



Supplementary Information for

Stem Cell-Derived Clade F AAV Mediate High Efficiency Homologous Recombination-Based Genome Editing

Laura J Smith, Jason Wright, Gabriella Clark, Taihra Ul-Hasan, Xiangyang Jin, Abigail Fong, Manasa Chandra, Thia St Martin, Hillard Rubin, David Knowlton, Jeff L. Ellsworth, Yuman Fong, Kamehameha K Wong Jr, Saswati Chatterjee*

*Corresponding author: Saswati Chatterjee, PhD
e-mail: schatterjee@coh.org

This PDF file includes:

Supplementary text: Materials and Methods
Figs. S1 to S11
Tables S1 to S8
References for SI reference citations

SI MATERIALS AND METHODS

AAV Vectors. All editing vectors were cloned into the pSaiLuc AAV2 vector backbone (1-3), using New England Biolabs (NEB) Gibson Assembly Cloning Kit, with primers designed using NEBuilder v.1.6.2 or by restriction cloning. The nucleotide substitution vector (AAVS1 NS) was created by site-directed mutagenesis of the Nhe1 site within the PPP1R12C homology arms of an AAV vector plasmid, to SphI using the QuikChange Mutagenesis Kit (Agilent Technologies). IL2RG-GFP and Rosa26-luciferase editing vectors were cloned by inserting synthesized sequences into the pSaiLuc backbone, which contains AAV2 ITRs. The details of the AAV editing vector genomes are shown in Table S1. The integrity of ITRs of every vector was confirmed by ITR sequencing (4). Editing vectors were packaged in different AAV capsids in 293 cells using HSV1 as a helper virus, as previously described (2, 5-7). Vector titers were determined by quantitative PCR (qPCR) using the following oligonucleotides:

qGFP Probe - FAM-AGCTACCAGTCCGCCCTGAGCAAAGA-TAMRA

qGFP Forward Primer – CTGCTGCCGACAACCA

qGFP Reverse Primer – TGTGATCGCGCTTCTCGTT

AAVS1 qHA-L Probe - FAM-TCGGCGCTGCACCA-TAMRA

AAVS1 qHA-L Reverse Primer – GATCCGCTCAGAGGACATCAC

AAVS1 qHA-L Forward Primer – CTCTTCCGGAGCACTTCCTT

Luc Probe - FAM-TTGCCTTCACTGATGCTCACATTGAGGT-TAMRA

Luc Forward Primer – CTCAAAGTATTCAAGCATAGGTGATGTC

Luc Reverse Primer – AACTGCACAAGGCCATGAAGA

Cells, Transductions and Flow Analysis. The human erythroleukemia cell line, K562 (ATCC® Number: CCL-243), the human hepatocellular carcinoma cell line, HepG2 (ATCC® HB-8065), and the fetal lung cell line, Wi-38 (ATCC® CCL-75) were obtained from American Type Culture Collection (ATCC) and cultured according to ATCC guidelines. All ATCC cell lines were authenticated by vendor using short tandem repeat profiling. Human hepatic sinusoidal endothelial cells (HSECs) were obtained from Creative Bioarray and cultured on Gelatin-coated plates using media containing SuperCult® Endothelial Cell Medium Supplement Kit (both from Creative Bioarray). Human skeletal muscle myoblasts (Lonza) were cultured in media containing SkGM™-2 Bullit™ (Lonza). Cytokine-primed peripheral blood CD34+ cells (PBSCs) were obtained with informed consent from healthy donors under a City of Hope IRB approved protocol as previously described (3). Cord blood (CB) CD34+ cells were obtained from ReachBio Research Labs and CB stem cells were cultured in low cytokines as previously described (3, 6). Mutant B lymphoblastoid and fibroblast cell lines (Table S6) were obtained from the Coriell Institute for Medical Research (Camden, NJ), and cultured according to their guidelines. Transductions were performed at one to three replicates per experiment using 1E4 to 1E5 cells at multiplicity of infection (MOI) ranging from 1E4 to 4E5 with AAVHSC editing vector averaging 4.65E11vgs/ml. Cells were harvested (adherent cells were trypsinized) 24 hours to 39 days post-transduction and washed with buffer containing 2% fetal calf serum and 0.1% sodium azide. Cell viability was determined by staining with 3µM of 4',6-diamidino-2-phenylindole (DAPI, Life Technologies). GFP expression and DAPI staining were assayed on a Cyan ADP Flow Cytometer (Dako) or Hypercyt Autosampler (Intellicyt). Live cells were initially gated on the basis of forward scatter and side scatter, and then by DAPI exclusion.

GFP expression was determined within the live cell population. Data was analyzed using FlowJo software (FlowJo, LLC).

Primary Human Liver Transductions. Normal human liver samples were obtained from discard tissues from biopsies performed at City of Hope Medical Center under an IRB approved protocol. Primary liver tissue culture protocol was adapted from published protocols (8, 9). Human liver tissue was immediately placed, washed and sliced in ice cold William's Medium E (WME, ThermoFisher Scientific), supplemented with streptomycin, penicillin, glutamine and glucose. Transductions were done in 25 μ l of fresh supplemented WME and 5 μ l of AAVS1 editing vector. 48 hours post-transduction, liver slices were treated with 2 μ l of 100X collagenase IV (ThermoFisher Scientific). Washed dissociated liver was analyzed by flow cytometry for GFP expression and viability by DAPI staining.

Nuclear Vector Genome Detection. Mixed donor CB CD34 $^{+}$ cells (AllCells-Boston) were thawed according to the vendor's specific protocol. Cells were cultured in low cytokines as previously described (3, 6). Cells were transduced immediately post thaw at a MOI of 1.5E5 with AAVHSC and AAV PPP1R12C-GFP vectors in triplicate. 48 hours post transduction, nuclei were isolated using Nuclei Isolation Kit: Nuclei EZ Prep (Millipore Sigma) according to the manufacturer's protocol. High molecular weight DNA was extracted from nuclear pellets. The number of vector genomes per nucleus was determined by real time PCR using primers and probes specific to GFP and the housekeeping gene hApoB.

Targeted Integration PCR (TI-PCR). High molecular weight DNA was isolated by standard procedures (10). Targeted integration (TI) assays were performed by amplification using a primer

specific for the chromosomal sequences downstream or upstream of the homology arms and a primer specific for the insert. Primers used were:

AAVS1-F Primer – CACTGTTCCCCTCCAGGCAGGTC

AAVS1-3 Primer – CAAACATGCTGCTGAAGTGG

GFP-R Primer - GTAACAGCTCCTGCCCTTGCTC

Sigma AAVS1 Forward Primer - GGCCCTGGCCATTGTCACTT

GFP Reverse Primer - AACGAGAAGCGCGATCACA

HindIII Reverse Primer – CCAATCCTGTCCCTAGTAAAGCTT

Amplifications were performed with Roche Expand Hifidelity PCR system. Cycling conditions used were: 5min at 95°C (1 cycle); 15 cycles of (30s at 95°C, 30s start at 62°C and decrease by 0.5°C per cycle; 3min at 68°C); 20 cycles of (30s at 95°C, 30s at 53°C, 3min at 68°C); 5min at 68°C (final extension cycle). Amplifications were also performed with Takara PrimeSTAR GXL. Cycling conditions used were: 2min at 94°C (1 cycle); 35 cycles of (20s at 96°C, 20s at 61°C, 1:45min at 72°C); 2min at 72°C (final extension cycle). PCR products were either directly sequenced or cloned using TOPO TA Cloning Kit for Sequencing (Life Technologies) and sequenced.

Droplet digital polymerase chain reaction (ddPCR) based allele quantitation assay. All experiments were run using the BioRad QX200™ Droplet Digital™ PCR System. Editing was determined by calculating the linkage between the genomic target and the inserted vector payload and was measured by detecting the amount of partitioned droplets that contain both the vector and the genome in relation to expectation of coincidence by chance employing the methods used for genetic linkage between variants described in Regan et al., 2015 (11). A

concentration of 0.1ng/ul of genomic DNA was analyzed across a minimum of 3 experiments per sample for linkage and measured using multiplexed ddPCR with a vector specific probe set and a genomic specific probe set. In order to verify if the linkage measurements were due to genomic integration of the vector into the target genomic site, each sample was then digested with BamHI, a restriction enzyme that de-couples the linked vector probe target from the genomic DNA probe target, acting as if no alleles had been edited. In each digested sample, linkage was measured in the digested and un-digested samples. The 0.95 and 0.99 confidence intervals were calculated for each sample.

ddPCR Standard Curve. In order to measure genome editing/linkage as per the method above against a known quantity of edited material, a standard DNA series was created. The standard consisted of 100 unedited genomes per ul, 1000 episomal vector genomes per ul and a range of cloned positive alleles at 1 per ul, 5 per ul, 10 per ul, 15 per ul, 20 per ul and 25 edited alleles per ul respectively. The amount of genetic linkage in each sample was measured and plotted against the known ratio of unedited to edited allele in each sample ($R^2 = 0.972$, Pearson correlation $p<0.001$).

AAVS1_Genomic_Probe

Sequence - GCG TTA GAG GGC AGA GTT C

Sequence 2 - AGC TCC CAT AGC TCA GTC T

Sequence 3 - /56-FAM/CA TTG TCA C/Zen/T TTG CGC TGC CCT C/3IABkFQ/

Sequence - GCG TTA GAG GGC AGA GTT C

Sequence 2 - AGC TCC CAT AGC TCA GTC T

Sequence 3 - /5HEX/CA TTG TCA C/Zen/T TTG CGC TGC CCT C/3IABkFQ/

AAVS1_Homology_Arm_Probe

Sequence - CAT CCG TCG GAG AAG GC

Sequence 2 - CCC ATG CCG TCT TCA CTC

Sequence 3 - /56-FAM/TA AGA AAC G/Zen/A GAG ATG GCA CAG GCC /3IABkFQ/

Sequence - TCG GAG AAG GCC ATC CTA A

Sequence 2 - GTC TTC ACT CGC TGG GTT C

Sequence 3 - /5HEX/AA ACG AGA G/Zen/A TGG CAC AGG CCC /3IABkFQ/

Vector_Probe

Sequence - GCA ATA GCA TCA CAA ATT TCA C

Sequence 2 - GAT CCA GAC ATG ATA AGA TAC ATT G

Sequence 3 - /56-FAM/TC ACT GCA T/Zen/T CTA GTT GTG GTT TGT CCA /3IABkFQ/

Sequence - GCA ATA GCA TCA CAA ATT TCA C

Sequence 2 - GAT CCA GAC ATG ATA AGA TAC ATT G

Sequence 3 - /5HEX/TC ACT GCA T/Zen/T CTA GTT GTG GTT TGT CCA /3IABkFQ/

In Vivo Editing. All animal care and experiments were performed under protocols approved by the City of Hope Institutional Animal Care and Use Committee. 6–8-week old male NOD.CB17-Prkdcscid/NCrCrl (NOD/SCID) mice were obtained from Charles River Laboratories (3, 6) and maintained in a pathogen-free facility at the Animal Resources Center of the Beckman Research Institute at City of Hope National Medical Center (Duarte, CA). Mice were placed on

sulfamethoxazole and trimethoprim oral pediatric antibiotic (Hi-Tech Pharmacal), 10 ml/500 ml H₂O, for at least 48 hours prior to injection. No exclusion criteria, specific randomization, or blinding was used for *in vivo* studies. Editing vectors were injected intravenously via tail vein. Serial bioluminescent imaging was performed starting at Day 3 post injection and continued for at least 120 days as previously described (3). Briefly, anesthetized mice were injected intraperitoneally with 0.15 mg/g of luciferin (Caliper Life Sciences). Images were taken 10 minutes after luciferin injection using a SPECTRAL LagoX imaging system (Spectral Instruments Imaging, LLC). Rosa26-edited mice were imaged both ventrally and dorsally using medium binning for 1 minute. Organs were harvested from mice injected with luciferin post whole body imaging and specific organs imaged using the SPECTRAL LagoX imaging system immediately *ex vivo post mortem*. Images were analyzed using AMIView software version 1.7.06. Specific organ vector genome biodistribution was assayed using RainDance RainDrop dPCR System. Specifically, genomic DNA was extracted from mouse tissue using PureLink® Genomic DNA Mini Kit (Thermo Fisher Scientific LLC) and sheared by sonication. Cycling conditions used were: 10min at 95°C (1 cycle); 45 cycles of (15s at 95°C, 15s at 55°C, 45s at 60°C); 10min at 98°C (final cycle). Vector genome copies per cell were calculated by the ratio of the percentage of luciferase positive droplets to the percentage of mouse ApoB positive droplets.

