

# Barcoding of Macaque Hematopoietic Stem and Progenitor Cells: A Robust Platform to Assess Vector Genotoxicity

Idalia M. Yabe,<sup>1,2</sup> Lauren L. Truitt,<sup>1</sup> Diego A. Espinoza,<sup>1</sup> Chuanfeng Wu,<sup>1</sup> Samson Koelle,<sup>1,3</sup> Sandhya Panch,<sup>1</sup> Marcus A.F. Corat,<sup>1,4</sup> Thomas Winkler,<sup>1</sup> Kyung-Rok Yu,<sup>1</sup> So Gun Hong,<sup>1</sup> Aylin Bonifacino,<sup>1</sup> Allen Krouse,<sup>1</sup> Mark Metzger,<sup>1</sup> Robert E. Donahue,<sup>1</sup> and Cynthia E. Dunbar<sup>1</sup>

<sup>1</sup>Hematology Branch, National Heart, Lung and Blood Institute, NIH, Bethesda, MD 20892, USA; <sup>2</sup>Department of Microbiology and Immunology, Georgetown University Medical Center, Washington, DC 20057, USA; <sup>3</sup>Department of Statistics, University of Washington, Seattle, WA 98195, USA; <sup>4</sup>Multidisciplinary Center for Biological Research, University of Campinas, Campinas, SP 13083-877, Brazil

**Gene therapies using integrating retrovirus vectors to modify hematopoietic stem and progenitor cells have shown great promise for the treatment of immune system and hematologic diseases. However, activation of proto-oncogenes via insertional mutagenesis has resulted in the development of leukemia. We have utilized cellular bar coding to investigate the impact of different vector designs on the clonal behavior of hematopoietic stem and progenitor cells (HSPCs) during *in vivo* expansion, as a quantitative surrogate assay for genotoxicity in a non-human primate model with high relevance for human biology. We transplanted two rhesus macaques with autologous CD34+ HSPCs transduced with three lentiviral vectors containing different promoters and/or enhancers of a predicted range of genotoxicities, each containing a high-diversity barcode library that uniquely tags each individual transduced HSPC. Analysis of clonal output from thousands of individual HSPCs transduced with these barcoded vectors revealed sustained clonal diversity, with no progressive dominance of clones containing any of the three vectors for up to almost 3 years post-transplantation. Our data support a low genotoxic risk for lentivirus vectors in HSPCs, even those containing strong promoters and/or enhancers. Additionally, this flexible system can be used for the testing of future vector designs.**

## INTRODUCTION

The first clinical trials demonstrating clear clinical improvement following transplantation of genetically modified hematopoietic stem and progenitor cells were reported 15 years ago, utilizing  $\gamma$ -retrovirus vectors based on the Moloney murine leukemia virus (MLV) to treat congenital immunodeficiency disorders.<sup>1,2</sup> However, soon thereafter, patients developed both myeloid and lymphoid acute leukemias due to insertional activation of nearby proto-oncogenes by vector enhancer sequences.<sup>3–7</sup> Lentivirus vectors (LVs) with deletion of viral enhancers were already in development, and they have subsequently been used in promising hematopoietic stem and progenitor cell (HSPC) gene therapy clinical trials; but, they have also been linked to at least one instance of clonal expansion.<sup>8–10</sup>

A number of murine models and *in vitro* assays have been explored for the prediction of HSPC genotoxicity, but each has potential limitations. Tracking of insertion sites following transplantation of transduced normal murine HSPCs requires large numbers of mice to track relatively low numbers of clones, and very long-term follow-up and/or secondary transplants are required to uncover clonal dominance and transformation.<sup>11–13</sup> Murine tumor-prone models have more rapid and penetrant tumor onset; however, the genes linked to tumors in these models do not closely match the responsible loci in human clinical trials, potentially decreasing pre-clinical utility.<sup>14,15</sup> *In vitro* immortalization of murine HPSCs has been developed as a more practical and rapid screening approach; however, this assay is difficult to quantitate; captures myeloid, but not lymphoid, transformation; and identifies only insertions activating *Evi1* pathways.<sup>16–18</sup>

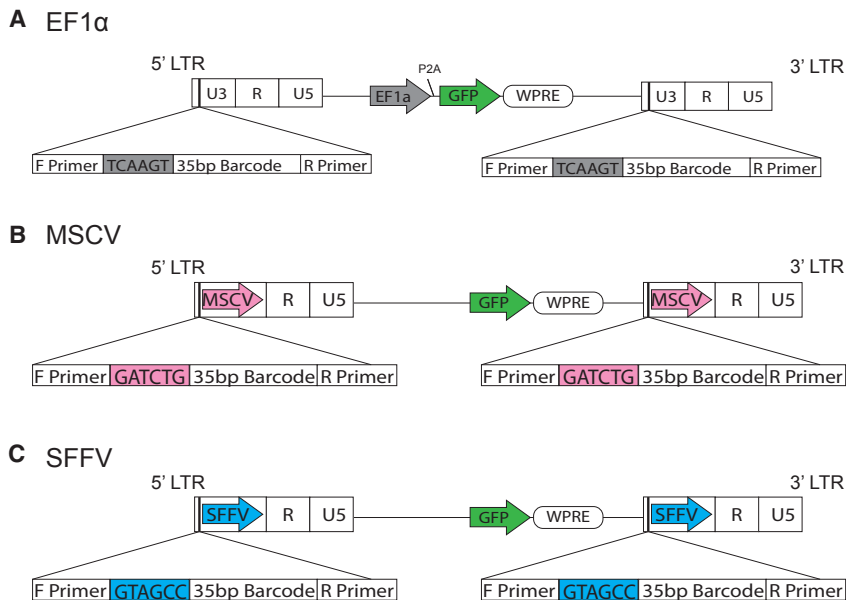
Both *in vivo* and *in vitro* murine assays may lack predictive value for humans due to differences in HSPC properties between species, including lifetime replicative demand, HSPC frequency, ease of immortalization, telomere length, and spectrum of common hematologic tumors.<sup>19</sup> Non-human primates (NHPs) are phylogenetically closely related to humans, and they have been predictive of human clinical results regarding gene transfer efficacy and insertion site patterns.<sup>19,20</sup> While NHP models are expensive and technically demanding, they can be used to study the behavior of thousands of vector insertion sites in a small cohort of animals over time. However, development of tumors in NHPs is rare and requires long-term follow-up;<sup>21,22</sup> thus, an approach to monitor genotoxic clonal expansion prior to actual malignant transformation would be of utility.

Received 30 August 2018; accepted 19 October 2018;  
<https://doi.org/10.1016/j.omtm.2018.10.009>

**Correspondence:** Cynthia E. Dunbar, National Heart, Lung and Blood Institute, NIH, Building 10 CRC Room 4E-5132, 9000 Rockville Pike, Bethesda, MD 20892, USA.

**E-mail:** [dunbarc@nhlbi.nih.gov](mailto:dunbarc@nhlbi.nih.gov)



**Figure 1. Design of Barcoded Vectors**

Three high-diversity barcode libraries were cloned upstream of the U3 region of the long terminal repeat (LTR) in three lentiviral vectors. The structure of the proviral integrated form is shown schematically for each vector: (A) Internal human elongation factor 1- $\alpha$  (EF1- $\alpha$ ) promoter driving GFP expression, (B) murine stem cell virus (MSCV) promoter/enhancer in the LTR driving GFP expression, and (C) spleen focus-forming virus (SFFV) viral promoter/enhancer driving GFP expression.

Insertion site identification and retrieval methodologies have evolved over time from the detection of restriction fragment length polymorphisms to PCR-based approaches for the retrieval of genomic sequences flanking proviral insertions. Linear amplification-mediated PCR (LAM-PCR) and modifications avoiding bias associated with restriction enzyme fragment generation have been the preferred methods for retrieving proviral integrations in both preclinical and clinical studies, and they have proven invaluable for identifying insertions associated with tumors, as well as providing an overview of insertion patterns with various vector classes.<sup>4,23,24</sup> However, these insertion retrieval approaches require large amounts of DNA and are semiquantitative, thus unable to reproducibly quantitate several fold changes in individual clonal contributions.<sup>25</sup> An alternative for clonal tracking is the inclusion of high-diversity barcodes within the vector backbone. In both murine<sup>26</sup> and NHP<sup>27</sup> studies, HSPC clonal dynamics have been mapped in a quantitative and efficient manner using this approach.

We now utilize barcoding to investigate the impact of specific promoter and enhancer elements within HIV-derived LVs on clonal dynamics. We focused on LVs because they are the current standard for efficient transduction of HSPCs in the clinical setting, having replaced MLV due to higher efficiency and lack of overt genotoxicity to date. We assayed vectors containing one of three promoter/enhancers with a range of predicted genotoxicity. Our goal was to develop an approach for screening genotoxicity in an animal model that is predictive of results in human HSPC gene transfer applications.

