ARTICLE

doi:10.1016/j.ymthe.2005.10.009

Integration Site Selection by HIV-Based Vectors in Dividing and Growth-Arrested IMR-90 Lung Fibroblasts

Angela Ciuffi,* Richard S. Mitchell,* Christian Hoffmann, Jeremy Leipzig, Paul Shinn,[1] Joseph R. Ecker,[1] and Frederic D. Bushman[†,‡]

Department of Microbiology, University of Pennsylvania School of Medicine, 3610 Hamilton Walk, Philadelphia, PA 19104-6076, USA
[1]Genomic Analysis Laboratory, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

**These authors contributed equally to this work.*

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

[‡]Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. DU665871 to DU667458.

Available online 1 December 2005

DNA integration is a defining step in the retroviral life cycle and the basis of stable gene transfer in retrovirus-based gene therapy. Previous studies of integration by HIV-based vectors have shown that integration is not random, but favored in active transcription units. Studies to date have focused on HIV integration in dividing cells, leaving open the question of whether integration target site selection might differ in nondividing cells. According to one idea, division of the host cell might be required for favored integration in transcription units, possibly as a result of chromatin remodeling during DNA replication. Here we have investigated this issue by comparing integration in dividing IMR-90 primary lung fibroblasts to integration in nondividing IMR-90 cells arrested in G1 by serum starvation and contact inhibition. We identified several differences in integration site selection in arrested versus dividing cells, including the frequency of integration in transcription units and in gene-rich regions. However, integration in nondividing cells was in fact more favored in transcription units, contrary to the idea that cell division was important for this bias. These data provide the first view of lentiviral integration in nondividing cells and help constrain models for the mechanism of favored integration in genes.

Key Words: HIV, lentivirus, vector, gene therapy, fibroblast, integrase, retrovirus, cell cycle, integration

INTRODUCTION

Integration of a DNA copy of the retroviral RNA genome into a chromosome of the host cell is a necessary and defining step in the retroviral life cycle [1,2]. The selection of target sites for integration is mostly independent of the sequence of the target DNA (though weak effects of sequence can be detected with careful study) [3–6]. Despite this, the sites of integration are not randomly placed in chromosomes, but instead are strongly influenced by local features, and the response to these features differs among retroviruses [7–9]. HIV favors integration in transcription units, and analysis of gene activity in infection target cells revealed that integration was particularly favored in active genes [7–9]. A study of SIV integration in macaques also showed favored integration in transcription units, suggesting that favored integration in active genes may be a shared feature of the lentiviruses [10]. Murine leukemia virus (MLV), in contrast, favors integration near gene transcription start sites in both humans and macaques [8,10]. Avian sarcoma-leukosis virus, in contrast, shows a third pattern, displaying a near-random distribution of integration sites, with no favoring of transcription start sites and only weak favoring of transcription units [9,11].

Integration target site selection has become an important issue in evaluating the safety of human gene therapy. The majority of gene therapy trials have involved the use of retroviral vectors to transduce therapeutic sequences and incorporate them stably into the target genome. Unfortunately, integration can also result in genomic toxicity. In the course of an otherwise quite successful gene therapy trial treating X-linked severe combined immunodeficiency, three patients developed a leukemia-like illness, and in two cases this was associated with

Copyright © The American Society of Gene Therapy
1525-0016/$30.00

integration of the gene therapy vector near the *LMO2* proto-oncogene [12,13]. This trial used an MLV-based vector for gene delivery, and the sites of integration were within the favored target region near transcription start sites, raising concerns that MLV integration site preferences might not be optimal for human gene therapy.

HIV-based vectors, which instead favor integration within transcription units, may thus be more attractive for human gene therapy, but the factors affecting targeting *in vivo* are just beginning to be worked out. The finding that transcriptional activity is favorable for integration suggests that integration site selection might differ among cell types because of cell-type-specific differences in gene activity. Such a trend was found to be significant in Mitchell *et al*. [9], though the magnitude of the effect was modest. Other studies have revealed reproducible cell-type-specific differences in the frequency of integration in transcription units, but for these as well the effects were also quantitatively modest [7–9].

Here we compared integration of an HIV-based vector in rapidly dividing cells to integration in cells arrested by serum starvation and contact inhibition. We reasoned that such an extreme difference in cell physiology might affect integration targeting. According to one idea, integration in active transcription units might be promoted by chromatin remodeling during rapid cell division and that conversely integration in nondividing cells would not show this trend. This issue is also of potential significance to the biology of HIV, for which nondividing macrophages are an important infection target in patients.

To investigate this issue, we studied integration in IMR-90 primary human lung fibroblasts [14], because IMR-90 cells can be efficiently arrested by serum starvation and contact inhibition and because analysis of primary cells is not complicated by chromosomal aneuploidy as is common in transformed cell lines. We found that integration was favored in transcription units in both dividing and nondividing IMR-90 cells, with the nondividing cells actually showing slightly higher frequency of integration in transcription units. Nondividing IMR-90 cells also showed more integration in actively transcribed genes and gene-rich regions. Thus the altered cell physiology associated with this form of cell cycle arrest did not diminish the proportion of integration sites accumulating in transcription units. These findings constrain models for the mechanisms guiding HIV integration and inform strategies for modulating integration site selection in cells.

