

GENE THERAPY FOR TREATING BETA-HEMOGLOBINOPATHIES

BACKGROUND

[0001] Beta-hemoglobinopathies, including sickle cell disease (SCD) and β -thalassemia, can be caused by genetic mutations in the β -globin gene (HBB). In addition to hematopoietic stem cell transplantation, gene therapy is one of the most promising treatments for these diseases. Gene therapy is a therapeutic strategy of human hereditary diseases, through gene addition or genome editing to treat hereditary diseases.

[0002] Previously, it was found that a few patients with naturally existing mutations in the HBB gene cluster or related genes maintain high γ -globin expression from their childhood to adulthood and do not show serious symptoms of anemia. This suggests that a therapeutic strategy for β -hemoglobinopathies is through reactivation of the expression of γ -globin. Currently, there are two common strategies to reactivate the expression of γ -globin. The first is to knock down a gene (*e.g.*, BCL11A) that suppresses the expression of γ -globin gene (HBG1/2) and the second is to disrupt the binding sequences of transcription factors at the promoters of the HBG1/2 genes. Different methods have been tested to implement the aforementioned therapeutic strategies, such as CRISPR/Cas (Clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein)-induced nucleotide insertions/deletions (indels) or base editor (BE)-induced point mutations.

[0003] CRISPR/Cas system has been the most prevalent genomic editing tool because of its convenience and high editing efficiency in living organisms. Directed by a guide RNA, the Cas nuclease can generate DNA double strand breaks (DSBs) at the targeted genomic sites in various cells (both cell lines and cells from living organisms). These DSBs are then repaired by the endogenous DNA repair system, which could be utilized to perform desired genome editing.

[0004] In general, two major DNA repair pathways can be activated by DSBs, non-homologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ can introduce random indels in the genomic DNA region around the DSBs, thereby leading to open reading frame (ORF) shift and ultimately gene inactivation. In contrast, when HDR is triggered, the genomic DNA sequence at target site could be replaced by the sequence of the exogenous donor DNA through a homologous recombination mechanism, which can be used to induce base substitutions. However, the practical efficiency of HDR-mediated base substitution is

low (normally < 5%) because the occurrence of homologous recombination is both cell type-specific and cell cycle-dependent and NHEJ is triggered more frequently than HDR.

[0005] Base editor (BE), which combines the CRISPR/Cas system with the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) cytosine deaminase family members, was recently developed to induce base substitutions with high efficiency. Through the fusion with Cas9 nickase (nCas9) or catalytically dead Cpf1 (dCpf1, also known as dCas12a), the cytosine (C) deaminase activity of APOBEC/AID family members can be purposely directed to the target genomic sites to induce C to Thymine (T) substitutions.

[0006] The safety and efficiency of gene editing are of great importance in gene therapy. Previous studies have reported that the DSBs induced by Cas9 nuclease can activate a p53-mediated DNA damage response pathway and then lead to cell death. Moreover, APOBEC/AID family members can trigger C-to-T base substitutions in single-stranded DNA (ssDNA) regions, which are formed randomly during various cellular processes including DNA replication, repair and transcription. Thus, the specificity of previous base editing systems is compromised, limiting the applications of BEs for therapeutic purposes.

SUMMARY

[0007] The instant disclosure, in some embodiments, describes improved gene therapy technologies useful for increasing the production of the γ -globin gene, which is useful for treating various hematological diseases, in particular inherited ones, such as beta-thalassemia and sickle cell anemia. Using a newly designed base editor, referred to as a transformer Base Editor (tBE), the present technology employs specifically designed guide RNA sequences to target the BCL11A erythroid enhancer or the C-terminal three tandem C2H2 zinc fingers (Znf4~6) for inactivation, or to the γ -globin promoter for activation. Such a base editor has improved efficiency and specificity, as demonstrated in the experimental examples.

[0008] In accordance with one embodiment of the present disclosure, provided is a base editing system, or one or more polynucleotides encoding the base editing system. In some embodiments, the base editing system comprises a CRISPR-associated (Cas) protein, a nucleobase deaminase, a single-guide RNA (sgRNA), and a helper single-guide RNA (hsgRNA). The sgRNA and the hsgRNA target sites at the BCL11A erythroid enhancer. In some embodiments, the sgRNA and the hsgRNA target sites at the BCL11A-binding motif in the γ -globin promoter. In some embodiments, the sgRNA and the hsgRNA target sites at the

ZBTB7A-binding motif in the γ -globin promoter. In some embodiments, the sgRNA and the hsgRNA target sites at a GATA1-half-E-box motif of BCL11A. In some embodiments, the sgRNA and the hsgRNA target sites at KLF1-binding motifs of BCL11A. In some embodiments, the sgRNA and the hsgRNA target sites at a GATA1-binding motif of NFIX. In some embodiments, the sgRNA and the hsgRNA target sites at the coding sequences of Znf4~6 in BCL11A. Example sgRNA and hsgRNA sequences are provided in **Tables 1-11**.

[0009] In one embodiment, provided is a method for promoting production of γ -globin in a human cell, comprising introducing into the cell a CRISPR-associated (Cas) protein, a nucleobase deaminase, a single-guide RNA (sgRNA), and a helper single-guide RNA (hsgRNA), wherein (a) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:1-10, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-28; (b) the sgRNA comprises the nucleic acid sequence of SEQ ID NO:29-30, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:31-36, (c) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:37-54, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:63-116, (d) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:117-122, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138, or (e) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:139-150, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-190, (f) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:353-430, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:431-628, In some embodiments, the Cas protein, the nucleobase deaminase, the sgRNA, and the hsgRNA are preferably introduced into the cell by one or more encoding polynucleotides.

[0010] In some embodiments, (a) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:1-10, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-28; (b) the sgRNA comprises the nucleic acid sequence of SEQ ID NO:29-30, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:31-36; (c) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:37-54, and the

hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:63-116; (d) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:118-122, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138; or (e) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:139-150, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-190. In some embodiments, the sgRNA comprises the nucleic acid sequence of SEQ ID NO:4, and the hsgRNA comprises the nucleic acid of SEQ ID NO:11. In some embodiments, the sgRNA comprises the nucleic acid sequence of SEQ ID NO:4, and the hsgRNA comprises the nucleic acid of SEQ ID NO:12.

[0011] In some embodiments, the sgRNA comprises the nucleic acid sequence of SEQ ID NO:30, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:31-36, preferably SEQ ID NO:33 or 34. In some embodiments, both sets of sgRNA/hsgRNA are included.

[0012] In some embodiments, the nucleobase deaminase is a cytidine deaminase. Non-limiting examples include APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F (A3F), APOBEC3G (A3G), APOBEC3H (A3H), APOBEC1 (A1), APOBEC3 (A3), APOBEC2 (A2), APOBEC4 (A4) and AICDA (AID).

[0013] In some embodiments, the base editing system further comprises a nucleobase deaminase inhibitor, fused to the nucleobase deaminase, via a protease cleavage site. In some embodiments, the nucleobase deaminase inhibitor is an inhibitory domain of a nucleobase deaminase. In some embodiments, the nucleobase deaminase inhibitor is an inhibitory domain of a cytidine deaminase. Non-limiting examples include SEQ ID NO: 192-193.

[0014] In some embodiments, the base editing system further comprises a protease that is capable of cleaving at the protease cleavage site. In some embodiments, the protease is selected from the group consisting of TuMV protease, PPV protease, PVY protease, ZIKV protease and WNV protease. In some embodiments, the protease cleaves the cleavage site only when the base editor is at the target site determined by the guide RNAs.

[0015] In some embodiments, the Cas protein is selected from the group consisting of SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9,

NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b, and RanCas13b.

[0016] In some embodiments, the Cas protein is catalytically impaired, such as nCas9 and dCpf1.

[0017] Also provided, in one embodiment, is a method of using the base editors, or one or more polynucleotides that encode the base editors, to promote production of γ -globin in a human cell, which may be an erythroid cell, a hematopoietic stem cell, or a stem cell, among others. In some embodiments, the method is carried out *ex vivo* or *in vivo* in a patient. In some embodiments, the patient suffers from β -thalassemia, sickle cell anemia, Haemoglobin C, or Haemoglobin E.

[0018] Yet another embodiment provides base editors that incorporate a cytidine deaminase inhibitor. Examples include hA3F-CDA1 and its analogs. In some embodiments, a fusion protein is provided, comprising: a first fragment comprising a cytidine deaminase or a catalytic domain thereof, a second fragment comprising a cytidine deaminase inhibitor comprising an amino acid sequence selected from the group consisting of SEQ ID NO:192, and 265-309 and sequences having at least 85% sequence identity to any of SEQ ID NO:192, and 265-309, and a protease cleavage site between the first fragment and the second fragment. Also provided are methods of using such fusion proteins for base editing and treatments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] FIG. 1: Editing efficiencies induced by tBE with the pairs of sgRNA-BCL11A-2 and its hsgRNAs targeting BCL11A erythroid enhancer region. FIG. 1A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-BCL11A-2, different hsgRNA-BCL11As and cytidine deaminase complex of tBE. FIG. 1B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA and YE1-BE4max with indicated sgRNA targeting BCL11A erythroid enhancer region.

[0020] FIG. 2: Editing efficiencies induced by tBE with the pairs of sgRNA-BCL11A-3 and its hsgRNAs targeting BCL11A erythroid enhancer region. FIG. 2A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid

expresses SpD10A nickase and another expresses sgRNA-BCL11A-3, different hsgRNA-BCL11As and cytidine deaminase complex of tBE. FIG. 2B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA and YE1-BE4max with indicated sgRNA targeting BCL11A erythroid enhancer region.

[0021] FIG. 3: Editing efficiencies induced by tBE with the pairs of sgRNA-BCL11A-4 and its hsgRNAs targeting BCL11A erythroid enhancer region. FIG. 3A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-BCL11A-4, different hsgRNA-BCL11As and cytidine deaminase complex of tBE. FIG. 3B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA and YE1-BE4max with indicated sgRNA targeting BCL11A erythroid enhancer region.

[0022] FIG. 4: Editing efficiencies induced by tBE with the pairs of sgRNA-BCL11A-5 and its hsgRNAs targeting BCL11A erythroid enhancer region. FIG. 4A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-BCL11A-5, different hsgRNA-BCL11As and cytidine deaminase complex of tBE. FIG. 4B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA and YE1-BE4max with indicated sgRNA targeting BCL11A erythroid enhancer region.

[0023] FIG. 5: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG and its hsgRNAs targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2). FIG. 5A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-HBG, hsgRNA-HBG and cytidine deaminase complex of tBE. FIG. 5B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA and YE1-BE4max with indicated sgRNA targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2).

[0024] FIG. 6: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-1 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-3/hsgRNA-BCL11A-1 targeting BCL11A erythroid enhancer region simultaneously. FIG. 6A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and

BCL11A erythroid enhancer regions simultaneously. FIG. 6B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0025] FIG. 7: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-2 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-3/hsgRNA-BCL11A-1 targeting BCL11A erythroid enhancer region simultaneously. FIG. 7A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 7B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0026] FIG. 8: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-3 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-3/hsgRNA-BCL11A-1 targeting BCL11A erythroid enhancer region simultaneously. FIG. 8A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoter and BCL11A erythroid enhancer regions simultaneously. FIG. 8B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0027] FIG. 9: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-1 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-3/hsgRNA-BCL11A-2 targeting BCL11A erythroid enhancer region simultaneously. FIG. 9A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoter and BCL11A erythroid enhancer regions simultaneously. FIG. 9B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of

sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0028] FIG. 10: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-2 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-3/hsgRNA-BCL11A-2 targeting BCL11A erythroid enhancer region simultaneously. FIG. 10A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 10B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0029] FIG. 11: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-3 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-3/hsgRNA-BCL11A-2 targeting BCL11A erythroid enhancer region simultaneously. FIG. 11A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 11B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0030] FIG. 12: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-1 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-4/hsgRNA-BCL11A-1 targeting BCL11A erythroid enhancer region simultaneously. FIG. 12A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 12B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters

regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0031] FIG. 13: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-2 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-4/hsgRNA-BCL11A-1 targeting BCL11A erythroid enhancer region simultaneously. FIG. 13A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 13B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0032] FIG. 14: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-3 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-4/hsgRNA-BCL11A-1 targeting BCL11A erythroid enhancer region simultaneously. FIG. 14A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 14B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0033] FIG. 15: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-1 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-4/hsgRNA-BCL11A-2 targeting BCL11A erythroid enhancer region simultaneously. FIG. 15A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 15B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0034] FIG. 16: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-2 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-4/hsgRNA-BCL11A-2 targeting BCL11A erythroid enhancer region simultaneously. FIG. 16A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 16B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0035] FIG. 17: Editing efficiencies induced by tBE with the pairs of sgRNA-HBG/hsgRNA-HBG-3 targeting the core binding sites of transcription factors in the promoter regions of γ -globin gene (HBG1/2), and the pairs of sgRNA-BCL11A-4/hsgRNA-BCL11A-2 targeting BCL11A erythroid enhancer region simultaneously. FIG. 17A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system to edit HBG1/2 promoters and BCL11A erythroid enhancer regions simultaneously. FIG. 17B: Sanger sequencing results show the base editing efficiencies induced by tBE with two indicated pairs of sgRNA/hsgRNA targeting the core binding sites of transcription factors in the promoters regions of γ -globin gene (HBG1/2) and the BCL11A erythroid enhancer region simultaneously.

[0036] FIG. 18: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-1-1 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region. FIG. 18A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-1-1, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 18B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region.

[0037] FIG. 19: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-1-2 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region. FIG. 19A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A

nickase and another expresses sgRNA-KLF1-1-2, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 19B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region.

[0038] FIG. 20: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-1-3 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region. FIG. 20A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-1-3, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 20B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region.

[0039] FIG. 21: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-2-1 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region. FIG. 21A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-2-1, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 21B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region.

[0040] FIG. 22: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-2-2 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region. FIG. 22A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-2-2, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 22B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locate in +55 kb DHS of BCL11A erythroid enhancer region.

[0041] FIG. 23: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-2-3 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region. FIG. 23A: Schematic diagram illustrating the

co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-2-3, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 23B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region.

[0042] FIG. 24: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-2-4 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region. FIG. 24A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-2-4, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 24B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in +55 kb DHS of BCL11A erythroid enhancer region.

[0043] FIG. 25: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-3-1 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in 1Mb upstream of BCL11A. FIG. 25A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-3-1, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 25B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in 1Mb upstream of BCL11A.

[0044] FIG. 26: Editing efficiencies induced by tBE with the pairs of sgRNA-KLF1-3-2 and its hsgRNAs targeting one of the three KLF1 binding motifs of BCL11A locates in 1Mb upstream of BCL11A. FIG. 26A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-KLF1-3-2, hsgRNA-KLF1 and cytidine deaminase complex of tBE. FIG. 26B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting one of the three KLF1 binding motifs of BCL11A locates in 1Mb upstream of BCL11A.

[0045] FIG. 27: Editing efficiencies induced by tBE with the pairs of sgRNA-GATA1-1 and its hsgRNAs targeting the GATA1-binding motif located in intron 4 of the NFIX gene. FIG.

27A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-GATA1-1, hsgRNA-GATA1 and cytidine deaminase complex of tBE. FIG. 27B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting GATA1-binding motif located in intron 4 of the NFIX gene.

[0046] FIG. 28: Editing efficiencies induced by tBE with the pairs of sgRNA-GATA1-2 and its hsgRNAs targeting the GATA1-binding motif located in intron 4 of the NFIX gene. FIG. 28A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-GATA1-2, hsgRNA-GATA1 and cytidine deaminase complex of tBE. FIG. 28B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting GATA1-binding motif located in intron 4 of the NFIX gene.

[0047] FIG. 29: Editing efficiencies induced by tBE with the pairs of sgRNA-ZBTB7A-1-1 and its hsgRNAs targeting one of the two ZBTB7A-binding motifs located in HBG1 enhancer. FIG. 29A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-ZBTB7A-1-1, hsgRNA-ZBTB7A and cytidine deaminase complex of tBE. FIG. 29B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the ZBTB7A-binding motif located in HBG1 enhancer.

[0048] FIG. 30: Editing efficiencies induced by tBE with the pairs of sgRNA-ZBTB7A-1-2 and its hsgRNAs targeting one of the two ZBTB7A-binding motifs located in HBG1 enhancer. FIG. 30A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-ZBTB7A-1-2, hsgRNA-ZBTB7A and cytidine deaminase complex of tBE. FIG. 30B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the ZBTB7A-binding motif located in HBG1 enhancer.

[0049] FIG. 31: Editing efficiencies induced by tBE with the pairs of sgRNA-ZBTB7A-1-3 and its hsgRNAs targeting one of the two ZBTB7A-binding motifs located in HBG1 enhancer. FIG. 31A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-ZBTB7A-1-3, hsgRNA-ZBTB7A and cytidine deaminase complex of tBE. FIG. 31B:

Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the ZBTB7A-binding motif located in HBG1 enhancer.

[0050] FIG. 32: Editing efficiencies induced by tBE with the pairs of sgRNA-ZBTB7A-1-5 and its hsgRNAs targeting one of the two ZBTB7A-binding motifs located in HBG1 enhancer. FIG. 32A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-ZBTB7A-1-5, hsgRNA-ZBTB7A and cytidine deaminase complex of tBE. FIG. 32B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the ZBTB7A-binding motif located in HBG1 enhancer.

[0051] FIG. 33: Editing efficiencies induced by tBE with the pairs of sgRNA-ZBTB7A-2 and its hsgRNAs targeting another one of the two ZBTB7A-binding motifs located in HBG1/2 promoter. FIG. 33A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-ZBTB7A-2, hsgRNA-ZBTB7A and cytidine deaminase complex of tBE. FIG. 33B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting another one of the two ZBTB7A-binding motifs located in HBG1/2 promoter.

[0052] FIG. 34: Identification of new cytidine deaminase inhibitors. FIG. 34a, Schematic diagrams illustrate the APOBEC family members that have single- or dual-CDA domains (left) and BEs that were constructed with one or two CDA domains of dual-domain APOBECs (right). FIG. 34b, Editing frequencies induced by the indicated BEs at one representative genomic locus. FIG. 34c, Statistical analysis of normalized editing frequencies, setting the ones induced by the single-CDA-containing BEs as 100%. n = 78 (hA3BCDA2-nSpCas9-BE, hA3B-BE3, mA3CDA1-nSpCas9-BE and mA3-BE3) or 74 (hA3FCDA2-nSpCas9-BE and hA3F-BE3) edited cytosines at seven on-target sites from three independent experiments shown in FIG. 34b. FIG. 34d, Schematic diagrams illustrate the fusion of different dCDI domains to the N-terminus of mA3CDA1-nSpCas9-BE (mA3CDA1-BE3). **e**, Editing frequencies induced by the indicated BEs at one representative genomic locus. **f**, Statistical analysis of normalized editing frequencies, setting the ones induced by the BEs without dCDI domain as 100%. n = 57 edited cytosines at five on-target sites from three independent experiments shown in FIG. 34e. FIG. 34b,e NT, non-transfected control. Data

are presented as mean $\pm$ s.d. from three independent experiments. FIG. 34c,f *P* value, two-tailed Student's *t* test. The median and interquartile range (IQR) are shown.