## RESULTS

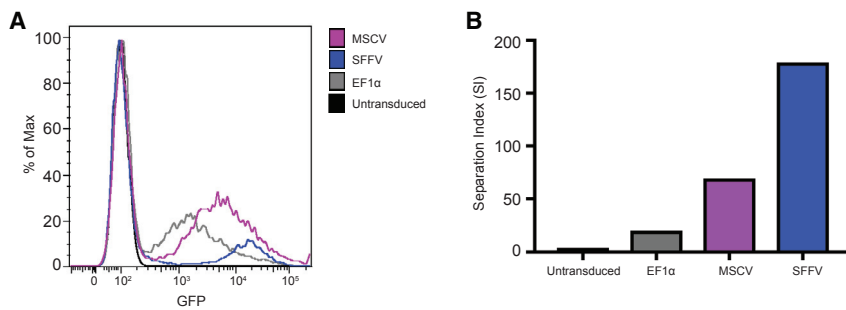
### Development of an *In Vivo* Platform to Study Relative Vector Toxicity

We generated three high-diversity barcode libraries containing one of three 6-bp library IDs followed by 35 random base pair barcodes and

flanked by sequences complementary to Illumina sequencing primers, and we inserted these cassettes into the U3 region of three different lentiviral vector plasmids (Figure 1).<sup>26,27</sup> Each vector also contains an internal GFP marker for the evaluation of transduction efficiencies and tracking. Following integration, the provirus contains the barcode in both long terminal repeats (LTRs) (Figure 1). The vectors were chosen to reflect a predicted range of genotoxicities. The elongation factor-1 alpha (EF1- $\alpha$ ) vector contains an internal relatively weak EF1- $\alpha$  promoter and no additional promoter/enhancer within the LTR, and it would be predicted to be the least genotoxic. The murine stem cell virus (MSCV) vector contains the medium-strength MSCV promoter/enhancer cloned into the U3 region of the LTR. The spleen focus-forming virus (SFFV)-barcoded lentiviral vector contains the very strong SFFV promoter/enhancer within the U3 region of the LTR, predicted to be the most genotoxic LV design based on *in vitro* immortalization and murine tumor-prone mouse assays.<sup>15,28</sup>

We confirmed that all three vectors contained the expected library ID and retained the ability to transduce and integrate in K562 cells, despite insertion of the barcode cassettes into the LTR (Figure 2). Promoter strength in the three vectors was assessed in transduced K562 populations by measuring GFP expression via flow cytometry. As expected, cells transduced with the SFFV vector had the brightest GFP expression, followed by the MSCV vector and then the EF1- $\alpha$  vector with stepwise dimmer GFP expression (Figure 2A). The larger the separation between GFP+ and GFP- populations, the higher the separation index (SI) value and the stronger the promoter/enhancer. As expected, SFFV-transduced K562 cells had the highest SI followed by MSCV and then EF1- $\alpha$  with the lowest SI value (Figure 2B).

Barcode library diversity for each vector construct was assessed via the transduction of K562 cells, barcode retrieval via PCR, and Illumina sequencing. Monte Carlo simulations were performed on K562 cells transduced with each vector preparation to ensure that each of the three barcoded libraries was diverse enough to result in a 95% chance that each barcode was present in a single engrafting HSPC (Figure S1). These simulations indicated that the EF1- $\alpha$ , MSCV, and SFFV lentiviral libraries could transduce 1,547, 1,986, and 1,943 long-term engrafting HSPCs, respectively, with >95%



**Figure 2. Analysis of Comparative Promoter Strength of Each Lentiviral Vector**

(A) Flow cytometric analysis of K562 cells transduced with the different vectors at the same MOI demonstrate varying intensities of GFP expression within the GFP+ cell population, with SFFV > MSCV > EF1 $\alpha$ . (B) Separation Index (SI) was evaluated to compare the distance between positive and negative GFP populations for each vector. The higher the SI value, the larger the separation between the peaks and the stronger the expression from each vector's promoter/enhancer.

confidence (i.e.,  $p < 0.05$ ) that >95% of individual barcodes would mark a single cell. Based on frequencies of long-term engrafting macaque HSPCs within CD34+ populations estimated via integration site retrieval, flow cytometric phenotyping for primitive markers, or immunodeficient mouse engraftment,<sup>23,29,30</sup> the diversities of the three libraries were sufficient to transduce the rhesus macaque CD34+ HSPC grafts utilized in our experimental design.

Peripheral blood CD34+ HSPCs collected following G-CSF/plerixafor mobilization from two rhesus macaques were split into three equal fractions, and each fraction was transduced independently with one of the three lentiviral vectors (Figure 3A). A transduction efficiency of no greater than 30% was targeted in order to favor transduction of individual HSPCs with no more than a single vector copy, containing a unique barcode.<sup>27,31</sup> The three transduced cell fractions were then combined and reinfused simultaneously into the autologous rhesus macaque recipient, following delivery of 1,000 rads ablative total body irradiation to the animal (Figure 3A). Information regarding CD34+ cell collection, transduction, and transplantation for both animals is summarized in Table 1. Following infusion of the transduced CD34+ cells, the animals recovered blood counts promptly. GFP expression in peripheral blood granulocytes, monocytes, B cells, and T cells was assessed over time by flow cytometry, and both animals showed relatively low but stable GFP percentages following engraftment (Figure 3B), well within the target range calculated by Poisson distribution analysis predicting the fraction of HSPCs containing a single barcode (Figure S2).

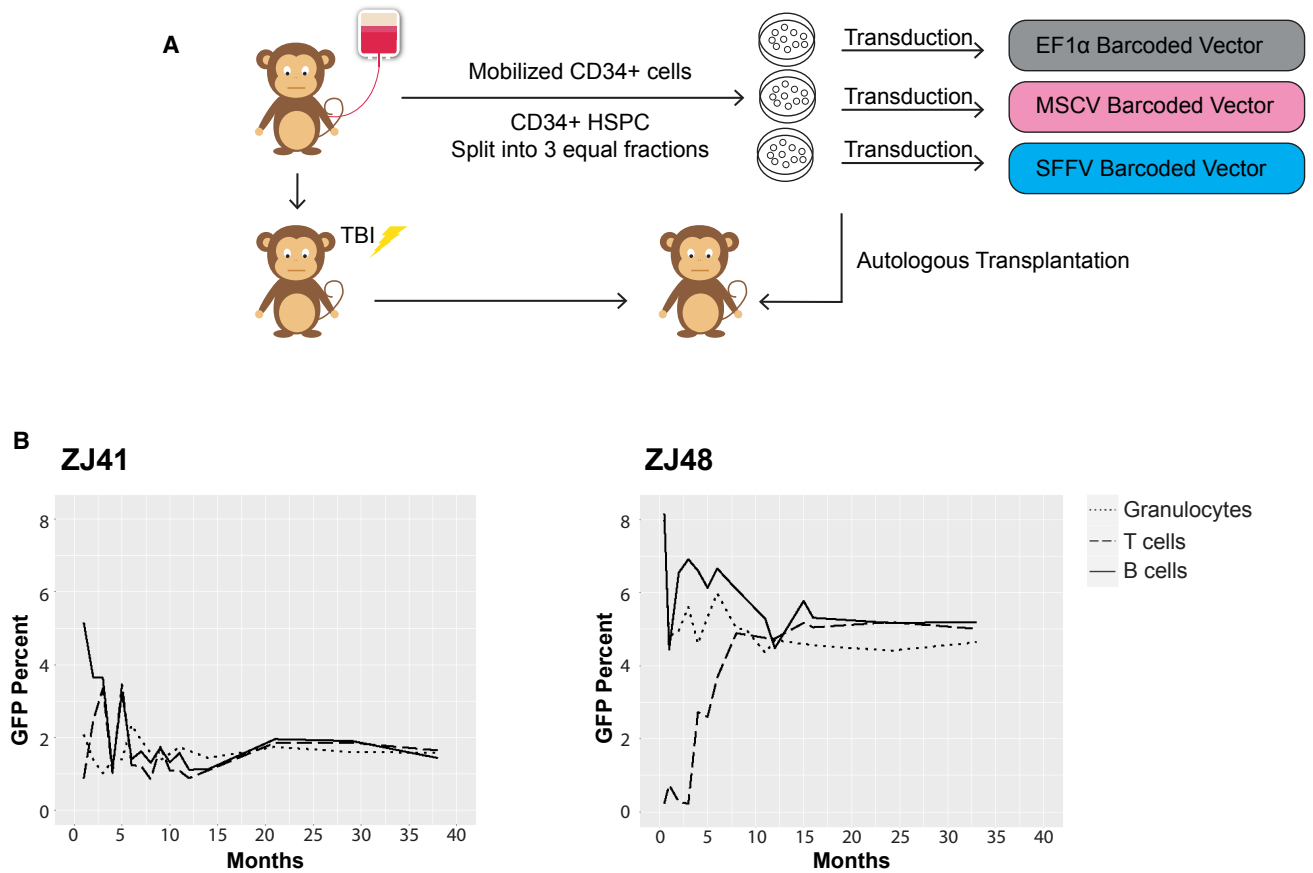
#### Post-transplantation Analysis of Vector Clonal Contributions over Time

To assess clonal patterns *in vivo* at time points from 1 month through 33 and 38 months post-transplantation, blood was obtained, and granulocytes, T cells, and B cells were lineage purified by flow cytometry, followed by DNA extraction, low-cycle PCR with primers bracketing the library IDs and barcodes, high-throughput Illumina sequencing, and data processing to retrieve and quantitate the read fraction for each individual barcode linked to one of the three vector-associated library IDs. To be included on a master clone list for tracking and analysis, a barcode had to be retrieved at a read count of at least 2,000 at one or more time points, as detailed in the [Materials and Methods](#), in order to exclude potential false barcodes and to account for sampling constraints.

