Molecular Therapy
Original Article

Aberrant Clonal Hematopoiesis following Lentiviral Vector Transduction of HSPCs in a Rhesus Macaque

Diego A. Espinoza,[1,8,9] Xing Fan,[1,9] Di Yang,[1,2,9] Stefan F. Cordes,[1] Lauren L. Truitt,[1] Katherine R. Calvo,[1] Idalia M. Yabe,[1] Selami Demirci,[3] Kristin J. Hope,[4] So Gun Hong,[1] Allen Krouse,[1] Mark Metzger,[1] Aylin Bonifacino,[1] Rong Lu,[5] Naoya Uchida,[3] John F. Tisdale,[3] Xiaolin Wu,[6] Suk See DeRavin,[7] Harry L. Malech,[7] Robert E. Donahue,[1] Chuanfeng Wu,[1] and Cynthia E. Dunbar[1]

[1]Translational Stem Cell Biology Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA; [2]Institute of Hematology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; [3]Sickle Cell and Vascular Biology Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA; [4]Stem Cell and Cancer Research Institute, McMaster University, Hamilton, ON, Canada; [5]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA, USA; [6]Cancer Research Technology Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, USA; [7]Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD, USA; [8]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Lentiviral vectors (LVs) are used for delivery of genes into hematopoietic stem and progenitor cells (HSPCs) in clinical trials worldwide. LVs, in contrast to retroviral vectors, are not associated with insertion site-associated malignant clonal expansions and, thus, are considered safer. Here, however, we present a case of markedly abnormal dysplastic clonal hematopoiesis affecting the erythroid, myeloid, and megakaryocytic lineages in a rhesus macaque transplanted with HSPCs that were transduced with a LV containing a strong retroviral murine stem cell virus (MSCV) constitutive promoter-enhancer in the LTR. Nine insertions were mapped in the abnormal clone, resulting in overexpression and aberrant splicing of several genes of interest, including the cytokine stem cell factor and the transcription factor *PLAG1*. This case represents the first clear link between lentiviral insertion-induced clonal expansion and a clinically abnormal transformed phenotype following transduction of normal primate or human HSPCs, which is concerning, and suggests that strong constitutive promoters should not be included in LVs.

INTRODUCTION

Lentiviral vectors (LVs) are successfully used for delivery of genetic material into hematopoietic stem and progenitor cells (HSPCs) in clinical trials for inherited non-malignant diseases.[1–8] In contrast to γ-retroviral vectors, which have been shown to induce insertion-associated leukemias in both clinical trials[9] and non-human primate models,[10] LVs have not been associated with malignant clonal expansions in clinical trials or large animal models.[11] LVs integrate preferentially within active genes rather than near transcriptional start sites[12,13] and have thus been considered less likely to activate nearby proto-oncogenes. However, reports from recent clinical trials suggest that LVs are capable of inducing insertion-site associated clonal (albeit non-malignant) expansions.[8,14] Long-term monitoring of transduced HSPC clones in humans and non-human primates remains critical for assessing the safety of LV gene therapies.

Our laboratory has used a rhesus macaque (RM) genetic barcoding model to study the output of lentivirally transduced transplanted HSPCs at a clonal level over time;[15–18] such a setting closely resembles that of lentiviral HSPC gene therapy trials in which autologous HSPCs are collected, transduced, and reinfused into the patient. In our published and unpublished studies, we tracked over 101,000 individual LV insertions in 10 transplanted macaques followed for up to 6 years and observed no clonal expansions associated with a transformed or aberrant hematologic phenotype. We now report the development of aberrant, markedly dysplastic fatal clonal hematopoiesis in one animal, affecting the erythroid, myeloid, and megakaryocytic lineages and originating from a single transduced HSPC clone containing 9 integrated LVs.

RESULTS

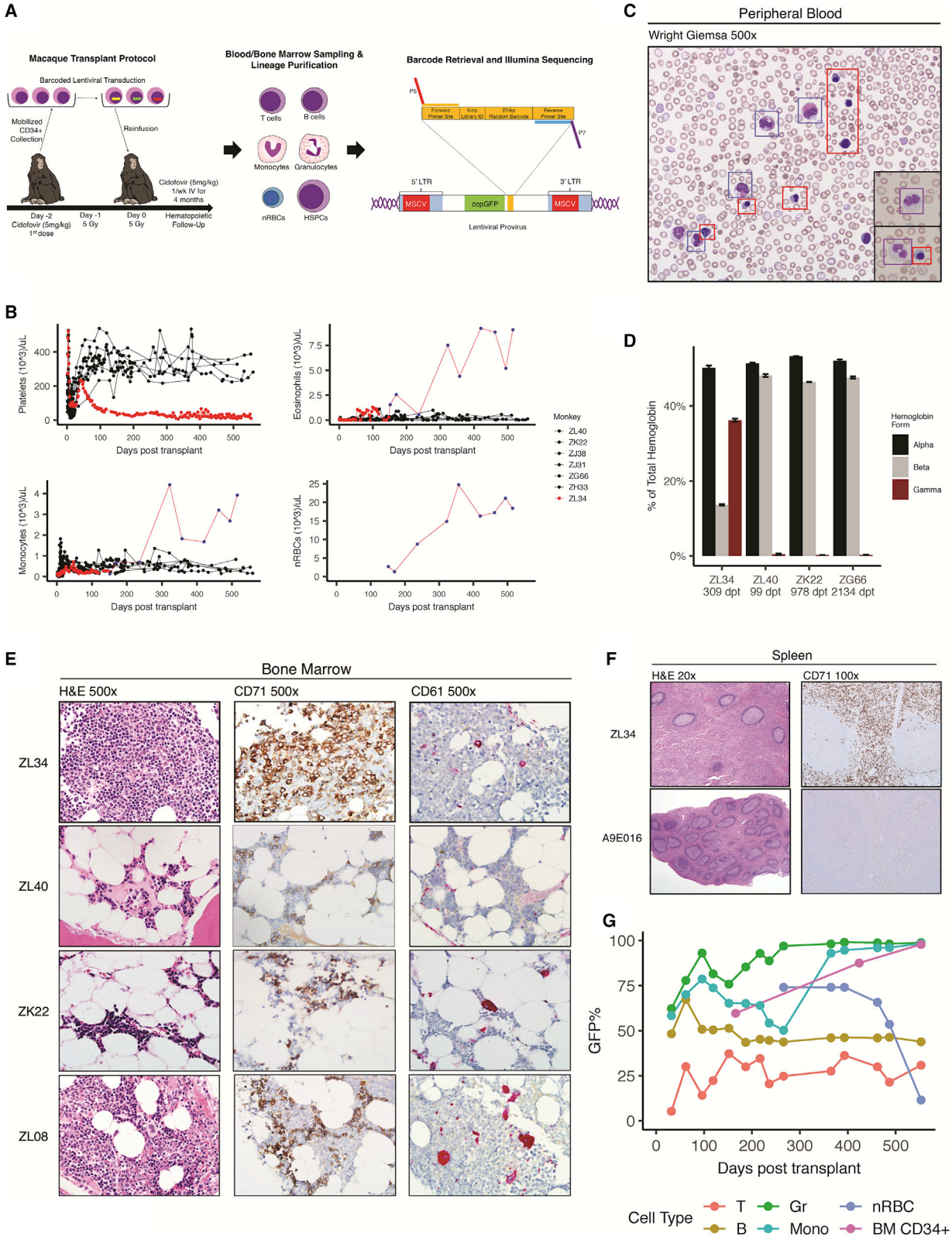
Lentiviral Transduction and Transplantation

We developed a lentiviral barcoding model to quantitatively study the output of thousands of individual HSPCs following autologous transplantation.[15–18] As part of our ongoing studies, we transplanted

Received 6 March 2019; accepted 4 April 2019;
https://doi.org/10.1016/j.ymthe.2019.04.003.

[9]These authors contributed equally to this work.

Correspondence: Chuanfeng Wu, Translational Stem Cell Biology Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA.
E-mail: wuc3@mail.nih.gov
Correspondence: Cynthia Dunbar, Translational Stem Cell Biology Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA.
E-mail: dunbarc@nhlbi.nih.gov



(legend on next page)

animal ZL34 with autologous CD34$^+$ HSPCs transduced with a lentiviral barcoding vector driving GFP expression via the strong murine retroviral murine stem cell virus (MSCV) promoter-enhancer inserted in the lentiviral long terminal repeat (LTR) (Figure 1A). The transduction or transplantation protocols utilized were unchanged from our prior experience,[15–18] and the MOI, CD34$^+$ transplanted dose (cells per kilogram), and GFP percentage (GFP%) of the infused cells were within the ranges utilized for prior animals (Table S1; n = 10). ZL34, along with another monkey, ZL40 (Table S1), received cytomegalovirus (CMV)-suppressing cidofovir for 4 months to study the effect of CMV reactivation on immune reconstitution. With the exception of ZL34, all other animals, including ZL40, also receiving cidofovir on the same schedule continued to show stable, highly polyclonal output from transduced HSPCs with no evidence of genotoxicity or malignant clonal expansions at the latest follow-up (4–70 months; median, 25 months), regardless of having received HSPCs transduced with vectors containing either an internal EF1α promoter (n = 3) or the LTR-embedded MSCV promoter (n = 7), as published previously.[15–17]

Development of Abnormal Hematopoiesis

ZL34 engrafted normally post-transplantation compared with other animals, promptly recovering neutrophils to more than 500/μL, platelets to more than 200,000/μL, and hemoglobin to more than 9 g/dL (Figures 1B, S1A, and S1B). However, ZL34's platelet count declined beginning 49 days post-transplantation to levels far below the range maintained in other transplanted macaques (Figure 1B), requiring platelet transfusions to treat bleeding. Waxing and waning eosinophilia without identified infectious, drug, or allergic etiology emerged 50 days post-transplantation and became persistent by day 300 (Figure 1B). Eosinophilia did not develop in other macaques on the same transplant regimen. The morphology of neutrophils became markedly dysplastic, with hypogranularity and bilobulated and/or hypolobulated nuclei, despite normal neutrophil counts and lack of circulating blasts (Figures 1C and S1A). Monocytes and basophils also rose to well above normal range by day 300 (Figures 1B and S1C). In contrast, lymphocyte counts remained within expected post-transplant ranges (Figure S1D).

