OMTM, Volume *8*

Supplemental Information

A Nonhuman Primate Transplantation Model to Evaluate Hematopoietic Stem Cell Gene Editing Strategies for β-Hemoglobinopathies

Olivier Humbert, Christopher W. Peterson, Zachary K. Norgaard, Stefan Radtke, and Hans-Peter Kiem

SUPPLEMENTAL MATERIALS AND METHODS

Statistical analysis of HbF expression

A robust MANOVA was conducted on the data using Munzel and Brunner's method,[48] implemented in R using the mulrank() function.[49] There was a significant main effect of the editing treatment on HbF expression as measured by % γ-globin (qPCR) and % HbF (flow), $F = 3.780$, $p < .01$. Univariate ANOVAs conducted using the aov() function in R confirm the significant effects observed in HbF as measured by % γ-globin (qPCR), $F(4) = 20.97$, $p < .01$, and % HbF (flow), $F(4) = 29.70$, $p < .01$. Constructing linear models of the % γ-globin and % HbF values using the lm() function from R and examining the contrasts between the editing treatments using the summary. lm() function, we observe that only the Bcl11a TALEN treatment resulted in a significantly different HbF when contrasted with the mock control as measured by % γ-globin (qPCR), $p < .001$. When examining HbF as measured by % F-Cell (flow), the Bcl11a TALEN and CD33 CRISPR treatments, $p < .001$ and $p < .05$, respectively, also appeared significant. A subsequent discriminant analysis using the lda() function in R reveals two discriminant functions. The first explained 98.06% of the variance, whereas the second explained only 1.94%. The coefficients of the discriminant functions revealed that function 1 produces similar coefficients of similar strength for % HbF ($b = 0.348$) and % γ-globin ($b = 0.397$). The second variate differentiates % HbF ($b = -0.250$) and % γ-globin ($b = 0.220$). The discriminant function plot (Figure S7) shows that the first function discriminated the Bcl11a TALEN treatment from all other treatments, while the second function differentiated the remaining treatments from each other.

MiSeq data analysis

For analysis of Bcl11a indels a custom-written computer script was used. Paired-end 150bp sequences were merged and adapter trimmed via SeqPrep (John St. John, https://github.com/jstjohn/SeqPrep, unpublished). Reads were then aligned to the wild-type template sequence.[50] Merged reads were filtered on the following criteria: (i) the 5′ and 3′ ends (23bp) must match the expected amplicon exactly, (ii) the read must not map to a different locus in the target genome as determined by Bowtie2[51] with default settings, and (iii) deletions must be <70% of the amplicon size or <70bp long. Indel events in aligned sequences were defined as previously described[52] except that (i) indels 1bp in length were also considered true indels to avoid undercounting real events, and (ii) indels were required to overlap the expected TALEN binding site gap +1bp on either side in order to be counted as a measure of TALEN-mediated gene disruption.

Mutant tracking bioinformatics

To track Bcl11a mutations over time, a Needleman-Wunsch aligner from the emboss suite (Needle, http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/needle.html) was first used to align the sequence reads to the reference amplicon. The options used with this aligner were: -gapopen 10.0, -gapextend 0.5, and -aformat3 sam. After adding a custom header, files were converted from sam to bam format using picard-tools (http://broadinstitute.github.io/picard/). Next, insertions and deletions were identified, classified, and counted using a custom R-script that relies heavily on software packages available through Bioconductor (https://www.bioconductor.org/). This script extracts the insertion, deletion, and substitution information associated with each read. The sequence is then reconstructed with '-' inserted where deletions occur and insertions recorded in a secondary column. The count for each mutation was tallied and recorded.

Any mutation found in only one read was removed from the analysis. A table containing mutation sequences, read count, and frequency for each mutation was then output for further analysis. In each sequencing run, a control sample consisting of GFP mRNA-electroporated cells from the same animal prior to transplantation determined the average frequency of mutation classes (insertion, deletion, substitution, insertion and substitution, etc.), and was used to perform a one-tailed binomial t-test on each mutation from the corresponding mutation class. Mutations from experimental samples were retained if they had a p-value < 0.05. In this study, we mainly focused on deletions as they were the most frequent mutations found in the TALEN target site. A custom R-script was used to generate the overall frequency of mutants in the sample, and track mutants across multiple samples. This clone-specific analysis pipeline quantified total mutation kinetics at Bcl11a (comparably to our bulk mutation analysis pipeline reinforcing the veracity of both data sets).

