High-resolution insertion-site analysis by linear amplification–mediated PCR (LAM-PCR)

Manfred Schmidt[1–3,5], Kerstin Schwarzwaelder[1], Cynthia Bartholomae[1], Karim Zaoui[2], Claudia Ball[1], Ingo Pilz[3], Sandra Braun[1], Hanno Glimm[1] & Christof von Kalle[1,4,5]

[1]Department of Translational Oncology, National Center for Tumor Diseases, Im Neuenheimer Feld 350, 69120 Heidelberg, Germany. [2]Department of Internal Medicine I, University Hospital of Freiburg, Hugstetter Strasse 55, 79106 Freiburg, Germany. [3]Institute for Molecular Medicine and Cell Research, University of Freiburg, Stefan-Meier-Strasse 17, 79104 Freiburg, Germany. [4]Division of Experimental Hematology, Cincinnati Children's Research Foundation, 3333 Burnet Avenue, Cincinnati, Ohio 45229, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to C.v.K. (christof.kalle@nct-heidelberg.de) or M.S. (manfred.schmidt@nct-heidelberg.de).

Integrating vector systems used in clinical gene therapy have proven their therapeutic potential in the long-term correction of immunodeficiencies[1–4]. The integration loci of such vectors in the cellular genome represent a molecular marker unique for each transduced cell and its clonal progeny. To gain insight into the physiology of gene-modified hematopoietic repopulation and vector-related influences on clonal contributions, we have previously introduced a technology—linear amplification–mediated (LAM) PCR—for detecting and sequencing unknown DNA flanking sequences down to the single cell level[5] (Supplementary Note online). LAM-PCR analyses have enabled qualitative and quantitative measurements of the clonal kinetics of hematopoietic regeneration in gene transfer studies, and uncovered the clonal derivation of non-leukemogenic and leukemogenic insertional side effects in preclinical and clinical gene therapy studies[4,6–8]. The reliability and robustness of this method results from the initial preamplification of the vector-genome junctions preceding nontarget DNA removal via magnetic selection. Subsequent steps are carried out on a semisolid streptavidin phase, including synthesis of double complementary strands, restriction digest, ligation of a linker cassette onto the genomic end of the fragment and exponential PCR(s) with vector- and linker cassette–specific primers. LAM-PCR can be adjusted to all unknown DNA sequences adjacent to a known DNA sequence. Here we describe the use of LAM-PCR analyses to identify 5′ long terminal repeat (LTR) retroviral vector adjacent genomic sequences (Fig. 1 and Box 1).

MATERIALS

REAGENTS

Taq DNA polymerase (Qiagen)
dNTPs (Fermentas)
Oligonucleotides and primers (MWG Biotech; **Supplementary Tables 1** and **2** online)
Magnetic particles: Dynabeads M-280 Streptavidin (Dynal)
Kilobase binder kit (Dynal)
Klenow polymerase (Roche)
Hexanucleotide mixture (Roche)
Restriction endonuclease(s) and incubation buffer(s) (New England Biolabs)
Fast-Link DNA ligation kit (Epicentre)
T4 DNA Ligase (New England Biolabs)
Spreadex EL1200 precast gel (Elchrom Scientific)
QIAquick PCR purification kit (Qiagen)
DNA extraction kit (Qiagen or Roche)

EQUIPMENT

Magnetic particle concentrator (MPC; MPC-E-1; Dynal)
Microcon-30 (Millipore)
Submerged gel electrophoresis apparatus SEA 2000 (Elchrom Scientific)
Thermocycler programmed with the desired protocols (Biometra)
Horizontal shaker KS 260 basic (IKA)
Gel documentation system (PeqLab)

*For a full list of materials, see **Supplementary Methods** online.*

PUBLISHED ONLINE 29 DECEMBER 2007; DOI:10.1038/NMETH1103

PROCEDURE

Linear PCR

1| Set up a 50-µl reaction linear PCR initiated from 5′-biotinylated vector-specific primer(s) as follows:

