Mol. Cells *33*, 525-531, May 31, 2012
DOI/10.1007/s10059-012-0038-x

Molecules and Cells

Retroviral Infection of hES Cells Produces Random-like Integration Patterns

Kwang-il Lim[1,2,*]

Retroviral integration provides us with a powerful tool to realize prolonged gene expressions that are often critical to gene therapy. However, the perturbation of gene regulations in host cells by viral genome integration can lead to detrimental effects, yielding cancer. The oncogenic potential of retroviruses is linked to the preference of retroviruses to integrate into genomic regions that are enriched in gene regulatory elements. To better navigate the double-edged sword of retroviral integration we need to understand how retroviruses select their favored genomic loci during infections. In this study I showed that in addition to host proteins that tether retroviral pre-integration complexes to specific genomic regions, the epigenetic architecture of host genome might strongly affect retroviral integration patterns. Specifically, retroviruses showed their characteristic integration preference in differentiated somatic cells. In contrast, retroviral infections of hES cells, which are known to display decondensed chromatin, produced random-like integration patterns lacking of strong preference for regulatory-element-rich genomic regions. Better identification of the cellular and viral factors that determine retroviral integration patterns will facilitate the design of retroviral vectors for safer use in gene therapy.

INTRODUCTION

Retroviruses have been considered as efficient gene delivery vehicles based on their several advantageous characteristics. They can be engineered to carry large foreign genes (up to 8 kb) and induce only minor or no immune response. Most of all, retroviruses allow stable gene expression through integration of their genomes into host genomes. However, due to their oncogenic potential, as shown by the incidence of leukemia in a recent severe combined immune deficiency (SCID) clinical trial in Europe (Hacein-Bey-Abina et al., 2003), the therapeutic use of retroviruses has been strictly limited. The oncogenic propensity of retroviruses mainly stems from their tendency to selectively integrate into the transcriptional start sites (TSSs) of genes (Lewinski et al., 2006; Wu et al., 2003), including oncogenes, in host genomes. This integration preference can lead to the deregulation of cell growth through trans-activation of oncogenes by retroviral promoters, ultimately transforming infected cells into cancer cells.

In parallel with attempts to shift such risky retroviral integration patterns to safer ones by molecular engineering (Lim et al., 2010), there have been significant efforts to elucidate how retroviruses, including lentiviruses, select integration spots in the host genome. It has been suggested that interactions among retroviral pre-integration complexes (PICs) and cellular proteins, such as transcription factors (TFs), determine viral integration sites via a mechanism in which cellular proteins tether PICs to certain genomic loci (Engelman and Cherepanov, 2008; Felice et al., 2009). However, only a fraction of host proteins associating with retroviral PICs have been identified, and other mechanisms that underlie such interactions are still poorly understood (Cattoglio et al., 2010). Moreover, cellular physiological states and cell-specific features, both of which determine the gene expression patterns of host cells as well as the epigenetic landscape of the host genome, may affect retroviral and lentiviral integrations, but the relevant studies have begun to be available only recently (Bartholomae et al., 2011; Biasco et al., 2011).

In this work I investigated how the overall chromatin architecture of host cells affected retroviral and lentiviral integration patterns by comparing retroviral and lentiviral infections of human embryonic stem (hES) cells with those of a differentiated somatic cell line. I found that the previously reported characteristic retroviral integration preference for gene regulatory regions enriched in TSSs and CpG islands was mostly abolished in hES cells, in which chromatin is known to be decondensed and highly plastic (Ang et al., 2011; Efroni et al., 2008). In addition, the inherent lentiviral integration preference for regions within genes was significantly reduced in hES cells compared with that in the somatic cell line, in which chromatin would be relatively much more condensed. I showed that direct interactions among viral PICs and cellular proteins were not the only factors determining integration preference; cellular physiological states and epigenetic features associated with cellular differentiation and development also affected viral genome integration patterns in host cells. Identification of the main determinants, both direct and indirect, of retroviral integration and a comprehensive understanding of the mechanisms underlying integration preference will aid in the effort to engineer retroviruses as therapeu-

[1]Departments of Chemical Engineering and Bioengineering and The Helen Wills Neuroscience Institute, University of California, Berkeley, USA,
[2]Department of Medical and Pharmaceutical Sciences, College of Science, Sookmyung Women's University, Seoul 140-742, Korea
*Correspondence: klim@sm.ac.kr

Received February 6, 2012; revised March 11, 2012; accepted March 14, 2012; published online April 20, 2012

Keywords: gene therapy, human embryonic stem cells, oncogenic potential, overall chromatin structure, retroviral integration patterns

tics with less oncogenic potential.

