ARTICLE

doi:10.1016/j.ymthe.2006.03.012

HIV Integration Site Selection: Targeting in Macrophages and the Effects of Different Routes of Viral Entry

Stephen D. Barr,[1,2] Angela Ciuffi,[1] Jeremy Leipzig,[1] Paul Shinn,[3] Joseph R. Ecker,[3] and Frederic D. Bushman[1,*]

[1]*Department of Microbiology, University of Pennsylvania School of Medicine, 3610 Hamilton Walk, Philadelphia, PA 19104-6076, USA*
[2]*Department of Medical Microbiology and Immunology, University of Alberta, 632 Heritage Medical Research Center, Edmonton, AB, Canada T6G 2S2*
[3]*Genomic Analysis Laboratory, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA*

**To whom correspondence and reprint requests should be addressed. E-mail: bushman@mail.med.upenn.edu.*

Available online 2 May 2006

We have studied the selection of HIV DNA integration sites in primary macrophages to investigate two questions. First, mature macrophages do not divide, allowing us to investigate whether HIV integration targeting differs between dividing cells and nondividing cells. We sequenced and analyzed 754 unique integration sites and found that integration in macrophages is favored in active transcription units (TUs), as was observed previously for other cell types. However, HIV integration in genes was slightly less favored in macrophages than in dividing PBMC or T cell lines. Second, we compared integration targeting by HIV-vector particles bearing either of two different envelope proteins (HIV R5 Env or VSV-G) to determine whether the mechanism of entry influenced subsequent integration targeting. Integration sites generated by HIV R5- or VSV-G-bearing particles showed no significant differences in their distributions in the human genome. Analysis of additional published integration site sequences also indicated that the route of entry did not affect integration site selection for other viral envelopes as well.

Key Words: HIV, lentivirus, vector, gene therapy, macrophages, integrase, retrovirus, envelope, VSV-G, integration

INTRODUCTION

The integration of HIV DNA in the chromosomes of macrophages is important in HIV-induced disease and in lentivirus-based gene therapy of this cell type (reviewed in [1–4]). In this paper, we (i) compare integration in macrophages to integration in other cell types and (ii) assess the role of the route of entry into macrophages in determining integration target site selection.

Previous studies have characterized HIV integration in several human cell types, revealing that HIV strongly favors integration in active transcription units (TUs) [5–10]. Studies of simple retroviruses, in contrast, revealed different patterns of integration site selection. Murine leukemia virus (MLV) favors integration near transcription start sites and shows only a slight preference for active TUs [6,11]. Avian sarcoma–leukosis virus (ASLV), in contrast, shows the most random distribution, favoring integration in active TUs only slightly and showing no preference for integration near transcription start sites [7,9,12].

The mechanisms guiding integration targeting are beginning to be clarified. Retroviral integration is not strongly sequence-specific with respect to the target DNA at the point of joining [13–16], but integration *in vivo* shows pronounced favored and disfavored chromosomal regions. Early studies suggested that relative exposure of integration target DNA might affect integration site selection [17–19]. Another model for target site selection invokes direct contacts between retroviral integration complexes and cellular proteins bound specifically at favored target sites [20–24]. In support of this idea, the cellular lens epithelium-derived growth factor (PSIP1/LEDGF/p75) protein, which binds tightly to HIV integrase [25–30], has been found to influence the placement of HIV integration sites *in vivo* [8]. However, knockdown of PSIP1/LEDGF/p75 did not fully eliminate favored integration in TUs, leaving open the possibility that additional factors modulate HIV integration.

Another possible factor in integration targeting is the point during the cell cycle at which integration takes

1525-0016/$30.00

place [24]. HIV can carry out integration regardless of the cell cycle stage, but for MLV, cells must pass through mitosis to permit integration [31,32]. One approach to investigating this issue is to analyze integration site selection during different phases of the cell cycle, taking advantage of the ability of HIV to infect cells regardless of the cell cycle stage. According to one possible model, chromatin remodeling during DNA replication promotes integration into TUs, so that integration in nondividing cells might not be favored in TUs. Recently, integration targeting was analyzed in IMR-90 lung fibroblasts that had been arrested by serum starvation and contact inhibition. In this model, it was found that HIV integration was still favored in active TUs [10]. The generality of this finding has been unknown, however, adding interest to a study of macrophages, which are a clinically relevant nondividing cell type.

Here we report an analysis of 754 sites of HIV integration in purified macrophages and a bioinformatic comparison to integration sites in T cells. We find that integration is indeed favored in TUs in macrophages, and comparison to transcriptional profiling data indicates that active TUs were favored. Comparison to integration site data sets from peripheral blood mononuclear cells (PBMC) and T cell lines revealed that the favoring of integration in TUs was slightly more pronounced in lymphoid cells, as was integration in features associated with gene-dense chromosomal regions.

