High Efficiency Restriction Enzyme–Free Linear Amplification-Mediated Polymerase Chain Reaction Approach for Tracking Lentiviral Integration Sites Does Not Abrogate Retrieval Bias

Chuanfeng Wu,* Alexander Jares,* Thomas Winkler, Jianjun Xie, Jean-Yves Metais, and Cynthia E. Dunbar

Abstract

Retroviral vectors are an efficient and widely employed means of introducing an exogenous expression cassette into target cells. These vectors have been shown to integrate semi-randomly into the cellular genome, and can be associated with genotoxicity due to impact on expression of proximate genes. Therefore, efficient and accurate integration site analysis, while quantifying contributions of individual vector-containing clones, is desirable. Linear amplification-mediated polymerase chain reaction (LAM-PCR) is a widely used technique for identifying integrated proviral and host genomic DNA junctions. However, LAM-PCR is subject to selection bias inherent in the reliance of the assay on the presence of a restriction enzyme–cutting site adjacent to a retrievable integration site, and it is further limited by an inability to discriminate prior to sequencing between the flanking genomic DNA of interest and uninformative internal vector DNA. We report a modified restriction enzyme–free LAM-PCR (Re-free LAM-PCR) approach that is less time and labor intensive compared to conventional LAM-PCR, but in contrast to some other nonrestrictive methods, compares in efficiency and sensitivity, excludes retrieval of uninformative internal vector sequences, and allows retrieval of integration sites unbiased by the presence of nearby restriction sites. However, we report that Re-free LAM-PCR remains inaccurate for quantitation of the relative contributions of individual integration site–containing clones in a polyclonal setting, suggesting that bias in LAM-PCR retrieval of integration sites is not wholly explained by restriction enzyme–related factors.

Introduction

INTEGRATING GAMMARETROVIRUS and lentivirus-derived gene transfer vectors have been widely employed in order to introduce an expression cassette into target cells, allowing stable expression of genes for experimental and clinical gene therapy applications (Cavazzana-Calvo *et al.*, 2000; Cartier *et al.*, 2009; Boztug *et al.*, 2010; Hacein-Bey-Abina *et al.*, 2010; Kang *et al.*, 2010). These vectors have been shown to integrate semi-randomly into the cellular genome and can influence expression of genes up to 100 kb away via activity of vector-encoded enhancers, interference with normal mRNA splicing, or direct disruption of genes or transcriptional control elements (Nienhuis *et al.*, 2006; Dropulic, 2011; Trobridge, 2011). Clinical gene therapy studies using integrating vectors for treatment of X-linked conditions have achieved clinical improvement in some patients, but also adverse events due to

insertional activation of proto-oncogenes and clonal expansion of modified cell populations. Indeed, malignant or pre-malignant uncontrolled clonal expansions in patients were documented in X-linked severe combined immunodeficiency, Wiskott-Aldrich syndrome, and X-linked chronic granulomatous disease trials (Hacein-Bey-Abina *et al.*, 2003; Ott *et al.*, 2006; Howe *et al.*, 2008). Consequently, the ability to identify and track proviral integration sites in individual transduced clones is critical for assessing the risk of insertional mutagenesis, understanding and potentially avoiding genotoxicity (Aiuti *et al.*, 2007), and tracking transformed individual clones to answer biologic questions.

Several methods for identifying and tracking vector integration sites have been developed based on polymerase chain reaction (PCR) over the past two decades. We have summarized the different methods and discussed their efficiency, sensitivity, and biases in our previous review (Wu

Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health (NIH), Bethesda, MD 20892.

*Chuanfeng Wu and Alexander Jares contributed equally to this work.

and Dunbar, 2011). All integration site tracking methods rely on proceeding away from known sequences in the proviral integrated genome into unknown adjacent genomic DNA, isolating the junction fragment, amplifying it, and sequencing it. Linear amplification-mediated PCR (LAM-PCR) is a well-established and widely used technique for isolating and sequencing proviral and host genomic DNA junctions (Schmidt *et al.*, 2007; Schmidt *et al.*, 2009). However, LAM-PCR is subject to selection bias inherent in the assay's use of restriction enzymes (Harkey *et al.*, 2007; Gabriel *et al.*, 2009). Furthermore, due to differences in efficiency of ligation and amplification depending on fragment length and potentially chromatin characteristics, there is evidence that LAM-PCR is insufficiently quantitative to draw conclusions regarding the relative frequencies of individual clonal contributions to a population of cells (Harkey *et al.*, 2007; Gabriel *et al.*, 2009), short of very marked skewing, which can be confirmed by Southern blot or allele-specific PCR. The assay is further limited by an inability to remove internal vector-amplified DNA, despite the fact that the internal band is cut out during the DNA gel extraction procedure. Also, the labor-intensive and time-consuming nature of LAM-PCR provides opportunity for improvement if the method is to be widely adopted for longitudinal gene therapy studies. Several assays that do not employ restriction enzymes, flanking-sequence exponential anchored PCR (FLEA-PCR) (Pule *et al.*, 2008), non-restrictive linear amplification-mediated PCR (nrLAM-PCR) (Paruzynski *et al.*, 2010), and transposase MuA based-PCR (Brady *et al.*, 2011), have been developed. However, none of these methods have been proven to accurately quantitate clonal contributions in a polyclonal setting. The ideal integration site detection method is technically straightforward, detects integration site-specific sequences efficiently, and provides quantitative information on relative clonal contributions in the sample.

