(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 June 2023 (08.06.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/099591 A1

(51) International Patent Classification:
C07K 14/805 (2006.01) *C12N 5/0789* (2010.01)
A61K 48/00 (2006.01) *C12N 15/90* (2006.01)
C12N 5/078 (2010.01) *C12N 5/074* (2010.01)

(21) International Application Number:
PCT/EP2022/083904

(22) International Filing Date:
30 November 2022 (30.11.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
21306681.4 01 December 2021 (01.12.2021) EP

(71) Applicants: **INSERM (INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE)** [FR/FR]; 101, rue de Tolbiac, 75013 Paris (FR). **ASSISTANCE PUBLIQUE-HÔPITAUX DE PARIS (APHP)** [FR/FR]; 3, avenue Victoria, 75004 Paris (FR). **FONDATION IMAGINE** [FR/FR]; 24, Boulevard du Montparnasse, 75015 Paris (FR). **UNIVERSITÉ PARIS CITÉ** [FR/FR]; 85 Boulevard Saint-Germain, 75006 Paris (FR).

(72) Inventors: **MICCIO, Annarita**; Institut Imagine, UMR 1163, Université Paris Descartes, 24 Boulevard du Montparnasse, 75015 Paris (FR). **ANTONIOU, Panagiotis**; UMR1163, Institut Imagine, 24 Boulevard du Montparnasse, 75015 Paris (FR).

(74) Agent: **INSERM TRANSFERT**; 7 rue Watt, 75013 Paris (FR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

WO 2023/099591 A1

(54) **Title:** METHODS FOR INCREASING FETAL HEMOGLOBIN CONTENT BY EDITING THE +55-KB REGION OF THE ERYTHROID-SPECIFIC BCL11A ENHANCER

(57) **Abstract:** Reactivation of fetal hemoglobin remains a critical goal in the treatment of patients with sickle cell disease and β-thalassemia. Several genome-editing strategies have been developed with the goal of reactivating the expression of fetal γ-globin as a potential therapy for β-hemoglobinopathies. BCL11Ais one of the major repressors of γ-globin. To avoid DSB-induced toxicity, the erythroid specific *BCL11A* enhancer and specifically the ATF4 binding site can be targeted using CBE- and ABE-mediated base-editing approaches in order to downregulate *BCL11A* expression and reactivate HbF. Here, the inventors exploited CBEs and ABEs to dissect the ATF4 binding site in SCD HSPCs and identify the critical base conversions that induce changes in enhancer activity, *BCL11A* downregulation, and consecutively, HbF reactivation and sickling phenotype rescue.

METHODS FOR INCREASING FETAL HEMOGLOBIN CONTENT BY EDITING THE +55-KB REGION OF THE ERYTHROID-SPECIFIC BCL11A ENHANCER

FIELD OF THE INVENTION:

The present invention is in the field of medicine, in particular haematology.

BACKGROUND OF THE INVENTION:

β-hemoglobinopathies, β-thalassemia and sickle cell disease (SCD), are monogenic diseases caused by mutations in the β-globin locus, affecting the synthesis or the structure of the adult hemoglobin (Hb). β-thalassemia is caused by mutations in the β-globin gene (*HBB*) locus that reduce (β^+) or abolish (β^0) the production of β-globin chains included in the adult hemoglobin (HbA) tetramer, leading to the precipitation of uncoupled α-globin chains, erythroid cell death and severe anemia (Taher et al., 2018). In SCD, an A>T mutation in the *HBB* gene causes the substitution of valine for glutamic acid at position 6 of the β-globin chain (β^S), which is responsible for deoxygenation-induced polymerization of sickle hemoglobin (HbS). This primary event drives red blood cell (RBC) sickling, hemolysis, vaso-occlusive crises, multi-organ damage, often associated with severely reduced life expectancy (Piel et al., 2017; Kato et al., 2018).

The only definitive cure for β-hemoglobinopathies is transplantation of allogeneic hematopoietic stem cells (HSCs) from an HLA-compatible donor, an option available to <30% of the patients (Chandrakasan and Malik, 2014). Gene therapy approaches based on the transplantation of autologous, genetically modified HSCs have been investigated as a treatment option for patients lacking a compatible donor (Cavazzana et al., 2017). Genome editing technology has been exploited to develop therapeutic approaches for β-hemoglobinopathies, based on direct gene correction. These approaches use designer nucleases, such as the CRISPR/Cas9 system that induces DNA double-strand breaks (DSBs) via a single guide RNA (gRNA) complementary to a specific genomic target (Cavazzana et al., 2017). The DSB can be repaired via homologous-directed repair (HDR), by providing a donor DNA template containing the wild type sequence. However, HDR-mediated gene correction is poorly efficient in HSCs.

The clinical history of β-hemoglobinopathies shows that the severity of both β-thalassemia and SCD is mitigated by the synthesis of the fetal γ-globin in adulthood (Forget, 1998). Fetal hemoglobin (HbF) compensates for the HbA deficiency in β-thalassemia, and γ-globin exerts a potent anti-sickling effect in SCD by replacing the mutant sickle β-chain (Cavazzana et al., 2017). In particular, single nucleotide polymorphisms (SNP) in *BCL11A* have been associated with elevated expression of HbF in the adult life (Canver et al., 2015). BCL11A is one of the major repressors of *HBG1/2*. Studies aimed to completely inactivate BCL11A showed that it adversely affects lymphoid development and vital HSC functions (Liu et al., 2003; Yu et al., 2012; Guda et al., 2015) and human RBC enucleation (Chang et al., 2017). On the contrary, precise *BCL11A* downregulation through targeting of its erythroid specific enhancers can derepress γ-globin without adverse effects (Wu et al., 2019). Several genome-editing strategies have been developed with the goal of reactivating the expression of fetal γ-globin as a potential therapy for both β-thalassemia and SCD, based on the disruption of *trans*-regulatory elements via the generation of deletions or insertions in patient hematopoietic stem/progenitor cells (HSPCs) and reactivate HbF expression in their erythroid progeny. CRISPR/Cas9 disruption of GATA1 and ATF4 activators binding sites (BS) within the erythroid specific *BCL11A* enhancers (+58-kb and +55-kb regions) efficiently reactivates HbF expression in erythroid cell lines and primary HSPCs (Wu et al., 2019; Huang et al., 2020).

It is noteworthy that HSCs are highly sensitive to DNA DSBs (Milyavsky et al., 2010) - especially in cases of multiple on-targets or concomitant on-target and off-target events. Even when highly specific gRNAs are used, Cas9/gRNA treatment of human HSPCs induces a DNA damage response that can lead to apoptosis (Cromer et al., 2018). CRISPR/Cas9 can cause P53-dependent cell toxicity and cell cycle arrest, resulting in the negative selection of cells with a functional P53 pathway (Haapaniemi et al., 2018). Furthermore, the generation of several on-target DSBs, simultaneous on-target and off-target DSBs, or even a single on-target DSB is associated with a risk of deletion, inversion and translocation (Kosicki et al., 2018). Hence, the development of novel, efficacious and safe therapeutic strategies for β-hemoglobinopathies based on precise base editing rather than on DSB-induced DNA repair has been preferential. The basic components of base editors are a catalytically disabled Cas9 nuclease and a deaminase; these eventually produce a C-G to T-A or A-T to G-C conversion (for CBEs and ABEs, respectively) (Rees and Liu, 2018). Base-editing approaches allow precise DNA repair virtually in the absence of DSBs, and thus eliminate the risks of DSB-induced apoptosis, translocations and insertions or deletions of large portions of DNA. Importantly, base editing

occurs in quiescent cells - suggesting that *bona fide* HSCs could be genetically modified using this novel technology (Yeh et al., 2018), and results in homogeneous, predictable base changes as compared to the heterogeneous and unpredictable mutagenesis induced by NHEJ. Base editing has been exploited to correct β-hemoglobinopathy-causing mutations (Antoniou et al., 2021; Newby et al., 2021). However, inducing HbF reactivation represents a universal approach that can be applied to all the patients regardless of the specific disease-causing mutation (Zeng et al., 2020; Antoniou et al., 2021).

SUMMARY OF THE INVENTION:

The present invention is defined by the claims. In particular, the present invention relates to methods for increasing fetal hemoglobin content in eukaryotic cells and uses thereof for the treatment of hemoglobinopathies.

DETAILED DESCRIPTION OF THE INVENTION:

To avoid DSB-induced toxicity, the erythroid specific *BCL11A* enhancer and specifically the ATF4 binding site can be targeted using CBE- and ABE-mediated base-editing approaches in order to downregulate *BCL11A* expression and reactivate HbF. Here, the inventors exploited CBEs and ABEs to dissect the ATF4 binding site in SCD HSPCs and identify the critical base conversions that induce changes in enhancer activity, *BCL11A* downregulation, and consecutively, HbF reactivation and sickling phenotype rescue.

Definitions:

As used herein, the term "**β-hemoglobinopathy**" has its general meaning in the art and refers to any defect in the structure or function of any hemoglobin of an individual, and includes defects in the primary, secondary, tertiary or quaternary structure of hemoglobin caused by any mutation, such as deletion mutations or substitution mutations in the coding regions of the HBB gene, or mutations in, or deletions of, the promoters or enhancers of such gene that cause a reduction in the amount of hemoglobin produced as compared to a normal or standard condition.

As used herein, the term "**sickle cell disease**" has its general meaning in the art and refers to a group of autosomal recessive genetic blood disorders, which results from mutations in a globin gene and which is characterized by red blood cells that assume an abnormal, rigid, sickle shape.

They are defined by the presence of βS-globin gene coding for a β-globin chain variant in which glutamic acid is substituted by valine at amino acid position 6 of the peptide: incorporation of the βS-globin in the Hb tetramers (HbS, sickle Hb) leads to Hb polymerization and to a clinical phenotype. The term includes sickle cell anemia (HbSS), sickle-hemoglobin C disease (HbSC), sickle beta-plus- thalassaemia (HbS/β+), or sickle beta-zerothalassaemia (HbS/β0).

As used herein, the term "**β-thalassemia**" refers to a hemoglobinopathy that results from an altered ratio of α-globin to β-like globin polypeptide chains resulting in the underproduction of normal hemoglobin tetrameric proteins and the precipitation of free, unpaired α-globin chains.

As used herein, the term "**hematopoietic stem cell**" or "**HSC**" refers to blood cells that have the capacity to self-renew and to differentiate into precursors of blood cells. These precursor cells are immature blood cells that cannot self-renew and must differentiate into mature blood cells. Hematopoietic stem progenitor cells display a number of phenotypes, such as Lin-CD34+CD38−CD90+CD45RA−, Lin-CD34+CD38−CD90−CD45RA−, Lin-CD34+CD38+IL-3aloCD45RA−, and Lin-CD34+CD38+CD10+(Daley et al., Focus 18:62-67, 1996; Pimentel, E., Ed., Handbook of Growth Factors Vol. III: Hematopoietic Growth Factors and Cytokines, pp. 1-2, CRC Press, Boca Raton, Fla., 1994). Within the bone marrow microenvironment, the stem cells self-renew and maintain continuous production of hematopoietic stem cells that give rise to all mature blood cells throughout life. In some embodiments, the hematopoietic progenitor cells or hematopoietic stem cells are isolated form peripheral blood cells.

As used herein, the term "**peripheral blood cells**" refer to the cellular components of blood, including red blood cells, white blood cells, and platelets, which are found within the circulating pool of blood. In some embodiments, the eukaryotic cell is a bone marrow derived stem cell.

As used herein the term "**bone marrow-derived stem cells**" refers to stem cells found in the bone marrow. Stem cells may reside in the bone marrow, either as an adherent stromal cell type that possess pluripotent capabilities, or as cells that express CD34 or CD45 cell-surface protein, which identifies hematopoietic stem cells able to differentiate into blood cells.

As used herein, the term "**mobilization**" or "**stem cell mobilization**" refers to a process involving the recruitment of stem cells from their tissue or organ of residence to peripheral

blood following treatment with a mobilization agent. This process mimics the enhancement of the physiological release of stem cells from tissues or organs in response to stress signals during injury and inflammation. The mechanism of the mobilization process depends on the type of mobilization agent administered. Some mobilization agents act as agonists or antagonists that prevent the attachment of stem cells to cells or tissues of their microenvironment. Other mobilization agents induce the release of proteases that cleave the adhesion molecules or support structures between stem cells and their sites of attachment.

As used herein, the term "**mobilization agent**" refers to a wide range of molecules that act to enhance the mobilization of stem cells from their tissue or organ of residence, e.g., bone marrow (e.g., CD34+ stem cells) and spleen (e.g., Hox11+ stem cells), into peripheral blood. Mobilization agents include chemotherapeutic drugs, e.g., cyclophosphamide and cisplatin; cytokines, and chemokines, e.g., granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), stem cell factor (SCF), Fms-related tyrosine kinase 3 (flt-3) ligand, stromal cell-derived factor 1 (SDF-1); agonists of the chemokine (C—C motif) receptor 1 (CCR1), such as chemokine (C—C motif) ligand 3 (CCL3, also known as macrophage inflammatory protein-1α (Mip-1α)); agonists of the chemokine (C—X—C motif) receptor 1 (CXCR1) and 2 (CXCR2), such as chemokine (C—X—C motif) ligand 2 (CXCL2) (also known as growth-related oncogene protein-β (Gro-β)), and CXCL8 (also known as interleukin-8 (IL-8)); agonists of CXCR4, such as CTCE-02142, and Met-SDF-1,; Very Late Antigen (VLA)-4 inhibitors; antagonists of CXCR4, such as TG-0054, plerixafor (also known as AMD3100), and AMD3465, or any combination of the previous agents. A mobilization agent increases the number of stem cells in peripheral blood, thus allowing for a more accessible source of stem cells for use in transplantation, organ repair or regeneration, or treatment of disease.

As used herein, the term "**isolated cell**" refers to a cell that has been removed from an organism in which it was originally found, or a descendant of such a cell. Optionally the eukaryotic cell has been cultured in vitro, e.g., in the presence of other cells. Optionally the eukaryotic cell is later introduced into a second organism or reintroduced into the organism from which it (or the cell from which it is descended) was isolated. As used herein, the term "isolated population" with respect to an isolated population of cells as used herein refers to a population of cells that has been removed and separated from a mixed or heterogeneous population of cells. In some

embodiments, an isolated population is a substantially pure population of cells as compared to the heterogeneous population from which the cells were isolated or enriched.

As used herein, the term "**alpha globin**" or "**α-globin**" has its general meaning in the art and refers to protein that is encoded in human by the HBA1 and HBA2 genes. The human alpha globin gene cluster located on chromosome 16 spans about 30 kb and includes seven loci: 5'-zeta - pseudozeta - mu - pseudoalpha-1 - alpha-2 - alpha-1 - theta - 3'. The alpha-2 (HBA2) and alpha-1 (HBA1) coding sequences are identical. These genes differ slightly over the 5' untranslated regions and the introns, but they differ significantly over the 3' untranslated regions. The ENSEMBL IDs (i.e. the gene identifier number from the Ensembl Genome Browser database) for HBA1 and HBA2 are ENSG00000206172 and ENSG00000188536 respectively.

As used herein, the term "**beta globin**" or "**β-globin**"" has its general meaning in the art and refers to a globin protein, which along with alpha globin (HBA), makes up the most common form of haemoglobin (Hb) in adult humans. Normal adult human Hb is a heterotetramer consisting of two alpha chains and two beta chains. HBB is encoded by the *HBB* gene on human chromosome 11. It is 146 amino acids long and has a molecular weight of 15,867 Da.

As used herein, the term "**gamma globin**" or "**γ-globin**" has its general meaning in the art and refers to protein that is encoded in human by the HBG1 and HBG2 genes. The HBG1 and HBG2 genes are normally expressed in the fetal liver, spleen and bone marrow. Two γ-globin chains together with two α-globin chains constitute fetal hemoglobin (HbF) which is normally replaced by adult hemoglobin (HbA) in the year following birth. The ENSEMBL IDs (i.e. the gene identifier number from the Ensembl Genome Browser database) for HBG1 and HBG2 are ENSG00000213934 and ENSG00000196565 respectively.

As used herein, the term "**transcriptional repressor**" has its general meaning in the art and refers a protein (transcription factor) that decreases gene transcription of a gene or set of genes. Most repressors are DNA-binding proteins that bind to enhancers or promoter-proximal elements. According to the present disclosure, the transcriptional repressor is BCL11A.

As used herein, the term "**BCL11A**" has its general meaning in the art and refers to the gene encoding for BAF chromatin remodeling complex subunit BCL11A (Gene ID: 53335). The

term is also known as EVI9; CTIP1; DILOS; ZNF856; HBFQTL5; BCL11A-L; BCL11A-S; BCL11a-M; or BCL11A-XL. Five alternatively spliced transcript variants of this gene, which encode distinct isoforms, have been reported. The protein associates with the SWI/SNF complex that regulates gene expression via chromatin remodeling. BCL11A is highly expressed in several hematopoietic lineages, and plays a role in the switch from γ- to β-globin expression during the fetal to adult erythropoiesis transition (*Sankaran VJ et al. "Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A", Science Science. 2008 Dec 19;322(5909):1839-42*).

As used herein, the term "**transcriptional activator**" has its general meaning in the art and refers to a protein that increases gene transcription of a gene or set of genes. Most activators are DNA-binding proteins that bind to enhancers or promoter-proximal elements. According to the present disclosure, the activator ATF4.

As used herein, the term "**ATF4**" has its general meaning in the art and refers to the activating transcription factor 4 (tax-responsive enhancer element B67) that is a protein that in humans is encoded by the *ATF4* gene (Gene ID:468). ATF4 is thus a transcriptional activator that was originally identified as a widely expressed mammalian DNA binding protein that could bind a tax-responsive enhancer element in the LTR of HTLV-1. The encoded protein was also isolated and characterized as the cAMP-response element binding protein 2 (CREB-2). The protein encoded by this gene belongs to a family of DNA-binding proteins that includes the AP-1 family of transcription factors, cAMP-response element binding proteins (CREBs) and CREB-like proteins. These transcription factors share a leucine zipper region that is involved in protein-protein interactions, located C-terminal to a stretch of basic amino acids that functions as a DNA binding domain. Two alternative transcripts encoding the same protein have been described. Two pseudogenes are located on the X chromosome at q28 in a region containing a large inverted duplication.

As used herein, the term "**transcriptional activator binding site**" refers to a site present on DNA whereby the transcriptional activator according to the present disclosure binds. According to the present invention, the base-editing enzyme of the present invention edits the genome sequence of the eukaryotic cell so that the activator is able to bind to its transcriptional activator binding site.

As used herein, the expression "**+55-kb region of the erythroid-specific BCL11A enhancer**" refers to the region depicted in **Figure 1** and having the nucleotide acid sequence as set forth in SEQ ID NO:1. The "**ATF4 binding site**" ranges from the nucleotide at position 21 to the nucleotide at position 31 in SEQ ID NO:1 (i.e. TTGCATCATCC (SEQ ID NO: 45)).

SEQ ID NO:1 > "+55-kb region of the erythroid-specific BCL11A enhancer. The ATF4 binding site is indicated in bold and underlined.

GAGCTCACAGCCTCCAAGCA**TTGCATCATCC**TGGTACCAGGAAGGC

As used herein, the term "**expression**" refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. Any method known in the art can be used to measure the expression of the gene (e. g. HPLC analysis of protein and RT-qPCR analysis of mRNA.) Typically, said methods are described in the EXAMPLE.

As used herein, the expression "**increasing the fetal hemoglobin content**" indicates that fetal hemoglobin is at least 5% higher in the eukaryotic cell treated with the gene editing platform, than in a comparable, eukaryotic cell, wherein a gene editing platform targeting an unrelated locus is present or where no gene editing platform is present. In some embodiments, the percentage of fetal hemoglobin expression in the eukaryotic cell is at least 10% higher, at least 20% higher, at least 30% higher, at least 40% higher, at least 50% higher, at least 60% higher, at least 70% higher, at least 80% higher, at least 90% higher, at least 1-fold higher, at least 2-fold higher, at least 5-fold higher, at least 10-fold higher, at least 100 fold higher, at least 1000-fold higher, or more than an eukaryotic cell.

As used herein, the expression "**repressing the expression BCL11A**" indicates the expression of BCL11A is at least 5% lower in the eukaryotic cell contacted with the gene editing platform of the present invention than in a comparable eukaryotic cell that was not contacted with said gene editing platform. In some embodiments, the percentage of BCL11A expression in the eukaryotic cell is at least 10% lower, at least 20% lower, at least 30% lower, at least 40% lower, at least 50% lower, at least 60% lower, at least 70% lower, at least 80% lower, at least 90% lower, at least 1-fold lower, at least 2-fold lower, at least 5-fold lower, at least 10 fold lower, at

least 100 fold lower, at least 1000-fold lower, or less than an eukaryotic cell that was not contacted with the gene editing platform.

As used herein, the terms "**polypeptide**", "**peptide**" and "**protein**" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, pegylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

As used herein, the term "**nucleic acid molecule**" or "**polynucleotide**" refers to a DNA molecule (for example, but not limited to, a cDNA or genomic DNA). The nucleic acid molecule can be single-stranded or double-stranded.

As used herein, the term "**isolated**" when referring to nucleic acid molecules or polypeptides means that the nucleic acid molecule or the polypeptide is substantially free from at least one other component with which it is associated or found together in nature.

As used herein, the term "**complementarity**" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base-pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions.

As used herein, the term "**stringent conditions**" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y.

As used herein, the term "**hybridization**" or "**hybridizing**" refers to a process where completely or partially complementary nucleic acid strands come together under specified hybridization conditions to form a double-stranded structure or region in which the two constituent strands are joined by hydrogen bonds. Although hydrogen bonds typically form between adenine and thymine or uracil (A and T or U) or cytosine and guanine (C and G), other base pairs may form (e.g., Adams et al., The Biochemistry of the Nucleic Acids, 11th ed., 1992).

As used herein, the term "**fusion polypeptide**" or "**fusion protein**" means a protein created by joining two or more polypeptide sequences together. The fusion polypeptides encompassed in this invention include translation products of a chimeric gene construct that joins the nucleic acid sequences encoding a first polypeptide, e.g., an RNA-binding domain, with the nucleic acid sequence encoding a second polypeptide, e.g., an effector domain, to form a single open-reading frame. In other words, a "fusion polypeptide" or "fusion protein" is a recombinant protein of two or more proteins which are joined by a peptide bond or via several peptides. The fusion protein may also comprise a peptide linker between the two domains.

As used herein the term "**wild type**" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms.