Background calculations for mutant tracking data

To empirically determine the background mutation rate at Bcl11a within our mutant tracking data, DNA was extracted from cells of the same animal electroporated with GFP mRNA**.** The estimated background rates of mutation were then applied in a one-tailed binomial test to determine the p-value associated with the likelihood of a mutation originating from an actual mutation event as opposed to arising through sequencing errors or other “noise.” Mutations with a p-value >0.01 were removed from the analysis. The frequency of the remaining mutants was calculated by dividing the count for each mutant by the total read count included in this step of the analysis (before removing low frequency mutants) for that tie point.

SUPPLEMENTAL REFERENCES

48. Munzel, U, and Brunner, E (2000). Nonparametric tests in the unbalanced multivariate one-way design. Biometrical Journal 42: 837-854.

49. Wilcox, R (2005). Introduction to robust estimation and hypothesis testing, Elsevier Academic Press, Burlington, MA, 608 pp.

50. Needleman, SB, and Wunsch, CD (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48: 443-453.

51. Langmead, B, and Salzberg, SL (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359.

52. Gabriel, R, Lombardo, A, Arens, A, Miller, JC, Genovese, P, Kaeppel, C, et al. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nature Biotechnology 29: 816-823.

Table S1. Primers used in this study

Purpose	Name	Sequence	Notes
Construction of donor AAV	Bcl11a-1000BglII-F	GGAGCAGATCTCCGCTGGTGATTATGTGTGC	amplify 5'
	Bcl11a.265-1000R	TGGCGTTACTGCAGCTAGCCAAGAGGCTCGGCTGTGGTTG	homology arm
	265 Bcl11aSFFV-F	CAACCACAGCCGAGCCTCTTGGCTAGCTGCAGTAACGCCA	amplify
	265 Bcl11aGFP-R	CAGGTGAGGAGGTCATGATCCTTACTTGTACAGCTCGTCCATG	SFFV.GFP/Mgmt
	Bcl11a.265+1000b-F	CATGGACGAGCTGTACAAGTAAGTGGGCAGTGCCAGATGAAC	amplify 5'
	Bcl11a+1000SalIbR	GAGCCGTCGACCGAGTAAGCATGTCTGTGCG	homology arm
Surveyor/SacI analysis	Bcl11a-F	CATAGGTGCATGCAGTCGTT	
	Bcl11a-R	AACAATCGTCATCCTCTGGC	
	CD33-F	GCTACTGCTGCCCCTGCTG	
	CD33-R	CTCCCAGTACCAGGGTCCCATC	
Miseq analysis	Miseq Bcl11a-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGATGTGCTTCTCCCCTTTCTGTCC	
	Miseq Bcl11a-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATTGCATTGTTTCCGTTTGTGCTCG	
Targeted integration analysis	Bcl11a pre-1000F	CAGGAAAGAAAGACATGTCTACC	validate
	SFFVmgmt 5'-R	CCATGCCTTGCAAAATGGCG	integration events

