Citation: Molecular Therapy — Methods & Clinical Development (2015) **2,** 14061; doi:10.1038/mtm.2014.61
© 2015 The American Society of Gene & Cell Therapy All rights reserved 2329-0501/15

www.nature.com/mtm

ARTICLE

High-throughput monitoring of integration site clonality in preclinical and clinical gene therapy studies

Frank A Giordano[1], Jens-Uwe Appelt[2,3], Barbara Link[2], Sebastian Gerdes[4], Christina Lehrer[2], Simone Scholz[2], Anna Paruzynski[2], Ingo Roeder[4], Frederik Wenz[1], Hanno Glimm[2], Christof von Kalle[2], Manuel Grez[5], Manfred Schmidt[2] and Stephanie Laufs[2]

Gene transfer to hematopoietic stem cells with integrating vectors not only allows sustained correction of monogenic diseases but also tracking of individual clones *in vivo*. Quantitative real-time PCR (qPCR) has been shown to be an accurate method to quantify individual stem cell clones, yet due to frequently limited amounts of target material (especially in clinical studies), it is not useful for large-scale analyses. To explore whether vector integration site (IS) recovery techniques may be suitable to describe clonal contributions if combined with next-generation sequencing techniques, we designed artificial ISs of different sizes which were mixed to simulate defined clonal situations in clinical settings. We subjected all mixes to either linear amplification–mediated PCR (LAM-PCR) or nonrestrictive LAM-PCR (nrLAM-PCR), both combined with 454 sequencing. We showed that nrLAM-PCR/454-detected clonality allows estimating qPCR-detected clonality *in vitro*. We then followed the kinetics of two clones detected in a patient enrolled in a clinical gene therapy trial using both, nrLAM-PCR/454 and qPCR and also saw nrLAM-PCR/454 to correlate to qPCR-measured clonal contributions. The method presented here displays a feasible high-throughput strategy to monitor clonality in clinical gene therapy trials is at hand.

Molecular Therapy — Methods & Clinical Development (2015) **2**, 14061; doi:10.1038/mtm.2014.61; published online 1 April 2015

INTRODUCTION

Hematopoietic stem cell (HSC) gene therapy has provided substantial benefits to patients suffering from severe combined immunodeficiency (SCID),[1–4] chronic granulomatous disease (CGD),[5] adrenoleukodystrophy,[6] Wiskott-Aldrich syndrome,[7] and human immunodeficiency virus.[8]

Nevertheless, this innovative and promising therapy bears the risk of inducing clonal imbalance and leukemia.[9–12] Such side effects eventually arise as direct consequence of (proto) oncogene activation by the promoter and enhancer elements located in the long terminal repeats (LTR) of the transfer vectors. A number of studies showed that integration of these vectors is nonrandom *a priori* (reviewed in ref. 13), but also that *in vivo* selection of clones that carry integrations located in recurrently targeted regions (so called "common" integration sites) occurs.[14,15] It is believed that only a small fraction of transplanted HSC actually carries survival-supporting integrations in "common" integration sites (IS) and even fewer might then contribute to hematopoiesis in a long-term manner.[16] Furthermore, to outcompete other "normal" clones, additional factors including mutations are likely necessary for conferring significant growth advantages to these few clones.[10,17] However, it remains unclear how clonal outgrowth is concerted, *e.g.*, after second-hits, and, even more important, how uncontrolled proliferation can be detected before the clinical appearance of the transplanted individuals starts to deteriorate.

Cells transduced with oncoretroviral and lentiviral vectors are individually marked by a highly characteristic insertion site profile. Autologous transplantation studies in animal models and clinical studies have the potential to shed light on issues such as the number of clones contributing to stable hematopoiesis, clonal succession, and lineage commitment. However, retroviral tracking studies are often limited by insensitive detection methods, low numbers of transplanted stem cells, and limited life span of immunodeficient mice.

Screening for (potentially) harmful HSC clones can be achieved using LAM- and LM-PCR (ligation-mediated PCR) combined with next-generation sequencing techniques, where a special emphasis lies on directly quantifying individual clonal contributions using the sequence retrieval frequencies ("read counts") of individual ISs over time.[18,19] Yet, although a suitable measure to identify dominant clones, LAM-/LM-PCR-based IS read counting may be a rough estimation and the mechanisms that skew the frequencies are to date not fully understood.[19,20] Therefore, it is broadly accepted that dominantly appearing clones have to be validated using an independent and precise method, the qPCR (quantitative real-time PCR)—a strategy that has already found its way into ongoing clinical gene therapy protocols.

[1]Department of Radiation Oncology, Universitätsmedizin Mannheim, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [2]Department of Translational Oncology, National Center for Tumor Diseases and German Cancer Research Center (DKFZ), Heidelberg, Germany; [3]CLC bio, Aarhus, Denmark; [4]Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universität, Dresden, Germany; [5]Institute for Biomedical Research, Georg-Speyer-Haus, Frankfurt, Germany.
Correspondence: S Laufs (stephanie.laufs@nct-heidelberg.de)
Received 20 June 2014; accepted 4 November 2014

As the number of users, the sample throughput and the variety of kits and machines for qPCR constantly rises, potential sources of error and artefact exist but are often neglected in daily laboratory practice.[21,22] Yet, in the light of the multitude of ongoing and novel clinical HSC gene transfer studies, it is of highest importance to minimize operational and technical variability and, in consequence, to apply uniform methodological and analytical standards.

Adhering to the *Minimum Information for Publication of Quantitative Real-Time PCR Experiments* (MIQE) guidelines,[22] we have previously established an integration-site specific qPCR that allows to accurately and sensitively trace individual growth kinetics of transduced HSC clones in clinical gene therapy applications.[23,24] However, quantification of individual clonal growth is dependent on the IS that is present in the clone of interest. Thus, establishing the assay can be challenging, especially since various vectors exhibit preferences for inserting in or near to repetitive elements.[25] Furthermore, large amounts of sample material are required.