Nucleated red blood cells (nRBCs) are normally found only in the bone marrow (BM) of non-human primates and humans, with nuclear extrusion occurring prior to release into the blood. Strikingly, large numbers of nRBCs were prematurely released into the blood in ZL34, first noted on day 151 and reaching concentrations of more than 25,000/μL (Figures 1B and 1C). No other transplanted macaques had detectable circulating nRBCs. ZL34's hemoglobin levels and RBC counts remained within the normal range until 1 year post-transplantation (Figures S1B and S1E). Analysis of globin chain ratios showed marked upregulation of γ-globin chains (and, thus, fetal hemoglobin) in ZL34's RBCs long-term, in contrast to the normal dominance of adult β chains by day 100 or later in other transplanted macaques (Figure 1D).[19]

Compared with healthy transplanted (ZL40 and ZK22) and non-transplanted (ZL08) RMs, ZL34's BM on day 424 was hypercellular, with increased eosinophils and marked erythroid dominance and markedly dysplastic uninuclear micro-megakaryocytes (Figure 1E). Blasts were not increased. The day 543 BM karyotype was normal in 20 of 20 metaphases. Splenic red pulp was markedly expanded because of extramedullary hematopoiesis, with infiltration of immature CD71$^+$ erythroid cells (Figure 1F), an abnormal finding in macaques or humans. Pathologically and clinically, this hematologic disorder would be classified by World Health Organization (WHO) criteria as an overlap myelodysplastic/myeloproliferative neoplasm with extramedullary hematopoiesis. Bleeding necessitated euthanasia of ZL34 on day 553.

Inclusion of GFP from the copepod *Pontellina plumata* (copGFP) in the vector allowed analysis of the level of engraftment with transduced cells. The percent of GFP$^+$ T and B cells remained generally stable over time and within the upper range of values observed in other barcoded RMs (Figures 1G and S1F).[15–18] In contrast, GFP%$^+$ granulocytes (including neutrophils and eosinophils) and monocytes approached 100%, coincident with development of markedly abnormal blood counts, suggesting expansion of transduced cells within these lineages. GFP expression in the circulating nRBCs was also very high but dropped precipitously before euthanasia, potentially because of transgene silencing. Other barcoded animals did not show expansion of GFP$^+$ cells of any lineage over time, with similar levels in all lineages other than delayed reconstitution of

Figure 1. Rhesus Macaque Autologous Transplantation Model and Development of Aberrant Hematopoiesis in Rhesus Macaque ZL34

(A) Transplantation timeline for ZL34. HSPCs were mobilized with G-CSF and AMD3100, collected via apheresis, enriched for CD34$^+$ HSPCs, and transduced with a barcoded lentiviral vector. Following total body irradiation, transduced autologous CD34$^+$ HSPCs were reinfused, and cidofovir was administered. The integrated form of the provirus is shown, detailing the position of the MSCV promoter and/or enhancer(s), copGFP marker gene, and high-diversity barcode library consisting of a 6-bp library ID and 27-bp random nucleotides, flanked by sequencing primer binding sites. (B) Platelet, eosinophil, monocyte, and nRBC concentrations in the blood over time in seven macaques receiving cells transduced with the barcoded LV containing the MSCV promoter and/or enhancer, including ZL34. Counts were determined by Coulter counter (red or black circles) or by automated PB smear scanning with Cellavision DM (blue dots, ZL34). (C) Wright's-stained ZL34 PB smear from day 424 post-transplantation, showing frequent eosinophils (blue squares), nucleated erythroid cells (red squares), which are not found in normal rhesus macaque blood, and hypogranular neutrophils with dysplastic nuclei (black insets, purple squares). (D) HPLC assay of hemoglobin chain composition in the blood partitioned into α-, β-, and γ-globin molecules for ZL34 309 days post-transplantation (dpt) and three barcoded macaques without clonal expansions (ZL40, ZK22, and ZG66 99–2134 days post-transplantation). ZL34 values correspond to a predicted fetal hemoglobin of 70%. Error bars denote SD for three technical replicates. (E) BM biopsies from ZL34 (day 424), two healthy transplanted macaques (ZK22 and ZL40), and one healthy non-transplanted macaque (ZL08). CD71, immunohistochemistry for the erythroid precursor-specific transferrin receptor (CD71); CD61, immunohistochemistry for the megakaryocyte-specific glycoprotein IIIa (CD61). Note that ZL34's megakaryocytes are small and uninuclear. (F) Spleen from ZL34 and a control-transplanted macaque (A9E016). (G) GFP%$^+$ cells by flow cytometry over time in ZL34 PB T cells, B cells, granulocytes, monocytes, and nRBCs and BM CD34$^+$ cells.

GFP^+ T cells because of total body irradiation (TBI) effects on the thymus (Figure S1F).[15–17]

The percent of GFP^+ $CD34^+$ BM HSPCs also increased over time, reaching over 90%. Flow cytometric analysis of $CD34^+$ subsets based on prior validation in macaques[20] showed that the most primitive $CD34^+CD90^+CD45RA^-$ population, highly enriched for multipotent long-lived HSCs, was most markedly GFP^+, with lower percentages in the $CD34^+CD90^-CD45RA^+$ population containing lymphoid precursors (Figures S2A–S2C).[20] Of note, the $CD34^+CD90^+CD45RA^-$ population was also markedly expanded in ZL34 BM compared with normal BM and present in the spleen, suggesting a differentiation block as well as extramedullary hematopoiesis (Figures S2A and S2B). These findings are consistent with development of a myeloproliferative neoplasm (MPN)/myelodysplastic syndromes (MDS) neoplastic syndrome resulting from abnormalities of transduced primitive HSPCs.

A Single Expanded Transduced Clone Was Responsible for Abnormal Hematopoiesis

We performed quantitative barcode retrieval from T cells, B cells, neutrophils, eosinophils, monocytes, nRBCs, and $CD34^+$ cells using primers flanking the barcode region (Figures 2A–2G).[15–18] Six barcodes contributed equally and accounted for 100% of barcodes from nRBCs at all time points. The same 6 clones appeared and expanded in neutrophils and monocytes beginning on day 187, replacing earlier highly polyclonal contributions. Eosinophils on day 494 contained only the same 6 dominant barcodes. In contrast, T and B cells maintained polyclonal diversity over time, with intermittent detection of much lower levels of the 6 barcodes, likely because of contamination with clonally expanded myeloid or erythroid cells during sorting and/or aberrant marker expression on dysplastic myeloid cells.

A seventh barcode was intermittently detected within the abnormal lineages because of a single mutation-related mismatch at the 3′ end of the primer annealing site (Figures 2A–2G), prompting redesign of an upstream primer to retrieve any additional insertions. Using this strategy, we recovered an additional 2 barcodes from circulating nRBCs for a total of 9 (Figure S3).

Because these expanded barcodes contributed roughly equally to every sample, we suspected multiple vector insertions within a single originating clonal HPSC. We plated ZL34 and control BM $CD34^+$ cells in semi-solid medium, allowing barcode retrieval from individual colony-forming units (CFUs). Of note, despite the erythroid predominance in the BM, a reduced ratio of erythroid and/or myeloid colonies was observed in ZL34 compared with normal BM (Figure S4). We plucked individual GFP^+ myeloid CFUs. All 9 barcodes were retrieved together from the majority of individual CFUs (Figure 2H). Rare CFUs lacked the 9 barcodes of interest, but each instead contained 1–4 other barcodes.

Vector copy numbers (VCNs) in samples from ZL34 and other lentivirally barcoded RMs was determined by qPCR. The normalized VCN per GFP^+ cell for nRBCs, eosinophils, monocytes, neutrophils, and $CD34^+$ HSPCs from late time points in ZL34 averaged 8.5 copies/cell (Figure 2I), which, along with the CFU barcode analysis, supports the conclusion that expansion from a single HSPC clone with 9 vector insertions was responsible for the myelodysplastic/myeloproliferative neoplasm in ZL34. The uninvolved B and T cell lineages in ZL34 had much lower VCNs per GFP^+ cell, as did myeloid and lymphoid lineages in other barcoded macaques (Figure 2I).

Retrieval of LV Insertion Sites for the 9 Dominant Barcodes

We recovered 9 LV insertions from circulating nRBCs using standard methodologies (Figure 3A; Table S2). In contrast, insertions retrieved from neutrophils on day 96 were polyclonal (Table S2). We were able to link each of the 9 dominant barcodes to one of the 9 nRBC insertions via PCR amplification and sequencing (Figure 3A). 4 insertions localized to introns within the *NCAM2*, *EIF3E*, *PLAG1*, and *IMMP2L* genes, with *PLAG1*, *NCAM2*, and *EIF3E* proviral insertions in the same orientation as endogenous transcripts and the *IMMP2L* insertion in the opposite orientation. Additional genes were located within 1 Mb windows surrounding each of the 9 insertions (Figures 3B and S5). One insertion was less than 20 kb downstream of *KITLG*, encoding stem cell factor (*SCF*), a cytokine of importance in hematopoiesis. None of the insertions localized to known CCCTC-binding factor (CTCF) binding sites or RM sequences homologous to open chromatin regions mapped in human hematopoietic cells (Figure S6).

