Dual α-globin and truncated erythropoietin receptor knock-in restores hemoglobin production in α-thalassemia major-derived hematopoietic stem and progenitor cells

Simon N. Chu[1,2,3], Eric Soupene[4], Beeke Wienert[5,] Han Yin[1,3], Devesh Sharma[1,2], Kun Jia[1,2], Shota Homma[6], Jessica P. Hampton[7], James M. Gardner[1,3], Bruce R. Conklin[5,8], Tippi C. MacKenzie[1,2], Matthew H. Porteus[7], & M. Kyle Cromer[1,2,9†]

[1]Department of Surgery, University of California, San Francisco, San Francisco, CA.
[2]Eli & Edythe Broad Center for Regeneration Medicine, University of California, San Francisco, San Francisco, CA.
[3]Diabetes Center, University of California, San Francisco, San Francisco, CA.
[4]Department of Pediatrics, University of California, San Francisco, Oakland, CA.
[5]Gladstone Institutes, San Francisco, CA.
[6]Department of Genetics, Stanford University, Stanford, CA.
[7]Department of Pediatrics, Stanford University, Stanford, CA.
[8]Department of Medicine, University of California, San Francisco, San Francisco, CA
[9]Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco, San Francisco, CA.

[†]Correspondence: kyle.cromer@uscf.edu

Abstract

Alpha-thalassemia is an autosomal recessive monogenic disease with increasing worldwide prevalence. The molecular basis is due to mutation or deletion of one or more duplicated α-globin genes, and disease severity is directly related to the number of allelic copies compromised. The most severe form, α-thalassemia major (αTM), results from loss of all four copies of α-globin and has historically resulted in fatality *in utero*. However, the advent of *in utero* transfusions now enables the survival of an increasing number of children with αTM. Postnatally, patients face challenges similar to β-thalassemia, including severe anemia and erythrotoxicity due to an imbalance of β-globin and α-globin chains. Hematopoietic stem cell transplantation (HSCT) as a therapeutic option is limited by donor availability and potential transplant-related complications. Despite progress in genome editing treatments for β-thalassemia, there is no analogous curative option for patients suffering from α-thalassemia. To address this, we designed a novel Cas9/AAV6-mediated genome editing strategy that integrates a functional α-globin gene into the β-globin locus in αTM patient-derived hematopoietic stem and progenitor cells (HSPCs). Incorporation of a truncated erythropoietin receptor transgene into the α-globin integration cassette dramatically increased erythropoietic output from edited HSPCs and led to the most robust production of α-globin, and consequently normal hemoglobin. By directing edited HSPCs toward increased production of clinically relevant RBCs instead of other divergent cell types, this approach has the potential to mitigate the limitations of traditional HSCT for the hemoglobinopathies, including low genome editing and low engraftment rates. These findings support development of a definitive *ex vivo* autologous genome editing strategy that may be curative for α-thalassemia.

Introduction

Alpha-thalassemia is one of the most common monogenic diseases in the world. It is currently estimated that 5% of the world's population harbors an α-thalassemia variant[1]. While the carrier frequency is highest in those with Southeast Asian heritage, data suggests that these rates are rising due to population growth, human migration, and advances in treating milder forms of the disease[2]. Humans harbor four copies of α-globin genes (a direct tandem repeat on chromosome 16), and disease severity directly correlates with the number of mutated or deleted alleles. The most severe form of the disease, in which all four copies are disrupted, is called α-thalassemia major (αTM). While this disease has historically been lethal *in utero*, patients may now survive to birth after *in utero* blood transfusions, often with excellent neurologic outcomes[3,4]. Postnatally, these patients present with a disease similar to β-thalassemia, with the absence of α-globin leading to an inability to form hemoglobin heterotetramers and subsequent severe anemia. Moreover, the accumulation of orphan β-globin chains, which form oxidized, covalently linked dimers and an unstable homotetramer (Hemoglobin H), leading to erythrotoxicity and hemolysis[5]. Similarly, these patients require chronic transfusions, which often results in iron overload and need for iron chelation therapy[6]. Although allogeneic-hematopoietic stem cell transplantation (HSCT) may provide a potential cure, suitable matched donors are only available in a minority of cases and carry a risk of immune rejection and graft-versus-host-disease[7]. Furthermore, while numerous gene therapy and genome editing strategies have been developed for patients with β-thalassemia[8,9,10] there are no such therapies for the most severe forms of α-thalassemia, indicating a major unmet medical need for this patient population.

Genome editing-based treatment strategies often attempt to correct the most common point mutations for a specific disorder, or to insert a functional transgene into the endogenous locus. However, these would not be viable options for αTM as the disease is most often caused by large deletions of all four copies of α-globin. In addition, there is evidence that disease severity is determined by both loss of α-globin leading to severe anemia and accumulation of toxic β-globin dimers in the absence of their binding partner[11]. As a result, an ideal α-thalassemia treatment strategy would address both factors in order to correct the β-globin:α-globin imbalance in patient-derived red blood cells (RBCs). Given prior success using Cas9/AAV6-mediated genome editing to integrate a copy of the *β-globin* gene at one of the α-globin loci for treatment of β-thalassemia[8], we hypothesized that an analogous strategy—integration of the *α-globin* gene at the *β-globin* locus—could be used to treat α-thalassemia.

In this work, we developed a novel Cas9/AAV6-mediated genome editing strategy to integrate a functional copy of the *α-globin* gene into the *β-globin* locus in αTM patient-derived

hematopoietic stem and progenitor cells (HSPCs) as a universal treatment strategy for the disorder - an approach that could be curative for any αTM patient regardless of the specific causative mutations or deletions. This approach allows both correction of the underlying disease as well as restoration of the β-globin:α-globin imbalance by placing the *α-globin* transgene under erythroid-specific expression of the *β-globin* locus. Taking a cue from human genetics, we also incorporated a truncated erythropoietin receptor (*tEPOR*) cDNA into the *α-globin* integration cassette, allowing for simultaneous correction of α-thalassemia and increased erythropoietic output from edited HSPCs[12]. By directing edited HSPCs toward increased production of clinically relevant RBCs instead of other divergent cell types, this approach has the potential to overcome many of the clinical challenges previously seen in HSCT for the hemoglobinopathies, including low editing and engraftment rates, as well as high morbidity and mortality from prerequisite myeloablative regimens[13,14,15]. Similar to our prior work with β-thalassemia[8], this strategy could be used in the development of an autologous-HSCT treatment for αTM, overcoming the shortage of matched donors, ameliorating the risk of immune rejection and graft-versus-host disease, and increasing the safety and efficacy of stem cell transplantation.

Results

Efficient cleavage with HBB intron-targeting CRISPR gRNAs

Due to the high editing frequencies achieved in primary HSPCs in prior work[8,16], we used a Cas9/AAV6-mediated genome editing strategy to knock an *α-globin* (*HBA*) transgene into the *β-globin* (*HBB*) locus in patient-derived HSPCs. To do so, we first designed and screened Cas9 guide RNAs (gRNAs) at the *HBB* locus. So that cleavage alone without homology-directed repair (HDR) would not disrupt β-globin production, we chose 22 candidate gRNAs located in intron 1 or intron 2 as well as the 3' UTR (Figure 1A and Supplemental Table 1). To determine cleavage frequencies, we delivered candidate gRNAs pre-complexed with Cas9 protein (forming a ribonucleoprotein (RNP) complex) to the human HUDEP-2 cell line via electroporation[17]. Three days post-targeting, genomic DNA was harvested and subjected to PCR amplification of the region surrounding the expected cleavage site. Insertion and deletion (indel) frequencies of the corresponding Sanger sequences were then quantified using ICE analysis[18]. We found that targeting the 5' and 3' UTR regions of *HBB* proved not to be feasible, as the former had significant homology to δ-globin and the latter had no gRNAs with detectable cleavage. We found that the most efficient gRNAs with greater than 75% indel frequencies corresponded sg7 in intron 1, and sg11 and 13 in intron 2 (Figure 1A and Supplemental Figure 1A)[19]. In spite of efficient cleavage, none of the 22 candidate gRNAs were found to noticeably disrupt β-globin production as seen by

Western blot (Supplemental Figure 1B). *In silico* off-target analysis of these guides using COSMID[20,21] revealed that sg7 had the most favorable predicted off-target profile (Figure 1B). On further evaluation of the 19 predicted off-target sites for sg7, all were found to reside in non-coding regions of the genome (Figure 1C).

Efficient integration of HDR templates at HBB locus

Following identification of effective *HBB* intron-targeting gRNAs, we proceeded to develop DNA repair templates packaged in AAV6 delivery vectors that could effectively mediate HDR at these cleavage sites (termed Intron 1 and Intron 2 donors; Figure 1D). Each integration cassette was comprised of a SFFV promoter driving expression of a GFP reporter to allow rapid readout of integration frequencies via flow cytometry. While the right homology arm of each DNA repair template corresponded to the ~900bp immediately downstream of the intron 1 or intron 2 Cas9 cut site, the left homology arm was split away from the cleavage sites to correspond to the ~900bp immediately upstream of the start codon of the endogenous *HBB* gene. As in a prior study[8], this split left homology arm is expected to allow promoterless integration cassettes—ultimately an *α-globin* transgene in this work—to be driven by the regulatory machinery of the endogenous locus.

Once DNA repair templates were assembled and packaged into AAV6 vectors, we tested each integration strategy by delivering the most effective gRNAs (sg7, 11, or 13) pre-complexed with high-fidelity Cas9 RNP[22] to wild-type (WT) human primary $CD34^+$-enriched HSPCs via electroporation. We then transduced the AAV6 vector corresponding to either intron 1 or intron 2 integration schemes immediately following electroporation. Five days later, we analyzed targeting frequencies by flow cytometry (Figure 1E and Supplemental Figure 2A) and found that the Intron 1 integration strategy was significantly more efficient (median of 17.1% GFP^+ cells) compared to Intron 2 integration (median of 12.5% for sg11 and 13.3% for sg13; $p<0.05$) (Figure 1F), perhaps due to the shorter distance that the left homology arm was split away from the Cas9 cleavage site. In addition, the Intron 1 integration scheme yielded a slightly higher mean fluorescence intensity (MFI) per edited cell compared to Intron 2 integration (Figure 1G). Given the high cleavage and integration frequency, as well as the most favorable off-target profile, we proceeded with further testing in primary WT HSPCs using sg7 and the Intron 1 integration strategy.

Integration at HBB drives erythroid-specific expression of transgenes

A key goal of this work is to achieve erythroid-specific expression of an *α-globin* transgene. To test whether the *HBB* Intron 1 integration strategy is able to achieve this, we designed custom promoterless integration cassettes which included either the full-length *HBA1* or *HBA2* transgene

(i.e., including all UTRs, exons, and introns), as well as a YFP reporter linked by a T2A cleavage peptide at the C-terminus of α-globin to allow quantification of HDR frequency by flow cytometry and to serve as a surrogate for α-globin protein production (Figure 2A). Because the specific UTRs flanking an integrated transgene can have a great bearing on expression[8], we also designed integration strategies that would allow our *HBA* transgene to be flanked by *HBB*, *HBA1*, and *HBA2* 5' and 3' UTRs.

To determine whether integration of these promoterless *α-globin* cassettes at *HBB* yielded erythroid-specific expression, we edited WT HSPCs as described before using electroporation-mediated delivery of Cas9:sg7 RNP followed by transduction with AAV6 vectors containing each repair template. We then performed an established, 14-day erythroid differentiation protocol[23] and quantified YFP expression and RBC differentiation (staining with antibodies for CD34, CD45, CD71, and GPA) via flow cytometry (Figure 2B and Supplemental Figure 2B). We found that all conditions efficiently differentiated into erythroid cells compared to mock (electroporation only) or AAV-only controls (Figure 2C). However, we did observe differences among the integration vectors in terms of HDR frequency and MFI of edited cells, which is an indirect readout for *α-globin* expression levels. We found that the vector integrating the *HBA1* transgene and *HBA1* UTRs achieved the most highly HDR-edited population of RBCs (median of 63% YFP^+ cells; Figure 2D) and the second highest MFI of the HDR-edited cell population (Figure 2E). As confirmation that these integration strategies achieved erythroid-specific transgene expression, we observed high YFP fluorescence in cells that acquired RBC markers ($CD71^+/GPA^+$) in all editing conditions but no fluorescence above background in cells retaining HSPC markers ($CD34^+/CD45^+$) (Figure 2F). We also tracked differentiation and YFP expression kinetics and found that all editing conditions proceed through RBC differentiation at a similar rate (Figure 2G). When monitoring expression of fluorescent reporters, we find that cells edited with the constitutive SFFV-GFP integration vector maintain high expression over the course of RBC differentiation, whereas all three conditions edited with HBA-2A-YFP constructs display increasing fluorescence that dovetails with RBC differentiation during the 14-day protocol (Figure 2H). Collectively, these results indicate that integration of the *HBA* transgene at *HBB* achieves high levels of erythroid-specific transgene expression.

α-globin integration in αTM ***HSPCs yields low levels of hemoglobin***

While T2A-YFP reporters allow rapid readout of HDR-editing frequency, to generate clinically relevant constructs we created *HBA1* integration vectors without T2A-YFP that were flanked by either *HBB* or *HBA1* UTRs (Figure 3A). To test these new vectors, we edited WT HSPCs as

before and found that both vectors mediate high frequencies of HDR editing (median of 36.6% and 42.3% HDR-edited alleles for *HBB* and *HBA1* UTR vectors, respectively) as measured by a custom droplet digital PCR (ddPCR) assay (Figure 3B).