RESULTS

Cell Cycle Arrest in IMR-90 Cells by Serum Starvation and Contact Inhibition

We wished to analyze the consequences of cell cycle arrest on integration site selection and so chose the IMR-90 model, in which integration in dividing and nondividing cells could be readily compared. We compared rapidly dividing IMR-90 cells to IMR-90 cells that were growth arrested in G1 by contact inhibition and serum starvation. We carried out cell cycle characterization of arrested and dividing cells using two-dimensional flow sorting as described in [15,16] (Fig. 1). We pulsed dividing and nondividing cells with BrdU for 90 min, harvested them, and fixed them in cold 70% ethanol. We then labeled the cells with anti-BrdU antibody and propidium iodide and analyzed them by FACS. The BrdU label allows new DNA synthesis to be quantified, while propidium iodide staining reports the total DNA content per cell.

Fig. 1A shows a FACS profile for dividing IMR-90 cells. The overall shape of the profile resembles an upside-down U. The G1 cells, low for both BrdU and propidium iodide staining, are seen in the lower left arm of the U. The middle crosspiece of the U, indicating DNA content between 2*n* and 4*n* and high BrdU staining, is composed of cells in S phase. G2/M cells form the right arm of the U (cells low in BrdU staining, 4*n* DNA content).

Characterization of arrested cells (Fig. 1B) shows a near complete absence of cells in S, indicating that the procedure for arresting the cell cycle was effective. About 90% of the cells were arrested in G1, with the remaining 10% arrested in G2/M. Thus the combination of serum

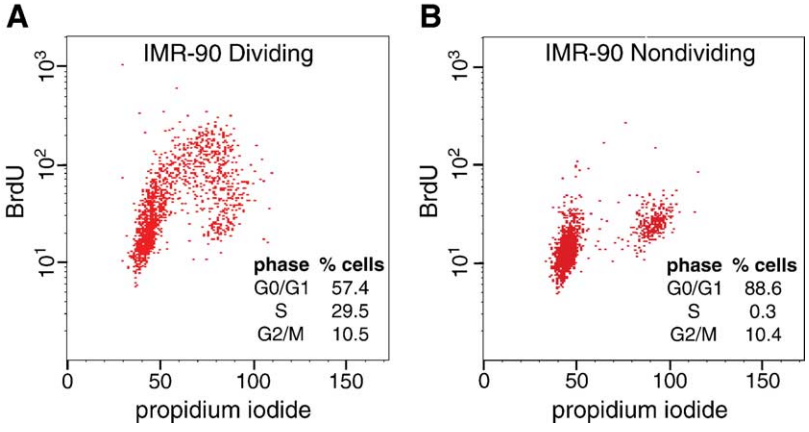

FIG. 1. Arrest of IMR-90 cells by serum starvation and contact inhibition. Cells were fixed, stained with anti-BrdU antibody and propidium iodide, and separated by FACS. (A) Dividing IMR-90 cells. (B) Nondividing IMR-90 cells.

Copyright © The American Society of Gene Therapy

doi:10.1016/j.ymthe.2005.10.009

starvation and contact inhibition caused essentially complete cell cycle arrest, with most of the cells in G1, as expected. In what follows we can be confident that all integration sites from the nondividing cell population were indeed from arrested cells. Most integration sites were likely isolated from cells arrested in G1, a smaller fraction may have been from cells arrested in G2/M.

Infection and Isolation of Integration Sites

We infected dividing and nondividing IMR-90 cells with an HIV-based vector that encodes a GFP marker that allows simple monitoring of transduction. Characterization of the percentage infection after 48 h by FACS showed that 30% of the dividing IMR-90 cells and 5% of the nondividing cells were transduced (data not shown). We harvested DNA from each cell population and cloned sequences flanking integrated proviruses using ligation-mediated PCR as described [7–9]. Briefly, genomic DNA was cleaved with restriction enzymes, a DNA adaptor was ligated to the DNA ends, and then PCR was carried out using primers that bound to the adaptor and the viral DNA end. We cloned, sequenced, and mapped PCR products to the human genome draft hg17. For nondividing IMR-90 cells, we determined 182 unique integration site sequences and compared these to 482 sites of integration in dividing IMR-90 cells reported previously [9].

Integration Site Distribution in Dividing and Nondividing IMR-90 Cells

We then analyzed the distributions of lentiviral integration sites in dividing and nondividing cells statistically. For comparison, we paired each integration site with 30 matched random locations in the human genome. We used the matching procedure to account for possible biases in the recovery of integration sites resulting from the use of restriction enzyme cleavage to clone integration sites. We chose each matched random control to be a random location in the human genome, but placed at the same distance from a restriction cleavage site as the experimental site. For statistical analysis, we pooled the experimental sites and compared them to pooled matched random controls.