MATERIALS AND METHODS

Cell culture

Human embryonic kidney (HEK) 293T cells were cultured in Iscove's modified Dulbecco's medium with 10% fetal bovine serum at 37°C and 5% CO_2. HSF6 human embryonic stem (hES) cells were kindly provided to the Schaffer lab by Susan Fisher at University of California at San Francisco. HSF6 cells were grown on a surface coated with hES cell-qualified Matrigel (BD Biosciences) in X-Vivo 10 Medium (Lonza) supplemented with 80 ng/ml FGF-2 (Peprotech) and 0.5 ng/ml TGF-β1 (R&D Systems). The preparation and coating with Matrigel was performed following the manufacturer's instructions. This culture system has previously been shown to support extended self-renewal of hES cells (Li et al., 2005; Zhang et al., 2006).

Virus packaging and infections

The retroviral and lentiviral plasmids, pCLPIT GFP and pFUGW, were used for virus vector packaging (Lois et al., 2002). For the retroviral vectors, pCLPIT GFP, pCMV gag-pol and pcDNA3 IVS VSV-G plasmids were introduced into the packaging cell line (HEK 293T cells) via calcium phosphate transfection. For the lentiviral vectors, pFUGW, pMDLg/pRRE, pRSV-Rev and pcDNA3 IVS VSV-G plasmids (Dull et al., 1998; Lois et al., 2002) were similarly introduced into HEK 293T cells via calcium phosphate transfection. Virus supernatant was twice harvested at two and three days post transfection and concentrated by ultracentrifugation. The concentrated viruses were used to infect cells at multiplicity of infection less than 1.

Mapping of viral integration sites

The relevant molecular biological preparation to isolate virus and host genome junctions was performed by following and modifying an established method (Wu et al., 2003). The genomic DNA of infected 293T cells and hES cells was isolated five to seven days post infection using the DNA Mini Kit (Qiagen) and then digested with either *MseI* or *HpyCH4III* [New England Biolabs (NEB)]. Digested genomic DNA fragments were ligated to pre-annealed Mse linker or Hpy linker DNA (Mse linker+, 5′-GTAATACGACTCACTATAGGGCTCCGCTT AAGGGAC-3′; Mse linker-, 5′-TAGTCCCTTAAGCGGAG-3′; Hpy linker+, 5′-GTAATACGACTCACTATAGGGCTCCGCTT AAGGGACN-3′; Hpy linker-, 5′-GTCCCTTAAGCGGAG-3′). The ligation products were used as templates for primary PCR with two primers annealing to the linkers and to either MLV LTR or HIV LTR (MLV-LTR, 5′-GTCTTGTGGTCTCGCTGTTCCT TGG-3′; HIV-LTR, 5′-AGTGCTTCAAGTAGTGTGTGCCCG-3′; linker primer, 5′-GTAATACGACTCACTATAGGGC-3′) under the following conditions: pre-incubation at 95°C for 2 min; 30 cycles each consisting of 30 s at 95°C, 30 s at 55°C and 2.5 min at 72°C; and a final extension step for 10 min at 72°C. The resulting primary PCR products were used for nested PCR with primers (MLV LTR nested, 5′-AAATTGGATCCGGTCTCCT CTGAGTGATTGACTACC-3′; HIV LTR nested, 5′-AAAAAG GATCCCCGTCTGTTGTGTGACTCTGGTAACT-3′; linker primer nested, 5′-AAATTAAGCTTAGGGCTCCGCTTAAGGGA C-3′) under the same conditions as the primary PCR. The PCR-amplified virus-host genome junctions were cloned into the pBS SK SP plasmid after restriction enzyme digestion with *Bam*HI and *Hin*dIII (NEB), and then sequenced with a primer (5′-CGCGTTGGCCGATTCATTAATG-3′). The host portions of the sequenced genome junctions were mapped to locations in the human genome, and the relevant genomic annotations were made using the GTSG-Quickmap tool (Appelt et al., 2009).

Generation of random integrations

Hypothetical random integration positions in the human genome were computational generated by GTSG-Quickmap tool, and the relevant genomic annotations were also made by the same tool (Appelt et al., 2009).

