagents were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 infected U937 Cells (U1), ACH-2 and J1.1 from Dr. Thomas Folks, J-Lat Full Length Cells (clone 8.4, 9.2, 10.6 and 15.4) and J-Lat Tat-GFP Cells (clone 82, A1, H7 and A72) from Dr. Eric Verdin. We thank Ashish Nair (University of Melbourne) for his help processing the human buffy coats.

## Competing interests

The authors declare that they have no competing interests.

## Data accessibility

The datasets supporting the conclusions of this article are available in the Sequence Read Archive (SRA), Accession number SRP078468 in <http://www.ncbi.nlm.nih.gov/sra>.

## Ethics approval and consent to participate

The use of blood samples from normal donors for this study was approved by the Melbourne human research ethics committee (HREC) 1443071. Donors were recruited by the Red Cross Blood Transfusion Service as normal blood donors and all provided written informed consent for the use of their blood products for the research.

## Funding

PUC; National Health and Medical Research Council, NHMRC, APP1058891. SRL; Delaney AIDS Research Enterprise, DARE, U19 AI96109-01, National Health and Medical Research Council, NHMRC APP1052979; 1058891. EM; Agency for Innovation by Science and Technology of the Flemish Government, IWT, 111286. LV; Research Foundation—Flanders, FWO, 1.8.020.09. N.00. WDS; Research Foundation—Flanders, FWO, 12G9716N.

Received: 14 July 2016 Accepted: 22 December 2016

Published online: 13 January 2017

## References

1. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci USA*. 1997;94(24):13193–7.
2. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*. 1997;278(5341):1295–300.
3. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*. 2013;155(3):540–51.
4. Spina CA, Anderson J, Archin NM, Bosque A, Chan J, Famiglietti M, et al. An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4<sup>+</sup>T cells from aviremic patients. *PLoS Pathog*. 2013;9(12):e1003834.

5. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*. 2014;345(6193):179–83.
6. Folks TM, Clouse KA, Justement J, Rabson A, Duh E, Kehrl JH, et al. Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. *Proc Natl Acad Sci USA*. 1989;86(7):2365–8.
7. Folks TM, Justement J, Kinter A, Dinarello CA, Fauci AS. Cytokine-induced expression of HIV-1 in a chronically infected promonocyte cell line. *Science*. 1987;238(4828):800–2.
8. Clouse KA, Powell D, Washington I, Poli G, Strebel K, Farrar W, et al. Monokine regulation of human immunodeficiency virus-1 expression in a chronically infected human T cell clone. *J Immunol*. 1989;142(2):431–8.
9. Perez VL, Rowe T, Justement JS, Butera ST, June CH, Folks TM. An HIV-1-infected T cell clone defective in IL-2 production and Ca<sup>2+</sup> mobilization after CD3 stimulation. *J Immunol*. 1991;147(9):3145–8.
10. Jordan A, Bisgrove D, Verdin E. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J*. 2003;22(8):1868–77.
11. 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*. 2001;20(7):1726–38.
12. Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, et al. HIV-1 integration landscape during latent and active infection. *Cell*. 2015;160(3):420–32.
13. Simonetti FR, Sobolewski MD, Fyne E, Shao W, Spindler J, Hattori J, et al. Clonally expanded CD4+T cells can produce infectious HIV-1 in vivo. *Proc Natl Acad Sci USA*. 2016;113(7):1883–8.
14. Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*. 2014;345(6196):570–3.
15. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*. 2002;110(4):521–9.
16. Vatakis DN, Kim S, Kim N, Chow SA, Zack JA. Human immunodeficiency virus integration efficiency and site selection in quiescent CD4+T cells. *J Virol*. 2009;83(12):6222–33.
17. Sunshine S, Kirchner R, Amr SS, Mansur L, Shakhbatyan R, Kim M, et al. HIV integration site analysis of cellular models of HIV latency with a probe-enriched next-generation sequencing assay. *J Virol*. 2016;90(9):4511–9.
18. Vandergeeten C, Fromentin R, Merlini E, Lawani MB, DaFonseca S, Bake-man W, et al. Cross-clade ultrasensitive PCR-based assays to measure HIV persistence in large-cohort studies. *J Virol*. 2014;88(21):12385–96.
19. Zhang L, Lewin SR, Markowitz M, Lin HH, Skulsky E, Karanicolos R, et al. Measuring recent thymic emigrants in blood of normal and HIV-1-infected individuals before and after effective therapy. *J Exp Med*. 1999;190(5):725–32.
20. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res*. 2005;33(Web Server issue):W741–8.
21. Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS ONE*. 2011;6(11):e28240.
22. Lenasi T, Contreras X, Peterlin BM. Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host Microbe*. 2008;4(2):123–33.
23. Ishida T, Hamano A, Koiwa T, Watanabe T. 5' Long terminal repeat (LTR)-selective methylation of latently infected HIV-1 provirus that is demethylated by reactivation signals. *Retrovirology*. 2006;3:69.
24. Marini B, Kertesz-Farkas A, Ali H, Lucic B, Lisek K, Manganaro L, et al. Nuclear architecture dictates HIV-1 integration site selection. *Nature*. 2015;521(7551):227–31.
25. Emiliani S, Fischle W, Ott M, Van Lint C, Amella CA, Verdin E. Mutations in the tat gene are responsible for human immunodeficiency virus type 1 postintegration latency in the U1 cell line. *J Virol*. 1998;72(2):1666–70.
26. Emiliani S, Van Lint C, Fischle W, Paras P Jr, Ott M, Brady J, et al. A point mutation in the HIV-1 Tat responsive element is associated with postintegration latency. *Proc Natl Acad Sci USA*. 1996;93(13):6377–81.
27. Cannon P, Kim SH, Ulich C, Kim S. Analysis of Tat function in human immunodeficiency virus type 1-infected low-level-expression cell lines U1 and ACH-2. *J Virol*. 1994;68(3):1993–7.
28. Butera ST, Roberts BD, Lam L, Hodge T, Folks TM. Human immunodeficiency virus type 1 RNA expression by four chronically infected cell lines indicates multiple mechanisms of latency. *J Virol*. 1994;68(4):2726–30.
29. Besansky NJ, Butera ST, Sinha S, Folks TM. Unintegrated human immunodeficiency virus type 1 DNA in chronically infected cell lines is not correlated with surface CD4 expression. *J Virol*. 1991;65(5):2695–8.
30. Olivares I, Sanchez-Jimenez C, Vieira CR, Toledano V, Gutierrez-Rivas M, Lopez-Galindez C. Evidence of ongoing replication in a human immunodeficiency virus type 1 persistently infected cell line. *J Gen Virol*. 2013;94(Pt 5):944–54.
31. Janas AM, Wu L. HIV-1 interactions with cells: from viral binding to cell-cell transmission. *Current protocols in cell biology/editorial board, Juan S Bonifacino [et al]*. 2009;Chapter 26:Unit 26 5.
32. Sowinski S, Jolly C, Berninghausen O, Purbhoo MA, Chauveau A, Kohler K, et al. Membrane nanotubes physically connect T cells over long distances presenting a novel route for HIV-1 transmission. *Nat Cell Biol*. 2008;10(2):211–9.
33. Serpente N, Hemar A, Cefai D, Dautry-Varsat A, Fagard R, Fischer S, et al. Transcriptional and post-transcriptional mechanisms are involved in the absence of CD4 surface expression in two HIV-1 chronically infected T cell lines. *Int Immunol*. 1993;5(8):939–47.
34. Cao YZ, Friedman-Kien AE, Huang YX, Li XL, Mirabile M, Moudgil T, et al. CD4-independent, productive human immunodeficiency virus type 1 infection of hepatoma cell lines in vitro. *J Virol*. 1990;64(6):2553–9.
35. Hoffman TL, LaBranche CC, Zhang W, Canziani G, Robinson J, Chaiken I, et al. Stable exposure of the coreceptor-binding site in a CD4-independent HIV-1 envelope protein. *Proc Natl Acad Sci USA*. 1999;96(11):6359–64.
36. Shan L, Yang HC, Rabi SA, Bravo HC, Shroff NS, Irizarry RA, et al. Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. *J Virol*. 2011;85(11):5384–93.
37. Ho WZ, Lai JP, Li Y, Douglas SD. HIV enhances substance P expression in human immune cells. *FASEB J*. 2002;16(6):616–8.
38. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res*. 2007;17(8):1186–94.
39. Stevens SW, Griffith JD. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J Virol*. 1996;70(9):6459–62.
40. Josefsson L, Palmer S, Faria NR, Lemey P, Casazza J, Ambrozak D, et al. Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4+ T-cells contain one HIV-1 DNA molecule. *PLoS Pathog*. 2013;9(6):e1003432.
41. Altfield M, Allen TM, Yu XG, Johnston MN, Agrawal D, Korber BT, et al. HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. *Nature*. 2002;420(6914):434–9.
42. Ott DE, Nigida SM Jr, Henderson LE, Arthur LO. The majority of cells are superinfected in a cloned cell line that produces high levels of human immunodeficiency virus type 1 strain MN. *J Virol*. 1995;69(4):2443–50.