We then quantified the relationship between integration sites in dividing and nondividing IMR-90 cells and analyzed them for significant differences. We carried out systematic comparisons in a semiautomated fashion across many types of genomic features. This analysis is presented in detail in the supplementary material and key conclusions are summarized below.

Integration frequency in transcription units is compared in Table 1. A complication in this analysis arises due to uncertainties in cataloging the collection of human genes. We therefore used six different sets of gene calls in our analysis to make sure that conclusions drawn were not dependent on the particular gene calls used. Integration in both dividing and nondividing IMR-90 cells was strongly favored in transcription units relative to the matched random control ($P < 0.0001$ for most comparisons). Taking the well-characterized RefSeq genes for comparison, 62% of sites from dividing IMR-90 cells and 71% of sites from nondividing IMR-90 were in transcription units, compared to 35% of the matched random sites. The conclusions were similar for other gene calls, though the exact values differed. A comparison of the frequency of integration in transcription units in dividing versus nondividing cells showed that the nondividing cells had a consistently higher proportion of integration sites in transcription units (Table 1). This trend achieved statistical significance for all the evidence-based gene calls ($P = 0.03$ to $P = 0.003$), with the computational GenScan calls the only ones returning an insignificant difference ($P = 0.4$). Thus integration was favored in transcription units in both dividing and nondividing IMR-90 cells, and unexpectedly the trend was significantly stronger in the nondividing cells.

Another approach to investigating this issue involves analyzing integration in gene-rich regions. To carry out this analysis, we partitioned the human chromosomes into intervals of 500 kb and quantified gene density and integration frequency in each. We then pooled the collection of integration sites for dividing and nondividing cells, divided the pools into 10 fractions based on the relative gene density in each interval, and then quantified the contributions of sites from dividing and nondividing cells for each interval (Fig. 2). This revealed a strong trend for sites from dividing IMR-90 to lie in the less gene-dense regions. Statistical analysis revealed the comparison to be highly significant ($P = 0.0001$). Thus

TABLE 1: Frequency of HIV integration in transcription units

Transcription unit catalog	IMR-90 dividing (P value vs random)	IMR-90 nondividing (P value vs random)	IMR-90 dividing matched random control	IMR-90 nondividing matched random control
Acembly	78.6% (<0.0001)	86.3% (<0.0001)	52.4%	52.3%
GenScan	74.9% (0.0089)	78.0% (<0.0001)	69.3%	69.7%
RefSeq	61.8% (<0.0001)	70.9% (<0.0001)	34.5%	35.2%
UniGene	65.4% (<0.0001)	74.7% (<0.0001)	40.2%	40.6%
Known	66.6% (<0.0001)	78.6% (<0.0001)	38.2%	38.8%
Ensembl	64.7% (<0.0001)	75.8% (<0.0001)	38.0%	38.0%

Copyright © The American Society of Gene Therapy

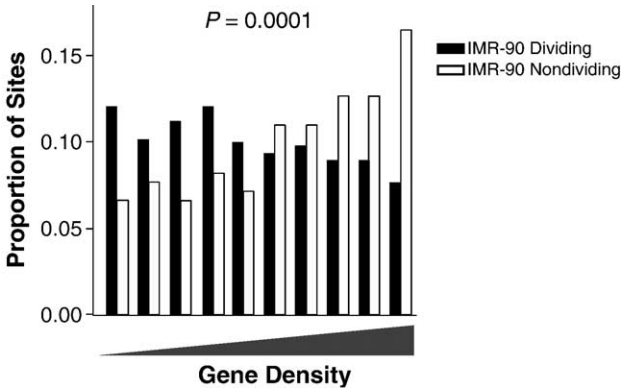

FIG. 2. Effects of gene density on integration in arrested and dividing IMR-90 cells. Genes hosting integration events in the IMR-90 dividing and nondividing cells were pooled and then divided into 10 equal portions based on the gene density in the surrounding 500 kb. The contributions of sites from the dividing cells and nondividing cells were then compared in each pool. The *P* value is obtained from the logistic regression of event type (dividing cell versus nondividing cell integration site) on a cubic B-spline basis (i.e., a third-order polynomial) for gene density. The cut values were group 1, 0 to 2.00×10^{-6}; group 2, 2.00×10^{-6} to 5.47×10^{-6}; group 3, 5.47×10^{-6} to 7.97×10^{-6}; group 4, 7.97×10^{-6} to 1.00×10^{-5}; group 5, 1.00×10^{-5} to 1.20×10^{-5}; group 6, 1.20×10^{-5} to 1.60×10^{-5}; group 7, 1.60×10^{-5} to 1.94×10^{-5}; group 8, 1.94×10^{-5} to 2.80×10^{-5}; group 9, 2.80×10^{-5} to 4.26×10^{-5}; group 10, 4.26×10^{-5} to 1.01×10^{-4}.

integration in the nondividing IMR-90 cells was favored in more gene-dense regions, paralleling the finding of favored integration in transcription units (Table 1).