Luc Forward Primer - CTCAAAGTATTCAAGCATAGGTGATGTC

Luc Reverse Primer - AACTGCACAAGGCCATGAAGA

Luc Probe - FAM-TTGCCTTCACTGATGCTCACATTGAGGT-TAMRA

mApoB Forward Primer - CGTGGGCTCCAGCATTCTA

mApoB Reverse Primer - TCACCAGTCATTCTGCCTTG

mApoB Probe - VIC-CCTTGAGCAGTGCCCGACCATTG-TAMRA

On Target Analysis. PCR products from GFP-specific TI-PCRs were used for on target analysis.

Briefly, K562 or PBSC were transduced with AAVS1 GFP editing vectors pseudotyped in AAVHSC7, AAVHSC15, or AAVHSC17 at MOIs ranging from 100K to 200K. Cells were collected between 4 to 39 days post-transduction and high molecular weight genomic DNA was used as template for TI-PCR. PCR products were cleaned using QIAquick PCR Purification Kit (Qiagen) and pooled. Pooled PCR products were sent to Genewiz and quality was assayed using electrophoresis and electropherogram analysis. A DNA library was generated and the paired ends were sequenced 2 X 150 bp using MiSeq platform (Illumina). A total of 1.9 million reads aligned to a reference AAVS1-GFP sequence using Bowtie-2 genome alignment tool (12). To capture the majority of potential indels, each read was queried for indels of 10bp or shorter, compared to the reference sequence (13). Multiple alignment analyses to quantify indels by these criteria were done using Geneious's iterative k-mer multiple alignment tool (<http://www.geneious.com>)(14). Sequences were verified by sequencing each ITR using primers (5' /5Biosg/CCATCTCATCCCTGCGTGTCTCCGTGGAGGAAGACGGAACCTG and 3' /5Biosg/CCATCTCATCCCTGCGTGTCTCCGGGGTCCTGGCTCTGCTCTAAGG) within the left and right homology arms <50bp from the ITR. A mixture of 1fmol of ITR containing plasmid DNA per 3.6pmol of untreated genomic DNA was used as a sequencing template. To create the sequencing libraries, each homology arm primer included a p7 Illumina handle and was biotinylated on the 5' terminus. Each biotinylated primer was used for 40 rounds of linear amplification of 100ng of plasmid/genomic DNA mixtures. Resultant biotinylated linear amplicon was purified by streptavidin purification using streptavidin dyna-bead magnetic separation. On-bead DNA was

treated with T4 RNA single-stranded ligase and 1nM of p5-Illumina adapter handle containing oligos (handle CCTCTCTATGGGCAGTCGGTGATG). The On-bead DNA was then used for a second round of amplification using p7 and p5 Illumina index oligos (P5 AATGATAACGGCGACCACCGAGATCTACACAAGTAGAGTCTTCCCTACACGACGCTTCCGATCTCCAT CTCATCCCTGCGTGTCTCC and P7 CAAGCAGAAGACGGCATACGAGATCATGATCGGTGACTGGAGTTAGACGTGTGCTTCCGATCTCC CTCTATGGGCAGTCGGTGATG) to create a 20pM sequencing library. Sequencing was performed using V2 300 cycle MiSeq kit (MS-102-2002). Sequencing results were aligned to a reference sequence of the ITR-containing plasmid using the Bowtie 2 read alignment software (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Each read with an intact barcode was included in the analysis. All reads with >50bp of alignment to the reference were tallied. Read alignment was visualized using the Geneious bioinformatics software package.

LAM-PCR. Detection of Rosa26 luciferase editing was performed via LAM-PCR. The following primers were designed immediately outside the editing homology arm for linear amplification:

Forward-LAM-Rosa1 Primer - CCGCCTAAAGAACGAGGCTGTGCTT

Forward-LAM-Rosa2 Primer - GGCGTGTGTTGGTGGCGTAAG

Reverse-LAM-Rosa1 Primer - CCTGGAAACCATTAAATGGTTAATAGCT

Linear amplification reactions were prepared with 50ug of genomic DNA from Rosa26-Luciferase mouse liver and ran using Takara LA-Taq (Takara Bio), with the program below: 2min at 95°C (1 cycle); 40 cycles of (30s at 95°C , 15s at 58°C, 2:30min at 72°C); 5min at 72°C (final extension). After electrophoresis on a 2% agarose gel, amplicons with sizes between 2-4 kb were isolated

using QIAGEN Gel extraction columns. Purified LAM-PCR amplicons were then used for nested PCR of Rosa-Luc integration using the following primer pairs:

PCR1: Forward- Rosa2 Primer - GGCGTGTTGGTGGCGTAAG

Rosa-Nested-Reverse 1 Primer - TTAAGCCTGCCAGAAGACT

PCR2: Forward- Rosa2 Primer - GGCGTGTTGGTGGCGTAAG

Rosa – Reverse 1 Primer - CCTGGAAACCATTAAATGGTTAATAGCT

Amplicons from each PCR were gel extracted and cloned in a TA-cloning system (pGem-T®, Promega) and sequenced by Sanger's method.

Gel Densitometry. ImageJ (<http://rsb.info.nih.gov/ij/index.html>) was used to compare the density of bands on an agar gel following PCR amplification (15). Step-by-step ImageJ analysis guidance is outlined in: <http://rsb.info.nih.gov/ij/docs/menus/analyze.html#gels>. Whole lane gel densitometry was measured between 0.5kb and 10kb of the DNA molecular weight marker. Densitometry measurements were collected as a histogram by size for each lane. A horizontal baseline reference was added to each histogram at a height of the density plot at 10kb (background density). Relative percent area of PCR peak densitometry was measured above the background threshold. Relative band intensity for each lane was measured compared to the no treatment.

Southern Blot Analysis. Southern blot protocol was adapted from published protocols (16). A biotin-labelled probe was synthesized by PCR using HotStartTaq (Qiagen) with 40% of dTTP in the reaction being replaced with biotin-16-dUTP (Biotium), and using the following primers directed to the luciferase ORF:

3'Luc-F Primer - ATT CAGCCATGCCAGAGAC

3'Luc-R Primer - CTAGACGGCGATTGCG

Mouse genomic DNA was prepared using the QiaAmp Fast DNA Tissue Kit (Qiagen). Fifty micrograms of DNA were digested overnight with Spel, precipitated and resuspended in 100 μ l TE with 10% glycerol. The entire sample was loaded onto a 0.8% agarose gel. DNA was transferred from a depurinated and denatured gel onto a Biodyne B Nylon Membrane (ThermoFisher Scientific). The DNA was crosslinked to the membrane on a UV Stratalinker 1800 (Stratagene). Pre-hybridized membrane was incubated overnight at 65°C with denatured probe. The washed blot was blocked in Odyssey Blocking Buffer (Li-Cor) with 1% SDS (1h/room temperature), then incubated in IRDye-800 streptavidin diluted 1:10,000 (45min/room temperature/with shaking). The blot was rinsed briefly in water, and then scanned on an Odyssey scanner.

Statistical Analysis. Significance for AAVHSC efficiency in comparison to other AAV clades shown in **Supplementary Fig. S2A** was determined by 2-tailed paired T-test using AAVHSCs as the comparison reference and outliers represented as individual circles. Boxplot depiction of individual AAVHSC editing compared to other AAV clades shown in **Supplementary Fig. S2A** show a median line, and quartiles are separated by boxes and whiskers. Branch distances in hierarchical clustering shown in **Supplementary Fig. S9B** are a relative representation of genome editing similarity across all queried serotypes. The statistical correlation between GFP expression and editing is p<0.05 (**Table S6**). Pearson test of correlation between percent GFP and percent editing was performed using RStudio (17, 18).

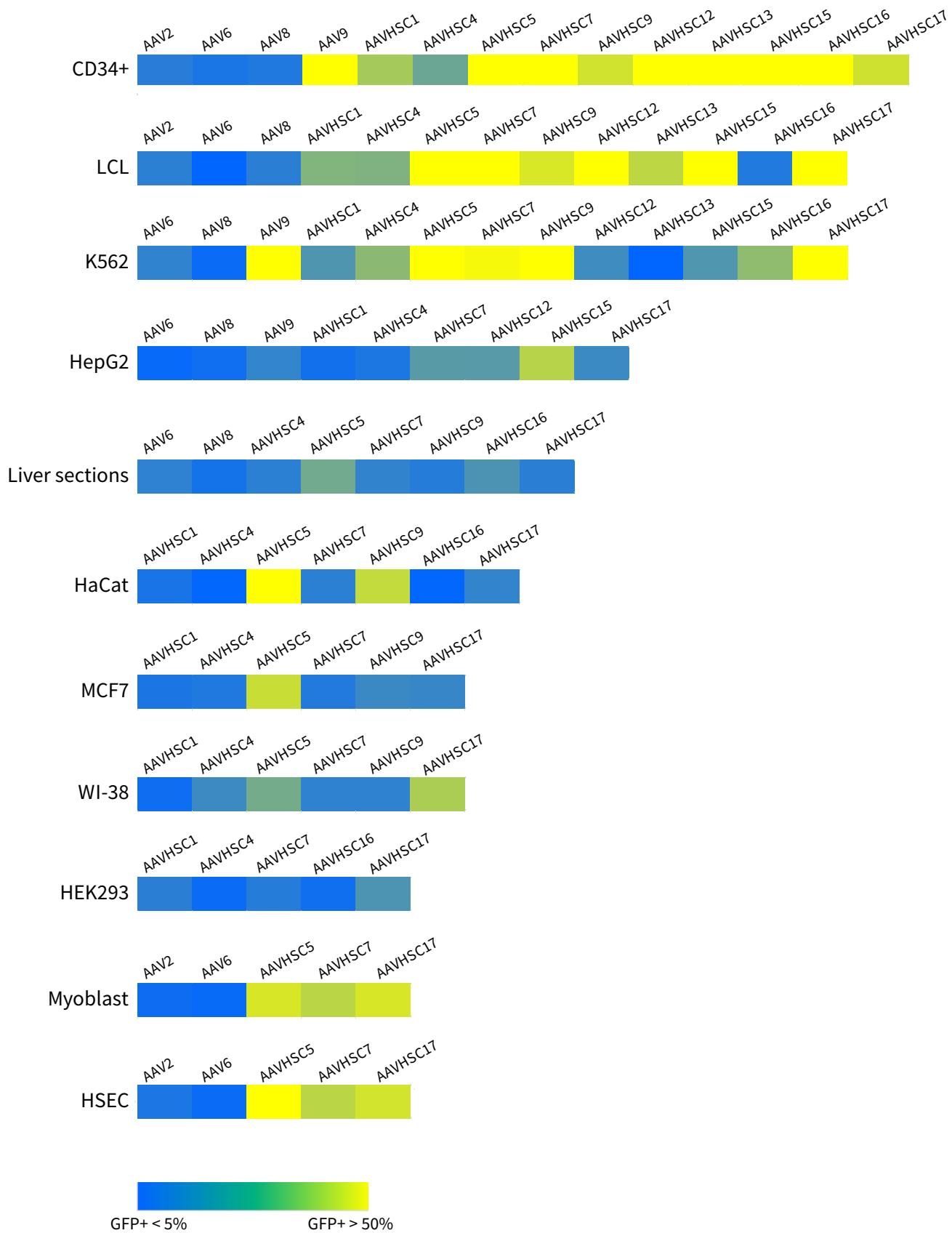


Figure S1

Fig. S1. AAVHSC and AAV9 mediate efficient genome editing in primary human cells and cell lines. Heat-map shows genome editing efficiency as measured by GFP expression in live cells after treatment with editing vectors that insert a promoterless GFP cassette into Intron 1 of the PPP1R12C gene within the AAVS1 safe harbor locus on the human chromosome 19. Genome editing efficiency was assayed across 11 cell and tissue types. Columns represent AAV serotypes tested. Data represents >906 individual measurements of genome editing.

