B) Among engrafted human cells, the distribution among B-cell, myeloid, or other (i.e. HSPC/RBC/T/NK/Pre-B) lineages. Bars represent median ± interquartile range. N=9 independent mouse transplantations.
C) Targeted allele frequency at *HBA1* determined by ddPCR among engrafted human cells in bulk sample and lineages. Bars represent median ± interquartile range. N=3 for mock treatment group and N=10 for targeted treatment group.

## METHODS

### *AAV6 vector design, production, and purification*

All AAV6 vectors were cloned into the pAAV-MCS plasmid (Agilent Technologies, Santa Clara, CA, USA), which contains inverted terminal repeats (ITRs) derived from AAV2. Gibson Assembly Mastermix (New England Biolabs, Ipswich, MA, USA) was used for the creation of each vector as per manufacturer's instructions. Cut site (CS) vectors were designed such that the left and right homology arms ("LHA" and "RHA", respectively) are immediately flanking the cut site at either *HBA2* or *HBA1* gene. Whole gene replacement (WGR) vectors have a LHA flanking the 5' UTR of either the *HBA2* or *HBA1* gene while the RHA immediately flanks downstream of its corresponding cut site. The LHA and RHA of every vector is 400bp, unless otherwise noted, with the vector name (*HBA2*/*HBA1* and CS/WGR) referencing the target integration type and homology arms used, respectively. Within Figure 1, CS and WGR vectors consisted of a SFFV-GFP-BGH expression cassette. An alternative promoter, UbC, was also used in creating a WGR vector for *HBA1* **(Supplemental Fig. 10)**. In Figure 2, WGR-T2A-YFP vectors consisted of the full-length *HBB* gene, unless noted, with a T2A-YFP expression cassette immediately following exon 3 of the *HBB* gene using the LHA and RHA described previously for WGR. These full-length HBB-T2A-YFP vectors were either flanked by 5' and 3' UTRs of HBB, *HBA2*, or *HBA1* as denoted in Figure 2a. In subsequent experiments, for targeting of SCD or β-thalassemia patient-derived $CD34^+$ HSPCs, WGR vectors were designed to target the *HBA1* site and contained a full-length *HBB* gene flanked by either *HBA1* UTRs or *HBB* UTRs. While the '*HBB* UTRs' and '*HBA1* UTRs' vector both share 400bp HAs, the '*HBA1* UTRs long HAs' vector was modified to have 880bp HAs. Few modifications were made to the production of AAV6 vectors as described[44]. 293T cells (Life Technologies, Carlsbad, CA, USA) were seeded in ten 15 $cm^2$ dishes with 13-15×$10^6$ cells per plate. 24h later, each dish was transfected with a standard polyethylenimine (PEI) transfection of 6μg ITR-containing plasmid and 22μg pDGM6 (gift from David Russell, University of Washington, Seattle, WA, USA), which contains the AAV6 cap genes, AAV2 rep genes, and Ad5 helper genes. After a 48-72h incubation, cells were lysed by 3 freeze-thaw cycles, treated with benzonase (Thermo Fisher Scientific, Waltham, MA, USA) at 250 U/mL, and the vector was then purified through an iodixanol gradient centrifugation at 48,000 RPM for 2.25h at 18°C. Afterwards, full capsids were

isolated at the 40–58% iodixanol interface and then stored at 80°C until further use. As an alternative method, AAVPro Purification Kit (All Serotypes)(Takara Bio USA, Mountain View, CA, USA) were also used following the 48-72 h incubation period, to extract full AAV6 capsids as per manufacturer's instructions. AAV6 vectors were titered using a Bio-Rad QX200 ddPCR machine and QuantaSoft software (Version 1.7; Bio-Rad, Hercules, CA, USA) to measure number of vector genomes as previously described[45].

***Culturing of CD34⁺ HSPCs***

Human CD34⁺ HSPCs were cultured as previously described[18,24,33,36,46,47]. CD34⁺ HSPCs were sourced from fresh cord blood (generously provided by Binns Family program for Cord Blood Research), frozen cord blood and Plerixafor- and/or G-CSF-mobilized peripheral blood (AllCells, Alameda, CA, USA and STEMCELL Technologies, Vancouver, Canada), frozen Plerixafor- and/or G-CSF-mobilized peripheral blood of patients with SCD, and frozen Plerixafor- and Filgrastim-mobilized peripheral blood from β-thalassemia patient (compound heterozygous - c.-138C>T and c.92+5G>C). The β-thalassemia-derived HSPCs were collected under protocol 14-H-0077, (registered on clinicaltrials.gov, NCT02105766), which was approved and renewed annually by the NHLBI IRB. The patient provided informed consent for the study. CD34⁺ HSPCs were cultured at $2.5\times10^5$–$5\times10^5$ cells/mL in StemSpan SFEM II (STEMCELL Technologies, Vancouver, Canada) base medium supplemented with stem cell factor (SCF)(100ng/mL), thrombopoietin (TPO)(100ng/mL), FLT3–ligand (100ng/mL), IL-6 (100ng/mL), UM171 (35nM), 20mg/mL streptomycin, and 20U/mL penicillin. The cell incubator conditions were 37°C, 5% $CO_2$, and 5% $O_2$.

***Genome editing of CD34⁺ HSPCs***

Chemically modified sgRNAs used to edit CD34⁺ HSPCs at either *HBA2* or *HBA1* were purchased from Synthego (Menlo Park, CA, USA) and TriLink BioTechnologies (San Diego, CA, USA) and were purified by high-performance liquid chromatography (HPLC). The sgRNA modifications added were the 2'-O-methyl-3'-phosphorothioate at the three terminal nucleotides of the 5' and 3' ends described previously[30]. The target sequences for sgRNAs were as follows: *sg1*: 5'-CTACCGAGGCTCCAGCTTAA-3'; *sg2*: 5'-GGCAGGAGGAACGGCTACCG-3'; *sg3*: 5'-GGGGAGGAGGGCCCGTTGGG-3'; *sg4*: 5'-CCACCGAGGCTCCAGCTTAA-3'; and *sg5*: 5'-GGCAAGAAGCATGGCCACCG-3'. All Cas9 protein (Alt-R S.p. Cas9 Nuclease V3) used was purchased from Integrated DNA Technologies (Coralville, Iowa, USA). The RNPs were complexed at a Cas9: sgRNA molar ratio of 1:2.5 at 25°C for 10min prior to electroporation. CD34⁺ cells were resuspended in P3 buffer (Lonza, Basel, Switzerland) with complexed RNPs and electroporated using the Lonza 4D Nucleofector (program DZ-100). Cells were plated at $2.5\times10^5$ cells/mL following electroporation in the cytokine-supplemented media described previously. Immediately following electroporation, AAV6 was supplied to the cells at $5\times10^3$-$1\times10^4$ vector

genomes/cell based on titers determined using a Bio-Rad QX200 ddPCR machine and QuantaSoft software (Version 1.7; Bio-Rad, Hercules, CA, USA).

***Indel frequency analysis by TIDE***

2-4d post-targeting, HSPCs were harvested and QuickExtract DNA extraction solution (Epicentre, Madison, WI, USA) was used to collect gDNA. The following primers were then used to amplify respective cut sites at *HBA2* and *HBA1* along with CleanAmp PCR 2x Master Mix (TriLink, San Diego, CA, USA) according to manufacturer's instructions: *HBA2* (sg1-3): forward: 5'-CCCGAAAGGAAAGGGTGGCG-3' reverse: 5'-TGGCACCTGCACTTGCACTG-3'; *HBA1* (sg4-5): forward: 5'-TCCGGGGTGCACGAGCCGAC-3', reverse: 5'-GCGGTGGCTCCACTTTCCCT-3'. PCR reactions were then run on a 1% agarose gel and appropriate bands were cut and gel-extracted using a GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to manufacturer's instructions. Gel-extracted amplicons were then Sanger sequenced with the following primers: *HBA2* (sg1-3): forward: 5'-GGGGTGCGGGCTGACTTTCT-3' reverse: 5'-CTGAGACAGGTAAACACCTCCAT-3'; *HBA1* (sg4-5): forward: 5'-TGGAGACGTCCTGGCCCC-3', reverse: 5'-CCTGGCACGTTTGCTGAGG-3'. Resulting Sanger chromatograms were the used as input for indel frequency analysis by TIDE (Version 3.2.0) as previously described[31].

***Gene targeting analysis by flow cytometry***

4-8d post-targeting with fluorescent gene replacement vectors, $CD34^+$ HSPCs were harvested and the percentage of edited cells was determined by flow cytometry. Cells were analyzed for viability using Ghost Dye Red 780 (Tonbo Biosciences, San Diego, CA, USA) and reporter expression was assessed using either the Accuri C6 flow cytometer and software (Version 9.4.11; BD Biosciences, San Jose, CA, USA) or FACS Aria II cytometer and FACS Diva software (Version 8.0.3; BD Biosciences, San Jose, CA, USA). The data was subsequently analyzed using FlowJo (Version 10.6.1; FlowJo LLC, Ashland, OR, USA).

***Allelic targeting analysis by ddPCR***

2-4d post-targeting, HSPCs were harvested and QuickExtract DNA extraction solution (Epicentre, Madison, WI, USA) was used to collect gDNA. gDNA was then digested using BAMH1-HF as per manufacturer's instructions (New England Biolabs, Ipswich, MA, USA). The percentage of targeted alleles within a cell population was measured with a Bio-Rad QX200 ddPCR machine and QuantaSoft software (Version 1.7; Bio-Rad, Hercules, CA, USA) using the following reaction mixture: 1-4µL of digested gDNA input, 10µL ddPCR SuperMix for Probes (No dUTP)(Bio-Rad, Hercules, CA, USA), primer/probes (1:3.6 ratio; Integrated DNA Technologies, Coralville, Iowa, USA), volume up to 20µL with $H_2O$. ddPCR droplet were then generated following the manufacturer's instructions (Bio-Rad, Hercules, CA, USA): 20µL of ddPCR reaction, 70µL droplet generation oil, and 40µL of droplet sample. Thermocycler (Bio-Rad, Hercules, CA, USA) settings

were as follows: 1. 98°C (10min), 2. 94°C (30s), 3. 57.3°C (30s), 4. 72°C (1.75min)(return to step 2 × 40–50 cycles), 5. 98°C (10 min). Analysis of droplet samples was done using the QX200 Droplet Digital PCR System (Bio-Rad, Hercules, CA, USA). To determine percentage of alleles targeted, the number of Poisson-corrected integrant copies/mL were divided by the number of Poisson-corrected reference DNA copies/mL. The following primers and 6-FAM/ZEN/IBFQ-labelled hydrolysis probes were purchased as custom-designed PrimeTime qPCR Assays from Integrated DNA Technologies (Coralvilla, IA, USA): All *HBA2*-GFP vectors (spans from BGH to outside 400bp *HBA2* RHA): forward: 5'-TAGTTGCCAGCCATCTGTTG-3', reverse: 5'-GGGGACAGCCTATTTTGCTA-3', probe: 5'-AAATGAGGAAATTGCATCGC-3'; All *HBA1*-GFP vectors (spans from BGH to outside 400bp *HBA1* RHA): forward: 5'- TAGTTGCCAGCCATCTGTTG-3', reverse: 5'-TAGTGGGAACGATGGGGGAT-3', probe: 5'-AAATGAGGAAATTGCATCGC-3'; *HBA2*-HBB-T2A-YFP vector (spans from YFP to outside 400bp *HBA2* RHA): forward: 5'-AGTCCAAGCTGAGCAAAGA-3', reverse: 5'-GGGGACAGCCTATTTTGCTA-3', probe: 5'-CGAGAAGCGCGATCACATGGTCCTGC-3'; All *HBA1*-HBB-T2A-YFP vectors (spans from YFP to outside 400bp *HBA1* RHA): forward: 5'-AGTCCAAGCTGAGCAAAGA-3', reverse: 5'-TAGTGGGAACGATGGGGGAT-3', probe: 5'-CGAGAAGCGCGATCACATGGTCCTGC-3'; *HBA1-HBB* vectors (with 400bp HAs, without T2A-YFP)(spans from *HBB* exon 3 to outside 400bp *HBA1* RHA): forward: 5'-GCTGCCTATCAGAAAGTGGT-3', reverse: 5'-TAGTGGGAACGATGGGGGAT-3', probe: 5'-CTGGTGTGGCTAATGCCCTGGCCC-3'; *HBA1-HBB* vector (with 880bp HAs, without T2A-YFP)(spans from *HBB* exon 3 to outside 880bp *HBA1* RHA): forward: 5'-GCTGCCTATCAGAAAGTGGT-3', reverse: 5'-ATCACAAACGCAGGCAGAG-3', probe: 5'-CTGGTGTGGCTAATGCCCTGGCCC-3'. The primers and HEX/ZEN/IBFQ-labelled hydrolysis probe purchased as custom-designed PrimeTime qPCR Assays from Integrated DNA Technologies (Coralvilla, IA, USA) were used to amplify the *CCRL2* reference gene: forward: 5'-GCTGTATGAATCCAGGTCC-3', reverse: 5'-CCTCCTGGCTGAGAAAAAG-3', probe: 5'-TGTTTCCTCCAGGATAAGGCAGCTGT-3'. Due to the length of the 'HBA1 UTRs long HAs' vector and to ensure episomal AAV is not detected, the ddPCR amplicon exceeds the template size recommended by the ddPCR manufacturer. Upon analysis of the data, the percentage of targeted alleles of this vector is underestimated. Therefore, in these instances a correction factor to account for this underestimation was determined by amplifying gDNA harvested from HSPCs targeted with *HBA1* UTRs vector with 400bp HAs with both sets of ddPCR primers and probes (those for vectors with 400bp and 880bp HAs) as well as *CCRL2* reference probes. The resulting correction factor was then applied to the targeted allele percentage from samples targeted with and amplified with primers and probe for 880bp HAs.

