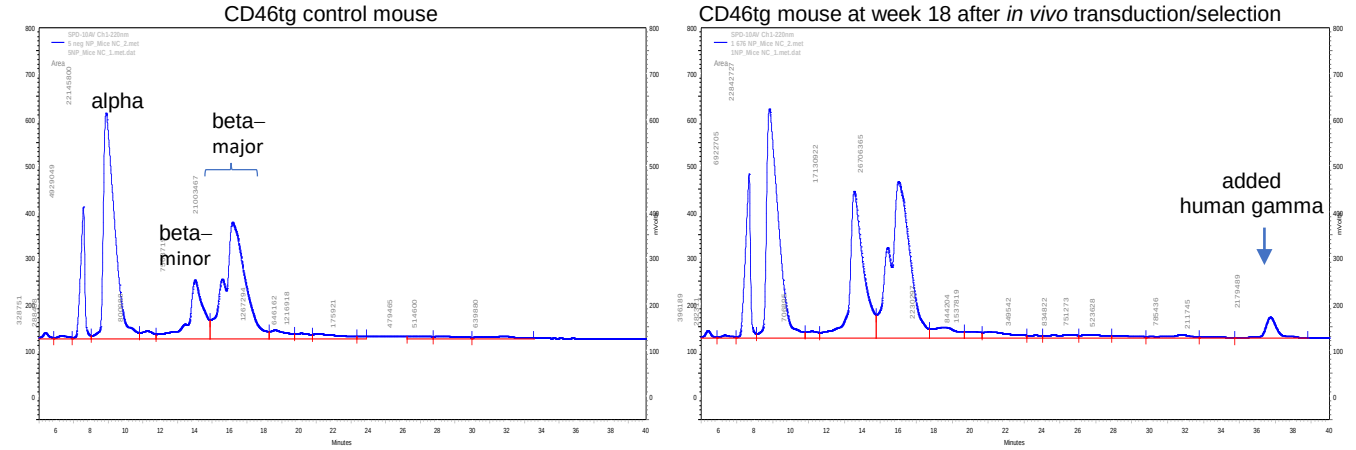


Figure S1. HPLC analysis of globin chains in RBCs from a hCD46tg control mouse and a representative CD46tg mouse after *in vivo* transduction/selection. The numbers (Volts) indicate the peak intensities. A total of 4 mice from each group was analyzed with similar results. The data are summarized in Fig.1E.

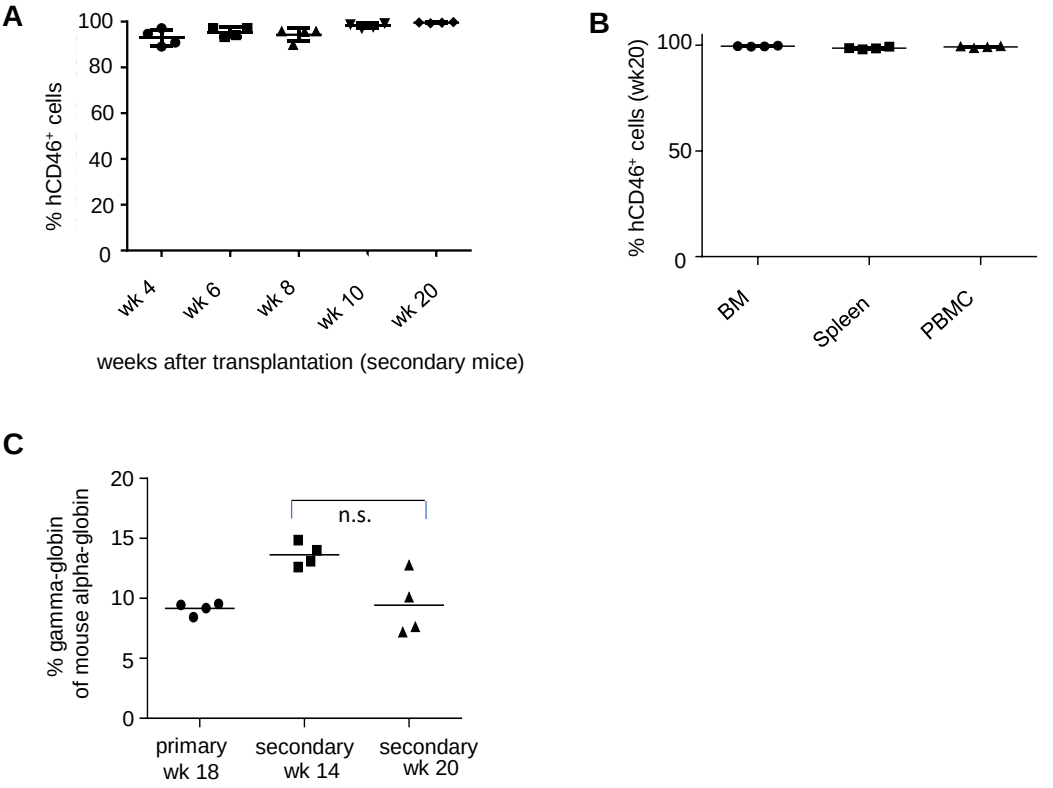
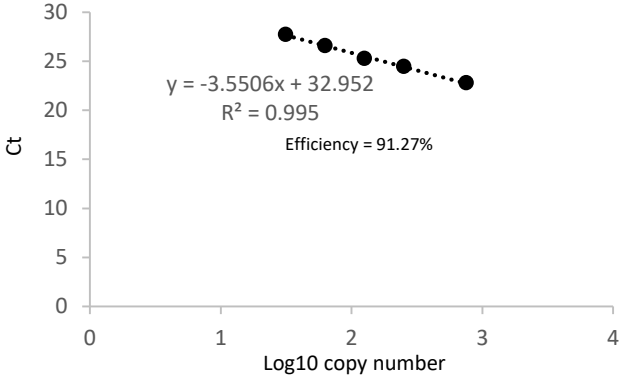