The overall level of contribution for each of the three vectors in a sample was assessed via normalization of the read number of each of the three library IDs to the total overall reads containing a valid library ID in that sample. We observed that the relative contributions of each of the three vectors to ongoing hematopoiesis were not uniform in either animal. In ZJ41, the SFFV vector accounted for over 90% of the reads in almost every sample in all three hematopoietic lineages, with far lower contributions from cells containing the MSCV or EF1- $\alpha$  vectors. In contrast, in animal ZJ48, the MSCV and EF1- $\alpha$  contributions were much higher and quite similar, with much lower contributions from the SFFV (Figure 4A). We have no clear explanation for the observed differences in transduction efficiency of engrafting cells between the different vectors within each animal; but, given that each vector performed well in at least one animal, the variability does not reflect vector design and instead likely resulted from some unknown characteristic of the actual viral preparation used for transplantation or variability in handling of the cell fractions during transduction with each vector. To rule out any technical variability or bias regarding barcode retrieval from each of the three vectors, a range of mixtures of single-copy barcoded K562-transduced cell clones for each of the three vectors was conducted (Figure S3). Barcode retrieval quantitation was closely correlated with the expected barcode frequency based on Pearson correlation coefficients of 0.9961, 0.999, and 0.9994 for EF1- $\alpha$ , MSCV, and SFFV, respectively, as our lab has previously reported for other vector designs.<sup>27</sup>

#### Clonal Diversity and Clonal Dynamics over Time

Quantitative tracking of clonal contributions from each unique barcode allows tracking of output from large numbers of individual HSPCs, and it allows for the assessment of clonal dominance linked to LV design. Both animals had individual barcode contributions assessed for each vector at each time point through 33–38 months. The number of unique barcodes obtained from granulocytes, B cells, and/or T cells at each time point for the given vectors was counted (Figure 4B). At month 1, a total of over 2,000 unique barcoded contributing HSPC clones, derived from all three vectors, was detected in both ZJ41 and ZJ48. By 2–3 months, the number of barcoded contributing clones detected decreased, as we have previously reported, as contributions from waves of abundant short-term repopulating HSPCs declined and were replaced by contributions from less abundant intermediate and long-term repopulating HSPCs.<sup>27,32</sup> Thereafter, clone numbers contributing at later time points remained



**Figure 3. Experimental Design and GFP Marking Post-transplantation**

(A) Experimental design. Rhesus macaque peripheral blood stem and progenitor cells were mobilized into the blood with G-CSF and plerixafor and collected via apheresis. CD34+ HSPCs were enriched via immunoabsorption. CD34+ cells were split into 3 equal aliquots and each fraction was transduced with one of three LVs as shown. Following transduction, the cells were collected, combined, and reinfused into the autologous macaque after the macaque completed total-body irradiation (TBI) of 1,000 rads. (B) Percentage of GFP<sup>+</sup> cells in peripheral blood granulocytes, B cells, and T cells tracked over time to the longest follow-up post-transplantation in animals ZJ41 (38 months) and ZJ48 (33 months).

relatively stable, with some late gradual decline in both animals, likely due to a loss of intermediate-contributing clones. The cumulative number of barcodes retrieved plateaued from all three vectors in both animals by 6–12 months post-transplantation, indicating clone retrieval is efficient and that a few dormant HSPCs were present that began contributing only at later time points (Figure 4C). As expected, application of different thresholds for the inclusion of clones in the enumeration of clone numbers altered the numbers of clones retrieved, but it did not change the overall patterns obtained over time, the critical issue for genotoxicity analysis (Figure S4).

Although each animal had differences in the absolute number of clones tracked among the three vectors, similar total numbers of clones were tracked and assessed for clonal expansion: 7,314 for SFFV, 5,488 for MSCV, and 4,450 for EF1- $\alpha$  when using a threshold of 2,000 reads at one or more time points to be included in the master clone list. While obtaining similar marking levels for all three vectors within an animal would have been ideal, it is not crucial since the

barcode permits tracking of contributions from each individual HSPC clone independently.

As an overall population measure of *in vivo* clonal diversity, both the Gini-Simpson's and Shannon's diversity indexes were calculated. Both Simpson and Shannon indexes measure diversity by taking into account the richness (the number of barcodes present) and evenness of the measured barcodes. However, the Gini-Simpson diversity index provides the option to calculate diversity solely on the evenness of the barcodes measured, independent of sample richness. Because vector marking was not equally balanced among the 3 different vectors in both animals, the Gini-Simpson diversity index provides a more applicable representation in the context of our data. As shown in Figure 4D, the Simpson diversity index for the vectors with reasonable HSPC engraftment in ZJ41 (SFFV) and ZJ48 (MSCV and EF1- $\alpha$ ) maintained a high-diversity index for all three lineages. For comparison, the more commonly used Shannon diversity index was also calculated and demonstrated similar results (Figure S5) as the

**Table 1. Macaque Information**

Animal ID	ZJ41	ZJ48
Sex	male	female
Date of birth	6/29/2011	7/12/2011
Weight (kg)	5.82	4.86
Mobilization cytokines	G-CSF, AMD3100	G-CSF, AMD3100
Date of transplant	2/11/2015	7/15/2015
Transduction efficiency	NA	SFFV (20.3%)
		MSCV (6.82%)
		EF1a (7.95%)
CD34+ cells transduced	3.60E+07	2.82E+07
CD34+ cells infused	6.00E+07	2.84E+07
Age day of transplant	43 months 13 days	48 months 3 days
Infused cells/kg	1.03E+07	5.84E+06
SFFV virus titer	8.10E+07	2.90E+08
MSCV virus titer	1.90E+08	1.30E+08
EF1a virus titer	2.50E+08	2.60E+08

Gini-Simpson, where all three libraries with reasonable HSPC engraftment maintained a high-diversity index over time, with no evidence for progression to oligoclonality with any of the three vectors.

#### Clonal Stability of Vectors by Their Top-Contributing Clones

To identify clonal expansion, we evaluated the top-contributing clones for each of the vectors in each lineage, focusing on the largest clones at the time point of longest follow-up as representing potentially expanding clones and tracking each of these clones over all time points. Due to their short half-life and lack of peripheral expansion, granulocytes best represent ongoing hematopoiesis from HSPCs in the context of this study; however, we also performed these analyses on B cells and T cells. Figure 5 shows the contributions mapped back over time from the top 20 clones detected in each lineage for each vector at the latest time point of 33 and 38 months, respectively, in the two animals. As shown in Figure S4, these plots are virtually identical when applying higher or lower thresholds for the inclusion of clones, indicating that the setting of thresholds only impacts clones contributing very low levels at all time points.

Long-term engrafting clones, as expected from our prior study of the dynamics of HSPC clonal reconstitution,<sup>32</sup> first appeared and began replacing contributions from shorter-term progenitors between 2 and 12 months. The contributions from these long-term clones increased sharply at the time of appearance as expected, and then they began to plateau. There was some relative increase in contributions for all these large long-term clones over time as short- and intermediate-term progenitors eventually disappeared (see Figure 4B), a process that appeared slower in ZJ41. Stabilization of lymphoid production from long-term HSPCs, particularly of T cells, was delayed compared to granulocytes, as previously reported,<sup>27,33</sup> likely due to thymic damage from irradiation. As seen in the stacked area plots

for all three vectors, the relative contributions from these clones to granulocytes, T cells, and B cells were then overall relatively stable for the duration of follow-up. No individual clone(s) showed a pattern of disproportionate outgrowth expected for genotoxic expansions, in contrast to the overall slow increases in relative contributions from long-term engrafting cells as compared to the intermediate- or short-term HSPCs observed. No clone size was larger than 4.84%.