[0053] FIG. 35: Characterization of new cytidine deaminase inhibitors. FIG. 35a, Schematic diagrams illustrate base editors constructed by fusing the indicated CDA domains to nSpCas9 and uracil DNA glycosylase inhibitor (UGI). The regulatory CDA domains are in grey shadow and the active CDA domains are in colors. NLS, nuclear localization sequence; XTEN and SGGs, linker peptides. FIG. 35b, C-to-T editing frequencies induced by the indicated BEs at six genomic loci. FIG. 35c, Schematic diagrams illustrate the fusion of different dCDI domains to the N-terminus of BE3 and hA3A-BE3. FIG. 35d, C-to-T editing frequencies induced by the indicated BEs at four genomic loci. FIG. 35b,d Data are presented as mean $\pm$ s.d. from three independent experiments.

[0054] FIG. 36: Editing efficiencies induced by tBE with the pairs of sgRNA-T43I-1~2, sgRNA-C747Y-G748K/R/E, sgRNA-S755N and theirs hsgRNAs targeting the coding sequences of Znf4 in BCL11A. FIG. 36A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-T43I-1~2, sgRNA-C747Y-G748K/R/E, sgRNA-S755N, different hsgRNA-T43I-1-1~3, hsgRNA-T43I-2-1~3, hsgRNA-C747Y-G748K/R/E-1~3, hsgRNA-S755N-1~3 and cytidine deaminase complex of tBE. FIG. 36B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf4 in BCL11A.

[0055] FIG. 37: Editing efficiencies induced by tBE with the pairs of sgRNA-L757F-1~2, sgRNA-L757F-T758I, sgRNA-V759I and theirs hsgRNAs targeting the coding sequences of Znf4 in BCL11A. FIG. 37A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-L757F-1~2, sgRNA-L757F-T758I, sgRNA-V759I, different hsgRNA-L757F-1-1~3, hsgRNA-L757F-2-1~3, hsgRNA-L757F-T758I-1~3, hsgRNA-V759I-1~2 and cytidine deaminase complex of tBE. FIG. 37B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf4 in BCL11A.

[0056] FIG. 38: Editing efficiencies induced by tBE with the pairs of sgRNA-H760Y, sgRNA-R761K, sgRNA-R761K-R762K, sgRNA-R762K-S763N and theirs hsgRNAs

targeting the coding sequences of Znf4 in BCL11A. FIG. 38A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-H760Y, sgRNA- R761K, sgRNA-R761K-R762K, sgRNA-R762K-S763N, different hsgRNA-H760Y-1~2, hsgRNA- R761K-1~2, hsgRNA-R761K-R762K-1~3, hsgRNA-R762K-S763N-1~3 and cytidine deaminase complex of tBE. FIG. 38B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf4 in BCL11A.

[0057] FIG. 39: Editing efficiencies induced by tBE with the pairs of sgRNA-H764Y and its hsgRNAs targeting the coding sequences of Znf4 in BCL11A , the pairs of sgRNA-G766N/S/D, sgRNA-G766N/S/D-E767K, sgRNA-R768K and theirs hsgRNAs targeting the coding sequences of the linker between BCL11A's Znf4 and Znf5. FIG. 39A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-H764Y, sgRNA-G766N/S/D, sgRNA-G766N/S/D-E767K, sgRNA-R768K, different hsgRNA-H764Y-1~2, hsgRNA-G766N/S/D-1~3, hsgRNA-G766N/S/D-E767K-1~2, hsgRNA-R768K-1~3 and cytidine deaminase complex of tBE. FIG. 39B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf4 in BCL11A and the coding sequences of the linker between BCL11A's Znf4 and Znf5.

[0058] FIG. 40: Editing efficiencies induced by tBE with the pairs of sgRNA-P769F/S/L and its hsgRNAs targeting the coding sequences of the linker between BCL11A's Znf4 and Znf5, the pairs of sgRNA-C775Y, sgRNA-A778V, sgRNA-A778T and theirs hsgRNAs targeting the coding sequences of Znf5 in BCL11A. FIG. 40A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-P769F/S/L, sgRNA-C775Y, sgRNA-A778V, sgRNA-A778T, different hsgRNA-P769F/S/L-1~3, hsgRNA-C775Y-1~3, hsgRNA-A778V-1~3, hsgRNA-A778T-1~3 and cytidine deaminase complex of tBE. FIG. 40B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of the linker between BCL11A's Znf4 and Znf5 and the coding sequences of Znf5 in BCL11A.

[0059] FIG. 41: Editing efficiencies induced by tBE with the pairs of sgRNA-A778V-A780V, sgRNA-C779Y-A780T, sgRNA-Q781*, sgRNA-S782N and theirs hsgRNAs targeting the coding sequences of Znf5 in BCL11A. FIG. 41A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-A778V-A780V, sgRNA-C779Y-A780T, sgRNA-Q781*, sgRNA-S782N, different hsgRNA-A778V-A780V-1~2, hsgRNA-C779Y-A780T-1~3, hsgRNA-Q781*-1~3, hsgRNA-S782N-1~3 and cytidine deaminase complex of tBE. FIG. 41B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf5 in BCL11A.

[0060] FIG. 42: Editing efficiencies induced by tBE with the pairs of sgRNA-S783N, sgRNA-L785F, sgRNA-L785F-T786I, sgRNA-R787K and theirs hsgRNAs targeting the coding sequences of Znf5 in BCL11A. FIG. 42A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-S783N, sgRNA-L785F, sgRNA-L785F-T786I, sgRNA-R787K, different hsgRNA-S783N-1, sgRNA-L785F-1~3, sgRNA-L785F-T786I-1~3, sgRNA-R787K-1 and cytidine deaminase complex of tBE. FIG. 42B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf5 in BCL11A.

[0061] FIG. 43: Editing efficiencies induced by tBE with the pairs of sgRNA-T791M-H792Y-1, sgRNA-T791M-H792Y-2, sgRNA-H792Y and theirs hsgRNAs targeting the coding sequences of Znf5 in BCL11A, the pairs of sgRNA-Q794* and its hsgRNAs targeting the coding sequences of the linker between BCL11A's Znf5 and Znf6 FIG. 43A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-T791M-H792Y-1~2, sgRNA-H792Y, sgRNA-Q794*, different hsgRNA-T791M-H792Y-1-1~3, hsgRNA-T791M-H792Y-2-1~3, hsgRNA-H792Y-1~3, hsgRNA-Q794*-1~3 and cytidine deaminase complex of tBE. FIG. 43B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf5 in BCL11A and the coding sequences of the linker between BCL11A's Znf5 and Znf6.

[0062] FIG. 44: Editing efficiencies induced by tBE with the pairs of sgRNA-G796K/R/E, sgRNA-G796K/R/E -D798N and theirs hsgRNAs targeting the coding sequences of the

linker between BCL11A's Znf5 and Znf6, the pairs of sgRNA-P808F/S/L, sgRNA-S813N and their hsgRNAs targeting the coding sequences of Znf6 in BCL11A. FIG. 44A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-T791M-H792Y-1~2, sgRNA-H792Y, sgRNA-Q794*, different hsgRNA-T791M-H792Y-1-1~3, hsgRNA-T791M-H792Y-2-1~3, hsgRNA-H792Y-1~3, hsgRNA-Q794*-1~3 and cytidine deaminase complex of tBE. FIG. 44B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of the linker between BCL11A's Znf5 and Znf6 and the coding sequences of Znf6 in BCL11A.

[0063] FIG. 45: Editing efficiencies induced by tBE with the pairs of sgRNA-S813N-2, sgRNA-E816K and their hsgRNAs targeting the coding sequences of Znf6 in BCL11A, the pairs of sgRNA-R826* and its hsgRNA targeting the coding sequences of the loop behind Znf6 in BCL11A. FIG. 45A: Schematic diagram illustrating the co-transfection of the plasmids for expressing tBE system. One plasmid expresses SpD10A nickase and another expresses sgRNA-S813N-2, sgRNA-E816K, sgRNA-R826*, different hsgRNA-S813N-2-1~2, hsgRNA-E816K-1, hsgRNA-R826*-1~3 and cytidine deaminase complex of tBE. FIG. 45B: Sanger sequencing results show the base editing efficiencies induced by tBE with indicated pairs of sgRNA/hsgRNA targeting the coding sequences of Znf6 in BCL11A and the coding sequences of the loop behind Znf6 in BCL11A.

DETAILED DESCRIPTION

Definitions

[0064] It is to be noted that the term “a” or “an” entity refers to one or more of that entity; for example, “an antibody,” is understood to represent one or more antibodies. As such, the terms “a” (or “an”), “one or more,” and “at least one” can be used interchangeably herein.

[0065] As used herein, the term “polypeptide” is intended to encompass a singular “polypeptide” as well as plural “polypeptides,” and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term “polypeptide” refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, peptides, dipeptides, tripeptides, oligopeptides, “protein”, “amino acid chain” or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of “polypeptide,” and the term

“polypeptide” may be used instead of, or interchangeably with any of these terms. The term “polypeptide” is also intended to refer to the products of post-expression modifications of the polypeptide, including without limitation glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or modification by non- naturally occurring amino acids. A polypeptide may be derived from a natural biological source or produced by recombinant technology, but is not necessarily translated from a designated nucleic acid sequence. It may be generated in any manner, including by chemical synthesis.

[0066] “Homology” or “identity” or “similarity” refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An “unrelated” or “non-homologous” sequence shares less than 40% identity, though preferably less than 25% identity, with one of the sequences of the present disclosure.

[0067] A polynucleotide or polynucleotide region (or a polypeptide or polypeptide region) has a certain percentage (for example, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 % or 99 %) of “sequence identity” to another sequence means that, when aligned, that percentage of bases (or amino acids) are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in Ausubel *et al.* eds. (2007) Current Protocols in Molecular Biology. Preferably, default parameters are used for alignment. One alignment program is BLAST, using default parameters.

[0068] The term “an equivalent nucleic acid or polynucleotide” refers to a nucleic acid having a nucleotide sequence having a certain degree of homology, or sequence identity, with the nucleotide sequence of the nucleic acid or complement thereof. A homolog of a double stranded nucleic acid is intended to include nucleic acids having a nucleotide sequence which has a certain degree of homology with or with the complement thereof. In one aspect, homologs of nucleic acids are capable of hybridizing to the nucleic acid or complement thereof. Likewise, “an equivalent polypeptide” refers to a polypeptide having a certain degree of homology, or sequence identity, with the amino acid sequence of a reference polypeptide.

In some aspects, the sequence identity is at least about 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%. In some aspects, the equivalent polypeptide or polynucleotide has one, two, three, four or five addition, deletion, substitution and their combinations thereof as compared to the reference polypeptide or polynucleotide. In some aspects, the equivalent sequence retains the activity (*e.g.*, epitope-binding) or structure (*e.g.*, salt-bridge) of the reference sequence.

[0069] The term “encode” as it is applied to polynucleotides refers to a polynucleotide which is said to “encode” a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for the polypeptide and/or a fragment thereof. The antisense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

Base Editors for Promoting Expression of Gamma-Globin

[0070] One embodiment of the present disclosure provides a newly designed base editor, referred to as transformer Base Editor (tBE), which can specifically edit cytosines in target regions with no observable off-target mutations. In the tBE system, a cytidine deaminase is fused with a nucleobase deaminase inhibitor to inhibit the activity of the nucleobase deaminase until the tBE complex is assembled at the target genomic site. In some embodiments, the tBE employs a sgRNA to bind at the target genomic site and a helper sgRNA (hsgRNA) to bind at a nearby region upstream to the target genomic site. The binding of two sgRNAs can guide the components of tBE to correctly assemble at the target genomic site for efficient base editing. Upon such assembly, a protease in the tBE system is activated, capable of cleaving the nucleobase deaminase inhibitor off from the nucleobase deaminase, which becomes activated.

[0071] The experimental example further tested a listing of designed sgRNA/hsgRNA sequences that target certain elements at the γ -globin promoter and/or other proteins whose expression impacts the expression of the γ -globin gene. For instance, the expression of the γ -globin is increased when the expression of BCL11A erythroid enhancer is impaired by a targeted mutation. Alternatively, when the BCL11A binding motif at the γ -globin promoter is mutated, the expression of the γ -globin gene can also be increased. Interestingly, the tBE technology can simultaneously target both the BCL11A's CREs and the BCL11A binding

motif at γ -globin promoter, which is contemplated to achieve even higher efficiency in activating γ -globin gene expression.

[0072] Moreover, sgRNA/hsgRNA sequences have also been designed and tested that target other protein factors that can influence the expression of γ -globin. For instance, KLF1 is an erythroid transcription factor that activates BCL11A expression directly by binding BCL11A's promoter; another protein, NFIX, regulates the expression of KLF1; yet, ZBTB7A (zinc finger and BTB domain containing 7A) binds a γ -globin promoter and represses its expression. Targeted genomic editing that disrupts the expression of any of these protein factors can lead to activation of the γ -globin, useful for treating diseases such as beta-thalassemia and sickle cell anemia. The data demonstrate that these designed sgRNA/hsgRNA sequences led to excellent editing efficiency and specificity.

[0073] In accordance with one embodiment of the present disclosure, therefore, provided is a base editing system, or one or more polynucleotides encoding the base editing system, useful for increasing the expression of the γ -globin gene in a target cell.

[0074] In some embodiments, the base editing system includes a CRISPR-associated (Cas) protein, a nucleobase deaminase, a single-guide RNA (sgRNA)/helper single-guide RNA (hsgRNA) pair targeting the BCL11A erythroid enhancer and/or the γ -globin promoter.

[0075] "Guide RNAs" are non-coding short RNA sequences which bind to the complementary target DNA sequences. A guide RNA first binds to the Cas enzyme and the gRNA sequence guides the complex via pairing to a specific location on the DNA, where Cas performs its endonuclease activity by cutting the target DNA strand. A "single guide RNA" frequently simply referred to as "guide RNA", refers to synthetic or expressed single guide RNA (sgRNA) that consists of both the crRNA and tracrRNA as a single construct. The tracrRNA portion is responsible for Cas endonuclease activity and the crRNA portion binds to the target specific DNA region. Therefore, the trans activating RNA (tracrRNA, or scaffold region) and crRNA are two key components and are joined by tetraloop which results in formation of sgRNA. Guide RNA targets the complementary sequences by simple Watson-Crick base pairing. TracrRNA are base pairs having a stemloop structure in itself and attaches to the endonuclease enzyme. crRNA includes a spacer, complementary to the target sequence, flanked region due to repeat sequences.

[0076] Example spacer sequences for the sgRNA/hsgRNA pair targeting the BCL11A erythroid enhancer are provided in **Tables 1-2**. In some embodiments, the sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:1-10, the hsgRNA includes the nucleic acid sequence of any one of SEQ ID NO:11-28. The sgRNA may any one of SEQ ID NO:2, 4, 6, 8, or 10, which is 20 nt in length. In some embodiments, the sgRNA includes at least a 10 nt fragment of any of these sequences, such as SEQ ID NO:1, 3, 5, 7, or 9. Such as apparent in these examples, the 10 nt fragment is preferably proximate to the PAM site. Such preference applies here as well in other examples as shown herein. The hsgRNA may include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:11, 13, 15, 17, 19, 21, 23, 25, and 27), or 20 nt in length (e.g., SEQ ID NO:12, 14, 16, 18, 20, 22, 24, 26 and 28). In some embodiments, the sgRNA includes the nucleic acid sequence of SEQ ID NO:1, and the hsgRNA comprises the nucleic acid of SEQ ID NO:17. In some embodiments, the sgRNA includes the nucleic acid sequence of SEQ ID NO:1, and the hsgRNA comprises the nucleic acid of SEQ ID NO:18. In some embodiments, the sgRNA includes the nucleic acid sequence of SEQ ID NO:4, and the hsgRNA comprises the nucleic acid of SEQ ID NO:11. In some embodiments, the sgRNA includes the nucleic acid sequence of SEQ ID NO:4, and the hsgRNA comprises the nucleic acid of SEQ ID NO:12.

[0077] Example spacer sequences for the sgRNA/hsgRNA pair targeting the γ -globin promoter (e.g., the BCL11A binding motif) are provided in **Tables 3-4**. In some embodiments, the sgRNA includes the nucleic acid sequence of SEQ ID NO:29-30, the hsgRNA includes the nucleic acid sequence of any one of SEQ ID NO:31-36. The hsgRNA may include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:31, 33 and 35), or 20 nt in length (e.g., SEQ ID NO:32, 34 and 36). In some embodiments, the sgRNA includes the nucleic acid sequence of SEQ ID NO:29-30, and the hsgRNA comprises the nucleic acid of SEQ ID NO:33. In some embodiments, the sgRNA includes the nucleic acid sequence of SEQ ID NO:29-30, and the hsgRNA comprises the nucleic acid of SEQ ID NO:34.

[0078] Example spacer sequences for the sgRNA/hsgRNA pair targeting one of the KLF1 motifs in BCL11A's CREs are provided in **Tables 5**. In some embodiments, the sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:37-42, the hsgRNA belonging to the same sub Table with its sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:55-62. The sgRNA may include a spacer (complementary region) that is about 10 nt

in length (e.g., SEQ ID NO:37, 39 and 41), or 20 nt in length (e.g., SEQ ID NO:38, 40 and 42). The hsgRNA may include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79), or 20 nt in length (e.g., SEQ ID NO:56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80).

[0079] Example spacer sequences for the sgRNA/hsgRNA pair targeting another of the KLF1 motifs in BCL11A's CREs are provided in **Tables 6**. In some embodiments, the sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:43-50, the hsgRNA belonging to the same sub Table with its sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:81-104. The sgRNA include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:43, 45, 47 and 49), or 20 nt in length (e.g., SEQ ID NO:44, 46, 48 and 50). The hsgRNA may include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101 and 103), or 20 nt in length (e.g., SEQ ID NO:82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102 and 104).

[0080] Example spacer sequences for the sgRNA/hsgRNA pair targeting yet another of the KLF1 motifs in BCL11A's CREs are provided in **Table 7**. In some embodiments, the sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:51-54, the hsgRNA belonging to the same sub Table with its sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:105-116. The sgRNA include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:51 and 53), or 20 nt in length (e.g., SEQ ID NO:52 and 54). The hsgRNA may include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:105, 107, 109, 111, 113 and 115), or 20 nt in length (e.g., SEQ ID NO:106, 108, 110, 112, 114 and 116).