Figure S2. Analysis of mice that received transplantations with bone marrow Lin$^-$ cells harvested at week 18 after in vivo transduction ("secondary recipients"). A) Engraftment measured in blood samples at the indicated time points based on the percentage of human CD46-postive cells in PBMCs. **B)** Engraftment in bone marrow, spleen, and PBMCs at week 20. **C)** Ratio of human gamma- to mouse alpha-globin protein measured by HPLC in RBCs. Each symbol represents an individual animal. Statistical analyses were done with the non-parametric Kruskal-Wallis test.



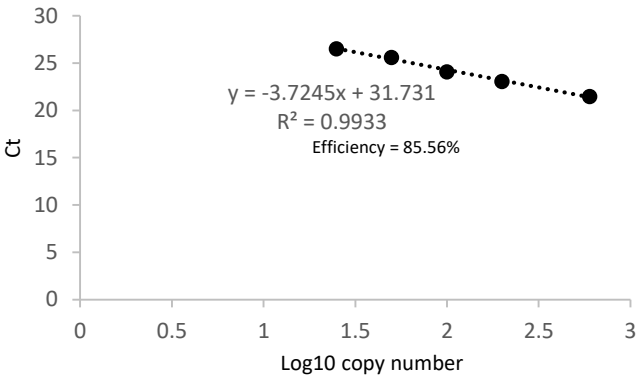


Figure S3. qPCR in single cell-derived progenitor colonies to measure the VCN (see Fig.3E). A) Standard curve of integrated transgene copy number. **B)** Standard curve for mouse GAPDH copy number. n=3 for each VCN.

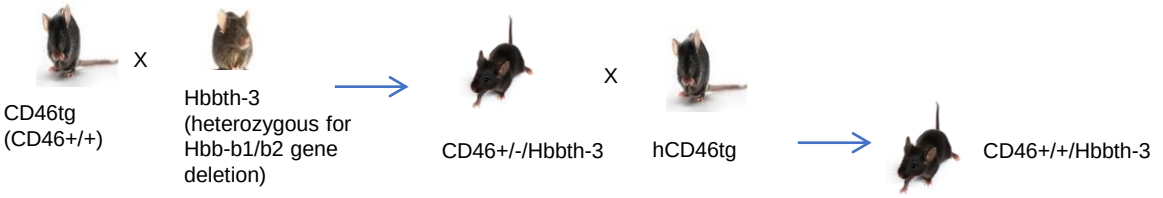


Figure S4. Generation of the CD46+/+/Hbbth-3 thalassemic model. Female CD46tg mice were bred with male Hbbth-3 mice. The F1 hybrid mice were back-crossed with hCD46+/+ mice to generate Hbbth-3 mice homozygous for hCD46+/+.

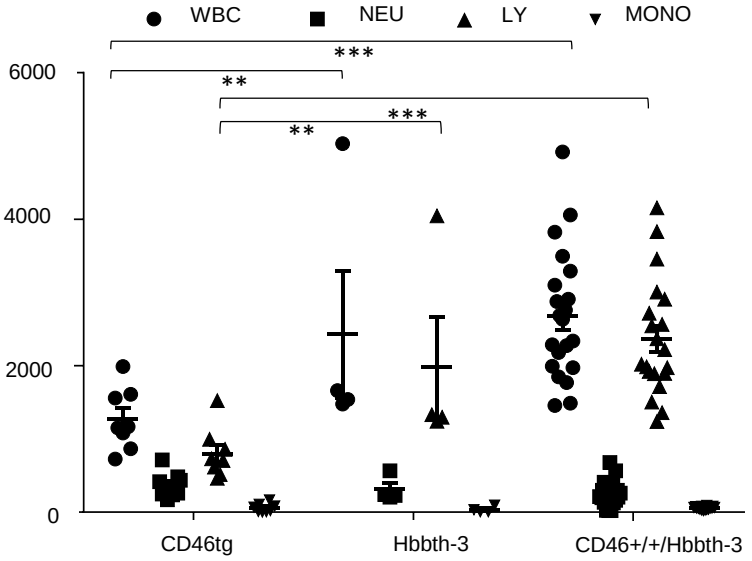


Figure S5. Analysis of white blood cells in thalassemic mice (Hbbth-3 and CD46+/+/Hbbth-3) compared to "healthy" CD46tg mice. WBCs: white blood cells, NEU: neutrophils, LY: lymphocytes, MONO: monocytes. *p≤0.05, ** p≤0.0002, ***p ≤0.00003. These are baseline levels in mice before treatment. (n=8 for CD46tg, n=4 for Hbb^{th3}, n=20 for $CD46^{++}/Hbb^{th3}$). Each symbol represents an individual animal. Statistical analyses were done with the non-parametric Kruskal-Wallis test.