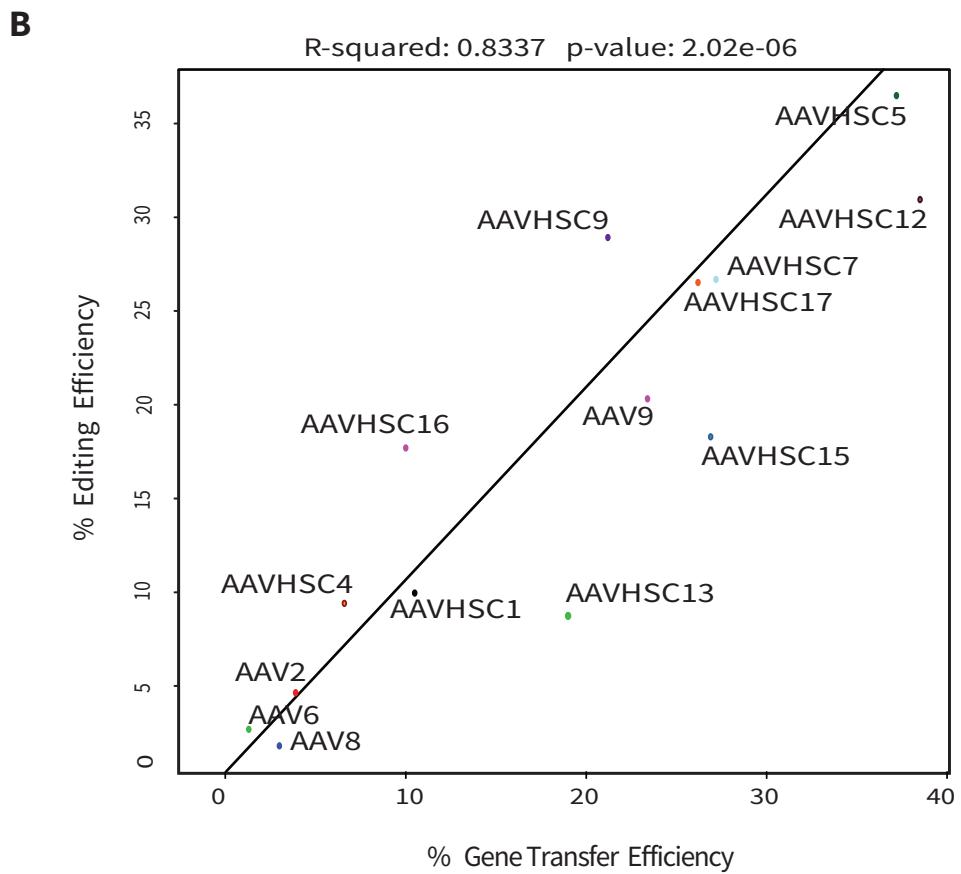
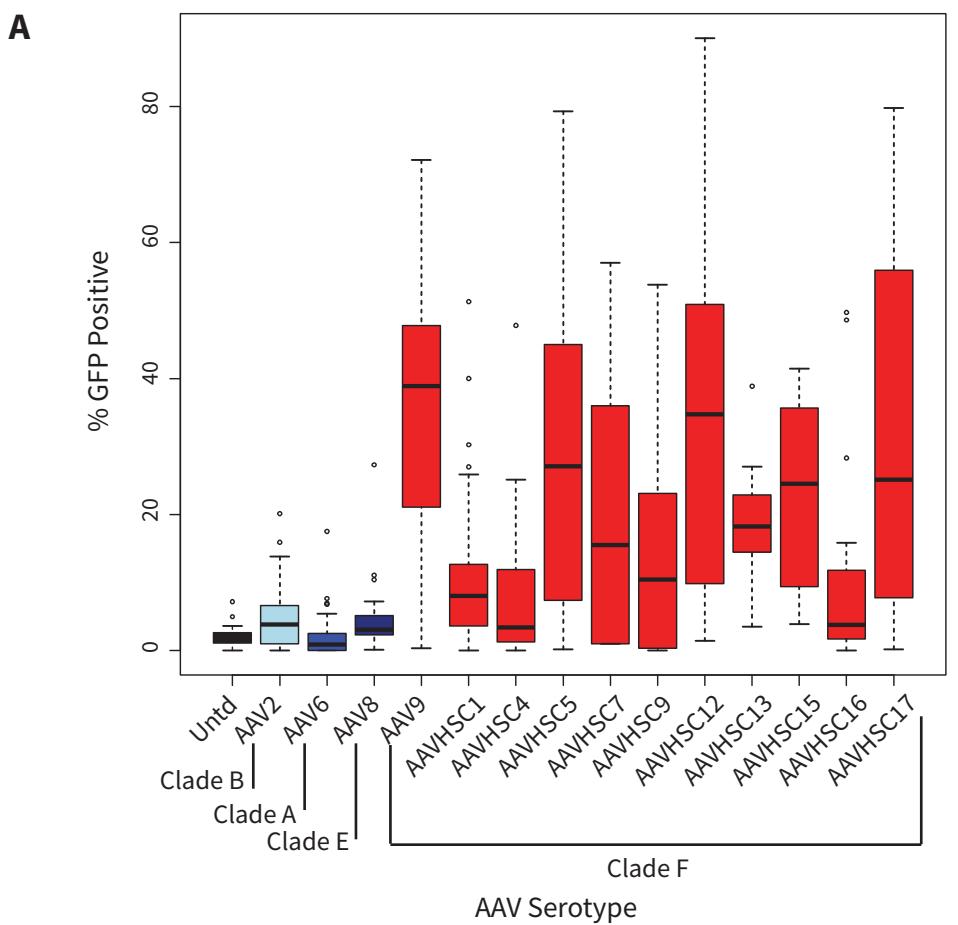


Figure S2

Fig. S2. Comparison of genome editing efficiency by AAV serotype. **(A)** Data shows genome editing efficiency as measured by GFP expression in live cells after treatment with PPP1R12C-GFP editing vectors. GFP was measured by flow cytometry. Genome editing efficiencies were pooled across all cell types studied. Data is displayed as boxplots, with a median line, and quartiles are separated by boxes and whiskers. Outliers are represented as individual circles. Sample sizes for the experimental groups are: Untransduced (untd), N=22; AAV2, N=20; AAV6, N=33; AAV8, N= 21; AAV9, N=30; AAVHSC1, N=28; AAVHSC4, N=29; AAVHSC5, N=58; AAVHSC7, N=56; AAVHSC9, N=31; AAVHSC12, N=21; AAVHSC13, N=12; AAVHSC15, N= 19; AAVHSC16, N=22; AAVHSC17, N=86. **(B)** Correlation of gene transfer and genome editing efficiencies by AAV serotype. Gene transfer efficiency was measured by fluorescent protein expression from a chicken beta actin (CBA) promoter within AAV vectors. Genome editing efficiency was measured by GFP expression after targeted insertion of a promoterless GFP ORF into Intron 1 of the PPP1R12C gene. The correlation coefficient and significance are noted.

