

(19)



(11)

EP 3 272 356 A1

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
24.01.2018 Bulletin 2018/04

(51) Int Cl.:
A61K 38/46 (2006.01) C12N 9/22 (2006.01)

(21) Application number: **17178186.7**

(22) Date of filing: **22.02.2013**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME

- **GROUDINE, Mark T.**
Seattle, WA 98112 (US)
- **STODDARD, Barry L.**
Bellevue, WA 98006 (US)
- **TAKEUCHI, Ryo**
Seattle, WA 98125 (US)

(30) Priority: **24.02.2012 US 201261603231 P**

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
13751886.6 / 2 836 226

(71) Applicant: **Fred Hutchinson Cancer Research Center**
Seattle, WA 98109 (US)

(74) Representative: **Griffin, Philippa Jane et al**
Mathys & Squire LLP
The Shard
32 London Bridge Street
London SE1 9SG (GB)

(72) Inventors:
• **BENDER, Michael A.**
Seattle, WA 98136 (US)

Remarks:

- This application was filed on 27-06-2017 as a divisional application to the application mentioned under INID code 62.
- Claims filed after the date of receipt of the divisional application (Rule 68(4) EPC).

(54) **COMPOSITIONS AND METHODS FOR THE TREATMENT OF HEMOGLOBINOPATHIES**

(57) Provided are compositions and methods for the treatment of hemoglobinopathies such as thalassemias and sickle cell disease. Compositions and methods include one or more endonuclease(s) or endonuclease fusion protein(s), including one or more homing endonuclease(s) and/or homing endonuclease fusion protein(s) and/or CRISPR endonuclease(s) and/or CRISPR endonuclease fusion protein(s): (a) to disrupt a Bcl1 1a coding region; (b) to disrupt a Bcl1 1a gene regulatory region; (c) to modify an adult human β -globin locus; (d) to disrupt a HbF silencing DNA regulatory element or pathway, such as a Bcl1 1a-regulated HbF silencing region; (e) to mutate one or more γ -globin gene promoter(s) to achieve increased expression of a γ -globin gene; (f) to mutate one or more δ -globin gene promoter(s) to achieve increased expression of a δ -globin gene; and/or (g) to correct one or more β -globin gene mutation(s).

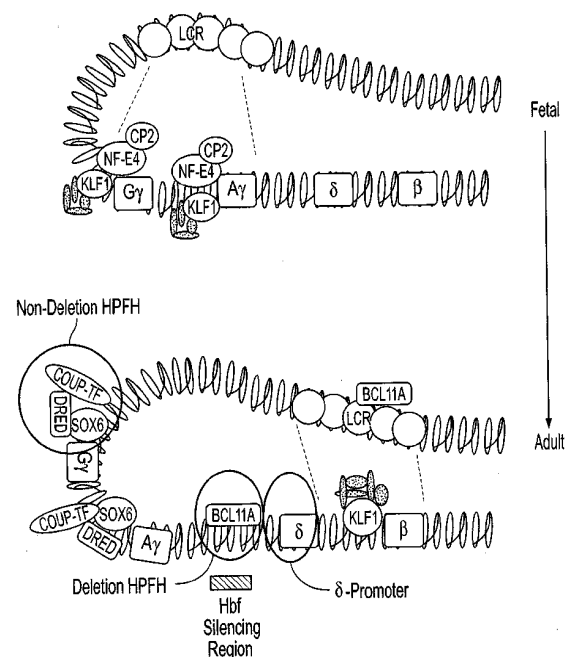


FIG. 1

EP 3 272 356 A1

Description**CROSS REFERENCE TO RELATED APPLICATIONS**

[0001] This application is being filed on 22 February 2013, as a PCT International patent application, and claims priority to U.S. Provisional Patent Application No. 61/603,231, filed February 24, 2012, the disclosure of which is hereby incorporated by reference herein in its entirety.

SEQUENCE LISTING

[0002] The present application includes a Sequence Listing in electronic format as a txt file entitled "sequence listing 54428.0006WOU1_ST25" and created on February 22, 2013 and which has a size of 174 kilobytes (KB). The contents of the txt file "sequence listing 54428.0006WOU1_ST25" are incorporated by reference herein.

BACKGROUND OF THE DISCLOSURE*Technical Field of the Disclosure*

[0003] The present disclosure relates, generally, to the treatment of genetic diseases. More specifically, the present disclosure provides endonuclease-based compositions and methods, including homing endonuclease- and Cas9 endonuclease-based compositions and methods, for altering the expression of globin genes, which compositions and methods are useful for the treatment of thalassemias, sickle cell disease, and other hemoglobinopathies.

Description of the Related Art

[0004] Hemoglobinopathies, such as thalassemias and sickle cell disease, are highly prevalent genetic red blood cell disorders that cause a significant health burden worldwide. Over 1,300,000 people with severe hemoglobin disorders are born each year. While 5% of people worldwide are carriers, the birth rates are 0.44 and 1.96 per thousand for clinically significant forms of thalassemia and sickle cell disease (SCD), respectively.

[0005] In the normal state, hemoglobins found in mammalian erythroid cells predominantly consist of heterotetramers of two α -like chains (polypeptides) and two β -like chains. The five genes of the β -globin locus reside in a cluster on chromosome 11. The genes are expressed in an erythroid, and developmentally stage specific manner; the ϵ , A_γ and G_γ , and δ and β genes being expressed primarily during the embryonic, fetal and post-natal periods respectively. At birth 95% of β -like chains are γ , with the rest being β . This ratio gradually inverts during the first year of life, explaining why phenotypes limited to the β -globin gene such as sickle cell and most β -thalassemias do not manifest until several months of age. Expression of the chromosome 16 based α -like genes differs; the embryonic ζ -gene parallels the expression of ϵ , but the twin α -genes are expressed from the fetal period onward. Thus α abnormalities manifest in utero, potentially with devastating consequences (e.g. hydrops fetalis). The resultant α -, β -heterotetramers are developmentally expressed; embryonic: Hb Gower1 ($\zeta_2\epsilon_2$), Hb Gower2 ($\alpha_2\epsilon_2$) and Hb Portland ($\zeta_2\gamma_2$); fetal: HbF (Fetal) ($\alpha_2\gamma_2$) and Adult: HbA2 ($\alpha_2\delta_2$) and HbA (Adult) ($\alpha_2\beta_2$).

[0006] β -thalassemia is caused by an abnormality in the adult β -globin locus, which results in an abnormal stoichiometry of β -like globin chains to α -like chains, resulting in the precipitation of the unpaired α -like chains. The severity of thalassemia is directly related to the degree of this globin chain imbalance. The ensuing damage mediated through several pathways including oxidation of cellular and membrane proteins culminates in ineffective erythropoiesis, apoptosis, and decreased red cell survival. Over 200 mutations have been described that are responsible for β -thalassemia.

[0007] Sickle cell disease is caused by a single nucleotide substitution within the β -globin gene, which results in glutamic acid being substituted by valine at amino acid position 6 of the peptide resulting in β^S . Hemoglobin S ($\alpha_2\beta^S_2$), which carries this mutation, is referred to as HbS, as opposed to the normal adult hemoglobin (HbA). Under conditions of low oxygen concentration HbS undergoes an allosteric change at which point it can polymerize. The deoxy- form of hemoglobin exposes a hydrophobic patch on the protein between the E and F helices. The hydrophobic valine at position 6 of the hemoglobin β -chain forms a hydrophobic patch which can associate with the hydrophobic patch of other hemoglobin S molecules causing hemoglobin S molecules to aggregate and form fibrous precipitates which, in turn, cause the red blood cells to adopt a sickle-shape and leads to alterations in numerous pathways that result in tissue damage via vaso-occlusion and hemolysis.

[0008] Although β -thalassemia and sickle cell disease (SCD) are quantitative and qualitative disorders, respectively, of the β -globin locus, the expression of normal β -like globin genes can ameliorate both diseases. In thalassemia, any improvement in the globin chain imbalance provides a selective advantage for each cell and results in clinical benefit. In sickle cell disease the presence of normal or certain mutant β -like chains can ameliorate the clinical phenotype by

competing for α -like chains more effectively than the mutant sickle cell chains thus reducing the amount of HbS, by forming hemoglobins that block the polymerization of Hbs (e.g. HbF) and increasing the amount of non-sickling hemoglobin per cell. For example, in sickle cell disease, fetal hemoglobin (HbF) levels of only 8% inhibit polymerization of HbS, which results in increased survival, while HbF levels of 20% provide nearly complete phenotypic correction. Critically, the progeny of donor erythroid cells containing normal HbA have a strong selective advantage following hematopoietic stem cell transplantation (HSCT) over endogenous derived cells containing HbS. A patient with 11% donor cells in the marrow had 35% donor BFUe and 73% donor erythrocytes, which resulted in transfusion independence. Thus, correction in a relatively small fraction of transplanted HSCs provides clinical benefit.

[0009] Severe forms of thalassemia require chronic transfusions, resulting in iron overload. Survival directly correlates with the efficacy of chelation, though cost, side effects, and compliance severely limit efficacy. The only FDA approved drug for SCD is hydroxyurea, which can attenuate morbidity and mortality. This treatment, however, is under-prescribed, compliance is poor, and it does not adequately protect health.

[0010] Hematopoietic cell transplantation (HCT) is an important therapeutic option for thousands of patients each year with hematologic malignancies and related disorders. According to the Center for International Blood and Marrow Transplant Research (CIBMTR), approximately 60,000 transplants were performed in 2009, an increase of over 15,000 transplants per year compared to a decade earlier. The effectiveness of transplantation is also increasing, with more recent outcomes demonstrating a significant reduction in the risk of relapse, non-relapse mortality, and overall mortality. Gooley et al., N. Engl. J. Med. 363:2091-101 (2011).

[0011] Allogeneic hematopoietic cell transplantation (HCT) from HLA-matched sibling or unrelated donors offers a cure for patients with hemoglobinopathies, but is limited by the need for a suitably matched related or unrelated donor and is complicated by graft versus host disease (GVHD) and infections. In addition, a major barrier is a high rate of graft failures, which is higher than observed for HCT for malignancies. Alternative approaches include performing HCT with donor cord blood cells, as cord blood donors can be identified for nearly all patients. Additional experimental approaches are focused on using a patient's own hematopoietic stem cells (HSCs) and inducing expression of the endogenous globin genes, or adding an exogenous β -like globin gene.

[0012] For many patients who are unable to find a donor, particularly those of ethnic minority or mixed race background, umbilical cord blood (CB) transplantation may offer the best hope for cure. A source of donor stem cells (easily collected at the time of birth without risk to the mother or infant), CB also has the advantage of being readily available and safely used in an HLA-mismatched setting without increasing the risk of GVHD.

[0013] Unfortunately, several factors, including the low cell dose available in many cord blood units lead to slow engraftment and an increase in transplant related mortality in adults and larger children. Significantly delayed hematopoietic recovery of both neutrophils and platelets is a known risk factor for cord blood transplant (CBT) recipients and is associated with the low total nucleated cell (TNC) and CD34⁺ cell doses provided in a single or double CB transplant. Similarly, these low cell numbers correlate with higher rates of graft failure, thus a particular concern in hemoglobinopathies where there is already high risk of graft failure. In fact, a recent analysis of adult single CBT recipients demonstrated that infused CD34⁺ cell dose is the most important predictor of myeloid engraftment.

[0014] Non-relapse mortality (NRM) is highest in double CBT (dCBT) recipients when compared to matched and mismatched unrelated donor recipients. Brunstein et al., Blood 116:4693-9 (2010). The majority of the NRM occurs within the first 100 days post transplant with infection being the most common cause of death. Importantly, an analysis of the risk factors for NRM among dCBT recipients revealed a higher risk in patients with delayed myeloid recovery (time to absolute neutrophil count (ANC) >500/ml) if the recovery was ≥ 26 days, the median time to engraftment in dCBT recipients. When, however, the analysis of risk factors for NRM was restricted to include only those dCBT recipients engrafting before day 26, no difference was found between the donor sources, emphasizing the important contribution of delayed engraftment to increased risk of NRM.

[0015] Moreover, an ANC of >100 on any given day post stem cell transplant has been previously shown to be a critical threshold for a decreased risk of mortality before day 100 post transplant (Offner et al., Blood 88:4058-62 (1996)). Thus, the significant delay in myeloid recovery that is observed in CBT recipients remains a critical barrier to successful outcomes in the CBT setting. The ability to increase not only the absolute number of CB progenitor cells available for transplantation, but also cells that can reliably result in more rapid myeloid recovery post-transplant, should improve overall survival for patients undergoing CBT. Strategies utilizing *ex vivo* expansion of cord blood stem/progenitor cells are being developed to overcome the low cell dose available in a cord blood graft with the goal of enhancing hematopoietic recovery and overall survival in CBT.

[0016] With the goal of overcoming the significant delay in neutrophil recovery that occurs following transplantation with umbilical cord blood (CB), the role of the Notch signaling pathway in regulating *ex vivo* expansion of hematopoietic stem/progenitor cells has been investigated to generate increased numbers of progenitor cells capable of rapid repopulation *in vivo*. A clinically feasible methodology utilizing an engineered Notch ligand (Delta1) has been developed, which results in a multi-log increase in the absolute numbers of CD34⁺ cells and a cellular therapy capable of rapid repopulation *in vivo*.

[0017] Infusion of expanded, partially HLA-matched cells results in a significant reduction in the median time to achieve an initial absolute neutrophil count (ANC) of 500/ml to just 11 days as compared to a median time of 25 days ($p < 0.0001$) in a concurrent cohort of 29 patients undergoing identical treatment but with two non-manipulated CB units. Although the number of patients treated was small (*i.e.* $n=14$), a significant effect on time to myeloid recovery was demonstrated,

as was the safety and clinical feasibility of this approach.

[0018] Despite tremendous investment of resources by many laboratories for over 30 years, there has been little progress in the development of therapeutic regimens for hemoglobinopathies, in large part due to the lack of identified drugable targets and the requirement for gene therapy vectors to persistently express at extremely high levels, while not leading to insertional mutagenesis. While increased expression of fetal hemoglobin (HbF) ameliorates both hemoglobinopathies, extensive research has not yielded viable new agents based on that observation. Hematopoietic stem cell (HSC) gene therapy with integrating lentiviral vectors is being pursued by several investigators. HSC gene therapy, however, requires high-level persistent expression and carries a substantial risk of insertional mutagenesis and leukemia.

[0019] What is critically needed in the art are compositions and methods, which exhibit improved efficacy for the treatment of hemoglobinopathies, including thalassemias and sickle cell disease while overcoming the safety concerns of existing therapeutic modalities.

SUMMARY OF THE DISCLOSURE

[0020] The present disclosure addresses these and other related needs in the art by providing, *inter alia*, compositions and methods for the treatment of hemoglobinopathies. Compositions and methods disclosed herein employ one or more polynucleotide that encodes one or more endonuclease(s) or endonuclease fusion protein(s), including one or more homing endonuclease(s) and/or homing endonuclease fusion protein(s) and/or one or more CRISPR endonucleases (*i.e.* Cas9 endonucleases in combination with one or more RNA guide strands) and/or CRISPR endonuclease fusion protein(s) (*i.e.* Cas9 endonuclease fusion protein(s) in combination with one or more RNA guide strands): (a) to disrupt a Bcl11a coding region or a Bcl11a gene regulatory region; (b) to disrupt a HbF silencing DNA regulatory element or pathway, such as a Bcl11a-regulated HbF silencing region; (c) to mutate one or more γ -globin gene promoter(s) to achieve increased expression of a γ -globin gene; (d) to mutate one or more δ -globin gene promoter(s) to achieve increased expression of a δ -globin gene; and/or (e) to correct one or more β -globin gene mutation(s).

[0021] Within a first embodiment, the present disclosure provides compositions and methods that comprise a polynucleotide that encodes one or more endonuclease(s), such as a homing endonuclease (HE) and/or a CRISPR endonucleases (*i.e.* Cas9 endonucleases in combination with one or more RNA guide strands) to achieve the targeted disruption of a sequence within a Bcl11a coding region or a Bcl11a gene regulatory region, thereby increasing to therapeutic levels the expression of an endogenous gene such as a γ - or a ϵ - globin gene. Within related aspects, the compositions of these embodiments comprise a polynucleotide that encodes one or more TALEN, one or more TALE-HE fusion protein, and/or one or more TREX2 protein.

[0022] Within a second embodiment, the present disclosure provides compositions and methods that comprise a polynucleotide that encodes one or more endonuclease(s), such as a homing endonuclease (HE) or a CRISPR endonucleases (*i.e.* Cas9 endonucleases in combination with one or more RNA guide strands) to achieve the targeted disruption of a key regulatory sequence within a β -globin gene locus, thereby increasing to therapeutic levels the expression of an endogenous gene such as a γ - or δ - globin gene. Within related aspects, the compositions of these embodiments comprise a polynucleotide that encodes one or more TALEN, one or more TALE-HE fusion protein, and/or one or more TREX2 protein.

[0023] Within certain aspects of this embodiment are provided HEs and CRISPR endonucleases that target a 3.6 kb region (SEQ ID NO: 1) within a β -globin gene locus (chr11:5212342-5215944 in HG18) that contains a binding site for the regulatory protein Bcl11a.

[0024] The homing endonucleases and CRISPR endonucleases described herein exhibit unique advantages over conventional gene targeting nucleases. Because they are broadly efficacious regardless of genotype, the homing and Cas9 endonucleases in combination with one or more RNA guide strands described herein are not patient specific, they provide clinical benefit in the heterozygotic state, and avoid the insertion of vector sequences.

[0025] Within a third embodiment, the present disclosure provides compositions and methods for recapitulating, via genome editing, one or more naturally-occurring mutation(s) within a patient's genome thereby providing clinical benefits including, for example, deletional or non-deletional forms of hereditary persistence of fetal hemoglobin (HPFH). More specifically, the present disclosure provides compositions and methods for achieving the direct correction of a thalassemia and/or a sickle cell disease (SCD) mutation through genome editing.

[0026] Within certain aspects of this embodiment, one or more homing endonuclease(s) is/are employed in combination with a normal or wild-type polynucleotide sequence (correction template) to permit the editing and/or repair of one or more genetic sequence, such as a β -like globin gene(s). These homing endonucleases permit the modification of key regulatory and/or coding sequences within a gene locus, exemplified herein by the human β -globin gene locus, through

the transient expression of a polynucleotide that includes one or more naturally occurring mutation(s). Within related aspects, the compositions of these embodiments comprise a polynucleotide that encodes one or more TALEN, one or more TALE-HE fusion protein, and/or one or more TREX2 protein.

[0027] More specifically, the present disclosure provides compositions and methods for genome editing, comprising one or more polynucleotides, each encoding a HE and a correction template, which may be employed to generate naturally-occurring mutations within stem cells, including, for example, hematopoietic stem cells (HSCs), embryonic stem (ES) cells, and induced pluripotent stem cells (iPSCs). Genome edited HSCs, ESs, and iPSCs, including autologous HSCs and iPSCs, may be transplanted into a patient to treat one or more hemoglobinopathies, such as a thalassemia and/or sickle cell disease.

[0028] The compositions and methods disclosed herein permit the efficient modification of HSCs, ESs, and iPSCs, through the transient expression of a polynucleotide encoding a HE with or without a targeting template, a Cas9 endonuclease, and/or an RNA guide strand, without the need for the persistent expression or insertion of an exogenous gene to achieve the amelioration of hemoglobinopathies in mature erythroid cells and in patient cells *in vivo*. Because these therapeutic methods do not require the integration and/or persistent expression of a transgene, the safety concerns associated with currently available gene therapy technologies are obviated.

[0029] Within a fourth embodiment, the present disclosure provides compositions and methods for the delivery of one or more homing endonuclease(s) and/or one or more Cas9 endonuclease(s) in combination with one or more RNA guide strands, each of which may be transiently expressed in targeted regions shown to have clinical benefit in humans. The endonuclease coding sequences described herein may be expressed in combination with, or fused to, a TAL effector nuclease (TALEN) coding sequence. Exemplified herein are TAL effector-HE (TALE-HE) fusion proteins and polynucleotides that encode those TALE-HE fusion proteins, which target critical genomic regions that influence fetal hemoglobin production.

[0030] Within certain aspects of these embodiments, a polynucleotide encoding one or more HE with or without a targeting template, one or more Cas9 endonuclease, one or more RNA guide strands, one or more TALEN, one or more TALE-HE fusion protein, and/or one or more TREX2 protein are operably linked to a promoter sequence within a viral vector to achieve the delivery and transient expression of a HE, a Cas9, an RNA guide strand, a TALEN, a TALE-HE fusion protein, and/or a TREX2 protein. Suitable viral vectors that may be satisfactorily employed for the delivery of HE, TALEN, TALE-HE fusion protein, and/or TREX2 protein may be selected from the group consisting of a coval pseudotyped lentiviral vector, a foamy virus vector, an adenoviral vector, and an adeno-associated viral (AAV) vector.

[0031] Within a fifth embodiment, the present disclosure provides compositions and methods comprising *ex-vivo* expanded modified hematopoietic stem cells (HSCs), which allow for efficient engraftment of corrected cells and the use of induced pluripotent stem cells (iPSCs) for screening and clinical application. Within certain aspects of these embodiments are provided compositions and methods for the efficient expansion of autologous HSCs, autologous gene-modified HSCs, iPSC-derived HSCs, and ES cells. Cord blood expansion methodology may be employed, which methodology utilizes Delta1 in serum free media supplemented with hematopoietic growth factors using mobilized peripheral blood CD34+ cells obtained from normal donors. These compositions and methods may be used in combination with one or more additional reagent to enhance the survival and proliferation of hematopoietic stem/progenitor cells. Within other aspects, these compositions and methods may employ endothelial cell co-cultures for the enhanced expansion of long-term repopulating cells, including corrected iPSC-derived HSCs.

[0032] Within a sixth embodiment, the present disclosure provides compositions and methods for providing supportive care, which compositions and methods comprise off-the-shelf cellular therapies that abrogate post-transplant neutropenia and improve outcome following transplantation of gene-corrected autologous HSCs. *Ex vivo* expanded, cryopreserved cord blood (CB) stem/progenitor cells may, for example, be administered as a means of supportive care to patients with thalassemia and/or sickle cell disease who are undergoing myeloablative HCT with autologous CD34+ gene corrected cells.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] Certain aspects of the present disclosure will be better understood in view of the following figures:

Figure 1 shows targets to increase expression of β -globin-like genes in adult erythroid tissues. Factors that are implicated in regulating the switch from a fetal expression pattern (two γ genes) to an adult program (δ and β) are displayed. (Adapted from Wilber et al., Blood 117(15):3945-3953 (2011)).

Figure 2 depicts three exemplary rare cleaving nuclease technologies.

Figure 3 is a graph showing that the risk of non-relapse mortality is highest among double CBT recipients. Non-relapse mortality after double CBT (DUCB), matched unrelated donor (MUD), mismatched unrelated donor (MMUD),

and matched related donor (SIB) transplant.

Figure 4 shows that a culture of CB progenitors with Delta1^{ext-IgG} results in more rapid neutrophil recovery in a myeloablative double CBT setting. The individual and median times (solid line) to ANC of $\geq 500/\mu\text{l}$ for patients receiving double unit CBT with two non-manipulated units ("conventional") versus with one *ex vivo* expanded unit and one non-manipulated unit ("expanded") is presented.

Figure 5 is a bar graph depicting the number of cord blood transplantations performed annually by disease indication.

Figure 6 (SEQ ID NO: 1) is the sequence of the 3.6 kb region for which the HbF silencing region falls, which spans chr11:5212342-5215944 in HG18.

Figure 7 (SEQ ID NO: 2) is the 350 base pair region spanning from a repeat element (chr11:5,213,912-5,214,261 in HG18), through the upstream French HPFH breakpoint known to disrupt the Bcl11a occupancy region within the HbF silencing region and that includes a GATA-1 binding motif, and from which exemplary homing endonucleases (HEs) of the present disclosure were designed.

Figure 8 (SEQ ID NO: 13) is the human beta-globin gene from 1 kb upstream of the cap through the polyA site, which spans from chr11:5203272-5205877 in HG18 (reverse strand).

Figure 9 (SEQ ID NO: 14) is a 606 bp region of the human beta-globin spanning from the promoter into Intron 2 (chr11:5204380-5204985 in HG18). This relatively small region contains the majority of mutations leading to severe thalassemia as well as the mutation causing sickle cell disease. This small region is readily amenable to homologous recombination resulting in gene correction.

Figure 10 (SEQ ID NO: 24) is the cDNA sequence for human Bcl11a cDNA (CCDS1862.1).

Figure 11 is a restriction map for the plasmid pET-21a(+).

Figure 12 is a restriction map for the plasmid pEndo (Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006).

Figure 13 is a schematic diagram of directed evolution for creating the BCL11A gene-targeting endonuclease. A constructed library was subjected to selection in IVC against a target site, a portion of which was replaced with the BCL11A gene target.

Figure 14 (SEQ ID NO: 28) is the nucleotide sequence of I-HjeMI (the parental enzyme for the BCL11A gene targeting nuclease), which is codon optimized for expression in *E. coli*.

Figure 15 (SEQ ID NO: 29) is the nucleotide sequence of I-HjeMI (the parental enzyme for the BCL11A gene targeting nuclease), which is codon optimized for mammalian expression.

Figure 16 (SEQ ID NO: 30) is the amino acid sequence of the homing endonuclease I-HjeMI.

Figure 17 (SEQ ID NO: 31) is the nucleotide sequence of a BCL11A gene targeting nuclease (Bcl11Ahje), which is based on the homing endonuclease I-HjeMI (obtained through directed evolution in IVC and in bacteria), which is codon optimized for expression in *E. coli*.

Figure 18 (SEQ ID NO: 32) is the nucleotide sequence of a BCL11A gene targeting nuclease based on the homing endonuclease I-HjeMI (obtained through directed evolution in IVC and in bacteria), which is codon optimized for mammalian expression.

Figure 19 (SEQ ID NO: 33) is the amino acid sequence of a BCL11A gene targeting nuclease based on the homing endonuclease I-HjeMI (obtained through directed evolution in IVC and in bacteria).

Figure 20 is a protein model showing the distribution of amino-acid residues different between the BCL11A gene-targeting endonuclease and its parental LHE I-HjeMI. Substituted residues of the BCL11A gene-targeting endonuclease are mapped on the crystal structure of I-HjeMI bound to its target site (PDB ID: 3UVF). D161 is deleted in the variant endonuclease.

Figure 21 is a bar graph showing the activity of a BCL11A gene-targeting endonuclease in a two-plasmid cleavage assay.

Figure 22A (SEQ ID NO: 34) nucleotide sequence of I-OnuI homing endonuclease (the parental enzyme for homing endonucleases targeting the HbF silencing region), codon optimized for expression in *E. coli*.

Figure 22B (SEQ ID NO: 15) is an amino acid sequence of I-OnuI homing endonuclease.

Figure 23 is an agarose gel showing the activity of an I-OnuI homing endonuclease targeting the HbF silencing region.

Figure 24 (SEQ ID NO: 35) is the nucleotide sequence of MegaTAL:5.5 RVD + Y2 I-Anil.

Figure 25 (SEQ ID NO: 36) is an amino acid sequence of MegaTAL:5.5 RVD + Y2 I-Anil.

Figure 26 (SEQ ID NO: 37) nucleotide sequence of Cas9 endonuclease (from Mali et al., Science (2013)).

Figure 27 (SEQ ID NO: 38) is the nucleotide sequence of an RNA Guide Strand for use with Cas9 endonuclease (from Mali et al., Science (2013)).

Figure 28 (SEQ ID NO: 62) is a nucleotide sequence of I-CpaMI homing endonuclease (ORF, codon optimized for mammalian expression).

Figure 29 (SEQ ID NO: 63) is an amino acid sequence of I-CpaMI homing endonuclease.

Figure 30 is an agarose gel showing the detection of targeted mutagenesis at the endogenous human BCL11A gene as described in Example 4.

Figure 31 is a restriction map for the plasmid pCcdB wt6 (Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006)).

Figure 32 (SEQ ID NO: 64) is a nucleotide sequence of a BCL11A gene targeting nuclease-encoding plasmid (pExodusBCL11Ahje).

Figure 33 (SEQ ID NO: 65) is a nucleotide sequence of TREX2-encoding plasmid (pExodus CMV.Trex2).

Figure 34 is a restriction map for the plasmid pExodusBCL11Ahje.

Figure 35 is a restriction map for the plasmid pExodus CMV.Trex2.

DETAILED DESCRIPTION OF THE DISCLOSURE

[0034] The present disclosure is directed, generally, to compositions and methods for the treatment of a genetic disease, such as a hemoglobinopathy, by the transient or persistent expression of a polynucleotide that encodes one or more endonuclease(s) or endonuclease fusion protein(s), including one or more homing endonuclease(s) and/or homing endonuclease fusion protein(s) and/or one or more Cas9 endonuclease(s) and/or Cas9 endonuclease fusion protein(s) in combination with one or more RNA guide strands: (a) to disrupt a Bcl11a coding region; (b) to disrupt a HbF silencing DNA regulatory element or pathway, such as a Bcl11a-regulated HbF silencing region; (c) to mutate one or more γ -globin gene promoter(s) to achieve increased expression of a γ -globin gene; (d) to mutate one or more δ -globin gene promoter(s) to achieve increased expression of a δ -globin gene; and/or (e) to correct one or more β -globin gene mutation(s). The compositions and methods disclosed herein find utility in the treatment of hemoglobinopathies, including β -thalassemia and sickle cell disease. The compositions and methods described herein may, optionally, comprise a polynucleotide that encodes one or more TALEN, one or more TALE-HE fusion protein, and/or one or more TREX2 protein.

[0035] The present disclosure will be better understood in view of the following definitions:

Definitions

[0036] As used herein, the term "hemoglobinopathy" refers to a class of genetic defects that result in an abnormal structure, abnormal function or altered expression of one or more of the globin chains of the hemoglobin molecule.

Hemoglobinopathies are inherited single-gene disorders. Common hemoglobinopathies include thalassemias and sickle-cell disease.

[0037] As used herein, the term "thalassemia" refers to a hemoglobinopathy that results from an altered ratio of α -like to β -like globin polypeptide chains resulting in the underproduction of normal hemoglobin tetrameric proteins and the accrual of free or unpaired α - or β - chains.

[0038] As used herein, the term "sickle-cell disease" 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 β^S -gene coding for a β -globin chain variant in which glutamic acid is substituted by valine at amino acid position 6 of the peptide, and second β -gene that has a mutation that allows for the crystallization of HbS leading to a clinical phenotype. The term "sickle-cell anaemia" refers to a specific form of sickle-cell disease in patients who are homozygous for the mutation that causes HbS. Other common forms of sickle cell disease include HbS / β -thalassemia, HbS / HbC and HbS / HbD. Table 1 discloses the nucleotide sequences encoding the initial amino acids of a wild-type and sickle cell β -globin chains

Table 1

β -globin chain	Sequence	Sequence Identifier
Wild-type	GTGCACCTCACTCCAGAGGAG	SEQ ID NO: 3
Sickle	GTGCACCTCACTCCAGTGGAG	SEQ ID NO: 4

[0039] As used herein, the term "hereditary persistence of fetal hemoglobin" or "HPFH" refers to, a benign condition in which significant fetal hemoglobin (hemoglobin F) production continues well into adulthood, disregarding the normal shutoff point.

[0040] As used herein, the term "globin" refers to a family of heme-containing proteins that are involved in the binding and transport of oxygen.

[0041] As used herein, the term "homing endonuclease" or "HE" refers to a class of restriction endonucleases that are characterized by recognition sequences that are long enough to occur only once in a genome and randomly with a very low probability (e.g., once every 7×10^{10} bp).

[0042] As used herein, the term "Transcription Activator-Like Effector Nuclease" or "TAL effector nuclease" or "TALEN" refers to a class of artificial restriction endonucleases that are generated by fusing a TAL effector DNA binding domain to a DNA cleavage domain.

[0043] As used herein, the term "three prime repair exonuclease 2" or "TREX2" refers to a nuclease having 3' exonuclease activity, which is typically involved in DNA replication, repair, and recombination.

[0044] As used herein, the term "Cas9 endonuclease" refers to an endonuclease that uses an RNA guide strand to target the site of endonuclease cleavage. The term "CRISPR endonuclease" refers to a Cas9 endonuclease in combination with an RNA guide strand. See, Jinek et al., Science 337:816-821 (2013); Cong et al., Science (Jan. 3, 2013) (Epub ahead of print); and Mali et al., Science (Jan. 3, 2013) (Epub ahead of print).

[0045] It will be understood that, unless indicated to the contrary, terms intended to be "open" (e.g., the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as "includes but is not limited to," etc.). Phrases such as "at least one," and "one or more," and terms such as "a" or "an" include both the singular and the plural.

[0046] It will be further understood that where features or aspects of the disclosure are described in terms of Markush groups, the disclosure is also intended to be described in terms of any individual member or subgroup of members of the Markush group. Similarly, all ranges disclosed herein also encompass all possible sub-ranges and combinations of sub-ranges and that language such as "between," "up to," "at least," "greater than," "less than," and the like include the number recited in the range and includes each individual member.

[0047] All references cited herein, whether *supra* or *infra*, including, but not limited to, patents, patent applications, and patent publications, whether U.S., PCT, or non-U.S. foreign, and all technical and/or scientific publications are hereby incorporated by reference in their entirety.

[0048] While various embodiments have been disclosed herein, other embodiments will be apparent to those skilled in the art. The various embodiments disclosed herein are for purposes of illustration and are not intended to be limiting, with the true scope and spirit being indicated by the claims.

Homing Endonucleases for AchievingHigh-efficiency, Multiplex Gene Disruption and Gene Editing Functions

[0049] As discussed above and exemplified below, the present disclosure provides compositions and methods comprising a polynucleotide that encodes one or more endonuclease(s), including one or more homing endonuclease(s) (HE(s)), such as one or more I-HjeMI homing endonuclease(s), I-CpaMI homing endonuclease(s), and/or I-OnuI homing endonuclease(s), and/or one or more Cas9 endonuclease in combination with one or more RNA guide strands, which may be transiently or persistently expressed in targeted cells shown to have clinical benefit in humans. Exemplary endonucleases target critical genomic regions that influence fetal hemoglobin production by: (a) disrupting a Bcl11a coding region or a Bcl11a gene regulatory region; (b) disrupting a HbF silencing DNA regulatory element or pathway, such as a Bcl11a-regulated HbF silencing region; (c) mutating one or more γ -globin gene promoter(s) to achieve increased expression of a γ -globin gene; (d) mutating one or more δ -globin gene promoter(s) to achieve increased expression of a δ -globin gene; and/or (e) correcting one or more β -globin gene mutation(s). The compositions and methods disclosed herein find utility in the treatment of hemoglobinopathies, including β -thalassemia and sickle cell disease.

[0050] The endonuclease coding sequences described herein may be expressed in combination with, or fused to, a DNA binding domain coding sequence, such as a TAL effector (TALE) coding sequence or a nuclease coding sequence such as a three prime repair exonuclease 2 (TREX2) coding sequence. Exemplified herein are TALE-HE fusion proteins and polynucleotides that encode one or more TALE-HE fusion protein(s).

[0051] Four protein scaffolds are known in the art for achieving targeted gene modification and disruption in eukaryotes: zinc finger nucleases (ZFNs), TAL effector nucleases (TALENs), homing endonucleases (HEs), and Cas9 endonucleases in combination with a RNA guide strand. The present disclosure employs TAL effector nucleases, homing endonucleases, and/or Cas9 endonucleases either alone or in combination. TAL nucleases offer more straightforward modular design and higher DNA recognition specificity than zinc finger nucleases while homing endonucleases, such as LAGLIDADG homing endonucleases (LHEs), offer highly specific cleavage profiles, compact structures, and, because they are compact monomeric proteins that do not require dimerization as do ZFNs and TALENs, the ability to be used in multiplex combinations. Accordingly, HEs and CRISPR endonucleases (i. e. Cas9 endonucleases in combination with one or more RNA guide strands) are extremely efficient in mediating gene disruption. Stoddard, *supra* and Mali et al., Science (2013), *supra*.

[0052] As part of the present disclosure, a critical region within the β -globin locus that suppresses HbF function has been identified. This region provides multiple targets for HE- and Cas9-mediated cleavage. Specifically-designed nucleases may be tested for activity against a cognate target site and for off-target activity against any closely related genomic targets. These HEs and Cas9 endonucleases in combination with one or more RNA guide strands may be engineered to avoid off-target genomic cleavage using the methods described in Stoddard, Structure 19:7-15 (2011) and Mali et al., Science (2013). HEs and Cas9 endonucleases in combination with one or more RNA guide strands that are disclosed herein are capable of directly targeting the γ - and δ -promoters and replacing a 606 bp region (SEQ ID NO: 14) that spans the majority of thalassemia mutations as well as the HbS mutation.

[0053] To facilitate the generation of large deletions spanning the HbF silencing region or subsets thereof one or more HEs and/or one or more Cas9 endonucleases in combination with one or more RNA guide strands may be co-transduced with a bridging oligonucleotide, which spans from the endonuclease cleavage site to the end of the target region. Chen et al., Nat. Methods 8(9):753-5 (2011). Higher frequency genome editing may be achieved by employing one or more HEs that bind to and cleave a sequence that flanks each side of the target region. Similarly, a HE and a mutagenizing oligonucleotide may be used to introduce promoter region mutations, which leads to elevated expression of a gamma or delta gene.

[0054] The presently disclosed HEs may be first evaluated in an erythroid cell line and in human CD34+ cells that are induced to differentiate to erythroid cells, thereby confirming the ability to alter globin gene expression. Depending upon the precise application contemplated, one, two, three, or more HEs may be delivered to facilitate the generation of larger deletions. The suitability of individual HEs can be assessed in additional culture and animal model assays to confirm their ability to target HSCs without compromising pluripotency and expansion potential, and to assess clinical benefit in hemoglobinopathy models. One or more HEs and one or more exonuclease, such as a TREX2 exonuclease or a TAL effector exonuclease, may be delivered to CD34+ HSC for the induction of targeted genetic deletions in critical regions for HbF.

[0055] Individual nucleases may be tested against a series of targets in a 350 bp region (SEQ ID NO: 2) defined by a region initiating at the edge of a repeat element and spanning through the upstream French HPFH breakpoint known to disrupt the Bcl11a occupancy region within the HbF silencing region and that includes a GATA-1 binding motif. Initial analyses have identified seven targets, evenly distributed throughout the region, which comprise DNA sequence modules for which pools of highly active endonuclease variants have been isolated and sequenced. Notably, one target overlaps with a potential Bcl11a binding motif and is adjacent to the GATA-1 motif. Successively larger deletions of a target region

may be achieved by transducing two, three, or more HEs. Alternatively, multiple targets for disrupting the Bcl11a gene have been identified on the 5'-end of the gene to ensure the elimination of gene function. Similarly, multiple optimal targets for Cas9 / RNA guide mediated disruption have been identified through the area that can be used singly, or in combination leading to larger deletions.

[0056] Individual HEs may be tested in transfected human cell lines using integrated genomic reporters, and may further employ additional selection steps to further optimize cleavage and gene conversion activities using protocols as described in Stoddard, *supra*. The validation and delivery of individual targeted HEs that are active against targets in the globin locus may be followed by vectorization of the nucleases in expression systems. For example, an expression system may be employed that links each HE to the coexpression of a nuclease, such as a TALEN and/or a TREX exonuclease, to achieve greatly enhanced gene disruption efficiency in transduced cells.

