Monocyte-derived macrophages exhibit distinct and more restricted HIV-1 integration site repertoire than CD4⁺ T cells

Yik Lim Kok¹,²*, Valentina Vongrad¹,²*, Mohamed Shilaih¹,²*, Francesca Di Giannonaro³, Herbert Kuster¹,², Roger Kouyos¹,², Huldrych F. Günthard¹,² & Karin J. Metzner¹,²

The host genetic landscape surrounding integrated HIV-1 has an impact on the fate of the provirus. Studies analysing HIV-1 integration sites in macrophages are scarce. We studied HIV-1 integration site patterns in monocyte-derived macrophages (MDMs) and activated CD4⁺ T cells derived from seven antiretroviral therapy (ART)-treated HIV-1-infected individuals whose cells were infected ex vivo with autologous HIV-1 isolated during the acute phase of infection. A total of 1,684 unique HIV-1 integration sites were analysed. Their distribution in the human genome and genetic features, and the effects of HIV-1 integrase polymorphisms on the nucleotide selection specificity at these sites were indistinguishable between the two cell types, and among HIV-1 isolates. However, the repertoires of HIV-1-hosting gene clusters overlapped to a higher extent in MDMs than in CD4⁺ T cells. The frequencies of HIV-1 integration events in genes encoding HIV-1-interacting proteins were also different between the two cell types. Lastly, HIV-1-hosting genes linked to clonal expansion of latently HIV-1-infected CD4⁺ T cells were over-represented in gene hotspots identified in CD4⁺ T cells but not in those identified in MDMs. Taken together, the repertoire of genes targeted by HIV-1 in MDMs is distinct from and more restricted than that of CD4⁺ T cells.

Received: 16 December 2015
Accepted: 21 March 2016
Published: 12 April 2016

¹Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland.
²Institute of Medical Virology, University of Zurich, Zurich, Switzerland.
³Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, Australia. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to K.J.M. (email: karin.metzner@usz.ch)
ex vivo infection of primary monocyte-derived macrophages (MDMs) is a practical solution. The study of Barr et al. examined HIV-1 integration site patterns in MDMs and compared them with those in CD4+ T cells obtained from other studies. They observed that HIV-1 integration events are found more often in long intergenic regions and, correspondingly, less often in gene-dense regions in MDMs compared to CD4+ T cells. In the study of Wellensiek et al., although they had examined HIV-1 integration site patterns in MDMs and CD4+ T cells, they focussed on the differences between adult and neonatal cells of each cell type rather than between cell types. To date, a systematic comparison of HIV-1 integration site patterns between MDMs and CD4+ T cells derived from the same experimental settings has never been done.

Here, we extended the characterization of HIV-1 integration site patterns in MDMs and CD4+ T cells in an experimental setting that mimics the in vivo situation. We used MDMs and CD4+ T cells derived from ART-treated HIV-1-infected individuals and infected these cells ex vivo with the individuals’ autologous HIV-1 isolated during the acute phase of their HIV-1 infection. We focussed on other aspects of HIV-1 integration site patterns, e.g. structural and functional gene clusters formed by HIV-1-hosting genes and the frequencies of HIV-1 integration events in certain gene subsets, and compared these features between the two cell types. We also examined the effects of two naturally occurring amino acid polymorphisms in the HIV-1 integrase on its nucleotide selection specificity at HIV-1 integration sites in primary cells. We showed that the repertoire of genes targeted by HIV-1 in MDMs was distinct and more restricted than that of CD4+ T cells.

Results

Characteristics of study participants and their autologous primary HIV-1 isolates. At the time of autologous HIV-1 isolation, the study participants were infected with HIV-1 for an estimated 15–153 days (Supplementary Table S1). All HIV-1-infected individuals are enrolled in the Zurich Primary HIV Infection (ZPHI) study20 and six of seven HIV-1-infected individuals started early ART during the acute phase of infection. The replication capacities of these primary HIV-1 isolates in MDMs were most comparable to that of HIV-1 JR-FL, a macrophage-tropic strain. The replication capacity of primary HIV-1 isolate 6 in MDMs was consistently the highest among the seven primary HIV-1 isolates (Supplementary Figure S1). All primary HIV-1 isolates were also replication competent in CD4+ T cells (Supplementary Figure S2).

At the time of MDMs and CD4+ T cell isolation from HIV-1-infected individuals, viral suppression ranged from 2.8–7.8 years (mean = 4.7 yr) (Supplementary Table S1). These cells were then infected ex vivo with autologous primary HIV-1 isolates or heterologous HIV-1 isolates. The non-restrictive linear amplification-mediated polymerase chain reaction (nrLAM-PCR) was used to amplify the 5’ HIV-1 integration junctions, which were analysed using our new Integration Site Analysis Pipeline (InStAP) (Supplementary Figure S3). Prior to these experiments, CD4+ T cells from three ART-treated HIV-1-infected individuals were activated and monitored for HIV-1 outgrowth. Reactivation of HIV-1 from latently infected cells did not occur within 7–8 days (data not shown). Thus, we are certain that the HIV-1 integration sites amplified from CD4+ T cells after two days of activation and a further two days of ex vivo infection were the result of these ex vivo infections. More importantly, no HIV-1 integration sites were found in mock infected cell samples of HIV-1-infected individuals.