We have also used the macrophage system to investigate whether the route of viral entry into cells influences the subsequent selection of sites for integration. Normally, HIV infection begins with fusion of the viral and cytoplasmic membranes mediated by HIV surface glycoproteins, resulting in introduction of the viral core into the cytoplasm. The pathway differs when infection is carried out by pseudotyped particles bearing the vesicular stomatitis virus glycoprotein (VSV-G) envelope. These particles are taken up by endocytosis, and fusion between the viral membrane and the vesicle membrane is induced by low pH. The route of the preintegration complex (PIC) to the nucleus is not fully clarified for either case and could differ. Of our integration sites from macrophages, 452 were generated by infection using particles bearing the VSV-G envelope and 302 were generated from particles bearing the CCR5-tropic (R5) HIV BaL envelope [33]. A detailed statistical comparison showed no significant differences between the two data sets. Comparison of published integration site data sets in which HIV or ASLV infection was mediated by still further envelope proteins also showed no strong effects on integration targeting, indicating that the envelope used for infection, and the route of entry, do not strongly affect integration site selection.

RESULTS

Bioinformatic Analysis of the Distribution of Integration Sites

We prepared separately HIV vector particles containing the VSV-G or HIV R5 envelope proteins. We used the particles to infect purified human macrophages. Six days after infection, we purified genomic DNA and cloned junctions between human and HIV-vector DNA using ligation-mediated PCR as described previously [5,6,9]. We processed raw DNA sequence reads to remove linker and viral sequences, condensed duplicates, and then determined the locations of unique integration sites on the human genome (draft hg17) using BLAT. We compared integration site sequences from macrophages to previously published integration site data sets (Table 1).

We also compared experimentally determined integration sites to matched random control sites generated *in silico*. The matching procedure used to generate the control sites was designed to account for possible bias due to the use of restriction enzyme digestion in cloning integration sites. That is, if the restriction enzyme recognition sites are not uniformly distributed in the human genome, then the integration sites recovered might not be representative of the full population of sites. To correct for this, we matched each experimentally determined site with 10 random sites *in silico* that were constrained to be the same number of bases from a restriction site as was the experimental site. For statistical analysis, we then compared the pooled experimental sites to the pooled matched random controls.

TABLE 1: Integration site data sets used in this study

Virus or vector	Envelope	Cell type/ integration target	No. of integration sites	Source
HIV vector	HIV (CCR5-tropic)	Macrophage	302	This work
HIV vector	VSV-G	Macrophage	452	This work
HIV-Tat vector	VSV-G	Jurkat	914	[36]
HIV vector	VSV-G	PBMC	542	[7]
HIV vector	VSV-G	SupT1	542	[5]
HIV	HIV (CXCR4-tropic)	SupT1	45	[14]
ASLV	ASLV subgroup A	293T-Tva	640	[7]
ASLV	VSV-G	HeLa	194	[12]

doi:10.1016/j.ymthe.2006.03.012

TABLE 2: Integration frequency in transcription units: comparison of HIV/VSV-G and HIV/R5

Transcription unit catalog	HIV/VSV-G (*P* value vs random)	HIV/R5 (*P* value vs random)	Matched random control	*P* value (HIV/VSV-G vs HIV/R5)
Acembly	76.3% (<0.0001)	77.5% (<0.0001)	52.8%	0.7125
GenScan	77.9% (0.0005)	76.5% (<0.0193)	70.2%	0.6561
RefSeq	61.1% (<0.0001)	62.9% (<0.0001)	35.4%	0.6080
UniGene	62.2% (<0.0001)	62.9% (<0.0001)	40.9%	0.8358
Known	65.7% (<0.0001)	66.9% (<0.0001)	39.5%	0.7373
Ensembl	67.0% (<0.0001)	65.9% (<0.0001)	39.1%	0.7448

Comparison of Integration in Macrophages by Particles Bearing the HIV R5 Or VSV-G Envelopes

Integration targeting in TUs by the HIV-vector particles bearing the VSV-G or HIV R5 envelope are compared in Table 2. A complication arises because of the incomplete annotation available for human TUs. For this reason, we assessed integration frequency in six different sets of gene annotation, which yielded generally similar conclusions, though the exact values differed. For example, using the well-characterized RefSeq gene set, 61.1% of the HIV/VSV-G sites were in TUs and 62.9% of the HIV/R5 sites were in TUs. Only 35.4% of the matched random controls were in TUs, a highly significant difference. Analysis of other sets of gene annotation also showed a much higher frequency of macrophage integration sites in TUs compared to the random controls. Thus comparisons of data sets for the two envelopes showed, regardless of the gene catalog used, that TUs were favored, but there was no significant difference between the HIV/R5 and the HIV/VSV-G data sets (Table 2).