[0057] The present disclosure also provides TALE-HE fusion proteins, and polynucleotides that encode TALE-HE fusion proteins, which exhibit the desired feature of restricting the recruitment and activity of engineered HEs to the desired target site, such as within a globin locus, through the synergistic recognition of adjacent DNA targets by the TALE and HE scaffolds. Such TALE-HE fusions combine the most favorable properties of each scaffold (*i.e.*, modular assembly of TALEs and nuclease specificity of HEs) while reducing nonspecific nuclease activity that is associated with traditional TALENs or zinc finger nucleases (ZFNs).

[0058] The high-resolution crystal structures have been determined for ten separate LAGLIDADG homing endonucleases (LHEs) in complex with their cognate DNA target sites. Stoddard, *Structure* 19:7-15 (2011) and Takeuchi et al., *Proc. Natl. Acad. Sci. U.S.A.* 108:13077-13082 (2011). Chimeric 'hybrids' of those LHEs have been constructed that, collectively, provide a broad range of LHE targeting proteins for gene-specific applications. Baxter et al., *Nucl. Acids Res.* 40(16):7985-8000 (2012).

[0059] HEs having suitable target sequence specificity may be identified by a yeast surface display strategy, combined with high-throughput cell sorting for desirable DNA cleavage specificity. A series of protein-DNA 'modules', which correspond to sequential pockets of contacts that extend across the entire target site, may be systematically randomized in separate libraries. Each library may then be systematically sorted for populations of enzymes that can specifically cleave each possible DNA variant within each module, and each sorted population deep-sequenced and archived for subsequent enzyme assembly and design. HEs that may be suitably employed in the compositions and methods of the present disclosure are commercially available (Pregenen, Seattle, WA).

[0060] Within certain aspects, the compositions and methods described herein may employ the co-expression of one or more HE, including, for example, one or more LHE, with a TREX2 3' exonuclease. In contrast to the 5' overhangs left by current versions of ZFNs and TALENs, HEs generate 3' overhangs at the site of targeted double-strand breaks, which results in an enhanced rate of end processing following HE cleavage. Near complete modification of a double strand break site in primary cells can be achieved through HE/TREX2 co-expression. Because of the way HE/TREX2 co-expression influences break processing, this combination achieves multiple targeted deletions in one region and increases the safety of nuclease-induced targeted gene disruption by diminishing break persistence and reducing the potential for large scale translocations mediated through alternative end joining pathways.

[0061] The crystal structure of a TAL effector (PthXo1) bound to its DNA target site has recently been determined. Mak et al., *Science* 335(6069):716-9 2012; e-pub 05Jan12 PubMed PMID: 22223736. These crystal structure data permit the precise definition of the boundaries of DNA recognition region and facilitates strategies for the creation of well-behaved TALEN-HE, or other TALEN-nuclease fusion construct, which may be applied to achieve a variety of complex genomic manipulations.

Genome Disruption to Bcl11a Gene Expression

[0062] Knockout of Bcl11a in a sickle cell mouse model ameliorates disease, supporting the clinical relevance of this pathway. Xu et al., *Science* 334:993-6 (2011). In addition, mice containing a YAC transgene spanning the human β -globin locus are used to model perturbations in Bcl11a mediated silencing of HbF. Heterozygous and homozygous knockout of the endogenous Bcl11a gene in these mice results in γ -globin mRNA comprising 20 and 76% of total β -like mRNA respectively, compared to 0.24% in controls. Sankaran et al., *Nature* 460:1093-7 (2009). This suggests Bcl11a acts as rheostat, modulating the degree of HbF suppression. Consistent with this, decrease of function mutations in Bcl11a result in elevated levels of HbF and a lessening of the clinical thalassemia and/or sickle cell disease phenotype. Galanello et al., *Blood* 114:3935-7 (2009).

[0063] Within certain embodiments, the present disclosure provides compositions and methods that comprise one or more endonuclease(s), including one or more homing endonuclease(s) (HE(s)), such as one or more I-HjeMI homing endonuclease(s), I-CpaMI homing endonuclease(s), and/or I-OnuI homing endonuclease(s), and/or one or more Cas9 endonuclease(s) to achieve the disruption of a sequence that encodes Bcl11a or its key regulatory sequences. As described in greater detail and exemplified herein, compositions and methods comprising one or more Cas9 endonucleases further comprise one or more Bcl11a gene-specific RNA guide strands to mediate the targeting of the Cas9

endonuclease to a Bcl11a gene sequence.

[0064] The Bcl11a gene has multiple exons spanning over 100kb and results in several splice variants that lead to proteins associated with different activities. As part of the present disclosure, several DNA targets have been identified that are transcribed into multiple Bcl11a splice variants. All disrupt the long (L) and extra-long (XL) forms, which are associated with the greatest HbF silencing activity, while one disrupts all forms of Bcl11a. These targets comprise DNA sequence modules for which pools of highly active endonuclease variants have been isolated and sequenced. The human Bcl11a cDNA sequence (CCDS1862.1) is presented herein as SEQ ID NO: 24 (Figure 10).

[0065] Thus, within certain aspects, the present disclosure provides compositions for achieving therapeutic levels of HbF, which compositions comprise a polynucleotide encoding one or more homing endonuclease (HE), which is capable of mediating the disruption of the nucleotide sequence within this 1.3 kb region, thereby preventing the binding of Bcl11a, and the formation of the corresponding repressive complex, and de-repressing γ -globin expression.

Genome Disruption to Block Bcl11a-mediated Silencing of HbF

[0066] As summarized above, within certain embodiments, the present disclosure provides compositions and methods for treating and/or ameliorating a genetic disease, such as a hemoglobinopathy, including a thalassemia and/or sickle cell disease. Certain aspects of these embodiments include the transient expression of a polynucleotide encoding one or more homing endonuclease(s) to disrupt a HbF silencing element or pathway within a β -globin gene locus or a δ -globin gene locus thereby increasing to therapeutic levels the expression of an endogenous gene such as a γ - or δ -globin gene.

[0067] The compositions disclosed herein comprise a polynucleotide encoding one or more homing endonuclease(s) (HE(s)) and one or more Cas9 endonuclease in combination with one or more RNA guide strands and, optionally, one or more transcription activator-like (TAL) effector(s), to achieve the targeted disruption of key regulatory sequences within the β -globin gene locus. More specifically, the compositions and methods disclosed herein achieve an increase in γ -globin gene expression, and consequent HbF protein production, by removing essential elements for Bcl11a binding to the HbF silencing region(s) within the β -globin gene locus.

[0068] During normal development, expression of an embryonic β -like gene (ϵ -globin) is sequentially replaced by a pair of γ -globin genes in the fetus and the δ - and β -globin genes in the adult. In adult erythroid tissues, the zinc finger protein Bcl11a binds to a region between the γ -globin and δ -globin genes within the β -globin gene locus thereby silencing the production of HbF. The importance of Bcl11a-mediated silencing of HbF is supported by knockdown of Bcl11a mRNA in human CD34 cells, which increases HbF levels to 24-36% of total β -like proteins. Sankaran et al., Science 322:1839-42 (2008). Removal of this region in the deletion form of HPFH, as well as the knockdown of Bcl11a, blocks Bcl11a-mediated HbF silencing and results in an elevated level of γ -globin gene expression and HbF protein production in adult erythroid tissues (Sankaran et al., N. Engl. J. Med. 365:807-14 (2011)).

[0069] While multiple mechanisms contribute to an elevation in HbF protein levels, it has been shown that a 3.6 kb region is key for HbF silencing (SEQ ID NO: 1). Sankaran et al., N. Engl. J. Med. 365:807-14 (2011). While there are several peaks of Bcl11a enrichment in the β -globin locus, the single peak in the 3.6 kb HbF silencing region stands out as proteins known to form a repressive complex with Bcl11a are bound in this region (GATA-1 and HDAC-1) and the chromatin is enriched for the repressive histone mark trimethylation of histone H3 on lysine27.

[0070] Notably this 3.6 kb region contains a single peak of Bcl11a binding downstream of the γ -gene. Multiple point mutations have been identified in the γ -globin promoters that result in HbF levels of 20-30% as a heterozygote, ameliorating thalassemia and SCD. These point mutations cluster to three regions all within 200 bp of the γ -cap sites: (1) -200, a GC-rich region bound by SP1 and a stage specific protein complex; (2) -175, bound by GATA-1; and (3) Oct1 and a CCAAT motif at -117 bound by several additional factors. Forget, Ann. NY Acad. Sci. 850:38-44 (1998).

[0071] Mutations within these three regions block the binding of a repressive complex in adult erythroid cells. Consequently, these regions are suitable targets for HE-mediated disruption and targeted mutation by the compositions and methods disclosed herein. The disruption of these regions leads to a decrease in repressive complexes, which results in an elevated level of γ -globin gene expression, and a corresponding increase in HbF protein production to levels that are sufficient to achieve therapeutic efficacy in methods for the treatment of hemoglobinopathies, including β -thalassemia and sickle cell disease.

[0072] A single peak of Bcl11a occupancy is present within the 3.6 kb HbF silencing region (Sankaran et al., N. Engl. J. Med. 365:807-14 (2011)) (SEQ ID NO: 1). This region of Bcl11a occupancy is disrupted by the upstream breakpoint of French HPFH Sankaran et al., N. Engl. J. Med. 365:807-14 (2011). Described herein is a 350 bp region initiating at the edge of a repeat element and spanning through the upstream French HPFH breakpoint known to disrupt the Bcl11a occupancy region within the HbF silencing region and that includes a GATA-1 binding motif (SEQ ID NO: 2). The base before the upstream French HPFH deletion is HG18 chr11:5,214,023. The GATA-1 motif spans chr11:5,214,200-5,214,206. Without being limited by theory, it is believed that GATA-1 and HDAC-1 form a repressive complex with Bcl11a when Bcl11a is bound within this 350 bp region and this leads to the formation of a repressive

complex that inhibits the expression of the γ -globin genes and, thereby, reduces cellular levels of HbF protein.

[0073] The HE-mediated disruption, which is achieved by the compositions and methods disclosed herein, occurs at high efficiency. Unlike shRNA knockdown approaches that are known in the art, the highly sequence specific disruption of the HbF silencing region, which is mediated by the homing endonucleases disclosed herein, avoids off target effects at other Bcl11a binding sites in the genome, and in other cell types, especially within B-cells where Bcl11a binding is required for normal development.

[0074] Thus, the homing endonucleases described herein exhibit unique advantages over conventional gene targeting nucleases. Because they are broadly efficacious regardless of genotype, the homing endonucleases described herein are not patient specific and provide clinical benefit in the heterozygotic state.

Recapitulation of Genetic Modifications for Correcting a Thalassemia or Sickle Cell Disease Mutation

[0075] Within other embodiments, the present disclosure provides compositions and methods for recapitulating, via genome editing, one or more naturally-occurring mutation(s) within a patient's genome to provide clinical benefits. More specifically, the present disclosure provides compositions and methods for achieving the direct correction of a thalassemia and/or sickle cell disease (SCD) mutation through genome editing.

[0076] The compositions and methods disclosed herein employ a correction template to achieve gene editing and correction to ameliorate hemoglobinopathies, including thalassemias and sickle cell disease, by enhancing the rate of homologous recombination (HR) between the correction template and the corresponding mutated sequence within a patient's genome. Exemplified herein are compositions and methods for correcting an underlying β -globin mutation, which provide clinical benefit in the heterozygotic state while avoiding the insertion of vector sequences. These compositions and methods may be used independently from or in combination with the compositions and methods described above for the disruption of Bcl11a-mediated gene silencing.

[0077] The present disclosure provides a robust set of technologies for genome editing that exploits the advantages of HEs, as compared to alternative platforms that are available in the art. These HEs can be combined with a TAL effector modular DNA binding platform to achieve additional therapeutic advantages.

[0078] While homologous recombination (HR) to edit genomes is powerful, it is inefficient. Introduction of a double stranded break at the region to be modified results in a tremendous increase in HR efficiency. Simultaneous introduction of a polynucleotide encoding a HE and a correction template, wherein the correction template comprises as little as 100 bp of flanking homology, allows an increased frequency of HR, thereby permitting genome editing as the corrective template is introduced.

[0079] The transduction of cells with a short synthesized correction template may also be employed for the efficient introduction of defined single base-pair mutations. Such approaches typically exploit a single HE. Alternatively HEs may be transduced that flank the region targeted for modification. Correction templates may be transduced by optimized methods as described herein. The design, transduction, and evaluation of HEs may be performed, as discussed in detail below, according to the methodology described Certo et al., Nat Methods 8:671-6 (2011) and Jarjour et al., Nucleic Acids Res 37:6871-80 (2009).

[0080] Within certain aspects of these embodiments, one or more homing endonuclease(s) is/are employed in combination with a normal or wild-type polynucleotide sequence to permit the editing and/or repair of one or more β -like globin gene(s). For example, the present disclosure provides compositions and methods for the treatment of hemoglobinopathies, which compositions and methods permit the modification of key regulatory and/or coding sequences within a gene locus, exemplified herein by the human β -globin gene locus, through the transient expression of a polynucleotide that includes one or more naturally occurring mutation(s).

[0081] More specifically, the present disclosure provides compositions and methods for genome editing, which may be employed to generate mutations that recapitulate naturally-occurring mutations within stem cells, including, for example, hematopoietic stem cells (HSCs), embryonic stem (ES) cells, and induced pluripotent stem cells (iPSCs). Genome edited HSCs, ESs, and iPSCs, including autologous HSCs and iPSCs, may be transplanted into a patient to treat one or more hemoglobinopathies, such as a thalassemia and/or sickle cell disease. The compositions and methods disclosed herein permit the efficient modification of HSCs, ESs, and iPSCs, without the need for the persistent expression or insertion of an exogenous gene to achieve the amelioration of hemoglobinopathies in mature erythroid cells and in patient cells *in vivo*.

[0082] Because these therapeutic methods do not require the integration and/or persistent expression of a transgene, the safety concerns associated with currently available gene therapy technologies are obviated. Within certain aspects of these embodiments, the compositions and methods employ one or more polynucleotide for the targeted disruption of Bcl11a-mediated silencing of HbF.

[0083] Exemplified herein are compositions and methods that permit the recapitulation of genetic modifications within one or more HbF silencing region(s) that is/are responsible for hereditary persistence of HbF (HPFH). Because such genetic modifications lead to increased expression of a therapeutically effective gene, the recapitulated genetic modifi-

cations need only be present as a heterozygote to achieve therapeutic efficacy.

[0084] The compositions and methods for ameliorating thalassemia and sickle cell disease that are disclosed herein achieve therapeutic efficacy by introducing one or more mutation that result in increased HbF and/or HbA₂ and/or HbA protein production. Exemplified herein are compositions and methods for recapitulating one or more naturally-occurring deletion(s) of the β -globin gene and/or regions, which activate γ -globin gene expression thereby increasing levels of fetal hemoglobin. Because a modest increase in HbF and/or HbA₂ protein production is sufficient to ameliorate these disease phenotypes, heterozygotic mutations are sufficient to achieve substantial therapeutic benefit.

[0085] Within certain aspects, the delivery of a correction template may be done in conjunction with the delivery of a selectable marker gene thereby permitting the selection of corrected cells *ex vivo* and *in vivo*, although such an approach requires long-term expression via integration of the selectable marker gene. Beard et al., J. Clin. Invest. 120:2345-54 (2010) and Munoz et al., Nucleic Acids Res. 39(2):729-743 (2011).

[0086] Activation of β -globin expression in adult tissues depends upon binding of KLF-1 at a CACCC box in its promoter. The δ -globin promoter lacks an intact CACCC box, KLF-1 is not bound and expression is limited to 2% of β -globin. Mutations of the δ -promoter that recapitulate the β -globin promoter by, for example, introducing an intact CACCC box, allow KLF-1 binding, and result in a therapeutically efficacious increase in δ -globin expression.

[0087] Within certain aspects of these embodiments, a non-deletion HPFH γ -globin promoter mutation may be generated. Only a single base pair must be modified to achieve efficacy. For example, a -175 T→C mutation (SEQ ID NO: 21) may be recapitulated to maximize the levels of HbF. Mutation of any of the four γ -globin genes will provide benefit, thus increasing potential targets.

Delivery of Homing Endonucleases, Cas9 Endonuclease, TAL Effector Nucleases, and TREX2 Exonucleases

[0088] Within further embodiments, the present disclosure provides systems, in particular non-integrating vector systems, for the delivery of one or more HE, Cas9, TALEN, and/or TREX2 nuclease described herein. There are three major challenges to the therapeutic gene editing of hematopoietic stem cells (HSCs): (1) nuclease reagents must be transiently delivered to HSCs; (2) gene editing efficiency in cells receiving a nuclease must be high; and (3) gene-edited HSCs must engraft to a level sufficient for therapeutic effect. These challenges may be overcome by employing various vectorization approaches.

[0089] Exemplified herein are cocl pseudotyped lentiviral vectors and foamy virus vectors for the efficient gene transfer to HSCs. Trobridge et al., Mol Ther 18:725-33 (2008). Alternatively, adenoviral vectors may be modified as previously described for use in gene transfer to HSCs. Wang et al., Exp. Hematol. 36:823-31 (2008) and Wang et al., Nat. Med. 17:96-104 (2011). Within other aspects of these embodiments, AAV-based vector systems may also be employed for the delivery of HEs, Cas9 (and/or RNA guide strands), TALE-HE, TALENs, and/or TREX2 nucleases.

[0090] AAV6-serotype recombinant AAV vectors provide a 4.5 kb payload, sufficient to deliver a promoter-HE-exonuclease or a promoter-TAL-HE fusion-exonuclease cassette in addition to a small recombination template. Alternatively, it can carry the small Cas9 polypeptide and guide RNAs. AAV6 provides efficient transduction of human CD34⁺ umbilical cord blood cells of all known AAV capsids and is able to mediate significant levels of transient gene expression in HSC. Self-complementary and single stranded AAV6 vectors may be employed for both gene knockout and recombination-based gene editing in HSC in cell lines and in primary CD34⁺ cells.

[0091] Adenoviral vectors with hybrid capsids are capable of efficiently transducing many types of hematopoietic cells including CD34⁺ cells. Improved transduction may be achieved with a chimeric adenoviral vector using the serotype 35 fiber (Ad5-F35) and the serotype 11 fiber (Ad5-F1) for efficient transduction of hematopoietic cells. Helper-dependent adenoviral vectors offer up to a 30 kb payload, along with transient gene expression in HSC, and can be used to deliver multiple HE/exonuclease cassettes, HE-TAL fusions, as well as very large recombination templates. Alternatively it can carry the small Cas9 polypeptide and guide RNAs. These modified chimeric adenovirus vectors may, therefore, be employed for both gene knockout and recombination-based gene editing in HSC.

[0092] Integration-deficient lentiviral and foamyviral vectors (IDLV and IDFV) provide 6 kb (IDLV) to 9 kb (IDFV) payloads, and have well documented capabilities to transduce human HSCs. Within certain aspects, both IDLV and IDFV vectors may be employed for gene knockout and recombination-based gene editing in HSC. IDLV with alternative promoter GFP cassettes provide efficient and high level expression in CD34⁺ HSC. High titer stocks may be achieved using a TFF purification step. Vectors with a set of promoter/GFP cassettes may be used to provide efficient and high level HE expression in CD34⁺ HSC and may be generated to express individual HEs, HE/Trex2, multiplex-HE (*i.e.*, two, three, or four HEs that are co-expressed), and multiplex-HE/TREX2 combinations. Multiplex HE expression permits multiple cleavage events in a critical region, which depending upon the precise application, may be desired to create increased HbF derepression. Such multiplex strategies are feasible with LHEs, because they function autonomously, and may be satisfactorily employed in combination with TREX2 co-expression to permit highly efficient and synchronous processing of closely-targeted double strand breaks. Alternatively it can carry the small Cas9 polypeptide and guide RNAs.

[0093] The efficiency of gene targeting, levels of globin gene expression in individual targeted cells as well as popu-

lations of cells and of their progeny, the effect of targeting on erythropoiesis and on stem cell function, and on hematologic parameters and organ function may be confirmed in model organisms.

[0094] Transductions may be followed by single-cell and bulk population assessments of modification efficiency and expression of β -like genes at the RNA and protein levels. Alterations in factor binding and chromatin structure may be assessed, as well as morphology, the extent of ineffective erythropoiesis and apoptosis. Candidates that score well in initial screens may be further assessed for effects on HSC pluripotency as well as the ability to ameliorate disease specific phenotypes *in vitro* and *in vivo*.

[0095] Initial screening of HE candidates and delivery systems may be performed in a mouse erythroleukemia cell line containing a single intact human chromosome 11 (N-MEL) and clinical grade CD34+ normal human HSCs with endpoints of assessing targeted mutation efficiency and globin gene expression. Both cell types can be induced to differentiate along an erythroid path during which expression of β -like genes is highly induced with high β - to γ - and δ -ratios allowing a quantitative assessment of effects globin gene regulation at a single-cell and population level. Second level assessments may include an analysis of the pluripotency of transduced CD34+ cells and erythropoiesis. Suitable assay systems may include culturing to assess long-term proliferative potential, analysis of myeloid and erythroid colonies for clonal analysis and transplant into NOD SCID gamma (NSG) mice followed by assessment of multilineage engraftment of primary and secondary recipients. Clinical effectiveness may be assessed simultaneously *in vitro* and *in vivo*.

[0096] Knockout of murine Bcl11a leads to a dramatic dose-dependent increase in γ -globin in mice containing a human β -globin locus and ameliorates the sickle phenotype in humanized mouse models. While both systems allow the analysis of globin gene expression, the sickle mice allow for the assessment of the improvement of phenotype in these mice with special attention to the hematologic parameters, liver and lung pathology, renal function and spleen size. Phenotypic improvement may be correlated to the number of HbF containing cells, the HbF/HbS ratio and expression patterns in single cell assays.

[0097] In addition, erythrocyte lifespan and morphology may be assessed by transducing human CD34+ HSCs from hemoglobinopathy patients. Cultured thalassemic cells show minimal expansion, a lack of hemoglobinization, evidence of ineffective erythropoiesis and increased apoptosis compared to normals. These features permit the quantitative assessment of expression levels and degree of erythropoiesis post-targeting. The degree of sickling of erythroid progeny of CD34+ cells under hypoxic conditions may also be assessed. CD34+ cells from patients may be transplanted into NSG mice, after which several features of abnormal erythropoiesis are recapitulated, allowing assessment of the effect of targeted mutagenesis.

Expansion of Autologous HSCs, ESs, and iPSC-derived HSCs

[0098] Within further embodiments, the present disclosure provides compositions and methods for the *ex-vivo* expansion of modified hematopoietic stem cells (HSCs) to allow for efficient engraftment of corrected cells and the use of induced pluripotent stem cells (iPSCs) for screening and clinical application. Within certain aspects of these embodiments are provided compositions and methods for the efficient expansion of autologous HSCs, autologous gene-modified HSCs, ESs, and iPSC-derived HSCs. Cord blood expansion methodology may be employed, which methodology utilizes Delta1 in serum free media supplemented with hematopoietic growth factors using mobilized peripheral blood CD34+ obtained from normal donors. These compositions and methods may be used in combination with one or more additional reagent to enhance the survival and proliferation of hematopoietic stem/progenitor cells. Within other aspects, these compositions and methods may employ endothelial cell co-cultures for the enhanced expansion of long-term repopulating cells, including corrected iPSC-derived HSCs.

[0099] For effective clinical translation of the presently disclosed gene correction strategies, the present disclosure provides methods for the *ex vivo* expansion of the absolute number of corrected autologous HSCs. Gene correction procedures are generally more efficient if done in a smaller scale and often only limited numbers of HSCs are available for correction. Thus, it is contemplated by the present disclosure that expansion methods may be employed to permit clinically feasible *ex vivo* expansion of corrected HSCs, ESs, and/or HSCs derived from induced pluripotent stem cells (iPSCs). Within certain aspects, the present disclosure provides methods for expanding hematopoietic stem/progenitor cells for therapeutic application by exploiting the role of Notch signaling in determining stem cell fate. Dahlberg et al., Blood 117:6083-90 (2010); Delaney et al., Nat Med 16:232-6 (2010); and Varnum-Finney et al., Nat Med 6:1278-81 (2000).

[0100] These methods permit the clinically relevant *ex vivo* expansion of cord blood stem/progenitor cells, and an expanded cellular therapy for treatment of myelosuppression in patients undergoing cord blood transplantation, by first using a partially HLA-matched fresh product (harvested post-culture and infused directly) and/or by using a previously expanded and cryopreserved product as an off-the-shelf non-HLA matched cellular therapy.

[0101] *Ex vivo* expansion of gene-corrected autologous HSCs enhances the safety and effectiveness of HSC-based gene therapy by permitting the transplantation of greater numbers of appropriately corrected repopulating cells to allow for rapid repopulation and ensures predominance of gene-corrected cells *in vivo*. Accordingly, the present disclosure provides compositions and methods for the supportive care via a third-party, non HLA-matched, donor *ex vivo* expanded

stem/progenitor cell, which is capable of providing rapid but transient myeloid recovery, essential to reduce the risk of early transplant related mortality secondary to infections that is observed after myeloablative T cell depleted autologous transplants. Delaney et al., Nat Med 16:232-6 (2010).

[0102] Agents that inhibit differentiation (e.g., the Notch ligand) may be combined with compositions and methods that enhance the proliferation and survival of early stem/progenitor cells thereby achieving improved Notch-mediated *ex vivo* expansion. Enhanced proliferation of cord blood stem/progenitor cells may be achieved by combining the Notch ligand, Delta1, with the aryl hydrocarbon receptor inhibitor (SR1) (Boitano et al., Science 329:1345-8 (2011)) or HoxB4 (Watts et al., Blood 116:5859-66 (2010) and Zhang et al., PLoSMed 3:e173 (2006)) to enhance proliferation and self-renewal of hematopoietic precursors, and with angiopoietin-like 5 to enhance their survival. Essential to the clinical application of gene therapy is the ability to expand long-term repopulating cells, assuring longevity of the corrected cell graft.

[0103] Akt-activated endothelial cells may be employed in co-culture systems to confirm expansion of gene-corrected cells. Butler et al., Cell Stem Cell 6:251-64 (2011). Expansion of gene-corrected cells depends upon endothelial cell-induced activation of Notch signaling in the hematopoietic precursors. A second critical aspect for clinical application is the genetic and epigenetic fidelity of the derived cells as compared to their normal counterparts to ensure appropriate behavior and lack of oncogenic potential *in vivo*. Importantly, genome-wide assessment of expanded cord blood stem/progenitor cells exhibit fidelity of the transcriptome, chromatin structure, and the DNA methylome in comparison with primary isolated CD34+ cells.

[0104] Expansion strategies in normal CD34+ cells may be employed in conjunction with defined methods that utilize CD34+ cells from patients with hemoglobinopathies. Cord blood expansion methodology may utilize Delta1 in serum free media supplemented with hematopoietic growth factors using mobilized peripheral blood CD34+ obtained from normal donors. Optimized *ex vivo* expansion conditions using established *in vitro* assays (immunophenotyping, growth, etc) and *in vivo* repopulating ability may be assessed using the NSG mouse model. Optimized conditions may be used in combination with compositions that include SR1 (aryl hydrocarbon receptor inhibitor), Hox proteins, or angiopoietins to enhance the proliferation and survival of early stem/progenitor cells. Promising combinations may be evaluated in progenitor cell *in vitro* assays and in the immunodeficient mouse model (NSG mice) and then extended from expansion of CD34+ from normal individuals to evaluate these methods for expansion of CD34+ cells from patients with thalassemia (and other hemoglobinopathies).

[0105] The transcriptional, genetic, and epigenetic fidelity of expanded cells with their normal counterpart HSCs may be assessed using genome wide approaches to assess the oncogenic potential of the generated cells. Following growth *in vivo* (after infusion), cells may be used to determine whether there are functionally significant aberrations that enhance *in vivo* growth of any affected clone(s), thereby allowing selective expansion and detection of rare cells.

Cellular Therapies to Abrogate Post-transplant Neutropenia and to Improve Outcome Following Transplantation of Gene-corrected Autologous HSCs

[0106] Within another embodiment, the present disclosure provides compositions and methods for providing supportive care, which compositions and methods comprise off-the-shelf cellular therapies that abrogate post-transplant neutropenia and improve outcome following transplantation of gene-corrected autologous HSCs. *Ex vivo* expanded, cryopreserved cord blood (CB) stem/progenitor cells may, for example, be administered as a means of supportive care to patients with thalassemia and/or sickle cell disease who are undergoing myeloablative HCT with autologous CD34+ gene corrected cells.

[0107] In studies aimed at developing an economically feasible "off-the-shelf" source of progenitor cells capable of providing rapid neutrophil recovery, a bank of pre-expanded, cryopreserved hematopoietic stem/progenitor cell products was generated -- each being derived from a single CB unit that can be held for future clinical use.

[0108] The safety of administering this "off the-shelf" non-HLA matched product to adults was demonstrated immediately following first salvage chemotherapy for relapsed/refractory AML, as well as in the myeloablative CBT setting in pediatric and adult patients with hematologic malignancy.

[0109] It has been hypothesized that this expanded cell product, which is devoid of T cells, can be infused as an off-the-shelf cellular therapy to provide rapid but temporary myeloid engraftment and to potentially facilitate autologous hematopoietic recovery in patients undergoing myeloablative HCT with autologous gene-corrected stem cell grafts, thereby reducing the infectious complications and risk of mortality.

[0110] Critical is the question of whether HLA-matching is required for safe infusion of an "off the-shelf" non-HLA matched product, which is devoid of T cells. Without the need for HLA matching, fresh CB units can be collected for immediate *ex vivo* expansion and the final product cryopreserved for future on demand use. Patient access to an off-the-shelf expanded CB product is dramatically enhanced as all of the expanded products banked would be potentially available for any given patient, regardless of HLA typing, race/ethnicity or location of the patient.

[0111] Moreover, the ability to create an off-the-shelf universal donor expanded cell therapy is not only promising to shorten the duration of severe neutropenia post HCT, it is also likely to enhance more broad areas of investigation

outside of stem cell transplantation, e.g., as a way of providing temporary myeloid engraftment for treatment of chemotherapy induced severe neutropenia, any acquired severe neutropenia or accidental radiation exposure.

[0112] *Ex vivo* expansion abrogates the risks of CBT by overcoming delayed hematopoietic recovery and a significant improvement in overall survival will result. A reduced risk of relapse has been observed in patients undergoing double CBT, and chronic GVHD is lower despite highly mismatched grafts. If the risk of early transplant related mortality can be reduced by infusion of *ex vivo* expanded cord blood progenitors to enhance hematopoietic recovery, overall survival is likely to exceed that seen with conventional unrelated donors.

[0113] Within further embodiments, the present disclosure provides cellular therapies to abrogate post-transplant neutropenia and to improve outcome following transplantation of gene-corrected autologous HSCs. Patients with thalassemia who undergo myeloablative HCT with autologous gene corrected cells are at increased risk of infections and mortality secondary to limiting numbers of CD34+ cells in the infused graft (until *ex vivo* expansion of these gene corrected cells to clinically feasible numbers are achieved). Infusion of a previously expanded and cryopreserved cord blood progenitor cell product as an off-the-shelf supportive care measure can be employed to reduce the risk of mortality by contributing to early, but transient, myeloid recovery until the long term graft contributes to hematopoietic recovery.

[0114] Patients who undergo myeloablative HCT experience severe pancytopenia as a direct consequence of the conditioning regimen, and all patients are at increased risk of infection and bleeding during this time. The time to hematopoietic recovery (of neutrophil and platelets) is directly influenced by the CD34+ cell dose, and thus, for those patients undergoing myeloablative HCT with umbilical cord blood where the stem cell dose is 1/10th of a conventional donor graft or with autologous CD34 enriched low cell dose grafts, the risk of transplant related mortality due to delayed hematopoietic recovery is even greater.

[0115] To overcome these risks and to increase the safety of these HCT approaches, there is a great need for novel therapies that can abrogate prolonged pancytopenia and facilitate more rapid hematopoietic recovery. As discussed above, such a strategy has been developed wherein the absolute number of marrow repopulating cord blood (CB) hematopoietic stem/progenitor cells (HSPC) can be increased by culture with the Notch ligand Delta1. Infusion of these partially HLA-matched *ex vivo* expanded CB cells into children or adults undergoing cord blood transplantation (CBT) has been demonstrated to be safe and can significantly shorten the time to reach an initial absolute neutrophil count of 500 from 26 to 11 days, as a result of rapid myeloid engraftment contributed by the expanded cells.

[0116] In more recent studies aimed at developing an economically feasible "off the-shelf" source of progenitor cells capable of providing rapid neutrophil recovery, we have generated a bank of pre-expanded cryopreserved hematopoietic stem/progenitor cell products, each derived from a single CB unit that can be held for future clinical use. We have now also demonstrated the safety of administering this "off-the-shelf" non-HLA matched product to adults immediately following first salvage chemotherapy for relapsed/refractory AML, as well as in the myeloablative CBT setting in pediatric and adult patients with hematologic malignancy. We hypothesize that this expanded cell product which is devoid of T cells can be infused as an off-the-shelf cellular therapy to provide rapid but temporary myeloid engraftment and to potentially facilitate autologous hematopoietic recovery in patients undergoing myeloablative HCT with autologous gene-corrected stem cell grafts, thereby reducing the infectious complications and risk of mortality.

[0117] Using the defined optimal methods for generation of *ex vivo* expanded cord blood stem/progenitor cells, a bank of off-the-shelf expanded cell products may be employed to determine the safety of infusing these cells as supportive care in an autologous gene-corrected HCT.

[0118] The present disclosure will be best understood in view of the following non-limiting Examples.

EXAMPLES

Example 1

Selection of Bcl11a Gene Targeting Homing Endonucleases Based on I-HjeMI, I-CpaMI, and I-OnuI Using In vitro Compartmentalization (IVC)

[0119] The open reading frame (ORF) of a parental LAGLIDADG homing endonuclease (LHE), I-HjeMI (Fig. 14; SEQ ID NO: 28; Jacoby et al., Nucl. Acids Res. 40(11):4954-4964 (2012); Taylor et al., Nucl. Acids Res. 40(Web Server issue):W110-6 (2012)), codon optimized for expression in *E. coli*, was cloned between the NcoI and NotI restriction sites of pET21-a(+) (Fig. 11; EMD Millipore (Novagen) division of Merck KGaA. To introduce site-directed saturation mutagenesis into the ORF of I-HjeMI, DNA fragments containing its partial ORF with approximately 20 base pairs of a region overlapped with flanking fragments on both sides were PCR-amplified using primers that contained degenerate codon 5'-NNK-3'(coding all 20 amino acids). Amino-acid residues mutated using such PCR primers are shown in Table 2. PCR products were purified by extraction from an agarose gel, and assembled in a subsequent round of PCR with a sequence containing 2 copies of target sites for variant endonucleases to be selected. Successfully assembled DNA fragment was again purified by gel extraction, and used as a library in *in vitro* compartmentalization (IVC).

[0120] Three rounds of IVC were conducted after each round of site-directed saturation mutagenesis in order to enrich variant nuclease genes with altered specificity. The oil-surfactant mixture (2% ABIL EM 90 (Evonik Industries AG Personal Care, Essen, North Rhine-Westphalia, Germany), 0.05% Triton X-100 in light mineral oil) was thoroughly mixed with the saturation buffer (100 mM potassium glutamate (pH 7.5), 10 mM magnesium acetate (pH 7.5), 1 mM dithiothreitol and 5 mg/ml bovine serum albumin), incubated at 37°C for 20 minutes, and centrifuged at 16,000 x g for 15 minutes at 4°C. Five hundred microliters of the upper phase was used to emulsify 30 µl of the *in vitro* protein synthesis mixture (25 µl of PURExpress (New England Biolabs, Ipswich, MA), 20 units of RNase inhibitor, 1 mg/ml bovine serum albumin, and 8 ng of a DNA library) by constant stirring at 1,400 r.p.m. for three and a half minutes on ice. The emulsion was incubated at 30°C for 4 hours, and then heated at 75°C for 15 minutes. Emulsified droplets were collected by centrifugation at 16,000 x g for 15 min at 4°C, and broken by an addition of phenol/chloroform/isoamyl alcohol. Nucleic acids were recovered by ethanol precipitation, and treated with RNase cocktail (Life Technologies Corporation (Invitrogen), Grand Island, New York). After purification using QIAquick PCR purification kit (Qiagen, Hilden, Germany), a DNA library was ligated with a DNA adaptor with a 4-base 3' overhang sequence complementary to the cohesive end of a target site generated by endonuclease variants expressed in emulsified droplets, and added to PCR mixture containing a pair of primers, one of which was specific for the ligated DNA adaptor in order to enrich genes of variant endonucleases linked to a cleaved target site. A PCR amplicon was gel-purified and the ORF of variant genes was further PCR-amplified to prepare a DNA library to be used in the subsequent round of IVC.

[0121] In the second round of IVC, an emulsion was made with 1 ng of a reconstructed library, and incubated at 42°C for 75 minutes before quenching *in vitro* transcription/translation reaction by heating at 75°C. The DNA library was recovered and active endonuclease genes were specifically enriched by PCR following ligation with a DNA adaptor as described above.

[0122] In the third round of IVC, an *in vitro* protein synthesis mixture containing 0.5 ng of a library fragment was emulsified in 4.5% Span 80/0.5 %Triton X-100/light mineral oil. The reaction ran at 42°C for 45 minutes and was heat-inactivated at 75°C. After extraction from emulsion, cleaved target site-associated endonuclease genes were PCR-amplified and subjected to the subsequent round of site-directed mutagenesis.

[0123] To redesign I-HjeMI variants that recognized the (-) and (+) half sites of the BCL11A gene target, 4 and 2 rounds of site-directed saturation mutagenesis were carried out, respectively (Fig. 13). A pool of variant nucleases targeting the former half-site was subjected to an additional (fifth) round of mutagenesis on the surface opposite to the protein-DNA interface, followed by 3 rounds of IVC (Table 2).

Table 2

<i>Amino-acid Positions Subjected to Saturation Mutagenesis in IVC</i>			
Round	Target Site*	Sequence Identifier	Amino Acid Residues
1	TTGAGGAGATGTCTCTGTTAAT	SEQ ID NO: 55	R61, R63, N64, E65, 166, M68, S70
2	TTGAGGTGATGTCTCTGTTAAT	SEQ ID NO: 56	Y20, S22, E33, G35, E37, S59, R61, R70, R72
3	TTGAAGTGATGTCTCTGTTAAT	SEQ ID NO: 57	Y20, S22, T24, T31, E33, G35, R72, R74
4	TCCAAGTGATGTCTCTGTTAAT	SEQ ID NO: 58	Y20, S22, T24, K26, G27, K28, T31, E33
5	TCCAAGTGATGTCTCTGTTAAT	SEQ ID NO: 59	S109, N110, A121, S123, N124, N135, S137
1	TTGAGGAGGTTTCTGTGTTAAT	SEQ ID NO: 60	S154, S168, D170, 1193, L195, R202, K204
2	TTGAGGAGGTTTCTCGGTGGTG	SEQ ID NO: 61	S154, L158, N159, D162, D163, 1166, 1168, K204, T206

* Underlined nucleotides differ from those in the target site for the parental LHE I-HjeMI.

DNA fragments that encoded the N-terminal and the C-terminal half-domains of I-HjeMI variant endonucleases responsible for the (-) and (+) half sites of the BCL11A gene target were assembled, and a pool of nucleases that cleaved the full BCL11A gene target site were selected through 3 rounds of IVC (Fig. 13).

