# Correction of RNA splicing defect in $\beta^{654}$-thalassemia mice using CRISPR/Cas9 gene-editing technology

Dan Lu,[1] Xiuli Gong,[1] Yudan Fang,[1] Xinbing Guo,[1] Yanwen Chen,[1] Fan Yang,[1] Guijun Zhao,[1] Qingwen Ma,[1] Yitao Zeng[1] and Fanyi Zeng[1,2,3]

[1]Shanghai Institute of Medical Genetics, Shanghai Children's Hospital, Shanghai Jiao Tong University; [2]Department of Histo-Embryology, Genetics and Developmental Biology, Shanghai Jiao Tong University School of Medicine and [3]Key Laboratory of Embryo Molecular Biology, Ministry of Health & Shanghai Key Laboratory of Embryo and Reproduction Engineering, Shanghai, China

**Correspondence:**

Fanyi Zeng
fzeng@vip.163.com

**Received:** December 22, 2020.
**Accepted**: September 23, 2021.
**Prepublished**: October 28, 2021.
https://doi.org/10.3324/haematol.2020.278238

# SUPPLEMENTARY APPENDIX

**Correction of RNA splicing defect in β$^{654}$-thalassemia mice using CRISPR/Cas9 gene-editing technology**

Dan Lu[1], Xiuli Gong[1], Yudan Fang[1], Xinbing Guo[1], Yanwen Chen[1], Fan Yang[1], Guijun Zhao[1], Qingwen Ma[1], Yitao Zeng[1], Fanyi Zeng[1,2,3]

1 Shanghai Institute of Medical Genetics, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai 200040, China

2 Department of Histo-Embryology, Genetics and Developmental Biology, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China

3 Key Laboratory of Embryo Molecular Biology, Ministry of Health & Shanghai Key Laboratory of Embryo and Reproduction Engineering, Shanghai 200040, China

## Online Supplementary Materials and Methods

### Construction of the CRISPR plasmids

Three sgRNAs targeting the DNA fragment containing both the IVS-2-654 C→T and IVS-2-579 were cloned into the pSpCas9 (BB)-2A-Puro (pX459) (Addgene plasmid #48139) backbone vector. The sequences of the guides are listed in Table S1. Other primers and sequences used throughout of this paper can also be found in Table S1.

### Off-target prediction analysis

The CRISPOR program (http://crispor.tefor.net)[1] was used to predict the potential off-target loci that may be affected by using the chosen CRISPR/Cas9 sgRNAs. The top ten potential gene loci (Table S2) were selected for analysis by PCR and targeted deep sequencing in 293T cells.

### *In vitro* transcription of sgRNAs

The DNA templates were prepared by PCR of pX459-sgRNA(G1/G2) plasmids as templet with specific primers (Table S1). The sgRNAs were *in vitro* transcribed with HiScribe™T7 Quick High Yield RNA Synthesis Kit (New England Biolabs) and purified with the MEGAclear kit (Life Technologies), according to manufacturer's instructions.

### Cell culture and transfection

293T cells were cultured using Gibco® DMEM, high glucose, supplemented

with 10% fetal bovine serum. 293T cells were seeded into 12-well plates to which a total of 1 μg of DNA plasmid pairs (pX459-sgRNA(G1/G2) plasmids, 0.5 μg for each plasmid) mixed with Lipofectamine 3000 (Invitrogen) were added according to the manufacturer's instructions. After 72 hours, genomic DNA was extracted from these cell lines.

**Sub-cloning and genotyping**

The PCR product was purified and ligated to pGEM-T vector and transformed to competent *E. coli* strain DH5α. After overnight culture at 37°C, randomly selected clones were sequenced by the Sanger method. The genotypes were determined by PCR of genomic DNA extracted from cells. ExTaq was activated at 95°C for 5 min, and PCR was performed for 34 cycles at 95°C for 30 sec, 58°C for 30 sec, and 72°C for 40 sec, with a final extension at 72°C for 7 min.

