ARTICLE

Genetic barcode sequencing for screening altered population dynamics of hematopoietic stem cells transduced with lentivirus

Daniela B Zanatta[1], Maristela Tsujita[2], Primavera Borelli[2], Rodrigo B Aguiar[1], Daniel G Ferrari[3] and Bryan E Strauss[1]

Insertional mutagenesis has been associated with malignant cell transformation in gene therapy protocols, leading to discussions about vector security. Therefore, clonal analysis is important for the assessment of vector safety and its impact on patient health. Here, we report a unique approach to assess dynamic changes in clonality of lentivirus transduced cells upon Sanger sequence analysis of a specially designed genetic barcode. In our approach, changes in the electropherogram peaks are measured and compared between successive time points, revealing alteration in the cell population. After *in vitro* validation, barcoded lentiviral libraries carrying IL2RG or LMO2 transgenes, or empty vector were used to transduce mouse hematopoietic (ckit+) stem cells, which were subsequently transplanted in recipient mice. We found that neither the empty nor IL2RG encoding vector had an effect on cell dynamics. In sharp contrast, the LMO2 oncogene was associated with altered cell dynamics even though hematologic counts remained unchanged, suggesting that the barcode could reveal changes in cell populations not observed by the frontline clinical assay. We describe a simple and sensitive method for the analysis of clonality, which could be easily used by any laboratory for the assessment of cellular behavior upon lentiviral transduction.

Molecular Therapy — Methods & Clinical Development (2014) **1**, 14052; doi:10.1038/mtm.2014.52; published online 19 November 2014

INTRODUCTION

Since the 1990s, gene therapy using hematopoietic stem cells (HSC) has been effectively applied to the treatment of primary immunodeficiencies. Despite the success of the therapy, some concerns have risen given the development of clonal vector-associated leukemias. The first cases of serious adverse events were reported in patients treated for X-linked severe combined immunodeficiency (X-SCID), where patients had developed leukemia initiated by the induction of LMO2 gene expression caused by the vector insertion.[1,2] Recently, in a gene therapy protocol for the treatment of Wiskott–Aldrich syndrome, several patients also developed leukemia related to LMO2 overexpression associated to the vector.[3,4] Clonal expansion events also have been described in clinical trials for the treatment of chronic granulomatous disease[5] and β-thalassemia.[6]

The vectors most commonly used in these therapies are derived from γ-retroviruses. These vectors have a tendency to insert their provirus in or near cellular promoters,[7] creating the unwanted opportunity for activation of neighboring oncogenes, such as LMO2 in the case of X-SCID gene therapy. The induction of a growth-promoting gene or positive selection for the therapeutic gene can contribute to the clonal expansion of transduced cells. Because of these concerns, the use of γ-retroviral vectors is decreasing, while the use of lentiviral vectors, regarded as safer, is gaining thrust.[8] Though the lentiviral provirus is inserted in host gene sequences, the GC-rich promoter regions are not favored and, thus, activation of endogenous genes and the consequences of insertional mutagenesis

should be reduced.[9] Even so, the use of lentiviral vectors is being closely studied in order to identify unwanted cellular responses initiated by components of the vector and the use of self-inactivating vectors and heterologous internal promoters with weak enhancer activity, such as EF1α, are favored.

Monitoring clonal expansion in populations of transduced cells could provide insight to the vector's impact on cellular behavior. In theory, clonal expansion should precede transformation and, as cited above, not all vectors/treatments actually lead to transformation. Therefore, waiting for transformation to occur as a sign of unwanted cellular activity may occur slowly or not at all. Alterations in the dynamics of cell populations that have been transduced may be difficult to perceive and measure. Unfortunately, there are few simple and quick methods that can reveal and quantify clonal expansion. Some currently used methods, such as LM-PCR and pyrosequencing, reveal population dynamics by detecting the integrated vector.[10–13] While effective, these approaches require skills, equipment, and resources that make them difficult to use in a less specialized laboratory.

We describe here a method for the evaluation of the dynamics of transduced cell populations that relies on simple and readily available techniques. A genetic barcode was designed for this purpose and inserted as a library in a lentiviral vector. While typical barcodes are designed with the four possible nucleotides (A, C, G, or T) for each base, our concept utilizes just two possible nucleotides (*e.g.,* A or C, C or G, G or T, etc.) at each position. When sequenced (Sanger method), the frequency of each base is reflected in the electropherogram by

[1]Center for Translational Investigation in Oncology/LIM 24, Cancer Institute of Sao Paulo, School of Medicine, University of Sao Paulo, Sao Paulo, Brazil; [2]Department of Clinical and Toxicologic Analyses, School of Pharmaceutical Sciences, University of Sao Paulo, Sao Paulo, Brazil; [3]Natural Computing Laboratory, School of Computing and Informatics, Mackenzie University, Sao Paulo, Brazil. Correspondence: BE Strauss (bstrauss@usp.br or bryan.strauss@icesp.org.br)
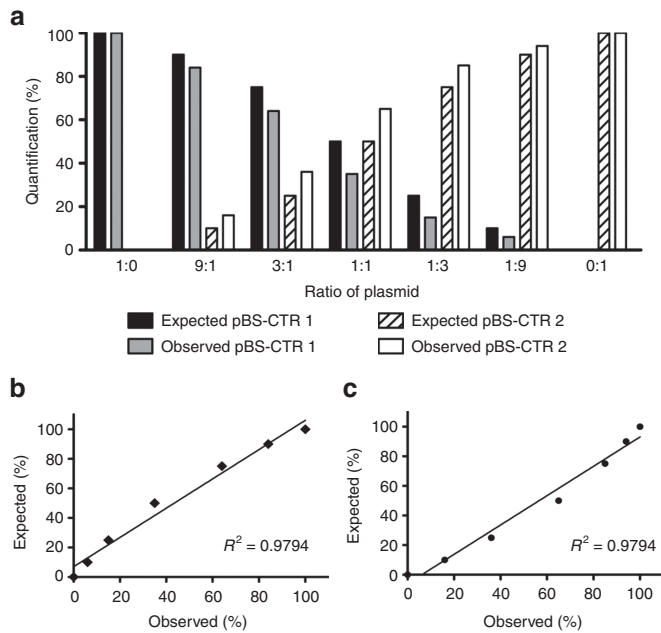
Figure 1 Sequencing of two mixed prototype barcodes yielded quantitative measurements. Two plasmid vectors containing known barcode sequences were mixed at defined proportions and diluted in gDNA. (**a**) Sequencing analysis of the barcode revealed the expected quantity of each plasmid. The values from analysis of the electropherogram were plotted for (**b**) clone pBS-CTR 1 and (**c**) pBS-CTR 2, showing close agreement between the experimental and expected results.
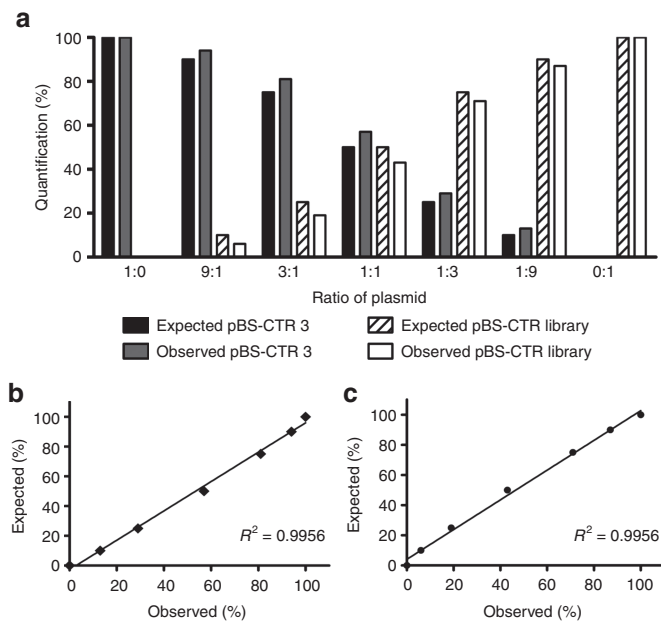

Figure 2 Quantitative results obtained from the prototype barcode library sequencing. A known barcode clone, pBS-CTR 3, was mixed at defined proportions with the plasmid library. (**a**) Sequencing of the barcode revealed the expected quantity of the clone and the library. The values for the (**b**) clone and (**c**) library were plotted, revealing close agreement between the experimental and expected results.

the area under each superimposed peak. If the areas change, we would know that the population dynamics have been altered. In this way, the sequencing of the barcode could reveal, characterize, and quantify the clonal expansion of a transduced cells population both *in vitro* and *in vivo*. As a model system, we have constructed lentiviral

vectors encoding the barcode library and expressing either the murine IL2RG or LMO2 cDNAs and used these to transduce normal hematopoietic stem cells that were then transplanted in recipient mice. Periodic sampling of peripheral blood and sequence analysis of the barcode revealed altered population dynamics in the presence of LMO2, but not IL2RG or the empty vector.