We determined whether *α-globin* integration can restore hemoglobin tetramer production in patient-derived HSPCs following erythroid differentiation. We isolated $CD34^+$ HSPCs from a bone marrow aspirate that was taken from a ~1-year-old patient with αTM. We then edited and differentiated HSPCs as before and found that all conditions were able to effectively differentiate into RBCs compared to the unedited control (Figure 3C). Following ddPCR analysis, we found that cells were edited at a median of 25.8% and 20.1% HDR-edited alleles for *HBB* and *HBA1* UTR vectors, respectively (Figure 3D). We then performed hemoglobin tetramer high-performance liquid chromatography (HPLC) analysis on edited patient-derived αTM cells at the end of erythroid differentiation and found that HDR-edited cells displayed a modest increase in all three alpha-containing hemoglobin [fetal hemoglobin (HbF), adult hemoglobin (HbA), and hemoglobin A2 (HbA2)] compared to WT (Figure 3E-H and Supplemental Figure 3). However, when compared to the robust HbF and HbA level obtained from RBCs derived from a healthy donor, single globin chain analysis found that the levels of α-globin produced were on average 2% of that found in WT erythroid cells (Figure 3I). While these results demonstrate for the first time that gene therapy or genome editing may be used to increase α-globin production in αTM patient-derived RBCs, it is unlikely that this low level of α-globin production will lead to a significant clinical benefit.

Dual integration of α-globin and tEPOR increases hemoglobin production in αTM ***HSPCs***

Following the finding that *α-globin* integration into *HBB* does not substantially restore hemoglobin production, we sought alternative strategies to increase the therapeutic potential of this approach. As described in a study of a cohort that included an Olympic gold medalist cross-country skier[24], a premature stop codon in the erythropoietin receptor (EPOR) was linked with a condition called congenital erythrocytosis, which is characterized by a non-pathogenic hyper-production of RBCs. In subsequent work[12], we demonstrated that genome editing-mediated integration of a truncated EPOR (*tEPOR)*-coding cDNA could increase erythropoietic output from primary human HSPCs. Thus, we hypothesized that co-expression of *tEPOR* cDNA from the current *α-globin* integration cassettes could lend a selective advantage to genome-edited and, therefore, α-globin-expressing erythroid cells. To test this, we added a *tEPOR* cDNA at the C-terminus of the α-globin integration cassette with expression linked by an internal ribosome entry site (IRES) (Figure 4A). We then edited WT HSPCs with our *HBA1*-UTRs. α-globin integration vector from Figure 3 and compared

it against the bicistronic *HBA*-IRES-*tEPOR* vector in WT HSPCs. In both instances, edited cells efficiently differentiated into erythroid cells compared to unedited controls (Figure 4B). When measuring HDR-editing frequencies by ddPCR, we found that cells edited with α-globin integration alone were edited at a median of 39.2% HDR-edited alleles at d0 and 38.4% at d14. However, HDR-editing frequencies in cells edited with the dual α-globin and *tEPOR* increased significantly over the course of erythroid differentiation, rising from a median of 29.3% HDR-edited alleles at d0 to 69.9% at d14 (*** p=0.0004; * p=0.01) (Figure 4C). As in prior work, these results suggest that cells expressing a *tEPOR* have a significant competitive advantage over unedited cells throughout the course of RBC differentiation. We next tested these vectors in αTM patient-derived HSPCs and again observed that edited and unedited cells were able to effectively undergo erythroid differentiation (Figure 4D). As with WT cells, we found that HDR-editing frequencies among cells targeted with the *α-globin* alone vector maintained a consistent HDR-editing rate over the course of RBC differentiation (Figure 4E). However, we again observed a dramatic increase in HDR-editing frequencies in cells edited with the α-globin+*tEPOR* cassette, rising from a median of 40.5% HDR-edited alleles at d4 to 60.5% at d16 (** p=0.0088; * p=0.0230)**.** In addition to the increase in HDR-editing frequencies, we also observed elevated cell counts over the course of RBC differentiation in conditions edited with the α-globin+*tEPOR* cassette (Figure 4F), indicating that *tEPOR* expression is driving an increased erythropoietic output from genome-edited HSPCs.

Next, to determine whether this strategy was able to restore α-globin production in the context of αTM, we performed HPLC Hb analysis on our edited patient-derived HSPCs following RBC differentiation, as before. We found that addition of *tEPOR* to the *α-globin* integration cassette dramatically increased the formation of hemoglobin tetramers, with a substantial increase in fetal hemoglobin (HbF), adult hemoglobin (HbA), and hemoglobin A2 (HbA2) (Figure 4G-I). Single globin chain HPLC confirmed that α-globin production was indeed elevated in HDR-edited samples (Supplemental Figure 4A-B), likely driving the increases in hemoglobin tetramer formation that we observed. Notably, the average percentage of α-globin production in α-globin+*tEPOR*-edited cells increased nearly 5-fold compared to the average percentage of α-globin integration alone (*** p<0.0001; ** p=0.0003) (Figure 4J); this increase corresponded to α-globin levels with a median of 10.3% of that found in wild-type erythroid cells. We next validated these results using Western blot for α-globin and β-globin and found that not only do cells edited with α-globin+*tEPOR* show a dramatic increase in α-globin compared to unedited and α-globin alone-edited patient-derived RBCs, but α-globin+*tEPOR*-edited cells also show a decrease in formation of toxic β-globin dimers to nearly WT levels (Figure 4K-L). Given these results, it is apparent that dual integration of α-globin+*tEPOR* has significant advantages over α-globin

integration alone: namely, an increase in α-globin production, a decrease in toxic β-globin dimer formation, and an increase in total erythropoietic output from edited HSPCs.

RNA-sequencing demonstrates restoration of oxygen transport and iron ion processes to* αTM *HSPCs

To define the cellular impact of our editing strategies, we used RNA-sequencing to reveal the transcriptional profile of αTM patient-derived HSPCs edited with α-globin as well as α-globin+*tEPOR* integration vectors in comparison to WT unedited HSPCs. Following editing and *in vitro* RBC differentiation using our described workflow, total RNA was isolated from edited αTM patient cells as well as a healthy WT control. Bulk RNA-sequencing was then performed, followed by comprehensive quality control prior to downstream statistical analysis (Figure 5A). All samples were found to have adequate data quality based on high mean and per-sequence quality scores (Supplemental Figure 5). Total normalized read counts across samples ranged from 17.6-22.0M and genomic alignment identified reads mapping to 17,776 unique genes. When examining normalized reads across the various samples, we found consistent expression across a number of RBC-associated genes in both healthy control and patient-derived cells, which is expected given the high frequency of erythroid differentiation observed in edited and unedited samples (Supplementary Figure 6). While we found that the most highly expressed genes by WT RBCs are the adult globins (*HBA1*, *HBA2*, and *HBB*), unedited "mock" αTM RBCs expressed no detectable *HBA1* or *HBA2* (Figure 5B). These findings were expected due to genetic confirmation of 4-gene α-globin deletion in this patient. When investigating the genes with the greatest rank change between WT RBCs compared to αTM RBCs, it is unsurprising that the top genes are *HBA1* and *HBA2*, as well with hemoglobin subunit mu (*HBM*), for which there is little regulatory information (Figure 5C). Despite the modest increase in α-globin production by *α-globin* integration alone according to HPLC, we observed substantial restoration of *HBA1* and *HBA2* expression compared to mock (*HBA1* and *HBA2* rising from undetectable in αTM mock to the 99.9th and 71.7th percentile of expressed genes, respectively) (Figure 5B). Comparing α-globin-integrated αTM cells to mock, we found that *HBA1* is the gene that displays the greatest rank change (Figure 5C). Predictably, this restoration of *α-globin* gene expression was further amplified in cells edited with the dual the α-globin+*tEPOR* vector, with *HBA1* and *HBA2* expression increasing to the 98.9th and 99.9th percentiles, respectively (Figure 5B). Because this editing strategy places *tEPOR* under the strong, erythroid-specific *HBB* promoter, we also found *EPOR* expression was elevated compared to all other conditions. These elevated expression levels are

confirmed by the fact that *HBA1*, *HBA2*, and *EPOR* are among the top four genes with the greatest rank change over unedited αTM cells (Figure 5C).

Shifting focus to the fetal and embryonic globins, we found that *γ-globin* genes (*HBG1* and *HBG2*) were more highly expressed in all αTM samples, regardless of editing, compared to the WT control (Figure 5D). This was expected as the αTM HSPCs were derived from a ~1-year-old patient, whereas the WT HSPCs were derived from mobilized peripheral blood of an adult donor. In fact, the repressor of γ-globin, *BCL11A*[25], was most highly expressed in WT cells compared to αTM cells, which may be responsible for the lower fetal hemoglobin levels in this older HSPC donor (Supplemental Figure 6). Interestingly, we found that ζ-globin was highly elevated compared to WT cells, and even more highly expressed than *HBG1* across all αTM samples (among the eight most highly expressed genes in unedited and edited conditions) in spite of normally being expressed during the first three months of gestation (Figure 5D). ζ-globin was also the gene with the greatest rank change in expression in all αTM conditions compared to WT control cells (Figure 5E).

To identify large-scale changes in biological processes occurring as a result of genome editing, gene ontology (GO) enrichment analysis was performed across the top 0.001% (eighteen) most highly expressed genes across all samples. Compared to the unedited mock αTM condition, α-globin integration alone led to enrichment of various iron ion-associated pathways to equivalent levels observed in the WT control (Figure 5F). Further improvements were observed in αTM cells edited with the dual α-globin+*tEPOR* integration strategy, showing enrichment of oxygen and gas transport as well as hydrogen peroxide-associated pathways to levels observed in the WT control (Figure 5F). In addition to assessing GO enrichment on the most highly expressed genes within a given sample, we also performed this analysis on the top 0.001% of genes with the greatest rank change across samples. Unsurprisingly, we found that gas transport and gas metabolism were the most significantly enriched pathways when comparing WT to unedited mock αTM cells (Figure 5G). When performing the opposite comparison, for genes with the greatest rank change in unedited αTM cells vs. WT, we found that the most significant pathways enriched were for response to incorrect and unfolded protein (Figure 5H), which may represent the cellular response to the formation of toxic β-globin dimers in absence of its α-globin binding partner. Interestingly, when comparing αTM cells edited with α-globin or α-globin+tEPOR integration strategies to WT cells, we saw no significantly enriched GO pathways (with the exception of “cellular response to heat” in the α-globin alone condition). Taken together with our Western blot data (Figure 4I), these findings suggest that while unedited αTM cells have highly elevated β-globin dimers, even a

modest increase in α-globin is able to mitigate the unfolded protein response that occurs in the context of αTM pathology. As with WT cells, we also found no significantly upregulated GO pathways when comparing α-globin-edited αTM cells to unedited αTM cells (Alpha-thal Supplemental Data). However, when comparing α-globin+tEPOR-edited cells to unedited αTM cells, we found a number of upregulated GO pathways, the most significantly enriched being associated with cytokine-mediated signaling (likely a result of elevated EPOR signaling) and hydrogen peroxide catabolism (Figure 5I).

Discussion

Given the increasing global burden of αTM and tremendous cost of disease management for patients and health care systems, there is an urgent need for curative treatments for this disease. To our knowledge, these results demonstrate for the first time that gene therapy or genome editing may be used to restore normal hemoglobin production to αTM patient-derived RBCs. This approach addresses the two major elements contributing to molecular pathology of the disease in a single genome editing event: successfully increasing α-globin production as well as reducing the formation of toxic β-globin dimers without noticeably disrupting β-globin formation. While the cells in this work were derived from patients with four-gene-deletion αTM, our editing strategy may also be effective in other clinical phenotypes of α-thalassemia such as in patients with inactivation of three α-globin genes or hemoglobin H disease, where most patients have received multiple transfusions by the age of 10[26].

In addition to these protein-level changes, deeper analysis reveals several interesting findings at a transcriptional level. The enrichment of pathways associated with various oxidative pathways in α-globin+*tEPOR*-edited cells suggests an improvement in the stability and function of edited RBCs. Most notably, hydrogen peroxide (H_2O_2) catabolism and metabolism has been identified as a critical process in preventing cellular injury in β-thalassemic erythrocytes[27], as the auto-oxidation of unpaired globin chains results in the generation and release of significant amounts of superoxide (O_2^-) and H_2O_2. These observations suggest that our novel genome editing strategy not only restores *HBA1* and *HBA2* gene expression, but also re-establishes essential antioxidant pathways that may be critical in addressing the pathophysiology of α-thalassemia. Moreover, the high expression of ζ-globin found in αTM cells hints at a possible compensatory mechanism for the loss of α-globin in the cell. These findings support the hypothesis that the presence of ζ-globin in patients with α-globin deletions may be protective, as spontaneous survivors of αTM without an *in utero* transfusion tend not to have deletions that extend into the ζ-

globin region[4]. While therapeutic approaches have been developed for β-thalassemia to increase HbF by phenocopying hereditary persistence of fetal hemoglobin (HPFH), either by negating the erythroid enhancer of *BCL11A* or inducing mutations in the γ-globin promoters, an analogous genome editing strategy has not yet been developed for α-thalassemia. This is in part due to the absence of a detailed understanding of the switch from ζ-globin to α-globin that transpires during embryonic development. Further investigation to identify the natural repressor of ζ-globin and approaches to ζ-globin upregulation as a potential treatment for α-thalassemia are ongoing[28].

In pursuit of restoring α-globin production in αTM cells, we discovered that pairing our α-globin transgene with a naturally occurring *EPOR* variant led to the greatest restoration of α-globin production by substantially increasing erythropoietic output of α-globin-expressing αTM patient-derived cells. By guiding edited HSPCs to prioritize the production of clinically relevant corrected RBCs over other cell types, this method may resolve certain clinical challenges previously encountered in HSCT for the hemoglobinopathies. A recent review of the literature found that of all allogeneic-HSCTs that had been reported for αTM, 20% developed transplant-related complications including graft failure and death[29]. Poor outcomes are often linked to low engraftment rates in the context of allogeneic-HSCT, leading to primary graft failure, as well as to low rates of potentially curative genome editing in the context of autologous-HSCT. As a solution, the concept of driving increased erythropoietic production via expression of *tEPOR* has the potential to lower editing and engraftment frequencies needed to correct RBC disorders. Moreover, as patients with severe anemia—such as those with αTM —have increased levels of circulating erythropoietin, a *tEPOR*-based genome editing strategy may lower the threshold for edited HSC engraftment needed to achieve a lifelong cure for disease. In doing so, this strategy has the potential to reduce or eliminate the need for high-morbidity myeloablation currently required to clear the HSC niche, which stand as a major barrier to HSCT safety and efficacy.

