ARTICLE

doi:10.1038/nature20134

CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells

Daniel P. Dever[1]*, Rasmus O. Bak[1]*, Andreas Reinisch[2], Joab Camarena[1], Gabriel Washington[1], Carmencita E. Nicolas[1], Mara Pavel-Dinu[1], Nivi Saxena[1], Alec B. Wilkens[1], Sruthi Mantri[1], Nobuko Uchida[3]†, Ayal Hendel[1], Anupama Narla[4], Ravindra Majeti[2], Kenneth I. Weinberg[1] & Matthew H. Porteus[1]

The β-haemoglobinopathies, such as sickle cell disease and β-thalassaemia, are caused by mutations in the β-globin (*HBB*) gene and affect millions of people worldwide. *Ex vivo* gene correction in patient-derived haematopoietic stem cells followed by autologous transplantation could be used to cure β-haemoglobinopathies. Here we present a CRISPR/Cas9 gene-editing system that combines Cas9 ribonucleoproteins and adeno-associated viral vector delivery of a homologous donor to achieve homologous recombination at the *HBB* gene in haematopoietic stem cells. Notably, we devise an enrichment model to purify a population of haematopoietic stem and progenitor cells with more than 90% targeted integration. We also show efficient correction of the Glu6Val mutation responsible for sickle cell disease by using patient-derived stem and progenitor cells that, after differentiation into erythrocytes, express adult β-globin (HbA) messenger RNA, which confirms intact transcriptional regulation of edited *HBB* alleles. Collectively, these preclinical studies outline a CRISPR-based methodology for targeting haematopoietic stem cells by homologous recombination at the *HBB* locus to advance the development of next-generation therapies for β-haemoglobinopathies.

Allogeneic haematopoietic stem-cell transplantation demonstrates that transplantation of haematopoietic stem cells (HSCs) with only a single wild-type *HBB* gene can cure the β-haemoglobinopathies. However, this transplantation technique is limited because of graft-versus-host disease and a lack of immunologically matched donors. An alternative to using allogeneic HSCs to cure the β-haemoglobinopathies is to use homologous recombination to modify the *HBB* gene directly in autologous HSCs[1,2]. In 1985, Smithies and colleagues[3] were able to modify the human *HBB* gene by homologous recombination in a human embryonic carcinoma cell line, albeit at an extremely low frequency (10^{-6}). The subsequent discoveries that a site-specific DNA double-strand break (DSB) could stimulate homologous-recombination-mediated correction of a reporter gene and that engineered nucleases could be used to induce this DSB, formed the foundation of using homologous-recombination-mediated genome editing using engineered nucleases to modify the *HBB* gene directly[4,5]. The ease of engineering as well as the robust activity of the CRISPR/Cas9 RNA-guided endonuclease system makes it a promising tool to apply to the continuing challenge of developing effective and safe homologous-recombination-mediated genome editing to cure β-haemoglobinopathies[6,7].

The CRISPR/Cas9 complex consists of the Cas9 endonuclease and a 100-nucleotide single-guide RNA (sgRNA). Target identification relies first on recognition of a 3-base-pair protospacer adjacent motif (PAM) and then on hybridization between a 20-nucleotide stretch of the sgRNA and the DNA target site, which triggers Cas9 to cleave both DNA strands[8]. DSB formation activates two highly conserved repair mechanisms: canonical non-homologous end-joining (NHEJ) and homologous recombination[9]. Through iterative cycles of break and NHEJ repair, insertions and/or deletions (INDELs) can be created at the site of the break. By contrast, genome editing by homologous recombination requires the delivery of a DNA donor molecule to serve as a homologous template, which the cellular recombination machinery uses to repair the break by a 'copy and paste' method[10]. For gene-editing purposes, the homologous recombination pathway can be exploited to make precise nucleotide changes in the genome[4]. One of the key features of precise genome editing, in contrast to viral-vector-based gene transfer methods, is that endogenous promoters, regulatory elements, and enhancers can be preserved to mediate spatiotemporal gene expression[1,11–13]. The CRISPR/Cas9 system is highly effective at stimulating DSBs in primary human haematopoietic stem and progenitor cells (HSPCs) when the sgRNA is synthesized with chemical modifications, and then electroporated into cells[14].

HSCs have the ability to repopulate an entire haematopoietic system[15], and several genetic[16–18] and acquired[19] diseases of the blood could potentially be cured by genome editing of HSCs. Recent studies have demonstrated efficient targeted integration in HSPCs by combining zinc-finger nuclease (ZFN) expression with exogenous homologous recombination donors delivered via single-stranded oligonucleotides[20], integrase-defective lentiviral vectors[21], or recombinant adeno-associated viral vectors of serotype 6 (rAAV6)[22,23]. While showing very positive results *in vitro*, collectively, these studies also suggested that targeting HSCs by homologous recombination at disease-causing loci is difficult in clinically relevant HSPCs.

In this study, we achieve efficient homologous-recombination-mediated editing frequencies at the *HBB* locus in $CD34^+$ HSPCs derived from mobilized peripheral blood using Cas9 ribonucleoproteins combined with rAAV6 homologous donor delivery. In brief, we demonstrate: (1) Cas9- and rAAV6-mediated *HBB* targeting in HSCs characterized by the identification of modified human cells in

[1]Department of Pediatrics, Stanford University, Stanford, California 94305, USA. [2]Department of Medicine, Division of Hematology, Cancer Institute, and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California 94305, USA. [3]Stem Cells, Inc. 7707 Gateway Blvd., Suite 140, Newark, California 94560, USA. [4]Division of Hematology/Oncology, Department of Pediatrics, Stanford University School of Medicine, Stanford, California 94035, USA. †Present address: Institute of Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California 94305, USA.
*These authors contributed equally to this work.

secondary transplants of immunodeficient mice; (2) efficient correction of the sickle cell disease (SCD)-causing Glu6Val mutation in HSPCs derived from several patients with SCD; and (3) development of a purification scheme using either fluorescence-activated cell sorting (FACS) or magnetic bead enrichment to create HSPC populations in which more than 85% of the cells have been modified by homologous-recombination-mediated targeted integration. This purification can be performed early in the manufacturing process when HSCs are still preserved, and may prove valuable in a clinical setting for removing untargeted HSPCs that will be in competition with homologous-recombination-edited HSPCs for engraftment and re-population after transplantation.

CRISPR/Cas9-mediated *HBB* gene editing in HSPCs

We confirmed high transduction of HSPCs using a self-complementary AAV6 (scAAV6) with an SFFV-GFP expression cassette[24] (Extended Data Fig. 1a). *HBB*-specific single-stranded AAV6 (ssAAV6) vectors were then produced containing SFFV-GFP flanked by arms homologous to *HBB* (Fig. 1a). To achieve gene editing at *HBB*, we used two different CRISPR platforms, which we have previously shown to be highly active in primary cells[14]. Both platforms use sgRNAs chemically modified at both termini with 2′-*O*-methyl-3′-phosphorothioate, and are delivered either in conjunction with Cas9 mRNA or as a ribonucleoprotein (RNP) complex. Both platforms yielded high INDEL frequencies when electroporated into HSPCs, with the RNP showing higher activity (Fig. 1b). By supplying ssAAV6 *HBB* donors after electroporation of Cas9 RNP, we achieved stable green fluorescent protein (GFP) expression in an average of 29% of HSPCs (Fig. 1c). We observed lower efficiencies using the mRNA platform (15%) (Fig. 1c). Cytotoxicity and off-target cleavage activity[14] was significantly decreased using the RNP system (Extended Data Fig. 1b–d).

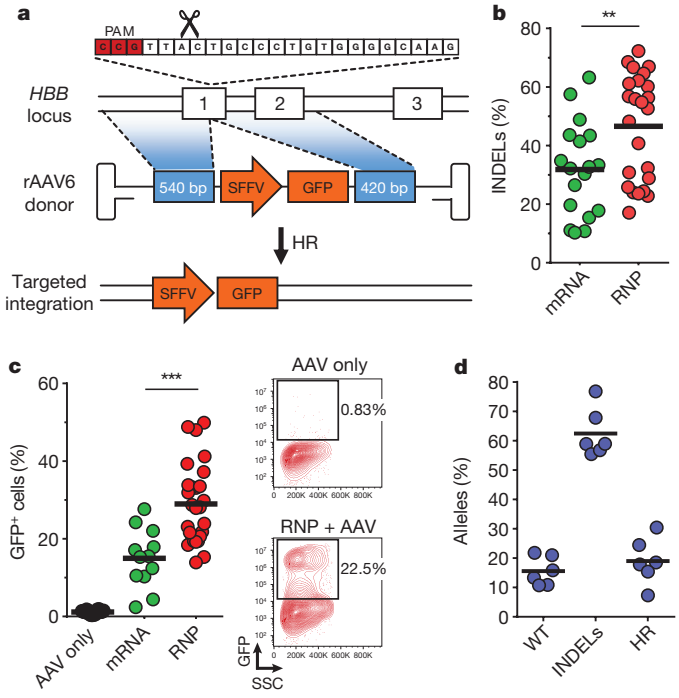


Figure 1 | CRISPR/Cas9 and rAAV6-mediated targeted integration at the *HBB* locus in human CD34$^+$ HSPCs. a, Schematic of targeted genome editing at the *HBB* locus using CRISPR/Cas9 and rAAV6. Site-specific DSBs are created by Cas9 (scissors) mainly between nucleotides 17 and 18 of the 20-bp target site, which is followed by the 5′-NGG-3′ PAM sequence (red). A DSB stimulates homologous recombination (HR) using the rAAV6 homologous donor as a repair template. White boxes, *HBB* exons; blue boxes, homology arms; orange boxes, SFFV-GFP-polyA expression cassette. **b**, HSPCs were electroporated with mRNA or RNP CRISPR system, and INDELs were analysed via TIDE software (n = number of data points within each group, all from different mobilized peripheral blood or cord blood donors; $^{**}P < 0.01$, unpaired t-test. Bars represent means). **c**, HSPCs electroporated as above and transduced with *HBB*-specific rAAV6s were analysed by flow cytometry 18–21 days after electroporation when GFP levels were found to be constant. Left, percentage of GFP$^+$ HSPCs. Right, representative FACS plots (n = number of data points within each group, all from different donors; $^{***}P < 0.001$, unpaired t-test. Bars represent means). **d**, HSPCs were treated as above but targeted with a rAAV6 Glu6Val donor. Frequencies of allele types were quantified by sequencing of a total of 600 clones from TOPO-cloned in-out PCRs ($n = 6$, all from different cord blood or bone marrow donors. Bars represent means). WT, wild type.

Because AAV genomes can be captured at the site of an off-target DSB via NHEJ[22,25,26], we performed experiments mismatching a nuclease with a non-homologous donor to see whether this occurs with our methodology. While approximately 20% of cells that received matched *HBB* nuclease and *HBB* donor maintained GFP expression after 18 days in culture, *IL2RG* nuclease combined with *HBB* donor resulted in 0.8% GFP$^+$ cells (Extended Data Fig. 1e–g), which was not significantly higher than the 0.5% GFP$^+$ cells observed when using the *HBB* AAV donor alone. These results demonstrate that capture of the *HBB* donor is an infrequent event using this system in human HSPCs. Furthermore, the observed GFP expression in 0.5% of HSPCs with the AAV donor alone suggests that random integration of rAAV6 is limited in human HSPCs. In fact, previous reports have shown AAV-mediated targeted integration without a DSB[27,28], and thus, these GFP$^+$ cells may be the result of on-target events.

SCD is caused by a single nucleotide mutation (A to T), which changes an amino acid (Glu6Val) at codon 6 of the *HBB* gene[29]. We created a 4.5-kb rAAV6 donor template that would introduce the Glu6Val mutation along with six other silent single nucleotide polymorphisms (SNPs) to prevent Cas9 re-cutting of homologous recombination alleles (Extended Data Fig. 2a–c). Using this rAAV6 donor with Cas9 RNP delivery, we measured an average allelic modification frequency of 19% in six different HSPC donors (Fig. 1d). These results confirm that the combined use of CRISPR with rAAV6 can precisely change the nucleotide at the position of the mutation that causes SCD.

Early enrichment of *HBB*-targeted HSPCs

Because HSCs differentiate and progressively lose their long-term repopulating capacity after culturing, the identification of gene-edited HSPCs early in the manufacturing process would be of great use. In experiments using a GFP-expressing rAAV6 donor we observed that while HSPCs receiving only the rAAV6 donor expressed low levels of GFP, HSPCs that also received Cas9 RNP generated a population at day 4 after electroporation that expressed much higher levels of GFP (Fig. 2a, left). We proposed that this GFP^{high} population was enriched for *HBB*-targeted cells. We therefore sorted and cultured the GFP^{high} population as well as the GFP^{low} and GFP$^-$ populations. While sorted GFP^{low} and GFP$^-$ populations were, respectively, approximately 25% and 1% GFP$^+$ after 15–20 days in culture, the GFP^{high} population was more than 95% GFP$^+$, suggesting that this population was indeed *HBB*-targeted (Fig. 2a, right). Linear regression showed that the percentage of GFP^{high}-expressing HSPCs at day 4 after electroporation strongly correlates with the total percentage of GFP$^+$ cells at day 18 (Extended Data Fig. 3).