The human genome is partitioned into gene-dense and gene-sparse regions, and several other characteristics also follow this trend—gene-dense regions are also more G/C-rich, rich in short interspersed nuclear elements (SINEs), sparse in long interspersed nuclear elements (LINEs), and enriched in Giemsa-dark bands [17,18]. To investigate whether these trends were reflected in the IMR-90 data, we carried out an analysis of integration frequency in repeated sequences (Table 2). For this analysis, we pooled unique integration sites with sites in identifiable repeats that had multiple best hits on the human genome and so could not be mapped to unique locations. Summed numbers of sites were 499 in dividing IMR-90 and 187 in nondividing IMR-90.

Integration in both data sets was favored in SINEs compared to the matched random controls, consistent with favored integration in gene-rich regions. This trend achieved statistical significance for *Alu* elements in the nondividing cell data set and for the mammalian interspersed repeats (MIRs) in the dividing cell data. LINEs tend to cluster in A/T-rich gene-sparse regions, and integration in both the dividing and the nondividing IMR-90 data sets was disfavored in LINEs. The trend achieved statistical significance only in the larger IMR-90 dividing cell data set (note that statistical significance is a function of both the strength of the trend and the number of measurements—thus a trend with the same difference in proportions will be more significant in a larger data set). Integration in both data sets was less frequent in endogenous retroviral sequences (LTR elements), which are also known to be found predominantly outside genes. This trend achieved significance in the larger dividing IMR-90 integration site data set. Both data sets showed slightly favored integration in DNA elements, and the trend achieved significance in the dividing cell data. These elements are fairly uniformly distributed relative to genes.

In summary, integration in both dividing and nondividing IMR-90 cells was strongly favored in transcription units. The effect was stronger in the nondividing cells, where integration in transcription units, gene-rich regions, and *Alu* repeats was significantly more frequent.

Genome-wide Transcriptional Analysis of Dividing and Arrested IMR-90 Cells

Next we characterized the changes in transcription resulting from G1 arrest in IMR-90 cells by transcriptional profiling. These data provide an overview of the physiological changes accompanying G1 arrest and allow the influence of transcription on integration site selection to be assessed. We isolated RNA from dividing and growth-arrested IMR-90 cells, labeled it, and applied it to two Affymetrix HU95Av2 chips (four chips total). This section describes the transcriptional changes detected—the next section considers the relationship to integration site selection.

TABLE 2: Genomic features at integration sites

Chromosomal feature	IMR-90 dividing (*P* value vs random)	IMR-90 nondividing (*P* value vs random)	IMR-90 dividing matched random control	IMR-90 nondividing matched random control
SINEs				
Alu	11.3% (0.3017)	20.9% (<0.0001)	9.8%	9.4%
MIR	5.0% (0.0122)	5.4% (0.1435)	3.0%	3.4%
DNA elements	5.2% (0.0137)	4.3% (0.3326)	3.2%	3.0%
LTR elements	5.4% (0.0053)	5.4% (0.1104)	9.0%	8.7%
LINEs	18.8% (0.0096)	17.6% (0.1518)	23.9%	22.1%
α satellite	0.0% (0.7161)	0.0% (0.6082)	0.0%	0.1%
Other	0.8% (0.1447)	1.1% (0.6262)	1.6%	1.5%

Copyright © The American Society of Gene Therapy

doi:10.1016/j.ymthe.2005.10.009

Analysis of the measured average difference values showed that the relative expressions of genes from the two samples were quite similar, with all pair-wise comparisons between microarrays showing correlation coefficients of 0.95 or higher. However, the pair-wise comparisons between arrays under the same conditions (cycling or G1-arrested) showed closer correlations (R = 0.98 or higher) than comparisons between conditions (R = 0.95 to 0.97), indicating consistent changes in transcription due to growth arrest.

We identified genes affected by growth arrest using the Significance Analysis of Microarray [19], accepting a false discovery rate of about 10% (δ = 1.90). This yielded 414 genes, of which 218 were increased in activity upon growth arrest and 196 were decreased (Fig. 3). We classified the types of genes affected by growth arrest according to their gene ontology using GoMiner [20] and EASE [21]. Groups of genes that were significantly changed relative to the representation on the microarray as a whole were identified using Fisher's exact test.

A large collection of genes for energy generation and macromolecular synthesis was downregulated by growth arrest. Twenty genes with oxidoreductase activity were downregulated ($P = 2 \times 10^{-6}$), including 3 genes involved in ATP biosynthesis (cytochrome c oxidase subunit VIc, NADH dehydrogenase Fe-S protein 4, NADH dehydrogenase flavoprotein 2). The "biosynthesis" category contained 19 downregulated genes (P = 0.02), including genes involved in cholesterol metabolism (6 genes, P = 0.0002) and lipid biosynthesis (9 genes, P = 0.0003). Notable among genes upregulated by growth arrest were 29 genes for "extracellular" components ($P = 4.5 \times 10^{-5}$) of which 20 encoded components of the "extracellular matrix" ($P = 1 \times 10^{-9}$). Other strongly affected classes included genes involved in cell adhesion (15 genes, P = 0.025) and calcium-binding proteins (16 genes, P = 0.0007).