10× Taq polymerase reaction buffer	5 µl	Template DNA (0.01–10,000 ng)	*x* µl
dNTPs (200 µM each)	1 µl	*Taq* polymerase (2.5 U/µl)	0.5 µl
Primer LTRI (0.5 pmol/µl)	0.5 µl	dH_2O	43 – *x* µl

The procedure described here starts with the analysis of extracted DNA. A comparison of common DNA extraction procedures (organic, anion-exchange or silica-based methods) did not reveal any differences on the qualitative outcome of the LAM-PCR technique. This protocol describes the procedure for one sample. When single magnetic separation units are used, do not analyze more than 20 samples in parallel. For high-throughput LAM-PCR analysis, we recommend the use of multipipettes and 96 well plate–based magnetic separators.

▲ **CRITICAL STEP**

2| Amplify nucleic acid fragments using the following PCR program. After the first 50 cycles of linear PCR, add 2.5 U *Taq* polymerase and carry out a second 50-cycle amplification using the same program.

Cycle number	Denaturation	Annealing	Polymerization	Final
1	5 min at 95 °C			
2–51	1 min at 95 °C	45 s at 60 °C	90 s at 72 °C	
52			10 min at 72 °C	
Hold				4 °C

When enough material is available (for example, monoclonal tumor sample), reduce the number of cycles from two 50-cycle amplifications to one 25-cycle reaction. The PCR parameters given here (and in Steps 21 and 26) allow the simultaneous detection of integration sites from minimal amounts of DNA, from as little as a single cell.

■ **PAUSE POINT** Do not store PCR products for several weeks at 4 °C as the sample volume will decrease as a result of evaporation, which may limit the efficiency of magnetic capture (Step 7).

Magnetic capture

3| Place 20 µl (200 µg of magnetic beads) of the magnetic bead solution at room temperature (18–25 °C) for 60 s on an MPC and discard the supernatant.

4| Resuspend the attached beads in 40 µl of PBS with 0.1% BSA, collect the beads using an MPC and discard the supernatant. Repeat this step once more.

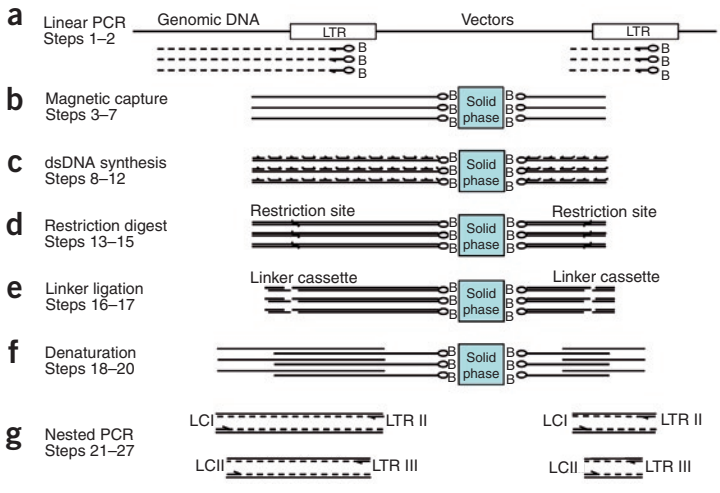

Figure 1 | Schematic outline of LAM-PCR to amplify 5′-LTR retroviral vector-genomic fusion sequences. (**a**–**g**) After preamplification of the target DNA (**a**) and its immobilization on paramagnetic beads (**b**), enzymatic reaction steps (**c**–**e**) generate target DNAs composed of a known linker sequence at the 5′ end of the unknown genomic flanking sequence. After denaturation of target DNA (**f**), nested PCR with linker- and vector-specific primers each complementary to one of the known ends of the target DNA can be applied (**g**). As the LTR is present at the 5′ and the 3′ end of the integrated vector, two distinct fragments are generated from each integration clone: the vector-genome junction of an integration clone as well as the internal vector sequence. B, biotinylated primer; LTR I, LTR primer for the linear PCR; LTR II, LTR primer for the first exponential PCR; LTR III, LTR primer for the second exponential PCR; LCI, linker cassette primer for the first exponential PCR; LCII linker cassette primer for the second exponential PCR.