To confirm that the GFP^{high} population was enriched for on-target integration, we used 'in-out PCR' (one primer binding the *HBB* locus outside the region of the homology arm and the other binding the integrated insert) to determine integration frequencies and allelic distribution in methylcellulose clones derived from the GFP^{high} population (95 clones). A total of 92% of clones had a targeted integration, with 38% containing biallelic integrations (Fig. 2b, Extended Data Fig. 4a–c and Supplementary Fig. 1a). This assay generates colonies from progenitor cells and the biallelic integration frequency could be different in HSCs. Nonetheless, these data show that the log-fold

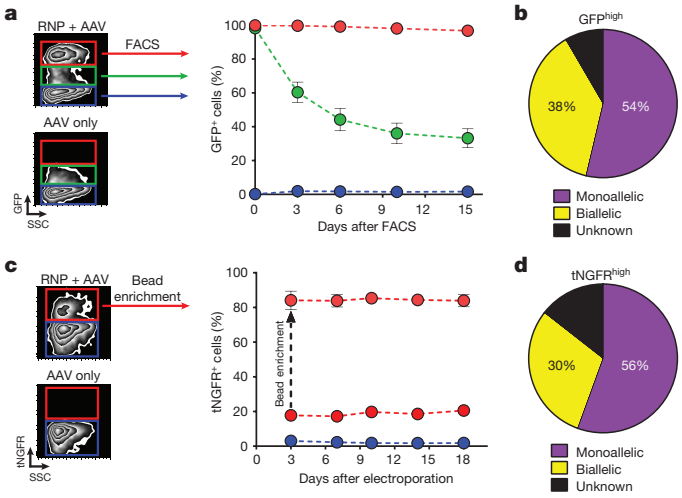


Figure 2 | Enrichment of *HBB*-targeted HSPCs using FACS and magnetic bead-based technologies. a, Left, representative FACS plots highlight the GFP^{high} population (red) generated by the addition of Cas9 RNP. Right, *HBB*-targeted HSPCs from GFP^{high} (red), GFP^{low} (green) and GFP^{-} (blue) fractions were sorted and monitored for GFP expression. Error bars represent s.e.m. ($n = 11$, all from unique mobilized peripheral blood or cord blood donors). **b**, PCR was performed on methylcellulose colonies from GFP^{high} HSPCs to detect targeted integration at the 3′ end. **c**, Left, representative FACS plots highlight the tNGFR^{high} population (red) generated by the addition of Cas9 RNP. Right, tNGFR^{high} (red) HSPCs were enriched using anti-CD271 (also known as LNGFR) magnetic microbeads and cultured for 18 days while monitoring tNGFR expression. Error bars represent s.e.m. ($n = 5$, all from unique cord blood donors). **d**, PCR was performed on tNGFR^{high}-derived methylcellulose colonies to detect targeted integrations at the 5′ end.

transgene expression shift after rAAV6 and RNP delivery is due to homologous recombination at the intended locus, and that the shift allows FACS-based enrichment of *HBB*-targeted HSPCs.

Although GFP is not a clinically relevant reporter gene, the truncated nerve growth factor receptor (tNGFR), in which the cytoplasmic intracellular signalling domain is removed, could be used to enrich for targeted HSPCs. tNGFR is expressed on the cell surface, thereby allowing antibody-mediated detection of gene marking, and has already been used in human clinical trials[30–33]. We examined whether we could enrich *HBB*-targeted HSPCs using tNGFR magnetic-bead-based separation technology. HSPCs that received RNP and the rAAV6 donor (with a tNGFR expression cassette) produced a tNGFR^{high} population that was not present in cells transduced with rAAV6 alone (Fig. 2c, left). We then enriched tNGFR^{high} cells using anti-NGFR magnetic microbeads and after 18 days in culture, on average 84% of HSPCs were tNGFR^{+} (Fig. 2c, right). We then performed in-out PCR on tNGFR^{high} methylcellulose clones to determine on-target integration frequencies, and found that 86% of clones had a targeted integration, with 30% having biallelic integrations (Fig. 2d, Extended Data Fig. 4d–f and Supplementary Fig. 1b).

We next evaluated progenitor cell capacity of GFP^{high} HSPCs using the colony-forming unit (CFU) assay. *HBB* GFP^{high} HSPCs formed all types of colonies (erythroid, granulocyte/macrophage, and multilineage) to a comparable extent as mock-electroporated cells (Extended Data Fig. 5a–c). We also evaluated frequencies of GFP^{high} cells in subpopulations of HSPCs[15,34–36] by immunophenotypic analysis, and observed a significant negative correlation between targeting frequencies and the immunophenotypic primitiveness of the analysed population (Extended Data Fig. 5d–f). To confirm these findings, we used another strategy to evaluate targeting rates in the primitive HSC population. HSPCs or HSCs were sorted from fresh cord blood, then subjected to homologous recombination experiments, and we observed a 38% reduction in targeting efficiencies in the HSCs compared to the heterogeneous HSPC population (Extended Data Fig. 5g). We next tested whether inefficient targeting of primitive cells could be due to reduced rAAV6 donor availability. HSCs and multipotent progenitors were transduced with scAAV6-SFFV-GFP, and results showed a fivefold reduction in the number of GFP^{+} HSCs and multipotent progenitors compared to the bulk CD34^{+} population (Extended Data Fig. 5h). Collectively, our data suggest that although HSCs are more difficult to target than progenitor cells, we are achieving homologous recombination frequencies above 4% and usually above 10%.

HBB targeting in long-term repopulating HSCs

The current gold standard for HSC function, defined by the capacity to self-renew and form differentiated blood cells, is *in vivo* engraftment into immunodeficient non-obese diabetic (NOD)- severe combined immunodeficiency (SCID) *Il2rg*$^{-/-}$ (NSG) mice. We used HSPCs derived from mobilized peripheral blood for such studies because of their high clinical relevance, although these cells have been shown to have reduced engraftment capacity in NSG mice compared to HSPCs derived from fetal liver, cord blood and bone marrow[22,37]. All transplanted mice displayed human engraftment in the bone marrow as measured by the presence of hCD45/HLA-ABC double-positive cells 16 weeks after transplant (Fig. 3a and Extended Data Fig. 6a, b). While we observed a decrease in human cell chimaerism for all treatment groups compared to the mock-treated group, all groups with nuclease-treated cells displayed similar chimaerism to the rAAV6-only group. We did measure a small, but not statistically significant, decrease for the RNP plus AAV GFP^{high} group compared to RNP plus AAV, which can be explained by transplantation of fewer total cells and fewer phenotypically identified long-term HSCs (Extended Data Fig. 6c). There was a significant decrease from RNP plus AAV input targeting frequencies (16% in the CD34^{+} mobilized peripheral blood HSPCs) compared to the percentage of GFP^{+} cells in the bone marrow at week 16 after transplantation (3.5%) (Fig. 3c). This decrease is consistent with previous publications, and immunophenotyping of input cells did in fact show an average of 4% targeting in the CD34^{+} CD38^{-} CD90^{+} CD45RA^{-} population (Extended Data Fig. 5f). Despite these observed reductions *in vivo*, our median rates of *HBB* gene targeting in human cells in the bone marrow (3.5%) seem to be higher than what was found by others, particularly Hoban *et al.* and Genovese *et al.*, in which most mice appeared to have less than 1% gene modification after transplant using ZFNs targeting *HBB* and *IL2RG*, respectively[20,21]. By contrast, mice transplanted with RNP plus GFP^{high} cells had a median of 90% GFP^{+} human cells at week 16 after transplant, with three mice showing more than 97% GFP^{+} human cells (Fig. 3a, c). We also observed both myeloid (CD33^{+}) and lymphoid (CD19^{+}) reconstitution with a median of 94% and 83% GFP^{+} cells, respectively (Fig. 3c), implicating targeting of multipotent HSCs. We detected 5% and 49% GFP^{+} human HSPCs (CD34^{+} CD10^{-}) in the bone marrow of mice transplanted with RNP plus AAV and RNP plus AAV GFP^{high} cells, respectively (Extended Data Fig. 7). Multi-parameter flow cytometric analysis showed no perturbations in lineage reconstitution and no evidence of abnormal haematopoiesis, a functional assessment of the safety of the editing procedure. To determine experimentally whether we targeted *HBB* in HSCs, we performed secondary transplants for the RNP plus AAV and RNP plus AAV GFP^{high} groups. For both groups, we detected human cells in the bone marrow of secondary recipients at weeks 12–14 after transplant, with 7% and 90% GFP^{+} cells for the RNP plus AAV and RNP plus AAV GFP^{high} groups, respectively (Fig. 3d). More importantly, we confirmed *HBB* on-target integration events in human cells sorted from the bone marrow of secondary recipients from RNP plus AAV GFP^{high} experimental groups (Fig. 3e and Supplementary Fig. 2). Collectively, these data confirm that our strategy targets the *HBB* gene in human HSCs.

We next scaled up the genome-editing process to resemble a more clinically relevant HSPC starting cell number. We electroporated 80 million mobilized peripheral-blood-derived HSPCs with the *HBB*

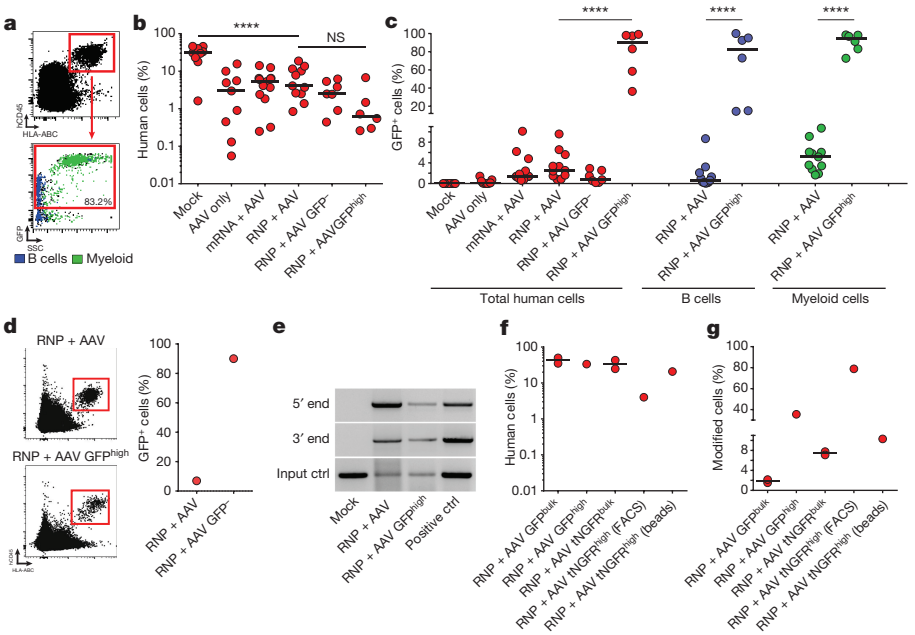


Figure 3 | *HBB* gene-targeted HSPCs display long-term and multi-lineage reconstitution in NSG mice. a, 16 weeks after transplantation, mouse bone marrow was analysed for human cell chimaerism and GFP expression by flow cytometry. Top, representative FACS plot from a mouse transplanted with RNP plus AAV GFP^{high} HSPCs showing engrafted human cells in the red gate. Bottom, representative FACS plot showing GFP-expressing human cells (red). CD19$^+$ B cells and CD33$^+$ myeloid cells are back-gated and shown in blue and green, respectively. **b**, Human engraftment in NSG mice from all experimental groups. Three different HSPC donors were used for engraftment studies (n = number of data points within group). ****$P < 0.0001$, NS (not significant) = $P \geq 0.05$, one-way analysis of variance (ANOVA) and Tukey's multiple comparison test. Bars represent median. **c**, Percentage GFP$^+$ cells in the total human population (red), CD19$^+$ B cells (blue), and CD33$^+$ myeloid cells (green) (n = number of data points within group). *$P < 0.05$, ****$P < 0.0001$, one-way ANOVA and Tukey's multiple comparison test for total human cells, and unpaired t-test with Welch's correction for B and myeloid cells. Bars represent median. **d**, Flow cytometric analysis of cell chimaerism and GFP expression 12–14 weeks after secondary transplantation. Left, representative FACS plot from a secondary mouse transplanted with RNP plus AAV (top) or RNP plus AAV GFP^{high} (bottom) cells showing engrafted human cells in the red gate. Right, percentage GFP$^+$ human cells in the bone marrow of secondary recipients. **e**, Gel images of in-out PCR analyses on sorted human cells from secondary recipients to analyse on-target integrations at the 5′ and 3′ ends. Input control (ctrl) PCR was performed on the human *CCR5* gene. Positive control is an HSPC sample targeted at *HBB* with SFFV-GFP-polyA. **f**, 80 million mobilized peripheral-blood-derived CD34$^+$ cells were electroporated with *HBB* RNP and transduced with *HBB* AAV6s. Bulk HSPCs or HSPCs enriched for targeting (by FACS or bead enrichment) were transplanted into the tail vein of sublethally irradiated mice. Then, 16 weeks after transplant, human cell chimaerism was analysed by flow cytometry (n = number of data points within group). **g**, Percentage GFP$^+$ and tNGFR$^+$ cells in the human population was analysed by flow cytometry (n = number of data points within group), bars represent median.