Recent studies have postulated that the progeny of a single gene-marked clone may be quantified by counting the number of corresponding IS sequences recovered by linker-mediated PCR strategies.[18,19,26] To test this hypothesis, we have constructed artificial IS standards (arIS) of different sizes (ranging from 50 to 500 base pairs of size), which we mixed in various ratios to model balanced and imbalanced clonality of hematopoiesis. We validated all arIS mixes using qPCR and subjected them to linear amplification-mediated PCR (LAM-PCR[27]) and nonrestrictive LAM-PCR (nrLAM-PCR[18]) followed by amplicon sequencing using a 454/Roche system. We then compared clonal contributions calculated from the number of recovered reads to qPCR data to assess the robustness of the individual estimation. Finally, we used sample material from a clinical gene therapy study on CGD[5] to prove the feasibility in the context of *in vivo* gene therapy.

RESULTS

Design and validation of internal standards for definition of size-adapted read count

In the current study, we sought to investigate whether combined IS detection/sequencing techniques allow estimating clonal contributions using predefined mixes of artificial internal standards (arIS) of different sizes (Supplementary Figure S1a). Each arIS can be cut using the restriction enzymes *Ple*I and *Bsm*AI and will deliver an LTR-genome fragment. We have mixed different ratios of the different arIS. In mix A7 the ratios of arIS 25, arIS 50, arIS 100, arIS 300 and arIS 500 were balanced with a 20% contribution of each arIS (Supplementary Figure S1b). In mixes H, I, and J, we wanted to create an 10:1:1:1:1 ratio setting for the clone delivering the smallest (arIS25 with 25bp in mix H), a clone delivering a medium-size (arIS 100 with 100 bp in mix I) and a clone delivering the largest LAM-CPR fragment (arIS500 with 500 bp in mix J). These different mixes closely reflect the situation in clinical IS analysis, where poly-, oligo-, and monoclonal patterns were observed. The exact proportions of each arIS in the mixtures were validated using qPCR. This method was validated before by our group allowing to detect the proportion of an individual plasmid down to 10 copies in a total plasmid number of 10^5 (ref. 24). Mix A7 should contain five arIS, each making up 20% of all detectable clones. In this balanced mix, clonal contributions ranged from 15.1 ± 0.8% to 25.6 ± 1.8%. In the unbalanced mixes, the dominant clone should make up 71.4% for arIS 25 in mix H, for arIS 100 in mix I and for arIS 500 in mix J, respectively. In all mixes, we verified the clonal contributions of the dominant clone which made up 69.8 ± 0.6% (mix H), 74.9 ± 1.9% (mix I) and 67.2 ± 1.3% (mix J). These qPCR data were then taken as the reference standard against which the LAM-/nrLAM-PCR retrieved ratios was compared. Thus, although we used two independent methods for the determination of concentrations (NanoDrop and Agilent), we still obtained a fluctuation of ratios when validating, revealing the high importance of a thorough qPCR validation after mixing of plasmids.

Consistency of amplicon sequencing-based clonal assessment

LAM-PCR as well as nrLAM-PCR were repeated twice to assess the reproducibility of amplicon sequencing-based clonal assessment (Supplementary Figure S2). When counting LAM-PCR-recovered amplicon sequence reads, we saw a correlation coefficient between the two different runs of $R^2 = 0.9294$ (Figure 1a). For nrLAM-PCR, the correlation was considerably higher ($R^2 = 0.9992$; Figure 1b).

Comparison of LAM-/nrLAM-PCR-based clonal assessment with qPCR data

To find out whether LAM-PCR and nrLAM-PCR amplicons were sequenced in relation to the individual clonal size and thus reflecting the clonal contribution, we compared LAM-based read counts and nrLAM-based read counts with qPCR data, respectively. While both techniques detected all arIS clones mixed, several differences were

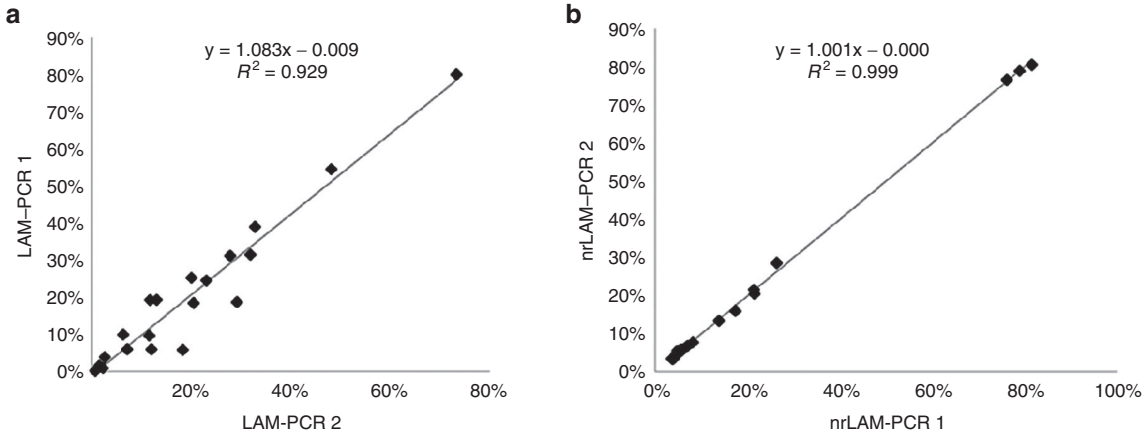

Figure 1 Consistency of duplicate data. LAM-PCR (**a**) and nrLAM-PCR (**b**) were performed twice on each of the different mixtures. Following 454-based sequencing and bioinformatical analysis of the read counts, the resulting clonal compositions of both duplicate runs were correlated in a scatter plot. For each correlation, the linear regression equation and the Pearson correlation coefficient (R^2) were calculated.