**Targeted deep sequencing**

DNA fragments containing the off-target sites were amplified from genomic DNA using KOD DNA polymerase (TOYOBP). Primers of targeted deep sequencing are listed in Table S1. Following amplification, the paired-end sequencing of PCR amplicons were gel-purified using QIAquick Gel Extraction Kit (Qiagen) and used for sequencing on Illumina Nextseq 500 (2×150) platform at Mingma, China. Data were analyzed using CRISPResso2.[2]

**Quantitative PCR**

Quantitative PCR **(**qPCR**)** was used to identify the $\beta^{654}$ mice. The primer pair Mhbb-QF1/R1 (Mhbb-QF1: 5′- TGGGCAGGCTGCTGGTTGTC -3′; Mhbb-QR1: 5′- CAAGTGATTCAGGCCATCGTT -3′), which can amplify a 152 bp product, was used to calculate the mouse β-major globin gene copy number in founder mice. The primer pair Mus TF-F/R (Mus TF-F: 5′- TGACTGCACCGGCAATTTC -3′; Mus TF-R: 5′- GGTACCCTCTGGAAGTTTAACGAA -3′), which can amplify a 92 bp product from the mouse transferrin gene, was used as an internal control. Each PCR amplification was performed in a 25 μL reaction volume containing 5 μL of template DNA (20 ng/μL), 1 μL of each primer, 12.5 μL of Power SYBR Green Mix, and 6.5 μL of distilled deionized water (dd$H_2O$) using the ABI7500 qPCR system.

**HPLC**

The samples for HPLC were prepared by collecting 50 μL whole blood and dissolving in 1mL pure water after filtration. 10 μL samples were loaded into ChromCore 300 C4 reversed-phase columns for polypeptides (300 Å, 5 μm, 4.6 mm × 250 mm). Individual globin chain levels were quantified on an Agilent 1260

instrument. A 40%-80% gradient mixture of 0.1% trifluoroacetic acid in water/acetonitrile was applied at a rate of 1 mL/min.

**Hematologic analysis**

Mouse peripheral blood samples were collected in heparinized microhematocrit tubes for hematologic analysis. 1-2 μL blood samples were prepared for blood smears stained with Wright-Giemsa (Baso, Zhuhai, China). The parameters examined include RBC count, hemoglobin (HGB) concentration, hematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and reticulocyte counts (RET) using a Hematology Analyzer (KX-21, Sysmex, Japan).

**Histopathology analysis**

Liver and spleen tissues from WT, $\beta^{654\text{-Ctrl}}$, and $\beta^{654\text{-E}}$ mice were embedded in paraffin, sliced to 4 μm sections, and stained with hematoxylin and eosin (Baso, Zhuhai, China). Bone marrow smears were stained with Wright-Giemsa (Baso, Zhuhai, China).

**Whole-genome sequencing and data analysis**

Genomic DNA was extracted from cells by using the DNeasy Blood and tissue kit (catalog number 69504, Qiagen) according to the manufacturer's instructions. WGS was performed at mean coverages of 50x by Illumina HiSeq X Ten. BWA (v0.7.12) was used to map qualified sequencing reads to the reference genome (mm10). The workflow of "Best Practice of GATK"[3] was used for sequence alignment to the reference genome (mm10) and variant (SNVs and indels) calling. The software involved includes BWA,[4] SAMtools,[5] and Genome Analysis Toolkit (GATK 4).[6] Structural variants (SV) were detected with Manta.[7] For analysis of sequence variations in $\beta^{654\text{-ER}}$ mice, the Cas-OFFinder[8] Web tool was used to identify candidate off-target sites with up to 3 mismatches.

**Statistical analysis**

All experimental data were analyzed using GraphPad Prism 5. A Student's t-test was used for intergroup comparisons. Probability ($P$) values $< 0.05$ was considered statistically significant.

**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-W245.
2. Clement K, Rees H, Canver MC, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol. 2019;37(3):224-226.
3. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1110):11.10.1-11.10.33.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-1760.
5. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-2079.
6. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-1303.
7. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32(8):1220-1222.
8. Bae S, Park J, Kim JS. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 2014;30(10):1473-1475.