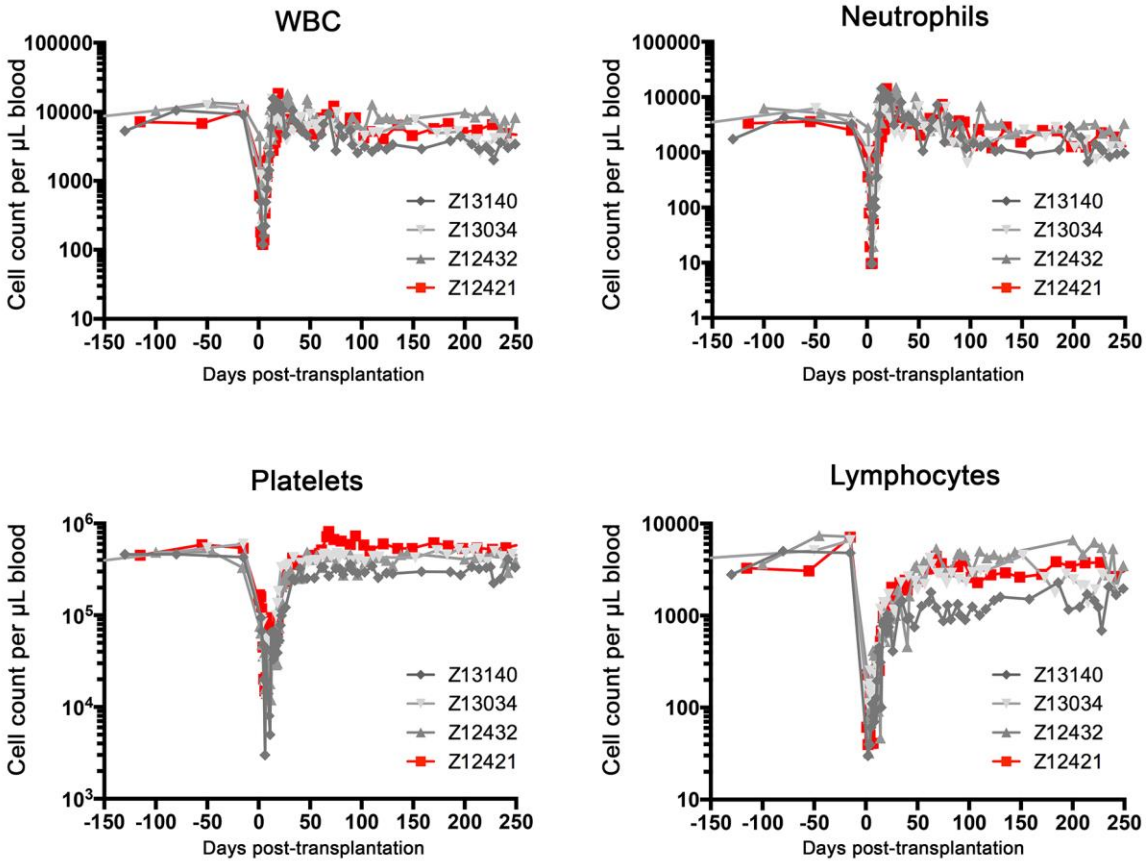


Figure S1. Complete blood cell count in transplanted pigtailed macaques. Day 0 represents time of transplantation.

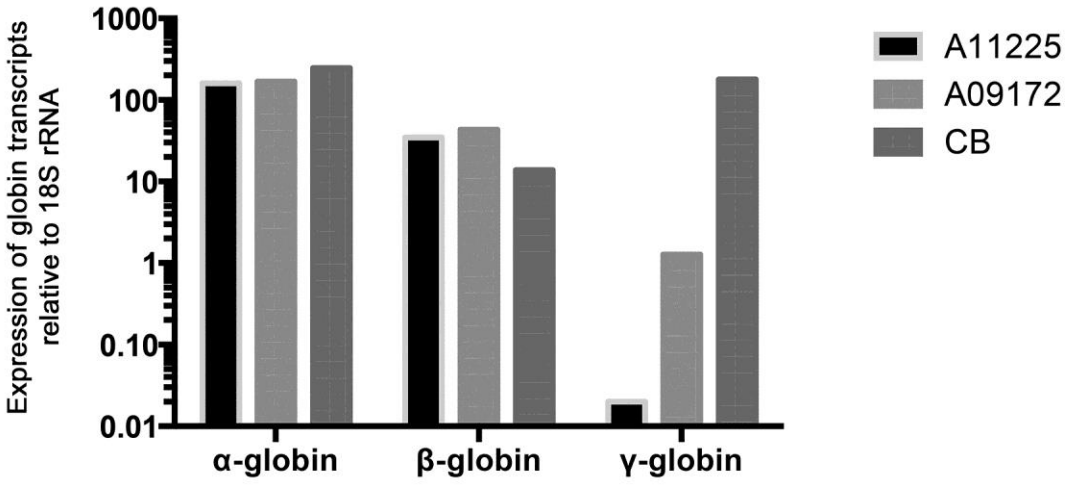


Figure S2. Quantitative PCR analysis of α-, β- and γ-hemoglobin measured from peripheral blood RNA isolated from an untransplanted control (A11225), a pigtailed macaque transplanted with γ-hemoglobin, lentiviral vector-modified HSCs (A09172) or from pigtailed macaque cord blood (CB). In all samples, hemoglobin expression was normalized to 18S ribosomal RNA expression.