Statistical calculations

The P values for the frequency of integrations within specific genomic regions relative to the corresponding frequency of the hypothetical random integrations or to that of other viral infections were calculated by the chi-square test.

RESULTS

Retroviral vectors preferred to integrate into TSSs in the genomes of a somatic cell line

As observed in previous studies (Lewinski et al., 2006; Lim et al., 2010), retroviral vectors based on murine leukemia virus (MLV) strongly preferred TSSs and CpG islands for genome integration upon infection of somatic human embryonic kidney (HEK) 293T cells (see Table 1A for a comparison of this infection with a hypothetical random integration case; The numbers of analyzed integration events are mentioned in the table footnotes.). MLV vector integrations took place within 5 kb of TSSs 62.5% of the time and either inside or within 1kb of CpG islands 40.6% of the time (significance of 5.6×10^{-22} and 4.0×10^{-71}, respectively, compared with the random integration pattern). Such integration patterns can lead to perturbation of genetic circuits of cells, potentially causing dysregulation of cell growth, because the favored genomic regions are enriched in gene regulatory elements. However, MLV overall disfavored repeat regions as consistent with a recent study (Table 1A; Lim et al., 2010). The MLV vector only integrated into repeat regions 12.5% of the time, which was significantly lower than the 48.9% that was expected to occur by random chance. Long interspersed nuclear elements (LINEs) were especially not favored. Finally, the virus did not show any significant integration preference for the regions within genes (Table 1A), which is also consistent with the previous study (Lim et al., 2010).

The characteristic integration preference of retroviruses for TSSs and CpG islands was almost abolished in hES cells

In contrast with the above results in differentiated somatic cells, MLV vectors did not show a strong integration preference for any particular genomic region in hES cells (Table 1B). The viruses only displayed a moderate preference for the regions inside or within 5 kb of CpG islands (Table 1B). The observed level of integration preference for CpG islands, which was only 2.3-fold higher than what was expected by random chance, was dramatically less than that observed in infected 293T cells (9.1-fold higher than that expected by chance, Table 1A). In hES cells, the retrovirus was no longer biased against integration into genomic repeat regions at a significant level ($P < 0.05$) (Table 1B). These surprising shifts in preference for or against the particular genomic regions in hES cells suggest that development-specific characteristics of host cells, such as epigenetic landscape and chromatin organization, would affect retroviral integration patterns.

The integration preference of lentiviruses for the regions within genes was significantly reduced in hES cells

Lentiviral vectors that were based on human immunodeficiency

Table 1. Global integration pattern of the MLV-based vector in HEK 293T cells (A) and hES cells (B)

A

	Percent (sample)	Percent (random)	P values (compared with random)
Within genes	59.38	44.61	9.29×10^{-2}
Within exons of genes	6.25	2.48	1.70×10^{-1}
Within introns of genes	53.13	42.12	2.07×10^{-1}
Within ± 5 kb of the TSS	62.50	10.43	5.58×10^{-22}
Inside or within 1 kb of CpG	40.63	1.56	3.95×10^{-71}
Inside or within 5 kb of CpG	65.63	7.18	1.50×10^{-37}
LINE	3.13	22.08	9.74×10^{-3}
SINE	3.13	13.66	8.27×10^{-2}
DNA	0.00	3.38	2.90×10^{-1}
LTR	6.25	9.11	5.74×10^{-1}
Total repeats	12.50	48.87	3.86×10^{-5}

B

	Percent (sample)	Percent (random)	P values (compared with random)	P values (compared with the case of 293T)
Within genes	47.62	44.61	6.95×10^{-1}	1.21×10^{-1}
Within exons of genes	2.38	2.48	9.67×10^{-1}	3.00×10^{-1}
Within introns of genes	45.24	42.12	6.82×10^{-1}	3.05×10^{-1}
Within ± 5 kb of the TSS	19.05	10.43	6.77×10^{-2}	6.00×10^{-9}
Inside or within 1 kb of CpG	4.76	1.56	9.40×10^{-2}	2.21×10^{-6}
Inside or within 5 kb of CpG	16.67	7.18	1.72×10^{-2}	2.37×10^{-11}
LINE	19.05	22.08	6.36×10^{-1}	3.14×10^{-9}
SINE	4.76	13.66	9.31×10^{-2}	5.44×10^{-1}
DNA	2.38	3.38	7.20×10^{-1}	N.A.
LTR	9.52	9.11	9.26×10^{-1}	3.81×10^{-1}
Total repeats	35.71	48.87	8.81×10^{-2}	5.39×10^{-6}