Effect of Insertions on Gene Expression

We performed bulk RNA sequencing (RNA-seq) on ZL34 peripheral blood (PB) nRBCs, BM nRBCs, and BM $CD34^+$ HSPCs obtained on days 424–553 post-transplantation as well as on control RM BM nRBCs and BM $CD34^+$ HSPCs. No control RM PB nRBCs were obtained because nRBCs are not found in normal PB. We asked whether expression of genes interrupted by or within 1 Mb of any of the 9 insertions (Figures S5 and S7) were differentially expressed in ZL34 compared with controls. We found 5 genes within these windows to be significantly differentially expressed, combining nRBC and CD34 data and regressing out the source (PB or BM) and cell type (HSPCs or nRBCs) so that only significant differences attributed to condition (ZL34 cells versus controls) were identified (log2 fold change more than 1 or less than −1; adjusted $p < 0.01$) (Figures 3C and 3D). *NCAM2* and *PLAG1* with intronic insertions in the same orientation were overexpressed, *KITLG* with an insertion 24 kB downstream was overexpressed, *ISOC1* with an insertion 18 kB upstream was overexpressed, and *TES* with an insertion 250 kB upstream was underexpressed. When restricting analysis to nRBCs, 4 genes were differentially expressed: *NCAM2*, *PLAG1*, *KITLG*, and *ISOC1* (Figures S8A and S8B).

When considering all recovered transcripts, adjacent to insertions or not, a number of additional genes were differentially expressed. *KITLG*, *NCAM2*, and *PLAG1* were among the top 100 in combined HSPC and nRBC analysis (Figure 3E; Table S3) and were either within the top 100 or differentially expressed within nRBCs alone (Figure S8C; Table S3). As expected from the hemoglobin analysis

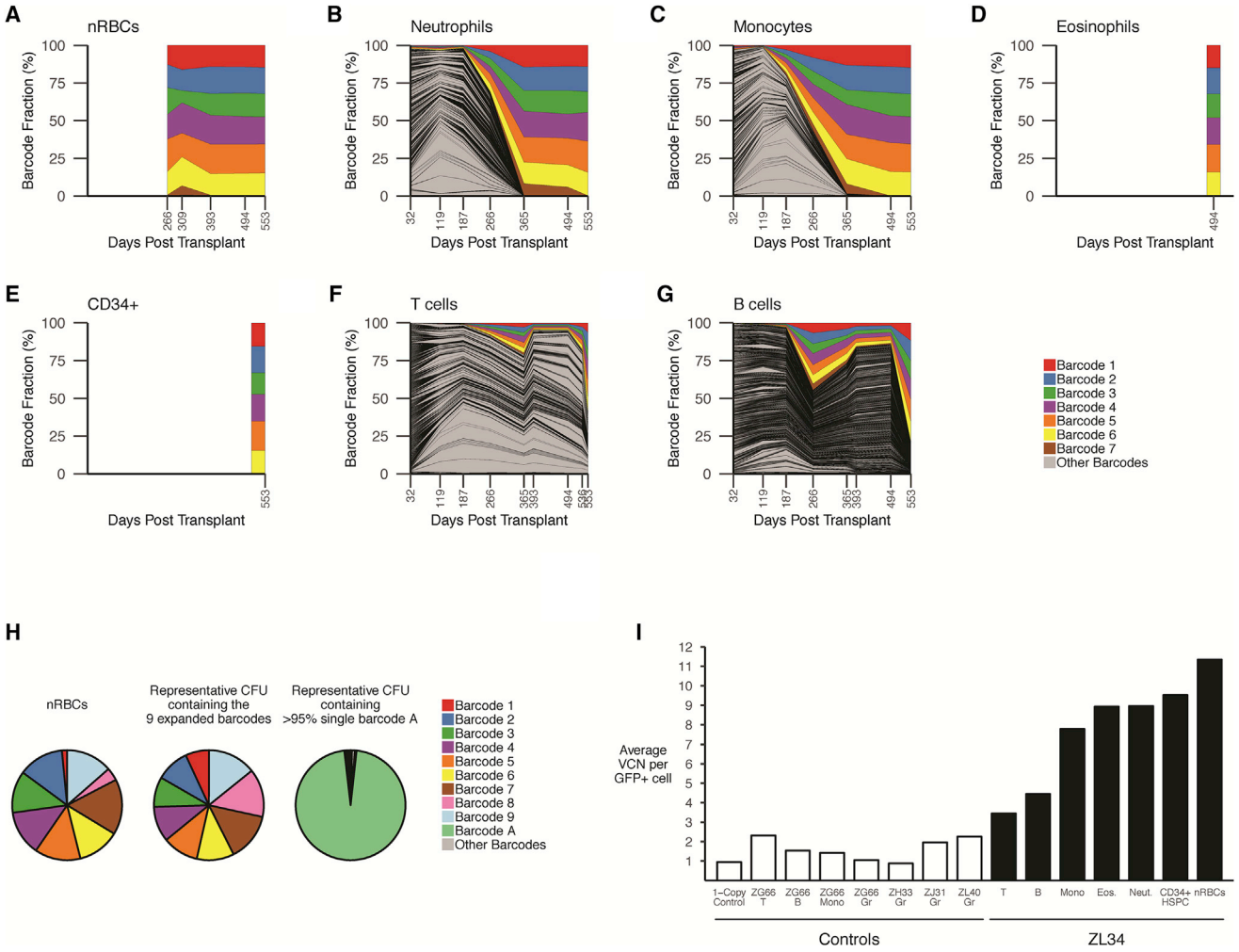


Figure 2. Clonal Tracking via Barcode Retrieval in ZL34

(A–G) Longitudinal clonal tracking of barcodes retrieved from nRBCs (A), neutrophils (B), monocytes (C), eosinophils (D), BM CD34$^+$ cells (E), T cells (F), and B cells (G). Each graph shows the fractional contributions from individual barcodes over time, with each colored ribbon representing the expanded barcodes detected via standard barcode recovery and all other barcodes shaded in gray. (H) Modified barcode recovery using the alternative forward primer on ZL34 nRBCs (393 days post-transplant), showing retrieval of 9 expanded barcodes. Modified barcode recovery using the same primers was performed on single-cell-derived GFP$^+$ CFUs from ZL34 BM CD34$^+$ cells (266 days post-transplant) plated at low density. Representative barcode analysis on a myeloid CFU containing all 9 barcodes is shown, matching those found in the circulating nRBC, confirming integration of the 9 barcodes in one original HSPC. For comparison, barcode analysis on a myeloid CFU from the same plate found to contain a single barcode, barcode A, is shown, confirming lack of cellular contamination across the plate that would complicate analysis. The small sector shown in gray represents multiple other barcodes with very low read counts, likely resulting from residual single cells not forming CFUs. (I) Vector copy number per GFP$^+$ cell quantified via qPCR on Gr (granulocytes), Mono (monocytes), T cells, B cells, Eos (eosinophils), nRBCs, and CD34$^+$ BM HSPCs from ZL34 (black bars, days 494–553 post-transplantation) and additional barcoded transplanted macaques (white bars, days 278–1,085 post-transplantation).

(Figure 1D), *HBG2* transcripts encoding the γ-globin chain were significantly upregulated in ZL34 nRBCs (Figure S8C).

SCF binds to the c-kit tyrosine kinase receptor on HSPCs and stimulates survival and proliferation.[21,22] SCF is normally produced by BM stromal elements, not HSPCs. We measured soluble SCF levels in ZL34 and control blood and in media conditioned by cultured ZL34 and normal PB mononuclear cells (MNCs) (Figure 4A). The secreted isoform of SCF was not detected above background in cultured samples. An increase in SCF was detected in ZL34 serum. Using an antibody that reacts with the transmembrane isoform of SCF, we found clear increased expression of cell-associated SCF in hematopoietic cells, as shown by immunohistochemistry on marrow sections and via fluorescence-activated cell sorting (FACS) (Figures 4B, 4C, and S9B).