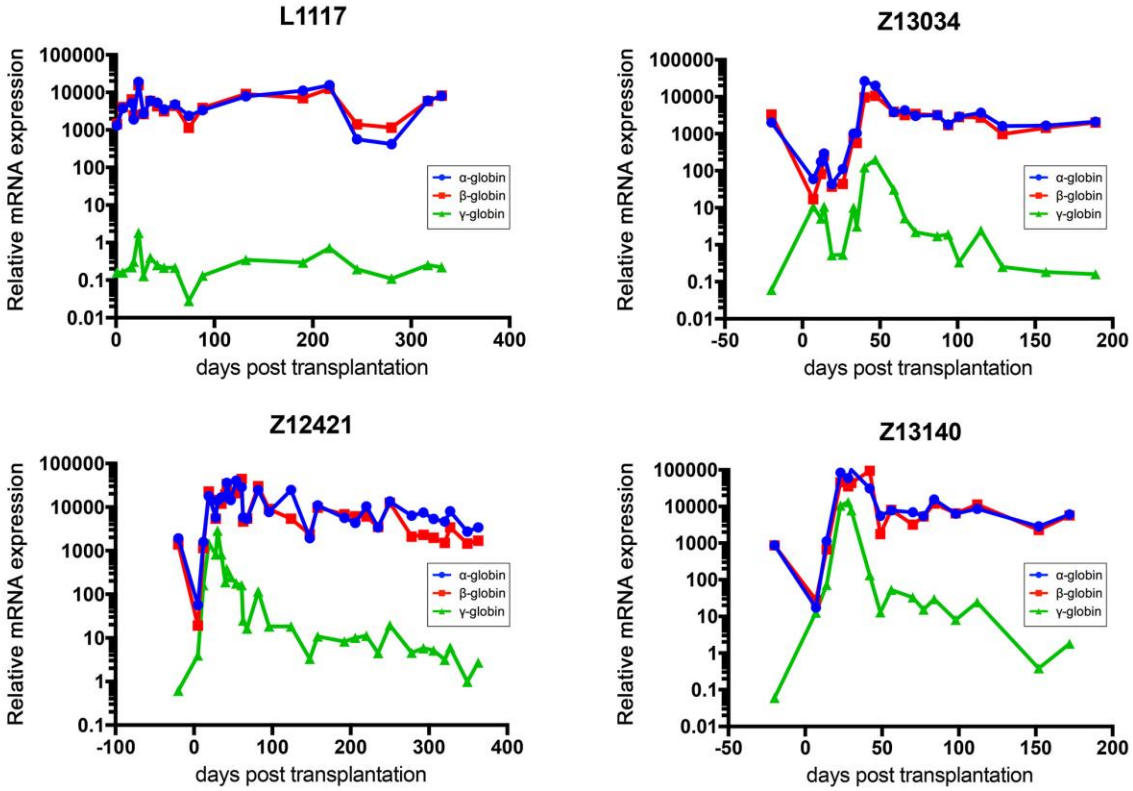


Figure S3. Quantitative PCR analysis of α-, β- and γ-hemoglobin measured in peripheral blood RNA of the indicated animals. For all time points, hemoglobin expression was normalized to 18S ribosomal RNA expression. L1117 is the untransplanted control, Z13034 and Z13140 are control transplants and Z12421 was transplanted with Bcl11a-edited HSCs.

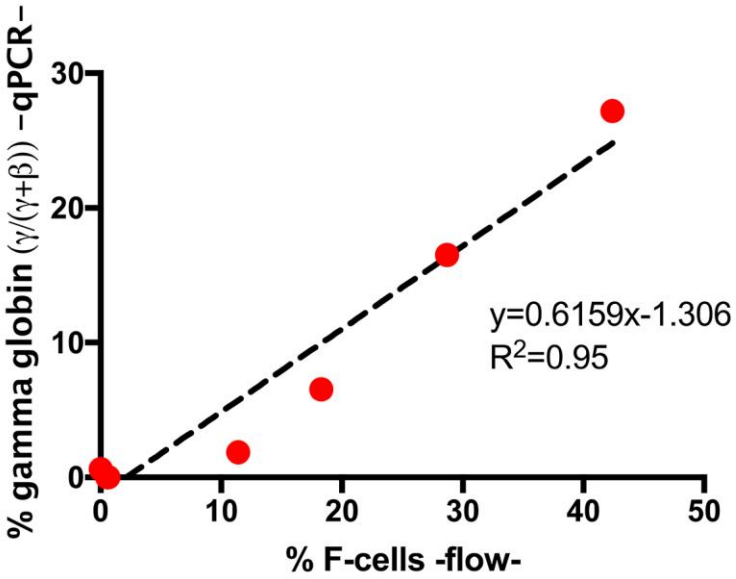


Figure S4. Correlation by linear regression analysis of γ-globin expression (defined as γ/(γ+β)) and F-cell frequencies in peripheral blood of transplanted animals Z12421, Z12432, Z13034 and Z13140 pre-transplantation and at peaked levels post-transplantation.