[0124] Table 3 presents exemplary BCL11a target sequences, which comprise DNA sequence modules for which

pools of highly active endonuclease variants (based upon homing endonucleases I-CpaMI and I-OnuI) have been isolated and their sequences determined.

Table 3

Base Homing Endonuclease I-CpaMI		
f 683/2508 chr2:60,542,847-60,542,868 Disrupts Bcl11a-X and XL forms	ATGGGATTCATATTGCAGACAA	SEQ ID NO: 25
Base Homing Endonuclease I-OnuI		
r 1588/2508 chr2:60,542,630-60,542,651 Disrupts Bcl11a-X and XL forms	AGCCATTGGATTCAACCGCAGC	SEQ ID NO: 26
f 525/2508 chr2:60,543,005-60,543,026 Disrupts all Bcl11a forms	caaaCAGccATTCAcCagTgcA	SEQ ID NO: 27

Example 2

Optimization of Activity of BCL11A Gene-Targeting I-HjeMI Variants Using Two-plasmid Gene Elimination Cleavage Assay in Bacteria

[0125] The activity of I-HjeMI variants obtained in selection using IVC display selections (disclosed in Example 1, above) was optimized using a two-plasmid selection system in bacterial cells according to the methodology of Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006). The ORF of the endonuclease genes was inserted between NcoI and NotI sites of the pENDO (Fig. 12, Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006)) expression plasmid. NovaXGF (EMD Millipore (Novagen)) competent cells harboring the pCcdB reporter plasmid (Fig. 31, Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006); Takeuchi et al., Nucl. Acids Res. 37(3):877-890 (2009); and Takeuchi et al., Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1107719108 (2011)) containing 4 copies of the BCL11A gene target were transformed with a pool of the pEndo plasmid encoding I-HjeMI variants. The transformants were grown in 2 x YT medium (16 g/L tryptone, 10 g/L yeast extract, and 5 g/L NaCl) at 37 °C for 30 min and then diluted 10-fold with 2 x YT medium supplemented with 100 µg/mL carbenicillin and 0.02 % L-arabinose (in order to preinduce expression of I-HjeMI variants). After the culture was grown at 30°C for 15 hours, the cells were harvested, resuspended in sterile water and spread on both nonselective (1 x M9 salt, 1 % glycerol, 0.8 % tryptone, 1 mM MgSO₄, 1 mM CaCl₂, 2 µg/mL thiamine, and 100 µg/mL carbenicillin) and selective plates (i.e. nonselective plates supplemented with 0.02% L-arabinose and 0.4 mM IPTG to induce expression of the toxic CcdB protein). After incubation at 30°C for 30-40 hours, the pEndo plasmid was extracted from the surviving colonies on the selective plates.

[0126] The ORFs encoding active I-HjeMI variants were amplified via error-prone PCR using the Gene Morph II Random Mutagenesis Kit (Agilent Technologies, Santa Clara, CA). After digestion with NcoI, NotI, and DpnI, the resulting fragments were recloned into the pEndo vector. The plasmid was subjected to 2 rounds of selection under the conditions where variant endonucleases were expressed at 30°C for 4 hours before plating. The N-terminal half and C-terminal half domains of the selected genes were shuffled using overlapping PCR, and again cloned into the pEndo vector. Transformed cells carrying both the pEndo plasmid and the pCcdB reporter were grown in 2 x YT medium containing 0.02 % L-arabinose at 37°C for an hour and then spread on selective plates at 37°C for 16-20 hours. After 2 rounds of selection at the same level of stringency, the pEndo plasmid was extracted from surviving colonies on the selective plates, and ORFs of the variant genes carried on the plasmid were sequenced.

Example 3

Activity of BCL11A Gene-targeting Endonucleases Tested in a Two-plasmid Cleavage Assay

[0127] Activity of an exemplary BCL11A gene-targeting endonuclease (BCL11Ahje; Fig. 17, SEQ ID NO: 31), its catalytically inactive variant (BCL11Ahje D18N), and its parental LHE I-HjeMI (Fig. 14, SEQ ID NO: 28) was measured in bacterial cells that harbor the pCcdB reporter plasmid (Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006)) containing 4 copies of either the target site for I-HjeMI (I-HjeMI target) or the BCL11A gene target (TCCAAGTGATGTCTCGGTGGTG (SEQ ID NO: 39; underlined nucleotides differ from those in the target site for the parental LHE I-HjeMI). The pCcdB reporter plasmid encodes "control of cell death B" ("ccdB", a toxic protein in bacteria, which is inducible by an addition of IPTG). Cleavage of the target sites in the reporter plasmid leads to RecBCD-mediated degradation of the reporter plasmid and corresponding cell survival on the selective medium containing IPTG. The

survival rate was determined by dividing the number of colonies on the selective plates by that on the nonselective plates. Error bars refer $\pm$ S.D. of 3 independent experiments.

Example 4

Detection of Targeted Mutagenesis at the Endogenous Human BCL11A Gene

[0128] HEK 293T cells (1.6×10^5) were seeded 24 hours prior to transfection in 12-well plates, and transfected with 0.5 μ g each of expression plasmids for the BCL11A gene targeting nuclease and TREX2. At 48 hours post transfection, transfected cells were lysed and genomic DNA was extracted using Quick-gDNA MiniPrep kit (Zymo Research). Approximately 500-bp fragment spanning the BCL11A gene target was PCR-amplified from 50 ng of the extracted genomes using a pair of the following primers: Bcl11A_up1, 5'- GCT GGA ATG GTT GCA GTA AC -3' (SEQ ID NO: 66); Bcl11A_down1, 5'- CAA ACA GCC ATT CAC CAG TG -3' (SEQ ID NO: 67). The PCR amplicon was incubated in 1 x NEB buffer 4 plus 1 x BSA (New England Biolabs) with or without 0.5 μ M of the BCL11A gene targeting nuclease that was purified from *E. coli* overexpressing the recombinant protein at 37 °C for 2 hours. The reaction was terminated by adding 5 x Stop solution (50 mM Tris-HCl (pH7.5), 5 mM EDTA, 0.5 % SDS, 25 % glycerol, 0.1 orange G and 0.5 mg/mL proteinase K). After incubation at room temperature for 15 minutes, a half of each sample was separated on a 1.6 % agarose gel containing ethidium bromide in TAE (upper panels). The rest of each sample was purified using DNA Clean & Concentrator-5 kit (Zymo Research), and used as a template in the second round of PCR with a pair of the following primers: Bcl11A_up2, 5'- CTG CCA GCT CTC TAA GTC TCC -3' (SEQ ID NO: 68); Bcl11A_down2, 5'- TGC AAC ACG CAC AGA ACA CTC -3' (SEQ ID NO: 69). The PCR product was again digested with the BCL11A gene targeting nuclease under the conditions described above, and analyzed on a 1.6 % agarose gel (lower panels) (See Figure 30).

Example 5

Selection of Fetal Hemoglobin Silencing Region Targeting Endonucleases based on I-Onul using In vitro Compartmentalization

[0129] Exemplary homing endonuclease (HE) target sequences, which are evenly distributed throughout the 350 bp region (SEQ ID NO: 2) that includes the region of Bcl11a occupancy within the HbF silencing region in adult erythroid cells that is disrupted in the French HPFH deletion, are presented in Table 4. These target sequences comprise DNA sequence modules for which pools of highly active endonuclease variants have been isolated and sequenced.

Table 4

Position	Chromosomal Location	Sequence	Sequence Identifier
Wild Type	N/A	TTTCCACTTATTCAACCTTTTA	SEQ ID NO:5
f 13/303	chr11:5,214,235-5,214,256	TGTGGCCCTATTCTTGTTCA	SEQ ID NO:6
f 79/303	chr11:5,214,169-5,214,190	CATTGTCACTTTCTCCCTACT	SEQ ID NO:7
f 143/303	chr11:5,214,105-5,214,126	TAAATACATTTCTTCACTAAG	SEQ ID NO:8
f 124/303	chr11:5,214,089-5,214,110	ACTAAGTGAGAATAATCTTTTA	SEQ ID NO:9
f 200/303	chr11:5,214,048-5,214,069	GCCACCACCTTTCTTGAATTAT	SEQ ID NO:10
f 211/303	chr11:5,214,037-5,214,058	TCTTGAATTATTCAATATCTTT	SEQ ID NO:11
f 274/303	chr11:5,213,974-5,213,995	TTAAAGGTCATTCATGGCTCCT	SEQ ID NO:12

[0130] Table 5 presents a region from -100 bp to 210 bp upstream of globin genes, which is identical for both α - and γ -globin genes and which contains many of the non-deletion HPFH mutations. Gene editing resulting in these mutations leads to decreased repression, thus activation, of a gamma gene and results in increased HbF.

EP 3 272 356 A1

Table 5

	Sequence	Sequence Identifier
5 Wild-type	TGGGGGCCCCCTTCCCCACACTATCTCAATGCAAATATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTGACCAATAGCCTTGACAA	SEQ ID NO:16
10 G-gamma - 202 C→G	TGGGGGCGCCTTCCCCACACTATCTCAATGCAAATATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTGACCAATAGCCTTGACAA	SEQ ID NO:17
15 G-gamma - 175 T→C	TGGGGGCCCCCTTCCCCACACTATCTCAATGCAAACATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTGACCAATAGCCTTGACAA	SEQ ID NO:18
G-gamma - 114 C→T	TGGGGGCCCCCTTCCCCACACTATCTCAATGCAAATATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTGACTAATAGCCTTGACAA	SEQ ID NO:19
20 A-gamma - 196 C→T	TGGGGGCCCCCTTCTCCACACTATCTCAATGCAAATATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTGACCAATAGCCTTGACAA	SEQ ID NO:20
25 A-gamma - 175 T→C	TGGGGGCCCCCTTCCCCACACTATCTCAATGCAAACATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTGACCAATAGCCTTGACAA	SEQ ID NO:21
A-gamma - 117 G→A	TGGGGGCCCCCTTCCCCACACTATCTCAATGCAAATATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTAACCAATAGCCTTGACAA	SEQ ID NO:22
30 A-gamma - 114→-102 deleted (deleted bases in bold)	TGGGGGCCCCCTTCCCCACACTATCTCAATGCAAATATCTG TCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTGGC CAGCCTTGCCTTGACCAATAGCCTTGACAA	SEQ ID NO:23

[0131] Table 6 presents amino acid positions within a parental LHE I-Onul homing endonuclease (Figure 22A, SEQ ID NO: 34) that were subjected to saturation mutagenesis in IVC (as described in Example 1, above) to create homing endonucleases that are targeted against a human fetal globin silencing region.

Table 6

<i>Amino-acid Positions Subjected to Saturation Mutagenesis in IVC to Create Targeted Homing Endonucleases against a Human Fetal Globin Silencing Region</i>			
Round	Target site*	Sequence Identifier	Amino-acid residues
45 1	TTTCCA <u>ATT</u> TATTCAACCTTTTA	SEQ ID NO: 40	L26, G44, Q46, A70, S72, S78, K80
2	TCT <u>GAA</u> TATTCAACCTTTTA	SEQ ID NO: 41	L26, R28, R30, N32, S40, E42, G44, K80, T82
50 1	TTTCCATTTATTCAAT <u>AT</u> TTTA	SEQ ID NO: 42	F182, N184, V199, S201, K225, K227, D236, V238
2	TTTCCATTTATTCAATAT <u>CTT</u>	SEQ ID NO: 43	F182, N184, I186, S190, K191, Q197, V199, V238, T240
*Underlined nucleotides differ from those in the target site for the parental LHE I-Onul.			

[0132] Figure 23 presents the results of a cleavage assay with 'half-targets' from a human fetal globin silencing region. The amplified bands contain both the cleaved half-sites (captured by ligation with complementary duplex oligonucleotides and corresponding overhanging ssDNA) and the sequences of the enzyme variants that are responsible for generation of cleaved DNA products. The final step upon completion of enrichment of the 'half-site' endonuclease libraries includes the assembly of DNA fragments that encode the N-terminal and C-terminal half domains of I-Onu1 homing endonuclease, which are responsible for the left (L) and right (R) half sites of the gGlobin silencing region target. Active I-Onu1 endonucleases are selected from a pool that cleaves the full-length human fetal globin silencing region target.

Example 6

MegaTALs Homing Endonucleases with N-terminal Fusions of TAL Anchors to Increase Specificity and Activity of a Gene-targeted Endonuclease

[0133] N-terminal fusions of TAL anchors can be employed to increase the specificity and activity of a gene-targeted endonuclease, including one or more homing endonucleases such as one or more of the I-HjeMI, I-CpaMI, and I-Onu1 homing endonucleases. MegaTALs are constructed using the Golden Gate assembly strategy described by Cermak et al., Nucl. Acids Res. 39:e82-e82 (2011), using an RVD plasmid library and destination vector (see, Fig. 24, SEQ ID NO: 35 and Fig. 25, SEQ ID NO: 36 for the nucleotide and amino acid sequences of MegaTAL:5.5 RVD + Y21-Anil).

[0134] Plasmids are modified to allow the assembly of TAL effector containing 1.5 to 10.5 TAL repeats and their corresponding RVDs ('Repeat Variable Diresidues,' which contact consecutive DNA bases in the 5' region of the target site and thereby define the cognate DNA sequence in that region). The pthX01 destination vector was modified to include a hemagglutinin (HA) tag immediately downstream of the NLS and to yield a TALEN scaffold that begins at residue 154 (relative to wild-type PthXo1 TAL effector) and ends 63 residues beyond the final 'half TAL repeat' sequence.

[0135] TAL effectors are built using the following RVDs to target each specific nucleotide: A - NI, C - HD, G - NN and T - NG. Following cloning of the TAL effector repeats into the destination vector, an individual protein linker ('Zn4'; VGGS) and the engineered homing endonuclease variants are cloned in place of the FokI nuclease catalytic domain, between engineered Xba-I and Sal-I restriction sites.

[0136] This is a 'model test case' MegaTAL (a fusion of a TAL effector at the N-terminal end of a single protein chain, fused via a flexible linker to the wild-type Y2 I-Anil homing endonuclease, which was originally described in Takeuchi et al., Nucl. Acids Res. 37(3): 877-890 (2009).

Example 7

A Cas9-based Endonuclease System for Disrupting a Bcl11a-regulated Fetal Hemoglobin (HbF) Silencing Region

[0137] The recent mechanistic understanding of the clustered regularly interspaced short palindromic repeat (CRISPR) system that bacteria use for adaptive immunity has led to the development of a powerful tool that allows for genome editing of mammalian cells, which can be employed in the compositions and methods for the treatment of hemoglobinopathies that are disclosed herein: (a) to disrupt a Bcl11a coding region; (b) to disrupt a HbF silencing DNA regulatory element or pathway, such as a Bcl11a-regulated HbF silencing region; (c) to mutate one or more γ -globin gene promoter(s) to achieve increased expression of a γ -globin gene; (d) to mutate one or more δ -globin gene promoter(s) to achieve increased expression of a δ -globin gene; and/or (e) to correct one or more β -globin gene mutation(s). The bacterial CRISPR system is described in Jinek et al., Science 337:816-821 (2013); Cong et al., Science (Jan. 3, 2013) (Epub ahead of print); and Mali et al., Science (Jan. 3, 2013) (Epub ahead of print).

[0138] The Cas9 protein generates a double stranded break at a specific site the location of which is determined by an RNA-guide sequence. All guide RNAs contain the same scaffold sequence that binds Cas9, as well as a variable targeting sequence having the structure G-N₂₀-GG, which provides Cas9-RNA complex cleavage specificity. Co-expression of the Cas9 protein and a guide RNA results in the efficient cleavage and disruption at a sequence-specific location within the human genome, which sequence-specific cleavage is defined by the guide RNA sequence. Co-expression of Cas9 and guide RNAs that are specific to multiple targets leads to efficient deletion of the intervening region between target sites.

[0139] Thus, within certain aspects of the present disclosure, Cas9-mediated genome editing is employed: (1) to disrupt the Bcl11a binding site within the HbF silencing region, (2) to disrupt Bcl11a gene function, and (3) to delete the entire HbF silencing region. Target regions are identified and guide RNAs are designed and generated based on consideration of the optimal guide RNA target sequences. Exemplified herein are guide RNAs that target the Bcl11a binding region within the HbF silencing region as well as the single GATA-1 binding motif. These guide RNAs are used singly or in combination to achieve the targeted disruption of the HbF silencing region. Cas9 with guide RNAs to additional regions within the HbF silencing regions that correspond to Bcl11a peaks of occupancy are also used singly and in

combination. Several pairs of guide RNAs flanking the Bcl11a binding site and GATA-1 motif, as well as the entire footprint, can also be co-expressed with Cas9 in order to generate deletions within the HbF silencing region.

[0140] The sequence of a human codon optimized Cas9 from Mali *et al.*, Science (Jan. 3, 2013) is presented in Fig. 26, SEQ ID NO: 37. The generic sequence of a guide RNA (Mali *et al.*) is presented in Fig. 27, SEQ ID NO: 38, the key sequence elements of which are presented in Table 7. Exemplary Cas9 Guide RNAs sequences of target-specific binding to and cleavage of the human fetal hemoglobin (HbF) silencing region (Fig. 6, SEQ ID NO: 1 and Fig. 7, SEQ ID NO: 2) are presented in Table 8.

Table 7

<i>Sequence Elements of a Generic Cas9 Guide RNA</i>		
Description	Sequence Identifier	Nucleotide Sequence
U6 Promoter Sequence	SEQ ID NO: 44	GGACGAAACACC
Generic Target-specific Sequence	SEQ ID NO: 45	GNNNNNNNNNNNNNNNNNNNN
Guide RNA Scaffold Sequence	SEQ ID NO: 46	GTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGT GGCACCGAGTCGGTGCT
Poly T Tail	SEQ ID NO: 47	NNNNTTTTTT

Table 8

<i>Target-specific Sequences for Exemplary Cas9 Guide RNAs</i>				
Designation	Description	Use	Sequence Identifier	Nucleotide Sequence
GGN20GG-B	Targets the GATA-1 recognition motif	Used singly, with #C to delete the putative Bcl11a and GATA-1 motifs jointly, or with #E, #F or #G to delete the entire Bcl11a ChIP peak and surrounding sequences	SEQ ID NO: 48	GCCATTTCTATTA TCAGACTTGG
GGN20GG-C	Targets the putative Bcl11a recognition motif	Used singly, with #B to delete the putative Bcl11a and GATA-1 motifs jointly	SEQ ID NO: 49	GCTGGGCTTCTGT TGCAGTAGGG
GGN20GG-D	Targets immediately adjacent to the putative Bcl11a recognition motif	Used singly, with #B to delete the putative Bcl11a and GATA-1 motifs jointly	SEQ ID NO: 50	GAAAAATGGGAGAC AAATAGCTGG
GGN20GG-E	Targets downstream of the Bcl11a binding peak	Used with #B and/or #H to delete the entire Bcl11a ChIP peak and surrounding sequences	SEQ ID NO: 51	GAATAATTCAAGA AAGGTGGTGG
GGN20GG-F	Targets downstream of the Bcl11a binding peak	Used with #B and/or #H to delete the entire Bcl11a ChIP peak and surrounding sequences	SEQ ID NO: 52	GATATTGAATAAT TCAAGAAAGG
GGN20GG-G	Targets downstream of the Bcl11a binding peak	Used with #B and/or #H to delete the entire Bcl11a ChIP peak and surrounding sequences	SEQ ID NO: 53	GCCTGAGATTCTG ATCACAAGGG
GGN20GG-H	Targets upstream of the Bcl11a binding peak	Used with #E, #F or #G to delete the entire Bcl11a ChIP peak and surrounding sequences	SEQ ID NO: 54	GGTAAATTCTTAA GGCCATGAGG

Example 8Vector Systems for Expressing Endonucleases

[0141] For NSG, sickle cell and thalassemia mouse models, human CD34 cells or mouse bone marrow nucleated cells are transduced along with a fluorescent marker allowing flow cytometry-based enrichment of cells prior to transplantation. Suitable transduction methods include the following:

[0142] AAV6 vectors. AAV6-serotype recombinant AAV vectors provide a 4.5 kb payload, sufficient to deliver a promoter-HE-exonuclease or promoter-TAL-HE fusion-exonuclease cassettes in addition to a small recombination template. Alternatively they can carry Cas9 and a guide RNA. In addition, we have preliminary data that show AAV6 provides the most efficient transduction of human CD34+ umbilical cord blood cells of all known AAV capsids, and is able to mediate significant levels of transient gene expression in HSC.

[0143] Modified Adenovirus vectors. Adenoviral vectors with hybrid capsids are capable of efficiently transducing many types of hematopoietic cells including CD34+ cells. Improved transduction with the chimeric vector using the serotype 35 fiber (Ad5-F35) was demonstrated by Dr. Rawlings (SCH) and more recent data suggest that the serotype 11 fiber (Ad5-F1 1) may be even more efficient in hematopoietic cells. Helper-dependent adenoviral vectors offer up to a 30 kb payload, along with transient gene expression in HSC, and can be used to deliver multiple HE/exonuclease cassettes, HE-TAL fusions, as well as very large recombination templates or a Cas9 expression cassette and multiple guide RNAs.

[0144] Integration-deficient Lentiviral and Foamyviral Vectors (IDLV and IDFV). These vectors provide 6 kb (IDLV) to 9kb (IDFV) payloads, and have well documented capabilities to transduce human HSCs. Both IDLV and IDFV vectors can be used for gene knockout and recombination-based gene editing in HSC. Drs. Rawlings (SCH) and Kiem (FHCRC) have generated and evaluated a series of IDLV with alternative promoter GFP cassettes and have determined constructs that provide efficient and high level expression in CD34+ HSC.

[0145] Direct nucleofection of plasmid and mRNA. Conditions for efficient transduction of N-MEL and CD34 cells have been defined using the Amaxa nucleofection system. Benefits include the lack of integration, and the ability to transduce multiple expression plasmids or RNA species simultaneously.

[0146] In parallel, sorted and un-sorted cells will be transplanted into separate mice. While the later transplants may contain low numbers of modified cells, human studies of post-transplant chimeras suggest that these cells will have a selective survival advantage and be enriched in the periphery. Regardless, single reticulocyte RNA and F-Cell analysis will allow assessment of gene disruption in cells even if present in low abundance.

[0147] Second level assessments will focus on the pluripotency of transduced CD34+ cells and erythropoiesis. Assays will include culturing to assess long-term proliferative potential, analysis of myeloid and erythroid colonies for clonal analysis and transplant into NOD *scid* gamma (NSG) mice followed by assessment of multi-lineage engraftment of primary and secondary recipients.

[0148] Typically stem cells are infused via tail vein injection after total body irradiation (275 rads for NSG mice or 1000 rads for C57 mice). Though efficient, this is effective, for most studies we will inject stem cells directly into the mouse femur as 50 fold fewer cells are required, ideal or assaying a potentially limited number of flow cytometry sorted and/or modified cells. After anesthesia and local analgesia are provided and anatomic landmarks defined, 0.5-1 million cells are directly injected into the femurs marrow space.

Example 9Characterization of Homing and Cas9 Endonucleases for Efficient Gene Targeting

[0149] For clinical impact efficient gene targeting is demonstrated by assessing levels of globin gene expression in individual targeted cells and in populations of cells, the effect of targeting on erythropoiesis and on stem cell function, and impact hematologic parameters and organ function in model organisms.

[0150] Transductions are followed by single cell and bulk population assessments of gene targeting efficiency and expression of all β -like genes at the RNA and protein levels. Alterations in factor binding and chromatin structure are assessed, as cell morphology and the extent of ineffective erythropoiesis and apoptosis. Candidate endonucleases that score well in initial screens are further assessed for effects on HSC pluripotency as well as the ability to ameliorate disease specific phenotypes *in vitro* and *in vivo*.

[0151] Initial screening of endonuclease candidates and delivery systems is performed in a mouse erythroleukemia cell line containing a single intact human chromosome 11 (N-MEL) and clinical grade CD34+ normal human HSCs with endpoints of assessing targeted mutation efficiency and globin gene expression. Both cell types can be induced to differentiate along an erythroid path during which expression of β -like genes is highly induced with a low γ -globin / γ - + β -globin RNA ratio. Using a HbF specific antibody, the percent of "F-cells" can be quantified. These low ratios are ideal as the systems are sensitive for detecting and quantifying even small increases in γ -globin mRNA as well as HbF

expression at the single cell and population level.

[0152] N-MEL cells are a derivative of murine erythroleukemia cells that contain a single intact human chromosome 11 that contains the β -globin locus. This erythroid cell line normally expresses low levels of mouse and human β -like globin genes, but can be induced to differentiate at which time globin expression is greatly increased.

[0153] N-MEL cells are efficiently transduced using the Amaxa nucleofection system. Using 2 μ g of plasmid DNA, Kit "L" and program A20 25% of cells are transduced. Infection at multiplicity of infection (MOI) of 20 with a RSCS-MCS-PG-WZ based lentiviral vector containing a homing endonuclease designed to disrupt the Bcl11a gene yields approximately 40% of N-MEL cells being transduced.

[0154] The efficiency of targeted disruption is assayed using the Cel-1 assay. The target region is amplified by PCR, heat-denatured, re-annealed and exposed to the enzyme Cel-1 that efficiently cleaves bubbles from mismatched regions as small as one base pair. If the targeted region has been mutated in any way, heteroduplexes of wild type and mutant strands are cleaved and detected by gel electrophoresis. This assay can be used on bulk populations of cells to estimate the efficiency of mutation, or on flow cytometry sorted individual cells in which case analysis of multiple cells provides an accurate assessment of mutation frequency.

[0155] Using routine quantitative Taqman RT-PCR (qRT-PCR) assays for RNA from a bulk population of cells, β -globin expression is induced 11-fold with differentiation and with a γ -globin / γ^- + β -globin ratio of 0.1%. After infection with a Bcl11a knockdown vector a 30-fold increase in this ratio can be obtained, thus demonstrating that this cell culture system provides an accurate readout of disruption of the Bcl11a mediated HbF silencing pathway.

[0156] Due to concerns that altering Bcl11a pathways may lead to a relative increase in γ -globin RNA but a decrease in globin gene expression, flow cytometry can be used to sort 1000 cell pools of cells that are lysed and the above qRT-PCR assays can be performed. Because RNA is directly compared from the same number of cells, any diminution in the amount of globin RNA is reflected by an increase in the Ct, providing a direct measure of both the level of β - and γ -expression, as well as the ratio of γ -globin / γ^- + β -globin.

[0157] To determine the percent of cells that show an altered γ -globin / γ^- + β -globin ratio after manipulation and to determine the range of change in expression observed, single cell assays are performed. Flow cytometry sorted individual cells are subjected to routine RT-PCR of γ^- and β -globin RNA simultaneously for several cycles and once adequate material is present to allow for accurate splitting of the sample, γ^- and β -globin are assessed by qRT-PCR as above.

[0158] Ultimately it is levels of HbF protein, not RNA that are therapeutic, thus single cell and bulk cell HbF assays are performed. Bulk populations of cells are assayed using HPLC after cell lysis and elution on a hemoglobin-dedicated column and the ratios of HbF to HbA and HbA2 are determined. The number of cells expressing HbF are compared pre- and post-transduction using glutaraldehyde fixing cells, permeabilizing with detergent and adding an HbF specific antibody followed by quantitation by flow cytometry.

[0159] Effects of mutations are assessed in modified cloned cells after isolation of single cells by flow cytometry. The above assays are performed. In addition, to show that targeted mutations disrupt binding of the Bcl11a repressive complex, cells are fixed in formaldehyde, chromatin isolated and sheared by sonication. Chromatin immune-precipitation (ChIP) is performed using commercially available antibodies to Bcl11a, GATA-1 and HDAC-1. Binding of these proteins to a target region in N-MEL cells, as well as erythroid-differentiated CD34 cells is described below. A lack of binding is assessed after targeted disruption.

Example 10

Clinical Glade CD34+ Hematopoietic Stem Cells

[0160] CD34+ cells from normal human donors are adhered to culture dishes with fibronectin peptide CH-296 and infected at an MOI of 20 twice, 8 hours apart, in media containing G-CSF, SCF, IL-3, IL-6, FLT-3, and TPO, with RSCS-MCS-PG-WZ based lentiviral vectors described above. This results in ~80% of cells being infected. Transduced cells are differentiated to erythroid cells using the protocol of Douy, Giarratana et al., Nat. Biotechnol. 23(1):69-74 (2005). The qRT-PCR assays above reveal a γ -globin / γ^- + β -globin ratio of 4%. This low ratio allows for the sensitive detection of increases that are secondary to genome editing.

[0161] To assess alterations in differentiation state, flow cytometry using multiple cell surface markers including CD34, CD71, and glycophorin are performed, as well as assessment of cell growth and morphology of cytopins after Wright-Giemsa staining. Disruption efficiency is assessed by Cel-1 assays on bulk populations and single cells as above. Additional assessment of globin gene expression and HbF is performed as above using qRT-PCR on 1000 cell pools and individual cells as well as HPLC and F-cell assays. ChIP is employed to assess binding of Bcl11a, GATA-1, and HDAC-1 to the target region.

[0162] To assess clinical effectiveness in human cells CD34+ human, HSCs from hemoglobinopathy patients are transduced and cultured using the same methods. In addition to the routine analysis for normal cells, additional disease-specific assessments are performed. Cultured thalassemic cells show minimal expansion, a lack of hemoglobinization,

evidence of ineffective erythropoiesis, and increased apoptosis compared to normal cells. This allows for quantitative assessments of improvements in expression and erythropoiesis post-targeting. Similarly, the degree of sickling of erythroid progeny of CD34 cells under hypoxic conditions is assessed. These transduced CD34+ cells from hemoglobinopathy patients are transplanted into NSG mice, after which several features of abnormal erythropoiesis are recapitulated, allowing assessment of the effect of targeted mutagenesis.

[0163] Clinical effectiveness is assessed *in vivo* in mouse models of hemoglobinopathies. Knockout of murine Bcl11a leads to a dramatic dose-dependent increase in γ -globin in mice containing a human β -globin locus on a transgene and ameliorates the sickle phenotype in humanized mouse models. While both systems allow the analysis of globin gene expression, the sickle mice allow for the assessment of the improvement of phenotype in these mice with special attention to the hematologic parameters especially the hematocrit, liver and lung pathology, renal function and spleen size. This can be correlated to the number of HbF containing cells, the HbF/S ratio and expression patterns in single cell assays similar to the above.

[0164] For RNA analysis total blood is used as it contains sufficient RNA-containing reticulocytes for analysis. For single cell analyses blood is stained with thiazole orange and RNA containing cells with the forward and side scatter profiles of red cells are collected. The erythrocyte lifespan is significantly reduced in the sickle cell mice and improvement in lifespan is assessed after intra-peritoneal injection of NHS-biotin with labels 100% of RBC. At time points, a microliter of blood is stained with strep-avidin FITC and the remaining percent of labeled RBC is assessed by flow cytometry. This is done with the mice from our BERK sickle cell mice. Pászty et al., Science 278(5339):876-878 (1997). In addition, the γ -globin / γ - + β -globin ratio is assessed in mice containing the A25 and A85 human β -globin YACs. Porcu et al., Blood 90(11):4602-4609 (1997).

[0165] Two thalassemic models are assessed, the Th-3 mouse heterozygotes (Yang et al., Proc. Natl. Acad. Sci. U.S.A. 92(25):11608-11612 (1995)) and mice heterozygous for a deletion of the LCR (Bender et al., Mol Cell. 5(2):387-393 (2000)). In each case, the focus of analysis for thalassemic mice will be hematocrit, assessment of ineffective erythropoiesis, spleen size, iron overload and organ morphology.

CLAUSES

[0166]

1. A composition for the treatment of a hemoglobinopathy, comprising one or more endonuclease(s) selected from the group consisting of a homing endonuclease (HE) and a CRISPR endonuclease, wherein each of said endonuclease(s) binds to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.

2. The composition of clause 1 wherein said endonuclease is a HE that binds to said Bcl11a coding region or said Bcl11a gene regulatory region.

3. The composition of clause 2 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

4. The composition of clause 3 wherein said HE is an I-HjeMI homing endonuclease.

5. The composition of clause 4 wherein said I-HjeMI homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 28 or SEQ ID NO: 29, wherein said amino acid substitution is selected from the group consisting of Y20, S22, T24, K26, G27, K28, T31, E33, G35, E37, S59, R61, R63, N64, E65, I66, M68, S70, R70, R72, R74, S109, N110, A121, S123, N124, N135, S137, S154, L158, N159, D162, D163, I166, I168, S168, D170, I193, L195, R202, K204, and T206.

6. The composition of clause 5 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32.

7. The composition of clause 6 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence of SEQ ID NO: 33 or a variant thereof wherein said I-HjeMI homing endonuclease can specifically bind to a Bcl11a coding region.

8. The composition of clause 1 wherein said endonuclease is a CRISPR endonuclease wherein said CRISPR comprises an RNA guide strand and a Cas9 endonuclease wherein said RNA guide strand mediates the binding of

said Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region.

9. The composition of clause 8 wherein said RNA guide strand comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.

10. The composition of clause 9 wherein said Cas9 endonuclease is encoded by the nucleotide sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 endonuclease.

11. The composition of clause 1 wherein said endonuclease is a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region.

12. The composition of clause 11 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

13. The composition of clause 12 wherein said HE is an I-Onul homing endonuclease.

14. The composition of clause 13 wherein said I-Onul homing endonuclease comprises the amino acid sequence encoded by a variant of the nucleotide sequence of SEQ ID NO: 34 that encodes an I-Onul homing endonuclease that can specifically bind to said fetal hemoglobin (HbF) silencing region.

15. The composition of clause 14 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82.

16. The composition of clause 15 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of F182, N184, I186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

17. The composition of clause 1, further comprising a TAL effector nuclease (TALEN), a TAL-HE, or a TREX2 protein.

18. A composition for the treatment of a hemoglobinopathy, comprising one or more polynucleotide encoding one or more endonuclease(s) selected from the group consisting of a homing endonuclease (HE) and a Cas9 endonuclease, wherein each of said endonuclease(s) binds to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.

19. The composition of clause 18 wherein said endonuclease is a HE that binds to said Bcl11a coding region or said Bcl11a gene regulatory region.

20. The composition of clause 19 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

21. The composition of clause 20 wherein said HE is an I-HjeMI homing endonuclease.

22. The composition of clause 21 wherein said I-HjeMI homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 28 or SEQ ID NO: 29, wherein said amino acid substitution is selected from the group consisting of Y20, S22, T24, K26, G27, K28, T31, E33, G35, E37, S59, R61, R63, N64, E65, I66, M68, S70, R70, R72, R74, S109, N110, A121, S123, N124, N135, S137, S154, L158, N159, D162, D163, I166, I168, S168, D170, I193, L195, R202, K204, and T206.

23. The composition of clause 22 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32.

24. The composition of clause 23 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence

of SEQ ID NO: 33 or a variant thereof wherein said I-HjeMI homing endonuclease can specifically bind to a Bcl11a coding region.

25. The composition of clause 24 wherein said endonuclease is a CRISPR endonuclease wherein said CRISPR comprises an RNA guide strand and a Cas9 endonuclease wherein said RNA guide strand mediates the binding of said Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region.

26. The composition of clause 25 wherein said RNA guide strand comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.

27. The composition of clause 26 wherein said Cas9 endonuclease is encoded by the nucleotide sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 endonuclease.

28. The composition of clause 18 wherein said endonuclease is a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region.

29. The composition of clause 28 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

30. The composition of clause 29 wherein said HE is an I-Onul homing endonuclease.

31. The composition of clause 30 wherein said I-Onul homing endonuclease comprises the amino acid sequence encoded by a variant of the nucleotide sequence of SEQ ID NO: 34 that encodes an I-Onul homing endonuclease that can specifically bind to said fetal hemoglobin (HbF) silencing region.

32. The composition of clause 31 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82.

33. The composition of clause 31 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of F182, N184, I186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

34. The composition of clause 18, further comprising a polynucleotide encoding a TAL effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

35. A polynucleotide encoding an endonuclease selected from the group consisting of a homing endonuclease (HE) and a CRISPR endonuclease, wherein said endonuclease binds to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.

36. The polynucleotide of clause 35 wherein said endonuclease is a HE that binds to said Bcl11a coding region or said Bcl11a gene regulatory region.

37. The polynucleotide of clause 36 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

38. The polynucleotide of clause 37 wherein said HE is an I-HjeMI homing endonuclease.

39. The polynucleotide of clause 38 wherein said I-HjeMI homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 28 or SEQ ID NO: 29, wherein said amino acid substitution is selected from the group consisting of Y20, S22, T24, K26, G27, K28, T31, E33, G35, E37, S59, R61, R63, N64, E65, I66, M68, S70, R70, R72, R74, S109, N110, A121, S123, N124, N135, S137, S154, L158, N159, D162, D163, I166, I168, S168, D170, I193, L195, R202, K204, and T206.

The polynucleotide of clause 24 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32.

40. The polynucleotide of clause 39 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence of SEQ ID NO: 33 or a variant thereof wherein said I-HjeMI homing endonuclease can specifically bind to a Bcl11a coding region.

41. The polynucleotide of clause 40 wherein said endonuclease is a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region.

42. The polynucleotide of clause 41 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-OnuI homing endonuclease.

43. The polynucleotide of clause 35 wherein said HE is an I-OnuI homing endonuclease.

44. The polynucleotide of clause 43 wherein said I-OnuI homing endonuclease comprises the amino acid sequence encoded by a variant of the nucleotide sequence of SEQ ID NO: 34 that encodes an I-OnuI homing endonuclease that can specifically bind to said fetal hemoglobin (HbF) silencing region.

45. The polynucleotide of clause 44 wherein said I-OnuI homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein each of said amino acid substitutions is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82.

46. The polynucleotide of clause 45 wherein said I-OnuI homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of F182, N184, I186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

47. The polynucleotide of clause 35 further comprising a polynucleotide encoding a TAL effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

48. A polynucleotide encoding an RNA guide strand that mediates the binding of a Cas9 endonuclease to a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.

49. The polynucleotide of clause 48 wherein said RNA guide strand mediates the specific binding of a Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region.

50. The polynucleotide of clause 49 wherein said fetal hemoglobin (HbF) silencing region comprises the nucleotide sequence of SEQ ID NO: 1.

51. The polynucleotide of clause 50 wherein said fetal hemoglobin (HbF) silencing region comprises the nucleotide sequence of SEQ ID NO: 2.

52. The polynucleotide of clause 50 or clause 51 wherein said RNA guide strand comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.

53. A vector system comprising a vector and a polynucleotide encoding an endonuclease selected from the group consisting of a homing endonuclease (HE) and a CRISPR endonuclease, wherein said endonuclease binds to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.

54. The vector system of clause 53 wherein said vector is selected from the group consisting of an AAV6, a modified adenovirus vector, an integration-deficient lentiviral vector (IDLV), and an integration-deficient foamyviral vector

(IDFV).

55. The vector system of clause 53 wherein said endonuclease is a HE that bind to said Bcl11a coding region or said Bcl11a gene regulatory region.

56. The vector system of clause 55 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

57. The vector system of clause 56 wherein said HE is an I-HjeMI homing endonuclease.

58. The vector system of clause 57 wherein said I- HjeMI homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 28 or SEQ ID NO: 29, wherein said amino acid substitution is selected from the group consisting of Y20, S22, T24, K26, G27, K28, T31, E33, G35, E37, S59, R61, R63, N64, E65, I66, M68, S70, R70, R72, R74, S109, N110, A121, S123, N124, N135, S137, S154, L158, N159, D162, D163, I166, I168, S168, D170, I193, L195, R202, K204, and T206. 25. The polynucleotide of clause 24 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32.