Here, we report a novel restriction enzyme-free LAM-PCR (Re-free LAM-PCR) method. Using a set of single integration site (single-copy) K562 clones transduced by an HIV-based lentivirus, we tested the efficiency, reliability, and sensitivity of the method and compared it to conventional LAM-PCR. We also investigate the nature of the integration site detection bias for each method in terms of adenine/thymine (A/T)-rich content for Re-free LAM-PCR and restriction enzyme sites for LAM-PCR, as well as the degree to which the methods are complementary in integrome coverage. We demonstrated that Re-free LAM-PCR effectively selects against internal vector sequences by using a blocking oligonucleotide-specific binding to the long terminal repeat (LTR)-proximal vector sequence. Re-free LAM-PCR is designed to be a less labor-intensive method, using less genomic DNA and no restriction enzyme digestion step. Re-free LAM-PCR should facilitate integration site analysis for assessing both retroviral safety and the biologic fate of transduced cells.

Materials and Methods

Lentiviral vector production and transduction of target cells

The HIV-derived replication-defective lentivirus vector pRRL.PPT.SF.IRES.GFP, which was modified from pRRL.PPT.SF.GFP (Schambach *et al.*, 2006) by including an internal ribosome entry site (IRES) sequence before the GFP

cassette, was used to produce lentiviral vectors as described (Hanawa *et al.*, 2002) via calcium phosphate transfection of vector and helper plasmids into 293T cells (Sigma-Aldrich, St. Louis, MO). K562 cells cultured in Roswell Park Memorial Institute (RPMI) 1640 medium plus 10% fetal bovine serum (FBS) were transduced by adding 6 μ g/ml polybrene (Millipore, Billerica, MA) and vector-containing virus supernatant at a multiplicity of infection of one to the cells, incubating for 16 hr, and then culturing for an additional 7 days before flow cytometric sorting for GFP expression.

Flow cytometry and cell sorting

Transduced K562 cells expressing low levels of GFP were sorted at single-cell frequency into a 96-well plate using a MoFlo Sorter (Cytomation, Carpinteria, CA). Individual clones were expanded and characterized. Flow cytometric analysis was performed with an LSR II instrument (Becton Dickinson, Franklin Lakes, NJ).

DNA extraction and southern blot analysis

Genomic DNA from single copy K562 clones and mixtures of clones was extracted using the DNeasy Blood & Tissue DNA Purification Kit according to the manufacturer's instructions (Qiagen, Valencia, CA), and quantified by NanoDrop (NanoDrop, Wilmington, DE). Ten micrograms of DNA from each clone was used for Southern blot analysis to characterize the copy number of integrated proviruses. Briefly, 10 μ g of genomic DNA was digested with Pci I (Thermo Fisher, Waltham, MA) and separated on a 0.8% Agarose gel. Pci I cuts once within the pRRL.PPT.SF.IRES. GFP vector sequence. Following the transfer to a nylon membrane, the DNA fragments were hybridized with a radio-labeled GFP cDNA probe generated by PCR from the original vector using primers specific to GFP (Supplementary Table 1; Supplementary Material available online at www.liebertonline.com/hum). The labeling reaction was performed using Amersham Ready-To-Go™ DNA Labeling Beads (GE Healthcare, Buckinghamshire, United Kingdom).

LAM-PCR

LAM-PCR was performed as previously described, with the primers and linker cassettes shown in Supplementary Table 1 (Schmidt *et al.*, 2007). One hundred nanograms of genomic DNA was linearly amplified using an HIV-3'-LTR-specific 5'-biotinylated primer. After the second strand synthesis by random priming, the DNA was digested with Tsp509I (TasI) (Thermo Fisher) and ligated to a linker cassette. Nested PCR was performed using HIV-3'-LTR-specific and linker-specific primers. The amplicons were purified from 2.5% low melting point agarose gels (NuSieve GTG, Cambrex, IA) using QIAGEN MiniElute Gel Extraction Kit (Qiagen) and cloned into pCR4-TOPO vector (Invitrogen, Carlsbad, CA) for sequencing with M13-primers using an ABI Prism Genetic Analyzer (Applied Biosystems, Foster City, CA). Spreadex gels (Elchrom Scientific, Cham, Switzerland) were used for analyzing the LAM-PCR products pattern in high resolution.

Modified restriction enzyme-free LAM-PCR

The workflow for this method is described in Figure 1, and all primers and oligonucleotides are listed in Supplementary

FIG. 1. Schematic of a linear amplification-mediated polymerase chain reaction (LAM-PCR) versus a restriction enzyme-free (Re-free) LAM-PCR. The Re-free LAM-PCR workflow does not include restriction enzyme digestion or linker ligation steps. Instead of a random hexanucleotide mix, Re-free LAM-PCR features a 5' known sequence (KS) extension at the 5' end of random hexamers. Blocking oligonucleotides specific to the long terminal repeat (LTR)-proximal vector sequence prevents the progression of the T7 DNA polymerase through the avidin-bound biotinylated LTR fragment, ensuring that internal vector sequences are not amplified in the subsequent exponential PCR.