The results show that the barcode is a sensitive and efficient tool for measuring the complexity of cell populations. This methodology will be valuable as a front-line test of new vectors and their components, such as promoters and therapeutic genes, contributing to the development of safer vectors with reduced probability of inducing clonal proliferation events.

RESULTS

Proof-of-principle of the barcode quantification method

To evaluate the ability of the technique to measure cell population complexity, a prototype barcode library was used. The barcode, named CTR (R = A/G), has this triplet repeated five times in its sequence and was cloned into the pBluescript II KS(-) phagemid vector, producing the library named pBS-CTR. After cloning, an individual plasmid clone was isolated and mixed at known proportions with a second isolated clone and diluted in gDNA, so that sequencing of the mixed plasmids should reveal the predicted proportion between each possible base, corresponding to each plasmid clone. The presence of gDNA is meant to simulate the milieu of DNA that would be derived from a transduced cell but does not contain any vector modification. The barcode region was sequenced, and the electropherogram amplitudes used to measure the amount of each clone present in the mix of plasmids (Figure 1a). We found a correlation (R^2 value) of over 97% between the experimental and the predicted results when two known clones were mixed at known ratios (Figure 1b,c).

In order to simulate altered population dynamics, a known clone sequence was mixed at different proportions with the plasmid library. The ratio of variable nucleotides seen in the library after sequencing was used as a baseline measurement. When mixed with a known clone, the expected values contributed by the clone and the library were calculated and compared to the experimental result obtained from sequencing (Figure 2a). In this case, we found a correlation over 99% between the obtained and the predicted results (Figure 2b,c).

Another prototype barcode library AGY (Y = T/C) was constructed and tested with the same methodology as described above (Supplementary Figures S1 and S2), and similar results were obtained. A set of electropherograms obtained from the AGY library mixed with a known clone are shown in Supplementary Figure S3, demonstrating the consistency of the sequencing and also permitting visual assessment of the change in peak areas in relation to the concentration of the clones. Examples of the calculations are also provided in the Supplementary Materials and Methods.

Taken together, these results indicate that Sanger sequencing of the barcode and analysis of the eletropherogram can be used to make a measurement of specific DNA elements in a mixed population.

Barcode evaluation in cell culture

Having validated the barcode technology, a more complex barcode containing 10 variable nucleotides was synthesized and inserted in the Lego-iG lentiviral vector,[14] which was slightly modified (Figure 3). The resulting plasmid library was used to produce virus-containing supernatant.

In this assay, we mixed NIH3T3 cells previously transduced by a single viral clone with cells transduced by the viral library, and the
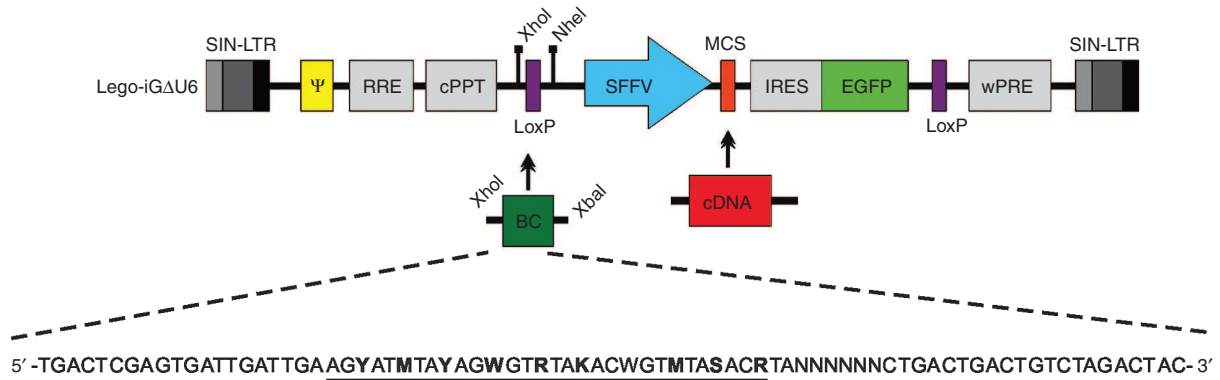
Figure 3 Schematic representation of lentiviral vectors encoding the barcode (with 10 variable nucleotides) as well as genes of interest. Note that the barcode contains *Xho*I and *Xba*I sites for cloning, the TGA stop codon present in all three reading frames both before and after the variable region (underlined), constituted of 10 positions that vary between two nucleotides (where Y=C/T, M=A/C, W=A/T, R=A/G, K=G/T, S=C/G) and 6 hypervariable sites (N=A/C/G/T) that add complexity but were not used in the quantitative analysis. The Lego-iG vector[14] was modified (deletion of the U6 promoter) to receive the barcode (BC). cDNAs of interest, murine IL2RG, or LMO2 were inserted in the polylinker prior to library construction. SIN-LTR, self-inactivating long terminal repeat; RRE, rev responsive element; cPPT, central polypurine tract, SFFV, used as a constitutive promoter, derived from the spleen focus forming virus LTR; IRES, internal ribosome entry site; eGFP, enhanced green fluorescent protein; wPRE, woodchuck hepatitis virus posttranscriptional regulatory element; LoxP, target of the CRE recombinase; MCS, multiple cloning site.
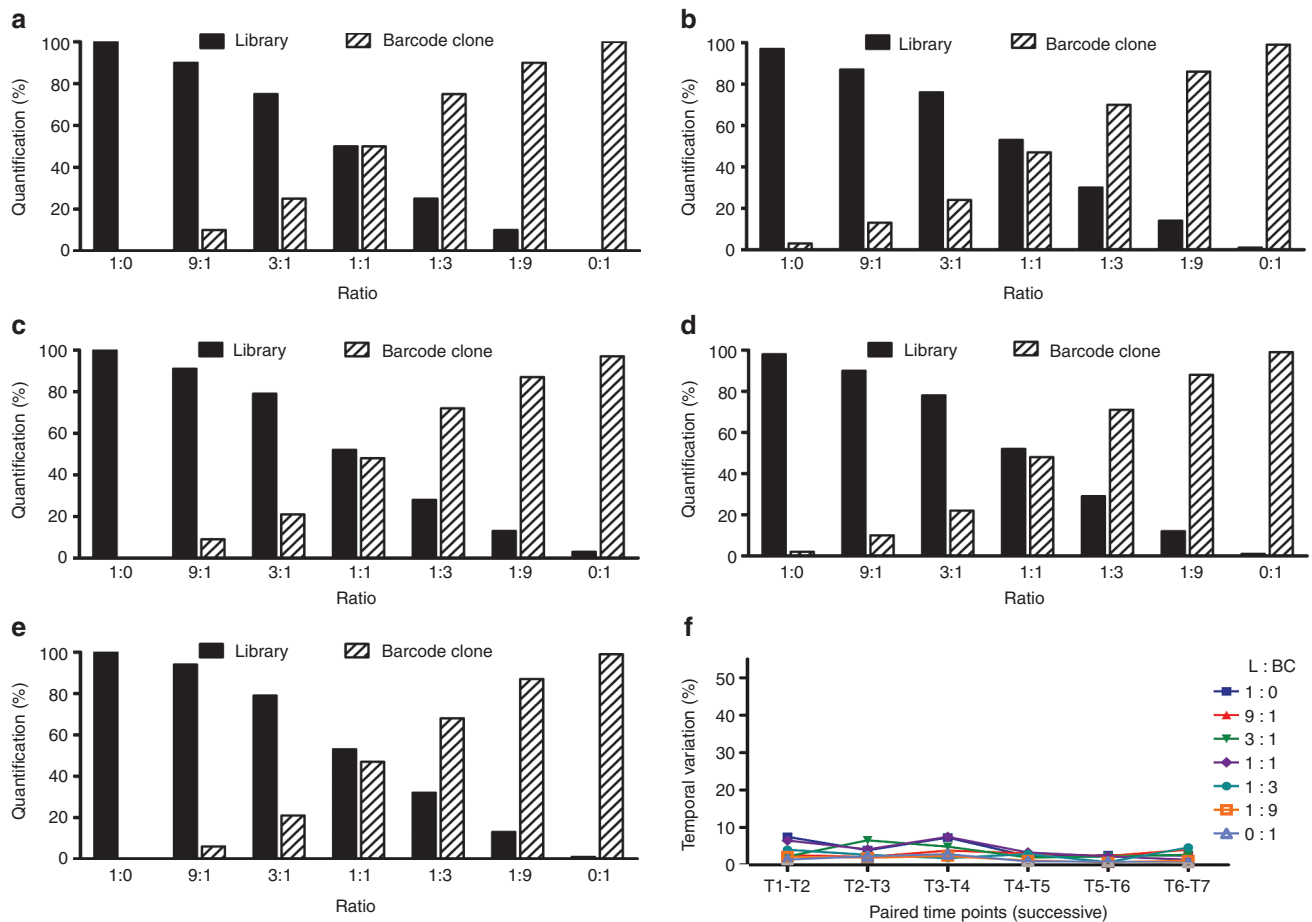