59. The vector system of clause 58 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32.

60. The vector system of clause 59 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence of SEQ ID NO: 33 or a variant thereof wherein said I-HjeMI homing endonuclease can specifically bind to a Bcl11a coding region.

61. The vector system of clause 53 wherein said endonuclease is a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region.

62. The vector system of clause 61 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

63. The vector system of clause 62 wherein said HE is an I-Onul homing endonuclease.

64. The vector system of clause 63 wherein said I-Onul homing endonuclease comprises the amino acid sequence encoded by a variant of the nucleotide sequence of SEQ ID NO: 34 that encodes an I-Onul homing endonuclease that can specifically bind to said fetal hemoglobin (HbF) silencing region.

65. The vector system of clause 64 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82.

66. The vector system of clause 64 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of F182, N184, I186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

67. The vector system of clause 53 further comprising a polynucleotide encoding a TAL effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

68. A vector system comprising a vector and a polynucleotide encoding an RNA guide strand that mediates the binding of a Cas9 endonuclease to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.

69. The vector system of clause 68 wherein said nucleotide sequence within an adult human β -globin locus is a fetal hemoglobin (HbF) silencing region.

70. The vector system of clause 69 wherein said fetal hemoglobin (HbF) silencing region comprises the nucleotide sequence of SEQ ID NO: 1.

71. The vector system of clause 69 wherein said fetal hemoglobin (HbF) silencing region comprises the nucleotide sequence of SEQ ID NO: 2.

72. The vector system of clause 70 or clause 71 wherein said RNA guide strand comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.

73. A cell comprising a polynucleotide encoding one or more endonuclease(s) selected from the group consisting of a homing endonuclease (HE) and a CRISPR endonuclease, wherein each of said endonuclease(s) binds to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.

74. The cell of clause 73 wherein said cell is a stem cell.

75. The cell of clause 73 wherein said cell is a stem cell selected from the group consisting of a hematopoietic stem cell (HSC), an induced pluripotent stem cell (iPSC), an embryonic stem (ES) cell, and an erythroid progenitor cell.

76. The cell of clause 73 wherein said cell is a stem cell selected from the group consisting of a hematopoietic stem cell (HSC), an induced pluripotent stem cell (iPSC), and an erythroid progenitor cell.

77. The cell of clause 73 wherein said endonuclease is a HE that can specifically bind to said Bcl11a coding region.

78. The cell of clause 77 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease, an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

79. The cell of clause 78 wherein said HE is an I-HjeMI homing endonuclease.

80. The cell of clause 79 wherein said I-HjeMI homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 28 or SEQ ID NO: 29, wherein said amino acid substitution is selected from the group consisting of Y20, S22, T24, K26, G27, K28, T31, E33, G35, E37, S59, R61, R63, N64, E65, I66, M68, S70, R70, R72, R74, S109, N110, A121, S123, N124, N135, S137, S154, L158, N159, D162, D163, I166, I168, S168, D170, I193, L195, R202, K204, and T206. 25.

81. The cell of clause 80 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32.

82. The cell of clause 81 wherein said I-HjeMI homing endonuclease comprises the amino acid sequence of SEQ ID NO: 33 or a variant thereof wherein said I-HjeMI homing endonuclease can specifically bind to a Bcl11a coding region.

83. The cell of clause 73 wherein said CRISPR endonuclease comprises an Cas9 endonuclease and an RNA guide strand wherein said RNA guide strand mediates the binding of said Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region.

84. The cell of clause 83 wherein said RNA guide strand comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.

85. The cell of clause 84 wherein said Cas9 endonuclease is encoded by the nucleotide sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 endonuclease.

86. The cell of clause 73 wherein said endonuclease is a HE that binds to a fetal hemoglobin (HbF) silencing region.

87. The cell of clause 86 wherein said HE is selected from the group consisting of an I-HjeMI homing endonuclease,

an I-CpaMI homing endonuclease, and an I-Onul homing endonuclease.

88. The cell of clause 87 wherein said HE is an I-Onul homing endonuclease.

5 89. The cell of clause 88 wherein said I-Onul homing endonuclease comprises the amino acid sequence encoded by a variant of the nucleotide sequence of SEQ ID NO: 34 that encodes an I-Onul homing endonuclease that can specifically bind to said fetal hemoglobin (HbF) silencing region.

10 90. The cell of clause 89 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82.

15 91. The cell of clause 89 wherein said I-Onul homing endonuclease comprises one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein said amino acid substitution is selected from the group consisting of F182, N184, I186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

20 92. The cell of clause 73, further comprising a polynucleotide encoding a TAL effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

93. A genome edited stem cell, wherein said genome edited stem cell is generated by the introduction of a homing endonuclease and a correction template.

25 94. The genome edited stem cell of clause 93 wherein said correction template comprises a nucleotide sequence that permits the modification of key regulatory or coding sequences within a globin gene locus.

30 95. The genome edited stem cell of clause 93 wherein said stem cell is selected from the group consisting of a hematopoietic stem cell (HSC), an induced pluripotent stem cell (iPSC), an embryonic stem (ES) cell, and an erythroid progenitor cell.

35 96. The genome edited stem cell of clause 93 wherein said stem cell is selected from the group consisting of a hematopoietic stem cell (HSC), an induced pluripotent stem cell (iPSC), and an erythroid progenitor cell.

40

45

50

55

EP 3 272 356 A1

SEQUENCE LISTING

<110> Fred Hutchinson Cancer Research Center
 5 <120> COMPOSITIONS AND METHODS FOR THE TREATMENT OF HEMOGLOBINOPATHIES
 <130> P43465EP-D1-PCT
 <140> EP13751886.6
 10 <141> 2013-02-22
 <150> US 61/603,231
 <151> 2012-02-24
 <160> 69
 15 <170> PatentIn version 3.5
 <210> 1
 <211> 3603
 <212> DNA
 20 <213> Homo sapiens
 <400> 1
 ccagtgcagca gggttggttta agataagcag ggtttcatta gtttgtgaga atgaaaaatg 60
 aaccttcatt ccactattcc cttaacttgc cctgagattg gctgttctgt catgtgtgtc 120
 25 ttgactcaga aaccctgttc tcctctacat atctcccccac cgcactctctt tcagcagttg 180
 tttctaaaaa tatcctccta gtttcatttt tgcagaagtg ttttaggcta atatagtggg 240
 atgtatctta gagtttaact tatttgtttc tgtcacttta tactaagaaa acttatctaa 300
 aagcagatgt tttacaagt tgactcaata taaagttctt ctttgccctc agagatTTTT 360
 gtctccaagg gaattttgag aggttggaat ggacaaatct attgctgcag tttaaacttg 420
 35 cttgcttcct ctttcttttg gtaaattctt cctataataa aactctaatt ttttattata 480
 ttgaaataaa tatccattaa aagaatattt aaaaaatgaa tagtgtttat ttaccagtta 540
 ttgaaatagg ttctggaaac atgaatttta aggttaacat tttaatgaca gataaaatca 600
 40 aatattatat acaaattttt tgaatgttta aaattatggt atgactaaag aaagaatgca 660
 aagtgaaaag tagatttacc atattcagcc agattaaatt taacgaagtt cctgggaata 720
 tgctagtaca gaacattttt acagatgtgt tcttaaaaaa aaatgtggaa ttagaccag 780
 45 gaatgaagat ccagtagtt tttcactctt ttctgaattc aaataatgcc acaatggcag 840
 acaaatacac acccatgagc atatccaaaa ggaaggattg aaggaaagag gaggaagaaa 900
 tggagaaaag aaggaaggaa gaggggaaga gagaggatgg aagggatgga ggagaagaag 960
 50 gaaaaataaa taatggagag gagaggagaa aaaaggaggg gagaggagag gagaagggat 1020
 agggaagaga aagagaaagg gaagggaaga gaggaagaa gagaagagga gagaaaagaa 1080
 acgaagagag gggaaggga ggaaaaaaaa gaggaaaaaa gagacaagag aagagataag 1140
 55 actgacagtt caaatttttg tggtgatatg gatcaataga aactcaaact ctgttggtga 1200

EP 3 272 356 A1

	cactgtacaa tagtataacc cctttggaaa acctttaata gtatccacaa atgctggatg	1260
	cttgataagt ctattaccta gcaattacat ttttagatat tcagaaacac atgcatgtgt	1320
5	gtatccaaaag acatgtatag aaatgcttat gacagcaata atcataaaaa cctcaaaccg	1380
	gtagccactt aaatgcttac caacagtaga attgataaat tacggtatag tcaaagaata	1440
	gaatattaca cagaaatgaa aagaatcaac tactgcttaa cacgtagcga tacaatgca	1500
10	ttttacagca tttggttgat taaaagtaac cagaggtgag ttcaaactat atgactttat	1560
	ttgtatatag aaagatggat gatgtgcctg agattctgat cacaagggga aatgttataa	1620
	aatagggtag agaggagcca tgaatgacct ttaaactttg ttacaagtta tttttctgta	1680
15	acctggaagc caacgaaaga tattgaataa ttcaagaaag gtggtggcat ggtttgattt	1740
	gtgtctttaa aagattattc tcacttagtg aagaaatgta ttttagaagt agagaaaatg	1800
20	ggagacaaat agctgggctt ctgttgacgt agggagaaa gtgacaatgc cttttctatt	1860
	atcagacttg gaccatgacg gtgatgtcag tcgtgaacac aagaataggg ccacatttgt	1920
	gagtttagtg gtacgataaa atcagaaata cagtcttgga tacattgtat tgtatgcact	1980
25	cttgtaaaat gcaaaaagat gtacttagat atgtggatct ggagctcaga aagaatacaa	2040
	ccaggtcaag aatacagaat ggaacagaac atacaagaac agatcataat gtgctgtgtg	2100
	aatcactacc actacctgtt aaaaatgaca gatgatgtac ttcacataa tctccttaaa	2160
30	atcttagaat gtgtttgtga gggaggaatt atgtttccaa ttcatatata agaaaattga	2220
	ttctaaaaaa aatgttaggt aaattcttaa ggccatgagg actgttatTT gatctttgtc	2280
35	tgTTaattcc aaagacttg cttttcactt taattctgtt ctacctgaaa tgattttaca	2340
	cattgggaga tctgggttaca tgtttattct atatggattg cattgagagg atttgtataa	2400
	cagaataagg tctttttttc ttttctcttc tgagatggag tttcatccct attgcccaag	2460
40	ctagagtgca atggtgcaat ctaggctcac cgcaacctct gcctcctggg ttcaagcaat	2520
	tctcctgcct cagccacctg aatagctggg actgcaggca tgcaccacac gcccggtga	2580
	ttttgtatTT ttagtagaga tggggtttca ccatgttggt caggctggtc ttgaactcct	2640
45	gacctcaagt gatctgcctg ccttggcctc ccaagtgct gggtttaca gcctgagcca	2700
	ccgcatccag ccaggataag gtctaaaagt ggaaagaata gcatctactc ttgttcagga	2760
	aacaatgagg acctgactgg gcagtaagag tgggtgattaa tagataggga caaattgaag	2820
50	cagaatcgaa ctgttgatta gaggtaggga aatgatttta atctgtgacc ttggtgaatg	2880
	ggcaagtagc tatctaata ctaaaatgga aaacactgga agagaaacag ttttagtata	2940
55	acaagtga aa taccatgct gagtctgagg tgcctatagg acatctatat aaataagccc	3000
	agtacattgt ttgatatatg ggtttggcac tgaggttgga ggtcagaggt tagaaatcag	3060

EP 3 272 356 A1

	agttgggaat tgggattata caggctgtat ttaagagttt agatataact gtgaatccaa	3120
	gagtgtgatg aatacaaagt taaatgaagg acctttaatg aacaccaaca tttaatgtga	3180
5	aatctcaagg aagtatgaag taagacatag tccccaaaat ccccgatgat tttagaactc	3240
	agtatcgatt ttaattagtg taatgccaat gtgggttaga atggaagtca acttgctgtt	3300
	ggtttcagag caggtaggag ataaggttct agattttgac acagtgaaaa gctgaaacaa	3360
10	aaaggaaaag gtagggtgaa agatgggaaa tgtatgtaag gaggatgagc cacatgggat	3420
	gggaggtata ctaaggactc tagggtcaga gaaatatggg ttatatcctt ctacaaaatt	3480
	cacattcttg gctgggtgtg gtggctcacg cctgtgatcc cagcactttc agaggccgag	3540
15	gaggggtggat cacctgatgt taggagttcg agatcagcct gaccaacatg gtgaaacccc	3600
	cta	3603
20	<210> 2 <211> 350 <212> DNA <213> Homo sapiens	
25	<400> 2 aaagatggat gatgtgcctg agattctgat cacaagggga aatgttataa aatagggtag	60
	agaggagcca tgaatgacct ttaaactttg ttacaagtta tttttctgta acctggaagc	120
30	caacgaaaga tattgaataa ttcaagaaag gtgggtggcat ggtttgattt gtgtctttaa	180
	aagattattc tcacttagtg aagaaatgta ttttagaagt agagaaaatg ggagacaaat	240
	agctgggctt ctgttgcagt agggaagaaa gtgacaatgc catttctatt atcagacttg	300
35	gaccatgacg gtgatgtcag tcgtgaacac aagaataggg ccacatttgt	350
40	<210> 3 <211> 21 <212> DNA <213> Homo sapiens	
	<400> 3 gtgcacctca ctccagagga g	21
45	<210> 4 <211> 21 <212> DNA <213> Homo sapiens	
50	<400> 4 gtgcacctca ctccagtgga g	21
55	<210> 5 <211> 22 <212> DNA <213> Homo sapiens	

	<400> 5	
	tttcactta ttcaaccttt ta	22
5	<210> 6	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 6	
	tgtggcccta ttcttgtgtt ca	22
	<210> 7	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 7	
	cattgtcact ttcttcccta ct	22
20	<210> 8	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 8	
	taaaatacat ttcttcacta ag	22
	<210> 9	
30	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 9	
35	actaagtgag aataatcttt ta	22
	<210> 10	
	<211> 22	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 10	
	gccaccacct ttcttgaatt at	22
45	<210> 11	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 11	
	tcttgaatta ttcaatatct tt	22
	<210> 12	
55	<211> 22	
	<212> DNA	
	<213> Homo sapiens	

EP 3 272 356 A1

	<400> 12		
	ttaaaggtca ttcattggctc ct		22
5	<210> 13		
	<211> 2606		
	<212> DNA		
	<213> Homo sapiens		
10	<400> 13		
	gcaatgaaaa taaatgtttt ttattaggca gaatccagat gctcaaggcc cttcataata	60	
	tccccagtt tagtagttgg acttagggaa caaaggaacc tttaatagaa attggacagc	120	
	aagaaagcga gcttagtgat acttgtgggc cagggcatta gccacaccag ccaccacttt	180	
15	ctgataggca gcctgcactg gtgggggtgaa ttctttgcc aagtgatggg ccagcacaca	240	
	gaccagcacg ttgcccagga gctgtgggag gaagataaga ggtatgaaca tgattagcaa	300	
	aagggcctag cttggactca gaataatcca gccttatccc aaccataaaa taaaagcaga	360	
20	atggtagctg gattgtagct gctattagca atatgaaacc tcttacatca gttacaattt	420	
	atatgcagaa atatttatat gcagagatat tgctattgcc ttaaccacaga aattatcact	480	
25	gttattcttt agaatggtgc aaagaggcat gatacattgt atcattattg ccctgaaaga	540	
	aagagattag ggaaagtatt agaaataaga taaacaaaaa agtatattaa aagaagaaag	600	
	catttttttaa aattacaaat gcaaaattac cctgattttg tcaatatgtg tacacatatt	660	
30	aaaacattac actttaaccc ataaatatgt ataattgatta tgtatcaatt aaaaataaaa	720	
	gaaaataaaag tagggagatt atgaatatgc aaataagcac acatatattc caaatagtaa	780	
	tgtactaggc agactgtgta aagttttttt ttaagttact taatgtatct cagagatatt	840	
35	tccttttggt atacacaatg ttaaggcatt aagtataata gtaaaaattg cggagaagaa	900	
	aaaaaaagaa agcaagaatt aaacaaaaga aaacaattgt tatgaacagc aaataaaaga	960	
	aactaaaacg atcctgagac ttccacactg atgcaatcat tcgtctgttt ccatttctaa	1020	
40	actgtaccct gttacttata cccttcctat gacatgaact taaccataga aaagaagggg	1080	
	aaagaaaaca tcaagcgtcc catagactca ccctgaagtt ctcaggatcc acgtgcagct	1140	
45	tgtcacagtg cagctcactc agtgtggcaa aggtgccctt gaggttgtcc aggtgagcca	1200	
	ggccatcact aaaggcaccg agcactttct tgccatgagc cttcacctta gggttgccca	1260	
	taacagcatc aggagtggac agatccccaaggactcaaa gaacctctgg gtccaagggg	1320	
50	agaccaccag cagcctaagg gtgggaaaat agaccaatag gcagagagag tcagtgccta	1380	
	tcagaaaccc aagagtcttc tctgtctcca catgccagct ttctattggg ctcttaaac	1440	
	ctgtcttgta accttgatac caacctgccc agggcctcac caccaacttc atccacgttc	1500	
55	accttgcccc acagggcagc aacggcagac ttctcctcag gagtcagatg caccatgggtg	1560	
	tctgtttgag gttgctagtg aacacagttg tgtcagaagc aaatgtaagc aatagatggc	1620	

EP 3 272 356 A1

	tctgccctga cttttatgcc cagccctggc tccctgcctc cctgctcctg ggagtagatt	1680
	ggccaaccct aggggtgtggc tccacagggg gaggtctaag tgatgacagc cgtacctgtc	1740
5	cttggtctct ctggcactgg cttaggaggt ggacttcaaa ccctcagccc tccctctaag	1800
	atatatctct tggccccata ccatcagtac aaattgctac taaaaacatc ctcccttgca	1860
	agtgtattta cgtaatatatt ggaatcacag cttggtaagc atattgaaga tcgttttccc	1920
10	aatcttctta ttacacaaat aagaagttga tgcactaaaa gtggaagagt tttgtctacc	1980
	ataattcagc tttgggatat gtagatggat ctcttcctgc gtctccagaa tatgcaaaat	2040
	acttacagga cagaatggat gaaaactcta cctcggttct aagcatatct tctccttatt	2100
15	tggattaaaa ctttctggta agaaaagaaa aaatatatat atatagtgt gtatatatac	2160
	acacatacat atacatatat atgcattcat ttgttgttgt ttttcttaat ttgctcatgc	2220
	atgctaataa attatgtcta aaaatagaat aaatacaaat caatgtgtct tgtgcattag	2280
20	ttacttatta ggttttggga aacaagagat aaaaaactag agacctctta atgcagtcaa	2340
	aaatacaaat aaataaaaag tcacttacaa cccaaagtgt gactatcaat ggggtaatca	2400
25	gtggtgtcaa ataggagggt aactggggac atctaactgt ttctgcctgg actaatctgc	2460
	aagagtgtct gggggaacaa aaagcctctg tgacttagaa agtaggggta ggaggggaaa	2520
	aggtcttcta cttggctcag attatctttt tccctctagtc cactaagaat actgcgtttt	2580
30	aaaatcattt ccttgattca agttcc	2606
	<210> 14	
	<211> 606	
35	<212> DNA	
	<213> Homo sapiens	
	<400> 14	
	cacctgaag ttctcaggat ccacgtgcag cttgtcacag tgcagctcac tcagtgtggc	60
40	aaagggtgcc ttgaggttgt ccagggtgagc caggccatca cttaaaggcac cgagcacttt	120
	cttgccatga gccttcacct tagggttgcc cataacagca tcaggagtgg acagatcccc	180
	aaaggactca aagaacctct gggccaagg gtagaccacc agcagcctaa ggggtggaaa	240
45	atagaccaat aggagagag agtcagtgcc tatcagaaac ccaagagtct tctctgtctc	300
	cacatgccca gtttctattg gtctccttaa acctgtcttg taaccttgat accaacctgc	360
	ccagggcctc accaccaact tcatccacgt tcaccttgcc ccacagggca gtaacggcag	420
50	acttctcctc aggagtcaga tgcaccatgg tgtctgtttg aggttgctag tgaacacagt	480
	tgtgtcagaa gcaaatgtaa gcaatagatg gctctgccct gacttttatg ccagccctg	540
55	gctcctgcc tccctgctcc tgggagtaga ttggccaacc ctagggtgtg gctccacagg	600
	gtgagg	606

EP 3 272 356 A1

<210> 15
 <211> 303
 <212> PRT
 <213> Ophiostoma novo-ulmi

<400> 15

Ser Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
 1 5 10 15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn
 20 25 30

Asn Lys Ser Ser Val Gly Tyr Ser Thr Glu Leu Gly Phe Gln Ile Thr
 35 40 45

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp
 50 55 60

Lys Val Gly Val Ile Ala Asn Ser Gly Asp Asn Ala Val Ser Leu Lys
 65 70 75 80

Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys
 85 90 95

Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Met Leu Phe Lys Gln
 100 105 110

Ala Phe Cys Val Met Glu Asn Lys Glu His Leu Lys Ile Asn Gly Ile
 115 120 125

Lys Glu Leu Val Arg Ile Lys Ala Lys Leu Asn Trp Gly Leu Thr Asp
 130 135 140

Glu Leu Lys Lys Ala Phe Pro Glu Ile Ile Ser Lys Glu Arg Ser Leu
 145 150 155 160

Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser
 165 170 175

Gly Glu Gly Cys Phe Phe Val Asn Leu Ile Lys Ser Lys Ser Lys Leu
 180 185 190

Gly Val Gln Val Gln Leu Val Phe Ser Ile Thr Gln His Ile Lys Asp
 195 200 205

Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile
 210 215 220

EP 3 272 356 A1

Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Val Thr
225 230 235 240

5 Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
245 250 255

10 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
260 265 270

15 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
275 280 285

Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe
290 295 300

20 <210> 16
<211> 110
<212> DNA
<213> Homo sapiens

25 <400> 16
tggggggcccc ttccccacac tatctcaatg caaatatctg tctgaaacgg tccctggcta 60
aactccaccc atggggttggc cagccttgcc ttgaccaata gccttgacaa 110

30 <210> 17
<211> 110
<212> DNA
<213> Homo sapiens

35 <400> 17
tggggggcgcc ttccccacac tatctcaatg caaatatctg tctgaaacgg tccctggcta 60
aactccaccc atggggttggc cagccttgcc ttgaccaata gccttgacaa 110

40 <210> 18
<211> 110
<212> DNA
<213> Homo sapiens

45 <400> 18
tggggggcccc ttccccacac tatctcaatg caaacatctg tctgaaacgg tccctggcta 60
aactccaccc atggggttggc cagccttgcc ttgaccaata gccttgacaa 110

50 <210> 19
<211> 110
<212> DNA
<213> Homo sapiens

55 <400> 19
tggggggcccc ttccccacac tatctcaatg caaatatctg tctgaaacgg tccctggcta 60

EP 3 272 356 A1

aactccaccc atgggttggc cagccttgcc ttgactaata gccttgacaa 110

<210> 20
 <211> 110
 <212> DNA
 <213> Homo sapiens

<400> 20
 tggggggcccc ttctccacac tatctcaatg caaatatctg tctgaaacgg tccctggcta 60
 aactccaccc atgggttggc cagccttgcc ttgaccaata gccttgacaa 110

<210> 21
 <211> 110
 <212> DNA
 <213> Homo sapiens

<400> 21
 tggggggcccc ttccccacac tatctcaatg caaacatctg tctgaaacgg tccctggcta 60
 aactccaccc atgggttggc cagccttgcc ttgaccaata gccttgacaa 110

<210> 22
 <211> 110
 <212> DNA
 <213> Homo sapiens

<400> 22
 tggggggcccc ttccccacac tatctcaatg caaatatctg tctgaaacgg tccctggcta 60
 aactccaccc atgggttggc cagccttgcc ttaaccaata gccttgacaa 110

<210> 23
 <211> 110
 <212> DNA
 <213> Homo sapiens

<400> 23
 tggggggcccc ttccccacac tatctcaatg caaatatctg tctgaaacgg tccctggcta 60
 aactccaccc atgggttggc cagccttgcc ttgaccaata gccttgacaa 110

<210> 24
 <211> 2508
 <212> DNA
 <213> Homo sapiens

<400> 24
 atgtctcgcc gcaagcaagg caaacccag cacttaagca aacgggaatt ctcgcccgag 60
 cctcttgaag ccattcttac agatgatgaa ccagaccacg gcccgttggg agctccagaa 120
 ggggatcatg acctcctcac ctgtgggcag tgccagatga acttcccatt gggggacatt 180
 cttatatttta tcgagcacia acggaaacaa tgcaatggca gcctctgctt agaaaaagct 240
 gtggataagc caccttcccc ttcaccaatc gagatgaaaa aagcatccaa tcccgtggag 300

EP 3 272 356 A1

	gttggcatcc aggtcacgcc agaggatgac gattgtttat caacgtcatc tagaggaatt	360
	tgccccaac aggaacacat agcagataaa cttctgcact ggaggggcct ctctccct	420
5	cgttctgcac atggagctct aatccccacg cctgggatga gtgcagaata tgccccgcag	480
	ggtatttgta aagatgagcc cagcagctac acatgtacaa cttgcaaaca gccattcacc	540
	agtgcattgt ttctcttgca acacgcacag aacactcatg gattaagaat ctacttagaa	600
10	agcgaacacg gaagtccctt gaccccgcg gttggatatcc cttcaggact aggtgcagaa	660
	tgtccttccc agccacctct ccatgggatt catattgcag acaataaccc ctttaacctg	720
	ctaagaatac caggatcagt atcgagagag gcttcgggcc tggcagaagg gcgctttcca	780
15	ccactcccc ccctgtttag tccaccaccg agacatcact tggaccccca ccgcatagag	840
	cgcttggggg cggaagagat ggccctggcc acccatcacc cgagtgcctt tgacagggtg	900
	ctgcgggtga atccaatggc tatggagcct ccgcacctgg atttctctag gagacttaga	960
20	gagctggcag ggaacacgtc tagccaccg ctgtcccgag gccggccag ccctatgcaa	1020
	aggttactgc aaccattcca gccaggtagc aagccgcctt tcctggcgac gccccccctc	1080
25	cctcctctgc aatccgcccc tcctccctcc cagcccccg tcaagtccaa gtcattgcag	1140
	ttctgcggca agacgttcaa atttcagagc aacctggtgg tgcaccggcg cagccacacg	1200
	ggcgagaagc cctacaagtg caacctgtgc gaccacgcgt gcacccaggc cagcaagctg	1260
30	aagcgccaca tgaagacgca catgcacaaa tcgtcccca tgacgggtcaa gtccgacgac	1320
	ggtctctcca ccgccagctc ccgggaaccc ggcaaccagc acttgggtggg cagcgccagc	1380
	agcgcgctca agtccgtggt ggccaagtgc aagagcgaga acgaccccaa cctgatcccg	1440
35	gagaacgggg acgaggagga agaggaggac gacgaggaag aggaagaaga ggaggaagag	1500
	gaggaggagg agctgacgga gagcgagagg gtggactacg gcttcgggct gagcctggag	1560
	gcggcgcgcc accacgagaa cagctcgcg ggcgcggtcg tgggcgtggg cgacgagagc	1620
40	cgcgccctgc ccgacgtcat gcagggcatt gtgctcagct ccatgcagca cttcagcgag	1680
	gccttccacc aggtcctggg cgagaagcat aagcgcggcc acctggccga ggccgagggc	1740
	cacagggaca cttgcgacga agactcgggt gccggcgagt cggaccgcat agacgatggc	1800
45	actgttaatg gcccgggctg ctccccgggc gactcggcct cggggggcct gtccaaaag	1860
	ctgctgctgg gcagccccag ctcgctgagc cccttctcta agcgcatcaa gctcgagaag	1920
	gagttcgacc tgcccccggc cgcgatgccc aacacggaga acgtgtactc gcagtggctc	1980
50	gccggctacg cggcctccag gcagctcaaa gatcccttcc ttagcttcgg agactccaga	2040
	caatcgctt ttgcctcctc gtcggagcac tcctcggaga acgggagttt gcgcttctcc	2100
55	acaccgccc gggagctgga cggagggatc tcggggcgca gcggcacggg aagtggaggg	2160
	agcacgcccc atattagtgg tccgggccc ggcaggccca gctcaaaaga gggcagacgc	2220

EP 3 272 356 A1

	agcgacactt gtgagtactg tgggaaagtc ttcaagaact gtagcaatct cactgtccac	2280
	aggagaagcc acacggggcga aaggccttat aaatgcgagc tgtgcaacta tgccgtgtgcc	2340
5	cagagtagca agctcaccag gcacatgaaa acgcatggcc aggtggggaa ggacgtttac	2400
	aaatgtgaaa tttgtaagat gccttttagc gtgtacagta ccctggagaa acacatgaaa	2460
	aaatggcaca gtgatcgagt gttgaataat gatataaaaa ctgaatag	2508
10		
	<210> 25	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 25	
	atgggattca tattgcagac aa	22
20		
	<210> 26	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 26	
25	agccattgga ttcaaccgca gc	22
30		
	<210> 27	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 27	
	caaacagcca ttcaccagtg ca	22
35		
	<210> 28	
	<211> 771	
	<212> DNA	
	<213> Artificial Sequence	
40		
	<220>	
	<223> Nucleotide Sequence of I-HjeMI, Codon Optimized for Expression in E. coli	
45	<400> 28	
	atgggatccc acatggacct gacctacgt tacctggttg gtctgttcga aggtgacggt	60
	tacttctcta tcaccaaaaa gggtaaatac ctgacctacg aactgggtat cgaactgtct	120
	atcaaagacg ttcagctgat ctacaaaatc aaagacatcc tgggtgttgg taaagtttct	180
50	ttccgtaaac gtaacgaaat cgaaatggtt tctctgcgta tccgtgacaa gaatcacctg	240
	aaaaacttca tcctgccgat cttcgacaaa taccgatgc tgtctaaca gacgtacgac	300
55	tacctgcgtt tcaaagacgc tctcctgtct aacattatct actctgacga tctgccggaa	360
	tacgctcgtt ctaacgaatc tatcaactct gttgactcta ttatcaacac ctcttacttc	420

EP 3 272 356 A1

tctgcttggc tggttggtt catcgaagct gaaggttgct tctctaccta caaactgaac 480
 aaagatgacg attacctgat cgcttctttc gacatcgctc agaaagacgg tgacatcctg 540
 5 atctctgcta tccacaaata cctgtctttc accacgaaaa tctacctgga caaaaccaac 600
 tgctctcgtc tgaaagtacg cgggtgtacgt tctgttaaaa acgtgggttaa attcatccag 660
 ggtgctccgg ttaaaactgct cggtaacaag aaactgcagt acaaactgtg gatcaaacag 720
 10 ctgcgtaaaa tctctcgta ctctgaaaaa atccagctgc cgtctaacta c 771

<210> 29
 <211> 771
 <212> DNA
 15 <213> Artificial Sequence
 <220>
 <223> Nucleotide Sequence of I-HjeMI, Codon Optimized for Expression in
 20 Mammals

<400> 29
 atgggcagcc acatggacct gacctacgcc tatctggtcg gcctgttcga gggcgacggc 60
 tatttttagca taaccaagaa gggcaagtat ctgacgtatg aactgggcat cgagctctcc 120
 25 atcaaggacg tgcagctcat ctacaagatc aaggacatcc tcggcgtggg caaagtgtct 180
 tttaggaaga ggaacgagat cgagatggtc agcctgcgaa tcagggacaa aaaccacctg 240
 aagaacttca tcctgcccac cttcgacaag taccatcatgc tgagcaacaa gcagtacgac 300
 30 tatctccgat tcaaggatgc cctcctgtcc aacatcatct atagcgacga cctgcccag 360
 tacgccagga gcaacgagtc aatcaatagc gtggacagca tcatcaacac ctcatacttc 420
 agcgcctggc tggttggctt catcgaggcc gagggctgct tcagcaccta caagctcaac 480
 35 aaggacgacg attatttgat cgcgagcttc gatatagccc agaaggacgg cgacattctc 540
 atctccgcga tccacaaata cctgagcttc acgacaaaaa tctacctgga caagaccaac 600
 tgtagcaggc tcaaggtcac cggcgtgagg agcgtcaaga acgtgggttaa gttcatccag 660
 40 ggtgcgccgg tcaagttgct gggtaacaag aagctgcagt acaaactttg gataaagcag 720
 ctgcgcaaga tctcccgata cagcgagaaa atccagctgc ccagtaacta c 771

45 <210> 30
 <211> 254
 <212> PRT
 <213> Trichoderma reesei

50 <400> 30

Met Gly Asp Leu Thr Tyr Ala Tyr Leu Val Gly Leu Phe Glu Gly Asp
 1 5 10 15

55 Gly Tyr Phe Ser Ile Thr Lys Lys Gly Lys Tyr Leu Thr Tyr Glu Leu
 20 25 30

EP 3 272 356 A1

	Gly	Ile	Glu	Leu	Ser	Ile	Lys	Asp	Val	Gln	Leu	Ile	Tyr	Lys	Ile	Lys	
			35					40					45				
5	Asp	Ile	Leu	Gly	Val	Gly	Lys	Val	Ser	Phe	Arg	Lys	Arg	Asn	Glu	Ile	
	50						55					60					
	Glu	Met	Val	Ser	Leu	Arg	Ile	Arg	Asp	Lys	Asn	His	Leu	Lys	Asn	Phe	
10	65					70					75					80	
	Ile	Leu	Pro	Ile	Phe	Asp	Lys	Tyr	Pro	Met	Leu	Ser	Asn	Lys	Gln	Tyr	
					85					90					95		
15	Asp	Tyr	Leu	Arg	Phe	Lys	Asp	Ala	Leu	Leu	Ser	Asn	Ile	Ile	Tyr	Ser	
				100					105						110		
	Asp	Asp	Leu	Pro	Glu	Tyr	Ala	Arg	Ser	Asn	Glu	Ser	Ile	Asn	Ser	Val	
20			115					120						125			
	Asp	Ser	Ile	Ile	Asn	Thr	Ser	Tyr	Phe	Ser	Ala	Trp	Leu	Val	Gly	Phe	
25		130					135					140					
	Ile	Glu	Ala	Glu	Gly	Cys	Phe	Ser	Thr	Tyr	Lys	Leu	Asn	Lys	Asp	Asp	
	145					150					155					160	
30	Asp	Tyr	Leu	Ile	Ala	Ser	Phe	Asp	Ile	Ala	Gln	Lys	Asp	Gly	Asp	Ile	
					165					170					175		
	Leu	Ile	Ser	Ala	Ile	His	Lys	Tyr	Leu	Ser	Phe	Thr	Thr	Lys	Ile	Tyr	
35				180					185					190			
	Leu	Asp	Lys	Thr	Asn	Cys	Ser	Arg	Leu	Lys	Val	Thr	Gly	Val	Arg	Ser	
40			195					200					205				
	Val	Lys	Asn	Val	Val	Lys	Phe	Ile	Gln	Gly	Ala	Pro	Val	Lys	Leu	Leu	
		210					215					220					
45	Gly	Asn	Lys	Lys	Leu	Gln	Tyr	Lys	Leu	Trp	Ile	Lys	Gln	Leu	Arg	Lys	
	225					230					235					240	
	Ile	Ser	Arg	Tyr	Ser	Glu	Lys	Ile	Gln	Leu	Pro	Ser	Asn	Tyr			
50					245					250							
	<210>	31															
	<211>	768															
55	<212>	DNA															
	<213>	Artificial Sequence															

<220>

<223> Nucleotide Sequence for a BCL11A Gene Targeting Nuclease Based on the Homing Endonuclease I-HjeMI; Codon Optimized for Expression in *E. coli* and Obtained through Directed Evolution in IVC and in Bacteria

<400> 31

```

atgggatccc acatggacct gacctacgct tacctggttg gtctgttcga aggtgacggt      60
tacttcacta tcgctaaagc gggtaagtac ctgaattacg agctgggtat cacgctgtct      120
atcaaagacg ctcagctgat ctacaaaatc aaagacatcc tgggtggttg taatgtttat      180
ttccggaaat ataggcagca tgaaatggtt tctctgcgga tccaggacaa gaatcacctg      240
aaaaacttca tcctgccgat cttcgacaaa taccgatgc tgtctaacaa gcagtacgac      300
tacctgcggt tcaaagacgc tctcctgtct aacattatct actctgacga tctgccggaa      360
tacgctcggt ctaacgaatc tatcaactct gttgactcta ttatcaacac ctcttacttc      420
tctgcttggc tggttggttt catcgaagct gaaggttgct tcacgaccta caaagcgagt      480
aaagataagt acctgacggc tgggttcagt atcgctcaga aagacggtga catcctgac      540
tctgctatcc acaaatacct gtctttcacc acgaaaccgt acaaggacaa aaccaactgc      600
tctcatctga aggtgaccgg tgtacgttct gttaacaacg tggttaaatt catccagggt      660
gctccggtta aactgctcgg taacaagaaa ctgcagtaca aactgtggat caaacagctg      720
cgtaaaatct ctcgttactc tgaaaaaatc cagctgccgt ctaactac      768

```

<210> 32

<211> 768

<212> DNA

<213> Artificial Sequence

<220>

<223> Nucleotide Sequence of a BCL11A Gene Targeting Nuclease Based on the Homing Endonuclease I-HjeMI; Codon Optimized for Expression in Mammals and Obtained through Directed Evolution in IVC and in Bacteria

<400> 32

```

atgggcagcc acatggacct gacctacgcc tatctggtcg gcctgttcga gggcgacggc      60
tattttacaa tagctaaggc cggcaagtat ctgaactacg agctgggcat cacactctcc      120
atcaaggacg ctcagctcat ctacaagatc aaggacatcc tcggcgtggg caacgtgtac      180
tttaggaagt acaggcaaca tgagatggtc agcctgcgaa tccaggacaa aaaccacctg      240
aagaacttca tcctgcccat cttcgacaag taccocatgc tgagcaacaa gcagtacgac      300
tacctgcgat tcaaggatgc ctcctgtcc aacatcatct atagcgacga cctgcccgag      360
tacgccagga gcaacgagtc aatcaatagc gtggacagca tcatcaacac ctcatacttc      420
agcgcttggc tggttggctt catcgaggcc gagggctgct tcaccaccta caaggcatcc      480
aaggacaagt acctgacagc gggcttctcc atagccaga aggacggcga cattctcatc      540

```

EP 3 272 356 A1

tccgcgatcc acaaatacct gagcttcacg accaaaccct acaaagacaa gaccaactgt 600
 agccacctca aggtcacggg cgtgaggagc gtcaataacg tggtaagtt catccagggt 660
 5 gcgcgggtca agctgctggg taacaagaag ctgcagtaca aactttggat aaagcagctg 720
 cgcaagatct cccgatacag cgagaaaatc cagctgcccc gtaactac 768

10 <210> 33
 <211> 256
 <212> PRT
 <213> Artificial Sequence

15 <220>
 <223> Amino Acid Sequence of a BCL11A Gene Targeting Nuclease Based on
 the Homing Endonuclease I-HjeMI