These transcriptional changes parallel a number of previous reports on the physiological responses of fibroblasts to serum. Cellular–extracellular matrix interactions have been implicated in regulating G1 arrest in fibroblasts [22], consistent with transcriptional responses reported here. Several previous genome-wide surveys have investigated the transcriptional response of fibroblasts to serum [23–25], yielding findings that parallel ours in a variety of respects. Much of the change in gene activity, particularly in genes important for remodeling of the extracellular environment, has been previously proposed to reflect a coupling of growth control in fibroblasts to functions important in wound repair [24,25]. Thus the transcriptional profiling data on the IMR-90 cells confirm that many of the expected changes in cell physiology took place upon cell cycle arrest and provide the background needed to analyze the relationship between integration site selection and gene activity.

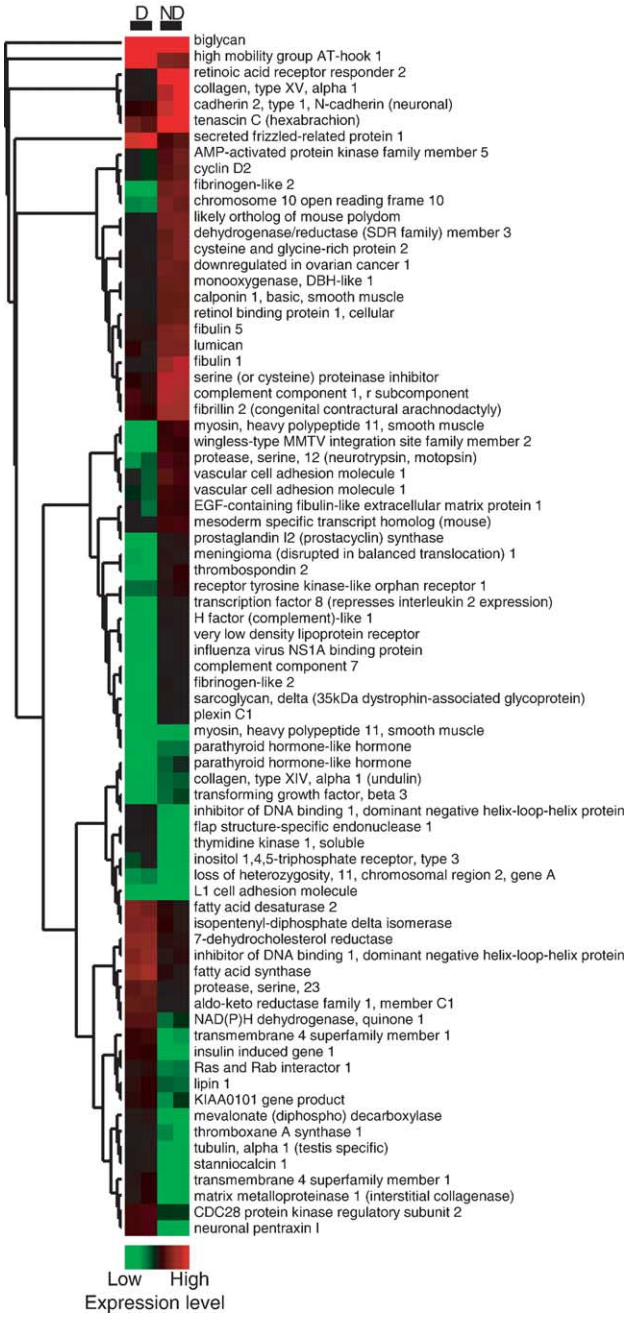

FIG. 3. Differences in gene activity between dividing (D) and nondividing (ND) IMR-90 cells. RNA from dividing IMR-90 cells was compared to that of nondividing cells using Affymetrix HU95Av2 microarrays (two chips each for dividing and nondividing cells). The 414 significantly affected genes were culled by demanding at least a threefold change in each direction, yielding 75 genes. Genes were clustered by expression pattern using hierarchical clustering with a Euclidean distance method. Intensity values are color coded from low (green) to high (red). Note that values were not divided by the row mean, thereby allowing comparison of absolute expression levels between rows.

Copyright © The American Society of Gene Therapy

Effects of Transcription on Integration Site Selection in G1-arrested IMR-90 Cells

We next assessed the relationship between transcriptional activity and integration site selection in the dividing and nondividing IMR-90 cells. We collected the genes hosting integration events that were queried by the HU95Av2 arrays and extracted their expression signals. A comparison of the medians revealed that the expression levels of genes targeted for integration in both data sets were significantly higher than for genes targeted in the matched random control ($P < 0.0001$ by Mann–Whitney test). The median for genes targeted in the nondividing cells was slightly higher than that in the dividing cells ($P = 0.01$, Mann–Whitney).