5| Wash the beads in 20 µl of binding solution (BS), collect beads using an MPC, discard the supernatant and resuspend the beads in 50 µl of BS.
Alternatively, wash the beads in 50 µl of 3 M LiCl-based binding and washing (BW) buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 3 M LiCl), collect them using an MPC, discard the supernatant and resuspend the beads in 50 µl of 6 M LiCl-based BW buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 6 M LiCl).

6| Transfer 50 µl of the magnetic bead solution to the tube containing the linear PCR product. Mix gently to avoid foaming BS.
The ratio of volumes of PCR product and magnetic bead solution must always be 1:1.

7| To covalently link the biotinylated PCR products to streptavidin-coupled beads (DNA-bead complexes), incubate the sample overnight at room temperature on a horizontal shaker at 300 r.p.m.
Incubation time may be 8–48 h. Incubation temperature may vary between 20 and 43 °C.
▲ **CRITICAL STEP**
■ **PAUSE POINT** Samples can be stored at 4 °C for several days.

8| Set up a 20-µl hexanucleotide priming reaction to generate double-stranded DNA.

Double-stranded DNA synthesis

ddH_2O	16.5 µl
10× hexanucleotide mixture	2 µl
dNTPs (200 µM each)	0.5 µl
Klenow polymerase (2 U/µl)	1 µl

BOX 1 PRIMERS AND RESTRICTION ENZYMES

The design of short LAM-PCR amplicons that do not exceed several hundred base pairs is critical for the efficiency of each analysis. Amplicons should be long enough to be resolved by electropheresis (>80 bp) and should encompass at least 20 bp of non-primer vector sequences. This will allow identification of amplicons that contain true retrovirus sequence.

Primers. Primers should have a (G+C) content of 50% (range: 40–60%) and a length of approximately 20 nucleotides (range: 18–25; Steps 1, 21 and 26). Avoid having 3 or more identical nucleotides at the 3′ end of primers. Primers need to be aligned to annotated assemblies of the genome of interest (template DNA of LAM-PCR) to avoid unspecific hybridization of the primers. An additional alignment search for repetitive elements is strongly recommended. A minimum of 6 nucleotides at the 3′ end of the designed primer should not reveal any similarity with repetitive elements. In all types of insertional vector analysis, it is important to understand the direction, structure and sequence a vector will have after its insertion in the target cell. Especially when designing primers for retrovirus LTRs, be aware of the correct sequence map. In 5′ to 3′ orientation, LTRs are composed of the following regions: unique 3′ (U3), redundancy (R), and unique 5′ (U5). Retroviral vector sequence maps commonly show the vector incorporated into a plasmid including full LTRs as cloned. Most retroviral vectors are composed of LTR sequences that are different between both vector ends to introduce extra sequences, delete parts of the LTR or simply to avoid sequence similarity that could result in unwanted recombinations. Here the cloning vector sequence map will always differ from the 'true' integrated proviral DNA. The integration process will transfer the U3 regions to the 5′ end of the vector and vice versa. After integration, both LTRs of the proviral DNA consist of the U5 region from the cloning vector 5′ LTR and the U3 region from the cloning vector 3′ LTR. This issue is decisive for a correct primer design. Most initial failures of obtaining meaningful amplicons by LAM-PCR that we have consulted on had deficiencies in this area.

Restriction enzymes. For short average amplicon lengths, we recommend using 'four-cutters' in the restriction digest during LAM-PCR (Step 13), which yield an average amplicon length of 128 bp plus the sequences needed for the LAM-PCR analysis on both ends of the template. In case of LTR vectors, choose restriction enzymes to avoid cuts too close (<40 bp) to the LTR within the vector (**Fig. 1**; in this case, an internal 3′ LTR–adjacent vector sequence is generated) because this generates short amplicons whose amplification can hamper and even outcompete larger integration site restricition fragments.