[0081] Example spacer sequences for the sgRNA/hsgRNA pair targeting a GATA1-binding motif of NFIX (Nuclear Factor IX) CRE are provided in **Table 8**. In some embodiments, the sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:117-122, the hsgRNA belonging to the same sub Table with its sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:123-138. The sgRNA include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:117, 119 and 121), or 20 nt in length (e.g., SEQ ID NO:118, 120 and 122). The hsgRNA may include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:123, 125, 127, 129, 131, 133, 135 and 137), or 20 nt in length (e.g., SEQ ID NO: 124, 126, 128, 130, 132, 134, 136 and 138).

[0082] Example spacer sequences for the sgRNA/hsgRNA pair targeting a ZBTB7A-binding motif of HBG1/2's CRE are provided in **Tables 9 and 10**. In some embodiments, the sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:139-150, the hsgRNA belonging to the same sub Table with its sgRNA includes the nucleic acid sequence of any one of SEQ ID NO:151-190. The sgRNA include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:139, 141, 143, 145, 147 and 149), or 20 nt in length (e.g., SEQ ID NO:140, 142, 144, 146, 148 and 150). The hsgRNA may include a spacer (complementary region) that is about 10 nt in length (e.g., SEQ ID NO:151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187 and 189), or 20 nt in length (e.g., SEQ ID NO:152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188 and 190).

[0083] Additional example spacer sequences for the sgRNA/hsgRNA pair targeting various other sites are provided in **Table 11**. Such example sites include, e.g., T743, T743, C747 and G748, S755, L757, L757, L757 and T758, V759, H760, R761, R761 and R762, R761 and S763, H764, G766, G766 and E767, R768, P769, C775, A778, A778, A778 and A780, C779 and A780, Q781, S782, S783, L785, L785 and T786, S783, T791 and H792, T791 and H792, H792, Q794, G796, G796 and D798, P808, S813, S813, E816, or R826 of BCL11A. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:353-430, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:431-628.

[0084] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:1-10, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-28.

[0085] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:29-30, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:31-36.

[0086] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:37-38, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:55-62. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:39-40, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ

ID NO:63-70. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:41-42, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:71-80.

[0087] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:43-44, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:81-104. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:45-46, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:81-104. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:47-48, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:81-104. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:49-50, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:87-98.

[0088] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:51-52, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:105-116. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:53-54, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:105-116.

[0089] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:117-118, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:119-120, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:121-122, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138.

[0090] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:139-140, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164. In some embodiments, the

sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:141-142, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:143-144, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:145-146, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:147-148, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:165-178.

[0091] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:149-150, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:179-190.

[0092] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:353-354, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:431-436. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:355-356, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:437-442. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:357-358, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:443-448.

[0093] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:359-360, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:449-454. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:361-362, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:455-460. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:363-364, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:461-466.

[0094] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:365-366, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:467-472. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:367-368, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:473-476. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:369-370, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:477-480.

[0095] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:371-372, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:481-484. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:373-374, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:485-488. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:375-376, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:489-492.

[0096] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:377-378, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:493-496. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:379-380, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:497-502. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:381-382, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:503-506.

[0097] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:383-384, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:507-512. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:385-386, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:513-518. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:387-388, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:519-524.

[0098] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:389-390, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:525-530. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:391-392, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:531-536. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:393-394, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:537-540.

[0099] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:395-396, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:541-546. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:397-398, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:547-552. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:399-400, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:553-558.

[0100] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:401-402, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:559-560. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:403-404, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:561-566. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:405-406, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:567-572.

[0101] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:407-408, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:573-574. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:409-410, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:575-580. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:411-412, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:581-586.

[0102] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:413-414, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:587-592. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:415-416, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:593-598. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:415-416, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:593-598.

[0103] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:417-418, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:599-602. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:419-420, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:603-606. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:421-422, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:607-612.

[0104] In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:423-424, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:613-616. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:425-426, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:617-620. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:427-428, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:621-622. In some embodiments, the sgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:429-430, and the hsgRNA includes a nucleic acid sequence selected from the group consisting of SEQ ID NO:623-628.

[0105] In some embodiments, the base editing system targets two or more of the above target sites, e.g., BCL11A, BCL11A binding motif of γ -globin, KLF1 binding motifs of BCL11A, GATA1-binding motif of NFIX, and/or ZBTB7A-binding motif of γ -globin.

[0106] In some embodiments, the base editing system targets both the BCL11A erythroid enhancer and the γ -globin promoter. Accordingly, two pairs of sgRNA/hsgRNA are included. In a particular example, the first sgRNA/hsgRNA pair includes spacers as described in SEQ ID NO:4 and 11 (or 12), and the second sgRNA/hsgRNA pair includes spacers as described in SEQ ID NO:30 and 33 (or 34).

[0107] The term “nucleobase deaminase” as used herein, refers to a group of enzymes that catalyze the hydrolytic deamination of nucleobases such as cytidine, deoxycytidine, adenosine and deoxyadenosine. Non-limiting examples of nucleobase deaminases include cytidine deaminases and adenosine deaminases.

[0108] “Cytidine deaminase” refers to enzymes that catalyze the irreversible hydrolytic deamination of cytidine and deoxycytidine to uridine and deoxyuridine, respectively. Cytidine deaminases maintain the cellular pyrimidine pool. A family of cytidine deaminases is APOBEC (“apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like”). Members of this family are C-to-U editing enzymes. Some APOBEC family members have two domains, one domain of APOBEC like proteins is the catalytic domain, while the other domain is a pseudocatalytic domain. More specifically, the catalytic domain is a zinc dependent cytidine deaminase domain and is important for cytidine deamination.

[0109] Non-limiting examples of APOBEC proteins include APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, and activation-induced (cytidine) deaminase (AID).

[0110] Various mutants of the APOBEC proteins are also known that have bring about different editing characteristics for base editors. For instance, for human APOBEC3A, certain mutants (e.g., W98Y, Y130F, Y132D, W104A, D131Y and P134Y) even outperform the wildtype human APOBEC3A in terms of editing efficiency or editing window. Accordingly, the term APOBEC and each of its family member also encompasses variants and mutants that have certain level (e.g., 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%) of sequence identity to the corresponding wildtype APOBEC protein or the catalytic domain and retain the cytidine deaminating activity. The variants and mutants can be derived with amino acid additions, deletions and/or substitutions. Such substitutions, in some embodiments, are conservative substitutions.

[0111] “Adenosine deaminase”, also known as adenosine aminohydrolase, or ADA, is an enzyme (EC 3.5.4.4) involved in purine metabolism. It is needed for the breakdown of adenosine from food and for the turnover of nucleic acids in tissues.

[0112] Non-limiting examples of adenosine deaminases include tRNA-specific adenosine deaminase (TadA), adenosine deaminase tRNA specific 1 (ADAT1), adenosine deaminase tRNA specific 2 (ADAT2), adenosine deaminase tRNA specific 3 (ADAT3), adenosine deaminase RNA specific B1 (ADARB1), adenosine deaminase RNA specific B2 (ADARB2), adenosine monophosphate deaminase 1 (AMPD1), adenosine monophosphate deaminase 2 (AMPD2), adenosine monophosphate deaminase 3 (AMPD3), adenosine deaminase (ADA), adenosine deaminase 2 (ADA2), adenosine deaminase like (ADAL), adenosine deaminase domain containing 1 (ADAD1), adenosine deaminase domain containing 2 (ADAD2), adenosine deaminase RNA specific (ADAR) and adenosine deaminase RNA specific B1 (ADARB1).

[0113] Some of the nucleobase deaminases have a single, catalytic domain, while others also have other domains, such as an inhibitory domain as currently discovered by the instant inventors. In some embodiments, therefore, the first fragment only includes the catalytic domain, such as mA3-CDA1, hA3F-CDA2 and hA3B-CDA2. In some embodiments, the first fragment includes at least a catalytic core of the catalytic domain.

[0114] The term “Cas protein” or “clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) protein” refers to RNA-guided DNA endonuclease enzymes associated with the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) adaptive immunity system in *Streptococcus pyogenes*, as well as other bacteria. Cas proteins include Cas9 proteins, Cas12a (Cpf1) proteins, Cas12b (formerly known as C2c1) proteins, Cas13 proteins and various engineered counterparts. Example Cas proteins include SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b, RanCas13b and those provided in **Table A** below.

Table A. Example Cas Proteins

Cas protein types	Cas proteins
Cas9 proteins	Cas9 from <i>Staphylococcus aureus</i> (SaCas9) Cas9 from <i>Neisseria meningitidis</i> (NmeCas9) Cas9 from <i>Streptococcus thermophilus</i> (StCas9) Cas9 from <i>Campylobacter jejuni</i> (CjCas9)
Cas12a (Cpf1) proteins	Cas12a (Cpf1) from <i>Acidaminococcus</i> sp BV3L6 (AsCpf1) Cas12a (Cpf1) from <i>Francisella novicida</i> sp BV3L6 (FnCpf1) Cas12a (Cpf1) from <i>Smithella</i> sp SC K08D17 (SsCpf1) Cas12a (Cpf1) from <i>Porphyromonas crevioricanis</i> (PcCpf1) Cas12a (Cpf1) from <i>Butyrivibrio proteoclasticus</i> (BpCpf1) Cas12a (Cpf1) from <i>Candidatus Methanoplasma termitum</i> (CmtCpf1) Cas12a (Cpf1) from <i>Leptospira inadai</i> (LiCpf1) Cas12a (Cpf1) from <i>Porphyromonas macacae</i> (PmCpf1) Cas12a (Cpf1) from <i>Peregrinibacteria bacterium</i> GW2011 WA2 33 10 (Pb3310Cpf1) Cas12a (Cpf1) from <i>Parcubacteria bacterium</i> GW2011 GWC2 44 17 (Pb4417Cpf1) Cas12a (Cpf1) from <i>Butyrivibrio</i> sp. NC3005 (BsCpf1) Cas12a (Cpf1) from <i>Eubacterium eligens</i> (EcCpf1)
Cas12b (C2c1) proteins	Cas12b (C2c1) <i>Bacillus hisashii</i> (BhCas12b) Cas12b (C2c1) <i>Bacillus hisashii</i> with a gain-of-function mutation (see, e.g., Strecker et al., Nature Communications 10 (article 212) (2019)) Cas12b (C2c1) <i>Alicyclobacillus kakegawensis</i> (AkCas12b) Cas12b (C2c1) <i>Elusimicrobia bacterium</i> (EbCas12b) Cas12b (C2c1) <i>Laceyella sediminis</i> (Ls) (LsCas12b)
Cas13 proteins	Cas13d from <i>Ruminococcus flavefaciens</i> XPD3002 (RfCas13d) Cas13a from <i>Leptotrichia wadei</i> (LwaCas13a) Cas13b from <i>Prevotella</i> sp. P5-125 (PspCas13b) Cas13b from <i>Porphyromonas gulae</i> (PguCas13b) Cas13b from <i>Riemerella anatipestifer</i> (RanCas13b)
Engineered Cas proteins	Nickases (mutation in one nuclease domain) Catalytically inactive mutant (dCas9; mutations in both of the nuclease domains) Enhanced variants with improved specificity (see, e.g., Chen <i>et al.</i> , Nature, 550, 407-410 (2017))

[0115] In some embodiments, the base editing system further includes a nucleobase deaminase inhibitor fused to the nucleobase deaminase. A “nucleobase deaminase inhibitor,” accordingly, refers to a protein or a protein domain that inhibits the deaminase activity of a nucleobase deaminase. In some embodiments, the second fragment includes at least an inhibitory core of the inhibitory protein/domain.

[0116] Two example nucleobase deaminase inhibitors are mA3-CDA2, hA3F-CDA1 and hA3B-CDA1 (sequences provided in **Table B**), which are the inhibitory domains of the

corresponding nucleobase deaminases. Additional nucleobase deaminase inhibitors have been identified in the protein databases as homologues of mA3-CDA2, hA3F-CDA1 and hA3B-CDA1 (see **Tables B1, B2 and B3**). Their biological equivalents (e.g., having at least about 80%, 85%, 90%, 95%, 97%, 98%, 99%, 99.5% sequence identity, or having one, two, or three amino acid addition/deletion/substitution, and having nucleobase deaminase inhibitor activity) can also be prepared with known methods in the art, such as conservative amino acid substitutions.

[0117] When the nucleobase deaminase inhibitor is included, it is fused to the nucleobase deaminase but is separated by a protease cleavage site. In some embodiments, the base editing system further includes the protease that is capable of cleaving the protease cleavage site.

[0118] The protease cleavage site can be any known protease cleavage site (peptide) for any proteases. Non-limiting examples of proteases include TEV protease, TuMV protease, PPV protease, PVY protease, ZIKV protease and WNV protease. The protein sequences of example proteases and their corresponding cleavage sites are provided in **Table B**.

Table B. Example Sequences

Name	Sequence	SEQ ID NO:
Mouse APOBEC3 cytidine deaminase domain 2	MSSSTLSNICLTGKLPETRFWVEGRMDPLSEEEFYQSQFYNQVRVKHLCY YHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAELFLDKIRSMELSQ VTITCYLTWSPCPNCAWQLAAFKRDRPDILHIYTSRLYFHWKRPFPQKG LCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRR LRRIKESWGLQDLVNDFGNLQLGPPMS	191
Human APOBEC3F cytidine deaminase domain 1	MKPHFRNTVERMYRDTFSYNFYNRPIILSRNTVWLCYEVKTKGSPSRPL DAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWVFSWTPCPD CVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIM DDEEFAYCWENFVYSEG	192
Human APOBEC3B cytidine deaminase domain 1	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLL WDTGVFRGQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWVFSWTPCP DCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKI MDYEEFAYCWENFVYNEGQ	193
TEV protease N-terminal domain	MGESLFKGRDYNPISSTICHLTNE SDGHTTSLYGIGFGPFIITNKHLEF RRNNGTLLVQSLHGVFKVKNNTTLQQLHLDGRDMIIRMPKDFPPFPQK LKFREPQREERICLVTTNFQT	194
TEV protease C-terminal domain	MKSMSMVSDTSCTFPSSDGIWFKHWIQTGDGQCGSPLVSTRDGFIVGI HSASNFTNTNNTFTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGGHK VFMVKPEEPFQPVKEATQ	195
TEV protease cleavage site	ENLYFQS	196
TuMV protease	MASSNSMFRGLRDYNPISNNICHLTNVSDGASNSLYGVGFGLIILTNRH LFERNNGELVIKSRHGEFVIKNTTQLHLLIPDRDLLLIRLPKDVPPFP QKLGRQPEKGERICMVGSNFQTKSITSIVSETSTIMPVENSQFWKHWI	197

	STKDGQCSPMVSTKDGKILGLHSLANFQNSINYFAAFPD DFAEKYLHT IEAHEWVKHWKYNTSAISWGS LN IQASQPSGLFKVSKLISDL DSTAVYA Q	
TuMV protease cleavage site	GGCSHQ S	198
PPV protease	MASSKSLFRGLRDYNPIASSICQLNNSSGARQSEMFG LGFGGLIVTNQH LFKRNDGELTIRSHHGEFVVKDTKTLKLLPCKGRDIVIIRLPKDFPFP RRLQFRTPPTTEDRVCLIGSNFQTKSISSTMSETSATYPVDNSHFWKHWI STKDGHCGLPIVSTRDGSILGLHSLANSTNTQNFYAAFPDNFETTYLSN QDNDNWIQWRYNPDEVCSWGLQLKRDIPQSPFTICKLLTDLDGEFVYT Q	199
PPV protease cleavage site	QVVVHQSK	200
PVY protease	MASAKSLMRGLRDNFPIAQTVCR LKVSVEYGASEMYGFGGAYIVANHH LFRSYNGSMEVQSMHGTFRVKNLHSLSVLP IKGRDIILIKMPKDFPVP QKLHFRAPTQNERICLVGTNFQEKYASSIIITETSTTYNIPGSTFWKHWI ETDNGHCGLPVVSTADGCIVGIHSLANNAHTTNYSAFDEDFESKYLRT NEHNEWVKS WVYNPDIVLWGPLKLDSTPKGLFKTTKLVQDLIDHDVVV EQ	201
PVY protease cleavage site	YDVRHQSR	202
ZIKV protease	MASDMYIERAGDI TWEKDAEVTGNSPRLDVALDESGDFSLVEEDGPPMR EGGGGSGGGGSGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVG VMQEGVFHTLWHTTKGAALMSGEGRLDPYWG DVKQDLVSYCGPWKLDAA WDGLSEVQLLAVPPGERARNIQTLP GIFTKDGDIGAVALDYPAGTSGS PILDKCGRVIGLYGNGVVIKNGSYVSAITQ GKREETPVECFE	203
ZIKV protease cleavage site	KERKRRGA	204
WNV protease	MASSTDMWIERTADISWESDAEITGSSERVDVRLDDDG NFQLMNDPGAP WKG GGGSGGGGVLWDTPSPKEYKKGDTTGTGVYRIMTRLLGSYQAGAG VMVEGVFHTLWHTTKGAALMSGEGRLDPYWG SVKEDRLCYGGPWKLQHK WNGQDEVQMIIVVEPGKNVKNVQTKPGVFKTPEGEIGAVTLDFPTGTSGS PIVDKNGDVIGLYGNGVIMPNGSYISAI VQGERMDEPIPAGEPEML	205
WNV protease cleavage site	KQKKRGGK	206
MS2	ACAUGAGGAUCACCCAUGU	207
sgRNA scaffold with 2×MS2	GUUUGAGAGCUAGGCCA ACAUGAGGAUCACCCAUGU CUGCAGGGCCUAG CAAGUUCAAUAAGGCUAGUCCGUUAUCAACUUGGCCA ACAUGAGGAUC ACCCAUGU CUGCAGGGCCAAUGGGCACCGAGUCGGUGC	208
PP7	GGAGCAGACGAUAUGGCGUCGCUCC	209
sgRNA scaffold with 2×PP7	GUUUGAGAGCUACCGGAGCAGACGAUAUGGCGUCGCUCCGGUAGCAAGU UCAAUAAGGCUAGUCCGUUAUCAACUUGGAGCAGACGAUAUGGCGUCG CUCCAAGUGGCACCGAGUCGGUGC	210
boxB	GCCUGAAGAAGGGC	211
sgRNA scaffold with 2×boxB	GUUUGAGAGCUAGGGCCCUGAAGAAGGGCCCUAGCAAGUUCAAUAAGG CUAGUCCGUUAUCAACUUGGGCCCUGAAGAAGGGCCCAUGGGCACCGA GUCGGUGC	212
MS2 coat protein (MCP)	MASNFTQFVLVDNNGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNME LTIP IFATNSDCELI VKAMQG LLKDGNIPI SAIAANS GIY	213
PP7 coat protein (PCP)	MGSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRRLRTASLRQN GAKTAYRVNLKLDQADVDSGLPKVRYTQVW SHDVTIVANSTEASRKSL YDLTKSLVATSQVEDLVVNLVPLGR	214
boxB coat protein (N22p)	MGNARTRRRERRAEKQAQWKAAN	215