In summary, we present results of a comprehensive genome editing approach for the treatment of α-thalassemia that restores α-globin production, normal hemoglobin formation, and addresses the adverse consequences of unpaired β-globin aggregates. We believe that these findings support development of a definitive *ex vivo* autologous genome editing strategy that may be curative for patients with α-thalassemia.

Methods:

HUDEP-2 cell culture and nucleofection

HUDEP-2 cells were grown in StemSpan SFEM (STEMCELL Technologies, Vancouver, Canada) supplemented with 50ng/mL stem cell factor, 3U/mL erythropoietin (both PeproTech, Rocky Hill, NJ, USA), 10^{-6}M dexamethasone, and 1µg/mL doxycycline (both Sigma-Aldrich, St. Louis, MO, USA). Media was made fresh before use and supplemented with 1× penicillin, streptomycin, and L-glutamine (Thermo Fisher Scientific, Waltham, MA, USA). Cells were maintained at a concentration of $5x10^5$ cells/mL. Cells were transfected with 100pmoles of Cas9 RNPs by nucleofection using a Neon Transfection System. 10^5 cells were resuspended in nucleofection buffer T (Neon Transfection Kit) and given one pulse of 1100 V for 20 ms. Cells were harvested 72 hours after nucleofection for downstream experiments. Synthetic chemically modified gRNAs used to edit HSPCs were purchased from Synthego (Redwood City, CA, USA) and were purified by HPLC. These modifications include 2′-O-methyl-3′-phosphorothioate at the three terminal nucleotides of the 5′ and 3′ ends described previously[30]**.** All gRNA sequences tested are shown in Supplemental Table 1.

Indel frequency analysis by ICE

HUDEP-2 cells were washed with PBS and DNA extracted using QuickExtract DNA extraction solution (Biosearch Technologies, Hoddesdon, UK) following the manufacturer's instructions. To quantify genome editing outcomes, a PCR was performed across the *HBB* gene spanning the gRNA cut sites. Amplicons were Sanger sequenced and indels quantified using ICE software[18] (Synthego).

***In vitro* culture of HSPCs**

Human $CD34^+$ HSPCs were cultured in conditions as previously described[16,31,32,33,34]. $CD34^+$ HSPCs were isolated from umbilical cord blood (provided by Stanford Binns Program for Cord Blood Research) or sourced from Plerixafor- and/or G-CSF-mobilized peripheral blood (AllCells, Alameda, CA, USA and STEMCELL Technologies). Bone marrow aspirates taken from patients with αTM under protocol no. 16-21157, which was approved and is renewed annually by the NHLBI Institutional Review Board (IRB). The patients provided informed consent for the study and samples were de-identified after collection. $CD34^+$ HSPCs were isolated using a EasySep Human CD34 Positive Selection Kit II according to the manufacturer's protocol (STEMCELL Technologies). $CD34^+$ HSPCs were cultured at $1×10^5$ cells/mL in StemSpan Serum-Free Expansion Medium II (STEMCELL Technologies) or Good Manufacturing Practice Stem Cell Growth Medium (SCGM, CellGenix, Freiburg, Germany) supplemented with a four human cytokine cocktail (PeproTech) that included 100ng/mL stem cell factor, 100ng/mL thrombopoietin,

100ng/mL Fms-like tyrosine kinase 3 ligand, 100ng/mL interleukin-6, 20mg/mL streptomycin, and 20U/mL penicillin. The cell incubator conditions were 37°C, 5% CO_2, and 5% O_2.

Genome editing of primary HSPCs

Cas9 protein was purchased from Integrated DNA Technologies (Coralville, IA, USA) or Aldevron (Fargo, ND, USA). Preceding electroporation, RNPs were complexed using above described gRNAs at a Cas9:gRNA molar ratio of 1:2.5 at 25°C for 10-20 min. Next, HSPCs were resuspended in P3 buffer (Lonza, Basel, Switzerland) with complexed RNPs and subsequently electroporated using a Lonza 4D Nucleofector (program DZ-100). Electroporated cells were then plated at 10^5 cells/mL in the previously described cytokine-supplemented media. Immediately following, AAV6 was added to cells at 5×10^3 vector genomes/cell based on titers determined using a Bio-Rad QX200 ddPCR machine and QuantaSoft software (v.1.7, Bio-Rad, Hercules, CA). The small molecule AZD-7648, a DNA-dependent protein kinase catalytic subunit inhibitor, was also added to cells to improve HDR frequencies as previously reported[35].

Gene-targeting analysis by flow cytometry

5d post-targeting with fluorescent reporter vectors (Figure 1D-G), $CD34^+$ HSPCs were harvested and the percentage of edited cells was determined by flow cytometry. This analysis was performed using either the Accuri C6 flow cytometer and software (v.9.4.11; BD Biosciences, San Jose, CA, USA) or the FACS Aria II cytometer and FACS Diva software (v.8.0.3; BD Biosciences). The data were subsequently analyzed using FlowJo (v.10.6.1; FlowJo LLC, BD Biosciences).

Allelic targeting analysis by ddPCR

Genome-edited HSPCs were harvested within 2d-4d post-electroporation and analyzed for modification frequencies of the alleles of interest. Our custom ddPCR analysis utilized a two-probe set where one probe binds the WT sequence (WT) and another binds unedited reference (REF) sequence outside of DNA repair template homology arms but residing on the same BamHI-digested gDNA fragment. Quantification of WT alleles, rather than HDR alleles, was necessary due to the large amplicons that would be required for in-out PCR given the long (~900bp) homology arms of our DNA repair templates. To calculate allele frequencies, we divided the WT copies/µL by the REF copies/µL to yield frequency of unedited WT alleles. Subtracting WT allele frequencies from 1 then yielded our HDR editing frequencies. The following primers and probes were used in the ddPCR reaction:

HBB WT unedited alleles:

Forward primer: 5'-ACGTGGATGAAGTTGGTGGT-3'

Reverse primer: 5'-CAGCATCAGGAGTGGACAGA-3'

Probe: 5'-6-FAM/AGAAACTGGGCATGTGGAGA/3IABkFQ-3'

HBB REF alleles:

Forward primer: 5'-CGATCACGTTGGGAAGCTAT-3'

Reverse primer: 5'-GAGTGTCTGGGGGAACAAAA-3'

Probe: 5'-6-FAM/CTTTTCCCCTCCTACCCCTA/3IABkFQ-3'

AAV6 vector design, production, and purification

Adeno-associated virus, serotype 6 (AAV6) vector plasmids were cloned into the pAAV-MCS plasmid (Agilent Technologies, Santa Clara, CA, USA), comprised of inverted terminal repeats (ITRs) derived from AAV2. Gibson Assembly Mastermix (New England Biolabs, Ipswich, MA, USA) was used for the creation of all DNA repair vectors according to manufacturer's instructions. AAV6 vector was produced with little variation as previously described[36]. 293T cells (Life Technologies, Carlsbad, CA, USA) were seeded in five 15 cm^2 dishes with 1.3-1.5×10^7 cells per plate 24hr pre-transfection. Each dish was then transfected with 112µg polyethylenimine (cat. 23966; Polysciences, Inc. Warrington, PA, USA), 6µg ITR-containing plasmid, and 22µg pDGM6 (gift from David Russell, University of Washington, Seattle, WA, USA), which expresses AAV6 cap, AAV2 rep, and Ad5 helper genes. Following a 48-72h incubation, cells were harvested, and vectors were purified using the AAVpro purification kit (cat.: 6666; Takara Bio, San Jose, CA, USA) as per manufacturer's instructions and stored at -80 °C until further use. AAV6 vectors were titered using ddPCR to measure the number of vector genomes as described above[37].

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

Following targeting, HSPCs derived from healthy donors or αTM patients were cultured for 14-16d at 37°C and 5% CO_2 in SFEM II medium (STEMCELL Technologies) as previously described[14,38]. SFEM II base medium was supplemented with 100U/mL penicillin–streptomycin, 10ng/mL stem cell factor, 1ng/mL interleukin-3 (PeproTech), 3U/mL erythropoietin (eBiosciences, San Diego, CA, USA), 200µg/mL transferrin (Sigma-Aldrich), 3% antibody serum (heat-inactivated; Atlanta Biologicals, Flowery Branch, GA, USA), 2% human plasma (from umbilical cord blood), 10µg/mL insulin (Sigma-Aldrich), and 3U/mL heparin (Sigma-Aldrich). 2-3d post electroporation, cells were transitioned into the first phase of differentiation media for a total of 7d

and maintained at 10^5 cells/mL. At d7, cells were transitioned to the above media without interleukin-3 and maintained at 10^5 cells/mL. At d11, cells were transitioned to the above media without interleukin-3 and with transferrin increased to 1mg/mL and were maintained at 10^6 cells/mL.

Immunophenotyping of differentiated erythrocytes

Differentiated erythrocytes were analyzed by flow cytometry for erythrocyte lineage-specific markers using a FACS Aria II (BD Biosciences). Edited and non-edited cells were analyzed using the following antibodies: hCD45 V450 (HI30; BD Biosciences), CD34 APC (561; BioLegend, San Diego, CA, USA), CD71 PE-Cy7 (OKT9; Affymetrix, Santa Clara, CA, USA), and CD235a PE (GPA) (GA-R2; BD Biosciences). To measure viability, cells were also stained with Ghost Dye Red 780 (Tonbo Biosciences, San Diego, CA, USA).

Hemoglobin tetramer analysis

Frozen pellets of approximately 10^6 cells *in vitro*-differentiated erythrocytes were thawed and lysed in 30µL of RIPA buffer with 1x Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific) for 5 minutes on ice. The mixture was vigorously vortexed and cell debris were removed by centrifugation at 13,000 RPM for 10mins at 4ºC. HPLC analysis of hemoglobins in their native form was performed on a cation-exchange PolyCAT A column ($35 \times 4.6mm^2$, 3µm, 1,500Å; PolyLC Inc., Columbia, MD, USA) using a PerkinElmer Flexar HPLC system (PerkinElmer, Waltham, MA, USA) at room temperature with detection at 415nm. Mobile phase A consisted of 20mM Bis-tris and 2mM KCN at pH 6.94, adjusted with HCl. Mobile phase B consisted of 20mM Bis-tris, 2mM KCN, and 200mM NaCl at pH 6.55. Hemolysate was diluted in buffer A prior to injection of 20µL onto the column with 8% buffer B and eluted at a flow rate of 2mL/min with a gradient made to 40% B in 6mins, increased to 100% B in 1.5mins, and returned to 8% B in 1min, and re-equilibrated for 3.5mins. Quantification of the area under the curve of the peaks were performed with TotalChrom software (PerkinElmer) and raw values were exported to GraphPad Prism (version 9, GraphPad Software, Boston, MA, USA) for plotting and further analysis.

Single globin chain analysis

Frozen pellets of approximately 10^6 cells *in vitro*-differentiated erythrocytes were thaw and lysed in 30µL of RIPA buffer with 1x Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific) for 5mins on ice. The mixture was vigorously vortexed and cell debris were removed by centrifugation at 13,000 RPM for 10mins at 4ºC. Analysis of globin chains was performed by reverse-phase

PerkinElmer Flexar HPLC system (PerkinElmer) with a Vydac 214TP C4 column (250 × 4.6mm^2, 5µm, 300Å; Avantor, Inc., Radnor Township, PA, USA) at 20°C and detection at 280nm. Mobile phase A consisted of 10% methanol made in acetonitrile and mobile B of 0.5% trifluoroacetic acid in water adjusted at pH 2.9 with NaOH. Hemolysate was diluted in B prior injection at a flow rate of 1ml/min in 49% A for 2mins, followed by 2mins in 50% A and a 20min gradient to 60% A. The column was then equilibrated to 49% A for 6mins. Quantification of the area under the curve of the peaks were performed with TotalChrom software and raw values were exported to GraphPad Prism (version 9, GraphPad Software) for plotting and further analysis.

Western blot to quantify globin protein

Samples were subjected to electrophoresis on 14% SDS-polyacrylamide gel electrophoresis under denaturing conditions. A previously frozen WT umbilical cord blood-derived sample was included as a positive control. Following electrophoresis, proteins were transferred to polyvinylidene difluoride membranes and blocked for 1h at room temperature using Tris-buffered saline with Tween 20 (TBS-Tween 20, 0.1%) containing 5% non-fat dry milk to prevent non-specific binding. The blocked membranes were then incubated with primary antibodies diluted in TBS-Tween 20 (0.1%) containing 1% non-fat dry milk. The antibodies used were anti-α (sc-514378), anti-β (sc-21757), and anti-γ (sc-21756) (Santa Cruz Biotechnology, Inc., Dallas, TX, USA). All primary antibodies were diluted at a 1/1000 ratio and incubated with membranes for 4 hours at room temperature with gentle shaking. After primary antibody incubation, membranes were washed three times with TBS-Tween 20 (0.1%) for 10 minutes each to remove any unbound primary antibodies. Membranes were then incubated with the secondary antibody, anti-mouse (sc-525408), diluted 1/2000 in TBS-Tween 20 (0.1%) containing 1% non-fat dry milk. This incubation lasted for 1 hour at room temperature with gentle shaking. Post-secondary antibody incubation, membranes were washed 3x with TBS-Tween 20 (0.1%) for 10mins each to remove any unbound secondary antibodies. Protein bands were visualized using a classic ECL detection system and chemiluminescent signal was captured on X-ray film. The intensity of the protein bands was quantified by Quantity One 1-D Analysis Software (Bio-Rad). Background subtraction was performed for each band, and the resulting values were normalized to a loading control to account for variability in protein loading.