We also compared integration frequency in repeated sequences for the two data sets (Table 3). In the human genome, gene-sparse regions are enriched in long interspersed nuclear elements (LINEs). Intergenic regions are enriched in LTR sequences derived from retroviral infections [9,34,35]. HIV integration sites from macrophages were less frequent in LTR elements, consistent with the enrichment of LTR sequences outside TUs. Integration in LINEs also was significantly disfavored. We detected no significant differences between the HIV/VSV-G and the HIV/R5 data sets except in the small miscellaneous group of "other" repeats, which is of uncertain significance. We conclude that there were no major differences in targeting to repeated sequences between the two data sets.

We compared the integration sites from the HIV/VSV-G and HIV/R5 data sets further in an automated fashion with respect to a variety of genomic features, including chromosome number, chromosomal banding pattern, gene boundaries, CpG islands, and gene activity (supplementary material). None of these comparisons showed a consistent significant difference between the two data sets. We thus conclude that the two routes of infection do not result in detectable differences in integration site selection in this model.

Comparison of HIV Integration Site Selection in Macrophages and T Cells

We then compared the distribution of integration sites made by HIV-vector infection of macrophages to the distribution in activated PBMC and T cell lines. Because the distributions of sites in macrophages were so similar in the HIV/VSV-G and HIV/R5 data sets, we pooled the two for subsequent analysis. Similarly, we pooled T cell data sets and compared the placement of sites from the two data sets relative to a variety of genomic features.

Table 4 compares integration frequency in TUs across the six types of gene catalogs. The percentage of integration sites in TUs was higher in T cells in each. In RefSeq genes, 72.9% of integration events from T cells were in TUs, compared to 61.8% in sites from macrophages, a statistically significant difference. The trend was significant for the other gene catalogs with the

TABLE 3: Integration frequency in repeated sequences: comparison of HIV/VSV-G and HIV/R5

Chromosomal feature	HIV/VSV-G (*P* value vs random)	HIV/R5 (*P* value vs random)	Matched random control	*P* value (HIV/VSV-G vs HIV/R5)
SINEs				
Alu	9.8% (0.9366)	11.9% (0.1923)	9.8%	0.3456
MIR	2.8% (0.8228)	3.6% (0.5436)	3.0%	0.5390
DNA elements	4.1% (0.1023)	3.6% (0.4101)	2.8%	0.7213
LTR elements	4.5% (0.0013)	2.0% (<0.0001)	8.8%	0.0576
LINEs	15.7% (0.0002)	17.2% (0.0159)	23.2%	0.5689
Satellites, α	0.0% (0.5872)	0.0% (0.6590)	0.1%	n/d
Satellites, other	0.0% (0.3634)	0.0% (0.4600)	0.2%	n/d
Other	2.2% (0.4656)	0.3% (0.0636)	1.7%	0.0361

n/d, not determined.

TABLE 4: Integration frequency in transcription units: comparison of macrophage and T cell data sets

Transcription unit catalog	HIV/macrophages (*P* value vs random)	HIV/T cells (*P* value vs random)	*P* value (HIV/macrophages vs HIV/T cells)
Acembly	76.8% (<0.0001)	86.8% (<0.0001)	<0.0001
GenScan	77.3% (<0.0001)	78.2% (<0.0001)	0.6311
RefSeq	61.8% (<0.0001)	72.9% (<0.0001)	<0.0001
UniGene	62.5% (<0.0001)	68.7% (<0.0001)	0.0018
Known	66.2% (<0.0001)	78.1% (<0.0001)	<0.0001
Ensembl	66.6% (<0.0001)	78.6% (<0.0001)	<0.0001

exception of the GenScan predictions, which are based solely on computational and not empirical data. Thus the frequency of integration in TUs appears to be modestly but significantly higher in T cells than in macrophages.

Another means of quantifying integration in TUs involves comparison to measures of gene density. Fig. 1 shows a comparison of integration frequency as a function of gene density for the macrophage and T cell data sets. As can be seen, integration sites in T cells are more often found in gene-rich regions than are the sites in macrophages (though both data sets did show enrichment in gene-dense regions relative to the matched random control).