<400> 33
 20 Met Gly Ser His Met Asp Leu Thr Tyr Ala Tyr Leu Val Gly Leu Phe
 1 5 10 15
 Glu Gly Asp Gly Tyr Phe Thr Ile Ala Lys Ala Gly Lys Tyr Leu Asn
 20 25 30
 25 Tyr Glu Leu Gly Ile Thr Leu Ser Ile Lys Asp Ala Gln Leu Ile Tyr
 35 40 45
 30 Lys Ile Lys Asp Ile Leu Gly Val Gly Asn Val Tyr Phe Arg Lys Tyr
 50 55 60
 Arg Gln His Glu Met Val Ser Leu Arg Ile Gln Asp Lys Asn His Leu
 35 65 70 75 80
 Lys Asn Phe Ile Leu Pro Ile Phe Asp Lys Tyr Pro Met Leu Ser Asn
 85 90 95
 40 Lys Gln Tyr Asp Tyr Leu Arg Phe Lys Asp Ala Leu Leu Ser Asn Ile
 100 105 110
 45 Ile Tyr Ser Asp Asp Leu Pro Glu Tyr Ala Arg Ser Asn Glu Ser Ile
 115 120 125
 Asn Ser Val Asp Ser Ile Ile Asn Thr Ser Tyr Phe Ser Ala Trp Leu
 50 130 135 140
 Val Gly Phe Ile Glu Ala Glu Gly Cys Phe Thr Thr Tyr Lys Ala Ser
 145 150 155 160
 55 Lys Asp Lys Tyr Leu Thr Ala Gly Phe Ser Ile Ala Gln Lys Asp Gly
 165 170 175

EP 3 272 356 A1

Asp Ile Leu Ile Ser Ala Ile His Lys Tyr Leu Ser Phe Thr Thr Lys
180 185 190

5 Pro Tyr Lys Asp Lys Thr Asn Cys Ser His Leu Lys Val Thr Gly Val
195 200 205

10 Arg Ser Val Asn Asn Val Val Lys Phe Ile Gln Gly Ala Pro Val Lys
210 215 220

Leu Leu Gly Asn Lys Lys Leu Gln Tyr Lys Leu Trp Ile Lys Gln Leu
225 230 235 240

15 Arg Lys Ile Ser Arg Tyr Ser Glu Lys Ile Gln Leu Pro Ser Asn Tyr
245 250 255

20 <210> 34
<211> 912
<212> DNA
<213> Artificial Sequence

25 <220>
<223> Nucleotide Sequence of I-OnuI, Codon Optimized for Expression in
E. coli

<400> 34
atgtccgcct acatgtcccg tcgcgagtc attaacccgt ggattctcac cggtttcgcc 60
30 gacgcggaag gctccttttt gctgcgcatc cgcaacaaca acaagtccag cgtcggctac 120
tccactgagc tcggcttcca aattacactt cataacaagg acaagagcat tcttgagaac 180
atccagtcaa catggaaggt gggcgtgatc gccaacagcg gtgacaacgc cgtgtcgtg 240
35 aaggtcacgc gttttgagga cctgaaggtc attatcgacc attttgaaaa ataccactg 300
attacgcaga agctcgggtga ctacatgctg tttaagcagg cgttttgcgt catggagaac 360
40 aaggagcatt tgaagattaa tggatcaag gagctggtgc gcattaaggc aaagctcaat 420
tggggtctga cggatgagct gaagaaggcc tttccggaga tcatctcgaa ggagcgctcc 480
ctcatcaaca agaacatccc taatttcaag tggctggcgg gttttacctc gggcgaggg 540
45 tgcttctttg ttaacctgat caagtcaaag tcgaagctag gtgtccaggt gcagctggtg 600
ttcagcatta cccaacacat caaggataag aacctcatga actctctgat tacctacttg 660
ggctgcggct acattaagga gaaaaacaag agtgagttct cctggcttga ctctgctg 720
50 acgaaattct ccgacatcaa cgacaagatc attccgggtct ttcaggaaaa cagctcatc 780
ggcgtgaagc tcgaggactt cgaggattgg tgtaaggctg ctaagctgat cgaggagaaa 840
aagcacctga cagaaagtgg cctggacgag atcaagaaga ttaagctgaa catgaacaag 900
55 ggcagagtat tc 912

<210> 35
 <211> 1998
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Nucleotide Sequence of MegaTAL:5.5 RVD + Y2 I-AniI

<400> 35

10	gtggatctac gcacgctcgg ctacagtcag cagcagcaag agaagatcaa accgaaggtg	60
	cggttcgacag tggcgcagca ccacgaggca ctgggtgggcc atggggtttac acacgcgcac	120
	atcggttgccg tcagccaaca cccggcagcg ttagggaccg tcgctgtcac gtatcagcac	180
15	ataatcacgg cggttgccaga ggcgacacac gaagacatcg ttggcgctcg caaacagtgg	240
	tccggcgccac gcgccctgga ggccttgctc acggatgcgg gggagttgag aggtccgccg	300
	ttacagttgg acacaggcca acttggtgaag attgcaaaac gtggcggcgt gaccgcaatg	360
20	gaggcagtg atgcatcgcg caatgcactg acgggtgccc ccctgaacct gaccccgac	420
	caagtgggtg ctatcgccag caacaatggc ggcaagcaag cgctcgaaac ggtgcagcgg	480
25	ctggttgccg tgctgtgcca ggaccatggc ctgactccgg accaagtggg ggctatcgcc	540
	agccacgatg gcggcaagca agcgctcgaa acgggtgcagc ggctggtgcc ggtgctgtgc	600
	caggaccatg gcctgacccc ggaccaagtg gtggctatcg ccagcaacat tggcggcaag	660
30	caagcgctcg aaacggtgca gcggctggtg ccggtgctgt gccaggacca tggcctgacc	720
	ccggaccaag tgggtggctat cgccagcaac aatggcggca agcaagcgct cgaaacggtg	780
	cagcggctgt tgccggtgct gtgccaggac catggcctga ctccggacca agtgggtggct	840
35	atgccagacc acgatggcgg caagcaagcg ctcgaaacgg tgcagcggct gttgccggtg	900
	ctgtgccagg accatggcct gaccccgac caagtgggtg ctatcgccag caacggtggc	960
40	ggcaagcaag cgctcgaaag cattgtggcc cagctgagcc ggcctgatcc ggcgttggcc	1020
	gcgttgacca acgaccacct cgtcgccttg gcctgcctcg gcggacgtcc tgccatggat	1080
	gcagtgaaaa agggattgcc gcacgcgccg gaattgatca gaagagtcaa tcgccgtatt	1140
45	ggcgaacgca cgtcccatcg cgttgcgata tctagagtgg gaggaagcga tcttacgtac	1200
	gcgtatattag ttggtctcta cgaaggggat ggatacttta gtatcaccaa gaaaggcaag	1260
	tacttgactt atgaattagg tattgagctg agcatcaaag acgtccaatt gatttacaag	1320
50	atcaagaaaa tcctaggtat tggcatcgta agcttcagga agagaaacga gattgaaatg	1380
	gttgccattga ggatccgtga taagaatcat ctaaaatcta agatattgcc tatatttgag	1440
55	aagtatccaa tgtttttccaa caaacagtac gactatttaa gattcaggaa tgcattgtta	1500
	tctggcatta tatacctaga agacttgcct gattacacta gaagtgcga accattgaat	1560

EP 3 272 356 A1

tctatagaat ccattatcaa cacatcatac ttctccgcct ggctagttgg atttatagaa 1620
gctgagggct gtttcagtgt gtacaagctg aacaaagacg atgactactt gattgcttca 1680
5 ttcgacattg cccaaagaga tggatgatc ttgatttcag caattaggaa gtacttaagt 1740
ttcactacta aggtttacct agacaagact aattgtagca aattgaaggt cactagtgtt 1800
agatccgtcg agaacatcat taagtttctg cagaatgctc ctgtcaaatt gttaggcaac 1860
10 aagaaactgc aatacaagtt gtgggtgaaa caactaagga agatttctag gtattccgag 1920
aagatcaaga ttccatcaaa ctacgtcgac cgagcatctt accgccattt ataccatata 1980
ttgttctgtt tttcttga 1998

15

<210> 36

<211> 665

<212> PRT

20 <213> Artificial Sequence

<220>

<223> Amino Acid Sequence of MegaTAL:5.5 RVD + Y2 I-AniI

<400> 36

25

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
1 5 10 15

30

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
20 25 30

35

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
35 40 45

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
50 55 60

40

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
65 70 75 80

45

Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
85 90 95

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
100 105 110

50

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
115 120 125

55

Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala
130 135 140

EP 3 272 356 A1

	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	145	150	155	160
5	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	165	170	175	
10	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	180	185	190	
	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	195	200	205	
15	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	210	215	220	
20	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	225	230	235	240
25	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	245	250	255	
	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	260	265	270	
30	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	275	280	285	
35	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	290	295	300	
40	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	305	310	315	320
	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	325	330	335	
45	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	340	345	350	
50	Leu	Gly	Gly	Arg	Pro	Ala	Met	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	355	360	365	
	Ala	Pro	Glu	Leu	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Gly	Glu	Arg	Thr	370	375	380	
55	Ser	His	Arg	Val	Ala	Ile	Ser	Arg	Val	Gly	Gly	Ser	Asp	Leu	Thr	Tyr	385	390	395	400

EP 3 272 356 A1

	Ala Tyr Leu Val Gly Leu Tyr Glu Gly Asp Gly Tyr Phe Ser Ile Thr	
	405	410 415
5	Lys Lys Gly Lys Tyr Leu Thr Tyr Glu Leu Gly Ile Glu Leu Ser Ile	
	420	425 430
10	Lys Asp Val Gln Leu Ile Tyr Lys Ile Lys Lys Ile Leu Gly Ile Gly	
	435	440 445
15	Ile Val Ser Phe Arg Lys Arg Asn Glu Ile Glu Met Val Ala Leu Arg	
	450	455 460
20	Lys Tyr Pro Met Phe Ser Asn Lys Gln Tyr Asp Tyr Leu Arg Phe Arg	
	485	490 495
25	Asn Ala Leu Leu Ser Gly Ile Ile Tyr Leu Glu Asp Leu Pro Asp Tyr	
	500	505 510
30	Thr Arg Ser Asp Glu Pro Leu Asn Ser Ile Glu Ser Ile Ile Asn Thr	
	515	520 525
35	Ser Tyr Phe Ser Ala Trp Leu Val Gly Phe Ile Glu Ala Glu Gly Cys	
	530	535 540
40	Phe Ser Val Tyr Lys Leu Asn Lys Asp Asp Asp Tyr Leu Ile Ala Ser	
	545	550 555 560
45	Phe Asp Ile Ala Gln Arg Asp Gly Asp Ile Leu Ile Ser Ala Ile Arg	
	565	570 575
50	Lys Tyr Leu Ser Phe Thr Thr Lys Val Tyr Leu Asp Lys Thr Asn Cys	
	580	585 590
55	Ser Lys Leu Lys Val Thr Ser Val Arg Ser Val Glu Asn Ile Ile Lys	
	595	600 605
	Phe Leu Gln Asn Ala Pro Val Lys Leu Leu Gly Asn Lys Lys Leu Gln	
	610	615 620
	Tyr Lys Leu Trp Leu Lys Gln Leu Arg Lys Ile Ser Arg Tyr Ser Glu	
	625	630 635 640
	Lys Ile Lys Ile Pro Ser Asn Tyr Val Asp Arg Ala Ser Tyr Arg His	
	645	650 655

Leu Tyr Pro Tyr Leu Phe Cys Phe Ser
660 665

5 <210> 37
 <211> 4146
 <212> DNA
 <213> Artificial Sequence

10 <220>
 <223> Engineered type II CRISPR system for human cells; expression
 format and full sequence of cas9 gene insert of Mali et al.

<400> 37

15 gccacccatgg acaagaagta ctccattggg ctcgatatcg gcacaaacag cgtcggctgg 60
 gccgtcatta cggacgagta caagggtgccg agcaaaaaat tcaaagttct gggcaatacc 120
 gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tggtcgactc cggggagacg 180
 20 gccgaagcca cgcgggtcaa aagaacagca cggcgagat atacccgcag aaagaatcgg 240
 atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300
 cataggctgg aggagtcctt tttggtggag gaggataaaa agcacgagcg ccaccaatc 360
 25 tttggcaata tcgtggacga ggtggcgtag catgaaaagt acccaaccat atatcatctg 420
 aggaagaagc ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgtg 480
 gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa cccagacaac 540
 30 agcgatgtcg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag 600
 aacccgatca acgcatccgg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa 660
 tcccggcggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt 720
 35 ggtaatctta tcgccctgtc actcgggctg accccaact ttaaactctaa cttcgacctg 780
 gccgaagatg ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg 840
 40 ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac 900
 gccattctgc tgagtgatat tctgcgagtg aacacggaga tcaccaaagc tccgctgagc 960
 gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt 1020
 45 gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc 1080
 tacgccgat acattgacgg cggagcaagc caggaggaat ttacaaatt tattaagccc 1140
 atcttgaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg 1200
 50 ttgcgcaaac agcgacttt cgacaatgga agcatcccc accagattca cctgggcgaa 1260
 ctgcacgcta tcctcaggcg gcaagaggat ttctaccct ttttgaaaga taacaggga 1320
 55 aagattgaga aaatcctcac atttcggata ccctactatg taggccccct cggccgggga 1380
 aattccagat tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc 1440

EP 3 272 356 A1

	gaggaagtcg tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt	1500
	gataaaaaatc tgcctaacga aaaggtgctt cctaaacact ctctgctgta cgagtacttc	1560
5	acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca	1620
	ttcctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa	1680
	gttaccgtga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt	1740
10	gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg	1800
	aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac	1860
	attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact	1920
15	tacgctcatc tcttcgacga caaagtcatg aaacagctca agaggcgccg atatacagga	1980
	tggggggcggc tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca	2040
20	atcctggatt ttcttaagtc cgatggattt gccaacggga acttcatgca gttgatccat	2100
	gatgactctc tcacctttaa ggaggacatc cagaaagcac aagtttctgg ccagggggac	2160
	agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg	2220
25	cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat	2280
	atcgttatcg agatggcccg agagaaccaa actaccaga agggacagaa gaacagtagg	2340
	gaaaggatga agaggattga agagggata aaagaactgg ggtcccaaact ccttaaggaa	2400
30	caccagttg aaaacaccca gcttcagaat gagaagctct acctgtacta cctgcagaac	2460
	ggcagggaca tgtacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg	2520
35	gatcatatcg tgcccagtc ttttctcaa gatgattcta ttgataataa agtggtgaca	2580
	agatccgata aaaatagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa	2640
	atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat	2700
40	aatctgacta aggttgaacg aggtggcctg tctgagttgg ataaagccgg cttcatcaaa	2760
	aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc	2820
	atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg	2880
45	aagtctaagc tgggtctcaga tttcagaaag gactttcagt tttataaggt gagagagatc	2940
	aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttatac	3000
	aaaaaatatc ccaagcttga atctgaatth gtttacggag actataaagt gtacgatgtt	3060
50	aggaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt	3120
	tacagcaata ttatgaatth tttcaagacc gagattacac tggccaatgg agagattcgg	3180
55	aagcgaccac ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg	3240
	gatttcgcca cagtcgggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc	3300

EP 3 272 356 A1

gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag 3360

ctgatcgcac gcaaaaaaga ttgggacccc aagaaatacg gcggattcga ttctcctaca 3420

5 gtcgcttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa 3480

agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc 3540

atcgactttc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt 3600

10 cccaagtact ctctctttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc 3660

gagctgcaga aaggtaacga gctggcactg ccctctaaat acgttaattt cttgtatctg 3720

gccagccact atgaaaagct caaagggtct ccogaagata atgagcagaa gcagctgttc 3780

15 gtggaacaac acaaacacta cttgatgag atcatcgagc aaataagcga attctccaaa 3840

agagtgatcc tcgccgacgc taacctcgat aagggtgcttt ctgcttaca taagcacagg 3900

gataagccca tcaggagca ggcagaaaac attatccact tgtttactct gaccaacttg 3960

20 ggcgcgctg cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct 4020

acaaaggagg tcctggacgc cacactgatt catcagtcga ttacggggct ctatgaaaca 4080

25 agaatcgacc tctctcagct cggtgagac agcagggctg accccaagaa gaagaggaag 4140

gtgtga 4146

<210> 38

30 <211> 455

<212> DNA

<213> Artificial Sequence

<220>

35 <223> Nucleotide Sequence of RNA Guide Strand for use with Cas9 Endonuclease

<220>

40 <221> misc_feature

<222> (320)..(338)

<223> n is a, c, g, or t

<400> 38

tgtacaaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg taccaaggctc 60

45 gggcaggaag agggcctatt tcccatgatt ccttcattatt tgcataatcg atacaaggct 120

gtagagaga taattagaat taatttgact gtaaacacaa agatattagt acaaaatacg 180

tgacgtagaa agtaataatt tcttgggtag tttgcagttt taaaattatg ttttaaaatg 240

50 gactatcata tgcttaccgt aacttgaaag tatttcgatt tcttggcttt atatatcttg 300

tggaaaggac gaaacaccgn nnnnnnnnnn nnnnnnnngt tttagagcta gaaatagcaa 360

gttaaaataa ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt 420

55 tctagacca gctttcttgt acaagtttg catta 455

<210> 39
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 5
 <220>
 <223> BCL11A gene target site
 <400> 39
 10 tccaagtgat gtctcgggtgg tg 22
 <210> 40
 <211> 22
 <212> DNA
 15 <213> Artificial Sequence
 <220>
 <223> Targeted Homing Endonuclease Target Site
 <400> 40
 20 tttccaatta ttcaaccttt ta 22
 <210> 41
 <211> 22
 <212> DNA
 25 <213> Artificial Sequence
 <220>
 <223> Targeted Homing Endonuclease Target Site
 <400> 41
 30 tcttgaatta ttcaaccttt ta 22
 <210> 42
 <211> 22
 <212> DNA
 35 <213> Artificial Sequence
 <220>
 <223> Targeted Homing Endonuclease Target Site
 <400> 42
 40 tttccattta ttcaatattt ta 22
 <210> 43
 <211> 22
 <212> DNA
 45 <213> Artificial Sequence
 <220>
 <223> Targeted Homing Endonuclease Target Site
 <400> 43
 50 tttccattta ttcaatatct tt 22
 <210> 44
 <211> 12

EP 3 272 356 A1

<212> DNA
 <213> Artificial Sequence

 <220>
 5 <223> U6 Promoter Sequence of a Generic Cas9 Guide RNA

 <400> 44
 ggacgaaaca cc 12

 10 <210> 45
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 15 <220>
 <223> Generic Target-specific Sequence of a Generic Cas9 Guide RNA

 <220>
 <221> misc_feature
 20 <222> (2)..(20)
 <223> n is a, c, g, or t

 <400> 45
 gnnnnnnnnnn nnnnnnnnnnn 20

 25 <210> 46
 <211> 77
 <212> DNA
 <213> Artificial Sequence

 30 <220>
 <223> Guide RNA Scaffold Sequence of a Generic Cas9 Guide RNA

 <400> 46
 35 gtttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60
 ggcaccgagt cggtgct 77

 <210> 47
 <211> 10
 40 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Poly T Tail of a Generic Cas9 Guide RNA
 45

 <220>
 <221> misc_feature
 <222> (1)..(4)
 50 <223> n is a, c, g, or t

 <400> 47
 nnnntttttt 10

 55 <210> 48
 <211> 23
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> Target-specific Sequence for Cas9 Guide RNA GGN20GG-B
 5
 <400> 48
 gccatttcta ttatcagact tgg 23

 <210> 49
 10 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 15 <223> Target-specific Sequence for Cas9 Guide RNA GGN20GG-C
 <400> 49
 gctgggcttc tgttgcagta ggg 23

 <210> 50
 20 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 25 <223> Target-specific Sequence for Cas9 Guide RNA GGN20GG-D
 <400> 50
 gaaaatggga gacaaatagc tgg 23

 <210> 51
 30 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 35 <223> Target-specific Sequence for Cas9 Guide RNA GGN20GG-E
 <400> 51
 gaataattca agaaaggtgg tgg 23

 <210> 52
 40 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 45 <223> Target-specific Sequence for Cas9 Guide RNA GGN20GG-F
 <400> 52
 50 gatattgaat aattcaagaa agg 23

 <210> 53
 <211> 23
 <212> DNA
 55 <213> Artificial Sequence
 <220>

EP 3 272 356 A1

	<223> Target-specific Sequence for Cas9 Guide RNA GGN20GG-G	
5	<400> 53 gcctgagatt ctgattcaciaa ggg	23
	<210> 54 <211> 23 <212> DNA <213> Artificial Sequence	
10	<220> <223> Target-specific Sequence for Cas9 Guide RNA GGN20GG-H	
15	<400> 54 ggtaaattct taaggccatg agg	23
	<210> 55 <211> 22 <212> DNA <213> Artificial Sequence	
20	<220> <223> Engineered target site	
25	<400> 55 ttgaggagat gtctctgtta at	22
	<210> 56 <211> 22 <212> DNA <213> Artificial Sequence	
30	<220> <223> Engineered target site	
35	<400> 56 ttgaggtgat gtctctgtta at	22
	<210> 57 <211> 22 <212> DNA <213> Artificial Sequence	
40	<220> <223> Engineered target site	
45	<400> 57 ttgaagtgat gtctctgtta at	22
	<210> 58 <211> 22 <212> DNA <213> Artificial Sequence	
50	<220> <223> Engineered target site	
55	<400> 58	

	tccaagtgat gtctctgtta at	22
5	<210> 59 <211> 22 <212> DNA <213> Artificial Sequence	
10	<220> <223> Engineered target site	
	<400> 59 tccaagtgat gtctctgtta at	22
15	<210> 60 <211> 22 <212> DNA <213> Artificial Sequence	
20	<220> <223> Engineered target site	
	<400> 60 ttgaggaggt ttctgtgtta at	22
25	<210> 61 <211> 22 <212> DNA <213> Artificial Sequence	
30	<220> <223> Engineered target site	
	<400> 61 ttgaggaggt ttctcggtgg tg	22
35	<210> 62 <211> 900 <212> DNA <213> Artificial Sequence	
40	<220> <223> Nucleotide Sequence of a I-CpaMI homing endonuclease; ORF, codon optimized for mammalian expression	
45	<400> 62 atgaacacca gctctagctt caatccctgg ttcttgaccg gctttagcga tgcagagtgc	60
	tctttcagca tcctgatata ggccaacagc aagtactcca ccggttgag gatcaagccc	120
50	gtgttcgcca tcggcttgca caagaaggac ctggagcttc tgaagagaat ccagagctat	180
	ctgggctgg gcaagataca cattcacggc aaagacagca ttcagttcag gattgacagc	240
	cccaaggagc tggaggtgat catcaaccac tttgagaact acccctggt aaccgccaag	300
55	tgggcccact acaccctctt taagaaggcc ctggacgtaa ttctgttgaa ggagcacctg	360
	agccagaagg gcctgcttaa actggtaggc attaaggcga gcctgaatct cgggttgaaac	420

EP 3 272 356 A1

ggcagcctca aggaggcgtt cccgaactgg gaagaactgc agatcgacag gccgagctac 480
 gtgttcaagg gcatccccga cccaactgg atcagcggct tcgctcagg cgatagcagc 540
 5 tttaatgtga aaatcagcaa ctccccacg tcaactgtca ataaaagggt gcagctgagg 600
 ttcggcacg gactgaacat cagagagaaa gcccttatcc aatacctggt ggcctacttt 660
 gacctgtcag acaacctgaa gaacatctac ttgcacctga acagcgcacg gttcgagggtg 720
 10 gtgaagttca gcgacatcac cgacaagatc atcccccttct tcgacaagta cagcatacaa 780
 ggcaagaaga gcctggacta catcaacttc aaggaagtgg ccgacattat caagagcaag 840
 aaccatctta ctagcgaggg cttccaggaa atcttggaca tcaaagccag tatgaacaag 900
 15
 <210> 63
 <211> 300
 <212> PRT
 <213> Cryphonectria parasitica
 20
 <400> 63
 Met Asn Thr Ser Ser Ser Phe Asn Pro Trp Phe Leu Thr Gly Phe Ser
 1 5 10 15
 25 Asp Ala Glu Cys Ser Phe Ser Ile Leu Ile Gln Ala Asn Ser Lys Tyr
 20 25 30
 30 Ser Thr Gly Trp Arg Ile Lys Pro Val Phe Ala Ile Gly Leu His Lys
 35 40 45
 Lys Asp Leu Glu Leu Leu Lys Arg Ile Gln Ser Tyr Leu Gly Val Gly
 35 50 55 60
 Lys Ile His Ile His Gly Lys Asp Ser Ile Gln Phe Arg Ile Asp Ser
 65 70 75 80
 40 Pro Lys Glu Leu Glu Val Ile Ile Asn His Phe Glu Asn Tyr Pro Leu
 85 90 95
 Val Thr Ala Lys Trp Ala Asp Tyr Thr Leu Phe Lys Lys Ala Leu Asp
 45 100 105 110
 Val Ile Leu Leu Lys Glu His Leu Ser Gln Lys Gly Leu Leu Lys Leu
 50 115 120 125
 Val Gly Ile Lys Ala Ser Leu Asn Leu Gly Leu Asn Gly Ser Leu Lys
 130 135 140
 55 Glu Ala Phe Pro Asn Trp Glu Glu Leu Gln Ile Asp Arg Pro Ser Tyr
 145 150 155 160

EP 3 272 356 A1

Val Phe Lys Gly Ile Pro Asp Pro Asn Trp Ile Ser Gly Phe Ala Ser
165 170 175

5 Gly Asp Ser Ser Phe Asn Val Lys Ile Ser Asn Ser Pro Thr Ser Leu
180 185 190

10 Leu Asn Lys Arg Val Gln Leu Arg Phe Gly Ile Gly Leu Asn Ile Arg
195 200 205

Glu Lys Ala Leu Ile Gln Tyr Leu Val Ala Tyr Phe Asp Leu Ser Asp
210 215 220

15 Asn Leu Lys Asn Ile Tyr Phe Asp Leu Asn Ser Ala Arg Phe Glu Val
225 230 235 240

20 Val Lys Phe Ser Asp Ile Thr Asp Lys Ile Ile Pro Phe Phe Asp Lys
245 250 255

25 Tyr Ser Ile Gln Gly Lys Lys Ser Leu Asp Tyr Ile Asn Phe Lys Glu
260 265 270

Val Ala Asp Ile Ile Lys Ser Lys Asn His Leu Thr Ser Glu Gly Phe
275 280 285

30 Gln Glu Ile Leu Asp Ile Lys Ala Ser Met Asn Lys
290 295 300

35 <210> 64
<211> 5264
<212> DNA
<213> Artificial Sequence

40 <220>
<223> Nucleotide sequence of BCL11A gene targeting nuclease-encoding
plasmid pExodusBCL11Ahje

<400> 64
gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg 60
45 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatatc cgttgacatt 240
50 gattattgac tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata 300
tggagttccg cgttacataa cttacggtaa atggcccggc tggctgaccg cccaacgacc 360
65 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 480

EP 3 272 356 A1

	atcatatgcc aagtagcccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	540
	atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca	600
5	tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg	660
	actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc	720
	aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg	780
10	gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca	840
	ctgcttactg gcttatcgaa attaatacga ctcaactatag ggagacccaa gctggctagc	900
	gtttaaactt aagcttggta ccgagctcgg atccactagt ccagtgtggg ggaattctgc	960
15	aggtacgttg acgccgccac catgggatat ccatacgatg tcccagatta tgcgccacct	1020
	aagaagaaac gcaaagtccc cgggggcagc cacatggacc tgacctacgc ctatctggtc	1080
	ggcctgttcg agggcgacgg ctatttttaca atagctaagg ccggcaagta tctgaactac	1140
20	gagctgggca tcacactctc catcaaggac gctcagctca tctacaagat caaggacatc	1200
	ctcggcgtgg gcaacgtgta ctttaggaag tacaggcaac atgagatggg cagcctgcga	1260
	atccaggaca aaaaccacct gaagaacttc atcctgcccc tcttcgacaa gtaccccatg	1320
25	ctgagcaaca agcagtagca ctacctgcga ttcaaggatg ccctcctgtc caacatcatc	1380
	tatagcgacg acctgcccga gtacgccagg agcaacgagt caatcaatag cgtggacagc	1440
30	atcatcaaca cctcatactt cagcgcctgg ctggttggct tcatcgaggc cgagggtgc	1500
	ttcaccacct acaaggcatc caaggacaag tacctgacag cgggcttctc catagcccag	1560
	aaggacggcg acattctcat ctccgcgatc cacaaatacc tgagcttcac gaccaaaacc	1620
35	tacaaagaca agaccaactg tagccacctc aagggtcaccg gcgtgaggag cgtcaataac	1680
	gtggttaagt tcatccaggg tgcgccggtc aagctgctgg gtaacaagaa gctgcagtag	1740
	aaactttgga taaagcagct gcgcaagatc tcccgatata gcgagaaaat ccagctgccc	1800
40	agtaactact aatctagagg gcccgtttaa acccgctgat cagcctcgac tgtgccttct	1860
	agttgccagc catctgttgt ttgccctcc cccgtgcctt ccttgacctt ggaaggtgcc	1920
	actcccactg tcctttccta ataaaatgag gaaattgcat cgcattgtct gagtaggtgt	1980
45	cattctattc tggggggtgg ggtggggcag gacagcaagg gggaggattg ggaagacaat	2040
	agcaggcatg ctgggggatgc ggtgggctct atggcttctg aggcggaaag aaccagctgg	2100
	ggctctaggg ggtatcccca cgcgccctgt agcggcgcat taagcgcggc ggggtgtggtg	2160
50	gttacgcgca gcgtgaccgc tacacttgcc agcgccttag cgcgcgctcc tttcgctttc	2220
	ttcccttcct ttctcgccac gttcgccggc tttcccgctc aagctctaaa tcgggggctc	2280
	cctttagggt tccgatttag tgctttacgg cacctcgacc ccaaaaaact tgattagggt	2340
55	gatggttcac gtagtgggcc atcgccctga tagacggttt ttcgcccttt gacgttggag	2400

EP 3 272 356 A1

	tccacgttct ttaatagtgg actcttgttc caaactggaa caaactcaa ccctatctcg	2460
	gtctattctt ttgatttata agggattttg ccgatttcgg cctattggtt aaaaaatgag	2520
5	ctgatttaac aaaaatttaa cgcgaattaa ttctgtggaa tgtgtgtcag ttaggggtgtg	2580
	gaaagtcccc aggcctccca gcaggcagaa gtatgcaaag cctatcagga catagcggtg	2640
	gctaccctgt atattgctga agagcttggc ggcgaatggg ctgaccgctt cctcgtgctt	2700
10	tacggtatcg ccgctcccga ttccgagcgc atcgcttct atcgcttct tgacgagttc	2760
	ttctgagcgg gactctgggg ttcgaaatga ccgaccaagc gacgccaac ctgccatcac	2820
	gagatttcga ttccaccgcc gccttctatg aaagggtggg cttcgggaatc gttttccggg	2880
15	acgccggctg gatgatcctc cagcgcgggg atctcatgct ggagttcttc gccaccacca	2940
	acttgtttat tgcagcttat aatggttaca aataaagcaa tagcatcaca aatttcacaa	3000
20	ataaagcatt tttttcactg cattctagtt gtggtttgtc caaactcatc aatgtatctt	3060
	atcatgtctg tataccgtcg acctctagct agagcttggc gtaatcatgg tcatagctgt	3120
	ttcctgtgtg aaattgttat ccgctcacia ttccacacia catacgagcc ggaagcataa	3180
25	agtgtaaagc ctgggggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac	3240
	tgcccgttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg	3300
	cggggagagg cggtttgcgt attgggcgct cttccgcttc ctgcgtcact gactcgctgc	3360
30	gctcggtcgt tcggtgcgg cgagcggat cagctcactc aaaggcggta atacggttat	3420
	ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca	3480
35	ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc	3540
	atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc	3600
	aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg	3660
40	gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta	3720
	ggtatctcag ttccggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg	3780
	ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac	3840
45	acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag	3900
	gcggtgctac agagtctctt aagtgggtggc ctaactacgg ctacactaga agaacagtat	3960
	ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagtgggt agctcttgat	4020
50	ccggcaaaaca aaccaccgct ggtagcggtt tttttgtttg caagcagcag attacgcgca	4080
	gaaaaaaagg atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga	4140
55	acgaaaactc acgttaaggg attttgggtca tgagattatc aaaaaggatc ttcacctaga	4200
	tccttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag taaacttggg	4260

EP 3 272 356 A1

ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcggt 4320
 catccatagt tgcctgactc cccgtcgtgt agataactac gatacgggag ggcttaccat 4380
 5 ctggccccag tgctgcaatg ataccgcgag acccacgctc accggctcca gatttatcag 4440
 caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgctt 4500
 ccatccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt 4560
 10 tgcgcaacgt tggtgccatt gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg 4620
 cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatcccc atgttgtgca 4680
 aaaaagcggg tagctccttc ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt 4740
 15 tatcactcat gggttatggca gcaactgcata attctcttac tgtcatgcca tccgtaagat 4800
 gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac 4860
 cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa 4920
 20 aagtgtcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt 4980
 tgagatccag ttcgatgtaa cccactcgtg caccacactg atcttcagca tcttttactt 5040
 tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa 5100
 25 gggcgacacg gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt 5160
 atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa 5220
 30 taggggttcc gcgcacattt ccccgaaaag tgccacctga cgtc 5264

 <210> 65
 <211> 5100
 <212> DNA
 35 <213> Artificial Sequence

 <220>
 <223> Nucleotide sequence of TREX2-encoding plasmid pExodus CMV.Trex2

 <400> 65
 40 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg 60
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggctcgt gagtagtgcg 120
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
 45 ttagggttag gcgttttgcg ctgcttcgag atgtacgggc cagatatacg cgttgacatt 240
 gattattgac tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata 300
 tggagttccg cgttacataa cttacggtaa atggcccgcg tggtgaccg cccaacgacc 360
 50 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 480
 55 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
 atgccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca 600

EP 3 272 356 A1

	tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg	660
	actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc	720
5	aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg	780
	gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaacca	840
	ctgcttactg gcttatcgaa attaatacga ctcaactatag ggagacccaa gctggctagc	900
10	atgtctgagc cacctcgggc tgagaccttt gtattcctgg acctagaagc cactgggctc	960
	ccaaacatgg accctgagat tgcagagata tccctttttg ctgttcaccg ctcttccctg	1020
	gagaaccacg aacgggatga ttctggttcc ttggtgctgc cccgtgttct ggacaagctc	1080
15	acactgtgca tgtgcccgga gcgccccttt actgccaagg ccagtgagat tactggtttg	1140
	agcagcgaaa gcctgatgca ctgcgggaag gctggtttca atggcgctgt ggtaaggaca	1200
20	ctgcagggct tcctaagccg ccaggagggc cccatctgcc ttgtggccca caatggcttc	1260
	gattatgact tcccactgct gtgcacggag ctacaacgtc tgggtgccca tctgccccaa	1320
	gacactgtct gcctggacac actgcctgca ttgcgggggc tggaccgtgc tcacagccac	1380
25	ggcaccaggg ctcaaggccg caaaagctac agcctggcca gtctcttcca cgcctacttc	1440
	caggctgaac ccagtgtctc ccattcagca gaaggatgat tgcacaccct gcttctgatc	1500
	ttcctgcatc gtgctcctga gctgctcgcc tgggcagatg agcaggcccg cagctgggct	1560
30	catattgagc ccatgtacgt gccacctgat ggtccaagcc tcgaagcctg aattctgcag	1620
	atatccagca cagtggcggc cgctcgagtc tagagggccc gtttaaaccg gctgatcagc	1680
35	ctcgactgtg ccttctagtt gccagccatc tgttgtttgc ccctcccccg tgccttcctt	1740
	gaccctggaa ggtgccactc ccactgtcct ttccctaataa aatgaggaaa ttgcatcgca	1800
	ttgtctgagt aggtgtcatt ctattctggg ggggtggggtg gggcaggaca gcaaggggga	1860
40	ggattgggaa gacaatagca ggcatgctgg ggatgcgggtg ggctctatgg cttctgaggc	1920
	ggaaagaacc agctggggct ctagggggta tccccacgcg ccctgtagcg gcgcattaag	1980
	cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccctagcgcc	2040
45	cgctcctttc gctttcttcc cttcctttct cgccacgttc gccggctttc cccgtcaagc	2100
	tctaaatcgg gggctccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa	2160
	aaaacttgat tagggatgat gttcacgtag tgggccatcg ccctgataga cggtttttcg	2220
50	ccctttgacg ttggagtcca cgttctttta tagtggactc ttgttccaaa ctggaacaac	2280
	actcaaccct atctcgggtct attcttttga tttataaggg attttgccga tttcggccta	2340
55	ttggttaaaa aatgagctga tttaacaaaa atttaacgcg aattaattct gtggaatgtg	2400
	tgtcagttag ggtgtggaaa gtccccaggc tccccagcag gcagaagtat gcaaagccta	2460

EP 3 272 356 A1

	tcaggacata gcgttggcta cccgtgatat tgctgaagag cttggcggcg aatgggctga	2520
	ccgcttcctc gtgctttacg gtatcgccgc tcccgattcg cagcgcatcg ccttctatcg	2580
5	ccttcttgac gagttcttct gagcgggact ctgggggttcg aaatgaccga ccaagcgacg	2640
	cccaacctgc catcacgaga tttcgattcc accgcgcct tctatgaaag gttgggcttc	2700
	ggaatcgttt tccgggacgc cggctggatg atcctccagc gcggggatct catgctggag	2760
10	ttcttcgccc accccaactt gtttattgca gcttataatg gttacaaata aagcaatagc	2820
	atcacaaatt tcacaaataa agcatttttt tcaactgcatt ctagttgtgg tttgtccaaa	2880
	ctcatcaatg tatcttatca tgtctgtata ccgtcgacct ctagctagag cttggcgtaa	2940
15	tcattggtcat agctgtttcc tgtgtgaaat tggttatccgc tcacaattcc acacaacata	3000
	cgagccggaa gcataaagt taaagcctgg ggtgcctaata gagtgagcta actcacatta	3060
	attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgcgtgccca gctgcattaa	3120
20	tgaatcggcc aacgcgcggg gagaggcggg ttgcgtattg ggcgtcttc cgcttcctcg	3180
	ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc tcaactcaaag	3240
	gcggtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat gtgagcaaaa	3300
25	ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc	3360
	cgccccctcg acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaaccgcaca	3420
	ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc tcctgttccg	3480
30	accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgtttct	3540
	catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt	3600
35	gtgcacgaac ccccggttca gcccgaccgc tgcgccttat ccggtaaacta tcgtcttgag	3660
	tccaacccgg taagacacga cttatcgcca ctggcagcag ccaactggtaa caggattagc	3720
	agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac	3780
40	actagaagaa cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga	3840
	gttggtagct cttgatccgg caaacaacc accgctggta gcgggttttt tgtttgcaag	3900
	cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg	3960
45	tctgacgctc agtggacga aaactcacgt taagggattt tggatcatgag attatcaaaa	4020
	aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata	4080
	tatgagtaaa cttggtctga cagttaccaa tgcttaatac gtgaggcacc tatctcagcg	4140
50	atctgtctat ttcgttcac catagttgcc tgactccccg tcgtgtagat aactacgata	4200
	cgggagggct taccatctgg cccagtgct gcaatgatac cgcgagacc acgctcaccg	4260
	gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgag aagtggctct	4320
55	gcaactttat ccgcctccat ccagtctatt aattgttgcc ggaagctag agtaagtagt	4380

EP 3 272 356 A1

	tcgccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt ggtgtcacgc	4440
	tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga	4500
5	tcccccatgt tgtgcaaaaa agcggtagc tccttcggtc ctccgatcgt tgtcagaagt	4560
	aagttggccg cagtgttata actcatgggt atggcagcac tgcataattc tcttactgtc	4620
	atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa	4680
10	tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtaa tacgggataa taccgcgcca	4740
	catagcagaa ctttaaaagt gctcatcatt ggaaaacggt cttcggggcg aaaactctca	4800
	aggatcttac cgctgttgag atccagttcg atgtaaccca ctctgtcacc caactgatct	4860
15	tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc	4920
	gcaaaaaagg gaataaggcg gacacggaaa tgttgaatac tcatactctt cttttttcaa	4980
20	tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt	5040
	tagaaaaata aacaaatagg ggttccgcgc acatttccc gaaaagtgcc acctgacgtc	5100
25	<210> 66 <211> 20 <212> DNA <213> Artificial Sequence	
30	<220> <223> PCR Primer	
	<400> 66 gctggaatgg ttgcagtaac	20
35	<210> 67 <211> 20 <212> DNA <213> Artificial Sequence	
40	<220> <223> PCR Primer	
	<400> 67 caaacagcca ttcaccagtg	20
45	<210> 68 <211> 21 <212> DNA <213> Artificial Sequence	
50	<220> <223> PCR Primer	
	<400> 68 ctgccagctc tctaagtctc c	21
55	<210> 69	

<211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR Primer

<400> 69
 tgcaacacgc acagaacact c

21

Claims

1. A composition comprising a CRISPR endonuclease for use in the treatment of a hemoglobinopathy, wherein said CRISPR endonuclease comprises a Cas9 endonuclease and one or more RNA guide strands that bind to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.
2. The composition for use according to Claim 1, wherein said one or more RNA guide strands mediates the binding of said Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region.
3. The composition for use according to Claim 2, wherein said one or more RNA guide strands comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.
4. The composition for use according to Claim 3, wherein said Cas9 endonuclease is encoded by the nucleotide sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 endonuclease.
5. The composition for use according to any of Claims 1-4, wherein the hemoglobinopathy is β -thalassemia or sickle cell disease.
6. A polynucleotide encoding an RNA guide strand that mediates the binding of a Cas9 endonuclease to a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.
7. The polynucleotide of Claim 6, wherein said RNA guide strand mediates the specific binding of a Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region.
8. The polynucleotide of Claim 7, wherein said fetal hemoglobin (HbF) silencing region comprises the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 2.
9. The polynucleotide of Claim 8, wherein said RNA guide strand comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.
10. A vector system comprising a vector that comprises a polynucleotide according to any of Claims 6-9.
11. A cell comprising: a polynucleotide encoding Cas9 endonuclease, and one or more RNA guide strands that bind to a nucleotide sequence selected from the group consisting of a Bcl11a coding region, a Bcl11a gene regulatory region, an adult human β -globin locus, a fetal hemoglobin (HbF) silencing region, a Bcl11a-regulated HbF silencing region, a γ -globin gene promoter, a δ -globin gene promoter, and a site of a β -globin gene mutation.
12. The cell of Claim 11, wherein said cell is a hematopoietic stem cell (HSC), an induced pluripotent stem cell (iPSC), or an erythroid progenitor cell.
13. The cell of Claim 11, wherein said one or more RNA guide strands mediates the binding of said Cas9 endonuclease

to a fetal hemoglobin (HbF) silencing region.