© 2015 The American Society of Gene & Cell Therapy

noted. In the majority of cases, the LAM-PCR-based estimation of clone copy numbers deviated significantly from the qPCR data, the assessment detected a dominant clone if the flanking site was not larger than 100 base pairs (Figure 2a). In line with the observation that larger fragments are weakly recovered using biotin-streptavidin capturing, the largest arIS was significantly underrepresented in all experiments and, most importantly, in the mix with arIS 500 being the dominant clone. The same mixtures were measured with nrLAM-PCR again compared with qPCR. In contrast to LAM-PCR data, nrLAM-PCR turned out to deliver close-to exact ratios in most clonal situations as measured by qPCR (Figure 2b). As expected, the subsequent correlation analyses of both methods revealed a

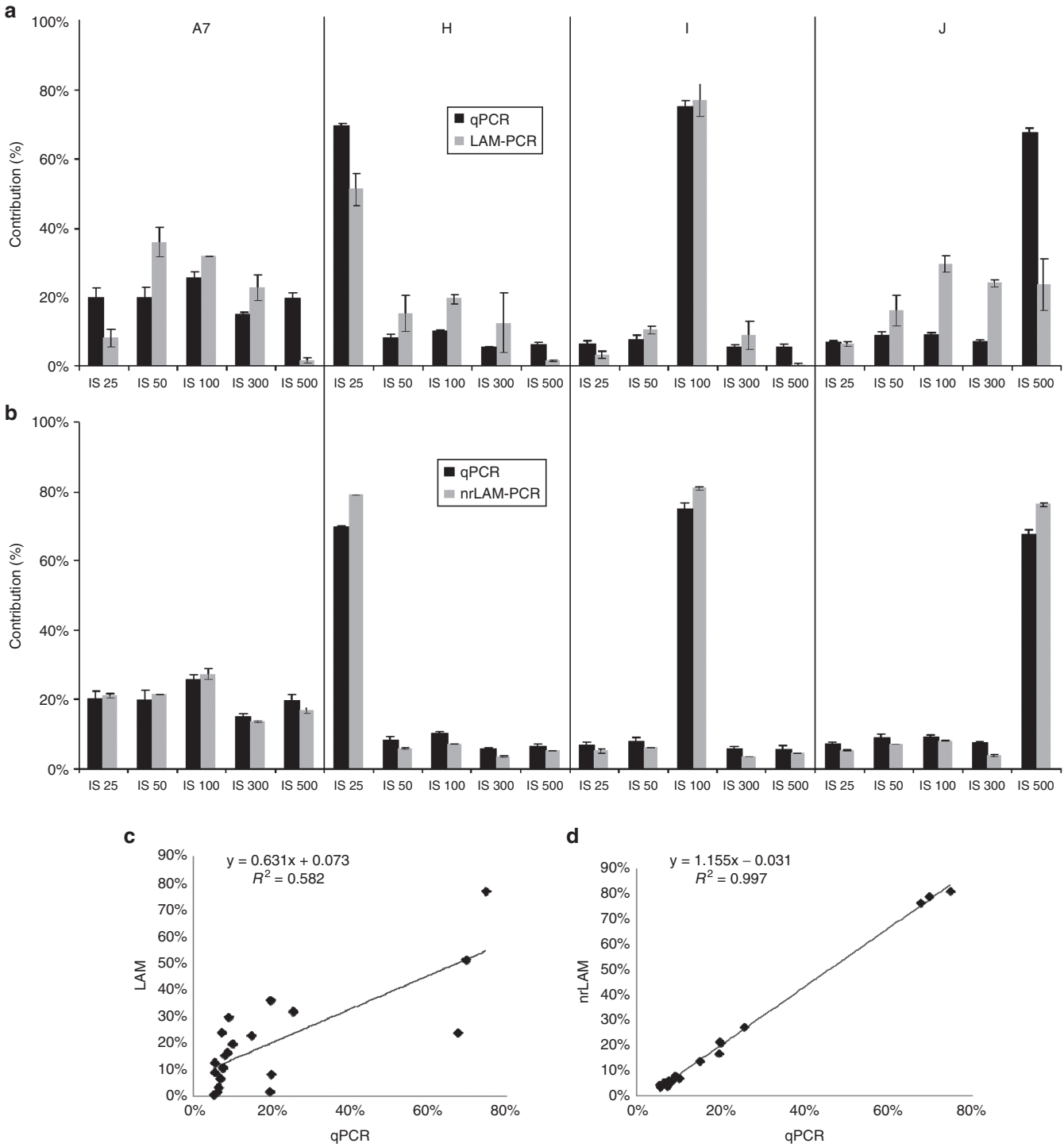

Figure 2 Clonal compositions detected with LAM-PCR and nrLAM-PCR compared to qPCR. LAM-PCR (**a**) and nrLAM-PCR (**b**) were performed twice and, following 454-based sequencing and bioinformatical analysis, clonal compositions were correlated with qPCR data (calculated from triplicates). Error bars represent standard deviations. The compositions of the mixes obtained with LAM-PCR (**c**) and nrLAM-PCR (**d**) -based clonal assessment were correlated in a scatter plot. For each correlation, the linear regression equation and the Pearson correlation coefficient (R^2) were calculated.

© 2015 The American Society of Gene & Cell Therapy

coefficient of determination for LAM-PCR versus qPCR of $R^2 = 0.5828$ and a considerably higher coefficient of determination for nrLAM-PCR versus qPCR of $R^2 = 0.9976$ (Figure 2c,d).

Detection of small clones and influence of background DNA

There is uncertainty whether or not background DNA may lower the detection limit of nrLAM-PCR and LAM-PCR. To assess the ability of both methods to quantify small clones, we established mixes L and O, containing minimal amounts of artIS100. Additionally, to find out whether the assays are biased by the presence of human DNA, we also spiked 100 ng of genomic background (BG) DNA into both mixes (Supplementary Figure S3). Both methods reliably detected smallest clonal contributions in the range below 1% (detection of artIS 100 in mix L: qPCR 0.7 ± 0.5%; LAM-PCR 1.4 ± 0.5%; nrLAM-PCR 0.7 ± 0.2%) without being significantly skewed by background DNA (detection of artIS 100 in mix L+BG: qPCR 0.3 ± 0.01%; LAM-PCR 0.5 ± 0.01%; nrLAM-PCR 0.4 ± 0.1%; Figure 3). Similar results were obtained in mixes with roughly 1% contribution of artIS100 (detection of artIS 100 in mix O: qPCR 1.9 ± 0.02%; LAM-PCR 4.7 ± 2.2%; nrLAM-PCR 1.2 ± 0.07%; detection of artIS 100 in mix O+BG: qPCR 1.1 ± 0.08%; LAM-PCR 1.3 ± 0.2%; nrLAM-1.2 ± 0.06%).