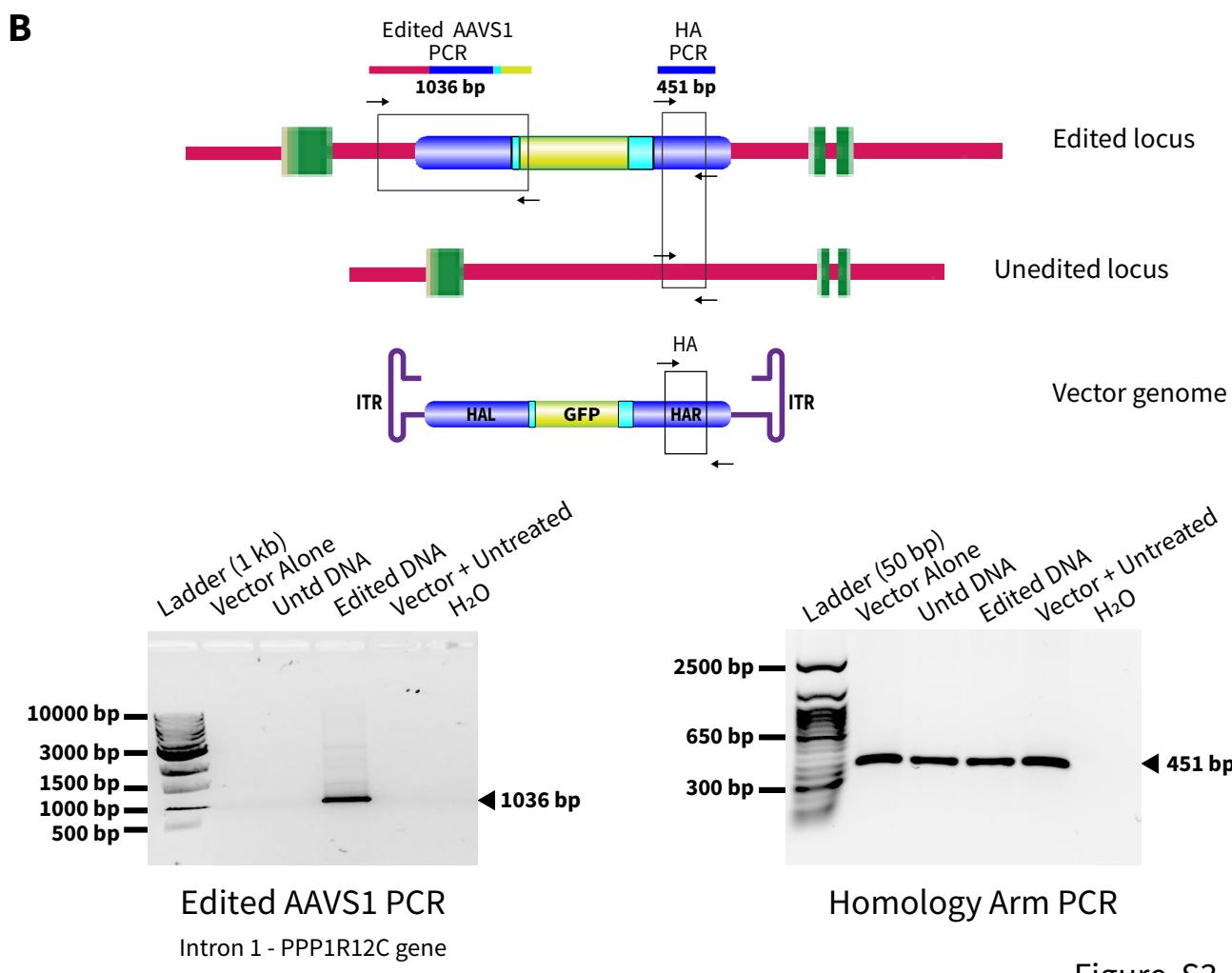
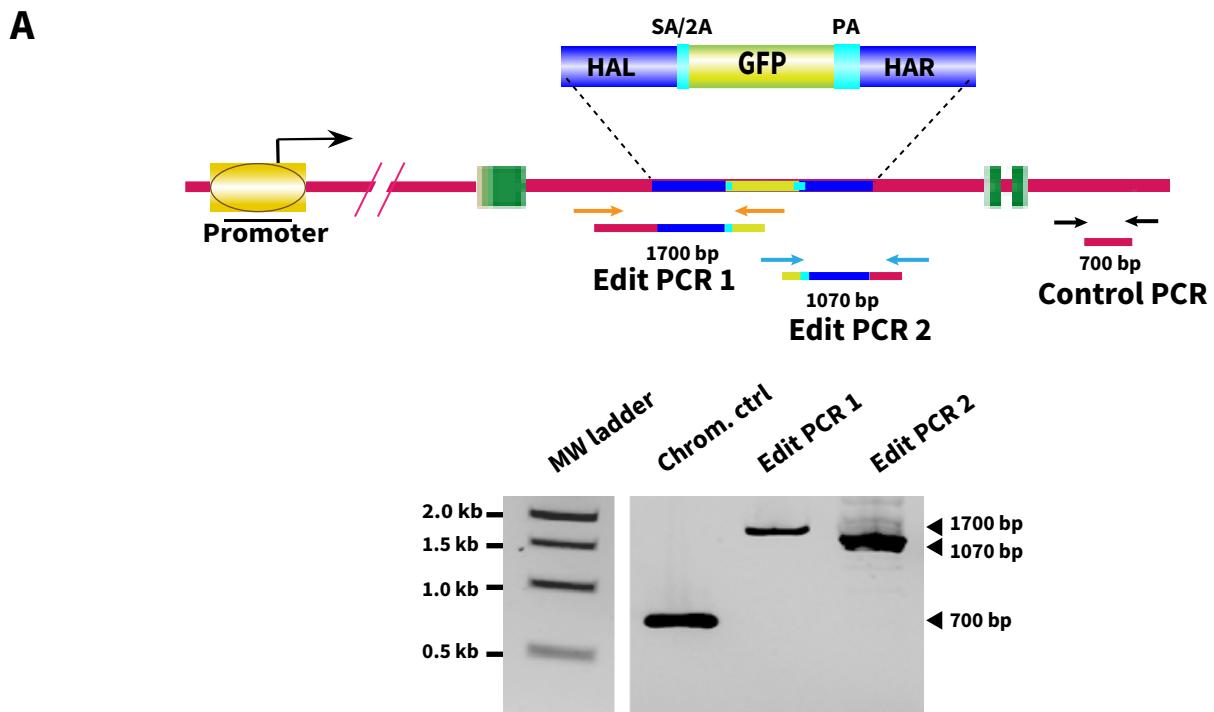
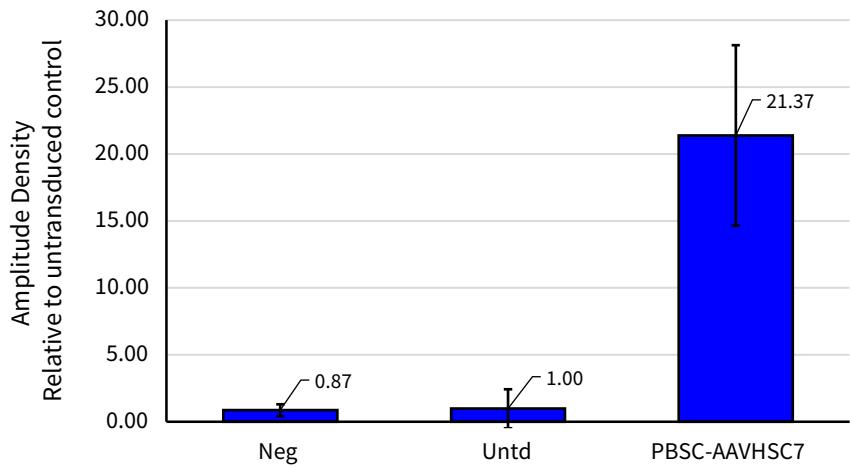


Figure S3

Fig. S3. Targeted Integration assays of the edited PPP1R12C locus. **(A)** Targeted integration analysis of edited K562 cells. Insert-specific and chromosome-specific primers used to amplify the 5' and 3' junction sequences of the edited region of PPP1R12C gene are depicted on the map. Gel shows the 1700bp 5'-junction amplicon, 1070bp 3'-junction amplicon and a 700bp chromosome-specific amplicon. Also shown is the molecular weight marker (MW). **(B)** Targeted integration assay of edited K562 cells showing the 1036bp 5'-junction of edited portion of the PPP1R12C gene. Samples of DNA isolated from AAVHSC viral prep (Vector alone), untreated human K562 genomic DNA (unttd DNA), genomic DNA from AAVHSC15-AAVS1 treated K562 cells (Edited DNA), and isolated vector DNA mixed with unedited genomic DNA (Vector + Untreated) were analyzed by end point PCR with primers diagnostic for homology arm, target DNA, editing and vector treatment. *Left gel*: Editing specific 1036bp PCR amplicon represents chromosomal sequences, including HAL and part of GFP. Absence of a 1036bp band in the Vector + Untreated DNA lane indicates that no crossover amplification occurred. *Right gel*: 451bp amplicon internal to HAR sequence. This amplifies both edited and unedited DNA and serves as a loading control.

A Densitometry of TI PCR Amplicons



B

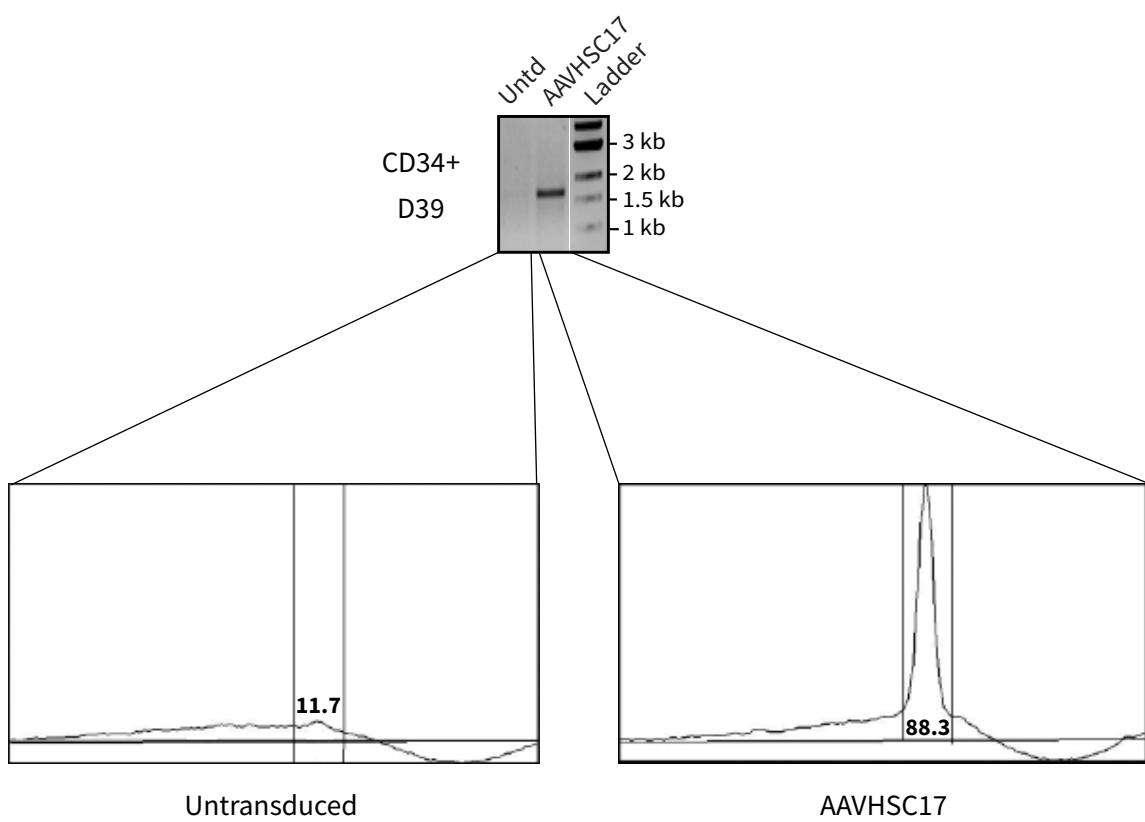
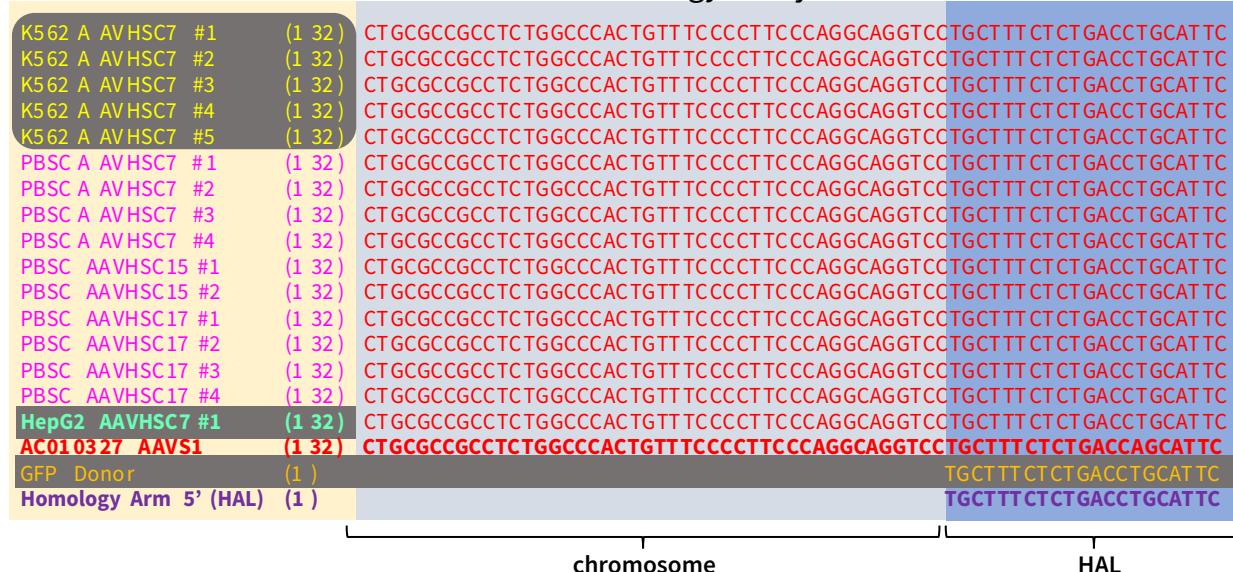


Figure S4

Fig. S4. Densitometry analysis of TI PCR assay of AAVHSC editing. **(A)** Mean densitometric measurements of each specified sample lane with relative band intensity measured compared to the no treatment. (Neg: negative control, N=3; Untd: untransduced, N=5; PBSC-AAVHSC7, N=6) **(B)** Day 39 untransduced and AAVHSC17 PPP1R12C-GFP transduced CD34⁺ cell sample lanes from Fig. 2B were analyzed for relative percent area of PCR peak densitometry above background threshold.