PLAG1 is a zinc-finger transcription factor and proto-oncogene upregulated in salivary gland tumors[23] and in murine and human myeloid

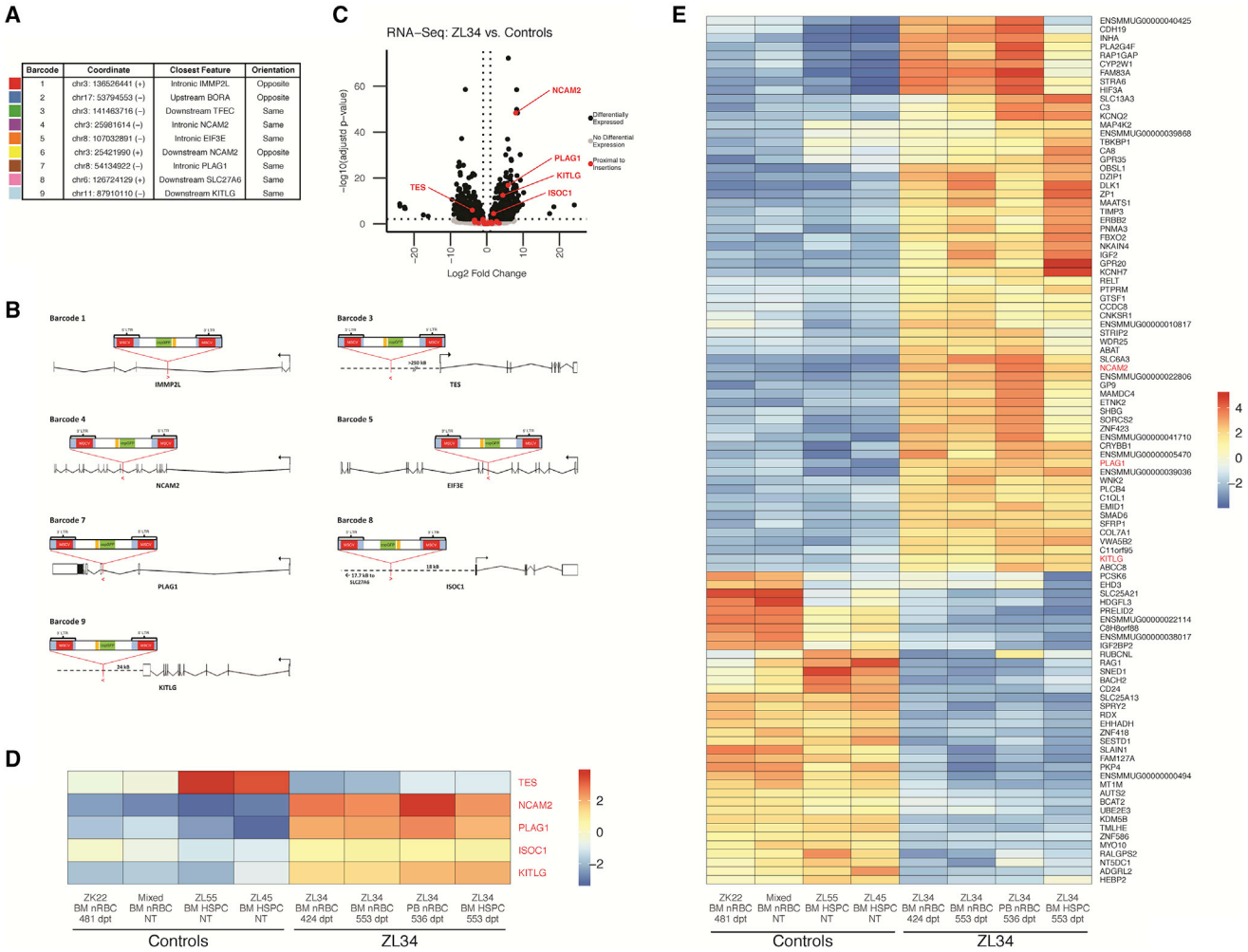


Figure 3. Identification of Insertion Sites and Effect on Gene Expression in Hematopoietic Cells

(A) Localization in the rhesus genome, closest gene, and proviral orientation relative to the closest gene for the 9 insertion sites corresponding to the 9 expanded barcodes. (B) Orientation and location of the provirus relative to the 5 proximally differentially expressed genes and 2 genes with intronic disruptions. (C) Volcano plot of RNA-seq data from normal control versus ZL34 nRBCs and CD34$^+$ cells showing all differentially expressed genes (black) as defined by adj. $p < 0.01$ and log2(fold change) > 1 or log2(fold change) < −1. Non-differentially expressed genes are shown in gray. Differentially expressed genes within 1 Mb upstream to 1 Mb downstream of any of the 9 insertion sites are colored in red and labeled. Two control BM nRBC samples (one from transplanted macaque ZK22 and one pooled from two non-transplanted macaques) and two control BM CD34$^+$ samples (from two different non-transplanted macaques) were analyzed along with two independent BM nRBC samples, one PB nRBC sample, and one CD34$^+$ sample from ZL34. (D) Mean-centered gene expression levels of *NCAM2*, *KITLG*, *PLAG1*, *ISOC1*, and *TES*. Expression levels are regularized log transformation (rlog) values from DESeq2. NT, non-transplanted. (E) Mean-centered gene expression levels of the top 100 differentially expressed genes (by adj. p value) in ZL34 compared with control nRBC and CD34$^+$ HSPC samples. Expression levels are rlog values from DESeq2. A complete list of differentially expressed genes is given in Table S3.

leukemias.[24,25] At the protein level, PLAG1 was overexpressed on a per-cell basis in ZL34 CD45$^-$ nRBCs, HSPCs, and monocytes but not T or B cells, as shown via intracellular FACS and, most strikingly, immunohistochemistry, revealing numerous intensely PLAG1-positive BM cells compared with controls (Figures 4B 4C, and S9B). Increased activity of PLAG1 was confirmed by significant increases in expression of known downstream targets, including *IGF2*, which encodes insulin-like growth factor 2 ($p = 3.66 \times 10^{-17}$) and *DLK1* ($p = 7.70 \times 10^{-31}$), both imprinted genes reported to be dysregulated in PLAG1-associated tumors (Figure 4D).[26–29] *MSI2*, a gene under direct regulation by PLAG1[30] whose product Musashi-2 is reported to promote HSPC expansion[31] and self-renewal was also upregulated ($p = 5.37 \times 10^{-5}$) (Figure 4D). Of note, mRNA encoding *HMGA2*, a transcription factor functionally upstream of PLAG1,[32] was expressed at a lower level in ZL34 nRBCs and HSPCs compared with control samples. Insertional overexpression of *HMGA2* has been previously linked to clonal expansion in a lentiviral gene therapy clinical trial,[8] suggesting that HMGA2 may stimulate HSPC clonal expansion via its effect on PLAG1 expression and activation of downstream targets active on HSPCs, such as *MSI2*, and that tonic low-level HMGA2

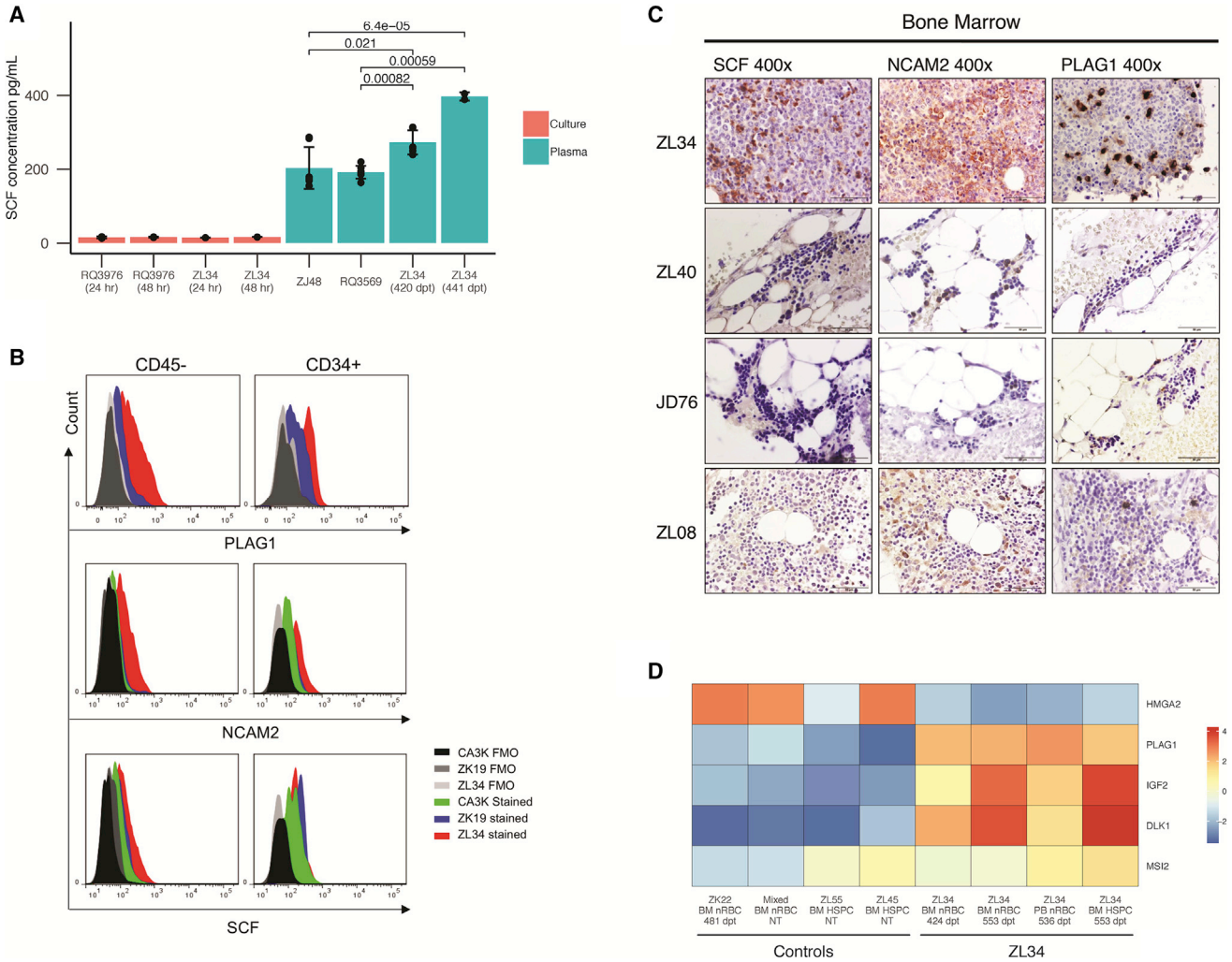


Figure 4. Analysis of Dysregulated Expression of *KITLG*, *NCAM2*, and *PLAG1*

(A) Stem cell factor (SCF) concentrations in serum and in culture media from high-density cultured PB MNCs for ZL34 and controls, as determined by ELISA. Error bars are shown for technical replicates. p values are shown and were determined by Welch's two-sample t test. (B) Expression of PLAG1, NCAM2, and SCF, assessed by flow cytometry in ZL34 and control non-transplanted normal macaque (ZK19) and a macaque recovering from malarial anemia (CA3K). Gating of BM CD45$^-$ cells was used as an alternative marker for nRBCs in these studies because of destruction of the CD71 epitope during permeabilization for intracellular staining. More than 85% of CD45$^-$ cells are nRBCs (Figure S9A). ZK19 is a non-transplanted normal monkey. (C) ZL34 and control BM samples (ZL08, non-transplanted; JD46 and ZL40, barcoded animals) stained via immunohistochemistry for PLAG1, SCF, and NCAM2. (D) Mean-centered gene expression levels for *PLAG1*, its putative upstream regulator *HMGA2*, and the PLAG1 downstream targets *IGF2*, *DLK1*, and *MSI2* in ZL34 and control samples. Expression levels are rlog values from DESeq2.

expression may be involved in normal HSPC homeostasis, controlled by feedback from downstream targets.