9| Collect the DNA-bead complexes (Step 7) using an MPC (60 s), discard the supernatant and resuspend the DNA-bead complexes in 100 µl of ddH_2O.

10| Collect the DNA-bead complexes using an MPC (60 s), discard the supernatant and resuspend the beads in 20 µl of hexanucleotide priming mixture.

11| Incubate the reactions at 37 °C for 1 h in a thermocycler.
As Klenow polymerase exhibits 3′ to 5′ exonuclease activity, do not exceed 1 h of incubation time.

12| Add 80 µl of ddH_2O to the hexanucleotide priming mixture, collect the DNA-bead complexes using an MPC (60 s), discard the supernatant and resuspend the DNA-bead complexes in 100 µl of ddH_2O.

Restriction digest

13| Collect the DNA-bead complexes using an MPC (60 s), discard the supernatant and resuspend the DNA-bead complexes in 20 µl of restriction digest mixture (17 µl of ddH_2O, 2 µl of 10× restriction enzyme buffer and 1 µl of restriction enzyme at 4 U/µl) to digest the unknown flanking sequence.
*The use of 'four-cutters' such as Tsp509I or MseI is strictly recommended (**Box 1** and **Supplementary Table 2**). Be sure that there is no restriction site located within the known sequence of interest.*

14| Incubate the restriction digest mixture at the temperature recommended by the manufacturer for 1 h in a thermocycler.
▲ **CRITICAL STEP**

15| Add 80 µl of ddH_2O to the restriction digest mixture, collect the DNA-bead complexes using an MPC (60 s), discard the supernatant and resuspend the DNA-bead complexes in 100 µl of ddH_2O.

Linker ligation

16| Set up a 10-µl ligation reaction to ligate a known oligonucleotide sequence (see **Box 2**) to the unknown flanking sequence as follows. Incubate the reactions at room temperature for 10 min to 1 h.

ddH_2O	5 µl
10× incubation buffer	1 µl
ATP (10 mM)	1 µl
Linker cassette	2 µl
'Fast Link' DNA ligase (2 U/µl)	1 µl

The ligation of a known oligonucleotide sequence to the unknown flanking sequence of the DNA-bead complexes allows additional exponential amplification of the genomic-vector fusion sequences (Steps 21–27) and visualization (Step 28).
▲ **CRITICAL STEP**

17| Add 90 µl of ddH_2O to the ligation reaction, collect the DNA-bead complexes using an MPC (60 s), discard the supernatant and resuspend the DNA-bead complexes in 100 µl of ddH_2O.

BOX 2 GENERATION OF A LINKER CASSETTE

(**i**) Set up the generation of an unidirectional double-stranded linker cassette using a 1.5 ml tube in a 200 µl of cassette ligation mixture (110 µl of Tris-HCl (250 mM; pH 7.5), 10 µl of $MgCl_2$ (100 mM), 40 µl of oligonucleotide LC1 (100 pmol/µl) and 40 µl of oligonucleotide LC2 (100 pmol/µl); **Supplementary Table 2**).
(**ii**) Incubate at 95 °C for 5 min in a heating block, and cool the reaction mixtures as slowly as possible (6–24 h).
(**iii**) Add 300 µl of ddH_2O to the linker cassette formation reaction and transfer the sample on a Microcon-30 column.
(**iv**) Centrifuge (15,000*g*) the sample at room temperature for 12 min.
(**v**) Place the reversed column onto a clean tube and centrifuge (1,000*g*) at room temperature for 3 min.
(**vi**) Adjust the volume of the concentrated sample with ddH_2O to a final volume of 80 µl and freeze; for example, 4 aliquots 20 µl each at –20 °C. Linker cassettes can be stored at –20 °C for years. Once linker cassettes are thawed, do not freeze them again.