***Off-target activity analysis by rhAmpSeq***

Predicted off-target sites for *HBA1* sg5 was identified using COSMID with up to three mismatches allowed in the 19 PAM-proximal bases and the PAM sequence NGG. rhAmpSeq targeted sequencing was performed for the 40 most highly predicted off-target sites as described

previously. The raw sequencing files have been uploaded to the public NCBI data repository: BioProject ID PRJNA691350; available at http://www.ncbi.nlm.nih.gov/bioproject/691350.

***In vitro differentiation of CD34⁺ HSPCs into erythrocytes***

Following targeting, HSPCs derived from healthy, SCD, or β-thalassemia patients were cultured for 14-16d at 37°C and 5% $CO_2$ in SFEM II medium (STEMCELL Technologies, Vancouver, Canada) as previously described[34,35]. SFEMII base medium was supplemented with 100U/mL penicillin–streptomycin, 10ng/mL SCF, 1ng/mL IL-3 (PeproTech, Rocky Hill, NJ, USA), 3U/mL erythropoietin (eBiosciences, San Diego, CA, USA), 200µg/mL transferrin (Sigma-Aldrich, St. Louis, MO, USA), 3% antibody serum (heat-inactivated from Atlanta Biologicals, Flowery Branch, GA, USA), 2% human plasma (umbilical cord blood), 10µg/mL insulin (Sigma-Aldrich, St. Louis, MO, USA) and 3U/mL heparin (Sigma-Aldrich, St. Louis, MO, USA). In the first phase, d 0-7 (day zero being 2d post-targeting) of differentiation, cells were cultured at $1\times10^5$ cells/mL. In the second phase, d7–10, cells were maintained at $1\times10^5$ cells/mL, and IL-3 was removed from the culture. In the third phase, d11–16, cells were cultured at $1\times10^6$ cells/mL, and transferrin was increased to 1 mg/mL within the culture medium.

***mRNA analysis***

Following differentiation of HSPCs into erythrocytes, cells were harvested and RNA was extracted using RNeasy Plus Mini Kit (Qiagen, Hilden, Germany). Subsequently, cDNA was made from approximately 100ng RNA using the iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad, Hercules, CA, USA). Expression levels of β-globin transgene and α-globin mRNA were quantified with a Bio-Rad QX200 ddPCR machine and QuantaSoft software (Version 1.7; Bio-Rad, Hercules, CA, USA) using the following primers and 6-FAM/ZEN/IBFQ-labelled hydrolysis probes purchased as custom-designed PrimeTime qPCR Assays from Integrated DNA Technologies (Coralvilla, IA, USA): *HBB*: forward: 5'-GAGAACTTCAGGCTCCTG-3', reverse: 5'-CGGGGGTACGGGTGCAGGAA-3', probe: 5'-TGGCCATGCTTCTTGCCCCT-3'; *HBA* (does not distinguish between *HBA2* and *HBA1*): forward: 5'-GACCTGCACGCGCACAAGCTT-3', reverse: 5'-GCTCACAGAAGCCAGGAACTTG-3', probe: 5'-CAACTTCAAGCTCCTAAGCCA-3'. To normalize for RNA input, levels of the RBC-specific reference gene *GPA* was determined in each sample using the following primers and HEX/ZEN/IBFQ-labelled hydrolysis probes purchased as custom-designed PrimeTime qPCR Assays from Integrated DNA Technologies (Coralvilla, IA, USA): forward: 5'-ATATGCAGCCACTCCTAGAGCTC-3', reverse: 5'-CTGGTTCAGAGAAATGATGGGCA-3', probe: 5'-AGGAAACCGGAGAAAGGGTA-3'. ddPCR reactions were created using the respective primers and probes and droplets were generated as described above. Thermocycler (Bio-Rad, Hercules, CA, USA) settings were as follows: 1. 98°C (10min), 2. 94°C (30s), 3. 59.4°C (30s), 4. 72°C (30s)(return to step 2 × 40–50 cycles), 5. 98°C (10 min). Analysis of droplet samples was done using the QX200 Droplet Digital PCR System (Bio-Rad, Hercules, CA, USA). To determine relative expression levels,

the number of Poisson-corrected HBA or HBB transgene copies/mL were divided by the number of Poisson-corrected GPA copies/mL.

***Immunophenotyping of differentiated erythrocytes***

HSPCs subjected to the above erythrocyte differentiation were analyzed at d14-16 for erythrocyte lineage-specific markers using a FACS Aria II and FACS Diva software (Version 8.0.3; BD Biosciences, San Jose, CA, USA). Edited and non-edited cells were analyzed by flow cytometry using the following antibodies: hCD45 V450 (1:50 dilution; 2µL in 100µL of pelleted RBCs in 1x PBS buffer; HI30; BD Biosciences, San Jose, CA, USA), CD34 APC (1:50 dilution; 561; BioLegend, San Diego, CA, USA), CD71 PE-Cy7 (1:500 dilution; OKT9; Affymetrix, Santa Clara, CA, USA), and CD235a PE (GPA)(1:500 dilution; GA-R2; BD Biosciences, San Jose, CA, USA).

***Steady-state hemoglobin tetramer analysis***

HSPCs subjected to the above erythrocyte differentiation were lysed using water equivalent to three volumes of pelleted cells. The mixture was incubated at room temperature for 15min, followed by 30s sonication. For separation of lysate from the erythrocyte ghosts, centrifugation was performed at 13,000 RPM for 5min. HPLC analysis of hemoglobins in their native form were analyzed on a weak cation-exchange PolyCAT A column (100 × 4.6-mm, 3µm, 1,000Å) (PolyLC Inc., Columbia, MD, USA) using a Shimadzu UFLC system at room temperature. Mobile phase A (MPA) consists of 20mM Bis-tris + 2mM KCN, pH 6.96. Mobile phase B (MPB) consists of 20mM Bis-tris + 2mM KCN + 200mM NaCl, pH 6.55. Clear hemolysate was diluted four times in MPA, and then 20µL was injected onto the column. A flow rate of 1.5mL/min and the following gradients were used in time (min)/%B organic solvent: (0/10%; 8/40%; 17/90%; 20/10%; 30/stop).

***Reverse-phase HPLC globin chain analysis***

Analysis of globin chains in $CD34^+$ cell-derived erythroblasts was performed by reverse-phase HPLC, as previously described[48,49]. In brief, the reverse-phase HPLC assay was carried out on an Agilent 1260 Infinity II HPLC system with Diode -Array Detector. The chromatographic column is Aeris™ 3.6µm WIDEPORE XB-C18 200Å, LC Column 250 × 4.6mm behind a securityGuard™ ULTRA cartridge (Phenomenex). Globin chains were separated using a gradient program of 41–47% solvent B (acetonitrile) mixing with solvent A (0.1% trifluoroacetic acid in HPLC grade water at pH 2.9) and quantified by the area under the curve of the corresponding peaks in reverse-phase HPLC chromatogram.

***Methylcellulose CFU assessment***

2d post-targeting, HSPCs were stained using CD34 APC (561; BioLegend, San Diego, CA, USA), Ghost Dye Red 780 (Tonbo Biosciences, San Diego, CA, USA) and live $CD34^+$ cells were sorted into 96-well plates containing MethoCult Optimum (STEMCELL Technologies, Vancouver, Canada).

After 12–16d, colonies were appropriately scored based on external appearance in a blinded fashion.

***CD34$^+$ HSPC transplantation into immunodeficient NSG mice***

Six- to eight-week-old female NSG mice (Jackson Laboratory, Bar Harbor, ME, USA) were irradiated using 200rads of radiation 12-24h prior to transplantation with targeted HSPCs (2d post-targeting) via intrafemoral or tail-vein injections. Approximately $2.5\times10^5$-$1.3\times10^6$ electroporated HSPCs (exact number noted in figures) were injected using an insulin syringe with a 27G, 0.5 inch (12.7mm) needle. Mouse were housed at an ambient temperature of 22$^o$C with 50% humidity and 12-hour light/12-hour dark cycles. This experimental protocol was approved by Stanford University's Administrative Panel on Laboratory Animal Care. All mouse studies reported in this paper were performed as a minimum of three separate experimental replicates of editing and transplantation. For sample size, we transplanted as many mice as was feasible to cover the non-Gaussian distribution that would be expected from experimental and donor variability, while also minimizing the total number of animals as per FDA's Center for Biologics Evaluation and Research guidelines.

***Assessment of Human Engraftment***

15-17wks post-transplantation of CD34$^+$-edited HSPCs, mice were euthanized and bone marrow was harvested from tibia, femurs, pelvis, sternum, and spine using a pestle and mortar. Mononuclear cells (MNCs) were enriched using a Ficoll gradient centrifugation (Ficoll-Paque Plus, GE Healthcare, Chicago, IL) for 25min at 2,000g at room temperature. The samples were then stained for 30min at 4°C with the following antibodies: monoclonal anti-human CD33 V450 (1:50 dilution; 6uL in 300uL of MNCs pelleted in MACS buffer (1x PBS, 2% FBS, 2mM EDTA); WM53; BD Biosciences, San Jose, CA, USA); HLA-ABC FITC (1:100 dilution; W6/32; BioLegend, San Diego, CA, USA); CD19 PerCp-Cy5.5 (1:20 dilution; HIB19; BD Biosciences); anti-mouse PE-Cy5 mTer119 (1:400 dilution; TER-119; eBiosciences, San Diego, CA, USA); anti-mouse CD45.1 PE-Cy7 (1:200 dilution; A20; eBiosciences, San Diego, CA, USA); hGPA PE (1:200 dilution; HIR2; eBiosciences, San Diego, CA, USA); hCD34 APC (1:50 dilution; 581; BioLegend, San Diego, CA, USA); and CD10 APC-Cy7 (1:20 dilution; HI10a; BioLegend, San Diego, CA, USA). Multi-lineage engraftment was established by the presence of myeloid cells (CD33$^+$) and B-cells (CD19$^+$) of engrafted human cells (CD45$^+$; HLA-A/B/C$^+$ cells). For GFP-expressing cells, HLA-FITC was not used in the cocktail. For secondary transplantation, only a portion of the primary mouse mononuclear population was stained, and the rest ($2.5\times10^5$ cells-$1.3\times10^6$ cells) were transplanted into six- to eight-week-old NSG mice post-irradiation conditioning. Cells were the assessed in same aforementioned manner 16wks post-transplantation into secondary mice.

***Statistical analysis***

All statistical tests on experimental groups were done using GraphPad Prism software (Version 8.4). The exact statistical tests used for each comparison are noted in the figure legends.

***Data availability***

High-throughput sequencing data generated in the Cas9 off-target activity assessment has been uploaded to the public NCBI data repository: BioProject ID PRJNA691350; available at http://www.ncbi.nlm.nih.gov/bioproject/691350. All other data supporting the findings of this study are either included in the published article and/or supplementary information files or (if too large to be included therein) are available from the corresponding author on reasonable request. These include but are not limited to Sanger sequencing, ddPCR, flow cytometry, and HPLC data that were used to generate the figures and conclusions in this study.

***Code availability***

No previously unreported computer code or algorithm was generated during the course of this study.

***Research animals***

All reported experiments were completed in compliance with the institutional Animal Care and Use Committee (IACUC Protocol #: D16-00134) administered at Stanford by the Administrative Panel on Laboratory Animal Care (APLAC Protocol #: 25065) in accordance with Stanford University policy.

***Human research participants***

We have complied with all relevant ethical regulations for the following sources of primary human cells used in this study. Wild-type $CD34^+$ HSPCs were sourced from fresh cord blood (generously provided by Binns Family program for Cord Blood Research; Stanford University IRB Protocol #: 33813). β-thalassemia-derived HSPCs were collected under protocol 14-H-0077, (registered on clinicaltrials.gov, NCT02105766), which was approved and renewed annually by the NHLBI IRB. We obtained informed consent from patients participating in the study. All human-derived samples were de-identified prior to use in laboratory studies.

***Ethics/competing interests statement***

The authors of this study also wish to declare the following conflicts of interest: M.H.P. is a member of the scientific advisory board of Allogene Therapeutics. M.H.P. is on the Board of Directors of Graphite Bio. M.H.P. has equity in CRISPR Tx. C.A.V., N.M.B., G.K., M.A.C., G.R.R., and M.A.B. are employees of Integrated DNA Technologies.

***Author contributions***

DPD and MHP supervised the project. MKC, JC, RM, VAS, JFT, DPD, and MHP designed experiments. MKC, JC, RMM, BJL, CAV, VTL, YZ, AG, FZ, EP, WS, ROB, and NU carried out experiments. MKC, JC, DPD, and MHP wrote the manuscript.