To investigate this issue more carefully, we divided all the genes on the array into eight categories, then distributed those genes that had hosted integration events into the same bins based on their expression levels, and summed the numbers in each category (Fig. 4). Statistical analysis across this distribution showed that the values again were significantly higher for genes targeted in nondividing cells, and this effect was particularly pronounced among the more highly expressed genes (categories 6 and higher).

Another approach to investigating this issue involves comparing integration frequency in intervals along the genome to expression intensity, which combines measures of gene density and expression values for each gene. We divided the genome into 500-kb intervals and quantified integration intensity and expression intensity for each. We then pooled the integration sites from dividing and nondividing cells and divided them into nine categories based on their expression intensity. Then we determined the proportion of integration sites from each data set for each category and plotted it (Fig. 5). Statistical analysis revealed a highly significant trend, with the integration sites from the nondividing cells significantly enriched in intervals of higher expression intensity ($P < 0.0001$).

In summary, integration in both dividing and nondividing IMR-90 cells was favored in active transcription units, and the trend was significantly stronger for integration in the nondividing cells. Combining measures of gene density and gene expression in the expression intensity parameter revealed a particularly notable difference between the data sets.

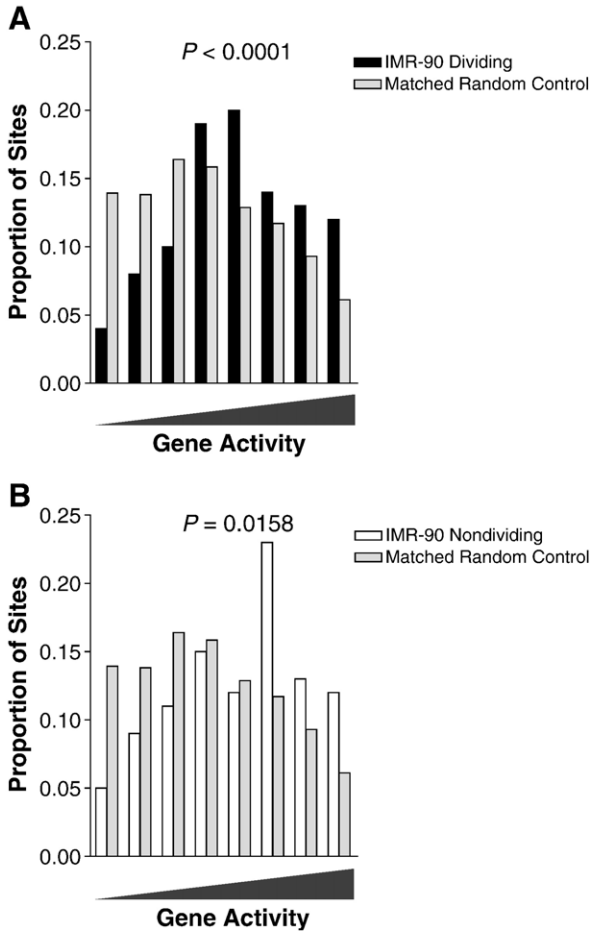

FIG. 4. Effects of gene activity on integration site selection. Expression levels were assayed in IMR-90 cells using Affymetrix HU95Av2 chips and scored using Affymetrix Microarray Suite 5.1 software. To assess the expression levels of genes hosting integration events, class boundaries were first generated by dividing all the genes on the array into eight classes by their relative level of expression. Genes that hosted integration events were then distributed into these classes, summed, and expressed as a proportion of the total number of integration sites in genes on the array. The leftmost class in each graph contains the most weakly expressed one-eighth of the genes, the rightmost class contains the most highly expressed one-eighth. Each was compared to genes ''targeted'' for integration in the matched random controls (controls from the dividing cells and nondividing cells were pooled for this analysis). (A) Expression levels of transcription units targeted for integration in the dividing IMR-90 cells compared to the matched random controls. (B) Expression levels of transcription units targeted for integration in the nondividing IMR-90 cells compared to the matched random controls.

Discussion

In this paper we compared integration by an HIV-based vector in dividing IMR-90 cells versus IMR-90 cells that were growth arrested by serum starvation and contact inhibition. These treatments predominantly arrested the cells in the G1 phase of the cell cycle, though a minor population arrested in G2/M could also be detected. Transcriptional profiling analysis confirmed that the arrested cells showed the physiological responses characteristic of serum withdrawal. Integration sites were characterized by cloning and sequencing junctions between viral and cellular DNA and mapping their locations on the human genome. Bioinformatic analysis revealed that integration in both dividing and nondividing cells was strongly favored in transcription units and that differences in the distributions between the two data sets were subtle. Thus