## Online Supplementary Data

### Table S1. Primer sequences.

| Primer name | Primer sequences |
|---|---|
| G1 | taaattgtaactgatgtaag |
| G2 | tgccctgaaagaaagagatt |
| G3 | tccctaatctctttctttca |
| T7-G1 | TAATACGACTCACTATAGGtaaattgtaactgatgtaag |
| T7-G2 | TAATACGACTCACTATAGGtgccctgaaagaaagagatt |
| T7-G3 | TAATACGACTCACTATAGGtcctaatctctttctttttca |
| G-R | AAAAAAGCACCGACTCGGTG |
| Mhbb-QF1 | TGGGCAGGCTGCTGGTTGTC |
| Mhbb-QR1 | CAAGTGATTCAGGCCATCGTT |
| Mus TF-F | TGACTGCACCGGCAATTTC |
| Mus TF-R | GGTACCCTCTGGAAGTTTAACGAA |
| β-L | GACCAAATCACGGTAATTTTGC |
| β-R | GGCAGAATCCAGATGCTCAA |
| HBG-L | GAGGAGAAGTCTGCCGTTAC |
| HBG-R | AGCCACACCAGCCACCACTT |
| GAPDH-L | AGGCCGGTGCTGAGTATGTC |
| GAPDH-R | TGCCTGCTTCACCACCTTCT |
| deep-OT1F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTATTTTCTACATAGTGACCC |
| deep-OT1R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNAAACATTCCTTACTGGC |
| deep-OT2F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTTCACATGTACCCCTAAAAC |
| deep-OT2R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNCTTTGGCTATTTGGATACCTT |
| deep-OT3F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGAAGCCCATAATATCTAC |
| deep-OT3R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGAACATAGGCTTACTGTCAAA |
| deep-OT4F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCAGGTCCCACTCTGTCTA |
| deep-OT4R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNTGGAAACTTGGATGCCTTGACTC |
| deep-OT5F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTACTCCAAGGCAACAATTAGA |
| deep-OT5R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGAGGGAAGGGTGATGGTAA |
| deep-OT6F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTTGAAGGCATTGCTCA |
| deep-OT6R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNTCAGGAGGTCCCACATATAAG |
| deep-OT7F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCAGCAACCAGAGCAGTTAT |
| deep-OT7R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGCCCAATCCTTACATATTA |
| deep-OT8F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCTCTACATAAAAGGAATCAA |
| deep-OT8R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGGACAGCAAAAGGACATAAG |
| deep-OT9F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTCACCCAAAACCTATGA |
| deep-OT9R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNTTTCCACCTTTGCACTAAGAT |
| deep-OT10F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTAGTGAGGGCTGTTATCT |
| deep-OT10R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNAAAGAATGCCATAGTATGAAT |
| deep-OT11F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGAATTCATGCCAAA |
| deep-OT11R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNATTGCTCTGGAATTAGAAGTT |
| deep-OT12F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGATTAAGATGCCCATAAGA |
| deep-OT12R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNTAGCCAAGAAGGAGGGATATT |
| deep-OT13F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGGAAGAGAATTTGACTGA |
| deep-OT13R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGGTGTCAAAATGTATCAAACAGTA |
| deep-OT14F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTGGGTTAAATCTGCTTAGT |
| deep-OT14R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNCAATTGGCGTACTGAT |
| deep-OT15F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTTTGTTTATCTGGAACTGTC |
| deep-OT15R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNAGCAGGAGAGAAGTAACTC |
| deep-OT16F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTCCAGGAGGTTGAGG |
| deep-OT16R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNATGAGATGAGCCAGTGATAGA |
| deep-OT17F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATGAAGGCATTGATGGATATA |
| deep-OT17R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGGGGACTAGCAAGTCAG |
| deep-OT18F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCAGTGGTCAGGCCTTAAT |
| deep-OT18R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNCCACAGACATGGCGTAACCTT |
| deep-OT19F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACACTCCACACCTTTGTTCCA |
| deep-OT19R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNAGTCTAGGCAGGTAGTCAGCTG |
| deep-OT20F | ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATACGAGGATGACCATATGTA |
| deep-OT20R | ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNAAAACTTAAGAACACGCTA |
| P5-index1-F | AATGATACGGCGACCACCGAGATCTACACTGGTCAACACTCTTTCCCTACACGAC |
| P5-index2-F | AATGATACGGCGACCACCGAGATCTACACACATCGACACTCTTTCCCTACACGAC |
| P5-index3-F | AATGATACGGCGACCACCGAGATCTACACCGATGTACACTCTTTCCCTACACGAC |
| P5-index4-F | AATGATACGGCGACCACCGAGATCTACACTGACCAACACTCTTTCCCTACACGAC |
| P7-adapter1-R | CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTG |
| P7-adapter2-R | CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTG |
| P7-adapter3-R | CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTG |
| P7-adapter4-R | CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTG |

* Lowercase letters represent sgRNA sequences.