- 5 **14.** The cell of Claim 13, wherein said one or more RNA guide strands comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, and SEQ ID NO: 54.

- 10 **15.** The cell of any of Claims 11-14, wherein said Cas9 endonuclease is encoded by the nucleotide sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 endonuclease.

10

15

20

25

30

35

40

45

50

55

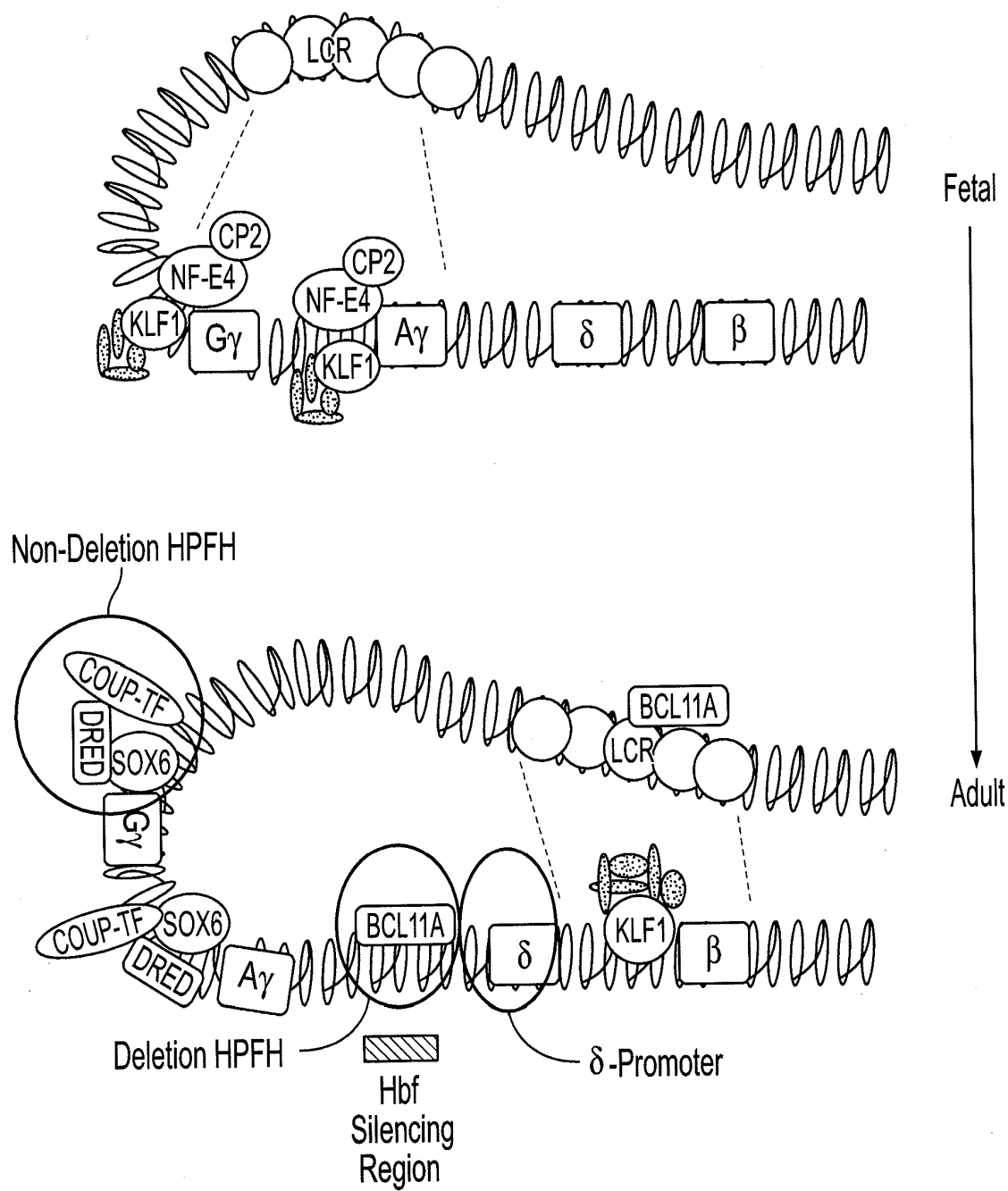


FIG. 1

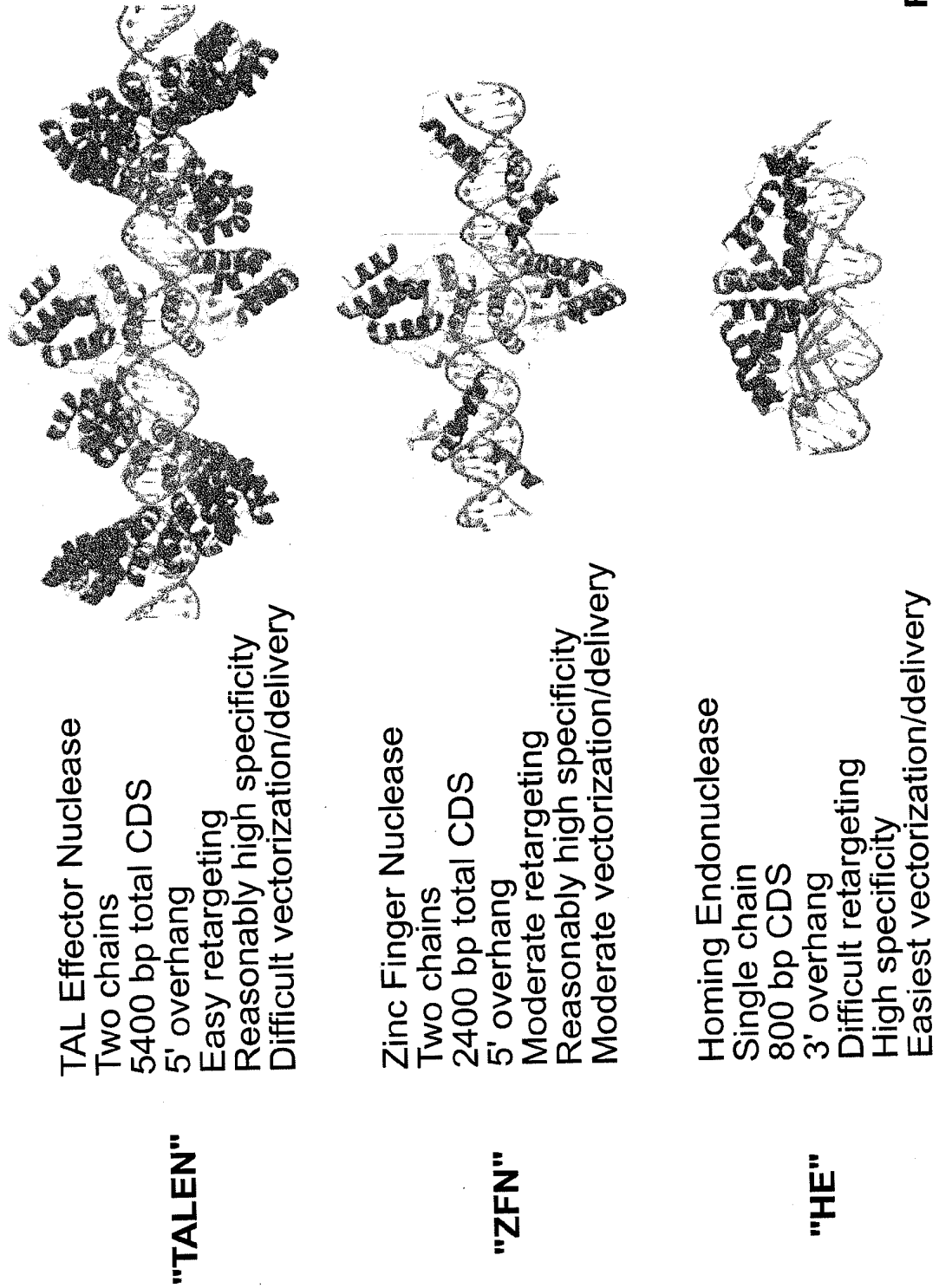
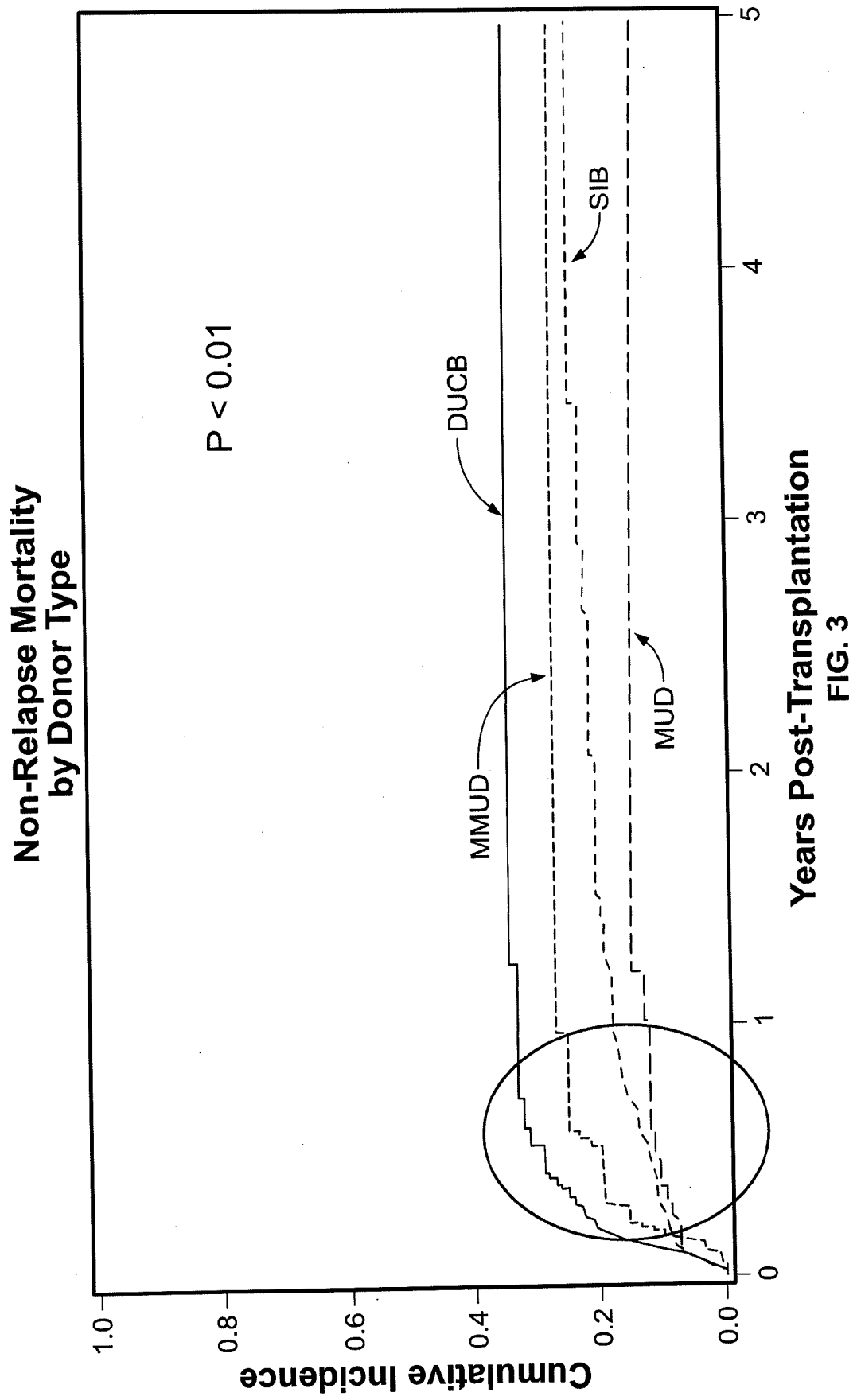


FIG. 2



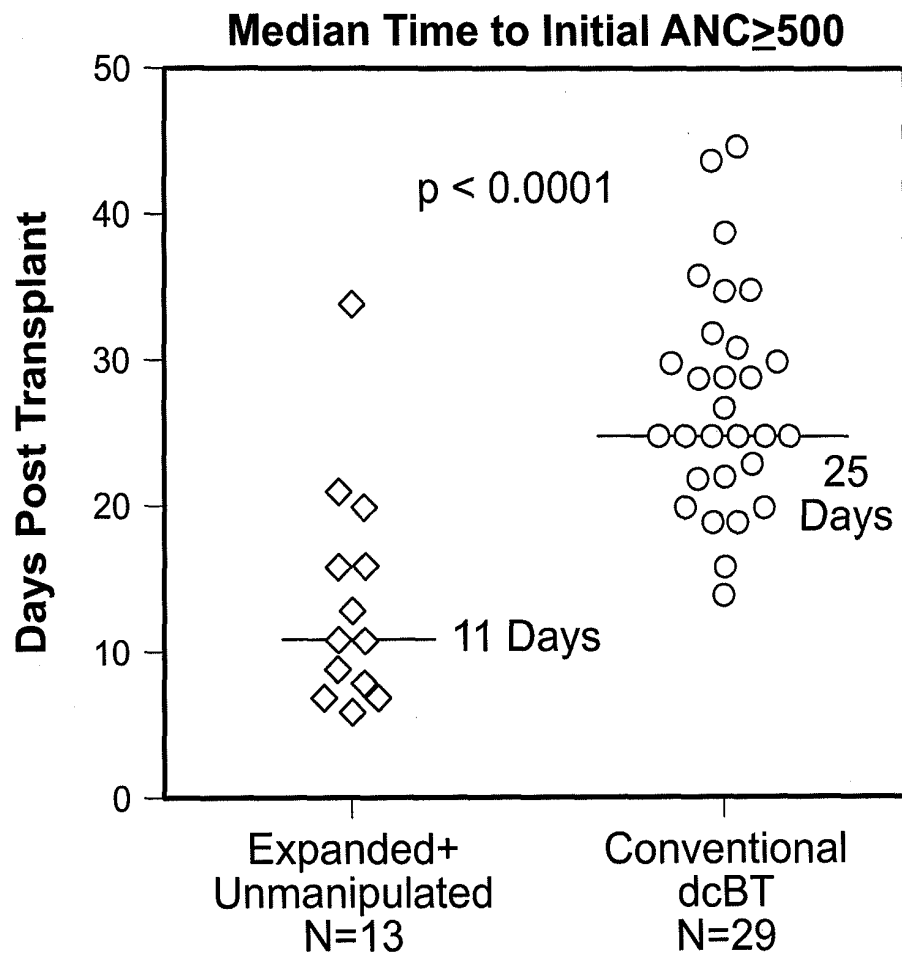
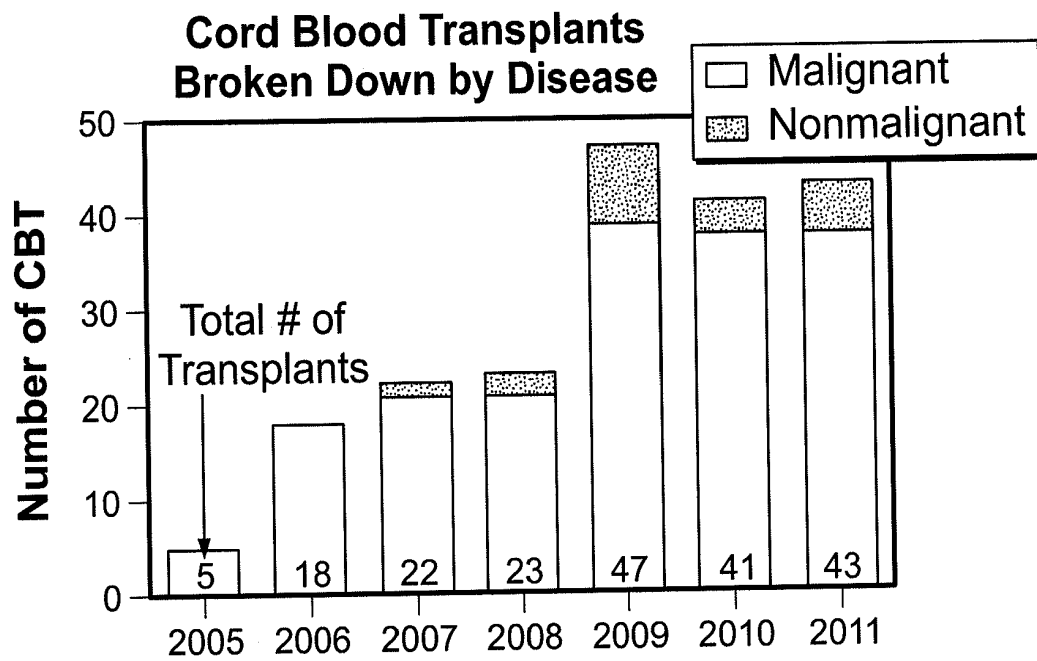


FIG. 4

**FIG. 5**

(SEQ ID NO: 1)

Nucleotide Sequence of a Human Fetal Hemoglobin (HbF) Silencing Region
(chr11:5212342-5215944 in HG18)

1	CCAGTGAGCAGGTTGGTTTAAGATAAGCAGGGTTTCATTAGTTTGTGAGA	50
51	ATGAAAAATGAACCTTCATTCCACTATTCCTTAACCTGCCCTGAGATTG	100
101	GCTGTTCTGTCATGTGTGTCTTGACTCAGAAACCCTGTTCTCCTCTACAT	150
151	ATCTCCCCACCGCATCTCTTTCAGCAGTTGTTTCTAAAAATATCCTCCTA	200
201	GTTTCATTTTTGCAGAAGTGTGTTTAGGCTAATATAGTGAATGTATCTTA	250
251	GAGTTTAACTTATTTGTTTCTGTCACTTTATACTAAGAAAACCTATCTAA	300
301	AAGCAGATGTTTTAACAAGTTGACTCAATATAAAGTTCTTCTTTGCCTCT	350
351	AGAGATTTTTGTCTCCAAGGGAATTTTGAGAGGTTGGAATGGACAAATCT	400
401	ATTGCTGCAGTTTAACTTGCTTGCTTCCTCCTTCTTTTGGTAAATTCTT	450
451	CCTATAATAAAACTCtaatTTTTtattatattgaaataaatatccattaa	500
501	aagaatattttaaaaaatGAATAGTGTTTATTTACCAGTTATTGAAATAGG	550
551	TTCTGGAAACATGAATTTTAAGGTTAACATTTTAATGACAGATAAAATCA	600
601	AATATTATATACAAATATTTTGAATGTTTAAAATTATGGTATGACTAAAG	650
651	AAAGAATGCAAAGTGAAAAGTAGATTTACCATATTCAGCCAGATTAAATT	700
701	TAACGAAGTTCCTGGGAATATGCTAGTACAGAACATTTTTACAGATGTGT	750
751	TCTTAAAAAAAATGTGGAATTAGACCCAGGAATGAAGATCCAGTAGTT	800
801	TTTCACTCTTTTCTGAATTCAAATAATGCCACAATGGCAGACAAATACAC	850
851	ACCCATGAGCATATCCAAaaggaaggattgaaggaaaggaggaggaaagaaa	900
901	tggagaaaggaaggaaggaagaggggaagagagaggatggaagggatgga	950
951	ggagaagaaggaaAAATAAATAATGgagaggagaggagaaaaaaggaggg	1000
1001	gagaggagaggagaagggatagggaagagaaagagaaaggggaagggaga	1050
1051	gaggaaagaagagaagaggagagaaaaagaaacgaagagaggggaagggaa	1100
1101	ggaaaaaaaagaggaaaaaagagacaagagaagagaTaagactgacagtt	1150
1151	caaattttggtggtgatatggatcaatagaaactcaaactctgttggtga	1200
1201	cactgtacaatagtataacccctttggaaaacctttaatagtatccacaa	1250
1251	atgctggatgcttgataagtcattacctagcaattacatttttagatat	1300

FIG. 6

1301	tcagaaacacatgcatgtgtgtatccaaagacatgtatagaaatgcttat	1350
1351	gacagcaataatcataaaacctcaaaccggtagccacttaaagcttac	1400
1401	caacagtagaattgataaattacggtatagtcaaagaatagaatattaca	1450
1451	cagaaatgaaaagaatcaactactgcttaacacgtagcgatacaaatgca	1500
1501	ttttacagcatttggttgattaaaagtaaccagaggtgagttcaaactat	1550
1551	atgacttttatttgatatatagAAAGATGGATGATGTGCCTGAGATTCTGAT	1600
1601	CACAAGGGGAAATGTTATAAAATAGGGTAGAGAGGAGCCATGAATGACCT	1650
1651	TTAAACTTTGTTACAAGTTATTTTTCTGTAACCTggaagccaacgaaaga	1700
1701	tattgaataattcaagaaaggtggtggcatggtttgatttggtgtctttaa	1750
1751	aagattattctcacttagtgaagaaatgtattttagaagtagagaaaatg	1800
1801	ggagacaaatagctgggcttctgttgacagtaGGGAAGAAAGTGACAATGC	1850
1851	CATTTCTATTATCAGACTTGGACCATGACGGTGATGTCAGTCGTGAACAC	1900
1901	AAGAATAGGGCCACATTTGTGAGTTTAGTGGTACGATaaaatcagaaata	1950
1951	cagtcttggtacattgtattgtatgcactcttgtaaaatgcaaaaagat	2000
2001	gtacttagatatgtggatctggagctcagAAAGAATACAACCAGGTCAAG	2050
2051	AATACAGAATGGAACAGAACATACAAGAACAGATCATAATGTGCTGTGTG	2100
2101	AATCACTACCACTACCTGTTAAAAATGACAGATGATGTACTTCATcaata	2150
2151	tctccttaaaatcttagaatgtgtttgtgagggaggaattatgtttccaa	2200
2201	ttcatatataagaaaattgattctaaaaaaaatgttaggtaaattcttaa	2250
2251	ggccaTGAGGACTGTTATTTGATCTTTGTCTGTTAATTCCAAAGACTTGG	2300
2301	CTTTTCACTTTAATTCTGTTCTACCTGAAATGATTTTACACATTGGGAGA	2350
2351	TCTGGTTACATGTTTATTCTATATGGATTGCATTGAGAGGATTTGTATAA	2400
2401	CAGAATAAGGTCTtttttttcttttcttcttgagatggagtttcatcct	2450
2451	attgccaagctagagtgcaatggtgcaatctaggctcaccgcaacctct	2500
2501	gcctcctgggttcaagcaattctcctgcctcagccacctgaatagctggg	2550
2551	actgcaggcatgcaccacacgcccggctgattttgtatttttagtagaga	2600
2601	tggggtttcaccatgttggtcaggctggtccttgaactcctgacctcaagt	2650
2651	gatctgcctgccttggcctcccaaagtgtgggtttacaagcctgagcca	2700
2701	ccgcatccagccAGGATAAGGTCTAAAAGTGGAAAGAATAGCATCTACTC	2750

FIG. 6 (Cont.)

2751	TTGTTTCAGGAAACAATGAGGACCTGACTGGGCAGTAAGAGTGGTGATTAA	2800
2801	TAGATAGGGACAAATTGAAGCAGAATCGAACTGTTGATTAgaggtaggga	2850
2851	aatgatttttaatctgtgaccttgggtgaatgggcaagtagctatctaata	2900
2901	ctaaaatggaaaacactggaagagaaaacagtttttagtataacaagtga	2950
2951	tacccatgctgagtctgaggtgcctataggacatctatataaataagccc	3000
3001	agtacattgtttgatatatgggtttggcactgaggttggaggtcagaggt	3050
3051	tagaaatcagagttgggaattgggattatacaggctgtatttaagagttt	3100
3101	agatataactgtgaatccaagagtgtgaTGAATACAAAGTTAAATGAAGG	3150
3151	ACCTTTAATGAACACCAACATTTAATGTGAAATCTCAAGGAAGTATGAAG	3200
3201	TAAGACATAGTCCCCAAAATCCCCGATGATTTTAGAACTCAGTATCGATT	3250
3251	TTAATTAGTGTAATGCCAATGTGGGTTAGAATGGAAGTCAACTTGCTGTT	3300
3301	GGTTTCAGAGCAGGTAGGAGATAAGGTTCTAGATTTTGACACAGTGAAAA	3350
3351	GCTGAAACAAAAAGGAAAAGGTAGGGTGAAAGATGGGAAATGTATGTAAG	3400
3401	GAGGATGAGCCACATGGTATGGGAGGTATACTAAGGACTCTAGGGTCAGA	3450
3451	GAAATATGGGTTATATCCTTCTACAAAATTCACATTCTTggctgggtgtg	3500
3501	gtggctcacgcctgtgatcccagcactttcagaggccgaggaggggtggat	3550
3551	cacctgatgttaggagttcgagatcagcctgaccaacatggtgaaacccc	3600
3601	cta	3603

FIG. 6 (Cont.)

(SEQ ID NO: 2)

Nucleotide Sequence of a 350 bp Portion of a Human HbF Silencing Region
(Including Sites for Bcl11a Occupancy)

1	AAAGATGGATGATGTGCCTGAGATTCTGATCACAAGGGGAAATGTTATAA	50
51	AATAGGGTAGAGAGGAGCCATGAATGACCTTTAACTTTGTTACAAGTTA	100
101	TTTTTCTGTAACTTggaagccaacgaaagatatgaataattcaagaaag	150
151	gtggtggcatggtttgatttggtgtctttaaaagattattctcacttagtg	200
201	aagaaatgtattttagaagtagagaaaatgggagacaaatagctgggctt	250
251	ctgttgcagtaGGGAAGAAAGTGACAATGCCATTTCTATTATCAGACTTG	300
301	GACCATGACGGTGATGTCAGTCGTGAACACAAGAATAGGGCCACATTTGT	350

FIG. 7

(SEQ ID NO: 13)

Human b-globin Gene from 1 kb Upstream of the
Cap Site through the PolyA Site

1	GCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCC	50
51	CTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACC	100
101	TTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTAGTGATACTTGTGGGC	150
151	CAGGGCATTAGCCACACCAGCCACCACTTTCTGATAGGCAGCCTGCACTG	200
201	GTGGGGTGAATTCCTTGCCAAAGTGATGGGCCAGCACACAGACCAGCACG	250
251	TTGCCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGCAA	300
301	AAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAA	350
351	TAAAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACC	400
401	TCTTACATCAGTTACAATTTATATGCAGAAATATTTATATGCAGAGATAT	450
451	TGCTATTGCCTTAACCCAGAAATTATCACTGTTATTCTTTAGAATGGTGC	500
501	AAAGAGGCATGATACATTGTATCATTATTGCCCTGAAAGAAAGAGATTAG	550
551	GGAAAGTATTAGAAATAAGATAAACAAAAAAGTATATTAAGAAGAAAG	600
601	CATTTTTTAAATTACAAATGCAAaattaccctgatttggtcaatatgtg	650
651	tacacatattaaaacattacactttaaccataaatatgtataatgatta	700
701	tgtatcaattaaaaataaaaGAAAATAAAGTAGGGAGATTATGAATATGC	750
751	AAATAAGCACACATATATTCCAAATAGTAATGTACTAGGCAGACTGTGTA	800
801	AAGTTTTTTTTTAAGTTACTTAATGTATCTCAGAGATATTTCTTTTGT	850
851	ATACACAATGTTAAGGCATTAAGTATAATAGTAAAAATTGCGgagaagaa	900
901	aaaaaaagaaagcaagaattaaacaaaagaaaacaattggttatgaacagc	950
951	aaataaaagaaactaaaaCGATCCTGAGACTTCCACACTGATGCAATCAT	1000
1001	TCGTCTGTTTCCATTCTAAACTGTACCCTGTTACTTATCCCCTTCCTAT	1050
1051	GACATGAACCTAACCATAGAAAAGAAGGGGAAAGAAAACATCAAGCGTCC	1100
1101	CATAGACTCACCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCACAGTG	1150
1151	CAGCTCACTCAGTGTGGCAAAGGTGCCCTTGAGGTTGTCCAGGTGAGCCA	1200
1201	GGCCATCACTAAAGGCACCGAGCACTTTCTTGCCATGAGCCTTCACCTTA	1250
1251	GGGTTGCCCATACAGCATCAGGAGTGGACAGATCCCCAAAGGACTCAAA	1300

FIG. 8

1301	GAACCTCTGGGTCCAAGGGTAGACCACCAGCAGCCTAAGGGTGGGAAAAT	1350
1351	AGACCAATAGGCAGAGAGAGTCACTGCCTATCAGAAACCCAAGAGTCTTC	1400
1401	TCTGTCTCCACATGCCCAGTTTCTATTGGTCTCCTTAAACCTGTCTTGTA	1450
1451	ACCTTGATACCAACCTGCCCAGGGCCTCACCACCAACTTCATCCACGTTC	1500
1501	ACCTTGCCCCACAGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAGATG	1550
1551	CACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGC	1600
1601	AAATGTAAGCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGC	1650
1651	TCCTGCCCTCCCTGCTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGC	1700
1701	TCCACAGGGTGAGGTCTAAGTGATGACAGCCGTACCTGTCTTGCTCTT	1750
1751	CTGGCACTGGCTTAGGAGTTGGACTTCAAACCCTCAGCCCTCCCTCTAAG	1800
1801	ATATATCTCTTGCCCCATACCATCAGTACAAATTGCTACTAAAAACATC	1850
1851	CTCCTTTGCAAGTGTATTTACGTAATATTTGGAAtcacagcttggttaagc	1900
1901	atattgaagatcgttttcccaattttcttattacacaaataagaagttga	1950
1951	tgcactaaaagtggaagagttttgtctaCCATAATTCAGCTTTGGGATAT	2000
2001	GTAGATGGATCTCTTCCTGCGTCTCCAGAATATGCAAATACTTACAGGA	2050
2051	CAGAATGGATGAAAACCTCTACCTCGGTTCTAAGCATATCTTCTCCTTATT	2100
2101	TGGATTAACCTTCTGGTAAGAAAAGAAAAAatatatatatatatgtgt	2150
2151	gtatatatacacacatacatatacatatatatGCATTCATTTGTTGTTGT	2200
2201	TTTTCTTAATTTGCTCATGCATGCTAATAAATTATGTCTAAAAATAGAAT	2250
2251	AAATACAAATCAATGTGCTCTGTGCATTAGTTACTTATTAGGTTTTGGGA	2300
2301	AACAAGAGATAAAAAACTAGAGACCTCTTAATGCAGTCAAAAATACAAAT	2350
2351	AAATAAAAAGTCACTTACAACCCAAAGTGTGACTATCAATGGGGTAATCA	2400
2401	GTGGTGTCAAATAGGAGGTAACTGGGGACATCTAACTGTTTCTGCCTGG	2450
2451	ACTAATCTGCAAGAGTGTCTGGGGGAACAAAAAGCCTCTGTGACTTAGAA	2500
2501	AGTAGGGGTAGGAGGGGAAAAGGTCTTCTACTTGGCTCAGATTATTTTTT	2550
2551	TCCTCTAGTCCACTAAGAATACTGCGTTTTAAATCATTTCTTGATTCA	2600
2601	AGTTCC	2606

FIG. 8 (Cont.)

(SEQ ID NO: 14)

606 bp Region of Human β -globin Spanning from the Promoter into Intron 2
(Mutations within this Region Result in Thalassemias and Sickle Cell Disease)

```

1  CACCCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCACAGTGCAGCTCAC   50
51 TCAGTGTGGCAAAGGTGCCCTTGAGGTTGTCCAGGTGAGCCAGGCCATCA  100
101 CTAAAGGCACCGAGCACTTTCTTGCCATGAGCCTTCACCTTAGGGTTGCC  150
151 CATAACAGCATCAGGAGTGGACAGATCCCCAAAGGACTCAAAGAACCTCT  200
201 GGGTCCAAGGGTAGACCACCAGCAGCCTAAGGGTGGGAAAATAGACCAAT  250
251 AGGCAGAGAGAGTCAGTGCCTATCAGAAACCCAAGAGTCTTCTCTGTCTC  300
301 CACATGCCCAGTTTCTATTGGTCTCCTTAAACCTGTCTTGTAACCTTGAT  350
351 ACCAACCTGCCCAGGGCCTCACCACCAACTTCATCCACGTTACCTTGCC  400
401 CCACAGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAGATGCACCATGG  450
451 TGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAAATGTAA  500
501 GCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGCTCCTGCCC  550
551 TCCCTGCTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGCTCCACAGG  600
601 GTGAGG                                                    606

```

FIG. 9

(SEQ ID NO: 24)

Nucleotide Sequence of Human Bcl11a cDNA (CCDS1862.1)

1	ATGTCTCGCCGCAAGCAAGGCAAACCCCAGCACTTAAGCAAACGGAATT	50
51	CTCGCCCGAGCCTCTTGAAGCCATTCTTACAGATGATGAACCAGACCACG	100
101	GCCCGTTGGGAGCTCCAGAAGGGGATCATGACCTCCTCACCTGTGGGCAG	150
151	TGCCAGATGAACTTCCCATTGGGGGACATTCTTATTTTTATCGAGCACAA	200
201	ACGGAAACAATGCAATGGCAGCCTCTGCTTAGAAAAAGCTGTGGATAAGC	250
251	CACCTTCCCCTTACCAATCGAGATGAAAAAAGCATCCAATCCCGTGGAG	300
301	GTTGGCATCCAGGTCACGCCAGAGGATGACGATTGTTTATCAACGTCATC	350
351	TAGAGGAATTTGCCCCAAACAGGAACACATAGCAGATAAACTTCTGCACT	400
401	GGAGGGGCCTCTCCTCCCCTCGTTCTGCACATGGAGCTCTAATCCCCACG	450
451	CCTGGGATGAGTGCAGAATATGCCCCGCAGGGTATTTGTAAAGATGAGCC	500
501	CAGCAGCTACACATGTACAACTTGCAAACAGCCATTACCAGTGCATGGT	550
551	TTCTCTTGCAACACGCACAGAACACTCATGGATTAAGAATCTACTTAGAA	600
601	AGCGAACACGGAAGTCCCCTGACCCCGCGGGTTGGTATCCCTTCAGGACT	650
651	AGGTGCAGAATGTCTTCCCAGCCACCTCTCCATGGGATTCATATTGCAG	700
701	ACAATAACCCCTTTAACCTGCTAAGAATACCAGGATCAGTATCGAGAGAG	750
751	GCTTCCGGCCTGGCAGAAGGGCGCTTTCACCCACTCCCCCCTGTTTAG	800
801	TCCACCACCGAGACATCACTTGACCCCCACCGCATAGAGCGCCTGGGGG	850
851	CGGAAGAGATGGCCCTGGCCACCCATCACCCGAGTGCCTTTGACAGGGTG	900
901	CTGCGGTTGAATCCAATGGCTATGGAGCCTCCCGCCATGGATTTCTCTAG	950
951	GAGACTTAGAGAGCTGGCAGGGAACACGTCTAGCCCACCGCTGTCCCCAG	1000
1001	GCCGGCCCAGCCCTATGCAAAGGTTACTGCAACCATTCCAGCCAGGTAGC	1050
1051	AAGCCGCCCTTCTGCGACGCCCCCCTCCCTCCTCTGCAATCCGCCCC	1100
1101	TCCTCCCTCCCAGCCCCCGGTCAAGTCCAAGTCATGCGAGTTCTGCGGCA	1150
1151	AGACGTTCAAATTTTCAAGCAACCTGGTGGTGCACCGGCGCAGCCACACG	1200
1201	GGCGAGAAGCCCTACAAGTGCAACCTGTGCGACCACGCGTGCACCCAGGC	1250
1251	CAGCAAGCTGAAGCGCCACATGAAGACGCACATGCACAAATCGTCCCCCA	1300
1301	TGACGGTCAAGTCCGACGACGGTCTCTCCACCGCCAGCTCCCCGGAACCC	1350

FIG. 10

1351	GGCACCAGCGACTTGGTGGGCAGCGCCAGCAGCGCGCTCAAGTCCGTGGT	1400
1401	GGCCAAGTTCAAGAGCGAGAACGACCCCAACCTGATCCCGGAGAACGGGG	1450
1451	ACGAGGAGGAAGAGGAGGACGACGAGGAAGAGGAAGAAGAGGAGGAAGAG	1500
1501	GAGGAGGAGGAGCTGACGGAGAGCGAGAGGGTGGACTACGGCTTCGGGCT	1550
1551	GAGCCTGGAGGCGGCGCGCCACCACGAGAACAGCTCGCGGGGCGCGGTCTG	1600
1601	TGGGCGTGGGCGACGAGAGCCGCGCCCTGCCCGACGTCATGCAGGGCATG	1650
1651	GTGCTCAGCTCCATGCAGCACTTCAGCGAGGCCTTCCACCAGGTCTCTGGG	1700
1701	CGAGAAGCATAAGCGCGGCCACCTGGCCGAGGCCGAGGGCCACAGGGACA	1750
1751	CTTGCGACGAAGACTCGGTGGCCGGCGAGTCGGACCGCATAGACGATGGC	1800
1801	ACTGTTAATGGCCGCGGCTGCTCCCCGGGCGAGTCGGCCTCGGGGGGCCT	1850
1851	GTCCAAAAGCTGCTGCTGGGCAGCCCCAGCTCGCTGAGCCCCTTCTCTA	1900
1901	AGCGCATCAAGCTCGAGAAGGAGTTCGACCTGCCCCGGCCGCGATGCCC	1950
1951	AACACGGAGAACGTGTACTCGCAGTGGCTCGCCGGCTACGCGGCCTCCAG	2000
2001	GCAGCTCAAAGATCCCTTCCTTAGCTTCGGAGACTCCAGACAATCGCCTT	2050
2051	TTGCCTCCTCGTCGGAGCACTCCTCGGAGAACGGGAGTTTGCGCTTCTCC	2100
2101	ACACCGCCCGGGGAGCTGGACGGAGGGATCTCGGGGCGCAGCGGCACGGG	2150
2151	AAGTGGAGGGAGCACGCCCCATATTAGTGGTCCGGGCCCCGGGCAGGCCCA	2200
2201	GCTCAAAAGAGGGCAGACGCGAGCGACACTTGTGAGTACTGTGGGAAAGTC	2250
2251	TTCAAGAACTGTAGCAATCTCACTGTCCACAGGAGAAGCCACACGGGCGA	2300
2301	AAGGCCTTATAAATGCGAGCTGTGCAACTATGCCTGTGCCCAGAGTAGCA	2350
2351	AGCTCACCAGGCACATGAAAACGCATGGCCAGGTGGGGAAGGACGTTTAC	2400
2401	AAATGTGAAATTTGTAAGATGCCTTTTAGCGTGACAGTACCCTGGAGAA	2450
2451	ACACATGAAAAAATGGCACAGTGATCGAGTGTTGAATAATGATATAAAAA	2500
2501	CTGAATAG	2508

FIG. 10 (Cont.)