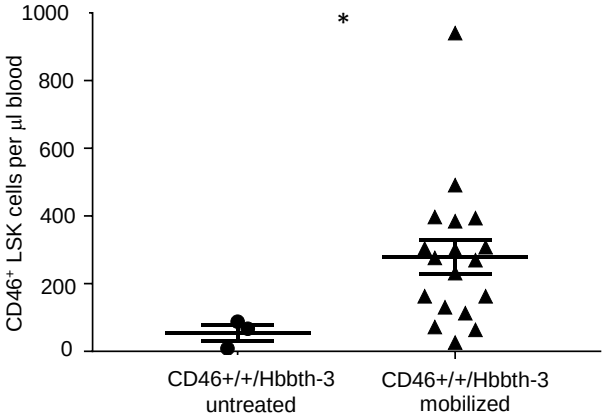


Figure S6. Mobilization of HSPCs in CD46+/+/Hbbth-3 mice. Shown are the numbers of mobilized LSK (Lineage$^-$/Sca-1$^+$/c-Kit$^+$/) cells in peripheral blood at 1 hour after the last AMD3100 injection. n=17 mobilized mice; n=3 untreated mice. Statistical analyses were done with the non-parametric Kruskal-Wallis test.

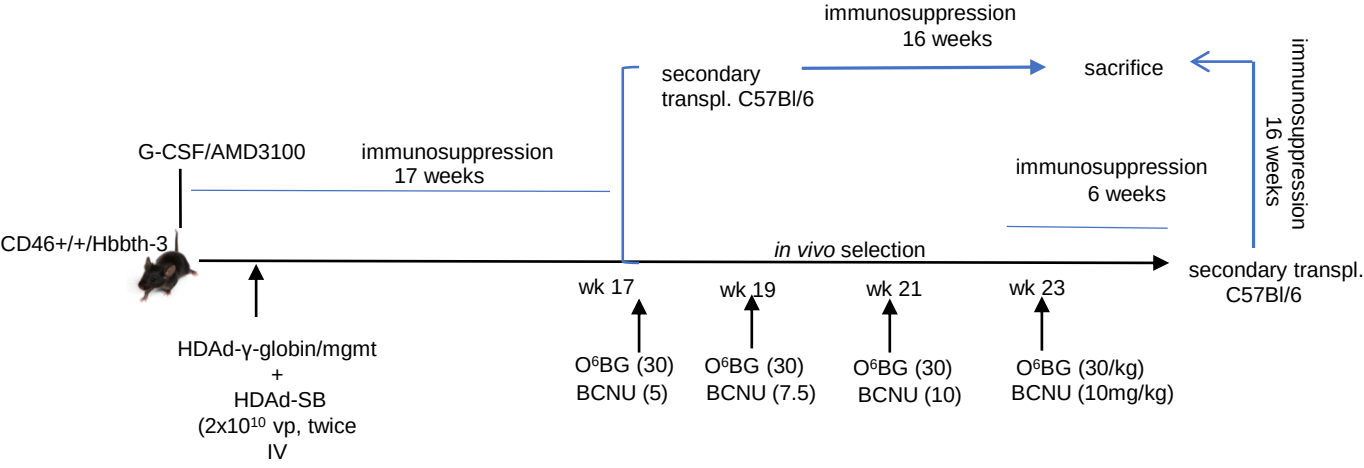


Figure S7. *In vivo* transduction/selection of mobilized CD46+/+/Hbbth-3 mice. *In vivo* transduction of mobilized CD46+/+/Hbbth-3 mice. HSPCs were mobilized by s.c. injections of human recombinant G-CSF for 6 days (days 1–6) followed by three s.c. injections of AMD3100/Plerixafor (days 5–7). 30 and 60 minutes after Plerixafor injection, animals were intravenously injected with a 1:1 mixture of HDAd-γ-globin/mgtm + HDAd-SB (2 injections, each $4x10^{10}$ vp). Following *in vivo* transduction, immuno-suppression was administered for 17 weeks to avoid immune responses against the human gamma-globin and MGMT(P140K) proteins. At week 17, treated mice either served as donors for secondary transplants or were subjected to *in vivo* selection with O^6-BG/BCNU. Secondary C57Bl/6 recipients were followed for 16 weeks under immunosuppression and then sacrificed. Mice subjected to *in vivo* selection received an escalating (5, 7.5, 10, 10mg/kg) O^6-BG/BCNU treatment every other week. Immuno-suppression was resumed two weeks after the last O^6-BG/BCNU dose. At week 29, mice were sacrificed and their bone marrow was transplanted into C57Bl/6 secondary recipients.