RNA-Sequencing of edited RBCs

Total RNA was extracted from frozen pellets of approximately 10^6 cells *in vitro*-differentiated erythrocytes using RNeasy Plus Micro Kit (Qiagen, Redwood City, CA, USA) according to the

manufacturers' instructions. Sequencing was provided by Novogene (Sacramento, CA, USA). Raw sequencing reads were analyzed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) to ensure adequate data quality. Reads were mapped to the GRCh38.p14 human genome assembly using the pseudo-alignment tool Salmon (https://github.com/COMBINE-lab/salmon). After pseudo-alignment, transcript levels count estimates were collapsed to the gene level, and then transformed using a regular log transformation. Relative expression levels of genes of interest were compared using these normalized values.

GO Enrichment Analysis

Gene sets were submitted to the GOrilla GO enrichment analysis tool (https://cbl-gorilla.cs.technion.ac.il/)[39] for pathway analysis. GOrilla reports the FDR q-value of each GO term following Benjamini-Hochberg correction for multiple testing, and those GO processes with FDR ≤ 0.01 were considered to be highly significant. For heatmap generation, the top 0.001% most highly expressed genes were analyzed for enriched GO processes across all samples. To generate the bubble plot, the top 0.001% of genes with the greatest rank change in expression were identified in each comparison group, using mock or WT as the denominator. Enrichment is defined by GOrilla using the formula (b/n) / (B/N) where N is the total number of genes, B is the total number of genes associated with a specific GO term, n is the number of genes in the top of the target gene set, and b is the number of genes in the intersection. Values were exported to GraphPad Prism (version 9, GraphPad Software) for plotting and further analysis.

Acknowledgements

The authors thank the following funding sources that made this work possible: University of California, San Francisco NIH T32 Research Training in Transplant Surgery Fellowship to S.N.C.; American Society of Gene & Cell Therapy Career Development Award to M.K.C.; NIH NHLBI (no. R01-HL161291) to M.K.C., M.H.P., and T.C.M. and funding from the UCSF Center for Maternal-Fetal Precision Medicine. We thank Emma Canepa and Billie Lianoglou for their assistance with patient sample collection and our patients for their gracious participation in research.

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. TCM is on the scientific advisory board of Acrigen and receives grant funding from Novartis, BioMarin, and Biogen.

M.K.C., B.W., T.C.M., and M.H.P. hold patent US-20220280571-A1 and provisional patent no. 63/236,178.

References

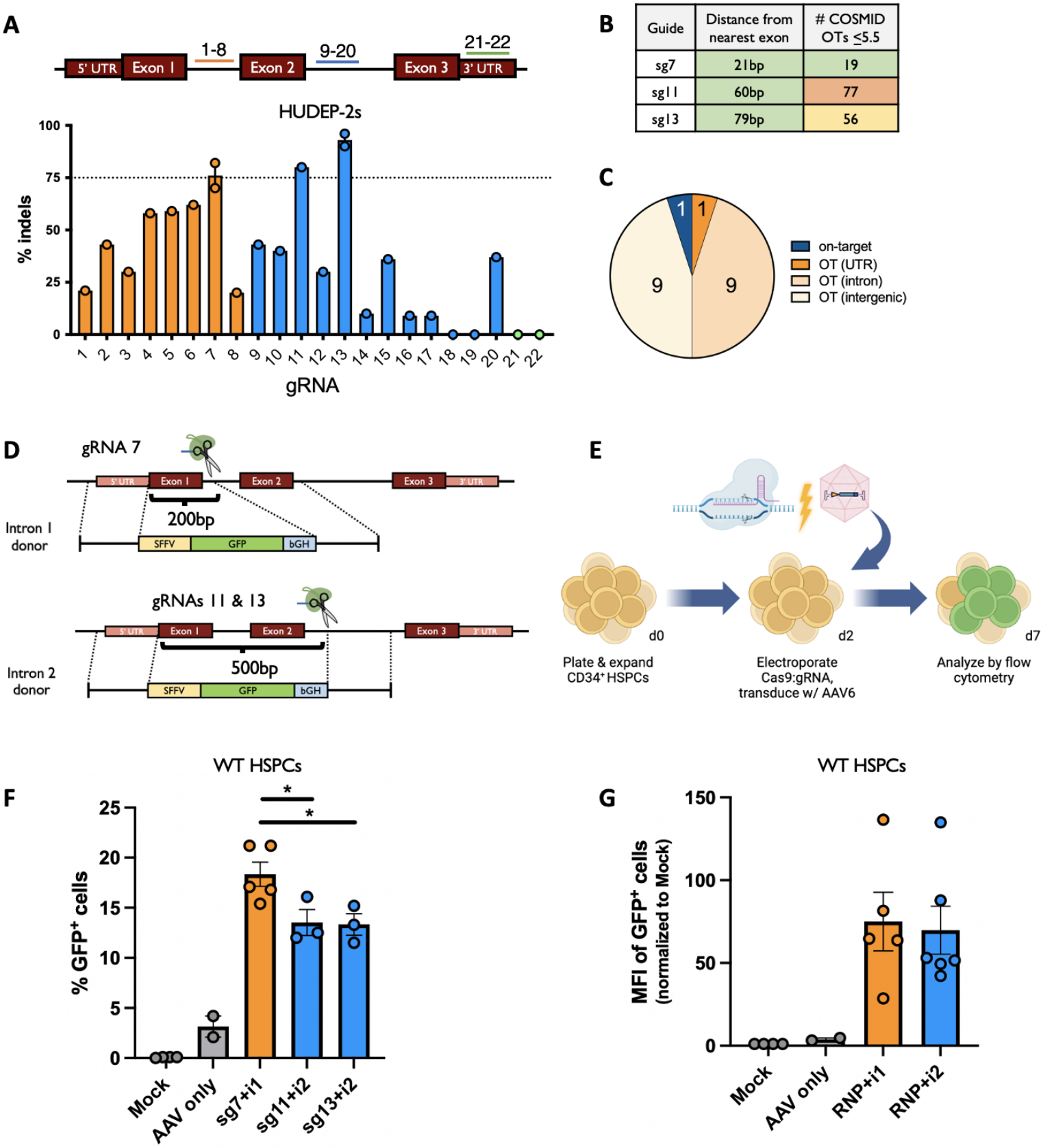
1. Vichinsky E. Complexity of alpha thalassemia: growing health problem with new approaches to screening, diagnosis, and therapy. *Ann N Y Acad Sci*. 2010;1202:180-187.
2. Piel FB, Weatherall DJ. The α-Thalassemias. *New England Journal of Medicine*. 2014;371(20):1908-1916.
3. Kreger EM, Singer ST, Witt RG, et al. Favorable outcomes after in utero transfusion in fetuses with alpha thalassemia major: a case series and review of the literature. *Prenat Diagn*. 2016;36(13):1242-1249.
4. Schwab ME, Lianoglou BR, Gano D, et al. The impact of in utero transfusions on perinatal outcomes in patients with alpha thalassemia major: the UCSF registry. *Blood Adv*. 2023;7(2):269-279.
5. Schrier SL, Centis F, Verneris M, Ma L, Angelucci E. The role of oxidant injury in the pathophysiology of human thalassemias. *Redox Rep*. 2003;8(5):241-245.
6. Baird DC, Batten SH, Sparks SK. Alpha- and Beta-thalassemia: Rapid Evidence Review. *Am Fam Physician*. 2022;105(3):272-280.
7. Fleischhauer K, Locatelli F, Zecca M, 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*. 2006;107(7):2984-2992.
8. Cromer MK, Camarena J, Martin RM, et al. Gene replacement of α-globin with β-globin restores hemoglobin balance in β-thalassemia-derived hematopoietic stem and progenitor cells. *Nat Med*. 2021;27(4):677-687.
9. Thompson AA, Walters MC, Kwiatkowski J, et al. Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia. *N Engl J Med*. 2018;378(16):1479-1493.
10. Frangoul H, Altshuler D, Cappellini MD, et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine*. 2021;384(3):252-260.
11. Schrier SL. Pathophysiology of thalassemia. *Curr Opin Hematol*. 2002;9(2):123-126.
12. Camarena J, Luna SE, Hampton JP, et al. Using human genetics to develop strategies to increase erythropoietic output from genome-edited hematopoietic stem and progenitor cells. Published online August 4, 2023:2023.08.04.552064.

13. Chakraborty N, Bilgrami S, Maness L, et al. Myeloablative chemotherapy with autologous peripheral blood stem cell transplantation for metastatic breast cancer: immunologic consequences affecting clinical outcome. *Bone Marrow Transplant*. 1999;24(8):837-843.
14. Bhatia S. Long-term health impacts of hematopoietic stem cell transplantation inform recommendations for follow-up. *Expert Rev Hematol*. 2011;4(4):437-452; quiz 453-454.
15. Inamoto Y, Shah NN, Savani BN, et al. Secondary solid cancer screening following hematopoietic cell transplantation. *Bone Marrow Transplant*. 2015;50(8):1013-1023.
16. Dever DP, Bak RO, Reinisch A, et al. CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature*. 2016;539(7629):384-389.
17. Kurita R, Suda N, Sudo K, et al. Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. *PLOS ONE*. 2013;8(3):e59890.
18. Conant D, Hsiau T, Rossi N, et al. Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J*. 2022;5(1):123-130.
19. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. 2014;42(22):e168.
20. Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol Ther Nucleic Acids*. 2014;3(12):e214.
21. Cromer MK, Majeti KR, Rettig GR, et al. Comparative analysis of CRISPR off-target discovery tools following ex vivo editing of CD34+ hematopoietic stem and progenitor cells. *Mol Ther*. 2023;31(4):1074-1087.
22. Vakulskas CA, Dever DP, Rettig GR, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med*. 2018;24(8):1216-1224.
23. Dulmovits BM, Appiah-Kubi AO, Papoin J, et al. Pomalidomide reverses γ-globin silencing through the transcriptional reprogramming of adult hematopoietic progenitors. *Blood*. 2016;127(11):1481-1492.
24. de la Chapelle A, Träskelin AL, Juvonen E. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. *Proc Natl Acad Sci U S A*. 1993;90(10):4495-4499.

25. Bauer DE, Orkin SH. Hemoglobin switching's surprise: the versatile transcription factor BCL11A is a master repressor of fetal hemoglobin. *Curr Opin Genet Dev*. 2015;33:62-70.
26. Vichinsky EP. Clinical Manifestations of α-Thalassemia. *Cold Spring Harb Perspect Med*. 2013;3(5):a011742.
27. Scott MD. H2O2 injury in beta thalassemic erythrocytes: protective role of catalase and the prooxidant effects of GSH. *Free Radic Biol Med*. 2006;40(7):1264-1272.
28. Gregory GL, Wienert B, Schwab M, et al. Investigating Zeta Globin Gene Expression to Develop a Potential Therapy for Alpha Thalassemia Major. *Blood*. 2020;136:3-4.
29. Horvei P, MacKenzie T, Kharbanda S. Advances in the management of α-thalassemia major: reasons to be optimistic. *Hematology Am Soc Hematol Educ Program*. 2021;2021(1):592-599.
30. Hendel A, Bak RO, Clark JT, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol*. 2015;33(9):985-989.
31. Bak RO, Dever DP, Porteus MH. CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nat Protoc*. 2018;13(2):358-376.
32. Bak RO, Dever DP, Reinisch A, Cruz Hernandez D, Majeti R, Porteus MH. Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. Levine RL, ed. *eLife*. 2017;6:e27873.
33. Charlesworth CT, Camarena J, Cromer MK, et al. Priming Human Repopulating Hematopoietic Stem and Progenitor Cells for Cas9/sgRNA Gene Targeting. *Mol Ther Nucleic Acids*. 2018;12:89-104.
34. Cromer MK, Vaidyanathan S, Ryan DE, et al. Global Transcriptional Response to CRISPR/Cas9-AAV6-Based Genome Editing in CD34+ Hematopoietic Stem and Progenitor Cells. *Mol Ther*. 2018;26(10):2431-2442.
35. Selvaraj S, Feist WN, Viel S, et al. High-efficiency transgene integration by homology-directed repair in human primary cells using DNA-PKcs inhibition. *Nat Biotechnol*. Published online August 3, 2023:1-14.
36. Khan IF, Hirata RK, Russell DW. AAV-mediated gene targeting methods for human cells. *Nat Protoc*. 2011;6(4):482-501.
37. Aurnhammer C, Haase M, Muether N, 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*. 2012;23(1):18-28.

38. Hu J, Liu J, Xue F, et al. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood*. 2013;121(16):3246-3253.
39. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009;10(1):48.
40. Allen F, Crepaldi L, Alsinet C, et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol*. 2019;37(1):64-72.

Figure 1



Guide	Distance from nearest exon	# COSMID OTs ≤5.5
sg7	21bp	19
sg11	60bp	77
sg13	79bp	56

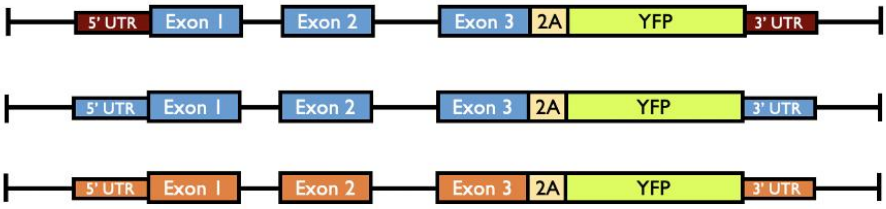
Figure 1. Efficient integration of HDR templates at *HBB* locus.