Figure 4 Sequencing of the barcode yielded quantitative results in a tissue culture based assay. NIH 3T3 cells were transduced with either a known lentiviral clone or with the lentiviral library. These cells were then mixed at known proportions, cultivated, and samples removed weekly for sequence analysis. The expected ratios (**a**) of the clone and library were compared with the experimental result generated from weekly collections of gDNA (shown in **b–e** are results form weeks 1, 3, 5, and 7, respectively). (**f**) Stability of lentiviral barcoded cells over time, where the data from all gDNA collections were used to generate a representation of barcode change (population dynamics) over time. Here, comparison of successive time points yielded little change, as seen by the small difference in peak ratios revealed by the sequencing eletropherograms, indicating a stable population of cells.

barcodes were monitored over time (Figure 4). We observed, at each of the seven weekly collections of gDNA, that the experimentally determined values for the barcode closely matched the expected values. This result implies that, as anticipated, neither the lentiviral vector nor the barcode changed the complexity (dynamics) of the cell population over time.
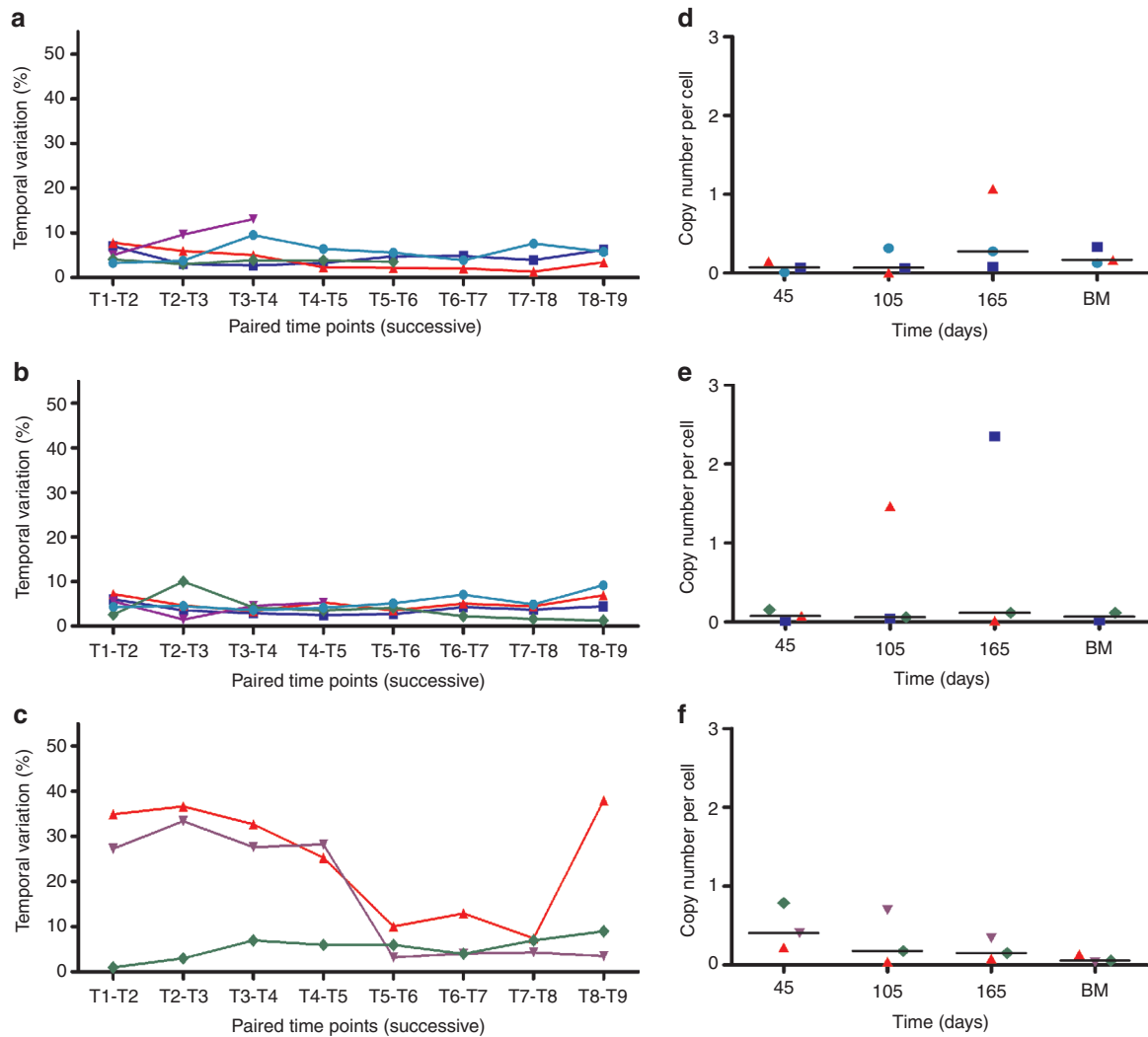
Figure 5 Altered population dynamics revealed through sequencing analysis of the barcode in an *in vivo* experiment. Hematopoietic stem cells were transduced with either the (**a**) lentiviral barcode library or a lentiviral library that contained the barcode as well as the (**b**) IL2RG or (**c**) LMO2 cDNA, before transplantation in recipient mice. Peripheral blood was collected at 2-week intervals from day 60 to day 165 and analyzed. White blood cells were not altered (Supplementary Figure S7), yet the LMO2 gene appeared to induce a change in the population complexity, as indicated by the elevated temporal variance. Panels **d**, **e**, and **f** show the quantification of integrated proviral copies present in peripheral blood collected from the library, IL2RG, and LMO2 groups, respectively, on days 45, 105, and 165 or from bone marrow (BM) harvested on day 240. For this, quantitative PCR was performed using primers specific for the amplification of eGFP present in the lentiviral vectors. Note that each symbol/color combination represents a single animal and permits the comparison between temporal variance and copy number within each experimental group.

The data described above was used to calculate the alterations in observed barcode values over time, and the results are present in Figure 4f. At each variable position, the difference between the peak areas was determined at each time point, and then the difference between successive time points was determined (temporal variance). With little alteration in the population dynamics, the temporal variance of $T_i–T_{i+1}$ would be near zero, yet significant alterations should be seen as higher values (please see Supplemental Materials and Methods for a sample calculation). Since only very small differences were observed, in agreement with Figure 4a, we interpret this as an indication that the cell population was not altered by the virus or the barcode.