***Acknowledgements***

The authors would like to thank the following funding sources that made this work possible: NIH NHLBI R01 - HL135607 that funded MHP and DPD as well as NIH NHLBI T32 - HL120824 that funded MKC.

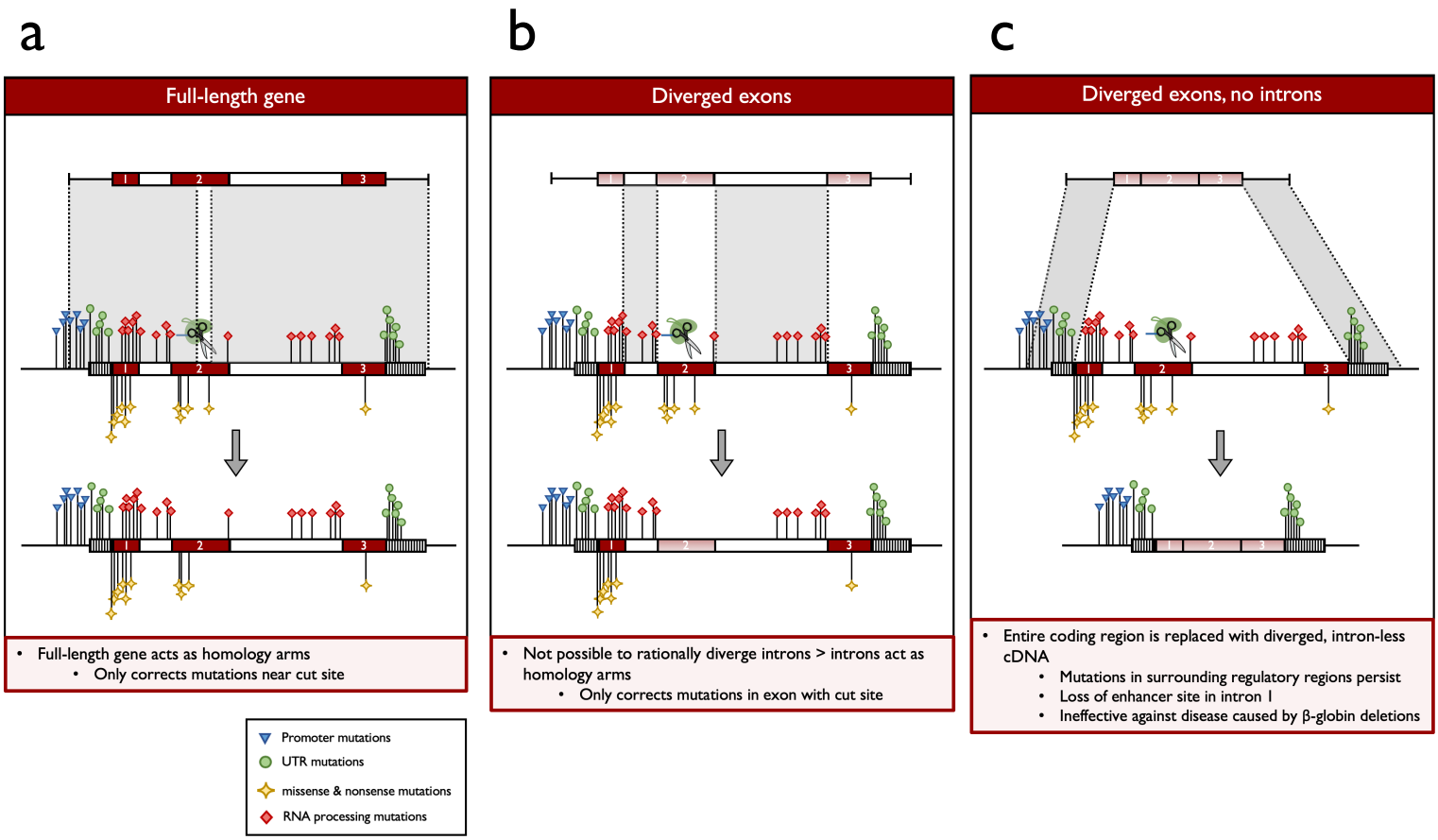
**REFERENCES**

1. Galanello, R. & Origa, R. Beta-thalassemia. *Orphanet J Rare Dis* **5**, 11 (2010).
2. Mentzer, W.C. & Kan, Y.W. Prospects for research in hematologic disorders: sickle cell disease and thalassemia. *JAMA* **285**, 640-642 (2001).
3. Ehlers, K.H., Giardina, P.J., Lesser, M.L., Engle, M.A. & Hilgartner, M.W. Prolonged survival in patients with beta-thalassemia major treated with deferoxamine. *J Pediatr* **118**, 540-545 (1991).
4. Mettananda, S., Gibbons, R.J. & Higgs, D.R. alpha-Globin as a molecular target in the treatment of beta-thalassemia. *Blood* **125**, 3694-3701 (2015).
5. Dye, D.E., Brameld, K.J., Maxwell, S., Goldblatt, J. & O'Leary, P. The impact of single gene and chromosomal disorders on hospital admissions in an adult population. *J Community Genet* **2**, 81-90 (2011).
6. Fleischhauer, K*., et al.* Graft rejection after unrelated donor hematopoietic stem cell transplantation for thalassemia is associated with nonpermissive HLA-DPB1 disparity in host-versus-graft direction. *Blood* **107**, 2984-2992 (2006).
7. Puthenveetil, G*., et al.* Successful correction of the human beta-thalassemia major phenotype using a lentiviral vector. *Blood* **104**, 3445-3453 (2004).
8. Negre, O*., et al.* Gene Therapy of the beta-Hemoglobinopathies by Lentiviral Transfer of the beta(A(T87Q))-Globin Gene. *Hum Gene Ther* **27**, 148-165 (2016).
9. Thompson, A.A*., et al.* Gene Therapy in Patients with Transfusion-Dependent beta-Thalassemia. *N Engl J Med* **378**, 1479-1493 (2018).
10. Breda, L*., et al.* Therapeutic hemoglobin levels after gene transfer in beta-thalassemia mice and in hematopoietic cells of beta-thalassemia and sickle cells disease patients. *PLoS One* **7**, e32345 (2012).
11. Cavazzana-Calvo, M*., et al.* Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* **467**, 318-322 (2010).
12. Marktel, S*., et al.* Intrabone hematopoietic stem cell gene therapy for adult and pediatric patients affected by transfusion-dependent ss-thalassemia. *Nat Med* **25**, 234-241 (2019).
13. Xu, L*., et al.* CRISPR-Edited Stem Cells in a Patient with HIV and Acute Lymphocytic Leukemia. *N Engl J Med* **381**, 1240-1247 (2019).

14. Stadtmauer, E.A*., et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**(2020).
15. Canver, M.C*., et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197 (2015).
16. Alter, B.P. Fetal erythropoiesis in stress hematopoiesis. *Exp Hematol* **7 Suppl 5**, 200-209 (1979).
17. Stamatoyannopoulos, G., Veith, R., Galanello, R. & Papayannopoulou, T. Hb F production in stressed erythropoiesis: observations and kinetic models. *Ann N Y Acad Sci* **445**, 188-197 (1985).
18. Bak, R.O., Dever, D.P. & Porteus, M.H. CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nat Protoc* **13**, 358-376 (2018).
19. Martin, R.M*., et al.* Highly Efficient and Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination. *Cell Stem Cell* **24**, 821-828 e825 (2019).
20. Pavel-Dinu, M*., et al.* Gene correction for SCID-X1 in long-term hematopoietic stem cells. *Nat Commun* **10**, 1634 (2019).
21. Gomez-Ospina, N*., et al.* Human genome-edited hematopoietic stem cells phenotypically correct Mucopolysaccharidosis type I. *Nat Commun* **10**, 4045 (2019).
22. Schiroli, G*., et al.* Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* **24**, 551-565 e558 (2019).
23. Schiroli, G*., et al.* Preclinical modeling highlights the therapeutic potential of hematopoietic stem cell gene editing for correction of SCID-X1. *Sci Transl Med* **9**(2017).
24. Dever, D.P*., et al.* CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384-389 (2016).
25. Pattabhi, S*., et al.* In Vivo Outcome of Homology-Directed Repair at the HBB Gene in HSC Using Alternative Donor Template Delivery Methods. *Mol Ther Nucleic Acids* **17**, 277-288 (2019).
26. DeWitt, M.A*., et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med* **8**, 360ra134 (2016).
27. De Ravin, S.S*., et al.* CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease. *Sci Transl Med* **9**(2017).
28. Thein, S.L. The molecular basis of beta-thalassemia. *Cold Spring Harb Perspect Med* **3**, a011700 (2013).
29. Weatherall, D. 2003 William Allan Award address. The Thalassemias: the role of molecular genetics in an evolving global health problem. *Am J Hum Genet* **74**, 385-392 (2004).
30. Hendel, A*., et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **33**, 985-989 (2015).
31. Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168 (2014).
32. Cradick, T.J., Qiu, P., Lee, C.M., Fine, E.J. & Bao, G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol Ther Nucleic Acids* **3**, e214 (2014).

33. Charlesworth, C.T.*, et al.* Priming Human Repopulating Hematopoietic Stem and Progenitor Cells for Cas9/sgRNA Gene Targeting. *Mol Ther Nucleic Acids* **12**, 89-104 (2018).
34. Dulmovits, B.M.*, et al.* Pomalidomide reverses gamma-globin silencing through the transcriptional reprogramming of adult hematopoietic progenitors. *Blood* **127**, 1481-1492 (2016).
35. Hu, J.*, et al.* Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood* **121**, 3246-3253 (2013).
36. Bak, R.O.*, et al.* Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *Elife* **6**(2017).
37. Andreani, M.*, et al.* Persistence of mixed chimerism in patients transplanted for the treatment of thalassemia. *Blood* **87**, 3494-3499 (1996).
38. Andreani, M.*, et al.* Long-term survival of ex-thalassemic patients with persistent mixed chimerism after bone marrow transplantation. *Bone Marrow Transplant* **25**, 401-404 (2000).
39. Ferrari, S.*, et al.* Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking. *Nat Biotechnol* **38**, 1298-1308 (2020).
40. Sharma, R., et al. TRACE-Seq Reveals Clonal Reconstitution Dynamics of Gene Targeted Human Hematopoietic Stem Cells. *biorxiv* (2020).
41. Magoc, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963 (2011).
42. McDermott, S.P., Eppert, K., Lechman, E.R., Doedens, M. & Dick, J.E. Comparison of human cord blood engraftment between immunocompromised mouse strains. *Blood* **116**, 193-200 (2010).
43. Wunderlich, M.*, et al.* AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. *Leukemia* **24**, 1785-1788 (2010).
44. Khan, I.F., Hirata, R.K. & Russell, D.W. AAV-mediated gene targeting methods for human cells. *Nat Protoc* **6**, 482-501 (2011).
45. Aurnhammer, C.*, et al.* Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. *Hum Gene Ther Methods* **23**, 18-28 (2012).
46. Cromer, M.K.*, et al.* Global Transcriptional Response to CRISPR/Cas9-AAV6-Based Genome Editing in CD34(+) Hematopoietic Stem and Progenitor Cells. *Mol Ther* **26**, 2431-2442 (2018).
47. Bak, R.O. & Porteus, M.H. CRISPR-Mediated Integration of Large Gene Cassettes Using AAV Donor Vectors. *Cell Rep* **20**, 750-756 (2017).
48. Park, S.H.*, et al.* Highly efficient editing of the beta-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease. *Nucleic Acids Res* **47**, 7955-7972 (2019).
49. Nemati, H., Bahrami, G. & Rahimi, Z. Rapid separation of human globin chains in normal and thalassemia patients by RP-HPLC. *Mol Biol Rep* **38**, 3213-3218 (2011).

Supplementary Fig. 1: Expected outcomes of introducing *HBB* transgene at endogenous locus




A) Expected outcome when integrating an un-diverged, full-length *HBB* (with introns) at the endogenous locus of HSPCs derived from patients with β-thalassemia. The varieties of disease-causing mutations are annotated in the figure legend.

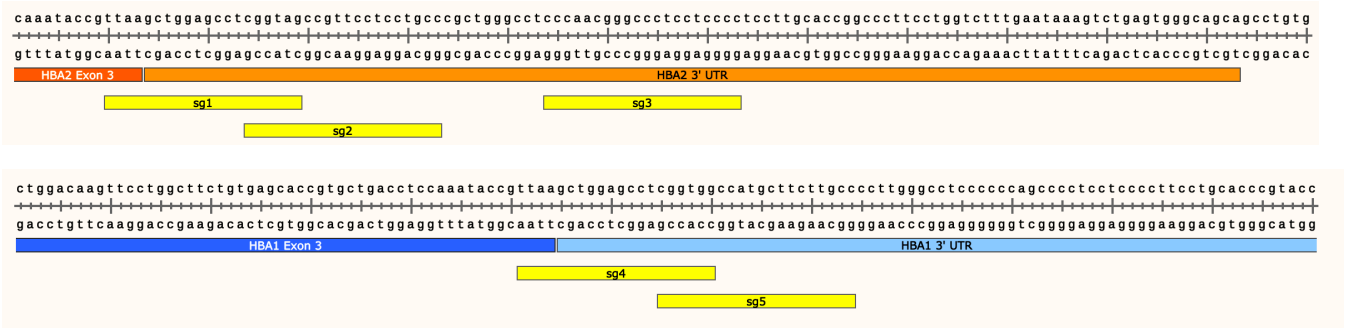
B) Expected outcome when integrating a diverged, full-length *HBB* (with introns) at the endogenous locus of HSPCs derived from patients with β-thalassemia.

