Molecular Therapy
Nucleic Acids
Original Article

Context base editing for splice correction of IVSI-110 β-thalassemia

Basma Naiisseh,[1] Panayiota L. Papasavva,[1] Nikoletta Y. Papaioannou,[1] Marios Tomazou,[2] Lola Koniali,[1] Xenia Felekis,[1] Constantina G. Constantinou,[1] Maria Sitarou,[3] Soteroula Christou,[4] Marina Kleanthous,[1] Carsten W. Lederer,[1,5] and Petros Patsali[1,5]

[1]Molecular Genetics of Thalassemia Department, The Cyprus Institute of Neurology & Genetics, 6 Iroon Avenue, Agios Dometios, Nicosia 2371, Cyprus; [2]Bioinformatics Department, The Cyprus Institute of Neurology & Genetics, Agios Dometios, Nicosia 2371, Cyprus; [3]Thalassemia Clinic Larnaca, State Health Services Organization, Larnaca 6301, Cyprus; [4]Thalassemia Clinic Nicosia, State Health Services Organization, Strovolos, Nicosia 2012, Cyprus

β-Thalassemia is brought about by defective β-globin (HBB [hemoglobin subunit β]) formation and, in severe cases, requires regular blood transfusion and iron chelation for survival. Genome editing of hematopoietic stem cells allows correction of underlying mutations as curative therapy. As potentially safer alternatives to double-strand-break–based editors, base editors (BEs) catalyze base transitions for precision editing of DNA target sites, prompting us to reclone and evaluate two recently published adenine BEs (ABEs; SpRY and SpG) with relaxed protospacer adjacent motif requirements for their ability to correct the common *HBB*$^{IVSI\text{-}110(G>A)}$ splice mutation. Nucleofection of ABE components as RNA into patient-derived CD34$^+$ cells achieved up to 90% editing of upstream sequence elements critical for aberrant splicing, allowing full characterization of the on-target base-editing profile of each ABE and the detection of differences in on-target insertions and deletions. In addition, this study identifies opposing effects on splice correction for two neighboring context bases, establishes the frequency distribution of multiple BE editing events in the editing window, and shows high-efficiency functional correction of *HBB*$^{IVSI\text{-}110(G>A)}$ for our ABEs, including at the levels of RNA, protein, and erythroid differentiation.

INTRODUCTION

Hereditary quantitative defects in β-globin chain synthesis cause blood disorders covered by the umbrella term β-thalassemia, characterized by clinical signs that include hepatosplenomegaly, anemia, and bone marrow expansion.[1] The highest national carrier rate for β-thalassemia in the EU at 12% is found on the island of Cyprus, where 76% of disease alleles bear the globally common and severe *β-globin* (*HBB*)$^{IVSI\text{-}110(G>A)}$ mutation (IVSI-110).[2] This G>A point mutation in intron 1 creates an aberrant splice acceptor site and results in the retention of a 19-nucleotide stretch of intronic sequence, including a premature termination codon, in a large proportion of the mature *HBB* mRNA.[3] Consequentially reduced quantity of HBB prompts homotetramerization and precipitation of surplus α-globin chains (HBA, encoded by the *HBA1* and *HBA2* genes) in precursor cells, which in IVSI-110 homozygotes leads to premature death of red blood cells and ineffective erythropoiesis.[4] As the only conventional curative treatment and despite considerable improvements in event- and disease-free survival, bone marrow transplantation is unsuitable for most patients owing to the limited availability of matching donors and potentially lethal complications, such as graft-versus-host disease, for suboptimal matches.[5] Frequency and severity of IVSI-110 have prompted the development of mutation-specific therapy approaches, which have been pursued by our group[6–8] and several others.[9–12] We explored efficient double-strand break (DSB)-based disruption of aberrant regulatory elements to restore normal splicing[7,8] and found by clonal analyses that deletion of longer upstream sequence stretches, but also of individual bases, may fully restore correct splicing.[7] Although designer nucleases have started to accelerate mutation-specific therapy development,[13] the proven inherent risks of genotoxicity for DSB-based approaches and the comparative safety of newer, DSB-independent editors may point the way for the next generation of advanced therapies.[13,14] Of these, base editors (BEs) combine high-efficiency precision editing with the safety of a DNA nickase and have proven to be effective for use in hematopoietic stem and progenitor cells (HSPCs) in experimental treatment for β-hemoglobinopathies, including the IVSI-110 mutation.[10] One class of these editors, adenine BEs (ABEs) prompt A>G transitions restricted to a small editing window close to the Cas-encoded protospacer adjacent motif (PAM). This study utilizes ABEs with relaxed PAM requirements[15] with the aim of correcting the IVSI-110 splice defect in patient-derived CD34$^+$ cells by altering an upstream sequence element that we have previously demonstrated to be critical for aberrant splicing and a suitable target for functional splice correction.[7,8] Such disruption of aberrant regulatory elements[7] allows greater flexibility than base editing directed at the primary mutation and might thus be more widely applicable across different intronic mutations.

Received 4 October 2023; accepted 28 March 2024;
https://doi.org/10.1016/j.omtn.2024.102183.

[5]These authors contributed equally

Correspondence: Carsten W. Lederer, PhD, Molecular Genetics of Thalassemia Department, The Cyprus Institute of Neurology & Genetics, 6 Iroon Avenue, Agios Dometios, Nicosia 2371, Cyprus.
E-mail: lederer@cing.ac.cy

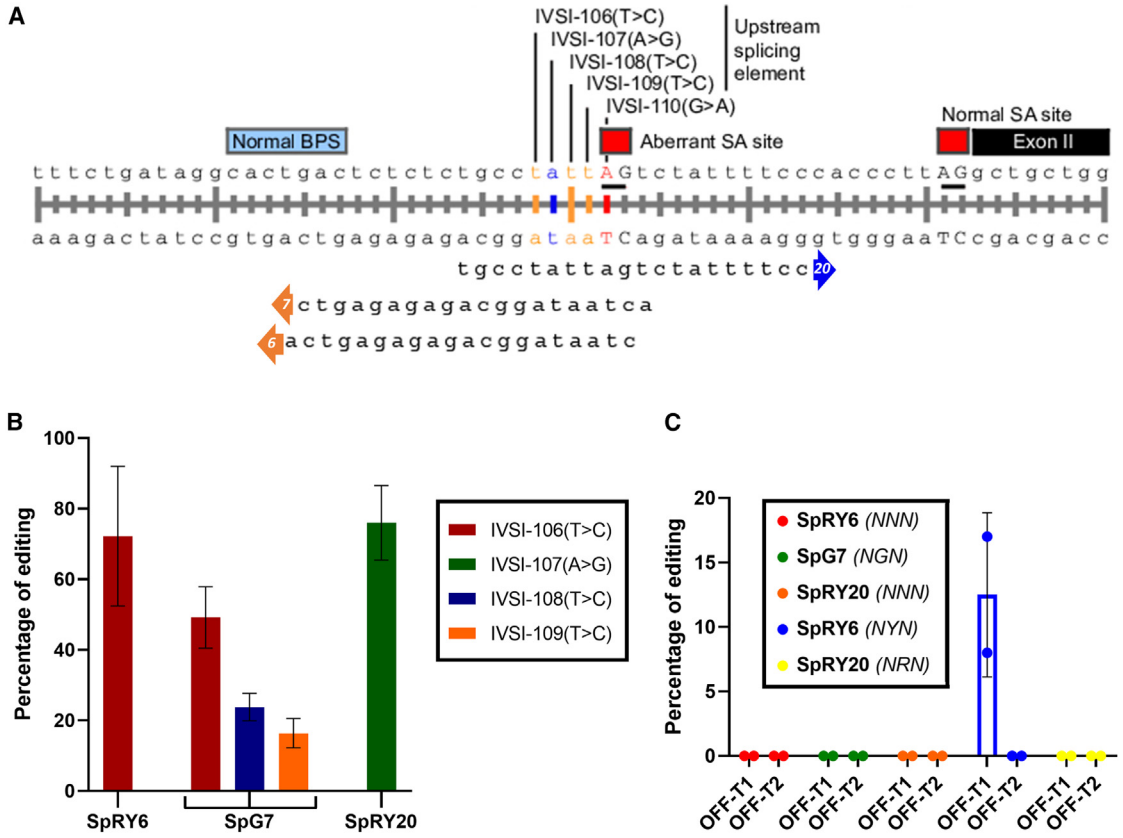


Figure 1. gRNA design strategy for targeting of the IVSI-110 region

(A) The three gRNAs capable of editing the region upstream of the IVSI-110 mutation are shown. Predicted editing windows extended from just upstream of the mutated IVSI-110G>A base (for gRNA20) to the third upstream T base (IVSI-106, for gRNAs 6 and 7). T>C base changes were detected for SpRY6 and SpG7 and are highlighted in orange, whereas an A>G base change was detected for SpRY20 and is highlighted in blue. (B) Bar graph depicting average editing efficiency of ABEs according to EditR analysis, across CD34$^+$ cells derived from different patients across independent experiments (n = 3 for SpRY20, n = 5 for other treatments). (C) Bar graph depicting the percentage of editing at the top two off-target sites (OFF-T1 and OFF-T2, respectively) for SpRY6, SpG7, and SpRY20, as indicated (n = 2). Off-target identities are detailed in Tables S2–S4. For (A) and (B), error bars show the standard deviation of the mean.

RESULTS

Novel ABEs show high DNA-editing efficiency

For use with near-PAMless BEs,[15] NGN-restricted SpG and NRN>NYN-restricted SpRY, gRNAs were designed to target the base four nucleotides upstream of the IVSI-110 mutation (Figure 1A) by placing it within the 4-nucleotide editing window,[16] hereafter referred to as IVSI-106, in a strategy suggested by our published clonal analyses of functional splice correction.[7] In order to avoid the suboptimal efficiency associated with plasmid-based expression,[17] we performed transient BE delivery by gRNA/mRNA nucleofection.[18] After testing different ABE/gRNA combinations for editing efficiency (data not shown), we proceeded with the gRNAs eliciting the highest efficiencies per ABE for analyses in thalassemic, patient-derived CD34$^+$ cells. According to EditR analysis,[19] the approach achieved up to 90% editing at the IVSI-106 target base with the more permissive SpRY ABE using gRNA6 and up to 60% on-target editing with the more stringent SpG ABE using gRNA7, albeit with bystander edits (<30%) of both other Ts in the base-editing window (Figure 1B). To explore additional editing of context bases, we employed SpRY ABEs with differently oriented gRNAs targeting the IVSI-107 adenine base adjacent to the IVSI-106 target. By evaluating different gRNA lengths, we observed a positive correlation for gRNA length and efficiency of editing for patient-derived CD34$^+$ cells (Figure S1), with optimized editing conditions and a gRNA length of 20 nucleotides (gRNA20) resulting in IVSI-107 editing of up to 88%. It is noteworthy in this context that SpRY6, while employing the less efficient NYN PAM (i.e., GTG) for its on-target site, shows efficiency similar to that of SpRY20, the activity of which is based on the NRN PAM (i.e., CAC), for which SpRY exerts its highest efficiency.[15]

In silico specificity predictions of gRNA/ABEs predicts higher safety for SpG7

In order to estimate the safety of our editors, we utilized the CRISPOR[20] *in silico* tool to help ascertain the predicted off-target events of our editors (Table 1; see Tables S2–S4 for full off-target information).[21] Based on CRISPOR and in line with its relatively

Table 1. *In silico* prediction of gRNA off-target effects

gRNA identity	gRNA sequence	MIT specificity score[24]	CrisprSCAN[25]	Off-targets[a]
gRNA7 with ABE_SpG (NGN)	**ACTAATAGGCAGAGAGAGTC**	24	42	0|1|2|64|8000|1|1|1|13
gRNA6 with ABE_SpRY (NYN)	**CTAATAGGCAGAGAGAGTCA**	17	43	0|1|4|116|14580|1|1|7|27
gRNA20 with ABE_SpRY (NRN)	**TGCCTATTAGTCTATTTTCC**	9	6	0|1|8|255|34890|0|0|14|24

[a]Based on analyses using CRISPOR, numbers shown indicate the number of (first row: total, second row: intragenic) off-targets with 0|1|2|3|4 gRNA mismatches.

stringent PAM (NGN), SpG7 had the lowest predicted number (67 total and 3 intragenic for up to 3 mismatches) of off-target hits. Off-target predictions for SpRY6 (121 total and 9 intragenic for up to 3 mismatches) and SpRY20 (264 total and 14 intragenic for up to 3 mismatches) give a high number of total potential off-target sites for SpRY20, but a comparably low number of potential intragenic off-target sites, in line with the high U:A content of the gRNA, which would typically favor intergenic sites.[22] Though based on the same ABE protein, off-target predictions for SpRY6 are approximately half those for SpRY20. We conducted investigation of the off-targeting activity of our ABEs by sequencing the top two predicted off-target regions, as ranked by the CRISPOR algorithm, for each editor with all its potential PAM(s), i.e., NGN for SpG7, NYN/NNN for SpRY6, and NRN/NNN for SpRY20. There were no off-targets detected for the five loci sequenced for SpG7 and SpRY20 (Figure 1C). Validating our assay, out of the four genes analyzed for SpRY6, the intronic *KIAA1328* site, predicted to be the top off-target for SpRY6 (NYN), showed up to 17% editing in ABE-treated patient cells (Figure S2), indicating superior specificity and safety for SpG7 and SpRY20. The off-target locus *KIAA1328* encodes the protein hinderin, which competes with the structural maintenance of chromosome (SMC) protein 3 for binding to the hinge domain of SMC1, thereby preventing their heterodimerization and consequent formation of the cohesion complex, necessary for sister chromatid separation and chromosomal stability.[23]

NGS analysis reveals editing profile of ABEs

Targeted deep sequencing of *HBB* accurately discerned that the editing efficiency of SpG7 as the cumulation of all three *cis* edits in the IVSI-110 upstream region was close to that of SpRY6, leading us to predict a similar effect by both editors on splice correction. This analysis also allowed us to determine the absolute and relative contribution of individual and combined editing events per allele. As shown in Figure 2, the edited bulk population consists of all possible combinations of editing events, including single (81% of total editing events), double (15%), and triple (4%) base edits. Based on our next-generation sequencing (NGS) data, a small but notable increase in insertion-deletion (indel) frequency is seen for SpRY6 and SpG7 compared to the control samples, so that the percentages of reads that contained at least one small insertion overlapping the gRNA binding site were, on average, 0.96% and 0.39% for SpRY6 and SpG7, respectively, while the observed percentage for the control samples was 0.23%. The indel rates normalized for segment length and corrected for baseline indel frequencies showed similar results (Figures S3–S7), with SpRY6 target rates being 3-fold higher at 4.1×10^{-4} insertion frequency per position compared to the control (1.3×10^{-4} insertion frequency per position). SpG7 insertion rates were approximately 1.42-fold higher, while SpRY20 remained equal or lower, compared to the control. Deletion rates were similar across all samples. In agreement with our results, increased indel frequency following base editing in the same *HBB* target region has also been noted by others.[10]

Post-editing functional analyses

All edited cells were subjected to induced erythroid differentiation for a period of 18 days post-editing and were functionally analyzed for hallmarks of β-thalassemia at different levels.

ABE editing of different context bases impacts RNA expression differentially

Analysis by RT-qPCR of normal and aberrant splicing at the RNA level, as a critical parameter of correction for IVSI-110,[7,8,26] indicates a significantly higher and comparable percentage of normal HBB mRNA for SpG7 (85% ± 6.7%) and SpRY6 (80% ± 4.8%) compared to the untreated control (56% ± 8.0%) (Figure 3A). However, for total *HBB* transcript levels normalized to *HBA* (Figure 3B), a quantitative increase of expression was observed by a factor of $1.9 \times 0.41^{\pm1} \pm 2.7$ for SpRY6, and a higher factor of $3.7 \times 0.64^{\pm1} \pm 2.1$ for SpG7, compared to control ($1.0 \pm 1.1 \times 0.48^{\pm1}$). Inversely, SpRY20-treated samples indicated a significantly lower percentage (30% ± 8.0%) of normal mRNA expression, as well as decreased overall *HBB* expression ($0.5 \pm 1.6 \times 0.24^{\pm1}$), as compared to untreated control. These results were confirmed via additional analyses based on RT-droplet digital (dd)PCR (Figure S8). Corresponding *in silico* predictions indicate an effect of the introduced base changes on DNA elements that regulate splicing and may elicit differences in transcript levels and ratios at the intron-exon border of IVSI (Figure S9). For SpRY20, in contrast to normal and other edited sequences, these predictions include the formation of a novel donor site between the aberrant and normal splice acceptors, the loss of a cryptic branchpoint site (BS), and a reduction in the overall ratio of splice enhancer:silencer elements as three *in silico* results that may guide future experimental substantiation.