**Table S2. Summary of predicted off-target sites by CRISPOR program in this study.**

| Site number | Location | Gene name | Sequences |
|---|---|---|---|
| G1 | chr11_5225863-5225883:+ | *HBB* | TAAATTGTAACTGATGTAAG |
| OT1 | chr3:146771381-146771403:+ | *PLSCR5\|RP11-649A16.1* | CAAACTATAACTAATGTAAG |
| OT2 | chr7:85759049-85759071:+ | *LINC00972\|GRM3* | AAAATCATAAATGATGTAAG |
| OT3 | chr3:175578526-175578548:- | *NAALADL2\|RNU4-91P* | TAAATAATAAATGATATAAG |
| OT4 | chr4:104416220-104416242:+ | *RP11-729M20.1\|CXXC4* | AAAATAGTAACAAATGTAAG |
| OT5 | chr8:119080608-119080630:- | *RP11-278I4.2\|COLEC10* | TAAATAATAAATGATGTAAA |
| OT6 | chr3:190711411-190711433:- | *RP11-95L3.2\|GMNC* | TAAATACTAACTGATATAAG |
| OT7 | chr4:117904768-117904790:+ | *AC108056.1\|NDST3* | CAAACTCTAACTGATGTAAA |
| OT8 | chr5:136856638-136856660:- | *CTB-1I21.1\|RNA5SP193* | TTAACTGTAACAAATGTAAG |
| OT9 | chr3:19296956-19296978:- | *KCNH8\|MIR4791* | TATATTGTATCTGATATAAT |
| OT10 | chr11:128088384-128088406:- | *RN7SKP279\|RP11-702B10.2* | TAAATTAAAACTGTTATAAG |
| G2 | chr11_5225995-5225975:- | *HBB* | TGCCCTGAAAGAAAGAGATT |
| OT11 | chr3:174204131-174204153:- | *RN7SKP234-NLGN1* | TGCTCCAAAAGAAAAAGATT |
| OT12 | chr12:93077796-93077818:+ | *Y_RNA-RP11-202G11.2* | GGACCAGAAAGAAAGAAATT |
| OT13 | chr1:245015063-245015085:- | *EFCAB2* | CACCCAGAAAGAAAGAGATT |
| OT14 | chr14:52992879-52992901:- | *FERMT2-DDHD1* | AGCCTCAAAAGAAAGAGATT |
| OT15 | chr11:113667535-113667557:- | *DRD2-TMPRSS5* | AATCCTGAAAGAAAAAGATT |
| OT16 | chr5:149321993-149322015:+ | *AFAP1L1* | TGCACTCAAAAAAAGAAATT |
| OT17 | chr11:105272413-105272435:+ | *RP11-94P11.4-Metazoa_SRP* | TTCCCCAAAAGAAAGAAATT |
| OT18 | chr12:62532287-62532309:+ | *MON2* | TTCCCTAAATGAAAGAAATT |
| OT19 | chr4:170395548-170395570:+ | *RP11-789C1.2-RP11-322J23.1* | TCTCCTGAAAGAAAGAGATT |
| OT20 | chr13:42045543-42045565:- | *RP11-187A9.3-DGKH* | TTACCTGAATGAAAAAGATT |

The red letters in the sequences represent mismatched bases.