As used herein, the term "**derived from**" refers to a process whereby a first component (e.g., a first molecule), or information from that first component, is used to isolate, derive or make a different second component (e.g., a second molecule that is different from the first).

As used herein, the "**percent identity**" between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = number of identical positions/total number of positions x 100), taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm, as described below. The percent identity between two amino acid sequences can be determined using the Needleman and Wunsch algorithm (Needleman, Saul B. & Wunsch, Christian D. (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins". Journal of Molecular Biology. 48 (3): 443–53.). The percent identity between two nucleotide or amino acid sequences may also be determined using for example algorithms such as EMBOSS Needle (pair wise alignment; available at www.ebi.ac.uk). For example, EMBOSS Needle may be used with a BLOSUM62 matrix, a "gap open penalty" of 10, a "gap extend penalty" of 0.5, a false "end gap penalty", an "end gap open penalty" of 10 and an "end gap extend penalty" of 0.5. In general, the "percent identity" is a function of the number of matching positions divided by the number of positions compared and multiplied by 100. For instance, if 6 out of 10 sequence positions are identical between the two compared sequences after alignment, then the identity is 60%. The % identity is typically determined over the whole length of the query sequence on which the analysis is performed. Two molecules having the same primary amino acid sequence or nucleic acid sequence are identical irrespective of any chemical and/or biological modification. According to the invention a first amino acid sequence having at least 90% of identity with a second amino acid sequence means that the first sequence has 90; 91; 92; 93; 94; 95; 96; 97; 98; 99 or 100% of identity with the second amino acid sequence.

As used herein, the term "**linker**" refers to any means, entity or moiety used to join two or more entities. A linker can be a covalent linker or a non-covalent linker. Examples of covalent linkers include covalent bonds or a linker moiety covalently attached to one or more of the proteins or domains to be linked. The linker can also be a non-covalent bond, e.g., an organometallic bond through a metal center such as platinum atom. For covalent linkages, various functionalities can be used, such as amide groups, including carbonic acid derivatives, ethers, esters, including organic and inorganic esters, amino, urethane, urea and the like. To provide for linking, the domains can be modified by oxidation, hydroxylation, substitution, reduction etc. to provide a site for coupling. Methods for conjugation are well known by persons skilled in the art and are encompassed for use in the present invention. Linker moieties include, but are not limited to,

chemical linker moieties, or for example a peptide linker moiety (a linker sequence). It will be appreciated that modification which do not significantly decrease the function of the RNA-binding domain and effector domain are preferred.

As used herein, the "**linked**" as used herein refers to the attachment of two or more entities to form one entity. A conjugate encompasses both peptide-small molecule conjugates as well as peptide-protein/peptide conjugates.

As used herein, the term "**base-editing enzyme**" refers to fusion protein comprising a defective CRISPR/Cas nuclease linked to a deaminase polypeptide. The term is also known as "**base-editor**". Two classes of base-editing enzymes--cytosine base-editing enzymes (CBEs) and adenine base-editing enzymes (ABEs)--can be used to generate single base pair edits without double stranded breaks. Typically, cytosine base-editing enzymes are created by fusing the defective CRISPR/Cas nuclease to a deaminase.

As used herein, the term "**deaminase**" refers to an enzyme that catalyses a deamination reaction. The term "**deamination**", as used herein, refers to the removal of an amine group from one molecule. In some embodiments, the deaminase is a cytidine deaminase, catalysing the hydrolytic deamination of cytidine or deoxycytidine to uracil or deoxyuracil, respectively. In some embodiments, the deaminase is an adenosine deaminase, catalysing the hydrolytic deamination of adenosine to inosine, which is treated like guanosine by the cell, creating an A to G (or T to C) change.

As used herein, the term "**nuclease**" includes a protein (i.e. an enzyme) that induces a break in a nucleic acid sequence, e.g., a single or a double strand break in a double-stranded DNA sequence.

As used herein, the term "**CRISPR/Cas nuclease**" has its general meaning in the art and refers to segments of prokaryotic DNA containing clustered regularly interspaced short palindromic repeats (CRISPR) and associated nucleases encoded by Cas genes. In bacteria the CRISPR/Cas loci encode RNA-guided adaptive immune systems against mobile genetic elements (viruses, transposable elements and conjugative plasmids). Three types of CRISPR systems have been identified. CRISPR clusters contain spacers, the sequences complementary to antecedent mobile elements. CRISPR clusters are transcribed and processed into mature CRISPR

(Clustered Regularly Interspaced Short Palindromic Repeats) RNA (crRNA). The CRISPR/Cas nucleases Cas9 and Cpf1 belong to the type II and type V CRISPR/Cas system and have strong endonuclease activity to cut target DNA. Cas9 is guided by a mature crRNA that contains about 20 nucleotides of unique target sequence (called spacer) and a trans-activating small RNA (tracrRNA) that also serves as a guide for ribonuclease III-aided processing of pre-crRNA. The crRNA:tracrRNA duplex directs Cas9 to target DNA via complementary base pairing between the spacer on the crRNA and the complementary sequence (called protospacer) on the target DNA. Cas9 recognizes a trinucleotide (NGG for S. Pyogenes Cas9) protospacer adjacent motif (PAM) to specify the cut site (the 3rd or the 4th nucleotide upstream from PAM).

As used herein, the term "**Cas9**" or "**Cas9 nuclease**" refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (e.g., a protein comprising an active or inactive DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3′ -5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, e.g., Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, e.g., "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti et al., J. J., McShan W. M., Ajdic D. J., Savic D. J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A. N., Kenton S., Lai H. S., Lin S. P., Qian Y., Jia H. G., Najar F. Z., Ren Q., Zhu H., Song L.,

White J., Yuan X., Clifton S. W., Roe B. A., McLaughlin R. E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C. M., Gonzales K., Chao Y., Pirzada Z. A., Eckert M. R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, S. pyogenes and S. thermophilus. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, the term "Cas9" refers to Cas9 from: Corynebacterium ulcerans (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); Campylobacter jejuni (NCBI Ref: YP_002344900.1); or *Neisseria. meningitidis* (NCBI Ref: YP_002342100.1). Typically the Cas9 nuclease comprises the amino acid sequence as set forth in SEQ ID NO: 2.

SEQ ID NO:2: Cas9 sequence

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD
NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVD
KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF
RKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY

EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

As used herein, the term "**defective CRISPR/Cas nuclease**" refers to a CRISPR/Cas nuclease having lost at least one nuclease domain.

As used herein, the term "**nickase**" has its general meaning in the art and refers to an endonuclease which cleaves only a single strand of a DNA duplex. Accordingly, the term "**Cas9 nickase**" refers to a nickase derived from a Cas9 protein, typically by inactivating one nuclease domain of Cas9 protein.

As used herein, the term "**guide RNA molecule**" generally refers to an RNA molecule (or a group of RNA molecules collectively) that can bind to a Cas9 protein and target the Cas9 protein to a specific location within a target DNA. A guide RNA can comprise two segments: a DNA-targeting guide segment and a protein-binding segment. The DNA-targeting segment comprises a nucleotide sequence that is complementary to (or at least can hybridize to under stringent conditions) a target sequence. The protein-binding segment interacts with a CRISPR protein, such as a Cas9 or Cas9 related polypeptide. These two segments can be located in the same RNA molecule or in two or more separate RNA molecules. When the two segments are in separate RNA molecules, the molecule comprising the DNA-targeting guide segment is sometimes referred to as the CRISPR RNA (crRNA), while the molecule comprising the protein-binding segment is referred to as the trans-activating RNA (tracrRNA).

As used herein, the term "**target nucleic acid**" or "**target**" refers to a nucleic acid containing a target nucleic acid sequence. A target nucleic acid may be single-stranded or double-stranded, and often is double-stranded DNA. A "target nucleic acid sequence," "target sequence" or "target region," as used herein, means a specific sequence or the complement thereof that one wishes to bind to using the CRISPR system as disclosed herein.

As used herein, the term "**target nucleic acid strand**" refers to a strand of a target nucleic acid that is subject to base-pairing with a guide RNA as disclosed herein. That is, the strand of a target nucleic acid that hybridizes with the crRNA and guide sequence is referred to as the "target nucleic acid strand." The other strand of the target nucleic acid, which is not complementary to the guide sequence, is referred to as the "non-complementary strand." In the case of double-stranded target nucleic acid (e.g., DNA), each strand can be a "target nucleic

acid strand" to design crRNA and guide RNAs and used to practice the method of this invention as long as there is a suitable PAM site.

As used herein, the term "**ribonucleoprotein complex**," or "**ribonucleoprotein particle**" refers to a complex or particle including a nucleoprotein and a ribonucleic acid. A "nucleoprotein" as provided herein refers to a protein capable of binding a nucleic acid (e.g., RNA, DNA). Where the nucleoprotein binds a ribonucleic acid it is referred to as "ribonucleoprotein." The interaction between the ribonucleoprotein and the ribonucleic acid may be direct, e.g., by covalent bond, or indirect, e.g., by non-covalent bond (e.g. electrostatic interactions (e.g. ionic bond, hydrogen bond, halogen bond), van der Waals interactions (e.g. dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions and the like).

As used herein, the term "**mutation**" has its general meaning in the art and refers to a substitution, deletion or insertion. The term "**substitution**" means that a specific amino acid residue at a specific position is removed and another amino acid residue is inserted into the same position. The term "**deletion**" means that a specific amino acid residue is removed. The term "insertion" means that one or more amino acid residues are inserted before or after a specific amino acid residue.

As used herein, the term "**mutagenesis**" refers to the introduction of mutations into a polynucleotide sequence.

As used herein, the term "**variant**" refers to a first composition (e.g., a first molecule), that is related to a second composition (e.g., a second molecule, also termed a "parent" molecule). The variant molecule can be derived from, isolated from, based on or homologous to the parent molecule. A variant molecule can have entire sequence identity with the original parent molecule, or alternatively, can have less than 100% sequence identity with the parent molecule. For example, a variant of a sequence can be a second sequence that is at least 50; 51; 52; 53; 54; 55; 56; 57; 58; 59; 60; 61; 62; 63; 64; 65; 66; 67; 68; 69; 70; 71; 72; 73; 74; 75; 76; 77; 78; 79; 80; 81; 82; 83; 84; 85; 86; 87; 88; 89; 90; 91; 92; 93; 94; 95; 96; 97; 98; 99; 100% identical in sequence compare to the original sequence.

As used herein, the term "**treatment**" or "**treat**" refer to both prophylactic or preventive treatment as well as curative or disease modifying treatment, including treatment of patient at risk of contracting the disease or suspected to have contracted the disease as well as patients who are ill or have been diagnosed as suffering from a disease or medical condition, and includes suppression of clinical relapse. The treatment may be administered to a subject having a medical disorder or who ultimately may acquire the disorder, in order to prevent, cure, delay the onset of, reduce the severity of, or ameliorate one or more symptoms of a disorder or recurring disorder, or in order to prolong the survival of a subject beyond that expected in the absence of such treatment. By "**therapeutic regimen**" is meant the pattern of treatment of an illness, e.g., the pattern of dosing used during therapy. A therapeutic regimen may include an induction regimen and a maintenance regimen. The phrase "**induction regimen**" or "**induction period**" refers to a therapeutic regimen (or the portion of a therapeutic regimen) that is used for the initial treatment of a disease. The general goal of an induction regimen is to provide a high level of drug to a patient during the initial period of a treatment regimen. An induction regimen may employ (in part or in whole) a "**loading regimen**", which may include administering a greater dose of the drug than a physician would employ during a maintenance regimen, administering a drug more frequently than a physician would administer the drug during a maintenance regimen, or both. The phrase "**maintenance regimen**" or "**maintenance period**" refers to a therapeutic regimen (or the portion of a therapeutic regimen) that is used for the maintenance of a patient during treatment of an illness, e.g., to keep the patient in remission for long periods of time (months or years). A maintenance regimen may employ continuous therapy (e.g., administering a drug at regular intervals, e.g., weekly, monthly, yearly, etc.) or intermittent therapy (e.g., interrupted treatment, intermittent treatment, treatment at relapse, or treatment upon achievement of a particular predetermined criteria [e.g., pain, disease manifestation, etc.]).

As used herein, the term "**therapeutically effective amount**" is meant a sufficient amount of population of cells to treat the disease at a reasonable benefit/risk ratio applicable to any medical treatment. It will be understood that the total usage compositions of the present invention will be decided by the attending physician within the scope of sound medical judgment. The specific therapeutically effective dose level for any particular patient will depend upon a variety of factors including the age, body weight, general health, sex and diet of the patient, the time of administration, route of administration, the duration of the treatment, drugs used in combination or coincidental with the population of cells, and like factors well known in the medical arts. In

some embodiments, the cells are formulated by first harvesting them from their culture medium, and then washing and concentrating the cells in a medium and container system suitable for administration (a "pharmaceutically acceptable" carrier) in a treatment-effective amount. Suitable infusion medium can be any isotonic medium formulation, typically normal saline, Normosol R (Abbott) or Plasma-Lyte A (Baxter), but also 5% dextrose in water or Ringer's lactate can be utilized. The infusion medium can be supplemented with human serum albumin. A treatment-effective amount of cells in the composition is dependent on the relative representation of the cells with the desired specificity, on the age and weight of the recipient, and on the severity of the targeted condition. This number of cells can be as low as approximately 10^3/kg, preferably $5x10^3$/kg; and as high as 10^7/kg, preferably 10^8/kg. The number of cells will depend upon the ultimate use for which the composition is intended, as will the type of cells included therein. Typically, the minimal dose is 2 millions of cells per kg. Usually 2 to 20 millions of cells are injected in the subject. The desired purity can be achieved by introducing a sorting step. For uses provided herein, the cells are generally in a volume of a liter or less, can be 500 ml or less, even 250 ml or 100 ml or less. The clinically relevant number of cells can be apportioned into multiple infusions that cumulatively equal or exceed the desired total amount of cells.

Methods:

Accordingly, the first object of the present invention relates to a method for increasing fetal hemoglobin content in a eukaryotic cell comprising the step of contacting the eukaryotic cell with a gene editing platform that consists of a (a) at least one base-editing enzyme and (b) least one guide RNA molecule for guiding the base-editing enzyme to at least one target sequence in the +55-kb region of the erythroid-specific BCL11A enhancer, thereby editing and disrupting the ATF4 binding site in said region so as to repress the expression of BCL11A and subsequently increase the expression of γ-globin.

In some embodiments, the eukaryotic cell is selected from the group consisting of hematopoietic progenitor cells, hematopoietic stem cells (HSCs), pluripotent cells (i.e. embryonic stem cells (ES) and induced pluripotent stem cells (iPS)). Typically, the eukaryotic cell results from a stem cell mobilization.

In some embodiments, the base-editing enzyme of the present invention comprises a defective CRISPR/Cas nuclease. The sequence recognition mechanism is the same as for the non-defective CRISPR/Cas nuclease. Typically, the defective CRISPR/Cas nuclease of the invention comprises at least one RNA binding domain. The RNA binding domain interacts with a guide RNA molecule as defined hereinafter. However, the defective CRISPR/Cas nuclease of the invention is a modified version with no nuclease activity. Accordingly, the defective CRISPR/Cas nuclease specifically recognizes the guide RNA molecule and thus guides the base-editing enzyme to its target DNA sequence.

In some embodiments, the defective CRISPR/Cas nuclease can be modified to increase nucleic acid binding affinity and/or specificity, alter an enzymatic activity, and/or change another property of the protein. In some embodiments, the nuclease domains of the protein can be modified, deleted, or inactivated. In some embodiments, the protein can be truncated to remove domains that are not essential for the function of the protein. In some embodiments, the protein is truncated or modified to optimize the activity of the RNA binding domain.

In some embodiments, the CRISPR/Cas nuclease consists of a mutant CRISPR/Cas nuclease i.e. a protein having one or more point mutations, insertions, deletions, truncations, a fusion protein, or a combination thereof. In some embodiments, the mutant has the RNA-guided DNA binding activity, but lacks one or both of its nuclease active sites. In some embodiments, the mutant comprises an amino acid sequence having at least 50% of identity with the wild type amino acid sequence of the CRISPR/Cas nuclease. Various CRISPR/Cas nucleases can be used in this invention. Non-limiting examples of suitable CRISPR/CRISPR/Cas nucleases include Cas3, Cas4, Cas5, Cas5e (or CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (or CasA), Cse2 (or CasB), Cse3 (or CasE), Cse4 (or CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4, and Cu1966. See e.g., WO2014144761 WO2014144592, WO2013176772, US20140273226, and US20140273233, the contents of which are incorporated herein by reference in their entireties.

In some embodiments, the CRISPR/Cas nuclease is derived from a type II CRISPR-Cas system. In some embodiments, the CRISPR/Cas nuclease is derived from a Cas9 protein. The Cas9 protein can be from *Streptococcus pyogenes, Streptococcus thermophilus, Streptococcus sp.,*

Nocardiopsis dassonvillei, Streptomyces pristinaespiralis, Streptomyces viridochromogenes, Streptomyces viridochromogenes, Streptosporangium roseum, Streptosporangium roseum, Alicyclobacillus acidocaldarius, Bacillus pseudomycoides, Bacillus selenitireducens, Exiguobacterium sibiricum, Lactobacillus delbrueckii, Lactobacillus salivarius, Microscilla marina, Burkholderiales bacterium, Polaromonas naphthalenivorans, Polaromonas sp., Crocosphaera watsonii, Cyanothece sp., Microcystis aeruginosa, Synechococcus sp., Acetohalobium arabaticum, Ammonifex degensii, Caldicelulosiruptor becscii, Candidatus Desulforudis, Clostridium botulinum, Clostridium difficile, Finegoldia magna, Natranaerobius thermophilus, Pelotomaculum thermopropionicum, Acidithiobacillus caldus, Acidithiobacillus ferrooxidans, Allochromatium vinosum, Marinobacter sp., Nitrosococcus halophilus, Nitrosococcus watsoni, Pseudoalteromonas haloplanktis, Ktedonobacter racemifer, Methanohalobium evestigatum, Anabaena variabilis, Nodularia spumigena, Nostoc sp., Arthrospira maxima, Arthrospira platensis, Arthrospira sp., Lyngbya sp., Microcoleus chthonoplastes, Oscillatoria sp., Petrotoga mobilis, Thermosipho africanus, or *Acaryochloris marina, inter alia*.

In some embodiments, the CRISPR/Cas nuclease is a mutant of a wild type CRISPR/Cas nuclease (such as Cas9) or a fragment thereof. In some embodiments, the CRISPR/Cas nuclease is a mutant Cas9 protein from *S. pyogenes*.

Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, e.g., Jinek et al., Science. 337:816-821(2012); Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) Cell. 28; 152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H841A completely inactivate the nuclease activity of S. pyogenes Cas9 (Jinek et al., Science. 337:816-821(2012); Qi et al., Cell. 28; 152(5):1173-83 (2013).

In some embodiments, the CRISPR/Cas nuclease of the present invention is nickase and more particularly a Cas9 nickase i.e. the Cas9 from *S. pyogenes* having one mutation selected from

the group consisting of D10A and H840A. In some embodiments, the nickase of the present invention comprises the amino acid sequence as set forth in SEQ ID NO: 3 or SEQ ID NO:4.

SEQ ID NO: 3> S. pyogenes nCas9 Protein Sequence having the D10A mutation

MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD
NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVD
KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF
RKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

SEQ ID NO: 4> S. pyogenes nCas9 Protein Sequence having the H840A mutation

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD
NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI
HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVD
KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLL
FKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF
RKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

In some embodiments, the Cas9 variants having mutations other than D10A or H840A are used, which e.g., result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (e.g., substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants of dCas9 are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about

95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO: 2 or 3. In some embodiments, variants of dCas9 are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 2 or 3, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.

According to the present invention, the second component of the base-editing enzyme herein disclosed comprises a non-nuclease DNA modifying enzyme that is a deaminase.

In some embodiments, the deaminase is a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC1 family deaminase. In some embodiments, the deaminase is an activation-induced cytidine deaminase (AID). In some embodiments, the deaminase is an ACF1/ASE deaminase.

In some embodiments, the deaminase is selected from the group consisting of AID: activation induced cytidine deaminase, APOBEC1: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1, APOBEC3A: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A, APOBEC3B: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B, APOBEC3C: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3C, APOBEC3D: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3D, APOBEC3F: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F, APOBEC3G: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G, APOBEC3H: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3H, ADA: adenosine deaminase, ADAR1: adenosine deaminase acting on RNA 1, Dnmt1: DNA (cytosine-5-)-methyltransferase 1, Dnmt3a: DNA (cytosine-5-)-methyltransferase 3 alpha, Dnmt3b: DNA (cytosine-5-)-methyltransferase 3 beta and Tet1: methylcytosine dioxygenase.

In some embodiments, the deaminase derives from the Activation Induced cytidine Deaminase (AID). AID is a cytidine deaminase that can catalyze the reaction of deamination of cytosine in the context of DNA or RNA. When brought to the targeted site, AID changes a C base to U base. In dividing cells, this could lead to a C to T point mutation. Alternatively, the change of

C to U could trigger cellular DNA repair pathways, mainly excision repair pathway, which will remove the mismatching U-G base-pair, and replace with a T-A, A-T, C-G, or G-C pair. As a result, a point mutation would be generated at the target C-G site. In some embodiments, the DNA modifying enzyme is AID*Δ that is an AID mutant with increased SHM activity whose Nuclear Export Signal (NES) has been removed (*Hess GT, Fresard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC: Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods 2016, 13(12):1036-1042*).

In some embodiments, the deaminase consists of a variant of the amino acid sequence as set forth in SEQ ID NO:5-15.