Copyright © The American Society of Gene Therapy

doi:10.1016/j.ymthe.2005.10.009

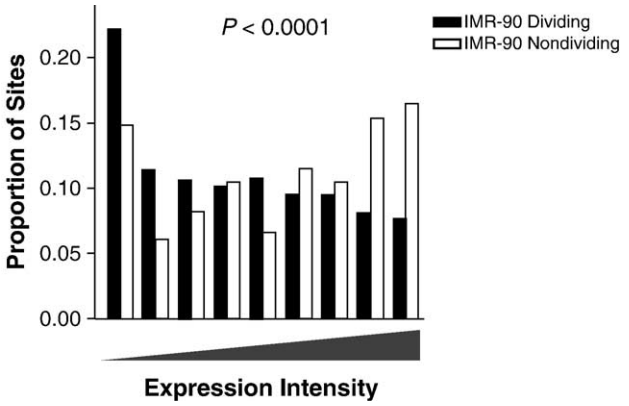

FIG.5. Effects of regional transcriptional intensity on integration. Expression intensity was quantified as follows. Transcriptional signals were quantified for 50,000 loci on the human genome, and the signal from each was scored as either in the upper half or in the lower half of all values on the array. Loci with expression values in the upper half were then collected and used in the subsequent analysis, allowing expression intensity to be estimated. The numbers of qualifying loci were then summed for 500-kb regions surrounding integration sites, and the distribution of values was quantified. Integration sites from dividing and nondividing cells were pooled and divided into nine bins based on expression intensity of the surrounding interval, and the contributions of integration sites from dividing and nondividing cells were compared in each bin. The *P* value is obtained from the logistic regression of event type (dividing cell versus nondividing cell integration site) on a cubic B-spline basis (i.e., a third-order polynomial) for transcriptional intensity. Transcriptional profiling data using the HU133Av2 chip probed with RNA from 293T cells were used for this analysis, since this allows about 50,000 loci to be queried, compared with only 12,000 for the HU95Av2 chip, providing much greater statistical power. The cut values were group 1, 0 to 1.33×10^{-6}; group 2, 1.33×10^{-6} to 2.33×10^{-6}; group 3, 2.33×10^{-6} to 4.17×10^{-6}; group 4, 4.17×10^{-6} to 6.00×10^{-6}; group 5, 6.00×10^{-6} to 7.66×10^{-6}; group 6, 7.66×10^{-6} to 1.01×10^{-5}; group 7, 1.01×10^{-5} to 1.40×10^{-5}; group 8, 1.40×10^{-5} to 2.39×10^{-5}; group 9, 2.39×10^{-5} to 5.70×10^{-5}.

arresting IMR-90 cells by this method resulted in only slight differences in integration site selection compared to dividing cells.

Evidently the hypothesis that cell cycle arrest would reduce integration in transcription units was incorrect. Additional studies will be required to determine whether this holds for other combinations of cell type, retroviral vector, and arrest protocol. We note that a study of integration by HIV in macrophages (S. Barr, A.C., J.L., P.S., J.R.E., F.D.B., unpublished data) also showed favored integration in transcription units despite the nondividing status of the target cells, suggesting that the observations in IMR-90 cells may hold for at least some other cell types. One hope in embarking on this study was that manipulating the cell cycle status of infection target cells might offer some measure of control over the placement of integration sites, which could have been of use in developing safer transduction methods for gene therapy, but so far this has not been supported by our studies.

An unexpected finding was that integration in the nondividing cells was in fact more frequent in transcription units than in the dividing cells (though for both dividing and nondividing cells the frequency of integration in transcription units was greater than for the matched random control). The mechanistic basis for this trend is unknown. One possible model would posit that there are changes in chromatin disfavorable to lentiviral integration in nondividing cells, but that these primarily affect regions outside transcription units. The net effect would thus be to increase the proportion of integration sites within transcription units. It is even possible that the nondividing state of the cells is disfavorable for integration generally, but that because integration in intergenic regions is more affected, there is a consequent increase in the proportion in transcription units. Additional work will be required to investigate this possibility.

MATERIALS AND METHODS

Cell culture. IMR-90 cells (initially at passage 36) were cultured in Dulbecco's modified Eagle's medium (DMEM; Invitrogen, Carlsbad, CA, USA) with 10% heat-inactivated fetal bovine serum (FBS). For growth arrest, the medium was removed from confluent dishes and replaced with DMEM plus 0.5% FBS (D0.5 medium), and cells were maintained in this state for 14 days, with medium being changed every 2 days. 293T cells were grown in DMEM with 10% FBS.

HIV-based vector infection. To produce HIV-based vector particles, 293T cells were cotransfected with three plasmids using the calcium phosphate method. One plasmid encoded the HIV vector segment (p156RRLsinPPTCMVGFPWPRE) [26], the second, the packaging construct (pCMVdeltaR9) [27], and the third, the gene for VSV-G (pMD.G) [27]. Forty-eight hours after transfection, the supernatants were collected, centrifuged to pellet cellular debris, and filtered through 0.45-μm filters. Viral particles were pelleted by ultracentrifugation in an SW40 rotor at 30K rpm, for 2 h at 4°C, and resuspended in 1 ml D0.5 medium. The viral supernatant was treated with 4 units/ml DNase I (Promega RQ1) for 90 min at 37°C. The vector was then diluted into D0.5 medium and DEAE-dextran added (final concentration 10 μg/ml). Infections were carried out at an m.o.i. of 10 (120 ng p24/10^6 cells).