In vivo population dynamics

We next sought to demonstrate the efficacy of the barcode in assessing cell population dynamics *in vivo* and without *a priori* knowledge of the barcode sequence. For this, we chose a model

of gene transfer to hematopoietic stem cells (ckit[+], collected from male donor mice) using the lentiviral library as a control or identical constructs encoding the barcode library as well as murine IL2RG or LMO2 cDNAs (transgene expression confirmed as shown in Supplementary Figure S4 and S5), followed by transplantation in irradiated isogenic female mice. Since LMO2 is known to be oncogenic, we expected to find alteration in the population dynamics of daughter cells arising from the stem cells modified with this transgene. However, the clinical observation in the X-SCID trials was expansion of a single cell clone. We feared that a large population of transplanted cells that overexpress LMO2 would yield poly-clonal expansion and thus not reproduce the clinical observation. For this reason, the transplanted mice received a limited number of cells expressing the gene of interest in order to minimize poly-clonal expansion. For the transplant, mice received a mixed population of transduced stem cells where 10% of these had been transduced with the vectors encoding the barcode plus the gene
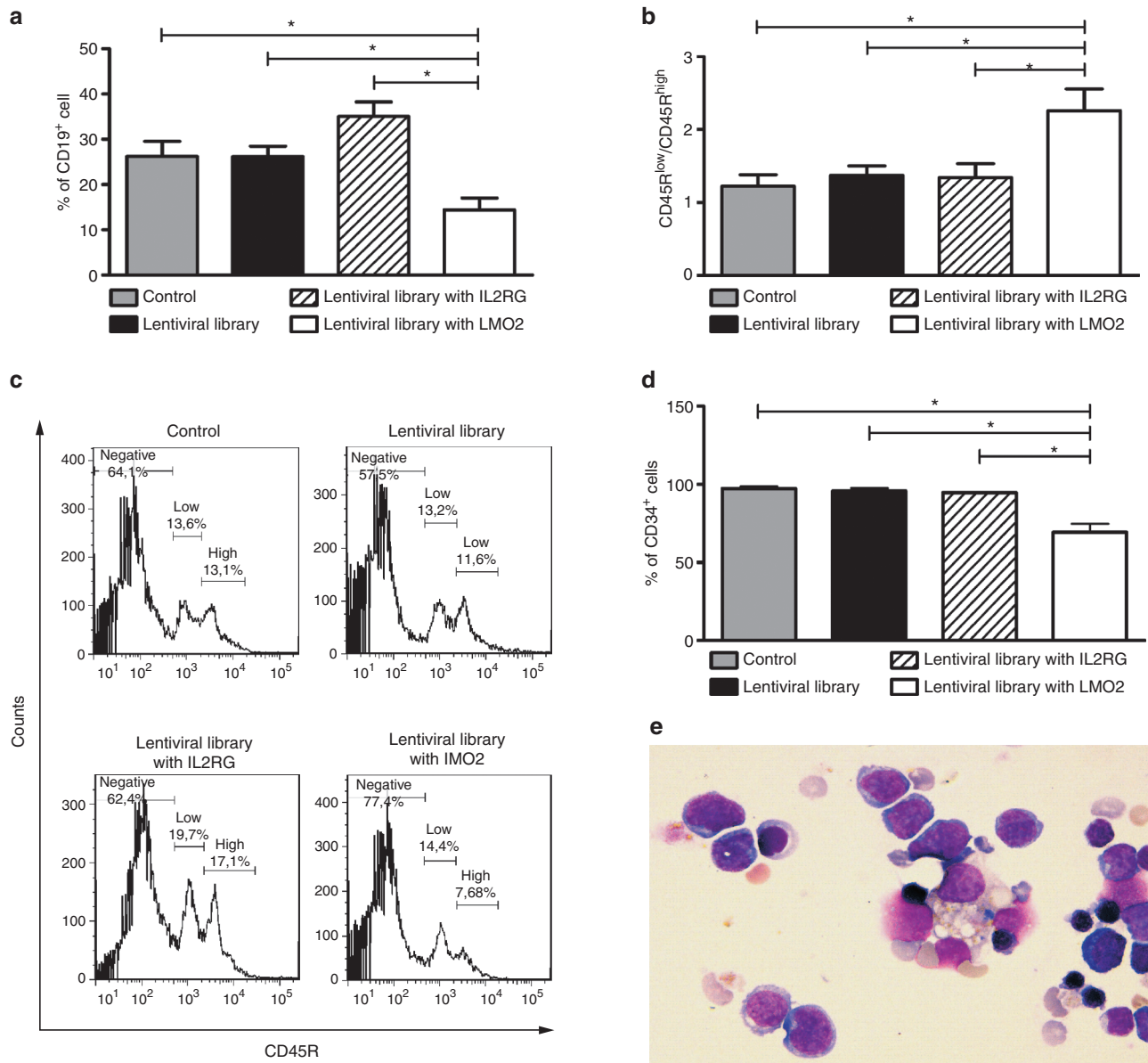
Figure 6 Immunophenotypic analysis of total bone marrow cells recovered from the transplant recipients shows alteration only in the LMO2 group. At day 240 after transplant, animals were euthanized and bone marrow collected for further analysis. Staining with (**a**) anti-CD19, (**b,c**) anti-CD45R (B220) or (**d**) anti-CD34 was performed. The percentage of CD19-positive cells is shown in **a**. The ratio between CD45R^{high} and CD45R^{low} values were determined and plotted in **b** (representative FACS data shown in **c**). The bone marrow was analyzed with the May–Grünwald–Giemsa (×100) staining method, revealing blasts in the LMO2 group consistent with leukemia. A representative image showing blasts among the bone marrow cells is presented in **e** (see Supplementary Materials and Methods for additional information). Significant difference indicated by *$P < 0.05$. Statistical analysis was performed by Mann–Whitney one tail test.

of interest (IL2RG or LMO2) and the remaining 90% of the cells had been transduced with only the barcode library. Transplanted cells of control mice were transduced only with the barcode library. In addition, irradiated mice were kept without transplant as control of engraftment and mice transplanted with nontransduced cells as control for the hematologic exams. Transduction efficiency was determined by flow cytometry detection of GFP in the hematopoietic stem cells just before the transplant, revealing 46.9, 38.9, and 43.3% GFP+ cells for the library, IL2RG, and LMO2 groups, respectively (Supplementary Figure S6).

Our sequencing based method was used to analyze the barcodes in peripheral blood gDNA that was collected at 2 week intervals from day 45 posttransplant to day 165 (Figure 5). The data

show that neither the lentiviral barcode library (Figure 5a) nor the lentiviral library with IL2RG (Figure 5b) impacted cell dynamics. In contrast, the quantification of the barcode in two of the animals that received LMO2 (Figure 5c) shows that the cell population was undergoing significant changes up to the fifth time point, but thereafter, the rate of change fell dramatically before starting to reveal change at the final time point. The remaining animal in the LMO2 group did not show large changes in temporal variance, yet blasts were observed when the bone marrow was examined, as described below. In addition, the N nucleotide positions in the barcode did not reveal complexity in the cell population for this animal. Taken together, we interpret these observations as indications that clonal expansion occurred before the first collection of
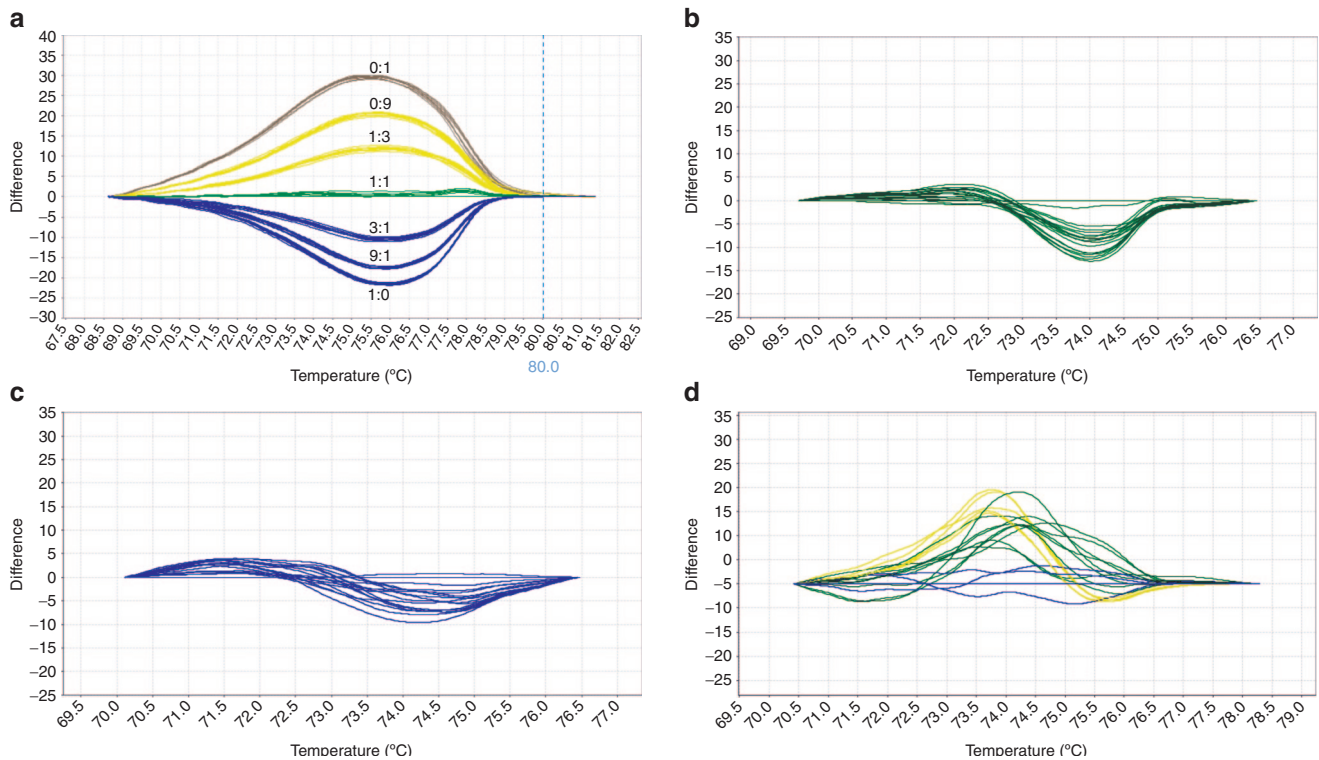
Figure 7 Difference plot from HRM analysis. (**a**) Cell culture experiment where NIH 3T3 cells transduced with either a known lentiviral clone or with the lentiviral library were mixed at defined proportions as indicated, cultivated, and samples removed weekly for sequence analysis. The HRM analyses were applied to samples from weeks 1, 3, 5, and 7, in triplicates. (**b**–**d**) Representative difference plots of the *in vivo* experiments. (**b**) Barcode library transduced only, (**c**) barcode library combined with IL2RG, and (**d**) barcode library combined with LMO2. The HRM analyses were applied to all collected samples (from day 45 to day 165), in duplicates. Analysis of the cell culture experiment revealed that both sequencing and HRM showed stable cell population dynamics, corroborating the result obtained by each approach. *In vivo* analysis showed that IL2RG or empty vector did not provide a selective advantage over time, as expected, since the melt curves exhibited similar shapes for each time point. However, the samples expressing the LMO2 transgene exhibited distinct melt curves, revealing changes in the cell population and indicating clonal expansion. HRM, high-resolution melting.