RNP system, transduced them with either SFFV-GFP or SFFV-tNGFR rAAV6, and then transplanted bulk RNP plus AAV, sorted GFP^{high} and tNGFR^{high} cells (enriched by FACS or magnetic microbeads). At 16 weeks after transplant, all mice displayed engraftment of edited human cells in the bone marrow (Fig. 3f). In this large-scale procedure, the human cell engraftment of treated cells was equivalent to 'mock' in our previous experiment (Fig. 3b). Although we observed reductions when comparing editing rates in the input cells to engrafted cells *in vivo*, the *HBB*-tNGFR mice showed a lower reduction (12% *in vitro* versus 7.5% *in vivo*) than the *HBB*-GFP mice (10% *in vitro* versus 1.9% *in vivo*), suggesting tNGFR might be a better transgene to evaluate editing of HSCs *in vivo* (Fig. 3g). Furthermore, mice transplanted with enriched targeted HSPCs displayed human cell-editing frequencies of 10–75% (three mice), with human engraftment levels ranging from 4% to 30% (Fig. 3f, g). These data suggest that our methodology can be translated to perform large-scale genome editing in HSCs at the *HBB* locus.

HBB gene correction in SCD HSPCs

We next tested our methodology to correct the disease-causing Glu6Val mutation in CD34$^+$ HSPCs derived from patients with SCD. We first confirmed high frequencies of INDELs (Fig. 4a) and homologous recombination using an SFFV-GFP donor (Fig. 4b) at the *HBB* locus in SCD HSPCs. We then produced a therapeutic rAAV6 donor (corrective SNP donor) designed to revert the Glu6Val mutation, while also introducing silent mutations to prevent Cas9 re-cutting and premature cross over (Extended Data Fig. 8a). Targeting SCD HSPCs with the corrective SNP donor reverted an average of 50% of the Glu6Val (HbS) alleles to wild-type (HbA) alleles (Fig. 4c), and analysis of methylcellulose clones showed that an average of 45% of clones had at least one HbA allele (Extended Data Fig. 8b). We next created an anti-sickling *HBB* cDNA therapeutic donor (HbAS3; ref. 38) using previously reported strategies[12,13] of knocking in divergent cDNAs into the gene start codon to preserve endogenous promoter/enhancer function, followed by a clinically relevant promoter (EF1α) driving tNGFR expression to allow for enrichment of edited cells (Extended Data Fig. 9a). Using this donor, we targeted an average of 11% of SCD patient-derived HSPCs (Fig. 4d) and confirmed seamless integration by in-out PCR (Extended Data Fig. 9b). Notably, we observed a tNGFR^{high} population as described previously (Fig. 2c), indicating the ability to enrich SCD-corrected HSPCs early in the culture process. We conclude that our methodology can correct the Glu6Val mutation using two different donor designs.

We next tested whether the *HBB*-edited SCD HSPCs maintained their erythroid differentiation potential by subjecting tNGFR^{high} and mock HSPCs to a 21-day *in vitro* erythroid differentiation protocol[39,40]. Flow cytometric analyses after erythroid differentiation showed a high proportion of GPA$^+$ CD45$^-$ CD71$^+$ CD34$^-$ cells, indicating the presence of mature differentiated erythrocytes that express haemoglobin[41] (Fig. 4e and Extended Data Fig. 10). To confirm that adult β-globin (HbA) or HbAS3 mRNA was transcribed from edited *HBB* alleles, we performed reverse transcription quantitative PCR

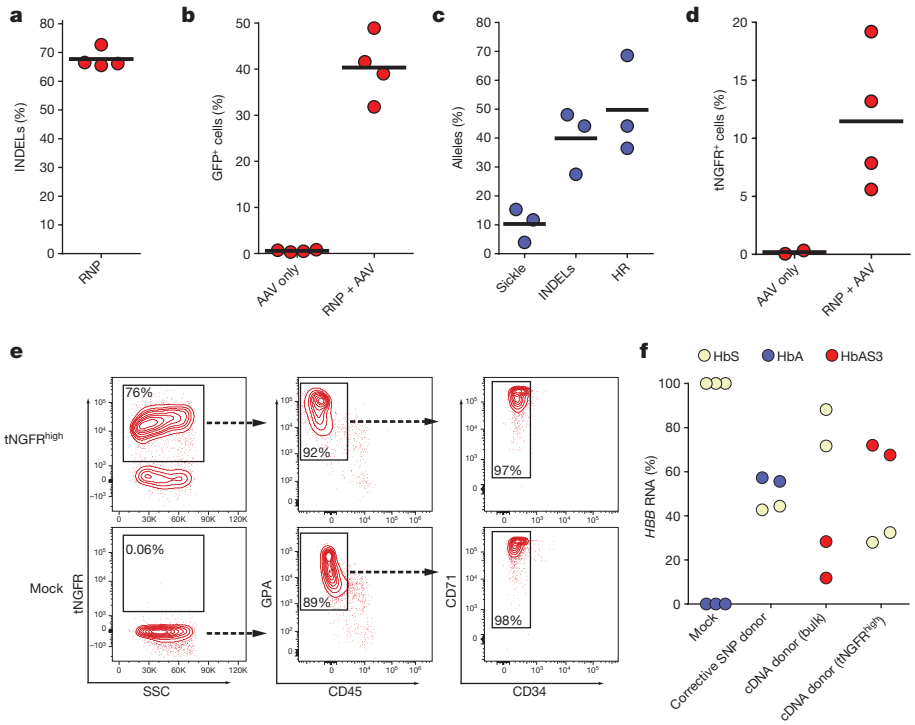


Figure 4 | Correction of the Glu6Val mutation in SCD patient-derived HSPCs. a, Genomic DNA from *HBB* RNP-treated SCD HSPCs was collected and INDELs were analysed via TIDE software ($n = 4$ different SCD patient donors). **b**, *HBB*-targeted SCD HSPCs were analysed for GFP expression by flow cytometry ($n = 4$ different SCD patient donors). **c**, SCD HSPCs were targeted with the rAAV6 corrective SNP donor. *HBB* allele types were analysed by sequencing TOPO-cloned PCR fragments derived from in-out PCR. Around 50–100 TOPO clones were analysed from each of three different HSPC donors ($n = 3$ different SCD patient donors). **d**, SCD HSPCs were targeted with the anti-sickling *HBB* cDNA-EF1α -tNGFR correction donor. Frequencies of tNGFR$^+$ cells were analysed by flow cytometry ($n = 3$ different SCD patient donors). **e**, SCD mock HSPCs and sorted SCD tNGFR^{high} HSPCs were differentiated into erythrocytes *in vitro*. Representative FACS plots from day 21 of differentiation show cell surface markers associated with erythrocytes (GPA$^+$ CD45$^-$ CD71$^+$ CD34$^-$). **f**, HbS, HbA and HbAS3 mRNA expression was quantified by RT–qPCR in erythrocytes differentiated from *HBB*-edited or mock SCD HSPCs. All mRNA transcript levels were normalized to the *RPLP0* input control ($n = 2$–3 different SCD patient donors). All bars represent means.

(RT–qPCR) on erythrocytes differentiated from edited SCD HSPCs. Erythrocytes edited with the corrective SNP donor expressed 56% HbA mRNA out of total β-globin mRNA, whereas erythrocytes edited with the cDNA donor (bulk) expressed 20% HbAS3 mRNA (Fig. 4f). Thus the percentage of HbAS3 mRNA (20%) matched or exceeded the percentage of cells modified by the tNGFR cassette (11%), suggesting functional expression of the AS3 cDNA from the endogenous HBB promoter. Notably, erythrocytes differentiated from enriched tNGFR^{high} HSPCs expressed 70% HbAS3 mRNA, confirming an enrichment of functionally corrected HSPCs.

Discussion

While our data support the idea that HSCs are more resistant to homologous recombination-mediated editing, they also show that it is possible to edit these cells at reasonable frequencies[21]. Enrichment of targeted cells resulted in the removal of most HSCs, leading to an overall eightfold decrease in the total number of HSCs in the transplanted enriched population (Extended Data Fig. 6c). Even though we transplanted eightfold fewer total HSCs in the enriched population than the non-enriched population (GFP^{high} versus RNP plus AAV), we found that absolute numbers of edited cells in the bone marrow of the mice from the enriched group was on average fivefold higher 16 weeks after transplant (Extended Data Fig. 6c). Thus, our enrichment strategy not only yields higher frequencies of modified cells in the transplanted mice, but also the absolute number of modified human cells in the mice was higher and can thus ameliorate the problem of inefficient HSC targeting. Recent advances in *ex vivo* HSC expansion protocols and identification of small molecule drugs, such as UM171 (ref. 42) that expands HSCs, might be combined with our strategy to generate a large and highly enriched population of genome-edited HSCs. Future studies will help to determine whether HSC expansion would be required for clinical translation of our enrichment model.

Our methodology sets the framework for CRISPR-mediated *HBB* targeting in HSCs that has the power to be translated to the clinic. Although GFP is an unsuitable marker for gene therapy, our enrichment protocol using tNGFR (Figs 2c, 3f, g) (or other similar signalling-inert cell surface markers) represents a strategy for the next generation of β-haemoglobinopathy therapies that are based on gene editing. These studies show that this methodology can enrich corrected SCD patient-derived HSPCs that can differentiate into erythrocytes that express *HBB* anti-sickling mRNA from the endogenous *HBB* promoter. This tNGFR selection strategy has the potential advantage over chemoselection strategies because it avoids exposing edited cells and patients to potentially toxic chemotherapy[13]. The strategy of knocking in a *HBB* cDNA along with a selectable marker to enrich for modified cells would be applicable to both SCD and almost all forms of β-thalassaemia. Furthermore, because we can efficiently scale up the genome-editing process to clinically relevant starting numbers, future studies will address the development of a current good manufacturing practice (cGMP)-compatible process for editing the *HBB* locus in HSPCs.

In conclusion, we believe that the presented methodology lays the foundation for CRISPR/Cas9-mediated genome editing therapies not only for the β-haemoglobinopathies, but also for a range of other haematological diseases, and generally advances HSC-based cell and gene therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 May; accepted 29 September 2016.
Published online 7 November 2016.