Clonal assessment using qPCR in a patient with CGD

To investigate if nrLAM-PCR-based clonal assessment is also applicable and reliable in an *in vivo* setting, we analyzed the clonal repertoire in a patient included in a clinical gene therapy study for X-linked CGD (X-CGD). We designed a qPCR for the quantification of two clones ("F02" and "A02") present in a patient.[5] One of the clones (F02) was shown to be gaining clonal dominance which in turn led to the development of MDS and myeloid leukemia.[17] Since the ISs have already been identified previously, we extracted the vector-flanking downstream sequences using the UCSC genome browser and constructed LTR-genome junction sequences (Supplementary Figure S4a,b), which were used to design primer sets that amplify the vector-genome junctions. Following cloning of the vector-genome junction, plasmid amplification and copy number calculation, testing of different primer pairs for specificity and cross-reactivity, adequate pairs were selected and standard curves were established as described before (Supplementary Figure S5).[23,24]

We first analyzed the individual kinetics in the peripheral blood of one patient from ~3 months (d119) until 3.5 years (d1309) after transplantation of gene-corrected HSC using qPCR (Figure 4a). Between d119 and d245, we detected clone 87429 F02 to make up between 0.5% (±0.01) and 0.7% (±0.01) of total hematopoiesis. Starting from d250-300, this clone expanded to 51.9% (±5.9) of total hematopoiesis at day 560 and 45.6% (±2.3) at d924. Interestingly, pancytopenia and mild hypocellularity in BM were first noted at d840 and d990, respectively.[17] On d990, loss of chromosome 7 was first noted. Since then, the clone reached 65.1% (±11.3) on d1309, when BM analyses were consistent with refractory cytopenia with multilineage dysplasia. The patient underwent allogenic stem cell transplantation about 6 weeks later (d1350). We also quantified the clonal contribution of clone A02 to all cells. Similar to clone F02, the clone initially contributed between 2.8% (±0.2) and 9.6% (±1.2) to total hematopoiesis from d119 to d245 and ranged between 20 and 30% between d560 and d990. Six weeks before allogenic stem cell transplantation (d1309), the clone abruptly retracted to 4.1% (±0.6).

Comparison of qPCR to LAM-PCR- and nrLAM-PCR-based clonal analysis

We next aimed to compare the qPCR obtained values with LAM-PCR and nrLAM-PCR-based clonal assessment in the analyzed CGD-patient. Since LAM-PCR and nrLAM-PCR will only recover IS-bearing cells/clones and deliver no information on the complete hematopoietic cell pool, we initially quantitatively assessed clonal contributions to the complete transgenic pool (expressed as $gp91^{phox}$-positive cells; Figure 4b,c). Of note, due to limited template material, LAM-PCR and nrLAM-PCR were only performed once. All three methods revealed a linear increase of the clonal contribution of F02 to the total transgenic population between d245 (qPCR: 0.9% ± 0.01; nrLAM-PCR: 0.5%; LAM-PCR: 0.2%) and d1378 (qPCR: 78.2% ± 7.5; nrLAM-PCR: 99.2%; LAM-PCR: 93.6%), indicating a constant loss of competing clones since this clone is not significantly growing in this period (as compared to the total cell pool, see above). A02 contributed to ~10–15% to the transgenic population until d1171, when it was seen to contribute a maximum of roughly 34% with qPCR (qPCR: 34.7% ± 7.5; nrLAM-PCR: 17.3%; LAM-PCR: 0%).

DISCUSSION

After the adverse effects seen in clinical gene therapy studies over the past decade, a thorough workup of the biological events that may cause clonal outgrowth has been initiated.[28] The increased safety profiles of novel, self-inactivating vectors that are now entering the stage have led to a variety of novel preclinical stem cell gene therapy studies and subsequently to an increased need of feasible high-throughput methods to monitor clonal compositions.

Several sophisticated and well-described techniques exist to detect and track integration-bearing cells, most of which include

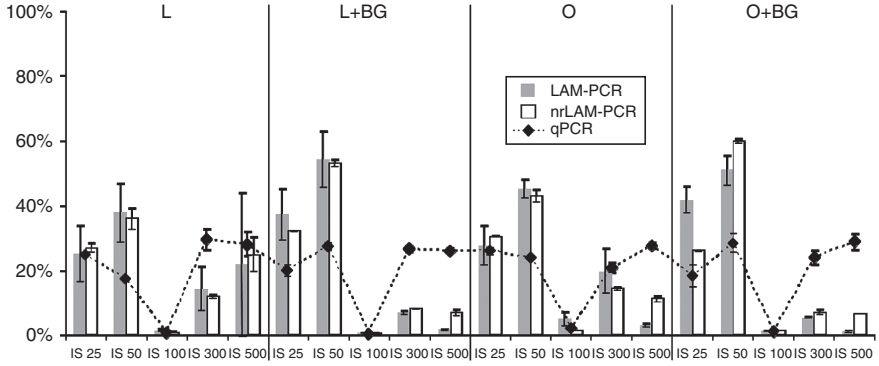

Figure 3 Clonal compositions detected with LAM-PCR and nrLAM-PCR compared to quantitative polymerase chain reaction (qPCR). LAM-PCR and nrLAM-PCR were performed twice and, following 454-based sequencing and bioinformatical analysis, clonal compositions were correlated with qPCR data. Mix L and mix O were analyzed with and without background DNA (BG).