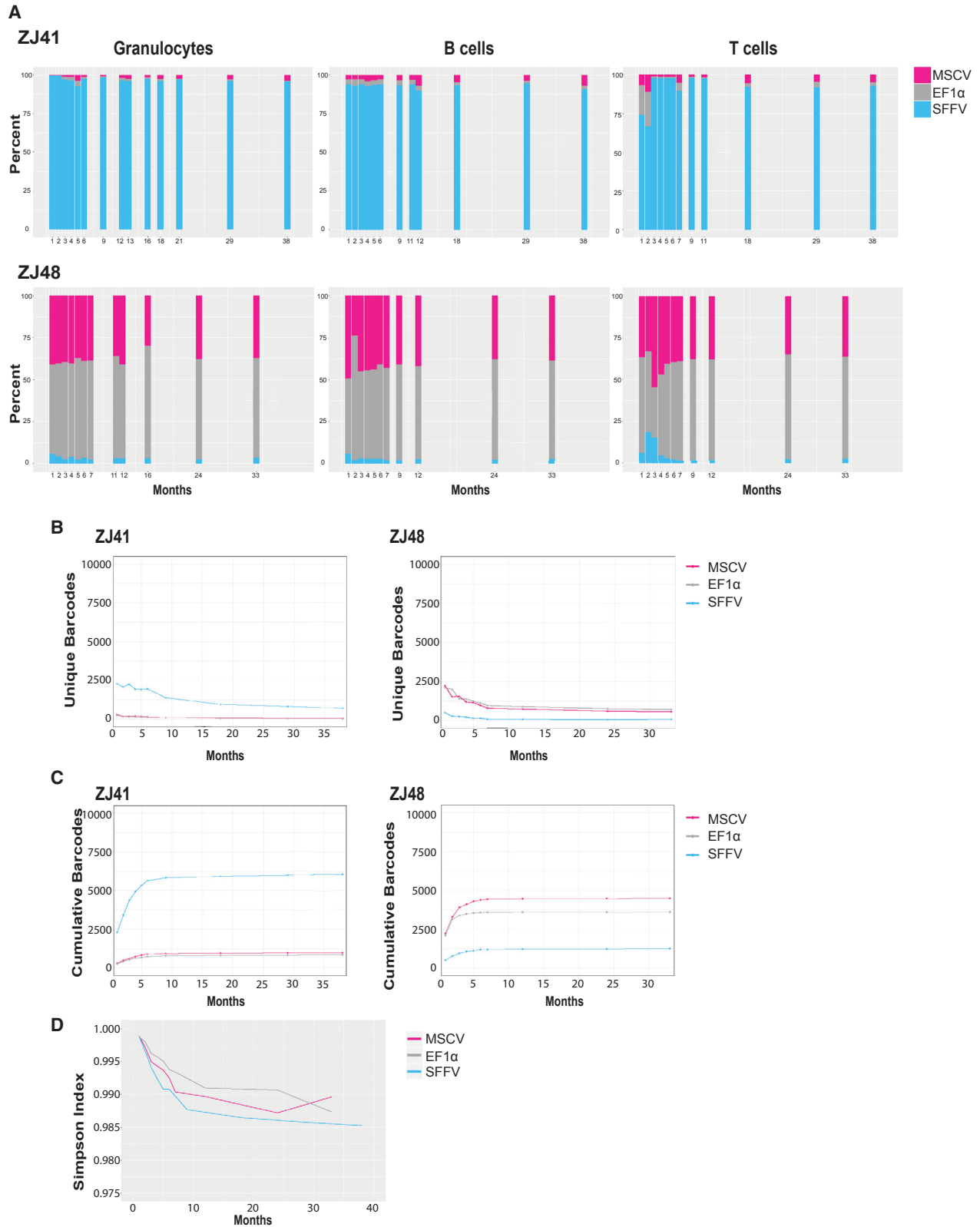
To assess comparative lineage contributions for these larger clones, we prepared heatmaps showing the contributions of each individual clone to each lineage over time for SFFV in ZJ41 and MSCV and EF1- $\alpha$  in ZJ48. These heatmaps show that the 20 largest clones in each lineage were primarily balanced multipotent clones (Figure 6), and they contributed at relatively similar levels to all lineages, in contrast to clones emerging and becoming malignant in the MLV clinical trials. In conclusion, we observed no patterns of clonal expansion of HSPC clones transduced with any of the three vectors consistent with genotoxicity.

#### DISCUSSION

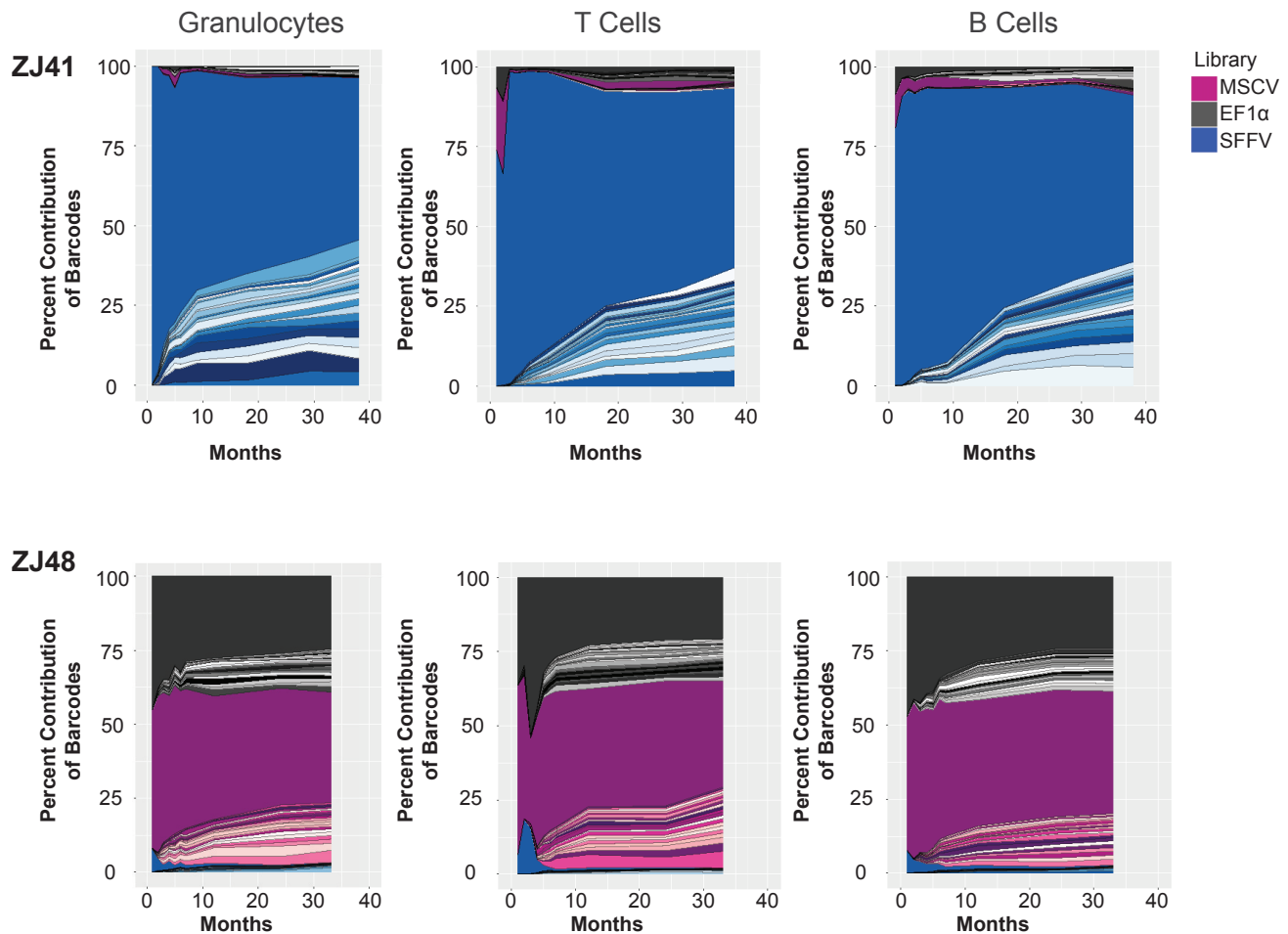
We developed a highly quantitative and sensitive technology to track the output of thousands of individual transduced HSPCs in the rhesus macaque model, and we applied this approach to assess potential genotoxicity of clinically relevant LVs with a range of predicted genotoxicity, based on murine and *in vitro* models. Inclusion of an internal EF1- $\alpha$  promoter was predicted to have the least genotoxic effect, followed by an MSCV promoter/enhancer located within the lentiviral LTR, and finally a very strong SFFV promoter/enhancer within the LTR.<sup>34,35</sup> In the animals included in this study, we did not achieve balanced initial levels of engrafted HSPCs transduced with each of the three vectors in the individual animals, due to unknown factors resulting in variability during *ex vivo* transduction or transplantation; however, a total of at least 4,450 clones was tracked for each of the three vectors between the two animals. Clonal diversity and stability were observed in the absence of any ongoing clonal expansions for 33 and 38 months follow-up to date in the two animals.