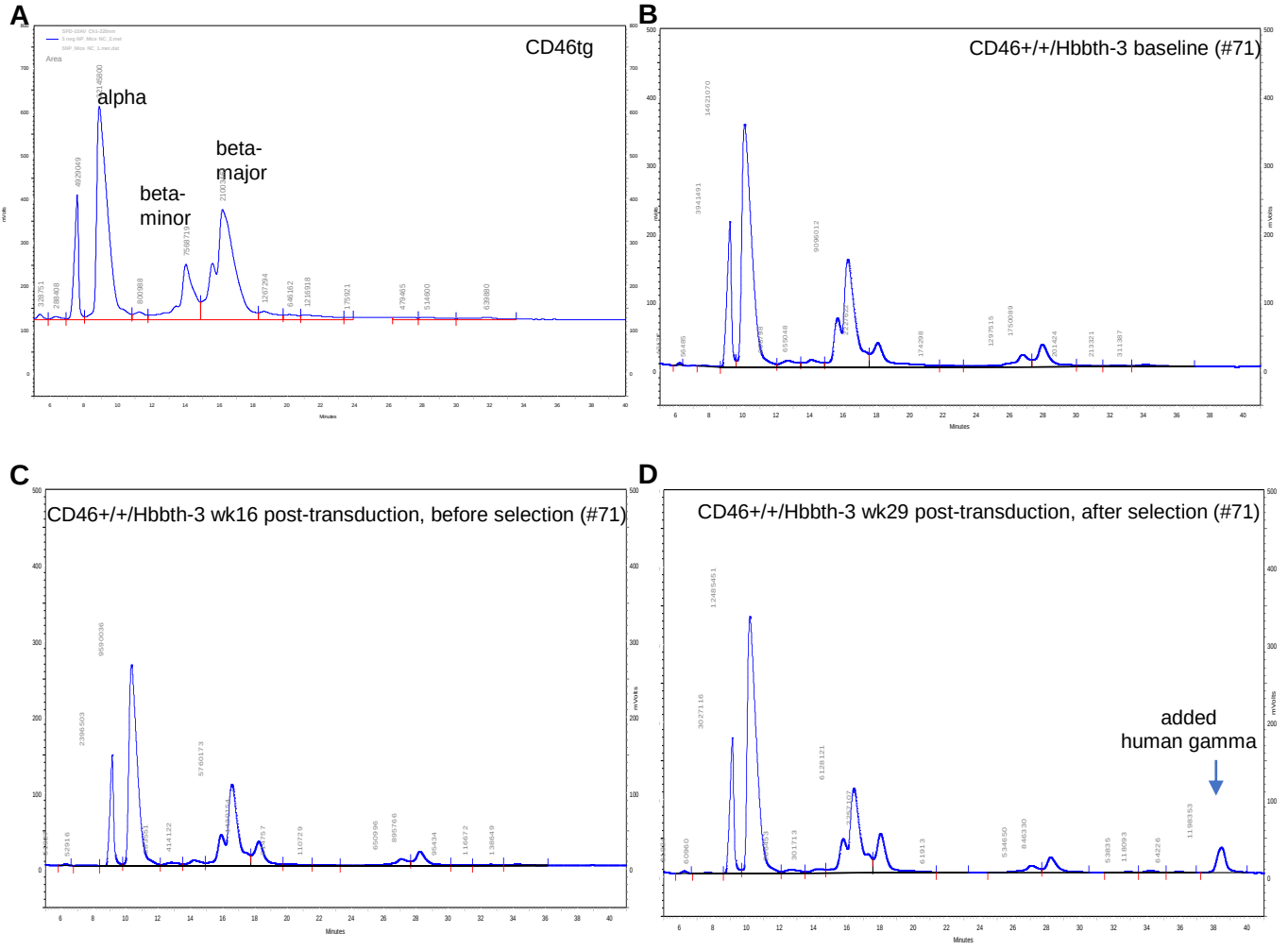


Figure S8. HPLC analysis of globin chains in RBCs. A) Representative chromatograms of mouse globin peaks in a control CD46tg mouse. The peaks for adult mouse alpha, beta-minor, and beta-major globin are labeled. **B-D)** Chromatogram of RBCs from a CD46+/+/Hbbth-3 mice (#71). Note that these mice are heterozygous for beta-minor and beta-major gene deletions. The extra peaks around 29 min could be associated with this. In D), the peak specific to human gamma-globin is labeled. Representative chromatograms are shown. The numbers (Volts) indicate the peak intensities.

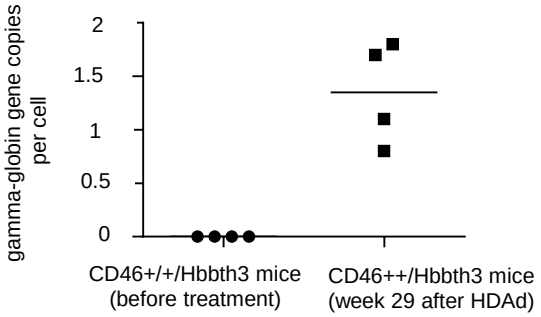


Figure S9. DNA analysis of treated CD46++/Hbbth-3 mice at week 29. Transgene (gamma-globin) copy number per bone marrow cell. Each symbol represents an individual animal.