The basic genetic requirements for HIV-1 integration were similar among individual HIV-1 isolates and between MDMs and CD4+ T cells. We obtained a total of 1,484 unique HIV-1 integration sites, derived from the seven selected HIV-1-infected individuals whose MDMs (n = 987) and CD4+ T cells (n = 497) were infected ex vivo with autologous primary HIV-1 isolates and heterologous HIV-1 isolates (Supplementary Table S2). Of these, 783 (79.3%) and 431 (86.7%) integration sites were found within genes in MDMs and CD4+ T cells, respectively. HIV-1 integration sites were also found mainly in gene-dense regions and away from centromeres, Giemsa dark regions, and the p arms of acrocentric chromosomes in both MDMs and CD4+ T cells (Fig. 1). The genetic distribution and genetic profiles of HIV-1 integration sites were consistent with previous studies18,22, and they were indistinguishable between MDMs and CD4+ T cells, as well as between pooled autologous and heterologous HIV-1 isolates (Supplementary Figure S4). These basic genetic requirements for HIV-1 integration were also similar among individual HIV-1 isolates (Supplementary Figure S5).

Amino acid residue polymorphisms in HIV-1 integrase altered its nucleotide selection specificity in primary cells but not the genetic patterns of HIV-1 integration sites. Recent reports have identified the amino acid residues in the HIV-1 integrase that dictate the nucleotide selection specificity at integration sites, i.e. S119, T122, R231, and K258, in cell lines23,24. Since the full-length genomes of all well characterized primary HIV-1 isolates were sequenced, we could analyse these amino acid residues of our seven primary HIV-1 isolates and two clonal HIV-1 strains. Six HIV-1 isolates had the most common amino acid residues in all four positions. Primary HIV-1 isolates 2 and 6 exhibited a polymorphism at positions 122 (T122I) and 119 (S119P), respectively. These two polymorphisms were present concurrently in primary HIV-1 isolate 5 (Supplementary Figure S6). We investigated the effects of S119P and T122I polymorphisms on the nucleotide selection specificity at HIV-1 integration sites independently by analysing the 20 nucleotide positions immediately upstream of HIV-1 integration junctions (nucleotide position 0) in both MDMs and CD4+ T cells.

We analysed two nucleotide positions that were previously reported to be altered by the S119P HIV-1 integrase variant25. S119P resulted in a significantly increased preference for adenines at positions −8 (p < 0.0001) and −9 (p < 0.01) in both MDMs and CD4+ T cells (Fig. 2a), consistent with previous reports26. T122I significantly decreased the preference of HIV-1 for thymines at position −8 in MDMs (p < 0.0001) (Fig. 2b), similar to the previously reported T122K27. Nonetheless, a consistent decrease in preference of HIV-1 for thymines at positions −4 and −8 was observed in both MDMs and CD4+ T cells (Fig. 2b). Primary HIV-1 isolate 5, which exhibited both polymorphisms
Figure 1. Chromosomal distribution of HIV-1 integration sites in monocyte-derived macrophages and CD4⁺ T cells infected ex vivo. Inner red lines indicate the position at which HIV-1 was found and outer blue peaks indicate gene densities. In each chromosome, red bands indicate centromeres, darker bands are AT-rich, and lighter bands are GC-rich.

Figure 2. Conservation and frequency plots of 20 host nucleotides upstream of integrated HIV-1. (a) Left panel shows the nucleotide selection specificity of HIV-1 with a serine residue at position 119 of the HIV-1 integrase protein in a T122 background. Right panel shows the nucleotide selection specificity of HIV-1 with serine substituted by a proline residue at position 119 of the HIV-1 integrase protein in a T122 background. (b) Left panel shows the nucleotide selection specificity of HIV-1 with a threonine residue at position 122 of the HIV-1 integrase protein in an S119 background. Right panel shows the nucleotide selection specificity of HIV-1 with threonine substituted by an isoleucine residue at position 122 of the HIV-1 integrase protein in an S119 background. χ² test: *p < 0.01 and **p < 0.0001 compared to the same nucleotide positions in the left panel.
in its integrase, could not be analysed due to the lack of sufficient HIV-1 integration sites. Both polymorphisms (S119P and T122I) dictated the nucleotide selection specificity at HIV-1 integration sites in primary cells.