peripheral blood at day 45. We did not anticipate significant clonal expansion before day 45 since data from the literature indicate this event occurring between weeks 5 and 31 after transplant[15] or the induction of leukemia only after 50 weeks.[16] However, our data suggest that the barcode approach may be especially useful for examining early time points. The average proviral copy number per cell was evaluated by real-time polymerase chain reaction (PCR), and the results show low copy number for all the groups and animals (Figure 5d–f).

The analysis of the white blood cells, lymphocytes, or neutrophils did not display major differences between the groups of animals over time (Supplementary Figure S7). In other words, these hematologic assays did not reveal any pathology, yet the analysis of the barcode revealed that the dynamics of the LMO2-transduced cells could be significantly impacted.

The animals were sacrificed at day 240 for collection of bone marrow cells. Bone marrow was analyzed by immunophenotyping upon staining with CD45R (B220), CD19, CD4, CD8, CD45, CD117, and CD34 antibodies followed by flow cytometry analysis. Cells from animals transduced with lentiviral library containing the LMO2 cDNA showed a lower percentage of CD19 marked cells in relation to the other groups (Figure 6a). In addition, the LMO2 group showed a decrease in the number of cells which express high levels of CD45R. That is to say, the ratio between CD45R^{high} and CD45R^{low} was significantly different between the LMO2 group and the others (Figure 6b,c). Moreover, the bone marrow cells of LMO2 mice presented weak expression of CD34 (Figure 6d). Markers CD117,

CD45, CD4, and CD8 revealed no significant difference in expression among the experimental groups.

In addition, bone marrow morphological study detected the presence of blasts (Figure 6e, Supplemental Materials and Methods) only in the bone marrow aspirates of LMO2 transduced animals, indicating that those animals may have been undergoing leukemogenesis.

High-resolution melting analysis corroborates the sequencing-based assays

The altered behavior observed in the barcoded cells expressing LMO2 was further validated through high-resolution melting (HRM) analysis. HRM is a technique allied with real-time PCR instruments, which make HRM easy to use, sensitive, and low cost, thus it is an interesting tool for molecular diagnosis. It is commonly used to detect small variations in nucleic acid sequences, allowing the detection of mutations, genotyping, and methylation analysis. HRM can also be used as a tool to prescreen library clones.[17] Therefore, we have used HRM to analyze the barcode region and thus to investigate population diversity in the *in vitro* and *in vivo* experiments. *In vitro*, the HRM difference plot of melt curves (Figure 7a) shows that each time point analyzed produced a unique melting plot that was easily distinguished from the others, as expected. Thus, this result demonstrates that a combined approach of DNA barcoding and HRM could be used to detect clonal changes in a cell population. This approach is especially interesting given that single base variations can be identified by HRM.

HRM was then used to analyze the barcode region recovered from the *in vivo* experiment (Figure 7b–d). Through this analysis, we were

able to detect altered population diversity in the cells transduced with the barcoded lentiviral vector containing the LMO2 oncogene (Figure 7d), since the result revealed a wide range of melt profiles in the same two animals that clearly showed changes in temporal variance. Whereas the HRM analysis of cells transduced with IL2RG (Figure 7c) or empty vector (Figure 7b) produced curves that were very similar in shape, as expected, clearly denoting that the samples were homogeneous and not indicating changes in the clonal diversity. Interestingly, the third animal from the LMO2 group that did not show large changes in temporal variance also showed consistent melt profiles by HRM when samples from different time points were assayed. In all, the results from both the barcode sequencing and HRM were in agreement, showing that either assay can reveal changes in transduced cell populations through analysis of the barcode.

DISCUSSION

We have presented here a simple method based on Sanger sequencing of a genetic barcode which reveals the dynamics of cell populations transduced with a lentiviral vector. We tested this technology *in vitro*, in tissue culture, and also *in vivo*. Throughout these assays, it was possible to use data generated from the barcode to analyze molecular and cellular populations and even reveal changes in their dynamics. Specifically, investigation of the lentiviral barcode in the LMO2 group from the *in vivo* assay showed a significant alteration that is consistent with clonal expansion. Interestingly, this change in population dynamics was revealed even when clinical analysis of peripheral blood showed no alteration, suggesting that the barcode could reveal changes in cell populations not observed by the front-line clinical assay. Bone marrow immunophenotyping indicated an accumulation of immature B-cells and myelogram analysis display the presence of blasts seen in the LMO2 group, which corroborate the observation of changes in the cell population. The *in vivo* assay also demonstrated that the lentiviral barcode or the lentiviral barcode with IL2RG cDNA seemed not to induce changes in the cell population. Taken together, this data suggests not only that the barcode method is useful for revealing changes in cell population, but also that the vector, the barcode, and the IL2RG cDNA were shown to be nongenotoxic in this assay.

There is no dispute that LMO2 is an oncogene and can contribute to leukemogenesis. For example, LMO2 was found to be present on several hematolymphoid neoplasias, such as follicular lymphomas, Burkitt lymphomas, lymphocyte-predominant Hodgkin lymphomas, mediastinal large B-cell lymphoma, diffuse large B-cell lymphoma, acute B-cell and myeloid leukemia.[18–20] At early stages of B-cell differentiation and maturation, LMO2 is expressed in the bone marrow and in germinal centers of secondary lymphoid organs and its expression is used to predict survival in patients with diffuse large B-cell lymphoma [21,22] as well as in patients with B-ALL, where LMO2 expression is related to the stage at which blasts undergo differentiation arrest.[23] Therefore, LMO2 behaved as expected in our assay, contributing to altered cell dynamics and, possibly, to abnormal B-cell development.

Changes in the B-cell population were revealed in the LMO2 transplant group upon staining with anti-CD19 antibody, whereas anti-CD45R staining revealed changes in leukocyte maturation. According to Shah *et al*.,[24] the intensity of CD45R increases throughout leukocyte maturation, so mature cells present high intensity (CD45R^high) and immature cells low intensity (CD45R^low) of this marker. Staining of this marker revealed that cell maturation was affected, accumulating immature cells in the bone marrow of the analyzed animals. Moreover, the weak expression of CD19 and CD34 markers in LMO2 expressing animals is consistent with some

cases of acute leukemia, where low atypical expression of these markers is observed.[25,26] These results, together with the bone marrow morphological assay, indicate that the animals expressing the LMO2 transgene probably acquired leukemia.

We admit that the *in vivo* assay would benefit from assessment of a larger number of animals. Additional assays, such as precise characterization of the leukemogenesis, also would have shed light on some of the biological questions surrounding oncogenesis due to LMO2 and/or the use of lentiviral vectors for gene transfer. These additional assays were not performed since we were focused on the barcode method and since there is no doubt that LMO2 can act as an oncogene. In addition, it was not our aim to test cooperation between LMO2 and integration, vector components, or IL2RG activity, all critical factors for understanding the adverse events seen in the SCID-X1 trials and all previously tested and published by other groups. Instead, we present an experimental approach that may well serve as a quick and easy means to measure the alteration in cellular dynamics upon transduction.