C) Expected outcome when integrating a diverged, *HBB* cDNA (without introns) at the endogenous locus of HSPCs derived from patients with β-thalassemia.

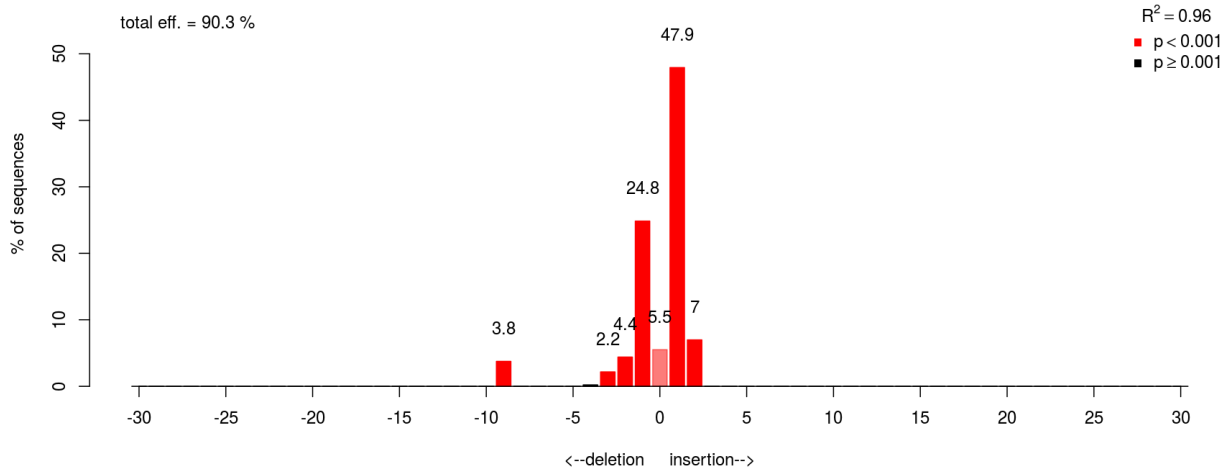
a

| Guide | Intended target | Sequence |
|---|---|---|
| sg1 | *HBA2* | CTACCGAGGCTCCAGCTTAANGG |
| sg2 | *HBA2* | GGCAGGAGGAACGGCTACCGNGG |
| sg3 | *HBA2* | GGGGAGGAGGGCCCGTTGGGNGG |
| sg4 | *HBA1* | CCACCGAGGCTCCAGCTTAANGG |
| sg5 | *HBA1* | GGCAAGAAGCATGGCCACCGNGG |

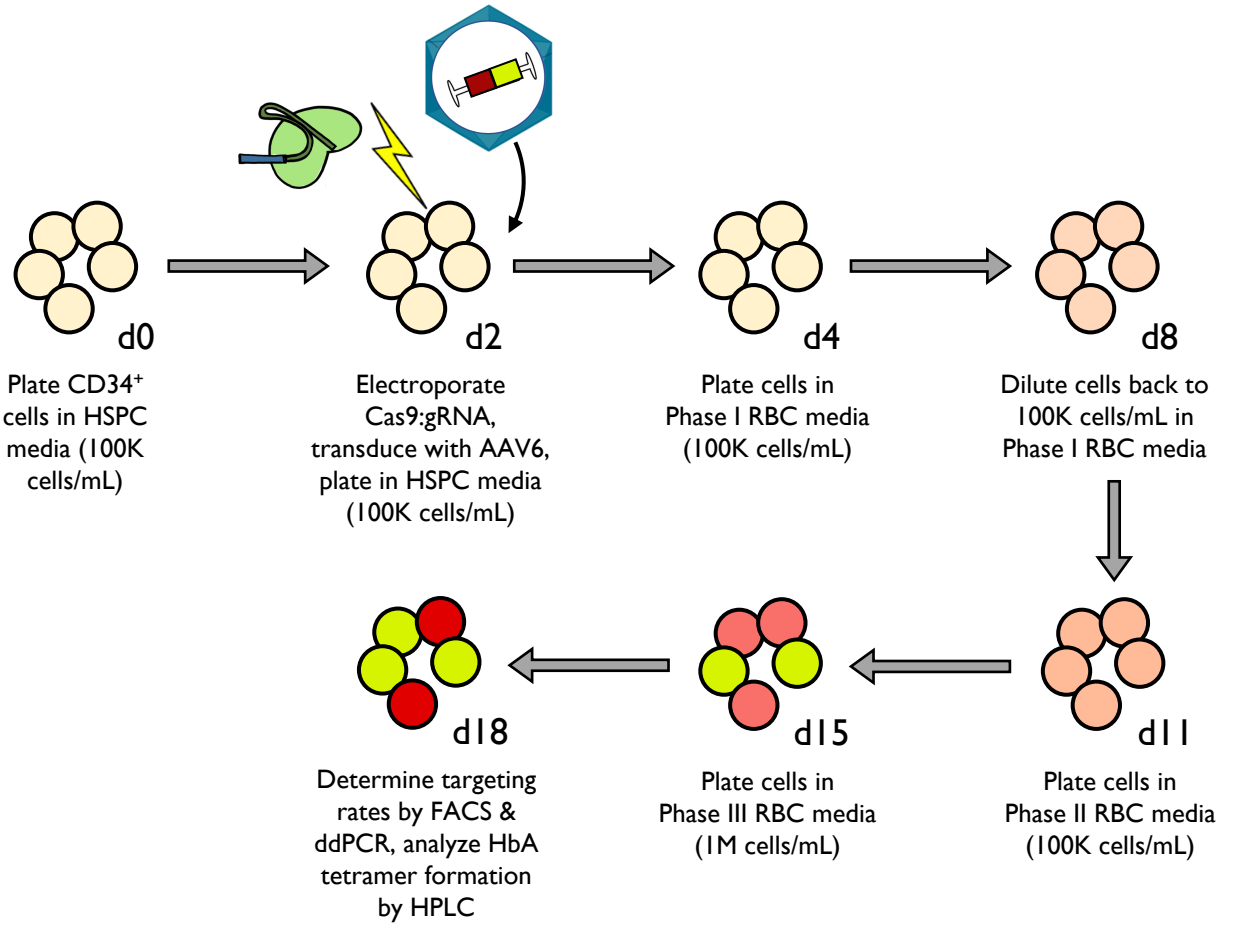
b




c

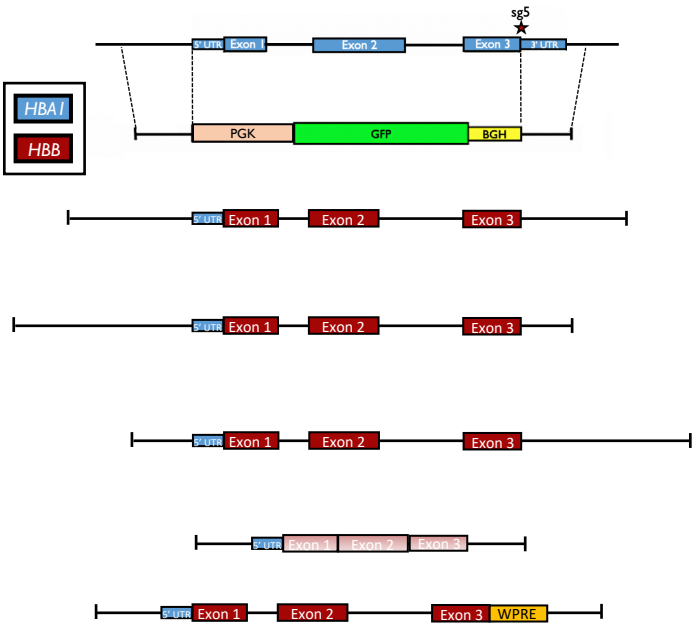


Supplementary Fig. 2: Timeline for targeting α-globin with promoterless reporter vectors




Timeline for targeting of HSPCs with HBB-T2A-YFP integration vectors, differentiation into RBCs, and subsequent analysis.

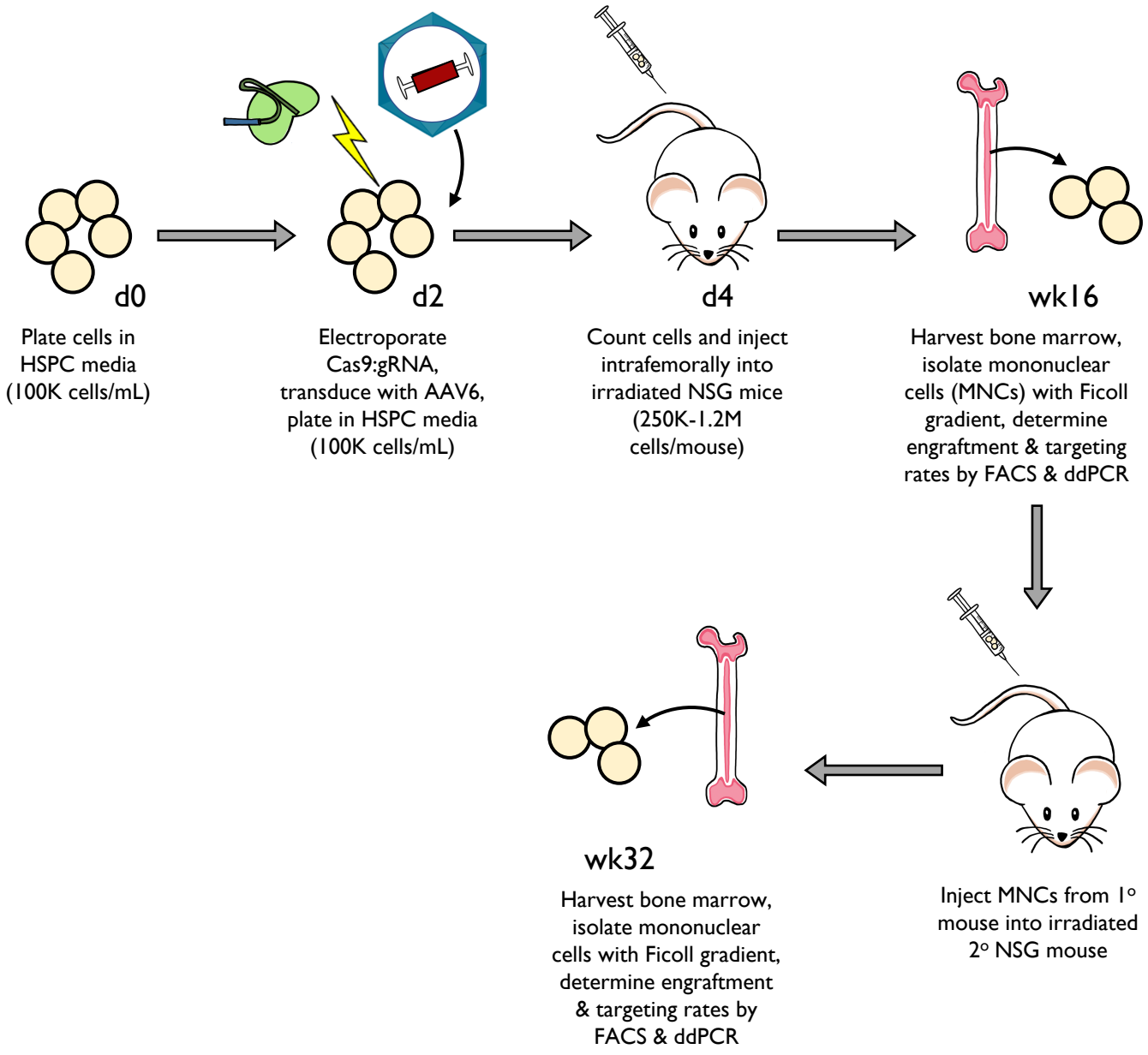
# Supplementary Fig. 3: WGR cassettes screened for development of clinical vector




| Vector | Rationale | Outcome |
|---|---|---|
| **S1** | To determine if PGK promoter is active at *HBA1* locus. | GFP expression was low, indicating that PGK is a weaker promoter at this locus than SFFV and UbC. |
| **S2** | To determine if even longer HAs (1360bp each) would improve targeting frequency. | Longer HAs reduced ddPCR efficiency, making it difficult to accurately determine targeting frequency. However, HPLC showed decreased HgbA production compared to "*HBA1* UTRs long HAs" vector (**Fig. 3E**), possibly due to incompletely packaged integration vectors. |
| **S3** | To determine if longer LHA (2320bp) with normal RHA (400bp) would improve targeting frequency. | Targeting frequency was further improved beyond 880bp HA vector, but HgbA production decreased, possibly due to incompletely packaged integration vectors. |
| **S4** | To determine if longer RHA (2320bp) with normal LHA (400bp) would improve targeting frequency. | Longer HAs decreased ddPCR efficiency, making it difficult to accurately determine targeting frequency. 5' end is too homologous and GC-rich to yield efficient ddPCR results. However, very low levels of HgbA were produced, possibly due to incompletely packaged integration vectors. |
| **S5** | To see if diverged codon, intron-less *HBB* can produce HgbA. | Extremely low levels of HgbA were produced, likely as a result of divergence and/or removal of regulatory regions in introns. |
| **S6** | To determine if 3' WPRE affects transgene expression. | Extremely low levels of HgbA were produced, likely due to disruption caused by WPRE. |