β-Globin and HbA levels are significantly increased after IVSI-106 targeting and significantly decreased after IVSI-107 targeting by ABEs

High-performance liquid chromatography (HPLC) analyses revealed a significantly increased HBB protein level relative to HBA for both SpRY6 (0.59 ± 0.10) and SpG7 (0.71 ± 0.01) (by reversed-phase

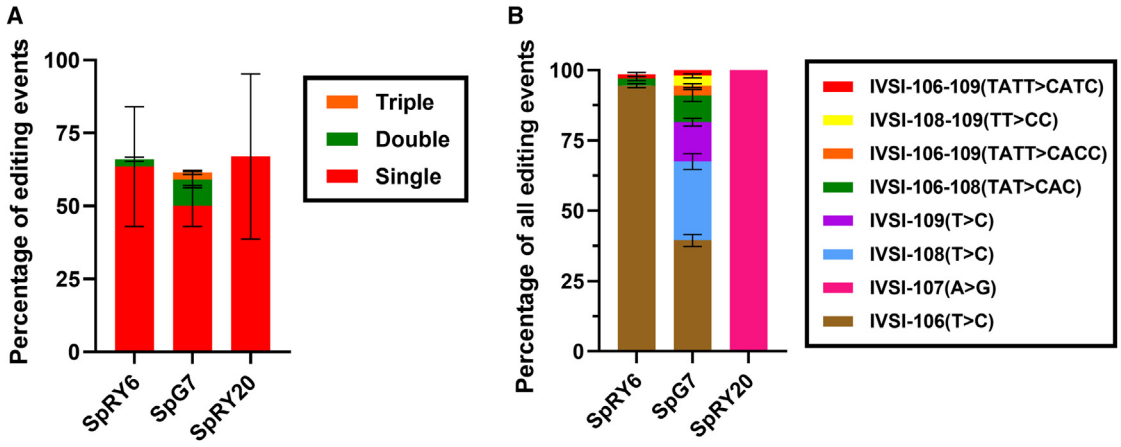


Figure 2. Next-generation DNA sequencing data of edited samples

(A) Graph illustrating frequency of editing events relative to homozygous IVSI-110(G>A) control. (B) Graph illustrating percentage contribution of editing events. For (A) and (B), error bars show the standard deviation of the mean.

[RP]-HPLC; Figures 4A–4C) compared to control cells (0.34 ± 0.14). A correspondingly significantly decreased HBD level relative to HBA is seen for SpG7 and SpRY6 (0.05 ± 0.02 for both) compared to untreated control (0.10 ± 0.01). Total HBG (as the sum of HBG2 and HBG1) relative to HBA is slightly decreased for SpG7 (0.16 ± 0.04) and SpRY6 (0.22 ± 0.06) compared to untreated control (0.29 ± 0.07). Assessment of proportions of total globin chains for the editors revealed significantly increased HBB levels for SpRY6 (0.71 ± 0.06) and SpG7 (0.77 ± 0.03) compared to untreated control (0.44 ± 0.10). As for HBG1, levels were significantly decreased for SpG7 (0.11 ± 0.02) and SpRY6 (0.15 ± 0.03) compared to untreated control (0.29 ± 0.07). In the case of HBG2, levels were decreased for SpG7 (0.07 ± 0.02) and SpRY6 (0.08 ± 0.03) compared to untreated control (0.15 ± 0.06). Finally, for SpG7, we detected a significantly decreased level of HBD (0.05 ± 0.02) compared to untreated controls (0.13 ± 0.03), with decreased expression also seen for SpRY6 (0.06 ± 0.03). As for the proportion of hemoglobins, HbA (by cation-exchange [CE]-HPLC; Figure 4D) was significantly increased following SpRY6 and SpG7 treatments compared to untreated control (40% ± 2.0%), with expression levels of 67% ± 4.5% and 75% ± 6.3%, respectively. Fetal hemoglobin (HbF) expression was also significantly decreased for both corrective editors at 21% ± 5.9% for SpRY6 and 14% ± 5.3% for SpG7 as compared to untreated control levels (40% ± 2.5%). HbA2 was likewise decreased in these samples, albeit not at a significant level, at 13% ± 3.5% for SpRY6 and 11% ± 1.6% for SpG7 compared to 19% ± 2.5% for the control. Compared to control values and consistent with decreased HBB/HBA transcript expression detected for SpRY20 treatment (Figure 3), protein levels of HBB relative to both HBA (0.19 ± 0.08) and HbA (15%) were decreased, with commensurate increased HbF levels (65%). In fact, compared to untreated control, SpRY20-treated cells expressed significantly higher levels of total HBG (0.50 ± 0.16) and notably lower levels of HBD (0.08 ± 0.03) relative to HBA. Expressed as proportions of β-like globin relative to HBA, SpRY20 showed a significantly decreased HBB (0.22 ± 0.09) and notably decreased HBD (0.09 ± 0.03), but significantly increased HBG1 (0.47 ± 0.10) and a notable increase in HBG2 (0.22 ± 0.04), compared to control values.

ABE-mediated changes of RNA and protein levels significantly impact erythroid morphology

Microscopy of cytocentrifuged samples during erythroid differentiation and differential scoring of cells in the erythroid lineage is highly informative for the restoration of normal hematopoiesis. Here, SpRY6- and SpG7-treated cells showed comparably greater hemoglobinization of 86% ± 11.9% and 84% ± 8.9%, respectively, compared to untreated control (82% ± 9.9%), and samples treated with these editors showed a significant shift toward late-stage cells (orthochromatic erythroblasts and reticulocytes) of 56% ± 15.4% and 52% ± 20.5%, respectively, compared to untreated control (37% ± 17.5%), indicating more advanced and higher rates of erythropoiesis (Figure 5). The basophilic subpopulation was slightly decreased for SpRY6 (14% ± 11.9%) and SpG7 (16% ± 8.9%) compared to untreated control (18% ± 9.8%). Likewise, the proportions of polychromatophilic cells for SpRY6 (30% ± 5.5%) and SpG7 (32% ± 15.2%) are decreased when compared to untreated controls (44% ± 14.5%). As for the more differentiated orthochromatic cells, a significantly increased percentage is seen for SpRY6 (51% ± 14.2%) and SpG7 (48% ± 19.8%) as compared to untreated controls (37% ± 18.2%). In the case of reticulocytes, the most differentiated erythroid subpopulation quantified, the percentage is increased for SpG7 (4% ± 1.6%) and SpRY6 (5% ± 2.8%) compared to untreated controls (1% ± 0.9%) but not at a statistically significant level. In contrast, SpRY20-treated cells exhibited delayed differentiation and lower hemoglobinization (73% ± 9.6%), and showed a significantly decreased late-stage differentiation (27% ± 15.4%), with a significantly increased basophilic erythroblast cell population (27% ± 9.6%), a slightly increased polychromatophilic subpopulation (46% ± 7.9%), and a decreased orthochromatic subpopulation (26% ± 15.1%) compared to untreated controls, indicating worsening of the β-thalassemia phenotype, in line with protein and RNA parameters.

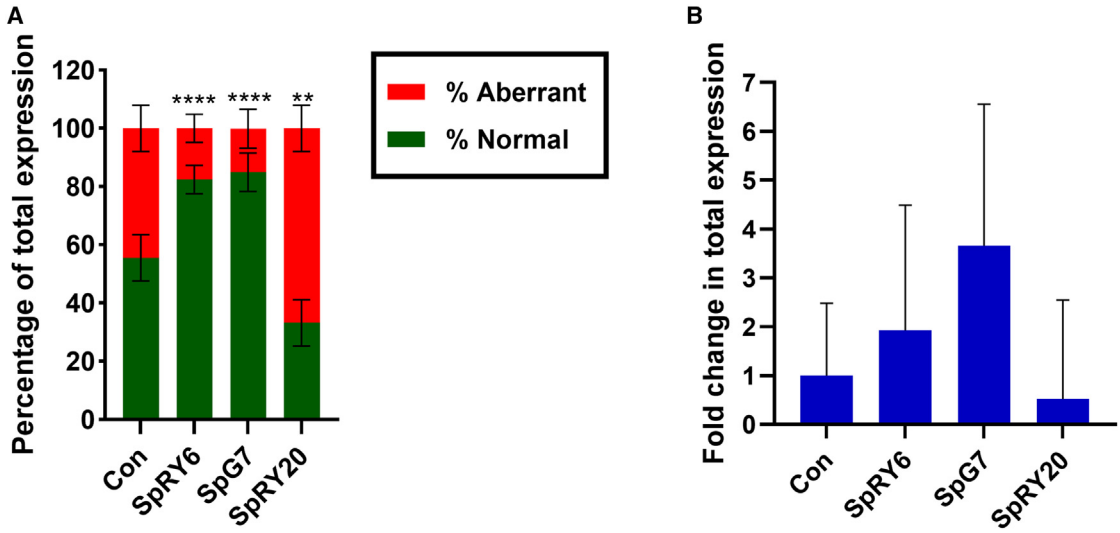


Figure 3. Assessment of splicing at the transcriptional level in patient-derived CD34⁺ cells on day 11 of erythroid differentiation

(A) Mean proportion of normal and aberrant HBB mRNA. The groupwise comparison to untreated control for aberrant RNA levels was performed using one-way ANOVA with Dunnett's multiple comparison analysis. ****$p < 0.0001$ and **$p = 0.0015$. (B) Relative mRNA expression levels of HBB to HBA. The groupwise comparison to untreated control was performed using Kruskal-Wallis analysis with Dunn's multiple comparisons test, and only the proportional upward variation is indicated. For (A) and (B), $n = 3$ for SpRY20, $n = 5$ for other treatments. For (A) and (B), error bars show the standard deviation of the mean.

Flow cytometric lineage and enucleation markers confirm the effect of ABEs on erythroid lineage differentiation

Analysis of erythroid differentiation via flow cytometry showed a gradual improvement of late-stage erythroid differentiation ($CD235a^+/CD36^-$) starting from day 11 of induced erythroid differentiation in SpRY6- and SpG7-edited samples compared to controls. In particular, we observed significantly higher percentages of $CD235a^+/CD36^-$ stained cells in SpRY6-edited samples (day 14: 44% ± 14.7%; day 18: 63% ± 11.7%) and SpG7 (day 14: 48% ± 11.2%; day 18: 63% ± 15.3%) compared to controls (day 14: 30% ± 15.4%; day 18: 38% ± 11.9%), indicating greater rates of late-stage differentiation and thus rescue of ineffective erythropoiesis in the SpRY6- and SpG7-treated samples (Figures 6A and S10). Specifically, our data indicate increased differentiation of SpG7-treated cells compared to its counterpart, SpRY6, in agreement with the detection of greater total HBB mRNA levels for the former. Additionally, significantly increased percentages of enucleated cells ($NucRed^-$) were seen for cells edited with SpRY6 (9.7% ± 4.2%) and SpG7 (9.7% ± 3.9%) compared to control cells (4.6% ± 2.7%) (Figures 6B and S11), corroborating the correction of erythrocyte formation. Opposing results were noted for SpRY20, with fewer enucleated cells (4.0% ± 3.6%) and fewer $CD235a^+/CD36^-$ cells (day 14: 26% ± 17.9%; day 18: 41% ± 26.0%) in comparison to untreated patient cells, albeit not at a statistically significant level.

DISCUSSION

We have presented here an approach for correction of the IVSI-110(G>A) β-thalassemia splicing mutation using our optimized ABEs, SpRY6 and SpG7, which substantially improve the disease phenotype of edited patient-derived $CD34^+$ cells by modifying a critical upstream splice element, IVSI-106, required for aberrant splicing, with demonstrable correction of disease hallmarks at the RNA, protein, morphology, and cytometry levels. Building on our earlier work,[7] these findings confirmed previous observations and thereby establish the identity of the IVSI-106 nucleotide rather than a change of position for neighboring elements as critical for regulation of *HBB* splicing.

Additionally, we identified an immediately adjacent splice element, IVSI-107, as required for normal instead of aberrant splicing, based on the observation that SpRY20-mediated A>G base transition of IVSI-107 aggravated those same β-thalassemia disease parameters corrected by the IVSI-106(T>C) transition. At a general level, our observation underscores the intricate balance of regulatory splice elements and the substantial and variable consequences different editing events may have in even well-characterized regulatory regions. More specifically, and on a route parallel to that chosen by the Miccio group in their elegant work on BE-based correction of the primary IVSI-110 mutation,[10] our study pinpoints and validates an alternative therapeutic target while opening up the path to exploring the interplay of factors affecting the splicing of *HBB* intron 1.

Interestingly, our NGS data showing the same-molecule distribution of base-editing events for SpG7 suggests that all combinations of three possible edits were represented, resembling chance distribution of edits within the editing window. This effectively resulted in an overall editing efficiency far higher than would be expected from the individual editing efficiencies per base and from the assumption of initial saturation of the most efficient editing event. As for similar analyses of on-target base editing for the IVSI-110(G>A) mutation by Hardouin et al.,[10] editing distribution was reproducible between different patients in the present study. However, whereas here over 80% of

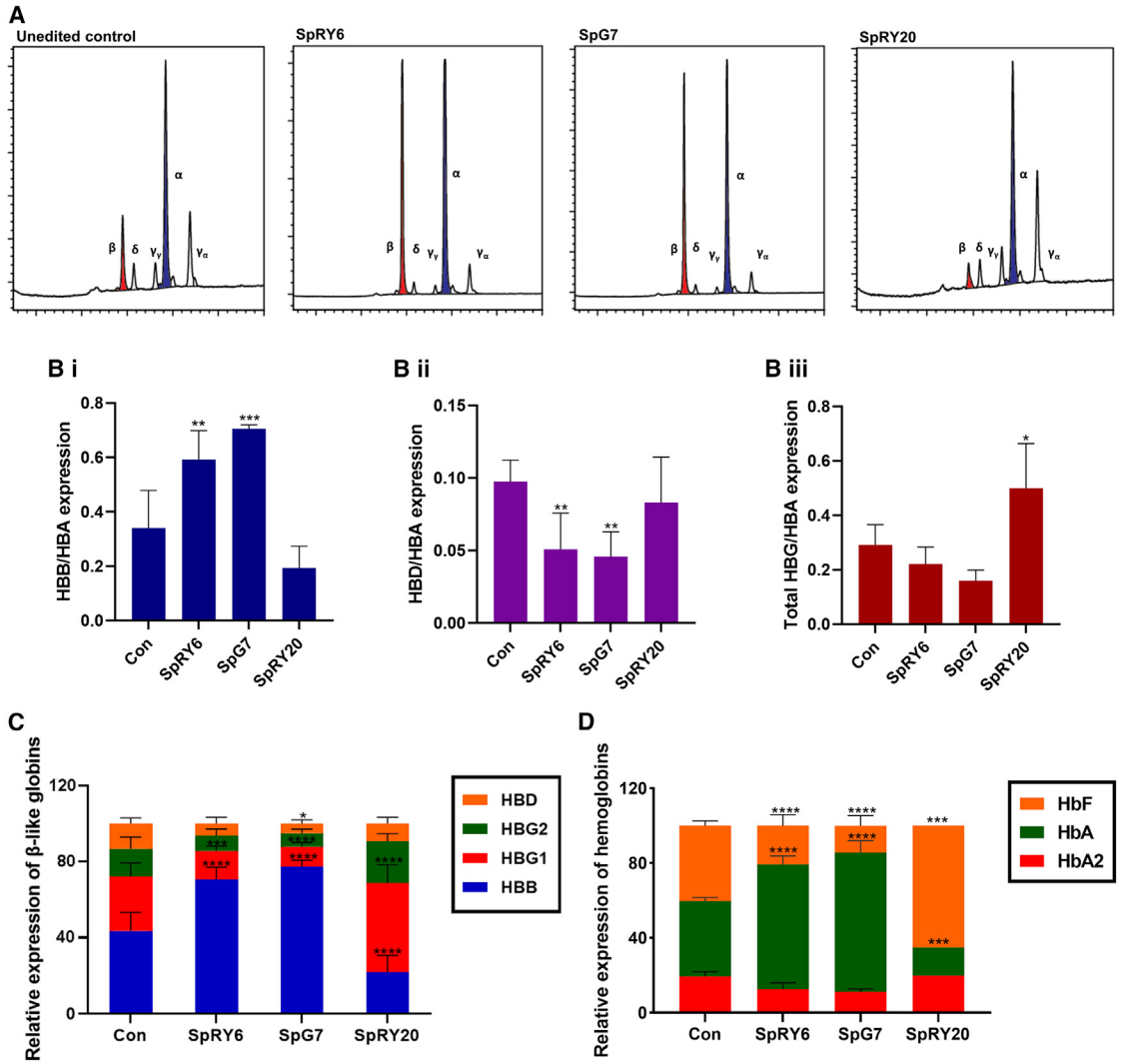


Figure 4. HPLC-based protein measurements of patient-derived CD34⁺ cells on day 14 of erythroid differentiation