SEQ ID NO:5 Human AID:
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAI
MTFKDYFYCWNT FVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

SEQ ID NO:6 Human APOBEC-3G
MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMR
FFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSL
CQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFN
NEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQD
YRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKH
CWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

SEQ ID NO:7 Human APOBEC-3F
MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMC
FLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLS
QAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNL
RKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTS
WSPCPECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVY
NDDEPFKPWKGLKYNFLFLDSKLQEILE

SEQ ID NO:8 Human APOBEC-3B:
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEM
CFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRL
SQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPL
VLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVT
WFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCW
DTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

SEQ ID NO:9 Human APOBEC-3C:
MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAER
CFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSL
SQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ

SEQ ID NO:10 Human APOBEC-3A:
MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGR
HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKE
ALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

SEQ ID NO:11 Human APOBEC-3H:

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLD
ETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGF
PKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV

SEQ ID NO:12 Human APOBEC-3D
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQ
EVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYY
RDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEA
MYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDIL
SPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMG
YKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

SEQ ID NO:13 Human APOBEC-1:
MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKK
FTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSG
VTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFR
LHLQNCHYQTIPPHILLATGLIHPSVAWR

SEQ ID NO:14 Human ADAT-2:
MEAKAAPKPAASGACSVSAEETEKWMEEAMHMAKEALENTEVPVGCLMVYNNEVVGKGRNEVNQTKNAT
RHAEMVAIDQVLDWCRQSGKSPSEVFEHTVLYVTVEPCIMCAAALRLMKIPLVVYGCQNERFGGCGSVL
NIASADLPNTGR PFQCIPGYRAEEAVEMLKTFYKQENPNAPKSKVRKKECQKS

SEQ ID NO:15 Human ADAT-1:
MWTADEIAQLCYEHYGIRLPKKGKPEPNHEWTLLAAVVKIQSPADKACDTPDKPVQVTKEVVSMGTGTK
CIGQSKMRKNGDILNDSHAEVIARRSFQRYLLHQLQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSS
HTPCGDASIIPMLEFEDQPCCPVFRNWAHNSSVEASSNLEAPGNERKCEDPDSPVTKKMRLEPGTAARE
VTNGAAHHQSFGKQKSGPISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSGKPGA
AFHQVGLLRVKPGRGDRTRSMSCSDKMARWNVLGCQGALLMHLLEEPIYLSAVVIGKCPYSQEAMQRAL
IGRCQNVSALPKGFGVQELKILQSDLLFEQSRSAVQAKRADSPGRLVPCGAAISWSAVPEQPLDVTANG
FPQGTTKKTIGSLQARSQISKVELFRSFQKLLSRIARDKWPHSLRVQKLDTYQEYKEAASSYQEAWSTL
RKQVFGSWIRNPPDYHQFK

In some embodiments, the deaminase is an adenosine deaminase. In some embodiments, the deaminase is an ADAT family deaminase. In some embodiments, the adenosine deaminase variant is a TadA deaminase. In some embodiments, the adenosine deaminase variant is a *Staphylococcus aureus* TadA, a *Bacillus subtilis* TadA, a *Salmonella typhimurium* TadA, a *Shewanella putrefaciens* TadA, a *Haemophilus influenzae* F3031 TadA, a *Caulobacter crescentus* TadA, or a *Geobacter sulfurreducens* TadA, or a fragment thereof. In some embodiments, the TadA deaminase is an E. coli TadA deaminase (ecTadA). In some embodiments, the TadA deaminase is a truncated E. coli TadA deaminase. For example, the truncated ecTadA may be missing one or more N-terminal amino acids relative to a full-length ecTadA. In some embodiments, the truncated ecTadA may be missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length ecTadA. In some embodiments, the truncated ecTadA may be missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20C- terminal amino acid residues relative to the full length ecTadA. In some embodiments, the TadA deaminase is TadA*7.10. In some embodiments, the TadA deaminase is a TadA*8 variant. For example, deaminase are described

in International PCT Application WO2018/027078, WO2017/070632, WO/2020/168132, WO/2021/050571 each of which is incorporated herein by reference for its entirety. Also, see Komor, A.C., et al.,"Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al.,"Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., et al.,"Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)), and Rees, H.A., et al.,"Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1, the entire contents of which are hereby incorporated by reference. An exemplary amino acid sequence for the wild type TadA(wt) adenosine deaminase is shown as SEQ ID NO: 16. In some embodiments, the amino acid sequence of the adenosine deaminase comprises at least 90% sequence identity to SEQ ID NO:16. In some embodiments, the amino acid sequence of the adenosine deaminase comprises the modification at position 82 as numbered in SEQ ID NO: 16. In some embodiments, the amino acid sequence comprises of the adenosine deaminase comprises a V82S modification, wherein position 82 is as numbered in SEQ ID NO: 16. In some embodiments, the amino acid sequence of the adenosine deaminase comprises the modification at position 166 as numbered in SEQ ID NO:16. In some embodiments, the amino acid sequence of the adenosine deaminase comprises a T166R modification, wherein position 166 is as numbered in SEQ ID NO: 16. In some embodiments, the amino acid sequence of the adenosine deaminase comprises modifications at positions 82 and 166 as numbered in SEQ ID NO: 16. In some embodiments, the amino acid sequence of the adenosine deaminase comprises V82S and T166R modifications, wherein positions 82 and 166 are as numbered in SEQ ID NO: 16. In some embodiments, the adenosine deaminase variant further comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, and Q154R. In some embodiments, the adenosine deaminase variant comprises a combination of alterations selected from the group consisting of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R. In some embodiments, the adenosine deaminase variant is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16,

TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24. In some embodiments, the adenosine deaminase is provided as a single (e.g., provided as a monomer) TadA variant as described above. In some embodiments, adenosine deaminase is provided as a heterodimer of a wild-type TadA (TadA(wt)) linked to a TadA variant as described above.

SEQ ID NO:16 > TadA sequence
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLSDFFRMRRQEIKAQKKAQSSTD

In some embodiments, the deaminase is fused to the N-terminus of the defective CRISPR/Cas nuclease. In some embodiments, the deaminase is fused to the C-terminus of the defective CRISPR/Cas nuclease. In some embodiments, the defective CRISPR/Cas nuclease and the deaminase are fused via a linker. In some embodiments, the linker comprises a (GGGGS)n (SEQ ID NO:17), a (G)n, an (EAAAK)n (SEQ ID NO: 18), a (GGS)n, an SGSETPGTSESATPES (SEQ ID NO: 19) motif (see, e.g., Guilinger J P, Thompson D B, Liu D R. Additional suitable linker motifs and linker configurations will be apparent to those of skill in the art. In some embodiments, suitable linker motifs and configurations include those described in Chen et al., Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013; 65(10):1357-69, the entire contents of which are incorporated herein by reference.

In some embodiments, the fusion protein may comprise additional features. Other exemplary features that may be present are localization sequences, such as nuclear localization sequences (NLS), cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable localization signal sequences and sequences of protein tags are provided herein, and include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable features will be apparent to those of skill in the art.

Various base-editing enzymes are known in the art (see e.g. Improving cytidine and adenine base-editing enzymes by expression optimization and ancestral reconstruction. Nat Biotechnol. 2018 May 29) and typically include those described in Table A.

Table A: some exemplary base-editing enzymes

Base-editing enzyme	References
ABEmax	Improving cytidine and adenine base-editing enzymes by expression optimization and ancestral reconstruction. Nat Biotechnol. 2018 May 29. pii: nbt.4172. doi: 10.1038/nbt.4172.
AncBE4max	Improving cytidine and adenine base-editing enzymes by expression optimization and ancestral reconstruction. Nat Biotechnol. 2018 May 29. pii: nbt.4172. doi: 10.1038/nbt.4172.
evoCDA1-BE4max-NG	Continuous evolution of base-editing enzymes with expanded target compatibility and improved activity. Nat Biotechnol. 2019 Jul 22. pii: 10.1038/s41587-019-0193-0. doi: 10.1038/s41587-019-0193-0.
evoFERNY-BE4max	Continuous evolution of base-editing enzymes with expanded target compatibility and improved activity. Nat Biotechnol. 2019 Jul 22. pii: 10.1038/s41587-019-0193-0. doi: 10.1038/s41587-019-0193-0.
CBE-NRCH	Miller SM, Wang T, Randolph PB, Arbab M, Shen MW, Huang TP, Matuszek Z, Newby GA, Rees HA, Liu DR. Continuous evolution of SpCas9 variants compatible with non-G PAMs. Nat Biotechnol. 2020 Apr;38(4):471-481. doi: 10.1038/s41587-020-0412-8. Epub 2020 Feb 10. PMID: 32042170; PMCID: PMC7145744.
CBE-SpG	Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science. 2020 Apr 17;368(6488):290-296. doi: 10.1126/science.aba8853. Epub 2020 Mar 26. PMID: 32217751; PMCID: PMC7297043.
ABE-SpRY (SEQ ID NO:18)	Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science. 2020 Apr 17;368(6488):290-296. doi: 10.1126/science.aba8853. Epub 2020 Mar 26. PMID: 32217751; PMCID: PMC7297043.

CBE-SpRY (SEQ ID NO:17)	Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science. 2020 Apr 17;368(6488):290-296. doi: 10.1126/science.aba8853. Epub 2020 Mar 26. PMID: 32217751; PMCID: PMC7297043.
ABE8e (SEQ ID NO: 37)	Lapinaite A, Knott GJ, Palumbo CM, Lin-Shiao E, Richter MF, Zhao KT, Beal PA, Liu DR, Doudna JA. DNA capture by a CRISPR-Cas9-guided adenine base editor. Science. 2020 Jul 31;369(6503):566-571. doi: 10.1126/science.abb1390. PMID: 32732424.
ABE8e-SpRY (SEQ ID NO: 38)	High-throughput continuous evolution of compact Cas9 variants targeting single-nucleotide-pyrimidine PAMs. Huang TP, Heins ZJ, Miller SM, Wong BG, Balivada PA, Wang T, Khalil AS, Liu DR. Nat Biotechnol. 2022 Sep 8. pii: 10.1038/s41587-022-01410-2. doi: 10.1038/s41587-022-01410-2. 10.1038/s41587-022-01410-2
ABE8e-NRCH (SEQ ID NO: 39)	pCMV-ABE8e-NRCH (Addgene plasmid # 165416 ; http://n2t.net/addgene:165416 ; RRID:Addgene_165416)
NG-ABE8e (SEQ ID NO: 40)	Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, Doudna JA and Liu DR. Nat Biotechnol (2020)

SEQ ID NO:17> amino acid sequence of CBE-SpRY

MKRTADGSEFESPKKKRKVSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQN
TNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLR
DLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFT
IALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAERTRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG
LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKA
IVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT
LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV
ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQK
GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTRLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS

GGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTD
ENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSKRTADGSEFEPKKKRKVGGGGSGATNFSLLKQAGDVEEN
PGPMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCF
SRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYN
SHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLL
EFVTAAGITLGMDELYK

SEQ ID NO:18> amino acid sequence of ABE- SpRY
MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEI
TEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTF
EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAERTRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA
SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV
KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD
KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD
SRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV
LSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTRLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSKRTADGSEFEPKKK
RKVGSGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
TGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNH
YLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKSGGSPKKKRKV

SEQ ID NO:37> amino acid sequence of ABE8e
MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAGSLMNVLNYPGMNHRVEI
TEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGL
AIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSKRTADGSEFEPKKKRKVGSGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDV

NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT
IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKSGGSPKKKR
KV

SEQ ID NO:38> amino acid sequence of ABE8e-SpRY
MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAGSLMNVLNYPGMNHRVEI
TEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGL
AIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAERTRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGF
LWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTRLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSKRTADGSEFEPKKKRKVGSGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDV
NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT
IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKSGGSPKKKR
KV

SEQ ID NO:39> amino acid sequence of ABE8e-NRCH
MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAGSLMNVLNYPGMNHRVEI
TEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGL
TIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGGHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGF
NSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETR
IDLSQLGGDSGGSKRTADGSEFEPKKKRKVGSGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDV
NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT
IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKSGGSPKKKR
KV

SEQ ID NO:40> amino acid sequence of NG-ABE8e
MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAGSLMNVLNYPGMNHRVEI
TEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGL
AIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGF
VSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSKRTADGSEFEPKKKRKVGSGATNFSLLKQAGDVEENPGPMVSKGEELFTGVVPILVELDGDV
NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERT
IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELX

The second component of the gene-editing platform disclosed herein consists of at least one guide RNA molecule suitable for guiding the base-editing enzyme to at least one target sequence located in the +55-kb region of the erythroid-specific BCL11A enhancer. The guide RNA molecule of the present invention thus comprises a guide sequence for providing the targeting specificity. It includes a region that is complementary and capable of hybridization to a pre-selected target site of interest in the +55-kb region of the erythroid-specific BCL11A enhancer. According to the present disclosure, the guide RNA targets the ATF4 binding site so as to edit said site and thus disrupting the binding of ATF4 to its binding site.

In some embodiment, this guide sequence can comprise from about 10 nucleotides to more than about 25 nucleotides. For example, the region of base pairing between the guide sequence and the corresponding target site sequence can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more than 25 nucleotides in length. In some embodiments, the guide sequence is about 17-20 nucleotides in length, such as 20 nucleotides.

Typically, a software program is used to identify candidate CRISPR target sequences on both strands of the DNA nucleic acid molecule based on desired guide sequence length and a CRISPR motif sequence (PAM) for a specified CRISPR enzyme. One requirement for selecting

a suitable target nucleic acid is that it has a 3′ PAM site/sequence. Each target sequence and its corresponding PAM site/sequence are referred herein as a Cas-targeted site. Type II CRISPR system, one of the most well characterized systems, needs only Cas 9 protein and a guide RNA complementary to a target sequence to affect target cleavage. For example, target sites for Cas9 from *S. pyogenes*, with PAM sequences NGG, may be identified by searching for 5′-Nx-NGG-3′ both on the input sequence and on the reverse-complement of the input. Since multiple occurrences in the genome of the DNA target site may lead to nonspecific genome editing, after identifying all potential sites, the program filters out sequences based on the number of times they appear in the relevant reference genome. For those CRISPR enzymes for which sequence specificity is determined by a "seed" sequence, such as the 11-12 bp 5′ from the PAM sequence, including the PAM sequence itself, the filtering step may be based on the seed sequence. Thus, to avoid editing at additional genomic loci, results are filtered based on the number of occurrences of the seed:PAM sequence in the relevant genome. The user may be allowed to choose the length of the seed sequence. The user may also be allowed to specify the number of occurrences of the seed:PAM sequence in a genome for purposes of passing the filter. The default is to screen for unique sequences. Filtration level is altered by changing both the length of the seed sequence and the number of occurrences of the sequence in the genome. The program may in addition or alternatively provide the sequence of a guide sequence complementary to the reported target sequence(s) by providing the reverse complement of the identified target sequence(s). Further details of methods and algorithms to optimize sequence selection can be found in U.S. application Ser. No. 61/836,080; incorporated herein by reference.

In some embodiment, the gene editing platform comprising a) a cytidine base-editing enzyme and b) and at least one guide RNA molecule suitable for introducing:

- one C>T mutation in the first cytidine residue of SEQ ID NO: 45 (TTGCATCATCC) and/or
- one C>T mutation in the second cytidine residue of SEQ ID NO: 45 (TTGCATCATCC) and/or
- one C>T mutation in the third cytidine residue of SEQ ID NO: 45 (TTGCATCATCC) and/or
- one C>T mutation in the fourth cytidine SEQ ID NO: 45 (TTGCATCATCC).

In some embodiments, the gene editing platform comprises a) a cytidine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 CBE I editing profile of the ATF4 binding site (i.e. TTGTATTATTT (SEQ ID NO:32)).

In some embodiments, the gene editing platform comprises a) a cytidine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 CBE II editing profile in the ATF4 binding site (i.e. TTGCATTATTT (SEQ ID NO:33)).

In some embodiment, the gene editing platform comprising a) a adenine base-editing enzyme and b) and at least one guide RNA molecule suitable for introducing:

- one A>G mutation in the first adenine residue in SEQ ID NO: 45 (TTGCATCATCC) and/or
- one A>G mutation in the second adenine residue of SEQ ID NO: 45 (TTGCATCATCC).

In some embodiments, the gene editing platform comprises a) an adenine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 ABE II profile in the ATF4 binding site (TTGCATCGTCC (SEQ ID NO:35)).

In some embodiments, the gene editing platform comprises a) an adenine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 ABE III profile in the ATF4 binding site (TTGCGTCGTCC (SEQ ID NO: 46)).

In some embodiments, the guide RNA targets a sequence selected from **Table 1** (see EXAMPLE).

In some embodiments, the gene editing platform comprises a) a base-editing enzyme that is an ABE-SpRY or a CBE-SpRY and b) and at least one guide RNA molecule that targets one sequence selected in **Table 1 or Table 4**.

In some embodiments, the gene editing platform comprises a) a base-editing enzyme and b) and at least one guide RNA molecule selected according to the combinations described in **Table 4**.

The guide RNA molecule of the present invention can be made by various methods known in the art including cell-based expression, in vitro transcription, and chemical synthesis. The ability to chemically synthesize relatively long RNAs (as long as 200 mers or more) using TC-RNA chemistry (see, e.g., U.S. Pat. No. 8,202,983) allows one to produce RNAs with special features that outperform those enabled by the basic four ribonucleotides (A, C, G and U). In particular, the RNA molecule of the present invention can be made with recombinant technology using a host cell system or an in vitro translation-transcription system known in the art. Details of such systems and technology can be found in e.g., WO2014144761 WO2014144592, WO2013176772, US20140273226, and US20140273233, the contents of which are incorporated herein by reference in their entireties.

In some embodiments, the guide RNA molecule may include one or more modifications. Such modifications may include inclusion of at least one non-naturally occurring nucleotide, or a modified nucleotide, or analogs thereof. Modified nucleotides may be modified at the ribose, phosphate, and/or base moiety. Modified nucleotides may include 2'-O-methyl analogs, 2'-deoxy analogs, or 2'-fluoro analogs. The nucleic acid backbone may be modified, for example, a phosphorothioate backbone may be used. The use of locked nucleic acids (LNA) or bridged nucleic acids (BNA) may also be possible. Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine, inosine, 7-methylguanosine.

In some embodiments, a plurality of guide RNA molecules are designed for targeting a plurality of sequences in the +55-kb region of the erythroid-specific BCL11A enhancer. In some embodiments, the gene editing platform disclosed herein thus comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 20 guide RNA molecules as disclosed herein.

In some embodiments, a plurality of base-editing enzyme along with a plurality of guide RNA molecules are designed for targeting a plurality of sequences in the +55-kb region of the erythroid-specific BCL11A enhancer. In some embodiments, the gene editing platform disclosed herein thus comprises 2, 3 or 4 base-editing enzymes and 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 20 RNA molecules as disclosed herein.

In some embodiments, the different components of the gene editing platform of the present invention are provided to the eukaryotic cell through expression from one or more expression vectors. For example, the nucleic acids encoding the guide RNA molecule or the base-editing

enzyme can be cloned into one or more vectors for introducing them into the eukaryotic cell. The vectors are typically prokaryotic vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding the guide RNA molecule or the base-editing enzyme herein disclosed. Preferably, the nucleic acids are isolated and/or purified. Thus, the present invention provides recombinant constructs or vectors having sequences encoding one or more of the guide RNA molecule or base-editing enzymes described above. Examples of the constructs include a vector, such as a plasmid or viral vector, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In some embodiments, the construct further includes regulatory sequences. A "**regulatory sequence**" includes promoters, enhancers, and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as inducible regulatory sequences. The design of the expression vector can depend on such factors as the choice of the eukaryotic cell to be transformed, transfected, or infected, the desired expression level, and the like. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Appropriate cloning and expression vectors for use with eukaryotic hosts are also described in e.g., Sambrook et al. (2001, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press). The vector can be capable of autonomous replication or integration into a host DNA. The vector may also include appropriate sequences for amplifying expression. In addition, the expression vector preferably contains one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell cultures, or such as tetracycline or ampicillin resistance in *E. coli*. Any of the procedures known in the art for introducing foreign nucleotide sequences into host cells may be used. Examples include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell.

In some embodiments, the different components of the gene editing platform of the present invention are provided to the population of cells through the use of an RNA-encoded system. For instance the CBE-SpRY and the ABE-SpRY are provided by mRNA sequences encoding proteins such as SEQ ID NO:17 and SEQ ID NO:18 respectively. In particular, the sequences of said mRNAs are provided as SEQ ID NO:30 and SEQ ID NO:31. In particular, the base-

editing system may be provided to the population of cells through the use of a chemically modified mRNA-encoded adenine or cytidine base editor together with modified guide RNA as described in *Jiang, T., Henderson, J.M., Coote, K. et al. Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. Nat Commun 11, 1979 (2020).* In particular, engineered RNA-encoded base-editing enzymes (e.g. ABE) system are prepared by introducing various chemical modifications to both mRNA that encoded the base-editing enzyme and guide RNA. In particular said modifications consist in uridine depleted mRNAs modified with 5-methoxyuridine: synonymous codons may be introduced to deplete uridines as much as possible without altering the coding sequence and replaced all the remaining uridines with 5-methoxyuridine. Said optimized base editing system exhibits higher editing efficiency at some genomic sites compared to DNA-encoded system. It is also possible to encapsulate the modified mRNA and guide RNA into lipid nanoparticle (LNP) for allowing lipid nanoparticle (LNP)-mediated delivery.

In some embodiments, the different components of the gene editing platform of the present invention are provided to the population of cells through the use of ribonucleoprotein (RNP) complexes. For instance. the base-editing enzyme can be pre-complexed with one or more guide RNA molecules to form a ribonucleoprotein (RNP) complex. The RNP complex can thus be introduced into the eukaryotic cell. Introduction of the RNP complex can be timed. The cell can be synchronized with other cells at G1, S, and/or M phases of the cell cycle. RNP delivery avoids many of the pitfalls associated with mRNA, DNA, or viral delivery. Typically, the RNP complex is produced simply by mixing the proteins (i.e. the base-editing enzyme) and one or more guide RNA molecules in an appropriate buffer. This mixture is incubated for 5-10 min at room temperature before electroporation. Electroporation is a delivery technique in which an electrical field is applied to one or more cells in order to increase the permeability of the cell membrane. In some embodiments, genome editing efficiency can be improved by adding a transfection enhancer oligonucleotide.

In some embodiments, a plurality of successive transfections are performed for reaching a desired level of mutagenesis in the cell.

A further object of the present invention relates to a method of treating a β-hemoglobinopathy in a subject in need thereof, the method comprising transplanting a therapeutically effective amount of a population of eukaryotic cells obtained by the method as above described.

In some embodiments, the population of cell is autologous to the subject, meaning the population of cells is derived from the same subject.

In some embodiments, the β-hemoglobinopathy is a sickle cell disease.

In some embodiments, the β-hemoglobinopathy is a β-thalassemia.