Table 1. The linear PCR reaction was initiated from the HIV-3'-LTR-specific 5'-biotinylated primer identical to that used in standard LAM-PCR. A library of random hexamers preceded by a 38 bp 5' unique nonhomologous known sequence (KS) with no homology to the mammalian genome or to the vector was then used for complementary strand synthesis of the linear product (Integrated DNA Technologies, Coralville, IA). A blocking 25-mer oligonucleotide fully complementary to the LTR-proximal vector sequence was used to prevent amplification of internal vector sequences. T7 DNA polymerase, which lacks a 5' → 3' exonuclease domain but with 3' → 5' exonuclease activity approximately 1,000-fold greater than that of Klenow polymerase, was used for priming the linear PCR products to generate double-stranded DNA. Finally, exponential nested PCR amplification was performed using primers specific to the KS and primers specific to the vector LTR. Reaction mixture systems and thermocycler conditions are described in Supplementary Table 2.

Integration sites analysis and statistics analysis

Sequences were analyzed using DNASTAR SeqManII software (Madison, Wisconsin), scanning for the pCR4-TOPO vector, assembling sequences with lengths greater

than 100 base pairs, with a minimum match size of 50 bp, and a percent match requirement of 95%. The trimmed sequences were aligned to the human genome (Feb.2009 GRCh37/hg19) using the BLAT search server. Integration sites were considered valid if the vector-genome junction sequence was completely present and the flanking genomic region had a unique sequence match of ≥ 95%. Unreadable, very short sequences and sequences only having KS or LTR reads were considered uninformative sequences and were not further analyzed. Graphical and statistical analysis was done by using Prism 4 GraphPad Software (La Jolla, CA).

Results

Generation of a series of single HIV-based lentivirus-integrated cell clones

In order to assess the efficiency and ability to quantify clonal contributions of Re-free LAM-PCR, we generated a series of K562 clones with single proviral integrants (Table 1), as confirmed by Southern blot analysis, showing single integration bands of different sizes (Fig. 2a). Integration sites for selected clones were identified and confirmed as unique by Re-free LAM-PCR (Fig. 2b and Table 1) using 100 ng of starting genomic DNA. Due to random priming of the 3'

TABLE 1. UNIQUE INTEGRATION SITE COORDINATES FOR SINGLE-COPY K562 CLONES

Clone ID	Score	Start	End	Qsize	Identity	Chromosome	Strand	Start	End
D5	493	1	493	493	100.00%	7_gl000195_random	–	62544	63036
D13	191	1	194	194	99.50%	7	–	110518361	110518555
D31	144	1	144	147	100.00%	1	+	155325908	155326051
D33	151	1	151	151	100.00%	5	–	32154680	32154830
D34	256	1	258	258	99.70%	2	+	78757576	78757833
D36	121	1	125	130	99.20%	21	+	48076128	48076254
D39	276	1	278	278	99.70%	5	–	887525	887802
D40	143	1	145	148	99.40%	7	+	11319076	11319220
D41	301	1	303	303	100.00%	19	+	40335360	40335664
D47	145	1	145	151	100.00%	11	+	75767866	75768010

hexamer of the KS anchor at varying distances from the LTR junction, we expected to generate multiple PCR products of different lengths, even from individual clones with single integrants. Instead of observing a continuous “smear” pattern when running the second exponential PCR product on a gel, a heterogeneous pattern of multiple discrete bands was observed (Fig. 2b). For instance, Re-free LAM-PCR on clone D31 resulted in discrete bands corresponding to different lengths from the LTR junction to the hexamer priming site, ranging from 50 bp to 162 bp (Fig. 2c). Using D31 clone genomic DNA alone as a starting template for Re-free LAM-PCR, by increasing incrementally the amount of DNA from 10 ng to 1 μ g, we observed more of a “smear” pattern in the 100–500 bp range, visualized on a Spreadex gel (Fig. 2d, left). For standard LAM-PCR, increased starting genomic DNA

led to an increased nonspecific, interband “smear” effect in the final nested PCR product (Fig. 2d, right).

Re-free LAM-PCR has comparable efficiency to regular LAM-PCR using relatively low amount of starting genomic DNA

We evaluated Re-free LAM-PCR integration site retrieval efficiency and the method’s accuracy in quantitatively tracking individual clonal contributions to a clonal mixture. We performed both Re-free LAM-PCR and traditional LAM-PCR on a mixture of equal amounts of DNA extracted from 10 K562 clones, each with a single, clone-specific integration site; this polyclonal sample is referred to as the 10 DNA mixture. In parallel, we also performed Re-free LAM-PCR

FIG. 2. Re-free LAM-PCR on single-copy K562 clones. **(a)** Ten K562 clones with single unique lentiviral integration site were identified by Southern blot following digestion with *Pci*I and hybridization with a green fluorescent protein (GFP) probe. **(b)** Re-free LAM-PCR products from the single-copy integrant clones run on an agarose gel. **(c)** Re-free LAM-PCR product lengths vary for the same clonal integration site, shown for one clone (D31) as a representative example, suggesting that KS-hexamer annealing occurs at varying distances from the LTR. **(d)** After the second exponential PCR step, different amounts of D31 DNA template showed different patterns of band amplification, with more of a band “smear” pattern for higher amounts of starting DNA (d, left). Restriction enzyme-based LAM-PCR band patterns were consistent across 500, 100, and 10 ng DNA starting amounts; 1 μ g DNA led to some nonspecific “smear” between bands (d, right).