NCAM2 is an adhesion molecule required for cell-cell or cell-matrix interactions during nervous system development but is not normally expressed in hematopoietic cells, and dysregulation has not been linked to hematologic disease. Both FACS and immunohistochemistry confirmed overexpression of NCAM2 at a protein level in CD34$^+$ cells, nRBCs, and monocytes (Figures 4B, 4C, and S9B). Decreased expression of the tumor suppressor *TES* has not been previously linked to hematologic malignancies. The *TES* gene itself was not directly disrupted by an insertion; thus, downregulation of this gene must have been indirect via disruption of a regulatory element. There is little information about the gene product of *ISOC1*. It is not expressed in hematopoietic cells and has no known link to hematologic disease.

Insertion-Related Aberrant Splicing

LVs have been reported to disrupt splicing via formation of hybrid vector-gene transcripts expressed from viral promoters, resulting in dysregulated expression, alterations in isoform, and/or loss of transcript regulatory sequences.[33,34] We first interrogated RNA-seq data

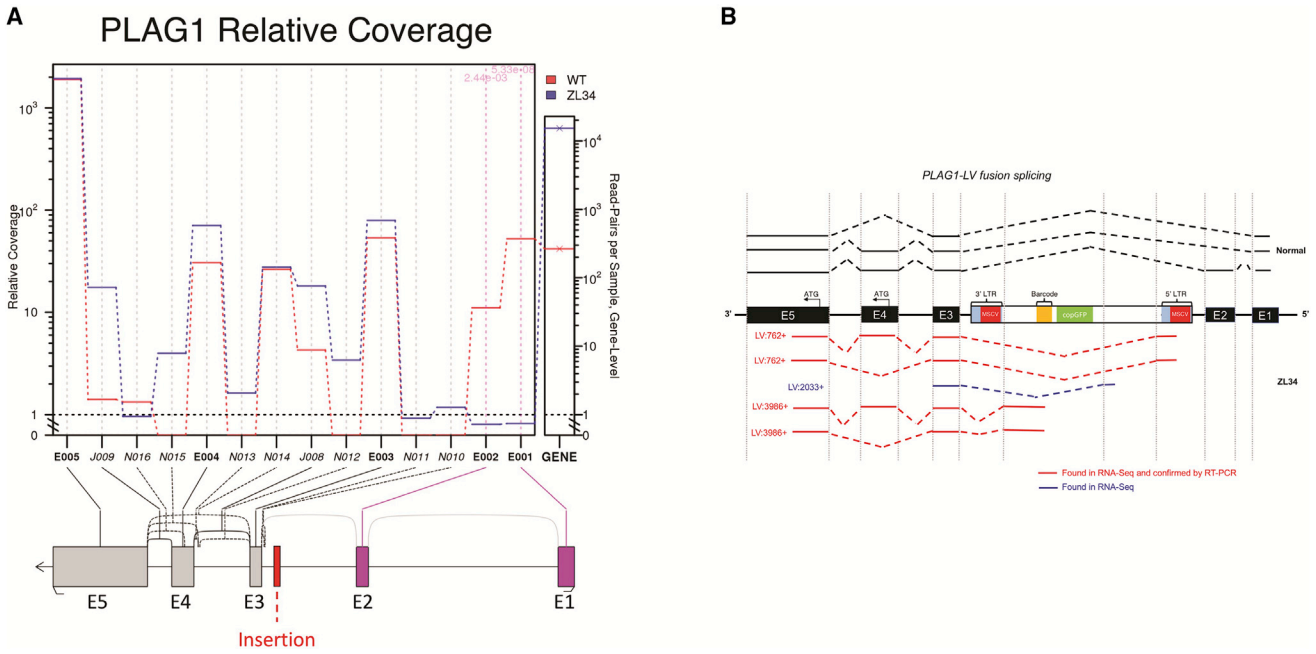


Figure 5. Abnormal Splicing Related to Lentiviral Insertions

(A) Differential expression of exons within ZL34 relative to control samples, as determined from RNA-seq data, showing significant downregulation of exons 1 and 2 within ZL34 *PLAG1* transcripts relative to control macaque nRBCs and HSPCs. (B) Hybrid vector-*PLAG1* mRNA species detected via RNA-seq fusion transcript analysis and/or confirmation by PCR and Sanger sequencing in ZL34 and control nRBCs. Lines denote transcripts detected in ZL34 (red) and normal controls (black). The results show the absence of exons 1 and 2 in ZL34 PLAG1 transcripts along with hybrid vector-*PLAG1* transcripts.

for any differentially spliced genes and found that ZL34 *PLAG1* transcripts contained significantly lower levels of exons E1 and E2 compared with controls (Figure 5A). We did not detect significant variable splicing for any other gene with intronic insertions (Table S4). However, several other genes of hematopoietic interest, including *HIF3A*, demonstrated both changes in isoform abundance (Table S4) and overall upregulation (Figure 3E). We next mined RNA-seq data for fusion transcripts. We detected fusions between vector and *PLAG1* sequences in ZL34 nRBCs and abundant hybrid forms containing vector sequences spliced to E3 before the translation start site,. These hybrid transcripts included isoforms with and without E4 (Figure 5B; Table S5), predicted to encode the two major isoforms of PLAG1 protein,[35] but no mutant fusion proteins. We also detected fusion *EIF3E* and *NCAM2* transcripts (Figure S10; Table S5); however, *EIF3E* was not overall upregulated in ZL34 (Figure S7). A subset of these hybrid transcripts was confirmed by RT-PCR (Figures 5B and S10). We did not detect hybrid transcripts for the other genes with intronic insertions or any non-LV fusion transcripts.

DISCUSSION

We describe the first case of lentiviral insertion-induced genotoxicity, resulting in hematopoietic clonal expansion and a neoplastic phenotype in a large animal or human. Following autologous transplantation of HSPCs transduced with a barcoded LV containing the MSCV promoter-enhancer within the LTR, one macaque developed a fatal myeloproliferative/myelodysplastic syndrome linked to clonal expansion from a single transduced HSPC.

The development of leukemias in clinical gene therapy trials and in primate models utilizing murine γ-retrovirus vectors (RVs) to transduce HSPCs stimulated a search for safer gene transfer systems.[9–11] Evidence implicated RVs activating nearby cellular proto-oncogenes with strong viral enhancers, exacerbated by RV integration favoring transcription start sites (TSSs).[13] LVs had endogenous enhancers removed and were shown to have an integration pattern favoring gene bodies rather than promoters, hypothesized to be less likely to activate proto-oncogenes.[36,37] Genotoxicity models, including *in vitro* immortalization of murine HSPCs and acceleration of leukemia in tumor-prone mice, predicted significantly lower genotoxicity from LVs versus RVs, even when strong viral promoters and/or enhancers derived from RVs were inserted within LV backbones, although inclusion of these RV elements within LVs, particularly within the viral LTR, did increase genotoxicity over LVs without these elements.[38–42] The majority of HSPC clinical gene therapy trials over the past decade have utilized LVs to try and decrease the risk of genotoxicity and take advantage of the increased efficiency of HSPC transduction associated with this vector class.[1,43]

In the majority of our barcoded macaques, including ZL34, we used a LV construct containing the MSCV viral promoter and/or enhancer inserted within the LTR. The original MSCV RV vector was derived

from a mutant murine myeloproliferative sarcoma virus (MSPV) with high expression in hematopoietic cells, and the MSCV promoter and/or enhancer was later shown to drive consistently high stable transgene expression in human HSPCs when included in LV vectors.[44,45] In a previous study, we reported development of acute myeloid leukemia (AML) following transduction of macaque HSPCs using an RV vector containing the MSCV promoter and/or enhancer.[10] However, in 8 macaques transplanted with HSPCs transduced with the same LV barcoding vector containing the MSCV sequences, we tracked over 65,000 individual clones for up to 5 years without clonal expansions. In addition, 3 macaques that received HSPCs transduced with a barcoded LV containing an internal EF1α promoter showed clonal patterns indistinguishable from the MSCV animals (Table S1).[15]