A

PPP1R12C Intron - Homology Arm junction



chromosome HAL

B

Homology Arm/GFP junction

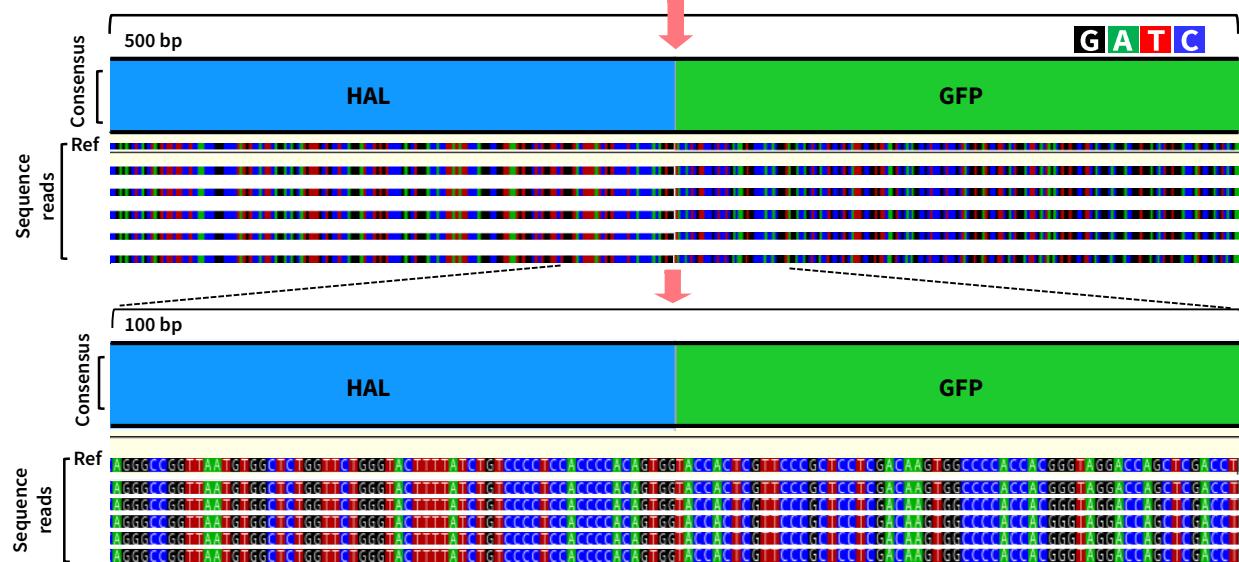
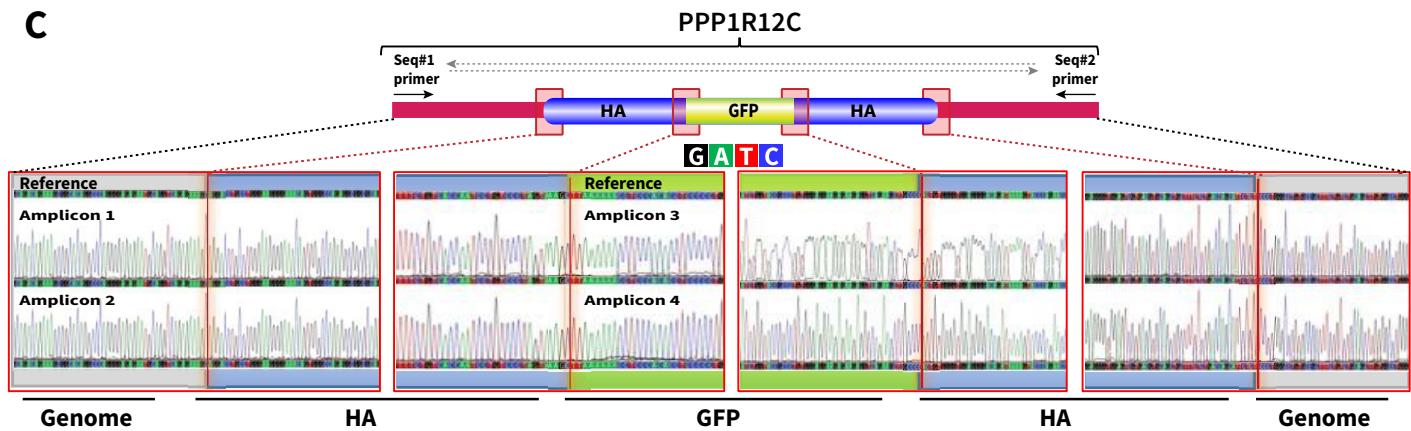
**C**

Figure S5

Fig. S5. Sequencing confirmation of accurate targeted insertion of the GFP cassette into Intron 1 of PPP1R12C. **(A)** Alignment of Sanger sequencing results for the homology arm-PPP1R12C target locus. The bottom three lines represent the reference sequences for the PPP1R12C gene/HA junction and the GFP sequence. **(B)** Alignment of HA/GFP insertion junction. Consensus sequences of the HA (blue) and the integrated GFP cDNA (green) are displayed as a stackplot of sequencing reads for the HA/PPP1R12C-GFP junction. Top line: Reference sequence. **(C)** Additional sequence confirmation of PPP1R12C genome editing including all junction regions. Sequencing alignments for chromosomal-HA junctions and GFP-HA junctions. Genomic sites where primers anneal to generate the sequenced amplicon are indicated (green arrows).

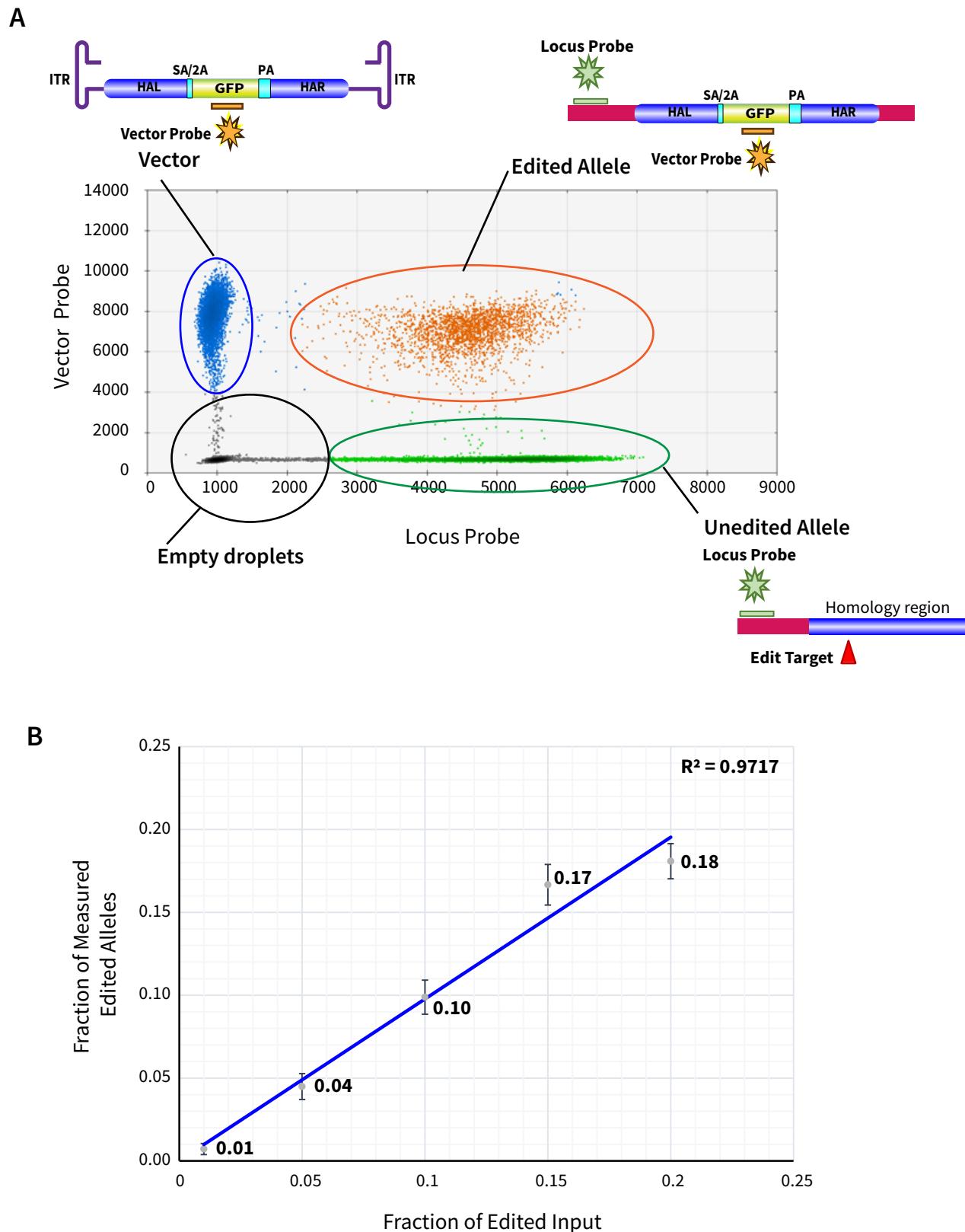


Figure S6

Fig. S6. Droplet digital PCR-based allele quantitation assay. **(A)** Edited alleles from CD34⁺ cells were quantitated by a droplet digital PCR (ddPCR) assay. The schema for the assay and a representative experiment are shown. The PPP1R12C gene was edited to insert the promoterless GFP cassette into intron 1. One set of GFP-specific primers and probes detected the GFP insert in the edited alleles and free vector. Another set of chromosome-specific primers and probes detected the PPP1R12C gene. Genomic DNA was highly diluted to prevent coincidence. Upper left quadrant depicts free vector (blue dots). Lower right quadrant depicts unedited chromosomes (green dots). Upper right quadrant depicts edited chromosomes, which display both the vector and the chromosome-specific signals (orange dots). Lower left quadrant depicts empty droplets (black dots). **(B)** In order to measure genome editing/linkage as per the method above against a known quantity of edited material a standard DNA series was created. The standard consisted of 100 unedited genomes per ul, 1000 episomal vector genomes per ul and a range of cloned positive alleles at 1 per ul, 5 per ul, 10 per ul, 15 per ul, 20 per ul and 25 edited alleles per ul respectively. The amount of genetic linkage in each sample was measured and plotted against the known ratio of unedited to edited allele in each sample ($R^2 = 0.972$, Pearson correlation $p < 0.001$). Details of that statistical analysis are shown in Table S4.