### The repertoire of HIV-1-hosting genes in MDMs was more restricted and distinct from that of CD4+ T cells.

Next, we assessed whether the annotated features of HIV-1-hosting genes, e.g. involvement in cellular and molecular processes, were different between the two cell types using the Database for Annotation, Visualization and Integrated Discovery (DAVID)\(^\text{25,26}\). The majority of HIV-1-hosting gene clusters over-represented in the human genome had annotations associated with structural motifs and domains of DNA and proteins in both cell types (Supplementary Figure S7). However, these gene clusters were more diverse in CD4+ T cells, which was also the case in most other categories of gene clusters (data not shown). MDMs had more over-represented gene clusters that were associated with protein metabolism, regulation, and transport whereas the proportion of over-represented gene clusters associated with a variety of other cellular processes, e.g. regulation of apoptosis, was higher in CD4+ T cells (Supplementary Figure S7). There were also gene clusters that were observed exclusively in either cell type, e.g. deubiquitination and regulation of axonogenesis in MDMs whereas T cell receptor signalling and cell cycle progression in CD4+ T cells (data not shown), which concordantly reflected the cellular and molecular functions of the corresponding cell types.

As the diversity of structural and functional HIV-1-hosting gene clusters appeared to be higher in CD4+ T cells, we investigated further the extent to which these gene clusters overlap between cell types in order to estimate the relative size of the repertoire of genes accessible to HIV-1 for integration. For this purpose, we analysed the top 40 gene clusters of each data set, including both over-represented and not over-represented gene clusters in the human genome. We scored gene clusters having the same or highly similar annotations between data sets, and observed that the gene cluster overlap was the highest (47.5%) between MDM/HIV-1\text{autologous} and MDM/HIV-1\text{heterologous}, and decreased by 7.5–15% between MDMs and CD4+ T cells. As expected, between CD4+ T cell/HIV-1\text{autologous} and CD4+ T cell/HIV-1\text{heterologous}, the extent of overlap was lower (32.5%) (Table 1). This apparent trend of gene cluster overlaps was reproducible when we included HIV-1 integration site data sets from two other studies: the one other publicly available MDM data set (Supplementary Table S2). To address this issue, we performed two analyses: (i) the proportion of HIV-1 integration sites forming the top 40 gene clusters were compared with the proportion of HIV-1 integration sites recovered from each HIV-1-infected individual’s sample and (ii) gene clusters formed with data from individuals 1 or 6 only, and when removed, were

| Data set 1                | Data set 2                | Top 40 DAVID gene cluster overlap (%) |
|--------------------------|--------------------------|--------------------------------------|
| MDM/HIV-1\text{a}        | MDM/HIV-1\text{h}        | 47.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 32.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 37.5                                 |
| MDM/HIV-1\text{a}        | MDM/HIV-1\text{h}        | 37.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 40.0                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 32.5                                 |
| MDMs\text{18}           | MDM/HIV-1\text{a}        | 37.5                                 |
| MDM/HIV-1\text{a}        | MDM/HIV-1\text{h}        | 37.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 27.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 22.5                                 |
| CD4+ T cells\text{27}    | MDM/HIV-1\text{a}        | 37.5                                 |
| MDM/HIV-1\text{a}        | MDM/HIV-1\text{h}        | 37.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 32.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 27.5                                 |
| CD4+ T cells\text{27}    | MDM/HIV-1\text{a}        | 22.5                                 |
| MDM/HIV-1\text{a}        | MDM/HIV-1\text{h}        | 20.0                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 17.5                                 |
| CD4+ T cell/HIV-1\text{a} | CD4+ T cell/HIV-1\text{h} | 26.0                                 |
| MDMs\text{18}           | CD4+ T cells\text{27}    | 25.0                                 |
| CD4+ T cells\text{27}    | CD4+ T cells\text{27}    | 25.0                                 |

Table 1. DAVID structural and functional gene clusters overlap between data sets. HIV-1- and HTLV-1-hosting genes were clustered based on their structural and functional annotations using DAVID. The top 40 clusters from all data sets were compared and clusters having the same or highly similar annotations were scored. MDM: Monocyte-derived macrophages; a: autologous; h: heterologous; HTLV-1: Human T-lymphotropic virus-1.
compared with the entire corresponding data sets. A high correlation ($R^2 = 0.9774$) was observed between the proportion of HIV-1 integration sites forming the top 40 gene clusters and that of HIV-1 integration sites recovered from each HIV-1-infected individual’s sample, indicating that each individual’s sample contributed towards the formation of the top 40 gene clusters, and the extent of their contributions were concordant with the proportion of HIV-1 integration sites recovered (Supplementary Figure S8). Additionally, data sets containing only HIV-1 integration sites derived from HIV-1-infected individuals 1 or 6 versus data sets without these HIV-1 integration sites overlapped with the entire corresponding data set to comparable levels (Supplementary Table S3), demonstrating our observations were neither HIV-1-infected individual-specific nor skewed by these two individuals who had the highest number of recovered HIV-1 integration sites.