(A) Schematic of *HBB*-targeting gRNAs and resulting indel frequencies in HUDEP-2 cells following electroporation-mediated delivery of Cas9:gRNA RNPs. Dotted line indicates indel frequencies at 75% threshold. Bars represent mean ± standard error of the mean (SEM). (B) *In silico* off-target analysis of gRNA 7, 11, 13 using COSMID. (C) Pie chart displaying genomic features at 19 predicted target sites for sg7. (D) Schematic of custom AAV6 DNA repair donors designed to mediate integration at intron 1 using sg7 or intron 2 using sg11 or 13. Bracket indicates distance between left and right homology arms. (E) Schematic of Cas9/AAV6 genome editing workflow in primary HSPCs. (F) Percentage of GFP^+ cells following editing in WT HSPCs was determined at d5 post-editing using flow cytometry. Bars represent mean ± SEM. * denotes $p<0.05$ by unpaired two-tailed t-test. (G) Mean fluorescence intensity (MFI) of GFP^+ cells from Figure 1F was determined by flow cytometry. Bars represent mean ± SEM.

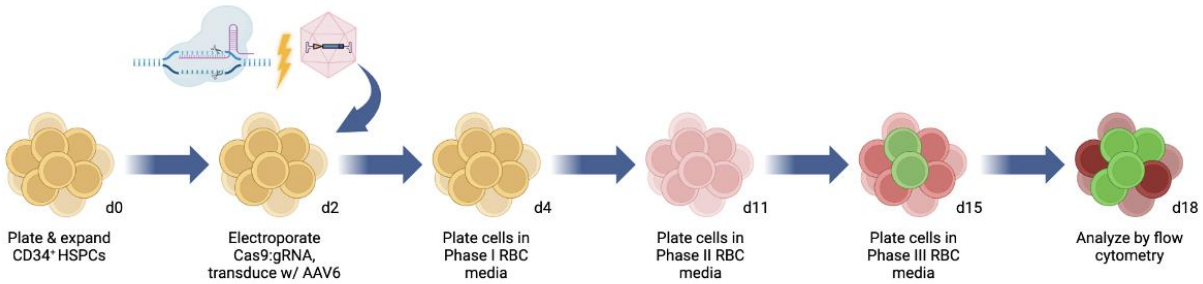
Figure 2

A

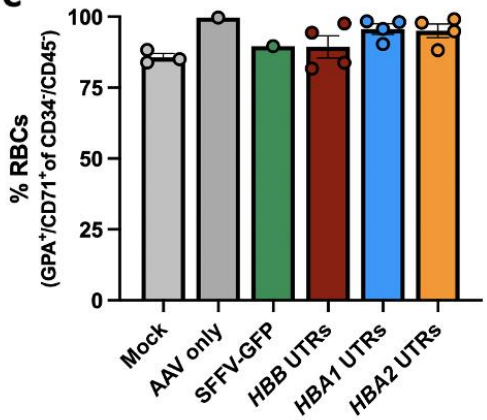
UTRs	Gene
HBB	*HBA1*
HBA1	*HBA1*
HBA2	*HBA2*



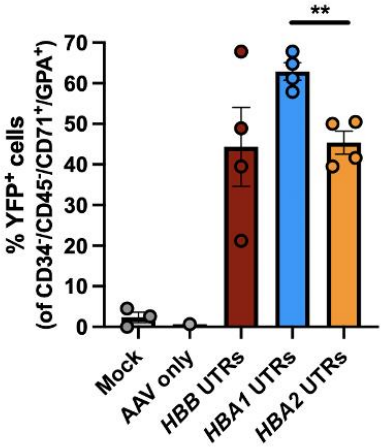
B



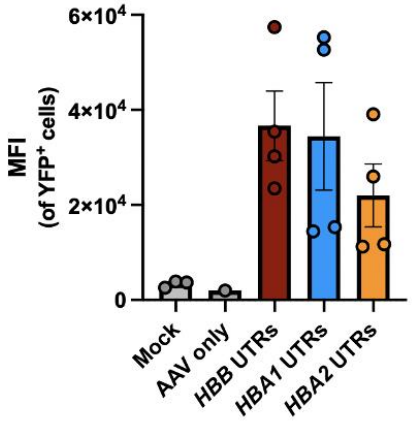
C



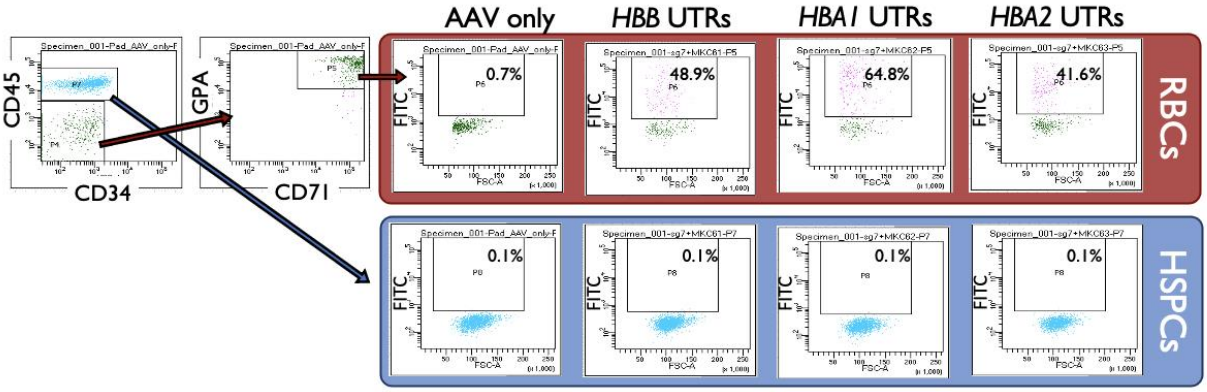
D



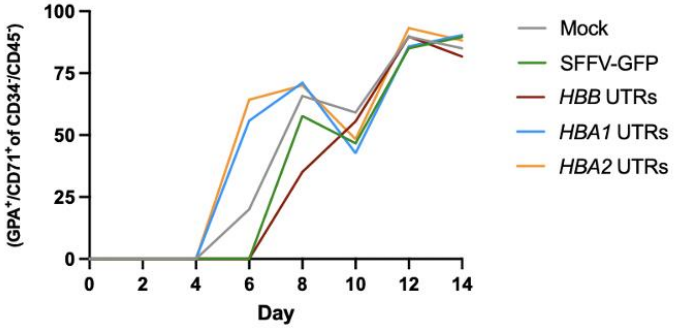
E



F



G



H

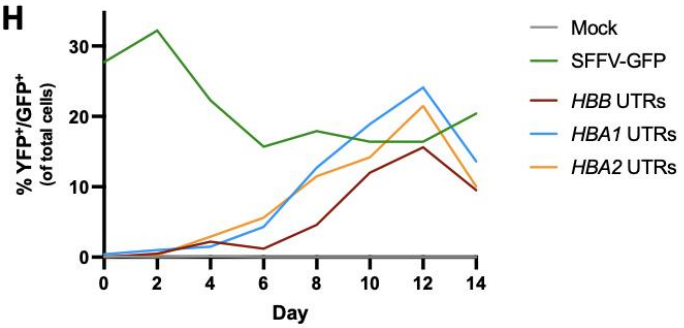


Figure 2. Integration at *HBB* drives erythroid-specific expression of transgenes.

(A) Schematic of custom AAV6 DNA repair donors designed to integrate promoterless *HBA-2A-YFP* transgenes at the start codon of *HBB*. Table to the left indicates whether cassette integrated *HBA1* or *HBA2* transgene as well as *HBB*, *HBA1*, or *HBA2* UTRs. (B) Schematic of Cas9/AAV6 genome editing workflow in primary HSPCs followed by *in vitro* RBC differentiation. (C) Percentage of $CD34^{-}/CD45^{-}$ HSPCs acquiring RBC surface markers—GPA and CD71—as determined by flow cytometry. Bars represent mean ± SEM. (D) Percentage of YFP^{+} cells in $CD34^{-}/CD45^{-}/CD71^{+}/GPA^{+}$ cells was determined at d14 of RBC differentiation using flow cytometry. Bars represent mean ± SEM. ** denotes $p<0.005$. (E) MFI of YFP^{+} cells from Figure 2D was determined by flow cytometry. Bars represent mean ± SEM. (F) At d11 of RBC differentiation, cells were stained for HSPC/RBC markers and analyzed by flow cytometry. Percentage of YFP^{+} cells are noted (N=1) (G) Percentage of $CD34^{-}/CD45^{-}$ cells that acquired CD71 and GPA RBC markers are plotted over the course of RBC differentiation (N=1). (H) Percentage of YFP^{+} or GFP^{+} cells are plotted over the course of RBC differentiation (N=1).

Figure 3

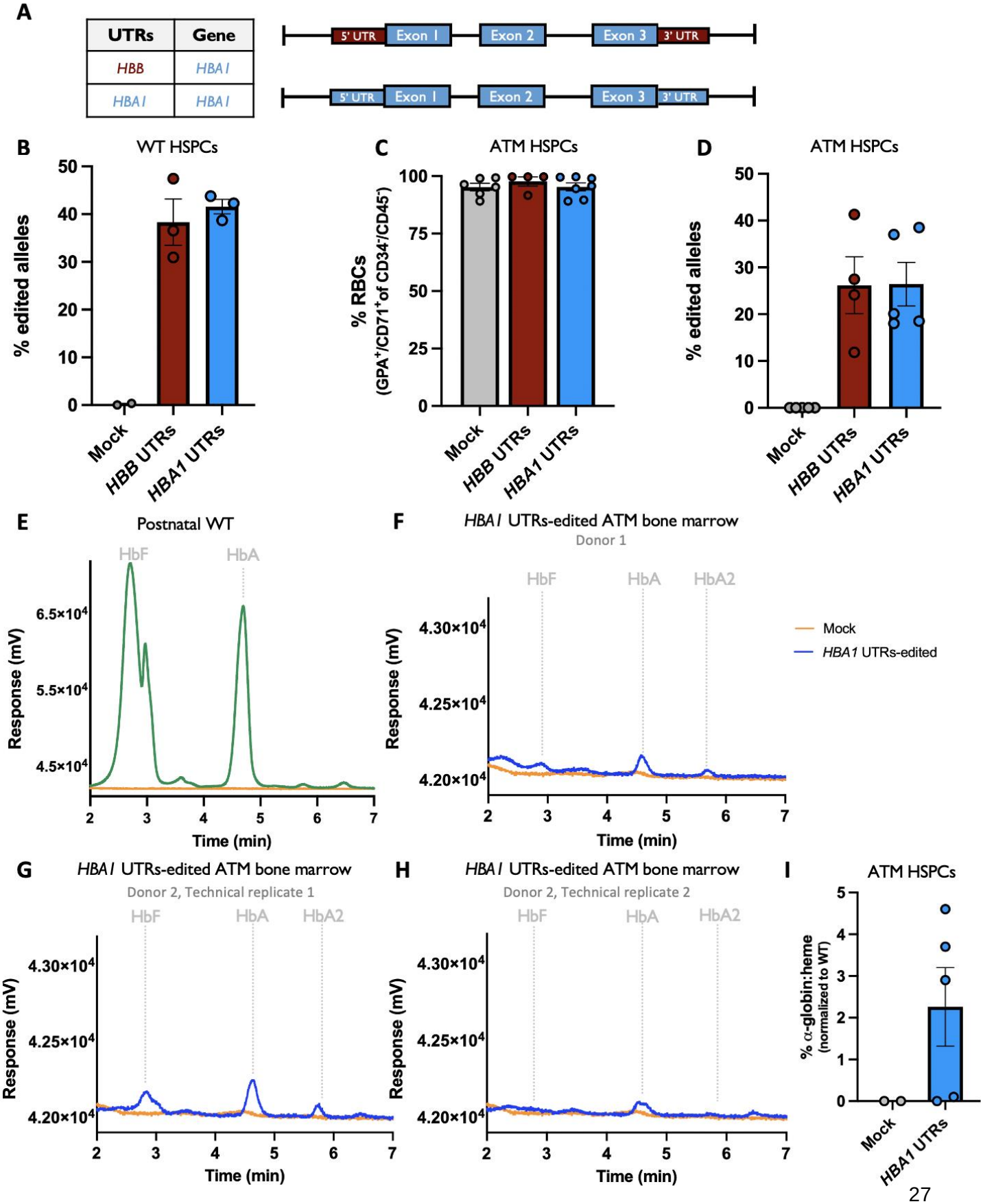

A
UTRs
Gene
HBB
HBA1
HBA1
HBA1
5' UTR
Exon 1
Exon 2
Exon 3
3' UTR
B
WT HSPCs
% edited alleles
Mock
HBB UTRs
HBA1 UTRs
C
ATM HSPCs
% RBCs
(GPA+/CD71+of CD34-/CD45-)
D
ATM HSPCs
% edited alleles
E
Postnatal WT
HbF
HbA
Response (mV)
Time (min)
F
HBA1 UTRs-edited ATM bone marrow
Donor 1
HbF
HbA
HbA2
Mock
HBA1 UTRs-edited
G
HBA1 UTRs-edited ATM bone marrow
Donor 2, Technical replicate 1
H
HBA1 UTRs-edited ATM bone marrow
Donor 2, Technical replicate 2
I
ATM HSPCs
% α-globin:heme
(normalized to WT)

Figure 3. α-globin integration yields low levels of hemoglobin in αTM HSPCs

(A) Schematic of custom AAV6 DNA repair donors designed to integrate promoterless *HBA* transgenes at the start codon of *HBB*. Table to the left indicates whether the cassette integrated *HBA1* transgene with *HBB* or *HBA1* UTRs. (B) Percentage of edited alleles in WT HSPCs at d14 of RBC differentiation. Bars represent mean ± SEM. (C) Percentage of $CD34^-/CD45^-$ αTM HSPCs acquiring RBC surface markers—GPA and CD71—as determined by flow cytometry. Bars represent mean ± SEM. (D) Percentage of edited alleles in αTM HSPCs at d14 of RBC differentiation. Bars represent mean ± SEM. (E) HPLC elution chromatogram displaying the hemoglobin tetramer profile from WT healthy control HSPCs following *in vitro* RBC differentiation. Time displayed on the x-axis represents retention time in minutes for each hemoglobin tetramer type to elute. Absorbance on the y-axis indicates the concentration of a particular hemoglobin tetramer. (F-H) HPLC elution chromatograms displaying the hemoglobin tetramer profile from αTM HSPCs that have undergone editing and RBC differentiation. Chromatograms represent 2 different donors as well as a technical replicate from donor 2 that was edited independently. (I) Percentage of α-globin:heme in mock and edited αTM cells normalized to WT is displayed. Bars represent mean ± SEM.