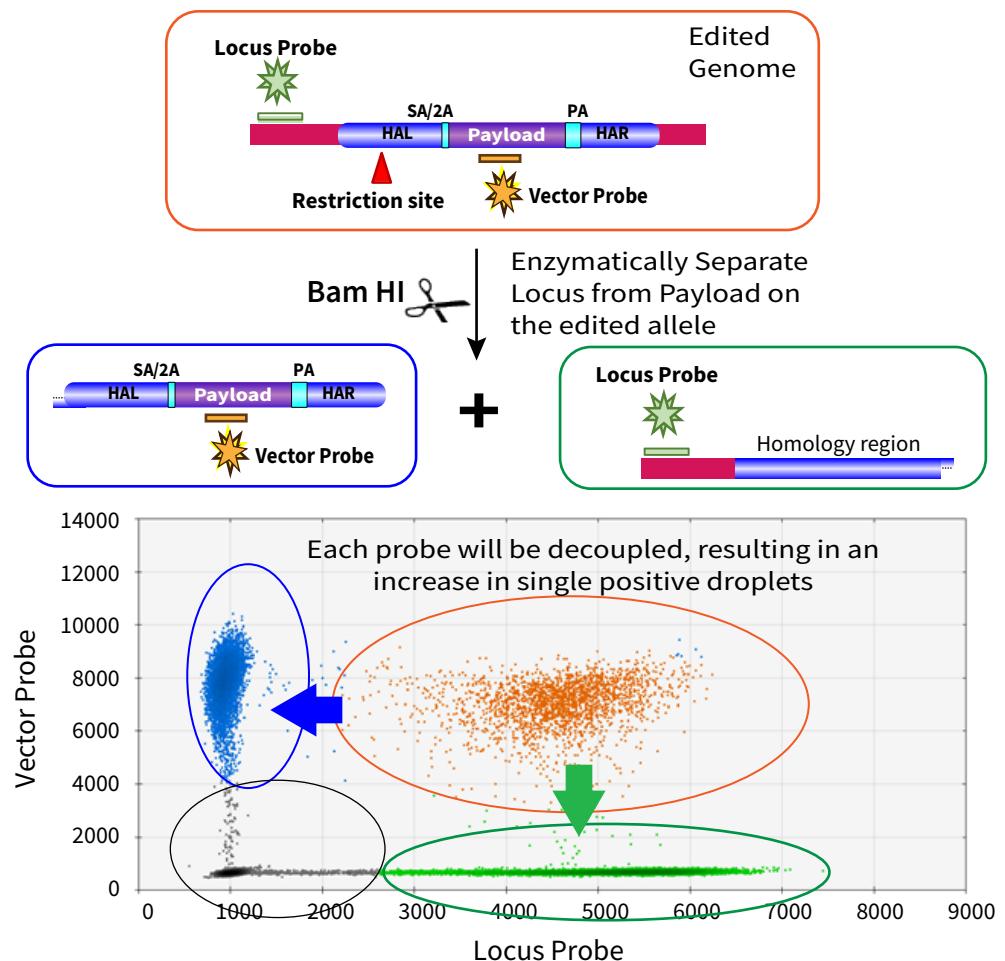
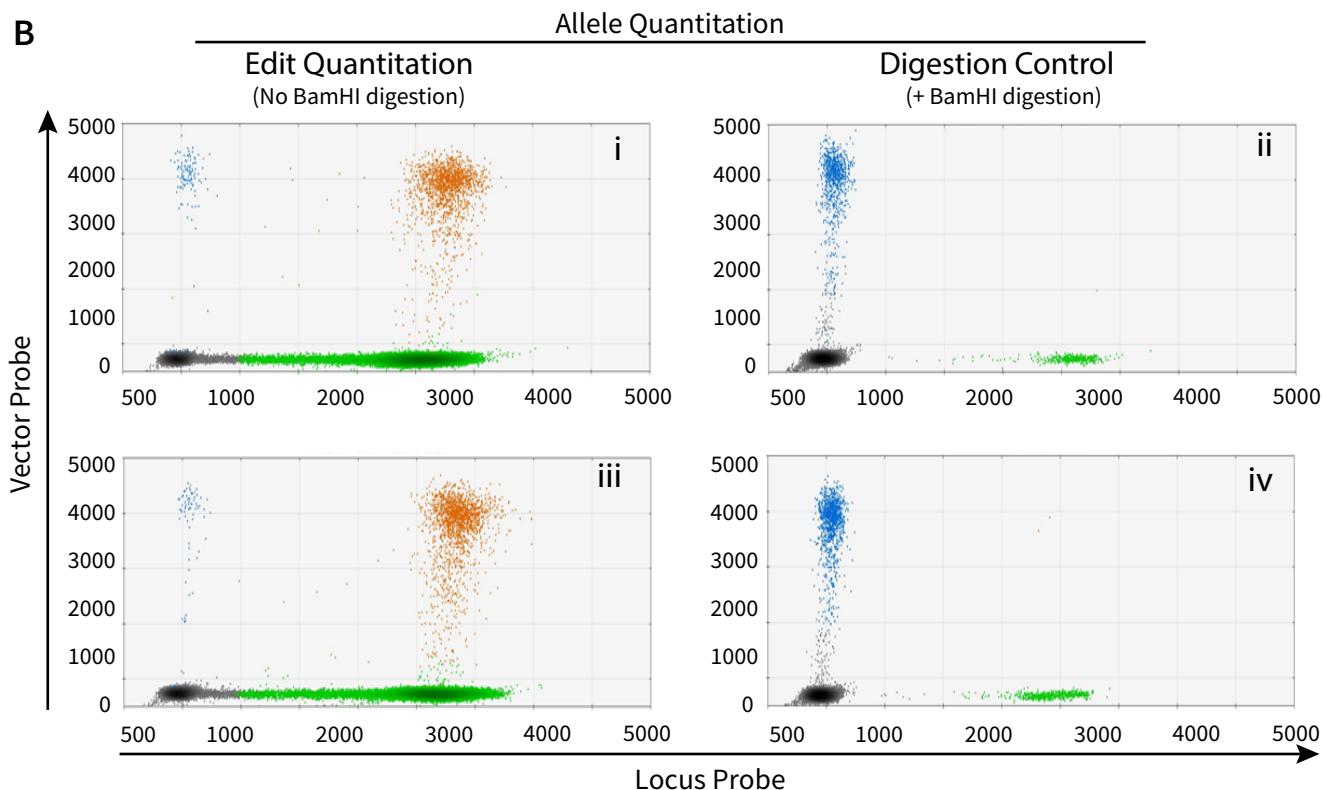
A**B****Figure S7**

Fig. S7. Specificity of ddPCR-based allele quantitation. **(A)** Schema of assay. One set of GFP specific primers and probes detected the GFP insert in the edited alleles and also the free vector. Another set of locus-specific primers and probes detected the PPP1R12C gene. Upper left quadrant depicts free vector (blue dots). Lower right quadrant depicts unedited chromosomes (green dots). Upper right quadrant depicts edited chromosomes, which display both the vector and the chromosome-specific signals (orange dots). Lower left quadrant depicts empty droplets (black dots). Digestion of the ddPCR product with a restriction enzyme that cuts between the binding site of the vector probe and the locus-specific probe is expected to separate the vector specific signal from the locus-specific signal only if they occur on the same DNA molecule. Thus, a reduction in the signal in the upper right quadrant (orange dots) after restriction enzyme digestion would indicate that the vector-specific sequences and the locus-specific sequences were coincident on the same chromosome, indicating successful editing. **(B)** Two representative experiments (experiment 1: panels i and ii; experiment 2: panels iii and iv) showing elimination of the vector/locus hybrid signals after restriction digestion (right - panels ii and iv) as compared with no digestion (left - panels i and iii). Eight replicates of this assay are shown in Table S5.

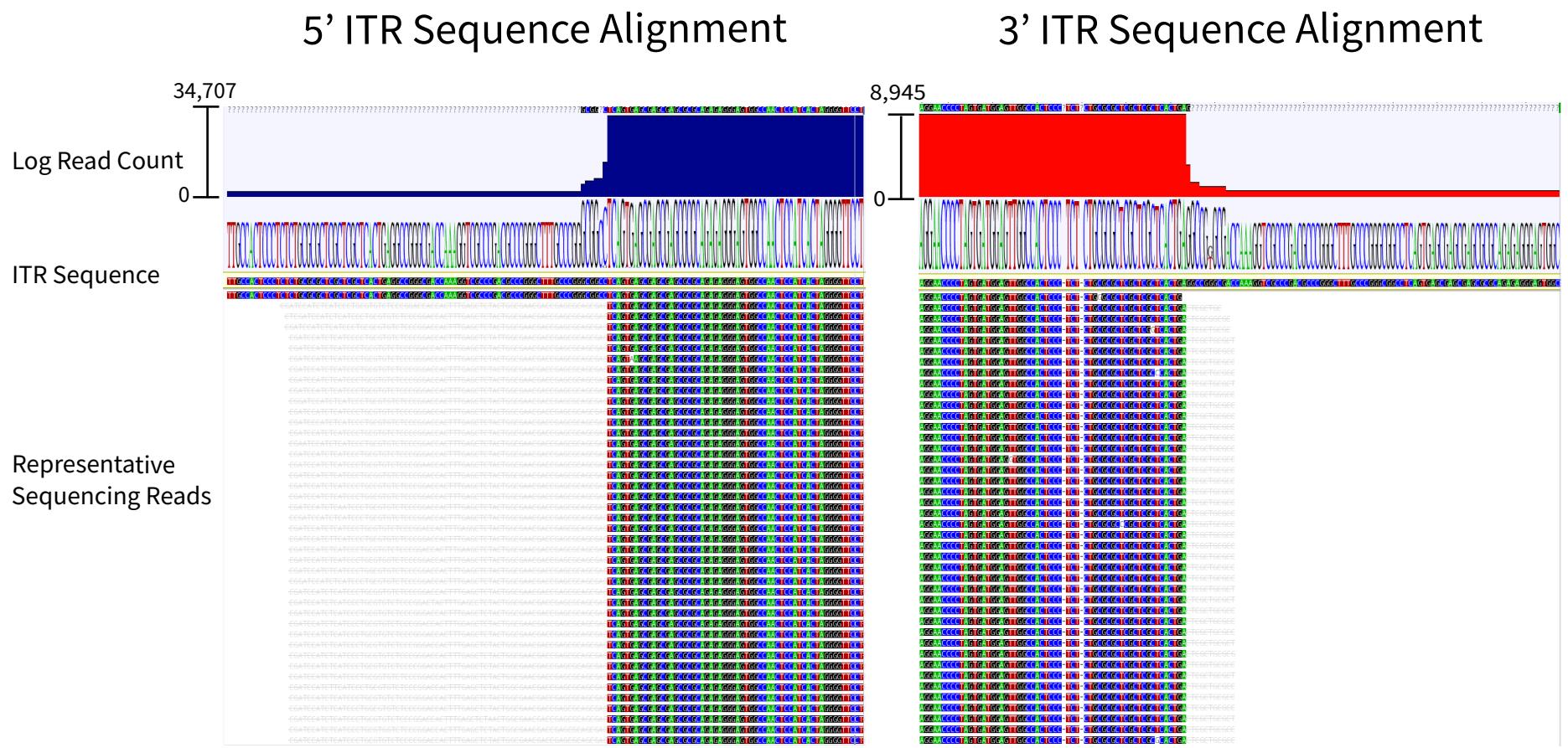


Figure S8

Fig. S8. Alignment of next generation sequencing (NGS) reads to the 3' and 5' AAV ITR sequences. In these libraries >34,000 and >8,000 reads covered at least 50bp of the 5' and 3' ITRs respectively. A small minority of reads spanned the entire ITR with the majority of sequences dropping off after ~60bp. Each ITR was sequenced using primers within the left and right homology arms <50bp from the ITR. A mixture of 1fmol of ITR containing plasmid DNA per 3.6pmol of untreated genomic DNA was used as a sequencing template. To create the sequencing libraries, each homology arm primer included a p7 Illumina handle and was biotinylated on the 5' terminus. Each biotinylated primer was used for 40 rounds of linear amplification of 100ng of plasmid/genomic DNA mixtures. Resultant biotinylated linear amplification was purified by streptavidin dyna-bead magnetic separation. On-bead DNA was treated with T4 RNA single-stranded ligase and 1nM of p5-Illumina adapter handle containing oligos. The On-bead DNA was then used for second round amplification using p7 and p5 Illumina index oligos to create a 20pM sequencing library. The resulting sequencing reads were aligned to a reference sequence of the ITR containing plasmid using

Bowtie 2 read alignment software (<http://bowtiebio.sourceforge.net/bowtie2/index.shtml>). Each read with an intact barcode was included in the analysis. All reads with >50bp of alignment to the reference were tallied. Read alignment was visualized using the Geneious bioinformatics software package.