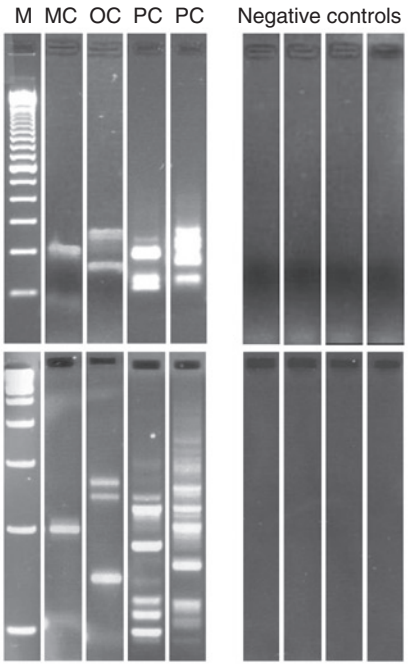

Figure 2 | LAM-PCR analysis of clinical samples. One microgram of DNA isolated from peripheral blood cells served as template for LAM-PCR[10]. High-resolution gels such as Spreadex (Elchrom Scientific) gels (bottom) provide greater resolution than 2% agarose gels (top) for separation of LAM-PCR products and to define the clonality of the analyzed sample(s). The number of visible bands correlates with the number of unique integration sites present in the sample. Shown are LAM-PCR products of a monoclonal sample (MC), an oligoclonal sample (OC) and two polyclonal samples (PC). Negative controls from left to right are untransduced genomic DNA, H_2O negative control of the LAM-PCR, H_2O negative control of the first exponential PCR, H_2O negative control of the second exponential PCR. M, 100 bp ladder.

Denaturation

18| Collect the DNA-bead complexes using an MPC (60 s), discard the supernatant and resuspend the beads in 5 µl of 0.1 N NaOH.
Always use fresh 0.1 N NaOH.

19| Incubate the denaturing reaction at room temperature for 10 min to 1 h.

20| Collect the DNA-bead complexes using an MPC (60 s), and transfer the supernatant that contains the unbiotinylated single-stranded (ss) DNA that is composed of linker sequence, genomic sequence and vector sequence to a new tube.
■ **PAUSE POINT** Denatured ssDNA can be stored at 4 °C for several days and at −20 °C for years.

Nested PCR

21| Set up a 50-µl reaction for each ssDNA to be amplified by a PCR using linker- and 5′-biotinylated vector-specific primers.

ddH_2O	40 µl
10× Taq polymerase reaction buffer	5 µl
dNTPs (200 µM each)	1 µl
Primer LTRII (50 pmol/µl)	0.5 µl
Primer LC I (50 pmol/µl)	0.5 µl
Template DNA	2 µl
Taq polymerase (2.5 U/µl)	1 µl

▲ **CRITICAL STEP**

22| Amplify nucleic acid fragments using the PCR program in Step 2, except using 35 cycles of denaturation, annealing and polymerization:
■ **PAUSE POINT** PCR products can be stored at 4 °C for several days and at −20 °C for years.

23| Collect the first PCR product using an MPC (Steps 3–7).
▲ **CRITICAL STEP**

24| Collect the DNA-bead complexes (Step 23) using an MPC (60 s), discard the supernatant and resuspend the DNA-bead complexes in 100 µl of ddH_2O.

25| Separate the non-biotinylated DNA strand using 20 µl of 0.1 N NaOH (Steps 18–20).

26| Set up a 50-µl reaction for each ssDNA sample to be amplified by PCR using linker- and vector-specific primers.

ddH_2O	40 µl
10× Taq polymerase reaction buffer	5 µl
dNTPs (200 µM each)	1 µl
Primer LTRIII (50 pmol/µl)	0.5 µl
Primer LC II (50 pmol/µl)	0.5 µl
Template DNA	2 µl
Taq polymerase (2.5 U/µl)	1 µl

▲ **CRITICAL STEP**

27| Amplify nucleic acid fragments using the same PCR program as in Step 22.
■ **PAUSE POINT** PCR products can be stored at 4 °C for several days and at −20 °C for years.