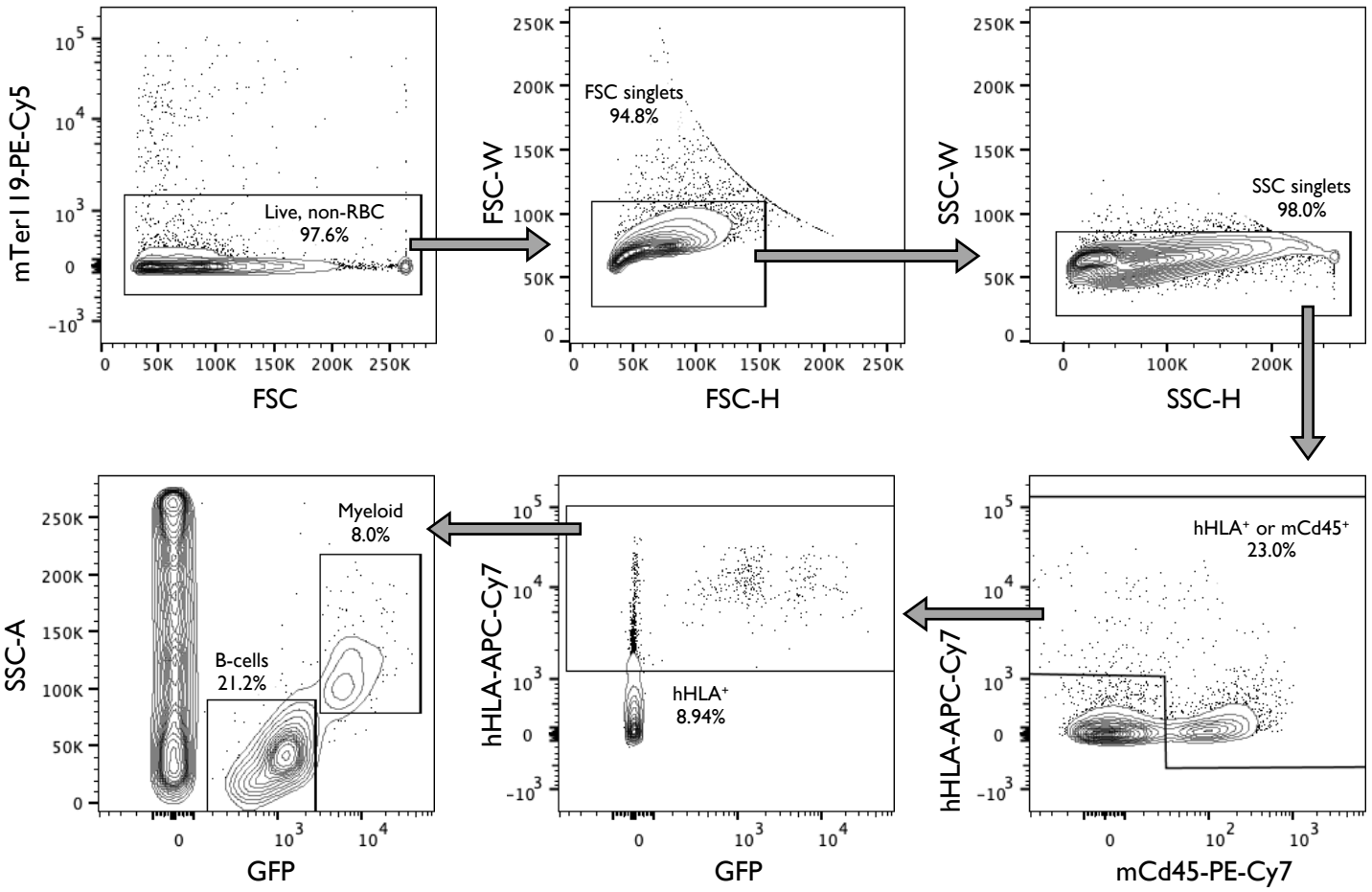
Schematics and corresponding rationale for design as well as eventual outcomes for Vectors S1-6.

Supplementary Fig. 4: Timeline for targeting HSPCs and transplanting into mice



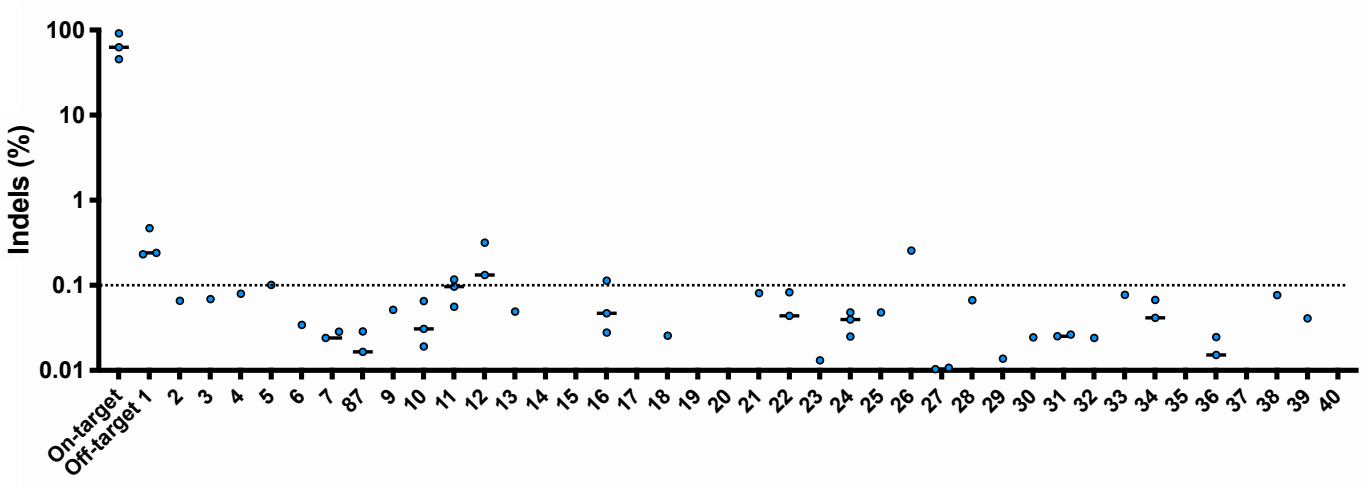

Timeline for targeting HSPCs with *HBB* integration vectors, transplantation into mice (both 1° and 2° engraftment), and subsequent analysis.

Supplementary Fig. 5: Staining and gating scheme used to analyze engraftment and targeting rates of human HSPCs into NSG mice




Representative flow cytometry staining and gating scheme used to analyze targeting and engraftment rates of human HSPCs transplanted into the bone marrow of NSG mice. This sample was targeted with a UbC-GFP integration at the *HBA1* locus. This demonstrates that only human cells (hHLA$^+$) are able to express GFP. Analysis was performed on BD FACS Aria II platform.

a




b

| | Chr | Start | Stop | Strand | Feature | Sequence | Gene |
|---|---|---|---|---|---|---|---|
| On | 16 | 177420 | 177441 | - | 3' UTR | GGCAAGAAGCATGGCCACCGNGG | *HBA1* |
| OT1 | 1 | 116989909 | 116989932 | - | 3' UTR | GCAAGAAGCAAGGGCCACCGNGG | *PTGFRN* |
| OT2 | 3 | 137162017 | 137162038 | - | intergenic | GGCA_GGAGCATGGCCTCCGNGG | - |
| OT3 | 13 | 113077881 | 113077902 | - | intron | GGCAAG_AGCCTGGCCCCCGNGG | *MCF2L* |
| OT4 | 15 | 88807360 | 88807381 | - | intron | GACACGAAGCATGGCCA_CGNGG | *ACAN* |
| OT5 | 1 | 148269712 | 148269734 | - | intergenic | GGCCAGAAGCCTGGCCAACGNGG | - |
| OT6 | 5 | 141389987 | 141390008 | - | exon | GGCCAGAATCATGGCCA_CGNGG | *PCDHGB4* |
| OT7 | 21 | 43355466 | 43355489 | - | intergenic | ACAAGCAGCATGGCCAACCGNGG | - |
| OT8 | 16 | 75538678 | 75538701 | + | 3' UTR | GCATGAAGCATGGATCACCGNGG | *TMEM231* |
| OT9 | 16 | 27302397 | 27302418 | + | intergenic | GGCAAGAAGCATGGAC_CCGNGG | - |
| OT10 | 4 | 13580208 | 13580231 | + | intron | GCAATGCAGCATGGCCACTGNGG | *BOD1L1* |
| OT11 | 15 | 98113993 | 98114015 | + | intergenic | GGTGAGAAGCATGGCCACCANGG | - |
| OT12 | 9 | 131578375 | 131578396 | - | 3' UTR | GGCA_GAAGCCTGGCCACAGNGG | *RAPGEF1* |
| OT13 | 14 | 104082149 | 104082170 | - | intergenic | GGCCAGAAGC_TGGCCACGGNGG | - |
| OT14 | 1 | 41291639 | 41291660 | - | intergenic | GGCA_GAAGCAAGGCCACAGNGG | - |
| OT15 | 20 | 19596791 | 19596812 | - | intron | TGCAA_AAGCATGGCCACCANGG | *SLC24A3* |
| OT16 | 2 | 120361718 | 120361739 | - | intergenic | GG_AAGCAGCATGGCCACCCNGG | - |
| OT17 | 8 | 137651409 | 137651430 | + | intergenic | GTCAAGAAGCATGGCTA_CGNGG | - |
| OT18 | 1 | 110306904 | 110306925 | + | intron | GGCAA_AACCATGGCCACCCNGG | *RBM15-AS1* |
| OT19 | 16 | 85546784 | 85546805 | + | intergenic | GGTAAGAAGC_TGGCCACCANGG | - |
| OT20 | 14 | 89983459 | 89983480 | - | intron | GGCAAGATGCA_GGCCACCANGG | *TDP1* |

| | Chr | Start | Stop | Strand | Sequence | Feature | Gene |
|---|---|---|---|---|---|---|---|
| OT21 | X | 35244229 | 35244251 | - | GGCAAGAAGCCTGGGCACCANGG | intergenic | - |
| OT22 | 17 | 17799461 | 17799482 | - | GGGCAGAAGCATGGCC_CCGNGG | intron | *RAI1* |
| OT23 | 3 | 152261101 | 152261122 | + | GGCAAGAAGC_TGGCAACCTNGG | intergenic | - |
| OT24 | 8 | 114006481 | 114006502 | + | GGCAGGAAGCATGGCC_CCANGG | intergenic | - |
| OT25 | 17 | 43862773 | 43862796 | - | GCAAGAAGGCATGGCCAGGGNGG | 3' UTR | *CD300LG* |
| OT26 | 2 | 132403081 | 132403103 | - | GGCAAGAAGCATGTCCAGTGNGG | intergenic | - |
| OT27 | 8 | 63020684 | 63020705 | + | GGCAAGAAGCATGGCT_CTGNGG | intron | *GGH* |
| OT28 | X | 38930273 | 38930294 | - | GGCAAGAAGCAT_GCCAAAGNGG | intergenic | - |
| OT29 | 3 | 66950667 | 66950688 | - | GGCAAGAAG_ATGGCCATCTNGG | intergenic | - |
| OT30 | 16 | 23501298 | 23501319 | - | GGCAAGAAGCATGGCT_CCANGG | intron | *GGA2* |
| OT31 | 4 | 108864307 | 108864328 | + | GGCAAGAAGCATGGC_AAAGNGG | intron | *COL25A1* |
| OT32 | 1 | 155655358 | 155655380 | + | GGCAAGCAGCTTGGCCACCGNGT | intergenic | - |
| OT33 | 17 | 81371898 | 81371919 | - | GGCCAGAAGC_TGGCCACCGNCG | intergenic | - |
| OT34 | 17 | 2693605 | 2693626 | - | GGCAGGAAGCATG_CCACCGNGC | intron | *CLUH* |
| OT35 | 9 | 119980958 | 119980979 | + | GGCAA_AAGCATGGACACCGNAG | intergenic | - |
| OT36 | 17 | 78854799 | 78854822 | + | GCAAGAAGCAATGGCAACCGNAG | 3' UTR | *TIMP2* |
| OT37 | 2 | 15692810 | 15692831 | - | GGCAAGAAGCAGGGCC_CCGNGA | intron | *LINC01804* |
| OT38 | 8 | 131002525 | 131002547 | - | GGCAGGAAGCATGGCCAC_ANAG | intron | *ADCY8* |
| OT39 | 3 | 81026992 | 81027014 | - | GGCAAGAAGCATTGCCACAGNTG | intron | *LINC02027* |
| OT40 | 21 | 24341987 | 24342010 | - | GCAAGCAAGCATGGCCACCTNAG | intergenic | - |