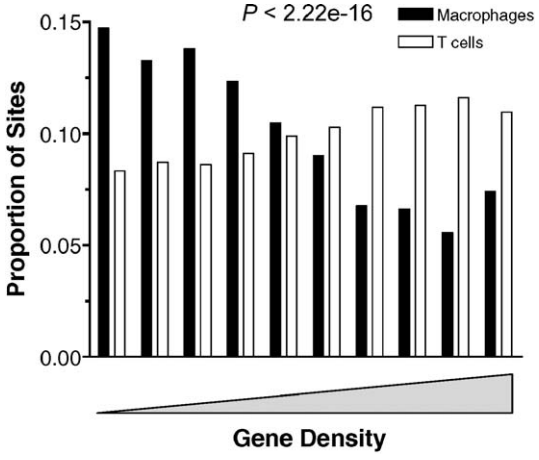

FIG. 1. Effects of gene density on integration in macrophages and T cells. The human genome was partitioned into segments of 8 Mb, then each segment was scored for the frequency of integration and relative gene density. Chromosomal segments hosting integration events in the macrophage and T cell data sets were pooled and then divided into 10 equal portions based on the local gene density. The contributions of sites from the macrophage and T cell data sets were then determined for each pool. The *P* value was obtained from the logistic regression of event type (macrophage versus T cell integration site) on a cubic B-spline basis (i.e., a third-order polynomial) for gene density. The cut values (gene calls per base pair; see the supplementary material) were group 1, 4.166667×10^{-8} to 1.500000×10^{-6}; group 2, 1.500000×10^{-6} to 2.018155×10^{-6}; group 3, 2.018155×10^{-6} to 2.498571×10^{-6}; group 4, 2.498571×10^{-6} to 3.066310×10^{-6}; group 5, 3.066310×10^{-6} to 3.946577×10^{-6}; group 6, 3.946577×10^{-6} to 5.351667×10^{-6}; group 7, 5.351667×10^{-6} to 7.150258×10^{-6}; group 8, 7.150258×10^{-6} to 1.005357×10^{-5}; group 9, 1.005357×10^{-5} to 1.495333×10^{-5}; group 10, 1.495333×10^{-5} to 2.956875×10^{-5}.

We then compared the macrophage and T cell data sets for the relative frequency of integration in repeated sequences. As with the macrophage integration sites, the T cell sites were less frequent in LINEs and LTR elements than were the matched random controls. Also, we found the T cell integration sites significantly more frequently in *Alu* repeats, which are enriched in gene-rich regions, consistent with the slightly stronger favoring of integration in TUs in the T cell data set (data not shown).

Then we compared the effects of gene activity on integration frequency for the macrophage and T cell data sets. Previous studies have established that gene activity positively correlates with integration in T cells [5–8,36]. We obtained transcriptional profiling data for gene activity in macrophages from two published studies [37,38]. The mean expression level of genes hosting integration events in macrophages was significantly higher than the mean for genes targeted by the matched random control, and this was true for both transcriptional profiling data sets analyzed ($P = 0.0002$, Mann–Whitney test).

Next, we compared integration frequency in the macrophage and T cell data sets relative to "transcriptional intensity," a score that combines the effects of gene density and transcriptional activity. As can be seen from Fig. 2, there was a consistent trend for integration in T cells to be more frequent in regions of relatively higher transcriptional intensity, though both the macrophage and the T cell data sets showed favoring of integration in regions of high transcriptional intensity compared to the matched random control.

A related means of examining the distribution of integration sites involves assessing integration frequency in intergenic regions as a function of their length (Fig. 3). Long intergenic regions define gene-sparse regions, and integration in these regions has been shown to correlate with lower proviral transcription [36]. We used the same statistical approach as above to compare integration frequency in the macrophage and T cell data sets. The sites from the macrophage data set were significantly enriched in the longer intergenic regions, consistent with their placement in regions of lower gene density and lower expression intensity. However, integration sites in long intergenic regions are less common in both the macrophage and the T cell data sets compared to the

doi:10.1016/j.ymthe.2006.03.012

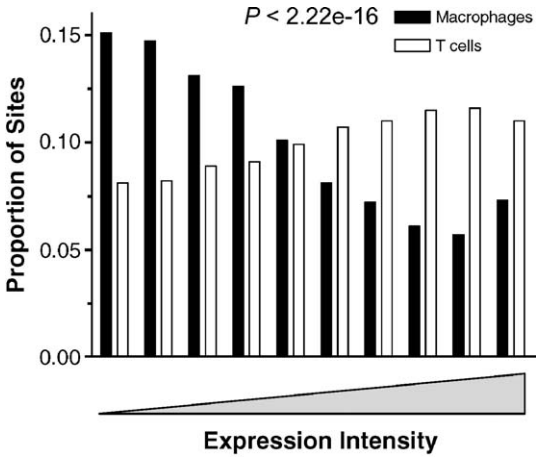

FIG. 2. Effects of expression intensity on integration in macrophages and T-cells. The values (integration frequency as a function of transcriptional intensity in each 8-Mb interval) were pooled for the macrophage and T cell data sets, then the pooled data were divided into 10 subsets based on their relative expression intensity. The contribution of macrophage or T cell integration sites to each interval was then quantified and statistical analysis carried out over the distribution. The *P* value was obtained from the logistic regression of event type (macrophage versus T cell integration site) on a cubic B-spline basis (i.e., a third-order polynomial) for expression intensity. The cut values (positive expression calls per base pair; see supplementary material) were group 1, 0 to 5.568452×10^{-7}; group 2, 5.568452×10^{-7} to 8.010417×10^{-7}; group 3, 8.010417×10^{-7} to 1.021131×10^{-6}; group 4, 1.021131×10^{-6} to 1.348958×10^{-6}; group 5, 1.348958×10^{-6} to 1.829018×10^{-6}; group 6, 31.829018×10^{-6} to 2.631667×10^{-6}; group 7, 2.631667×10^{-6} to 3.683810×10^{-6}; group 8, 3.683810×10^{-6} to 5.134045×10^{-6}; group 9, 5.134045×10^{-6} to 7.011769×10^{-6}; group 10, 7.011769×10^{-6} to 1.286742×10^{-5}.