FIG. 3. Re-free LAM-PCR versus LAM-PCR for clonal detection in a polyclonal setting. Re-free LAM-PCR and LAM-PCR were performed using a DNA mixture with equal DNA contributions from 10 K562 single-copy clones (10% each). Integration site detection was not uniform for either the single reaction on 100-ng genomic DNA or a pooled set of triplicate 100-ng samples for either methodology (a and c). Re-free LAM-PCR had lower frequencies of amplification of internal vector sequences and uninformative sequences as compared to LAM-PCR. Increased reaction number didn't increase the efficiency of the integration site detection, and the D13 integration site was not detected in all runs (a). In a second set of experiments (b and d), we mixed the 10 K562 single-copy clones in equal cell ratios prior to DNA extraction, which could be a better model of *in vivo* polyclonality and avoids issues with uniformly mixing genomic DNA.

and traditional LAM-PCR on DNA extracted after mixing the 10 cell populations in equal amounts; this polyclonal sample is referred to as the 10 cell mixture, and is postulated to most accurately reflect the clinical sample (Supplementary Table 3).

Starting with a relatively small amount of genomic DNA (100 ng), 96 TOPO4 TA clones were sequenced for each LAM-PCR or Re-free LAM-PCR reaction, whether DNA mixture or cell mixture. For each assay, we performed single reaction runs and triple reaction runs, the latter pooling three independent replicates before the first nested PCR. Integration site recovery analysis for both methods was summarized in Supplementary Table 3. In the cell mixture experiments, which model clinical settings more closely, the average recovery efficiency across single reactions was $65.00\% \pm 5.0\%$ (mean $\pm$ SEM, $n=2$) for Re-free LAM-PCR and $70\% \pm 0\%$, ($n=2$) for regular LAM-PCR. This difference was not statistically significant ($p=0.4226$). In the DNA mixture experiments, the average recovery efficiency across single reactions was $85.00\% \pm 5.0\%$ ($n=2$) for Re-free LAM-PCR and $55.00\% \pm 5.0\%$ ($n=2$) for regular LAM-PCR, and likewise, the difference was not found to be statistically significant, although there was a strong trend toward increased efficiency for Re-free PCR ($p=0.0513$). In order to compare the two methods, we combined the data across DNA and cell mixtures: average single reaction recovery efficiency of the 10 single integrant clones was $75.00\% \pm 6.455\%$ (mean $\pm$ SEM, $n=4$) for Re-free LAM-PCR and $62.50\% \pm 4.787\%$ (mean $\pm$ SEM, $n=4$) for regular LAM-PCR. Although the average of the clonal detection efficiency for Re-free LAM-PCR was higher than for regular LAM-PCR, the difference between

these two methods was not statistically significant ($p=0.1708$). Furthermore, no internal vector sequences were detected in the Re-free LAM-PCR method (Fig. 3a and b; Supplementary Table 3), suggesting that the blocking oligonucleotides were binding to the LTR-proximal internal vector sequence during the priming step, preventing the progression of the T7 polymerase during complementary strand synthesis, and consequently efficiently blocking amplification of internal vector sequences.

For Re-free LAM-PCR, we found that performing the reactions in triplicate did not increase clone-specific integration site detection ($75.00\% \pm 5\%$ mean $\pm$ SEM, $n=2$) compared to a single reaction system ($75.00\% \pm 6.455\%$, $n=4$). For both single and triple Re-free LAM-PCR reaction systems, no amplification of internal vector sequences was observed. For regular LAM-PCR, pooling three replicates at the first nested PCR stage results in the retrieval of 9/10 ($90.00\% \pm 0.00\%$, $n=2$) clone-specific integration sites (Fig. 3c and d), which corresponds to significantly ($p=0.0187$) better recovery efficiency than the average single reaction system ($62.50\% \pm 4.787\%$, $n=4$).

Re-free LAM-PCR and regular LAM-PCR are not quantitative

Contrary to our expectations, the frequency of integration site retrieval using Re-free LAM-PCR did not reflect the equal contribution of each clone in the starting DNA pool. We predicted 10% retrieval for each clone-specific integration site, which would indicate unbiased, uniform detection of proviral integrants. However, individual integration site

FIG. 4. A/T content analysis for the sequence of clone D13 and D41. Analysis of A/T content reveals that the D13 clone is high in A/T content (80% A/T in the 195 bp between LTR junction and the KS) and contains 6 Tsp509I restriction sites (a), whereas the flanking genomic DNA for D41 integration site has moderate levels of A/T content (53.2%) and contains no Tsp509I restriction site within 305 bp of the integration site (b).

detection frequencies ranged from 0% to 27% of retrieved sequences from the 10 clone DNA mixture sample (Fig. 3a), and from 0% to 62.5% of retrieved sequences from the 10 clone cell mixture sample (Fig. 3b). The results indicated that Re-free LAM-PCR did not accurately quantify relative clonal contributions in a polyclonal setting. As expected, regular LAM-PCR did not quantify relative clonal contributions either (Fig. 3c and d). Additionally, uninformative sequences were far more frequent ($p=0.0006$) in conventional LAM-PCR-based sequencing ($36.81\% \pm 3.894\%$, mean $\pm$ SEM, $n=6$) than for Re-free LAM-PCR ($17.01\% \pm 2.603\%$, $n=6$). Internal vector sequences were also more prevalent ($p<0.001$) in LAM-PCR sequencing results ($14.41\% \pm 0.626\%$, $n=6$) than for Re-free LAM-PCR, where internal vector sequences were absent.