The utility of this type of preclinical *in vivo* NHP model for insertion-related clonal expansions depends on whether such expansions predict and generally precede later overt clinical transformation to myelodysplastic syndrome (MDS) or leukemia. In trials for both chronic granulomatous disease (CGD) and Wiskott-Aldrich syndrome (WAS) utilizing  $\gamma$ -retrovirus vectors with very strong viral promoter/enhancers, myeloid transformations were preceded by a steady increase in contributions from clones harboring insertions activating MECOM or other proto-oncogenes, preceding overt MDS or acute myeloid leukemia (AML) by a year or more.<sup>7,36,37</sup> Clinical transformation was accompanied by the acquisition of second hits such as gross chromosomal losses or translocations. Acute T cell leukemias occurring in  $\gamma$ -retroviral clinical trials have been less clearly preceded by long periods of detectable clonal expansions.<sup>36</sup> The initial report from the French X-linked severe combined immunodeficiency (X-SCID) trial detected an expanding clone 2 years before leukemia in one patient; however, the British X-SCID





(legend on next page)



**Figure 5. Clonal Stability of Top-Contributing Clones**

Stacked area plots delineating the percentage contribution of the top 20 contributing clones in each lineage for each vector at the time of longest follow-up (38 months for ZJ41 and 33 months for ZJ48), tracked over all time points. Each of the top 20 clones is shown as a separate shaded region in the stack. Contributions from non-top 20 clones are shown as the remaining single solid color.

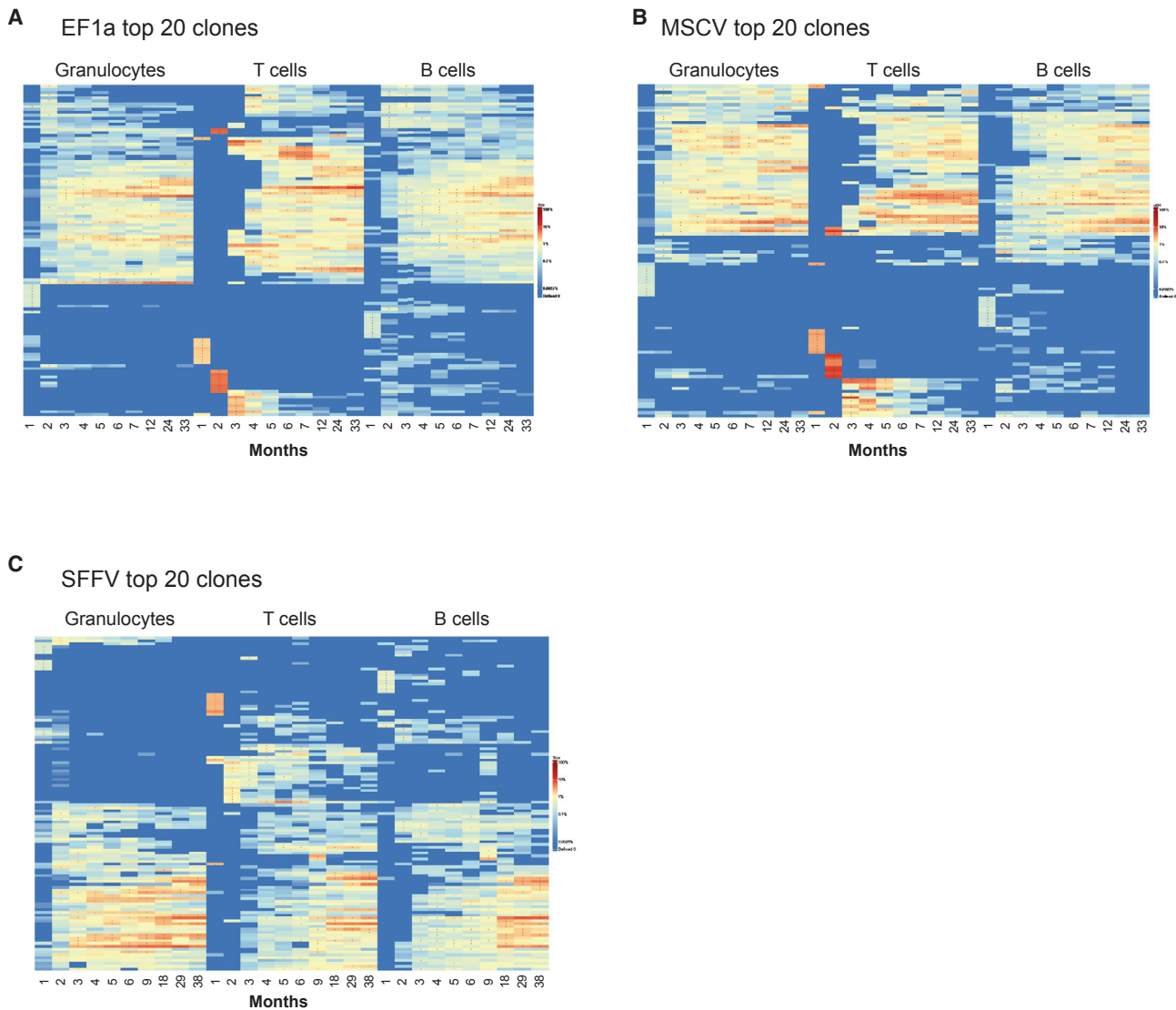
and German WAS trials reported much shorter windows of detection prior to overt T cell acute lymphoblastic leukemia (T-ALL).<sup>4,5,36</sup> These and other studies using vector insertion site (VIS) retrieval analyses for detection of dominant clones sampled patients infrequently, demonstrated variable sensitivity levels, and were hampered by the need for complex statistical methodologies to estimate clonal contribution over time; thus, clonal expansion prior to overt leukemia may have been missed.<sup>25,38</sup> Our goal was to apply highly sensitive and

quantitative barcoding based on the hypothesis that clonal expansion could be detected earlier and more reproducibly using this methodology.

Prior studies of NHPs transplanted with HSPCs transduced with integrating retroviral vectors and followed long term suggested that these models closely mimic genotoxicity observed in human clinical trials. Analyses of VIS from human trials and NHP

**Figure 4. Vector Contributions and Clonal Diversity of EF1- $\alpha$ , MSCV, and SFFV Barcoded Vectors**

(A) The percentage contribution of each vector (EF1- $\alpha$ , MSCV, and SFFV) to total transduced hematopoiesis in granulocytes, B cells, and T cells purified from the blood of animals ZJ41 and ZJ48 over time post-transplantation. The read number of each vector's library ID in a sample over the total read numbers for all library IDs in the sample  $\times 100$  is plotted. (B) Number of unique barcodes for each vector (EF1- $\alpha$ , MSCV, and SFFV) retrieved from all lineages (granulocytes, B cells, and T cells) at a single time point for each vector in ZJ41 and ZJ48. A threshold of 2,000 reads was applied to establish a master list of barcodes for each vector within each animal: to include a barcode on the master list, it must have contributed 2,000 reads at at least one time point. Once established on the master list, this barcode would be counted even at time points where it contributed less than 2,000 reads. The same barcode found contributing to more than one lineage was counted only once. (C) Cumulative number of unique barcodes retrieved for each vector (EF1- $\alpha$ , MSCV, and SFFV) from all three lineages combined (granulocytes, B cells, and T cells) over time in ZJ41 and ZJ48. (D) Simpson's diversity index for all lineages combined (granulocytes, B cells, and T cells) combined in ZJ41 and ZJ48.



**Figure 6. Heatmaps Showing Lineage Contributions of Top-Contributing Clones**

Each row in the heatmap corresponds to a barcode and each column to a sample. The barcodes are ordered by unsupervised hierarchical clustering of the Euclidean distance between barcodes' log fractional abundances in the samples. The color gradient depicts the log fractional contribution of individual barcodes to each sample. The contributions from the 20 most abundant barcodes identified in samples from every lineage and time point are plotted over all lineages and time points. The top 20 most abundant barcodes in an individual sample are designed with asterisks, and each column thus contains 20 asterisks. Each row (barcode) contains a minimum of one asterisk, because a barcode must have been a top 20 contributor to at least one sample to be included in the heatmap. (A) Top 20 EF1- $\alpha$  clones at each time point in each lineage mapped for animal ZJ48. (B) Top 20 MSCV clones at each time point in each lineage mapped for animal ZJ48. (C) Top 20 SFFV clones at each time point in each lineage mapped for animal ZJ41. Short-term lineage-restricted progenitors contributing at months 1–3 can be seen being replaced by long-term multilineage progenitors.

experiments have demonstrated close similarities in integration patterns between species for both retrovirus vectors (RVs) and LVs.<sup>39–42</sup> Both humans and rhesus macaques transplanted with MLV-transduced HSPCs have developed clonal expansions linked to insertions in or near the MECOM (MDS1/EV11) gene complex, primarily in myeloid cells.<sup>20,37</sup> NHPs transplanted with MLV-transduced CD34+ cells have developed clonal dominance preceding to overt MDS and/or leukemia due to insertional mutagenesis.<sup>21,22,43</sup>

Although overt leukemia with RV transduction of HSPCs in large animal models has been an infrequent event,<sup>44</sup> most studies involved follow-ups shorter than relevant for clinical trials, relatively low transduction efficiency (median of 5% gene marking), and did not clearly provide data demonstrating whether overt malignancy was preceded in time by a long period of detectable clonal expansion. Our group previously demonstrated effective LV gene transfer to repopulating NHP HSPCs and long-term polyclonal



hematopoiesis, without evidence of marked clonal expansion, even utilizing a vector with an extremely strong internal viral promoter/enhancer.<sup>45</sup>