(A) Representative chromatograms of RP-HPLC-based detection of human globin chains in patient cells indicating induction of HBB after SpRY6 or SpG7 treatment and reduction of HBB with SpRY20. (B) Bar graphs depicting proportion of HBB-like globin species relative to HBA expressed post-treatment for all editors and untreated controls. Groupwise comparisons were performed using one-way ANOVA with Dunnett's multiple comparison test for HBB (navy, i) and HBG (maroon, iii) and with Holm-Sidak's multiple comparisons test for HBD (purple, ii). For HBB/HBA, ***$p < 0.0001$ and **$p = 0.0031$; for HBD/HBA, **$p = 0.0085$ for SpRY6 and **$p = 0.0060$ for SpG7; for total HBG/HBA, *$p = 0.0108$. (C) Bar graph depicting proportion of each globin species post-treatment. Groupwise comparisons were performed using mixed-effects analysis with Dunnett's multiple comparison test. ****$p < 0.0001$, ***$p = 0.0004$, and *$p = 0.0454$. All as measured by RP-HPLC ($n = 3$ for SpRY20, $n = 5$ for other treatments). (D) Bar graph depicting post-editing expression of hemoglobin species after control, SpRY6, and SpG7 treatments ($n = 3$) as measured by CE-HPLC. SpRY20 data are shown as a single analysis owing to insufficient material for replicate samples. Groupwise comparisons were performed using parametric mixed-effects analysis with Dunnett's multiple comparison test. ****$p < 0.0001$ and ***$p = 0.0002$. For (B) through (D), error bars show the standard deviation of the mean, and only the upward variation is given.

editing events for SpG7, upstream of the aberrant splice site, were single base edits, in the previous study, 75% of editing events at the IVSI-110 locus were double or triple editing events proximal to the normal splice acceptor. For both studies, highly efficient functional correction renders it unlikely that individual editing events may have detrimental rather than corrective effects, but corresponding functional dissection of individual edits would call for clonal analyses, which are outside the scope of also the current study. Instead, our analyses prioritized evaluation of the therapeutic potential for SpG7, due to its more potent correction, lower indel rates, and lower predicted and observed off-target activity compared to SpRY6, toward its full preclinical evaluation with long-term repopulating cells in NBSGW mice and comparison to equivalent DSB-based editors as the current clinical benchmark.[7,27,28] In the same vein and with a focus on on-target events, the present study performed an experimental evaluation of off-target sites for the top two predictions per editor and, in the

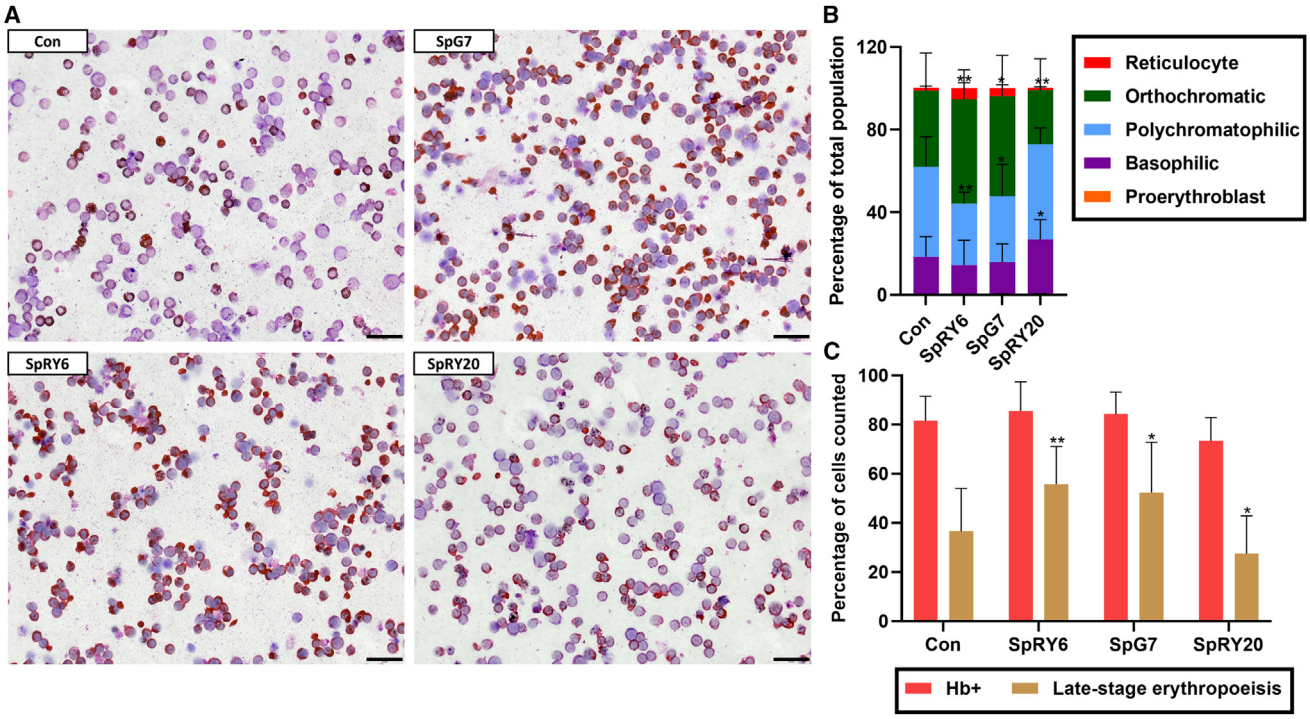


Figure 5. Cytocentrifugation and scoring of patient-derived CD34⁺ cells on day 11 of erythroid differentiation

(A) Representative cytocentrifugation images of stained treated and control cells. Size marker: 20 μm. (B) Bar chart representing differential scoring of the proportion of erythroid subpopulations. Groupwise comparison was made to untreated control using parametric mixed-effects analysis with Dunnett's multiple comparison test. For basophilic: $^{*}p = 0.0283$. For polychromatophilic: $^{**}p = 0.0030$ and $^{*}p = 0.0151$. For orthochromatophilic: $^{**}p = 0.0032$ for SpRY6, $^{**}p = 0.0016$ for SpRY20, and $^{*}p = 0.0171$. (C) Bar chart indicating the number of cells in late-stage erythropoiesis (equal to the sum of orthochromatophilic cells and reticulocytes from B) and hemoglobin-positive (Hb^{+}) cells, equivalent to late-stage erythropoietic cells plus polychromatophilic (dianisidine-staining cells). Analyses are of patient-derived cells following treatment ($n = 3$ for SpRY20, $n = 5$ for other treatments). Groupwise comparison was made to untreated controls and significance obtained using parametric mixed-effects analysis with Dunnett's multiple comparison test. $^{**}p = 0.0028$ and $^{*}p = 0.0137$ for SpG7 and $^{*}p = 0.0295$ for SpRY20. For (B) and (C), error bars show the standard deviation of the mean, and only the upward variation is given.

process, detected a significant off-target site for SpRY6, *KIAA1328*, and consequently highlighted SpG7 as safer for therapeutic application. While such experimental evaluation of off-target sites, albeit limited, compares favorably to application of only *in silico* analyses for *ex vivo* cultures elsewhere,[29] more comprehensive genotoxicity studies, as well as upscaling of manipulations and use of good-manufacturing-practice–level reagents and environments will be required to provide further safety and efficiency data toward clinical translation of our approach. Of note, SpRY20, though based on the same enzyme as SpRY6, in contrast to the latter, does not show a marked increase in on-target indel frequency. This may be due to the complementary orientation of gRNA20 compared to gRNA6, which, together with additional sequence context effects, might lead to alternative interactions, editing intermediates, and therefore nickase activity, along with inadvertent DSB behavior.

Intriguingly, the detrimental effect of SpRY20 can be explained by at least two mutually inclusive explanations. BSs are a requirement for splicing catalysis and have been found to be numerous in most human introns.[30] In the case of *HBB* intron 1, the dominant and strongest branchpoint is at position −37, but a cryptic branchpoint at position −24 exists, albeit with a lower binding energy.[30] In both cases, lariat formation occurs at an adenine[31] that bulges from the U2/BS helix, later taking on the role of reactive nucleophile,[32] which in the case of *HBB* intron 1 is base IVSI-107. SpRY20 action changes this residue and thereby limits the spliceosome to using only the distal BS. In fact, Reed et al. observed that the substitution of a guanine in place of the BS adenine (A>G, akin to action by SpRY20) altered 3′ splice site selection.[31] Moreover, the IVSI-107(A>G) base change predictably gives rise to a transcript containing a recognition motif (UGUU) for two members of the CELF family of RNA-binding proteins. CELF1 and -2 can repress or activate splicing[33] and can do the latter by displacing silencer elements in the polypyrimidine tract[34] or physically occluding *trans*-acting factors involved in branchpoint recognition.[30] Thereby, the creation of this tetranucleotide, which is sufficient for recognition[35] by the CELF proteins, might favor aberrant splicing, possibly by displacing an inhibitory element or recruiting enhancing factors exacerbating the IVSI-110 phenotype. Altogether, aggravation specifically of the characteristic hallmarks of $HBB^{IVSI\text{-}110(G>A)}$ thalassemia by SpRY20, combined withits off-target activity being being

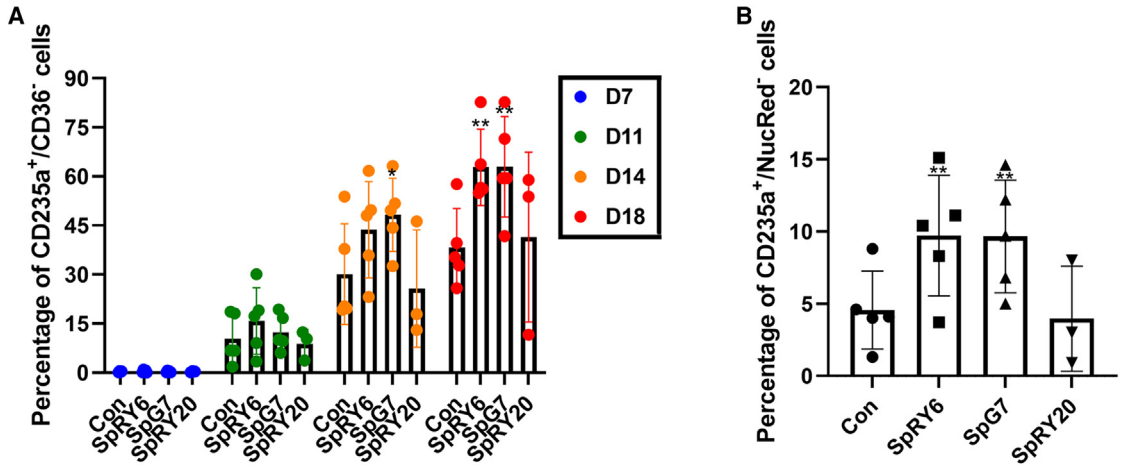


Figure 6. Flow cytometry analysis of edited vs. control cells undergoing differentiation

(A) CD235a$^+$/CD36$^-$ staining. Cells stained for differentiation markers were scored on four collection timepoints (days 7, 11, 14, and 18). Groupwise comparisons were performed using two-way ANOVA with Dunnett's multiple comparison test. Day 14: $^*p = 0.0379$; day 18: $^{**}p = 0.0032$ for SpRY6 and $^{**}p = 0.0030$ for SpG7. (B) CD235a$^+$/NucRed$^-$ staining. Corresponding representative flow cytometry scatterplots are shown in Figures S10 and S11 and corresponding data for cell proliferation in Figure S12. On the final day of differentiation (day 18), cells were additionally stained with the nucleation marker NucRed; the chart shows the proportion of enucleated cells in the CD235a$^+$ cell population. Significance was obtained using two-way ANOVA with Dunnett's multiple comparison test. $^{**}p = 0.0036$ for SpRY6 and $^{**}p = 0.0039$ for SpG7. Statistical analyses were performed in comparison to control for both analyses of live patient cells following treatment ($n = 3$ for SpRY20, $n = 5$ for other treatments). For (A) and (B), error bars show the standard deviation of the mean.

below the detection level for the top two off-target sites, point to SpRY20 on-target action rather than to off-target action as the basis of its effects. In the same vein, HbF induction by SpRY20 may be explained by decreased expression of β-globin, which is commonly associated with detection of increased γ-globin levels, for which a range of complementary mechanisms have been proposed at the transcriptional, translational, and post-translational levels.[36–39] The interconnectedness of, first, the expression of different genes on the β-globin locus and, second, sequence elements and steps of gene expression, from DNA accessibility to post-translational action, for each gene on that locus, poses a persistent challenge to distinguishing cause and effect for corresponding expression events. However, beyond the scope of the current study, surer insights into mechanisms underlying the effects of SpRY20 and of the therapeutic BEs presented here may be gained by further experimentation, such as by the modification of individual bases and protein factors suspected of involvement, in conjunction with monitoring of the resulting transcripts.

Many advanced treatment options are being developed for hemoglobinopathies and have achieved validation of gene addition and genome-editing approaches in clinical trials, where prolonged or persistent transfusion dependence for severe β-thalassemia genotypes is still a frequent occurrence. The ongoing CRISPR-Therapeutics/Vertex *ex vivo* genome-editing trial utilizes efficient genome disruption to induce HbF with good efficacy data but with the inherent risk of mutagenesis typical of DSB-based DNA-editing approaches. To address the issue of safety, clinical trials aiming for HbF induction now also include BEs with high on-target fidelity.[40] However, therapies based on HbF may be more affected by genetic modifiers, such as those affecting γ-globin expression, than therapies based on HbA, which has evolved for optimal expression and function in adults.

The present study addresses many concerns over advanced therapies and their safety in the therapy of β-thalassemia by applying DSB-independent and efficient base editing to restore HbA and normal phenotypes based on methodology of virus- and DNA-free, highly transient delivery of editors for cell manipulation. Our approach is tailored for the *HBB*$^{IVSI\text{-}110(G>A)}$ mutation and, in many affected individuals, may thus prove superior to universal therapeutic approaches. Beyond *HBB*$^{IVSI\text{-}110(G>A)}$, exploitation of context bases for phenotypic correction gives greater flexibility in the choice of efficient and safe DNA editors than does targeting the causative mutation alone. DSB-independent genome editors might therefore be of utility as a therapeutic approach in a range of other disorders caused by abnormal splicing, in particular where the underlying mutation itself is inaccessible to direct editing.