In both the DNA mixture and cell mixture Re-free LAM-PCRs, the K562 D13 clone was not found, but the D41 clone was consistently found at higher than expected frequency (Fig. 3a and b). Re-free LAM-PCR appears to be efficient for qualitative integration sites detection, generally allowing identification of the majority of clones contributing, but detection frequencies do not reflect actual quantitative clonal contributions, with evidence of bias as some integration sites were consistently detected more frequently, whereas one integration site could not be detected in the clonal mixture despite multiple replicates. The D41 clone retrieved at high frequency by Re-free LAM-PCR was never detected by LAM-PCR (Fig. 3c and d).

Combining Re-free LAM-PCR and LAM-PCR allows for broader coverage of the integrome if enough DNA sample is available

Neither standard LAM-PCR performed with a single enzyme Tsp509I nor Re-free LAM-PCR detected the individual integration site from all 10 K562 clones from a polyclonal mixture. LAM-PCR did not detect clone D41 and Re-free LAM-PCR did not detect clone D13 in the context of a mixture of clones, despite being retrieved by Re-free LAM-PCR on D13 DNA alone (Fig. 3). We sought to determine whether the D41 and D13 detection patterns observed in Re-free LAM-PCR and LAM-PCR in a 10 clone mixture setting were caused by subthreshold amplification of flanking genomic DNA or by issues with genomic access. We reduced the number of clones in the mixture from 10 to 5 clones and performed Re-free LAM-PCR and LAM-PCR (Supplementary Fig. 1a and b). Clone D13 was also not detected in Re-free LAM-PCR in the 5 clone mixture.

From the previous successful D13-only Re-free LAM-PCR used to initially identify the integration site in this clone, analysis of the 195 bp flanking genomic sequence from the LTR to the KS-hexamer priming site revealed high A/T content: 80% in this D13 clone (Fig. 4a). For clone D41, still not detected in the 5 clone mixture by LAM-PCR, further

FIG. 5. Tsp509I restriction sites and A/T content analysis of all the 10 clones. Distance between the integration site and the 5' Tsp509I restriction sites within 500 bp in the genomic (a) may link to limited LAM-PCR access to the sequence flanking the provirus (b), detection frequencies summarized from all runs of LAM-PCR. Increased A/T content of flanking genomic sequence is correlated with decreased frequency of integration site detection in Re-free LAM-PCR (c).

FIG. 6. Sensitivity assay for Re-free LAM-PCR compared with regular LAM-PCR. **(a)** A serial dilution of 10 clone cell DNA against a nontransduced K562 DNA background revealed different patterns of flanking genomic DNA amplification by Re-free LAM-PCR. Different second-nested PCR product lengths were visualized on a Spreadex gel. A sharp drop-off of PCR product was observed at 0.1% of transduced-cell DNA. Whereas 100% transduced-cell DNA mapped 7/10 clone-specific integration sites, 1% transduced cell DNA mapped less: 4/10. **(b)** The same serial dilution repeated with LAM-PCR control yielded a similar drop-off of PCR product at 0.1% of transduced-cell DNA, however, there was less variation in flanking genomic DNA amplification patterns at $\geq 1\%$ transduction level, and the LAM-PCR specific internal vector band was present. For LAM-PCR, whereas 100% transduced-cell DNA mapped 7/10 clone-specific integration sites, 1% transduced-cell DNA mapped less: 5/10. In an artificial "clonal dominance" model, a serial dilution of clone D33 DNA resulted in consistent clone-specific integration site detection at transduced-DNA content $\geq 1\%$ for both Re-free LAM-PCR **(c)** and control LAM-PCR **(d)**. D33 flanking genomic sequences account for the vast majority of the sequencing data for both methods (panels c and d).

analysis of the flanking genomic sequence revealed that there was no 5' Tsp509I restriction site within 305 bp from the LTR, and the nearest 5' Tsp509I was found at 474 bp away from the integration site (Figs. 4b and 5a). However, combined sequencing results from LAM-PCR and Re-free LAM-PCR detected all the clones in our artificial polyclonal setting. It is possible that combining these two approaches may provide broader retrieval of the integrome in polyclonal samples.

Re-free LAM-PCR selection bias may be linked to A/T-content of the flanking genomic sequence

Analysis of the LTR-proximal 5' flanking genomic sequence revealed that if the Tsp509I restriction site was further upstream, the integration site would be less likely detected by regular LAM-PCR (Fig. 5a and b). Conversely, for Re-free LAM-PCR, increased A/T content of an integration site's 5' flanking genomic sequence is correlated with decreased detection of the site (Fig. 5c). Four out of five clones (D13, D34, D39, and D40) that contain more than 60% A/T content were detected at the lowest frequency (0%–2.6%) (Fig. 5c).