This barcode approach may be applicable to *in vitro* assays of genotoxicity, though this was not directly tested in our work. One such assay, IVIM, involves the *in vitro* cultivation of transduced HSCs and interpretation of their transformation or loss of dependence on cytokines as a sign of genotoxicity.[27,28] We feel that our approach offers a distinct benefit since alteration in cellular dynamics may be detected before the emergence of transformed clones. That is to say, the barcode approach may be especially well suited for evaluating earlier events during clonal expansion or cases where full blown transformation is not reached. This feature may be useful when testing specific vector components, such as the promoter, which may contribute to clonal expansion due to activation of genes neighboring the insertion site but may not be sufficient to promote transformation. Our approach may facilitate the comparison of promoters with strong enhancers, such as the SFFV promoter used here, and those with weaker enhancer activity, such as the EF1α promoter often favored for clinical application.

The barcode concept presented here may prove to be a convenient tool for analyzing the cellular response to the vector and its components. Moreover, this simple method does not rely on the development of full blown leukemia in order to reveal alterations in cellular behavior. This may be useful when comparing impact of major components, such as gammaretrovirus versus lentiviral vectors, or of specific vector sequences, such as the promoter, regulatory elements, marker genes, and therapeutic genes. Since the method relies on technologies available to most all laboratories, including clinical centers, this barcode approach may be more accessible, cost effective, and convenient than other technologies that have been described.

For example, a technology based on labeled vectors similar to those used here has been applied to clonal analysis of HSCs.[10] Although the barcode was analyzed by conventional sequencing, the clonal analysis was assessed through cell isolation and subcloning. In other words, complex cellular environments were reduced to clonal populations before analysis. In our approach, the cell population is studied as a whole. Even so, our barcode design could certainly be used in defined cell populations.

Another report has combined the technology of lentiviral vectors containing a barcode library with high-throughput sequencing to track *in vivo* differentiation of HSCs. In this study, it was possible to identify and quantify each bar code recovered from the cell population. The technology was highly sensitive and could be applied in the early identification of oncogenic events in gene therapy protocols, allowing the identification of the integration

8

site and malignant cells.[12] Even though the clonal analysis could be assessed, high-throughput sequencing is relatively costly, and some laboratories may rely on contract services, rather than performing such assays in-house. In comparison, our approach is easily accessible and implemented and may serve as a stepping stone to high-throughput methods.

Recently, Cornils *et al.*[13] have developed a multicolor technique allied with barcode labeling, combined to next-generation sequencing, to monitor individual cell clones. In this work, the clonality was confirmed by analysis of the hematopoietic organs directly. In comparison, our simplified approach assessed the population dynamics through blood sampling and did not require the sacrifice of the animal.

Linker-mediated PCR (LM-PCR) has been used widely to study population dynamics and to determine vector integration sites in transduced cells.[11,29–31] This assay utilizes sophisticated recombinant DNA techniques in order to reveal the provirus–genomic DNA junction. Since several manipulations of the gDNA and PCR products are made, the results yield a qualitative assessment of the number of integration sites, and this is usually visualized as PCR products separated by agarose gel electrophoresis or quantified by next-generation sequencing. While extremely valuable information is generated by this technique, it may be too cumbersome for some labs to perform. While our barcode approach does not replace LM-PCR, it does offer a quick and easy alternative to generate preliminary data that may then justify the investment in LM-PCR.

The design of our barcode also includes some features that may aid in both the construction of the library as well as contribute to the safe use of the barcode *in vivo*. For example, cloning into non-identical, but compatible, restriction sites (such as *Xba*I in the barcode and *Nhe*I in the vector) may be exploited to reduce vectors devoid of the barcode. Upon ligation, but before transformation, the reaction product may be digested with *Nhe*I, thus linearizing only the empty vectors and preventing their propagation in the transformed bacteria. This "kill site" approach was used in the construction of our libraries. In addition, the barcode contained the TGA stop codon in all three reading frames, both before and after the variable sequence. This should ensure that no unwanted peptides are translated. With a similar intent, we also designed the barcode to be devoid of ATG start codons in the variable region.

Clinical application of any barcoded vector implies that a library of sequences must be safely transferred to the patient. We can only speculate on what the issues surrounding such a situation would be, but we call attention to the generation of unwanted peptides, insertion of sequences that attract unwanted molecular responses, such as methylation, and issues regarding the GMP production of nonclonal vectors that contain a loosely defined region in their genome. In our approach, we took measures to ensure that peptides would not be expressed by including stop codons in all three reading frames both before and after the variable region and by excluding the ATG start codon. Since methylation is typically associated with CpG islands, we were careful to not overrepresent these bases in the barcode. Limiting the variable bases to just two options helps to define the sequence in the barcode. Perhaps for clinical application, it would be best to avoid repeated Ns since this would control the generation of start codons and CpG islands. Clinical safety of our barcode approach was not directly examined, but we have no reason to believe that any of the animals suffered due the presence of the barcode in their hematopoietic cells. Insertional mutagenesis is the major concern regarding lenti/retroviral gene therapy products, yet we observed alteration in cellular dynamics only in animals transplanted with cells transduced with the lentiviral vector containing

the LMO2 oncogene, even though a strong enhancer/promoter was present in all conditions. This situation points to the activity of the oncogene as being more critical than either the barcode or the promoter encode by the vector. For clinical production of vectors containing nonclonal sequences, we can only imagine that the vector should be as well defined as possible and that expectations for quality control/assurance should be adjusted to accommodate a final product that is by its very nature variable.

The barcode technology applied here may prove to be useful in non-hematologic applications as well, such as the heterogeneity of solid tumors and the contribution of specific clones to the formation of metastasis. Alternatively, clonality of reprogrammed cells (iPS) may be revealed by transduction of the starting cell population with a vector encoding the barcode or by including the barcode in the reprogramming vectors. Certainly, the safety of using an integrating vector for cellular reprogramming has been questioned,[32–34] but the technology described here may also help to quantify the true impact of the vector on reprogrammed cells.

The barcode sequencing approach described here may prove to be a very useful tool for the development of safer vectors, enabling *in vitro* and *in vivo* preclinical assays of vector toxicity that do not require isolation of clonal cell populations or access to highly specialized resources.

MATERIAL AND METHODS

Library construction

For the construction of the CTR barcode library, the following oligonucleotide was synthesized (International DNA Technologies, Coralville, IA): 5′-GGCAGATCTCTRCTRCTRCTRCTRNANAGAATTCA GG-3′. The complementary sequence (dsDNA) was generated by mixing 0.15 nmol of the barcode oligonucleotide and 0.6 nmol of the primer 5′-CCTGAATTCT-3′, heating the primer mixture for 5 minutes at 95 °C before annealing at 60 °C followed by second strand synthesis by a T4 DNA polymerase (New England Biolabs, Ipswich, MA) reaction. The prototype barcode digested with *Bgl*II and *Eco*RI was cloned into the *Bam*HI and *Eco*RI sites of the pBluescript II KS(-) phagemid vector. Note that the CTR triplet (R=A or G) is present five times and that the NAN (N=any base) variable site is present to increase sequence variation but was not used in the measure of complexity.

The lentiviral barcode was synthetized (5′-TGACTCGAGTGATT GATTGAAGYATMTAYAGWGTRTAKACWGTMTASACRTANNNNNNCT GACTGACTGTCTAGACTAC-3′) and the dsDNA was generated as above, but using the primer 5′- GTAGTCTAGACAGTCAGTCAG -3′. Next, the dsDNA barcode was cloned into the Lego-iG vector[14] kindly provided by Dr Kristoffer Riecken (University Medical Center, Hamburg, Germany). Before introduction of the barcode library into Lego-iG, this vector was modified by removing the U6 promoter. The lentiviral barcode digested with *Xho*I and *Xba*I was cloned into the *Xho*I and *Nhe*I sites of the modified Lego-iG. The murine cDNAs, IL2RG, or LMO2 were isolated from pCMV-SPORT6 clone ID 4924951 and 5349568 (Thermo Scientific Open Biosystems, Pittsburgh, PA) and inserted in the polylinker prior to library construction.