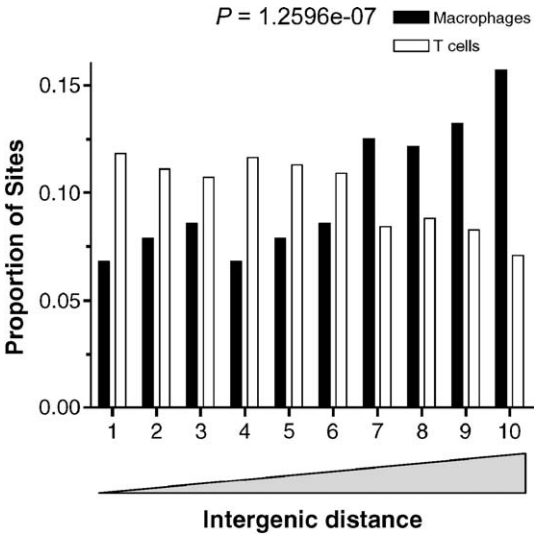

FIG. 3. Comparison of integration frequency in macrophages and T cells in intergenic regions. Integration sites in intergenic regions from macrophages and T cells were distributed into bins based on the length of the intergenic region hosting the integration event. Integration sites from the two data sets were pooled and then divided into 10 categories based on the relative length of the intergenic region hosting the integration event. The contribution of each data set to the collection of sites in each interval was then assessed and analyzed statistically. The *P* value was obtained from the logistic regression of event type (macrophage versus T cell integration site) on a cubic B-spline basis (i.e., a third-order polynomial) for length of the intergenic region.

matched random controls ($P = 1.2 \times 10^{-47}$ for T cells and $P = 6.7 \times 10^{-3}$ for macrophages).

Chromosomal banding patterns also correlate with the above collection of features, with more gene-rich and actively transcribed regions enriched in Giemsa-dark bands [35]. The Giemsa banding pattern was correlated with integration site distribution in T cells and macrophages. Both favored integration in Giemsa-dark regions, but the trend was slightly stronger in the T cell data ($P = 1.02 \times 10^{-11}$) (supplementary material).

In summary, we found integration sites from both the macrophage and the T cell data sets more frequently in a collection of features associated with regions that were gene dense, low in LTR elements and LINEs, higher in transcriptional intensity, enriched in shorter intergenic regions, and Giemsa dark. These trends, though seen in both data sets compared to random, were more pronounced for the T cell data set.

Integration Site Selection Analyzed by Chromosome

Integration site selection in whole chromosomes differed to a surprising degree between the macrophage and the T cell data sets. Previous work indicated that integration site selection was not random with respect to whole chromosomes [5,7,39]. Relative gene density differed among the chromosomes, and this appeared to account in part for the differences. The plot in Fig. 4 shows that integration in the T cell data set was strongly favored in the gene-dense chromosomes (e.g., 16,17,19,22) [35,40]. For integration in macrophages, some of the gene-dense

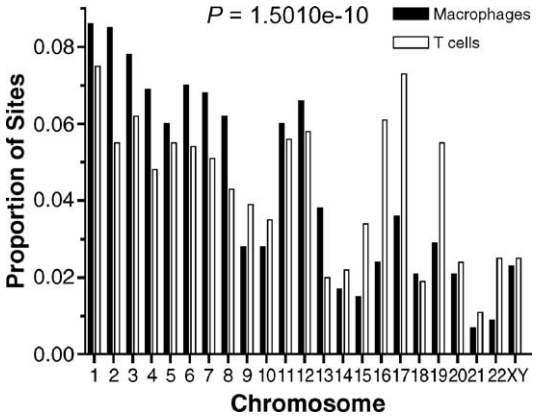

FIG. 4. Chromosomal preferences for integration in macrophages and T cells. The proportions of integration events for the macrophage and T cell data sets are shown for each chromosome. The *P* value was determined using a likelihood ratio statistic for a logistic regression model as described in the supplementary material.

chromosomes were favored for integration, though not as strongly. The most notable trend was that integration was more frequent in the longer (lower numbered) chromosomes, which are larger integration targets. Thus the relative responses to gene density and transcriptional intensity are affecting integration site distribution, but there may be additional unidentified factors affecting distribution of sites.