Because RV promoters and/or enhancers included in LVs promote high constitutive transgene expression, two clinical trials for adrenoleukodystrophy have included an RV promoter and/or enhancer similar to MSCV, termed MND, in the LV to drive high expression of the transgene.[4,5,46,47] In contrast to the vector utilized in ZL34, the adrenoleukodystrophy trial LV placed the RV promoter and/or enhancer internally instead of within the LV LTR.[4,5,46,47] No clinical trials to date have utilized an LV with an RV promoter and/or enhancer within the LTR. Tumor-prone mouse models have suggested that an internal strong promoter and/or enhancer is less genotoxic than the same promoter and/or enhancer placed within the LV LTR.[39] So far, no clonal expansions or hematopoietic toxicity have been reported in the 23 patients enrolled in these trials, with many years of follow-up. Other clinical trials using LVs containing endogenous or weaker constitutive promoters have reported polyclonal hematopoiesis for years post-transplantation.[6,48–51] However, a thalassemia patient developed a marked clonal expansion linked to insertion of an LV containing an internal erythroid-specific promoter, resulting in increased expression and aberrant splicing of *HMGA2*.[8] In contrast to our macaque, the clonal expansion in this patient was transient and not associated with any documented hematologic abnormalities. Recently, persistence and massive clonal expansion of a chimeric antigen receptor (CAR)-T cell clone in a patient was linked to an LV insertion inactivating the epigenetic regulator TET2.[14]

The expanded neoplastic clone in our macaque had 9 insertions, with 4 of the 9 insertions located within introns. Ongoing clinical trials for disorders requiring high transgene expression, such as sickle cell disease, are targeting and achieving VCNs of 4 or greater *in vivo* long-term to achieve disease correction; thus, a fraction of HSPCs must have VCNs as high as 9. We believe that the most compelling candidate insertion driving the HSPC phenotype was within the gene encoding the zinc-finger protein PLAG1. *PLAG1* has been identified as a cooperating neoplastic "hit" via a mutagenesis screen in *Cbfb-MYH11* murine leukemias and is upregulated in 20% of human AML.[24] Recent studies have reported a role for PLAG1 in stimulating expression of Musashi-2, a protein that stimulates HSPC expansion.[30] Detection of abundant aberrant splicing and *LV-PLAG1* mRNA fusions corroborates previous reports of LV-associated aberrant splicing.[8,33,34,52,53] Of note, the LV-*PLAG1* splicing abrogated expression of non-translated E1 and E2 while maintaining expression of exons 3, 4, and 5, encompassing all translated sequences. We detected transcripts both with and without E4, predicting translation of both the short and long isoforms of PLAG1 protein.[26] Functional differences between the two PLAG1 isoforms are poorly understood.

We observed upregulation of expression of known downstream targets of PLAG1 in ZL34 cells, most notably *IGF2* and *MSI2*.[30] Imprinting-related downregulation of *IGF2* at the *IGF2/H19* locus maintains HSPC quiescence, with loss of imprinting resulting in murine HSPC proliferation and exhaustion.[54] *MSI2* encodes the Musashi-2 RNA-binding protein, which promotes the expansion of long-term repopulating HSCs[31] and regulates self-renewal.[55] It is interesting to note that *HMGA2* was downregulated in ZL34 hematopoietic cells compared with those from normal controls. *HMGA2 is* a transcription factor that functions as a critical upstream regulator of *PLAG1* expression.[32] An LV insertion resulting in overexpression of *HMGA2* was linked to marked clonal expansion in an LV gene therapy trial noted above,[8] and, based on the HSPC expansion seen in our animal, it seems possible that dysregulation of PLAG1 and downstream targets may have contributed to the expansion, although, to our knowledge, no aberrant hematologic phenotype developed in that patient. The downregulation of *HMGA2* in ZL34 HSPCs coincident with upregulation of *PLAG1* and downstream targets suggests that HMGA2 may be involved in normal hematopoiesis and regulated by feedback. Finally, the elevated levels of *IGF2* mRNA resulting from *PLAG1* overexpression may also have played a role in the persistently high fetal hemoglobin (HbF) levels observed in ZL34.[56]

Although we suspect that genotoxicity linked to the *PLAG1* insertion likely played a key role in ZL34's phenotype, we cannot rule out roles for other insertions. Given the doses transplanted, HSPCs with LV insertions in *PLAG1* have almost certainly been delivered to our other macaques without development of abnormal clonal hematopoiesis. An insertion just downstream of *KITLG* resulted in upregulation of SCF, with a small increase in serum SCF. Cell-associated SCF protein was increased in HSPCs also expressing the KIT receptor, suggesting autocrine signaling, shown previously for other cytokines to drive proliferation in hematopoietic cells.[57] We administered pharmacologic doses of SCF to macaques and did not observe eosinophilia or release of nRBCs into the circulation; thus, overexpression of SCF was unlikely to be the sole driver of ZL34's aberrant hematopoiesis. However, it is of interest to note that production of fetal hemoglobin increases in baboons administered SCF[58] and in human erythroid cultures supplemented with SCF,[59] suggesting that dysregulation of SCF may have contributed to elevated HbF in ZL34. Insertions upregulating expression or leading to aberrant splicing of *NCAM2*, *ISOC1*, or *EIF3E* seem less likely to be contributing to clonal expansion and the phenotype, given the lack of known roles for these genes in hematopoiesis or leukemogenesis. Expression of testin (*TES*) RNA was decreased, a gene with tumor suppressor activities, potentially via inhibition of cell migration and metastasis.[60] It is intriguing to speculate

that loss of TES expression might have contributed to abnormal release of nRBCs into the blood. The constraints of the primate model and our inability to derive immortalized lines from ZL34 precluded further direct investigations of the contribution of each specific insertion to the hematologic phenotype.

In conclusion, this fatal genotoxic clonal expansion adverse event linked to LV insertions is concerning. The implications of this event must be balanced against the encouraging safety and efficacy of LV HSPC gene therapy clinical trials in patients with life-threatening diseases. It is also important to note that the vector resulting in toxicity in our macaque model included a strong viral promoter and/or enhancer within the LTR, a design shown to be associated with increased genotoxicity in preclinical models and not utilized in any clinical trial to date. However, we believe our findings imply that strong viral promoters should be avoided in LVs whenever possible. In addition, approaches to decrease LV-related aberrant splicing should be pursued, given our evidence that upregulation of *PLAG1* was driven primarily by hybrid transcripts. Targeted gene editing approaches avoid semi-random vector insertion and have the potential to be a less genotoxic technology, although these approaches are in their infancy and may be associated with a different set of off-target genotoxicities.[61–63]

MATERIALS AND METHODS

Autologous Transplantation of Lentivirally Barcoded HSPCs

$CD34^+$ HSPCs were collected from ZL34, a 4-year-old male RM, following mobilization with granulocyte colony-stimulating factor (G-CSF)/AMD1300, transduced (MOI = 25) with a barcoded lentivirus library, and reinfused into the macaque following TBI as described in a protocol approved by the National Heart, Lung, and Blood Institute (NHLBI) Animal Care and Use Committee, following all applicable animal care regulations.[15–18] ZL34 received weekly dosing with 5 mg/kg cidofovir to suppress CMV reactivation.

Barcode Retrieval

DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN). 200 ng was used for barcode retrieval PCR with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) via 28 cycles of 98°C for 10 s, 70°C for 30 s, 72°C for 30 s, and then 72°C for 10 min. Indexed or non-indexed forward primers and indexed reverse primers were added (Table S6) for multiplex sequencing. Following gel purification, 15–24 multiplexed samples were used to create a DNA library for sequencing on an Illumina HiSeq2500/3000. Custom Python code (https://github.com/d93espinoza/barcode_extracter) was used for extraction and processing of barcodes from FASTQ files as described previously.[15–18] For modified barcode recovery, "Starcode" software (https://github.com/gui11aume/starcode) was used for extraction and processing of barcodes. Visualization was performed using custom R code.

Cell Purification and FACS Analysis

PB and BM samples were separated into granulocyte and mononuclear cell fractions via centrifugation over lymphocyte separation medium (MP Biomedical) and stained with antibodies (Table S7) prior to flow cytometric analysis and sorting to purities of more than 98%. $CD34^+$ cells were purified by magnetic bead immunoselection (Miltenyi Biotec) to more than 95% purity. Neutrophils and eosinophils within the granulocyte pellet were separated based on CD33 expression. $CD45^-CD71^+$ flow gating was used to purify nRBCs for all figures, with the exception of those requiring intracellular PLAG1/NCAM2/SCF FACS staining (Figures 4B and S9), in which solely $CD45^-$ gating (>85% nRBCs) was used.

Colony Formation Assays

100 $CD34^+$ BM cells purified from ZL34 BM were plated in 1 mL methylcellulose medium (STEMCELL Technologies, catalog number H4435) and cultured for 14 days. Single widely separated colonies were plucked for DNA extraction in 20 μL DirectPCR lysis reagent buffer (Viagen Biotech, catalog number 401-E) with 1 μL proteinase K and 2 mg/mL RNase.

Vector Integration Site Retrieval

Vector integration sites were identified and retrieved as described previously.[51] Briefly, genomic DNA was sheared (Covaris) to an average size of 300–500 bp, end-repaired, and dA-tailed. T-linkers (Oligoseq) were ligated to the resulting fragments. PCR was performed using one LTR-specific primer and one linker-specific primer, followed by nested primers containing Illumina sequencing adaptors. The resulting library was sequenced on an Illumina MiSeq. Integration site junctions were trimmed with custom Perl scripts and mapped to rheMac8 using Blat.