Barcode evaluation experiments

Plasmid DNA (total prototype library or individual clones) was purified with the Wizard Plus Miniprep DNA Purification System (Promega, Madison, WI). Genomic DNA (gDNA) was extracted from HT1080 (human fibrosarcoma) cell line using EZNA Tissue DNA Kit (Omega Bio-Tek, Norcross, GA). The assays in Figure 1 were performed first preparing individual stocks of plasmid/gDNA (0.1 ng/µl

of plasmid to 100 ng/µl of gDNA) for two distinct pBS-CTR clones containing known barcode sequence and then mixing at the ratios of 0:1; 1:9; 1:3; 1:1; 3:1; 9:1; and 1:0 to a final DNA concentration of 10 ng/µl in TE (10 mmol/l Tris, pH 8.0, 1 mmol/l ethylenediaminetetraacetic acid, pH 8.0). For the assays in Figure 2, a pBS-CTR clone was mixed with the total library at the proportions indicated above to a final DNA concentration of 10 ng/µl in TE. In either case, 1 µl of each DNA mix was submitted to PCR amplification and sequencing of the PCR product, as described below.

For the cell culture experiment, NIH 3T3 cells (immortalized mouse fibroblasts) were transduced with either lentivirus containing the total library or with lentivirus containing a single, clonal, known barcode sequence. A week after transduction these two cultures were trypsinized, counted, analyzed for eGFP expression by flow cytometry (to identify transduced cells), and mixed at the proportions of transduced cells of 0:1; 1:9; 1:3; 1:1; 3:1; 9:1; and 1:0, before plating 1×10^6 cells per 10 cm dish. Cells were maintained in complete medium DMEM (Invitrogen Life Technologies, Grand Island, NY) supplemented with 10% bovine calf serum (GE Healthcare Life Sciences HyClone, Logan, UT), 100 µg/ml gentamycin, 50 µg/ml ampicillin, and 2.5 µg/ml amphotericin B, at 37 °C, in a humidified atmosphere of 5% CO_2. Cells were passaged as necessary and, at weekly intervals, gDNA was isolated and the remaining cells maintained in culture. For the sequencing reaction, 100 ng of gDNA was submitted to PCR amplification followed by sequencing, as described below.

Barcode sequencing
The barcode prototype was amplified using the primers M13-20 and M13-R of the pBluescript II KS(-) phagemid vector. To amplify the barcode sequence from the lentiviral vector, primers complementary to internal vector sequences adjacent to the cloning site of the barcode were used (forward: 5′-GCAGGGGAAAGAATAGTAGA-3′ and either reverse: 5′-CGGTGACCATCTGTTCTTG-3′ or 5′-GGGGT TGTGAGCTCTTTTAT-3′). Subsequent to the amplification reaction, a portion was verified by agarose gel electrophoresis, and the remaining product purified with the EZNA Cycle-Pure Kit (Omega Bio-Tek), DNA concentration determined (Thermo Scientific, Wilmington, DE), and 10 ng of the PCR product was sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) with the same primers used in the PCR as well as 5′-TAAGAAACTGCTGAGGGTTG-3′ and analyzed on a 3500 XL Genetic Analyzer (Applied Biosystems).

Electropherogram analysis
The electropherogram from the sequencing step was analyzed with Polyphred to calculate the numerical values relative to the areas under the curves (peaks) of the detected nucleotides. From the full sequence of the barcode, only the paired degenerative nucleotides (not N) were considered for the analysis. For these nucleotides, the values were normalized by the sum of all possible nucleotides present at the degenerate position. The normalization gives a percent value of each possible base for each unique degenerate position of the barcode sequence.

The experimentally obtained values from the sequencing reactions were compared to the expected values, based on the serial dilution used, for each degenerative base. In the prototype barcode experiment (Figure 1), the expected values were based on the dilution proportion of the two clones; *i.e.*, in the proportion of 0:1, we expected 100% of the second clone, and in the proportion of 1:9, we expected 10% of the first clone and 90% of the second one, and

so on. The obtained values were calculated using the mean of the percent values of the known nucleotides of each clone sequence.

For the expected values in Figures 2 and 4, a baseline value for the library was first determined empirically, that is, sequencing the 100% library condition, and then, the expected values for the clone were determined based on the proportion used in the mix with the library. The obtained values for the clone in the mixture were calculated using the mean of the percent values of the known nucleotides of its sequence and removing the mean expected value for the library at the same nucleotides.

To reveal population dynamics over time, the absolute difference between peak areas for each degenerate base (not Ns) was determined at each time point. Then, the absolute difference between successive time points (T_i-T_{i+1}) for each base was determined, and the average difference was calculated and plotted for each successive pair of time points, revealing the temporal variance.

Mice
Female and male C57Bl/6 mice (~7 weeks old) were obtained under specific pathogen-free conditions from the Centro de Bioterismo, FM-USP, and housed in micro-isolator cages. All animal experiments were approved by institutional ethics committees (School of Medicine, USP).

HSC collection, transduction, and engraftment
Bone marrow cells (BMC) were harvested from the femurs and tibias of 8-week-old male mice. BMC were flushed from the bones upon washing using phosphate-buffered saline with 0.5% fetal bovine serum and 2 mmol/l ethylenediaminetetraacetic acid. Red blood cells and bone debris were removed by density gradient centrifugation using Ficoll-Paque (GE Healthcare Life Sciences, Uppsala, Sweden). The cells were then c-kit enriched with CD117 microbeads (AutoMACS; Miltenyi Biotec, San Diego, CA).

Purified cells were transduced in lots of 1×10^5 cells during 20 hours in serum-free IMDM medium (Invitrogen Life Technologies) supplemented with 1% penicillin/streptomycin and cytokines (100 ng/ml murine stem cell factor, 100 ng/ml mFlt3L, 100 ng/ml hIL-11, 20 ng/ml mIL-3; Peprotech, Mexico City, Mexico) as per Almarza *et al.*[35] Cells were transduced in the presence of 8 ng/µl polybrene to facilitate virus infection and with a multiplicity of infection of 50 transducing units/cell. The integrated proviral copy number was verified through quantitative PCR using primers specific for the amplification of eGFP present in the lentiviral vectors. The number of proviral copies was determined in comparison with a standard curve where plasmid DNA encoding the lentiviral vector was used at known concentrations. The calculation of proviral copies was performed as per the Clontech Lenti-X Integration Site Analysis manual, though this kit was not used. The qPCR was performed with SYBR Master Mix (Invitrogen Life Technologies) in a 7500 Real-Time PCR System (Invitrogen Life Technologies).

The day after transduction, cells were counted, pooled, harvested, and engrafted into lethally irradiated (8.5 Gy) isogenic female animals of same age as the donors. Immediately after irradiation, animals were injected intravenously (tail vein) with 5×10^4 HSC cells (with or without transduction, as indicated) in 100 µl of saline solution. In order to serve as a control of the experimental procedure, irradiated animals that did not receive HSC injection were also maintained under the same conditions. All animals were given 2% sucrose and tetracycline (10 mg/ml) in the drinking water during 60 days after transplant to aid in their recovery and as a preventative measure to avoid infections.

All experimental groups were designed to contain five animals at the beginning of the experiments, but one animal that received LMO2 transduced cells died within 2 weeks; 2 animals died during blood collections: one animal from the LMO2 group (60 days after engraftment); one from IL2RG (105 days after engraftment); and two from the Library group (105 and 120 days after engraftment). Since all groups suffered loss, including those that revealed no change in cell dynamics, the cause was likely due to handling and not related to the vector itself.

In vivo barcode analysis

Blood sampling started 45 days after the transplant procedure. Each collection was made from the retro-orbital plexus, and samples were collected at 15-day intervals for gDNA extraction. In addition, at 30-day intervals, peripheral blood was also used for differential white blood cell counts, performed manually with Panotic stained blood smears by the Centro de Bioterismo, FM-USP. Animals (three from each group) were euthanized 240 days after transplant for bone marrow collection.

Immunophenotyping

Flow cytometry was used to perform immunophenotypic analysis of total BMC. Before labeling, red cells were lysed using a hypotonic NaCl solution and washed twice with phosphate-buffered saline/0.5% fetal bovine serum/2 mmol/l ethylenediaminetetraacetic acid. Then, cell suspensions (1×10^6 cells) were incubated with antibodies conjugated to phycoerythrin and specific for CD45R/220, CD19, CD4, CD8, CD45, or CD117; or CD34 conjugated to allophycocyanin. For all antibodies, a correspondent negative control was assessed. After incubation for 20 minutes at room temperature in the dark, the cells were washed with phosphate-buffered saline containing 1% bovine serum albumin, followed by fixation in 1% formaldehyde. All antibodies were products of BD Biosciences (San Jose, CA). 20.000 cells were acquired on a FACSCanto II flow cytometer (Becton Dickinson, Franklin Lakes, NJ) using 488 and 633 nm excitation lasers. Data were analyzed with FlowJo 7.6 software (TreeStar, Ashland, OR).