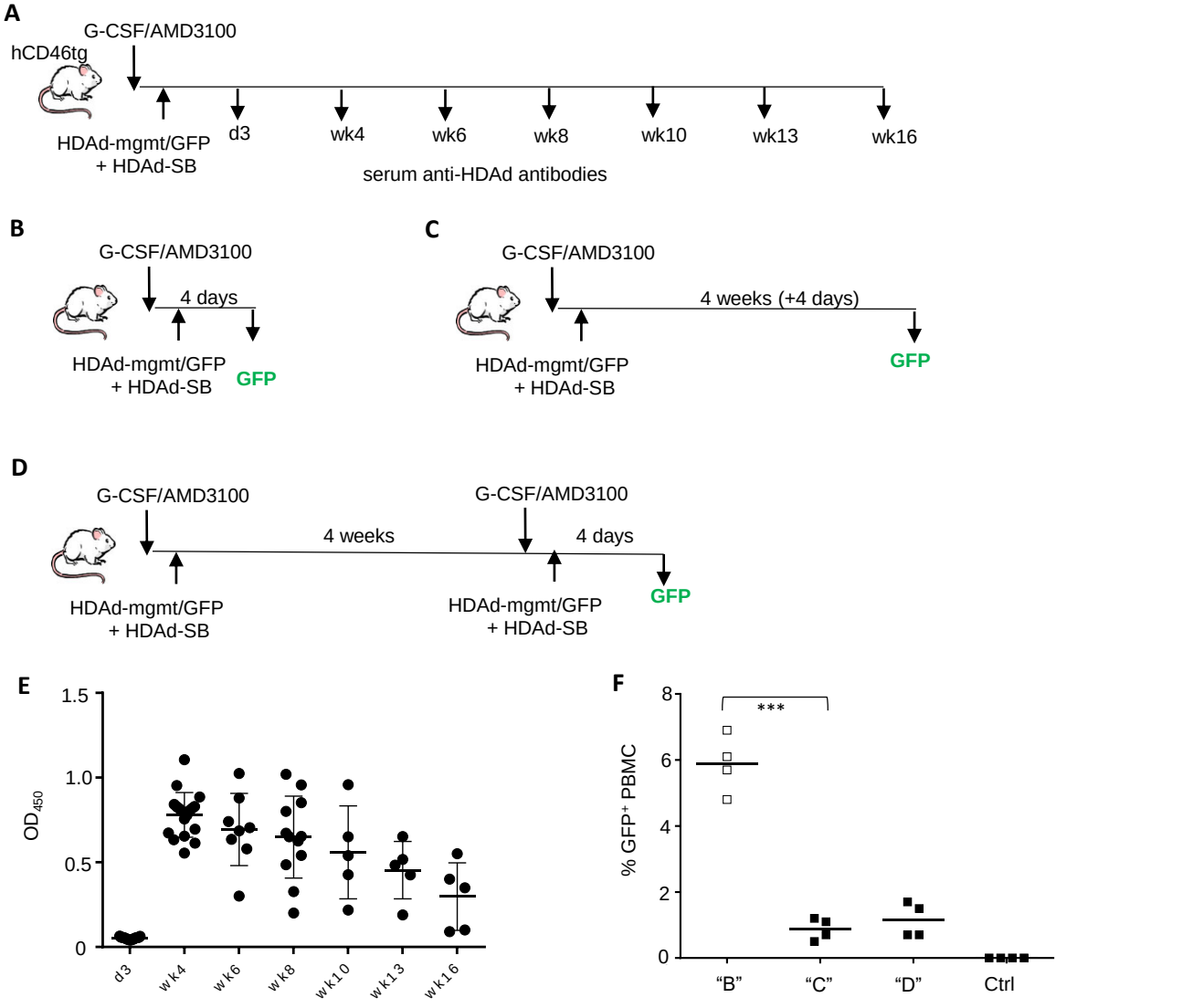


Figure S10. Effect of anti-HDAd5/35++ antibodies on a second round of transduction. A) CD46tg mice were mobilized and injected with HDAd-mgmt/GFP + HDAd-SB. Serum samples were collected as indicated. **B and C)** Flow cytometry analysis of PBMCs at day 4 and week 4 after mobilization/transduction. **D)** Second round of mobilization/transduction at week 4 and subsequent GFP analysis. **E)** anti-HDAd5/35++ antibody titers based on OD_{450}. An OD_{450}=0.2 titer is considered to be neutralizing. **F)** Percentage of GFP-positive PBMCs measured in different cohorts (see B-D). Ctrl are untreated CD46tg mice. Each symbol in E) and F) represents an individual animal. Statistical analyses were done with the non-parametric Kruskal-Wallis test.