Kits

This invention further provides kits containing reagents for performing the above-described methods, including all component of the gene editing platform as disclosed herein for performing mutagenesis. To that end, one or more of the reaction components, e.g., guide RNA molecules, and nucleic acid molecules encoding for the base-editing enzymes for the methods disclosed herein can be supplied in the form of a kit for use. In some embodiments, the kit comprises one or more base-editing enzymes and one or more guide RNA molecules. In some embodiments, the kit can include one or more other reaction components. In some embodiments, an appropriate amount of one or more reaction components is provided in one or more containers or held on a substrate. Examples of additional components of the kits include, but are not limited to, one or more host cells, one or more reagents for introducing foreign nucleotide sequences into host cells, one or more reagents (e.g., probes or PCR primers) for detecting expression of the guide RNA or base-editing enzymes or verifying the target nucleic acid's status, and buffers or culture media for the reactions. The kit may also include one or more of the following components: supports, terminating, modifying or digestion reagents, osmolytes, and an apparatus for detection. The components used can be provided in a variety of forms. For example, the components (e.g., enzymes, RNAs, probes and/or primers) can be suspended in an aqueous solution or as a freeze-dried or lyophilized powder, pellet, or bead. In the latter case, the components, when reconstituted, form a complete mixture of components for use in an assay. The kits of the invention can be provided at any suitable temperature. For example, for storage of kits containing protein components or complexes thereof in a liquid, it is preferred that they are provided and maintained below 0° C., preferably at or below −20° C., or otherwise in a frozen state. The kits can also include packaging materials for holding the container or combination of containers. Typical packaging materials for such kits and systems include solid matrices (e.g., glass, plastic, paper, foil, micro-particles and the like) that hold the

reaction components or detection probes in any of a variety of configurations (e.g., in a vial, microtiter plate well, microarray, and the like). The kits may further include instructions recorded in a tangible form for use of the components.

The invention will be further illustrated by the following figures and examples. However, these examples and figures should not be interpreted in any way as limiting the scope of the present invention.

FIGURES:

Figure 1. sgRNA design to target the +55-kb region of the erythroid-specific *BCL11A* enhancer.

A-B. Schematic representation of part of the *BCL11A* gene on chromosome 2, depicting exons 2, 3 and 4 (Ex 2, 3 and 4) and the DNaseI hypersensitive sites +62-kb, +58-kb and +55-kb. The sequence of the +55-kb region of the erythroid-specific *BCL11A* enhancer is depicted. ATF4 transcriptional activator BS is highlighted in bold. Target sequences of the sgRNAs used with base editing enzymes, are reported (highlighted with arrows) and aligned with the DNA sequence that they bind to in a stranded-oriented way. Oval shape indicates transcriptional activator and ATF4.

Figure 2. RNA-mediated base editing of the erythroid-specific *BCL11A* enhancer in SCD HPSCs-derived erythroblasts.

A. Experimental protocol used for base editing experiments in SCD HSPCs. A base editor mRNA and a sgRNA were co-transfected in SCD HSPCs. Cells were differentiated into mature RBCs using a three-phase erythroid differentiation protocol or plated in methylcellulose-containing medium under conditions supporting erythroid (BFU-E) and granulo-monocytic (CFU-GM) differentiation.

B. C-G to T-A or A-T to G-C base editing efficiency, calculated by the EditR software, in erythroblasts differentiated from SCD HSPCs edited in the +55-kb region and subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). CBE I and II profiles were generated by CBE-SpRY and ATF4_BS_1 and ATF4_BS_2 sgRNA respectively, and ABE I and II profiles were generated by ABE-SpRY and ATF4_BS_1 and ATF4_BS_2 sgRNA respectively.

D. Frequency of InDels, measured by TIDE analysis, in control, base- and Cas9- edited samples subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). **** p≤0.0001 (Ordinary One-way ANOVA).

Figure 3. **RNA-mediated base editing of *BCL11A* enhancer in SCD HSPCs does not affect progenitors and reactivates HbF in erythroid colonies**.

A. C-G to T-A or A-T to G-C base editing efficiency, calculated by the EditR software, in BFU-E and CFU-GM pooled colonies subjected to Sanger sequencing. BFU-E and CFU-GM derived from SCD HSPCs edited in the +55-kb region. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). +55: CBE I and II profiles were generated by CBE-SpRY, and ABE I and II profiles were generated by ABE-SpRY.

B-C. Frequency of InDels, measured by TIDE analysis, for control, base- and Cas9- edited BFU-E (**B**) and CFU-GM (**C**) pooled samples subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). *** p≤0.001; **** p≤0.0001 (Ordinary One-way ANOVA).

D. CFC frequency for control and edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

E. Analysis of HbF and HbS by CE-HPLC in BFU-E. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

F. Expression of γ-globin chains measured by RP-HPLC in BFU-E. γ-globin expression was normalized to α-globin. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

Figure 4. **Disruption of the erythroid-specific *BCL11A* enhancer does not impact erythroid differentiation process of SCD HSPCs**.

A. Frequency of enucleated cells at day 6, 13, 16 and 19 of erythroid differentiation, as measured by flow cytometry analysis of DRAQ5 nuclear staining in control (untreated, or mock-transfected with TE buffer, or transfected with a BE mRNA only, or transfected with a BE mRNA and a sgRNA targeting the unrelated *AAVS1* locus) and edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

B-D. Frequency of $CD71^+$ (B), $CD36^+$ (C) and GPA^+ (D) cells at day 6, 13 and 19 of erythroid differentiation, as measured by flow cytometry analysis. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

E. Frequency of α4-$Integrin^+$, $BAND3^+$ and α4-$Integrin^+$/$BAND3^+$ in $7AAD^-$/GPA^+ cells at day 6, 13 and 19 of erythroid differentiation, as measured by flow cytometry analysis. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

Figure 5. **Disruption of the erythroid-specific *BCL11A* enhancer reactivates HbF and ameliorates the sickling phenotype in RBCs derived from base edited SCD HSPCs**.

A. Analysis of HbF and HbS by CE-HPLC in SCD patient RBCs. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). * $p \leq 0.05$; *** $p \leq 0.001$ (Two-way ANOVA).

B. RT-qPCR analysis of β^S- and γ-globin mRNA levels in SCD patient erythroblasts at day 13 of erythroid differentiation. β^S- and γ-globin mRNA expression was normalized to α-globin mRNA and expressed as percentage of the β^S-+γ- globins mRNA. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). ** $p \leq 0.01$; **** $p \leq 0.0001$ (Two-way ANOVA).

C. Expression of γ-globin chains measured by RP-HPLC in SCD patient RBCs at the end of the erythroid differentiation. γ-globin expression was normalized to α-globin. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

D. Frequency of HbF and HbS expressing cells in GPA^+ population for control and edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

E. Frequency of sickling cells upon O_2 deprivation in control and edited samples. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors).

Figure 6. HbF reactivation in single erythroid progenitors.

A. *HBG* mRNA relative expression in single BFU-E colonies (Ctrl n=23; CBE I mono n=9; CBE I bi n=11; CBE II mono n=4; CBE II bi n=11; ABE I mono n=9; ABE I bi n=15; ABE II mono n=6; ABE II bi n=2; 1 donor) bearing different editing profiles and either monoallelic (mono) or biallelic (bi) editing. *HBG* mRNA expression was normalized to *HBA1/2* mRNA and expressed as percentage of the total *HBB*+*HBG* mRNA. BFU-Es derived from SCD HSPCs that were either mock-transfected with TE buffer, or transfected with a BE mRNA only, or transfected with a BE mRNA and a sgRNA targeting the unrelated *AAVS1* locus were used as

negative controls (Ctrl). * p≤0.05; ** p≤0.01; *** p≤0.001; **** p≤0.0001 (One-way ANOVA).

B. Correlation between *HBG* mRNA relative expression and base-editing efficiency in single BFU-E colonies (Ctrl n=23; CBE I n=20; CBE II n=15; ABE I n=24; ABE II n=8; 1 donor) bearing different editing profiles. *HBG* mRNA expression was normalized to *HBA1/2* mRNA and expressed as percentage of the total *HBB+HBG* mRNA. Base-editing efficiency was calculated by the EditR software in samples subjected to Sanger sequencing. BFU-Es derived from SCD HSPCs that were either mock-transfected with TE buffer, or transfected with a BE mRNA only, or transfected with a BE mRNA and a sgRNA targeting the unrelated *AAVS1* locus were used as negative controls. ns p>0.05; * p≤0.05; ** p≤0.01; *** p≤0.001; **** p≤0.0001 (CBE I: R^2=0.3989, $Y = 0.4685*X + 29.98$, p<0.0001 non-zero slope significancy; CBE II: R^2=0.4072, $Y = 0.3532*X + 31.75$, p<0.0001 non-zero slope significancy; ABE I: R^2=0.07206, $Y = 0.1431*X + 31.24$, p=0.0651 non-zero slope significancy; ABE II: R^2=0.4710, $Y = 0.7134*X + 30.87$, p<0.0001 non-zero slope significancy; Multiple t-test).

Figure 7. Disruption of the erythroid-specific *BCL11A* enhancer in SCD HSPC-derived cells with AncBE4max-OPT.

C-G to T-A base editing efficiency, calculated by the EditR software, in DNA samples obtained from cells cultured in the HSPC medium and subjected to Sanger sequencing. SCD HSPCs were edited in the +55-kb region with either CBE-SpRY or AncBE4max-OPT. Data are expressed as single values (n=1 donor).

Figure 8. Disruption of the erythroid-specific *BCL11A* enhancer in SCD HSPC-derived cells with a highly processive ABE.

A. A-T to G-C base editing efficiency, calculated by the EditR software, in SCD HSPC-derived cells (erythroid liquid cultures, BFU-E colonies, CFU-GM colonies) subjected to Sanger sequencing. SCD HSPCs were edited in the +55-kb region with ABE8e. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors).

B. Analysis of HbF and HbS by CE-HPLC in SCD-derived BFU-E. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). **** p≤0.0001 (Two-way ANOVA).

EXAMPLE:

Methods

HSPC purification and culture

We obtained human non-mobilized peripheral blood $CD34^+$ HSPCs from SCD patients. SCD samples eligible for research purposes were obtained from the "Hôpital Necker-Enfants malades" Hospital (Paris, France). Written informed consent was obtained from all adult subjects. All experiments were performed in accordance with the Declaration of Helsinki. The study was approved by the regional investigational review board (reference: DC 2014-2272, CPP Ile-de-France II "Hôpital Necker-Enfants malades"). HSPCs were purified by immunomagnetic selection with AutoMACS (Miltenyi Biotec) after immunostaining with the CD34 MicroBead Kit (Miltenyi Biotec). Forty-eight hours before transfection, $CD34^+$ cells were thawed and cultured at a concentration of $5x10^5$ cells/ml in the "HSPC medium" containing StemSpan (STEMCELL Technologies) supplemented with penicillin/streptomycin (Gibco), 250 nM StemRegenin1 (STEMCELL Technologies), and the following recombinant human cytokines (PeproTech): human stem cell factor (SCF) (300 ng/ml), Flt-3L (300 ng/ml), thrombopoietin (TPO) (100 ng/ml), and interleukin-3 (IL-3) (60 ng/ml).

Plasmids

Plasmids used in this study include:
pCAG-CBE4max-SpRY-P2A-EGFP (RTW5133) (Addgene #139999),
pCMV-T7-SpRY-P2A-EGFP (RTW4830) (Addgene #139989).
pCMV_AncBE4max_P2A_GFP (Addgene #112100),
ABE8e (Addgene #138489).

A DNA fragment (3'UTR+poly-A) containing two copies of the 3' untranslated region (UTR) of the *HBB* gene and a poly-A sequence of 96 adenines was purchased by Genscript. Similarly, another DNA fragment containing the uridine-depleted coding sequence of pCAG-CBE4max-SpRY-P2A-EGFP was created (CBE-SpRY_U-delp).

The CBE-SpRY-OPT plasmid was created by inserting the 3'UTR+poly-A fragment in the pCAG-CBE4max-SpRY-P2A-EGFP (Addgene #140003) plasmid, and by replacing the CBE4max-SpRY coding sequence with the CBE-SpRY_U-delp fragment. CBE-SpRY-OPT plasmid contains a T7 promoter followed by a G nt allowing efficient capping.

The ABE-SpRY-OPT plasmid was created by inserting the 3'UTR+poly-A fragment in the pCMV-T7-SpRY-P2A-EGFP (RTW4830) (Addgene #139989) plasmid.

The AncBE4max-OPT plasmid was created by inserting point mutations in the PAM-identification domain of the Cas9 nickase of the CBE-SpRY-OPT plasmid that allow the recognition of the NGG PAM and by inserting point mutations in the deaminase domain of the CBE-SpRY-OPT plasmid that allow the ancestral reconstitution of the enzyme.

sgRNA design

We manually designed sgRNAs targeting the +55-kb region of *BCL11A* (Table 1). To generate the sgRNA expression plasmid, oligonucleotides were annealed to create the sgRNA protospacer and the duplexes were ligated into the Bbs I-digested MA128 plasmid (provided by M. Amendola, Genethon, France). For RNA-mediated base editing we used chemically modified synthetic sgRNAs harboring 2′-O-methyl analogs and 3′-phosphorothioate nonhydrolyzable linkages at the first three 5′ and 3′ nucleotides (Synthego).

Table 1. gRNA target sequences.

gRNA	Target sequence (5' to 3')	Position (hg19)	Strand
ATF4 _bs_1	CATTGCATCATCCTGGTACC (SEQ ID NO:19)	chr2: 60725617-60725636	+
ATF4_bs_2	GCATCATCCTGGTACCAGGA (SEQ ID NO:20)	chr2: 60725621-60725640	+
ATF4_bs_3	CTCCAAGCATTGCATCATCC (SEQ ID NO:21)	chr2: 60725610-60725629	+

mRNA in vitro transcription

10 µg of base editor expressing plasmids were digested overnight with 20 Units of a restriction enzyme that cuts once right after the poly-A tail. The linearized plasmids were purified using a PCR purification kit (QIAGEN #28106) and were eluted in 30 µl of DNase/RNase-free water. 1 µg of linearized plasmid was used as template for the in vitro transcription reaction (MEGAscript, Ambion #AM1334). The in vitro transcription protocol was modified as follows. The GTP nucleotide solution was used at a final concentration of 3.0 mM instead of 7.5 mM and the anti-reverse cap analog N7-Methyl-3'-O-Methyl-Guanosine-5'-Triphosphate-5'-Guanosine (ARCA, Trilink #N-7003) was used at a final concentration of 12.0 mM resulting in a final ratio of Cap:GTP of 4:1 that allows efficient capping of the mRNA. The incubation time for the in vitro reaction was reduced to 30 minutes. For constructs without lacking a poly-A tail in the plasmid (ABE8e), an additional step of polyadenylation was performed using manufacturer's guidelines (Poly-A tailing kit, Ambion). mRNA was precipitated using lithium

chloride and resuspended in TE buffer in a final volume that allowed to achieve a concentration of >1 µg/µl. The mRNA quality was checked using Bioanalyzer (Agilent).

RNA transfection

$2x10^5$ $CD34^+$ HSPCs per condition were transfected with 3.0 µg of the enzyme encoding mRNA, respectively, and a synthetic sgRNA at a final concentration of 2.3 µM. We used the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the CA137 program (Nucleofector 4D). Untransfected cells or cells transfected with TE buffer or with the enzyme-encoding mRNA only, or with the enzyme-encoding mRNA and a sgRNA targeting the AAVS1 locus, served as negative controls.

Ribonucleoprotein (RNP) transfection

RNP complexes were assembled at room temperature using a 90 µM Cas9-GFP protein and a 180 µM synthetic sgRNA (ratio Cas9:sgRNA of 1:2). $CD34^+$ HSPCs ($2x10^5$ cells/condition) were transfected with RNP complexes using the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the CA137 program (Nucleofector 4D) in the presence of a transfection enhancer (IDT). Untransfected cells or cells transfected with TE buffer or with the enzyme-encoding mRNA only, or with the enzyme-encoding mRNA and a sgRNA targeting the AAVS1 locus, served as negative controls.

HSPC differentiation

Transfected $CD34^+$ HSPCs were differentiated into mature RBCs using a three-phase erythroid differentiation protocol, as previously described(Giarratana et al., 2005; Weber et al., 2020). During the first phase (day 0 to day 6), cells were cultured in a basal erythroid medium supplemented with 100 ng/ml recombinant human SCF (PeproTech), 5 ng/ml recombinant human IL-3 (PeproTech), 3 IU/ml EPO Eprex (Janssen-Cilag) and 10^{-6} M hydrocortisone (Sigma). During the second phase (day 6 to day 9), cells were co-cultured with MS-5 stromal cells in the basal erythroid medium supplemented with 3 IU/ml EPO Eprex (Janssen-Cilag). During the third phase (day 9 to day 20), cells were co-cultured with stromal MS-5 cells in a basal erythroid medium without cytokines. Erythroid differentiation was monitored by flow cytometry analysis of CD36, CD71, GYPA, BAND3 and α4-Integrin erythroid surface markers and of enucleated cells using the DRAQ5 double-stranded DNA dye. 7AAD was used to identify live cells.

Colony-forming cell (CFC) assay

$CD34^+$ HSPCs were plated at a concentration of $1x10^3$ cells/mL in a methylcellulose-based medium (GFH4435, Stem Cell Technologies) under conditions supporting erythroid and granulo-monocytic differentiation. BFU-E and CFU-GM colonies were counted after 14 days. Colonies were randomly picked and collected as bulk populations (containing at least 25 colonies) to evaluate base editing efficiency, globin expression by RT-qPCR and RP-HPLC and hemoglobin expression by CE-HPLC. BFU-Es were randomly picked and collected as single colonies to evaluate base-editing efficiency and globin expression by RT-qPCR.

Evaluation of editing efficiency

Base editing efficiency and InDel frequency were evaluated in HSPC-derived erythroid cells at the end of the first phase of differentiation and in BFU-E and CFU-GM 14 days after plating. Genomic DNA was extracted from control and edited cells using PURE LINK Genomic DNA Mini kit (LifeTechnologies), or Quick-DNA/RNA Miniprep (ZYMO Research), following manufacturers' instructions. To evaluate base editing efficiency at sgRNA target sites, we performed PCR followed by Sanger sequencing and EditR analysis (Kluesner et al., 2018). TIDE analysis (Tracking of InDels by Decomposition) was also performed in order to evaluate the percentage of InDels in edited samples (Brinkman et al., 2014).

Table 2. Primers used to detect base editing and InDels events.

Amplified region	F/R	Sequence (5' to 3')
+55-kb region	F	CCCTATCAGTGCCGACCAAG (SEQ ID NO: 22)
	R	GTGAGTAGGTAGAGGGGTCAGA (SEQ ID NO:23)

F, forward primer; R, reverse primer.

RT-qPCR

Total RNA was extracted from SCD HSPCs differentiated towards the erythroid lineage (day 13) using RNeasy micro kit (QIAGEN), and from BFU-E pools using Quick-DNA/RNA Miniprep (ZYMO Research). RNA was treated with DNase using the DNase I kit (Invitrogen), following manufacturer's instructions. Mature transcripts were reverse-transcribed using SuperScript First-Strand Synthesis System for RT-qPCR (Invitrogen) with oligo (dT) primers. RT-qPCR was performed using the iTaq universal SYBR Green master mix (Biorad) and the Viia7 Real-Time PCR system (ThermoFisher Scientific), or the CFX384 Touch Real-Time PCR Detection System (Biorad).

Table 3. Primers used for RT-qPCR.

Amplified region	F/R	Sequence (5' to 3')
HBA	F	CGGTCAACTTCAAGCTCCTAA (SEQ ID NO:24)
	R	ACAGAAGCCAGGAACTTGTC (SEQ ID NO:25)
HBB	F	GCAAGGTGAACGTGGATGAAGT (SEQ ID NO:26)
	R	TAACAGCATCAGGAGTGGACAGA (SEQ ID NO:27)
HBG1+HBG2	F	CCTGTCCTCTGCCTCTGCC (SEQ ID NO:28)
	R	GGATTGCCAAAACGGTCAC (SEQ ID NO:29)

F, forward primer; R, reverse primer.

Flow cytometry analysis

HSPC-derived erythroid cells were fixed with 0.05 % cold glutaraldehyde and permeabilized with 0.1 % TRITON X-100. After fixation and permeabilization, cells were stained with an antibody recognizing GYPA erythroid surface marker (PE-Cy7-conjugated anti-GYPA antibody, 563666, BD Pharmingen) and either an antibody recognizing HbF (FITC-conjugated anti-HbF antibody, clone 2D12 552829 BD), or an antibody recognizing HbS (anti-HbS antibody, H04181601, BioMedomics) followed by the staining with a secondary antibody recognizing rabbit IgG (BV421-conjugated anti-rabbit IgG, 565014, BD). Flow cytometry analysis of CD36, CD71, GYPA, BAND3 and α4-Integrin erythroid surface markers was performed using a V450-conjugated anti-CD36 antibody (561535, BD Horizon), a FITC-conjugated anti-CD71 antibody (555536, BD Pharmingen), a PE-Cy7-conjugated anti-GYPA antibody (563666, BD Pharmingen), a PE-conjugated anti-BAND3 antibody (9439, IBGRL) and an APC-conjugated anti-CD49d antibody (559881, BD). Flow cytometry analysis of enucleated or viable cells was performed using double-stranded DNA dyes (DRAQ5, 65-0880-96, Invitrogen and 7AAD, 559925, BD, respectively). Flow cytometry analyses were performed using Fortessa X20 (BD Biosciences) or Gallios (Beckman coulter) flow cytometers. Data were analyzed using the FlowJo (BD Biosciences) software.

RP-HPLC analysis of globin chains

RP-HPLC analysis was performed using a NexeraX2 SIL-30AC chromatograph and the LC Solution software (Shimadzu). A 250x4.6 mm, 3.6 µm Aeris Widepore column (Phenomenex) was used to separate globin chains by HPLC. Samples were eluted with a gradient mixture of solution A (water/acetonitrile/trifluoroacetic acid, 95:5:0.1) and solution B (water/acetonitrile/trifluoroacetic acid, 5:95:0.1). The absorbance was measured at 220 nm.

CE-HPLC analysis of hemoglobin tetramers

Cation-exchange HPLC analysis was performed using a NexeraX2 SIL-30AC chromatograph and the LC Solution software (Shimadzu). A 2 cation-exchange column (PolyCAT A, PolyLC, Columbia, MD) was used to separate hemoglobin tetramers by HPLC. Samples were eluted with a gradient mixture of solution A (20mM bis Tris, 2mM KCN, pH=6.5) and solution B (20mM bis Tris, 2mM KCN, 250mM NaCl, pH=6.8). The absorbance was measured at 415 nm.