a

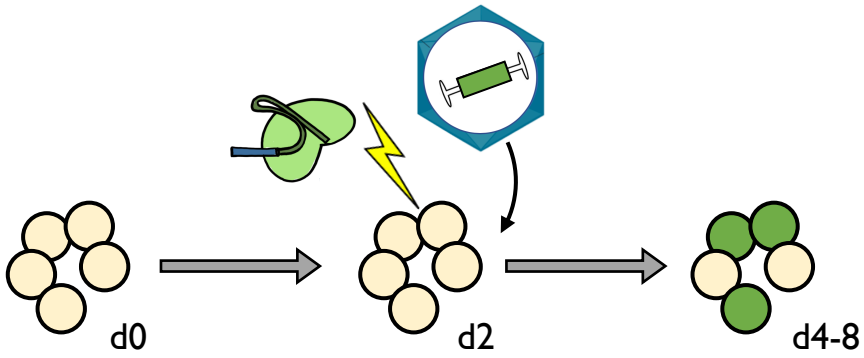
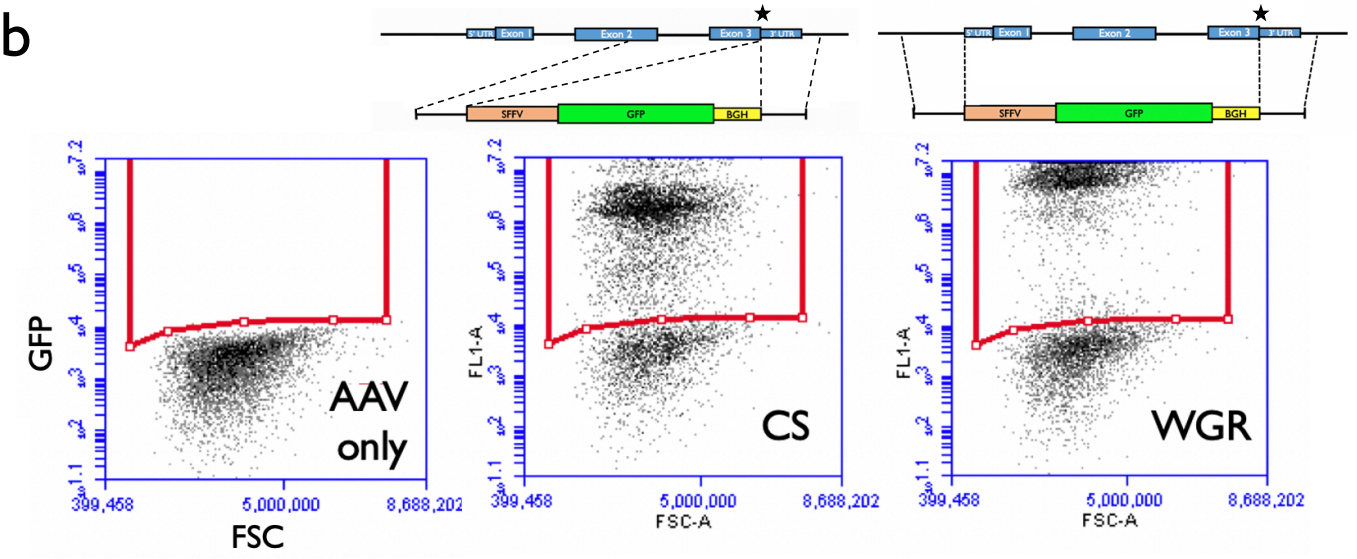



Plate CD34⁺ cells in HSPC media (100K cells/mL)

Electroporate Cas9:gRNA, transduce with AAV6, plate in HSPC media (100K cells/mL)

Determine targeting rates by FACS & ddPCR

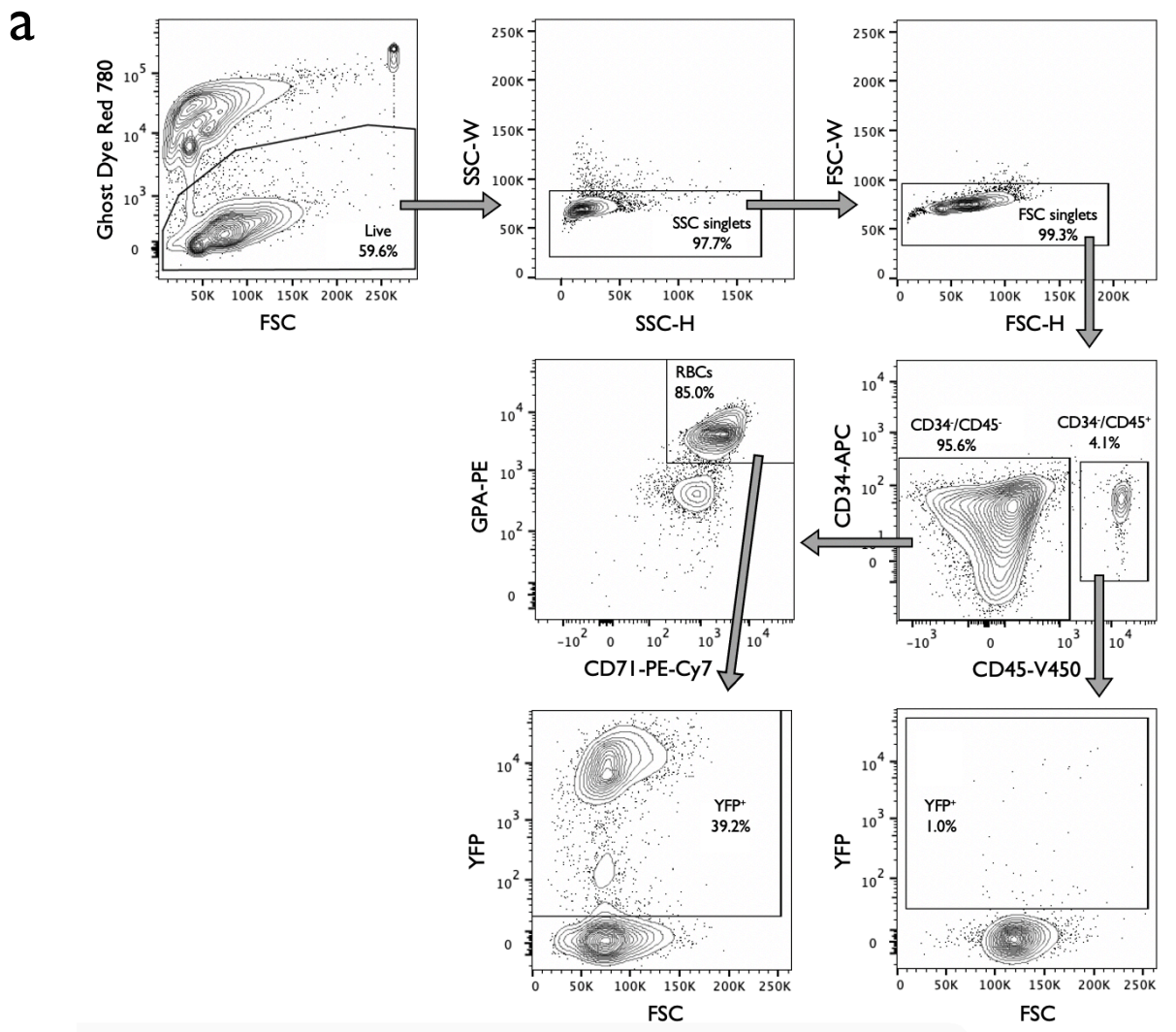
b

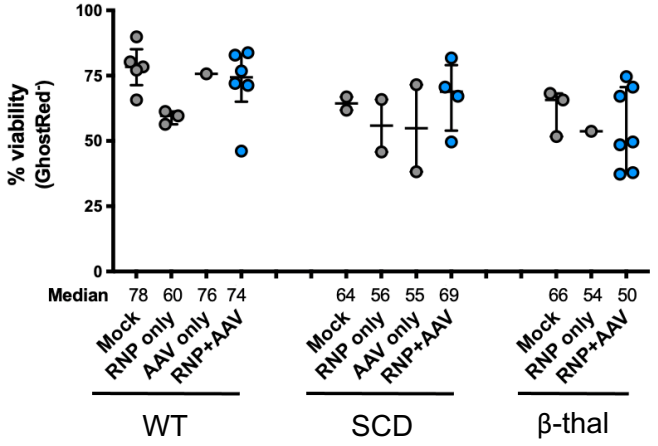



Median MFI = 0 (N = 4)

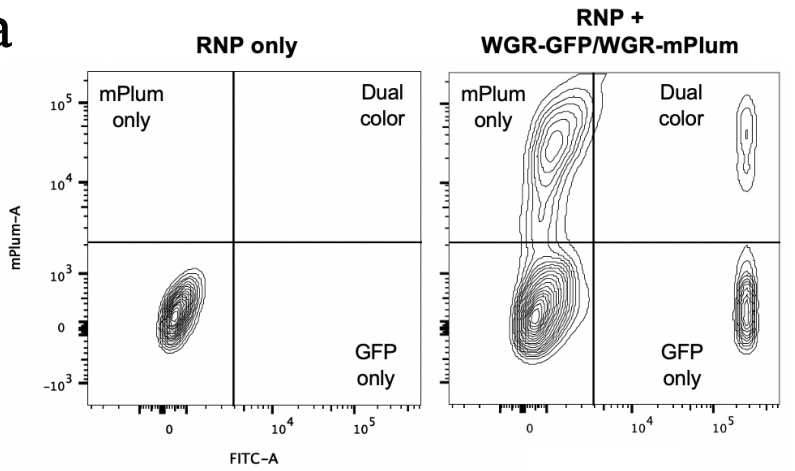
Median MFI = 2.1e6 (N = 12)

Median MFI = 5.0e6 (N = 12)


a
Ghost Dye Red 780
FSC
Live
59.6%
SSC-W
SSC-H
SSC singlets
97.7%
FSC-W
FSC-H
FSC singlets
99.3%
GPA-PE
CD71-PE-Cy7
RBCs
85.0%
CD34-APC
CD45-V450
CD34-/CD45-
95.6%
CD34+/CD45+
4.1%
YFP
YFP+
39.2%
YFP+
1.0%
b
HBA1 UTRs
HBA2 UTRs
HBB UTRs
YFP+
39.2%
YFP+
10.1%
YFP+
30.6%


% viability
(GhostRed⁻)
100
75
50
25
0
Median
78
60
76
74
64
56
55
69
66
54
50
Mock
RNP only
AAV only
RNP+AAV
Mock
RNP only
AAV only
RNP+AAV
Mock
RNP only
RNP+AAV
WT
SCD
β-thal

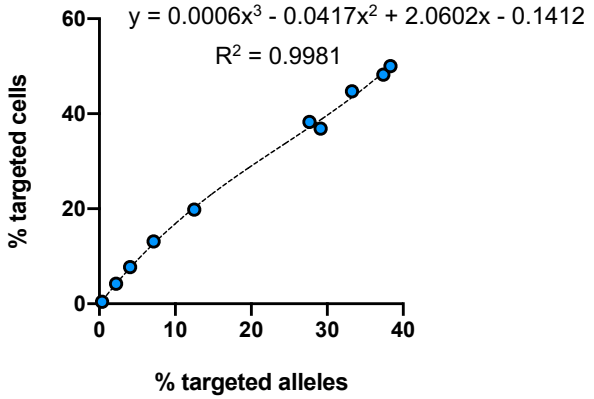
a



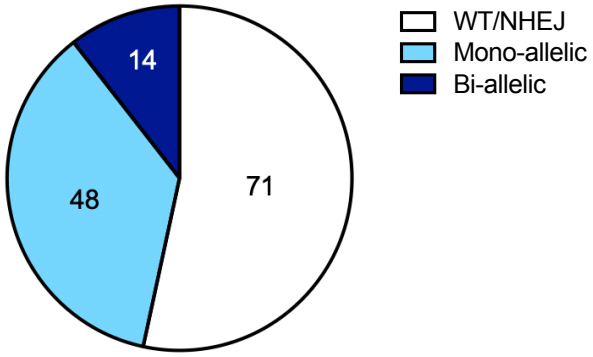

b

| MOI | % mPlum only | % GFP only | % dual color | % edited cells | % edited alleles |
|---|---|---|---|---|---|
| 0 | 0.1 | 0.1 | 0.2 | 0.4 | 0.7 |
| 100 | 3.4 | 0.7 | 0.1 | 4.2 | 4.4 |
| 300 | 5.1 | 2.4 | 0.2 | 7.7 | 8.1 |
| 500 | 17.9 | 11.8 | 8.6 | 38.3 | 55.4 |
| 1000 | 7.3 | 5.2 | 0.6 | 13.1 | 14.3 |
| 1000 | 22.4 | 11.4 | 10.9 | 44.7 | 66.5 |
| 2500 | 12.3 | 4.9 | 2.6 | 19.8 | 25.0 |
| 2500 | 23.9 | 11.0 | 13.3 | 48.2 | 74.8 |
| 5000 | 28.9 | 7.8 | 13.3 | 50.0 | 76.6 |
| 10000 | 18.8 | 7.4 | 10.7 | 36.9 | 58.3 |