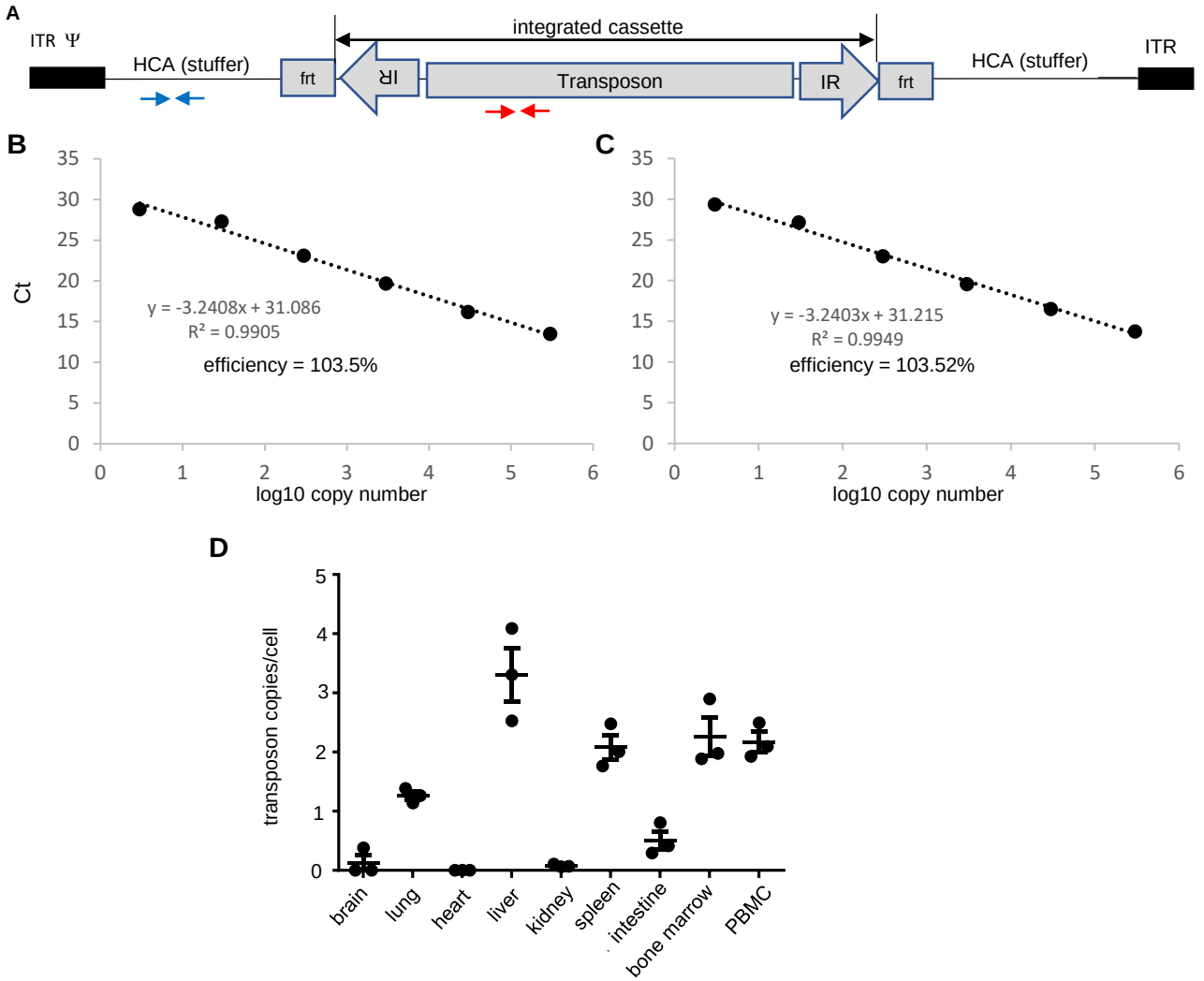


Figure S11. Vector DNA biodistribution at week 18 after HDAd injection (10 weeks in vivo selection) A) Primer design. The red primers are specific to the transgene cassette and will detect both integrated and episomal vector DNA. The blue primers will detect vector stuffer DNA derived from plasmid pHCA. Upon SB100x-mediated integration, the corresponding target region for the blue primers will be lost. The blue primers are therefore used to measure episomal vector copies. **B)** Standard curve of integrated transgene copy number. **C)** Standard curve for HCA (episomal vector) copy number. **D)** Integrated transgene copy number per cell. Episomal vector copies (blue primers) were subtracted from total vector copies (red primers). The vector-specific signals were normalized to *GAPDH*. Each symbol represents an individual animal.

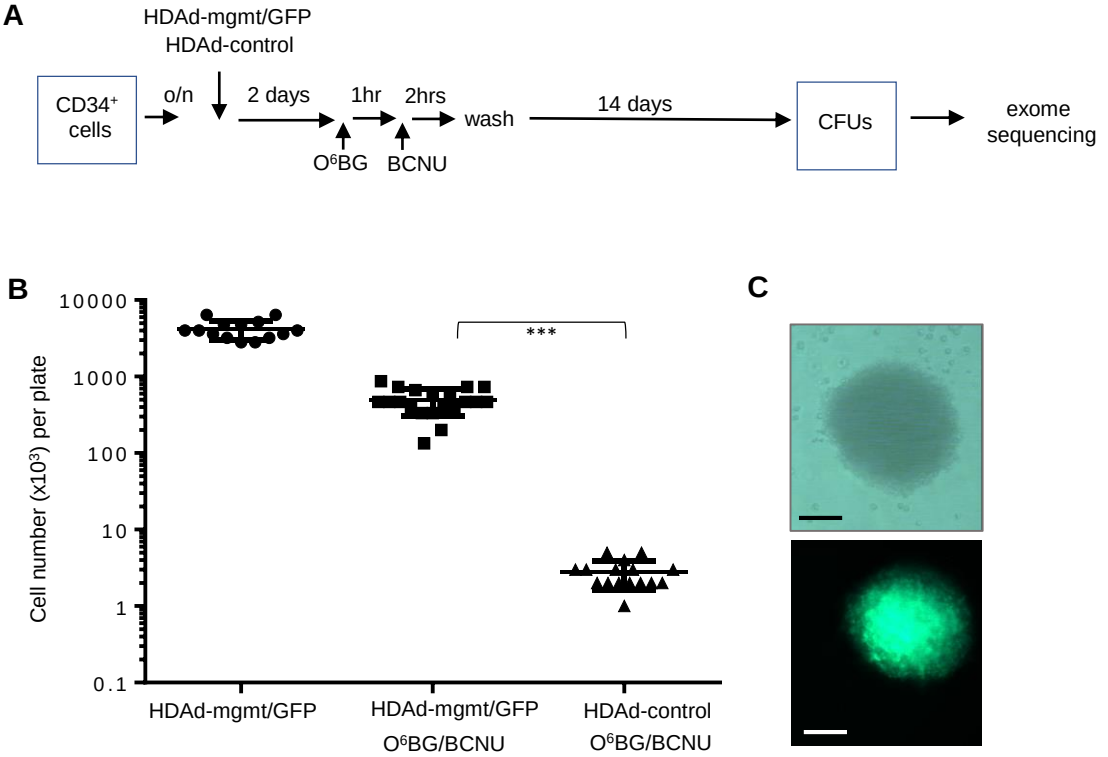


Figure S12. In vitro assay to assess the mutagenicity of O⁶BG/BCNU treatment. A) After overnight recovery from cryopreservation, CD34⁺ cells were transduced with HDAd-mgmt/GFP or HDAd control at an MOI of 3000 vp/cell which mediated GFP expression in ~50% of cells two days later. Cells were then treated with 10μM O⁶BG followed by 25μM BCNU (or DMSO solvent) for 2 hours. After washing, cells were plated in methylcellulose for CFU assay (3000 cells per 35mm dish). Colonies and pooled cells were counted 14 days later and genomic DNA subjected to whole exome sequencing.. **B)** Numbers of pooled cells per plate. Each symbol represents the cell number in an individual 35mm dish. Statistical analyses were done with the non-parametric Kruskal-Wallis test. **C)** Representative colony from the HDAd-mgmt/GFP + O⁶BG/BCNU group. It demonstrates GFP expression in the majority of cells with GFP fading at the colony periphery due to the loss of episomal viral genomes. The scale bar is 1mm.