UGI	TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDES TDENVMLLTSDAPEYKPWALVIQDSNGENKIKML	216
P2A	GSGATNFSLLKQAGDVEENPGP	217
T2A	GSGEGRGSLLTCGDVEENPGP	218
E2A	GSGQCTNYALLKLAGDVESNPGP	219

Table B1: mA3CDA2 Core Sequence Related Domains

Name	Sequence	SEQ ID NO:
Mouse APOBEC3 cytidine deaminase domain 2 core (AA282-AA355)	SEKKGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRP DLILHIYTSRLYFHWKRPFQKGLC	220
Mus spicilegus A3 (AA248-AA321)	SEKKGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRP DLIPHIYTSRLYFHWKRPFQKGLC	221
Cricetulus longicaudatus A3 (AA249-AA322)	SEKKGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWRLAAFKRDRP DLILHIYTSRLYFHWKRPFQKGLC	222
Mus terricolor A3 (AA248-AA321)	SEKKGQHAIEILFLNKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKKDRP DLILHIYTSRLYFHWKRPFQKGLC	223
Mus caroli A3 (AA260-AA333)	SKKKGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDHP DLILHIYTSRLYFHWKRPFQKGLC	224
Mus pahari A3 (AA263-AA336)	SKKKGQHAIEILFLEKIRSMELSMRITCYLTWSPCPNCAWQLAAFKQDRP DLILHIYTSRLYFHWRRIFQKGLC	225
Mus shortridgei A3 (AA233-AA306)	SKKKGQHAIEILFLEKIRSMELSMRITCYLTWSPCPNCAWQLAAFKQDRP DLILHIYTSRLYFHWRRIFQKGLC	226
Mus setulosus A3 (AA29-AA302)	SKKKGQHAIEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLETFKKDRP DLILHIYTSRLYFHWKRAFQEGLC	227
Grammomys surdaster A3 (AA270-AA344)	SKKKGPHAEILFLDKMWSMEELSQVRITCYLTWSPCPNCARQLAAFKKDH PGLILRIYTSRLYFYWRRKFQKGLC	228
Rattus norvegicus A3 (AA256-AA328)	KKGEQHVEILFLEKMRSMELSQVRITCYLTWSPCPNCARQLAAFKKDHPD LILRIYTSRLYFYWRKKFQKGLC	229
Mastomys coucha A3 (AA258-AA331)	SKKGRQHAIEILFLEKVRSMQLSQVRITCYLTWSPCPNCAWQLAAFKMDHP DLILRIYASRLYFHWRRAFQKGLC	230
Cricetulus griseus A3B (AA235- AA307)	NKKGKHAIEILFIDEMRSLELGQVQITCYLTWSPCPNCAQELAAFKSDHPD LVLRIYTSRLYFHWRRKYQEGLC	231
Peromyscus leucopus A3 (AA266-AA338)	NKKGKHAIEILFIDEMRSLELGQARITCYLTWSPCPNCAQKLAAFKKDHPD LVLRVYTSRLYFHWRRKYQEGLC	232
Mesocricetus auratus A3 (AA268-AA340)	NKKDKHAIEILFIDKMRSLCQVRITCYLTWSPCPNCAQELAAFKKDHPD LVLRIYTSRLYFHWRRKYQEGLC	233
Microtus ochrogaster A3B (AA266- AA338)	NKKGKHAIEILFIDEMRSKLSQERITCYLTWSPCPNCAQELAAFKRDHPG LVLRIYASRLYFHWRRKYQEGLC	234
Nannospalax galili A3 (AA231-AA302)	NKRAKHAIEILLIDMMSMELGQVQITCYITWSPCPTCAQELAAFKQDHPD LVLRIYASRLYFHWKRKFQKGL	235
Meriones unguiculatus A3 (AA233-AA305)	NKKGRHAIEICLIDEMRSLGLGKAQITCYLTWSPCRKCAQELATFKKDHPD LVLRVYASRLYFHWRSKYQQGLC	236
Dipodomys ordii A3 (AA256-AA330)	NKKGGHAEIRFIERIRSMGLDPSQDYQITCYLTWSPCLDCAFKLAKLKKD FPRLTLRIFTSRLYFHWIRKFQKGL	237
Jaculus jaculus A3	NKKGKHAEARFVDMRSMQLDHALITCYLTWSPCLDCSQKLAALKRDHPG	238

(AA303-AA374)	LTLRIFTSRLYFHWVKKFQEG	
Chinchilla lanigera A3H (AA86-AA161)	SPQKGHHAESRFIKRISMDLDRSRSYQITCFLTWSPCPSCAQELASFKR AHPHLRFQIFVSRLYFHWKRSYQAGL	239
Heterocephalus glaber A3 (AA277-AA350)	KKGYHAESRFIKRICSMGLDQDSYQVTCTFLTWSPCPHCAQELVSFKRAH PHLRLQIFTARLFFHWKRSYQEG	240
Octodon degus A3 (AA256-AA329)	KKGQHAEIFRIFERIHSMALDQARSYQITCFLTWSPCPFCAQELASFKSTH PRVHLQIFVSRLYFHWKRSYQEG	241
Urociellus parryi A3 (AA256-AA330)	NKKGHHAEIFRIKIRSLDLQDSQNYEVTCTFLTWSPCPDCAQELVALTRS HPHVRLRLFTSRLYFHWFSFQEG	242
Aotus nancymae A3H (AA75-AA146)	NRHAEICFIDEIESMGLDKTQCYEVTCTFLTWSPCPSCAQKLAFTKAQVH LNLRIFFASRLYYHWRSSYQKGL	243
Cebus capucinus imitator A3H (AA55-AA126)	NRHAEICFIDEIESMGLDKTQCYEVTCTFLTWSPCPSCAQKLVAFAKAQDH LNLRIFFASRLYYHWRSSYQKGL	244
Saimiri boliviensis boliviensis A3H (AA56-AA125)	HVEICFIDKIASMGLDKTQCYDVTCTFLTWSPCPSCAQKLAFAKAQDHLN LRIFASRLYYHWRSSYQKGL	245
Homo sapiens A3H (AA49-AA123)	NKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSCAWELVDFIKA HDHLNLGIFASRLYYHWCKPQQKGL	246
Homo sapiens ARP10 (AA48-AA123)	ENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSCAWELVDFIK AHDHLNLGIFASRLYYHWCKPQQKGL	247
Pan paniscus A3H (AA49-AA123)	NKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSCAWKLVDIFIQA HDHLNLRIFFASRLYYHWCKPQQEG	248
Symphalangus syndactylus A3H (AA49-AA123)	NKKKRHAEIFRINKIKSMGLDETQCYQVTCYLTWSPCPSCAWELVDFIKA HDHLNLGIFASRLYYHWCRHQEG	249
Macaca mulatta A3H (AA49-AA123)	NKKKDHAEIFRINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKA HRHLNLRIFFASRLYYHWRPNYQEG	250
Theropithecus gelada A3H (AA54-AA128)	NKKKEHAEIFRINKIKSMGLDETQCYQVTCYLTWSPCPSCAGKLVDIFIQA HHHLNLRIFFASRLYYHWRPNYQEG	251
Mandrillus leucophaeus A3H (AA49-AA123)	NKKKHHAEIFRINKIKSMGLDETQCYQVTCYLTWSPCPSCARELVDFIKA HRHLNLRIFFASRLYYHWRPHYQEG	252
Bos grunniens A3 (AA74-AA148)	NKKQRHAEIFRIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFI TR NNHLKLEIFASRLYFHWIKPFKMG	253
Bubalus bubalis A3 (AA74-AA148)	NKKQRHAEIFRIDKINSLDLNPSQSYKIICYITWSPCPNCASELVDFI TR NDHLDLQIFASRLYFHWIKPFKRG	254
Odocoileus virginianus texanus A3H (AA209-AA283)	NKKQRHAEIFRIDKINSLDLNRRQSYKIICYITWSPCPRCASELVDFITG NDHLNLQIFASRLYFHWKPFQRGL	255
Sus scrofa A3 (AA51-AA125)	NKKKRHAEIFRIDKINSLDLNQNQCYRIICYVTWSPCHNCAKELVDFI SN RHHLSQLFASRLYFHWVRCYQRGL	256
Ceratotherium simum simum A3B (AA232-AA306)	NKKKRHAEIFRIDKIKSLGLDRVQSYEITCYITWSPCPTCALELVAFTRD YPRLSLQIFASRLYFHWRRRSIQGL	257
Equus caballus A3H (AA79-AA153)	NKKKRHAEIFRIDKINSLGLDQDSYETCYVTWSPCATCACKLIKFTRK FPNLSLRIFVSRLYYHWRQNNQGL	258
Enhydra lutris kenyonii A3B (AA243-AA316)	KKKRHAEIFRIDSIRALQLDQSQRFETCYLTWSPCPTCAKELAMFVQDH PHISLRLFASRLYFHWKRYQEG	259
Leptonychotes weddellii A3H (AA50-AA123)	KKKRHAEIFRIDNIKALRLDTSQRFETCYVTWSPCPTCAKELVAFVRDH RHISLRLFASRLYFHWLRENKGL	260
Ursus arctos horribilis A3F (AA552-AA626)	NKKKRHAEIFRIDKIRSLQRDSSQTFETCYVTWSPCFTCAEELVAFVRD HPHVRLRLFASRLYFHWLRKYQEG	261

Panthera leo bleyenberghi A3H (AA50-AA124)	NKKRRHAEICFIDKIKSLTRDTSQRFEIICYITWSPCPFCAEELVAFVKD NPHLSLRIFASRLYVHWRWKYQQGL	262
Panthera tigris sumatrae A3H (AA50-AA124)	NKKRRHAEICFIDKIKSLTRDTSQRFEIICYITWSPCPFCAEELVAFVKD NPHLSLRIFASRLYVHWRWKYQQGL	263
Tupaia belangeri A3 (AA46-AA120)	NKKHRHAEVRFIAKIRSMSLDLDQKHQLTCYLTWSPCPSCAQELVTFMAE SRHLNLQVFVSRLYFHWQRDFQQGL	264

Table B2: hA3FCDA1 Core Sequence Related Domains

Name	Sequence	SEQ ID NO:
Pan troglodytes A3F (AA29-AA136)	RRNTVWLCYEVKTKGPSRPLDTKIFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKCFQITWFVSWTPCPDCVAKLAEF LAEHPNVTLTISAARLYYYWER DYRRALCR	265
Pan paniscus A3F (AA29-AA136)	RRNTVWLCYEVKTKGPSRPLDTKIFRGQVYFQFENHAEMCFLSWFCGNQ LPAYKCFQITWFVSWTPCPDCVAKLAEF LAEHPNVTLTISAARLYYYWER DYRRALCR	266
Colobus angolensis palliatu A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMTWGAIFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKCFQITWFVSWTPCPDCVGKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	267
Macaca mulatta A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMTWDTKIFRGQVYSKEPHEHAEMCFLSRFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	268
Macaca fascicularis A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSVPTWGTKIFRGQVYSKEPHEHAEMCFLSWFCGNQ LPTYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	269
Rhinopithecus roxellana A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMTWGAIFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	270
Rhinopithecus bieti A3F (AA18-AA125)	RRNTVWLCYEVKTRGPSMTWGAIFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	271
Rhinopithecus roxellana A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMTWGAIFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	272
Macaca mulatta A3F (AA40-AA147)	RRNTVWLCYEVKTRGPSMTWDTKIFRGQVYSKEPHEHAEMCFLSRFCGNQ LPAYKRFQITWFVSWTPCPCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	273
Trachypithecus francoisi A3F (AA40-AA147)	RRNTVWLCYEVKTRGPSMTWGAIFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	274
Gorilla gorilla A3F (AA29-AA127)	RRNTVWLCYEVKTKGPSRPLDAKIFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKCFQITWFVSWTPCPDCVAKLAEF LAEHPNVTLTISAARLYYYWE	275
Papio anubis A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMTWDAKIFRGQVYFQPYHAEMCFLSRFCGNQ LPAYKRFQITWFVSWTPCPDCVVKTEFLAEHPNVTLTISAARLYYYWET DYRRALCR	276
Pongo abelii A3F (AA29-AA136)	RRNTVWLCYKVKTKGPSRPLNAKIFRGQVYFEPQYHAEMCFLSWFCGNQ LSAYERFQITWFVSWTPCPDCVAMLAEF LAEHPNVTLTISAARLYYYWER DYRGALRR	277
Macaca leonina A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMTWGTKIFRGQVCFEPQYHAEMCFLSRFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	278
Macaca nemestrina A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMTWGTKIFRGQVCFEPQYHAEMCFLSRFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWET DYRRALCR	279
Homo sapiens A3B (AA30-AA137)	RSYTWLCYEVKIKGRSNLLWDTGVFRGQVYFEPQYHAEMCFLSWFCGNQ LPAYKCFQITWFVSWTPCPDCVAKLAEF LSEHPNVTLTISAARLYYYWER DYRRALCR	280
Gorilla gorilla gorilla	RSYNWLCYEVKIKGRSNLLWNTGVFRGQMYSQPEHHAEMCFLSWFCGNQ	281

A3B (AA30-AA137)	LPAYKCFQITWVSWTPCPDCVAKLAEF LAEYPNVTLTISAARLYYYWER DYRRALCR	
Pan troglodytes A3B (AA30-AA137)	RSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMCFLSWFCGNQ LSAYKCFQITWVSWTPCPDCVAKLAEF LAEHPNVTLTISAARLYYYWER DYRRALCR	282
Theropithecus gelada A3F (AA29-AA136)	RRNTVWLCYEVKTRGPSMP TWGTKIFRGQVYFQPQYHAEMCFLSRFCGNQ LPAYKRFQITWVSWNCPDCVAKVIEFLAEHPNVTLTISAARLYYYWGR DWRRALRR	283
Mandrillus leucophaeus A3F (AA29-AA130)	RRNTVWLCYEVKTRGPSMP TWGTKIFRGQVYFQPQYHAEMCFLSWFCGNQ LPAYKRFQITWVSWTPCPDCVVKVAEF LAEHPNVTLTISAARLYYYWET DY	284
Gorilla gorilla gorilla A3B (AA30-AA137)	RSYTWLCYEVKIRGRSNLLWDTGVFRGQMYSQPEHHAEMCFLSWFCGNQ LPAYKCFQITWVSWTPCLDCVAKLAEF LAEYPNVTLTISTARLYYYWER DYRRALCR	285
Pan paniscus A3B (AA30-AA137)	RSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMYFLSWFCGNQ LSAYKCFQITWVSWTPCPDCVAKLAEF LAEHPNVTLTISAARLYYYWER DYRRALCR	286
Hylobates moloch A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYFQPEYHAEMCFLSWFCGNQ LPAYKRFQITWVSWTPCPDCVAKVAEF LAEHPNVTLTISAARLYYYWEK DWQRALCR	287
Symphalangus syndactylus A3G (AA22-AA129)	RRNTVWLCYEVKTKDPSRRLDTKIFRGKVYFQLENHAEMCFLSWFCGNQ LPANRCFQITWVSWNCPCLPCVAKVTKFLAEHPNVTLTISAARLYYYRAR DWRRALRR	288
Macaca mulatta A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYSKPEHHAEMCFLSWFCGNQ LPAHKRFQITWVSWTPCPDCVAKVAEF LAEYPNVTLTISAARLYYYWET DYRRALCR	289
Chlorocebus sabaeus A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYSKPEHHAEMCFLSWFCGNQ LPAHKRFQITWVSWTPCPDCVAKVAEF LAEYPNVTLTISAARLYYYWET DYRRALCR	290
Nomascus leucogenys A3B (AA30-AA137)	RRSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYFQPEYHAEMCFLSWFCGN QLPAYKRFQITWVSWTPCPDCVAKVAVFLAEHPNVTLTISAARLYYYWE KDWQRALCR	291
Trachypithecus francoisi A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGNQ LPAYKRFQITWVSWTPCPDCVAKLAEF LTHEPNVTLTISAARLYYYRGR DWRRALCR	292
Trachypithecus francoisi A3F (AA22-AA129)	RSYTWLCYEVKKRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGNQ LPAYKRFQITWVSWTPCPDCVAKVAEF LAEHPKVTLTISAARLYYYWDR DWRRALCR	293
Rhinopithecus bieti A3F (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGNQ LPAYKRFQITWVSWTPCPDCVAKVAEF LTHEPNVTLTISAARLYYYRGR DWRRALCR	294
Rhinopithecus roxellana A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGNQ LPAYKRFQITWVSWTPCPDCVAKVAEF LTHEPNVTLTISAARLYYYRGR DWRRALCR	295
Pongo abelii A3F (AA29-AA128)	RRNYTWLCYEVKIRKDP SKLAWDTGVFRGQVYSQPEHHAEMCFLSWFCGN QLSAYERFQITWVSWTPCPDCVAKLAEF LAEHPNVTLTVSAARLYYYWE	296
Macaca mulatta A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQ LPAYKRFQITWVSWNCPDCVAKVIEFLAEHPNVTLTISTARLYYYWGR DWQRALCR	297
Macaca leonina A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQ LPAYKRFQITWVSWNCPDCVVKVIEFLAEHPNVTLTISTARLYYYWGR DWQRALCR	298
Macaca nemestrina A3B (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCFLSRFCGNQ LPAYKRFQITWVSWNCPDCVAKVTEFLAEHPNVTLTISTARLYYYWGR DWQRALCR	299
Macaca mulatta A3D (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYFQPQYHAEMCFLSWFCGNQ LPAYKRFQITWVSWNCPDCVAKVTEFLAEHPNVTLTISVARLYYYRGR DWRRALCR	300
Pongo abelii A3F (AA29-AA149)	RRNYTWLCYEVKIRKDP SKLAWDTGVFRGQVLPKLQSNHREVVYFEPQYH AEMCFLSWFCGNQLSAYERFQITWVSWTPCPDCVAMLAEF LAEHPNVTL TVSAARLYYYWERDYRGALRR	301
Erythrocebus patas	RRYTWLCYEVKIKKDP SKLPWDTGVFQGGVVRPKFQSNRRYEVYFQPQYHA	302