Other assays used to assess genotoxicity *in vitro* or in murine models have various strengths and limitations, as summarized in the **Introduction**. However, these approaches suggested that the SFFV promoter/enhancer within a  $\gamma$ -retroviral backbone is extremely genotoxic compared to other RV constructs, correctly predicting the extremely high rate of leukemogenesis in the WAS trial utilizing this vector design as compared to other RV clinical gene therapy trials using weaker viral promoters and/or enhancers.<sup>17,36</sup> The initial study comparing RV versus LV with an internal strong promoter in a murine tumor-prone model demonstrated no leukemia induction over background for the LV, but a follow-up study did find detectable tumor acceleration when the strong SFFV viral promoter was included in the LV LTR, albeit 10-fold less than with the same SFFV design in an RV, informing our choice of this design as the potentially most genotoxic LV in the current NHP study.<sup>15,46</sup> *In vitro* immortalization assays comparing different strengths and locations of promoter/enhancers also demonstrated genotoxicity of the SFFV cassette within an LV LTR or internally, compared to almost undetectable genotoxicity with internal weaker promoters such as phosphoglycerate kinase (PGK).<sup>28</sup> A recent paper from Aranyosy and coworkers applied a competitive barcoding approach in a murine HSPC transplantation model, very similar to the approach in our rhesus macaque model, to study patterns of reconstitution and clonal dynamics with various RV and LV constructs.<sup>47</sup> Unfortunately, in this model, intermouse variation was very high, and long-term engraftment was generated from a very small number of clones in most animals, precluding genotoxicity comparisons.

Clinical trials utilizing LV transduction of HSPCs in a total of over 30 patients to treat WAS<sup>38,39,48</sup> metachromatic leukodystrophy (MLD),<sup>49</sup> adrenoleukodystrophy (ALD),<sup>8,50</sup> and sickle cell anemia (SCA)<sup>51</sup> have not demonstrated any consistent or persistent clonal expansions following retrieval of many hundreds of thousands of individual VIS; but, tracking has been limited to a relatively few time points, and it has utilized semiquantitative VIS retrieval. These trials used vectors with internal endogenous promoters or promoter/enhancers (WAS and SCA), weak internal constitutive promoters such as EF1- $\alpha$  or PGK (MLD), or a strong internal retroviral promoter (ALD). However, a single thalassemia patient receiving HSPCs transduced with an LV driving a hemoglobin gene from a strong erythroid-specific promoter/enhancer has shown prolonged but eventually transient clonal expansion of cells, linked to an LV insertion resulting in aberrant splicing and overexpression of the gene HMGA2.<sup>9</sup> LV insertions have been shown to result in aberrant splicing and chimeric transcripts.<sup>9,15,52</sup>

In sum, these clinical trials, and our similar findings in the NHP model using a more quantitative tracking methodology, are encouraging in terms of the relative lack of genotoxicity of LVs, even those containing strong promoter/enhancer elements such as an LTR-in-

serted SFFV cassette. However, the risks associated with any integrating vector, including LV, are finite, and genotoxic risks must be considered in the context of disease severity and alternative therapies. All these trials also reported very encouraging clinical results regarding disease amelioration, suggesting that LV HSPC gene therapy approaches will be applied to larger numbers of patients and disease indications in the near future, making safety predictions even more important. While our model did not detect differential genotoxic risk between the LVs utilized in this study, we believe the approach could be useful in the future to test novel vector platforms or pretransplant cell expansion approaches. Sensitivity might be enhanced or clonal expansions accelerated by the application of proliferative stress post-transplant, for instance, with busulfan, which we have previously shown can induce clonal dominance.<sup>22</sup>

## MATERIALS AND METHODS

### Construction and Production of Barcoded LVs

The generation and utilization of high-diversity 35-bp barcode libraries preceded by a 6-bp library ID and flanked by PCR primer-binding sites has been previously described.<sup>26</sup> Barcode library cassettes containing primer-binding sites for PCR amplification flanking the library ID and barcodes were subcloned into the U3 region 40 bp from the 5' end of the LTR in each of three LV plasmids as follows: EF1- $\alpha$ -library ID TCAAGT, vector plasmid pCDH-EF1-MCS-T2A-copGFP (System Biosciences, Palo Alto, CA, CD526A-1); MSCV-library ID GATCTG, vector plasmid pCDH-MCS-T2A-copGFP-MSCV (System Biosciences, CD523A-1); and SFFV-library ID GTAGCC, vector plasmid LV.SF.LTR (obtained from Eugenio Montini, Ospedale San Raffaele Srl<sup>15</sup>).

Lentiviral particles were produced as described using the  $\chi$ HIV system designed to overcome a block to HIV transduction of rhesus macaque target cells.<sup>53</sup> Retention of sufficient barcode diversity within each barcoded vector library was confirmed via Monte Carlo simulations using custom Python code to calculate the number of engrafting cells that could be transduced with 95% certainty that each unique barcode is only present in a single transduced cell 95% of the time, as described and using code available at <https://github.com/dunbarlabNIH>.<sup>27,32</sup> Vector libraries were only utilized for transplantation experiments if sufficiently diverse.

### K562 Validation of Quantitative Simultaneous Barcode Retrieval from 3 Lentiviral Libraries

K562 cells were transduced separately with each of the 3 different vectors at an MOI of 0.5 in the presence of 4  $\mu$ g/mL protamine sulfate (Sigma, St. Louis, MO) in RPMI 1640 medium (Life Technologies, Grand Island, NY), supplemented with 0.5 mg/mL penicillin-streptomycin-glutamine (Life Technologies, Grand Island, NY) and 10% fetal bovine serum (Sigma, St. Louis, MO). Cells were cultured at 37°C in the presence of 5% CO<sub>2</sub> followed by single-cell sorting on GFP-expressing cells to obtain single-cell clones. DNA was extracted from the different K562 clones and subjected to Southern blotting as previously described.<sup>27</sup> Clones containing single vector copies based on the presence of a single band on Southern blot were identified.

**Table 2. Fluorochrome-Conjugated Antibodies**

Antigen	Antibody Clone	Conjugate Fluorophore	Company and Catalog Number
CD3	SP34-2	APC-Cy7	BD Pharmingen (557757)
CD20	2H7	PE-Cy5	BD Pharmingen (555624)
CD14	TUK4	Pacific Blue	Invitrogen (MHCD1428)
CD16	3G8	APC	BioLegend (302012)
CD56	B159	PE	BD Pharmingen (555516)
CD33	AC104.3E3	PE	Miltenyi Biotec (130-091-732)

APC, allophycocyanin.

Single-copy clones were mixed at different ratios of cell numbers and then diluted into untransduced K562 cells to create mixtures at a range of vector ratios and overall marking levels, corresponding to the overall levels of marking and ratios between barcode libraries (vectors) obtained *in vivo* in ZJ48 and ZJ41.

#### Transduction and Autologous Transplantation of Rhesus Macaque CD34+ HSPCs

The NHLBI Animal Care and Use Committee approved all animal studies. Rhesus macaque HSPCs were mobilized from the bone marrow by treatment with 4 days of granulocyte colony-stimulating factor (G-CSF) 10 mg/kg/day (Amgen, Thousand Oaks, CA) and AMD3100 (plerixafor) 1 mg/kg (Sigma, St. Louis, MO) on the morning of the fifth day, 3–4 hr prior to apheresis. Collection of peripheral blood mononuclear cells (PBMCs) via apheresis and enrichment of CD34+ HSPCs via immunoselection were performed as previously described.<sup>54</sup> CD34+ cells were split into three equal fractions, and, following overnight culture on RetroNectin-coated plates (Takara, T100B, Mountain View, CA) in X-VIVO 10 media (Lonza, Rockland, ME) supplemented with 100 ng/mL recombinant human flt3 ligand, stem cell factor, and thrombopoietin (Miltenyi Biotec, Auburn, CA), each fraction was transduced with one of the three vectors at an MOI of 25 in the presence of 4 µg/mL protamine sulfate (Sigma, St. Louis, MO). 24 hr later, all three fractions of transduced cells were collected from the plates, pooled, and reinfused into the autologous rhesus macaque. Each animal received total-body irradiation at a dose of 500 cGy/day for 2 days prior to the day of cell reinfusion. Aliquots of cells were removed prior to pooling and reinfusion, and they were cultured for an additional 48 hr prior to flow cytometric analysis of GFP expression. Cell aliquots from the end of transduction were also removed for DNA extraction and barcode recovery.

#### Cell Processing and Flow Cytometric Sorting

Peripheral blood was separated into mononuclear cells and granulocytes via centrifugation over Ficoll (MP Biomedicals, Solon, OH); incubated with ammonium-chloride-potassium (ACK) lysis buffer (Quality Biological, Gaithersburg, MD) to remove red blood cells; and labeled with fluorophore-conjugated monoclonal antibodies against CD3, CD20, CD14, CD56, CD16, and CD33 (antibody clone and source information summarized in Table 2), allowing sorting of individual granulocyte, monocyte, T cell, and B cell lineages on a

FACSAriaII flow cytometer (BD Biosciences, San Jose, CA). DNA was extracted from sorted cell pellets using the DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA).