Sickling assay

HSPC-derived mature RBCs obtained at the end of the erythroid differentiation, were incubated under gradual hypoxic conditions (20% O_2 for 20 min; 10% O_2 for 20 min; 5% O_2 for 20 min; 0% O_2 for 60-80 min) and a time course analysis of sickling was performed in real time by video microscopy. Images were captured every 20 min using an AxioObserver Z1 microscope (Zeiss) and a 40x objective. Throughout the time course, images were captured and then processed with ImageJ to determine the percentage of non-sickle RBCs per field of acquisition in the total RBC population. More than 400 cells were counted per condition.

Results

RNA-mediated base editing in SCD HSPCs disrupts the +55-kb region of the erythroid-specific *BCL11A* enhancer

To disrupt the *BCL11A* erythroid enhancer, we targeted the ATF4 BS in the +55-kb region (**Figure 1**). We mainly took advantage of CBE-SpRY and ABE-SpRY enzymes, given their PAMless nature that allowed us to design sgRNAs by placing the base editing window within different strands and positions of the ATF4 BS, thus allowing different types of conversions (**Figure 1**). To establish a clinically relevant method to deliver the base editing system in primary HSPCs and achieve high editing efficiencies coupled with minimal toxicity, we optimized a protocol based on transfection of mRNA encoding base editors and synthetic modified sgRNAs. First, we optimised the plasmids encoding CBE-SpRY and ABE-SpRY for *in vitro* transcription and mRNA production. In particular, we inserted two copies of the 3' untranslated region (UTR) of the *HBB* gene (which has been shown to increase the half-life of mRNA and improve protein levels(Ross and Sullivan, 1985; Karikó et al., 1999; Holtkamp et al., 2006)) and a poly-A sequence after the 3' UTR to further stabilize the mRNA(Gallie, 1991) in CBE-SpRY and ABE-SpRY constructs (SEQ ID NO:30 and SEQ ID NO:31). The optimized

plasmids, CBE-SpRY-OPT and ABE-SpRY-OPT were used for *in vitro* transcription and mRNA production.

SEQ ID NO:30 > mRNA sequence encoding for CBE-SpRY. The different regions of the mRNA are indicated as follows :*Kozak sequence* - **CBE-SpRY coding sequence** - *HBB 3'UTR (1st copy)* - *HBB 3'UTR (2nd copy)* - Poly-a tail - Residual nucleotides (till the restriction enzyme site)

*GGGAGAGCCGCCACC***AUGAAGCGGACCGCCGACGGCAGCGAGUUCGAGAGCCCCAAGAAGAAGCGGAAGGUGAGC**
AGCGAGACCGGCCCCGUGGCCGUGGACCCCACCCUGCGGCGGCGGAUCGAGCCCCACGAGUUCGAGGUGUUCUUC
GACCCCCGGGAGCUGCGGAAGGAGACCUGCCUGCUGUACGAGAUCAACUGGGGCGGCCGGCACAGCAUCUGGCGG
CACACCAGCCAGAACACCAACAAGCACGUGGAGGUGAACUUCAUCGAGAAGUUCACCACCGAGCGGUACUUCUGC
CCCAACACCCGGUGCAGCAUCACCUGGUUCCUGAGCUGGAGCCCCUGCGGCGAGUGCAGCCGGGCCAUCACCGAG
UUCCUGAGCCGGUACCCCCACGUGACCCUGUUCAUCUACAUCGCCCGGCUGUACCACCACGCCGACCCCCGGAAC
CGGCAGGGCCUGCGGGACCUGAUCAGCAGCGGCGUGACCAUCCAGAUCAUGACCGAGCAGGAGAGCGGCUACUGC
UGGCGGAACUUCGUGAACUACAGCCCCAGCAACGAGGCCCACUGGCCCCGGUACCCCCACCUGUGGGUGCGGCUG
UACGUGCUGGAGCUGUACUGCAUCAUCCUGGGCCUGCCCCCCUGCCUGAACAUCCUGCGGCGGAAGCAGCCCCAG
CUGACCUUCUUCACCAUCGCCCUGCAGAGCUGCCACUACCAGCGGCUGCCCCCCCACAUCCUGUGGGCCACCGGC
CUGAAGAGCGGCGGCAGCAGCGGCGGCAGCAGCGGCAGCGAGACCCCCGGCACCAGCGAGAGCGCCACCCCCGAG
AGCAGCGGCGGCAGCAGCGGCGGCAGCGACAAGAAGUACAGCAUCGGCCUGGCCAUCGGCACCAACAGCGUGGGC
UGGGCCGUGAUCACCGACGAGUACAAGGUGCCCAGCAAGAAGUUCAAGGUGCUGGGCAACACCGACCGGCACAGC
AUCAAGAAGAACCUGAUCGGCGCCCUGCUGUUCGACAGCGGCGAGACCGCCGAGCGGACCCGGCUGAAGCGGACC
GCCCGGCGGCGGUACACCCGGCGGAAGAACCGGAUCUGCUACCUGCAGGAGAUCUUCAGCAACGAGAUGGCCAAG
GUGGACGACAGCUUCUUCCACCGGCUGGAGGAGAGCUUCCUGGUGGAGGAGGACAAGAAGCACGAGCGGCACCCC
AUCUUCGGCAACAUCGUGGACGAGGUGGCCUACCACGAGAAGUACCCCACCAUCUACCACCUGCGGAAGAAGCUG
GUGGACAGCACCGACAAGGCCGACCUGCGGCUGAUCUACCUGGCCCUGGCCCACAUGAUCAAGUUCCGGGGCCAC
UUCCUGAUCGAGGGCGACCUGAACCCCGACAACAGCGACGUGGACAAGCUGUUCAUCCAGCUGGUGCAGACCUAC
AACCAGCUGUUCGAGGAGAACCCCAUCAACGCCAGCGGCGUGGACGCCAAGGCCAUCCUGAGCGCCCGGCUGAGC
AAGAGCCGGCGGCUGGAGAACCUGAUCGCCCAGCUGCCCGGCGAGAAGAAGAACGGCCUGUUCGGCAACCUGAUC
GCCCUGAGCCUGGGCCUGACCCCCAACUUCAAGAGCAACUUCGACCUGGCCGAGGACGCCAAGCUGCAGCUGAGC
AAGGACACCUACGACGACGACCUGGACAACCUGCUGGCCCAGAUCGGCGACCAGUACGCCGACCUGUUCCUGGCC
GCCAAGAACCUGAGCGACGCCAUCCUGCUGAGCGACAUCCUGCGGGUGAACACCGAGAUCACCAAGGCCCCCCUG
AGCGCCAGCAUGAUCAAGCGGUACGACGAGCACCACCAGGACCUGACCCUGCUGAAGGCCCUGGUGCGGCAGCAG
CUGCCCGAGAAGUACAAGGAGAUCUUCUUCGACCAGAGCAAGAACGGCUACGCCGGCUACAUCGACGGCGGCGCC
AGCCAGGAGGAGUUCUACAAGUUCAUCAAGCCCAUCCUGGAGAAGAUGGACGGCACCGAGGAGCUGCUGGUGAAG
CUGAACCGGGAGGACCUGCUGCGGAAGCAGCGGACCUUCGACAACGGCAGCAUCCCCCACCAGAUCCACCUGGGC
GAGCUGCACGCCAUCCUGCGGCGGCAGGAGGACUUCUACCCCUUCCUGAAGGACAACCGGGAGAAGAUCGAGAAG
AUCCUGACCUUCCGGAUCCCCUACUACGUGGGCCCCCUGGCCCGGGGCAACAGCCGGUUCGCCUGGAUGACCCGG
AAGAGCGAGGAGACCAUCACCCCCUGGAACUUCGAGGAGGUGGUGGACAAGGGCGCCAGCGCCCAGAGCUUCAUC
GAGCGGAUGACCAACUUCGACAAGAACCUGCCCAACGAGAAGGUGCUGCCCAAGCACAGCCUGCUGUACGAGUAC
UUCACCGUGUACAACGAGCUGACCAAGGUGAAGUACGUGACCGAGGGCAUGCGGAAGCCCGCCUUCCUGAGCGGC
GAGCAGAAGAAGGCCAUCGUGGACCUGCUGUUCAAGACCAACCGGAAGGUGACCGUGAAGCAGCUGAAGGAGGAC
UACUUCAAGAAGAUCGAGUGCUUCGACAGCGUGGAGAUCAGCGGCGUGGAGGACCGGUUCAACGCCAGCCUGGGC
ACCUACCACGACCUGCUGAAGAUCAUCAAGGACAAGGACUUCCUGGACAACGAGGAGAACGAGGACAUCCUGGAG
GACAUCGUGCUGACCCUGACCCUGUUCGAGGACCGGGAGAUGAUCGAGGAGCGGCUGAAGACCUACGCCCACCUG
UUCGACGACAAGGUGAUGAAGCAGCUGAAGCGGCGGCGGUACACCGGCUGGGGCCGGCUGAGCCGGAAGCUGAUC
AACGGCAUCCGGGACAAGCAGAGCGGCAAGACCAUCCUGGACUUCCUGAAGAGCGACGGCUUCGCCAACCGGAAC
UUCAUGCAGCUGAUCCACGACGACAGCCUGACCUUCAAGGAGGACAUCCAGAAGGCCCAGGUGAGCGGCCAGGGC
GACAGCCUGCACGAGCACAUCGCCAACCUGGCCGGCAGCCCCGCCAUCAAGAAGGGCAUCCUGCAGACCGUGAAG
GUGGUGGACGAGCUGGUGAAGGUGAUGGGCCGGCACAAGCCCGAGAACAUCGUGAUCGAGAUGGCCCGGGAGAAC
CAGACCACCCAGAAGGGCCAGAAGAACAGCCGGGAGCGGAUGAAGCGGAUCGAGGAGGGCAUCAAGGAGCUGGGC
AGCCAGAUCCUGAAGGAGCACCCCGUGGAGAACACCCAGCUGCAGAACGAGAAGCUGUACCUGUACUACCUGCAG
AACGGCCGGGACAUGUACGUGGACCAGGAGCUGGACAUCAACCGGCUGAGCGACUACGACGUGGACCACAUCGUG
CCCCAGAGCUUCCUGAAGGACGACAGCAUCGACAACAAGGUGCUGACCCGGAGCGACAAGAACCGGGGCAAGAGC
GACAACGUGCCCAGCGAGGAGGUGGUGAAGAAGAUGAAGAACUACUGGCGGCAGCUGCUGAACGCCAAGCUGAUC
ACCCAGCGGAAGUUCGACAACCUGACCAAGGCCGAGCGGGGCGGCCUGAGCGAGCUGGACAAGGCCGGCUUCAUC
AAGCGGCAGCUGGUGGAGACCCGGCAGAUCACCAAGCACGUGGCCCAGAUCCUGGACAGCCGGAUGAACACCAAG
UACGACGAGAACGACAAGCUGAUCCGGGAGGUGAAGGUGAUCACCCUGAAGAGCAAGCUGGUGAGCGACUUCCGG
AAGGACUUCCAGUUCUACAAGGUGCGGGAGAUCAACAACUACCACCACGCCCACGACGCCUACCUGAACGCCGUG

GUGGGCACCGCCCUGAUCAAGAAGUACCCCAAGCUGGAGAGCGAGUUCGUGUACGGCGACUACAAGGUGUACGAC
GUGCGGAAGAUGAUCGCCAAGAGCGAGCAGGAGAUCGGCAAGGCCACCGCCAAGUACUUCUUCUACAGCAACAUC
AUGAACUUCUUCAAGACCGAGAUCACCCUGGCCAACGGCGAGAUCCGGAAGCGGCCCCUGAUCGAGACCAACGGC
GAGACCGGCGAGAUCGUGUGGGACAAGGGCCGGGACUUCGCCACCGUGCGGAAGGUGCUGAGCAUGCCCCAGGUG
AACAUCGUGAAGAAGACCGAGGUGCAGACCGGCGGCUUCAGCAAGGAGAGCAUCCGGCCCAAGCGGAACAGCGAC
AAGCUGAUCGCCCGGAAGAAGGACUGGGACCCCAAGAAGUACGGCGGCUUCCUGUGGCCCACCGUGGCCUACAGC
GUGCUGGUGGUGGCCAAGGUGGAGAAGGGCAAGAGCAAGAAGCUGAAGAGCGUGAAGGAGCUGCUGGGCAUCACC
AUCAUGGAGCGGAGCAGCUUCGAGAAGAACCCCAUCGACUUCCUGGAGGCCAAGGGCUACAAGGAGGUGAAGAAG
GACCUGAUCAUCAAGCUGCCCAAGUACAGCCUGUUCGAGCUGGAGAACGGCCGGAAGCGGAUGCUGGCCAGCGCC
AAGCAGCUGCAGAAGGGCAACGAGCUGGCCCUGCCCAGCAAGUACGUGAACUUCCUGUACCUGGCCAGCCACUAC
GAGAAGCUGAAGGGCAGCCCCGAGGACAACGAGCAGAAGCAGCUGUUCGUGGAGCAGCACAAGCACUACCUGGAC
GAGAUCAUCGAGCAGAUCAGCGAGUUCAGCAAGCGGGUGAUCCUGGCCGACGCCAACCUGGACAAGGUGCUGAGC
GCCUACAACAAGCACCGGGACAAGCCCAUCCGGGAGCAGGCCGAGAACAUCAUCCACCUGUUCACCCUGACCCGG
CUGGGCGCCCCCCGGGCCUUCAAGUACUUCGACACCACCAUCGACCCCAAGCAGUACCGGAGCACCAAGGAGGUG
CUGGACGCCACCCUGAUCCACCAGAGCAUCACCGGCCUGUACGAGACCCGGAUCGACCUGAGCCAGCUGGGCGGC
GACAGCGGCGGCAGCGGCGGCAGCGGCGGCAGCACCAACCUGAGCGACAUCAUCGAGAAGGAGACCGGCAAGCAG
CUGGUGAUCCAGGAGAGCAUCCUGAUGCUGCCCGAGGAGGUGGAGGAGGUGAUCGGCAACAAGCCCGAGAGCGAC
AUCCUGGUGCACACCGCCUACGACGAGAGCACCGACGAGAACGUGAUGCUGCUGACCAGCGACGCCCCCGAGUAC
AAGCCCUGGGCCCUGGUGAUCCAGGACAGCAACGGCGAGAACAAGAUCAAGAUGCUGAGCGGCGGCAGCGGCGGC
AGCGGCGGCAGCACCAACCUGAGCGACAUCAUCGAGAAGGAGACCGGCAAGCAGCUGGUGAUCCAGGAGAGCAUC
CUGAUGCUGCCCGAGGAGGUGGAGGAGGUGAUCGGCAACAAGCCCGAGAGCGACAUCCUGGUGCACACCGCCUAC
GACGAGAGCACCGACGAGAACGUGAUGCUGCUGACCAGCGACGCCCCCGAGUACAAGCCCUGGGCCCUGGUGAUC
CAGGACAGCAACGGCGAGAACAAGAUCAAGAUGCUGAGCGGCGGCAGCAAGCGGACCGCCGACGGCAGCGAGUUC
GAGCCCAAGAAGAAGCGGAAGGUGGGCGGCGGCGGCAGCGGCGCCACCAACUUCAGCCUGCUGAAGCAGGCCGGC
GACGUGGAGGAGAACCCCGGCCCCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGCGUGGUGCCCAUCCUGGUG
GAGCUGGACGGCGACGUGAACGGCCACAAGUUCAGCGUGAGCGGCGAGGGCGAGGGCGACGCCACCUACGGCAAG
CUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUGGUGACCACCCUGACCUAC
GGCGUGCAGUGCUUCAGCCGGUACCCCGACCACAUGAAGCAGCACGACUUCUUCAAGAGCGCCAUGCCCGAGGGC
UACGUGCAGGAGCGGACCAUCUUCUUCAAGGACGACGGCAACUACAAGACCCGGGCCGAGGUGAAGUUCGAGGGC
GACACCCUGGUGAACCGGAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGGGCCACAAGCUG
GAGUACAACUACAACAGCCACAACGUGUACAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUCAAG
AUCCGGCACAACAUCGAGGACGGCAGCGUGCAGCUGGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGC
CCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGAGCGCCCUGAGCAAGGACCCCAACGAGAAGCGGGAC
CACAUGGUGCUGCUGGAGUUCGUGACCGCCGCCGGCAUCACCCUGGGCAUGGACGAGCUGUACAAGUGA*GCUAGA*
AGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCCUAAGUCCAACUACUAAACUGGGGGAUA
UUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAAACAUUUAUUUUCAUUGCUGCGCUAGAAGCUCGCU
UUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCCUAAGUCCAACUACUAAACUGGGGGAUAUUAUGAAG
GGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAAACAUUUAUUUUCAUUGCUGCGGGACAUUCUUAAUUAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAGGGCUUAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUG
CUUUUUUUCUAGA

SEQ ID NO:31 > mRNA sequence encoding for ABE-SpRY. The different regions of the mRNA are indicated as follows :*T7 promoter* - **ABE-SpRY coding sequence -** *HBB 3'UTR (1st copy) - HBB 3'UTR (2nd copy)* - Poly-a tail - Residual nucleotides (till the restriction enzyme site)

*GGGAGAGCCGCCACC***AUGAAACGGACAGCCGACGGAAGCGAGUUCGAGUCACCAAAGAAGAAGCGGAAAGUCUCU**
GAAGUCGAGUUUAGCCACGAGUAUUGGAUGAGGCACGCACUGACCCUGGCAAAGCGAGCAUGGGAUGAAAGAGAA
GUCCCCGUGGGCGCCGUGCUGGUGCACAACAAUAGAGUGAUCGGAGAGGGAUGGAACAGGCCAAUCGGCCGCCAC
GACCCUACCGCACACGCAGAGAUCAUGGCACUGAGGCAGGGAGGCCUGGUCAUGCAGAAUUACCGCCUGAUCGAU
GCCACCCUGUAUGUGACACUGGAGCCAUGCGUGAUGUGCGCAGGAGCAAUGAUCCACAGCAGGAUCGGAAGAGUG
GUGUUCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCUCCCUGAUGGAUGUGCUGCACCACCCCGGCAUGAAC
CACCGGGUGGAGAUCACAGAGGGAAUCCUGGCAGACGAGUGCGCCGCCCUGCUGAGCGAUUUCUUUAGAAUGCGG
AGACAGGAGAUCAAGGCCCAGAAGAAGGCACAGAGCUCCACCGACUCUGGAGGAUCUAGCGGAGGAUCCUCUGGA
AGCGAGACACCAGGCACAAGCGAGUCCGCCACACCAGAGAGCUCCGGCGGCUCCUCCGGAGGAUCCUCUGAGGUG
GAGUUUUCCCACGAGUACUGGAUGAGACAUGCCCUGACCCUGGCCAAGAGGGCACGCGAUGAGAGGGAGGUGCCU
GUGGGAGCCGUGCUGGUGCUGAACAAUAGAGUGAUCGGCGAGGGCUGGAACAGAGCCAUCGGCCUGCACGACCCA
ACAGCCCAUGCCGAAAUUAUGGCCCUGAGACAGGGCGGCCUGGUCAUGCAGAACUACAGACUGAUUGACGCCACC
CUGUACGUGACAUUCGAGCCUUGCGUGAUGUGCGCCGGCGCCAUGAUCCACUCUAGGAUCGGCCGCGUGGUGUUU