Using the National Centre for Biotechnology Information (NCBI) database of published protein interactions between HIV-1 and its cellular host$^{29}$, we investigated the frequency of HIV-1 integration events in a subset of genes encoding HIV-1-interacting proteins. We hypothesized that this frequency would be different between MDMs and CD4$^+$ T cells given the lower HIV-1-hosting gene cluster overlap between them (Table 1). Here, we expanded our analysis and included a total of 11 HIV-1 integration site data sets: two MDM data sets ($n = 987$) and two CD4$^+$ T cell data sets ($n = 497$) derived from this study and one other MDM data set ($n = 754$) and six other CD4$^+$ T cell data sets ($n = 7,557$) obtained elsewhere$^{14,18,27,30–33}$ (Table 2). We observed an increase of HIV-1 integration events in genes encoding HIV-1-interacting proteins in MDMs ($1.2 \times$) and a significant increase in CD4$^+$ T cells ($1.4 \times$, $p < 0.001$) compared to three sets of 5,000 random genes generated in silico. Details of data sets derived from other studies (open symbols) are found in Table 2. HTLV-1 integration site data set served as an independent control. Two-tailed Mann-Whitney U test with 95% confidence interval was used to test for statistical significance.

### Table 2. Data sets examined in this study.

| Cell type         | Virus | Number of integration sites reported | Source |
|-------------------|-------|--------------------------------------|--------|
| MDMs HIV-1        |       | 987                                  | This study |
| CD4$^+$ T cells HIV-1 |   | 497                                  | This study |
| MDMs HIV-1        |       | 754                                  | 18     |
| CD4$^+$ T cells HIV-1 |   | 457                                  | 32     |
| CD4$^+$ T cells HIV-1 |   | 2661                                 | 33     |
| CD4$^+$ T cells HIV-1 |   | 2410                                 | 30     |
| PBMCs HIV-1       |       | 534                                  | 27     |
| SupT1 cells HIV-1 |       | 524                                  | 14     |
| Jurkat cells HIV-1|       | 971                                  | 31     |
| Jurkat cells and PBMCs HTLV-1 | | 579                                  | 28     |

### Figure 3. Frequency of HIV-1 integration events in genes encoding HIV-1-interacting proteins.

A total of three MDM data sets (two derived from this study; closed circles) and eight CD4$^+$ T cell data sets (two derived from this study; closed squares) were compared to three sets of 5,000 random genes generated in silico. Details of data sets derived from other studies (open symbols) are found in Table 2. HTLV-1 integration site data set served as an independent control. Two-tailed Mann-Whitney U test with 95% confidence interval was used to test for statistical significance.
Given the consistent trend of gene cluster overlaps, a rigorous assessment to show the robustness of our analysis, and a significantly different (p < 0.05) over-representation of HIV-1 integration events in genes encoding HIV-1-interacting proteins between MDMs and CD4+ T cells, we demonstrated that the repertoires of genes accessible to HIV-1 for integration in MDMs and CD4+ T cells were distinct and that of MDMs’ was more restricted.

Genes linked to clonal expansion of latently HIV-1-infected CD4+ T cells were targeted more frequently in CD4+ T cells but not monocyte-derived macrophages. Expansion of the viral reservoir by means of clonal expansion of latently HIV-1-infected CD4+ T cells harbouring HIV-1 proviruses in genes of certain functional nature has recently been suggested27,30,32. These genes are referred to as genes linked to clonal expansion of latently HIV-1-infected CD4+ T cells here.

We compiled genes linked to clonal expansion of latently HIV-1-infected CD4+ T cells identified in two studies27,30 and collapsed them into a total of 196 unique genes. Next, we divided each HIV-1 integration site data set into two groups; one containing unique protein-coding genes in which HIV-1 was found ≥ 2 times, i.e. gene hotspots (Fig. 4) and the other, only once (Fig. 5a). We included seven HIV-1 integration site data sets for this analysis: two derived from this study (data sets of the same cell type combined; n = 1,484) and one other MDM data set (n = 754) and four other CD4+ T cell data sets (n = 4,613) obtained elsewhere14,18,31–33 (Table 2).

By comparing the frequencies of HIV-1-hosting genes linked to clonal expansion of latently HIV-1-infected CD4+ T cells in gene hotspots and genes with single HIV-1 integration events, we observed a highly significant increase of genes linked to clonal expansion of latently HIV-1-infected CD4+ T cells in CD4+ T cells (2.5×, p < 0.001), but not in MDMs (Fig. 5b), showing again that the repertoires of genes accessible to HIV-1 for integration in the two cell types were distinct.