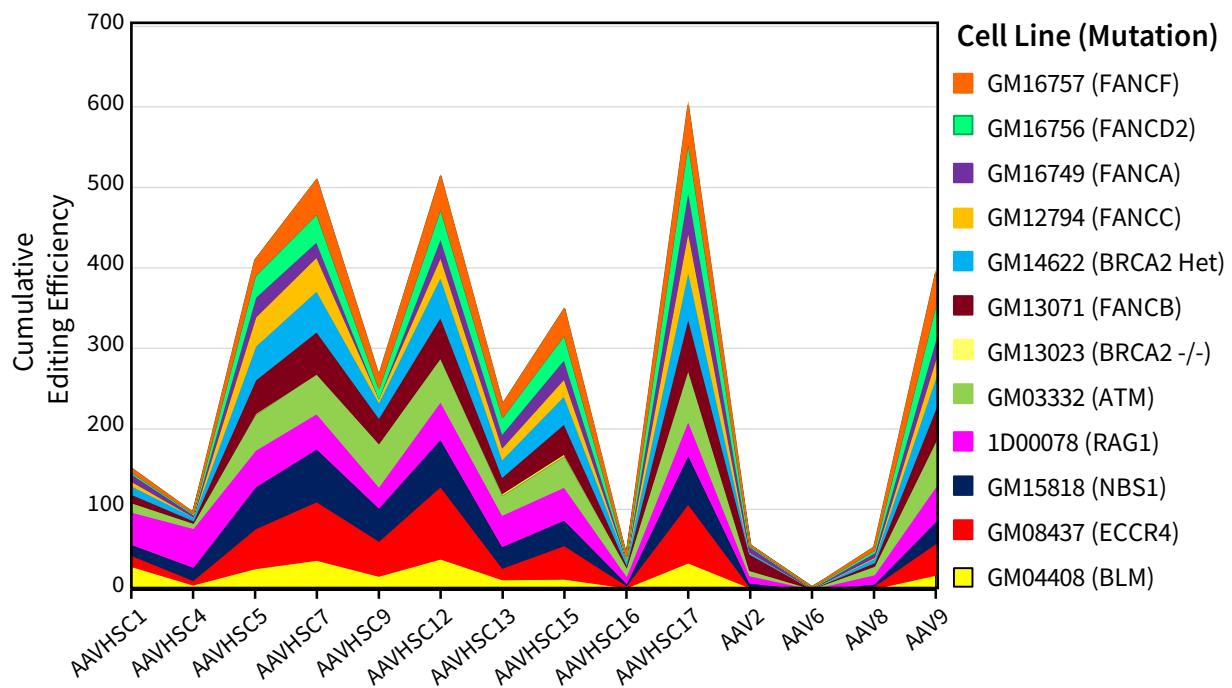
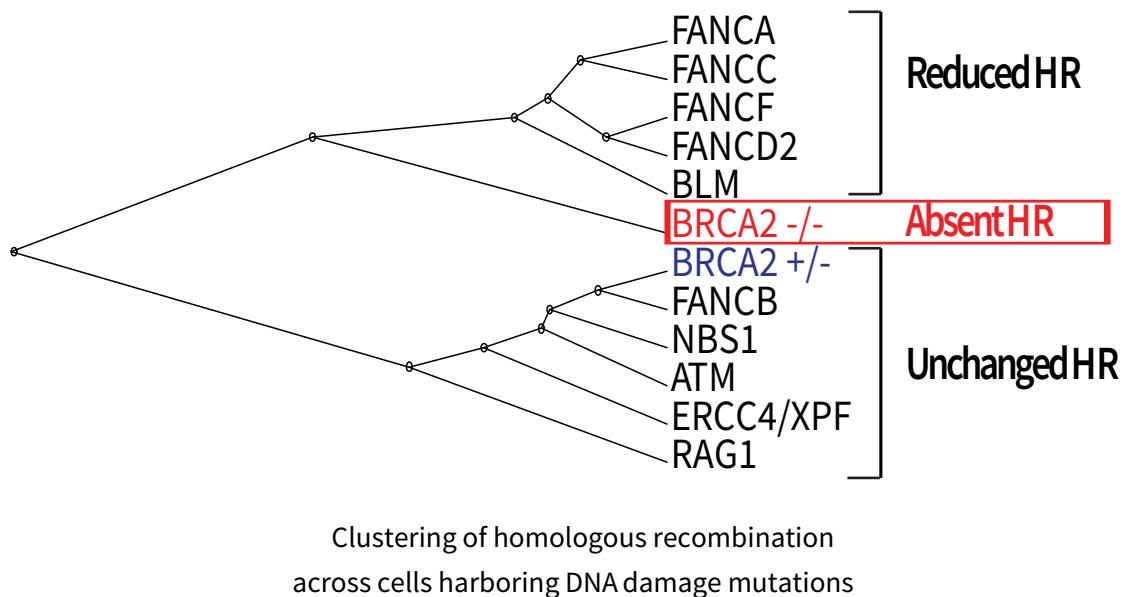
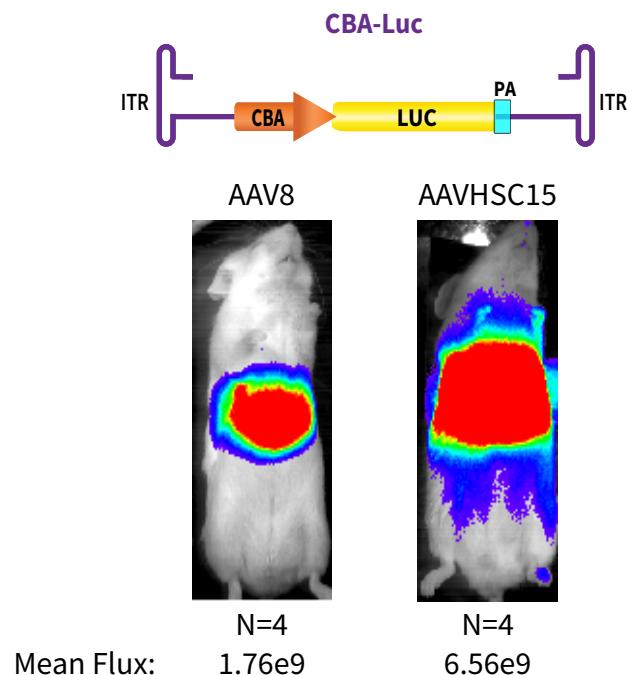
A**B**

Figure S9

Fig. S9. Role of DNA repair genes in genome editing by AAVHSC. **(A)** Cumulative genome editing efficiency in LCL bearing mutations of the DNA repair genes. Editing was measured by flow cytometry for GFP expression following targeted insertion of the promoterless GFP ORF into intron 1 of PPP1R12C gene. **(B)** Hierarchical clustering of genome editing efficiencies across LCLs bearing mutations in DNA repair genes. Branch distances are a relative representation of genome editing similarity across all queried serotypes.

A Gene Transfer



B Gene Editing

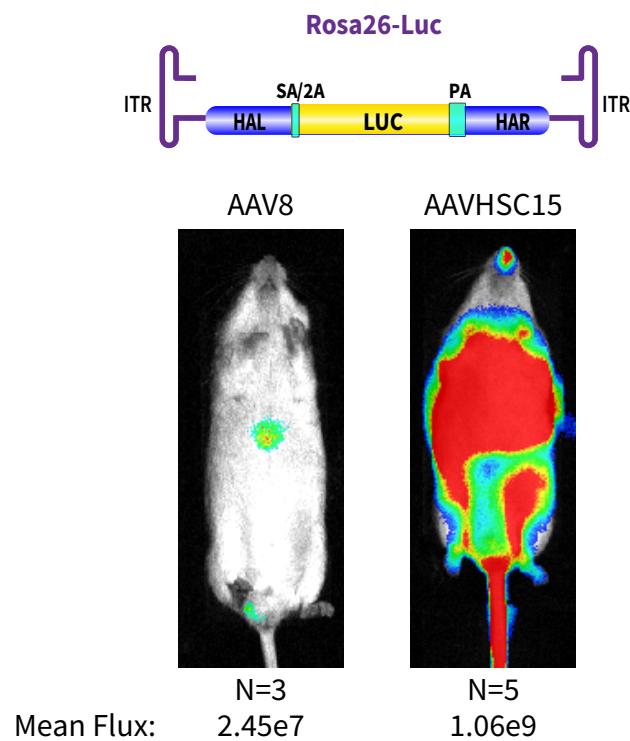
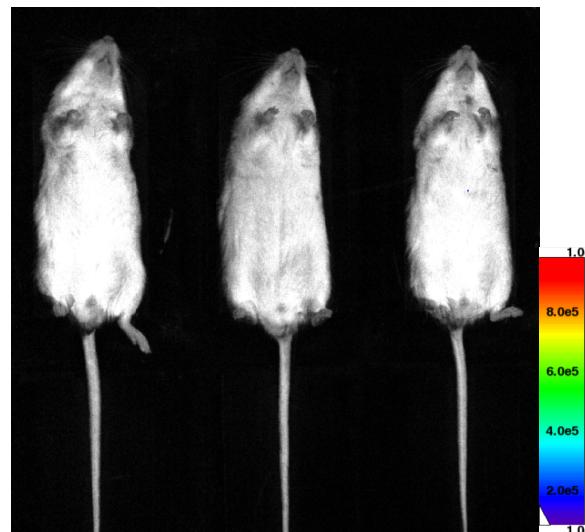


Figure S10

Fig. S10. A comparison of *in vivo* gene transfer and genome editing efficiencies between AAVHSC15 and AAV8. **(A)** *In vivo* bioluminescent imaging of luciferase expression in mice injected with AAV8 (left) and AAVHSC15 (right) gene transfer vectors encoding the luciferase gene under the control of a CBA promoter. Mice were injected intravenously with AAV vectors at 1e11 vg/mouse. AAV8: N=4; AAVHSC15: N=4. **(B)** *In vivo* bioluminescent imaging of luciferase expression in mice injected with AAV8 (left) and AAVHSC15 (right) genome editing vectors containing a promoterless luciferase ORF targeted to Intron 1 of the mouse Rosa26 gene. Mice were injected intravenously with AAV vectors at 5e11 vg/mouse. AAV8: N=3; AAVHSC15: N=5. The mean whole body flux (photons/sec/cm²) is displayed for each group at 50 - 60 days after IV injection.

A

AAVHSC15 no HA



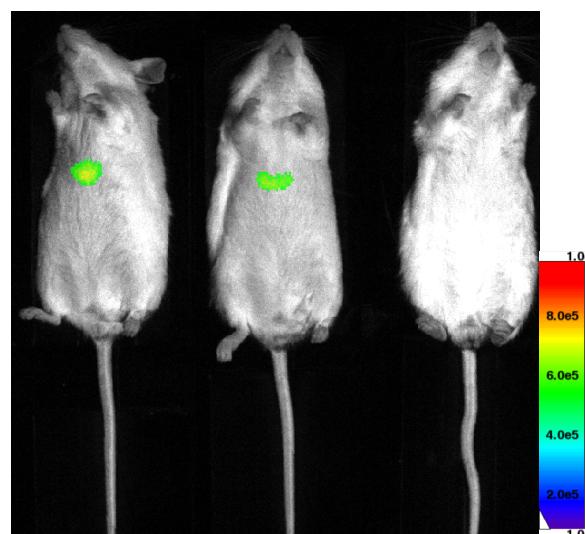
(N=3)

B

Rosa26-Luc



AAV8 Rosa26-Luc



(N=3)

Figure S11

Fig. S11. Individual control groups of mice for *in vivo* genome editing. **(A)** AAVHSC15 noHA: This vector lacked HA and only contained the promoterless luciferase bounded by an SA/2A sequence and a pA sequence. Mice were injected with 5e11 vg by tail vein injection. Imaging was performed 35 – 60 days post-injection. N=3. Each individual AAVHSC15 noHA-treated mouse is shown. Lack of editing indicates that HAs are required for gene targeting. **(B)** AAV8 Rosa26-Luc: This vector was identical to the AAVHSC15 Rosa26-Luc vector injected in the experimental mice shown in Fig. 5B, except for the use of the AAV8 capsid instead of the AAVHSC15 capsid. Mice were injected with 5e11 vg by tail vein injection. Imaging was performed 35 – 63 days post-injection. N=3. Each individual AAV8-treated mouse is shown.