Further Analysis of Retroviral Integration Following Entry Mediated by Different Envelopes

The above data on integration by HIV vectors entering macrophages via the VSV-G or HIV R5 envelope indicated that the route of entry had little or no effect on integration targeting. Does this hold for other envelopes and other retroviruses? This question can be addressed by further analysis of published integration site data sets in which different envelopes were used for infection. A complication arises because different human cell types show reproducible differences in the frequency of integration in TUs [7,8]. However, integration in human SupT1 cells has been studied for both HIV vectors bearing the VSV-G envelope [5] and for authentic HIV bearing an X4-tropic envelope [14]. Another comparison is possible for integration by ASLV, though this involves comparing entry of an ASLV-based vector pseudotyped with VSV-G in HeLa cells [12] to entry using the natural subgroup A envelope into 293T cells engineered to express the Tva receptor [7].

We used automated comparisons to assess differences in integration frequency for each pair of data sets across a variety of genomic features, including chromosomal distribution, TUs, CpG islands, gene density, expression intensity, proximity to gene boundaries, cytobands, and G/C percentage. We compared different types of gene catalogs, and for some quantitative measures we also analyzed different length genomic intervals. No strongly significant differences were seen for either the HIV data sets in SupT1 cells or the ASLV data sets in HeLa and 293T cells (data not shown). For a small fraction of comparisons, the *P* values were slightly below 0.05, suggesting possible significance. However, so many pair-wise comparisons were evaluated that the biological relevance of these weak trends is questionable.

We conclude that the HIV/SupT1 and ASLV data sets did not show strong differences in integration site selection resulting from entry mediated by different retroviral envelopes, consistent with our findings in the macrophage model.

Discussion

We present 754 new integration site sequences made by infection of macrophages with an HIV-based vector. We find that in macrophages integration was favored in TUs. Comparison to transcriptional profiling data indicated that active TUs were particularly favored. In the one previous study of integration in nondividing cells, HIV-vector integration in growth-arrested IMR-90 lung fibroblasts was also found to be favored in TUs [10]. The findings presented here are consistent with integration in nondividing cells being generally favored in active TUs, as was seen with dividing cells. We note, however, that both macrophages and IMR-90 cells were arrested predominantly in G0/G1, leaving open the question of whether arrest at other cell cycle stages might affect targeting.

The frequency of integration in TUs was higher in T cells, and this correlated with increased integration in (1) SINEs, (2) regions of high transcriptional intensity, (3) short intergenic regions, and (4) Giemsa dark regions. The frequency of integration in these features differs among cell types, but reproducible differences are seen when a single cell type is analyzed repeatedly (see [8]). Why integration in T cells is consistently more frequent in TUs compared to other cell types is unclear, though we note that T cells have higher levels of the LEDGF/p75 expression than do monocytes or macrophages, and a previous study has suggested that LEDGF/p75 is partially responsible for targeting integration to TUs [8].

In this study, we also compared the possible influence of different routes of entry on integration site selection, but found no major differences due to entry mediated by the VSV-G or HIV R5 envelopes. Similarly, comparison of (1) HIV infection mediated by VSV-G or HIV/X4 envelopes or (2) ASLV infection mediated by VSV-G or ASLV-subgroup A envelopes also showed no major differences in integration site selection. Thus the available data suggest that integration targeting is not strongly affected by the route of entry. One simple model to explain these observations would be that, following entry by different routes, the sorting pathways traversed by PICs converge prior to integration.

Materials and Methods

Cell culture. Peripheral blood mononuclear cells were obtained by Ficoll–Hypaque separation from whole blood of healthy volunteers. Monocytes were positively selected in a cell sorter using anti-CD14. The purity of the sorted monocytes was >99%. Monocytes were cultured for 24 h at 37°C with 5% CO_2 at a concentration of 2×10^6 cells/ml in R-10 medium (RPMI medium supplemented with 10% heat-inactivated fetal bovine serum and 50 μg/ml gentamycin (Invitrogen)). Monocytes were added (2×10^6 per well) to a six-well plate in R-10 medium supplemented with 50 ng/ml recombinant human granulocyte/macrophage colony-stimulating factor (GM-CSF; Peprotech, Rocky Hill, NJ, USA) and incubated at 37°C with 5% CO_2 for 6 days. On the sixth day, the medium was aspirated and the cells were washed three times with prewarmed R-10 medium. The resulting differentiated macrophages, displaying the characteristic "fried egg" appearance, were incubated in R-10 medium lacking GM-CSF for a further 24 h before infection.