GGCGUGAGGAACGCAAAAACCGGCGCCGCAGGCUCCCUGAUGGACGUGCUGCACUACCCCGGCAUGAAUCACCGC
GUCGAAAUUACCGAGGGAAUCCUGGCAGAUGAAUGUGCCGCCCUGCUGUGCUAUUUCUUUCGGAUGCCUAGACAG
GUGUUCAAUGCUCAGAAGAAGGCCCAGAGCUCCACCGACUCCGGAGGAUCUAGCGGAGGCUCCUCUGGCUCUGAG
ACACCUGGCACAAGCGAGAGCGCAACACCUGAAAGCAGCGGGGGCAGCAGCGGGGGGUCAGACAAGAAGUACAGC
AUCGGCCUGGCCAUCGGCACCAACUCUGUGGGCUGGGCCGUGAUCACCGACGAGUACAAGGUGCCCAGCAAGAAA
UUCAAGGUGCUGGGCAACACCGACCGGCACAGCAUCAAGAAGAACCUGAUCGGAGCCCUGCUGUUCGACAGCGGC
GAAACAGCCGAGAGAACCCGGCUGAAGAGAACCGCCAGAAGAAGAUACACCAGACGGAAGAACCGGAUCUGCUAU
CUGCAAGAGAUCUUCAGCAACGAGAUGGCCAAGGUGGACGACAGCUUCUUCCACAGACUGGAAGAGUCCUUCCUG
GUGGAAGAGGAUAAGAAGCACGAGCGGCACCCCAUCUUCGGCAACAUCGUGGACGAGGUGGCCUACCACGAGAAG
UACCCCACCAUCUACCACCUGAGAAAGAAACUGGUGGACAGCACCGACAAGGCCGACCUGCGGCUGAUCUAUCUG
GCCCUGGCCCACAUGAUCAAGUUCCGGGGCCACUUCCUGAUCGAGGGCGACCUGAACCCCGACAACAGCGACGUG
GACAAGCUGUUCAUCCAGCUGGUGCAGACCUACAACCAGCUGUUCGAGGAAAACCCCAUCAACGCCAGCGGCGUG
GACGCCAAGGCCAUCCUGUCUGCCAGACUGAGCAAGAGCAGACGGCUGGAAAAUCUGAUCGCCCAGCUGCCCGGC
GAGAAGAAGAAUGGCCUGUUCGGAAACCUGAUUGCCCUGAGCCUGGGCCUGACCCCCAACUUCAAGAGCAACUUC
GACCUGGCCGAGGAUGCCAAACUGCAGCUGAGCAAGGACACCUACGACGACGACCUGGACAACCUGCUGGCCCAG
AUCGGCGACCAGUACGCCGACCUGUUUCUGGCCGCCAAGAACCUGUCCGACGCCAUCCUGCUGAGCGACAUCCUG
AGAGUGAACACCGAGAUCACCAAGGCCCCCCUGAGCGCCUCUAUGAUCAAGAGAUACGACGAGCACCACCAGGAC
CUGACCCUGCUGAAAGCUCUCGUGCGGCAGCAGCUGCCUGAGAAGUACAAAGAGAUUUUCUUCGACCAGAGCAAG
AACGGCUACGCCGGCUACAUUGACGGCGGAGCCAGCCAGGAAGAGUUCUACAAGUUCAUCAAGCCCAUCCUGGAA
AAGAUGGACGGCACCGAGGAACUGCUCGUGAAGCUGAACAGAGAGGACCUGCUGCGGAAGCAGCGGACCUUCGAC
AACGGCAGCAUCCCCCACCAGAUCCACCUGGGAGAGCUGCACGCCAUUCUGCGGCGGCAGGAAGAUUUUUACCCA
UUCCUGAAGGACAACCGGGAAAAGAUCGAGAAGAUCCUGACCUUCCGCAUCCCCUACUACGUGGGCCCUCUGGCC
AGGGGAAACAGCAGAUUCGCCUGGAUGACCAGAAAGAGCGAGGAAACCAUCACCCCCUGGAACUUCGAGGAAGUG
GUGGACAAGGGCGCUUCCGCCCAGAGCUUCAUCGAGCGGAUGACCAACUUCGAUAAGAACCUGCCCAACGAGAAG
GUGCUGCCCAAGCACAGCCUGCUGUACGAGUACUUCACCGUGUAUAACGAGCUGACCAAAGUGAAAUACGUGACC
GAGGGAAUGAGAAAGCCCGCCUUCCUGAGCGGCGAGCAGAAAAAGGCCAUCGUGGACCUGCUGUUCAAGACCAAC
CGGAAAGUGACCGUGAAGCAGCUGAAAGAGGACUACUUCAAGAAAAUCGAGUGCUUCGACUCCGUGGAAAUCUCC
GGCGUGGAAGAUCGGUUCAACGCCUCCCUGGGCACAUACCACGAUCUGCUGAAAAUUAUCAAGGACAAGGACUUC
CUGGACAAUGAGGAAAACGAGGACAUUCUGGAAGAUAUCGUGCUGACCCUGACACUGUUUGAGGACAGAGAGAUG
AUCGAGGAACGGCUGAAAACCUAUGCCCACCUGUUCGACGACAAAGUGAUGAAGCAGCUGAAGCGGCGGAGAUAC
ACCGGCUGGGGCAGGCUGAGCCGGAAGCUGAUCAACGGCAUCCGGGACAAGCAGUCCGGCAAGACAAUCCUGGAU
UUCCUGAAGUCCGACGGCUUCGCCAACAGAAACUUCAUGCAGCUGAUCCACGACGACAGCCUGACCUUUAAAGAG
GACAUCCAGAAAGCCCAGGUGUCCGGCCAGGGCGAUAGCCUGCACGAGCACAUUGCCAAUCUGGCCGGCAGCCCC
GCCAUUAAGAAGGGCAUCCUGCAGACAGUGAAGGUGGUGGACGAGCUCGUGAAAGUGAUGGGCCGGCACAAGCCC
GAGAACAUCGUGAUCGAAAUGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAAUG
AAGCGGAUCGAAGAGGGCAUCAAAGAGCUGGGCAGCCAGAUCCUGAAAGAACACCCCGUGGAAAACACCCAGCUG
CAGAACGAGAAGCUGUACCUGUACUACCUGCAGAAUGGGCGGGAUAUGUACGUGGACCAGGAACUGGACAUCAAC
CGGCUGUCCGACUACGAUGUGGACCAUAUCGUGCCUCAGAGCUUUCUGAAGGACGACUCCAUCGACAACAAGGUG
CUGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGUGCCCUCCGAAGAGGUCGUGAAGAAGAUGAAGAAC
UACUGGCGGCAGCUGCUGAACGCCAAGCUGAUUACCCAGAGAAAGUUCGACAAUCUGACCAAGGCCGAGAGAGGC
GGCCUGAGCGAACUGGAUAAGGCCGGCUUCAUCAAGAGACAGCUGGUGGAAACCCGGCAGAUCACAAAGCACGUG
GCACAGAUCCUGGACUCCCGGAUGAACACUAAGUACGACGAGAAUGACAAGCUGAUCCGGGAAGUGAAAGUGAUC
ACCCUGAAGUCCAAGCUGGUGUCCGAUUUCCGGAAGGAUUUCCAGUUUUACAAAGUGCGCGAGAUCAACAACUAC
CACCACGCCCACGACGCCUACCUGAACGCCGUCGUGGGAACCGCCCUGAUCAAAAAGUACCCUAAGCUGGAAAGC
GAGUUCGUGUACGGCGACUACAAGGUGUACGACGUGCGGAAGAUGAUCGCCAAGAGCGAGCAGGAAAUCGGCAAG
GCUACCGCCAAGUACUUCUUCUACAGCAACAUCAUGAACUUUUUCAAGACCGAGAUUACCCUGGCCAACGGCGAG
AUCCGGAAGCGGCCUCUGAUCGAGACAAACGGCGAAACCGGGGAGAUCGUGUGGGAUAAGGGCCGGGAUUUUGCC
ACCGUGCGGAAAGUGCUGAGCAUGCCCCAAGUGAAUAUCGUGAAAAAGACCGAGGUGCAGACAGGCGGCUUCAGC
AAAGAGUCUAUCAGACCCAAGAGGAACAGCGAUAAGCUGAUCGCCAGAAAGAAGGACUGGGACCCUAAGAAGUAC
GGCGGCUUCCUGUGGCCCACCGUGGCCUAUUCUGUGCUGGUGGUGGCCAAAGUGGAAAAGGGCAAGUCCAAGAAA
CUGAAGAGUGUGAAAGAGCUGCUGGGGAUCACCAUCAUGGAAAGAAGCAGCUUCGAGAAGAAUCCCAUCGACUUU
CUGGAAGCCAAGGGCUACAAAGAAGUGAAAAAGGACCUGAUCAUCAAGCUGCCUAAGUACUCCCUGUUCGAGCUG
GAAAACGGCCGGAAGAGAAUGCUGGCCUCUGCCAAGCAGCUGCAGAAGGGAAACGAACUGGCCCUGCCCUCCAAA
UAUGUGAACUUCCUGUACCUGGCCAGCCACUAUGAGAAGCUGAAGGGCUCCCCCGAGGAUAAUGAGCAGAAACAG
CUGUUUGUGGAACAGCACAAGCACUACCUGGACGAGAUCAUCGAGCAGAUCAGCGAGUUCUCCAAGAGAGUGAUC
CUGGCCGACGCUAAUCUGGACAAAGUGCUGUCCGCCUACAACAAGCACCGGGAUAAGCCCAUCAGAGAGCAGGCC
GAGAAUAUCAUCCACCUGUUUACCCUGACCAGACUGGGAGCCCCUAGAGCCUUCAAGUACUUUGACACCACCAUC
GACCCCAAGCAGUACAGAAGCACCAAAGAGGUGCUGGACGCCACCCUGAUCCACCAGAGCAUCACCGGCCUGUAC
GAGACACGGAUCGACCUGUCUCAGCUGGGAGGUGACUCUGGCGGCUCAAAAAGAACCGCCGACGGCAGCGAAUUC
GAGCCCAAGAAGAAGAGGAAAGUCGGAAGCGGAGCUACUAACUUCAGCCUGCUGAAGCAGGCUGGAGACGUGGAG
GAGAACCCUGGACCUAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGAC

GGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUG AAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCACCCUGACCUAUGGAGUGCAG UGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAG GAGCGCACCAUCUUCUUCAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUG GUGAACCGCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGGGGCACAAGCUGGAGUACAAC UACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUCAAGAUCCGCCAC AACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUG CUGCCCGACAACCACUACCUGAGCACCCAGUCCGCCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUC CUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUCUGGUGGUUCUCCCAAG AAGAAGAGGAAAGUCUAA*GCUAGAAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCCUAA GUCCAACUACUAAACUGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAAACAUUUAUUU UCAUUGCUGCGCUAGAAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCCUAAGUCCAACU ACUAAACUGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAAACAUUUAUUUUCAUUGCU GCGGGACAUUCUUAAUU*AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGCUUAAGUUAAAAUAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUUUUUCUAGA

In vitro transcribed CBE-SpRY-OPT or ABE-SpRY-OPT mRNAs were transfected in SCD HSPCs (3 non-mobilized donors) in combination with chemically modified sgRNAs (**Figure 2A**). Transfected SCD HSPCs were differentiated towards the erythroid lineage (**Figure 2A**). CBE-SpRY coupled with ATF4_bs_1 or ATF4_bs_2 sgRNAs led to disruption of the ATF4 BS with efficiencies of 47.0% ± 7.0 [mean ± standard error of the mean (SEM)] and 46.7% ± 3.8 and editing profiles, here after named +55 CBE I or +55 CBE II, respectively TTGTATTATTT (SEQ ID NO:32) and TTGCATTATTT (SEQ ID NO:33), respectively; **Figure 2B**). ABE-SpRY coupled with ATF4_bs_1 or ATF4_bs_2 sgRNAs led to disruption of the ATF4 BS with efficiencies of 63.7% ± 2.2 and 24.3% ± 2.7 and editing profiles, here after named +55 ABE I or +55 ABE II, respectively (TTGCGTCATCC (SEQ ID NO:34) and TTGCATCGTCC (SEQ ID NO:35), respectively; **Figure 2B**). In parallel, to compare base editing and Cas9 nuclease strategies, we transfected RNP complexes containing Cas9 nuclease and ATF4_bs_3 sgRNA (Huang et al., 2020), disrupting the ATF4 BS in the +55-kb region of the erythroid-specific *BCL11A* enhancer with efficiencies of 86.8% ± 1.6 (here after named +55 Cas9) (**Figure 2C)**. TIDE analysis confirmed the DSB-free nature of base editors, as we detected no InDels in base edited samples (**Figure 2C**). In conclusion, we were able to efficiently target the ATF4 BS at the +55-kb *BCL11A* enhancer region, by targeting alternative bases and without causing DSBs in SCD HSPCs.

RNA-mediated base editing of *BCL11A* enhancer in SCD HSPCs does not affect progenitor viability and differentiation, and reactivates HbF in erythroid colonies

Transfected SCD HSPCs were subjected to the CFC assay (**Figure 2A**). Base editing efficiencies in BFU-E and CFU-GM pooled colonies were similar to those observed in liquid erythroid cultures, for sgRNAs targeting the +55-kb region, while editing efficiency tended to

be lower and more variable in CFU-GM compared to BFU-E pools (**Figure 3A**). InDels were absent in base edited samples, as measured by TIDE analysis (**Figure 3B and C**). In Cas9 nuclease-treated samples, Indel efficiencies in BFU-E were similar to those observed in in erythroid bulks and higher than in CFU-GM (**Figure 3B and C**). Neither the RNA transfection of SCD HSPCs, nor the editing by itself, affected erythroid or non-erythroid progenitors, as indicated by the CFC assay. Indeed, no significant differences, in terms of numbers of erythroid BFU-E and granulomonocytic CFU-GM colonies, were observed between control and base edited samples (**Figure 4D**). These results demonstrate that our therapeutic strategy does not affect progenitors' viability and differentiation towards the erythroid and granulomonocytic lineages.

A first insight into the therapeutic potential of our strategy was the HbF reactivation in erythroid BFU-E pools, upon RNA-mediated base editing of the erythroid-specific *BCL11A* enhancer. In particular, CE-HPLC analysis showed that base edited samples expressed HbF levels intermediate between control (22.1% ± 7.4 in untreated samples) and +55 Cas9 samples (49.7% ± 2.2) (**Figure 3E**). There was a tendency of higher HbF levels in the +55 CBE samples, as compared to the +55 ABE samples (41.9% ± 4.6, 39.1% ± 3.1, 35.3% ± 7.1 and 34.5% ± 10.8 for +55 CBE I, II, ABE I and II profiles respectively) (**Figure 3E**). These results were confirmed by RP-HPLC detecting single globin chains (**Figure 3F**). Overall, these data indicate that base editing-mediated disruption of the +55-kb erythroid-specific *BCL11A* enhancer reactivates γ-globin genes in erythroid BFU-E colonies derived from base edited SCD HSPCs, with variable levels of expression, based on the type of modification.

Disruption of the erythroid-specific *BCL11A* enhancer reactivates HbF in RBCs derived from base edited SCD HSPCs

Erythroid BFU-E colonies bear high levels of HbF background and contain only a small proportion of fully differentiated erythroid cells, therefore do not allow us to precisely identify the most potent editing profile in terms of HbF reactivation in enucleated RBCs. Therefore, we assessed the HbF reactivation achieved in mature RBCs, derived from fully differentiated control and transfected SCD HSPCs (**Figure 2A**).

First, we assessed the effect of our strategy on the erythroid differentiation process to evaluate the safety of these therapeutic strategies. To this aim, we measured the frequency of enucleated

cells (DRAQ5$^-$ cells) along the erythroid differentiation. In all the samples, we observed an increase of the enucleation rate along the differentiation with mature enucleated erythrocytes reaching up to ~60% of the total cells at the end of the differentiation. Importantly, no significant difference in the enucleation rate was found between control and treated samples (**Figure 4A**), indicating that our RNA-mediated base editing of the *BCL11A* erythroid-specific enhancer does not affect the enucleation process. To better assess the impact of our procedure on SCD HSPC erythroid differentiation, we evaluated the expression of erythroid surface markers (CD36, CD71, GPA, BAND3 and α4-Integrin) along the erythroid differentiation by flow cytometry. Along the differentiation, we observed a similar increase in the frequency of cells expressing GPA and Band 3, and a similar decrease in the frequency of cells expressing CD36, CD71 and α4-Integrin, in control and edited samples (**Figure 4B-E**). Summarizing, targeting the ATF4 BS in the erythroid-specific *BCL11A* enhancer did not impair SCD HSPC erythroid differentiation.

We then evaluated the levels of HbF reactivation in RBCs derived from control and edited HSPCs. CE-HPLC analysis revealed HbF reactivation in all the samples (**Figure 5A**). CBE-treated samples showed higher HbF levels (24.5% ± 7.2 and 23.6 ± 6.4, for CBE I and II profiles, respectively) compared to ABE-treated samples (16.5% ± 5.0 and 15.6% ± 4.1, for ABE I and II profiles, respectively) (**Figure 5A**). Of note, cells carrying the ABE I profile showed modest HbF levels despite of the high base editing efficiency, suggesting that the target base is not essential for ATF4 binding. On the contrary, cells carrying the ABE II profile reactivated HbF despite the low base editing efficiency, indicating that the converted base is critical for ATF4 binding. Disruption of the ATF4 BS with either CBEs or Cas9 led to similar levels of HbF (29.3% ± 6.7 for Cas9-treated samples), even though the editing efficiency in Cas9 nuclease treated samples was two-fold higher than in base edited samples (**Figure 5A**). These results were confirmed by RT-qPCR and RP-HPLC at mRNA and single globin chain levels (**Figure 5B and C**). Flow cytometry analysis confirmed these data, with +55 CBE I, II and ABE I samples showing a high proportion of F-cells (56.9% ± 13.6, 55.2% ± 11.9 and 54.0% ± 13.4 respectively), approaching that observed in the Cas9 nuclease-treated samples (62.3% ± 10.3) (**Figure 5D**). The frequency of HbS expressing cells, however, was similar between control and all the edited samples (**Figure 5D**). Finally, a sickling assay was performed in control and edited samples. High frequencies of corrected cells were observed for base edited samples (up to 56.1% ± 25.5) (**Figure 5E**).

Overall, these data show that base editing mediated disruption of the ATF4 activator BS in the +55-kb region leads to HbF reactivation and ameliorated the pathological sickling phenotype. Interestingly, the different bases or combinations of bases have a different role in the binding of these TFs, with some of them being more critical and thus representing potent base editor targets for down-regulating BCL11A and reactivating HbF.

HbF reactivation in single erythroid progenitors

Transfected SCD HSPCs were subjected to the CFC assay (**Figure 2A**). To accurately compare the efficacy of the various editing approaches, we measured γ-globin expression at the clonal level in BFU-Es. We observed significant γ-globin reactivation in colonies bearing biallelic editing profiles for CBE I and CBE II groups, while for colonies of the same groups bearing monoallelic editing there was a trend for higher γ-globin expression as compared to the controls (**Figure 6A**). Interestingly, the ABE II profile allowed significantly higher γ-globin expression not only in colonies bearing biallelic edits, but also in colonies with monoallelic edits, highlighting the ability of a single A>G mutation (in position 8 of the ATF4 binding site) to induce high HbF levels (**Figure 6A**). On the contrary, the ABE I profile, bearing a different A>G mutation (in position 6 of the ATF4 binding site), did not show any significant γ-globin reactivation even in colonies having biallelic edits (**Figure 6A**). Similarly, we observed a positive correlation between base-editing efficiency and γ-globin expression in CBE I, CBE II and ABE II groups (**Figure 6B**). Generation of the ABE II profile was the most potent event in terms of γ-globin reactivation (**Figure 6B**). CBE I and CBE II profiles showed similar γ-globin levels, while ABE I showed no significant γ-globin reactivation (**Figure 6B**). Overall, these data demonstrate the capacity of the different editing profiles to reactivate HbF through the disruption of the ATF4 transcriptional activator BS in the +55-kb region of the erythroid-specific *BCL11A* enhancer.

Disruption of the erythroid-specific *BCL11A* enhancer in SCD HSPCs with more precise and highly processive base editors.

In an effort to generate the best performing profiles (CBE I and ABE II) in a more efficient and precise way, we exploited BE enzymes recognizing NGG PAMs. This would minimize potential off-target effects thanks to the PAM recognition, as compared to the near PAMless SpRY-based enzymes used in the previous experiments. In this frame, we used the AncBE4max enzyme that recognizes NGG PAMs. We firstly optimised the AncBE4max plasmid for *in vitro*

transcription and mRNA production, similarly by adding two copies of the 3' UTR of *HBB* and a poly-A sequence after the 3' UTR. The optimized plasmid, AncBE4max-OPT was used for *in vitro* transcription and mRNA production. *In vitro* transcribed AncBE4max-OPT or CBE-SpRY-OPT mRNAs were transfected in SCD HSPCs (1 non-mobilized donor) in combination with chemically modified ATF4_bs_1 sgRNAs (**Figure 2A**). Transfected SCD HSPCs were cultures in the "HSPC medium". Both enzymes resulted in similar levels of C>T conversion, demonstrating that we can reproduce the CBE I profile also with a more precise base editor (**Figure 7**).

In parallel, we used the NGG-recognizing ABE8e enzyme in order to create the ABE II profile more efficiently and in a more precise manner. We transfected SCD HSPCs (2 non-mobilized donors) with ABE8e mRNA in combination with chemically modified ATF4_bs_2 sgRNAs (**Figure 2A**). Transfected SCD HSPCs were differentiated towards the erythroid lineage, or subjected to the CFC assay (**Figure 2A**). We observed high rates of base conversion in both erythroid liquid cultures and BFU-E or CFU-GM colonies (up to ~90%; **Figure 8A**). Of note, we were not able to precisely create the ABE II profile (1 single A>G mutation in position 8 of the ATF4 binding site), but thanks to the high processivity of the enzyme we generated a new editing profile, named ABE III, which bears two A>G mutations in position 6 and 8 (**Figure 8A**). The generation of the ABE III profile was associated with very high HbF in erythroid BFU-E colonies very efficiently thanks to the high levels of BE efficiency obtained using ABE8e (**Figure 8A and B**).

Of note, alternative sgRNAs were designed to precisely reproduce at high efficiency the ABE II profile harboring 1 single A>G mutation in position 8 of the ATF4 binding site (**Table 4**).

Table 4: optimized combinations for generating the +55kb ABE II profile:

sgRNA	Target sequence (5' to 3')	Enzyme(s)	Position (hg19)	Strand
ATF4_bs_4	CATCATCCTGGTACCAGGAA (SEQ ID NO: 41)	ABE8e-NRCH, NG-ABE8e, ABE8e-SpRY	chr2: 60725622-60725641	+
ATF4_bs_5	ATCATCCTGGTACCAGGAAG (SEQ ID NO: 42)	ABE8e-SpRY	chr2: 60725623-60725642	+

ATF4_bs_6	TCATCCTGGTACCAGGAAGG (SEQ ID NO: 43)	ABE8e-SpRY	chr2: 60725624-60725643	+
ATF4_bs_7	CATCCTGGTACCAGGAAGGC (SEQ ID NO: 44)	ABE8e-SpRY	chr2: 60725625-60725644	+

In conclusion, we were able to efficiently target the ATF4 BS at the +55-kb *BCL11A* enhancer region, by using more precise and highly efficient BE enzymes in SCD HSPCs.

REFERENCES:

Throughout this application, various references describe the state of the art to which this invention pertains. The disclosures of these references are hereby incorporated by reference into the present disclosure.

Antoniou, P., Miccio, A., and Brusson, M. (2021). Base and Prime Editing Technologies for Blood Disorders. *Front. Genome Ed.* 3, 1. doi:10.3389/fgeed.2021.618406.

Brinkman, E. K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42, e168–e168. doi:10.1093/nar/gku936.

Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjana, N. E., Shalem, O., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 527, 192–197. doi:10.1038/nature15521.

Cavazzana, M., Antoniani, C., and Miccio, A. (2017). Gene Therapy for β-Hemoglobinopathies. *Mol. Ther. J. Am. Soc. Gene Ther.* 25, 1142–1154. doi:10.1016/j.ymthe.2017.03.024.

Chandrakasan, S., and Malik, P. (2014). Gene Therapy for Hemoglobinopathies. *Hematol. Oncol. Clin. North Am.* 28, 199–216. doi:10.1016/j.hoc.2013.12.003.

Chang, K.-H., Smith, S. E., Sullivan, T., Chen, K., Zhou, Q., West, J. A., et al. (2017). Long-Term Engraftment and Fetal Globin Induction upon BCL11A Gene Editing in Bone-Marrow-Derived CD34+ Hematopoietic Stem and Progenitor Cells. *Mol. Ther. Methods Clin. Dev.* 4, 137–148. doi:10.1016/j.omtm.2016.12.009.

Cromer, M. K., Vaidyanathan, S., Ryan, D. E., Curry, B., Lucas, A. B., Camarena, J., et al. (2018). Global Transcriptional Response to CRISPR/Cas9-AAV6-Based Genome Editing in CD34+ Hematopoietic Stem and Progenitor Cells. *Mol. Ther. J. Am. Soc. Gene Ther.* 26, 2431–2442. doi:10.1016/j.ymthe.2018.06.002.

Forget, B. G. (1998). Molecular basis of hereditary persistence of fetal hemoglobin. *Ann. N. Y. Acad. Sci.* 850, 38–44. doi:10.1111/j.1749-6632.1998.tb10460.x.

Gallie, D. R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.* 5, 2108–2116. doi:10.1101/gad.5.11.2108.

Giarratana, M.-C., Kobari, L., Lapillonne, H., Chalmers, D., Kiger, L., Cynober, T., et al. (2005). Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat. Biotechnol.* 23, 69–74. doi:10.1038/nbt1047.