Gene hotspots identified for HIV-1 integration. We identified a total of 75 (9.5%) and 33 (7.7%) genes in which HIV-1 was found ≥ 2 times in MDMs and CD4+ T cells, respectively (Fig. 4). Furthermore, ~70% of these genes were identified in ≥ 2 HIV-1-infected individuals. As the probability of getting the same gene ≥ 2 times in multiple data sets was low34, these genes were considered as hotspots for HIV-1 integration here. Three gene hotspots were identified ≥ 4 times in both cell types, i.e. MECP2, KANSL1, and UBE2G1, and they were also listed in other studies (Table 3). In most cases, HIV-1 proviruses integrated into the same introns (Table 3), and they spanned 36 kb, 139 kb, and 22 kb in MECP2, KANSL1, and UBE2G1, respectively.

Discussion
We compared HIV-1 integration site patterns between macrophages and CD4+ T cells derived from seven ART-treated HIV-1-infected individuals whose cells were infected ex vivo with autologous, as well as heterologous, primary HIV-1 isolates generated during the acute phase of infection. This unique experimental set-up allowed us to replicate the in vivo situation for the examination of HIV-1 integration site patterns arising from the two major cellular targets of HIV-1 and potentially among different HIV-1 isolates.
The various genetic features at HIV-1 integration sites were similar among different HIV-1 isolates, regardless of whether the virus was autologous or heterologous. However, given that the gene clusters consistently overlapped to a higher extent in MDMs, the repertoire of genes accessible to HIV-1 for integration in MDMs appeared to be more restricted than and distinct from that of CD4+ T cells. This distinction between the two cell types was again reflected in the different frequencies of HIV-1-hosting genes in two gene subsets: genes encoding HIV-1-interacting proteins and those linked to clonal expansion of latently HIV-1-infected CD4+ T cells. Firstly, we observed an over-representation of HIV-1 integration events in the former gene subset in both MDMs and CD4+ T cells, but to different extents. Despite the heterogeneity of the NCBI HIV-1 Human Interaction Database29, its representativeness and specificity were demonstrated by the substantially lower frequency of HIV-1-interacting genes in random and HTLV-1 data sets. Secondly, HIV-1 integration into genes linked to clonal expansion of latently HIV-1-infected CD4+ T cells was over-represented in gene hotspots in CD4+ T cells but not in MDMs. The gene cluster distribution of our CD4+ T cell data set was consistent with that derived from the study of Wagner et al.27, whereas such consistency was somewhat lacking between our MDM data sets with that derived from the study of Barr et al.18. Comparison of HIV-1 integration sites between MDMs and CD4+ T cells derived from the same experimental settings is important, as evident from the absolute percentages of the top 40 DAVID gene cluster overlaps; MDMs from the study of Barr et al.18 and CD4+ T cells from the study of Wagner et al.27 had the same absolute values when compared to MDMs from this study. Nonetheless, the trend of gene cluster
overlaps was consistent across all data sets that were derived from the same experimental settings, i.e. gene cluster overlaps were the highest between MDM data sets, followed by between MDM and CD4$^+$ T cell data sets, and the lowest was observed between CD4$^+$ T cell data sets.

We focussed on gene hotspots that were identified at least four times in both cell types, i.e. MECP2, KANSL1, and UBE2G1. The products of these genes are involved in transcriptional and post-translational regulations of gene$^{35-37}$. Interestingly, downregulation of MECP2 has been reported to enhance replication of HIV-1, in Jurkat cells$^{38}$. Although it was unclear if these gene hotspots, as well as genes encoding HIV-1-interacting proteins were highly accessible prior to or upon HIV-1 infection, it is conceivable that they were transcriptionally active enough to allow integration$^{4,15,18,33}$ and/or located in the vicinity of the nuclear pore, close to the incoming HIV-1 DNA$^{34}$. The model of Marini et al.$^{34}$ describes that specific chromatin organization in the nucleus results in certain genes being closer to the nuclear pore, thus recurrently targeted by HIV-1 for integration. This model could be envisaged to be applicable down to the levels of genes, in which certain intragenic regions might be closer and more exposed to the incoming HIV-1 DNA. This would explain why HIV-1 proviruses that were found in the three gene hotspots preferred integration into the same introns in most cases.

The effect of T122I polymorphism on HIV-1 nucleotide selection specificity has not been reported previously. This natural and the second most abundant amino acid polymorphism at position 122 consistently altered the nucleotide selection specificity of HIV-1, demonstrating its role in integration site targeting. The effect of S119P polymorphism on HIV-1 nucleotide selection at the site of integration in cell lines has been described by Demeulemeester et al.$^{23}$, and our observation in primary cells was consistent with theirs. Certain amino acid polymorphisms of the HIV-1 integrase, especially at position 119, have been associated with rapid disease progression$^{32}$. However, association of natural integrase variants with disease progression could not be examined in our study participants because all of them have been successfully treated. Although these HIV-1 integrase polymorphisms altered the nucleotide selection specificity at integration sites, they did not have an effect on various genetic features at HIV-1 integration sites$^{23}$. It is likely that the retargeting of HIV-1 was restricted by the repertoire of accessible genes in MDMs and CD4$^+$ T cells.