A

Sample #1: untreated CD34⁺ cells

Total Aligned reads	Percent Aligned Reads	Targeted Aligned Reads	Read Enrichment	Padded Target Aligned Reads	Padded Read Enrichment
46,870,836	80.51%	38,437,631	82.01%	40,158,769	85.68%

Total Aligned Bases	Percent Aligned Bases	Targeted Aligned Bases	Base Enrichment	Padded Target Aligned Bases	Padded Base Enrichment
6,544,191,633	75.45%	4,308,625,487	65.64%	5,541,019,474	84.67%

Sample #2: selected CD34⁺ cells

Total Aligned reads	Percent Aligned Reads	Targeted Aligned Reads	Read Enrichment	Padded Target Aligned Reads	Padded Read Enrichment
47,858,908	81.07%	39,945,698	81.38%	40,463,838	84.55%

Total Aligned Bases	Percent Aligned Bases	Targeted Aligned Bases	Base Enrichment	Padded Target Aligned Bases	Padded Base Enrichment
6,590,512,869	74.93%	4,339,416,710	65.84%	5,523,089,486	83.80%

Figure S13. Whole exome sequencing of CD34⁺ cells that survived drug treatment vs untreated CD34⁺ cells. Sample sequences were compared to a Homo sapiens reference genome (UCSC hg19) **A)** Sample sequencing information. **B)** Enrichment sequencing report shown in a pdf (Fig.S13B_pdf). **C)** List of de novo mutations in selected CD34⁺ cells shown in an excel file (Fig.S13_xlx). The file contains chromosome number, genome position, reference allele, variant allele, variant type, sequence context, consequence, dbSNP ID, COSMIC ID, ClinVar, Gene ID, variant quality, variant frequency, total depth, reference allele depth, variant allele depth, and strand bias.

Suppl. Methods

Flow cytometry: Cells were resuspended at $1x10^6$ cells/100 µL in PBS supplemented with 1 % FCS and incubated with FcR blocking reagent (Miltenyi Biotech, Auburn CA) for ten minutes on ice. Next the staining antibody solution was added in 100 µL per 10^6 cells and incubated on ice for 30 minutes in the dark. After incubation, cells were washed once in FACS buffer (PBS, 1%FBS). For secondary staining the staining step was repeated with a secondary staining solution. After the wash, cells were resuspended in FACS buffer and analyzed using a LSRII flow cytometer (BD Biosciences, San Jose, CA). Debris was excluded using a forward scatter-area and sideward scatter-area gate. Single cells were then gated using a forward scatter-height and forward scatter-width gate. Flow cytometry data were then analyzed using FlowJo (version 10.0.8, FlowJo, LLC). The following antibodies were used: anti-mouse Ter-119-APC (clone: Ter-119, Biolegend, San Diego, cat# 116212), biotin-conjugated lineage detection cocktail (Miltenyi Biotec, San Diego, cat#130-092-613); anti-mouse LY-6A/E (Sca-1)-PE-Cyanine7 (clone D7, eBioscience, San Diego, cat# 25-5981-82); anti-mouse CD117 (c-Kit)-PE (Clone 2B8, eBioscience, San Diego, cat# 12-1171-83); anti-mouse CD3-APC (clone 17A2, Invitrogen, Waltham, MA, cat# 17-0032-82); anti-mouse CD19-PE-Cyanine7 (clone eBio1D3, eBioscience, San Diego, cat# 25-0193-82); anti-mouse Ly-66 (Gr-1)-PE, (clone RB6-8C5, eBioscience, San Diego, CA, cat #12-5931-82); anti-human CD46-APC (clone E4.3, BD Pharmingen, San Diego, CA, cat# 564253).

Colony forming unit assay. 1500 of lineage-negative, GFP-positive cells were plated in triplicates in ColonyGEL 1202 mouse complete medium (ReachBio, Seattle WA) or $5x10^4$ bone marrow cells were plated in duplicate, in methylcellulose medium with recombinant cytokines (Methocult M3434, Stem Cell Technologies Inc., Vancouver, CA). Cells were incubated for 7-12 days at 37°C in 5% CO_2. Colonies were enumerated using a Leica MS 5 dissection microscope (Leica Microsystems).

Integration site analysis. Amplification of genomic DNA junctions was performed by linear amplification-mediated PCR and bioinformatic analysis of integration sites was performed as described previously (1).
To analyze the sequencing data, sample specific barcoded sequencing reads were demultiplexed using CASAVA, an Illumina software package. The quality of sequencing runs of resulting fastq files was evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads starting with the barcode 5'-GTATGTAAACTTCCGACTTCAA-3' that follows the TA dinucleotide, which is characteristic of SB integration, were aligned against the latest version of mouse reference genome (GRCm38/mm10 (Dec, 2011)), using Bowtie2 (2). Only reads that mapped exactly to a unique position in the reference genome were kept for further analysis, with the threshold of at least 3 uniquely mapped reads per locus. To analyze the distribution of integrations, annotations of exons and CDS of the corresponding reference genome were downloaded and the percentage of integration sites overlapping with the given genomic coordinates were analyzed using BEDTools (3). Chromosomal distributions of integration sites were visualized on an ideogram using the NCBI Genome Decoration Page (http://www.ncbi.nlm.nih.gov/genome/tools/gdp/). We have indexed and created normal, shuffled and randomized of mouse genome and counted the number of integrations for each window and plotted the density.