Preparation of HIV vector particles. HIV-1 vector particles were generated by calcium phosphate-mediated transfection of 293T cells with three plasmids: p156RRLsinPPTCMVGFPWPRE [41] (encoding the HIV vector segment), pCMVdeltaR9 [42] (the packaging construct), and pMD.G [42] (encoding the VSV-G envelope) or pBaL [33] (encoding the HIV envelope

doi:10.1016/j.ymthe.2006.03.012

specific for CCR5). Forty-eight hours after transfection, supernatants containing the viral particles were harvested and centrifuged for 5 min at 350*g* at 4°C to pellet cell debris. Supernatants were then filtered through a 0.45-μm filter and concentrated by ultracentrifugation at 23,000*g* for 2 h at 4°C. The viral pellets were resuspended in ~1/100 volume of fresh medium. The concentration of HIV-1 vector particles in stocks was determined by p24 ELISA.

Virus infection. Prior to infection, HIV-1 vector particles were digested with DNase I (0.2 units/μl) for 1 h at 37°C. An equivalent of 250 ng of p24 HIV/VSV-G or 1000 ng of p24 HIV/R5 in 1 ml of R-10 medium (without GM-CSF) was added to each well containing a monolayer of macrophages and incubated at 37°C with 5% CO_2 for 24 h. After 24 h, an additional 1 ml of R-10 medium (without GM-CSF) was added to each well and incubated for a further 6 days. After 6 days, cells were scraped, centrifuged for 5 min at 350*g*, and washed twice with PBS. Cells were tested for GFP expression by flow cytometry, revealing that 26% of the macrophages infected with HIV/VSV-G were positive, while 2.5% of cells infected with HIV/R5 were positive.

Isolation of integration sites. Integration sites were cloned by ligation-mediated PCR essentially as described previously [5,6]. Briefly, DNA was extracted using the DNeasy Tissue Kit (Qiagen, Valencia, CA, USA) 6 days postinfection and digested with *Avr*II, *Nhe*I, and *Spe*I. Linker DNA was ligated to the digested ends, followed by DNA amplification with two rounds of PCR, TOPO-TA/XL cloning, and high-throughput sequencing. Average lengths of human DNA sequences were 545 bp for the HIV/VSV-G data set and 543 bp for the HIV/R5 data set. Primer sequences are available as supplementary information in [9]. Integration site sequences are deposited at NCBI under accession numbers DX571401 to DX571726 (HIV-R5) and DX571727 to DX572203 (HIV-VSV).

Bioinformatic analyses. Integration site locations were identified using the BLAT feature in the Human Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway) against the May 2004 freeze of the human genome sequence (hg17). Integration site sequences were judged to be of acceptable quality if (1) the match to the genome began within 3 bp of the 5′-CA-3′ terminus of the viral DNA, (2) the match proximal to the LTR end showed an identity of at least 98%, and (3) the match yielded a unique best hit using default parameters in the client-server BLAT ranking. An integration site was scored as present in a TU if it was mapped in DNA between the base pairs encoding the 5′ and 3′ ends of the transcribed region as specified in the various gene catalog annotations (http://www.genome.ucsc.edu/cgi-bin/hgGateway). Analysis of previously published retroviral integration sites into the human genome was updated using the hg17 draft. New matched random controls were also generated as described in the supplementary material. "DNA elements" in Table 3 indicates remnants of DNA transposons. Detailed description of our bioinformatic methods can be found in the supplementary material.

ACKNOWLEDGMENTS

We thank Charles Berry for key software and help with statistical analysis and members of the Bushman laboratory for helpful discussions. This work was supported by NIH Grants AI52845 (to F.D.B.) and PAR-03-138 (to Carl June), the James B. Pendleton Charitable Trust, Robin and Frederic Withington (to F.D.B.), and the Fritz B. Burns Foundation (to J.R.E.). S.B. was supported in part by fellowships from the Natural Sciences and Engineering Research Council of Canada and the Alberta Heritage Foundation for Medical Research. A.C. was supported in part by a fellowship from the Swiss National Science Foundation.

RECEIVED FOR PUBLICATION JANUARY 4, 2006; REVISED FEBRUARY 14, 2006; ACCEPTED MARCH 5, 2006.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ymthe.2006.03.012.