1. Naldini, L. *Ex vivo* gene transfer and correction for cell-based therapies. *Nat. Rev. Genet.* **12,** 301–315 (2011).
2. Porteus, M. Genome editing: a new approach to human therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **56,** 163–190 (2016).
3. Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature* **317,** 230–234 (1985).
4. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300,** 763 (2003).
5. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl Acad. Sci. USA* **91,** 6064–6068 (1994).
6. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157,** 1262–1278 (2014).
7. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346,** 1258096 (2014).
8. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
9. Kass, E. M. & Jasin, M. Collaboration and competition between DNA double-strand break repair pathways. *FEBS Lett.* **584,** 3703–3708 (2010).
10. Porteus, M. H. Towards a new era in medicine: therapeutic genome editing. *Genome Biol.* **16,** 286 (2015).
11. Woods, N. B., Bottero, V., Schmidt, M., von Kalle, C. & Verma, I. M. Gene therapy: therapeutic gene causing lymphoma. *Nature* **440,** 1123 (2006).
12. Hubbard, N. *et al.* Targeted gene editing restores regulated CD40L expression and function in X-HIGM T cells. *Blood* **127,** 2513–2522 (2016).
13. Voit, R. A., Hendel, A., Pruett-Miller, S. M. & Porteus, M. H. Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res.* **42,** 1365–1378 (2014).
14. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33,** 985–989 (2015).
15. Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A. M. & Peault, B. Isolation of a candidate human hematopoietic stem-cell population. *Proc. Natl Acad. Sci. USA* **89,** 2804–2808 (1992).
16. Mukherjee, S. & Thrasher, A. J. Gene therapy for PIDs: progress, pitfalls and prospects. *Gene* **525,** 174–181 (2013).
17. Cavazzana-Calvo, M. *et al.* Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature* **467,** 318–322 (2010).
18. Naldini, L. Gene therapy returns to centre stage. *Nature* **526,** 351–360 (2015).
19. Jenq, R. R. & van den Brink, M. R. Allogeneic haematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. *Nat. Rev. Cancer* **10,** 213–221 (2010).
20. Hoban, M. D. *et al.* Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood* **125,** 2597–2604 (2015).
21. Genovese, P. *et al.* Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510,** 235–240 (2014).
22. Wang, J. *et al.* Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat. Biotechnol.* **33,** 1256–1263 (2015).
23. De Ravin, S. S. *et al.* Targeted gene addition in human $CD34^+$ hematopoietic cells for correction of X-linked chronic granulomatous disease. *Nat. Biotechnol.* **34,** 424–429 (2016).
24. Sather, B. D. *et al.* Efficient modification of *CCR5* in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci. Transl. Med.* **7,** 307ra156 (2015).
25. Wang, J. *et al.* Highly efficient homology-driven genome editing in human T cells by combining zinc-finger nuclease mRNA and AAV6 donor delivery. *Nucleic Acids Res.* **44,** e30 (2016).
26. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* **36,** 767–773 (2004).
27. Russell, D. W. & Hirata, R. K. Human gene targeting by viral vectors. *Nat. Genet.* **18,** 325–330 (1998).
28. Barzel, A. *et al.* Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* **517,** 360–364 (2015).
29. Hoban, M. D., Orkin, S. H. & Bauer, D. E. Genetic treatment of a molecular disorder: gene therapy approaches to sickle cell disease. *Blood* **127,** 839–848 (2016).
30. Bonini, C. *et al.* HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukemia. *Science* **276,** 1719–1724 (1997).
31. Ciceri, F. *et al.* Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haemopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I-II study. *Lancet Oncol.* **10,** 489–500 (2009).
32. Oliveira, G. *et al.* Tracking genetically engineered lymphocytes long-term reveals the dynamics of T cell immunological memory. *Sci. Transl. Med.* **7,** 317ra198 (2015).
33. Bonini, C. *et al.* Safety of retroviral gene marking with a truncated NGF receptor. *Nat. Med.* **9,** 367–369 (2003).
34. Seita, J. & Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2,** 640–653 (2010).
35. Majeti, R., Park, C. Y. & Weissman, I. L. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* **1,** 635–645 (2007).
36. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10,** 120–136 (2012).
37. Gu, A. *et al.* Engraftment and lineage potential of adult hematopoietic stem and progenitor cells is compromised following short-term culture in the presence of an aryl hydrocarbon receptor antagonist. *Hum. Gene Ther. Methods* **25,** 221–231 (2014).
38. Levasseur, D. N. *et al.* A recombinant human hemoglobin with anti-sickling properties greater than fetal hemoglobin. *J. Biol. Chem.* **279,** 27518–27524 (2004).
39. Dulmovits, B. M. *et al.* Pomalidomide reverses γ-globin silencing through the transcriptional reprogramming of adult hematopoietic progenitors. *Blood* **127,** 1481–1492 (2016).
40. 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).
41. Romero, Z. *et al.* β-globin gene transfer to human bone marrow for sickle cell disease. *J. Clin. Invest.* **123,** 3317–3330 (2013).
42. Fares, I. *et al.* Cord blood expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science* **345,** 1509–1512 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements D.P.D. was supported by a Stanford Child Health Research Institute (CHRI) Grant and Postdoctoral Award. R.O.B. was supported by an Individual Postdoctoral grant (DFF–1333-00106B) and a Sapere Aude, Research Talent grant (DFF–1331-00735B), both from the Danish Council for Independent Research, Medical Sciences. M.H.P. acknowledges the support of the Amon Carter Foundation, the Laurie Kraus Lacob Faculty Scholar Award in Pediatric Translational Research and NIH grant support PN2EY018244, R01-AI097320 and R01-AI120766. We thank D. Russell for the pDGM6 plasmid, H.-P. Kiem for scAAV6, G. de Alencastro and M. Kay for help with AAV production, the Binns Program for Cord Blood Research at Stanford University for cord-blood-derived $CD34^+$ HSPCs. We also thank Lonza (A. Toell and G. Alberts) for donating the LV unit for performing large-scale genome-editing studies. We further thank members of the Porteus laboratory, D. DiGiusto and M. G. Roncarolo for input, comments and discussion.

Author Contributions D.P.D. and R.O.B. contributed equally to this work as well as performed and designed most of the experiments. D.P.D., R.O.B., A.R. and C.E.N. designed, executed and analysed engraftment studies. J.C. performed some *HBB* homologous recombination, NHEJ and tNGFR experiments. G.W. performed *in vitro* erythrocyte differentiation experiments. M.P.-D. assisted with large-scale electroporation experiments. N.S. ran and analysed HSPC immunophenotyping experiments. A.B.W. scored and collected methylcellulose clones. N.U. sorted freshly isolated HSCs before targeting. S.M. purified $CD34^+$ HSPCs from peripheral blood of patients with SCD. A.N. was the attending physician who oversaw transfusions of patients with SCD. A.H., R.M. and K.I.W. contributed to experimental design and data interpretation. M.H.P directed the research and participated in the design and interpretation of the experiments and the writing of the manuscript. D.P.D. and R.O.B. wrote the manuscript with help from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H.P. (mporteus@stanford.edu).

Reviewer Information *Nature* thanks B. Ebert, L. Naldini and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

AAV vector production. AAV vector plasmids were cloned in the pAAV-MCS plasmid (Agilent Technologies) containing inverted terminal repeats from AAV serotype 2. The *HBB* rAAV6 GFP and tNGFR donor contained promoter, MaxGFP or tNGFR, and BGH polyA. The left and right homology arms for the GFP and tNGFR *HBB* donors were 540 bp and 420 bp, respectively. The Glu6Val rAAV6 donor contained 2.2 kb of sequence homologous to the sequence upstream of Glu6Val. The nucleotide changes are depicted in Extended Data Fig. 2. Immediately downstream of the last nucleotide change was 2.2 kb of homologous *HBB* sequence. *HBB* cDNA contained same homology arms as GFP and tNGFR donors above except the left homology arm was shortened to end at the sickle mutation. Sequence of full *HBB* cDNA is depicted in (Extended Data Fig. 9b). The sickle corrective donor used in the SCD-derived HSPCs in Fig. 4 had a total of 2.4 kb sequence homology to *HBB* with the SNPs shown in Extended Data Fig. 8a in the centre. scAAV6 carrying the SFFV promoter driving GFP was provided by H.-P. Kiem. AAV6 vectors were produced as described with a few modifications[43]. In brief, 293FT cells (Life Technologies) were seeded at 13×10^6 cells per dish in ten 15-cm dishes one day before transfection. One 15-cm dish was transfected using standard PEI transfection with 6 μg ITR-containing plasmid and 22 μg pDGM6 (a gift from D. Russell), which contains the AAV6 cap genes, AAV2 rep genes, and adenovirus helper genes. Cells were incubated for 72 h until collection of AAV6 from cells by three freeze–thaw cycles followed by a 45 min incubation with TurboNuclease at 250 U ml^{-1} (Abnova). AAV vectors were purified on an iodixanol density gradient by ultracentrifugation at 237,000*g* for 2 h at 18 °C. AAV vectors were extracted at the 60–40% iodixanol interface and dialysed three times in PBS with 5% sorbitol in the last dialysis using a 10K MWCO Slide-A-Lyzer G2 Dialysis Cassette (Thermo Fisher Scientific). Vectors were added pluronic acid to a final concentration of 0.001%, aliquoted, and stored at −80 °C until use. AAV6 vectors were titred using quantitative PCR to measure number of vector genomes as described previously[44].

$CD34^+$ haematopoietic stem and progenitor cells. Frozen $CD34^+$ HSPCs derived from bone marrow or mobilized peripheral blood were purchased from AllCells and thawed according to manufacturer's instructions. $CD34^+$ HSPCs from cord blood were either purchased frozen from AllCells or acquired from donors under informed consent via the Binns Program for Cord Blood Research at Stanford University and used fresh without freezing. $CD34^+$ HSPCs from patients with SCD were purified within 24 h of the scheduled apheresis. For volume reduction via induced rouleaux formation, whole blood was added 6% Hetastarch in 0.9% sodium chloride injection (Hospira, Inc.) in a proportion of 5:1 (v/v). Following a 60–90-min incubation at room temperature, the top layer, enriched for HSPCs and mature leukocytes, was carefully isolated with minimal disruption of the underlying fraction. Cells were pelleted, combined, and resuspended in a volume of PBS with 2 mM EDTA and 0.5% BGS directly proportional to the fraction of residual erythrocytes—typically 200–400 ml. Mononuclear cells (MNCs) were obtained by density gradient separation using Ficoll and $CD34^+$ HSPCs were purified using the $CD34^+$ Microbead Kit Ultrapure (Miltenyi Biotec) according to manufacturer's protocol. Cells were cultured overnight and then stained for CD34 and CD45 using APC anti-human CD34 (clone 561; Biolegend) and BD Horizon V450 anti-human CD45 (clone HI30; BD Biosciences), and a pure population of HSPCs defined as $CD34^{bright}/CD45^{dim}$ were obtained by cell sorting on a FACS Aria II cell sorter (BD Biosciences). All $CD34^+$ HSPCs were cultured in StemSpan SFEM II (StemCell Technologies) supplemented with SCF (100 ng ml^{-1}), TPO (100 ng ml^{-1}), Flt3 ligand (100 ng ml^{-1}), IL-6 (100 ng ml^{-1}), and StemRegenin1 (0.75 mM). Cells were cultured at 37 °C, 5% CO_2 and 5% O_2.

Electroporation and transduction of cells. The *HBB* and *IL2RG* synthetic sgRNAs used were purchased from TriLink BioTechnologies with chemically modified nucleotides at the three terminal positions at both the 5′ and 3′ ends. Modified nucleotides contained 2′-*O*-methyl-3′-phosphorothioate and the sgRNAs were HPLC-purified. The genomic sgRNA target sequences, with PAM in bold, are: *HBB*: 5′-CTTGCCCCACAGGGCAGTAA**CGG**-3′ (refs 45, 46); *IL2RG*: 5′-TGGTAATGATGGCTTCAACA**TGG**-3′. Cas9 mRNA containing 5-methylcytidine and pseudouridine was purchased from TriLink BioTechnologies. Cas9 protein was purchased from Life Technologies. Cas9 RNP was made by incubating protein with sgRNA at a molar ratio of 1:2.5 at 25 °C for 10 min immediately before electroporation. $CD34^+$ HSPCs were electroporated 1–2 days after thawing or isolation. $CD34^+$ HSPCs were electroporated using the Lonza Nucleofector 2b (program U-014) and the Human T Cell Nucleofection Kit (VPA-1002, Lonza) as we have found this combination to be superior in optimization studies. The following conditions were used: 5×10^6 cells ml^{-1}, 300 μg ml^{-1} Cas9 protein complexed with sgRNA at 1:2.5 molar ratio, or 100 μg ml^{-1} synthetic chemically modified sgRNA with 150 μg ml^{-1} Cas9 mRNA (TriLink BioTechnologies, non-HPLC purified). Following electroporation, cells were incubated for 15 min at 37 °C after which they were added AAV6 donor vectors at an MOI (vector genomes/cell) of 50,000–100,000 and then incubated at 30 °C or 37 °C overnight (if incubated at 30 °C, plates were then transferred to 37 °C) or targeting experiments of freshly sorted HSCs (Extended Data Fig. 5g), cells were electroporated using the Lonza Nucleofector 4D (program EO-100) and the P3 Primary Cell Nucleofection Kit (V4XP-3024). For the electroporation of 80 million $CD34^+$ HSPCs, the Lonza 4D-Nucleofector LV unit (program DZ-100) and P3 Primary Cell Kit were used. Subsequently, we have found no benefit to the 30 °C incubation and now perform all of our manufacturing at 37 °C.

Measuring targeted integration of fluorescent and tNGFR donors. Rates of targeted integration of GFP and tNGFR donors were measured by flow cytometry at least 18 days after electroporation. Targeted integration of a tNGFR expression cassette was measured by flow cytometry of cells stained with APC-conjugated anti-human CD271 (NGFR) antibody (BioLegend, clone: ME20.4). For sorting of GFP^{high} or $tNGFR^{high}$ populations, cells were sorted on a FACS Aria II SORP using the LIVE/DEAD Fixable Blue Dead Cell Stain Kit (Life Technologies) to discriminate live and dead cells according to manufacturer's instructions.