Anti-adenovirus serum IgG antibody titers: The assay was performed as described elsewhere (4). In brief, ELISA plates were coated with $5x10^8$ vp of HDAd-γ-globin/mgmt in PBS at 4°C overnight. Mouse serum samples were diluted 1:800 in blocking buffer (StartingBlock, Thermo Scientific, Rockford, IL). Binding was detected using goat anti-mouse immunoglobulin G (IgG)-horseradish peroxidase (BD Pharmingen, San Jose, CA, cat# 554002) and 1-Step™ Ultra TMB-ELISA (Thermo Scientific, Rockford, IL). The OD450 was measured.

Vector genome biodistribution: Total DNA from tissues was extracted by Quick-DNA miniprep kit (Zymo Research). Viral DNA extracted from HDAd-GFP/mgmt or HDAd-globin/mgmt virus were serially diluted and served as standard curve. qPCR was conducted with 9.6 ng DNA in duplicate in a 10 µl reaction using the power SYBR Green PCR master mix on a StepOnePlus real-time PCR system (Applied Biosystems). GFP (forward 5'- TCGTGACCACCCTGACCTAC-3'; reverse 5'- GGTCTTGTAGTTGCCGTCGT-3') and human gamma-globin (forward 5'- GTGCTTGAAGGGGAACAACTAC-3'; reverse 5'- CCTGGCCTCCAGATAACTACAC-3') primers were used to measure integrated vector copy number. Primers that bind to the HDAd vector DNA but outside of the transposon (forward 5'- GGAAGTAGGTGCTGCCTGAG-3', reverse 5'- CCCTTTGCCTGTGTGATTTT-3') were used to measure episomal vector copy number (see Figure S10A). Mouse GAPDH primers were used to normalize loading (forward 5'- TTCCATCCTCCAGAAACCAG-3; reverse 5'- GTTCTTCTCGGGCAAAAATG-3').

Vector copy number per single cell/colony): Lineage-negative bone marrow cells were isolated from total mouse bone marrow cells by MACS using the Lineage Cell Depletion kit from Miltenyi Biotech. 750 cells/dish were plated in ColonyGEL™ 1202 Mouse Complete Medium (Reachbio, Seattle), 15 days later, well-isolated colonies were aspirated carefully with a pipette tip and washed with of phosphate-buffered saline. Cell pellet were incubated with 10 µl proteinase K (ThermoFisher) in lysis buffer (50mM KCl, 50mM Tris-HCl (pH8.0), 2.5mM EDTA, 0.45% NP-40, 0.45% Tween-20) at 55°C overnight, followed by 10 min at 95 °C. Samples were diluted to 100-200 µl and 4.8 µl DNA was used in a 10 µl reaction. Transgene-specific primers (see "Vector Biodistribution") were used to measure integrated vector copy number. Mouse *GAPDH* (see above primer sequence) was used as a single copy control gene (2 copies/cell). Transgene-specific signals were normalized to *GAPDH* (taken as 2 copies per cell, which allowed us to calculate the cell number in the given DNA preparation). Mouse bone marrow DNA with a VCN =2.5 was serially diluted and served as standard curve.

$CD34^+$ cell culture: $CD34^+$ cells from G-CSF-mobilized adult donors were obtained from the FHCRC Hematopoietic Cell Procurement and Processing Services. Cells were recovered from frozen stocks and incubated overnight in StemSpan H3000 (STEMCELL Technologies, Vancouver, Canada) with penicillin/streptomycin, Flt3 ligand (Flt3L, 25 ng/ml), interleukin 3 (10 ng/ml), thrombopoietin (TPO) (2 ng/ml), and stem cell factor (SCF) (25 ng/ml). Cytokines and growth factors were from Peprotech (Rocky Hill, NJ). (These cytokine/growth factor concentrations are 50% of those used for pre-stimulation of cells

in the context of lentivirus transduction.) CD34$^+$ cells were transduced with HDAd5/35++ vectors in low attachment 12 well plates at MOIs indicated in the figure legends.

Exome sequencing Genomic DNA was isolated from CD34$^+$ cells using the Qiagen DNAeasy Blood & Tissue Kit. Exome sequencing was performed by Omega Bioservices (Norcross, GA) using the Illumina Nextera Rapid Capture Exome kit with 100x reads per sample; read format 2x150bp; sequencing platform 4000/X Ten. The following bioinformatics software was used: Enrichment (BaseSpace Workflow) version 3.0.0.; Isas analysis software version 2.10.12; SAMtools version 1.2; Isaac aligner softe ware version iSAAC-03.16.02.20; Picard (calculate HS metrics) version 1.130; Starling Variant Caller version 2.4.6; Manta (SV caller) version 0.28.0; Illumina Annotation Engine version 1.3.5.2; Pluggable Universal Metrics Analyzer (PUMA) version 00.15.11.02; PUMA Metrics version 1.013.0.

1. Richter M, Saydaminova K, Yumul R, Krishnan R, Liu J, Nagy EE, Singh M, Izsvak Z, Cattaneo R, Uckert W, et al. In vivo transduction of primitive mobilized hematopoietic stem cells after intravenous injection of integrating adenovirus vectors. *Blood.* 2016;128(18):2206-17.
2. Langmead B, and Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357-9.
3. Quinlan AR, and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841-2.
4. Seregin SS, Appledorn DM, McBride AJ, Schuldt NJ, Aldhamen YA, Voss T, Wei J, Bujold M, Nance W, Godbehere S, et al. Transient pretreatment with glucocorticoid ablates innate toxicity of systemically delivered adenoviral vectors without reducing efficacy. *Mol Ther.* 2009;17(4):685-96.