Re-free LAM-PCR detects integration sites in samples containing as little as 1% transduced cells

The sensitivity of the method was studied by serially diluting 10 clone cell mixture DNA into a nontransduced K562 cell DNA background, while keeping total genomic DNA at 100 ng for the initial first linear PCR. For the purposes of our study, we adjusted the percentage of transduced cells in a range from 100% to 0.1% (shown in Fig. 6). For both Re-free LAM-PCR and LAM-PCR, there was an observed steep drop-off of second nested PCR product from 1% transduced cells to 0.1%, as visualized on a Spreadex gel by SYBR Green staining (Invitrogen).

Sequencing analysis for both methods was conducted by obtaining 96 sequences for each sample run. The Re-free LAM-PCR data revealed that the recovery efficiency was decreased to 40% in the sample containing 1% transduced 10 cell mixture DNA compared to 70% recovery in the 100% transduced sample (Fig. 6a). For regular LAM-PCR, the equivalent drop of recovery efficiency was 70% to 50% (Fig. 6b). Consequently, we observed similar sensitivity for both Re-free LAM-PCR and LAM-PCR. The results indicated that both methods (starting with 100 ng genomic DNA) detect integration sites in samples where as few as 1% of cells are transduced; however, polyclonal integrome coverage was markedly decreased if the marking level falls below that threshold.

Given concerns regarding clonal expansion in gene therapy applications, we sought to test Re-free LAM-PCR's ability to detect a "dominant clone," which would present as an increasing share of the clinical cell sample and would be reflected in the DNA extracted from the sample. Since the D33 clone appeared to be detected by both Re-free LAM-PCR and regular LAM-PCR, we performed the same serial dilution as for the 10 clone sensitivity assay, with the D33 DNA fraction varying from 100% to 0.1% of the starting genomic DNA sample, against a backdrop of nontransduced K562 DNA. Using LAM-PCR and Re-free LAM-PCR, the goal was to determine at which point a clone can be deemed dominant by looking at sequencing data as a function of graduated

clonal frequencies. As shown in the Figures 6c and d, the lowest amount of D33 transduced clone that allows for detection by either LAM-PCR or Re-free LAM-PCR was 1 ng, corresponding to the 1% of the 100-ng starting DNA sample.

Discussion

Following transplantation of transduced cells, quantitative assessment of the clonal contributions of individual cells and their progeny, based on identification of unique proviral integration sites can provide important insights into both the biology and the safety of diverse gene transfer and gene therapy applications utilizing integrating vectors. For instance, detecting clonal expansion over time may predict genotoxicity, and comparison of clonal contributions to different hematopoietic lineages can confirm stem cell transduction or help clarify hematopoietic ontogeny (Dunbar, 2007). Accurate identification of all vector integration sites is also required for screening potential "safe harbor" genetic modification of patient-specific induced pluripotent stem cells for regenerative medicine applications (Papapetrou *et al.*, 2011).

A key goal of integration sites mapping is informative and efficient identification of all genomic integrants using the smallest amount of starting DNA, as availability of clinical DNA samples may be limited and repeated sampling impossible (Bleier *et al.*, 2008). LAM-PCR relies on adjacent restriction enzyme sites in order to access genomic insertions, since some integration sites occur too close or too far from any specific restriction enzyme site, resulting in fragments that are too small to resolve, or alternatively, too long to be amplified, thus limiting the analysis to a subset of clones in a mixture (Harkey *et al.*, 2007; Gabriel *et al.*, 2009). In a "multiarm" approach, a combination of the five most potent four-cutter restriction enzymes gives access to 88.7% of the analyzable genome, however, performing LAM-PCR with five different enzymes is more labor-intensive and impractical in terms of the large amounts of DNA required (Gabriel *et al.*, 2009). Indeed, a multiarm approach that pools X amount of enzymatic reactions would require X times starting genomic DNA, depleting limited amounts of clinical material.

Without the restriction enzyme digestion step, and with an abridged priming step prior to nested PCR, our Re-free LAM-PCR requires less labor, time, and less expense, in addition to being more DNA-efficient. A single reaction of Re-free LAM-PCR, with a starting DNA amount of 100 ng, is capable of reaching up to 90% efficiency in detecting clones in a 10-clone test polyclonal setting. In contrast, three pooled reactions, each with 100 ng of starting DNA, are required to attain up to 90% efficiency by regular LAM-PCR on the same clonal mixture. Several other approaches without restriction enzyme digest have been shown to access the integrome with a higher reported quantity of starting genomic DNA, for example, up to 1 μ g (Pule *et al.*, 2008; Paruzynski *et al.*, 2010). Re-free LAM-PCR allows for high genomic coverage and retrieval efficiency with a comparably low amount of starting genomic DNA. Sensitivity assay data indicates that Re-free LAM-PCR can track clone-specific integration sites at transduction levels as low as 1%, comparable to LAM-PCR.