Positive selection and enrichment of $tNGFR^+$ HSPCs. Positive selection of targeted HSPCs was performed using the CD271 (tNGFR) Microbead Kit (Miltenyi Biotech), according to the manufacturer's instructions 72 h after electroporation. In brief, $tNGFR^+$ cells were magnetically labelled with CD271 Microbeads after which the cell suspension was loaded onto an equilibrated MACS column inserted in the magnetic field of a MACS separator. The columns were washed three times, and enriched cells were eluted by removing the column from the magnetic field and eluting with PBS. Enrichment was determined by flow cytometry during culture for 2–3 weeks by FACS analysis every 3 days.

Immunophenotyping of targeted HSPCs. Collected wells were stained with LIVE/DEAD Fixable Blue Dead Cell Stain (Life Technologies) and then with anti-human CD34 PE-Cy7 (581, BioLegend), CD38 Alexa Fluor 647 (AT1, Santa Cruz Biotechnologies), CD45RA BV 421 (HI100, BD Biosciences), and CD90 BV605 (5E10, BioLegend) and analysed by flow cytometry. For sorting of $CD34^+$ or $CD34^+$ $CD38^-$ $CD90^+$ cells, cord-blood-derived $CD34^+$ HSPCs were stained directly after isolation from blood with anti-human CD34 FITC (8G12, BD Biosciences), CD90 PE (5E10, BD Biosciences), CD38 APC (HIT2, BD Bioscience), and cells were sorted on a FACS Aria II (BD Bioscience), cultured overnight, and then electroporated with *HBB* RNP and transduced with *HBB* GFP rAAV6 using our optimized parameters.

Measuring targeted integration of the E6V donor. For assessing the allele modification frequencies in samples with targeted integration of the Glu6Val rAAV6 donor, PCR amplicons spanning the targeted region (see Extended Data Fig. 2a) were created using one primer outside the donor homology arm and one inside: *HBB*_outside 5′-GGTGACAATTTCTGCCAATCAGG-3′ and *HBB*_inside: 5′-GAATGGTAGCTGGATTGTAGCTGC-3′. The PCR product was gel-purified and re-amplified using a nested primer set (*HBB*_nested_fw: 5′-GAAGATATGCTTAGAACCGAGG-3′ and *HBB*_nested_rv: 5′-CCACATGCCCAGTTTCTATTGG-3′) to create a 685-bp PCR amplicon (see Extended Data Fig. 2a) that was gel-purified and cloned into a TOPO plasmid using the Zero Blunt TOPO PCR Cloning Kit (Life Technologies) according to the manufacturer's protocol. TOPO reactions were transformed into XL-1 Blue competent cells, plated on kanamycin-containing agar plates, and single colonies were sequenced by McLab by rolling circle amplification followed by sequencing using the following primer: 5′-GAAGATATGCTTAGAACCGAGG-3′. For each of the six unique $CD34^+$ donors used in this experiment, 100 colonies were sequenced. Additionally, 100 colonies derived from an AAV-only sample were sequenced and detected no integration events.

Measuring INDEL frequencies. INDEL frequencies were quantified using the TIDE software[47] (tracking of indels by decomposition) and sequenced PCR products obtained by PCR of genomic DNA extracted at least 4 days after electroporation as previously described[14].

Methylcellulose CFU assay. The CFU assay was performed by FACS sorting of single cells into 96-well plates containing MethoCult Optimum (StemCell Technologies) 4 days after electroporation and transduction. After 12–16 days, colonies were counted and scored based on their morphological appearance in a blinded fashion.

Genotyping of methylcellulose colonies. DNA was extracted from colonies formed in methylcellulose from FACS sorting of single cells into 96-well plates. In brief, PBS was added to wells with colonies, and the contents were mixed and transferred to a U-bottomed 96-well plate. Cells were pelleted by centrifugation at 300*g* for 5 min followed by a wash with PBS. Finally, cells were resuspended in 25 μl QuickExtract DNA Extraction Solution (Epicentre) and transferred to PCR plates, which were incubated at 65 °C for 10 min followed by 100 °C for 2 min. Integrated or non-integrated alleles were detected by PCR. For detecting *HBB* GFP integrations at the 3′ end, two different PCRs were set up to detect

integrated (one primer in insert and one primer outside right homology arm) and non-integrated alleles (primer in each homology arm), respectively (see Extended Data Fig. 4a). *HBB*_int_fw: 5′-GTACCAGCACGCCTTCAAGACC-3′, *HBB*_int_rv: 5′-GATCCTGAGACTTCCACACTGATGC-3′, *HBB*_no_int_fw: 5′-GAAGATATGCTTAGAACCGAGG-3′, *HBB*_no_int_rv: 5′-CCACATGCCCAGTTTCTATTGG-3′. For detecting *HBB* tNGFR integrations at the 5′ end, a 3-primer PCR methodology was used to detect the integrated and non-integrated allele simultaneously (see Extended Data Fig. 4d). *HBB*_outside_5′Arm_fw: 5′-GAAGATATGCTTAGAACCGAGG-3′, SFFV_rev: 5′-ACCGCAGATATCCTGTTTGG-3′, *HBB*_inside_3′Arm_rev: 5′-CCACATGCCCAGTTTCTATTGG-3′. Note that for the primers assessing non-integrated alleles, the Cas9 cut site is at least 90 bp away from the primer-binding sites and since CRISPR/Cas9 generally introduces INDELs of small sizes, the primer-binding sites should only very rarely be disrupted by an INDEL.

Transplantation of CD34$^+$ HSPCs into NSG mice. For *in vivo* studies, 6 to 8 week-old NSG mice were purchased from the Jackson laboratory (Bar Harbour). The experimental protocol was approved by Stanford University's Administrative Panel on Laboratory Animal Care. For transplant data in Fig. 3a–c, sample sizes were not chosen to ensure adequate power to detect a pre-specified effect size. Four days after electroporation/transduction or directly after sorting, 500,000 cells (or 100,000–500,000 cells for the GFP^{high} group) were administered by tail-vein injection into the mice after sub-lethal irradiation (200 cGy) using an insulin syringe with a 27 gauge × 0.5 inch (12.7 mm) needle. For transplant data in Fig. 3f, g, three days after electroporation, 400,000–700,000 bulk HSPCs or HSPCs enriched for targeting (FACS or bead-enrichment) were transplanted as described above. Mice were randomly assigned to each experimental group and evaluated in a blinded fashion. For secondary transplants, human cells from the RNP plus AAV group were pooled and CD34$^+$ cells were selected using a CD34 bead enrichment kit (MACS CD34 MicroBead Kit UltraPure, human, Miltenyi Biotec), and finally cells were injected into the femurs of female secondary recipients (3 mice total). Because GFP^{high} mice had low engraftment, they were not CD34$^+$-selected, but total mononuclear cells were filtered, pooled, and finally injected into the femur of two secondary recipients.

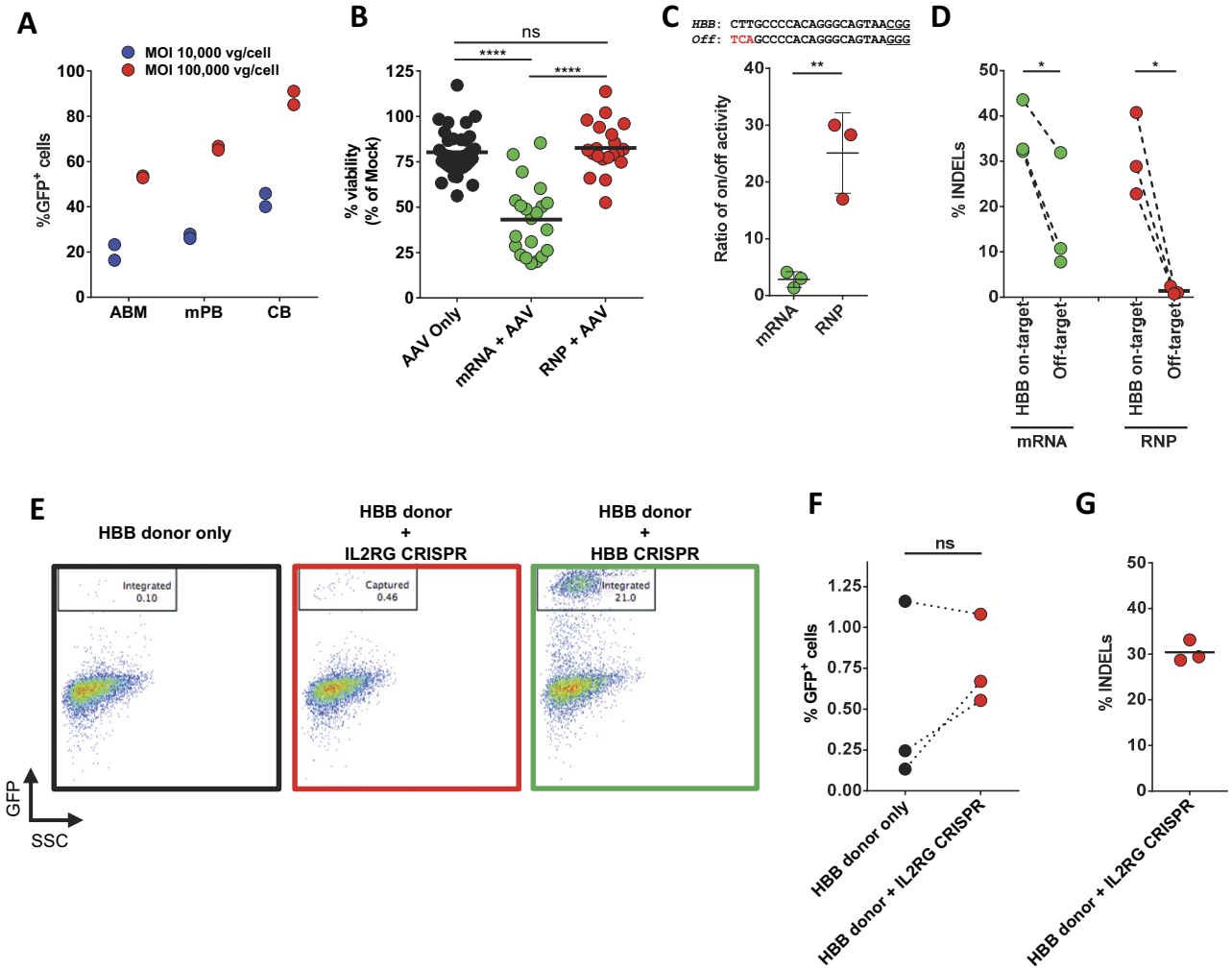
Assessment of human engraftment. At week 16 after transplantation, mice were euthanized, mouse bones (2× femur, 2× tibia, 2× humerus, sternum, 2× pelvis, spine) were collected and crushed using mortar and pestle. MNCs were enriched using Ficoll gradient centrifugation (Ficoll-Paque Plus, GE Healthcare) for 25 min at 2,000*g*, room temperature. Cells were blocked for nonspecific antibody binding (10% v/v, TruStain FcX, BioLegend) and stained (30 min, 4 °C, dark) with monoclonal anti-human CD45 V450 (HI30, BD Biosciences), CD19 APC (HIB19, BD Biosciences), CD33 PE (WM53, BD Biosciences), HLA-ABC APC-Cy7 (W6/32, BioLegend), anti-mouse CD45.1 PE-Cy7 (A20, eBioScience), anti-mouse PE-Cy5 mTer119 (TER-119, eBioscience) antibodies. Normal multi-lineage engraftment was defined by the presence of myeloid cells (CD33$^+$) and B-cells (CD19$^+$) within engrafted human CD45$^+$ HLA-ABC$^+$ cells. Parts of the mouse bone marrow were used for CD34-enrichment (MACS CD34 MicroBead Kit UltraPure, human, Miltenyi Biotec) and the presence of human HSPCs was assessed by staining with anti-human CD34 APC (8G12, BD Biosciences), CD38 PE-Cy7 (HB7, BD Biosciences), CD10 APC-Cy7 (HI10a, BioLegend), and anti-mouse CD45.1 PE-Cy5 (A20, eBioScience) and analysed by flow cytometry. The estimation of the total number of modified human cells in the bone marrow at week 16 after transplant was calculated by multiplying the percentage engraftment with the percentage GFP$^+$ cells among engrafted cells. This number was multiplied by the total number of MNCs in the bone marrow of a NSG mouse (1.1×10^8 per mouse) to give the total number of GFP$^+$ human cells in the total bone marrow of the transplanted mice. The total number of MNCs in the bone marrow of a NSG mouse was calculated by counting the total number of MNCs in one femur in four NSG mice. The total number of MNCs in one mouse was then calculated assuming one femur is 6.1% of the total marrow as found previously[48].