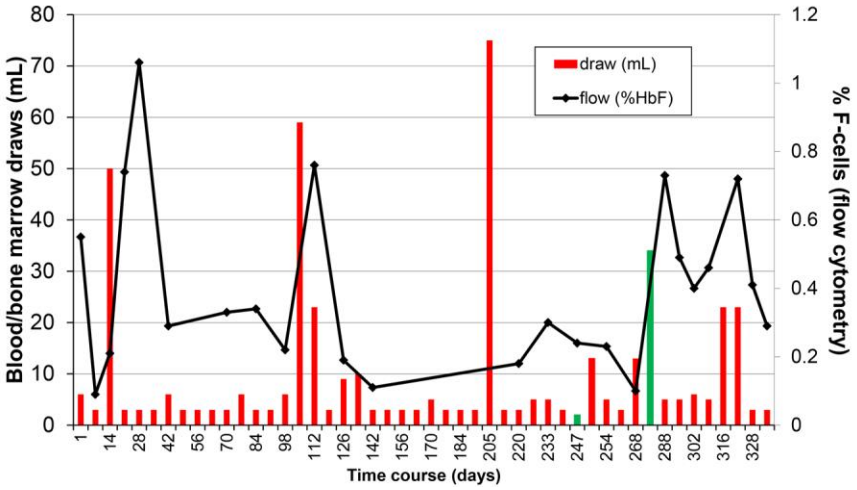


Figure S5. HbF response to repeated blood and bone marrow draws. Blood (red) and bone marrow (green) draw volumes are represented with bar graph (left y-axis) and the frequency of F-cells in peripheral blood is shown with black line (right y-axis).

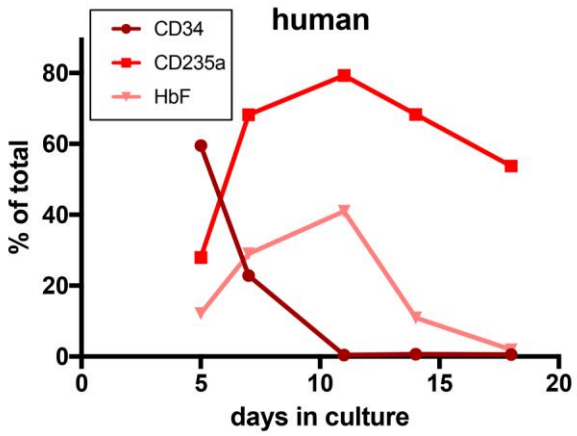


Figure S6. Characterization of in vitro erythroid differentiation of human mobilized peripheral blood CD34+ cells by flow cytometry analysis using the markers described in the legend.