#### Barcode Retrieval

500 ng DNA samples from each cell type were amplified via 28 cycles of PCR using primers flanking the library ID and barcode. Primers added a 6-bp sample ID to facilitate multiplex sequencing. Samples were run on a 2.5% agarose gel, and the resulting 157-bp amplified fragment was gel purified (MinElute Gel Extraction Kit, QIAGEN, Valencia, CA). Equal amounts of product were pooled and prepared for high-throughput sequencing, allowing pooling of 12–20 multiplexed samples per lane in the sequencing flow cell of the Illumina HiSeq 2000 sequencer.<sup>27</sup>

#### Barcode Retrieval and Data Analysis

After de-multiplexing, 4 million reads (selected randomly to avoid lane position bias) sequencing output from each sample were analyzed using custom python code as described.<sup>32</sup> In brief, reads were extracted via the presence of one of the three library IDs, reads with abundance lower than 100 reads were discarded (a step shown previously not to change results, excluding reads almost certainly resulting from PCR and sequencing artifacts and speeding processing time from days to hours for each file), reads with 0–2 mismatches or insertions or deletions (indels) were counted as identical and were combined to determine the read number for each barcode with a library ID within a sample. The fractional contribution of each barcode to a sample was then calculated from the ratio of the individual barcode's read number divided by the total number of valid barcode reads in the sample. A read number threshold of 2,000 was applied to exclude false barcodes resulting from sequencing artifacts or the presence of false library IDs in the genome.

A master list of barcodes was then generated for each animal and each vector, consisting of any barcode with a read abundance of over 2,000 in any sample for that animal and vector. Once a barcode was included on the master list, all analyses included the read numbers for that barcode, even if the read number in other samples fell below 2,000. The threshold of 2,000 was chosen conservatively and previously validated for this rhesus macaque model.<sup>32</sup> Figure S4 shows the impact of applying lower and higher thresholds, which impact on enumeration of overall clone numbers but do not impact on tracking of high contributing or potentially expanding clones. Distinguishing very low contributing clones from false barcodes or recurrent sequencing artifacts is impossible, but real clones contributing at a very low level in every sample are unlikely to be biologically important.

Using our methodology on primate samples with this level of marking, we have previously shown that lower thresholds result in the counting of false barcodes and sampling issues, and higher thresholds result in the exclusion of valid and reproducibly retrieved clones.<sup>32</sup>

Stacked area plots were made using R studio (Integrated Development for RStudio, 2015, Boston, MA; <https://www.rstudio.com/>)

and custom R code. Heatmaps were produced using custom R code. All custom Python and R code can be accessed at <https://github.com/dunbarlabNIH>. Prism 6.0 (GraphPad, La Jolla, CA) was used to make dot plot diagrams.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and can be found with this article online at <https://doi.org/10.1016/j.omtm.2018.10.009>.

#### AUTHOR CONTRIBUTIONS

I.M.Y., C.W., and C.E.D. contributed to the design and conceptualization of the research. L.L.T., D.A.E., and S.K. provided software and conducted formal analyses. L.L.T. performed visualizations. I.M.Y., C.W., S.P., M.A.F.C., T.W., K.-R.Y., S.G.H., A.B., A.K., M.M., and R.E.D. conducted investigations. C.E.D. provided supervision. I.M.Y. and C.E.D. wrote the manuscript.

#### CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

#### ACKNOWLEDGMENTS

This research was supported by funding from the Division of Intramural Research at the National Heart, Lung and Blood Institute (NHLBI) at the NIH. We thank the staff of the NHLBI Primate Program for their assistance in animal care; the NHLBI DNA Sequencing and Genomics Core for DNA sequencing; and Philip J. McCoy, Venina Dominical, and Keyvan Keyvanfar from NHLBI flow cytometry cores. For technical assistance, the authors thank Stephanie Sellers.

#### REFERENCES

- Aiuti, A., Slavin, S., Aker, M., Ficara, F., Deola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabucchi, A., Carlucci, F., et al. (2002). Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* 296, 2410–2413.
- Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., et al. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 288, 669–672.
- Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* 118, 3132–3142.
- Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302, 415–419.
- Howe, S.J., Mansour, M.R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., Brugman, M.H., Pike-Overzet, K., Chatters, S.J., de Ridder, D., 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.
- Nienhuis, A.W., Dunbar, C.E., and Sorrentino, B.P. (2006). Genotoxicity of retroviral integration in hematopoietic cells. *Mol. Ther.* 13, 1031–1049.
- Stein, S., Ott, M.G., Schultze-Strasser, S., Jauch, A., Burwinkel, B., Kinner, A., Schmidt, M., Krämer, A., Schwäble, J., Glimm, H., 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.
- Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., Veres, G., Schmidt, M., Kutschera, I., Vidaud, M., Abel, U., Dal-Cortivo, L., Caccavelli, L., et al. (2009).

- Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* 326, 818–823.
- Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., Denaro, M., Brady, T., Westerman, K., et al. (2010). Transfusion independence and HMGA2 activation after gene therapy of human  $\beta$ -thalassaemia. *Nature* 467, 318–322.
- De Ravin, S.S., Wu, X., Moir, S., Anaya-O'Brien, S., Kwatema, N., Littel, P., Theobald, N., Choi, U., Su, L., Marquesen, M., et al. (2016). Lentiviral hematopoietic stem cell gene therapy for X-linked severe combined immunodeficiency. *Sci. Transl. Med.* 8, 335ra57.
- Kustikova, O., Fehse, B., Modlich, U., Yang, M., Düllmann, J., Kamino, K., von Neuhoff, N., Schlegelberger, B., Li, Z., and Baum, C. (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* 308, 1171–1174.
- Kustikova, O.S., Schiedlmeier, B., Brugman, M.H., Stahlhut, M., Bartels, S., Li, Z., and Baum, C. (2009). Cell-intrinsic and vector-related properties cooperate to determine the incidence and consequences of insertional mutagenesis. *Mol. Ther.* 17, 1537–1547.
- Negre, O., Bartholomae, C., Beuzard, Y., Cavazzana, M., Christiansen, L., Courne, C., Deichmann, A., Denaro, M., de Dreuzy, E., Finer, M., et al. (2015). Preclinical evaluation of efficacy and safety of an improved lentiviral vector for the treatment of  $\beta$ -thalassaemia and sickle cell disease. *Curr. Gene Ther.* 15, 64–81.
- Cesana, D., Ranzani, M., Volpin, M., Bartholomae, C., Duros, C., Artus, A., Merella, S., Benedicenti, F., Sergi, L., Sanvito, F., et al. (2014). Uncovering and dissecting the genotoxicity of self-inactivating lentiviral vectors in vivo. *Mol. Ther.* 22, 774–785.
- Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M., et al. (2009). The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J. Clin. Invest.* 119, 964–975.
- Du, Y., Jenkins, N.A., and Copeland, N.G. (2005). Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow progenitor cells. *Blood* 106, 3932–3939.
- Modlich, U., Bohne, J., Schmidt, M., von Kalle, C., Knöss, S., Schambach, A., and Baum, C. (2006). Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood* 108, 2545–2553.
- Modlich, U., Navarro, S., Zychlinski, D., Maetzig, T., Knoess, S., Brugman, M.H., Schambach, A., Charrier, S., Galy, A., Thrasher, A.J., et al. (2009). Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. *Mol. Ther.* 17, 1919–1928.
- Larochelle, A., and Dunbar, C.E. (2013). Hematopoietic stem cell gene therapy: assessing the relevance of preclinical models. *Semin. Hematol.* 50, 101–130.
- Calmels, B., Ferguson, C., Laukkanen, M.O., Adler, R., Faulhaber, M., Kim, H.J., Sellers, S., Hematti, P., Schmidt, M., von Kalle, C., et al. (2005). Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. *Blood* 106, 2530–2533.
- Seggewiss, R., Pittaluga, S., Adler, R.L., Guenaga, F.J., Ferguson, C., Pilz, I.H., Ryu, B., Sorrentino, B.P., Young, W.S., 3rd, Donahue, R.E., et al. (2006). Acute myeloid leukemia is associated with retroviral gene transfer to hematopoietic progenitor cells in a rhesus macaque. *Blood* 107, 3865–3867.
- Xie, J., Larochelle, A., Maric, I., Faulhaber, M., Donahue, R.E., and Dunbar, C.E. (2010). Repetitive busulfan administration after hematopoietic stem cell gene therapy associated with a dominant HDAC7 clone in a nonhuman primate. *Hum. Gene Ther.* 21, 695–703.
- Schmidt, M., Zickler, P., Hoffmann, G., Haas, S., Wissler, M., Muessig, A., Tisdale, J.F., Kuramoto, K., Andrews, R.G., Wu, T., et al. (2002). Polyclonal long-term repopulating stem cell clones in a primate model. *Blood* 100, 2737–2743.
- Biasco, L., Baricordi, C., and Aiuti, A. (2012). Retroviral integrations in gene therapy trials. *Mol. Ther.* 20, 709–716.
- Brady, T., Roth, S.L., Malani, N., Wang, G.P., Berry, C.C., Leboulch, P., Hacein-Bey-Abina, S., Cavazzana-Calvo, M., Papapetrou, E.P., Sadelain, M., et al. (2011). A method to sequence and quantify DNA integration for monitoring outcome in gene therapy. *Nucleic Acids Res.* 39, e72.