Differentiation of CD34$^+$ HSPCs into erythrocytes *in vitro*. SCD patient-derived HSPCs were cultured in three phases following targeting at 37 °C and 5% CO_2 in SFEM II media according to previously established protocols[39,40]. Media was supplemented with 100 U ml^{-1} penicillin/streptomycin, 2 mM L-glutamine, 40 μg ml^{-1} lipids, 100 ng ml^{-1} SCF, 10 ng ml^{-1} IL-3 (PeproTech), 0.5 U ml^{-1} erythropoietin (eBiosciences), and 200 μg ml^{-1} transferrin (Sigma Aldrich). In the first phase, corresponding to days 0–7 (day 0 being day 4 after electroporation), cells were cultured at 10^5 cells ml^{-1}. In the second phase, corresponding to days 7–11, cells were maintained at 10^5 cells ml^{-1} and erythropoietin was increased to 3 U ml^{-1}. In the third and final phase, days 11–21, cells were cultured at 10^6 cells ml^{-1} with 3 U ml^{-1} of erythropoietin and 1 mg ml^{-1} of transferrin. Erythrocyte differentiation of edited and non-edited HSPCs was assessed by flow cytometry using the following antibodies: hCD45 V450 (HI30, BD Biosciences), CD34 FITC (8G12, BD Biosciences), CD71 PE-Cy7 (OKT9, Affymetrix), and CD235a PE (GPA) (GA-R2, BD Biosciences).

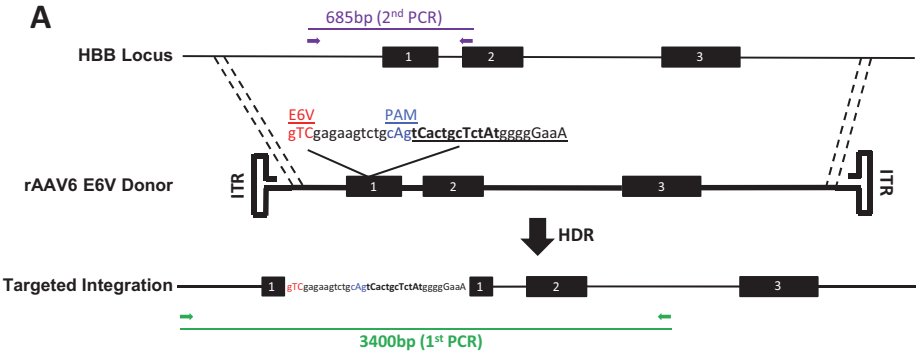
Assessment of mRNA levels in differentiated erythrocytes. RNA was extracted from 100,000–250,000 differentiated erythrocytes between days 16–21 of erythroid differentiation using the RNeasy Mini Kit (Qiagen) and was DNase-treated with RNase-Free DNase Set (Qiagen). cDNA was made from 100 ng RNA using the iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad). Levels of HbS, HbA (from corrective SNP donor), and HbA-AS3 (anti-sickling *HBB* cDNA donor) were quantified by qPCR using the following primers and FAM/ZEN/IBFQ-labelled hydrolysis probes purchased as custom-designed PrimeTime qPCR Assays from IDT: HbS primer (fw): 5′-TCACTAGCAACCTCAAACAGAC-3′, HbS primer (rv): 5′-ATCCACGTTCACCTTGCC-3′, HbS probe: 5′-TAACG GCAGACTTCTCCACAGGAGTCA-3′, HbA primer (fw): 5′-TCACTAGCAACCT CAAACAGAC-3′, HbA primer (rv): 5′-ATCCACGTTCACCTTGCC-3′, HbA probe: 5′-TGACTGCGGATTTTTCCTCAGGAGTCA-3′, HbAS3 primer fw: 5′-GTGTATCCCTGGACACAAAGAT-3′, HbAS3 primer (rv): 5′-GGGC TTTGACTTTGGGATTTC-3′, HbAS3 probe: 5′-TTCGAAAGCTTCGGC GACCTCA-3′. Primers for HbA and HbS are identical, but probes differ by six nucleotides, and therefore it was experimentally confirmed that these two assays do not cross-react with targets. To normalize for RNA input, levels of the reference gene *RPLP0* was determined in each sample using the IDT predesigned RPLP0 assay (Hs.PT.58.20222060). qPCR reactions were carried out on a LightCycler 480 II (Roche) using the SsoAdvanced Universal Probes Supermix (BioRad) following manufacturer's protocol and PCR conditions of 10 min at 95 °C, 50 cycles of 15 s at 95 °C and 60 s at 58 °C. Relative mRNA levels were determined using the relative standard curve method, in which a standard curve for each gene was made from serial dilutions of the cDNA. The standard curve was used to calculate relative amounts of target mRNA in the samples relative to levels of *RPLP0*.

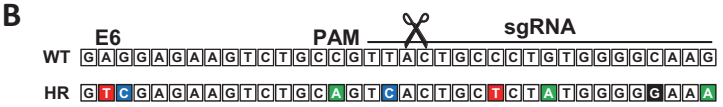
Data availability. The authors declare that the data supporting the findings of this study are available within the paper.

43. Khan, I. F., Hirata, R. K. & Russell, D. W. AAV-mediated gene targeting methods for human cells. *Nat. Protocols* **6,** 482–501 (2011).
44. 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).
45. Hendel, A. *et al.* Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Reports* **7,** 293–305 (2014).
46. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β-globin and *CCR5* genes have substantial off-target activity. *Nucleic Acids Res.* **41,** 9584–9592 (2013).
47. 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).
48. Boggs, D. R. The total marrow mass of the mouse: a simplified method of measurement. *Am. J. Hematol.* **16,** 277–286 (1984).



Extended Data Figure 1 | High tropism of rAAV6 for CD34$^+$ HSPCs, and viability and specificity assessment of gene editing in CD34$^+$ HSPCs. a, CD34$^+$ HSPCs were transduced with a scAAV6 expressing GFP from an SFFV promoter at multiplicities of infections (MOIs) of 10,000 or 100,000 viral genomes (vg) per cell for 48 h and then analysed for percentage GFP$^+$ expression by flow cytometry using a non-transduced sample to set the GFP$^+$ gate at <0.1% GFP$^+$ cells. scAAV was used because it eliminates second-strand synthesis as a confounder of actual transduction. Results are from two independent experiments from at least two donors and error bars represent s.d. ABM, adult bone marrow; CB, cord blood; mPB: mobilized peripheral blood. **b**, CD34$^+$ HSPCs were electroporated with the *HBB* CRISPR system (mRNA or RNP delivery) or without (AAV only), and then transduced with *HBB* rAAV6 donor vectors at an MOI of 100,000 vg per cell. Day 4 after electroporation, cells were analysed by flow cytometry and live cells were gated in high forward scatter (FSC) and low side scatter (SSC). Percentage of cells in FSC/SSC gate (that is, percentage viability) is shown relative to that of mock-electroporated cells. Each data point represents a unique CD34$^+$ HSPC donor. **c**, Top, sgRNA target sequences at the *HBB* on-target site and a highly complementary off-target site (Chr9:101833584–101833606) are shown. PAM sequences are underlined and red sequence highlights the three mismatches of the off-target site. Bottom, HSPCs were electroporated with either the mRNA or RNP-based CRISPR system, and 4 days post electroporation genomic DNA was extracted and analysed for INDEL frequencies using TIDE at the on-target *HBB* and the off-target site. Results are shown as the ratio of on- to off-target activity highlighting the increased specificity of the RNP system. Averages from three different CD34$^+$ HSPC donors are shown and error bars represent s.e.m. $^{**}P < 0.01$, unpaired Student's *t*-test. **d**, INDEL frequencies for the data presented in **c**. $^{*}P < 0.05$, paired Student's *t*-test. **e**, Representative FACS plots showing stable GFP rates at day 18 after electroporation in donor-nuclease mismatch experiments. Mismatching nuclease and donor (red box) leads to infrequent end-capture events compared to on-target homologous recombination events observed with matched nuclease and homologous rAAV6 donor (green box). HSPCs were electroporated with 15 μg Cas9 mRNA and either *HBB* or *IL2RG* 2′-*O*-methyl-3′-phosphorothioate-modified sgRNA, then transduced with *HBB*-GFP rAAV6 donor followed by 18 days of culture. **f**, End-capture experiments were performed in three replicate experiments each in three unique CD34$^+$ HSPC donors. ns (not significant) = $P \geq 0.05$, paired Student's *t*-test. Activity of the *IL2RG* CRISPR was confirmed by quantification of INDELs at the *IL2RG* target site using TIDE analysis.

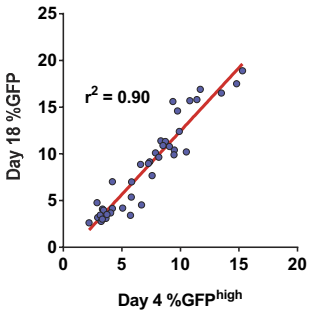




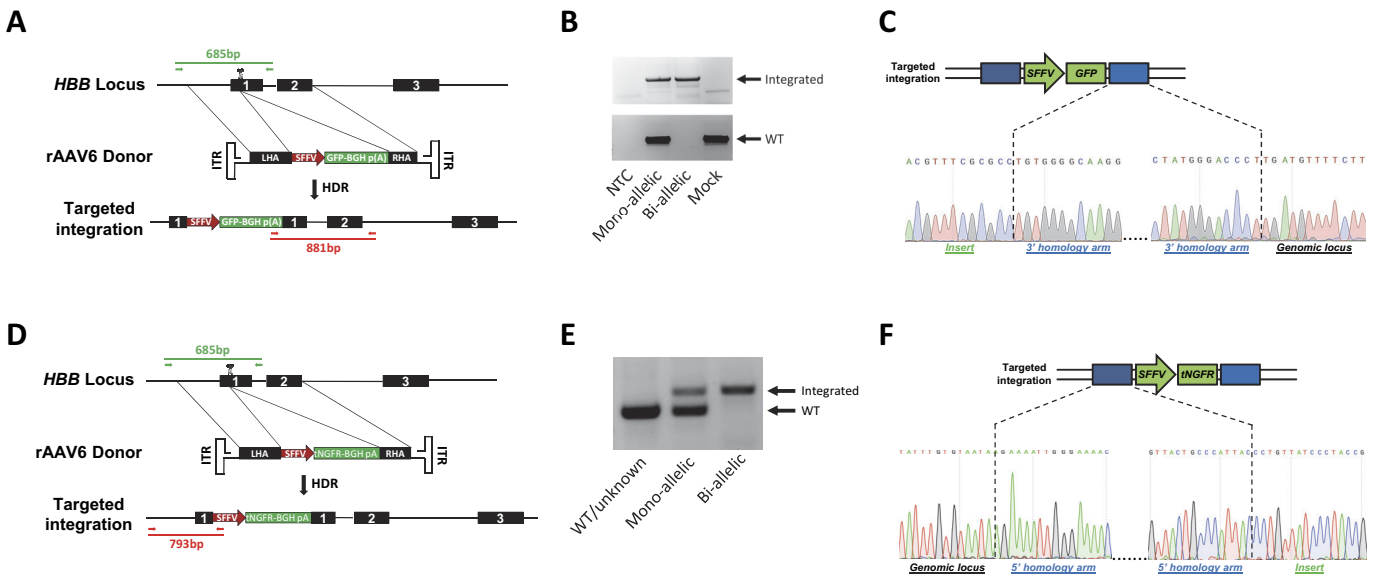
C

```
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTA---CCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -3
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCC---------CTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -9
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTA---CCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -3
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGT----GCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -4
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT---GCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -3
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGC--------CCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -8
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTA-------------CAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -13
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTA-TGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -1
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGT-ACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -1
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGT-----------GGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -11
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCC---------CTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -9
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCT--------TGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -8
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTA--GCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -2
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTA------TGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -6
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCAA-------TGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -9/+2
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTT-CTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -1
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGC------TGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -6
TCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGG--CTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT -3/+1
```