28| For visualization of the LAM-PCR products, separate 10 µl of the second exponential PCR product on a 2% agarose gel. Then separate another 10 µl on a high-resolution Spreadex gel according to the manufacturer's recommendations (**Fig. 2**).
The 2% agarose gel electrophoresis is not sufficient to separate the LAM-PCR amplicons derived from an oligoclonal or polyclonal sample efficiently. Separation on a high-resolution gel allows a resoluton of 4–10 base pairs to visualize different integration clones as distinct bands.
➨ **TROUBLESHOOTING**

29| Sequence LAM-PCR amplicons either by 'picking' individual bands or by conventional shotgun cloning (for example, using TOPO TA vector (Invitrogen) according to the manufacturer's recommendations) and sequencing.
➨ **TROUBLESHOOTING**

TROUBLESHOOTING TABLE

PROBLEM	SOLUTION
Step 28 No bands are visible.	Make sure the experiment was designed correctly. Repeat the first and second exponential PCR using 2 µl of denatured ssDNA (Steps 21–28).
Bands are visible in the LAM-PCR DNA control and/or H_2O control (**Fig. 2**).	Repeat first and second exponential PCR with 1–2 µl of denatured ssDNA (Steps 21–28). If bands are still present, samples were contaminated during LAM-PCR. Discard working solutions (such as primers) and repeat LAM-PCR with fresh aliquots.
Bands are visible in the first and/or second exponential PCR H_2O control (see **Fig. 2**, samples 7 and 8).	Repeat first and second (first H_2O control contaminated) or second (second H_2O control contaminated) exponential PCR with 1–2 µl of denatured ssDNA (Steps 21–28). When bands are still present, discard PCR working solutions and repeat exponential PCRs (for example, with 1 µl of ssDNA) with fresh aliquots.
Step 29 Sequenced LAM-PCR amplicons are unspecific.	Pay close attention to the correct design of the experiment. Design new primers.

CRITICAL STEPS

Step 1 The designed primers (**Fig. 1** and **Supplementary Table 1**) should hybridize close to the 5′ or 3′ end of the vector. When two different primers are used for the linear PCR (for example, to allow a fast and efficient evaluation of new primers), reduce the primer stock concentration (0.125 pmol each). Input DNA varies, as for a successful analysis the vector copy number within the sample is crucial. As an example, 1–10 ng of template DNA of a polyclonal clinical sample reaches efficient results when 100% of cells were transduced. In the case of a minor proportion of transduced cells present in the DNA sample (for example, 1–0.1%), minimal DNA template for LAM-PCR should be 100 ng. The use of a genomic control DNA and H_2O is strictly recommended (**Fig. 2**).

Step 7 Never freeze samples in Steps 7–18, otherwise the beads will loose their magnetic properties.

Step 14 When the restriction enzyme does not exhibit star activity, the incubation time can be extended for several hours. To guarantee efficient reaction steps during LAM-PCR, in particular if polyclonal samples are analyzed, the use of a 'four-cutter' is recommended.

Step 16 When freshly produced linker cassettes are used, we recommend performing the generation of the linker cassette either before setting up LAM-PCR or during linear PCR (Step 2) and magnetic capture (Step 7; see **Box 2**). Once an aliquot of the linker cassette is thawed, never freeze it again. Instead of Fast-Link DNA ligase, also conventionally T4 DNA Ligase can be used according to the manufacturer's recommendations. The use of the correct linker cassette is crucial (see **Supplementary Table 2**).

Step 21 Do not use more than 2.0 µl of the denatured ssDNA as template, otherwise NaOH will inhibit the first exponential PCR. Alternatively, desalt the ssDNA using a Microcon-30 column before the first PCR. Primer concentration range should be 0.1–1 µM.