Isolation of integration sites. Integration site clones were prepared as described in Schroder *et al*. [7]. DNA was isolated from cells 48 h after infection using the Qiagen DNeasy kit. Genomic DNA was cleaved with *Avr*II, *Spe*I, and *Nhe*I and then ligated to adaptors (sequences of DNA oligonucleotides used in this study are as described in [7]). Sequences at junctions between host and viral DNA were then amplified using one primer complementary to the adaptor and one complementary to the viral LTR. A second round of nested PCR was carried out subsequently. The PCR products were then cloned using the Topo XL and TA cloning kits, and individual colonies were grown into microtiter plates and analyzed by high-throughput sequencing.

Transcriptional profiling. RNA was purified using TRIzol, labeled according to the Affymetrix protocol, and then applied to Affymetrix HU95Av2 chips. Signal values were generated using Affymetrix Microarray Suite 5.1.

Bioinformatic analysis. Sequences were trimmed to remove linker and viral DNA sequences using custom software and aligned on the human genome (hg17) using BLAT. Local features were cataloged and used to generate Tables 1 and 2. Matched random controls were generated computationally as described in the supplementary data. For the analysis in Fig. 5 and the supplementary data, a larger 293T transcriptional profiling data set using the HU133Aplus2 chip was used, thereby increasing the statistical power of the analysis. In this work, we have

Copyright © The American Society of Gene Therapy

not controlled for inflation of type 1 error due to multiple comparisons. Our approach has been to restrict interpretation to highly significant trends seen consistently across multiple measures; i.e., we demand levels of significance that would very likely survive correction for multiple comparison, though to work out exactly the statistical corrections would be very difficult. For one thing, it is difficult to judge exactly what qualifies as independent comparisons in genomic data, since the distributions of many genomic features are correlated with one another. Microarray data can be found at GEO GSE3555.

ACKNOWLEDGMENTS

We thank members of the Bushman laboratory for helpful discussions. This work was supported by NIH Grant AI52845, the James B. Pendleton Charitable Trust, and Robin and Frederic Withington (to F.D.B.) and the Fritz B. Burns Foundation (to J.R.E.). A.C. was supported in part by a fellowship from the Swiss National Science Foundation.

RECEIVED FOR PUBLICATION OCTOBER 3, 2005; REVISED OCTOBER 27, 2005; ACCEPTED OCTOBER 27, 2005.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ymthe.2005.10.009.

REFERENCES

1. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). *Retroviruses.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Bushman, F., *et al.* (2005). Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* **3:** 848–858.
3. 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.
4. 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.
5. 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.
6. 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:** 5211–5214.
7. Schroder, A., *et al.* (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110:** 521–529.
8. 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.
9. Mitchell, R., *et al.* (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2:** E234.
10. Hematti, P., *et al.* (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* **2:** E423.
11. Narezkina, A., *et al.* (2004). Genome-wide analyses of avian sarcoma virus integration sites. *J. Virol.* **78:** 11656–11663.
12. Hacein-Bey-Abina, S., *et al.* (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **348:** 255–256.
13. Hacein-Bey-Abina, S., *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302:** 400–401.
14. Nichols, W. W., *et al.* (1977). Characterization of a new human diploid cell strain, IMR-90. *Science* **196:** 60–63.
15. Wade, M., and Allday, M. J. (2000). Epstein–Barr virus suppresses a G(2)/M checkpoint activated by genotoxins. *Mol. Cell. Biol.* **20:** 1344–1360.
16. Dolbeare, F., Gratzner, H., Pallavicini, M. G., and Gray, J. W. (1983). Flow cytometric measurement of total DNA content and incorporated bromodeoxyuridine. *Proc. Natl. Acad. Sci. USA* **80:** 5573–5577.
17. Venter, J. C. (2001). The sequence of the human genome. *Science* **291:** 1304–1351.
18. Lander, E., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
19. Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98:** 5116–5121.
20. Zeeberg, B. R., *et al.* (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* **4:** R28.
21. Hosack, D. A., Dennis, G., Jr., Sherman, B. T., Lane, H. C., and Lempicki, R. A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol.* **4:** R70.
22. Gadbois, D. M., Bradbury, E. M., and Lehnert, B. E. (1997). Control of radiation-induced G1 arrest by cell-substratum interactions. *Cancer Res.* **57:** 1151–1156.
23. Chang, H. Y., *et al.* (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc. Natl. Acad. Sci. USA* **99:** 12877–12882.
24. Iyer, V. R., *et al.* (1999). The transcriptional program in the response of human fibroblasts to serum. *Science* **283:** 83–87.
25. Chang, H. Y., *et al.* (2004). Gene expression signature in fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2:** E7.
26. 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.
27. Naldini, L., *et al.* (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272:** 263–267.

Copyright © The American Society of Gene Therapy