Guda, S., Brendel, C., Renella, R., Du, P., Bauer, D. E., Canver, M. C., et al. (2015). miRNA-embedded shRNAs for Lineage-specific BCL11A Knockdown and Hemoglobin F Induction. *Mol. Ther. J. Am. Soc. Gene Ther.* 23, 1465–1474. doi:10.1038/mt.2015.113.

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 24, 927–930. doi:10.1038/s41591-018-0049-z.

Holtkamp, S., Kreiter, S., Selmi, A., Simon, P., Koslowski, M., Huber, C., et al. (2006). Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* 108, 4009–4017. doi:10.1182/blood-2006-04-015024.

Huang, P., Peslak, S. A., Lan, X., Khandros, E., Yano, J. A., Sharma, M., et al. (2020). The HRI-regulated transcription factor ATF4 activates BCL11A transcription to silence fetal hemoglobin expression. *Blood* 135, 2121–2132. doi:10.1182/blood.2020005301.

Karikó, K., Kuo, A., and Barnathan, E. (1999). Overexpression of urokinase receptor in mammalian cells following administration of the in vitro transcribed encoding mRNA. *Gene Ther.* 6, 1092–1100. doi:10.1038/sj.gt.3300930.

Kato, G. J., Piel, F. B., Reid, C. D., Gaston, M. H., Ohene-Frempong, K., Krishnamurti, L., et al. (2018). Sickle cell disease. *Nat. Rev. Dis. Primer* 4, 1–22. doi:10.1038/nrdp.2018.10.

Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., et al. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J.* 1, 239–250. doi:10.1089/crispr.2018.0014.

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771. doi:10.1038/nbt.4192.

Liu, P., Keller, J. R., Ortiz, M., Tessarollo, L., Rachel, R. A., Nakamura, T., et al. (2003). Bcl11a is essential for normal lymphoid development. *Nat. Immunol.* 4, 525–532. doi:10.1038/ni925.

Milyavsky, M., Gan, O. I., Trottier, M., Komosa, M., Tabach, O., Notta, F., et al. (2010). A Distinctive DNA Damage Response in Human Hematopoietic Stem Cells Reveals an Apoptosis-Independent Role for p53 in Self-Renewal. *Cell Stem Cell* 7, 186–197. doi:10.1016/j.stem.2010.05.016.

Newby, G. A., Yen, J. S., Woodard, K. J., Mayuranathan, T., Lazzarotto, C. R., Li, Y., et al. (2021). Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* 595, 295–302. doi:10.1038/s41586-021-03609-w.

Piel, F. B., Steinberg, M. H., and Rees, D. C. (2017). Sickle Cell Disease. *N. Engl. J. Med.* 376, 1561–1573. doi:10.1056/NEJMra1510865.

Rees, H. A., and Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 19, 770–788. doi:10.1038/s41576-018-0059-1.

Ross, J., and Sullivan, T. D. (1985). Half-lives of beta and gamma globin messenger RNAs and of protein synthetic capacity in cultured human reticulocytes. *Blood* 66, 1149–1154.

Taher, A. T., Weatherall, D. J., and Cappellini, M. D. (2018). Thalassaemia. *The Lancet* 391, 155–167. doi:10.1016/S0140-6736(17)31822-6.

Weber, L., Frati, G., Felix, T., Hardouin, G., Casini, A., Wollenschlaeger, C., et al. (2020). Editing a γ-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. *Sci. Adv.* 6, eaay9392. doi:10.1126/sciadv.aay9392.

Wu, Y., Zeng, J., Roscoe, B. P., Liu, P., Yao, Q., Lazzarotto, C. R., et al. (2019). Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med.* 25, 776–783. doi:10.1038/s41591-019-0401-y.

Yeh, W.-H., Chiang, H., Rees, H. A., Edge, A. S. B., and Liu, D. R. (2018). In vivo base editing of post-mitotic sensory cells. *Nat. Commun.* 9, 2184. doi:10.1038/s41467-018-04580-3.

Yu, Y., Wang, J., Khaled, W., Burke, S., Li, P., Chen, X., et al. (2012). Bcl11a is essential for lymphoid development and negatively regulates p53. *J. Exp. Med.* 209, 2467–2483. doi:10.1084/jem.20121846.

Zeng, J., Wu, Y., Ren, C., Bonanno, J., Shen, A. H., Shea, D., et al. (2020). Therapeutic base editing of human hematopoietic stem cells. *Nat. Med.* 26, 535–541. doi:10.1038/s41591-020-0790-y.

CLAIMS:

1. A method for increasing fetal hemoglobin content in a eukaryotic cell comprising the step of contacting the eukaryotic cell with a gene editing platform that consists of a (a) at least one base-editing enzyme and (b) least one guide RNA molecule for guiding the base-editing enzyme to at least one target sequence in the +55-kb region of the erythroid-specific BCL11A enhancer, thereby editing and disrupting the ATF4 binding site in said region so as to repress the expression of BCL11A and subsequently increase the expression of γ-globin.

2. The method of claim 1 wherein the eukaryotic cell is selected from the group consisting of hematopoietic progenitor cells, hematopoietic stem cells (HSCs), pluripotent cells (i.e. embryonic stem cells (ES) and induced pluripotent stem cells (iPS)).

3. The method of claim 1 or 2 wherein the base-editing enzyme comprises a nickase and more particularly a Cas9 nickase.

4. The method of claim 3 wherein the nickase comprises the amino acid sequence as set forth in SEQ ID NO: 3 or SEQ ID NO:4.

5. The method of claim 1 or 2 wherein the base-editing enzyme is a cytidine deaminase or an adenosine deaminase.

6. The method according to any one of claims 1 to 5 wherein the base-editing enzyme is selected from the group consisting of ABE-SpRY (SEQ ID NO:18), CBE-SpRY (SEQ ID NO:17), ABE8e (SEQ ID NO: 37), ABE8e-SpRY (SEQ ID NO: 38), ABE8e-NRCH (SEQ ID NO: 39), and NG-ABE8e (SEQ ID NO: 40).

7. The method according to any one of claims 1 to 6 wherein the guide RNA targets the ATF4 binding site so as to edit said site and thus disrupting the binding of ATF4 to its binding site.

8. The method according to any one claims 1 to 7 wherein the gene editing platform comprises a) a cytidine base-editing enzyme and b) and at least one guide RNA molecule suitable for introducing:

- one C>T mutation in the first cytidine residue of SEQ ID NO: 45 (TTGCATCATCC) and/or

- one C>T mutation in the second cytidine residue of SEQ ID NO: 45 (TTGCATCATCC) and/or

- one C>T mutation in the third cytidine residue of SEQ ID NO: 45 (TTGCATCATCC) and/or

- one C>T mutation in the fourth cytidine SEQ ID NO: 45 (TTGCATCATCC).

9. The method of claim 8 wherein the gene editing platform comprises a) a cytidine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 CBE I editing profile of the ATF4 binding site (i.e. TTGTATTATTT (SEQ ID NO:32)).

10. The method of claim 8 wherein the gene editing platform comprises a) a cytidine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 CBE II editing profile in the ATF4 binding site (i.e. TTGCATTATTT (SEQ ID NO:33)).

11. The method according to any one claims 1 to 7 wherein the gene editing platform comprises a) a adenine base-editing enzyme and b) and at least one guide RNA molecule suitable for introducing:

- one A>G mutation in the first adenine residue in SEQ ID NO: 45 (TTGCATCATCC) and/or

- one A>G mutation in the second adenine residue of SEQ ID NO: 45 (TTGCATCATCC).

12. The method of claim 11 wherein the gene editing platform comprises a) an adenine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 ABE II profile in the ATF4 binding site (TTGCATCGTCC (SEQ ID NO:35)).

13. The method of claim 11 wherein the gene editing platform comprises a) an adenine base-editing enzyme and b) and at least one guide RNA molecule suitable for generating the +55 ABE III profile in the ATF4 binding site (TTGCGTCGTCC (SEQ ID NO: 46)).

14. The method according to any one of claims 1 to 13 wherein the guide RNA targets a sequence selected from Table 1.

15. The method according to any one of claims 1 to 13 wherein the gene editing platform comprises a) a base-editing enzyme that is an ABE-SpRY or a CBE-SpRY and b) and at least one guide RNA molecule that targets one sequence selected in Table 1 or Table 4.

16. The method according to any one of claims 1 to 13 wherein the gene editing platform comprises a) a base-editing enzyme and b) and at least one guide RNA molecule selected according to the combinations described in Table 4.

17. The method according to any one of claims 1 to 16 wherein the gene editing platform comprises a plurality of guide RNA molecules that are designed for targeting a plurality of sequences in the +55-kb region of the erythroid-specific BCL11A enhancer.

18. The method according to any one of claims 1 to 17 1 wherein the different components of the gene editing platform are provided to the population of cells through the use of an RNA-encoded system.

19. A method for increasing fetal hemoglobin levels in a subject in need thereof, the method comprising transplanting a therapeutically effective amount of a population of eukaryotic cells obtained by the method according to any one of claims 1 to 18.

20. A method of treating a β-hemoglobinopathy in a subject in need thereof, the method comprising transplanting a therapeutically effective amount of a population of eukaryotic cells obtained by the method any one of claims 1 to 18.

21. The method of claim 20 wherein the hemoglobinopathy is sickle cell disease or β-thalassemia.

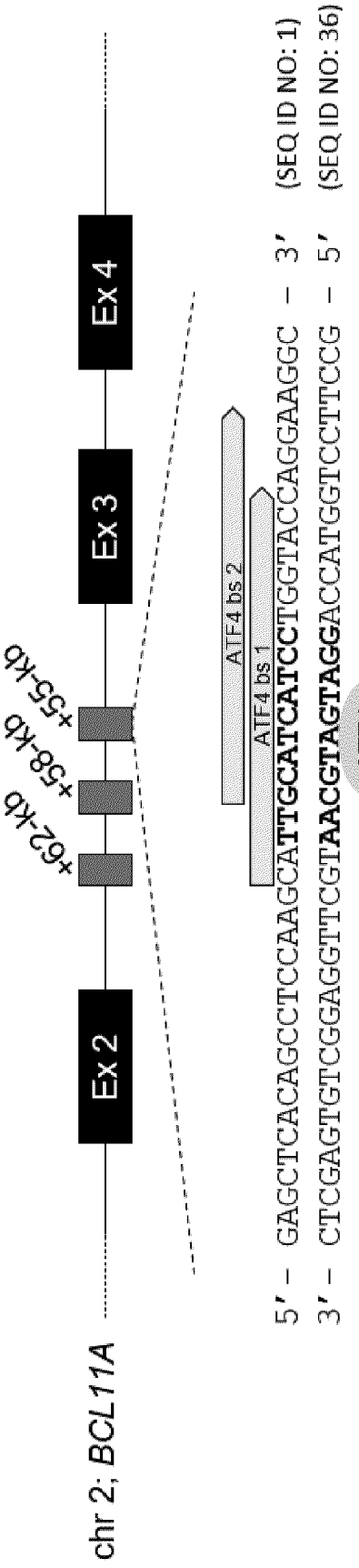

chr 2; BCL11A
Ex 2
Ex 3
Ex 4
+62-kb
+58-kb
+55-kb
ATF4 bs 2
ATF4 bs 1
ATF4
5′ – GAGCTCACAGCCTCCAAGCATTGCATCATCCTGGTACCAGGAAGGC – 3′ (SEQ ID NO: 1)
3′ – CTCGAGTGTCGGAGGTTCGTAACGTAGTAGGACCATGGTCCTTCCG – 5′ (SEQ ID NO: 36)

Figure 1

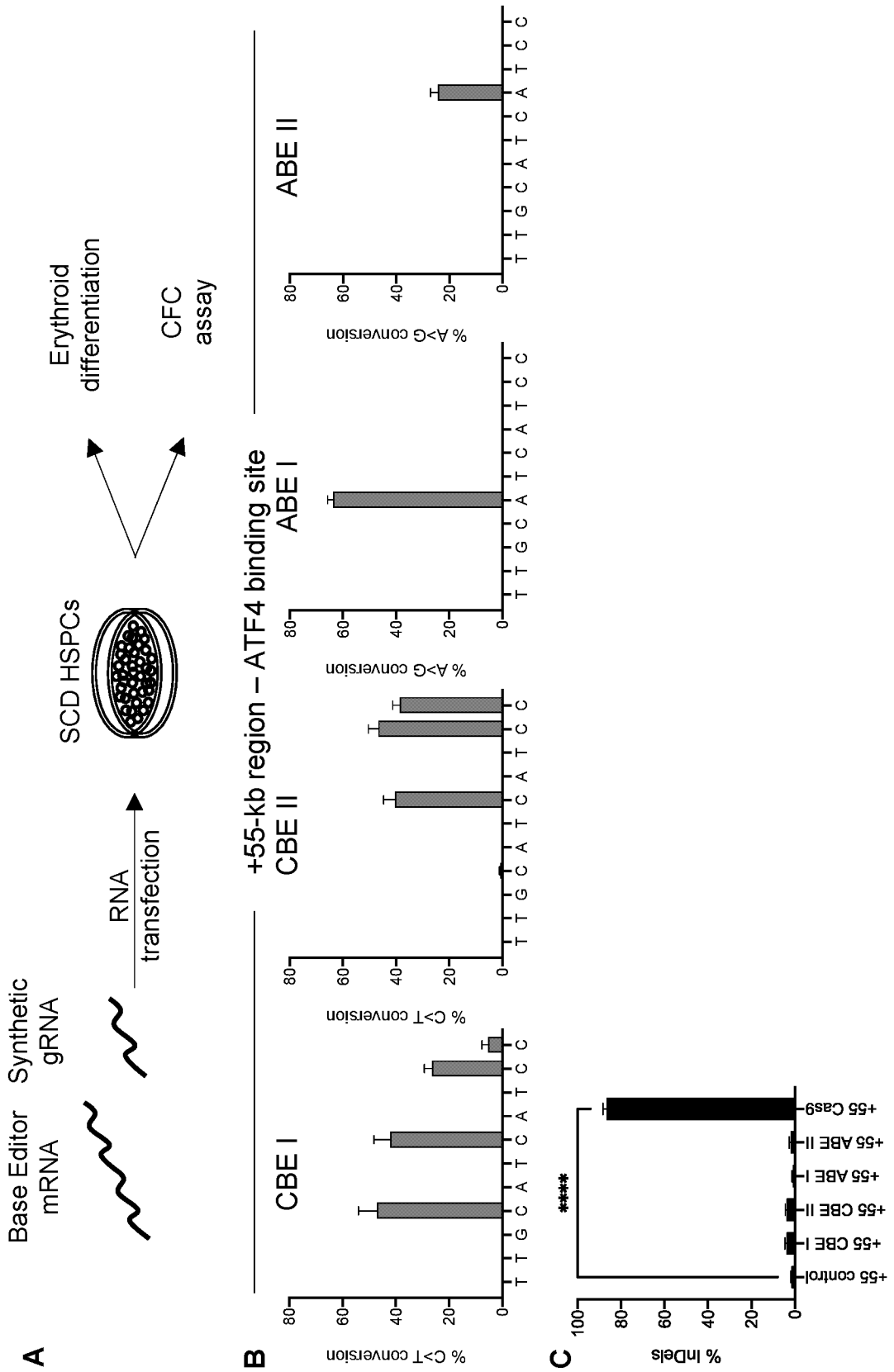

A
Base Editor mRNA
Synthetic gRNA
RNA transfection
SCD HSPCs
Erythroid differentiation
CFC assay
B
+55-kb region – ATF4 binding site
CBE I
CBE II
ABE I
ABE II
% C>T conversion
% A>G conversion
T T G C A T C A T C C C
C
% Indels
****
+55 control
+55 CBE I
+55 CBE II
+55 ABE I
+55 ABE II
+55 Cas9

Figure 2

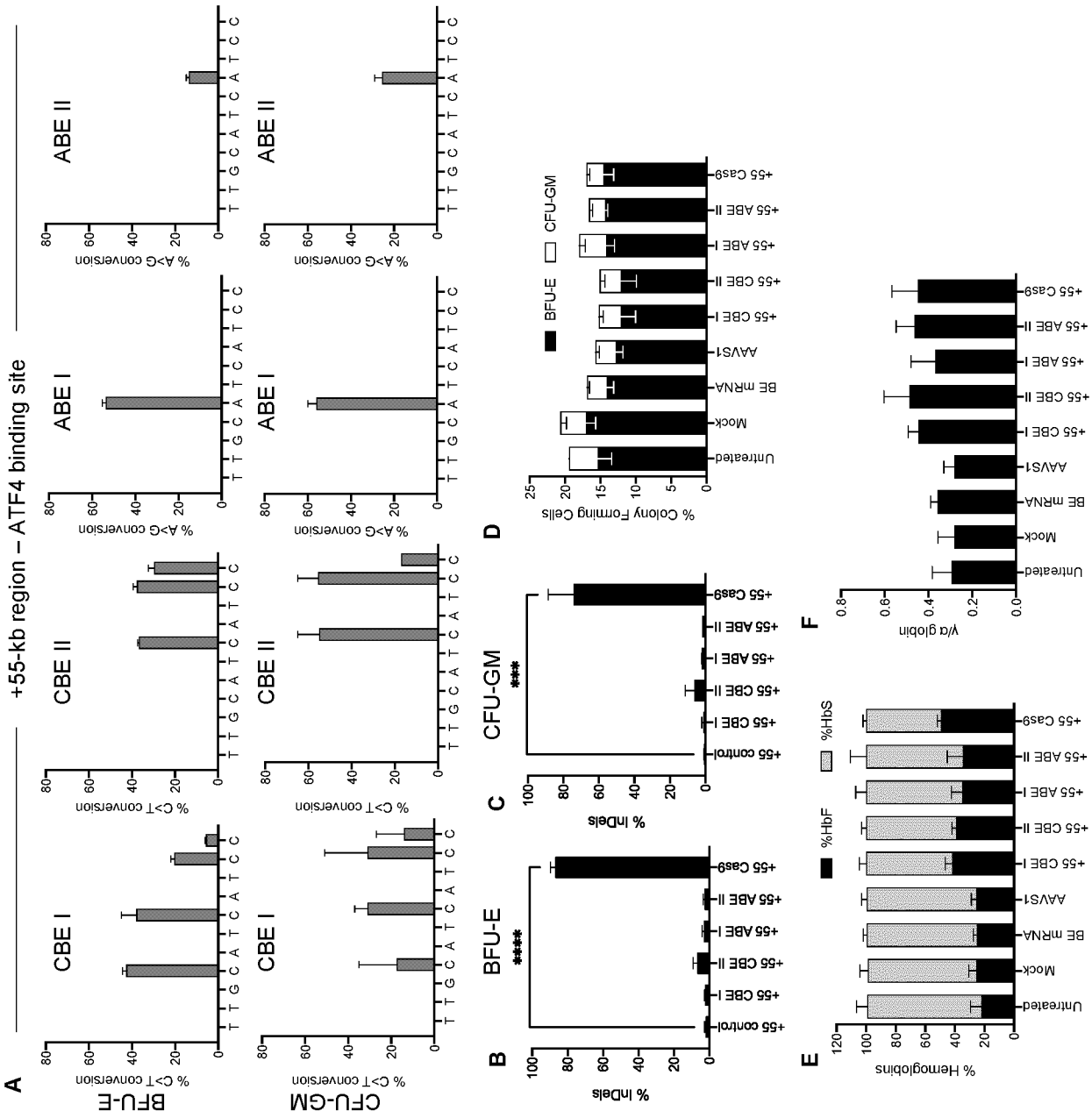

A
+55-kb region – ATF4 binding site
BFU-E
CFU-GM
CBE I
CBE II
ABE I
ABE II
% C>T conversion
% A>G conversion
T T G C A T C A T C C C
B
BFU-E
% Indels
****
+55 control
+55 CBE I
+55 CBE II
+55 ABE I
+55 ABE II
+55 Cas9
C
CFU-GM
% Indels
***
D
BFU-E
CFU-GM
% Colony Forming Cells
Untreated
Mock
BE mRNA
AAVS1
+55 CBE I
+55 CBE II
+55 ABE I
+55 ABE II
+55 Cas9
E
%HbF
%HbS
% Hemoglobins
F
γ/α globin

Figure 3

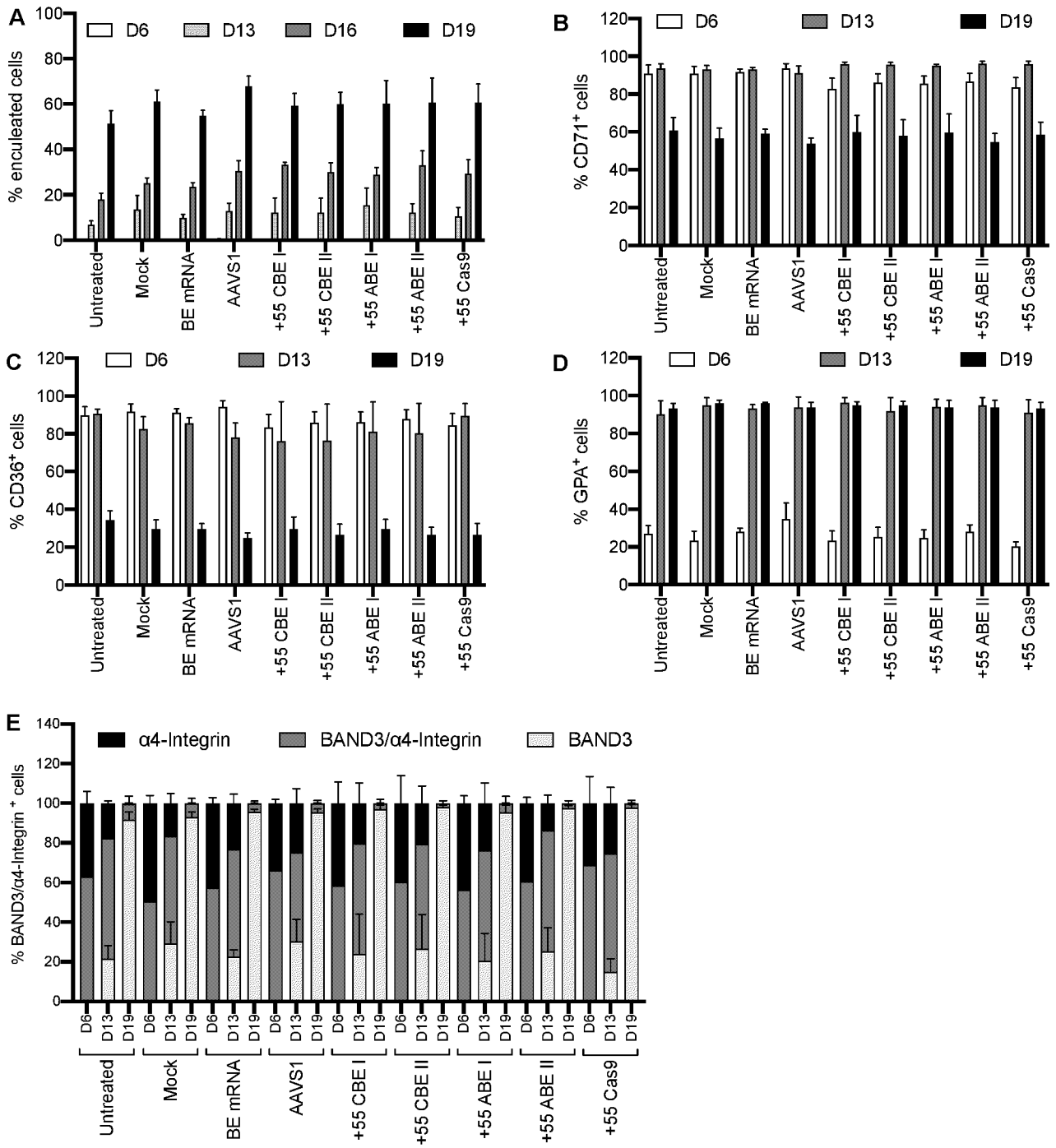

A
% enculeated cells
D6
D13
D16
D19
B
% CD71+ cells
C
% CD36+ cells
D
% GPA+ cells
Untreated
Mock
BE mRNA
AAVS1
+55 CBE I
+55 CBE II
+55 ABE I
+55 ABE II
+55 Cas9
E
% BAND3/α4-Integrin + cells
α4-Integrin
BAND3/α4-Integrin
BAND3