MATERIALS AND METHODS

Isolation and culture of PB-derived hCD34$^+$ HSPCs

Utilization of primary samples was based on informed consent according to the Declaration of Helsinki and was covered by bioethics approval from the Cyprus National Bioethics Committee (license numbers ΕΕΒΚ/ΕΠ/2018/52 and ΕΕΒΚ/ΕΠ/2022/05). Isolation of CD34$^+$ cells from peripheral blood (PB) mononuclear cells was performed using the CD34$^+$ MicroBead Kit, LS MiniMACS Column, by magnetic separation using an autoMACS Separator (Miltenyi Biotec, Bergisch Gladbach, Germany) from five IVSI-110-homozygous patients with fully informed consent. Isolated cells were then expanded and, after editing,

induced to differentiate into the erythroid lineage in a three-phase program in which medium contents and concentration levels varied. Briefly, in phase 1 of differentiation (days 1–7), 0.2–1.0 × 10^4 cells/mL were maintained in Iscove's modified Dulbecco's medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 5% human serum (Sigma-Aldrich, St. Louis, MO, USA), 2 IU/mL heparin (Sigma-Aldrich), 10 μg/mL insulin (Sigma-Aldrich), 330 μg/mL holo-transferrin (Sigma-Aldrich), 1 μM hydrocortisone (Medochemie, Limassol, Cyprus), 100 ng/mL stem cell factor (SCF; R&D Systems, Minneapolis, MN, USA), 5 ng/mL interleukin-3 (IL-3; R&D Systems), 3 U/mL erythropoietin (Sandoz GmbH, Kundl, Austria), 1× penicillin-streptomycin (Thermo Fisher Scientific), and GlutaMAX (Thermo Fisher Scientific). In phase 2 (days 7–11), IL-3 and hydrocortisone were omitted, and cell concentration was kept at 1–2 × 10^5 cells per mL. In phase 3 (days 11–18), SCF was omitted compared to phase 2, and the concentration was maintained at 1–2 × 10^6 cells per mL.

Design and synthesis of gRNAs

gRNAs were manually designed based on the availability of an appropriate PAM adjacent to the target site and on positioning of the editing window for the desired base substitution.[16]

All gRNAs were purchased from Synthego (Redwood City, CA, US) for use in combination with *in-vitro*-synthesized Cas mRNA (for BEs).

In vitro transcription

The Cas mRNA for BEs was synthesized with the HiScribe T7 Anti-Reverse Cap Analog mRNA kit with tailing (New England Biolabs [NEB], Ipswich, MA, US) following the manufacturer's instructions, as described elsewhere.[18]

Cloning procedure

Original plasmids were received from Addgene (Watertown, MA, USA) and cultured, purified, and verified by analytical restriction digest using standard methods.[41] Excision of GFP from the nearly PAM-less pCMV-T7-ABEmax(7.10)-SpG-P2A-EGFP and pCMV-T7-ABEmax(7.10)-SpRY-P2A-EGFP plasmids (Addgene #140002 and #140003; gifts from Benjamin Kleinstiver) involved isolation of the ABEs from their backbone and insertion into a classical, GFP-free BE backbone (pCMV_ABEmax; Addgene #112093; gift from David Liu). Resulting bacterial clones were confirmed by DNA sequencing, deposited to Addgene, and are available under accession numbers #195278 and #195279.

Nucleofection

Nucleofection of *in-vitro*-transcribed mRNA was performed using the Lonza 4D-Nucleofector with the P3 Primary Cell 4D-Nucleofector X Kit (both Lonza, Basel, Switzerland) with the CA-137 program according to our published procedures[18] and manufacturer's instructions, respectively.

PCR

Amplification was performed using the Q5 High-Fidelity PCR Kit (NEB) and the Veriti thermal cycler (TFS, Waltham, Massachusetts, USA) according to the manufacturer's instructions. Primer sequences (Metabion International AG, Martinsried, Germany) can be found in Table S1.

gDNA extraction and sequencing

Genomic DNA from eukaryotic cells was extracted using QIAmp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Cycle sequencing was performed using the BigDye Terminator v.1.1 Cycler sequencing kit (Applied Biosystems, Waltham, Massachusetts, USA) and the Veriti thermal cycler (TFS). Performa DTR Gel Filtration Cartridges (Edge Biosystems, Gaithersburg, Maryland, USA) were used for purification of sequencing products, all in accordance with manufacturers' instructions, and analyzed on a Hitachi 3031xl Genetic Analyzer with Sequence Detection Software v.5.2 (Applied Biosystems).

Targeted deep sequencing

The Devyser Thalassemia kit was used to analyze *HBB* and *HBA* gene clusters and to generate a target-specific library, following the manufacturer's protocol (Dvysr, Hägersten, Sweden). This was then subjected to NGS using the MiSeq Sequencing platform, employing the MiSeq Reagent MicroKit v.2 (Illumina, San Diego, CA, USA) (300 cycles) with a high coverage of 200×.

HPLC

RP-HPLC of globin chains and CE-HPLC of hemoglobins were performed by modification of our published procedures and manufacturer's instructions, respectively.[26,42]

Cell counting

Cell counting was performed manually using a Bright-Line hemocytometer (Sigma-Aldrich, Munich, Germany), and cell viability was measured using trypan blue solution, 0.4% (Thermo Fisher Scientific). Cells were counted under a light microscope (Eclipse TS100, Nikon, Chiyoda, Japan).

Cytocentrifugation and microscopy

For cell preparations on slides, the Tharmac Cellspin II cytocentrifuge with an EASY rotor (Tharmac, Wiesbaden, Germany) was used, followed by dianisidine and standard May-Grünwald/Giemsa (Fluka, Munich, Germany) staining, as described elsewhere.[26] The proliferation rate for post-edited cells on the final day of differentiation is shown in Figure S12.

Assessment of RNA expression

RNA was isolated using TRIzol Reagent (Thermo Fisher Scientific) following the associated user manual offered by the company (MAN0001271). cDNA was then created and used as template for qPCR analysis as described elsewhere.[4]

For ddPCR, experiments were performed using the QX200 AutoDG ddPCR system (Bio-Rad Laboratories, Hercules, CA, USA) following the manufacturer's protocol. Briefly, each 25 μL reaction included 12.5 μL ddPCR Supermix for probes (no dUTP), 450 nM of each

primer, and 2 μL of the sample. Droplet generation was performed on the AutoDG instrument, followed by PCR amplification on a BioRad T100 Thermal Cycler (Bio-Rad Laboratories) using the following thermal protocol: 10 min at 95°C, 40 cycles of 30 s at 95°C, 60 s at 50°C, and final hold at 4°C. Each PCR step included a 2°C/s ramp rate. The plate was read using the QX200 droplet reader and analyzed using the QuantaSoft software (v.1.7.4.0917) (Bio-Rad Laboratories).

Primer and probe sequences (Metabion International AG) can be found in Table S1.

Flow cytometry analysis

Antibodies for early, anti-Hu CD36 APC and late, anti-Hu CD235a PE (both from ExBio, Vestec, Czech Republic) were used for surface marker analysis. NucRed Live 647 ReadyProbe Reagent (Thermo Fisher Scientific) staining was also utilized. Samples were run on the BD FACSAria III (BD Biosciences, Franklin Lakes, NJ, USA), and post-run analysis was performed on the FCS Express 7 software.

Statistical and bioinformatics analyses

Statistical analyses, such as normality tests (Shapiro-Wilks) and correspondingly parametric or non-parametric groupwise comparisons, as specified for each figure panel, were performed using GraphPad Prism 9 software (GraphPad, La Jolla, CA, USA).

For DNA-level evaluation of base editing, the EditR software (https://moriaritylab.shinyapps.io/editr_v10/) was used.[19]

Splice site analyses were performed using the Human Splicing Finder software (https://www.genomnis.com/access-hsf).[43]

NGS data processing was performed using third-party tools and custom scripts in R programming language v.4.3.1.[44] NGS bioinformatics analysis was performed using a combination of third-party publicly available software and custom-written scripts. Specifically, all FASTQ files were processed using the nf-core/Sarek pipeline written in the nextflow framework.[45] Reads underwent quality control, filtering, and trimming of residual adapter sequences following alignment to the GRCh37 genome using BWA-MEMMEM.[46] The aligned BAM files were marked for duplicates and processed with samtools v.1.10[47] to extract the reads aligned to the region of interest (ROI), which was defined as the 20 bases targeted by the gRNA ± 50 bases upstream and downstream of the target site.

The resulting ROI BAM files were parsed and processed in R using the Rsamtools,[48] tidyverse,[49] and matrixStats[50] packages. Specifically, using the CIGAR notation, all reads were represented in a matrix format of dimensions $d * w$ where d is the depth of sequencing and w is the number of bases at the ROI. Base compositions were calculated as proportions for each position (column):

$$P_{i,j} = \frac{C_i}{d_j} \quad i \in \{A, T, C, G, N\}, j \in \{1, 2, 3 \ldots w\},$$

where Pi,j is the proportion of the base i calculated as the counts of i over the depth dj at the jth position.

Similarly, proportions of 7-nucleotide sequences overlapping the region of the targeted bases at chr11: 5,248,050–5,248,056 were calculated using the *table()* function in R.

Alignments at the ROI that comprised at least one small indel were extracted and analyzed further to calculate their frequencies at the target site (gRNA 20 bp segments) and flanking regions (sequences in the ROI that do not overlap the gRNA binding site). The rates were calculated as follows:

$$R_{x,r} = \frac{N_{xr}}{l_r \cdot \overline{d_r}} \quad x \in \{Ins, Del\}, r \in \{Target, Flanks\},$$

where N_{xr} is the number of reads in region r, which contains an indel over the mean depth $\overline{d_r}$ and length lr (in bases) of the specific region. Fold changes of R_{xr} were calculated by first correcting for baseline rates by subtracting the respective flanking region rates $R_{x,Flanks}$ (assumed as baseline) from the $R_{x,Target}$ of each sample and then dividing by the corresponding corrected mean rate of the controls.

The alignments, base and read proportions, and indel rates were visualized using the ggplot2 package in R. All NGS processing was performed on a Linux (Ubuntu 20.04.6 LTS) server with a 64-core CPU and 1 TB RAM.

DATA AND CODE AVAILABILITY

All data are available in the article. Correspondence and requests for materials should be addressed to C.W.L. Reprints and permissions information is available.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2024.102183.

ACKNOWLEDGMENTS

The project was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (EXCELLENCE/1216/0092 and EXCELLENCE/0421/0086), by a Telethon Cyprus PhD studentship, by an A. G. Leventis Foundation educational grant, by COST Action CA21113 (GenE-HumDi), supported by COST (European Cooperation in Science and Technology), and by core funding of the Cyprus Institute of Neurology & Genetics. Reagent and equipment support by the project "New Infrastructure for the Diagnosis and Therapy of Patients" was funded by the Norway Grants 2014–2021. We are grateful to Thessalia Papasavva for advice on NGS analyses and to all of our patients for

generously donating their samples and for making our work possible. The graphical abstract was in part created using BioRender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, formal analysis, data curation, and visualization, B.N., P.P., and C.W.L.; investigation and methodology, B.N., P.L.P., N.Y.P., M.T., L.K., X.F., C.G.C., P.P., and C.W.L.; resources, M.S., S.C., M.K., P.P., and C.W.L.; supervision, M.K., P.P., and C.W.L.; writing – original draft preparation, B.N., M.T., P.P., and C.W.L.; writing – review & editing, B.N., P.L.P., M.T., P.P., and C.W.L.; project administration, M.K., P.P., and C.W.L.; funding acquisition: M.K., P.P., and C.W.L. All authors have read and agreed to the published version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Origa, R. (2017). β-Thalassemia. Genet. Med. *19*, 609–619. https://doi.org/10.1038/gim.2016.173.
2. Kountouris, P., Lederer, C.W., Fanis, P., Feleki, X., Old, J., and Kleanthous, M. (2014). IthaGenes: An interactive database for haemoglobin variations and epidemiology. PLoS One *9*, e103020. https://doi.org/10.1371/journal.pone.0103020.
3. Busslinger, M., Moschonas, N., and Flavell, R.A. (1981). β+ Thalassemia: Aberrant splicing results from a single point mutation in an intron. Cell *27*, 289–298. https://doi.org/10.1016/0092-8674(81)90412-8.
4. Patsali, P., Papasavva, P., Christou, S., Sitarou, M., Antoniou, M.N., Lederer, C.W., and Kleanthous, M. (2020). Relative and absolute quantification of aberrant and normal splice variants in hbbivsi−110 (G > a) β-thalassemia. Int. J. Mol. Sci. *21*, 1–17. https://doi.org/10.3390/ijms21186671.
5. Angelucci, E. (2010). Hematopoietic stem cell transplantation in thalassemia. Hematology (United States) *2010*, 456–462. https://doi.org/10.1182/asheducation-2010.1.456.
6. Patsali, P., Papasavva, P., Stephanou, C., Christou, S., Sitarou, M., Antoniou, M.N., Lederer, C.W., and Kleanthous, M. (2018). Short-hairpin RNA against aberrant HBBIVSI-110(G>A) mRNA restores β-globin levels in a novel cell model and acts as mono- and combination therapy for β-thalassemia in primary hematopoietic stem cells. Haematologica *103*, e419–e423. https://doi.org/10.3324/haematol.2018.189357.
7. Patsali, P., Mussolino, C., Ladas, P., Floga, A., Kolnagou, A., Christou, S., Sitarou, M., Antoniou, M.N., Cathomen, T., Lederer, C.W., and Kleanthous, M. (2019). The Scope for Thalassemia Gene Therapy by Disruption of Aberrant Regulatory Elements. J. Clin. Med. *8*, 1959. https://doi.org/10.3390/jcm8111959.
8. Patsali, P., Turchiano, G., Papasavva, P., Romito, M., Loucari, C.C., Stephanou, C., Christou, S., Sitarou, M., Mussolino, C., Cornu, T.I., et al. (2019). Correction of IVS I-110(G>A) β-thalassemia by CRISPR/Cas- And TALEN-mediated disruption of aberrant regulatory elements in human hematopoietic stem and progenitor cells. Haematologica *104*, E497–E501. https://doi.org/10.3324/haematol.2018.215178.
9. El-Beshlawy, A., Mostafa, A., Youssry, I., Gabr, H., Mansour, I.M., El-Tablawy, M., Aziz, M., and Hussein, I.R. (2008). Correction of aberrant pre-mRNA splicing by antisense oligonucleotides in β-thalassemia Egyptian patients with IVSI-110 mutation. J. Pediatr. Hematol. Oncol. *30*, 281–284. https://doi.org/10.1097/MPH.0b013e3181639afe.
10. Hardouin, G., Antoniou, P., Martinucci, P., Felix, T., Manceau, S., Joseph, L., Masson, C., Scaramuzza, S., Ferrari, G., Cavazzana, M., and Miccio, A. (2023). Adenine base editor–mediated correction of the common and severe IVS1-110 (G>A) β-thalassemia mutation. Blood *141*, 1169–1179. https://doi.org/10.1182/blood.2022016629.
11. Wu, Y., Zeng, J., Roscoe, B.P., Liu, P., Yao, Q., Lazzarotto, C.R., Clement, K., Cole, M.A., Luk, K., Baricordi, C., et al. (2019). Highly efficient therapeutic gene editing of human hematopoietic stem cells. Nat. Med. *25*, 776–783. https://doi.org/10.1038/s41591-019-0401-y.
12. Puthenveetil, G., Scholes, J., Carbonell, D., Qureshi, N., Xia, P., Zeng, L., Li, S., Yu, Y., Hiti, A.L., Yee, J.K., and Malik, P. (2004). Successful correction of the human β-thalassemia major phenotype using a lentiviral vector. Blood *104*, 3445–3453. https://doi.org/10.1182/blood-2004-04-1427.
13. Papasavva, P., Kleanthous, M., and Lederer, C.W. (2019). Rare Opportunities: CRISPR/Cas-Based Therapy Development for Rare Genetic Diseases. Mol. Diagn. Ther. *23*, 201–222. https://doi.org/10.1007/s40291-019-00392-3.
14. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826. https://doi.org/10.1038/nbt.2623.
15. Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science *368*, 290–296. https://doi.org/10.1126/science.aba8853.
16. Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788. https://doi.org/10.1038/s41576-018-0059-1.
17. Youn, H., and Chung, J.K. (2015). Modified mRNA as an alternative to plasmid DNA (pDNA) for transcript replacement and vaccination therapy. Expet Opin. Biol. Ther. *15*, 1337–1348. https://doi.org/10.1517/14712598.2015.1057563.
18. Papaioannou, N.Y., Patsali, P., Naiisseh, B., Papasavva, P.L., Koniali, L., Kurita, R., Nakamura, Y., Christou, S., Sitarou, M., Mussolino, C., et al. (2023). High-efficiency editing in hematopoietic stem cells and the HUDEP-2 cell line based on in vitro mRNA synthesis. Front. Genome Ed. *5*, 1141618. https://doi.org/10.3389/fgeed.2023.1141618.
19. Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahante, J.E., Webber, B.R., and Moriarity, B.S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. CRISPR J. *1*, 239–250. https://doi.org/10.1089/crispr.2018.0014.
20. Concordet, J.P., and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. *46*, W242–W245. https://doi.org/10.1093/nar/gky354.
21. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9 Synthesis of an arrayed sgRNA library targeting the human genome. Nat. Biotechnol. *34*, 184–191. https://doi.org/10.1038/nbt.3437.Optimized.
22. Qi, W.-H., Yan, C.-C., Li, W.-J., Jiang, X.-M., Li, G.-Z., Zhang, X.-Y., Hu, T.-Z., Li, J., and Yue, B.-S. (2016). Distinct patterns of simple sequence repeats and GC distribution in intragenic and intergenic regions of primate genomes. Aging (Albany NY) *8*, 2635–2654. https://doi.org/10.18632/aging.101025.
23. Patel, C.A., and Ghiselli, G. (2005). Hinderin, a five-domains protein including coiled-coil motifs that binds to SMC3. BMC Cell Biol. *6*, 3. https://doi.org/10.1186/1471-2121-6-3.
24. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832. https://doi.org/10.1038/nbt.2647.
25. Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods *12*, 982–988. https://doi.org/10.1038/nmeth.3543.
26. Loucari, C.C., Patsali, P., Van Dijk, T.B., Stephanou, C., Papasavva, P., Zanti, M., Kurita, R., Nakamura, Y., Christou, S., Sitarou, M., et al. (2018). Rapid and Sensitive Assessment of Globin Chains for Gene and Cell Therapy of Hemoglobinopathies. Hum. Gene Ther. Methods *29*, 60–74. https://doi.org/10.1089/hgtb.2017.190.
27. Turchiano, G., Andrieux, G., Klermund, J., Blattner, G., Pennucci, V., el Gaz, M., Monaco, G., Poddar, S., Mussolino, C., Cornu, T.I., et al. (2021). Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. Cell Stem Cell *28*, 1136–1147.e5. https://doi.org/10.1016/j.stem.2021.02.002.
28. Frangoul, H., Altshuler, D., Cappellini, M.D., Chen, Y.-S., Domm, J., Eustace, B.K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., et al. (2021). CRISPR-Cas9