The frequencies (%) of MLV integrations (sample) in infected HEK 293T cells (Table 1A, 32 integration events) and hES cells (Table 1B, 42 integration events), and one million hypothetical random integrations (random), within the indicated genomic regions, are shown. The relevant P values for the frequency of MLV integrations within the indicated genomic regions compared with the corresponding frequency of the hypothetical random integrations or with that of other viral infections were calculated by the chi-square test. In these calculations the numbers of viral integration events of samples were considered. The integration frequencies for samples that were significantly higher or lower than those with which they were compared are highlighted in gray ("Compared with random" indicates that samples were compared with the hypothetical random integrations; "Compared with the case of 293T" indicates that samples were compared with the infections of HEK 293T cells.). LINE and SINE indicate long and short interspersed nuclear elements, respectively. DNA and LTR indicate DNA transposon and long terminal repeat retrotransposon elements, respectively. N.A. denotes not available.

virus 1 (HIV-1) had a strong integration preference for the regions within genes in 293T cells as consistent with previously published results (Table 2A) (Cattoglio et al., 2010; Mitchell et al., 2004; Schroder et al., 2002). Compared with the events expected by random chance, lentiviral integrations occurred 1.9-fold more often (85.4% vs. 44.6%) in the regions within genes (significance of 1.3×10^{-8}) (Table 2A). Interestingly, lentiviral vectors preferred to integrate within introns rather than within exons (Table 2A). Additionally, there was a significantly higher integration preference for short interspersed nuclear elements (SINEs) than that expected by random chance (31.3% vs. 13.7%). However, unlike MLV in 293T cells, the lentiviral vectors did not show strong integration preferences for TSSs and CpG islands (Tables 1A and 2A) (Kim et al., 2008; Lim et al., 2010). Such a preference for integrating within genes is a characteristic of the lentiviruses.

This distinct lentiviral integration pattern was less significant in hES cells (Table 2B). The frequency of integration events within genes decreased from 85.4% in infected 293T cells to 65.3% in infected hES cells (significance of 8.6×10^{-5}). In addition, the lentiviral preference for SINEs, which was clearly detected in 293T cells, was not observed in hES cells. Except for significantly disfavoring the regions inside LTRs, the lentiviral integration pattern in hES cells was overall closer to the pattern of hypothetical random integrations than that in 293T cells (Tables 2A and 2B). This finding is similar to that observed above for MLV-based vectors (Tables 1A and 1B) and suggests that the unique integration patterns of retroviruses and lentiviruses favoring certain genomic regions would depend upon specific genome-modifications or chromatin organization patterns in differentiated cells. In the absence of these genomic features, some cellular factors would not be available for the PICs of retroviruses and lentiviruses to select their preferred host genomic spots during integrations.

Retroviral integrations were highly concentrated around TSSs in 293T cells, but not in hES cells

To more quantitatively analyze the potential effects of retroviral integration patterns on gene regulation I marked the retroviral integration spots in the host genomes using a relative length

Table 2. Global integration patterns of the HIV-1 based vector in HEK 293T cells (A) and hES cells (B)

A

	Percent (sample)	Percent (random)	P values (compared with random)	P values (compared with MLV)
Within genes	85.42	44.61	1.29×10^{-8}	2.40×10^{-4}
Within exons of genes	2.08	2.48	8.60×10^{-1}	2.33×10^{-1}
Within introns of genes	83.33	42.12	7.34×10^{-9}	2.75×10^{-5}
Within $\pm$ 5 kb of the TSS	4.17	10.43	1.56×10^{-1}	6.94×10^{-17}
Inside or within 1 kb of CpG	0.00	1.56	3.83×10^{-1}	9.96×10^{-9}
Inside or within 5 kb of CpG	6.25	7.18	8.03×10^{-1}	4.63×10^{-18}
LINE	18.75	22.08	5.78×10^{-1}	5.14×10^{-10}
SINE	31.25	13.66	3.87×10^{-4}	4.65×10^{-29}
DNA	2.08	3.38	6.19×10^{-1}	N.A.
LTR	4.17	9.11	2.34×10^{-1}	5.51×10^{-1}
Total repeats	56.25	48.87	3.06×10^{-1}	4.95×10^{-20}