REFERENCES

1. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Stebbing, J., Gazzard, B., and Douek, D. C. (2004). Where does HIV live? *N. Engl. J. Med.* **350:** 1872–1880.
3. Verani, A., Gras, G., and Pancino, G. (2005). Macrophages and HIV-1: dangerous liaisons. *Mol. Immunol.* **42:** 195–212.
4. Sharova, N., Swingler, C., Sharkey, M., and Stevenson, M. (2005). Macrophages archive HIV-1 virions for dissemination in trans. *EMBO J.* **24:** 2481–2489.
5. Schroder, A., Shinn, A., Chen, H., Berry, C., Ecker, J. R., and Bushman, F. D. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110:** 521–529.
6. Wu, X., Li, Y., Crise, B., and Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300:** 1749–1751.
7. Mitchell, R., *et al.* (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2:** E234.
8. Ciuffi, A., *et al.* (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* **11:** 1287–1289.
9. Barr, S. D., Leipzig, J., Shinn, P., Ecker, J. R., and Bushman, F. D. (2005). Integration targeting by avian sarcoma–leukosis virus and human immunodeficiency virus in the chicken genome. *J. Virol.* **79:** 12035–12044.
10. Ciuffi, A., *et al.* (2006). Integration site selection by HIV-based vectors: targeting in dividing and nondividing IMR-90 lung fibroblasts. *Mol. Ther.* **13:** 366–373.
11. Hematti, P., *et al.* (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* **2:** E423.
12. Narezkina, A., *et al.* (2004). Genome-wide analyses of avian sarcoma virus integration sites. *J. Virol.* **78:** 11656–11663.
13. Stevens, S. W., and Griffith, J. D. (1996). Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J. Virol.* **70:** 6459–6462.
14. Carteau, S., Hoffmann, C., and Bushman, F. D. (1998). Chromosome structure and HIV-1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J. Virol.* **72:** 4005–4014.
15. Holman, A. G., and Coffin, J. M. (2005). Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. USA* **102:** 6103–6107.
16. Wu X., Li Y., Crise B., Burgess S. M., and Munroe D. J. (2005). Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* **79.**
17. Panet, A., and Cedar, H. (1977). Selective degradation of integrated murine leukemia proviral DNA by deoxyribonucleases. *Cell* **11:** 933–940.
18. Vijaya, S., Steffan, D. L., and Robinson, H. L. (1986). Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin. *J. Virol.* **60:** 683–692.
19. Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* **61:** 336.
20. Sandmeyer, S. (2003). Integration by design. *Proc. Natl. Acad. Sci. USA* **100:** 5586–5588.
21. Zhu, Y., Dai, J., Fuerst, P. G., and Voytas, D. F. (2003). Controlling integration specificity of yeast retrotransposon. *Proc. Natl. Acad. Sci. USA* **100:** 5891–5895.
22. Boeke, J. D., and Devine, S. E. (1998). Yeast retrotransposons: finding a nice quiet neighborhood. *Cell* **93:** 1087–1089.
23. Bushman, F. D. (2003). Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* **115:** 135–138.
24. Bushman, F., *et al.* (2005). Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* **3:** 848–858.
25. Cherepanov, P., *et al.* (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* **278:** 372–381.
26. Maertens, G., *et al.* (2003). LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* **278:** 33528–33539.
27. Turlure, F., Devroe, E., Silver, P. A., *et al.* and Engelman, A. (2004). Human cell proteins and human immunodeficiency virus DNA integration. *Front. Biosci.* **9:** 3187–3208.
28. Llano, M., *et al.* (2004). LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.* **78:** 9524–9537.
29. Llano, M., Delgado, S., Vanegas, M., and Poeschla, E. M. (2004). LEDGF/p75 prevents proteasomal degradation of HIV-1 integrase. *J. Biol. Chem.* **279:** 55570–55577.
30. Cherepanov, P., Ambrosio, A. L., Rahman, S., Ellenberger, T., and Engelman, A. (2005). Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl. Acad. Sci. USA* **102:** 17308–17313.
31. Roe, T., Reynolds, T. C., Yu, G., and Brown, P. O. (1993). Integration of murine leukemia virus DNA depends on mitosis. *EMBO J.* **12:** 2099–2108.
32. Lewis, P., Hensel, M., and Emerman, M. (1992). Human immunodeficiency virus infection of cells arrested in the cell cycle. *EMBO J.***:** 3053–3058.

33. Hwang, S. S., Boyle, T. J., Lyerly, H. K., and Cullen, B. R. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253:** 71–74.
34. Smit, A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9:** 657–663.
35. Lander, E., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
36. Lewinski, M., *et al.* (2005). Genome-wide analysis of chromosomal features repressing HIV transcription. *J. Virol.* **79:** 6610–6619.
37. Zhao, L., *et al.* (2004). The 5-lipoxygenase pathway promotes pathogenesis of hyperlipidemia-dependent aortic aneurysm. *Nat. Med.* **10:** 966–973.
38. Chaussabel, D., Semnani, R. T., McDowell, M. A., Sacks, D., Sher, A., and Nutman, T. B. (2003). Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. *Blood* **102:** 672–681.
39. Laufs, S., *et al.* (2003). Retroviral vector integration occurs in preferred genomic targets in human bone marrow-repopulating cells. *Blood* **101:** 2191–2198.
40. Venter, J. C. (2001). The sequence of the human genome. *Science* **291:** 1304–1351.
41. Follenzi, A., Ailes, L. E., Bakovic, S., Gueuna, M., and Naldini, L. (2000). Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat. Genet.* **25:** 217–222.
42. Naldini, L., *et al.* (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272:** 263–267.