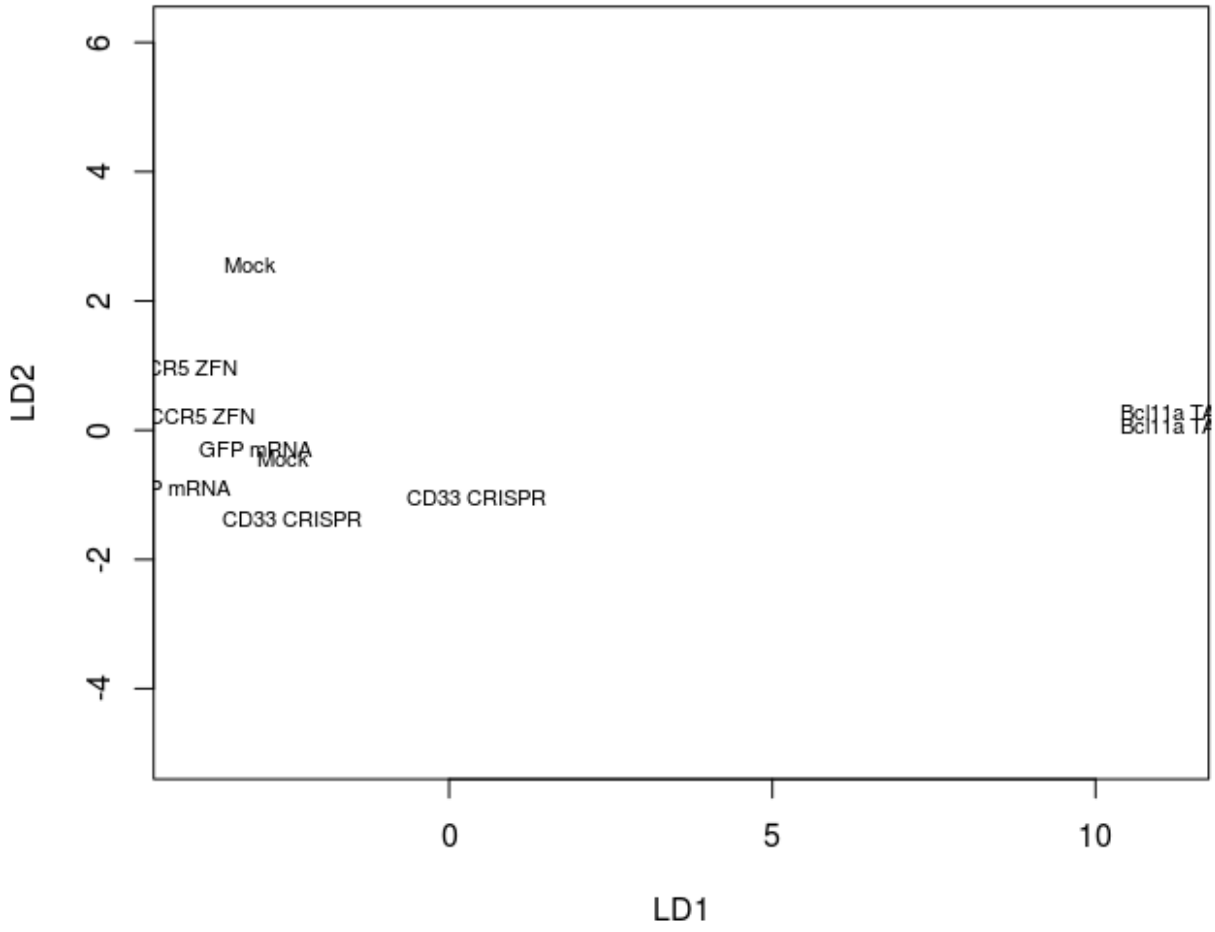


Figure S7. Combined Groups Plot. Plot of the variate scores for each complete pair (% γ-globin and % HbF) of observations labelled according to editing treatment. Variate one (LD1) discriminates the Bcl11a TALEN group from the others while variate two (LD2) separates out the remaining groups.