B

	Percent (sample)	Percent (random)	P values (compared with random)	P values (compared with the case of 293T)
Within genes	65.31	44.61	3.56×10^{-3}	8.62×10^{-5}
Within exons of genes	2.04	2.48	8.43×10^{-1}	9.85×10^{-1}
Within introns of genes	63.27	42.12	2.73×10^{-3}	1.64×10^{-4}
Within $\pm$ 5 kb of the TSS	12.24	10.43	6.78×10^{-1}	4.69×10^{-3}
Inside or within 1 kb of CpG	0.00	1.56	3.78×10^{-1}	N.A.
Inside or within 5 kb of CpG	8.16	7.18	7.90×10^{-1}	5.80×10^{-1}
LINE	24.49	22.08	6.84×10^{-1}	3.03×10^{-1}
SINE	20.41	13.66	1.69×10^{-1}	1.02×10^{-1}
DNA	8.16	3.38	6.39×10^{-2}	2.85×10^{-3}
LTR	0.00	9.11	2.67×10^{-2}	1.44×10^{-1}
Total repeats	53.06	48.87	5.57×10^{-1}	6.53×10^{-1}

The frequencies (%) of HIV-1 integrations (sample) in infected HEK 293T cells (Table 2A, 48 integration events) and in hES cells (Table 2B, 49 integration events), and one million hypothetical random integrations (random), within the indicated genomic regions, are shown. The relevant P values for the frequency of HIV-1 integrations within the indicated genomic regions compared with the corresponding frequency of the hypothetical random integrations or with that of other viral infections were calculated by the chi-square test. In these calculations the numbers of viral integration events of samples were considered. The integration frequencies for samples that were significantly higher or lower than those with which they were compared are highlighted in gray ("Compared with MLV" indicates that samples were compared with the case of the MLV-infected HEK 293T cells). N.A. denotes not available.

scale of transcriptional units (Fig. 1). As observed in previous studies (Hematti et al., 2004), retroviral integration sites were highly concentrated in the regions near TSSs in 293T cells (Fig. 1A). This result indicates that retroviral PICs would be guided or tethered to TSSs, regions that are enriched in gene regulatory elements, through interactions with host proteins involved in gene regulation. In contrast, such a concentrated pattern completely disappeared in hES cells, alternatively forming a spreading-out integration pattern (Fig. 1B). This observed change in integration pattern suggests that the potential oncogenic risk of retroviral vectors will be greatly reduced when they are used to manipulate hES cells, as integrations into hES cells pose a decreased risk of perturbing the gene regulation that often occurs (via enhancer effects of viral genomes) when viral integrations take place near TSSs (Hacein-Bey-Abina et al., 2003; Kiem et al., 2010).

Genes that are involved in apoptosis, anti-apoptosis and cell cycle control were targeted by retroviruses and lentiviruses

To determine which cellular pathways or functions were affected by infection with retroviruses and lentiviruses I functionally clustered the genes harboring the viral genome, as well as genes in the vicinity of the viral integrants (Tables S1 and S2), using DAVID 6.7, a gene ontology tool (Huang da et al., 2009a; 2009b). It was noticeable in the retroviral infection of 293T cells that a fraction of the genes that were located near the retroviral integrants were involved in programmed cell death or apoptosis. These results are consistent with a previously published work (Cattoglio et al., 2007). My gene ontology analysis further showed that in hES cells, retroviral integrations targeted genes involved in cell cycle control. This finding suggests that retroviral infections may lead to dysregulation of cell growth and survival in both differentiated and undifferentiated cells. Retroviruses did not integrate directly into oncogenes in 293T cells or hES cells in my experiments, but they did integrate into the regions near oncogenes, such as DEK and C16orf75.