In summary, we demonstrated in various analyses using our own data sets and of others that the repertoires of HIV-1-hosting genes targeted by HIV-1 in MDMs and CD4$^+$ T cells were distinct in terms of their structural and functional classes and relative sizes. These cell type-dependent HIV-1 integration site patterns might account for the differences in the life cycle of HIV-1 in these cell populations.

Methods

HIV-1-infected study participants and HIV-1 negative donors. HIV-1-infected individuals are enrolled in the Zurich Primary HIV Infection (ZPHI) study, which is a non-randomized, observational, open-label, and single-centre study (www.clinicaltrials.gov; ID: NCT00537966; registered 25th September 2007$^{20}$). For the current sub-study, the inclusion criteria were: (i) HIV-1-infected individuals had at least 1.5 years of viral suppression of <20 HIV-1 RNA copies/mL plasma with antiretroviral therapy (ART) at the time of cell sample collection, and (ii) their primary HIV-1 isolates were capable of replication in monocyte-derived macrophages (MDMs).

Ethics, consent, and permission. This study was approved by the Ethics Committee of the University Hospital Zurich and written informed consent was obtained from all HIV-1-infected individuals and HIV-1 negative donors. All experiments were performed in accordance with the relevant guidelines and regulations.

Generation and sequencing of primary HIV-1 isolates and clonal HIV-1 strains. The generation of primary HIV-1 isolates has been described elsewhere$^{39}$. Briefly, CD4$^+$ T cells from HIV-1-infected individuals in the acute phase of infection were co-cultured with activated donors’ peripheral blood mononuclear cells (PBMCs) until the HIV-1 p24 levels of the culture supernatant was >2,000 pg/mL prior to harvesting of the virus supernatant. HIV-1 p24 levels were measured with an enzyme-linked immunosorbent assay (ELISA$^{40}$). The 50% tissue culture infectious dose (TCID$_{50}$) of the primary HIV-1 isolates was determined by co-culturing quadruplicates of 5-fold dilutions of primary HIV-1 isolates and scoring the number of p24-positive wells, and the TCID$_{50}$ was subsequently calculated using the method of Reed and Muench$^{41}$.

HIV-1JR.Fl$^{42}$ and HIV-1NL4-3$^{43}$ were generated by transfecting HEK 293T cells with pJR-FL and pNL4-3, respectively, and the viral supernatant was harvested and cleared of cells with a 0.22 μm filter 48 hours post-transfection. pJR-FL$^{42}$ was obtained from Irvin SY Chen and Yoshio Koyanagi, and pNL4-3$^{42}$ was obtained from Malcolm Martin through the NIH AIDS Reagent Programme, Division of AIDS, NIAID, NIH.

We had previously developed a sensitive and robust method to amplify and sequence the full-length genomes of primary HIV-1 isolates with minimal artefactual in vitro recombination$^{44}$. The seven selected primary HIV-1 isolates were all confirmed to be of subtype B as determined by their pol nucleotide sequences using the REGA HIV-1 subtyping tool V3$^{45}$.