Figure 4

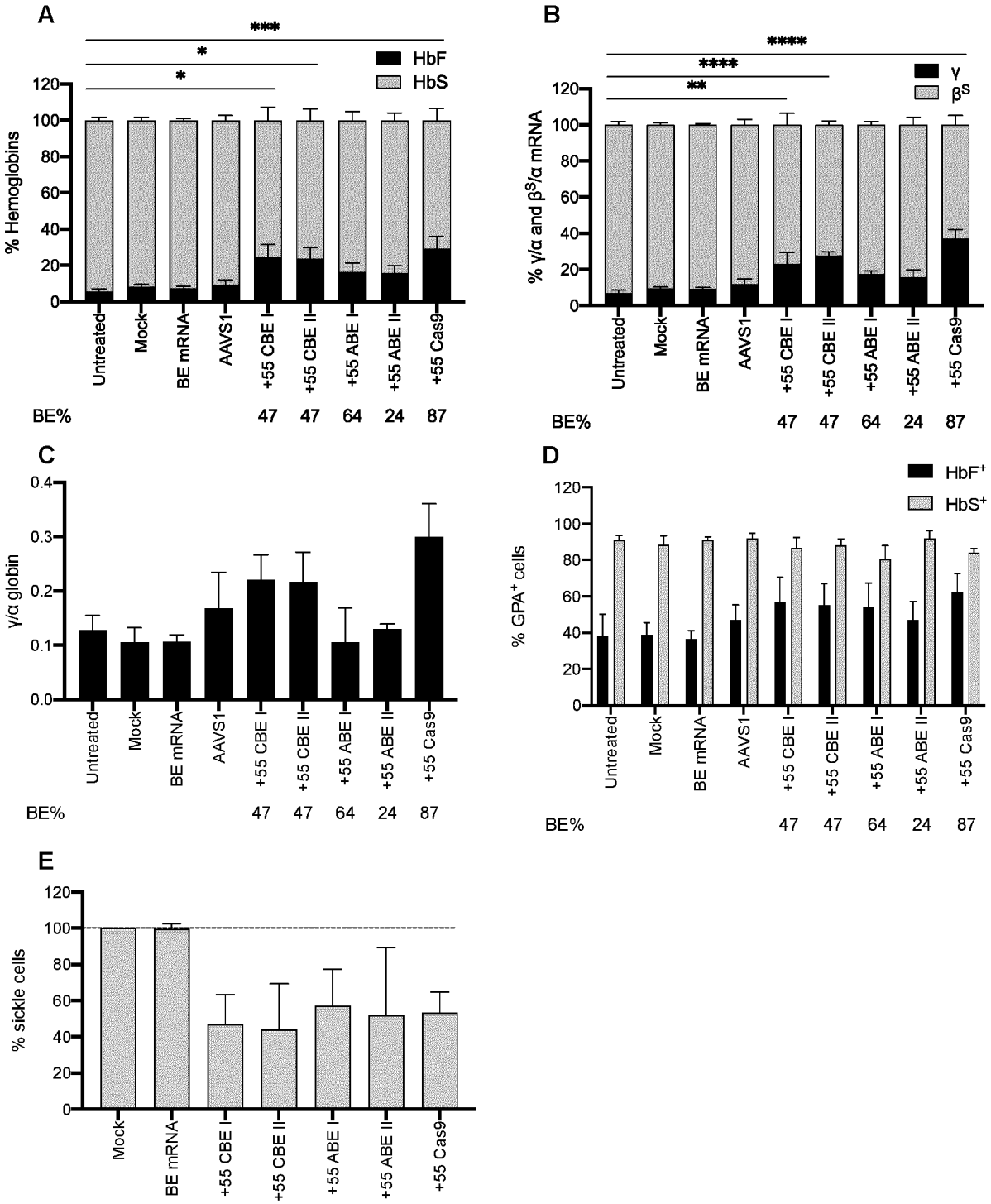
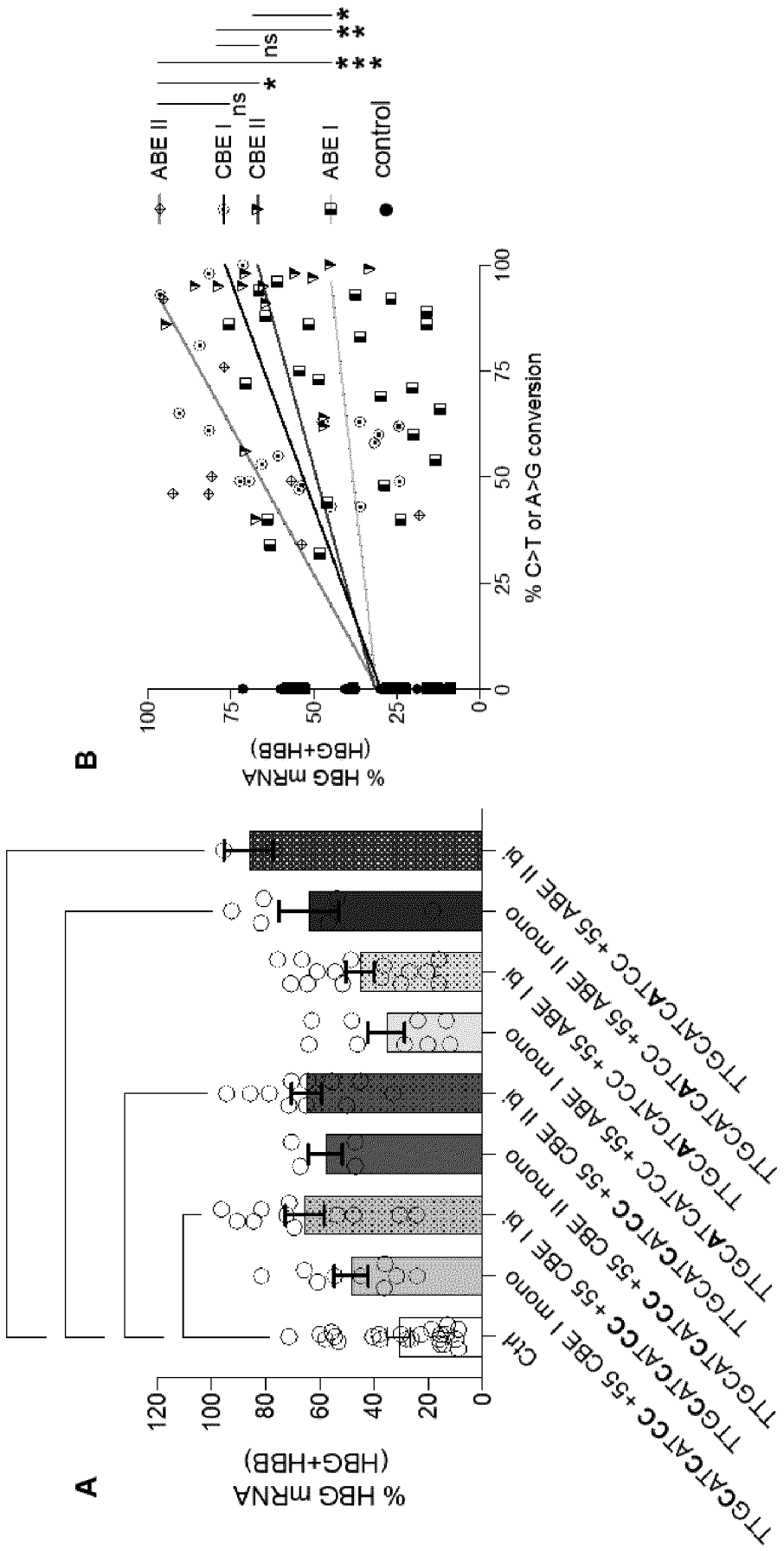

A
***
*
*
HbF
HbS
% Hemoglobins
120
100
80
60
40
20
0
Untreated
Mock
BE mRNA
AAVS1
+55 CBE I
+55 CBE II
+55 ABE I
+55 ABE II
+55 Cas9
BE%
47
47
64
24
87
B
****
****
**
γ
βS
% γ/α and βS/α mRNA
C
γ/α globin
0.4
0.3
0.2
0.1
0.0
D
HbF+
HbS+
% GPA+ cells
E
% sickle cells

Figure 5



Figures 6A and 6B

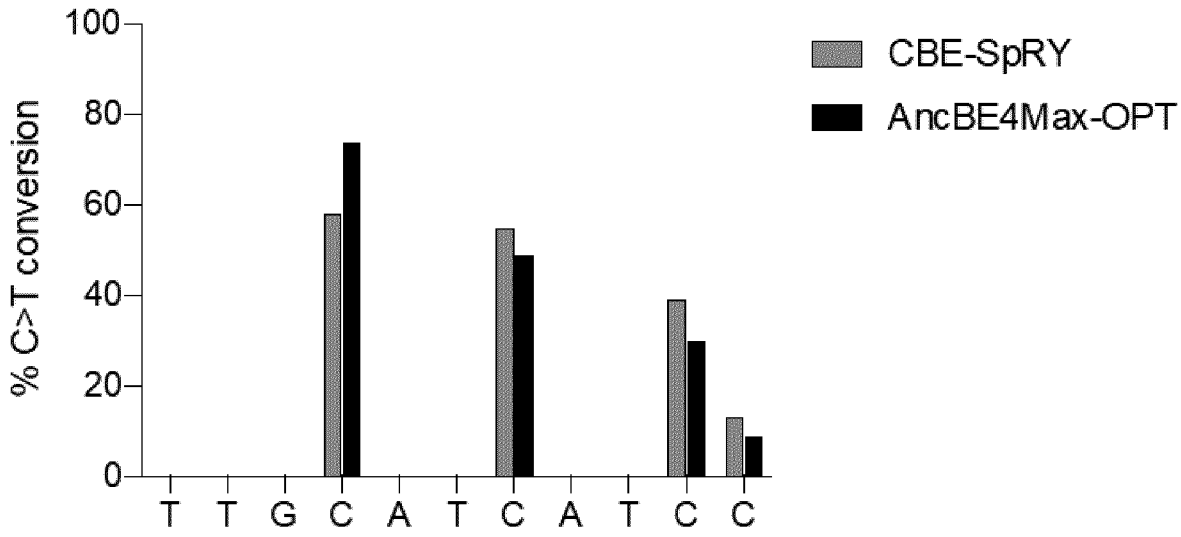
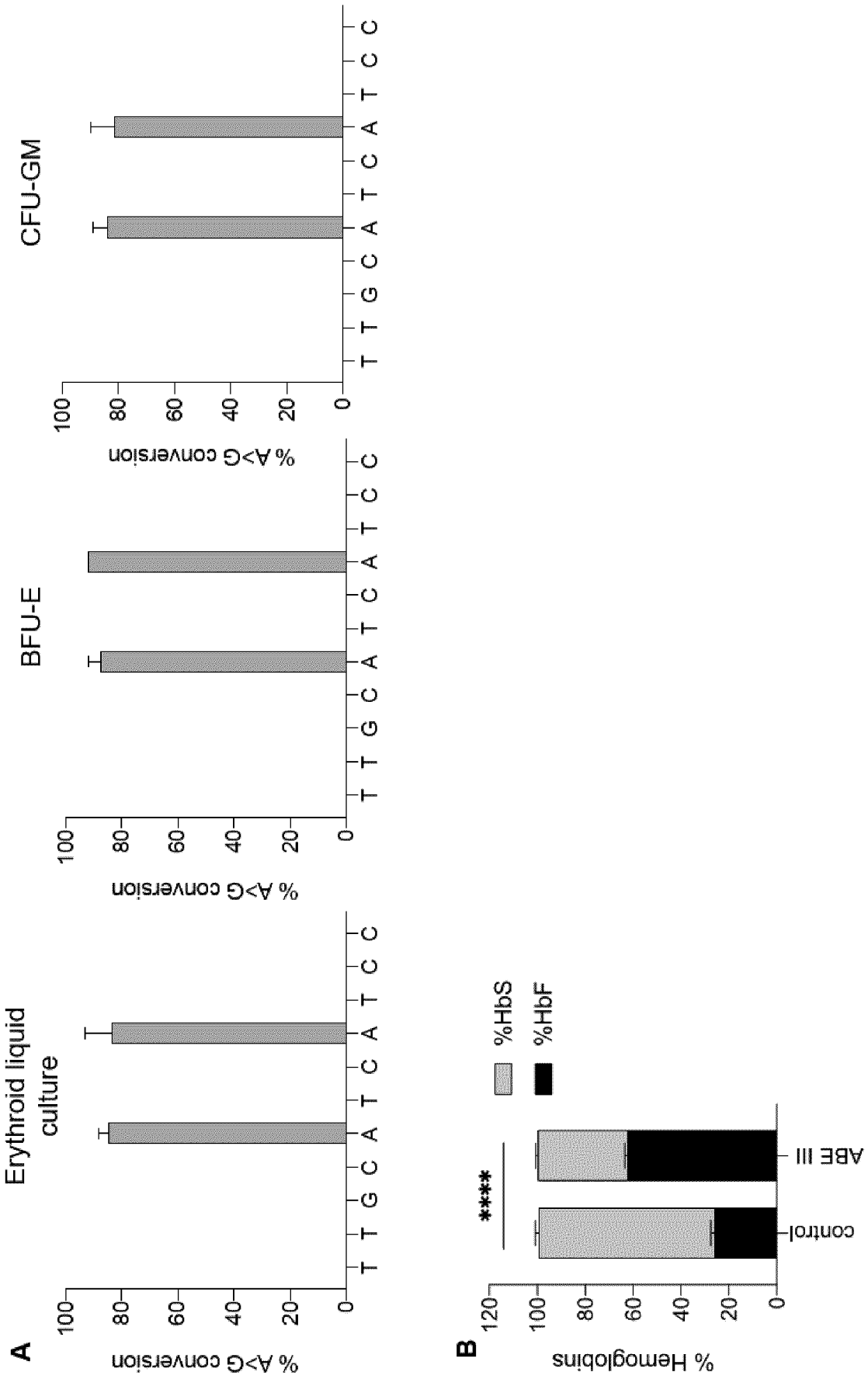

100
80
60
40
20
0
% C>T conversion
T T G C A T C A T C C
CBE-SpRY
AncBE4Max-OPT

Figure 7



Figures 8A and 8B

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2022/083904

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/805 A61K48/00 C12N5/078 C12N5/0789 C12N15/90
C12N5/074
ADD.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2021/067233 A1 (CHILDRENS MEDICAL CT CORP [US]) 8 April 2021 (2021-04-08) Disruption of BCL11A +55 enhancer in order to increase expression of gamma-globin in beta-globinopathies.; paragraphs [0007] - [0019], [0026], [0036] Base editing in region of +58 enhancer in human CD34+HSPCs; paragraph [0353] - paragraph [0390]; examples 1,2	1-21

-/--

[x] Further documents are listed in the continuation of Box C.
[x] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 March 2023	21/03/2023

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bretherick, James

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ZENG JING ET AL: "Therapeutic base editing of human hematopoietic stem cells", NATURE MEDICINE, NATURE PUBLISHING GROUP US, NEW YORK, vol. 26, no. 4, 16 March 2020 (2020-03-16) , pages 535-541, XP037090965, ISSN: 1078-8956, DOI: 10.1038/S41591-020-0790-Y [retrieved on 2020-03-16] cited in the application abstract Using modified A3A base editor for editing BCL11A erythroid enhancer at position +58 Base editing viable alternative to nuclease editing of HSC-targeted therapeutic genome modification.Demonstrate that highly efficient, specific and disease-ameliorating base editing in human HSCs is feasible with RNP delivery and might encourage therapeutic application of base editing for a range of disorders in which corrected or augmented hematopoiesis could be beneficial.; page 540, column 2, paragraph 2 - paragraph 4; figures 1,2 ----- -/--	1-21

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HUANG PENG ET AL: "The HRI-regulated transcription factor ATF4 activates BCL11A transcription to silence fetal hemoglobin expression", BLOOD , vol. 135, no. 24 11 June 2020 (2020-06-11), pages 2121-2132, XP055920666, US ISSN: 0006-4971, DOI: 10.1182/blood.2020005301 Retrieved from the Internet: URL:https://watermark.silverchair.com/bloodbld2020005301.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAA-kwggPlBgkqhkiG9w0BBwagggPWMIID0gIBADCCA8sGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMZus6sbxa1o-dTTg6AgEQgIIDnBOj_lRmk7SkzvnpWG4tmdct_sPX3u_kAxeq18bgkJQQE8uZxng58XIHuUJVWubkSlwpd525qrVG_EWWa abstract CRISPR-Cas9 guided loss-of-function genetic screen of 1446 transcription factors using HbF levels as readout, which uncovered ATF4 as a novel g-globin regulator. By combining chromatin immunoprecipita-tion sequencing (ChIP-seq), Capture-C, and gene editing techniques, we determined that ATF4 binds to the BCL11A155 enhancer to augment enhancer-promoter contacts and stimu-late BCL11A transcription. HRI knockout mice display normal levels of both Bcl11a and Bcl11a target genes, suggesting that the HRI-ATF4-BCL11A pathway is not conserved in mice. ATF4 profiling and gene editing in murine cells showed that a cor-responding ATF4 element in the murine Bcl11a enhancer is largely dispensable for Bcl11a transcription, providing an ex-planation for the species-selective effect. Our study thus un-covers a mechanism by which HRI regulates HbF levels.; page 2121, column 2, paragraph 2 ATF4 regulates BCL11A through the155 enhancer To determine the function of ATF4 binding at the BCL11A 155 element, we disrupted it via CRISPR-Cas9	1-21

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	mediated KO in pooled HUDEP2 cells (Figure 3D; supplemental Figure 4A). Both ChIP-seq and ChIP quantitative polymerase chain reaction verified specific and complete abrogation of ATF4 binding at the BCL11A 155 enhancer (Figure 3C-E; supplemental Figure 3D). As a consequence, BCL11A mRNA levels were reduced by ;50% (Figure 4A; supplemental Figure 4B). Similar results were obtained when disrupting the ATF4 element in primary eryth- roblasts (editing efficiency of 80% based on TIDE analysis; Figure 4B; supplemental Figure 4A,C).29 Subsequently, both g-globin mRNA levels and F-cell numbers were significantly elevated (Figure 4A-C; supplemental Figure 4B-E). In primary cells, the degree of BCL11A loss and g-globin induction was comparable to that achieved with disruption of the GATA motif at the 158 enhancer. Finally, forced expression of BCL11A-ER restored g-globin levels in pooled HUDEP2 cells with the ATF4 element disrupted (supplemental Figure 4G). Together, these results indicate that ATF4 regulates BCL11A transcription through binding at the BCL11A 155 enhancer in HUDEP2 and primary human erythroid cells.; page 2127, column 1, paragraph 1 - column 2, paragraph 2; figures 3-6 In summary, our study uncovers a major regulatory signaling pathway that extends linearly from HRI to ATF4 to BCL11A to g-globin (Figure 6D). This species-selective enhancer function is relevant when considering mouse models, such as the Townes or BERK mouse models, for the preclinical testing of g-globin in- ducers. Our findings suggest that lack of a phenotypic effect in mice should not discourage clinical advancement of therapeutic approaches with promising results in human cell systems.;	
1	page 2129, column 2 - page 2131, column 1,	

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	paragraph 3 Discussion Using a domain-focused CRISPR screen, we identified ATF4 as a novel regulator of g-globin through direct transcriptional control of BCL11A. Depletion of ATF4 or removal of the ATF4 element within the BCL11A 155 enhancer reduced BCL11A transcrip- tion. Removal of the corresponding ATF4 binding element at Bcl11a 155 enhancer in murine cells lowered Bcl11a levels only very modestly and insufficiently to reactivate BCL11A-controlled murine embryonic globin genes. Hence, the dependence of BCL11A transcription on binding of ATF4 to the BCL11A 155 enhancer varies between species. Together, these data de- lineate a linear HRI.ATF4.BCL11A.g-globin pathway with a species-selective component at the ATF4-BCL11A juncture (Figure 6E). The BCL11A 158 enhancer is a validated target for therapeutic genome editing.7,8 Our data show that disruption of the ATF4 element at the BCL11A 155 enhancer has an effect size not dissimilar to that of perturbation of the BCL11A158 enhancer inprimary erythroblasts, suggesting the BCL11A +55 enhancer could serve as an additional target for therapeutic genome editing.; page 2127, column 2, paragraph 3 - page 2129, column 1, paragraph 1; figure 6A -----	

1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2022/083904

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed.

   b. [ ] furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

      [ ] accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. [ ] With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2022/083904

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2021067233 A1	08-04-2021	EP 4017980 A1	29-06-2022
		US 2022380757 A1	01-12-2022
		WO 2021067233 A1	08-04-2021