Analysis of Vector Integration Sites Relative to Known Regulatory Sequences

The locations of vector integration sites were compared with CTCF binding sites as determined by chromatin immunoprecipitation sequencing (ChIP-seq). FASTQ files from four ChIP-seq experiments using hepatocytes of three male macaques were obtained (ArrayExpress: E-MTAB-437).[64] We used STAR[65] to map the reads to the rheMac8/Mmu8_8.0.1 reference genome and MACS2[66] to predict the genomic locations of CTCF binding sites from the aligned reads. Finally, the Bioconductor package GenomicRanges[67] was used to compute overlaps between vector integration sites and CTCF binding sites.

Locations of vector integration sites were also more broadly compared with the locations of other regulatory elements, as determined by an assay for transposase-accessible chromatin using sequencing (ATAC-seq) data. ATAC-seq data from 18 human hematopoietic lineages were downloaded (GEO: GSE96772).[68] We translated the ATAC-seq peaks from human genomic coordinates (aligned to hg19) into macaque genomic coordinates (aligned to rheMac8) using the University of California, Southern California (UCSC) batch coordinate conversion program Liftover and the appropriate chain file from the USCS Genome Browser website (http://genome.ucsc.edu). Again we used the Bioconductor package

GenomicRanges[67] to compute overlaps between vector integration sites and ATAC-seq peaks.

Hemoglobin Chain Measurements

Whole-blood cell pellets were lysed in 100 μL high-pressure liquid chromatography (HPLC)-grade water by pulse-vortexing three times for 30 s, followed by one cycle of freeze-thawing and centrifugation at 16,000 × *g* for 15 min. 10 μL of 100 mM Tris(2-chloroethyl) phosphate (TCEP; Thermo Fisher Scientific) was added, and after 5 min of incubation at room temperature, 85 μL of 0.1% trifluoroacetic acid (TFA)/32% acetonitrile was added. 10-μL samples were analyzed at a 0.7 mL/min flow rate for 50 min using the Agilent 1100 HPLC (Agilent Technologies) equipped with the Aeris 3.6-μm Widepore C4 200 (250 × 4.6 mm, Phenomenex, Torrance, CA) reverse-phase column using solvent A (0.12% TFA in water) and solvent B (0.08% TFA in acetonitrile). A starting concentration of solvent B was designated as 35% for the separation of globin proteins and the percentage of solvent B was changed as follows: 3 min at up to 41.2%, 3 min at up to 41.6%, 5 min at up to 42%, 4 min at up to 42.4%, 6 min at up to 42.8%, 6 min at up to 44.4%, 6 min at up to 47%, 7 min at up to 75%, and re-equilibration for 10 minutes at 35%. The globin types were detected at 215 nm and confirmed by an Agilent HPLC-6224 mass spectrometer equipped with an electrospray ionization (ESI) interface and a time-of-flight (TOF) mass detector as described previously.[69,70]

Soluble SCF Detection Using ELISA

Plasma was obtained by centrifuging whole blood at 2,000 × *g* for 15 min at 4°C. 1 × 10^7 PB MNCs were cultured for 24 and 48 h in RPMI10 medium supplemented with 10% fetal bovine serum. The stem cell factor human SimpleStep ELISA Kit (Abcam, Cambridge, MA; catalog number 176109) was used to measure SCF levels in plasma and conditioned culture medium.

RNA-Seq and Analysis

$CD45^-CD71^+$ nRBC and $CD34^+$ cells from ZL34 (samples collected on days 410–540 post-transplantation) or control monkey PB or BM (4 age-matched untransplanted animals and 1 transplanted animal day 465 post-transplantation) were purified by FACS (nRBCs) or immunoselection ($CD34^+$ cells). Total RNA was extracted with RNAzol RT (MRC, OH, USA). RNA libraries were prepared using the Illumina TruSeq Stranded Total RNA kit with Ribo-Zero and sequenced on an Illumina HiSeq 3000.

RNA-seq data analysis was performed using STAR,[65] DESeq2,[71] and custom R code. Genome assembly Mmul_8.0.1 was used. Macaque annotation from Ensembl release 92 was used. Analysis of nRBC and HSPC data together was performed in DESeq2 using the design formula ~source + cell type + condition, where source was PB or BM, cell type was HPSC or nRBC, and condition was ZL34 or control. nRBC-only analysis was performed using the design formula ~source + condition. Built-in Cook's cutoff was used in DESeq2 to alleviate the effect of outliers on resulting differentially expressed genes.

Histologic and Immunohistochemical Analyses

BM biopsies were obtained from the posterior iliac crests. Following fixation and decalcification, sections were prepared. Spleen samples were obtained at the time of autopsy. Antibodies utilized for immunohistochemical staining are listed in Table S7.

Analysis of Differential Usage of Splice Variants

We implemented a pipeline to detect fusion genes in parallel with computing differential usage of gene features and detects. In brief, STAR v 2.6.0c[65] was used to align reads to a combination *Macaca mulatta*, (ENSEMBL Mmul_8.0.1) plus custom vector reference genomes and a combination *Macaca mulatta* annotation (ENSEMBL, Mmul_8.0.1.92 general transfer format) plus custom annotation for the vector sequence. Then STAR-Fusion (https://github.com/STAR-Fusion/STAR-Fusion/wiki) was used to compute fusion genes using a custom genome resource library for the macaque and vector reference genomes, which was built using FusionFilter. Differential usage of gene features was computed using QoRTs[72] and JunctionSeq.[73] Our implementation of differential gene expression relied on DESeq2.[71] All components of our pipeline were unified under custom shell and R scripts that were run on the NIH High Performance Computing cluster. All of our source code and scripts are available via BitBucket (https://bitbucket.org/DunbarLab_Releases/variant-splicing_fusion-genes).

Data Availability

All data and code used in this study (RNA-seq and barcode data) will be made available upon request from the corresponding authors.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.ymthe.2019.04.003.

AUTHOR CONTRIBUTIONS

Conceptualization, C.E.D. and C.W.; Analytics, D.A.E., S.F.C., and L.L.T.; Investigation, D.A.E., X.F., D.Y., C.W., K.R.C., I.M.Y., S.D., X.W., and S.S.D.; Resources, R.L., H.L.M., N.U., J.F.T., and K.J.H.; Animal Support, S.G.H., A.K., M.M., and R.E.D.; Writing, C.E.D., D.A.E., and C.W.; Supervision, C.E.D., C.W., and J.F.T.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

This research was supported by the NHLBI Division of Intramural Research. Di Yang was funded by the Scientific Research Training Program for Young Talents sponsored by Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, China. We thank Keyvan Keyvanfar, the NHLBI FACS and DNA Sequencing and Genomics Cores, the NIH Biowulf High-Performance Computing Resource, NIH veterinary pathology, and NHLBI animal care staff for assistance. We acknowledge Patrick Duffy and Amber Raja for supplying marrow from animal CA3K.