Gene Editing for Sickle Cell Disease and β-Thalassemia. N. Engl. J. Med. *384*, 252–260. https://doi.org/10.1056/nejmoa2031054.

29. Xu, S., Luk, K., Yao, Q., Shen, A.H., Zeng, J., Wu, Y., Luo, H.Y., Brendel, C., Pinello, L., Chui, D.H.K., et al. (2019). Editing aberrant splice sites efficiently restores β-globin expression in b-thalassemia. Blood *133*, 2255–2262. https://doi.org/10.1182/blood-2019-01-895094.

30. Pineda, J.M.B., and Bradley, R.K. (2018). Most human introns are recognized via multiple and tissue-specific branchpoints. Genes Dev. *32*, 577–591. https://doi.org/10.1101/gad.312058.118.

31. Reed, R., and Maniatis, T. (1988). The role of the mammalian branchpoint sequence in pre-mRNA splicing. Genes Dev. *2*, 1268–1276. https://doi.org/10.1101/gad.2.10.1268.

32. Cretu, C., Gee, P., Liu, X., Agrawal, A., Nguyen, T.V., Ghosh, A.K., Cook, A., Jurica, M., Larsen, N.A., and Pena, V. (2021). Structural basis of intron selection by U2 snRNP in the presence of covalent inhibitors. Nat. Commun. *12*, 4491. https://doi.org/10.1038/s41467-021-24741-1.

33. Goo, Y.H., and Cooper, T.A. (2009). CUGBP2 directly interacts with U2 17S snRNP components and promotes U2 snRNA binding to cardiac troponin T pre-mRNA. Nucleic Acids Res. *37*, 4275–4286. https://doi.org/10.1093/nar/gkp346.

34. Gromak, N., Matlin, A.J., Cooper, T.A., and Smith, C.W.J. (2003). Antagonistic regulation of α-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. RNA *9*, 443–456. https://doi.org/10.1261/rna.2191903.

35. Jensen, K.B., Musunuru, K., Lewis, H.A., Burley, S.K., and Darnell, R.B. (2000). The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. Proc. Natl. Acad. Sci. USA *97*, 5740–5745. https://doi.org/10.1073/pnas.090553997.

36. Hardison, R.C. (2022). Promoter competition in globin gene control. Blood *139*, 2089–2091. https://doi.org/10.1182/blood.2022015642.

37. Rees, D.C., Porter, J.B., Clegg, J.B., and Weatherall, D.J. (1999). Why are hemoglobin F levels increased in HbE/beta thalassemia? Blood *94*, 3199–3204.

38. Efremov, D.G., Dimovski, A.J., Sukarova, E., Schilirò, G., Zisovski, N., Efremov, G.D., Burrone, O.R., and Huisman, T.H. (1994). Gamma-mRNA and Hb F levels in beta-thalassaemia. Br. J. Haematol. *88*, 311–317. https://doi.org/10.1111/j.1365-2141.1994.tb05024.x.

39. Qiu, C.C., and Abraham, E.C. (1988). Evidence for posttranslational control of fetal hemoglobin synthesis. Am. J. Hematol. *29*, 58–59. https://doi.org/10.1002/ajh.2830290116.

40. Han, W., Qiu, H.-Y., Sun, S., Fu, Z.-C., Wang, G.-Q., Qian, X., Wang, L., Zhai, X., Wei, J., Wang, Y., et al. (2023). Base editing of the HBG promoter induces potent fetal hemoglobin expression with no detectable off-target mutations in human HSCs. Cell Stem Cell *30*, 1624–1639.e8. https://doi.org/10.1016/j.stem.2023.10.007.

41. Green, M.R., and Sambrook, J. (2012). Molecular Cloning: A Laboratory Manual Fourth (Cold Spring Harbor Laboratory Press, Cold Spring Harbor).

42. Ou, C.N., Buffone, G.J., Reimer, G.L., and Alpert, A.J. (1983). High-performance liquid chromatography of human hemoglobins on a new cation exchanger. J. Chromatogr. *266*, 197–205. https://doi.org/10.1016/s0021-9673(01)90893-3.

43. Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, e67. https://doi.org/10.1093/nar/gkp215.

44. Ihaka, R., and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. J. Comput. Graph Stat. *5*, 299–314. https://doi.org/10.2307/1390807.

45. Garcia, M., Juhos, S., Larsson, M., Olason, P.I., Martin, M., Eisfeldt, J., DiLorenzo, S., Sandgren, J., Díaz De Ståhl, T., Ewels, P., et al. (2020). Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Res. *9*, 63. https://doi.org/10.12688/f1000research.16665.2.

46. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

47. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

48. Morgan, M. (2011). An Introduction to Rsamtools. Ind Commer Train *10*, 11–18.

49. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.

50. Bengtsson, H., Ahlmann-Eltze, C., Bravo, H.C., Gentleman, R., Gleixner, J., Hickey, P., Hossjer, O., Jaffee, H., Jiang, D., Langfelder, P., et al. (2023). matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). Version 1.2.0. https://github.com/HenrikBengtsson/matrixStats.

OMTN, Volume *35*

Supplemental information

Context base editing for splice correction of IVSI-110 β-thalassemia

Basma Naiisseh, Panayiota L. Papasavva, Nikoletta Y. Papaioannou, Marios Tomazou, Lola Koniali, Xenia Felekis, Constantina G. Constantinou, Maria Sitarou, Soteroula Christou, Marina Kleanthous, Carsten W. Lederer, and Petros Patsali

Supplemental Tables

Table S1. Primer identities for DNA and RNA expression analyses

Primer Name	Application	Sequence
wtHBB_Probe_ZNA	RT-qPCR/ddPCR	6-FAM-TGGG(PDC)AGG(PDC)TG(PDC)TG-ZNA-3-BHQ-1 VIC-TAAGGGTGGGAAAATAGA-MGB
IVSI-110_MGB_Probe	RT-qPCR/ddPCR	VIC-TAAGGGTGGGAAAATAGA-MGB
hHBB_FW	RT-qPCR	GCCCTGGCCCACAAGTATC
hHBB_RV	RT-qPCR	GCCCTTCATAATATCCCCCAGTT
hHBA_FW	RT-qPCR	GGTCAACTTCAAGCTCCTAAGC
hHBA_RV	RT-qPCR	GCTCACAGAAGCCAGGAACTTG
hHBB_EX1_FW_3	RT-qPCR/ddPCR	GGGCAAGGTGAACGTG
hHBB_EX2_RV_1	RT-qPCR/ddPCR	GGACAGATCCCCAAAGGAC
HBB_CD34_FW	PCR/Sequencing	TGAGGAGAAGTCTGCCGTTAC
HBB_CD34_RV	PCR/Sequencing	CAGCTCACTCAGTGTGGC
SpG7_OFF1_RP11-275O18.1-RP11-362A1.1_FW	PCR/Sequencing	AATACTCTCCTTTCATGTGCCATG
SpG7_OFF1_RP11-275O18.1-RP11-362A1.1_RV	PCR/Sequencing	ATGGCTTCTCTCTTGTGAGGAG
SpG7_OFF2/SpRY6(NNN)_OFF1_ YWHAEP4/AC097467.2-RP11-27G13.4_FW	PCR/Sequencing	CCCAGGAGAGCATGGAAAGG
SpG7_OFF2/SpRY6(NNN)_OFF1_ YWHAEP4/AC097467.2-RP11-27G13.4_RV1	PCR/Sequencing	AGGAACTCTCAAGAACTGCAAGC
SpG7_OFF2/SpRY6(NNN)_OFF1_ YWHAEP4/AC097467.2-RP11-27G13.4_RV2	PCR/Sequencing	GCAATAGGCGAGACATCACCG
SpRY6(NNN)_OFF2_ RP11-781M16.2_FW	PCR/Sequencing	ACCATACCAGTACCTTGGC
SpRY6(NNN)_OFF2_ RP11-781M16.2_RV	PCR/Sequencing	CTGACTGGCTGGAAGGAG
SpRY6(NYN)_OFF1_ KIAA1328_FW	PCR/Sequencing	AGACACAAAGGTTGAAAGTTAAAAGG
SpRY6(NYN)_OFF1_ KIAA1328_RV	PCR/Sequencing	TCATATATTTTGAGGTGCTGTTTGG
SpRY6(NYN)_OFF2_ RPS23P3-RNU6-699P_FW	PCR/Sequencing	ACAAGTGCCCTCACTGTCA
SpRY6(NYN)_OFF2_ RPS23P3-RNU6-699P_RV	PCR/Sequencing	ACCTGAGCATCCCCTGAGTC
SpRY20(NNN)_OFF1_ NCOA7-AS1-NCOA7_FW1	PCR/Sequencing	ACCACATTTCATTATTTTGTTGAC
SpRY20(NNN)/(NRN)_OFF1_ NCOA7-AS1-NCOA7_FW2	PCR/Sequencing	AATTATGCTCATAGTGAGGG
SpRY20(NNN)/(NRN)_OFF1_ NCOA7-AS1-NCOA7_RV	PCR/Sequencing	CAATCTAGGGAGGGCAG
SpRY20(NNN)_OFF2_ SNRPGP2-RP11-61D1.2_FW	PCR/Sequencing	ATGGGTGAGTTTATGACTCTTGAGAG
SpRY20(NNN)_OFF2_ SNRPGP2-RP11-61D1.2_RV	PCR/Sequencing	TGTTTCTGGGGCAGCAGTC
SpRY20(NRN)_OFF2_ AC009541.1-hsa-mir-490_FW	PCR/Sequencing	GCTCAGCCATTCTACTTCTAGGCA
SpRY20(NRN)_OFF2_ AC009541.1-hsa-mir-490_RV	PCR/Sequencing	AGCGGTATTTTCTCAGGATGGTGG

Table S2. Top 10 off-target predictions for SpG7 (ACTAATAGGCAGAGAGAGTCAGT) as ranked by CRISPOR.[1] PAM = NGN.

Off-target Sequence	Mismatch Positions	Mismatch Count	MIT Off-target	CFD Off-target	Chromo-some	Strand	Locus Description
ATTAATAGGAAAAGAGAGTCAGG	.*.....*.*.....	3	1.275483	0.735354	Chr12	+	intergenic:RP11-275O18.1-RP11-362A1.1
ACTAAAAGGCAGAGAGAGACAGG	.....*.....*.	2	2.105189	0.247619	Chr4	-	intergenic:YWHAEP4/AC097467.2-RP11-27G13.4
GCTGATAGGCAGAGAGAGACAGG	*..*.....*.	3	1.127119	0.178571	Chr16	-	intergenic:RP11-53L24.1-RPSAP56
GCAAATAAGAAGAGAGAGTCAGA	*.*....*.*.....	4	1.299248	0.042989	Chr13	-	intergenic:LINC00333-LINC00375
ATTAATTGGAAGAGAGAGTCAGA	.*....*..*.....	3	1.680987	0.023937	Chr14	+	intron:EML5
AGTATTAGGCAGAGAGAGTCAGA	.*..*.....	2	5.722892	0.017567	Chr4	+	exon:RP11-781M16.2
ATTTATAGGCTGAGAGAGTCTGA	.*.*.....*.....	3	1.483122	0.012361	Chr9	-	intergenic:RNA5SP279-RP11-443B9.1
TCTCCTAGGAAGAGAGAGTCTGA	*..**....*.....	4	1.317696	0.010621	Chr10	-	intergenic:LINC00841-AL512640.1
AGAAATAGGCAGAGAGAGACAGA	.**.....*.	3	1.040778	0.009859	Chr16	-	intergenic:AC136932.2-LINC00273/AC136932.1
AGAGATAGGGAGAGAGAGTCAGA	.***.....*.....	4	1.239513	0.008387	Chr19	+	intergenic:TUBB4A-TNFSF9

Yellow highlight identifies genes sequenced for off-target activity.

Table S3.1. Top 10 off-target predictions for SpRY6 (CTAATAGGCAGAGAGAGTCA) as ranked by CRISPOR.[1] PAM = NNN

Off-target Sequence	Mismatch Positions	Mismatch Count	MIT Off-target	CFD Off-target	Chromosome	Strand	Locus Description
CTAAAAGGCAGAGAGAGACAGGG	....*.....*..	2	2.165116	0.333333	Chr4	-	intergenic:YWHAEP4/AC097467.2-RP11-27G13.4
GTATTAGGCAGAGAGAGTCAGAG	*..*.....	2	5.722892	0.150637	Chr4	+	exon:RP11-781M16.2
CTCCTAGGCTGAGAGAGTCAAGG	..**.....*.....	3	2.309774	0.133465	Chr6	-	intergenic:RP3-404K8.2-HDGFL1
GTCACAGGCAGAGAGAGTCACAG	*.*.*.....	3	2.392593	0.101449	Chr11	-	exon:CWF19L2
ACCATAGGCAGAGAGAGTCATAG	***.....	3	2.287424	0.095238	Chr15	+	intron:GRAMD2
GTAATAGGCAGAAAGAGTCAAGA	*.....*.....	2	3.91117	0.058528	Chr8	+	intergenic:RP11-192P9.1-TRPS1
CTAGTAGACCGAGAGAGTCACAG	...*...*.*.....	3	2.34257	0.054012	Chr5	-	intergenic:IPO11/KIF2A-IPO11
CTACAAGGCTGAGAGAGTCAAAG	...**....*.....	3	2.34257	0.040369	Chr5	+	intergenic:CTC-505O3.3-CTC-505O3.2
CCAATAGGCAGAGAGAGTCAGTG	.*.....	1	100	0.033395	Chr11	+	exon:HBB
CTGATAGGCAGAGAGAGACAGGA	..*.....*..	2	2.62276	0.028292	Chr16	-	intergenic:RP11-53L24.1-RPSAP56

Yellow highlight identifies genes sequenced for off-target activity.