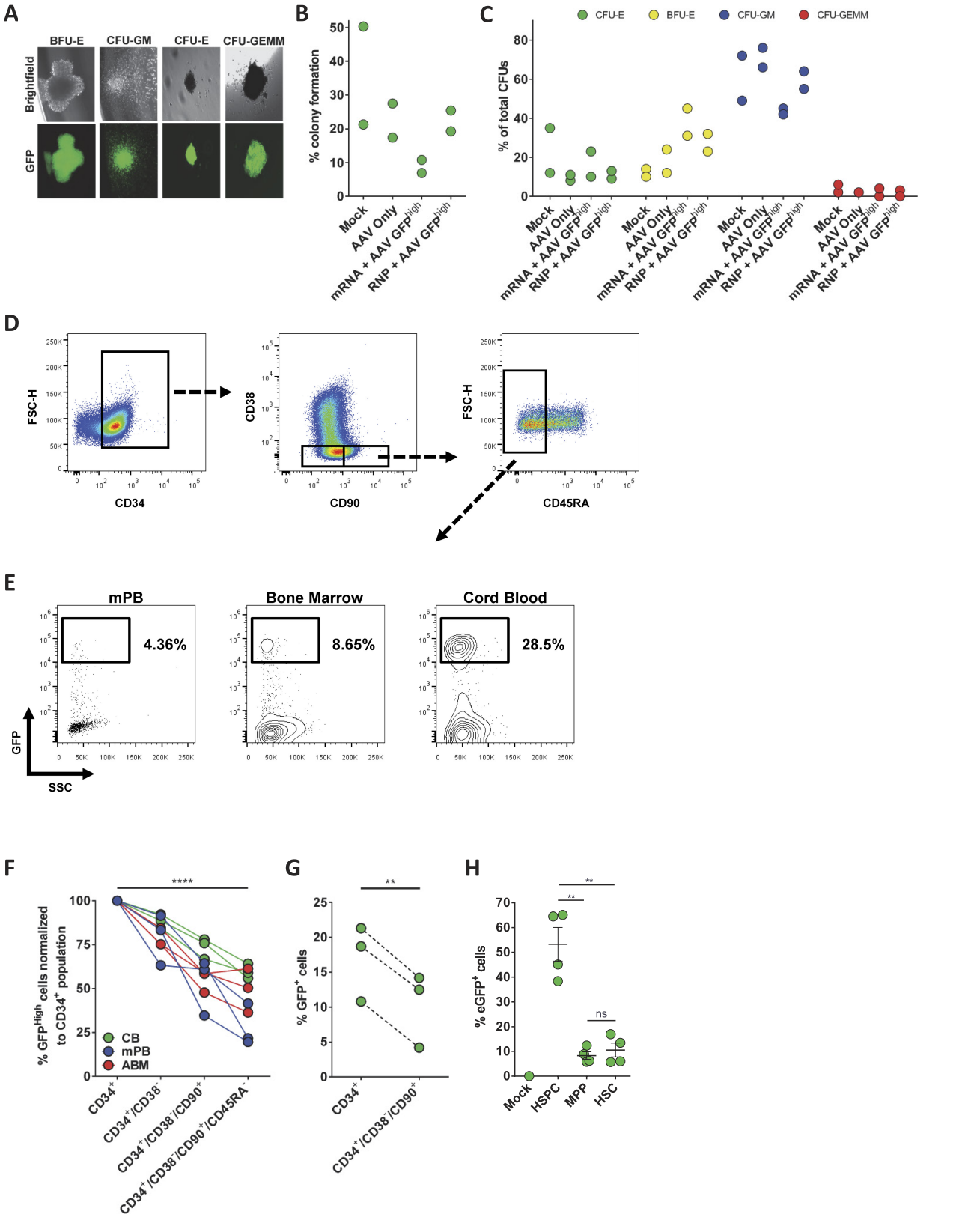
Extended Data Figure 2 | Schematic of targeting rAAV6 Glu6Val homologous donor to the *HBB* locus. a, The human *HBB* locus on chromosome 11 is depicted at the top of the schematic and consists of three exons (black boxes) and two introns. The rAAV6 Glu6Val donor includes the Glu to Val mutation at codon 6, which is the amino acid change causing SCD. Other SNPs (all SNPs are capitalized) were introduced to PAM site (blue) and sgRNA-binding site (bold) to prevent re-cutting following homologous recombination in HSPCs. To analyse targeted integration frequencies in HSPCs, a two-step PCR was performed. First, a 3,400-bp in-out PCR (green) was performed followed by a nested 685-bp PCR (purple) on a gel-purified fragment from the first PCR. This second PCR fragment was cloned into TOPO vectors, which were sequenced to determine the allele genotype (wild type, INDEL or homologous recombination). **b**, The sequence of a wild-type *HBB* allele aligned with the sequence of an allele that has undergone homologous recombination. **c**, Representative INDELs from the data represented in Fig. 1d. The *HBB* reference sequence is shown in green.



Extended Data Figure 3 | Linear regression model shows that the day 4 GFP^{high} population is a reliable predictor of targeting frequencies. Day 4 GFP^{high} percentages (*x* axis) were plotted against day 18 total GFP^{+} percentages (*y* axis), and linear regression was performed. Data were generated from experiments including a total of 38 different CD34^{+} HSPC donors, treated with either 15 μg or 30 μg Cas9 RNP to generate data points with a wider distribution of targeting frequencies.

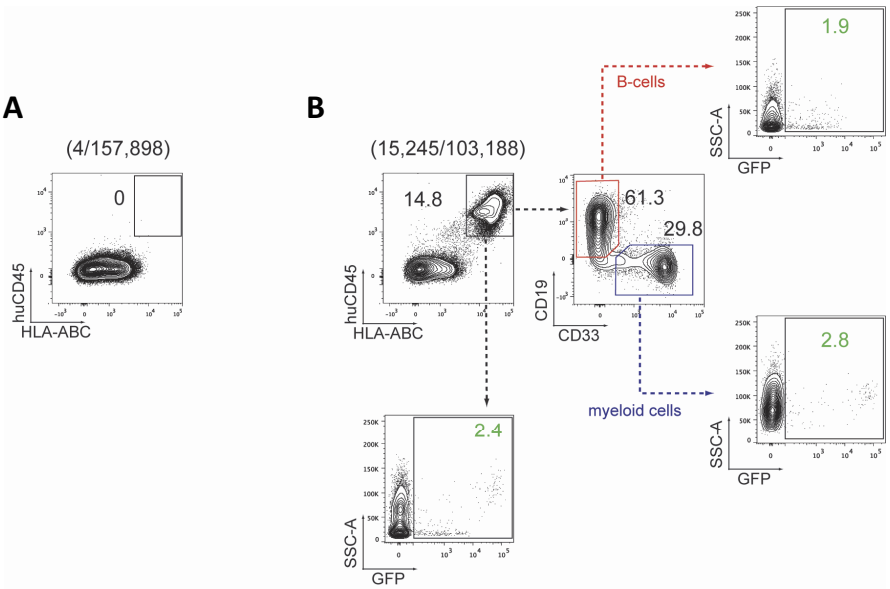


Extended Data Figure 4 | Overview of PCR genotyping of methylcellulose colonies with homologous recombination of the GFP and tNGFR donor at the *HBB* locus. a, The *HBB* locus was targeted by creating a DSB in exon 1 via Cas9 (scissors) and supplying a rAAV6 GFP donor template. Alleles with integrations were identified by PCR (red, 881 bp) on methylcellulose-derived colonies using an in-out primer set. Wild-type alleles were identified by PCR (green, 685 bp) using primers flanking the sgRNA target site. **b**, Representative genotyping PCRs showing mono- and biallelic clones as well as a clone derived from mock-treated cells. NTC, non-template control (see Supplementary Fig. 1a for uncropped gel). **c**, Representative Sanger sequence chromatograms for junctions between right homology arm (blue) and insert (green) or genomic locus, highlighting seamless homologous recombination. **d**, The *HBB* locus was targeted by creating a DSB in exon 1 via Cas9 (scissors) and supplying a rAAV6 tNGFR donor template. Genotypes were assessed by a three-primer genotyping PCR on methylcellulose-derived colonies using an in-out primer set (red, 793 bp) and a primer set flanking the sgRNA target site (green, 685 bp). Note that the two forward primers are the same. **e**, Representative genotyping PCRs showing a wild-type/unknown, mono-, and biallelic clone (see Supplementary Fig. 1b for uncropped gel). **f**, Representative Sanger sequence chromatograms for junctions between left homology arm (in blue) and insert (in green) or genomic locus highlighting seamless homologous recombination.



Extended Data Figure 5 | See next page for caption.

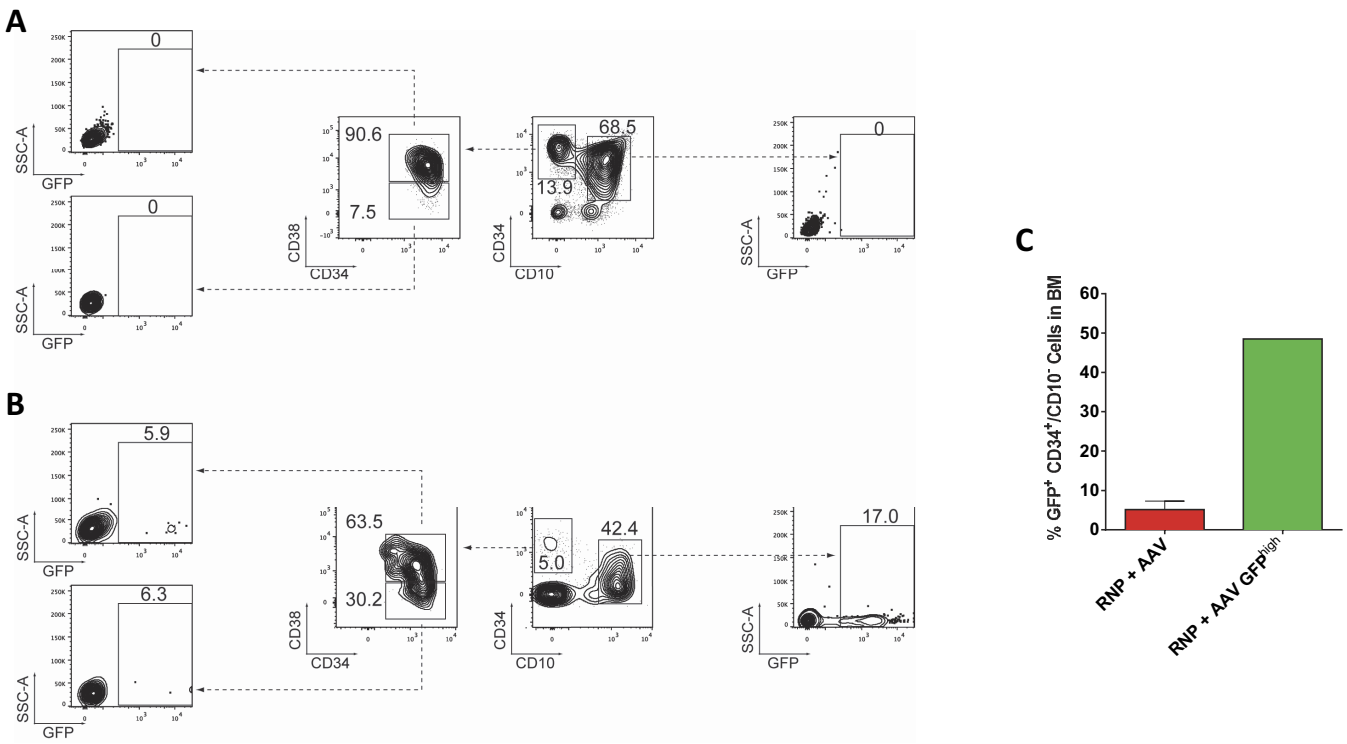
Extended Data Figure 5 | Haematopoietic progenitor CFU assay and targeting in different HSPC subpopulations. a, GFP^{high} HSPCs were single-cell-sorted into 96-well plates containing methylcellulose. Representative images from fluorescence microscopy show lineage-restricted progenitors (BFU-E, CFU-E, CFU-GM) and multipotent progenitors (CFU-GEMM) with GFP expression. **b**, CFUs were counted 14 days after sorting and shown relative to the total number of cells sorted (percentage colony formation) ($n = 2$ different HSPC donors). **c**, Colonies were scored according to their morphology: (1) CFU-erythroid (CFU-E); (2) burst forming unit-erythroid (BFU-E); (3) CFU-granulocyte/macrophage (CFU-GM); and (4) CFU-granulocyte/erythrocyte/macrophage/megakaryocyte (CFU-GEMM) ($n = 2$ different HSPC donors). **d**, Representative FACS plots at day 4 after electroporation of $CD34^+$ HSPCs showing the gating scheme for analysing targeting frequencies in different HSPC subsets (Extended Data Fig. 5f). Cells were immunophenotyped for CD34, CD38, CD90 and CD45RA expression and relevant FACS gates are indicated. **e**, Representative FACS plots showing GFP^{high} cells in the $CD34^+$ $CD38^-$ $CD90^+$ $CD45RA^-$ population of HSPCs derived from mobilized peripheral blood, bone marrow, or cord blood. **f**, 500,000 HSPCs isolated from mobilized peripheral blood, adult bone marrow, or cord blood were electroporated with RNP and transduced with GFP rAAV6 donor. At day 4 after electroporation, cells were phenotyped by flow cytometry for the cell surface markers CD34, CD38, CD90 and CD45RA (Extended Data Fig. 5d, e). Percentage GFP^{high} cells in the indicated subpopulations are shown (data points represent unique donors, $n = 3$ per HSPC source), $^{****}P < 0.0001$, paired Student's t-test. **g**, $CD34^+$ or $CD34^+$ $CD38^-$ $CD90^+$ cells were sorted directly from freshly isolated cord blood $CD34^+$ HSPCs, cultured overnight, and then electroporated with RNP and transduced with GFP rAAV6. Bars show average percentage GFP^+ cells at day 18 after electroporation. ($n = 3$ from different HSPC donors), $^{**}P < 0.01$, paired Student's t-test. **h**, Multipotent progenitor (MPP) ($CD34^+$ $CD38^-$ $CD90^-$ $CD45RA^-$) and HSC ($CD34^+$ $CD38^-$ $CD90^+$ $CD45RA^-$) populations were sorted from fresh cord-blood-derived $CD34^+$ HSPCs and immediately after sorting, cells were transduced with scAAV6-SFFV-GFP at an MOI of 100,000 vg per cell along the bulk HSPC population. scAAV6 was used because it eliminates second-strand synthesis as a confounder of actual transduction, although the activity of the SFFV promoter may not be equivalent in each population, thus potentially underestimating the degree of transduction of MPPs and HSCs. Two days later, transduction efficiencies were measured by FACS analysis of GFP expression using non-transduced cells (mock) to set the GFP^+ gate. Error bars represent s.e.m., $n = 4$, two different HSPC donors. $^{**}P < 0.01$; NS, not significant $= P \geq 0.05$, unpaired t-test with Welch's correction.