REFERENCES

1. Naldini, L., Trono, D., and Verma, I.M. (2016). Lentiviral vectors, two decades later. Science *353*, 1101–1102.
2. 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.
3. Biffi, A., Montini, E., Lorioli, L., Cesani, M., Fumagalli, F., Plati, T., Baldoli, C., Martino, S., Calabria, A., Canale, S., et al. (2013). Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. Science *341*, 1233158.
4. 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.
5. 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.
6. Thompson, A.A., Walters, M.C., Kwiatkowski, J., Rasko, J.E.J., Ribeil, J.A., Hongeng, S., Magrin, E., Schiller, G.J., Payen, E., Semeraro, M., et al. (2018). Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia. N. Engl. J. Med. *378*, 1479–1493.
7. 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.
8. 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 β-thalassaemia. Nature *467*, 318–322.
9. 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.
10. 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.
11. Rivière, I., Dunbar, C.E., and Sadelain, M. (2012). Hematopoietic stem cell engineering at a crossroads. Blood *119*, 1107–1116.
12. Schröder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521–529.
13. Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749–1751.
14. Fraietta, J.A., Nobles, C.L., Sammons, M.A., Lundh, S., Carty, S.A., Reich, T.J., Cogdill, A.P., Morrissette, J.J.D., DeNizio, J.E., Reddy, S., et al. (2018). Disruption of TET2 promotes the therapeutic efficacy of CD19-targeted T cells. Nature *558*, 307–312.
15. 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.
16. 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.
17. Wu, C., Espinoza, D.A., Koelle, S.J., Potter, E.L., Lu, R., Li, B., Yang, D., Fan, X., Donahue, R.E., and Dunbar, C.E. (2017). Geographic clonal tracking in macaques provides insights into HSPC migration and differentiation. J. Exp. Med. *215*, 217–232.
18. Yu, K.-R., Espinoza, D.A., Wu, C., Truitt, L., Shin, T.-H., Chen, S., Fan, X., Yabe, I.M., Panch, S., Hong, S.G., et al. (2018). The impact of aging on primate hematopoiesis as interrogated by clonal tracking. Blood *131*, 1195–1205.
19. Humbert, O., Peterson, C.W., Norgaard, Z.K., Radtke, S., and Kiem, H.P. (2017). A Nonhuman Primate Transplantation Model to Evaluate Hematopoietic Stem Cell Gene Editing Strategies for β-Hemoglobinopathies. Mol. Ther. Methods Clin. Dev. *8*, 75–86.
20. Radtke, S., Adair, J.E., Giese, M.A., Chan, Y.Y., Norgaard, Z.K., Enstrom, M., Haworth, K.G., Schefter, L.E., and Kiem, H.P. (2017). A distinct hematopoietic stem cell population for rapid multilineage engraftment in nonhuman primates. Sci. Transl. Med. *9*, 9.
21. Li, C.L., and Johnson, G.R. (1994). Stem cell factor enhances the survival but not the self-renewal of murine hematopoietic long-term repopulating cells. Blood *84*, 408–414.
22. Leary, A.G., Zeng, H.Q., Clark, S.C., and Ogawa, M. (1992). Growth factor requirements for survival in G0 and entry into the cell cycle of primitive human hemopoietic progenitors. Proc. Natl. Acad. Sci. USA *89*, 4013–4017.
23. Kas, K., Voz, M.L., Röijer, E., Aström, A.K., Meyen, E., Stenman, G., and Van de Ven, W.J. (1997). Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. Nat. Genet. *15*, 170–174.
24. Castilla, L.H., Perrat, P., Martinez, N.J., Landrette, S.F., Keys, R., Oikemus, S., Flanegan, J., Heilman, S., Garrett, L., Dutra, A., et al. (2004). Identification of genes that synergize with Cbfb-MYH11 in the pathogenesis of acute myeloid leukemia. Proc. Natl. Acad. Sci. USA *101*, 4924–4929.
25. Landrette, S.F., Kuo, Y.H., Hensen, K., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Perrat, P.N., Van de Ven, W.J.M., Delwel, R., and Castilla, L.H. (2005). Plag1 and Plagl2 are oncogenes that induce acute myeloid leukemia in cooperation with Cbfb-MYH11. Blood *105*, 2900–2907.
26. Declercq, J., Van Dyck, F., Braem, C.V., Van Valckenborgh, I.C., Voz, M., Wassef, M., Schoonjans, L., Van Damme, B., Fiette, L., and Van de Ven, W.J. (2005). Salivary gland tumors in transgenic mice with targeted PLAG1 proto-oncogene overexpression. Cancer Res. *65*, 4544–4553.
27. Declercq, J., Skaland, I., Van Dyck, F., Janssen, E.A., Baak, J.P., Drijkoningen, M., and Van de Ven, W.J. (2008). Adenomyoepitheliomatous lesions of the mammary glands in transgenic mice with targeted PLAG1 overexpression. Int. J. Cancer *123*, 1593–1600.
28. Van Dyck, F., Scroyen, I., Declercq, J., Sciot, R., Kahn, B., Lijnen, R., and Van de Ven, W.J. (2008). aP2-Cre-mediated expression activation of an oncogenic PLAG1 transgene results in cavernous angiomatosis in mice. Int. J. Oncol. *32*, 33–40.
29. Voz, M.L., Agten, N.S., Van de Ven, W.J., and Kas, K. (2000). PLAG1, the main translocation target in pleomorphic adenoma of the salivary glands, is a positive regulator of IGF-II. Cancer Res. *60*, 106–113.
30. Belew, M.S., Bhatia, S., Keyvani Chahi, A., Rentas, S., Draper, J.S., and Hope, K.J. (2018). PLAG1 and USF2 Co-regulate Expression of Musashi-2 in Human Hematopoietic Stem and Progenitor Cells. Stem Cell Reports *10*, 1384–1397.
31. Rentas, S., Holzapfel, N., Belew, M.S., Pratt, G., Voisin, V., Wilhelm, B.T., Bader, G.D., Yeo, G.W., and Hope, K.J. (2016). Musashi-2 attenuates AHR signalling to expand human haematopoietic stem cells. Nature *532*, 508–511.
32. Klemke, M., Müller, M.H., Wosniok, W., Markowski, D.N., Nimzyk, R., Helmke, B.M., and Bullerdiek, J. (2014). Correlated expression of HMGA2 and PLAG1 in thyroid tumors, uterine leiomyomas and experimental models. PLoS ONE *9*, e88126.
33. Cesana, D., Sgualdino, J., Rudilosso, L., Merella, S., Naldini, L., and Montini, E. (2012). Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. J. Clin. Invest. *122*, 1667–1676.
34. Moiani, A., Paleari, Y., Sartori, D., Mezzadra, R., Miccio, A., Cattoglio, C., Cocchiarella, F., Lidonnici, M.R., Ferrari, G., and Mavilio, F. (2012). Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. J. Clin. Invest. *122*, 1653–1666.
35. Debiec-Rychter, M., Van Valckenborgh, I., Van den Broeck, C., Hagemeijer, A., Van de Ven, W.J., Kas, K., Van Damme, B., and Voz, M.L. (2001). Histologic localization of PLAG1 (pleomorphic adenoma gene 1) in pleomorphic adenoma of the salivary gland: cytogenetic evidence of common origin of phenotypically diverse cells. Lab. Invest. *81*, 1289–1297.
36. 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.

37. Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. *2*, E234.

38. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. Dunbar, C.E., High, K.A., Joung, J.K., Kohn, D.B., Ozawa, K., and Sadelain, M. (2018). Gene therapy comes of age. Science *359*, 359.

44. Grez, M., Akgün, E., Hilberg, F., and Ostertag, W. (1990). Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. Proc. Natl. Acad. Sci. USA *87*, 9202–9206.

45. Ramezani, A., Hawley, T.S., and Hawley, R.G. (2000). Lentiviral vectors for enhanced gene expression in human hematopoietic cells. Mol. Ther. *2*, 458–469.

46. Challita, P.M., Skelton, D., el-Khoueiry, A., Yu, X.J., Weinberg, K., and Kohn, D.B. (1995). Multiple modifications in cis elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma cells. J. Virol. *69*, 748–755.

47. Robbins, P.B., Skelton, D.C., Yu, X.J., Halene, S., Leonard, E.H., and Kohn, D.B. (1998). Consistent, persistent expression from modified retroviral vectors in murine hematopoietic stem cells. Proc. Natl. Acad. Sci. USA *95*, 10182–10187.

48. 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.

49. 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.

50. Hacein-Bey Abina, S., Gaspar, H.B., Blondeau, J., Caccavelli, L., Charrier, S., Buckland, K., Picard, C., Six, E., Himoudi, N., Gilmour, K., et al. (2015). Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome. JAMA *313*, 1550–1563.

51. De Ravin, S.S., Wu, X., Moir, S., Anaya-O'Brien, S., Kwatemaa, 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.

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. Heckl, D., Schwarzer, A., Haemmerle, R., Steinemann, D., Rudolph, C., Skawran, B., Knoess, S., Krause, J., Li, Z., Schlegelberger, B., et al. (2012). Lentiviral vector induced insertional haploinsufficiency of Ebf1 causes murine leukemia. Mol. Ther. *20*, 1187–1195.

54. Venkatraman, A., He, X.C., Thorvaldsen, J.L., Sugimura, R., Perry, J.M., Tao, F., Zhao, M., Christenson, M.K., Sanchez, R., Yu, J.Y., et al. (2013). Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. Nature *500*, 345–349.

55. Hope, K.J., Cellot, S., Ting, S.B., MacRae, T., Mayotte, N., Iscove, N.N., and Sauvageau, G. (2010). An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. Cell Stem Cell *7*, 101–113.

56. de Vasconcellos, J.F., Tumburu, L., Byrnes, C., Lee, Y.T., Xu, P.C., Li, M., Rabel, A., Clarke, B.A., Guydosh, N.R., Proia, R.L., and Miller, J.L. (2017). IGF2BP1 overexpression causes fetal-like hemoglobin expression patterns in cultured human adult erythroblasts. Proc. Natl. Acad. Sci. USA *114*, E5664–E5672.

57. Dunbar, C.E., Browder, T.M., Abrams, J.S., and Nienhuis, A.W. (1989). COOH-terminal-modified interleukin-3 is retained intracellularly and stimulates autocrine growth. Science *245*, 1493–1496.

58. Lavelle, D., Molokie, R., Ducksworth, J., and DeSimone, J. (2001). Effects of hydroxyurea, stem cell factor, and erythropoietin in combination on fetal hemoglobin in the baboon. Exp. Hematol. *29*, 156–162.

59. Aerbajinai, W., Zhu, J., Kumkhaek, C., Chin, K., and Rodgers, G.P. (2009). SCF induces gamma-globin gene expression by regulating downstream transcription factor COUP-TFII. Blood *114*, 187–194.

60. Li, X. (2008). Tes, a potential Mena-related cancer therapy target. Drug Discov. Ther. *2*, 1.

61. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nat. Med. *24*, 927–930.

62. Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., et al. (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat. Med. *24*, 939–946.

63. Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat. Biotechnol. *36*, 765–771.

64. Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335–348.

65. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

66. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

67. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118.

68. Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., Greenleaf, W.J., et al. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell *173*, 1535–1548.e16.

69. Taggart, C., Cervantes-Laurean, D., Kim, G., McElvaney, N.G., Wehr, N., Moss, J., and Levine, R.L. (2000). Oxidation of either methionine 351 or methionine 358 in alpha 1-antitrypsin causes loss of anti-neutrophil elastase activity. J. Biol. Chem. *275*, 27258–27265.

70. Apffel, A., Fischer, S., Goldberg, G., Goodley, P.C., and Kuhlmann, F.E. (1995). Enhanced sensitivity for peptide mapping with electrospray liquid chromatography-mass spectrometry in the presence of signal suppression due to trifluoroacetic acid-containing mobile phases. J. Chromatogr. A *712*, 177–190.

71. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

72. Hartley, S.W., and Mullikin, J.C. (2015). QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC Bioinformatics *16*, 224.

73. Hartley, S.W., and Mullikin, J.C. (2016). Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucleic Acids Res. *44*, e127.