**Table S3. List of key parameters of mice gene-editing process.**

| Item | No. |
|---|---|
| No. of embryos microinjected | 142 |
| No. of embryos transferred | 123 |
| No. of the live-born mice | 56 |
| Birth Rate | 45.53% (56/123) |
| No. of mice tested at 19 days after birth | 37 |
| No. of $\beta^{654\text{-}E}$ mice | 12 |
| The survival rate of $\beta^{654\text{-}E}$ mice | 32.43% (12/37) |
| No. of $\beta^{654\text{-}ER}$ or $\beta^{654\text{-}ENR}$ mice | 10 |
| Gene editing rate | 83.33% (10/12) |

**Table S4. Hematologic analyses of offspring from $\beta^{654\text{-}ER}$ mice.**

| Group | N | RBC ($10^6$/μL) | HGB (g/L) | HCT (%) | MCV (fL) | MCH (pg) | MCHC (g/L) | RET (%) |
|---|---|---|---|---|---|---|---|---|
| **F1** | 21 | 9.1±1.1[a] | 132.3±20.2[a] | 44.9±6.5[a] | 49.2±1.7[a] | 14.5±0.8[a] | 294.6±10.9 [a] | 3.2±0.9[a] |
| **F2** | 13 | 9.7±1.2[a] | 144.8±17.8[a] | 48.5±5.8[a] | 50.3±1.5[a] | 15.0±0.4[a] | 298.7±5.1[a] | 3.0±0.7[a] |
| **WT** | 28 | 9.2±1.1[a] | 136.6±18.6[a] | 44.6±5.7[a] | 48.3±1.0[a] | 14.8±0.4[a] | 305.9±8.7[a] | 3.3±0.6[a] |
| **$\beta^{654\text{-}Ctrl}$** | 21 | 6.5±0.7 | 82.3±7.6 | 25.7±2.3 | 39.4±2.2 | 12.6±0.8 | 320.8±14.3 | 19.2±3.6 |

Values represent mean ± SD; N: number of mice tested; RBC: red blood cell; HGB: hemoglobin; HCT: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RET: reticulocyte. Statistically significant differences, WT or $\beta^{654\text{-}ER}$, compared to the $\beta^{654\text{-}Ctrl}$ group: [a] $P<0.01$.

**Table S5. Initial and filtered unique variant counts from whole-genome sequencing.**

| Group | Sample | Counts of SV | Unique SV | Counts of Indel | Unique Indel | Counts of SNV | Unique SNV |
|---|---|---|---|---|---|---|---|
| $\beta^{654\text{-}Ctrl}$ | $\beta^{654\text{-}Ctrl}$_1 | 7280 | - | 347386 | - | 1121679 | - |
| | $\beta^{654\text{-}Ctrl}$_2 | 7327 | - | 353393 | - | 1153867 | - |
| | $\beta^{654\text{-}Ctrl}$_3 | 10035 | - | 413460 | - | 1370339 | - |
| $\beta^{654\text{-}ER}$ | Sample_52 | 9252 | 5707 | 385124 | 213082 | 1070986 | 355264 |
| | Sample_53 | 6646 | 4506 | 290055 | 170524 | 640101 | 212543 |
| | Sample_59 | 7716 | 5396 | 386931 | 264064 | 1003649 | 562068 |
| | Sample_64 | 8641 | 5734 | 404374 | 253982 | 1083585 | 516555 |
| | Sample_87 | 16038 | 14187 | 290845 | 190791 | 661374 | 297755 |
| | Sample_90 | 16994 | 15724 | 264252 | 184865 | 499451 | 279619 |
| | Sample_92 | 15802 | 12247 | 422284 | 253057 | 1245989 | 558789 |