© 2015 The American Society of Gene & Cell Therapy

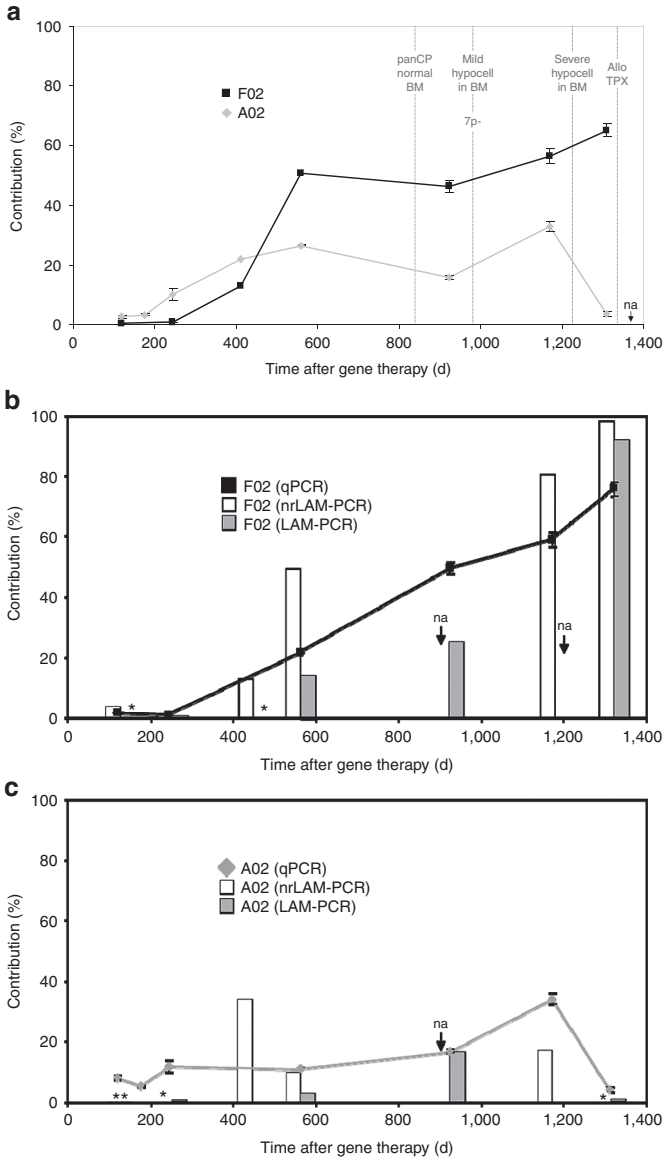

Figure 4 Quantitative polymerase chain reaction (qPCR)-measured kinetics of two clones present in a gene therapy patient. (**a**) *In vivo* growth kinetics of two clones (F02, black curves and A02, gray curves) were followed using qPCR for more than 1,300 days after HSC gene therapy. Here, the clonal contributions in relation to all DNA containing cells in the peripheral blood (*i.e.*, numbers were normalized to hEpoR) are shown. na, sample not analyzed. Correlation of clonal compositions detected. qPCR, LAM-PCR, and nrLAM-PCR-based clonal assessments of clone F02 (**b**) and clone A02 (**c**) are compared. qPCR was performed in triplicates (error bars denoting standard deviations), LAM-PCR and nrLAM-PCR were performed once due to limited availability of template material. * Clone not detected at this time point. na, sample not analyzed.

initial restriction digests, biotin capturing of vector-flanking sites, linker ligation, and fragment amplification.[18,27,29–35] Yet, the use of these techniques for estimating the clonal kinetics of gene-transduced cells relies on three assumptions: (i) a low multiplicity of infection results in ≤1 integration per stem cell, (ii) the genomic position of the integration in the genome of such a progenitor "marks" the clones derived thereof and (iii) for each of the individual clones, the number of recovered identical (redundant) IS sequences (*e.g.*, by next-generation sequencing) rises direct proportionally with its growth. Especially, the latter point is currently controversially discussed as biased IS recovery due to the use of restriction enzymes was reported.[19,20] For example, in a recent study, we investigated MSCV-MGMT^{P140K}-transduced bone marrow progenitors from mice that underwent alkylator treatment.[19] Monoclonality in one mouse was discovered using LM-PCR in combination with (454-based) next-generation sequencing and subsequently confirmed by qPCR. However, when measuring contributions of two nondominant clones, read count analysis showed to considerably underestimate clonal contributions, indicating a skewed image of clonality, most likely due to lack of restriction enzyme cleavage sites around the inserted vectors as already seen in our analyses of the clonal composition of peripheral blood from a patient of the British SCID-X study.[20]

With the development of a nonrestrictive LAM-PCR (nrLAM-PCR), IS recovery has become independent of restriction motif distribution in the host cellular DNA that is flanking an integration.[6,18,36,37] We therefore hypothesized that each of the multiple integrations present in a complex sample should be thus having equal opportunities to be drawn by nrLAM-PCR, allowing a good estimation of the contributions of each clone. To test this hypothesis, we constructed artificial ISs (arIS) of different lengths that are flanked by restriction motifs in 5′ and 3′ direction. These arIS were cloned into standard plasmids and several mixes were made to simulate balanced clonality, as well as oligoclonality of small, medium and large fragment-delivering integrations (*i.e.*, clones). Here, especially LAM-PCR showed the expected restriction and amplification bias.[20] Several reasons could explain this: First, the recovering rate by biotin/streptavidin-based DNA capturing will be considerably lower for larger fragments due to secondary structures that either prevent efficient primer binding or sterically interfere with the magnetic capturing process. Second, the amplification efficiency during nested PCR (after ligation of the adapter) will be eventually lower for larger fragments than for smaller fragments. Although this bias might be neglectable in the range up to 1,000 bp, LAM-PCR has to be seen as a per se competitive reaction. Consequently, the parallel amplification of insertion sites will always result in favored amplification of smaller fragments. Third, the 454-based sequencing which we used to estimate clonality here, also relies on a primer-based immobilization of ssDNA fragments in the reaction chamber, which putatively also favors enrichment of easily accessible (smaller) fragments over larger, potentially convoluted fragments. Consequently, oligoclonal fragments were detected in the two mixes with fragments of 25 and 100 bp flanking site length and underrepresented in an oligoclonal mix with a dominant 500 bp flanking site. This is fully in line with the above mentioned *in vivo* findings and again underlines the necessity to use multiple different enzymes to enable higher genomic accessibility in gene therapy studies.[20] In contrast, nrLAM-PCR results revealed the nature of the clonal mixes we established (R^2~1) and nrLAM-PCR derived clonal contributions were also nearly exactly confirmed in a duplicate run (R^2~1), proving the high reliability and high reproducibility of the assay.