In LAM-PCR, even after the excision of the gel band corresponding to amplified internal vector, undesirable internal vector sequences amounted to about 16% of the

shotgun-sequencing product. On the other hand, no internal vector sequences were detected in the final Re-free LAM-PCR shotgun sequencing product. These results suggest that Re-free LAM-PCR–blocking oligonucleotides, specific to the LTR-proximal vector sequence, select effectively against subsequent nested PCR amplification of internal sequences. As shotgun sequencing is replaced in modern molecular biology applications, future high-throughput sequencing of Re-free LAM-PCR–nested PCR products will be enhanced by the absence of internal sequences, making longitudinal *in vivo* follow-ups more informative, as only a limited amount of sample DNA is often available.

To our surprise and disappointment, Re-free LAM-PCR did not provide accurate quantitative information on clonal contributions, suggesting that integration site detection bias is not solely the result of restriction enzyme-related factors in terms of distance from restriction enzyme sites or efficiency of digestion (Harkey *et al.*, 2007). However, Re-free LAM-PCR was able to detect a clonal integration site (D41) that was not accessible to LAM-PCR on repeated runs, due to the lack of an LTR-proximal Tsp509I restriction site. Indeed, previous studies (Harkey *et al.*, 2007) have determined that while the Tsp509I AA|TT restriction motif is the most widely distributed and efficient, it still results in 10% of the genome being inaccessible to LAM-PCR–based integration site retrieval. Since the D13 clonal integration site, located in an A/T rich region and undetected by Re-free LAM-PCR in the mixture samples, is readily accessible via LAM-PCR, our results suggested that both methods present distinct biases, which prevent the detection of potential integration sites of interest. In the past, increasing the number of LAM-PCR repeats, and using various restriction enzymes, a laborious and time-consuming process, achieved increased integration site detection. As an alternative, we suggest instead performing one Re-free LAM-PCR run and one regular LAM-PCR run, each with 100ng starting genomic DNA, as a means of increasing recovery efficiency. If less DNA is available, Re-free LAM-PCR provides a labor-saving means of mapping qualitatively the majority of integrants, retrieving around 75% of total integration sites. Re-free LAM-PCR and LAM-PCR combined can provide for complementary and presumably more complete genomic coverage in situations where more sample DNA is available.

Furthermore, Re-free LAM-PCR efficiency and quantitative potential may be feasible with improvements in polymerase technology, allowing access and efficient priming and extension across a wider range of GC- and AT-rich templates and amplicons, as well as improved tolerance to common PCR inhibitors, possibly leading to fuller genomic access in nonrestriction enzyme qualitative integration sites detection. We noticed that clone D47 was also significantly under-represented in Re-free LAM-PCR analyses. However, the A/T content for the 250 bps surrounding the D47 integration site is a moderate 58.4%. This observation suggests that factors beyond A/T content, such as flanking DNA secondary structure motifs, could play a role in restricting access to the integrome.

Of the 10 single copy K562 clones, D13 and D40 were located on chromosome 7, while clones D33 and D39 were both located on chromosome 5. In a genome-wide analysis of lentiviral integration sites using next generation sequencing technology, chromosomes 7 and 5 were found to be over-

represented as sites of lentivector integration in tetraploid K562 cells compared to control 293T cells (Ustek *et al.*, 2012). In our study, D13, D40, and D39 clones were under-represented, whereas D33 can be easily detected in the Re-free LAM-PCR method. Since we used tetraploid karyotype-abnormal K562 cells, the possibility that chromosomal integration preferences would differ from normal primary cells is conceivable, as is an impact on retrieval using Re-free LAM-PCR.

Genetic engineering of the hematopoietic stem cell (HSC) provides a clinically accessible paradigm that has broader implications for stem cell engineering as a whole. With widespread awareness of the risks of certain aspects of vector design, such as LTR-driven transgene expression, current second- and third-generation vectors appear to be safer. Nonetheless, since lentiviral vectors still have not been targeted to specific loci or genomic regions (Riviere *et al.*, 2012), there is a key need for both quantitative (e.g., clonal dominance) and qualitative integration sites information. The recent development of HSC barcoding with a unique proviral barcode per transplanted cell (Lu *et al.*, 2011) constitutes a promising avenue for quantification of repopulating clones for real-time monitoring of clonality. Given that sequencing flanking genomic DNA at the LTR junction remains necessary to ascertain the medium- to long-term implications of a proviral integration (Glimm *et al.*, 2011), vector design that includes a barcode with the transgene could be complemented by Re-free LAM-PCR–based sequencing of the flanking genomic DNA of an integration site in order to determine whether a clone's biology had been altered due to the proviral location in the genome.

Acknowledgments

This research was supported by the Intramural Research Programs of the National Heart, Lung, and Blood Institute, National Institutes of Health, and the NIH Center for Regenerative Medicine (NIH CRM) funding. We thank Leigh Samsel and the staff in the Flow Cytometry Core Facility at NHLBI for their assistance.

Author Disclosure Statement

The authors declare no competing financial interests.