A3DE (AA30-AA149)	EMCFLSWFCGNQLPAYKHFQITWVSWNCPDCVAKVTEFLAEHPNVTLT ISAARLYYYWGKDWRRALCR	
Pan troglodytes A3B (AA1-AA79)	MYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWVSWTPCPDCVAKLAKFL AEHPNVTLTISAARLYYYWERDYRRALCR	303
Macaca mulatta A3DE (AA30-AA149)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVRPKLQSNRRYEVYFQPYHA EMCFLSWFCGNQLPAYKRFQITWVSWNCPDCVAKVTEFLAEHPNVTLT ISAARLYYYWGKDWRRALRR	304
Ptilocolobus tephrosceles A3F (AA30-AA137)	RRYTWLCYEVKIMKDH SKLPWYTG VFRGQVYFEPQNHAE MCFLSWFCGNQ LPAYECCQITWVSWTPCPDCVAKVTEFLAEHPNVTLTISAARLYYYRGR DWRRALRR	305
Macaca leonina A3D (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWYTG VFRGQVYFEPQYHAEMCFLSWFCGNQ LPANKRFQITWVSWNCPDCVAKVTEFLAEHPNVTLTISVARLYYYRGK DWRRALRR	306
Macaca nemestrina A3D (AA30-AA137)	RSYTWLCYEVKIRKDP SKLPWDTGVFRDQVYFEPQYHAEMCFLSWFCGNQ LPANKRFQITWVSWNCPDCVTKVTEFLAEHPNVTLTISVARLYYYRGK DWRRALRR	307
Chlorocebus aethiops A3DE (AA30-AA149)	RRYTWLCYEVKIKKDP SKLPWDTGVFPQVVRPKFQSNRRYEVYFQPYHA EMYFLSWFCGNQLPAYKHFQITWVSWNCPDCVAKVTEFLAEHRNVTLT ISAARLYYYWGKDWRRALCR	308
Macaca mulatta A3D (AA30-AA158)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQVRPKLQSNRRYELSNWECRKH VYFQPYHAEMCFLSWFCGNQLPANKRFQITWVSWNCPDCVAKVTEFL AEHPNVTLTISAARLYYYWGKDWRRALRR	309

Table B3: hA3BCDA1-Related Domains

Name	Sequence	SEQ ID NO:
Gorilla A3B (AA29-AA138)	GRSYNWLCYEVKIKRGRSNLLWNTGVFRGQMYSQPEHHAEMCFLSWFCGN QLPAYKCFQITWVSWTPCPDCVAKLAFLAEHPNVTLTISTARLYYYWE RDYRRALCRL	310
Pan paniscus A3B (AA29-AA138)	GRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMYFLSWFCGN QLPAYKCFQITWVSWTPCPDCVAKLAFLAEHPNVTLTISAARLYYYWE RDYRRALCRL	311
Pan troglodytes A3B (AA29-AA138)	GRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMCFLSWFCGN QLSAYKCFQITWVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWE RDYRRALCRL	312
Gorilla A3F (AA30-AA137)	RNTVWLCYEVKTKGFSRRLD KIFRGQVYFEPQYHAEMCFLSWFCGNQL PAYKCFQITWVSWTPCPDCVAKLAFLAEHPNVTLTISAARLYYYWE	313
Pan troglodytes A3F (AA30-AA137)	RNTVWLCYEVKTKGFSRRLD KIFRGQVYFEPQYHAEMCFLSWFCGNQL PAYKCFQITWVSWTPCPDCVAKLAFLAEHPNVTLTISAARLYYYWERD YRRALCRL	314
Human sapiens A3F (AA30-AA137)	RNTVWLCYEVKTKGFSRRLD KIFRGQVYSQPEHHAEMCFLSWFCGNQL PAYKCFQITWVSWTPCPDCVAKLAFLAEHPNVTLTISAARLYYYWERD YRRALCRL	315
Macaca leonine A3F (AA30-AA137)	RNTVWLCYEVKTRGFSMP TWGTKIFRGQVCFEPQYHAEMCFLSRFCGNQL PAYKRFQITWVSWTPCPDCVAKVAEFLAEHPNVTLTISAARLYYYWETD YRRALCRL	316
Macaca nemestrina A3F (AA30-AA137)	RNTVWLCYEVKTRGFSMP TWGTKIFRGQVCFEPQYHAEMCFLSRFCGNQL PAYKRFQITWVSWTPCPDCVAKVAEFLAEHPNVTLTISAARLYYYWETD YRRALCRL	317
Rhinopithecus roxellana A3F (AA30-AA137)	RNTVWLCYEVKTRGFSMP TWGAKIFRGQVYFEPQYHAEMCFLSWFCGNQL PAYKRFQITWVSWTPCPDCVAKVAEFLAEHPNVTLTISAARLYYYWETD YRRALCRL	318
Mandrillus leucophaeus A3F (AA30-AA130)	RNTVWLCYEVKTRGFSMP TWGTKIFRGQVYFEPQYHAEMCFLSWFCGNQL PAYKRFQITWVSWTPCPDCVVKVAEFLAEHPNVTLTISAARLYYYWETD Y	319
Macaca mulatta A3F (AA30-AA137)	RNTVWLCYEVKTRGFSMP TWDTKIFRGQVYSKPEHHAEMCFLSRFCGNQL PAYKRFQITWVSWTPCPDCVAKVAEFLAEHPNVTLTISAARLYYYWETD YRRALCRL	320
Theropithecus gelada A3F (AA30-AA137)	RNTVWLCYEVKTRGFSMP TWGTKIFRGQVYFEPQYHAEMCFLSRFCGNQL PAYKRFQITWVSWNCPDCVAKVIEFLAEHPNVTLTISAARLYYYWGRD WRRALRL	321

Cercocebus atys A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWYTG VFRGQVYSKPEHHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVIEFLAEHPNVTLTISAA RLYYYWS RDWQRALCRL	322
Macaca fascicularis A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVIEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	323
Macaca mulatta A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVIEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	324
Macaca leonina A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVIEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	325
Mandrillus leucophaeus A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWYTG VFRGQVYSKPEHHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVIEFLAEHPNVTLTIFTARLYYYWG RDWQRALCRL	326
Macaca nemestrina A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSKPEHHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVTEFLAEHPNVTLTISTARLYYYWG RDWQRALCRL	327
Rhinopithecus bieti A3F (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGN QLPAYKRFQITWFVSWTPCPDCVAKVAEFLTEHPNVTLTISAA RLYYYRG RDWRRALCRL	328
Rhinopithecus roxellana A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYSEPEHHAEMYFLSWFCGN QLPAYKRFQITWFVSWTPCPDCVAKVAEFLTEHPNVTLTISAA RLYYYRG RDWRRALCRL	329
Chlorocebus sabaeus A3B (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYSKPEHHAEMCF LSWFCGN QLPAYKRFQITWFVSWTPCPDCVAKVAEFLAEHPNVTLTISAA RLYYYWE TDYRRALCRL	330
Nomascus leucogenys A3B (AA30-AA138)	RSYTWLCYEVKIRKDP SKLPWDTGVFRGQMYFQPEYHAEMCF LSWFCGNQ LPAYKRFQITWFVSWTPCPDCVAKVAVFLAEHPNVTLTISAA RLYYYWEK DWQRALCRL	331
Cercocebus atys A3F (AA29-AA138)	GRSYTWLCYEVKIKYP SKLLWDTGVFQGGVYFQPYHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVTEFLAEHPNVTLTISAA RLYYYWE KDXRRALRRL	332
Papio anubis A3F (AA29-AA138)	GRSYTWLCYEVKIKEDP SKLLWDTGVFQGGVYFQPYHAEMCF LSRFCGN QLPAYKRFQITWFVSWNPCDCVAKVTEFLAEHPNVTLTISAA RLYYYWG RDWRRALRRL	333
Chlorocebus aethiops A3D (AA29-AA150)	GRRYTWLCYEVKIKDP SKLPWDTGVFPQGVPRPKFQSNRRYEVYFQPYH AEMYFLSWFCGNQLPAYKHFIQITWFVSWNPCDCVAKVTEFLAEHRNVTL TISAA RLYYYWGKDWRRALCRL	334
Chlorocebus sabaeus A3D (AA29-AA134)	GRRYTWLCYEVKIKDP SKLPWDTGVFPQGPYHAEMYFLSWFCGNQLPA YKHFIQITWFVSWNPCDCVAKVTEFLAEHRNVTLTISAA RLYYYWGKDW RRALCRL	335
Chlorocebus sabaeus A3F (AA29-AA150)	GRRYTWLCYEVKIKDP SKLPWDTGVFPQGVPRPKFQSNRRQKVYFQPYH AEMYFLSWFCGNQLPAYKHFIQITWFVSWNPCDCVAKVTEFLAEHRNVTL TISAA RLYYYWGKDWRRALCRL	336
Erythrocebus patas A3D (AA29-AA150)	GRRYTWLCYEVKIKDP SKLPWDTGVFPQGVPRPKFQSNRRYEVYFQPYH AEMCF LSWFCGNQLPAYKHFIQITWFVSWNPCDCVAKVTEFLAEHPNVTL TISAA RLYYYWGKDWRRALCRL	337
Macaca fascicularis A3D (AA29-AA159)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVPRPKLQSNRRYELSNWECK R VYFQPYHAEMYFLSWFCGNQLPANKRFQITWFASWNPCDCVAKVTEF LAEHPNVTLTISAA RLYYYRGKDWRRALRRL	338
Macaca fascicularis A3F (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYFQPYHAEMYFLSWFCGN QLPANKRFQITWFASWNPCDCVAKVTEFLAEHPNVTLTISAA RLYYYRG KDWRRALRRL	339
Macaca nemestrina A3D (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRDQVYFQPYHAEMCF LSWFCGN QLPANKRFQITWFVSWNPCDCVTKVTEFLAEHPNVTLTISAA RLYYYRG KDWRRALRRL	340
Macaca leonina A3D (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWYTG VFRGQVYFQPYHAEMCF LSWFCGN QLPANKRFQITWFVSWNPCDCVAKVTEFLAEHPNVTLTISAA RLYYYRG KDWRRALRRL	341
Macaca mulatta A3D (AA29-AA138)	GRSYTWLCYEVKIRKDP SKLPWDTGVFRGQVYFQPYHAEMCF LSWFCGN QLPAYKRFQITWFVSWNPCDCVAKVTEFLAEHPNVTLTISAA RLYYYRG KDWRRALCRL	342
Gorilla A3D (AA29-AA150)	GRSYTWLCYEVKIRGSSNLLWNTGVFRGPVPPKLQSNHRQEVYFQFENH AEMCF LSWFCGNRLPANRRFIQITWFVSWNPCCLPCVVKVTKFLAEHPNVTL TISAA RLYYYRDREWRRLRRL	343

Pan paniscus A3D (AA29-AA150)	GRSYTWLCYEVKIKRGCSNLIWDTGVFRGPVLPKLQSNHRQEVYFQFENH AEMCFFSWFCGNRLPANRRFQITWVSVWNPCLPCVVKVTKFLAEHPNVTL TISAARLYYYQDREWRRVLRL	344
Pan troglodytes A3D (AA29-AA150)	GRSYTWLCYEVKIKRGCSNLIWDTGVFRGPVLPKLQSNHRQEVYFQFENH AEMCFFSWFCGNRLPANRRFQITWVSVWNPCLPCVVKVTKFLAEHPNVTL TISAARLYYYQDREWRRVLRL	345
Homo sapiens A3D (AA29-AA150)	GRSYTWLCYEVKIKRGCSNLIWDTGVFRGPVLPKLQSNHRQEVYFQFENH AEMCFLSWFCGNRLPANRRFQITWVSVWNPCLPCVVKVTKFLAEHPNVTL TISAARLYYYRDRDWRWVLLRL	346
Nomascus leucogenys A3D (AA29-AA150)	GRSYTWLCYEVKIKRDKPSKLPWDKGVFRGQVLPKQSNHRQEVYFQLENH AEMCFLSWFCGNQLPANRRFQITWVSVWNPCLPCVAKVTEFLAEHPNVTL TISAARLYYYRGRDWRRALRL	347
Saimiri boliviensis A3C (AA29-AA138)	GKKYTWLCYEVKIKKDTSKLPWNITGVFRGQVNFNPEHHAEMYFLSWFRGK LLPACKRSQITWVSVWNPCLYCVAKVAEFLAEHPNVTLTIVSTARLYCYWK KDWRRALRKL	348
Saimiri boliviensis A3F (AA29-AA138)	GKKYTWLCYEVKIKKDTSKLPWNITGVFRGQVNFNPEHHAEMYFLSWFRGK LLPACKRSQITWVSVWNPCLYCVAKVAEFLAEHPNVTLTIVSTARLYCYWK KDWRRALRKL	349
Piliocolobus tephrosceles A3F (AA36-AA145)	GRRYTWLCYEVKIMKDHSLPWPYTGIVFRGQVYFEPQNHAEMLSWFCGN QLPAYECCQITWVSWTPCPDCVAKVTEFLAEHPNVTLTISAARLYYYRG RDWRRALRL	350
Colobus angolensis palliatus A3F (AA29-AA138)	GRRYTWLCYEVKISKDPSKLPWDITGIVFRGQVYFEPQYHAEMCFLSWYCGN QLPAYKCFQITWVSWTPCPDCVGKVAEFLAEHPNVTLTISAARLYYYWE TDYRRALCRL	351
Pongo abelii A3F (AA30-AA150)	RNYTWLCYEVKIRKDPKSLAWDTGVFRGQVLPKLQSNHRREVYFEPQYHA EMCFLSWFCGNQLSAYERFQITWVSWTPCPDCVAMLAFLAEHPNVTLT VSAARLYYYWERDYRGALRL	352

[0119] In some embodiments, the protease cleavage site is a self-cleaving peptide, such as the 2A peptides. “2A peptides” are 18-22 amino-acid-long viral oligopeptides that mediate “cleavage” of polypeptides during translation in eukaryotic cells. The designation “2A” refers to a specific region of the viral genome and different viral 2As have generally been named after the virus they were derived from. The first discovered 2A was F2A (foot-and-mouth disease virus), after which E2A (equine rhinitis A virus), P2A (porcine teschovirus-1 2A), and T2A (thossea asigna virus 2A) were also identified. A few non-limiting examples of 2A peptides are provided in SEQ ID NO:217-219.

[0120] In some embodiments, the protease cleavage site is a cleavage site (e.g., SEQ ID NO:196) for the TEV protease. In some embodiments, the TEV protease provided in the base editing system includes two separate fragments, each of which on its own is not active. However, in the presence of the remaining fragment of the TEV protease, they will be able to execute the cleavage. Such an arrangement provides additional control and flexible of the base editing capabilities. The TEV fragments may be the TEV N-terminal domain (e.g., SEQ ID NO:194) or the TEV C-terminal domain (e.g., SEQ ID NO:195).

[0121] Various arrangement of the proteins/fragments can be made for a fusion protein in the disclosed base editing systems. Non-limiting examples include, from N-terminal side to C-terminal side:

- (1) first fragment (e.g., catalytic domain) – protease cleavage site – second fragment (e.g., inhibitory domain);
 - (2) first fragment (e.g., catalytic domain and Cas protein) – protease cleavage site – second fragment (e.g., inhibitory domain);
 - (3) first fragment (e.g., catalytic domain, Cas protein and TEV N-terminal domain) – protease cleavage site (e.g., TEV cleavage site) – second fragment (e.g., inhibitory domain);
 - (4) second fragment (e.g., inhibitory domain) – protease cleavage site (e.g., TEV cleavage site) – first fragment (e.g., catalytic domain, Cas protein and TEV N-terminal domain);
- and
- (5) second fragment (e.g., inhibitory domain) – protease cleavage site (e.g., TEV cleavage site) – first fragment (e.g., Cas protein, catalytic domain, and TEV C-terminal domain).

[0122] Such fusion proteins may include other fragments, such as uracil DNA glycosylase inhibitor (UGI) and nuclear localization sequences (NLS).

[0123] The “Uracil Glycosylase Inhibitor” (UGI), which can be prepared from *Bacillus subtilis* bacteriophage PBS1, is a small protein (9.5 kDa) which inhibits *E. coli* uracil-DNA glycosylase (UDG) as well as UDG from other species. Inhibition of UDG occurs by reversible protein binding with a 1:1 UDG:UGI stoichiometry. UGI is capable of dissociating UDG-DNA complexes. A non-limiting example of UGI is found in *Bacillus phage AR9* (YP_009283008.1). In some embodiments, the UGI comprises the amino acid sequence of SEQ ID NO:216 or has at least at least 70%, 75%, 80%, 85%, 90% or 95% sequence identity to SEQ ID NO:216 and retains the uracil glycosylase inhibition activity.

[0124] The fusion protein, in some embodiments, may include one or more nuclear localization sequences (NLS).

[0125] A “nuclear localization signal or sequence” (NLS) is an amino acid sequence that tags a protein for import into the cell nucleus by nuclear transport. Typically, this signal consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface. Different nuclear localized proteins may share the same NLS. An NLS has

the opposite function of a nuclear export signal (NES), which targets proteins out of the nucleus. A non-limiting example of NLS is the internal SV40 nuclear localization sequence (iNLS).

[0126] In some embodiments, a peptide linker is optionally provided between each of the fragments in the fusion protein. In some embodiments, the peptide linker has from 1 to 100 amino acid residues (or 3-20, 4-15, without limitation). In some embodiments, at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the amino acid residues of peptide linker are amino acid residues selected from the group consisting of alanine, glycine, cysteine, and serine.

Nucleobase Deaminase Inhibitors Fusion Proteins, tBEs

[0127] As demonstrated in the experimental examples, hA3F-CDA1 has been identified as an excellent cytidine deaminase inhibitor. Analogs of hA3F-CDA1 are shown in Table B2, as well as those having at least 70%, 75%, 80%, 85% 90%, 95%, 97%, 98%, or 99% sequence identity to hA3F-CDA1 or any of those in Table B2.

[0128] Accordingly, a fusion protein is designed that can be used to generate a base editor with improved base editing specificity and efficiency. In one embodiment, the present disclosure provides a fusion protein that includes a first fragment comprising a nucleobase deaminase (e.g., a cytidine deaminase) or a catalytic domain thereof, a second fragment comprising a nucleobase deaminase inhibitor, and a protease cleavage site between the first fragment and the second fragment. In some embodiments, the nucleobase deaminase inhibitor is hA3F-CDA1 (SEQ ID NO:192), or any of its analogs, such as those shown in Table B2, as well as those having at least 70%, 75%, 80%, 85% 90%, 95%, 97%, 98%, or 99% sequence identity to hA3F-CDA1 or any of those in Table B2.

[0129] A base editor that incorporates such a fusion protein has reduced or even no editing capability and accordingly will generate reduced or no off-target mutations. Upon cleavage of the protease cleavage site and release of the nucleobase deaminase inhibitor from the fusion protein at a target site, the base editor that is at the target site will then be able to edit the target site efficiently.

[0130] In some embodiments, the fusion protein further includes a clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) protein, optionally in the first fragment, next to the nucleobase deaminase or the catalytic domain thereof.