Interestingly, in 293T cells multiple genes that harbored the lentiviral integrants were involved in chromatin and chromosomal organizations. Similar results have previously been observed in HIV-infected hematopoietic progenitor cells (HPCs) (Cattoglio et al., 2010). In addition, lentiviral vectors integrated

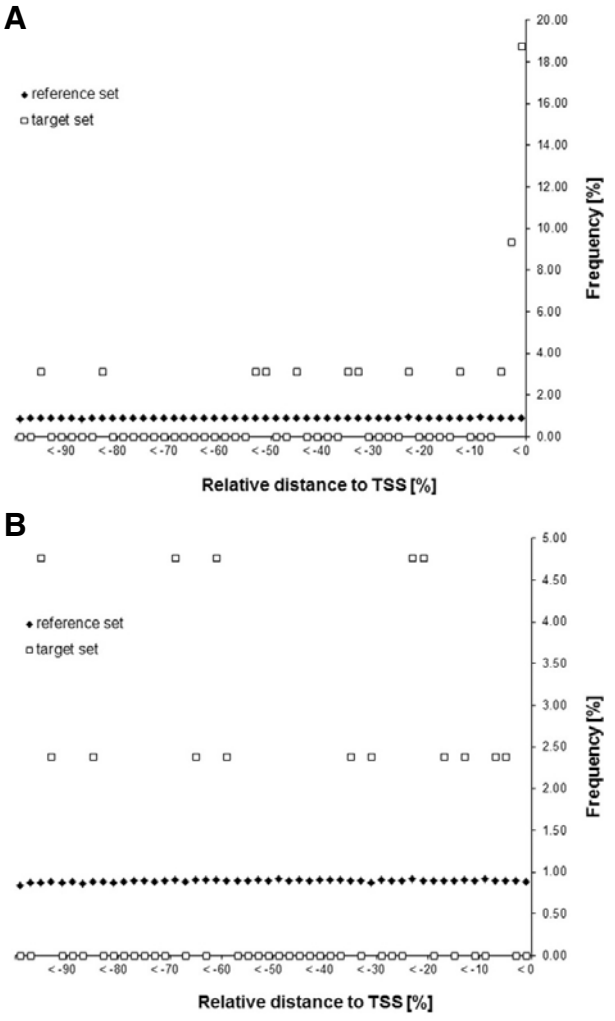

Fig. 1. The relative distance of retroviral integrants to transcriptional start sites (represented as percent of gene length). (A) Infected 293T cells. (B) Infected hES cells. The reference set and the target set denote hypothetical random integrations and experimental retroviral integrations, respectively. Given a range of the relative distances to TSS, the frequencies of genomic loci that harbored random integrations or viral integrations are shown.

into genes that were involved in apoptosis or programmed cell death, thereby potentially affecting cell survival, rather than integrating near the genes as shown in retroviral infections. Lentiviral vector infections may lead to both negative and positive regulation of apoptosis, through integrating near both relevant groups of genes. In infected hES cells the genes neighboring lentiviral integrants were involved in cell cycle control. In summary, both retroviral and lentiviral vectors can dysregulate cell growth and survival by integrating into or near genomic regions enriched in genes that are involved in such cellular processes. On the other hand, lentiviral vectors often integrated directly into oncogenes (as defined in the Sanger Cancer Gene Census), including BCL9 and NUP98 in 293T cells and TSC1, FNBP1, TCF3 and RUNX1 in hES cells. Whether integrating into oncogenes can reduce their potential oncogenic effects, through insertional disruption of the cancer genes, needs to be carefully assessed.

DISCUSSION

The potential genotoxicity arising from the characteristic integration preference of retroviruses for TSSs and CpG islands, both of which are enriched in gene regulatory elements, has limited the use of retroviruses in gene therapy applications. To engineer safer viral vectors for therapeutic use requires a better understanding of how viruses select integration loci in host genomes. Only a few cellular tethering molecules, including LEDGF/p75 and SWI/SNF (Engelman and Cherepanov, 2008; Kalpana et al., 1994), have been shown to guide the lentiviral PIC to specific genomic regions. A more comprehensive list of virus genome tethering molecules and other cell factors that also affect retroviral and lentiviral integration patterns needs to be determined.

In this work I showed that certain cellular environments could alter viral integration patterns. As observed in previous studies (Lewinski et al., 2006; Lim et al., 2010), MLV strongly preferred the genomic regions within or near TSSs and CpG islands in a differentiated cell line (Table 1A). In addition, MLV overall avoided repeat regions, such as LINEs and SINEs, consistent with previous results (Hematti et al., 2004). Surprisingly, the characteristic retroviral integration pattern almost completely disappeared in hES cells (Table 1B). In hES cells MLV did not show any significant preference for the regions within 5 kb of TSSs and inside or within 1 kb of CpG islands. The virus showed only a moderate level of integration preference for the regions within 5 kb of CpG islands, but this preference was 3.9-fold lower than that in 293T cells. Such a drastically reduced or abolished integration preference for the regions enriched in gene regulatory elements indicates a much lower oncogenic risk from the MLV infection of hES cells. The MLV integration bias against repeat regions was similarly abolished in infected hES cells (Table 1). Overall, the MLV integration pattern in hES cells was nearly statistically indistinguishable from a hypothetical random integration pattern.