Figure 4

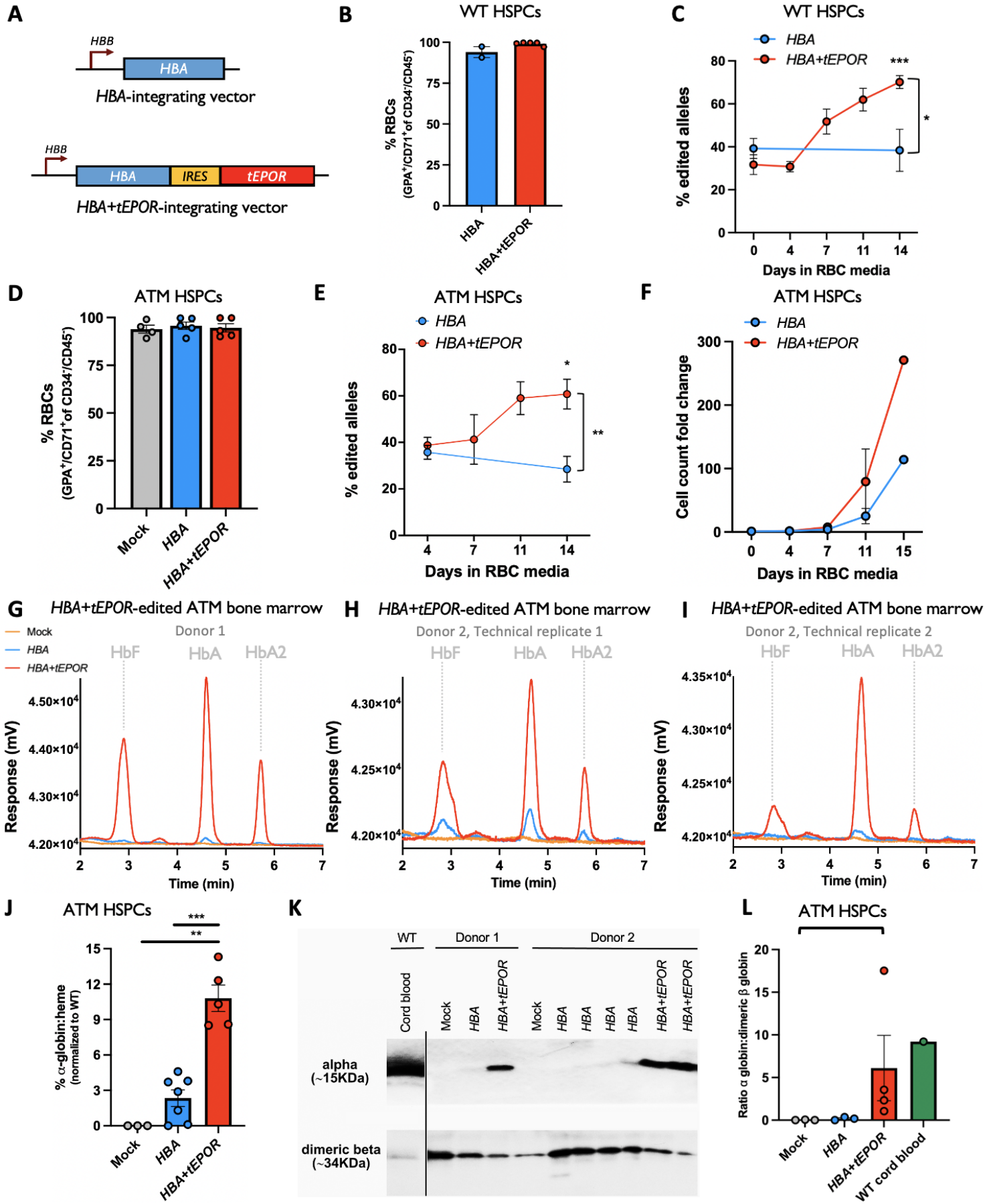
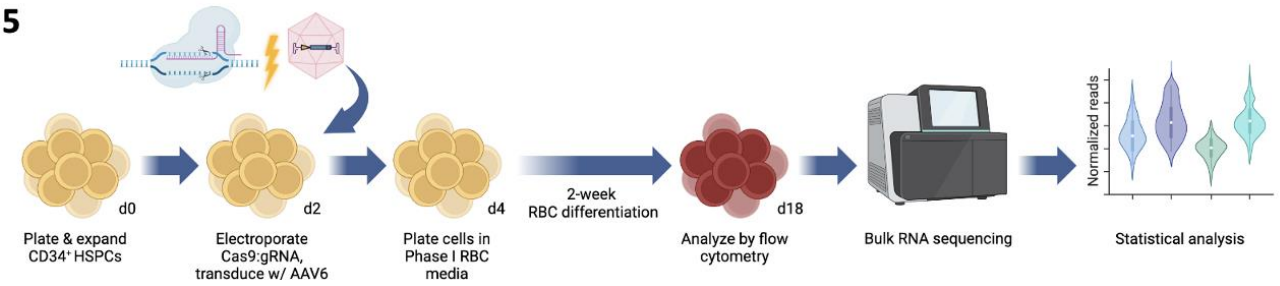

A
HBB
HBA
HBA-integrating vector
HBB
HBA
IRES
tEPOR
HBA+tEPOR-integrating vector
B
WT HSPCs
% RBCs (GPA+/CD71+of CD34-/CD45-)
HBA
HBA+tEPOR
C
WT HSPCs
HBA
HBA+tEPOR
% edited alleles
Days in RBC media
D
ATM HSPCs
% RBCs (GPA+/CD71+of CD34-/CD45-)
Mock
HBA
HBA+tEPOR
E
ATM HSPCs
HBA
HBA+tEPOR
% edited alleles
Days in RBC media
F
ATM HSPCs
HBA
HBA+tEPOR
Cell count fold change
Days in RBC media
G
HBA+tEPOR-edited ATM bone marrow
Donor 1
Mock
HBA
HBA+tEPOR
HbF
HbA
HbA2
Response (mV)
Time (min)
H
HBA+tEPOR-edited ATM bone marrow
Donor 2, Technical replicate 1
HbF
HbA
HbA2
Response (mV)
Time (min)
I
HBA+tEPOR-edited ATM bone marrow
Donor 2, Technical replicate 2
HbF
HbA
HbA2
Response (mV)
Time (min)
J
ATM HSPCs
% α-globin:heme (normalized to WT)
Mock
HBA
HBA+tEPOR
K
WT
Donor 1
Donor 2
Cord blood
Mock
HBA
HBA+tEPOR
Mock
HBA
HBA
HBA
HBA
HBA+tEPOR
HBA+tEPOR
alpha (~15KDa)
dimeric beta (~34KDa)
L
ATM HSPCs
Ratio α globin:dimeric β globin
Mock
HBA
HBA+tEPOR
WT cord blood

Figure 4. Dual integration of α-globin and *tEPOR* increases hemoglobin production in αTM HSPCs

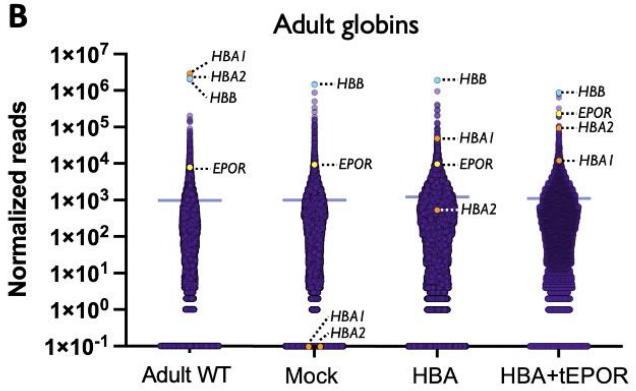
(A) Schematic of custom AAV6 donors designed to integrate promoterless *HBA* and *HBA+tEPOR* transgenes at start codon of HBB. Both vectors are flanked by *HBA1* UTRs. (B) Percentage of $CD34^-/CD45^-$ WT HSPCs acquiring RBC surface markers—GPA and CD71—as determined by flow cytometry. Bars represent mean ± SEM. (C) Percentage of edited alleles in WT HSPCs over the course of RBC differentiation. Bars represent mean ± SEM. *** denotes p=0.0004 and * denotes p=0.01. (D) Percentage of $CD34^-/CD45^-$ αTM HSPCs acquiring GPA and CD71 as determined by flow cytometry. Bars represent mean ± SEM. (E) Percentage of edited alleles in αTM HSPCs over the course of RBC differentiation. Bars represent mean ± SEM. ** denotes p=0.0088 and * denotes p=0.0230. (F) Cell count fold change over the course of RBC differentiation. Bars represent mean ± SEM. (G-I) Hemoglobin tetramer HPLC elution chromatograms. Chromatograms represent 2 different donors as well as a technical replicate from donor 2 that was edited independently. (J) Percentage of α-globin:heme in mock and edited αTM cells normalized to WT is displayed. *HBA* values shown in Figure 3I are included for comparison. Bars represent mean ± SEM. *** denotes p<0.0001; ** denotes p=0.0003. (K) Western blot of mock and edited αTM HSPCs at the end of RBC differentiation compared to WT umbilical cord blood-derived RBCs. 2 different donors as well as technical replicates from donor 2 are displayed. (L) Ratio of α-globin to dimeric β-globin quantification from Western blot. Bars represent mean ± SEM.

Figure 5

A



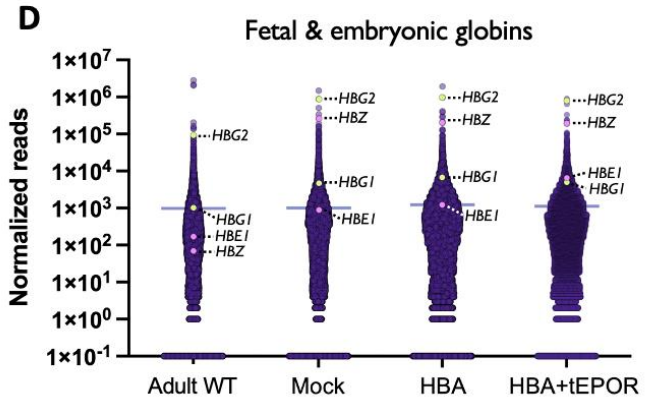
B



C

Rank	WT Mock	HBA Mock	HBA+tEPOR Mock
1	*HBA1*	*HBA1*	*HBA2*
2	*HBA2*	*GPI*	*SCD*
3	*HBM*	*LPIN2*	*HBA1*
4	*TRIB3*	*BNIP3*	*EPOR*
5	*H1F0*	*RPL9*	*INSIG1*

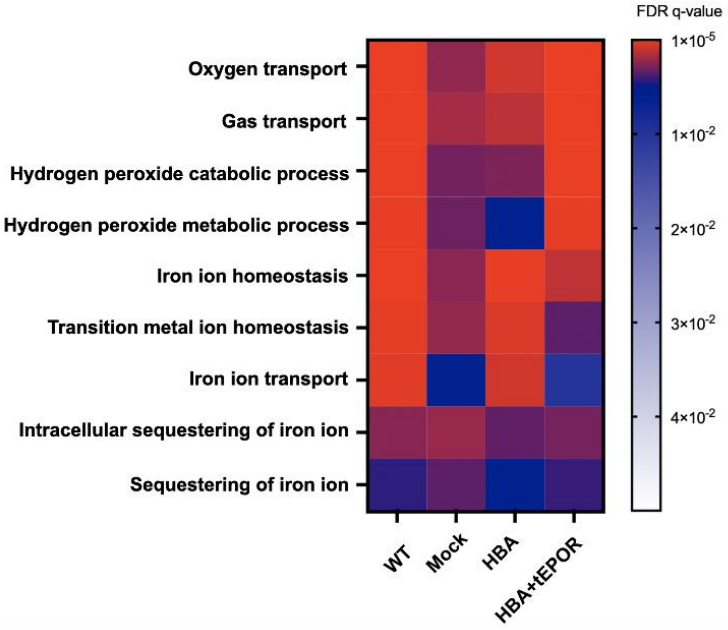
D



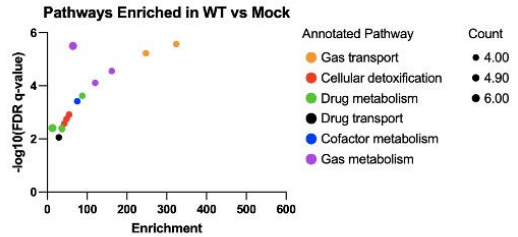
E

Rank	Mock WT	HBA WT	HBA+tEPOR WT
1	*HBZ*	*HBZ*	*HBZ*
2	*HSPA1A*	*HSPA1A*	*SCD*
3	*NME4*	*HPD*	*HSPA1A*
4	*HSPA1B*	*SYNPO2L*	*EPOR*
5	*SYNPO2L*	*NME4*	*INSIG1*

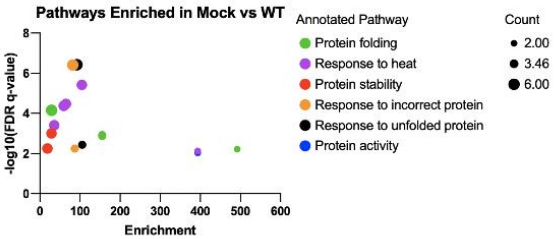
F



G



H



I

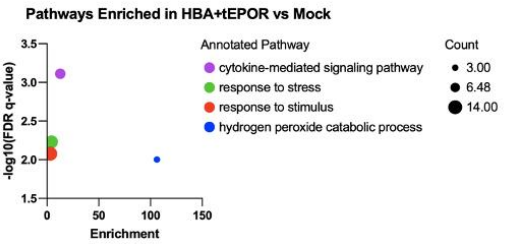


Figure 5. Bulk RNA-sequencing of edited αTM HSPCs demonstrates restoration of α-globin gene expression and oxygen transport and iron ion metabolic processes

(A) Schematic of bulk RNA-sequencing workflow. (B) Normalized reads across all samples displayed with specific genes *HBA1*, *HBA2*, *HBB*, and *EPOR* highlighted. A count of "0.1" was assigned to *HBA1* and *HBA2* for display purposes. (C) Table displaying top 5 genes with the greatest rank change comparing WT and edited αTM cells to mock. (D) Normalized reads across all samples displayed with fetal and embryonic globin genes *HBG1*, *HBG2*, *HBE1*, and *HBZ* highlighted. (E) Table displaying top 5 genes with the greatest rank change comparing mock and edited αTM cells to WT. (F) Heatmap of FDR q-values across 9 commonly enriched GO processes across conditions. Input for GO analysis was the top 0.001% most highly expressed genes within each condition. (G-I) Bubble plots depicting enriched GO processes across treatment comparisons; y-axis displays significance by $-\log_{10}$(FDR q-value) and x-axis displays enrichment. Color of bubble pertains to GO process and size of bubble pertains to number of genes represented in the respective GO process. Input for GO analysis was the top 0.001% of genes showing greatest rank change for each comparison (shown in label above each panel).