Table S3.2. Top 7[†] off-target predictions for SpRY6 (CTAATAGGCAGAGAGAGTCA) as ranked by CRISPOR.[1] PAM = NYN

Off-target Sequence	Mismatch Positions	Mismatch Count	MIT Off-target	CFD Off-target	Chromosome	Strand	Locus Description
CCAATAGGCAGAGAGAGTCAGTG	.*.....	1	100	0.033395	Chr11	+	exon:HBB
ACAATAGGCAGAGAGAGTCAATT	**.....	2	5.21978	0	Chr18	-	intron:KIAA1328
CTAATAGGGTGAGAGAGTCAGCA	.....**.....	2	2.937332	0	Chr4	+	intergenic:RPS23P3-RNU6-699P
CTAAGAGGCAGAGAGAATCAACC	....*.....*...	2	3.890957	0	Chr7	+	intergenic:AC011288.2-AC005019.3
CTGATAGGCTGAGAGAGTCAACA	..*.....*.....	2	6.438065	0	Chr8	+	intergenic:NIPAL2-KB-1458E12.1
CTGGGAGGCAGAGAGAGTCAACC	..***.....	3	2.287424	0	Chr10	-	intergenic:RP11-810B23.1-MTND5P17
CTCTCAGGCAGAGAGAGTCAGTT	..***.....	3	2.287424	0	Chr2	+	intergenic:FAM84A/AC011897.1-AC068286.1

Yellow highlight identifies genes sequenced for off-target activity.

[†]*For this PAM, Off-targets 8-10 are not given by the software. Additionally, the top ranked predicted gene (HBB) was considered to be on-target and therefore skipped in our targeted sequencing assay.*

Table S4.1. Top 10 off-target predictions for SpRY20 (TGCCTATTAGTCTATTTTCCCAC) as ranked by CRISPOR.[1] PAM = NNN

Off-target Sequence	Mismatch Positions	Mismatch Count	MIT Off-target	CFD Off-target	Chromosome	Strand	Locus Description
TGTCTATTATTCTATTTTCCAGG	..*.....*.....	2	6.438065	0.55859375	Chr6	+	intergenic:NCOA7-AS1-NCOA7
TACCTATTAGTCAATTTTCCCTG	.*.....*.....	2	3.604411765	0.022823	Chr18	-	intergenic:SNRPGP2-RP11-61D1.2
TTCCTATTAGTCTATTTTTCAGT	.*.....*.	2	5.541666667	0.004737	Chr13	-	intergenic:RPL7P45-DAOA-AS1
TGCCTATTGGTCTATTTTCCCAC	.....*.....	1	61.1	0	Chr11	-	exon:HBB
TGCCTATTTGTCTATTTTCCTCT	.....*.....	1	61.1	0	Chr2	+	intergenic:ERMN-FAM133DP
AGCCTATTAGTCTATTTCCCTCT	*.....*..	2	3.448148148	0	Chr9	+	intron:RAPGEF1
TGCCTAGGAGTCTATTTTCCCCA	.....**.....	2	3.56510989	0	Chr6	-	intergenic:PRIM2-GAPDHP41
TGCCTTTTACTCTATTTTCCAAT	.....*...*.....	2	3.350283228	0	Chr4	+	intergenic:FAT4-RP11-399F2.2
TGCCCATTAGTCTATTTTCAACA	....*.....*	2	5.659285714	0	Chr10	+	intergenic:YWHAZP5-RP11-56I23.1
TGCCTATCAGTCTATTTTCATTC	.....*.....*	2	4.214361702	0	Chr1	-	intergenic:TMA16P2-MAST2

Yellow highlight identifies genes sequenced for off-target activity.

Table S4.2. Top 10 off-target predictions for SpRY20 (TGCCTATTAGTCTATTTTCCCAC) as ranked by CRISPOR.[1] PAM = NRN

Off-target Sequence	Mismatch Positions	Mismatch Count	MIT Off-target	CFD Off-target	Chromosome	Strand	Locus Description
TGTCTATTATTCTATTTTCCAGG	..*.....*.....	2	6.438065	0.55859375	Chr6	+	intergenic:NCOA7-AS1-NCOA7
GGTTTATTAGTCTATTTTCCTGG	*.**.....	3	2.287424	0.526086956	Chr7	+	intergenic:AC009541.1-hsa-mir-490
TGCCTATTTTTCTATTTTCCCAG	.....**.....	2	2.937332	0.126388889	Chr7	-	intergenic:AC006007.1-RN7SL207P
TACATATTAATCTATTTTCCAGC	.*.*.....*.....	3	2.461181	0.014778827	Chr21	+	intergenic:RNU1-139P-RPL37P4
ATCATATTAGTCTATTTTCCTGC	**.*.....	3	2.319902	0.011908559	Chr1	+	intergenic:RP11-445J9.1-RP11-113I24.1
GGTGTATTAGTCTATTTTCCTGC	*.**.....	3	2.287424	0.007306763	ChrX	-	intergenic:RPS6KA3-RN7SKP183
TTCCTATTAGTCTATTTTTCAGT	.*.....*.	2	5.541667	0.004737198	Chr13	-	intergenic:RPL7P45-DAOA-AS1
TGCCTATTGGTCTATTTTCCCAC	.....*.....	1	61.1	0	Chr11	-	exon:HBB
TGCCTTTTACTCTATTTTCCAAT	.....*...*.....	2	3.350283	0	Chr4	+	intergenic:FAT4-RP11-399F2.2
CTCCCATTAGTCTATTTTCCCAT	**..*.....	3	2.426564	0	Chr13	+	intergenic:SRGNP1-RNY3P3

Yellow highlight identifies genes sequenced for off-target activity.

Supplemental Figures

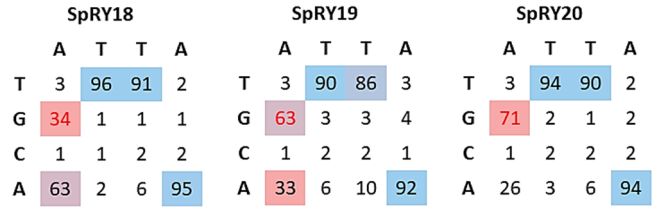


Figure S1. Output of EditR software showing efficiency of induced IVSI-107(A>G) base changes. *Editing efficiency (A>G base change) indicated in red numbers. gRNAs differ in the length of their PAM-distal, 5' sequence, as indicated in their names (gRNA for SpRY18 with 18 nt total length, up to 20 nt length for SpRY20).*

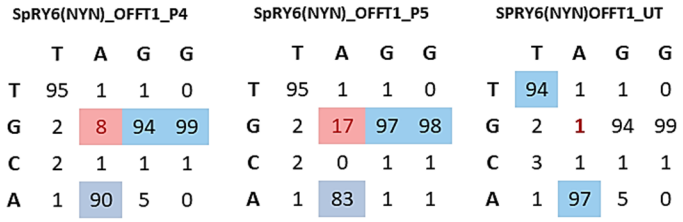


Figure S2. Output of EditR software showing off-target activity of SpRY6 editor in top predicted gene for NYN (KIAA1328, Table S3). *Editing efficiency (A>G base change) indicated in red numbers. Data shown for two patients (Patients 4 and 5, indicated as P4 and P5).*

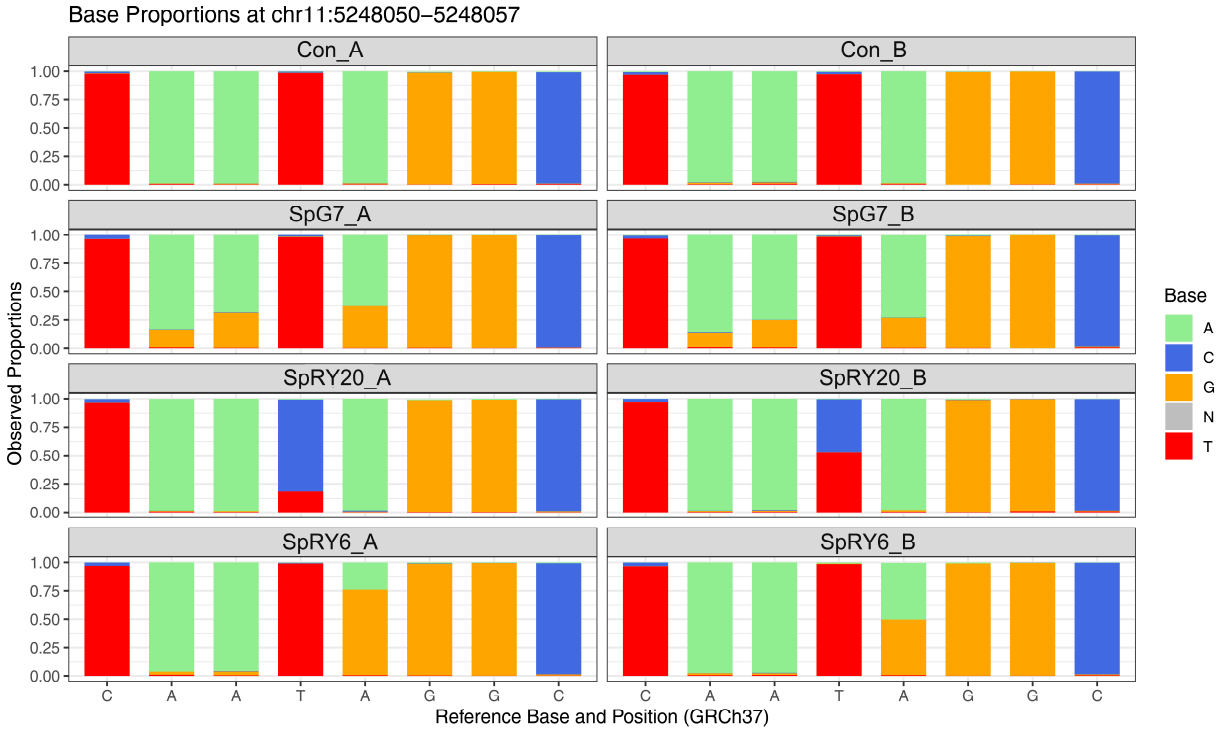


Figure S3. Plot illustrating the base composition of positions in the target region *indicating comparable editing distribution for two independent samples.*

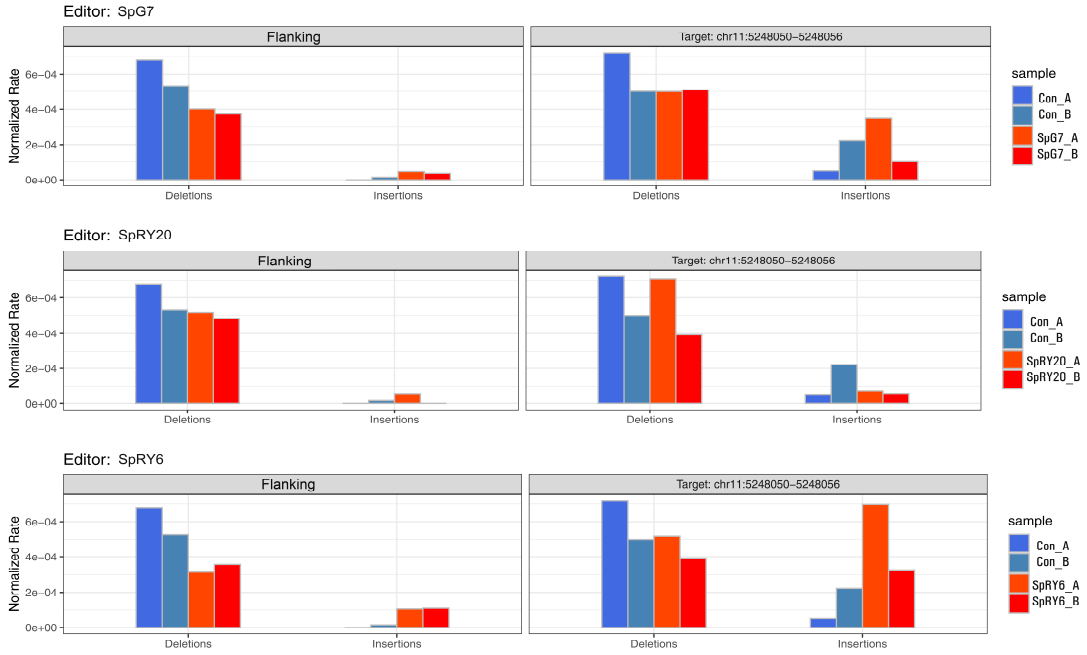


Figure S4. Bar plots illustrating normalized rate of indels *for reads overlapping target and flanking sequences of the region of interest (ROI) of two patients (A=Patient 5, B=Patient 4).*

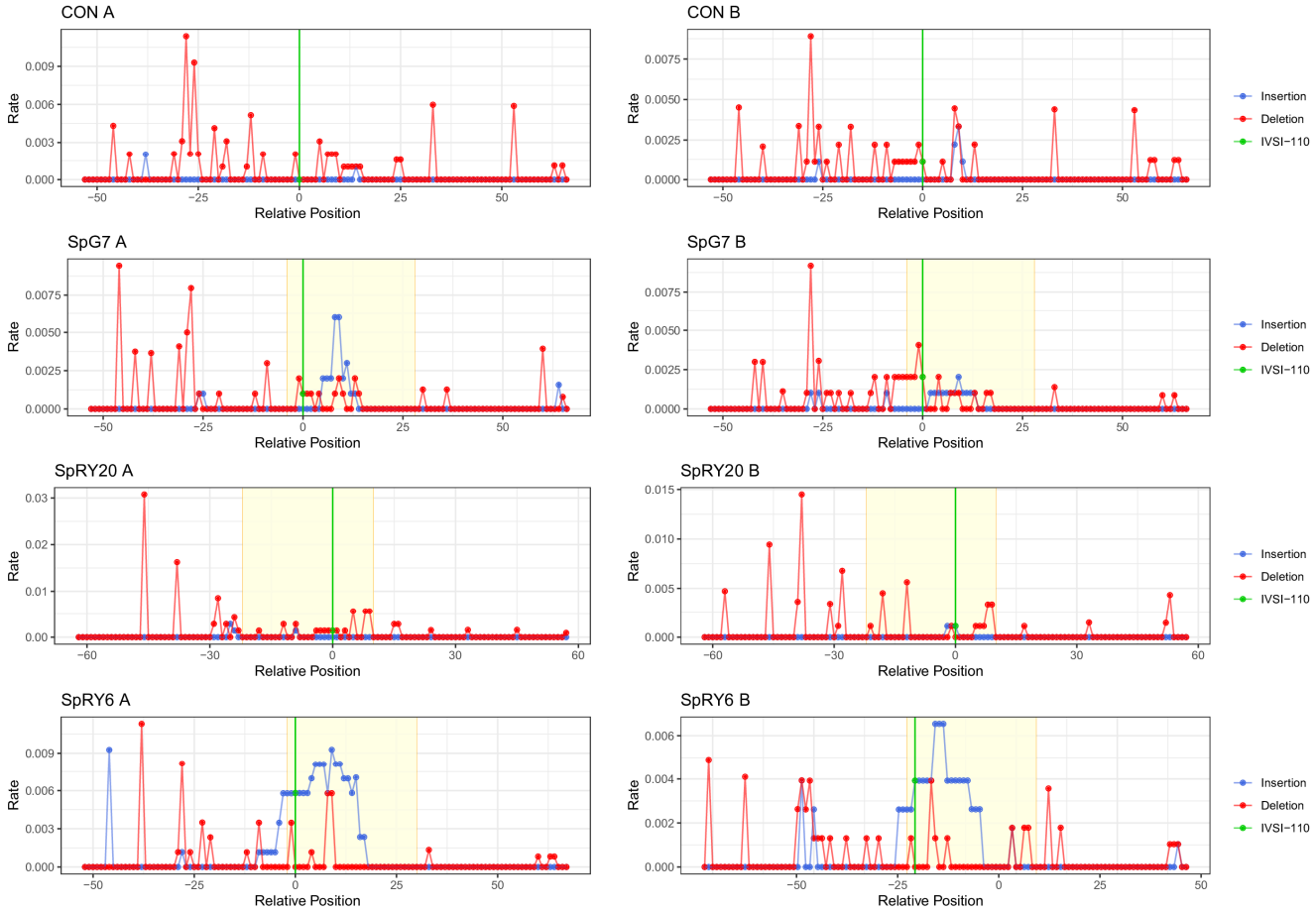


Figure S5. Illustration of insertion (blue) and deletion (red) rates *at each position of the targeted region (shaded yellow) with the IVS1-110 indicated in green.*