**Table S6. Thirty-five off-target sites for G1/G2 in the mouse genome predicted by Cas-OFFinder.**

| No. | sgRNA | DNA | Chromosome | Position | Direction | Mismatches |
|---|---|---|---|---|---|---|
| 1 | G1 | TAAATTGTAACTGATAAAATTGG | chr15 | 19512164 | - | 3 |
| 2 | G1 | TGATTTGTGACTGATGTAAGAGG | chr15 | 24488540 | + | 3 |
| 3 | G1 | AAAATTGTAACTGATGTTCGTGG | chr5 | 40501464 | + | 3 |
| 4 | G1 | GAGATTGTCACTGATGTAAGTGG | chr5 | 64689128 | - | 3 |
| 5 | G1 | CAAATTGTAGCTGATGTAACAGG | chr7 | 138239140 | + | 3 |
| 6 | G1 | TAAATTGTAGCTGATGTAAGTGG | chr2 | 52791351 | + | 1 |
| 7 | G1 | TAATTAGTAGCTGATGTAAGAGG | chr2 | 157363184 | - | 3 |
| 8 | G1 | TAAAATGTAAATGATGAAAGAGG | chr4 | 145683976 | - | 3 |
| 9 | G1 | TGATTTGTATCTGATGTAAGAGG | chr17 | 31247452 | - | 3 |
| 10 | G1 | TAAATTGTCACTTATGTTAGTGG | chrX | 103390000 | + | 3 |
| 11 | G1 | CAAATTGTATCTGATGTAAGAGG | chrX | 108632259 | + | 2 |
| 12 | G1 | TATAATGTAACTGATGTAAGTGG | chr6 | 114706092 | + | 2 |
| 13 | G1 | AACATTGTAACTGAGGTAAGGGG | chr11 | 10553586 | - | 3 |
| 14 | G1 | TAAAGTTTAACTAATGTAAGTGG | chr10 | 36097147 | + | 3 |
| 15 | G1 | TAAATTGAAACTAATGTAACTGG | chr10 | 40986901 | - | 3 |
| 16 | G1 | TAAAGCGTAACTGAAGTAAGCGG | chr10 | 123030322 | - | 3 |
| 17 | G1 | TAATTGCTAACTGATGTAAGAGG | chr13 | 30225663 | - | 3 |
| 18 | G1 | TAAATTGTAACTCAGTTAAGAGG | chr3 | 37931054 | + | 3 |
| 19 | G1 | TAATTAGTAATTGATGTAAGAGG | chr3 | 128640890 | - | 3 |
| 20 | G2 | TGCACTGAAAGAAAAAGATGTGG | chr15 | 92531701 | - | 3 |
| 21 | G2 | TGCACTCAAAGAAAGAGAATTGG | chr5 | 12423332 | - | 3 |
| 22 | G2 | TGCCCTGAAAGAAACTAATTAGG | chr1 | 26809785 | - | 3 |
| 23 | G2 | CGGCCTGAAAGAAAGAAATTGGG | chr1 | 86425826 | + | 3 |
| 24 | G2 | TTCAGTGAAAGAAAGAGATTGGG | chr1 | 128424331 | + | 3 |
| 25 | G2 | TGCCCTGAACAAAAGCGATTAGG | chr4 | 58124943 | - | 3 |
| 26 | G2 | TACCTTGAAGGAAAGAGATTAGG | chr17 | 4392949 | - | 3 |
| 27 | G2 | TTACCTGAAAGAAAGAGACTGGG | chr14 | 32074650 | - | 3 |
| 28 | G2 | TGCCCAAAAAGAAAGAGAATGGG | chr14 | 61447679 | + | 3 |
| 29 | G2 | TGCTTTGAAAGAAAGAAATTTGG | chr6 | 107501455 | + | 3 |
| 30 | G2 | TGTCTTGAAAGAAAGAGAGTGGG | chr11 | 79258171 | - | 3 |
| 31 | G2 | TCCCGTGAAAGAAAGAAATTTGG | chr10 | 42844492 | - | 3 |
| 32 | G2 | AGCCCTGAAAGACATAGATTCGG | chr10 | 71530156 | + | 3 |
| 33 | G2 | TGACCTGAAAAAAAGAAATTTGG | chr10 | 113588503 | - | 3 |
| 34 | G2 | TGCCTTGAAAGAAAGAGAGGGGG | chr18 | 35302981 | - | 3 |
| 35 | G2 | TGGCCTGAAAGAAAGAGAACAGG | chr3 | 75965220 | + | 3 |

The red letters in the sequences represent mismatched bases.