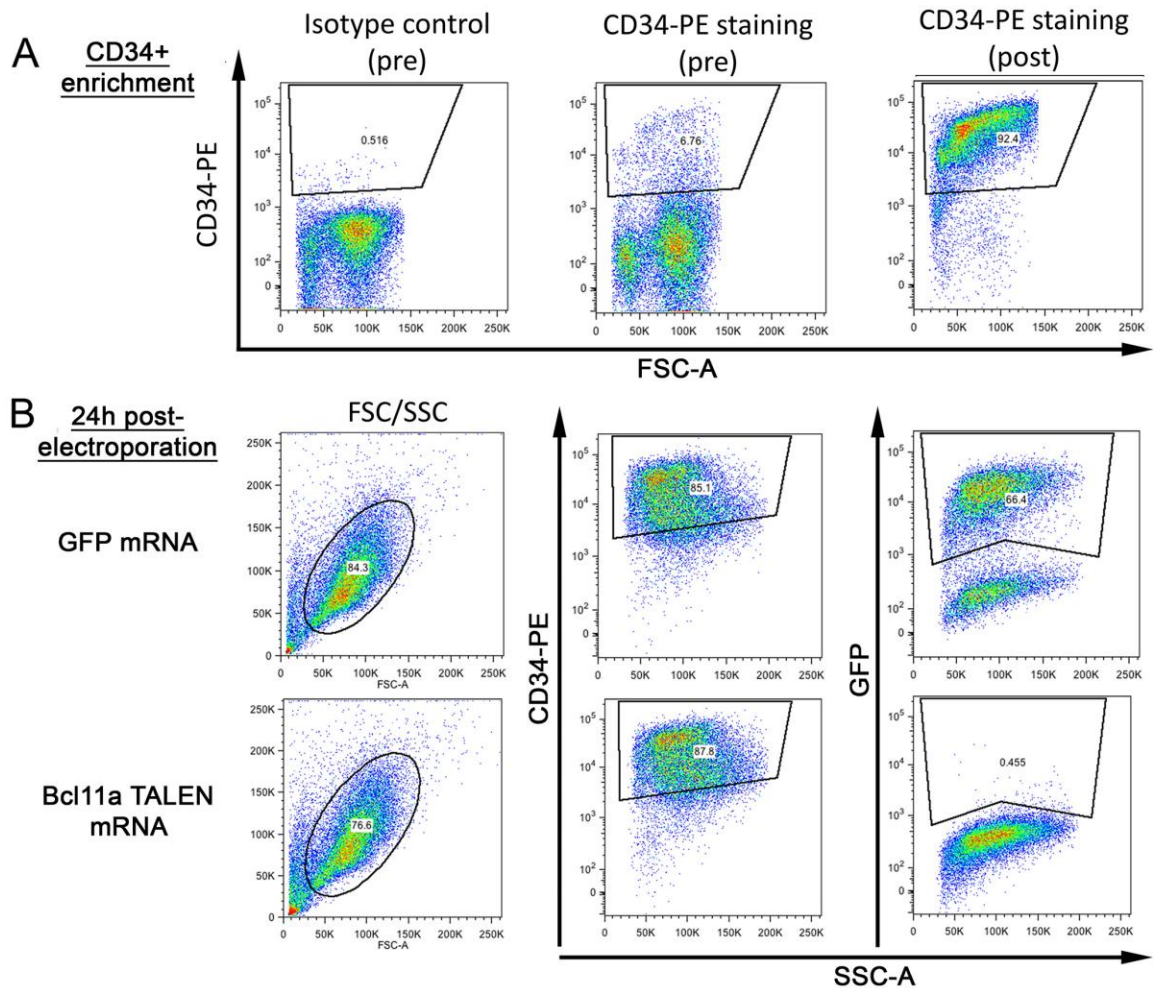
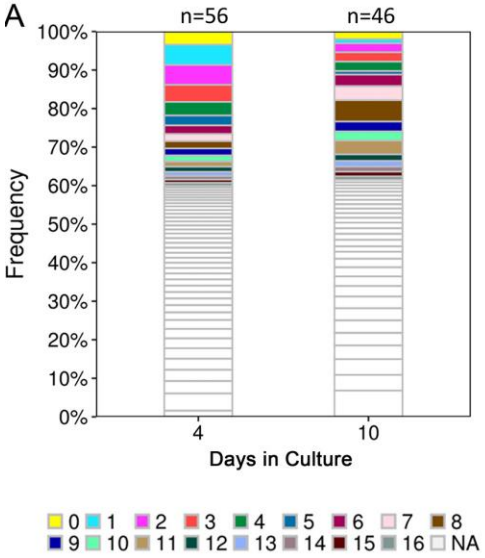


Figure S8. Flow cytometry analysis of CD34+ cells isolated from animal Z12421 before and after electroporation and prior to infusion. **A)** CD34+ enrichment of bone marrow collected from animals Z12421 showing fraction of CD34+ cells before (pre) and after (post) enrichment. **B)** Determination of the fraction of live cells (FCS/SSC), CD34+ purity (CD34PE), and fraction of GFP+ cells (GFP) in Z12421 CD34+ cells electroporated with GFP or Bcl11a TALEN mRNA at 24h post-treatment.

A



B

Color	Deletion Size	Sequence
0	181	----------
1	2	gatgaaccagaccacggcccgttgggag--ccagaaggggatcatgacctcctcac
2	12	gatgaaccagaccacggc----------ccagaaggggatcatgacctcctcac
3	1	gatgaaccagaccacggcccgtt-ggagctccagaaggggatcatgacctcctcac
4	4	gatgaaccagaccacggcccgttgggagc----gaaggggatcatgacctcctcac
5	1	gatgaaccagaccacggcccgttgggagct-cagaaggggatcatgacctcctcac
6	27	gatgaacc----------aggggatcatgacctcctcac
7	2	gatgaaccagaccacggcccgtt--gagctccagaaggggatcatgacctcctcac
8	6	gatgaaccagaccacggcccgttggg------agaaggggatcatgacctcctcac
9	1	gatgaaccagaccacggcccgttgggagctccag-aggggatcatgacctcctcac
10	13	gatgaaccagaccacgg----------ccagaaggggatcatgacctcctcac
11	15	gatgaaccagaccacggcccgtt----------ggatcatgacctcctcac
12	8	gatgaaccagaccacggcccg--------tccagaaggggatcatgacctcctcac
13	1	gatgaaccaga-cacggcccgttgggagctccagaaggggatcatgacctcctcac
14	1	gatgaaccagaccacggcccgttgggagctccagaa-gggatcatgacctcctcac
15	1	gatgaa-cagaccacggcccgttgggagctccagaaggggatcatgacctcctcac
16	1	gatgaaccagaccacggcccgttgggagctccagaaggggatcatgacctcTtc-c