What factors in hES cells are responsible for the shift in the integration pattern? Because strong integration preference of retroviruses for particular genomic regions is often consistently maintained independent of cell types when they are at a similar development stage (Mitchell et al., 2004), I propose that a cellular factor highly varying dependent upon the development stage, the genomic organization of host cells, made such a difference. Retroviruses selected broader regions of the genome in hES cells, which are known to have a decondensed form of genome. Such a genomic organization of hES cells is linked to their undifferentiated pluripotent state. hES cell chromatin has high plasticity, co-exists in open conformation, and undergoes hyperdynamic interactions with chromatin proteins (Ang et al., 2011; Fisher and Fisher, 2011; Mattout and Meshorer, 2010; Orkin and Hochedlinger, 2011). The chromatin of hES cells is dynamically associated with structural proteins having a much lower rigidity compared with that of differentiated somatic cells (Efroni et al., 2008; Jorgensen et al., 2007; Mattout and Meshorer, 2010). In addition, the lack of certain types of lamin A/C proteins and the structural features of nuclear lamina (dynamic rather than static) further contribute to the plasticity of the hES cell genome. These various genomic features of hES cells lead to a globally active gene transcription profile with expression of a more diverse set of genes (Efroni et al., 2008; Mattout and Meshorer, 2010). The open and hyperdynamic nature of the chromatin, unique in hES cells, would allow much broader genomic regions accessible, so that retroviruses could integrate into more diverse genomic loci having different features without strong integration propensity for the regions within or near TSSs

and CpG islands. The open genomic organization and resulting global transcription patterns of hES cells also promoted a higher expression of repeat regions than that observed in differentiated cells (Joffe et al., 2010), which have a different pattern of chromatin organization. This higher expression might be related to the increased frequency of retroviral integrations into the repeat regions in the hES cells, a level that was found to be close to that of random integrations (Table 1).

Similarly, a broader integration pattern was also observed in lentivirus-infected hES cells. As reported previously (Cattoglio et al., 2010), HIV-1 based lentiviral vectors strongly prefer to integrate within genes, especially within introns. Therefore, lentiviral integrations would lead to the dysregulation of genes by perturbing posttranscriptional processes, such as splicing, through integration into the introns of genes (Cattoglio et al., 2010) rather than by frameshifting protein coding through integration into the exons of genes. This strong integration preference was significantly reduced in hES cells (Table 2) from 85.4% to 65.3% (significance of 8.6×10^{-5}). Although lentiviral integration patterns were less changed than retroviral integration patterns, the alternation in integration preferences for both viruses demonstrates that the decondensed and open context of genome in hES cells can allow, in general, viral access to genomic loci with more diverse features. Different levels of dependence of viral integration patterns on the chromatin organization of host cells are expected, as MLV inherently favors open chromatin unlike lentiviruses (Lewinski et al., 2006).

In this study I showed that the well-known characteristic retroviral integration preferences are clearly observed only in differentiated somatic cells in which chromatin is tightly condensed and organized. These results suggest that in addition to interactions among viral PIC components and cellular proteins, the condensed genomic features of host cells are also necessary to yield such characteristic integration patterns of retroviruses. However, I cannot exclude the possibility that hES cell-specific molecules are responsible for the shift in retroviral integration patterns. Before retroviruses can be safely used for therapeutic applications, a better understanding of the factors underlying the selection of integration loci is needed.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

I thank David Schaffer for his support that was critical for me to realize my ideas in the course of this project. I also thank Phung Gip and Lauren Little in the Schaffer lab for their technical advice. This work was supported by the SRC Research Center for Women's Diseases of Sookmyung Women's University (2011). I declare that no conflict of interest exists.

REFERENCES

Ang, Y.S., Gaspar-Maia, A., Lemischka, I.R., and Bernstein, E. (2011). Stem cells and reprogramming: breaking the epigenetic barrier? Trends Pharmacol. Sci. *32*, 394-401.

Appelt, J.U., Giordano, F.A., Ecker, M., Roeder, I., Grund, N., Hotz-Wagenblatt, A., Opelz, G., Zeller, W.J., Allgayer, H., Fruehauf, S., et al. (2009). QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. Gene Ther. *16*, 885-893.