Ex vivo HIV-1 infection of monocyte-derived macrophages and CD4$^+$ T cells. Blood samples of ART-treated HIV-1-infected individuals were first depleted of CD8$^+$ T cells using RosetteSep Human CD8 Depletion Cocktail (Stemcell Technologies). CD8$^-$ PBMCs were isolated by density gradient centrifugation using Lymphoprep (Stemcell Technologies). CD14$^+$ monocytes were positively isolated by magnetic-activated cell sorting using CD14 microbeads (Miltenyi Biotec) whereas the remaining CD8$^-$ CD14$^+$ fraction was treated functionally as CD4$^+$ T cells. The median purities of CD14$^+$ monocytes and CD4$^+$ T cells were 95.4% and 98.8%, respectively, as measured by flow cytometry. CD4$^+$ cells were induced to differentiate into MDMs in RPMI-1640 (10% (v/v) fetal bovine serum (FBS), 100 U/mL Penicillin, and 100 μg/mL Streptomycin) supplemented with 5% (v/v) human serum (Sigma) and 20 ng/mL macrophage-colony stimulating factor (M-CSF) (Peprotech) for eight days. CD4$^+$ T cells were activated in RPMI-1640 (10% (v/v) FBS, 100 U/mL Penicillin, and 100 μg/mL...
Amplification and sequencing of 5′ HIV-1 integration junctions. The 5′ HIV-1 integration junctions were amplified with the non-restrictive linear amplification-mediated PCR (nLAM-PCR)3 from a total of 0.2–1.0 μg of genomic DNA extracted from cells infected e viv o with HIV-1 isolates. The procedure was essentially the same as described elsewhere21, except: (i) the 3′ primers used for linear, first exponential, and second exponential amplifications were 5′CCCTGGTGTGTAGTTCTGCC3′, 5′CAATCAGGGGAAGTAGGCTTTGG3′, and 5′TCGTGGCAAGTCAGATGTGTTAAATAGACAGN0–3, GGTGTTAGACCCACAGATCAAG3′, respectively, (ii) the 5′ primer used for second exponential amplification was 5′GTCTCTGCTGGGCTCAGGATG TGTTAAAGACCAGN0–3, GATCTGAATTCAGTGGCACAG3′, (iii) the cycling conditions for linear amplification were 2× (95°C for 2 min followed by 50 cycles of 95°C for 45 s, 57°C for 45 s, and 72°C for 8 s) and (iv) The cycling conditions for first and second exponential amplifications were 95°C for 2 min followed by 40 cycles of 95°C for 45 s, 57°C for 45 s, and 72°C for 8 s, and a final extension at 72°C for 2 min. The 5′ HIV-1 LTR-specific primers were designed based on the most conserved regions of the consensus sequence of HIV-1 subtype B in the Los Alamos HIV Sequence Database. The N0–3 random nucleotides in the primers used for the second exponential amplification were introduced to increase nucleotide diversity to enable DNA cluster generation during DNA sequencing with the Illumina MiSeq platform.

5′ HIV-1 integration junctions were sequenced with the Illumina MiSeq platform. Illumina sequencing adapters were added to 1 ng of purified PCR products in a PCR using the following primers: 5′AATGATACGGCGAC CACCGAGATCTACAC(Index)TGGTCGGCAGCGTC3′ and 5′CAACGAGAAGACGGCATACGAGAT(Index) GTCTCTGCTGGGCTCAGGATG TGTTAAAGACCAGN0–3, GATCTGAATTCAGTGGCACAG3′. The cycling conditions were 72°C for 3 min, 95°C for 30 s, 12 cycles of 95°C for 10 s, 55°C for 30 s, and 72°C for 30 s, and a final extension at 72°C for 5 min. A total of 12–14 pM of purified PCR products were sequenced with the MiSeq Reagent Kit v2 (300 or 500 cycles) with 8% PhiX.

Post-sequencing processing and mapping to the human genome assembly. Sequencing reads were filtered to remove low quality reads. True 5′ HIV-1 integration junctions were identified using an in-house bioinformatic pipeline, i.e. Integration Site Analysis Pipeline (InStAP), and satisfied the following criteria: (i) presence of both the nLAM-PCR adaptor and 5′ HIV-1 LTR, (ii) host DNA was within 3 nucleotides from the end of 5′ HIV-1 LTR, and (iii) ≥85% of the lengths of the reads mapped to the human assembly GRCh37. p13 with ≥98% identity. Contaminants were accounted for and removed from all samples in which they were identified except the sample that was sequenced first and/or with the highest number of read count. Manuscript on the details of InStAP is in preparation.

HIV-1 integration site data sets derived from this study and obtained elsewhere for comparison are shown in Table 2. All HIV-1 integration sites derived from this study are listed in Supplementary Table S4.

Bioinformatic databases. Annotated features of HIV-1-hosting genes were assessed using the Database for Annotation, Visualization and Integrated Discovery (DAVID)25,26. DAVID groups genes, which are related by their structural and functional annotations, into clusters and calculates an enrichment score to indicate if the gene clusters were significantly over-represented in a particular genomic background. Clusters of HIV-1-hosting genes that were significantly over-represented in the human genome are defined as having a DAVID enrichment score >1.3 with the stringency level in DAVID set to high. By definition, a gene cluster having a DAVID enrichment score >1.3 was equivalent to an overall geometric mean <0.05 for all its gene members, and a high stringency would consider a gene cluster as biologically significant only when the kappa coefficient, κ ≥ 0.85 for at least three genes with at least three related annotations between any two.

The National Centre for Biotechnology Information (NCBI) catalogues and provides access to a database of published interactions between HIV-1 proteins and its cellular host34. We used the November 2014 release (the latest at time of analysis) of the HIV-1 Human Interaction Database, which documented a total of 3,337 unique RefSeq protein-coding genes whose products interact physically, as well as indirectly, with HIV-1. This number of genes represented ~16.15% of all unique RefSeq protein-coding genes in the human genome (20,664 as of 15th of December 2014).

Statistical analyses. Two-tailed Fisher's exact test, χ² test, and Mann-Whitney U test with a 95% confidence interval were used at where indicated.