[0131] When the fusion protein is used, *in vitro*, *ex vivo*, or *in vivo*, to conduct gene/base editing in a cell, two additional molecules can be introduced. In one example, one molecule (B) is a single guide RNA (sgRNA) that further incorporates a tag sequence that can be recognized by an RNA recognition peptide. The sgRNA, alternatively, can be replaced by a crRNA that targets the target site and a CRISPR RNA (crRNA) alone, or in combination with a trans-activating CRISPR RNA (tracrRNA). Examples of tag sequences and corresponding RNA recognition peptides include MS2/MS2 coat protein (MCP), PP7/PP7 coat protein (PCP), and boxB/boxB coat protein (N22p), the sequences of which are provided herein. The molecule (B) may be provided as a DNA sequence encoding the RNA molecule.

[0132] The other additional molecule (C), in some embodiments, includes a second TEV protease fragment coupled to the RNA recognition peptide (e.g., MCP, PCP, N22p). The first TEV fragment and the second TEV fragment, in some embodiments, when present together, are able to cleave a TEV protease site.

[0133] Such co-presence can be triggered by the molecule (C) binding to the molecule (B) by virtue of the tag sequence-RNA recognition protein interaction. Meanwhile, the fusion protein (A) and the molecule (B) will be both present at the target genome locus for gene editing. Therefore, the molecule (B) brings both of the TEV protease fragments from the fusion protein (A) and molecule (C) together, which will activate the TEV protease, leading to removal of the nucleobase deaminase inhibitor from the fusion protein and activation of the base editor. It can be readily appreciated that such activation only occurs at the target genome site, not at off-target single-stranded DNA regions. As such, base editing does not occur at the single-stranded DNA regions that sgRNA does not bind to.

Gene Therapy

[0134] The disclosed base editing system can be used to engineer a target cell. If used *in vitro* or *ex vivo*, the gene therapy approach can increase the expression of γ -globin in the target cell. If used *in vivo*, the gene therapy approach can treat diseases associated with insufficient production or dysfunction of hemoglobins. Example diseases include β -thalassemia, sickle cell anemia, Haemoglobin C, and Haemoglobin E.

[0135] In some embodiments, each component of the base editing system can be introduced to the target cell individually, or in combination. For instance, a fusion protein may be packaged into nanoparticle such as liposome. In another example, a guide RNA and a protein may be combined into a complex for introduction.

[0136] In some embodiments, some or all of the components of the base editing system can be introduced as one or more polynucleotides encoding them. These polynucleotides may be constructed as plasmids or viral vectors, without limitation.

[0137] In an example *ex vivo* approach, CD34+ hematopoietic stem and progenitor cells (HSPCs) can be collected from a patient by apheresis after mobilization with either filgrastim and plerixafor (in a patient with β -thalassemia) or plerixafor alone (in a patient with SCD) after a minimum of 8 weeks of transfusions of packed red cells targeting a level of sickle hemoglobin of less than 30%. The HSPCs can then be edited with the disclosed gene editing technology, along with the designed sgRNA, to produce edited cells. DNA sequencing can be used to evaluate the percentage of allelic editing at the on-target site.

[0138] Prior to infusion of the edited cells, the patient can be given a pharmacokinetically adjusted busulfan myeloablation. The edited cells can be administered through intravenous infusion.

EXAMPLES

Example 1. Base Editing at BCL11A/gamma-Globin

[0139] This example tested a newly designed base editor, referred to as transformer Base Editor (tBE), which can specifically edit cytosines in target regions with no observable off-target mutations, to edit the BCL11A gene which is useful for treating β -hemoglobinopathies.

[0140] The tBE fuses a base editor with a cytidine deaminase inhibitor to inhibit the activity of the cytidine deaminase until the tBE complex is assembled at the target genomic site. The tBE employs a sgRNA (about 20 nt) to bind at the target genomic site and a helper sgRNA (hsgRNA, 10-20 nt) to bind at a nearby region upstream to the target genomic site. The binding of two sgRNAs can guide the components of tBE to correctly assemble at the target genomic site for efficient base editing.

[0141] To test whether the tBE can perform high-specificity and high-efficiency base editing in BCL11A erythroid enhancer region, we designed 45 pairs (5 sgRNA x 9 hsgRNA, as listed in **Tables 1 and 2**) of sgRNA/hsgRNAs to target the BCL11A erythroid enhancer region (the core BCL11A erythroid enhancer). For comparison, we co-transfected the sgRNAs in sgRNA/hsgRNA pairs with a previously reported BE, YE1-BE4max and a single sgRNA targeting the same genomic site with tBE (**FIG. 1-4**).

Table 1. sgRNA targeting BCL11A erythroid enhancer

sgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA-BCL11A-1	cuuuuauacac	1	cuaacaguugcuuuuauacac	2
sgRNA-BCL11A-2	cacaggcucc	3	uugcuuuuauacacaggcucc	4
sgRNA-BCL11A-3	ggcuccagga	5	uuuuauacacaggcuccagga	6
sgRNA-BCL11A-4	gcuccaggaa	7	uuuauacacaggcuccaggaa	8
sgRNA-BCL11A-5	aggaaggguu	9	cacaggcuccaggaaggguu	10

Table 2. hsgRNA targeting BCL11A erythroid enhancer

hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
hsgRNA-BCL11A-1	uaacacacca	11	cucuuaagacauaacacacca	12
hsgRNA-BCL11A-2	auaacacacc	13	acucuuaagacauaacacacc	14
hsgRNA-BCL11A-3	aaacaacuu	15	caccagggucaauacaacuu	16
hsgRNA-BCL11A-4	acaacuuuga	17	cagggucaauacaacuuuga	18
hsgRNA-BCL11A-5	cuuugaagcu	19	gucaauacaacuuugaagcu	20
hsgRNA-BCL11A-6	aagcuagucu	21	uacaacuuugaagcuagucu	22
hsgRNA-BCL11A-7	gcuagucuag	23	caacuuugaagcuagucuag	24
hsgRNA-BCL11A-8	gucuagugca	25	uuugaagcuagucuagugca	26
hsgRNA-BCL11A-9	gcaagcuaac	27	cuagucuagugcaagcuaac	28

[0142] We extracted genomic DNA 72 hours after transfecting the plasmids into cells, and compared the C-to-T editing efficiencies of these BEs at target sites. From Sanger sequencing results, we found both tBE and YE1-BE4max induced gene editing in BCL11A erythroid enhancer region. It's worth noting that tBE induced similar or higher base editing efficiencies than YE1-BE4max at some target sites, such as the target sites for sgRNA-BCL11A-3/hsgRNA-BCL11A-2 (**FIG. 2B**) and sgRNA-BCL11A-4/hsgRNA-BCL11A-2 (**FIG. 3B**). These results indicate that tBE could perform highly efficient base editing in BCL11A erythroid enhancer region.

[0143] Next, we tested whether the tBE can perform high-specificity and high-efficiency base editing at the BCL11A binding motif in HBG1/2 promoters. We designed 3 pairs (1 sgRNA x 3 hsgRNA, as listed in **Tables 3 and 4**) of sgRNA/hsgRNA to target the BCL11A binding motif in HBG1/2 promoters (**FIG. 5**). After 72 hours of plasmid transfection, we extracted genomic DNA from transfected cells. From the sanger sequencing results, we found both tBE and YE1-BE4max induced gene editing in the BCL11A binding motif of HBG1/2 promoters. Also, tBE induced similar or higher base editing efficiencies than YE1-BE4max at some target sites, such as the target sites for sgRNA-HBG/hsgRNA-HBG-1/2 (**FIG. 5B**).

Table 3. sgRNA targeting HBG1/2 promoters

sgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA-HBG	agccuugaca	29	cuugaccaauagccuugaca	30

Table 4. hsgRNA targeting HBG1/2 promoters

hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
hsgRNA-HBG-1	acuccaccca	31	cccuggcuaaacuccaccca	32
hsgRNA-HBG-2	cuccacccau	33	ccuggcuaaacuccacccau	34
hsgRNA-HBG-3	acccaugggu	35	gcuaaacuccacccaugggu	36

[0144] We then tested whether the tBE can perform high-specificity and high-efficiency at the BCL11A erythroid enhancer and HBG1/2 promoter regions simultaneously by modifying the plasmids of tBE system (**FIG. 6A, 7A, 8A, 9A, 10A, 11A, 12A, 13A, 14A, 15A, 16A, 17A**). Based on the experimental results obtained at these two sites, we chose the pairs of sgRNA/hsgRNA with relatively high editing efficiency (sgRNA-HBG/hsgRNA-HBG-1/2/3, sgRNA-BCL11A-3/hsgRNA-BCL11A-1/2 and sgRNA-BCL11A-4/hsgRNA-BCL11A-1/2) and transfected these plasmids in various combinations. After 72 hours of plasmid transfection, we extracted genomic DNA from transfected cells. From the sanger sequencing results, we found that the co-transfection of tBE with sgRNA-HBG/hsgRNA-HBG-2 and sgRNA-BCL11A-4/hsgRNA-BCL11A-1 induced the highest C-to-T mutation frequencies at both target sites (**FIG. 13B**).

[0145] This example used a highly precise and efficient base editing system (tBE) to perform base editing at the therapeutic genomic sites of the β -hemoglobinopathies. Furthermore, the tBE system, which contains Cas9 nickase (D10A), is less toxic than Cas9 nuclease as Cas9 nickase activates a lower level of p53 pathway than Cas9 nuclease. In addition, this example

achieved high specificity and efficiency base editing individually or simultaneously at two therapeutic target sites, which can reactive a high expression level of γ -globin. This example therefore demonstrates a clinical use of tBE, especially in the gene therapies of the β -hemoglobinopathies.

Example 2. Base Editing at Other Sites Impacting gamma-Globin Expression

[0146] The expression of BCL11A may be impacted by other cis elements or protein factors. There are three DNase I hypersensitive sites (DHSs), referred to as DHSs +62, +58, and +55 based on distance in kilobases from the transcription start site (TSS) of BCL11A. KLF1 is a key erythroid transcription factor that activates BCL11A directly by binding BCL11A's promoter. Furthermore, there is a GATA1-binding motif located in intron 4 of the NFIX gene, which could regulate the expression of BCL11A indirectly by influencing the expression of KLF1. In addition, ZBTB7A (zinc finger and BTB domain containing 7A), a repressor protein, could bind the HBG1/2 enhancer/promoter by identifying a conserved motif and repress the expression of HBG.

[0147] To test whether the tBE can perform high-specificity and high-efficiency base editing at the three KLF1 binding motifs of BCL11A (two core KLF1 binding motifs locate in +55 kb DHS of BCL11A erythroid enhancer region, the other one locates in 1 Mb upstream of BCL11A). We designed 50 pairs of sgRNA/hsgRNAs (Tables 5-7) to target the three KLF1 binding motifs (**FIG. 18-26**), and extracted genomic DNA 72 hours after transfecting them with tBE into cells. Sanger sequencing results (**FIG. 18-26**) demonstrate that the tBE, with the designed sgRNA/hsgRNA, efficiently induced gene editing at the three KLF1 binding motifs of BCL11A, which can be useful for activating the expression of the γ -globin gene. And, those bold marked hsgRNAs together with their corresponding sgRNAs worked efficiently at their target sites.

Table 5. sgRNA/hsgRNA targeting KLF1-binding motif 1 of BCL11A

Table 5.1 sgRNA-KLF1-1-1 and its hsgRNAs targeting KLF1-binding motif 1 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-1-1	gugaucuugu	37	ggcacacccugugaucuugu	38
hsgRNA	hsgRNA-KLF1-1-4	caccuucuca	55	gugagcucucccaccuucuca	56
	hsgRNA-KLF1-1-5	caaugcuugg	57	caggauaugcaaugcuugg	58
	hsgRNA-KLF1-1-6	augcaaugcu	59	uaccaggaugcaaugcu	60

	hsgRNA-KLF1-1-7	uccugguacc	61	cccaugccuuccugguacc	62
--	-----------------	------------	----	---------------------	----

Table 5.2 sgRNA-KLF1-1-2 and its hsgRNAs targeting KLF1-binding motif 1 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-1-2	ugugaucuug	39	gggcacacccugugaucuug	40
hsgRNA	hsgRNA-KLF1-1-4	caccuucuca	63	gugagcucacccaccuucuca	64
	hsgRNA-KLF1-1-5	caaugcuugg	65	caggauaugcaaugcuugg	66
	hsgRNA-KLF1-1-6	augcaaugcu	67	uaccaggaugaugcaaugcu	68
	hsgRNA-KLF1-1-7	uccugguacc	69	cccaugccuuccugguacc	70

Table 5.3 sgRNA-KLF1-1-3 and its hsgRNAs targeting KLF1-binding motif 1 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-1-3	ugagaaggug	41	gggugugcccugagaaggug	42
hsgRNA	hsgRNA-KLF1-1-8	ggcuggacag	71	accaggcugggcuggacag	72
	hsgRNA-KLF1-1-9	caggcugggc	73	gaugcacaccaggcugggc	74
	hsgRNA-KLF1-1-10	caccaggcu	75	acaagaugcacaccaggcu	76
	hsgRNA-KLF1-1-11	acaccaggc	77	cacaagaugcacaccaggc	78
	hsgRNA-KLF1-1-12	augcacaccc	79	agcacacaagaugcacaccc	80

Table 6. sgRNA/hsgRNA targeting KLF1-binding motif 2 of BCL11A

Table 6.1 sgRNA-KLF1-2-1 and its hsgRNAs targeting KLF1-binding motif 2 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-2-1	augcacaccc	43	agcacacaagaugcacaccc	44
hsgRNA	hsgRNA-KLF1-2-1	gaccgcucac	81	cagccuuggggaccgcucac	82
	hsgRNA-KLF1-2-2	cacagccuug	83	gaaggcugggcacagccuug	84
	hsgRNA-KLF1-2-3	cucccuaccg	85	gugccgacaacucccuaccg	86
	hsgRNA-KLF1-2-10	gacccuauuc	99	ucccuaccgcgacccuauuc	100
	hsgRNA-KLF1-2-11	ccccuauucag	101	ccuaccgcgacccuauucag	102
	hsgRNA-KLF1-2-12	cuaucagugc	103	accgcgacccuauucagugc	104

Table 6.2 sgRNA-KLF1-2-2 and its hsgRNAs targeting KLF1-binding motif 2 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-2-2	caggcugggc	45	gaugcacaccaggcugggc	46
hsgRNA	hsgRNA-KLF1-2-1	gaccgcucac	81	cagccuuggggaccgcucac	82
	hsgRNA-KLF1-2-2	cacagccuug	83	gaaggcugggcacagccuug	84
	hsgRNA-KLF1-2-3	cucccuaccg	85	gugccgacaacucccuaccg	86
	hsgRNA-KLF1-2-10	gacccuauuc	99	ucccuaccgcgacccuauuc	100

	hsgRNA-KLF1-2-11	ccccauacag	101	ccuaccgcgaccccuacag	102
	hsgRNA-KLF1-2-12	cuaucagugc	103	accgcgaccccuacagugc	104

Table 6.3 sgRNA-KLF1-2-3 and its hsgRNAs targeting KLF1-binding motif 2 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-2-3	ggcuggacag	47	accaggcugggacgacag	48
hsgRNA	hsgRNA-KLF1-2-1	gaccgcucac	81	cagccuuggggacgcucac	82
	hsgRNA-KLF1-2-2	cacagccuug	83	gaaggcugggcacagccuug	84
	hsgRNA-KLF1-2-3	cucccuaccg	85	gugccgacaacucccuaccg	86
	hsgRNA-KLF1-2-10	gaccccuac	99	ucccuaccgcgaccccuac	100
	hsgRNA-KLF1-2-11	ccccauacag	101	ccuaccgcgaccccuacag	102
	hsgRNA-KLF1-2-12	cuaucagugc	103	accgcgaccccuacagugc	104

Table 6.4 sgRNA-KLF1-2-4 and its hsgRNAs targeting KLF1-binding motif 2 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-2-4	ugugcuuggu	49	gugcaucuugugugcuuggu	50
hsgRNA	hsgRNA-KLF1-2-4	gugaucuugu	87	ggcacaccucugaucuugu	88
	hsgRNA-KLF1-2-5	ugugaucuug	89	gggcacaccucugaucuug	90
	hsgRNA-KLF1-2-6	caccuucuca	91	gugagcucucccaccuucuca	92
	hsgRNA-KLF1-2-7	caaugcuugg	93	cagggaugaugcaaugcuugg	94
	hsgRNA-KLF1-2-8	augcaaugcu	95	uaccagggaugaugcaaugcu	96
	hsgRNA-KLF1-2-9	uccugguacc	97	cccauugccuuccugguacc	98

Table 7. sgRNA/hsgRNA targeting KLF1-binding motif 3 of BCL11A

Table 7.1 sgRNA-KLF1-3-1 and its hsgRNAs targeting KLF1-binding motif 3 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-3-1	cccaccugc	51	ucaccaggaccccaccugc	52
hsgRNA	hsgRNA-KLF1-3-1	cagccaggac	105	ggccaucugccagccaggac	106
	hsgRNA-KLF1-3-2	ucugccagcc	107	aagguggccaucugccagcc	108
	hsgRNA-KLF1-3-3	agcaagaagg	109	gugaaaacgcagcaagaagg	110
	hsgRNA-KLF1-3-4	cgcagcaaga	111	caugugaaaacgcagcaaga	112
	hsgRNA-KLF1-3-5	aaacgcagca	113	ugccaugugaaaacgcagca	114
	hsgRNA-KLF1-3-6	gugaaaacgc	115	ucucugccaugugaaaacgc	116

Table 7.2 sgRNA-KLF1-3-2 and its hsgRNAs targeting KLF1-binding motif 3 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-KLF1-3-2	gcugaaaccc	53	acccaccugcugaaaccc	54

hsgRNA	hsgRNA-KLF1-3-1	cagccaggac	105	ggccaucugccagccaggac	106
	hsgRNA-KLF1-3-2	ucugccagcc	107	aagguggccaucugccagcc	108
	hsgRNA-KLF1-3-3	agcaagaagg	109	gugaaaacgcagcaagaagg	110
	hsgRNA-KLF1-3-4	cgcagcaaga	111	caugugaaaacgcagcaaga	112
	hsgRNA-KLF1-3-5	aaacgcagca	113	ugccaugugaaaacgcagca	114
	hsgRNA-KLF1-3-6	gugaaaacgc	115	ucucugccaugugaaaacgc	116

[0148] We also tested whether the tBE can perform base editing at the GATA1-binding motif located in intron 4 of the NFIX gene. We designed more than 20 pairs of sgRNA/hsgRNAs to target the GATA1-binding motif (FIG. 27-28) and extracted genomic DNA 72 hours after transfecting them with tBE into cells. Sanger sequencing results (FIG. 27-28) demonstrate that the tBE, with the designed sgRNA/hsgRNA, efficiently induced gene editing at the GATA1-binding motif located in the NFIX gene, which can be useful for activating the expression of the γ -globin gene.