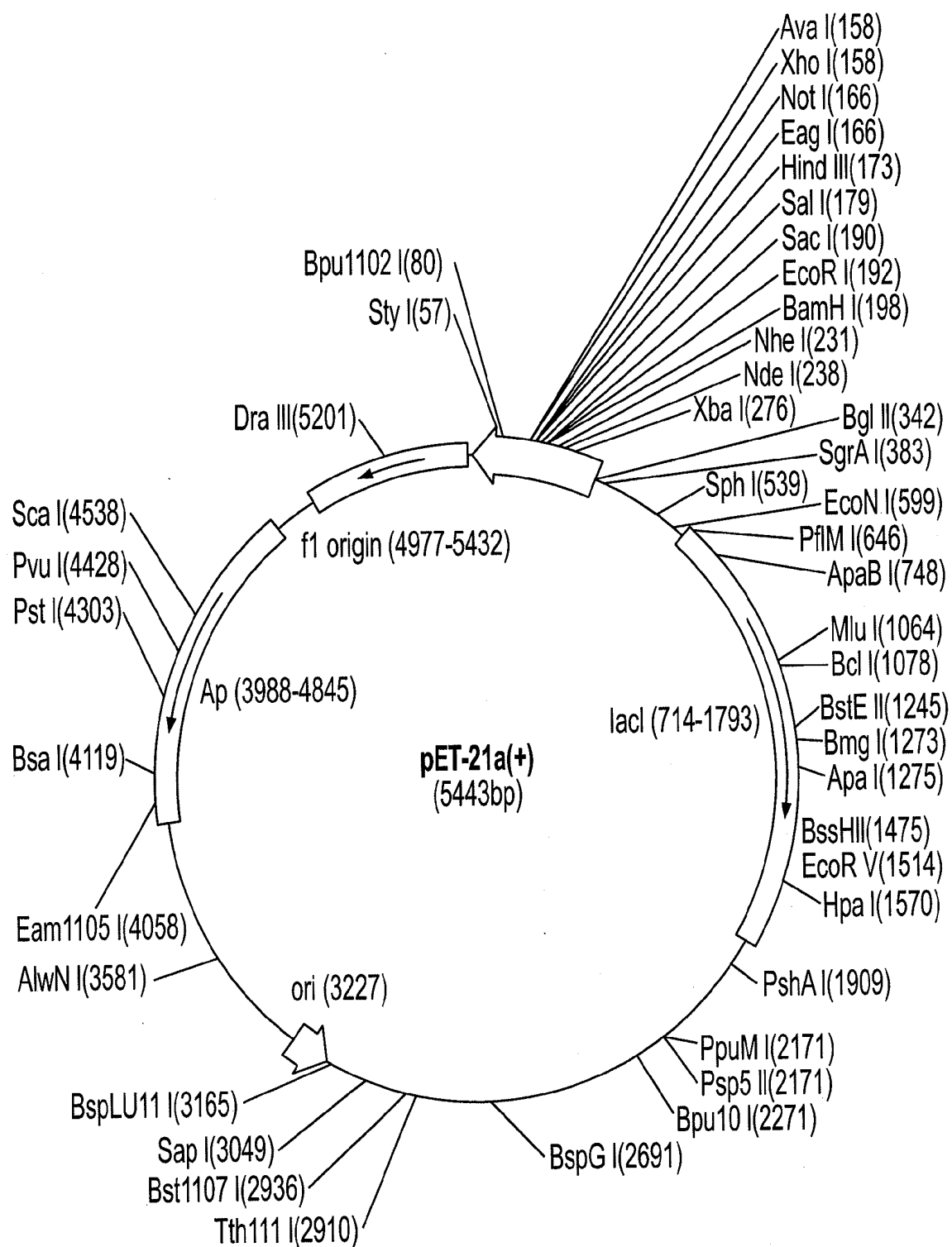


FIG. 11

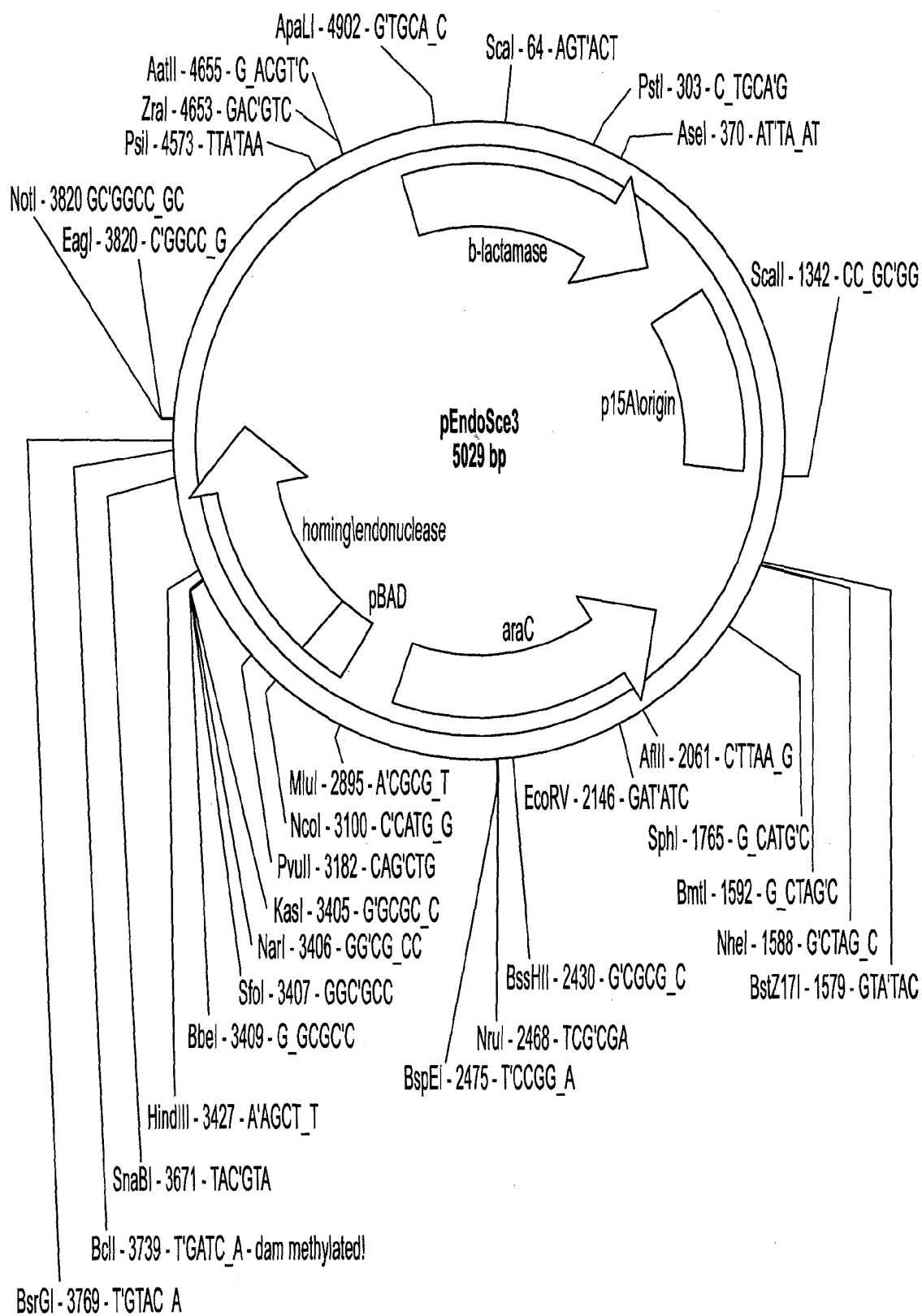


FIG. 12

Directed Evolution for Creating a BCL11A Gene-targeting Endonuclease

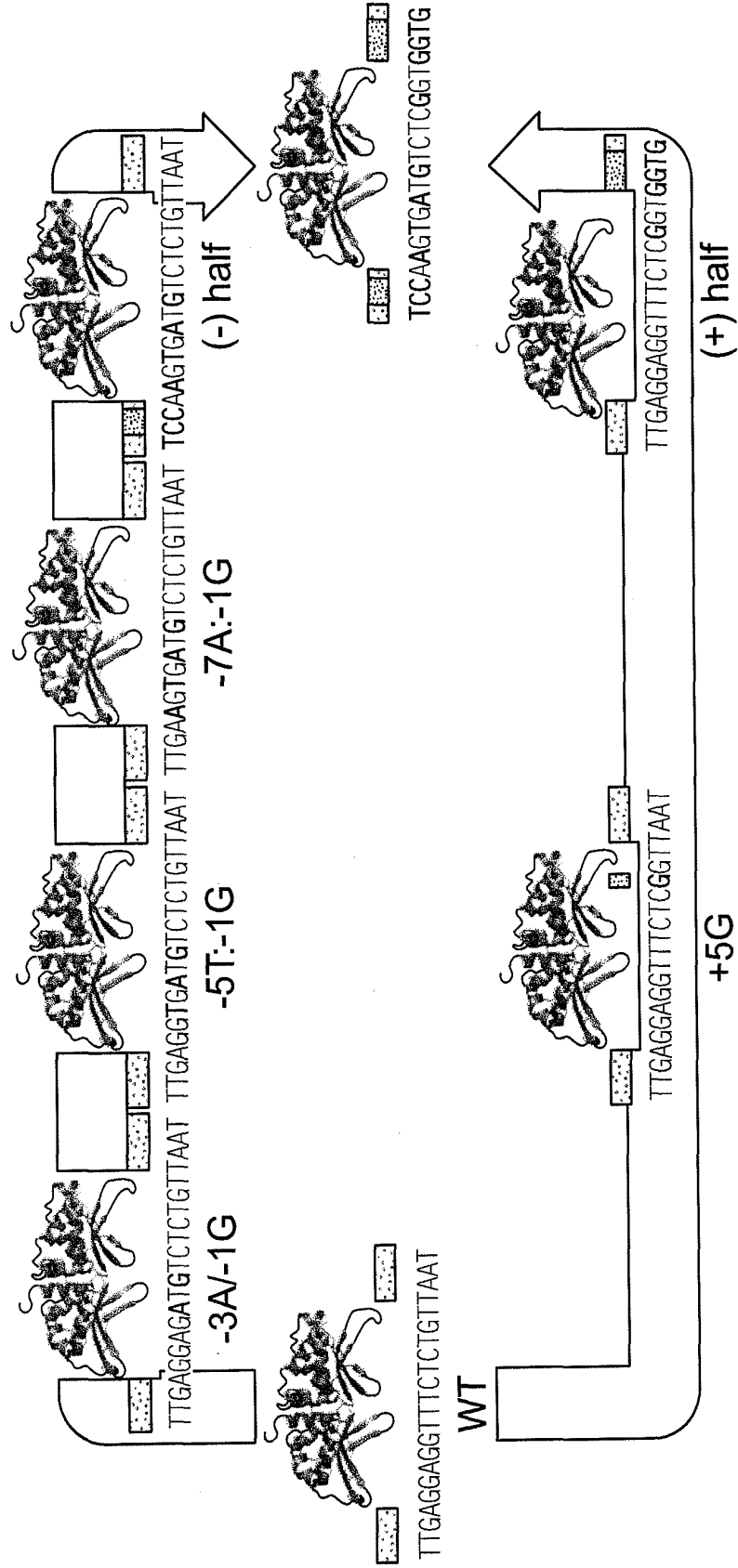


FIG. 13

(SEQ ID NO: 28)

Nucleotide Sequence of I-HjeMI, Codon
Optimized for Expression in E.coli

```

1  ATGGGATCCACATGGACCTGACCTACGCTTACCTGGTTGGTCTGTTCGA   50
51 AGGTGACGGTTACTTCTCTATCACCAAAAAGGGTAAATACCTGACCTACG  100
101 AACTGGGTATCGAACTGTCTATCAAAGACGTTCACTGATCTACAAAATC  150
151 AAAGACATCCTGGGTGTTGGTAAAGTTTCTTTCCGTAAACGTAACGAAAT  200
201 CGAAATGGTTTCTCTGCGTATCCGTGACAAGAATCACCTGAAAACTTCA  250
251 TCCTGCCGATCTTCGACAAATACCCGATGCTGTCTAACAAGCAGTACGAC  300
301 TACCTGCGTTTCAAAGACGCTCTCCTGTCTAACATTATCTACTCTGACGA  350
351 TCTGCCGGAATACGCTCGTTCTAACGAATCTATCAACTCTGTTGACTCTA  400
401 TTATCAACACCTCTTACTTCTCTGCTTGGCTGGTTGGTTTCATCGAAGCT  450
451 GAAGGTTGCTTCTCTACCTACAAACTGAACAAAGATGACGATTACCTGAT  500
501 CGCTTCTTTTCGACATCGCTCAGAAAGACGGTGACATCCTGATCTCTGCTA  550
551 TCCACAAATACCTGTCTTTCACCACGAAAATCTACCTGGACAAAACCAAC  600
601 TGCTCTCGTCTGAAAGTGACCGGTGTACGTTCTGTAAAAACGTGGTTAA  650
651 ATTCATCCAGGGTGCTCCGGTTAAACTGCTCGGTAACAAGAAACTGCAGT  700
701 ACAAACTGTGGATCAAACAGCTGCGTAAAATCTCTCGTTACTCTGAAAAA  750
751 ATCCAGCTGCCGTCTAACTAC                                     771

```

FIG. 14

(SEQ ID NO: 29)

Nucleotide Sequence of I-HjeMI, Codon Optimized for
Expression in Mammals

1	ATGGGCAGCCACATGGACCTGACCTACGCCTATCTGGTCGGCCTGTTCGA	50
51	GGGCGACGGCTATTTTAGCATAACCAAGAAGGGCAAGTATCTGACGTATG	100
101	AACTGGGCATCGAGCTCTCCATCAAGGACGTGCAGCTCATCTACAAGATC	150
151	AAGGACATCCTCGGCGTGGGCAAAGTGTCTTTTAGGAAGAGGAACGAGAT	200
201	CGAGATGGTCAGCCTGCGAATCAGGGACAAAAACCACTGAAGAACTTCA	250
251	TCCTGCCCATCTTCGACAAGTACCCCATGCTGAGCAACAAGCAGTACGAC	300
301	TATCTCCGATTCAAGGATGCCCTCCTGTCCAACATCATCTATAGCGACGA	350
351	CCTGCCCCGAGTACGCCAGGAGCAACGAGTCAATCAATAGCGTGGACAGCA	400
401	TCATCAACACCTCATACTTCAGCGCCTGGCTGGTTGGCTTCATCGAGGCC	450
451	GAGGGCTGCTTCAGCACCTACAAGCTCAACAAGGACGACGATTATTTGAT	500
501	CGCGAGCTTCGATATAGCCCAGAAGGACGGCGACATTCTCATCTCCGCGA	550
551	TCCACAAATACCTGAGCTTCACGACCAAAATCTACCTGGACAAGACCAAC	600
601	TGTAGCAGGCTCAAGGTCACCGGCGTGAGGAGCGTCAAGAACGTGGTTAA	650
651	GTTTCATCCAGGGTGCGCCGGTCAAGTTGCTGGGTAACAAGAAGCTGCAGT	700
701	ACAAACTTTGGATAAAGCAGCTGCGCAAGATCTCCCGATACAGCGAGAAA	750
751	ATCCAGCTGCCCAGTAACTAC	771

FIG. 15

(SEQ ID NO: 30)

Amino Acid Sequence of the Homing Endonuclease I-HjeMI

1	MGDLTYAYLVGLFEGDGYFSITKKGKYLTYELGIELSIKDVQLIYKIKDI	50
51	LGVGKVSFRKRNEIEMVSLRIRDKNHLKNFILPIFDKYPMLSNKQYDYLR	100
101	FKDALLSNIISDDLPEYARSNESINSVDSIINTSYFSAWLVGFI EAEGC	150
151	FSTYKLNKDDDYLIASF DIAQKGDGILISAIHKYLSFTTKIYLDKTNCSR	200
201	LKVTGVR SVKNVVKFIQGAPVKLLGNKKLQYKLWIKQLRKISRYSEKIQ L	250
251	PSNY	254

FIG. 16

(SEQ ID NO: 31)

Nucleotide Sequence for a BCL11A Gene Targeting Nuclease
Based on the Homing Endonuclease I-HjeMI
(Codon Optimized for Expression in E.coli and Obtained
through Directed Evolution in IVC and in Bacteria)

```

1  ATGGGATCCACATGGACCTGACCTACGCTTACCTGGTTGGTCTGTTCGA   50
51  AGGTGACGGTTACTTCACTATCGCTAAAGCGGGTAAGTACCTGAATTACG  100
101 AGCTGGGTATCACGCTGTCTATCAAAGACGCTCAGCTGATCTACAAAATC  150
151 AAAGACATCCTGGGTGTTGGTAATGTTTATTTCCGAAATATAGGCAGCA  200
201 TGAAATGGTTTCTCTGCGGATCCAGGACAAGAATCACCTGAAAACTTCA  250
251 TCCTGCCGATCTTCGACAAATACCCGATGCTGTCTAACAAGCAGTACGAC  300
301 TACCTGCGTTTCAAAGACGCTCTCCTGTCTAACATTATCTACTCTGACGA  350
351 TCTGCCGGAATACGCTCGTTCTAACGAATCTATCAACTCTGTTGACTCTA  400
401 TTATCAACACCTCTTACTTCTCTGCTTGGCTGGTTGGTTTCATCGAAGCT  450
451 GAAGGTTGCTTCACGACCTACAAAGCGAGTAAAGATAAGTACCTGACGGC  500
501 TGGGTTCAAGTATCGCTCAGAAAGACGGTGACATCCTGATCTCTGCTATCC  550
551 ACAAATACCTGTCTTTCACCACGAAACCGTACAAGGACAAAACCAACTGC  600
601 TCTCATCTGAAGGTGACCGGTGTACGTTCTGTTAACAACGTGGTTAAATT  650
651 CATCCAGGGTGCTCCGGTTAAACTGCTCGGTAACAAGAAACTGCAGTACA  700
701 AACTGTGGATCAAACAGCTGCGTAAAATCTCTCGTTACTCTGAAAAAATC  750
751 CAGCTGCCGTCTAACTAC                                     768

```

FIG. 17

(SEQ ID NO: 32)

Nucleotide Sequence of a BCL11A Gene Targeting Nuclease
Based on the Homing Endonuclease I-HjeMI

(Codon Optimized for Expression in Mammals and Obtained
through Directed Evolution in IVC and in Bacteria)

```

1  ATGGGCAGCCACATGGACCTGACCTACGCCTATCTGGTCGGCCTGTTCGA   50
51  GGGCGACGGCTATTTTAcaATAgCtAAGgccGGCAAGTATCTGAacTACG   100
101 AGCTGGGCATCacaCTCTCCATCAAGGACGctCAGCTCATCTACAAGATC   150
151 AAGGACATCCTCGGCGTGGGCAAcGTGtaCTTTAGGAAGtacAggcAaca   200
201 tGAGATGGTCAGCCTGCGAATCcaGGACAAAAACCACCTGAAGAACTTCA   250
251 TCCTGCCCATCTTCGACAAGTACCCCATGCTGAGCAACAAGCAGTACGAC   300
301 TACCTGCGATTCAAGGATGCCCTCCTGTCCAACATCATCTATAGCGACGA   350
351 CCTGCCCAGTACGCCAGGAGCAACGAGTCAATCAATAGCGTGGACAGCA   400
401 TCATCAACACCTCATACTTCAGCGCCTGGCTGGTTGGCTTCATCGAGGCC   450
451 GAGGGCTGCTTCAcCACCTACAAGgcatcCAAGGACaAgTACCTGAcaGC   500
501 GgGCTTctccATAGCCCAGAAGGACGGCGACATTCTCATCTCCGCGATCC   550
551 ACAAATACCTGAGCTTCACGACCAAAccCTACaaaGACAAGACCAACTGT   600
601 AGCcacCTCAAGGTCACCGGCGTGAGGAGCGTCAAtAACGTGGTTAAGTT   650
651 CATCCAGGGTGCGCCGGTCAAGCTGCTGGGTAAACAAGAAGCTGCAGTACA   700
701 AACTTTGGATAAAGCAGCTGCGCAAGATCTCCCGATACAGCGAGAAAATC   750
751 CAGCTGCCCAGTAACFAC                                         768

```

FIG. 18

(SEQ ID NO: 33)

**Amino Acid Sequence of a BCL11A Gene Targeting
Nuclease Based on the Homing Endonuclease I-HjeMI**

1	MGSHMDLTYAYLVGLFEGDGYFTIAKAGKYLNyelGITLSIKDAQLIYKI	50
51	KDILGVGNVYFRKYRQHEMVSLRIQDKNHLKNFILPIFDKYPMLSNKQYD	100
101	YLRFKDALLSNIISDDLPEYARSNESINSVDSIINTSYFSAWLVGFI EA	150
151	EGCFTTYKASKDKYLTAGFSIAQKGDILISAIHKYLSFTTKPYKDKTNC	200
201	SHLKVTGVRSVNNVVKFIQGAPVKLLGNKKLQYKLWIKQLRKISRYSEKI	250
251	QLPSNY	256

FIG. 19

Distribution of Amino-acid Residues Different between
the BCL11A Gene-targeting Endonuclease and its
Parental LHE I-HjeM1

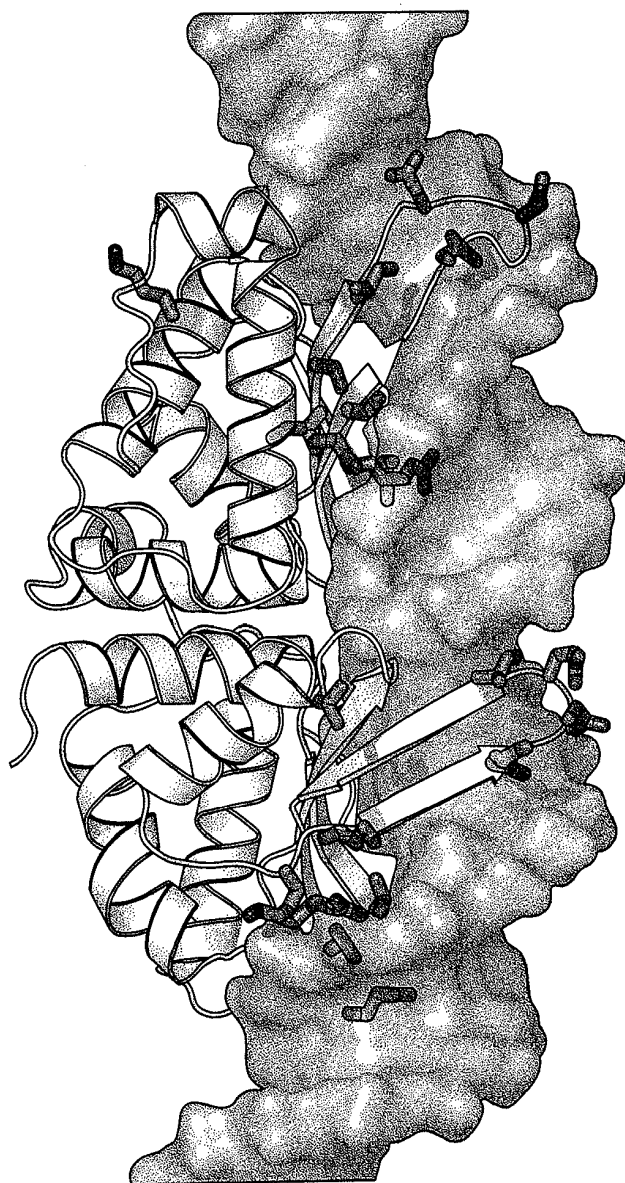
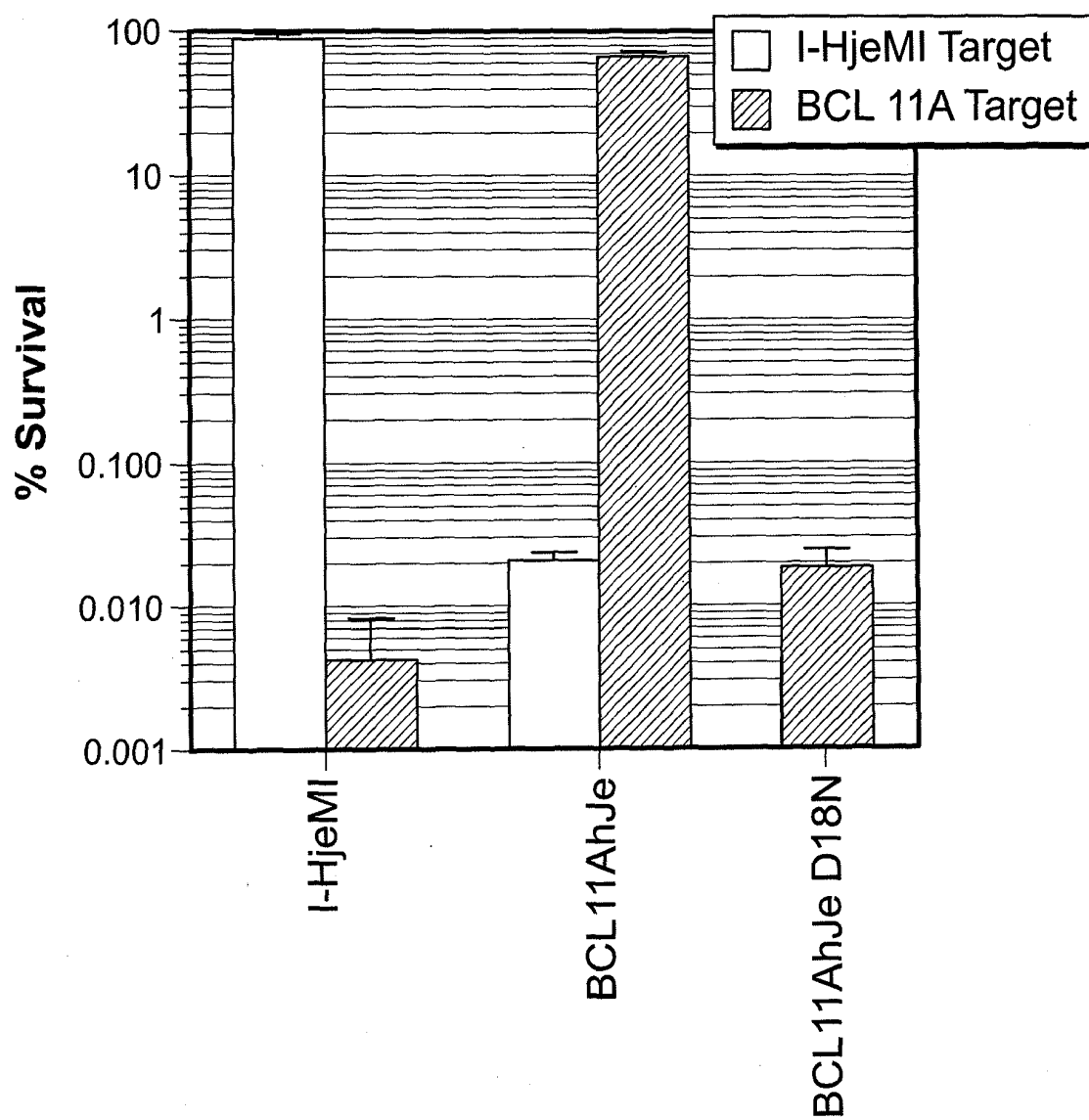


FIG. 20

Activity of a BCL11A Gene-targeting Endonuclease in a Two-plasmid Cleavage Assay

**FIG. 21**

(SEQ ID NO: 34)

Nucleotide Sequence of I-OnuI, Codon Optimized for Expression in *E. coli*
(parental enzyme for HEs targeting the HbF silencing)

```

1  ATGTCCGCCTACATGTCCCGTCGCGAGTCCATTAACCCGTGGATTCTCAC   50
51 CGGTTTCGCCGACGCGGAAGGCTCCTTTTTGCTGCGCATCCGCAACAACA  100
101 ACAAGTCCAGCGTCGGCTACTCCACTGAGCTCGGCTTCCAAATTACACTT  150
151 CATAACAAGGACAAGAGCATTCTTGAGAACATCCAGTCAACATGGAAGGT  200
201 GGGCGTGATCGCCAACAGCGGTGACAACGCCGTGTCGCTGAAGGTCACGC  250
251 GTTTTGAGGACCTGAAGGTCATTATCGACCATTTTGAAAAATACCCACTG  300
301 ATTACGCAGAAGCTCGGTGACTACATGCTGTTTAAGCAGGCGTTTTGCGT  350
351 CATGGAGAACAAGGAGCATTTGAAGATTAATGGTATCAAGGAGCTGGTGC  400
401 GCATTAAGGCCAAAGCTCAATTGGGGTCTGACGGATGAGCTGAAGAAGGCC  450
451 TTTCCGGAGATCATCTCGAAGGAGCGCTCCCTCATCAACAAGAACATCCC  500
501 TAATTTCAAGTGGCTGGCGGGTTTTACCTCGGGCGAGGGTTGCTTCTTTG  550
551 TTAACCTGATCAAGTCAAAGTCGAAGCTAGGTGTCCAGGTGCAGCTGGTG  600
601 TTCAGCATTACCCAACACATCAAGGATAAGAACCTCATGAACTCTCTGAT  650
651 TACCTACTTGGGCTGCGGCTACATTAAGGAGAAAAACAAGAGTGAGTTCT  700
701 CCTGGCTTGACTTCGTCGTCACGAAATTCTCCGACATCAACGACAAGATC  750
751 ATTCCGGTCTTTTCAAGGAAAACACGCTCATCGGCGTGAAGCTCGAGGACTT  800
801 CGAGGATTGGTGTAAGGTCGCTAAGCTGATCGAGGAGAAAAAGCACCTGA  850
851 CAGAAAGTGGCCTGGACGAGATCAAGAAGATTAAGCTGAACATGAACAAG  900
901 GGCAGAGTATTC                                           912

```

FIG. 22A

(SEQ ID NO: 15)

Amino Acid Sequence of I-OnuIhoming endonuclease.

```
SAYMSRRESI NPWILTFAD AEGSFLLRIR NNNKSSVGYS TELGFQITLH NKDKSILENI 60
QSTWKVGVIA NSGDNAVSLK VTRFEDLKVI IDHFEKYPLI TQKLG DYMLF KQAF CVMENK 120
EHLKINGIKE LVRIKAKLNW GLTDELKKAF PEIISKERSL INKNIPNFKW LAGFTSGEGC 180
FFVNLIKSKS KLG VQVQLVF SITQHIKDN LMNSLITYLG CGYIKEKNKS EFSWLDFVVT 240
KFSDINDKII PVFQENTLIG VKLEDFEDWC KVAKLIEEKK HLTESGLDEI KIKLNMNKG 300
RVF 303
```

FIG. 22B

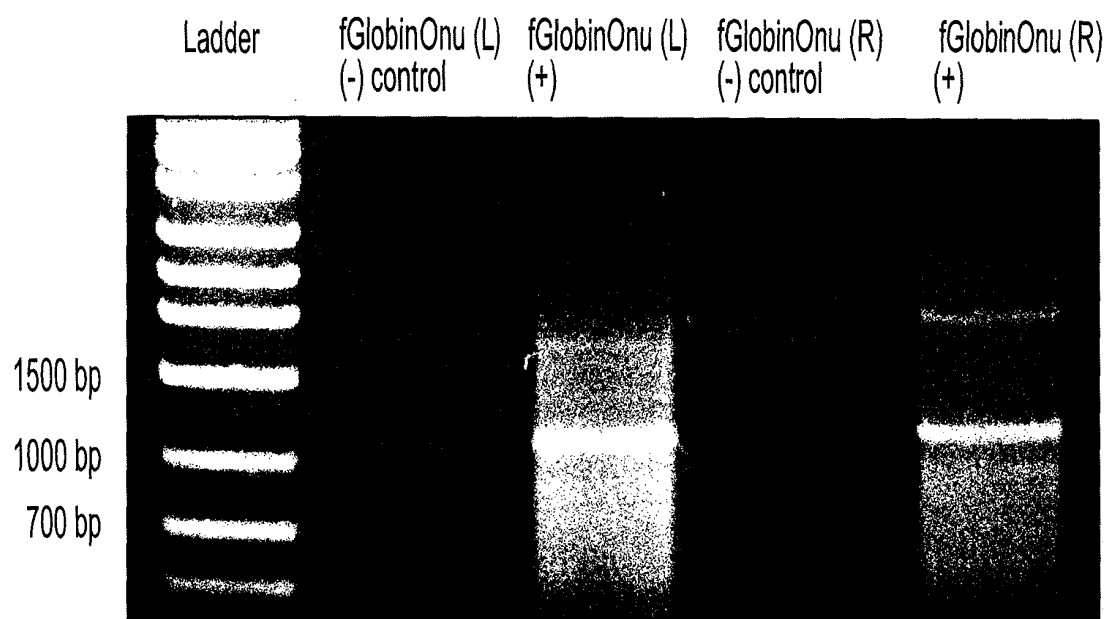


FIG. 23

(SEQ ID NO: 35)

Nucleotide Sequence of MegaTAL:5.5 RVD + Y2 I-AniI

```

1  gtggatctacgcacgctcggctacagtcagcagcagcaagagaagatcaa  50
51 accgaaggtgcgttcgacagtgggcgcagcaccacgaggcactggtgggcc  100
101 atgggtttacacacgcgcacatcggttgcgctcagccaacacccggcagcg  150
151 ttagggaccgctcgctgtcacgtatcagcacataatcacggcgttgccaga  200
201 ggcgacacacgaagacatcggttggcgtcggcaaacagtgggtccggcgcac  250
251 gcgccctggaggccttgctcacggatgcgggggagttgagaggtccgccg  300
301 ttacagttggacacaggccaacttgtgaagattgcaaacgtggcggcgct  350
351 gaccgcaatggaggcagtgcatgcatcgcgcaatgcactgacgggtgccc  400
401 ccctgaacctgaccccggaaccaagtgggtggctatcgccagcaacaatggc  450
451 ggcaagcaagcgctcgaaacgggtgcagcggctggtgccggtgctgtgcc  500
501 ggaccatggcctgactccggaccaagtgggtggctatcgccagccacgatg  550
551 gcggcaagcaagcgctcgaaacgggtgcagcggctggtgccggtgctgtgc  600
601 caggaccatggcctgaccccggaaccaagtgggtggctatcgccagcaacat  650
651 tggcggcaagcaagcgctcgaaacgggtgcagcggctggtgccggtgctgt  700
701 gccaggaccatggcctgaccccggaaccaagtgggtggctatcgccagcaac  750
751 aatggcggcaagcaagcgctcgaaacgggtgcagcggctggtgccggtgct  800
801 gtgccaggaccatggcctgactccggaccaagtgggtggctatcgccagcc  850
851 acgatggcggcaagcaagcgctcgaaacgggtgcagcggctggtgccggtg  900

```

FIG. 24

901 ctgtgccaggaccatggcctgaccccggaaccaagtggaggctatcgccag 950
 951 caacgggtggcggcaagcaagcgctcgaaagcattgtggcccagctgagcc 1000
 1001 ggcctgatccggcggttgccgcgttgaccaacgaccacctcgtcgcccttg 1050
 1051 gcctgcctcggcggacgtcctgccatggatgcagtgaaaaagggattgcc 1100
 1101 gcacgcgccggaattgatcagaagagtcaatcgccgtattggcgaacgca 1150
 1151 cgtcccatcgcggttgcgatatctagagtgggaggaagcgatcttacgtac 1200
 1201 gcgtatttagttgggtctctacgaagggatggatactttagtatcaccaa 1250
 1251 gaaaggcaagtacttgacttatgaattaggtattgagctgagcatcaaag 1300
 1301 acgtccaattgatttacaagatcaagaaaatcctaggtattggcatcgta 1350
 1351 agcttcaggaagagaaaacgagattgaaatggttgattgaggatccgtga 1400
 1401 taagaatcatctaaaatctaagatattgcctatatattgagaagtatccaa 1450
 1451 tgttttccaacaaacagtacgactatttaagattcaggaatgcattgtta 1500
 1501 tctggcattatatacctagaagacttgccctgattacactagaagtgcga 1550
 1551 accattgaattctatagaatccattatcaacacatcatacttctccgcct 1600
 1601 ggctagttggatttatagaagctgagggtgtttcagtgtgtacaagctg 1650
 1651 aacaaagacgatgactacttgattgcttcattcgacattgccccaaagaga 1700
 1701 tggtgatatcttgatttcagcaattaggaagtacttaagtttcactacta 1750
 1751 aggtttacctagacaagactaattgtagcaaattgaaggtcactagtgtt 1800
 1801 agatccgtcgagaacatcattaagtttctgcagaatgctcctgtcaaatt 1850
 1851 gttaggcaacaagaaactgcaatacaagttgtggttgaaacaactaagga 1900
 1901 agatttctaggtattccgagaagatcaagattccatcaaactacgtcgac 1950
 1951 cgagcatcttaccgccatttataccatatttgttctgtttttcttga 1998

FIG. 24 (Cont.)