Graphical plots. Some of the graphical plots were generated using the following online platforms: Plotly (https://plot.ly/), WebLogo46,47, and Circos48.

References
1. Dalgleish, A. G. et al. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 312, 763–767 (1984).
2. Klatzmann, D. et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312, 767–768 (1984).
3. Koenig, S. et al. Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. Science (New York, N.Y.) 233, 1089–1093 (1986).
42. Koyanagi, Y.
40. Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4.
37. Watanabe, T. K.
38. Chiang, K., Liu, H. & Rice, A. P. miR-132 enhances HIV-1 replication.
29. Ako-Adjei, D.
32. Ikeda, T., Shibata, J., Yoshimura, K., Koito, A. & Matsushita, S. Recurrent HIV-1 integration at the BACH2 locus in resting CD4
31. Lewinski, M. K.
34. Marini, B.
25. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional
16. Stevens, S. W. & Griffith, J. D. Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs
10. Brinkmann, R.
18. Barr, S. D.
14. Schroder, A. R.
13. Carteau, S., Hoffmann, C. & Bushman, F. Chromosome structure and human immunodeficiency virus Type I infection.
23. Demeulemeester, J.
25. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional
12. Koppensteiner, H., Brack-Werner, R. & Schindler, M. Macrophages and their relevance in Human Immunodeficiency Virus Type I research.
17. Stevens, S. W. & Griffith, J. D. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration.
10.1093/nar/gku1126 (2015).
8. Pelchen-Matthews, A., Kramer, B. & Marsh, M. Infectious HIV-1 assembles in late endosomes in primary macrophages.
7. Raposo, G. et al. Human macrophages accumulate HIV-1 particles in MHC II compartments. Traffic, 3, 718–729 (2002).
6. Deng, H. et al. Identification of a major co-receptor for primary isolates of HIV-1. Nature 381, 661–666, doi: 10.1038/381661a0 (1996).
5. Maddon, P. J. et al. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell 47, 333–348 (1986).
4. Alkhatib, G. et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science (New York, N.Y.) 272, 1955–1958 (1996).
3. Aoki-Adjei, D. et al. HIV-1 integration site distributions in resting and activated CD4+ T cells from human immunodeficiency virus type 1-infected individuals carry integrated HIV-1 genomes with higher transcription in cerebrospinal fluid specimens compared with peripheral blood mononuclear cells. J. Virol. 78, 2465–2474, doi: 10.1128/jvi.00279-04 (2004).
2. Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4.
1. Koyanagi, Y. et al. Development of a functional assay for CD4 modulation of HIV-1 infection. Nature 424, 596–598, doi: 10.1038/nature01770 (2003).
45. Pineda-Pena, A. C. et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect Genet Evol* **19**, 337–348, doi: 10.1016/j.meegid.2013.04.032 (2013).

46. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic acids research* **18**, 6097–6100 (1990).

47. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome research* **14**, 1188–1190, doi: 10.1101/gr.849004 (2004).

48. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639–1645, doi: 10.1101/gr.092759.109 (2009).

**Acknowledgements**

We are grateful to individuals enrolled in the ZPHI study for their commitment and biological samples. We thank Dominique Braun and Christina Grube (USZ, Zurich, Switzerland) for patient care, Fabienne-Desirée Geissberger (UZH, Zurich, Switzerland) for technical assistance, Frederic Bushman and Nirav Malani (UPenn, Pennsylvania, USA), Anne Arens (Qiagen, Schwetzingen, Germany), and Kim D. Pruitt (NCBI, NLM, NIH, Maryland, USA) for their assistance in InSPiRD, HISAP, and NCBI databases, respectively. The following reagents were obtained through the NIH AIDS Reagent Programme, Division of AIDS, NIAID, NIH: pJR-FL from Irvin SY Chen and Yoshio Koyanagi and pNL4-3 from Malcolm Martin. This study has been financed by SNF grant 310030_141067 (to K.J.M. and H.F.G.), University of Zurich’s clinical research priority programme: viral infectious diseases – ZPHI (to H.F.G.), and an unrestricted research grant from the Vontobel Stiftung (to K.J.M. and H.F.G.).

**Author Contributions**

Y.L.K. and V.V. designed the study, performed the experiments, interpreted the data, and wrote the manuscript. M.S., V.V. and R.K. developed InStAP. FDG provided the HIV-1 integrase sequences of primary HIV-1 isolates. H.K. performed the experiment on the reactivation of *in vivo* HIV-1 from ART-treated HIV-1-infected individuals’ CD4+ T cells. H.F.G. participated in the design of the study. K.J.M. designed the study and edited the manuscript.

**Additional Information**

**Supplementary information** accompanies this paper at http://www.nature.com/srep

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** 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; doi: 10.1038/srep24157 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/