26. Lu, R., Neff, N.F., Quake, S.R., and Weissman, I.L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* 29, 928–933.
27. Wu, C., Li, B., Lu, R., Koelle, S.J., Yang, Y., Jares, A., Krouse, A.E., Metzger, M., Liang, F., Loré, K., et al. (2014). Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell* 14, 486–499.
28. Zychlinski, D., Schambach, A., Modlich, U., Maetzig, T., Meyer, J., Grassman, E., Mishra, A., and Baum, C. (2008). Physiological promoters reduce the genotoxic risk of integrating gene vectors. *Mol. Ther.* 16, 718–725.
29. Larochele, A., Savona, M., Wiggins, M., Anderson, S., Ichwan, B., Keyvanfar, K., Morrison, S.J., and Dunbar, C.E. (2011). Human and rhesus macaque hematopoietic stem cells cannot be purified based only on SLAM family markers. *Blood* 117, 1550–1554.
30. Radtke, S., Adair, J.E., Giese, M.A., Chan, Y.Y., Norgaard, Z.K., Enstrom, M., Haworth, K.G., Scheffer, L.E., and Kiem, H.P. (2017). A distinct hematopoietic stem cell population for rapid multilineage engraftment in nonhuman primates. *Sci. Transl. Med.* 9, eaan1145.
31. Fehse, B., Kustikova, O.S., Bubenheim, M., and Baum, C. (2004). Po(s)on—it's a question of dose. *Gene Ther.* 11, 879–881.
32. Koelle, S.J., Espinoza, D.A., Wu, C., Xu, J., Lu, R., Li, B., Donahue, R.E., and Dunbar, C.E. (2017). Quantitative stability of hematopoietic stem and progenitor cell clonal output in rhesus macaques receiving transplants. *Blood* 129, 1448–1457.
33. Loré, K., Seggewiss, R., Guenaga, F.J., Pittaluga, S., Donahue, R.E., Krouse, A., Metzger, M.E., Koup, R.A., Reilly, C., Douek, D.C., and Dunbar, C.E. (2006). In vitro culture during retroviral transduction improves thymic repopulation and output after total body irradiation and autologous peripheral blood progenitor cell transplantation in rhesus macaques. *Stem Cells* 24, 1539–1548.
34. Ryu, B.Y., Evans-Galea, M.V., Gray, J.T., Bodine, D.M., Persons, D.A., and Nienhuis, A.W. (2008). An experimental system for the evaluation of retroviral vector design to diminish the risk for proto-oncogene activation. *Blood* 111, 1866–1875.
35. Zhou, S., Mody, D., DeRavin, S.S., Hauer, J., Lu, T., Ma, Z., Hacein-Bey Abina, S., Gray, J.T., Greene, M.R., Cavazzana-Calvo, M., et al. (2010). A self-inactivating lentiviral vector for SCID-X1 gene therapy that does not activate LMO2 expression in human T cells. *Blood* 116, 900–908.
36. Braun, C.J., Boztug, K., Paruzynski, A., Witzel, M., Schwarzer, A., Rothe, M., Modlich, U., Beier, R., Göhring, G., Steinemann, D., et al. (2014). Gene therapy for Wiskott-Aldrich syndrome—long-term efficacy and genotoxicity. *Sci. Transl. Med.* 6, 227ra33.
37. Ott, M.G., Schmidt, M., Schwarzwaldler, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kühlcke, K., Schilz, A., Kunkel, H., et al. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EV11, PRDM16 or SETBP1. *Nat. Med.* 12, 401–409.
38. Biasco, L., Pellin, D., Scala, S., Dionisio, F., Basso-Ricci, L., Leonardelli, L., Scaramuzza, S., Baricordi, C., Ferrua, F., Cicalese, M.P., et al. (2016). In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell* 19, 107–119.
39. Aiuti, A., Biasco, L., Scaramuzza, S., Ferrua, F., Cicalese, M.P., Baricordi, C., Dionisio, F., Calabria, A., Giannelli, S., Castiello, M.C., et al. (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* 341, 1233151.
40. Beard, B.C., Dickerson, D., Beebe, K., Gooch, C., Fletcher, J., Okbinoglu, T., Miller, D.G., Jacobs, M.A., Kaul, R., Kiem, H.P., and Trobridge, G.D. (2007). Comparison of HIV-derived lentiviral and MLV-based gammaretroviral vector integration sites in primate repopulating cells. *Mol. Ther.* 15, 1356–1365.
41. Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., 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.
42. Hematti, P., Hong, B.K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I.E., Eckfeldt, C.E., Sharma, Y., Schmidt, M., et al. (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* 2, e423.
43. Zhang, X.B., Beard, B.C., Trobridge, G.D., Wood, B.L., Sale, G.E., Sud, R., Humphries, R.K., and Kiem, H.P. (2008). High incidence of leukemia in large animals after stem cell gene therapy with a HOXB4-expressing retroviral vector. *J. Clin. Invest.* 118, 1502–1510.
44. Kiem, H.P., Sellers, S., Thomasson, B., Morris, J.C., Tisdale, J.F., Horn, P.A., Hematti, P., Adler, R., Kuramoto, K., Calmels, B., et al. (2004). Long-term clinical and molecular follow-up of large animals receiving retrovirally transduced stem and progenitor cells: no progression to clonal hematopoiesis or leukemia. *Mol. Ther.* 9, 389–395.
45. Kim, Y.J., Kim, Y.S., Larochele, A., Renaud, G., Wolfsberg, T.G., Adler, R., Donahue, R.E., Hematti, P., Hong, B.K., Roayaei, J., et al. (2009). Sustained high-level polyclonal hematopoietic marking and transgene expression 4 years after autologous transplantation of rhesus macaques with SIV lentiviral vector-transduced CD34+ cells. *Blood* 113, 5434–5443.
46. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., Sergi Sergi, L., Benedicenti, F., Ambrosi, A., Di Serio, C., et al. (2006). Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat. Biotechnol.* 24, 687–696.
47. Aranyosy, T., Thielecke, L., Glauche, I., Fehse, B., and Cornils, K. (2017). Genetic Barcodes Facilitate Competitive Clonal Analyses In Vivo. *Hum. Gene Ther.* 28, 926–937.
48. Morris, E.C., Fox, T., Chakraverty, R., Tendeiro, R., Snell, K., Rivat, C., Grace, S., Gilmour, K., Workman, S., Buckland, K., et al. (2017). Gene therapy for Wiskott-Aldrich syndrome in a severely affected adult. *Blood* 130, 1327–1335.
49. Sessa, M., Lorioli, L., Fumagalli, F., Acquati, S., Redaelli, D., Baldoli, C., Canale, S., Lopez, I.D., Morena, F., Calabria, A., et al. (2016). Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an ad-hoc analysis of a non-randomised, open-label, phase 1/2 trial. *Lancet* 388, 476–487.
50. Eichler, F., Duncan, C., Musolino, P.L., Orchard, P.J., De Oliveira, S., Thrasher, A.J., Armant, M., Dansereau, C., Lund, T.C., Miller, W.P., et al. (2017). Hematopoietic Stem-Cell Gene Therapy for Cerebral Adrenoleukodystrophy. *N. Engl. J. Med.* 377, 1630–1638.
51. Ribeil, J.A., Hacein-Bey-Abina, S., Payen, E., Magnani, A., Semeraro, M., Magrin, E., Caccavelli, L., Neven, B., Bourget, P., El Nemer, W., et al. (2017). Gene Therapy in a Patient with Sickle Cell Disease. *N. Engl. J. Med.* 376, 848–855.
52. Scholz, S.J., Fronza, R., Bartholomä, C.C., Cesana, D., Montini, E., von Kalle, C., Gil-Farina, I., and Schmidt, M. (2017). Lentiviral Vector Promoter is Decisive for Aberrant Transcript Formation. *Hum. Gene Ther.* 28, 875–885.
53. Uchida, N., Washington, K.N., Hayakawa, J., Hsieh, M.M., Bonifacino, A.C., Krouse, A.E., Metzger, M.E., Donahue, R.E., and Tisdale, J.F. (2009). Development of a human immunodeficiency virus type 1-based lentiviral vector that allows efficient transduction of both human and rhesus blood cells. *J. Virol.* 83, 9854–9862.
54. Donahue, R.E., Kuramoto, K., and Dunbar, C.E. (2005). Large animal models for stem and progenitor cell analysis. *Curr. Protoc. Immunol. Chapter 22*, Unit 22A.1.