Step 23 Magnetic capture of the first exponential PCR product has proven highest specificity, in particular if a new experiment including new primers is set up. Alternatively, purified (for example, using QIAprep PCR purification kit) or 0.01–1 µl of unpurified first exponential PCR product may be used as template for the second exponential PCR.

Step 26 Do not use more than 2.0 µl of the denatured ssDNA as template. Otherwise NaOH will inhibit the second exponential PCR. Alternatively, desalt the ssDNA using a Microcon-30 column before the second PCR. Primer concentration range: 0.1–1 µM.

COMMENTS

A variety of PCR technologies have been established to characterize unknown flanking DNA regions (for example, ligation-mediated (LM)-PCR[9]). However, none of these methods have the potential to characterize highly complex samples with multiple target sequences and a sensitivity down to the single-cell level. In contrast, LAM-PCR has sufficient sensitivity, specificity and robustness to enable the molecular analysis of gene transfer studies using integrating vector systems (oncoretro-, foamy-, lentiviral and AAV vector integration), highlighted in the first discovery of clinical vector–induced side effects from minimal sample amounts from patients[4,6–8].

For the *in vivo* monitoring of the progeny of individual transduced cell clones (for example, detection of a clone in preclinical or clinical gene therapy in different sorted hematopietic lineages over time), a tracking PCR based on the LAM-PCR sequencing data output can be accomplished. Using primers specific for the vector LTR in combination with primers hybridizing at the genomic flank of the clone of interest, a fast and efficient identification of the clone(s) can be performed. For optimal sensitivity and specificity, design of a nested PCR and a magnetic capture step of the first PCR product (LTR primer is biotinylated) is recommended. Additional quantification of individual clones can be accomplished in different ways. First, a quantitative-competetive tracking PCR approach makes use of designed partially deleted standards that equal the clones of interest. The addition of a defined copy number of the standard to the wild-type DNA allows simultaneous amplification in the same PCR[4,6]. Based on the signal intensity of the standard and the wild-type DNA, calculations of the individual clone contribution within the sample can be assessed. Second, high-throughput LAM-PCR analyses in a clinical gene therapy trial of chronic granulomatous disease (CGD) have provided a reliable measurement of quantitative clone contribution based on the retrieval frequency of individual LAM-PCR amplicons: the more often one integration site is sequenced, the more this integration site is present in the analyzed sample[4]. To keep potential biases as small as possible, we recommend repetitive analysis of the same sample and inclusion of two or more distinct restriction enzymes during LAM-PCR.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft, the German Ministry of Education and Research, the European Union and the US National Institutes of Health. We thank all present and former members and collaborators of our laboratory who have contributed work or materials to the many experiments, time and efforts toward establishing and optimizing this protocol.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Cavazzana-Calvo, M. *et al.* Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* **288**, 669–672 (2000).
2. Aiuti, A. *et al.* Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* **296**, 2410–2413 (2002).
3. Gaspar, H.B. *et al.* Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* **364**, 2181–2187 (2004).
4. Ott, M.G. *et al.* Correction of X-linked chronic granulomatous disease by gene therapy is augmented by insertional activation of MDS/EVI1, PRDM16 or SETBP1. *Nat. Med.* **12**, 401–409 (2006).
5. Schmidt, M. *et al.* Polyclonal long-term repopulating stem cell clones in a primate model. *Blood* **100**, 2737–2743 (2002).
6. Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419 (2003).
7. Deichmann, A. *et al.* Vector integration is non-random, clustered and influences the in vivo fate of lymphopoiesis in SCID-X1 gene therapy. *J. Clin. Invest.* **117**, 2225–2232 (2007).
8. Schwarzwaelder, K. *et al.* Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *J. Clin. Invest.* **117**, 2241–2249 (2007).
9. Mueller, P.R. & Wold, B. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**, 780–786 (1989).
10. Glimm, H. *et al.* Efficient marking of human cells with rapid but transient repopulating activity in autografted recipients. *Blood* **106**, 893–898 (2005).