High-resolution melt curve analysis

The PCR for HRM curve analysis using the MeltDoctor HRM master mix (Applied Biosystems) was performed on a 7500 Fast Real-Time PCR System (Applied Biosystems) in a 20 µl reaction volume that included 20 ng of genomic DNA, 5 µmol/l of each primer (forward: 5'-CGGTGACCCTCGAGTGATT-3' and reverse: 5'-GCCTTGCAAAATGGC GTTAC-3') and 10 µl of MeltDoctor HRM master mix. Dimethyl sulfoxide was added in an effort to increase the amplification of the GC-rich fragments by changing the melting behavior of the DNA. PCR cycling conditions were as delineated by Applied Biosystems: an initial denaturation at 95 °C, followed by 40 cycles of 95 °C for 15 seconds and anneal/extend at 60 °C for 1 minute, followed by one cycle of 95 °C for 15 seconds and 60 °C for 15 seconds for heteroduplex formation. Each DNA sample was analyzed in triplicate (for in vitro experiments) or duplicate (for in vivo experiments). All melting curves were analyzed using the High Resolution Melt Software v3.0.1 (Applied Biosystems). Melting data was normalized, temperature shifted, and displayed as difference plots. Each time point sample was compared to the first collected sample of each animal or to the 1:1 proportion for the cell culture experiment.

Statistical analysis

The GraphPad Prism 5 software (La Jolla, CA) was used for statistical analyses. All comparisons were conducted through a nonparametric statistical test using Mann–Whitney, one-tail, analysis.

REFERENCES

1 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.

2 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.

3 Fischer, A, Hacein-Bey-Abina, S and Cavazzana-Calvo, M (2011). Gene therapy for primary adaptive immune deficiencies. J Allergy Clin Immunol 127: 1356–1359.

4 Braun, CJ, Boztug, K, Paruzynski, A, Witzel, M, Schwarzer, A, Rothe, M et al. (2014). Gene therapy for Wiskott-Aldrich syndrome–long-term efficacy and genotoxicity. Sci Transl Med 6: 227ra33.

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 Cavazzana-Calvo, M, Payen, E, Negre, O, Wang, G, Hehir, K, Fusil, F et al. (2010). Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. Nature 467: 318–322.

7 Wu, X, Li, Y, Crise, B and Burgess, SM (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science 300: 1749–1751.

8 Baum, C, Schambach, A, Bohne, J and Galla, M (2006). Retrovirus vectors: toward the plentivirus? Mol Ther 13: 1050–1063.

9 Baum, C, von Kalle, C, Staal, FJ, Li, Z, Fehse, B, Schmidt, M et al. (2004). Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. Mol Ther 9: 5–13.

10 Gerrits, A, Dykstra, B, Kalmykowa, OJ, Klauke, K, Verovskaya, E, Broekhuis, MJ et al. (2010). Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood 115: 2610–2618.

11 Bartholomae, CC, Glimm, H, von Kalle, C and Schmidt, M (2012). Insertion site pattern: global approach by linear amplification-mediated PCR and mass sequencing. Methods Mol Biol 859: 255–265.

12 Lu, R, Neff, NF, Quake, SR and Weissman, IL (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat Biotechnol 29: 928–933.

13 Cornils, K, Thielecke, L, Hüser, S, Forgber, M, Thomaschewski, M, Kleist, N et al. (2014). Multiplexing clonality: combining RGB marking and genetic barcoding. Nucleic Acids Res 42: e56.

14 Weber, K, Bartsch, U, Stocking, C and Fehse, B (2008). A multicolor panel of novel lentiviral "gene ontology" (LeGO) vectors for functional gene analysis. Mol Ther 16: 698–706.

15 Kustikova, OS, Schiedlmeier, B, Brugman, MH, Stahlhut, M, Bartels, S, Li, Z et al. (2009). Cell-intrinsic and vector-related properties cooperate to determine the incidence and consequences of insertional mutagenesis. Mol Ther 17: 1537–1547.

16 Woods, NB, Bottero, V, Schmidt, M, von Kalle, C and Verma, IM (2006). Gene therapy: therapeutic gene causing lymphoma. Nature 440: 1123.

17 Pepers, BA, Schut, MH, Vossen, RH, van Ommen, GJ, den Dunnen, JT and van Roon-Mom, WM (2009). Cost-effective HRMA pre-sequence typing of clone libraries; application to phage display selection. BMC Biotechnol 9: 50.

18 Natkunam, Y, Zhao, S, Mason, DY, Chen, J, Taidi, B, Jones, M et al. (2007). The oncoprotein LMO2 is expressed in normal germinal-center B cells and in human B-cell lymphomas. Blood 109: 1636–1642.

19 Cobanoglu, U, Sonmez, M, Ozbas, HM, Erkut, N and Can, G (2010). The expression of LMO2 protein in acute B-cell and myeloid leukemia. Hematology 15: 132–134.

20 Agostinelli, C, Paterson, JC, Gupta, R, Righi, S, Sandri, F, Piccaluga, PP et al. (2012). Detection of LIM domain only 2 (LMO2) in normal human tissues and haematopoietic and non-haematopoietic tumours using a newly developed rabbit monoclonal antibody. Histopathology 61: 33–46.

21 Lossos, IS, Czerwinski, DK, Alizadeh, AA, Wechser, MA, Tibshirani, R, Botstein, D *et al*. (2004). Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med* **350**: 1828–1837.

22 Natkunam, Y, Farinha, P, Hsi, ED, Hans, CP, Tibshirani, R, Sehn, LH *et al*. (2008). LMO2 protein expression predicts survival in patients with diffuse large B-cell lymphoma treated with anthracycline-based chemotherapy with and without rituximab. *J Clin Oncol* **26**: 447–454.

23 Malumbres, R, Fresquet, V, Roman-Gomez, J, Bobadilla, M, Robles, EF, Altobelli, GG *et al*. (2011). LMO2 expression reflects the different stages of blast maturation and genetic features in B-cell acute lymphoblastic leukemia and predicts clinical outcome. *Haematologica* **96**: 980–986.

24 Shah, VO, Civin, CI and Loken, MR (1988). Flow cytometric analysis of human bone marrow. IV. Differential quantitative expression of T-200 common leukocyte antigen during normal hemopoiesis. *J Immunol* **140**: 1861–1867.

25 Foucar, K, Reichard, K and Czuchlewski, D. (2010) *Bone Marrow Pathology, vol. 1*. American Society for Clinical Pathology, Chicago. ch. 18, pp. 376–431.

26 Sarry, JE, Murphy, K, Perry, R, Sanchez, PV, Secreto, A, Keefer, C *et al*. (2011). Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J Clin Invest* **121**: 384–395.

27 Modlich, U, Bohne, J, Schmidt, M, von Kalle, C, Knöss, S, Schambach, A *et al*. (2006). Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood* **108**: 2545–2553.

28 Modlich, U, Navarro, S, Zychlinski, D, Maetzig, T, Knoess, S, Brugman, MH *et al*. (2009). Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. *Mol Ther* **17**: 1919–1928.

29 Abel, U, Deichmann, A, Nowrouzi, A, Gabriel, R, Bartholomae, CC, Glimm, H *et al*. (2011). Analyzing the number of common integration sites of viral vectors–new methods and computer programs. *PLoS One* **6**: e24247.

30 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.

31 Schmidt, M, Schwarzwaelder, K, Bartholomae, CC, Glimm, H and von Kalle, C (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.

32 Csete, M (2010). Translational prospects for human induced pluripotent stem cells. *Regen Med* **5**: 509–519.

33 Robbins, RD, Prasain, N, Maier, BF, Yoder, MC and Mirmira, RG (2010). Inducible pluripotent stem cells: not quite ready for prime time? *Curr Opin Organ Transplant* **15**: 61–67.

34 Jalving, M and Schepers, H (2009). Induced pluripotent stem cells: will they be safe? *Curr Opin Mol Ther* **11**: 383–393.

35 Almarza, E, Zhang, F, Santilli, G, Blundell, MP, Howe, SJ, Thornhill, SI *et al*. (2011). Correction of SCID-X1 using an enhancerless Vav promoter. *Hum Gene Ther* **22**: 263–270.

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