After these encouraging findings we set out to test whether the assay is also revealing a reliable clonal image in an *in vivo* setting. Two patients were enrolled in the German CGD-trial and the analysis of retroviral insertion sites by LAM-PCR has been described before.[5] In one patient, the polyclonal population became clonally restricted and dominated by clones containing retroviral integrations in the MDS1–EVI1 gene locus. We used peripheral blood samples of this patient to quantitatively assess the growth of two clones, one of which was etiologically linked to the development of MDS and AML. The integrations present in the clones were described earlier which

© 2015 The American Society of Gene & Cell Therapy

enabled us to quickly establish site-specific qPCRs which we used as gold standard. Using this "gold standard" we saw the clone that was etiologically linked to the development of MDS and AML made up more than 50% of total hematopoiesis at times long before pancytopenia and hypocellularity in BM were first noted.[17] One might suggest that there is a "threshold clonality" located between 50 and 65% (the timepoint of loss of chromosome 7p), which if crossed, causes the initial symptoms of the disease. This correlates to data from clonality screens in patients with RUNX-mutated acute myeloid leukemia, where clonal sizes reached up to 50% at the time of the initial diagnosis.[38]

When correlating qPCR data to nrLAM-PCR-based clonal assessment, we were only able to perform one run since template material of the clinical study was scarce. However, based on our findings from the duplicate runs, we assumed that even repetitive sampling will not reveal significant deviations. Although the correlation was much weaker than in the highly artificial *in vitro* setting, possibly indicating the strong influence of background DNA on the assay, oligoclonal outgrowth (of clone F02) as well as normal clonal fluctuations (clone A02) were distinguishable. In contrast to recent comparisons of qPCR- and LM-PCR/454-based clonal assessments, clonality is rather "over-estimated" by using nrLAM-PCR and 454 sequencing.

Taken together, we here present a valid strategy for large-scale clonal assessments in gene therapy studies. Our findings on two clones present in a gene therapy patient strongly advocate the necessity to tightly track and robustly quantify suspicious clones in current and future gene therapy studies.

MATERIALS AND METHODS

Design of artificial IS standards

Five artificial IS standards (arIS) were designed (Supplementary Figure S1a), each consisting of the 3′ end of a MLV-derived LTR followed by a 50–500 bp spanning flanking sequence that is full-length mapping to the human genome (Supplementary Table S1). The flanking sequences had to fulfil several criteria: a BLAST or BLAT of a sequence[39,40] had to retrieve an unambiguous (max score >> 2nd score) or unique hit in the human genome, and match with 100% identity (Supplementary Table S2). Furthermore, the sequences had to have low rates of repetitive element content. All arIS were then *in vitro*-synthesized into the multiple cloning site of a standard cloning vector (Eurofins MWG GmbH, Ebersberg, Germany) with *Ple*I and *Bsm*AI sites flanking each arIS ensuring that the whole arIS fragment can be cut out using any of these two enzymes. Each sequence was validated by Sanger sequencing. Finally, the plasmids were amplified in One shot *Escherichia coli* (Invitrogen GmbH, Karlsruhe, Germany) and plasmid preparations were serially diluted to a stock concentration of 10^7 copies/µl. Each plasmid elution was double checked in quality by a NanoDrop Photometer and a 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany). Absolute numbers of arIS plasmid copies were then calculated based on the concentration and the molecular weight of the plasmid: $MW \equiv BPS_{Plasmid+Insert} \cdot 330Da \cdot 2$. Next, the weight per molecule (WPM in g/molecule) was calculated by applying the formula $WPM \equiv \frac{MW}{6,022 \cdot 10^{23}}$. The concentration c (in ng/µl) of the plasmid stock was used to calculate the copy number (1/µl) as $CN \equiv \frac{c}{WPM}$.

Mixtures of artificial IS standards

We established eight different mixes of arIS (A7, H, I, J, L, L+BG, O, O+BG), whereas A7 reflected balanced clonal contributions and mixes H, I and J clonal imbalances. Mixes L, L+BG, O, and O+BG were established to simulate minor clonal contributions (<1%) either with (+BG) or without addition of 100 ng/reaction background DNA. Each mixture was confirmed by quantitative real-time PCR in triplicates.

Real-time quantitative PCR of arIS

We used a modification of a quantitative real-time PCR (qPCR) assay as described before by our group[23,24] to quantify the different mixes described above. In brief, serially diluted arIS mixes, ranging from 10^7 to 10^2 copies/µl, were generated and assayed in triplicates on a 384-well plate in a qPCR (LightCycler 480 Real-Time PCR System, Roche Diagnostics, Mannheim, Germany) using the QuantiTect SYBR Green PCR master mix (Qiagen, Hilden, Germany) and 5 pmol of corresponding primers (Supplementary Table S3). The following protocol was used: activation 15 minutes at 90 °C, followed by 40 amplification cycles (95 °C for 15 seconds and 60 °C for 60 seconds). Crossing point (C_p) values were determined automatically using the second derivative maximum mode (see LightCycler 480 Handbook for details).

Recovery of ISs with LAM-PCR and nrLAM-PCR

LAM-PCR and nrLAM-PCR were performed as extensively described before.[18,27] For the recovery of arIS fragments with LAM-PCR, we used *Ple*I and *Bsm*AI. The primers arIS-fw-bio were used for initial capturing of cleaved fragments, arIS-fw-1 and LC1 for all first exponential PCRs, arIS-fw-2 and LC2 for second exponential PCRs and bar-coded primers TitA1-4 together with TitB-linker for preparation of 454-sequencing (Supplementary Table S4).

IS counting based on 454 reads

All LAM- and nrLAM-PCR-recovered amplicons were purified and "tagged" for subsequent 454-sequencing (Roche/454) using a bar coding system suggested by the manufacturer. Each run delivered several thousand sequence reads, which we initially subjected to trimming and processing tools described before.[26,41,42] In brief, the tools map each sequence read to the most recent version of the human genome and subsequently count identical (redundant) reads, which allows to subsequently calculate read ratios.[19]

qPCR on patient samples

Genomic DNA was extracted from whole blood using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). Integration-site specific qPCR was then carried out with 100 ng DNA in a volume of 10 µl under similar conditions applied for the qPCR of arIS (see above). Two virus-genome junctions (*i.e.*, two clones) detected with nrLAM-PCR were amplified from nested PCR products (see above) with the LTR-specific forward primer arIS-fw and the two flanking site-specific reverse primers F02-A (5′-GGAATAAGTAACAGCAGATGGTTGAG) and A02-A (5′-GTTTTCAGAAAAAAAAG TTCTACAGG). To relate the frequency of each clone on the total cell pool we used a qPCR fitted to a genomic fragment of the human erythropoietin receptor (EpoR) gene using the primers Epo-fw (5′-CTGCTGCCAGCTTTGAGTACACTA) and Epo-rv (5′-GAGATGCCAGAGTCAGATACCACAA). To analyze the frequency of each clone in relation to the transgenic pool we amplified a part of the $gp91^{phox}$ cDNA using the primers gp91-f (5′-GGTTTTGGCGATCTCAACAGAA-3′) and gp91-r (5′-TGTATT GTCCCACTTCCATTTTGAA-3′). Following cloning of all PCR fragments (pGEM-T Easy Vector System, Promega, Mannheim, Germany) and transformation of plasmids into chemically competent *E. coli*