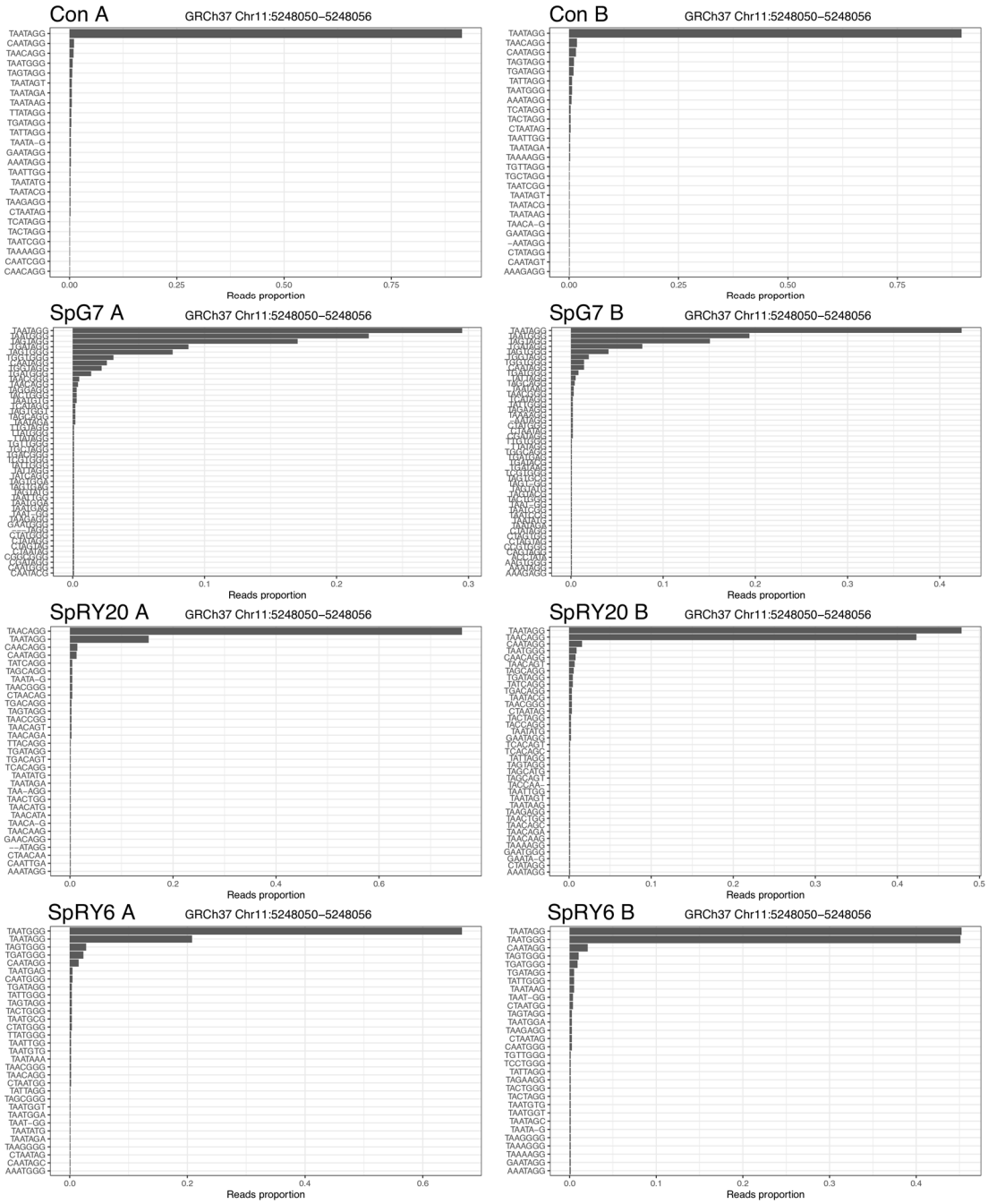


Figure S6. Plot showing relative proportion of 7-base oligonucleotides overlapping the target site as extracted from single reads. *(Chr11:5248050 = IVSI-110; Chr11:5248051 = IVSI-109; Chr11:5248052 = IVSI-108; Chr11:5248053 = IVSI-107; Chr11:5248054 = IVSI-106).*

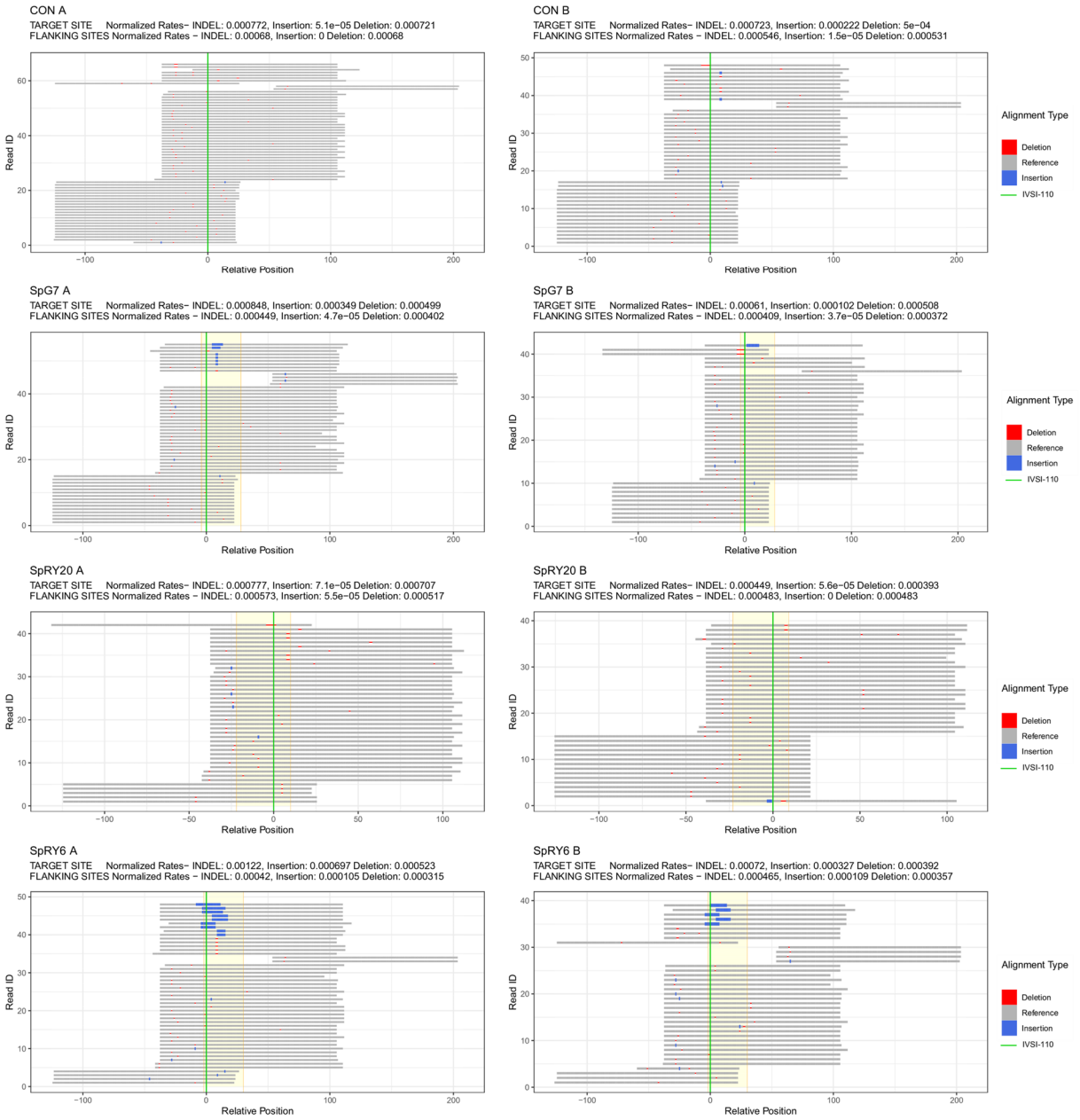


Figure S7. Representative alignment of reads overlapping the region of interest (ROI) ranked based on the size of the indel (large to small) in the target region. Grey bars indicate read segments matching the reference genome, and the yellow shaded region marks the binding loci of the gRNA. The green line indicates the IVSI-110(G>A) mutation site.

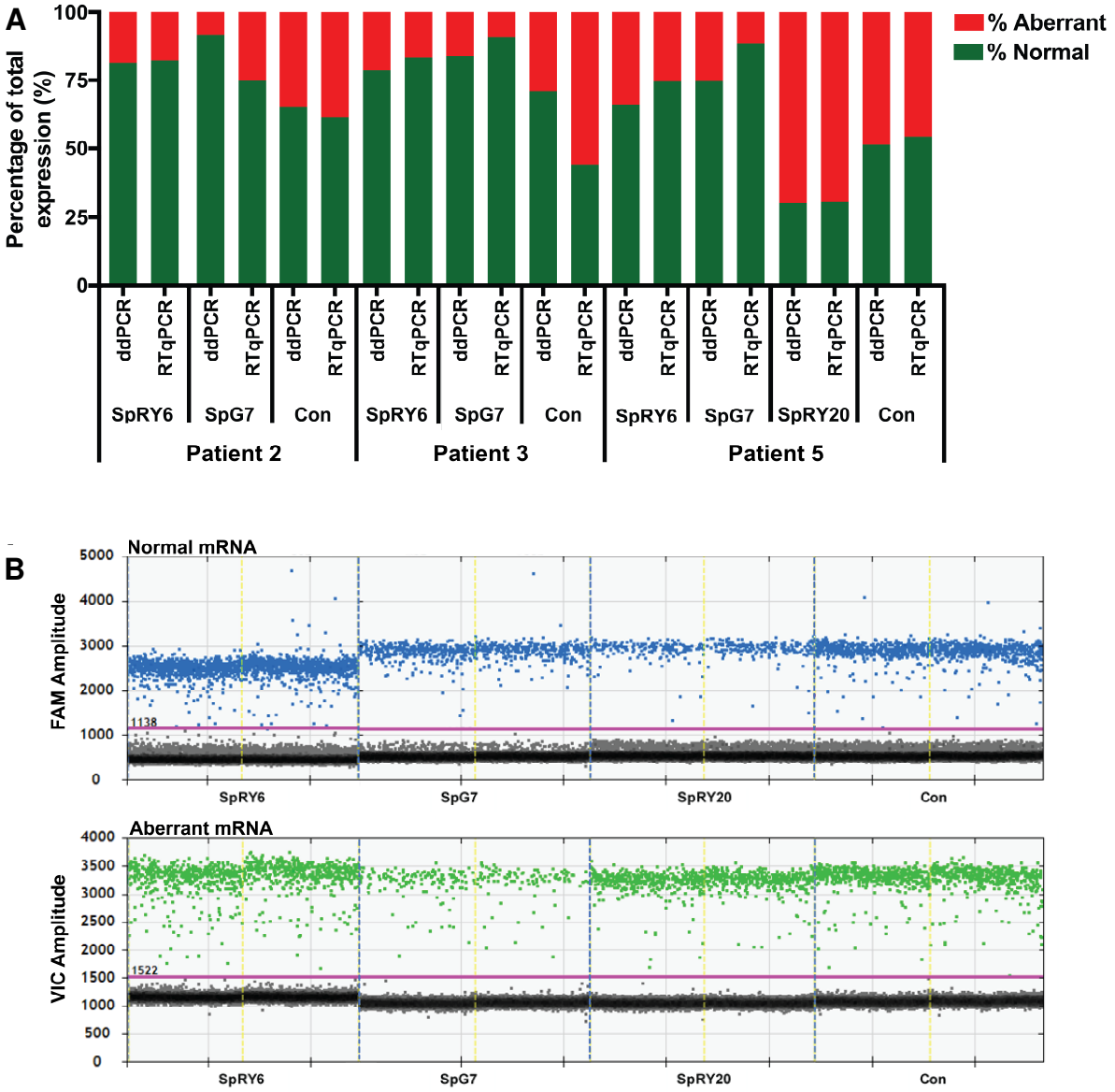
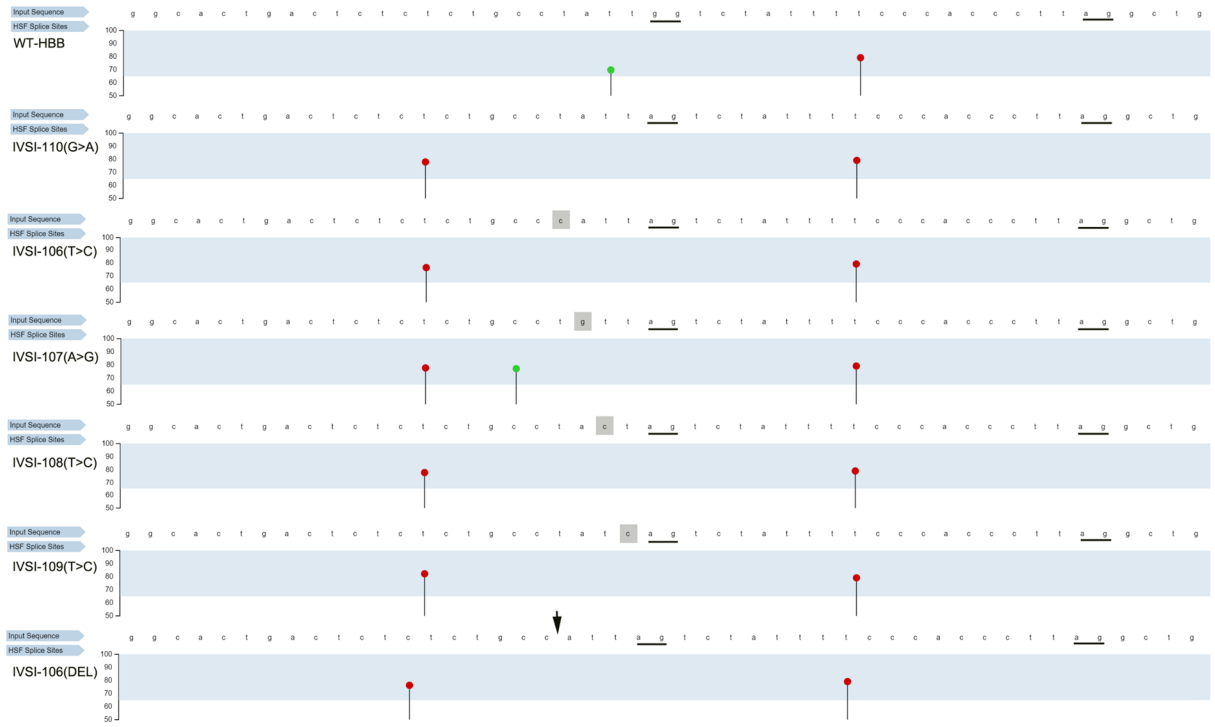
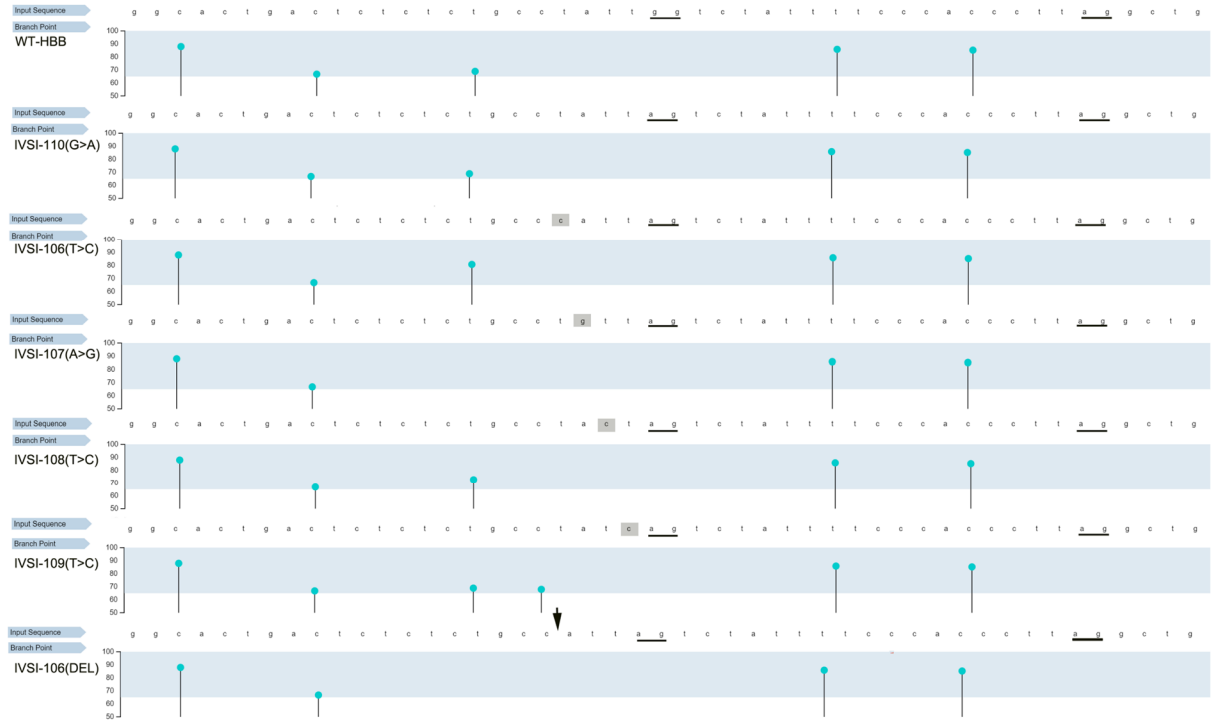