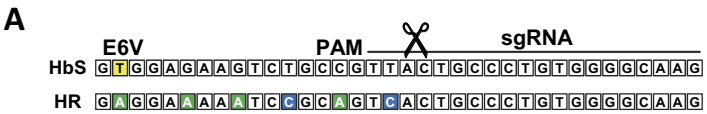
C

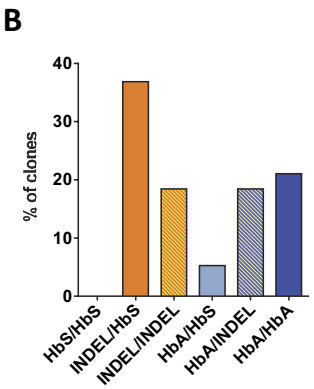
Sample Name	% engraftment (average per mouse)	% GFP⁺ cells among human cells (average per mouse)	Total live cells injected per mouse (average per mouse)	Average total HSCs injected per mouse ($CD34^+/CD38^-/CD90^+/CD45RA^-$)	Estimated total modified human cells in BM per mouse at Week 16
RNP + AAV	6.99 (+/-1.8)	3.46 (+/-0.82)	500,000	24,000 (+/-1100)	270,000 (+/-33,000)
RNP + AAV GFP^{high}	1.67 (+/-1.0)	78.8 (+/-10.6)	350,000 100,000 x1 250,000 x2 500,000 x3	2,900 (+/-770)	1,460,000 (+/-390,000)

Extended Data Figure 6 | Analysis of human engraftment.
a, Representative FACS plot from the analysis of the bone marrow of a control mouse not transplanted with human cells. Mice were euthanized and bone marrow was collected from femur, tibia, hips, humerus, sternum and vertebrae. Cells were subject to Ficoll density gradient to isolate mononuclear cells, which were analysed for human engraftment by flow cytometry. Human engraftment was delineated as hCD45/HLA-ABC double positive. From a total of 157,898 cells, 4 were found within the human cell gate of a non-injected control mouse, showing the very limited background human staining. **b**, Representative FACS plots showing gating scheme for analyses of NSG mice transplanted with human cells and analysed as described in **a**. Representative plots are from one mouse from the RNP plus rAAV6 experimental group. As above, human engraftment was delineated as hCD45/HLA-ABC double positive. B cells were marked by CD19 expression, and myeloid cells identified by CD33 expression. GFP expression was analysed in total human cells (2.4%), B cells (1.9%) and myeloid cells (2.8%). The GFP brightness in B cells is lower than in myeloid cells, suggesting that the SFFV promoter is not as active in the B-cell lineage compared to the myeloid lineage (see also Fig. 3a). **c**, Overview of engraftment for RNP plus AAV and RNP plus AAV GFP^{high} experimental groups. Average engraftment frequencies and percentage GFP^+ human cells ± s.e.m. are shown. Total number of cells transplanted was the same (500,000) for all mice in the RNP group, whereas in the GFP^{high} group, one mouse was transplanted with 100,000 cells, two mice with 250,000 cells, and three mice with 500,000 cells. The total number of HSCs transplanted per mouse (±s.e.m.) was calculated based on the frequencies of GFP^+ cells in the $CD34^+$ $CD38^-$ $CD90^+$ $CD45RA^-$ subset analysed by flow cytometry (see Extended Data Fig. 5f) directly before injection. The total number of modified human cells in the bone marrow at week 16 after transplant per mouse (±s.e.m.) was estimated based on calculations presented in the Methods. This shows that the enrichment not only resulted in a higher percentage of edited cells (column 3) but also resulted in an absolute higher number (column 6) of edited cells.

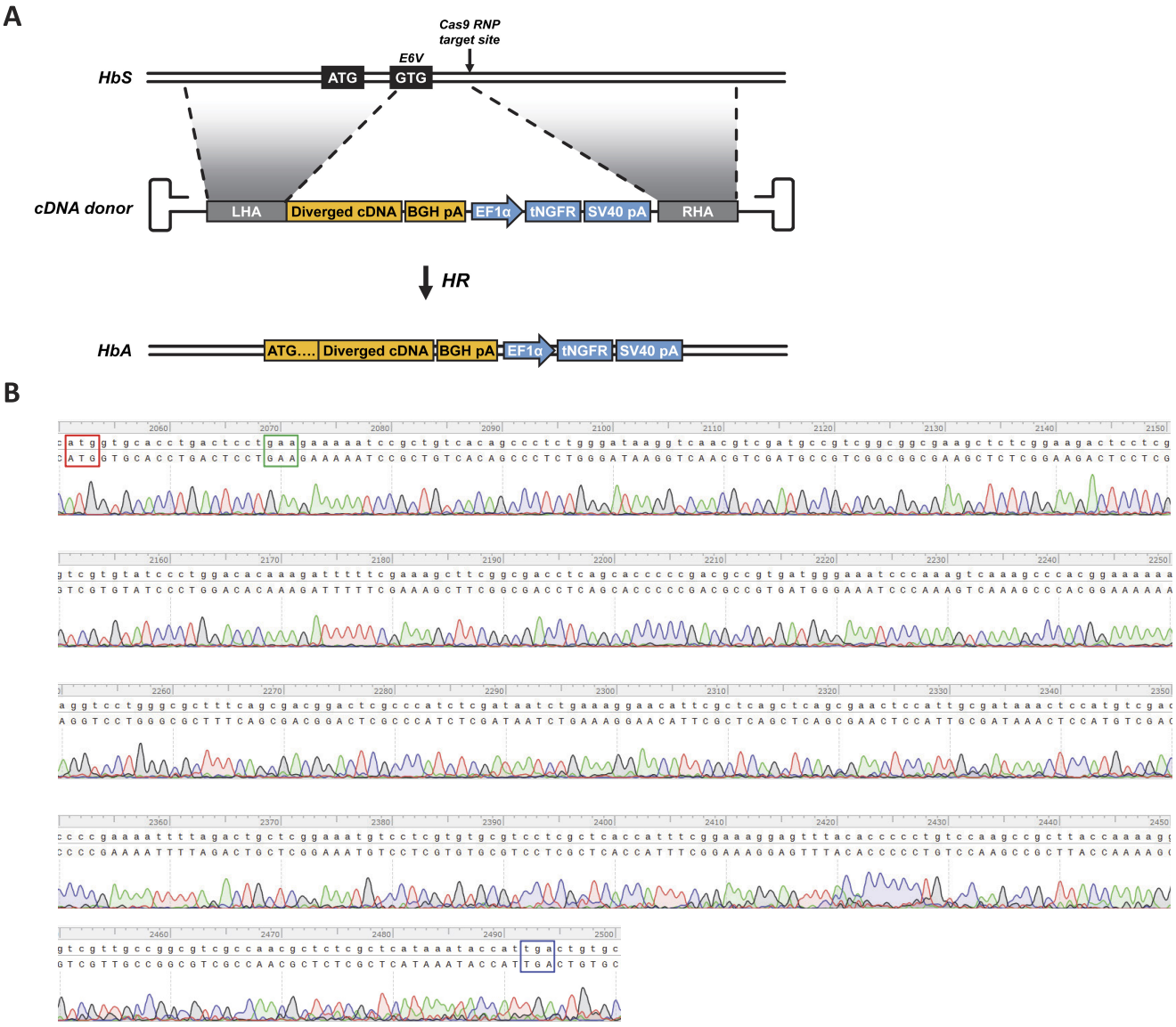


Extended Data Figure 7 | Genome-edited human HSPCs in the bone marrow of NSG mice at week 16 after transplantation. **a**, **b**, Representative FACS plots from the analysis of NSG mice from the mock (**a**) or RNP plus AAV (**b**) experimental group at week 16 after transplantation. Mice were euthanized and bone marrow was obtained, MNCs were isolated via Ficoll density gradient, after which human $CD34^+$ cells were enriched by magnetic-activated cell sorting (MACS), and finally cells were stained with anti-CD34, anti-CD38 and anti-CD10 antibodies to identify human GFP^+ cells in the $CD34^+$ $CD10^-$ and $CD34^+$ $CD10^-$ $CD38^-$ populations (note that CD10 was included as a negative discriminator for immature B cells). **c**, Collective data from the analysis of GFP^+ cells in the human $CD34^+$ $CD10^-$ population from the RNP plus AAV ($n = 11$) and RNP plus AAV GFP^{high} ($n = 6$) experimental groups. For the RNP plus AAV GFP^{high} group, cells from all six mice were pooled before analysis and thus, no error bar is available. Error bar on RNP group represents s.e.m.

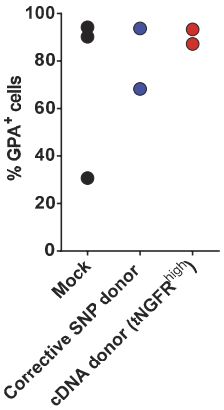




Extended Data Figure 8 | Correction of the sickle cell mutation in patient-derived CD34$^+$ HSPCs. a, Schematic overview of the sequence of the sickle cell allele aligned with the sequence of an allele that has undergone homologous recombination using the corrective SNP donor. The Glu6Val mutation in patients with SCD (A>T) is highlighted in yellow. The sgRNA recognition sequence, the PAM site and the cut site (scissors) are shown. The donor carries synonymous nucleotide changes between the sickle nucleotide and the cut site to avoid premature crossover during homologous recombination. Synonymous changes are also added to the PAM and an early nucleotide in the sgRNA target site to avoid subsequent re-cutting and potential inactivation of the corrected allele. **b**, HSPCs from two different patients with SCD were targeted with the corrective SNP donor and seeded in methylcellulose. After 14 days, in-out PCR amplicons from a total of 38 clones were sequenced and genotypes were extracted from sequence chromatograms.



Extended Data Figure 9 | Targeting *HBB* with a cDNA donor and a tNGFR expression cassette. a, Schematic representation of the AAV6 donor encoding a functional *HBB* anti-sickling cDNA (Gly16Asp, Glu22Ala, Thr87Gln) followed by an expression cassette for tNGFR. The left homology arm stops just before the sickle mutation (A→T) followed by the remaining *HBB* cDNA, which has been diverged from the endogenous sequence by introducing synonymous mutations at codon wobble positions. The *HBB* cDNA expression cassette is followed by an EF1α promoter driving tNGFR expression (*HBB* cDNA-EF1α-tNGFR), which constitutes a clinically compatible expression cassette enabling enrichment and tracking of modified cells. **b**, Chromatogram from sequencing of an in-out PCR amplicon on $CD34^+$ cells derived from patients with SCD, electroporated with *HBB* Cas9 RNP and transduced with rAAV6 *HBB* cDNA donor. PCR was performed on genomic DNA extracted 4 days after electroporation of a bulk sample. Chromatogram shows the sequence of the full *HBB* cDNA with start codon, sickle-cell codon position (containing a corrected Glu codon), and stop codon highlighted in red, green and blue boxes, respectively.



Extended Data Figure 10 | Edited HSPCs from patients with SCD differentiate into erythrocytes that express glycophorin A. $CD34^+$ HSPCs derived from patients with SCD were edited with *HBB* Cas9 RNP and either the corrective SNP donor or the cDNA donor. Four days after electroporation, cells edited with the cDNA donor were sorted for $tNGFR^+$ cells. This population as well as the populations edited with the corrective SNP donor and mock-electroporated cells were subjected to a 21-day erythrocyte differentiation protocol, followed by staining for glycophorin A (GPA). All data points within experimental groups are derived from experiments in cells from different patients with SCD, $n = 3$ (mock) and $n = 2$ (SNP and cDNA donor).