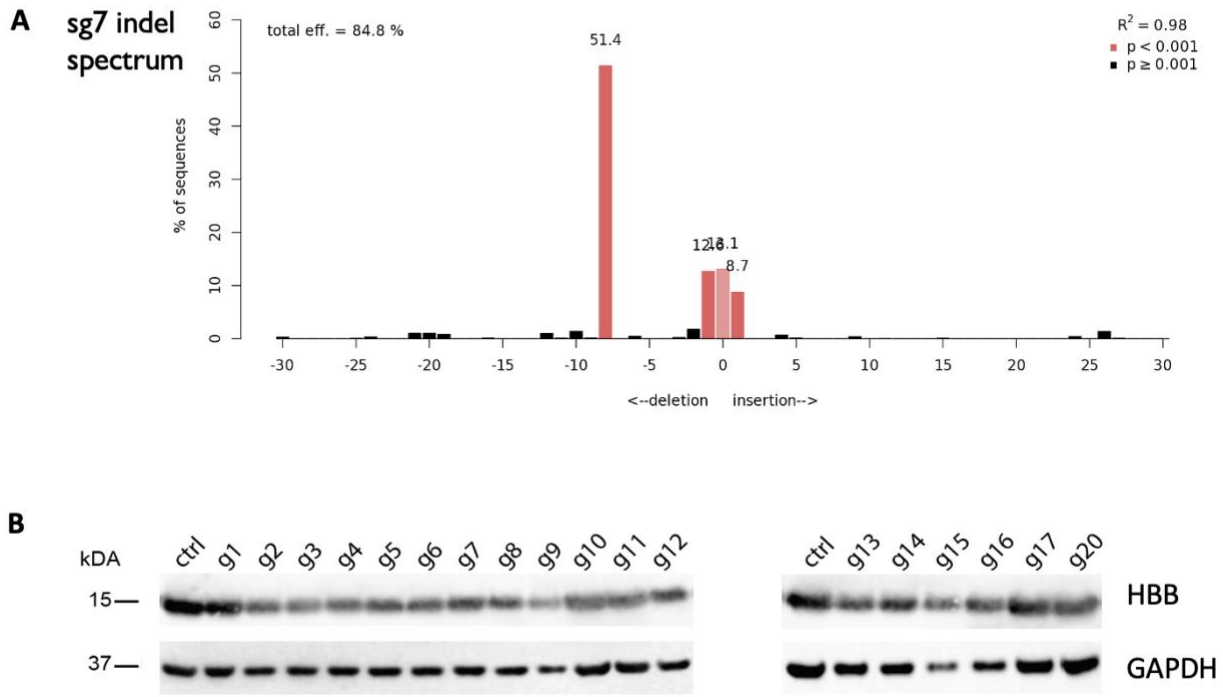
Supplemental Table 1

Guide	Location	Sequence (PAM)	% indels	Primary indel	FORECasT prediction	Nearest exon
1	intron 1	TGGTATCAAGGTTACAAGAC(AGG)	21	-5	+1	Exon 1; 19bp
2	intron 1	AAGGTTACAAGACAGGTTTA(AGG)	43	+1	+1	Exon 1; 26bp
3	intron 1	TCCACATGCCCAGTTTCTAT(TGG)	30	-10	-8	Exon 1; 41bp
4	intron 1	TAAGGAGACCAATAGAAACT(GGG)	58	+1	-1	Exon 1; 44bp
5	intron 1	ACCAATAGAAACTGGGCATG(TGG)	59	-7	+1	Exon 1; 51bp
6	intron 1	TGGAGACAGAGAAGACTCTT(GGG)	62	-2	-2	Exon 2; 59bp
7	intron 1	GGGTGGGAAAATAGACCAAT(AGG)	76	-8	-8	Exon 2; 21bp
8	intron 1	CACTGACTCTCTCTGCCTAT(TGG)	20	-3	+1	Exon 2; 25bp
9	intron 2	ACTTAACCATAGAAAAGAAG(GGG)	43	-3	-3	Exon 2; 37bp
10	intron 2	TTCTTTCCCCTTCTTTTCTA(TGG)	40	+1	-5	Exon 2; 42bp
11	intron 2	TATGGTTAAGTTCATGTCAT(AGG)	80	-5	-5	Exon 2; 60bp
12	intron 2	TTAAGTTCATGTCATAGGAA(GGG)	30	-4	-1	Exon 2; 65bp
13	intron 2	TAGGAAGGGGATAAGTAACA(GGG)	93	+1	-1	Exon 2; 79bp
14	intron 2	AACAGGGTACAGTTTAGAAT(GGG)	10	-17	-2	Exon 2; 95bp
15	intron 2	GACGAATGATTGCATCAGTG(TGG)	36	+1	+1	Exon 2; 123bp
16	intron 2	TGCATCAGTGTGGAAGTCTC(AGG)	9	+1	+1	Exon 2; 133bp
17	intron 2	AAGTATAATAGTAAAAATTG(CGG)	9	+1	+1	Exon 2; 224bp
18	intron 2	TTTTGTTATACACAATGTTA(AGG)	0	-	-	Exon 2; 251bp
19	intron 2	TTAACATTGTGTATAACAAA(AGG)	0	-	-	Exon 2; 264bp
20	intron 2	ATTCCAAATAGTAATGTACT(AGG)	37	+1	+1	Exon 2; 328bp
21	3' UTR	AAGGAACCTTTAATAGAAAT(TGG)	0	-	-	Exon 3; 23bp
22	3' UTR	TGCTGTCCAATTTCTATTAA(AGG)	0	-	-	Exon 3; 28bp

Supplemental Table 1

22 candidate gRNAs targeting the HBB locus. 2nd column displays their respective location in intron 1 or intron 2. 3rd column displays their respective sequence with PAM sequence in parentheses. Columns display % indels found in gRNA screen, primary indel represented, FORECasT-predicted primary indel[40], and nearest exon feature and distance (from predicted cut site).

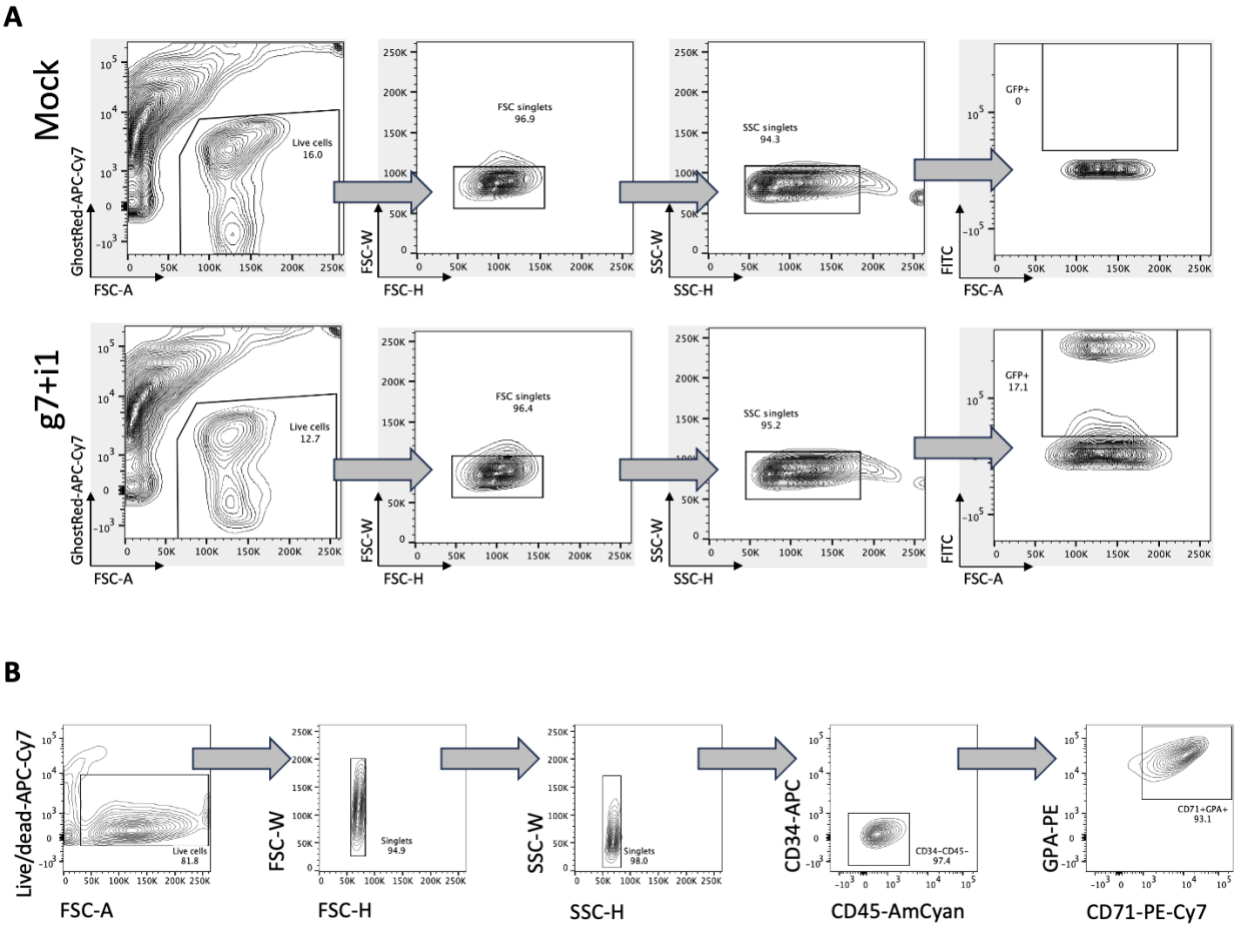
Supplemental Figure 1



Supplemental Figure 1

(A) Representative indel spectrum introduced by Cas9 when complexed with sg7 which targeted intron 1 of *HBB*. Analysis generated by TIDE software. (B) Western blot for HBB following editing with each gRNA (g1-20) in HUDEP-2 cells with GAPDH loading control. Unedited control (ctrl) is in leftmost lane. Protein ladder was used to determine kDA, which is annotated to the left of the gels.

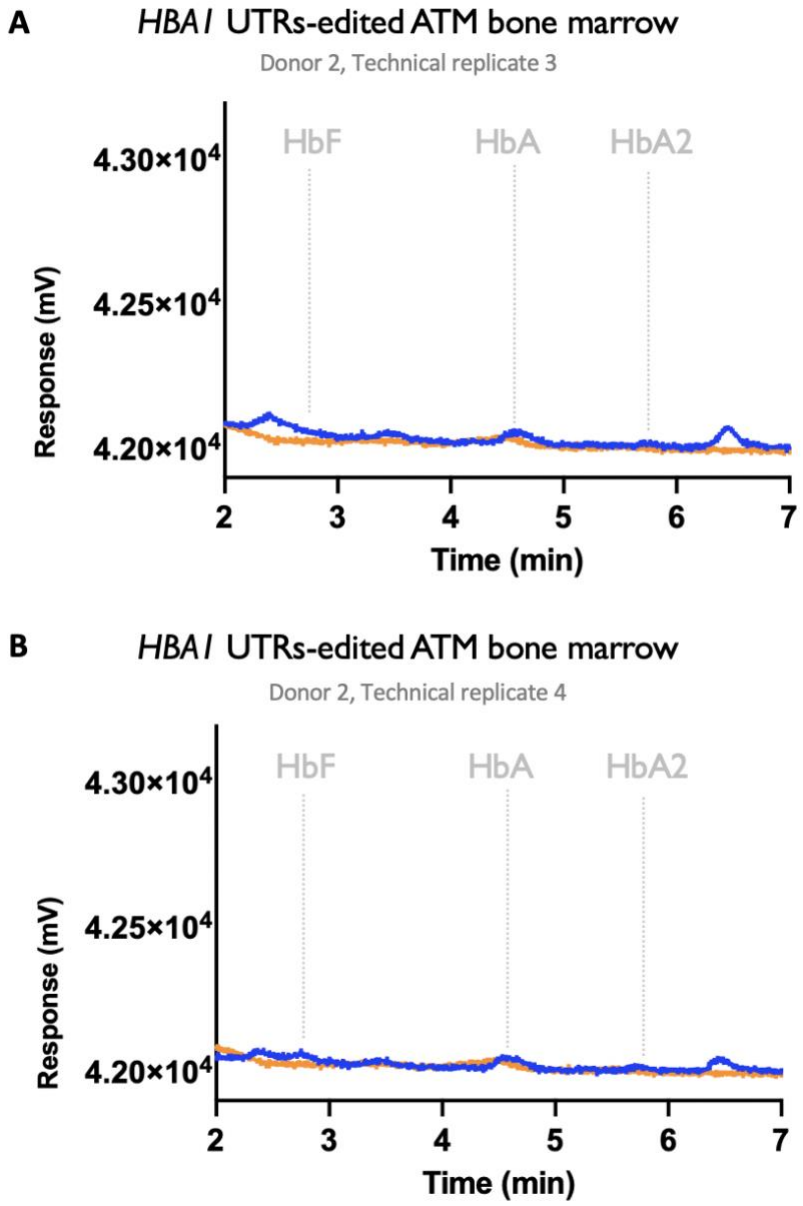
Supplemental Figure 2



Supplemental Figure 2

(A) Representative gating strategy for analysis of HDR integration frequencies in WT HSPCs. (B) Representative gating strategy for analysis of RBC differentiation from WT HSPCs.