Table S1. Details of Editing Vector Genomes

EDITING VECTORS	HOMOLOGY ARMS	EXPRESSION CASSETTE	TOTAL AAV GENOME
PPP1R12C-GFP	HAL: 800b; HAR: 800b	SA/2A-GFP-pA – 1064b	3095b
IL2RG-GFP 002 (ATG)	HAL: 800b; HAR: 800b	GFP-pA – 1077b	3063b
IL2RG-GFP 004 (Intron 7)	HAL: 800b; HAR: 800b	SA/2A-GFP-pA – 1084b	3070b
Nucleotide Substitution	HAL: 583b; HAR: 1045b	Restriction Site – 6b	2011b
Rosa-Luciferase	HAL: 800b; HAR: 800b	SA/2A-Luciferase-pA – 1830b	3966b

Note: HAL: Left Homology Arm; HAR: Right Homology Arm.

Table S2. Lack of Toxicity of AAVHSC Editing in Primary CD34⁺ Cells and Cell Lines *In Vitro*

CELL TYPE	RANGE OF MOI	AVERAGE VIABILITY OF CONTROL CELLS	AVERAGE VIABILITY OF AAVHSC EDITING VECTOR TRANSDUCED CELLS
CD34	1E4 to 1.5E5	56.3% + 27.0%	54.2% + 24.2%
K562	5E4 to 4E5	88.5% + 5.7%	87.8% + 3.1%
LCL	1.5E5	94.5% + 4.9%	91.4% + 9.1%

Note: Primary human CD34+ cells, K562 or B-lymphoblastoid cell lines were transduced with AAVHSC editing vectors at the indicated MOIs. Viability was assessed by DAPI staining and flow cytometry. Differences in viability between control and transduced cells were not significant with p values ≥ 0.5 .

Table S3. GFP Expression in Long-Term Cultures of CD34+ cells treated with AAVHSC

PPP1R12C-GFP Editing Vectors

	% GFP at Day 20	% GFP at Day 39
AAVHSC7	3.93	7.52
AAVHSC15	5.48	4.7
AAVHSC17	9.8	3.18

Note: Pooled cord blood CD34⁺ cells were transduced with PPP1R12C-GFP editing vectors packaged in the capsids indicated in column 1 at MOI: 1.5E5 and assayed for long term editing by flow cytometry analysis for GFP expression. TI analysis of cells from the AAVHSC17-treated culture is shown in Fig. 2B.

Table S4. Statistical Analysis of Standard Curve for ddPCR

Pearson Product Moment Correlation - Ungrouped Data		
Statistic	Standard Control	Measured Allele Frequency
Mean	8.5	8.3333333333333
Covariance	62.2	
Correlation	0.986018866397566	
Determination	0.972233204891942	
T-Test	11.8345661804367	
p-value (2 sided)	0.000291841684463132	
p-value (1 sided)	0.000145920842231566	
95% CI of Correlation	[0.873225486052631, 0.998536515175426]	
Degrees of Freedom	4	
Number of Observations	6	

Table S5. Quantitation of Colocalization of Locus and Vector Probes Before and After Digestion

	Allele Quantitation	Digestion Quantitation
Sample 1	0.0175	0.01
Sample 2	0.04875	0.012
Sample 3	0.0458	0.001
Sample 4	0.2825	0.0004
Sample 5	0.265	0.0008
Sample 6	0.384	0.01
Sample 7	0.374	0.002
Sample 8	0.339	0.0118

Note: Each sample represents CD34⁺ cells treated with AAVHSC17 PPP1R12C-GFP editing vector.

Editing was analyzed by ddPCR-based allele quantitation as described (Fig. S6). All samples were analyzed with and without a digestion control (Fig. S7B). The digestion control cleaves between the hybridization sites for the locus probe and the vector probe. The reduction of the signal after digestion indicates that the locus probe and the vector bound to the same DNA molecule and indicates successful editing of the allele.

Table S6. Correlation of Editing Quantitation with GFP Expression

Pearson Product Moment Correlation - Ungrouped Data		
Statistic	Fraction Edited Alleles	Fraction GFP Positive
Mean	0.256113806428571	0.216442857142857
Covariance	0.0240737319516286	
Correlation	0.871522073854803	
p-value (2 sided)	0.0105915534075008	
p-value (1 sided)	0.0052957767037504	
95% CI of Correlation	[0.344679360826762, 0.980844984363137]	
Degrees of Freedom	5	
Number of Observations	7	

Note: The statistical correlation between GFP expression and editing is p<0.05. Pearson test of correlation between percent GFP and percent editing was performed using RStudio (17, 18).

Table S7. DNA Repair Mutant Lymphoblastoid Lines: Detailed description of DNA repair mutant cell lines used. Also shown are the specific mutations and genes.

	Cell line	Gene	Cell type	Genotype	Mutation
BRCA2-KO	EUFA423	FANCD1/ BRCA2	Fibroblast	Compound Heterozygote	AT insertion at 7691 and A insertion at 9900
BRCA2 WT	GM08437	XPF/ECCR4	Fibroblast	Compound Heterozygote	T deletion at 1969 resulting in a frameshift and A>G transition at 1699 resulting in a substitution of threonine for alanine at codon 567
RAG1	1D00078	RAG1	LCL	Compound Heterozygote	C>T transition at 1106 resulting in a stop at codon 332 and G>A transition at 1794 resulting in a substitution of histidine for arginine at codon 561
ERCC4	GM08437	XPF/ECCR4	Fibroblast	Compound Heterozygote	T deletion at 1969 resulting in a frameshift and a A>G transition at 1699 resulting in a substitution of threonine for alanine at codon 567
ATM	GM03332	ATM	LCL	Homozygote	G>A transition at 7913 resulting in a substitution of a termination codon for tryptophan at codon 2638 in exon 55
NBS1	GM15818	NBS1	LCL	Homozygote	Deletion of 5 bp at 657 resulting in a premature termination at codon 218 in exon 6
FANCB	GM13071	FANCB	LCL	Homozygote	T insertion at 1838 resulting in a frameshift and a premature stop codon in exon 8 (19)
BRCA2 Heterozygous	GM14622	BRCA2	LCL	Heterozygote	TT deletion at 6503 resulting in a frameshift and a truncation at codon 2099 in exon 11
BRCA2 Homozygous	GM13023	FANCD1	LCL	Compound Heterozygote	G>T transition at 8415 resulting in a substitution of a lysine to an asparagine at codon 2729 and a C>A at 8732 resulting in a premature stop at codon 2835 (20)
BLM	GM04408	BLM	LCL	Compound Heterozygote	6-bp deletion and 7-bp insertion at 2281 resulting in a frameshift and a stop codon and a 1-bp deletion at 3233 resulting in a frameshift and premature termination
FANCD2	GM16756	FANCD2	LCL	Compound Heterozygote	A>G transition at 376 resulting in a substitution of glycine for serine at codon 126 and abnormal splicing leading to the insertion of 13 bp from intron 5 into the mRNA via the utilization of a cryptic splice site and a G>A transition at 3707 resulting in a substitution of histidine for arginine at codon 1236
FANCF	GM16757	FANCF	LCL	Compound Heterozygote	47 bp deletion at 349 and C>T transition at 16 resulting in a premature termination at codon 6
FANCC	GM12794	FANCC	LCL	Homozygote	1-bp deletion at 322 in exon 1 which results in a frameshift
FANCA	GM16749	FANCA	LCL	Homozygote	Deletion of exons 18 to 28 (21)
ATM	GM14158	ATM	LCL	Heterozygote	A insertion at 5712 at codon 1904 in exon 40 which results in a frameshift and a truncated protein
BRCA1	GM13710	BRCA1	LCL	Heterozygote	Arginine to glycine substitution at codon 1443 in exon 13
BRCA1	GM14090	BRCA1	LCL	Heterozygote	AG deletion at 185 at codon 23 in exon 3
BRCA1	GM14091	BRCA1	LCL	Heterozygote	C insertion at 5382 at codon 1755 in exon 20

Table S8: Luciferase Expression and Vector Genome Quantitation in Tissues Isolated from Mice Injected IV with AAVHSC15-Rosa26-Luc Editing Vector

ORGAN	FLUX*	VG / CELL**
Liver	3.82E+07	0.737
Muscle	3.64E+06	0.398
Heart	1.09E+05	0.317
Brain	7.94E+04	0.007
Kidney	2.49E+05	ND
Lung	3.69E+05	ND

Notes: *Average flux (photons/sec/cm²) from 3 mice injected with 5e11 vg/mouse of AAVHSC15-Rosa26-Luc. **VG/Cell: Quantitation by ddPCR; ddPCR was performed with a luciferase specific set of primers and probes. ApoB primers and probe set was used for quantitation per cell. Mice were harvested 161 to 181 days post-injection.

References for SI reference citations

1. Chatterjee S, Johnson P, & Wong K (1992) Dual-target inhibition of HIV-1 in vitro by means of an adeno-associated virus antisense vector. *Science* 258(5087):1485-1488.
2. Chatterjee S & Wong JKK (1993) Adeno-Associated Viral Vectors for the Delivery of Antisense RNA. *Methods* 5(1):51-59.
3. Smith LJ, *et al.* (2014) Gene transfer properties and structural modeling of human stem cell-derived AAV. *Molecular therapy : the journal of the American Society of Gene Therapy* 22(9):1625-1634.
4. Mroske C, Rivera H, Ul-Hasan T, Chatterjee S, & Wong KK (2012) A capillary electrophoresis sequencing method for the identification of mutations in the inverted terminal repeats of adeno-associated virus. *Hum Gene Ther Methods* 23(2):128-136.
5. Fisher-Adams G, Wong KK, Jr., Podsakoff G, Forman SJ, & Chatterjee S (1996) Integration of adeno-associated virus vectors in CD34+ human hematopoietic progenitor cells after transduction. *Blood* 88(2):492-504.
6. Santat L, *et al.* (2005) Recombinant AAV2 transduction of primitive human hematopoietic stem cells capable of serial engraftment in immune-deficient mice. *Proc Natl Acad Sci U S A* 102(31):11053-11058.
7. Paz H, *et al.* (2007) Quiescent subpopulations of human CD34-positive hematopoietic stem cells are preferred targets for stable recombinant adeno-associated virus type 2 transduction. *Hum Gene Ther* 18(7):614-626.

8. de Graaf IA, *et al.* (2010) Preparation and incubation of precision-cut liver and intestinal slices for application in drug metabolism and toxicity studies. *Nat Protoc* 5(9):1540-1551.
9. Olinga P & Schuppan D (2013) Precision-cut liver slices: a tool to model the liver ex vivo. *J Hepatol* 58(6):1252-1253.
10. Sambrook J (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press).
11. Regan JF, *et al.* (2015) A rapid molecular approach for chromosomal phasing. *PLoS One* 10(3):e0118270.
12. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
13. Sanjana NE, *et al.* (2016) High-resolution interrogation of functional elements in the noncoding genome. *bioRxiv*.
14. Kearse M, *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.
15. Gassmann M, Grenacher B, Rohde B, & Vogel J (2009) Quantifying Western blots: pitfalls of densitometry. *Electrophoresis* 30(11):1845-1855.
16. Zavala AG, Kulkarni AS, & Fortunato EA (2014) A dual color Southern blot to visualize two genomes or genic regions simultaneously. *J Virol Methods* 198:64-68.
17. Wessa P (2015) Pearson Correlation (v1.0.10) in Free Statistics Software (v1.1.23-r7) (Office for Research Development and Education).

18. Borghers E, and Wessa, P. (Statistics - Econometrics - Forecasting. (Office for Research Development and Education).
19. Meetei AR, *et al.* (2004) X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet* 36(11):1219-1224.
20. Ikeda H, *et al.* (2003) Genetic reversion in an acute myelogenous leukemia cell line from a Fanconi anemia patient with biallelic mutations in BRCA2. *Cancer Res* 63(10):2688-2694.
21. Joenje H, *et al.* (2000) Complementation analysis in Fanconi anemia: assignment of the reference FA-H patient to group A. *Am J Hum Genet* 67(3):759-762.