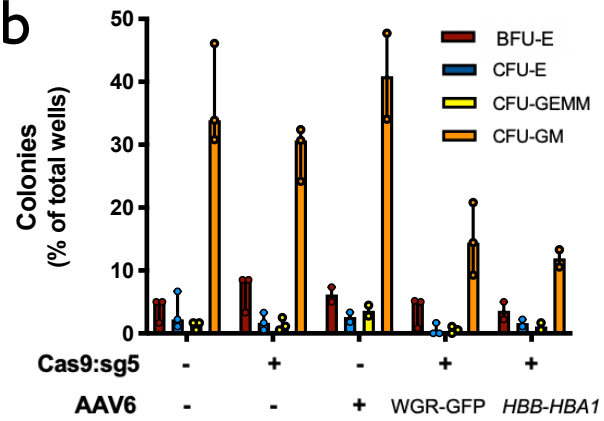
c



a



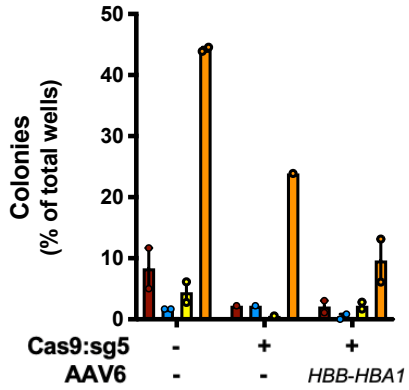

b




c

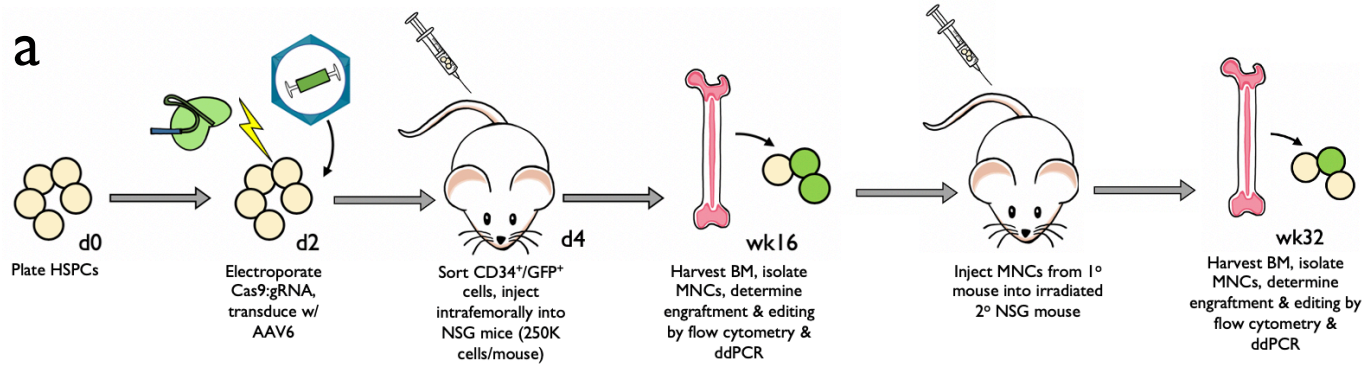
| | Mock | RNP only | AAV only | WGR-GFP | HBB-*HBA1* |
|---|---|---|---|---|---|
| **BFU-E** | 8.6% | 17.3% | 11.6% | 18.7% | 19.7% |
| **CFU-E** | 7.3% | 4.8% | 4.9% | 2.8% | 9.1% |
| **CFU-GEMM** | 2.9% | 3.6% | 6.8% | 2.9% | 6.1% |
| **CFU-GM** | 81.3% | 74.3% | 76.8% | 75.6% | 65.1% |

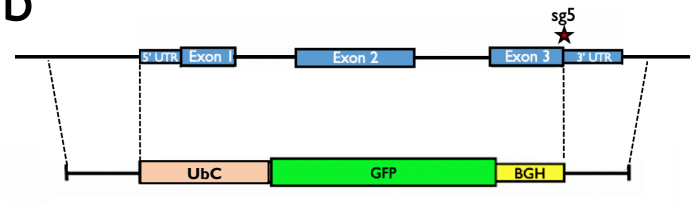
d

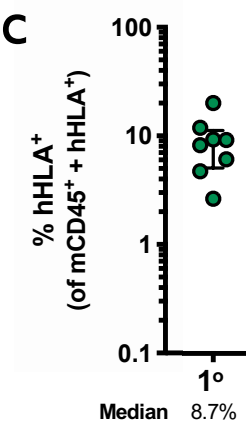


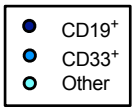

e

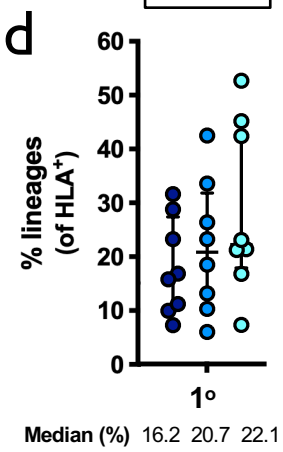
| | Mock | RNP only | *HBA1* UTRs long HAs | Edited alleles |
|---|---|---|---|---|
| **BFU-E** | 13.8% | 7.7% | 14.0% | 31.8% |
| **CFU-E** | 2.9% | 7.7% | 2.1% | 100.0% |
| **CFU-GEMM** | 7.4% | 1.9% | 16.4% | 30.0% |
| **CFU-GM** | 75.9% | 82.7% | 67.5% | 34.8% |

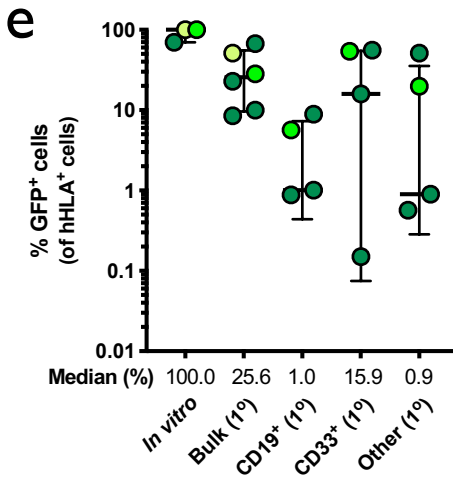

a
d0
Plate HSPCs
d2
Electroporate Cas9:gRNA, transduce w/ AAV6
d4
Sort CD34+/GFP+ cells, inject intrafemorally into NSG mice (250K cells/mouse)
wk16
Harvest BM, isolate MNCs, determine engraftment & editing by flow cytometry & ddPCR
Inject MNCs from 1° mouse into irradiated 2° NSG mouse
wk32
Harvest BM, isolate MNCs, determine engraftment & editing by flow cytometry & ddPCR

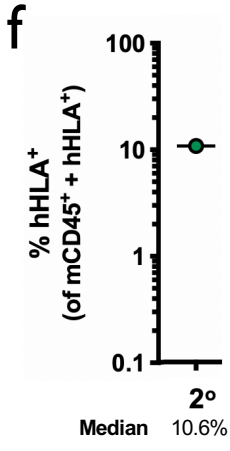

b
sg5
5' UTR
Exon 1
Exon 2
Exon 3
3' UTR
UbC
GFP
BGH

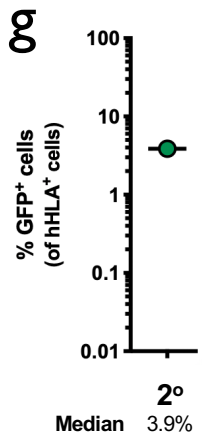

c
% hHLA+ (of mCD45+ + hHLA+)
100
10
1
0.1
1°
Median 8.7%


CD19+
CD33+
Other

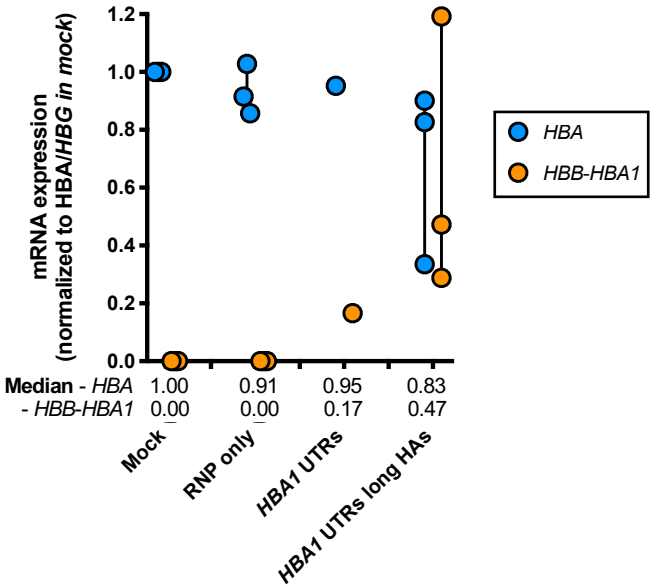

d
% lineages (of HLA+)
60
50
40
30
20
10
0
1°
Median (%) 16.2 20.7 22.1


e
% GFP+ cells (of hHLA+ cells)
100
10
1
0.1
0.01
Median (%) 100.0 25.6 1.0 15.9 0.9
In vitro
Bulk (1°)
CD19+ (1°)
CD33+ (1°)
Other (1°)

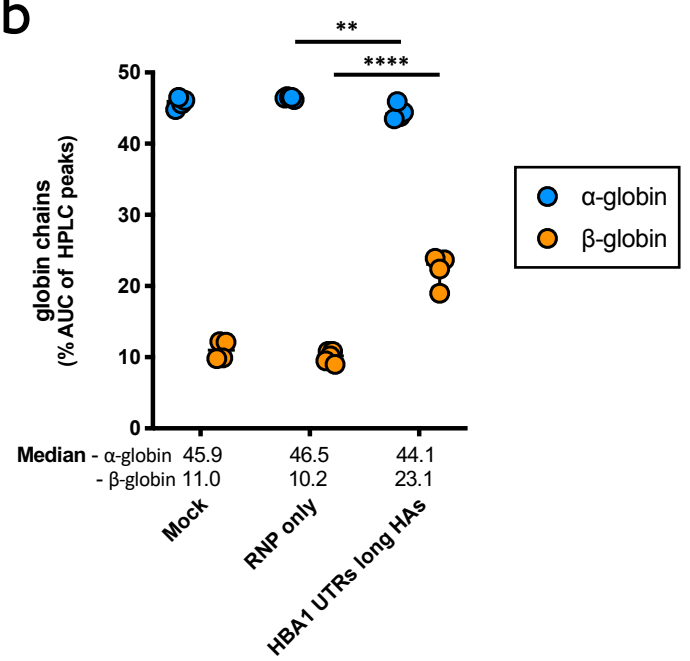

f
% hHLA+ (of mCD45+ + hHLA+)
100
10
1
0.1
2°
Median 10.6%


g
% GFP+ cells (of hHLA+ cells)
100
10
1
0.1
0.01
2°
Median 3.9%

a


mRNA expression
(normalized to HBA/HBG in mock)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
HBA
HBB-HBA1
Median - HBA 1.00 0.91 0.95 0.83
- HBB-HBA1 0.00 0.00 0.17 0.47
Mock
RNP only
HBA1 UTRs
HBA1 UTRs long HAs

b


**
****
globin chains
(% AUC of HPLC peaks)
50
40
30
20
10
0
α-globin
β-globin
Median - α-globin 45.9 46.5 44.1
- β-globin 11.0 10.2 23.1
Mock
RNP only
HBA1 UTRs long HAs

# Inventory of Supporting Information

**Manuscript #:** NMED-A99635B-Z

**Corresponding author name(s):** Matthew H. Porteus

Please complete each of the Inventory Tables below to outline your Extended Data and Supplementary Information items.

There are four sections:

- *Extended Data*
- *Supplementary Information: Flat Files*
- *Supplementary Information: Additional Files*
- *Source Data*

Each section includes specific instructions. Please complete these tables as fully as possible. We ask that you avoid using spaces in your file names, and instead use underscores, i.e.: Smith_ED_Fig1.jpg not Smith ED Fig1.jpg

Please note that titles and descriptive captions will only be lightly edited, so please ensure that you are satisfied with these prior to submission.