© 2015 The American Society of Gene & Cell Therapy

(DH5α Subcloning Efficiency, Invitrogen, Carlsbad, CA), insert-positive plasmids were extracted and purified after overnight culture (peqGOLD Mini Prep Kit, Peqlab, Erlangen, Germany). Plasmid DNA concentration was measured in triplicates using DNA spectrophotometry (NanoDrop Technologies, Wilmington, DE). Absolute numbers of F02, A02, EpoR, and gp91^{phox} plasmid copies were then calculated as described above.

Mathematical considerations for clonal size calculations *in vivo*

Clonal contributions have to be calculated, whereas the individual contribution C_{Corr} of a clone (the copy count is expressed as N_{Clone}) to the corrected cell pool $N_{Gp91}{}^{phox}$ should be expressed as

$$C_{Corr} = \frac{N(Clone)}{N(Gp91^{phox})} \cdot 100\% \text{ with the corresponding standard deviation}$$

$$SD\{C_{corr}\} \approx \left(\frac{N_{clone}}{N_{Gp91^{phox}}}\sqrt{\left(\frac{SD\{N_{clone}\}^2}{N_{clone}^2} + \frac{SD\left\{N_{Gp91^{phox}}\right\}^2}{N_{Gp91^{phox}}^2}\right)}\right).$$

C_{Total} represents the share of a clone on all DNA-containing cells (0.5 N_{hEpoR} due to biallelic presence of the human EpoR gene in euploid human cells) with $C_{Total} = \frac{N(Clone)}{0.5N(EpoR)} \cdot 100\%$ and the corresponding standard deviation $SD\{C_{total}\} \approx \left(\frac{N_{clone}}{0.5N_{EpoR}}\sqrt{\left(\frac{SD\{N_{clone}\}^2}{N_{clone}^2} + \frac{SD\left\{0.5N_{EpoR}\right\}^2}{0.5N_{EpoR}^2}\right)}\right)$.