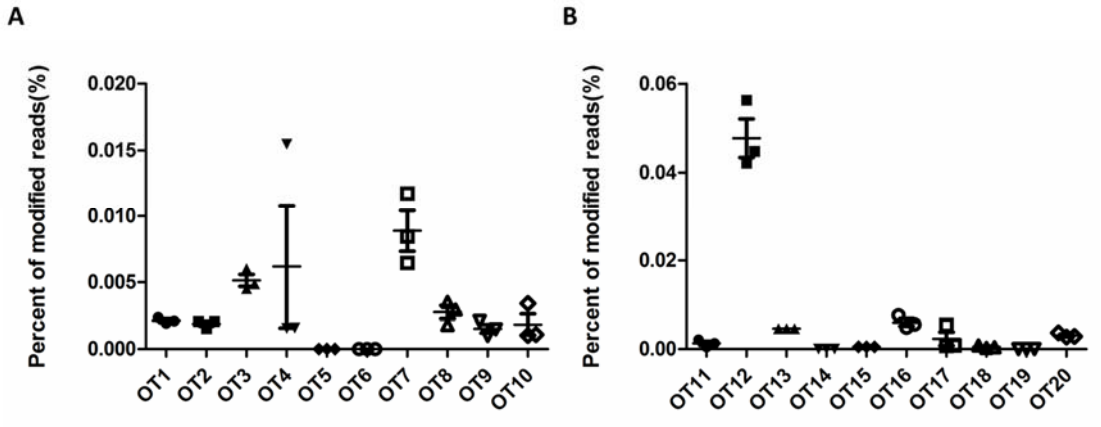


**Figure S1. Deep sequencing analysis of potential off-target sites in 293T cells edited by sgRNA G1+G2.** Predicted off-target sites OT1 to OT20 are listed in Table S2. (A) Deep sequencing analysis of top 10 off-target sites for sgRNA G1. (B) Deep sequencing analysis of top 10 off-target sites for sgRNA G2.

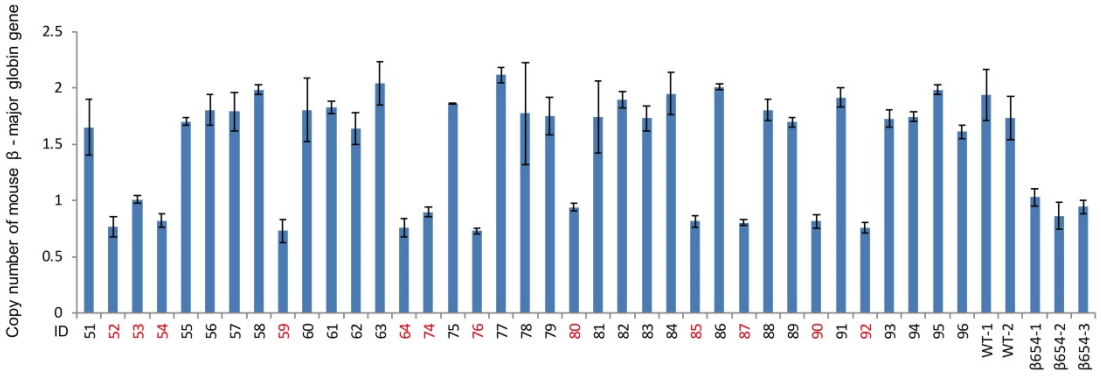


**Figure S2. Copy number analysis for mouse β-major globin gene by qPCR.** WT-1, WT-2: wild-type mice; $\beta^{654}$-1~3: $\beta^{654\text{-Ctrl}}$ mice. Copy number = 2, correlates to wild-type mouse. Copy number = 1, represents one mouse β-major globin gene, and correlates to $\beta^{654}$ mice. The mouse ID numbers in red represent $\beta^{654\text{-E}}$ mice subjected to gene editing.