Bartholomae, C.C., Arens, A., Balaggan, K.S., Yanez-Munoz, R.J., Montini, E., Howe, S.J., Paruzynski, A., Korn, B., Appelt, J.U., Macneil, A., et al. (2011). Lentiviral vector integration profiles differ in rodent postmitotic tissues. Mol. Ther. *19*, 703-710.

Biasco, L., Ambrosi, A., Pellin, D., Bartholomae, C., Brigida, I., Roncarolo, M.G., Di Serio, C., von Kalle, C., Schmidt, M., and Aiuti, A. (2011). Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell. EMBO Mol. Med. *3*, 89-101.

Cattoglio, C., Facchini, G., Sartori, D., Antonelli, A., Miccio, A., Cassani, B., Schmidt, M., von Kalle, C., Howe, S., Thrasher, A.J., et al. (2007). Hot spots of retroviral integration in human CD34+ hematopoietic cells. Blood *110*, 1770-1778.

Cattoglio, C., Pellin, D., Rizzi, E., Maruggi, G., Corti, G., Miselli, F., Sartori, D., Guffanti, A., Di Serio, C., Ambrosi, A., et al. (2010). High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. Blood *116*, 5507-5517.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J. Virol. *72*, 8463-8471.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437-447.

Engelman, A., and Cherepanov, P. (2008). The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. PLoS Pathog. *4*, e1000046.

Felice, B., Cattoglio, C., Cittaro, D., Testa, A., Miccio, A., Ferrari, G., Luzi, L., Recchia, A., and Mavilio, F. (2009). Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS One *4*, e4571.

Fisher, C.L., and Fisher, A.G. (2011). Chromatin states in pluripotent, differentiated, and reprogrammed cells. Curr. Opin. Genet. Dev. *21*, 140-146.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415-419.

Hematti, P., Hong, B.K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I.E., Eckfeldt, C.E., Sharma, Y., Schmidt, M., et al. (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol. *2*, e423.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44-57.

Joffe, B., Leonhardt, H., and Solovei, I. (2010). Differentiation and large scale spatial organization of the genome. Curr. Opin. Genet. Dev. *20*, 562-569.

Jorgensen, H.F., Azuara, V., Amoils, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B.S., Ramsahoye, B., Merkenschlager, M., and Fisher, A.G. (2007). The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. Genome Biol. *8*, R169.

Kalpana, G.V., Marmon, S., Wang, W., Crabtree, G.R., and Goff, S.P. (1994). Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science *266*, 2002-2006.

Kiem, H.P., Ironside, C., Beard, B.C., and Trobridge, G.D. (2010). A retroviral vector common integration site between leupaxin and zinc finger protein 91 (ZFP91) observed in baboon hematopoietic repopulating cells. Exp. Hematol. *38*, 819-22, 822.e1-3.

Kim, S., Kim, N., Dong, B., Boren, D., Lee, S.A., Das Gupta, J., Gaughan, C., Klein, E.A., Lee, C., Silverman, R.H., et al. (2008). Integration site preference of xenotropic murine leukemia virus-related virus, a new human retrovirus associated with prostate cancer. J. Virol. *82*, 9964-9977.

Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS Pathog. *2*, e60.

Li, Y., Powell, S., Brunette, E., Lebkowski, J., and Mandalam, R. (2005). Expansion of human embryonic stem cells in defined serum-free medium devoid of animal-derived products. Biotechnol. Bioeng. *91*, 688-698.

Lim, K.I., Klimczak, R., Yu, J.H., and Schaffer, D.V. (2010). Specific insertions of zinc finger domains into Gag-Pol yield engineered

retroviral vectors with selective integration properties. Proc. Natl. Acad. Sci. USA *107*, 12475-12480.
Lois, C., Hong, E.J., Pease, S., Brown, E.J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science *295*, 868-872.
Mattout, A., and Meshorer, E. (2010). Chromatin plasticity and genome organization in pluripotent embryonic stem cells. Curr. Opin. Cell Biol. *22*, 334-341.
Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. *2*, E234.
Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell *145*, 835-850.
Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.
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.
Zhang, Y.W., Denham, J., and Thies, R.S. (2006). Oligodendrocyte progenitor cells derived from human embryonic stem cells express neurotrophic factors. Stem Cells Dev. *15*, 943-952.