References

- Aiuti, A., Cassani, B., Andolfi, G., *et al.* (2007). Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. *J. Clin. Invest.* 117, 2233–2240.
- Bleier, S., Maier, P., Allgayer, H., *et al.* (2008). Multiple displacement amplification enables large-scale clonal analysis following retroviral gene therapy. *J. Virol.* 82, 2448–2455.
- Boztug, K., Schmidt, M., Schwarzer, A., *et al.* (2010). Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N. Engl. J. Med.* 363, 1918–1927.
- Brady, T., Roth, S.L., Malani, N., *et al.* (2011). A method to sequence and quantify DNA integration for monitoring outcome in gene therapy. *Nucleic Acids Res.* 39, e72.
- Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., *et al.* (2009). Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* 326, 818–823.

- Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., *et al.* (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 288, 669–672.
- Dropulic, B. (2011). Lentiviral vectors: their molecular design, safety, and use in laboratory and preclinical research. *Hum. Gene Ther.* 22, 649–657.
- Dunbar, C. E. (2007). The yin and yang of stem cell gene therapy: insights into hematopoiesis, leukemogenesis, and gene therapy safety. *Hematology Am. Soc. Hematol. Educ. Program* 2007, 460–465.
- Gabriel, R., Eckenberg, R., Paruzynski, A., *et al.* (2009). Comprehensive genomic access to vector integration in clinical gene therapy. *Nat. Med.* 15, 1431–1436.
- Glimm, H., Ball, C.R., and von Kalle, C. (2011). You can count on this: barcoded hematopoietic stem cells. *Cell. Stem Cell* 9, 390–392.
- Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302, 415–419.
- Hacein-Bey-Abina, S., Hauer, J., Lim, A., *et al.* (2010). Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 363, 355–364.
- Hanawa, H., Kelly, P.F., Nathwani, A.C., *et al.* (2002). Comparison of various envelope proteins for their ability to pseudotype lentiviral vectors and transduce primitive hematopoietic cells from human blood. *Mol. Ther.* 5, 242–251.
- Harkey, M.A., Kaul, R., Jacobs, M.A., *et al.* (2007). Multiarm high-throughput integration site detection: limitations of LAM-PCR technology and optimization for clonal analysis. *Stem Cells Dev.* 16, 381–392.
- Howe, S.J., Mansour, M.R., Schwarzwaelder, K., *et al.* (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* 118, 3143–3150.
- Kang, E.M., Choi, U., Theobald, N., *et al.* (2010). Retrovirus gene therapy for X-linked chronic granulomatous disease can achieve stable long-term correction of oxidase activity in peripheral blood neutrophils. *Blood* 115, 783–791.
- Lu, R., Neff, N.F., Quake, S.R., and Weissman, I.L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* 29, 928–933.
- Nienhuis, A.W., Dunbar, C.E., and Sorrentino, B.P. (2006). Genotoxicity of retroviral integration in hematopoietic cells. *Mol. Ther.* 13, 1031–1049.
- Ott, M.G., Schmidt, M., Schwarzwaelder, K., *et al.* (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat. Med.* 12, 401–409.
- Papapetrou, E.P., Lee, G., Malani, N., *et al.* (2011). Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. *Nat. Biotechnol.* 29, 73–78.
- Paruzynski, A., Arens, A., Gabriel, R., *et al.* (2010). Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nat. Protoc.* 5, 1379–1395.
- Pule, M.A., Rousseau, A., Vera, J., *et al.* (2008). Flanking-sequence exponential anchored-polymerase chain reaction amplification: a sensitive and highly specific method for detecting retroviral integrant-host-junction sequences. *Cytotherapy* 10, 526–539.
- Riviere, I., Dunbar, C.E., and Sadelain, M. (2012). Hematopoietic stem cell engineering at a crossroads. *Blood* 119, 1107–1116.
- Schambach, A., Bohne, J., Chandra, S., *et al.* (2006). Equal potency of gammaretroviral and lentiviral SIN vectors for expression of O6-methylguanine-DNA methyltransferase in hematopoietic cells. *Mol. Ther.* 13, 391–400.
- Schmidt, M., Schwarzwaelder, K., Bartholomae, C., *et al.* (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat. Methods* 4, 1051–1057.
- Schmidt, M., Schwarzwaelder, K., Bartholomae, C.C., *et al.* (2009). Detection of retroviral integration sites by linear amplification-mediated PCR and tracking of individual integration clones in different samples. *Methods Mol. Biol.* 506, 363–372.
- Trobridge, G.D. (2011). Genotoxicity of retroviral hematopoietic stem cell gene therapy. *Expert Opin. Biol. Ther.* 11, 581–593.
- Ustek, D., Sirma, S., Gumus, E., *et al.* (2012). A genome-wide analysis of lentivector integration sites using targeted sequence capture and next generation sequencing technology. *Infect. Genet. Evol.* 12, 1349–1354.
- Wu, C., and Dunbar, C.E. (2011). Stem cell gene therapy: the risks of insertional mutagenesis and approaches to minimize genotoxicity. *Front Med.* 5, 356–371.

Address correspondence to:
Dr. Cynthia E. Dunbar
CRC-Building 10, Room 4E-5132
National Institutes of Health
10 Center Drive
Bethesda, MD 20852

E-mail: dunbarc@nhlbi.nih.gov

Received for publication April 13, 2012;
accepted after revision September 18, 2012.

Published online: September 19, 2012.