Table 8. sgRNA/hsgRNA targeting GATA1-binding motif of NFIX

Table 8.1 sgRNA-GATA-1 and its hsgRNAs targeting GATA1-binding motif of NFIX

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-GATA-1	gguggcacac	117	ccagcuauaggguggcacac	118
hsgRNA	hsgRNA-GATA-1	cacagcuggu	123	gacagcugugcacagcuggu	124
	hsgRNA-GATA-2	ugugcacagc	125	uggggacagcugugcacagc	126
	hsgRNA-GATA-3	ugcggccaug	127	ggaggcacugugcggccaug	128
	hsgRNA-GATA-4	ugugcggccca	129	ggggaggcacugugcggccca	130
	hsgRNA-GATA-5	ggaacagcug	131	ugcaggacaggggaacagcug	132
	hsgRNA-GATA-6	agggaaacagc	133	gcugcaggacagggaaacagc	134
	hsgRNA-GATA-7	gcugcaggac	135	aagcagcccagcugcaggac	136
	hsgRNA-GATA-8	gccagcugc	137	caauuaagcagcccagcugc	138

Table 8.2 sgRNA-GATA-2 and its hsgRNAs targeting GATA1-binding motif of NFIX

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-GATA-2	ggcacacagc	119	gcuaucagguggcacacagc	120
hsgRNA	hsgRNA-GATA-1	cacagcuggu	123	gacagcugugcacagcuggu	124
	hsgRNA-GATA-2	ugugcacagc	125	uggggacagcugugcacagc	126
	hsgRNA-GATA-3	ugcggccaug	127	ggaggcacugugcggccaug	128
	hsgRNA-GATA-4	ugugcggccca	129	ggggaggcacugugcggccca	130
	hsgRNA-GATA-5	ggaacagcug	131	ugcaggacaggggaacagcug	132
	hsgRNA-GATA-6	agggaaacagc	133	gcugcaggacagggaaacagc	134

	hsgRNA-GATA-7	gcugcaggac	135	aagcagcccagcugcaggac	136
	hsgRNA-GATA-8	gccagcugc	137	caauaagcagcccagcugc	138

Table 8.3 sgRNA-GATA-3 and its hsgRNAs targeting GATA1-binding motif of NFIX

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-GATA-3	acacagcugc	121	aucagguggcacacagcugc	122
hsgRNA	hsgRNA-GATA-1	cacagcuggu	123	gacagcugugcacagcuggu	124
	hsgRNA-GATA-2	ugugcacagc	125	uggggacagcugugcacagc	126
	hsgRNA-GATA-3	ugcgcccaug	127	ggaggcacugugcgcccaug	128
	hsgRNA-GATA-4	ugugcgccca	129	ggggaggcacugugcgccca	130
	hsgRNA-GATA-5	ggaacagcug	131	ugcaggacagggaaacagcug	132
	hsgRNA-GATA-6	agggaacagc	133	gcugcaggacagggaaacagc	134
	hsgRNA-GATA-7	gcugcaggac	135	aagcagcccagcugcaggac	136
	hsgRNA-GATA-8	gccagcugc	137	caauaagcagcccagcugc	138

[0149] In addition, we tested whether the tBE can perform base editing at the two ZBTB7A-binding motifs located in the HBG1/2 promoter/enhancer. We designed 41 pairs of sgRNA/hsgRNAs to target the ZBTB7A-binding motifs (**FIG. 29-33**) and extracted genomic DNA 72 hours after transfecting them with tBE into cells. Sanger sequencing results (**FIG. 29-33**) demonstrate that the tBE, with the designed sgRNA/hsgRNA, efficiently induced gene editing at the two ZBTB7A-binding motifs located in HBG1/2 promoter/enhancer, which can be useful for activating the expression of the γ -globin gene.

Table 9. sgRNA/hsgRNA targeting ZBTB7A-binding motif 1 of HBG1/2

Table 9.1 sgRNA-ZBTB7A-1-1 and its hsgRNAs targeting ZBTB7A-binding motif 1 of HBG1/2

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-ZBTB7A-1-1	ccucccgug	139	auuucucuuccucccgug	140
hsgRNA	hsgRNA-ZBTB7A-1-1	agaagcagag	151	uuuccuuaucaagaagcagag	152
	hsgRNA-ZBTB7A-1-2	uuuccuuau	153	auaaaauuuuuccuuau	154
	hsgRNA-ZBTB7A-1-3	ucauaagagc	155	uuagaugagcucuaagagc	156
	hsgRNA-ZBTB7A-1-4	aaaaguaauu	157	gaggcuuuugaaaaguaauu	158
	hsgRNA-ZBTB7A-1-5	ucugugggg	159	acugaccuuauucugugggg	160
	hsgRNA-ZBTB7A-1-6	uuauucugug	161	ccaacugaccuuauucugug	162
	hsgRNA-ZBTB7A-1-7	accuuauucg	163	cucccaacugaccuuauucg	164

Table 9.2 sgRNA-ZBTB7A-1-2 and its hsgRNAs targeting ZBTB7A-binding motif 1 of HBG1/2

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-ZBTB7A-1-2	gugaggacac	141	uuuccuccggugaggacac	142
hsgRNA	hsgRNA-ZBTB7A-1-1	agaagcagag	151	uuuccuuaucaagaagcagag	152
	hsgRNA-ZBTB7A-1-2	uuuccuuau	153	auaaaauuuuuuccuuau	154
	hsgRNA-ZBTB7A-1-3	ucauaagagc	155	uuagaugagcucuaaagagc	156
	hsgRNA-ZBTB7A-1-4	aaaaguaauu	157	gaggcuuuugaaaaguaauu	158
	hsgRNA-ZBTB7A-1-5	ucuguggggg	159	acugaccuuauucuguggggg	160
	hsgRNA-ZBTB7A-1-6	uuauucugug	161	ccaacugaccuuauucugug	162
	hsgRNA-ZBTB7A-1-7	accuuauucg	163	cucccaacugaccuuauucg	164

Table 9.3 sgRNA-ZBTB7A-1-3 and its hsgRNAs targeting ZBTB7A-binding motif 1 of HBGI/2

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-ZBTB7A-1-3	gaggacacag	143	uccuccggugaggacacag	144
hsgRNA	hsgRNA-ZBTB7A-1-1	agaagcagag	151	uuuccuuaucaagaagcagag	152
	hsgRNA-ZBTB7A-1-2	uuuccuuau	153	auaaaauuuuuuccuuau	154
	hsgRNA-ZBTB7A-1-3	ucauaagagc	155	uuagaugagcucuaaagagc	156
	hsgRNA-ZBTB7A-1-4	aaaaguaauu	157	gaggcuuuugaaaaguaauu	158
	hsgRNA-ZBTB7A-1-5	ucuguggggg	159	acugaccuuauucuguggggg	160
	hsgRNA-ZBTB7A-1-6	uuauucugug	161	ccaacugaccuuauucugug	162
	hsgRNA-ZBTB7A-1-7	accuuauucg	163	cucccaacugaccuuauucg	164

Table 9.4 sgRNA-ZBTB7A-1-4 and its hsgRNAs targeting ZBTB7A-binding motif 1 of HBGI/2

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-ZBTB7A-1-4	ggacacagug	145	cucccgugaggacacagug	146
hsgRNA	hsgRNA-ZBTB7A-1-1	agaagcagag	151	uuuccuuaucaagaagcagag	152
	hsgRNA-ZBTB7A-1-2	uuuccuuau	153	auaaaauuuuuuccuuau	154
	hsgRNA-ZBTB7A-1-3	ucauaagagc	155	uuagaugagcucuaaagagc	156
	hsgRNA-ZBTB7A-1-4	aaaaguaauu	157	gaggcuuuugaaaaguaauu	158
	hsgRNA-ZBTB7A-1-5	ucuguggggg	159	acugaccuuauucuguggggg	160
	hsgRNA-ZBTB7A-1-6	uuauucugug	161	ccaacugaccuuauucugug	162
	hsgRNA-ZBTB7A-1-7	accuuauucg	163	cucccaacugaccuuauucg	164

Table 9.5 sgRNA-ZBTB7A-1-5 and its hsgRNAs targeting ZBTB7A-binding motif 1 of HBGI/2

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-ZBTB7A-1-5	gaaagagaaa	147	ucaccgggaggaaagagaaa	148
hsgRNA	hsgRNA-ZBTB7A-1-8	uugcagaugg	165	ucuccuggauugcagaugg	166
	hsgRNA-ZBTB7A-1-9	ggauugcaga	167	uucucuuccggauugcaga	168
	hsgRNA-ZBTB7A-1-10	guucucuucc	169	cguggucagggguucucuucc	170

	hsgRNA-ZBTB7A-1-11	cucgugguca	171	ugaaggcugacucgugguca	172
	hsgRNA-ZBTB7A-1-12	acucgugguc	173	cugaaggcugacucgugguc	174
	hsgRNA-ZBTB7A-1-13	ggcugacucg	175	cauuucugaaggcugacucg	176
	hsgRNA-ZBTB7A-1-14	acauuucuga	177	guuuuucucacauuucuga	178

Table 10. sgRNA/hsgRNA targeting ZBTB7A-binding motif 2 of HBGI/2

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-ZBTB7A-2	acuaucuaa	149	ccuuccccacacuaucuaa	150
hsgRNA	hsgRNA-ZBTB7A-2-1	auccucuu	179	aaauagcaguauccucuu	180
	hsgRNA-ZBTB7A-2-2	aguauccuc	181	aaaaauagcaguauccuc	182
	hsgRNA-ZBTB7A-2-3	aaaaaaaauu	183	caaaggcuauaaaaaaaauu	184
	hsgRNA-ZBTB7A-2-4	aacaaggcaa	185	acugaaucggaacaaggcaa	186
	hsgRNA-ZBTB7A-2-5	aaucggaaca	187	ggaauagcugaauucggaaca	188
	hsgRNA-ZBTB7A-2-6	augacugaau	189	aaaaacuggaauagacugaau	190

Table 11. sgRNA/hsgRNA targeting the zinc finger structure of BCL11A

Table 11.1 sgRNA-T743I-1 and its hsgRNAs targeting T743 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-T743I-1	gugaguacug	353	agcgacacuugugaguacug	354
hsgRNA	hsgRNA-T743I-1-1	gggcccgggc	431	uuagugguccggcccgggc	432
	hsgRNA-T743I-1-3	gguccgggcc	433	ccauuuagugguccgggcc	434
	hsgRNA-T743I-1-5	auuagugguc	435	cacgccccauuuagugguc	436

Table 11.2 sgRNA-T743I-2 and its hsgRNAs targeting T743 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-T743I-2	ugaguacugu	355	gcgacacuugugaguacugu	356
hsgRNA	hsgRNA-T743I-2-1	gggcccgggc	437	uuagugguccggcccgggc	438
	hsgRNA-T743I-2-3	gguccgggcc	439	ccauuuagugguccgggcc	440
	hsgRNA-T743I-2-5	auuagugguc	441	cacgccccauuuagugguc	442

Table 11.3 sgRNA-C747Y-G748K/R/E and its hsgRNAs targeting C747 and G748 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-C747Y-G748K/R/E	uacucacaag	357	uuucccacaguacucacaag	358
hsgRNA	hsgRNA-C747Y-G748K/R/E-2	cuguggacag	443	guggcuucuccuguggacag	444
	hsgRNA-C747Y-G748K/R/E-3	uccuguggac	445	guguggcuucuccuguggac	446

	hsgRNA-C747Y-G748K/R/E-4	cuucuccugu	447	gcccuguguggcuccuccugu	448
--	--------------------------	------------	-----	-----------------------	-----

Table 11.4 sgRNA-S755N and its hsgRNAs targeting S755 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-S755N	caguucuuga	359	gagauugcuacaguucuuga	360
hsgRNA	hsgRNA-S755N-1	uucgcccug	449	uauaaggccuuucgcccug	450
	hsgRNA-S755N-2	cuuucgccc	451	uuuauaaggccuuucgccc	452
	hsgRNA-S755N-3	gccuuucgcc	453	cauuuauaaggccuuucgcc	454

Table 11.5 sgRNA-L757F-1 and its hsgRNAs targeting L757 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-L757F-1	cacuguccac	361	guagcaaucucacuguccac	362
hsgRNA	hsgRNA-L757F-1-1	ugaguacugu	455	gcgacacuugugaguacugu	456
	hsgRNA-L757F-1-2	gugaguacug	457	agcgacacuugugaguacug	458
	hsgRNA-L757F-1-3	gcucaaaaga	459	ggcaggcccagcucaaaaga	460

Table 11.6 sgRNA-L757F-2 and its hsgRNAs targeting L757 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-L757F-2	uguccacagg	363	gcaaucucacuguccacagg	364
hsgRNA	hsgRNA-L757F-2-1	ugaguacugu	461	gcgacacuugugaguacugu	462
	hsgRNA-L757F-2-2	uugugaguac	463	gcagcgacacuugugaguac	464
	hsgRNA-L757F-2-3	gacacuugug	465	cagacgcagcgacacuugug	466

Table 11.7 sgRNA-L757F-T758I and its hsgRNAs targeting L757 and T758 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-L757F-T758I	ccacaggaga	365	aucucacuguccacaggaga	366
hsgRNA	hsgRNA-L757F-T758I-1	acugugggaa	467	acuugugaguacugugggaa	468
	hsgRNA-L757F-T758I-2	gacacuugug	469	cagacgcagcgacacuugug	470
	hsgRNA-L757F-T758I-3	cagcgacacu	471	agggcagacgcagcgacacu	472

Table 11.8 sgRNA-V759I and its hsgRNAs targeting V759 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-V759I	agauugcuac	367	guggacagugagauugcuac	368
hsgRNA	hsgRNA-V759I-1	gcuuuuauaa	473	ugcacagcucgcuuuuauaa	474
	hsgRNA-V759I-2	cgcauuuaua	475	uugcacagcucgcuuuuaua	476

Table 11.9 sgRNA-H760Y and its hsgRNAs targeting H760 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-H760Y	agaagccaca	369	uguccacaggagaagccaca	370
hsgRNA	hsgRNA-H760Y-1	ugaguacugu	477	gcgacacuugugaguacugu	478
	hsgRNA-H760Y-2	gugaguacug	479	agcgacacuugugaguacug	480

Table 11.10 sgRNA-R761K and its hsgRNAs targeting R761 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-R761K	acagugagau	371	ucuccuguggacagugagau	372
hsgRNA	hsgRNA-R761K-1	uugcacagcu	481	acaggcauaguugcacagcu	482
	hsgRNA-R761K-2	auaguugcac	483	gggcacaggcauaguugcac	484

Table 11.11 sgRNA-R761K-R762K and its hsgRNAs targeting R761 and R762 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-R761K-R762K	guggacagug	373	ggcuucuccuguggacagug	374
hsgRNA	hsgRNA-R761K-R762K-1	uugcacagcu	485	acaggcauaguugcacagcu	486
	hsgRNA-R761K-R762K-2	auaguugcac	487	gggcacaggcauaguugcac	488

Table 11.12 sgRNA-R762K-S763N and its hsgRNAs targeting R761 and S763 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-R762K-S763N	cuguggacag	375	guggcuucuccuguggacag	376
hsgRNA	hsgRNA-R762K-S763N-1	uugcacagcu	489	acaggcauaguugcacagcu	490
	hsgRNA-R762K-S763N-2	auaguugcac	491	gggcacaggcauaguugcac	492

Table 11.13 sgRNA-H764Y and its hsgRNAs targeting H764 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-H764Y	cacgggcgaa	377	ggagaagccacacgggcgaa	378
hsgRNA	hsgRNA-H764Y-1	ugaguacugu	493	gcgacacuugugaguacugu	494
	hsgRNA-H764Y-2	gugaguacug	495	agcgacacuugugaguacug	496

Table 11.14 sgRNA-G766N/S/D and its hsgRNAs targeting G766 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-G766N/S/D	gcuucuccug	379	cgcccugugggcuucuccug	380
	hsgRNA-G766N/S/D-1	cucugggcac	497	gagcuugcuacucugggcac	498

hsgRNA	hsgRNA-G766N/S/D-2	uugcuacucu	499	ccuggugagcuugcuacucu	500
	hsgRNA-G766N/S/D-3	cuugcuacuc	501	gccuggugagcuugcuacuc	502

Table 11.15 sgRNA-G766N/S/D-E767K and its hsgRNAs targeting G766 and E767 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-G766N/S/D-E767K	uggcuucucc	381	uucgcccugugggcuucucc	382
hsgRNA	hsgRNA-G766N/S/D-E767K-1	caggcrauagu	503	acucugggcacaggcrauagu	504
	hsgRNA-G766N/S/D-E767K-2	gcacaggcrau	505	gcuaucucugggcacaggcrau	506

Table 11.16 sgRNA-R768K and its hsgRNAs targeting R768 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-R768K	uucgcccugug	383	uauaaggccuuucgcccugug	384
hsgRNA	hsgRNA-R768K-1	cucugggcac	507	gagcuugcuacucugggcac	508
	hsgRNA-R768K-2	uugcuacucu	509	ccuggugagcuugcuacucu	510
	hsgRNA-R768K-3	cuugcuacuc	511	gccuggugagcuugcuacuc	512

Table 11.17 sgRNA-P769F/S/L and its hsgRNAs targeting P769 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-P769F/S/L	aaaugcgagc	385	aaggccuuauaaaugcgagc	386
hsgRNA	hsgRNA-P769F/S/L-1	gcaaucucac	513	aagaacugugagcaaucucac	514
	hsgRNA-P769F/S/L-2	caagaacugu	515	ggaaagucuucaagaacugu	516
	hsgRNA-P769F/S/L-3	cuucaagaac	517	gugggaaagucuucaagaac	518

Table 11.18 sgRNA-C775Y and its hsgRNAs targeting C775 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-C775Y	cgcauuuaua	387	uugcacagcucgcauuuaua	388
hsgRNA	hsgRNA-C775Y-1	uugcuacucu	519	ccuggugagcuugcuacucu	520
	hsgRNA-C775Y-2	cuugcuacuc	521	gccuggugagcuugcuacuc	522
	hsgRNA-C775Y-3	uucaugugcc	523	gccaugcguuuucaugugcc	524

Table 11.19 sgRNA-A778V and its hsgRNAs targeting A778 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-A778V	ugcccagagu	389	acuaugccugugcccagagu	390
hsgRNA	hsgRNA-A778V-1	acgggcgaaa	525	gagaagccacacgggcgaaa	526
	hsgRNA-A778V-2	cacgggcgaa	527	ggagaagccacacgggcgaa	528

	hsgRNA-A778V-3	gccacacggg	529	cacaggagaagccacacggg	530
--	-----------------------	-------------------	-----	-----------------------------	-----

Table 11.20 sgRNA-A778T and its hsgRNAs targeting A778 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-A778T	uugcacagcu	391	acaggcauaguugcacagcu	392
hsgRNA	hsgRNA-A778T-1	uucaugugcc	531	gccaugcguuuucaugugcc	532
	hsgRNA-A778T-2	cguuuucaug	533	ccuggccaugcguuuucaug	534
	hsgRNA-A778T-3	ugcguuuuc	535	caccuggccaugcguuuuc	536