If you have any questions about any of the information contained in this inventory, please contact the Editorial Assistant: medicine@us.nature.com



nature research

## 1. Extended Data

### Complete the Inventory below for all Extended Data figures.

- Keep Figure Titles to one sentence only
- Upload your files as 'Figure Files' in our Manuscript Tracking system
- File names should include the Figure Number. i.e.: *Smith_ED_Fig1.jpg*
- Please be sure to include the file extension in the Filename. Note that Extended Data files must be submitted as .jpg, .tif or .eps files *only*, and should be approximately 10MB
- All Extended Data figure legends must be provided in the Inventory below and should not exceed 300 words each *(if possible)*
- Please include Extended Data *ONLY* in this table

| **Figure #** | **Figure title**<br>One sentence only | **Filename**<br>This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg* | **Figure Legend**<br>If you are citing a reference for the first time in these legends, please include all new references in the main text Methods References section, and carry on the numbering from the main References section of the paper. If your paper does not have a Methods section, include all new references at the end of the main Reference list. |
|---|---|---|---|
| **Extended Data Fig. 1** | Analysis of Cas9 sgRNAs targeting α-globin loci | Cromer_ED_Fig1.pdf | A) Table with guide RNA sequences. PAM shown in grey, and differences between *HBA1* and *HBA2* are highlighted in red for each guide.<br>B) Schematic depicting locations of all five guide sequences at genomic loci.<br>C) Representative indel spectrum of *HBA1*-specific sg5 generated by TIDE software. |
| **Extended Data Fig. 2** | Off-target analysis of *HBA1*-specific sg5 | Cromer_ED_Fig2.pdf | A) Summary of rhAmpSeq targeted sequencing results at on-target and 40 most highly predicted off-target sites by COSMID for *HBA1* sg5. Values are indel frequency for RNP treatment after subtraction of indel frequency for Mock |

| | | | |
|---|---|---|---|
| | | | treatment at each locus for each experimental replicate. N=3 biologically independent HSPC donors, though not all values are displayed since some were <0.01% after subtraction of Mock indel frequencies. Bars represent median.<br>B) List of genomic coordinates for forty most highly predicted off-target sites by COSMID for *HBA1* sg5. |
| **Extended Data Fig. 3** | Targeting α-globin with GFP integration vectors | Cromer_ED_Fig3.pdf | A) Timeline for editing and analysis of HSPCs targeted with SFFV-*GFP* integration vectors.<br>B) Depicted are representative flow cytometry images for human HSPCs 14d post-editing. This indicates that WGR integration yields a greater MFI per $GFP^+$ cell than CS integration at the *HBA1* locus. Analysis was performed on BD Accuri C6 platform. Median MFI across all replicates is shown below each flow cytometry image, and schematics of integration vectors are shown above. |
| **Extended Data Fig. 4** | Staining and gating scheme used to analyze editing and differentiation rates of RBCs | Cromer_ED_Fig4.pdf | A) Representative flow cytometry staining and gating scheme for human HSPCs targeted at *HBA1* with HBB-T2A-YFP (*HBA1* UTRs) and differentiated into RBCs. This indicates that only RBCs ($CD34^-$/$CD45^-$/$CD71^+$/$GPA^+$) are able to express the promoterless YFP marker. Analysis was performed on BD FACS Aria II platform.<br>B) Representative flow cytometry images of RBCs ($CD34^-$/$CD45^-$/$CD71^+$/$GPA^+$) derived from HSPCs targeted with *HBA1* UTRs, *HBA2* UTRs, and *HBB* UTRs vector. AAV only controls were used for each vector to establish gating scheme, leading to slight variation in positive/negative cut-offs across images. |
| **Extended Data Fig. 5** | Viability of HSPCs post-editing | Cromer_ED_Fig5.pdf | HSPC viability was quantified 2-4d post-editing by flow cytometry. Depicted are the percentage of cells that stained negative for GhostRed viability dye. All cells were edited with our optimized *HBB* WGR vector using standard conditions (i.e. electroporation of Cas9 RNP+sg5, 5K MOI of AAV, and no AAV wash at 24h). Bars represent median ± interquartile range. |

| | | | |
|---|---|---|---|
| | | | WT: N=5 for mock, N=3 for RNP only, N=1 for AAV only, and N=6 for RNP+AAV treatment group; SCD: N=2 for each treatment group with exception of RNP+AAV with N=4; β-thal: N=3 for mock, N=1 for RNP only, and N=7 for RNP+AAV treatment group. |
| **Extended Data Fig. 6** | Relationship between % edited alleles and % edited cells | Cromer_ED_Fig6.pdf | A) Representative flow cytometry plots of HSPCs simultaneously targeted at *HBA1* with GFP (shown in Fig. 1c) and mPlum integration cassettes.<br>B) Table showing % of populations targeted with GFP only, mPlum only, and both colors. Percent of edited cells was then converted to % edited alleles by the following equation: (total % targeted cells + (% dual color)*2)/2 = total % targeted alleles.<br>C) Percent edited cells is plotted against % edited alleles for data shown in panel B. A polynomial regression ($R^2$ = 0.9981) was used to determine an equation to convert between % edited alleles and % edited cells. |
| **Extended Data Fig. 7** | Colony-forming ability of edited HSPCs | Cromer_ED_Fig7.pdf | A) Distribution of genotypes of methylcellulose colonies displayed in Panels B and D. Numbers of clones corresponding to each category are included in the pie chart.<br>B) *In vitro* (pre-engraftment) live $CD34^+$ HSPCs from healthy donors were single-cell sorted into 96-well plates containing semisolid methylcellulose media for colony forming assays. 14d post-sorting cells were analyzed for morphology. Depicted are number of colonies formed for each lineage (CFU-E = erythroid lineage; CFU-GEMM = multi-lineage; or CFU-GM = granulocyte/macrophage lineage) divided by the total number of wells available for colonies. Columns represent median $\pm$ interquartile range. N=3 experimental replicates with a minimum of 3 96-well methylcellulose-coated plates for mock, RNP only, and WGR-GFP AAV6 treatment groups; N=2 for AAV only and *HBB-HBA1* AAV6 treatment groups. |

| | | | |
|---|---|---|---|
| | | | C) Percent distribution of each lineage among all colonies for each treatment for Panel B.<br>D) As above, *in vitro* (pre-engraftment) live $CD34^+$ β-thalassemia HSPCs were sorted into 96-well plates for colony forming assays. Depicted are number of colonies formed for each lineage (B = BFU-E and C = CFU-E (erythroid lineage); GE = CFU-GEMM (multi-lineage); or GM = CFU-GM (granulocyte/macrophage lineage)) divided by the total number of wells available for colonies. Columns represent median ± interquartile range. For Mock and RNP+AAV, N=2 experimental replicates with a minimum of 3 96-well methylcellulose-coated plates for each treatment; N=1 experimental replicate with 3 plates for RNP only treatment.<br>E) Percent distribution of each lineage among all colonies for each treatment for Panel D. |
| **Extended Data Fig. 8** | Engraftment into NSG mice of human HSPCs targeted with GFP at α-globin locus | Cromer_ED_Fig8.pdf | A) Timeline for targeting of HSPCs with UbC-GFP integration vector, transplantation into mice (both $1^o$ and $2^o$ engraftment), and subsequent analysis.<br>B) AAV6 DNA repair donor design schematic to introduce a UbC-GFP-BGH integration is depicted at the *HBA1* locus.<br>C) 16 weeks after bone marrow transplantation of targeted human $CD34^+$ HSPCs into NSG mice, bone marrow was harvested and rates of engraftment were determined ($1^o$). Depicted is the percentage of $mTerr119^-$ cells (non-RBCs) that were $hHLA^+$ from the total number of cells that were either $mCd45^+$ or $hHLA^+$. Bars represent median ± interquartile range. N=8 biologically independent NSG mouse transplantations.<br>D) Among engrafted human cells, the distribution among $CD19^+$ (B-cell), $CD33^+$ (myeloid), or other (i.e. HSPC/RBC/T/NK/Pre-B) lineages are indicated. Bars |

| | | | |
|---|---|---|---|
| | | | represent median ± interquartile range. N=8 biologically independent NSG mouse transplantations.<br>E) Percentage of $GFP^+$ cells among pre-transplantation (in vitro, post-sorting) and successfully engrafted populations, both bulk HSPCs and among lineages. Bars represent median ± interquartile range. N=3 independent HSPC donors from *in vitro* experiments that were transplanted into N=6 individual NSG mice, from which N=4 individual mice were lineage sorted and analyzed. Various green shades correspond to each particular HSPC donor.<br>F) Following primary engraftments, engrafted human cells were transplanted a second time into the bone marrow of NSG mice. 16 weeks post-transplantation, bone marrow was harvested and rates of engraftment were determined ($2^o$). Depicted is the percentage of $mTer119^-$ cells (non-RBCs) that were $hHLA^+$ from the total number of cells that were either $mCd45^+$ or $hHLA^+$. N=1 NSG mouse transplantation.<br>G) Percentage of $GFP^+$ cells among successfully engrafted population from the secondary transplant depicted in Panel F. N=1 NSG mouse transplantation. |
| **Extended Data Fig. 9** | Characterization of targeted β-thalassemia HSPCs | Cromer_ED_Fig9.pdf | A) Following differentiation of targeted HSPCs into RBCs, mRNA was harvested and converted into cDNA. Expression of *HBA* (does not distinguish between *HBA1* and *HBA2*) and *HBB* transgene were normalized to *HBG* expression. Bars represent median ± interquartile range. N=3 biologically independent editing experiments for all treatment groups with exception of *HBA1* UTRs with N=1. |

| | | | |
|---|---|---|---|
| | | | B) Summary of reverse-phase globin chain HPLC results showing % AUC of β-globin and α-globin. Bars represent median ± interquartile range. Bars represent median ± interquartile range. N=3 biologically independent erythroid differentiation experiments for all treatment groups with exception of RNP only with N=5. **: P<0.005; ***: P<0.0001 using unpaired two-tailed t test without adjustment for multiple comparisons. |
| **Extended Data Fig. 10** | | | |

***Delete rows as needed to accommodate the number of figures (10 is the maximum allowed).***

## 2. Supplementary Information:

## A. Flat Files

**Complete the Inventory below for all additional textual information and any additional Supplementary Figures, which should be supplied in one combined PDF file.**

- **Row 1:** A combined, flat PDF containing any Supplementary Text, Discussion, Notes, Additional Supplementary Figures, Supplementary Protocols, simple tables, and all associated legends. Only one such file is permitted.
- **Row 2:** Nature Research's Reporting Summary; if previously requested by the editor, please provide an updated Summary, fully completed, without any mark-ups or comments. **(Reporting Summaries are not required for all manuscripts.)**

| Item | Present? | Filename | A brief, numerical description of file contents. |
|---|---|---|---|

| | | | |
|---|---|---|---|
| | | **This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf** | **i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*** |
| **Supplementary Information** | Yes | NMED-A99635B-Z_SI_Cromer.pdf | Supplementary Figures 1-5 |
| **Reporting Summary** | Yes | NMED-A99635B-Z_Reporting_summary_Cromer.pdf | |

## B. Additional Supplementary Files

**Complete the Inventory below for all additional Supplementary Files that cannot be submitted as part of the Combined PDF.**

- Do not list Supplementary Figures in this table (see section 2A)
- Where possible, include the title and description within the file itself
- Spreadsheet-based tables and data should be combined into a workbook with multiple tabs, not submitted as individual files.
- Compressed files are acceptable where necessary. ZIP files are preferred.
- Please note that the *ONLY* allowable types of additional Supplementary Files are:
    - Supplementary Tables ○ Supplementary Audio ○ Supplementary Videos ○ Supplementary Software
    - Supplementary Data, for example: raw NMR Data, Cryo-EM Data, Computational Data, Crystallographic Data, etc.

| **Type** | **Number**<br>If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc. | **Filename**<br>This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_ Supplementary_Video_1.mov* | **Legend or Descriptive Caption**<br>Describe the contents of the file |
|---|---|---|---|
| Choose an item. | | | |
| Choose an item. | | | |
| Choose an item. | | | |
| Choose an item. | | | |
| Choose an item. | | | |
| Choose an item. | | | |

***Add rows as needed to accommodate the number of files.***

## 3. Source Data

## Complete the Inventory below for all Source Data files.

- Acceptable types of Source Data for Main Figures and Extended Data Figures are:
  - Statistical Source Data
    - Plain Text (ASCII, TXT) or Excel formats only
    - One file for each relevant Figure, containing all source data
  - Full-length, unprocessed Gels or Blots
    - JPG, TIF, or PDF formats only
    - One file for each relevant Figure, containing all supporting blots and/or gels
- 'Source Data' is only allowed for Main Figures and Extended Data Figures.
  - Include Unprocessed Gels or Blots for Supplementary Figures as additional Supplementary Figures.
  - Include Statistical Source Data for Supplementary Figures as 'Supplementary Data' files and list them in section 2B, above.
  - Please see this example of Source Data in a publication.

| **Parent Figure or Table** | **Filename**<br>**This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_SourceData_Fig1.xls,* or *Smith_ Unmodified_Gels_Fig1.pdf*** | **Data description**<br>**i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.** |
|---|---|---|
| **Source Data Fig. 1** | Cromer_SourceData_Fig1.xls | Statistical Source Data |
| **Source Data Fig. 2** | Cromer_SourceData_Fig2.xls | Statistical Source Data |
| **Source Data Fig. 3** | Cromer_SourceData_Fig3.xls | Statistical Source Data |
| **Source Data Fig. 4** | Cromer_SourceData_Fig4.xls | Statistical Source Data |
| **Source Data Fig. 5** | Cromer_SourceData_Fig5.xls | Statistical Source Data |
| **Source Data Fig. 6** | Cromer_SourceData_Fig6.xls | Statistical Source Data |
| **Source Data Fig. 7** | | |
| **Source Data Fig. 8** | | |

| | | |
|---|---|---|
| **Source Data Extended Data Fig. 1** | | |
| **Source Data Extended Data Fig. 2** | Cromer_SourceData_ED2.xls | Statistical Source Data |
| **Source Data Extended Data Fig. 3** | | |
| **Source Data Extended Data Fig. 4** | | |
| **Source Data Extended Data Fig. 5** | Cromer_SourceData_ED5.xls | Statistical Source Data |
| **Source Data Extended Data Fig. 6** | | |
| **Source Data Extended Data Fig. 7** | | |
| **Source Data Extended Data Fig. 8** | Cromer_SourceData_ED8.xls | Statistical Source Data |
| **Source Data Extended Data Fig. 9** | Cromer_SourceData_ED9.xls | Statistical Source Data |
| **Source Data Extended Data Fig. 10** | | |