Figure S9. Bcl11a deletion signature in infusion product. **A)** Tracking of deletion signatures in TALEN-treated CD34+ cells from animal Z12421 cultured for 4 and 10 days. Number of unique deletion signatures detected at each time point is shown on top; deletion signatures appearing at both time points are shown as colored boxes while unique deletion signatures are shown in white. **B)** Sequences of deletion signatures from panel (A) with the size of deletion (given in number of nucleotides).

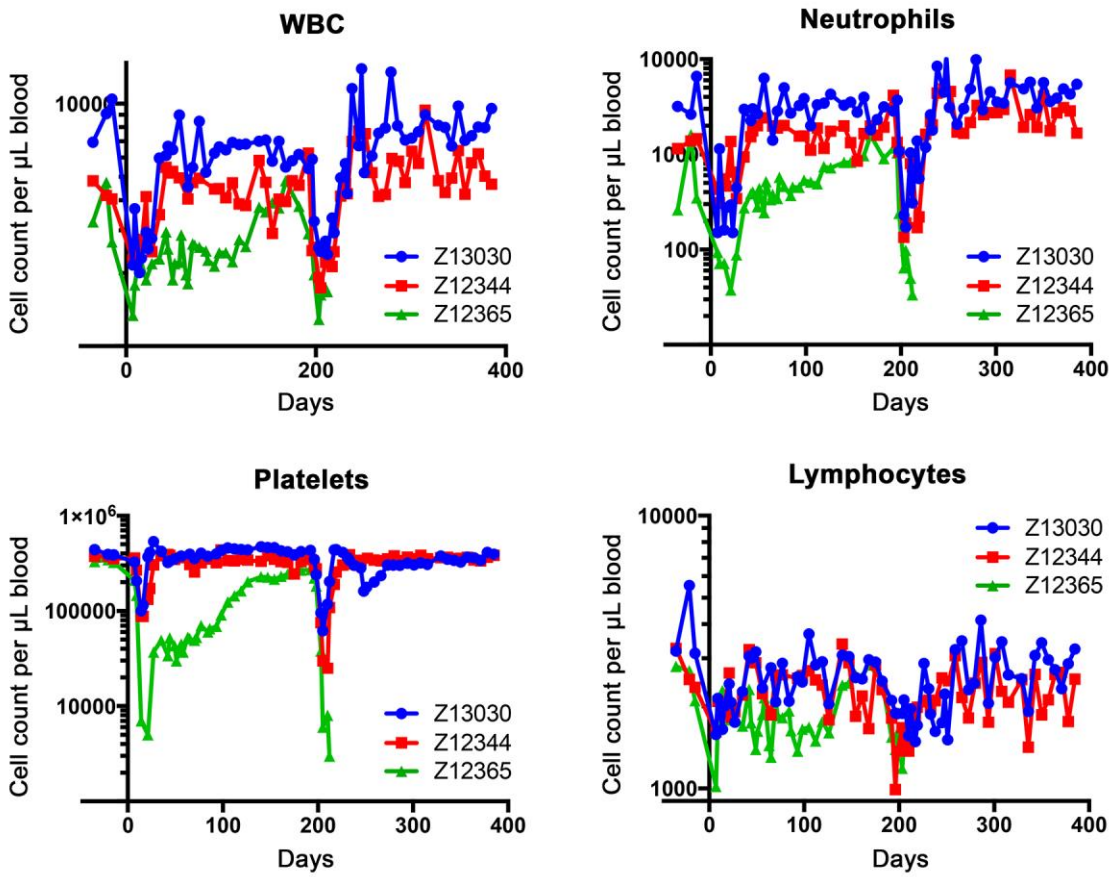


Figure S10. Complete blood count in three transplanted pigtailed macaques treated with 2 rounds of chemoselection with O6BG/BCNU at day 0 and day 183.