Figure S8. Assessment of splicing at the transcript level in patient-derived CD34$^+$ cells on day 11 of erythroid differentiation via multiplex RT-qPCR and droplet digital PCR. (A) Percentage of normal and aberrant HBB mRNAs of three patients (2, 3, and 5), as measured by two different quantitative technologies, RT-qPCR and (RT-)ddPCR. Of the samples shown, SpRY20 was only available for Patient 5. (B) FAM and VIC droplet amplitudes of Patient 5, as exported from ddPCR QuantaSoft software. The pink line represents the threshold that separates the positive and negative droplet clusters. The vertical yellow line separates two same-plate technical replicates for the same sample. Blue droplets are positive for the normal mRNA and green for the aberrant mRNA. Grey droplets do not contain the specific target. For ddPCR data not amplitude height but the ratio of droplets classified as positive and negative are the basis for quantification. Statistical analysis of the percentage of normal HBB mRNA in SpG7- (n=3) and SpRY6- (n=3) base edited samples relative to untreated control (UT; n=3) using parametric paired repeated measures one-way ANOVA with an assumption of sphericity and followed by Dunnett's multiple comparisons test. Statistical analysis showed a significant increase in the percentage of normal HBB mRNA measured via RTqPCR in SpG7 (84.73±8.67%; *p = 0.0242), SpRY6 (79.97±4.76%; *p = 0.0412) relative to UT (53.23±8.65%). Similarly, statistical analysis showed a significant increase in the percentage of normal HBB mRNA measured via RT-ddPCR in SpG7 (83.27±8.46%; ** p = 0.0040), SpRY6 (75.3±8.18%; *p = 0.0223) relative to UT (62.53±10.1%).

A


Input Sequence
HSF Splice Sites
WT-HBB
IVSI-110(G>A)
IVSI-106(T>C)
IVSI-107(A>G)
IVSI-108(T>C)
IVSI-109(T>C)
IVSI-106(DEL)

B


Input Sequence
Branch Point
WT-HBB
IVSI-110(G>A)
IVSI-106(T>C)
IVSI-107(A>G)
IVSI-108(T>C)
IVSI-109(T>C)
IVSI-106(DEL)

C

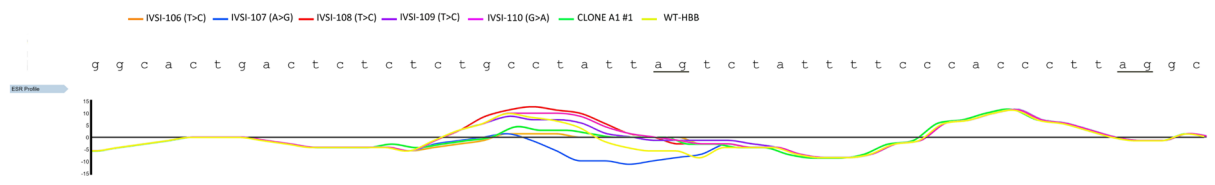

IVSI-106 (T>C)
IVSI-107 (A>G)
IVSI-108 (T>C)
IVSI-109 (T>C)
IVSI-110 (G>A)
CLONE A1 #1
WT-HBB
ESR Profile

Figure S9. *In silico* prediction of corrective effects on the splicing machinery. We employed the Human Splice Finder (HSF) software,[2] in order to elucidate the mechanism underlying the successful restoration of patient cell phenotype upon manipulation of the upstream intronic sequence using our ABEs for IVSI-106|108|109 (T>C) substitutions, as well as to explain the cause of deterioration of the IVSI-110 thalassemia phenotype upon IVSI-107(A>G) substitution, (A) *Representation of predicted donor and acceptor sites* across the sequence input, which is that of the HBB exon-intron junction, wherein the IVSI-110 mutation resides. Splice acceptor sites are indicated in red, donor sites in green. Base changes are highlighted in gray, acceptor sites are underlined in black, and an arrow is used to indicate the missing base in Clone A1 #2. The HSF software accurately predicted the gain of an acceptor site upon insertion of the IVSI-110(G>A) mutation. In the case of IVSI-107(A>G) (SpRY20 base editing), the formation of a novel donor site between the aberrant and normal splice acceptor sites is predicted in comparison to all other listed edits and the normal HBB locus. (B) *Representation of predicted branchpoint sites* across the sequence input, which is that of the HBB exon-intron junction wherein the IVSI-110 mutation resides. Base changes are highlighted in gray, acceptor sites are underlined in black, and an arrow is used to indicate the missing base in Clone A1 #2. The IVSI-107(A>G) base transition (SpRY20 base editing) and the IVSI-106 deletion (corresponding to the functionally corrected clone A1 #2 from our previous work)[3] are predicted to lose the same branch point site, indicating functional inutility of this site for normal HBB splicing. IVSI-109, on the other hand, appears to gain an extra branchpoint site. (C) *ESR profile of the HBB locus input sequence*, showing similar profiles for SpRY6 and corrective clone A1 #2. ESR = exonic splice enhancers ÷ exonic splicing silencers. Aberrant and normal splice acceptor sites are underlined. A change in trend is indicated when a base is edited, or deleted, as with our restorative clone A1 #2.[3] Compared to the curve for A1 #2 as reference for correction of splicing, a similar lowering to baseline is noticed upon IVSI-106 mutation (SpRY6), whereas the curve dips well below the baseline upon IVSI-107 mutation (SpRY20).

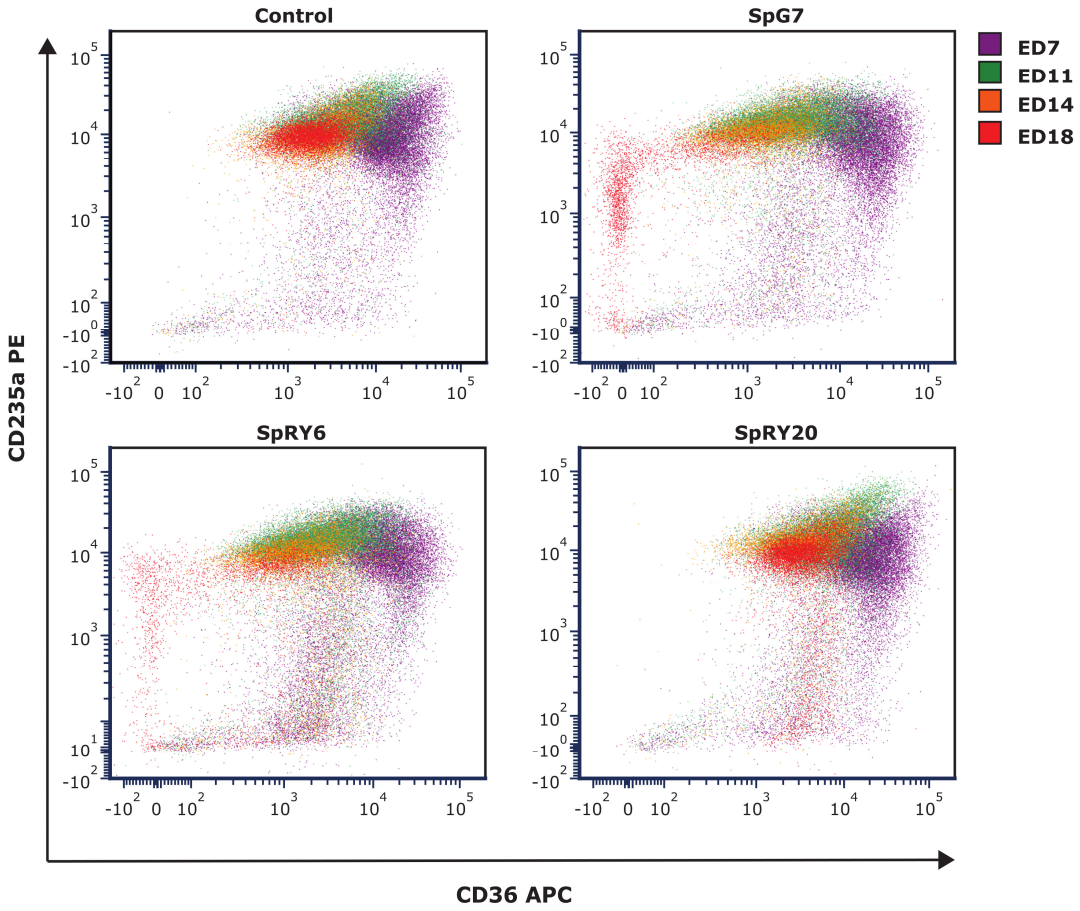


Figure S10. Representative flow cytometry color dot plots illustrating genome-edited and control cells stained for differentiation markers CD235a and CD36. Colors indicate stages of differentiation as different collection timepoints (ED7: purple; ED11: green; ED14: orange; ED18: red).

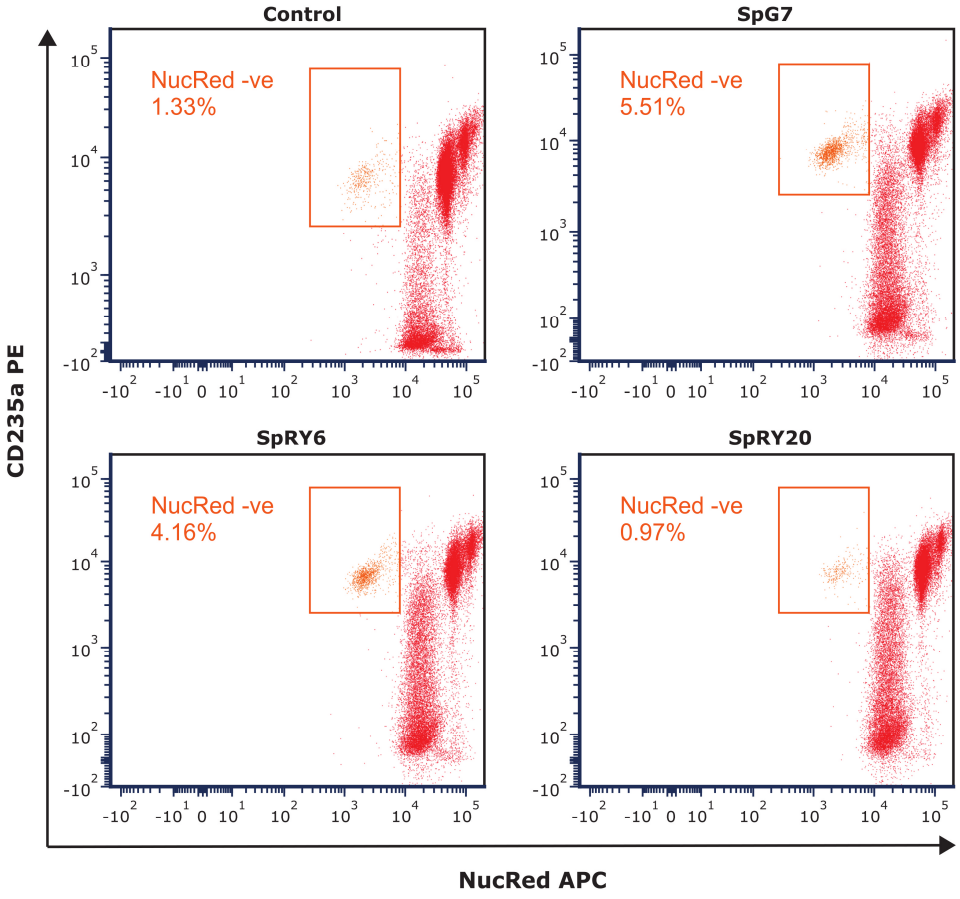


Figure S11. Representative flow cytometry color dot plots illustrating genome-edited and control cells stained for differentiation marker CD235a and nucleation marker NucRed on ED18. Percentage of enucleated erythroid cells were gated as CD235a^{+}/NucRed^{-} (in orange).

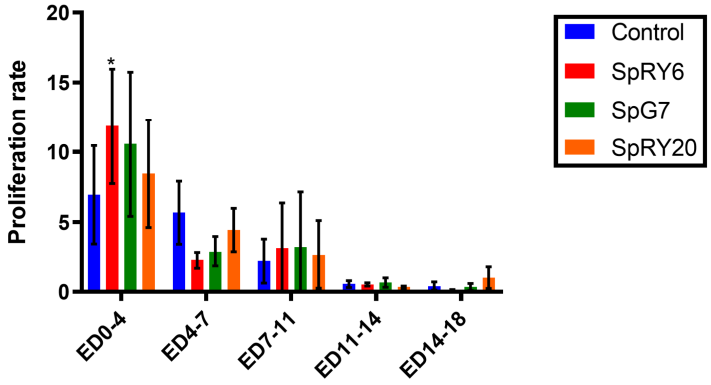


Figure S12. Proliferation rate of cells during erythroid differentiation at different time points (ED0-4, ED4-7, ED7-11, ED11-14 and ED14-18). Cell counts indicate a proliferation advantage in cells treated with corrective SpRY6 and SpG7, as compared to control, in the initial stage of differentiation. Contrarily, the final stage shows higher proliferation and corresponding delayed differentiation of the SpRY20-treated cells, in line with the treatment's worsening of the thalassemic disease phenotype. Groupwise comparisons with Control were performed in triplicate and significance obtained using two-way ANOVA with Dunnett's multiple comparison test. SpRY6 for ED0-4: *p = 0.0431.

Supplemental References

1. Concordet JP, Haeussler M. CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 2018;46(W1):W242–5.

2. Desmet FO, Hamroun D, Lalande M, Collod-Bëroud G, Claustres M, Béroud C. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):1–14.

3. Patsali P, Mussolino C, Ladas P, Floga A, Kolnagou A, Christou S, Sitarou M, Antoniou MN, Cathomen T, Lederer CW, et al. The Scope for Thalassemia Gene Therapy by Disruption of Aberrant Regulatory Elements. J Clin Med. 2019 Nov;8(11).