ACKNOWLEDGMENTS

This work was supported by grant 0315-452-C of the Federal Ministry of Education and Research (F.G., S.G., I.R., S.L.), grant RO3500/1-1 of the Deutsche Forschungsgemeinschaft (I. R.) and a grant of the Prof. Dr. Karl- und Gerhard Schiller Foundation (F.G., S.L.). F.G. received a fellowship of the Peter and Traudl Engelhorn Foundation. F.G., M.S., and S.L. designed and supervised the project; C.L., S.S., A.P., and B.L. performed molecularbiological experiments, S.G., J.-U.A., and I.R. performed bioinformatical analyses and statistics; F.W., H.G., C.v.K., and M.G. interpreted the data and contributed to the writing of the manuscript; F.G. and S.L. wrote the paper; all authors commented on the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Hacein-Bey-Abina, S, Le Deist, F, Carlier, F, Bouneaud, C, Hue, C, De Villartay, JP *et al.* (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* **346**: 1185–1193.
2. Gaspar, HB, Parsley, KL, Howe, S, King, D, Gilmour, KC, Sinclair, J *et al.* (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* **364**: 2181–2187.
3. Aiuti, A, Cattaneo, F, Galimberti, S, Benninghoff, U, Cassani, B, Callegaro, L *et al.* (2009). Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med* **360**: 447–458.
4. Aiuti, A, Slavin, S, Aker, M, Ficara, F, Deola, S, Mortellaro, A *et al.* (2002). Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* **296**: 2410–2413.
5. Ott, MG, Schmidt, M, Schwarzwaelder, K, Stein, S, Siler, U, Koehl, U *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.
6. Cartier, N, Hacein-Bey-Abina, S, Bartholomae, CC, Veres, G, Schmidt, M, Kutschera, I *et al.* (2009). Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* **326**: 818–823.
7. Boztug, K, Schmidt, M, Schwarzer, A, Banerjee, PP, Díez, IA, Dewey, RA *et al.* (2010). Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N Engl J Med* **363**: 1918–1927.
8. Mitsuyasu, RT, Merigan, TC, Carr, A, Zack, JA, Winters, MA, Workman, C *et al.* (2009). Phase 2 gene therapy trial of an anti-HIV ribozyme in autologous CD34+ cells. *Nat Med* **15**: 285–292.
9. Hacein-Bey-Abina, S, Von Kalle, C, Schmidt, M, McCormack, MP, Wulffraat, N, Leboulch, P *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**: 415–419.
10. Howe, SJ, Mansour, MR, Schwarzwaelder, K, Bartholomae, C, Hubank, M, Kempski, H *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.
11. Hacein-Bey-Abina, S, von Kalle, C, Schmidt, M, Le Deist, F, Wulffraat, N, McIntyre, E *et al.* (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **348**: 255–256.
12. Hacein-Bey-Abina, S, Garrigue, A, Wang, GP, Soulier, J, Lim, A, Morillon, E *et al.* (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**: 3132–3142.
13. Lewinski, MK and Bushman, FD (2005). Retroviral DNA integration–mechanism and consequences. *Adv Genet* **55**: 147–181.
14. Deichmann, A, Hacein-Bey-Abina, S, Schmidt, M, Garrigue, A, Brugman, MH, Hu, J *et al.* (2007). Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J Clin Invest* **117**: 2225–2232.
15. Schwarzwaelder, K, Howe, SJ, Schmidt, M, Brugman, MH, Deichmann, A, Glimm, H *et al.* (2007). Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *J Clin Invest* **117**: 2241–2249.
16. Baum, C, Düllmann, J, Li, Z, Fehse, B, Meyer, J, Williams, DA *et al.* (2003). Side effects of retroviral gene transfer into hematopoietic stem cells. *Blood* **101**: 2099–2114.
17. Stein, S, Ott, MG, Schultze-Strasser, S, Jauch, A, Burwinkel, B, Kinner, A *et al.* (2010). Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat Med* **16**: 198–204.
18. Paruzynski, A, Arens, A, Gabriel, R, Bartholomae, CC, Scholz, S, Wang, W *et al.* (2010). Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nat Protoc* **5**: 1379–1395.
19. Giordano, FA, Sorg, UR, Appelt, JU, Lachmann, N, Bleier, S, Roeder, I *et al.* (2011). Clonal inventory screens uncover monoclonality following serial transplantation of MGMT P140K-transduced stem cells and dose-intense chemotherapy. *Hum Gene Ther* **22**: 697–710.
20. Gabriel, R, Eckenberg, R, Paruzynski, A, Bartholomae, CC, Nowrouzi, A, Arens, A *et al.* (2009). Comprehensive genomic access to vector integration in clinical gene therapy. *Nat Med* **15**: 1431–1436.
21. Baker, M (2011). qPCR: quicker and easier but don't be sloppy. *Nat Meth* **8**: 207–212.
22. Bustin, SA, Benes, V, Garson, JA, Hellemans, J, Huggett, J, Kubista, M *et al.* (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**: 611–622.
23. Bozorgmehr, F, Laufs, S, Sellers, SE, Roeder, I, Zeller, WJ, Zeller, WJ *et al.* (2007). No evidence of clonal dominance in primates up to 4 years following transplantation of multidrug resistance 1 retrovirally transduced long-term repopulating cells. *Stem Cells* **25**: 2610–2618.
24. Nagy, KZ, Laufs, S, Gentner, B, Naundorf, S, Kuehlcke, K, Topaly, J *et al.* (2004). Clonal analysis of individual marrow-repopulating cells after experimental peripheral blood progenitor cell transplantation. *Stem Cells* **22**: 570–579.
25. Bushman, F, Lewinski, M, Ciuffi, A, Barr, S, Leipzig, J, Hannenhalli, S *et al.* (2005). Genome-wide analysis of retroviral DNA integration. *Nat Rev Microbiol* **3**: 848–858.
26. Appelt, JU, Bartholomae, CC, Gabriel, R, Paruzynski, A, Gustafson, D, Cartier, N *et al.* (2011). Bioinformatical clonality analysis of next generation sequencing derived viral vector integration sites. *Hum Gene Ther Methods* **23**: 111–118.
27. Schmidt, M, Schwarzwaelder, K, Bartholomae, C, Zaoui, K, Ball, C, Pilz, I *et al.* (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* **4**: 1051–1057.
28. Biasco, L, Baricordi, C and Aiuti, A (2012). Retroviral integrations in gene therapy trials. *Mol Ther* **20**: 709–716.
29. Laufs, S, Gentner, B, Nagy, KZ, Jauch, A, Benner, A, Naundorf, S *et al.* (2003). Retroviral vector integration occurs in preferred genomic targets of human bone marrow-repopulating cells. *Blood* **101**: 2191–2198.
30. Li, Z, Düllmann, J, Schiedlmeier, B, Schmidt, M, von Kalle, C, Meyer, J *et al.* (2002). Murine leukemia induced by retroviral gene marking. *Science* **296**: 497.
31. Kustikova, OS, Baum, C and Fehse, B (2008). Retroviral integration site analysis in hematopoietic stem cells. *Hematopoietic Stem Cell Protocols* **430**: 255–267.
32. Wu, C, Jares, A, Winkler, T, Xie, J, Metais, JY and Dunbar, CE (2013). High efficiency restriction enzyme-free linear amplification-mediated polymerase chain reaction approach for tracking lentiviral integration sites does not abrogate retrieval bias. *Hum Gene Ther* **24**: 38–47.
33. Cornils, K, Bartholomae, CC, Thielecke, L, Lange, C, Arens, A, Glauche, I *et al.* (2013). Comparative clonal analysis of reconstitution kinetics after transplantation of hematopoietic stem cells gene marked with a lentiviral SIN or a ?-retroviral LTR vector. *Exp Hematol* **41**: 28–38.e3.
34. Uren, AG, Mikkers, H, Kool, J, van der Weyden, L, Lund, AH, Wilson, CH *et al.* (2009). A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nat Protoc* **4**: 789–798.
35. Brady, T, Roth, SL, Malani, N, Wang, GP, Berry, CC, Leboulch, P *et al.* (2011). A method to sequence and quantify DNA integration for monitoring outcome in gene therapy. *Nucleic Acids Res* **39**: e72.

© 2015 The American Society of Gene & Cell Therapy

36. Biffi, A, Montini, E, Lorioli, L, Cesani, M, Fumagalli, F, Plati, T *et al.* (2013). Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* **341**: 1233158.
37. Arens, A, Appelt, JU, Bartholomae, CC, Gabriel, R, Paruzynski, A, Gustafson, D *et al.* (2012). Bioinformatic clonality analysis of next-generation sequencing-derived viral vector integration sites. *Hum Gene Ther Methods* **23**: 111–118.
38. Kohlmann, A, Nadarajah, N, Alpermann, T, Grossmann, V, Schindela, S, Dicker, F *et al.* (2014). Monitoring of residual disease by next-generation deep-sequencing of RUNX1 mutations can identify acute myeloid leukemia patients with resistant disease. *Leukemia* **28**: 129–137.
39. Altschul, SF, Gish, W, Miller, W, Myers, EW and Lipman, DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
40. Kent, WJ (2002). BLAT–the BLAST-like alignment tool. *Genome Res* **12**: 656–664.
41. Appelt, JU, Giordano, FA, Ecker, M, Roeder, I, Grund, N, Hotz-Wagenblatt, A *et al.* (2009). QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. *Gene Ther* **16**: 885–893.
42. Giordano, FA, Hotz-Wagenblatt, A, Lauterborn, D, Appelt, JU, Fellenberg, K, Nagy, KZ *et al.* (2007). New bioinformatic strategies to rapidly characterize retroviral integration sites of gene therapy vectors. *Methods Inf Med* **46**: 542–547.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

Supplementary Information accompanies this paper on the *Molecular Therapy—Methods & Clinical Development* website (http://www.nature.com/mtm)

© 2015 The American Society of Gene & Cell Therapy