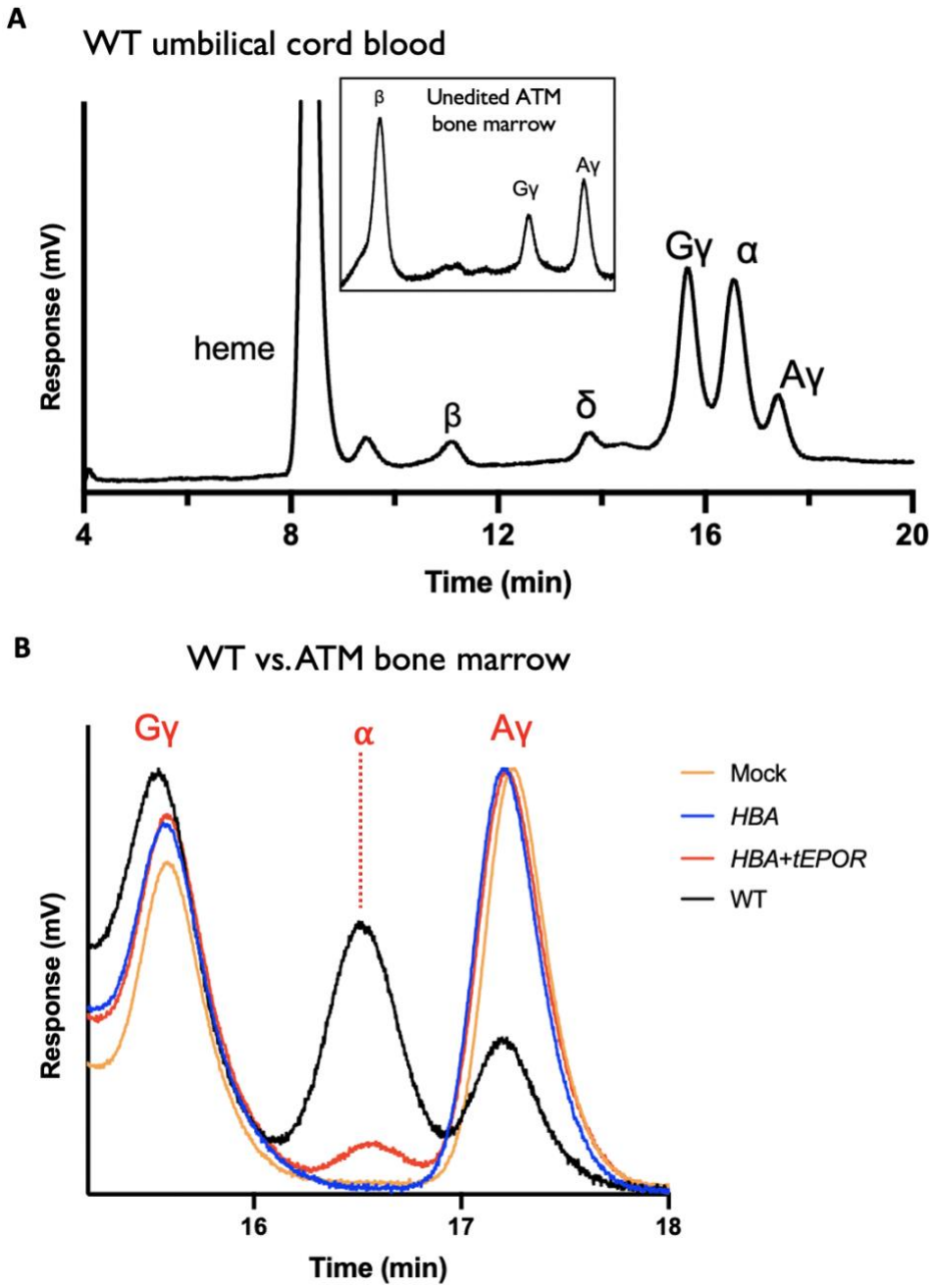
Supplemental Figure 3



Supplemental Figure 3

(A) Hemoglobin tetramer HPLC for *HBA1* UTRs-edited αTM bone marrow-derived RBCs. Figure displays results for donor 2, technical replicate 3. (B) As above, for donor 2, technical replicate 4.

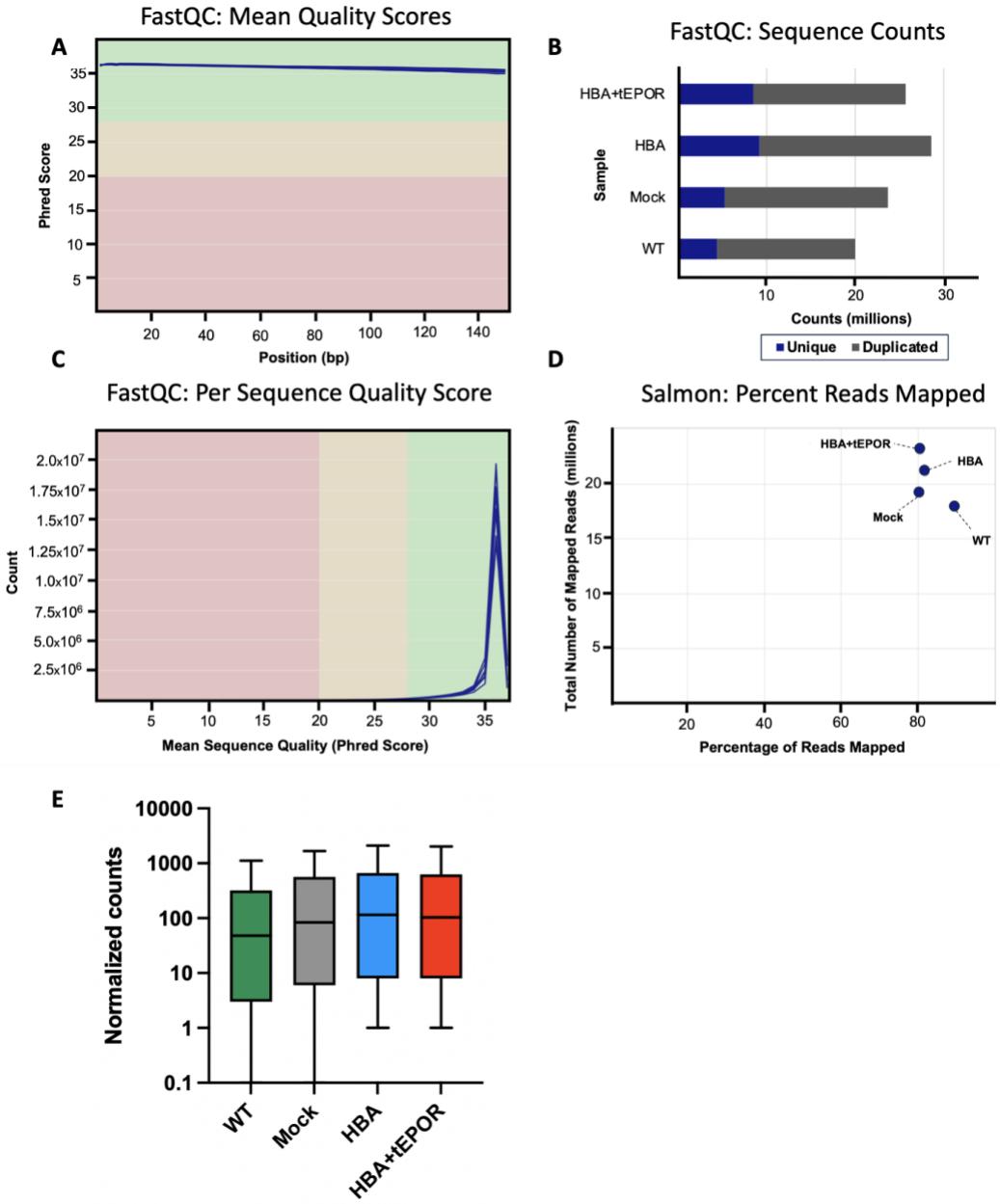
Supplemental Figure 4



Supplemental Figure 4

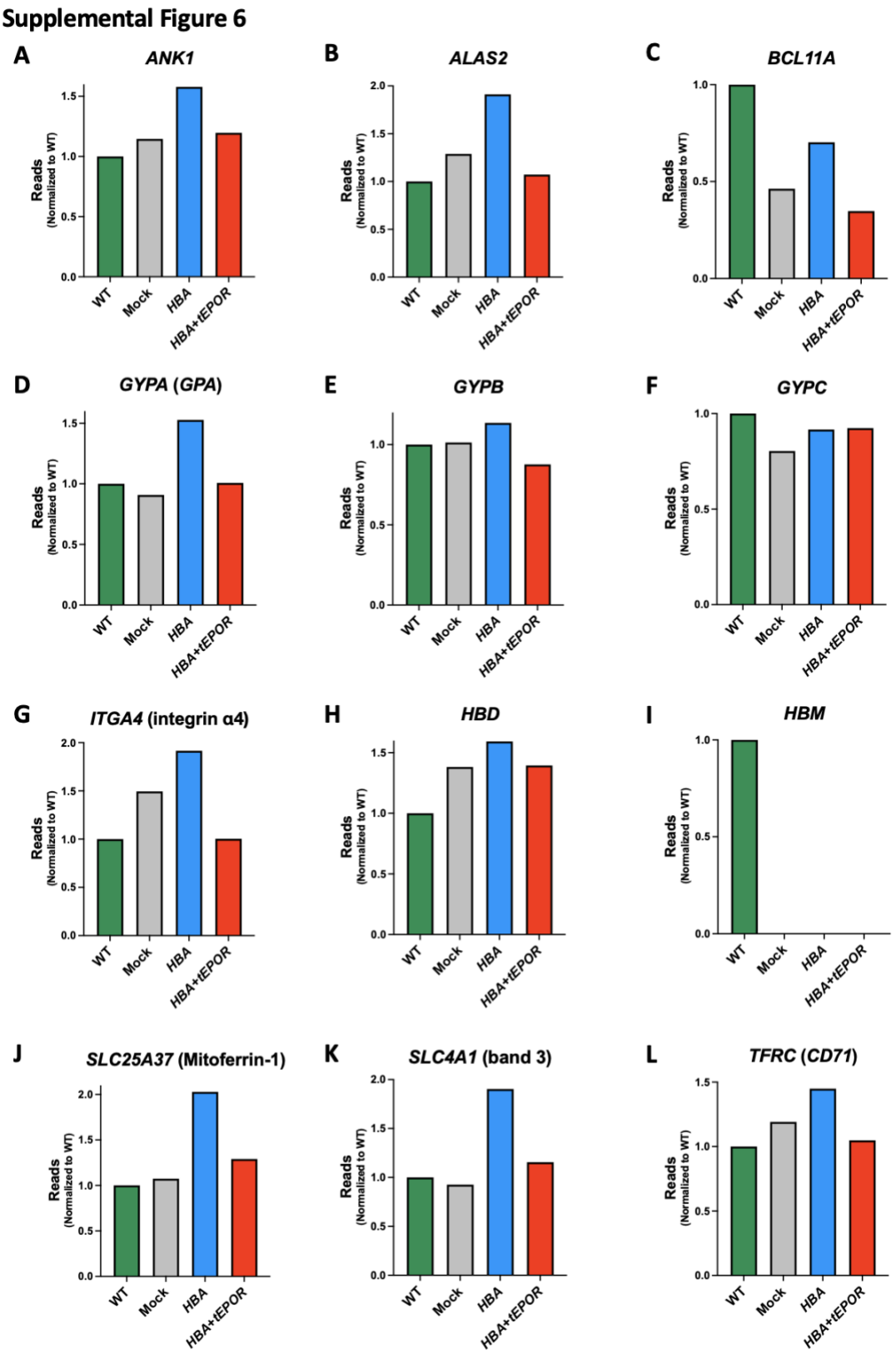
(A) Outset image displays single globin chain profiles of WT umbilical cord blood-derived RBCs. Inset image displays single globin chain profile from unedited αTM patient cells. (B) Single globin chain profile of αTM patient cells from bone marrow that have undergone editing compared to mock and WT.

Supplemental Figure 5



Supplemental Figure 5

(A) Mean quality score for each nucleotide position along a read per sample. (B) Total amount of reads per sample, with unique and duplicated reads annotated within each sample. (C) Number of reads with an adequate mean sequence quality score in each sample. (D) Percentage of sequencing reads mapped by Salmon software compared to the absolute number of reads mapped. (E) Median normalized read counts for each sample. Box displays interquartile range and whiskers show 10-90th percentiles.



Supplemental Figure 6

(A-L) Individual select genes are displayed by reads normalized to WT.