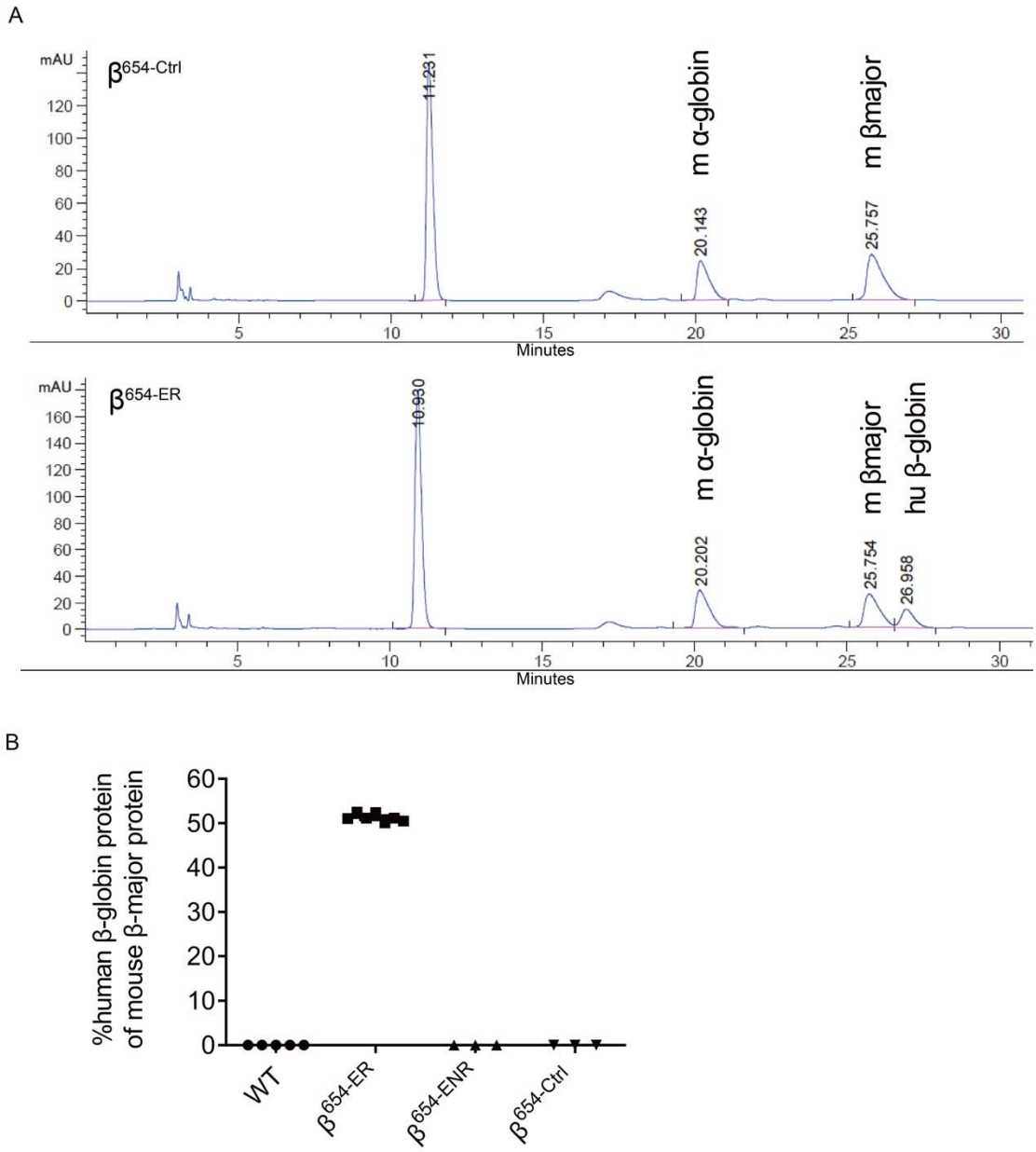


**Figure S3. HPLC analysis of globin chains in RBCs.** (A) Representative chromatograms of mice. Upper panel: Data from representative chromatograms of $\beta^{654\text{-Ctrl}}$ mice. The peaks of mouse α-globin (m α-globin) and mouse βmajor globin (m βmajor) are indicated in the upper panel. Lower panel: data from representative chromatograms of effective gene-edited $\beta^{654\text{-ER}}$ mice. The peaks of human β-globin (hu β-globin), mouse α-globin (m α-globin), and mouse β major globin (m βmajor) are indicated in the lower panel. (B) Summary of human β-globin levels to mouse β major globin chains in individual mice. WT, n=5; $\beta^{654\text{-ER}}$, n=7 (mouse IDs: 52, 53, 64, 92, 87, 90 and 59); $\beta^{654\text{-ENR}}$, n=3 (mouse IDs: 54, 74, and 76).