Table 11.21 sgRNA-A778V-A780V and its hsgRNAs targeting A778 and A780 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-A778V-A780V	cagaguagca	393	ugccugugcccagaguagca	394
hsgRNA	hsgRNA-A778V-A780V-1	acgggcgaaa	537	gagaagccacacgggcgaaa	538
	hsgRNA-A778V-A780V-2	cacgggcgaa	539	ggagaagccacacgggcgaa	540

Table 11.22 sgRNA-C779Y-A780T and its hsgRNAs targeting C779 and A780 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-C779Y-A780T	auaguugcac	395	gggcacaggcauaguugcac	396
hsgRNA	hsgRNA-C779Y-A780T-1	uucaugugcc	541	gccaugcguuuucaugugcc	542
	hsgRNA-C779Y-A780T-2	cguuuucaug	543	ccuggccaugcguuuucaug	544
	hsgRNA-C779Y-A780T-3	ugcguuuuc	545	caccuggccaugcguuuuc	546

Table 11.23 sgRNA-Q781 and its hsgRNAs targeting Q781 of BCL11A*

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-Q781*	caagcucacc	397	cccagaguagcaagcucacc	398
hsgRNA	hsgRNA-Q781*-1	cacgggcgaa	547	ggagaagccacacgggcgaa	548
	hsgRNA-Q781*-2	gaagccacac	549	guccacaggagaagccacac	550
	hsgRNA-Q781*-3	agaagccaca	551	uguccacaggagaagccaca	552

Table 11.24 sgRNA-S782N and its hsgRNAs targeting S782 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-S782N	gcacaggcau	399	gcuacucugggcacaggcau	400
hsgRNA	hsgRNA-S782N-1	ccuggccaug	553	uccuuccccaccuggccaug	554
	hsgRNA-S782N-2	caccuggcc	555	cguccuuccccaccuggcc	556
	hsgRNA-S782N-3	uuccccaccu	557	guaaacguccuuccccaccu	558

Table 11.25 sgRNA-S783N and its hsgRNAs targeting S783 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-S783N	cucugggcac	401	gagcuugcuacucugggcac	402
hsgRNA	hsgRNA-S783N-1	cuuccccacc	559	uguaaacguccuuccccacc	560

Table 11.26 sgRNA-L785F and its hsgRNAs targeting L785 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-L785F	accaggcaca	403	uagcaagcucaccaggcaca	404
hsgRNA	hsgRNA-L785F-1	aaaugcgagc	561	aaggccuauaaaugcgagc	562
	hsgRNA-L785F-2	uauaaaugcg	563	cgaaaggccuauaaaugcg	564
	hsgRNA-L785F-3	cuuauaaaug	565	ggcgaaaggccuauaaaug	566

Table 11.27 sgRNA-L785F-T786I and its hsgRNAs targeting L785 and T786 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-L785F-T786I	cacaugaaaa	405	gcucaccaggcacaugaaaa	406
hsgRNA	hsgRNA-L785F-T786I-1	ugugcaacua	567	aaaugcgagcugugcaacua	568
	hsgRNA-L785F-T786I-2	augcgagcug	569	ggccuauaaaugcgagcug	570
	hsgRNA-L785F-T786I-3	aaaugcgagc	571	aaggccuauaaaugcgagc	572

Table 11.28 sgRNA-R787K and its hsgRNAs targeting S783 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-R787K	cuugcuacuc	407	gccuggugagcuugcuacuc	408
hsgRNA	hsgRNA-R787K-1	cuuccccacc	573	uguaaacguccuuccccacc	574

Table 11.29 sgRNA-T791M-H792Y-1 and its hsgRNAs targeting T791 and H792 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-T791M-H792Y-1	caggugggga	409	aacgcauggccaggugggga	410
hsgRNA	hsgRNA-T791M-H792Y-1-1	ugcccagagu	575	acuaugccugugcccagagu	576
	hsgRNA-T791M-H792Y-1-2	cugugcccag	577	gcaacuaugccugugcccag	578
	hsgRNA-T791M-H792Y-1-3	gccugugccc	579	gugcaacuaugccugugccc	580

Table 11.30 sgRNA-T791M-H792Y-2 and its hsgRNAs targeting T791 and H792 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-T791M-H792Y-2	ggccaggugg	411	gaaaacgcauggccaggugg	412

hsgRNA	hsgRNA-T791M-H792Y-2-1	ugcccagagu	581	acuaugccugugcccagagu	582
	hsgRNA-T791M-H792Y-2-2	cugugcccag	583	gcaacuaugccugugcccag	584
	hsgRNA-T791M-H792Y-2-3	gccugugccc	585	gugcaacuaugccugugccc	586

Table 11.31 sgRNA-H792Y and its hsgRNAs targeting H792 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-H792Y	agguggggaa	413	acgcaugggccagguggggaa	414
hsgRNA	hsgRNA-H792Y-1	ugcccagagu	587	acuaugccugugcccagagu	588
	hsgRNA-H792Y-2	cugugcccag	589	gcaacuaugccugugcccag	590
	hsgRNA-H792Y-3	gccugugccc	591	gugcaacuaugccugugccc	592

Table 11.32 sgRNA-Q794* and its hsgRNAs targeting Q794 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-Q794*	uggggaagga	415	cauggccagguggggaagga	416
hsgRNA	hsgRNA-Q794*-1	cagaguagca	593	ugccugugcccagaguagca	594
	hsgRNA-Q794*-2	ugcccagagu	595	acuaugccugugcccagagu	596
	hsgRNA-Q794*-3	cugugcccag	597	gcaacuaugccugugcccag	598

Table 11.33 sgRNA-G796K/R/E and its hsgRNAs targeting G796 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-G796K/R/E	ccuggccaug	417	uccuucccccaccuggccaug	418
hsgRNA	hsgRNA-G796K/R/E-1	cacgcuaaaa	599	ggguacuguaacacgcuaaaa	600
	hsgRNA-G796K/R/E-2	acacgcuaaaa	601	aggguacuguaacacgcuaaaa	602

Table 11.34 sgRNA-G796K/R/E-D798N and its hsgRNAs targeting G796 and D798 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-G796K/R/E-D798N	caccuggcca	419	cguccuucccccaccuggcca	420
hsgRNA	hsgRNA-G796K/R/E-D798N-1	cacgcuaaaa	603	ggguacuguaacacgcuaaaa	604
	hsgRNA-G796K/R/E-D798N-2	acacgcuaaaa	605	aggguacuguaacacgcuaaaa	606

Table 11.35 sgRNA-P808F/S/L and its hsgRNAs targeting P808 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-P808F/S/L	uagcguguac	421	agaugccuuuagcguguac	422

hsgRNA	hsgRNA-P808F/S/L-1	uggggaagga	607	cauggccagguggggaagga	608
	hsgRNA-P808F/S/L-2	agguggggaa	609	acgcauggccagguggggaa	610
	hsgRNA-P808F/S/L-3	ggccaggugg	611	gaaaacgcauggccaggugg	612

Table 11.36 sgRNA-S813N-1 and its hsgRNAs targeting S813 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-S813N-1	cacgcuaaaa	423	ggguacugucacgcuaaaa	424
hsgRNA	hsgRNA-S813N-1-1	ucgaucacug	613	uauucaacacucgaucacug	614
hsgRNA	hsgRNA-S813N-1-2	acucgaucac	615	auuauucaacacucgaucac	616

Table 11.37 sgRNA-S813N-2 and its hsgRNAs targeting S813 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-S813N-2	acacgcuaaaa	425	agguacugucacgcuaaaa	426
hsgRNA	hsgRNA-S813N-2-1	ucgaucacug	617	uauucaacacucgaucacug	618
hsgRNA	hsgRNA-S813N-2-2	acucgaucac	619	auuauucaacacucgaucac	620

Table 11.38 sgRNA-E816K and its hsgRNAs targeting E816 of BCL11A

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-E816K	guacuguaca	427	uuucuccaggguacuguaca	428
hsgRNA	hsgRNA-E816K-1	auucaacacu	621	uuauucauuauucaacacu	622

Table 11.39 sgRNA-R826 and its hsgRNAs targeting R826 of BCL11A*

	sgRNA/ hsgRNA	10nt	SEQ ID NO:	20nt	SEQ ID NO:
sgRNA	sgRNA-R826*	uguugaauaa	429	agugaucgaguguugaauaa	430
hsgRNA	hsgRNA-R826*-1	aguaccugg	623	uagcgugucaguaccugg	624
hsgRNA	hsgRNA-R826*-2	acaguaccu	625	uuuagcgugucaguaccu	626
hsgRNA	hsgRNA-R826*-3	uacaguacc	627	uuuuagcgugucaguacc	628

Example 3. Identification of a New Deaminase Inhibitor

[0150] A tBE includes, along with a base editor, a cytidine deaminase inhibitor to inhibit the activity of the cytidine deaminase. The inhibitor can be cleaved once the tBE complex is assembled at the target genomic site. This example tested a newly identified cytidine deaminase inhibitor, hA3F-CDA1.

[0151] As illustrated in FIG. 34a, each of hA3B, hA3D, hA3F and hA3G contains a CDA1 domain, which was contemplated to correspond to the mouse mA3-CDA2 domain, which we

have confirmed as a cytidine deaminase inhibitor. Base editing constructs with or without these potential inhibitors were prepared (panel on the right).

[0152] The editing frequencies of these base editors were measured at a representative genomic locus, and the results are charted in **FIG. 34b**. Statistical analysis of normalized editing frequencies was carried out (**FIG. 34c**), setting the ones induced by the single-CDA-containing BEs as 100%. $n = 78$ (hA3BCDA2-nSpCas9-BE, hA3B-BE3, mA3CDA1-nSpCas9-BE and mA3-BE3) or 74 (hA3FCDA2-nSpCas9-BE and hA3F-BE3) edited cytosines at seven on-target sites from three independent experiments. The positive control mA3-BE3 exhibited the highest inhibitory effect and hA3F-CDA1 was the best among the tested candidate inhibitors.

[0153] Each of mA3-CDA2, hA3F-CDA1 and hA3B-CDA1 was fused to mA3-CDA1 (mA3CDA1-nSpCas9-BE) to prepare three tBE (**FIG. 34d**). As shown in the editing frequencies (**FIG. 34e**) and statistical analysis results (**FIG. 34e**), hA3F-CDA1 again outperformed hA3B-CDA1 in terms of the inhibitory activities.

[0154] Similar constructs were prepared and tested for their inhibitory effects on C-to-T editing frequencies at six genomic loci (**FIG. 35a**) or C-to-T editing frequencies at four genomic loci (**FIG. 35c**). Again, hA3F-CDA1 exhibited the best inhibitory effects.

[0155] This example, therefore, identifies hA3F-CDA1 as an excellent cytidine deaminase inhibitor, suitable for preparing transformer Base Editors (tBE).

* * *

[0156] The present disclosure is not to be limited in scope by the specific embodiments described which are intended as single illustrations of individual aspects of the disclosure, and any compositions or methods which are functionally equivalent are within the scope of this disclosure. It will be apparent to those skilled in the art that various modifications and variations can be made in the methods and compositions of the present disclosure without departing from the spirit or scope of the disclosure. Thus, it is intended that the present disclosure cover the modifications and variations of this disclosure provided they come within the scope of the appended claims and their equivalents.

[0157] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

CLAIMS

What is claimed is:

1. A method for promoting production of γ -globin in a human cell, comprising introducing into the cell a CRISPR-associated (Cas) protein, a nucleobase deaminase, a single-guide RNA (sgRNA), and a helper single-guide RNA (hsgRNA), wherein
 - (a) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:1-10, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-28;
 - (b) the sgRNA comprises the nucleic acid sequence of SEQ ID NO:29-30, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:31-36,
 - (c) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:37-54, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:63-116,
 - (d) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:117-122, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138,
 - (e) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:139-150, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-190, or
 - (f) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:353-430, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:431-628, andwherein the Cas protein, the nucleobase deaminase, the sgRNA, and the hsgRNA are preferably introduced into the cell by one or more encoding polynucleotides.
2. The method of claim 1, wherein the sgRNA comprises the nucleic acid sequence of SEQ ID NO:4, and the hsgRNA comprises the nucleic acid of SEQ ID NO:11.
3. The method of claim 1, wherein the sgRNA comprises the nucleic acid sequence of SEQ ID NO:30, and the hsgRNA comprises the nucleic acid of SEQ ID NO:33.

4. The method of claim 1, wherein the sgRNA and the hsgRNA comprise (b) and at least one pair selected from (a) and (c)-(e).

5. The method of claim 1, wherein the sgRNA and the hsgRNA comprise (a) and (b).

6. The method of claim 1, wherein:

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:38, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:55-62, preferably SEQ ID NO:57 or 59;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:40, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:63-70, preferably SEQ ID NO:57 or 59;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:42, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:71-80, preferably SEQ ID NO:71, 73, 77 or 79;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:44, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:81-104, preferably SEQ ID NO:81, 85 or 101;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:46, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:81-86 and SEQ ID NO:99-104, preferably SEQ ID NO:81, 85 or 101;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:48, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 81-86 and SEQ ID NO:99-104, preferably SEQ ID NO:85;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:50, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:87-98, preferably SEQ ID NO:87, 89, 91 or 93;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:52, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:105-116, preferably SEQ ID NO:111, 113 or 115; or

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:54, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:105-115.

7. The method of claim 1, wherein:

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:118, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138, preferably SEQ ID NO:127 or 129;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:120, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138, preferably SEQ ID NO:123, 127 or 129; or

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:122, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138.

8. The method of claim 1, wherein:

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:140, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164, preferably SEQ ID NO:153, 155, 157, 159 or 163;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:142, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164, preferably SEQ ID NO:159, 161 or 163;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:144, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164, preferably SEQ ID NO:151, 161 or 163;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:146, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-164;

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:148, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:165-178, preferably SEQ ID NO:165, 169, 171, 173, 175 or 177; or

the sgRNA comprises the nucleic acid sequence of SEQ ID NO:150, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:179-190, preferably SEQ ID NO:185, 187 or 189.

9. The method of any one of claims 1-8, wherein the nucleobase deaminase is a cytidine deaminase.

10. The method of claim 9, wherein the cytidine deaminase is selected from the group consisting of APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F

(A3F), APOBEC3G (A3G), APOBEC3H (A3H), APOBEC1 (A1), APOBEC3 (A3), APOBEC2 (A2), APOBEC4 (A4) and AICDA (AID).

11. The method of any one of claims 1-10, further comprising introducing into the cell a nucleobase deaminase inhibitor, fused to the nucleobase deaminase, via a protease cleavage site.

12. The method of claim 11, wherein the nucleobase deaminase inhibitor is an inhibitory domain of a nucleobase deaminase.

13. The method of claim 11, wherein the nucleobase deaminase inhibitor is an inhibitory domain of a cytidine deaminase.

14. The method of claim 11, wherein the nucleobase deaminase inhibitor comprises an amino acid sequence selected from SEQ ID NO:191-193.

15. The method of any one of claims 1-14, further comprising introducing into the cell a protease that is capable of cleaving at the protease cleavage site.

16. The method of claim 15, wherein the protease is selected from the group consisting of TuMV protease, PPV protease, PVY protease, ZIKV protease and WNV protease.

17. The method of any one of claims 1-16, wherein the Cas protein is selected from the group consisting of SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b, and RanCas13b.

18. The method of claim 17, wherein the Cas protein is catalytically impaired.

19. The method of claim 18, wherein the Cas protein is nCas9 or dCpf1.

20. The method of any one of claims 1-19, wherein the cell is an erythroid cell, a hematopoietic stem cell, or a stem cell.
21. The method of any one of claims 1-20, wherein the cell is *ex vivo*, or *in vivo* in a human patient.
22. The method of claim 21, wherein the patient suffers from β -thalassemia, sickle cell anemia, Haemoglobin C, or Haemoglobin E.
23. One or more polynucleotides encoding a CRISPR-associated (Cas) protein, a nucleobase deaminase, a single-guide RNA (sgRNA), and a helper single-guide RNA (hsgRNA), wherein
- (a) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:1-10, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-28;
 - (b) the sgRNA comprises the nucleic acid sequence of SEQ ID NO:29-30, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:31-36,
 - (c) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:37-54, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:63-116,
 - (d) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:118-122, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:123-138, or
 - (e) the sgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:139-150, and the hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:151-190.
24. The one or more polynucleotides of claim 23, wherein the sgRNA comprises the nucleic acid sequence of SEQ ID NO:4, and the hsgRNA comprises the nucleic acid of SEQ ID NO:11.
25. The one or more polynucleotides of claim 24 or 25, wherein the one or more polynucleotides further encode a second sgRNA and a second hsgRNA, wherein the second

sgRNA comprises the nucleic acid sequence of SEQ ID NO:30, and the second hsgRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:31-36, preferably SEQ ID NO:33.

26. The one or more polynucleotides of claim 23, wherein the sgRNA comprises the nucleic acid sequence of SEQ ID NO:30, and the hsgRNA comprises the nucleic acid of SEQ ID NO:33.

27. A fusion protein comprising:
a first fragment comprising a cytidine deaminase or a catalytic domain thereof,
a second fragment comprising a cytidine deaminase inhibitor comprising an amino acid sequence selected from the group consisting of SEQ ID NO:192, and 265-309 and sequences having at least 85% sequence identity to any of SEQ ID NO:192, and 265-309, and
a protease cleavage site between the first fragment and the second fragment.

28. The fusion protein of claim 27, wherein the cytidine deaminase is selected from the group consisting of APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F (A3F), APOBEC3G (A3G), APOBEC3H (A3H), APOBEC1 (A1), APOBEC3 (A3), APOBEC2 (A2), APOBEC4 (A4) and AICDA (AID).

29. The fusion protein of claim 27 or 28, wherein the first fragment further comprises a clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) protein.

30. The fusion protein of claim 29, wherein the Cas protein is selected from the group consisting of SpCas9, FnCas9, St1Cas9, St3Cas9, NmCas9, SaCas9, AsCpf1, LbCpf1, FnCpf1, VQR SpCas9, EQR SpCas9, VRER SpCas9, SpCas9-NG, xSpCas9, RHA FnCas9, KKH SaCas9, NmeCas9, StCas9, CjCas9, AsCpf1, FnCpf1, SsCpf1, PcCpf1, BpCpf1, CmtCpf1, LiCpf1, PmCpf1, Pb3310Cpf1, Pb4417Cpf1, BsCpf1, EeCpf1, BhCas12b, AkCas12b, EbCas12b, LsCas12b, RfCas13d, LwaCas13a, PspCas13b, PguCas13b, and RanCas13b.

31. The fusion protein of any one of claims 27-29, wherein the catalytic domain of the cytidine deaminase and the cytidine deaminase inhibitor are not different domains of a same cytidine deaminase.