(SEQ ID NO: 36)

Amino Acid Sequence of MegaTAL:5.5 RVD + Y2 I-AniI

1	VDLRTLGYSSQQQKEIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAA	50
51	LGTVAVTYQHIITALPEATHEDIVGVGKQWSGARALEALLTDAGELRGPP	100
101	LQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTDPQVVAIASNNG	150
151	GKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC	200
201	QDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN	250
251	NGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV	300
301	LCQDHGLTPDQVVAIASNGGKQALESIVAQLSRPDPALAALTNDHLVAL	350
351	ACLGGRPAMD AVKKGLPHAPELIRRVNRRIGERTSHRVAISR VGGS DLT Y	400
401	AYLVGLYEGDGYFSITKKGKYLT YELGIELSIKDVQLIYKIKKILGIGIV	450
451	SFRKRNEIEMVALRIRDKNHLKSKILPIFEKYPMFSNKQYDYLRFRNALL	500
501	SGIIYLEDLPDYTRSDEPLNSIESIINTSYFSAWL VGFIEAEGCFSVYKL	550
551	NKDDDYLIASF DIAQRDGDILISAIRK YLSFTTKVYLDKTNCSK LKVTSV	600
601	RSVENIIKFLQNA PVKLLGNKKLQYKLWLKQLRKISRYSEKIKIPS NYVD	650
651	RASYRHLYPYLFCFS	665

FIG. 25

(SEQ ID NO: 37)

Nucleotide Sequence of Cas9 Endonuclease

(from Mali *et al.*, *Science* (2013))

1	gccaccATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAG	50
51	CGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAAT	100
101	TCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATT	150
151	GGCGCCCTCCTGTTCTGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAA	200
201	AAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACC	250
251	TGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTC	300
301	CATAGGCTGGAGGAGTCCTTTTTTGGTGGAGGAGGATAAAAAGCACGAGCG	350
351	CCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGT	400
401	ACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAG	450
451	GCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTTCG	500
501	GGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCTG	550
551	ACAAACTCTTTATCCAACCTGGTTCAGACTTACAATCAGCTTTTCGAAGAG	600
601	AACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAG	650
651	GCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGG	700
701	AGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTG	750
751	ACCCCCAATTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCA	800
801	ACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGA	850
851	TCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAGAACCTGTCAGAC	900
901	GCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGC	950
951	TCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACT	1000
1001	TGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAG	1050
1051	GAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGG	1100
1101	CGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGAAA	1150
1151	AAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTG	1200
1201	TTGCGCAAACAGCGCACTTTTCGACAATGGAAGCATCCCCACCAGATTCA	1250
1251	CCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCT	1300

FIG. 26

1301	TTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCCGGATA	1350
1351	CCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGAT	1400
1401	GA CTCGCAAATCAGAAGAGACCATCACTCCCTGGAAC TTCGAGGAAGTCG	1450
1451	TGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTT	1500
1501	GATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTA	1550
1551	CGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAG	1600
1601	AAGGGATGAGAAAGCCAGCATTCTGTCTGGAGAGCAGAAGAAAGCTATC	1650
1651	GTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAA	1700
1701	AGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCG	1750
1751	GAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCAGATCTCCTG	1800
1801	AAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT	1850
1851	TCTTGAGGACATTGTCCTCACCCCTTACGTTGTTTGAAGATAGGGAGATGA	1900
1901	TTGAAGAACGCTTGAAAACCTTACGCTCATCTCTTCGACGACAAAGTCATG	1950
1951	AAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAA	2000
2001	ACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAGACAATCCTGGATT	2050
2051	TTCTTAAGTCCGATGGATTTGCCAACCGAACTTCATGCAGTTGATCCAT	2100
2101	GATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGG	2150
2151	CCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAG	2200
2201	CTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTC	2250
2251	AAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCG	2300
2301	AGAGAACCAACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGA	2350
2351	AGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAATCCTTAAGGAA	2400
2401	CACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTA	2450
2451	CCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATC	2500
2501	GGCTCTCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTTCTCAA	2550
2551	GATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAAATAGAGG	2600
2601	GAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATT	2650
2651	ATTGGCGGCAGCTGCTGAACGCCAACTGATCACACAACGGAAGTTTCGAT	2700
2701	AATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGG	2750
2751	CTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGG	2800
2801	CCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAAATGACAAA	2850
2851	CTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGA	2900

FIG. 26 (Cont.)

2901	TTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACC	2950
2951	ACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATC	3000
3001	AAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGT	3050
3051	GTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGG	3100
3101	CCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACC	3150
3151	GAGATTACACTGGCCAATGGAGAGATTGGAAGCGACCACTTATCGAAAC	3200
3201	AAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTGCGCA	3250
3251	CAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACC	3300
3301	GAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAA	3350
3351	CAGCGACAAAGCTGATCGCACGAAAAAAGATTGGGACCCCAAGAAATACG	3400
3401	GCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAA	3450
3451	GTGGAGAAAAGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGG	3500
3501	CATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTC	3550
3551	TCGAGGCGAAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAGCTT	3600
3601	CCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGAAACGAATGCTCGC	3650
3651	TAGTGCGGGCGAGCTGCAGAAAAGTAACGAGCTGGCACTGCCCTCTAAAT	3700
3701	ACGTTAATTTCTTGATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT	3750
3751	CCCGAAGATAATGAGCAGAAGCAGCTGTTGTTGGAACAACACAAACACTA	3800
3801	CCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCC	3850
3851	TCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGG	3900
3901	GATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCT	3950
3951	GACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAG	4000
4001	ACAGAAAGCGGTACACCTCTACAAAGGAGGTCTGGACGCCCACTGATT	4050
4051	CATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCT	4100
4101	CGGTGGAGACAGCAGGGGCTGACCCCAAGAAGAAGAGGAAGGTGTGA	4046

FIG. 26 (Cont.)

(SEQ ID NO: 38)

Nucleotide Sequence of RNA Guide Strand for use with Cas9 Endonuclease
(from Mali *et al.*, *Science* (2013))

```

1  TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGG  50
51  TACCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATT  100
101 TGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACT  150
151 GTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT  200
201 TCTTGGGTAGTTTGCAGTTTTAAATTATGTTTTAAAATGGACTATCATA  250
251 TGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATCTTG  300
301 TGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTA  350
351 GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG  400
401 CACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGG  450
451 CATTA 455

```

FIG. 27

(SEQ ID NO: 62)

Nucleotide Sequence of a I-CpaMI homing endonuclease

(ORF, codon optimized for mammalian expression)

ATGAACACCA GCTCTAGCTT CAATCCCTGG TTCCTGACCG GCTTTAGCGA TGCAGAGTGC 60
 TCTTTCAGCA TCCTGATACA GGCCAACAGC AAGTACTCCA CCGGTTGGAG GATCAAGCCC 120
 GTGTTGCGCA TCGGCTTGCA CAAGAAGGAC CTGGAGCTTC TGAAGAGAAT CCAGAGCTAT 180
 CTGGGCGTGG GCAAGATACA CATTACGGC AAAGACAGCA TTCAGTTCAG GATTGACAGC 240
 CCCAAGGAGC TGGAGGTGAT CATCAACCAC TTTGAGAACT ACCCCCTGGT AACCGCCAAG 300
 TGGGCCGACT ACACCCTCTT TAAGAAGGCC CTGGACGTAA TTCTGTTGAA GGAGCACCTG 360
 AGCCAGAAGG GCCTGCTTAA ACTGGTAGGC ATTAAGGCGA GCCTGAATCT CGGGTTGAAC 420
 GGCAGCCTCA AGGAGGCGTT CCCGAACTGG GAAGAACTGC AGATCGACAG GCCGAGCTAC 480
 GTGTTCAAGG GCATCCCCGA CCCCAACTGG ATCAGCGGCT TCGCGTCAGG CGATAGCAGC 540
 TTTAATGTGA AAATCAGCAA CTCCCCCAGC TCACTGCTCA ATAAAAGGGT GCAGCTGAGG 600
 TTCGGCATCG GACTGAACAT CAGAGAGAAA GCCCTTATCC AATACCTGGT GGCCTACTTT 660
 GACCTGTCAG ACAACCTGAA GAACATCTAC TTCGACCTGA ACAGCGCACG GTTCGAGGTG 720
 GTGAAGTTCA GCGACATCAC CGACAAGATC ATCCCCTTCT TCGACAAGTA CAGCATACAA 780
 GGCAAGAAGA GCCTGGACTA CATCAACTTC AAGGAAGTGG CCGACATTAT CAAGAGCAAG 840
 AACCATCTTA CTAGCGAGGG CTTCCAGGAA ATCTTGACA TCAAAGCCAG TATGAACAAG 900

FIG. 28

(SEQ ID NO: 63)

Amino Acid Sequence of I-CpaMI homing endonuclease

MNTSSSFNPW FLTGFSDAEC SFSILIQANS KYSTGWRIKP VFAIGLHKKD LELLKRIQSY 60
LGVGKIHIHG KDSIQFRIDS PKELEVIINH FENYPLVTAK WADYTLFKKA LDVILLKEHL 120
SQKGLLKLVG IKASLNLGLN GSLKEAFPW EELQIDRPSY VFKGIPDPNW ISGFASGDSS 180
FNVKISNSPT SLLNKRQRL FGIGLNIREK ALIQYLWAYF DLSDNLKNIY FDLNSARFEV 240
VKFSDITDKI IPFFDKYSIQ GKKSLEYINF KEVADIISK NHLTSEGFQE ILDIKASMNK 300

FIG. 29

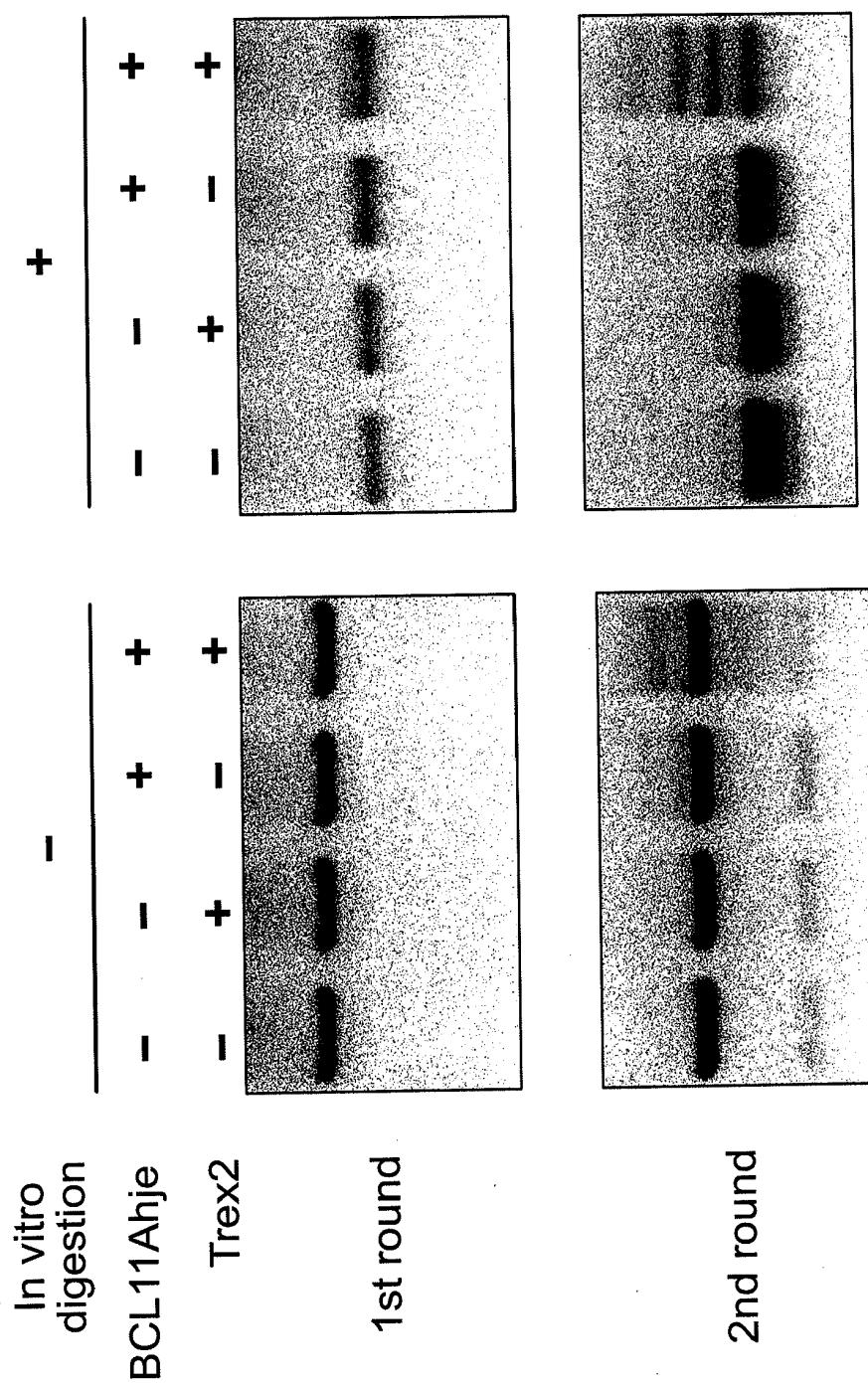


FIG. 30

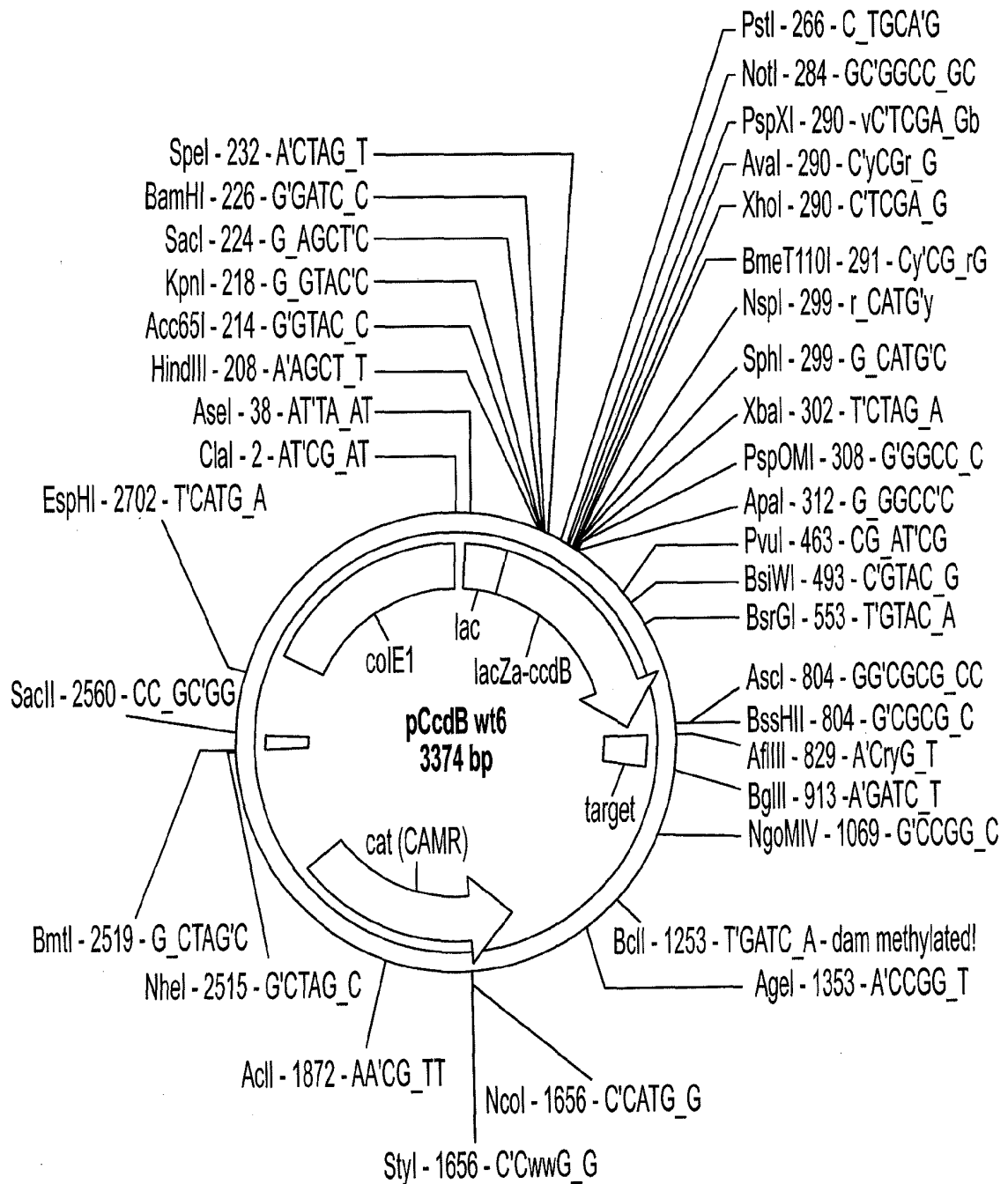


FIG. 31

(SEQ ID NO: 64)

Nucleotide sequence of BCL11A gene targeting nuclease-encoding plasmid

(pExodusBCL11Ahje)

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT
CTGCTCCCTGCTTGTGTGTTGGAGGTGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGA
CAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAA
CTTACGGTAAATGGCCCGCTGGCTGACCGCCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTAGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTT
TGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACC
AAAAATCAACGGGACTTTCCAAAATGTCGTAACTACCGCCCATTTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAG
GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAACGACTCACTATAG
GGAGACCCAAGCTGGCTAGCGTTTAACTTAAGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGGAAATCTGCG
AGGTACGTTGACGCCGCCACCATGGGATATCCATACGATGTCCAGATTATGCGCCACCTAAGAAGAAACGCAAGTCCC
CGGGGGCAGCCACATGGACCTGACCTACGCCATCTGGTGGGCTGTTTCGAGGGCGACGGCTATTTTAcAATAgCtAAGg
ccGGCAAGTATCTGAacTACGAGCTGGGCATCaCaCTCTCCATCAAGGACGctCAGCTCATCTACAAGATCAAGGACATC
CTCGGGCTGGGCAAcGTGtaCTTTAGGAAGtacAggcAacatGAGATGGTCAGCTGCGAATCcaGGACAAAAACCACT
GAAGAATTCATCCTGCCCATCTTCGACAAGTACCCCATGCTGAGCAACAAGCAGTACGACTACCTGCGATTCAAGGATG
CCCTCCTGTCCAACATCATCTATAGCGACGACCTGCCGAGTACGCCAGGAGCAACGAGTCAATCAATAGCGTGGACAGC
ATCATCAACACCTCATACTTCAGCGCCTGGCTGGTTGGCTTCATCGAGGCCGAGGGCTGCTTCAcCACCTACAAGgcat
CAAGGACaAgTACCTGAcAaGCGgGCTTctccATAGCCCAGAAGGACGGCGACATTCTCATCTCCGCGATCCACAAATACC
TGAGCTTCACGACCAAAccCTACaaGACAAGACCAACTGTAGCcacCTCAAGGTCAACGGCGTGAGGAGCGTCAAtAAC
GTGGTTAAGTTCATCCAGGGTGCGCGGTCAAGCTGCTGGGTAAACAAGAAGCTGCAGTACAAACTTTGGATAAAGCAGCT
GCGCAAGATCTCCCGATACAGCGAGAAAATCCAGCTGCCCAGTAACTACTAATCTAGAGGGCCCGTTTAAACCCGCTGAT
CAGCCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTCCTTCTTGACCCTGGAAGGTGCC
ACTCCCACTGTCTTCTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCGATGCTGGGGATGCGGTGGGCTCTATGGCTTCTG
AGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCGCTCCTTTCGCTTCTCTCCCTCTCTTCGCCCAC
GTTTCGCGCGCTTTCGCCGTCAAGCTCTAAATCGGGGGCTCCCTTATAGGTTCCGATTTAGTGCTTTACGGCACCTCGACC
CCAAAAAAGCTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTCGCGTTGGAG
TCCACGTTCTTTAATAGTGGACTCTTGTTCAAAATGGAACAACACTCAACCTTATCTCGGTCTATTCTTTGATTATA
AGGGATTTTGCCGATTTGCGGCTATTGGTTAAAAATGAGCTGATTTAACAAAAATTTAACCGCAATTAATTGTGTGGAA
TGTGTGTGAGTTAGGGTGTGGAAAGTCCCAGGCTCCCAGCAGGCGAGAAGTATGCAAGCCTATCAGGACATAGCGTTG
GCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGCCGCTCCCGA
TTCGCGAGCGCATCGCTTCTATCGCTTCTTGACGAGTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGC
GACGCCCCAAGCTGCCATCACGAGATTCGATTCCACCGCGGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGG
ACGCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTCTTCGCCCAACCCAACTTGTTTTATTGCAGCTTAT
AATGGTTACAAAATAAGCAATAGCATCACAAATTTACAAAAAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCT

FIG. 32

CAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGT
 TTCTGTGTGAAATTGTTATCCGCTCACAATTCACACACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCC
 TAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCA
 TTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGC
 GCTCGGTGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC
 GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
 GCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACC
 AGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTCTC
 CTTTCGGGAAGCGTGGCGCTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGG
 CTGTGTGCACGAACCCCCGTTACGCCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
 ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGTACAGAGTTCTTG
 AAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA
 AAGAGTTGGTAGCTCTTGATCCGGCAAACAACCAACCGCTGGTAGCGGTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
 GAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAACTCACGTTAAGGG
 ATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
 TATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTT
 CATCCATAGTTGCCTGACTCCCCGTGCTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATG
 ATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
 TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTT
 TGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCC
 CAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGCTCCTCCGATCGTTGTGAG
 AAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTGATGCCATCCGTAAGAT
 GCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCG
 TCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGTGCTCATCATTGGAAACGTTCTTCGGGGCGAAAACT
 CTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGCATCTTTTACTT
 TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAATGTTGA
 ATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG
 TATTTAGAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAAGTGCCACCTGACGTC

FIG. 32 (Cont.)

(SEQ ID NO: 65)

Nucleotide sequence of TREX2-encoding plasmid (pExodus CMV.Trex2)

gacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagtt
 aagccagtatctgctccctgcttgtgtgttgagggtcgctgagtagtgcgcgagcaaaatttaagctaca
 acaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggtaggcgttttgcgctgcttcgcg
 atgtacgggcccagatatacgcgcttgacattgattattgactagttattaatagtaataattacggggtc
 attagttcatagcccatatatggagttccgcgcttacataacttacggtaaatggcccgctggctgaccg
 cccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgcccaatagggaactttcc
 attgacgtcaatgggtggagtatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgcc
 aagtacgccccctattgacgtcaatgacggtaaatggcccgctggcattatgccagtagcatgacctta
 tgggactttcctacttggcagtagcatctacgtattagtcacgctattaccatgggtgatgcggttttggc
 agtagcatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccacccattgacgtcaa
 tgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtacaactccgccccattgacg
 caaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactagagaacca
 ctgcttactggcttatcgaaattaatacgactcactataggagagaccaagctggctagcatgtctgagc
 cacctcgggctgagacctttgtattcctggacctagaagccactgggctcccaaactggacctgagat
 tgcagagatatccctttttgctgttcaccgctcttccctggagaaccagaacgggatgattctggttcc
 ttggtgctgccccgtgttctggacaagctcacactgtgcatgtgcccgagcgccccctttactgccaagg
 ccagtgagattactggtttgagcagcgaaagcctgatgcactgcgggaaggctggtttcaatggcgctgt
 ggtaaggacactgcagggttcctaagccgccaggaggccccatctgccttgtggccacaatggcttc
 gattatgacttcccactgctgtgcacggagctacaacgtctgggtgcccatctgccccagacactgtct
 gcctggacacactgcctgcattgcggggcctggaccgtgtcacagccacggcaccagggtcaaggccg
 caaaagctacagcctggccagctcttccaccgctacttccaggctgaaccagtgctgccattcagca
 gaagggtgatgtgcacacctgcttctgatcttccctgcacgtgctcctgagctgctcgctgggcagatg

FIG. 33

agcaggcccgagctgggctcatattgagcccatgtacgtgccacctgatggccaagcctcgaagcctg
 aattctgcagatatccagcacagtggcgccgctcgagtctagagggcccggttaaaccgctgatcagc
 ctcgactgtgccttctagttgccagccatctgttgtttgccctccccgtgccttccttgaccctggaa
 ggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcatt
 ctattctggggggtgggggtggggcaggacagcaaggggaggattgggaagacaatagcaggcatgctgg
 ggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcg
 ccctgtagcggcgcatataagcgcggcggtgtggtggttacgcgcagcgtgaccgctacacttgccagcg
 ccctagcggccgctcctttcgctttcttcccttcttctcgcacgttcgcgggctttccccgtcaagc
 tctaaatcgggggtccctttagggttccgatttagtgctttacggcacctcgacccccaaaaaacttgat
 tagggtgatgggtcacgtagtgggccatgcacctgatagacgggttttcgccctttgacgttggagtcca
 cgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggtctattcttttga
 ttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcg
 aattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtat
 gcaaagcctatcaggacatagcgttggtacccgtgatattgctgaagagcttgccggcgaatgggctga
 ccgcttcctcgtgctttacgggtatcgccgctcccgattcgagcgcacgccttctatcgcttcttgac
 gagtcttctgagcgggactctggggttcgaaatgaccgaccaagcgcgccccaacctgccatcacgaga
 tttcgattccaccgccccttctatgaaagggtgggcttcggaatcgttttccgggacgcccggctggatg
 atcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatg
 gttacaaataaagcaatagcatcacaatttcacaaataaagcatttttttactgcattctagttgtgg
 tttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaa
 tcatggatcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacagagccggaa
 gcataaagtgtaaagcctggggtgcctaatagagtgaagtaactcacattaattgcgttgcgctcactgcc
 cgctttccagtcgggaaacctgtcggtgccagctgcattaatgaatcggccaacgcgcggggagaggcggt
 ttgcgtattgggcgctcttccgcttctcgtcactgactcgctgcgctcggtcggttcggctgcggcgag
 cggtatcagctcactcaaaggcggttaatacggttatccacagaatcaggggataacgcaggaaagaacat
 gtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcttgctggcggttttccataggctc
 cgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccgcagaggactataaa

FIG. 33 (Cont.)

gataccaggcggtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggata
 cctgtccgccttttctcccttcgggaagcgtggcgcttttctcatagctcacgctgtaggtatctcagttcg
 gtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccggttcagcccgaccgctgcgccttat
 ccggtaactatcgtcttgagtccaaccggtaagacacgacttatcgccactggcagcagccactggtaa
 caggattagcagagcgaggtatgtaggcggctacagagttcttgaagtggcctaactacggctac
 actagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagct
 cttgatccggcaaaaccaccgctggtagcgggttttttgtttgcaagcagcagattacgcgcagaaa
 aaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaaacgaaaactcacgt
 taagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtt
 ttaaataatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatacagtgaggcacc
 tatctcagcgatctgtctatttcgttcacatagttgcctgactccccgctgctgtagataactacgata
 cgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagatt
 tatcagcaataaaccagccagccggaagggccgagcgcagaagtggcctgcaactttatccgcctccat
 ccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttggt
 gccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaac
 gatcaaggcgagttacatgatcccccagttgtgcaaaaaagcggttagctccttcggctcctccgatcgt
 tgtcagaagtaagtggccgcagtggttatcactcatggttatggcagcactgcataattctcttactgtc
 atgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgc
 ggcgaccgagttgctcttgccggcgctcaatacgggataataccgcgccacatagcagaactttaaaagt
 gctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctggttgagatccagttcg
 atgtaaccactcgtgcaccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaa
 aaacaggaaggcaaaatgccgcaaaaaaggggaataagggcgacacggaaatggtgaatactcatactctt
 ctttttcaatattattgaagcatttatcaggggtattgtctcatgagcggatacatatttgaatgtatt
 tagaaaaataaacaataggggttccgcgcacatttccccgaaaagtgccacctgacgct

FIG. 33 (Cont.)

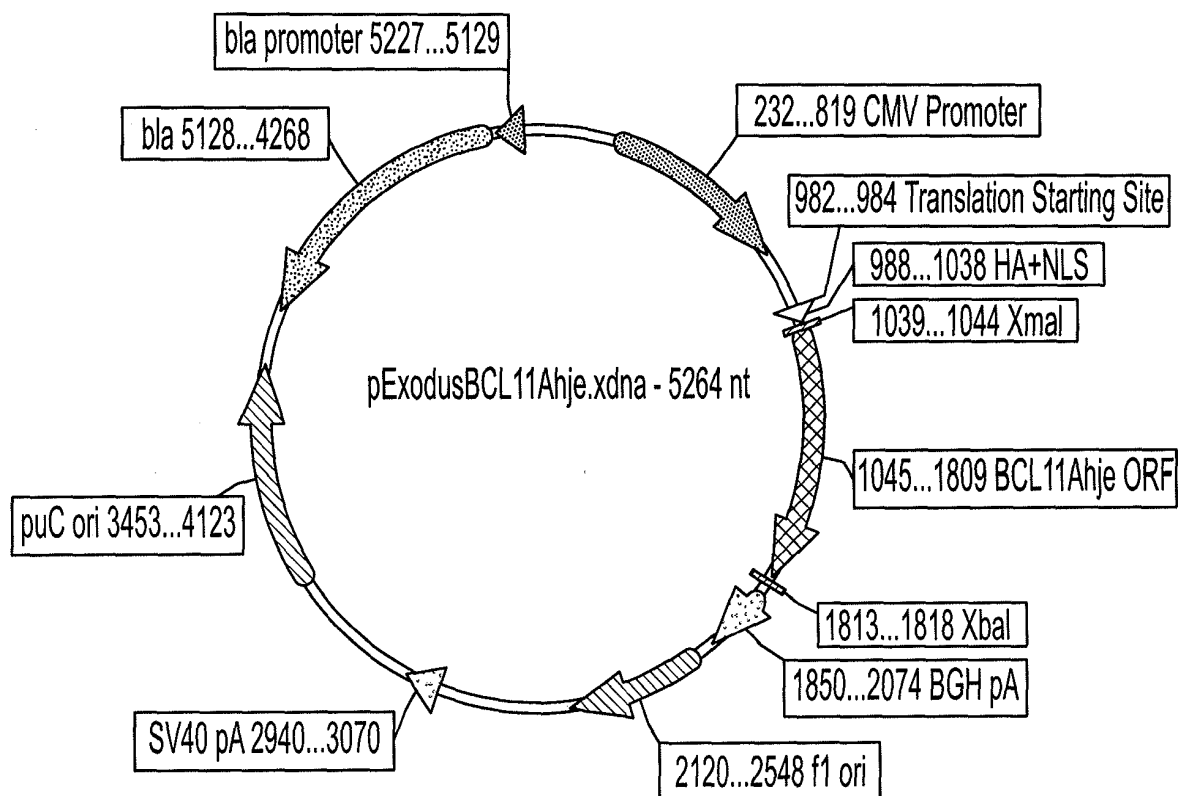
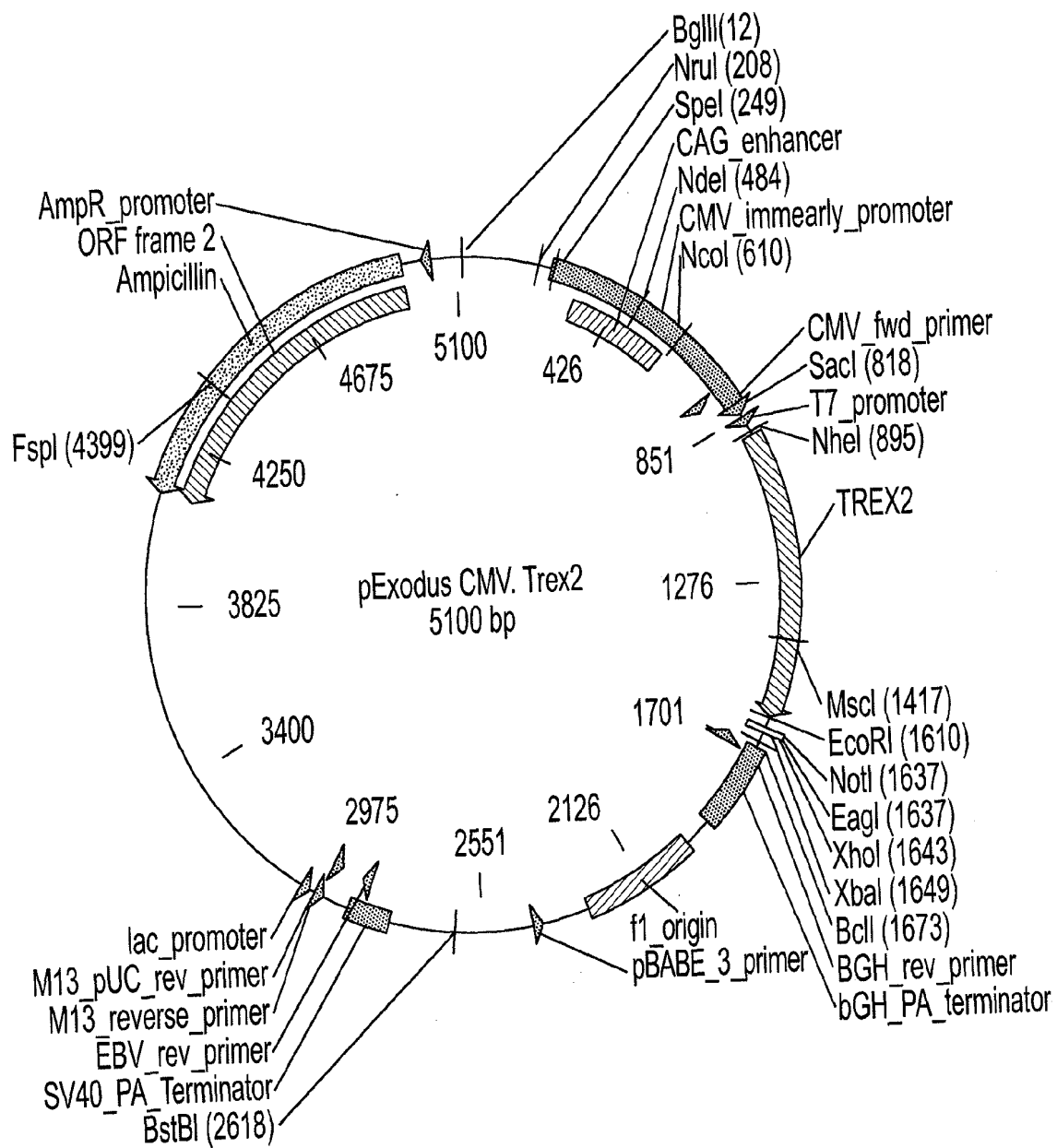


FIG. 34

**FIG. 35**



EUROPEAN SEARCH REPORT

Application Number
EP 17 17 8186

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	SANKARAN V G ET AL: "Transcriptional silencing of fetal hemoglobin by BCL11A", 1 August 2010 (2010-08-01), ANNALS OF THE NEW YORK ACADEMY OF SCIENCES, NEW YORK ACADEMY OF SCIENCES, US, PAGE(S) 64 - 68, XP002743270, ISSN: 0077-8923 [retrieved on 2010-08-02] * the whole document *	1-15	INV. A61K38/46 C12N9/22
A	WO 2010/030963 A2 (CHILDRENS MEDICAL CENTER [US]; ORKIN STUART H [US]; SANKARAN VIJAY G []) 18 March 2010 (2010-03-18) * paragraphs [0009] - [0016] *	1-15	
E	WO 2014/036219 A2 (SANGAMO BIOSCIENCES INC [US]) 6 March 2014 (2014-03-06) * paragraphs [0018] - [0023] *	1,4-6, 10-12,15	
			TECHNICAL FIELDS SEARCHED (IPC)
			C12N A61K
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 27 October 2017	Examiner Barnas, Christoph
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	

EPO FORM 1503 03.02 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 17 17 8186

5

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

27-10-2017

10

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010030963 A2	18-03-2010	BR PI0919017 A2	30-05-2017
		CA 2737180 A1	18-03-2010
		DK 2334794 T3	20-02-2017
		EP 2334794 A2	22-06-2011
		EP 3208339 A1	23-08-2017
		ES 2615829 T3	08-06-2017
		MX 345116 B	17-01-2017
		PT 2334794 T	15-02-2017
		US 2011182867 A1	28-07-2011
		US 2013129629 A1	23-05-2013
		US 2017152513 A1	01-06-2017
		WO 2010030963 A2	18-03-2010
		ZA 201101964 B	30-05-2012

WO 2014036219 A2	06-03-2014	AU 2013308770 A1	05-03-2015
		CA 2882499 A1	06-03-2014
		CL 2015000450 A1	10-07-2015
		CN 104704110 A	10-06-2015
		EA 201590167 A1	30-11-2015
		EP 2890780 A2	08-07-2015
		HK 1210502 A1	22-04-2016
		JP 2015533786 A	26-11-2015
		KR 20150047498 A	04-05-2015
		PH 12015500433 A1	20-04-2015
		SG 10201701601W A	27-04-2017
		SG 11201500852W A	29-04-2015
		US 2014080216 A1	20-03-2014
		US 2014093913 A1	03-04-2014
		WO 2014036219 A2	06-03-2014

40

45

50

55

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 61603231 PCT [0001]

Non-patent literature cited in the description

- GOOLEY et al. *N. Engl. J. Med.*, 2011, vol. 363, 2091-101 [0010]
- BRUNSTEIN et al. *Blood*, 2010, vol. 116, 4693-9 [0014]
- OFFNER et al. *Blood*, 1996, vol. 88, 4058-62 [0015]
- WILBER et al. *Blood*, 2011, vol. 117 (15), 3945-3953 [0033]
- DOYON et al. *J. Am. Chem. Soc.*, 2006, vol. 128 (7), 2477-2484 [0033] [0125] [0127]
- MALI et al. *Science*, 2013 [0033] [0051] [0052]
- JINEK et al. *Science*, 2013, vol. 337, 816-821 [0044] [0137]
- CONG et al. *Science*, 03 January 2013 [0044] [0137]
- MALI et al. *Science*, 03 January 2013 [0044] [0137] [0140]
- STODDARD. *Structure*, 2011, vol. 19, 7-15 [0052] [0058]
- CHEN et al. *Nat. Methods*, 2011, vol. 8 (9), 753-5 [0053]
- TAKEUCHI et al. *Proc. Natl. Acad. Sci. U.S.A.*, 2011, vol. 108, 13077-13082 [0058]
- BAXTER et al. *Nucl. Acids Res.*, 2012, vol. 40 (16), 7985-8000 [0058]
- MAK et al. *Science*, 2012, vol. 335 (6069), 716-9 [0061]
- XU et al. *Science*, 2011, vol. 334, 993-6 [0062]
- SANKARAN et al. *Nature*, 2009, vol. 460, 1093-7 [0062]
- GALANELLO et al. *Blood*, 2009, vol. 114, 3935-7 [0062]
- SANKARAN et al. *Science*, 2008, vol. 322, 1839-42 [0068]
- SANKARAN et al. *N. Engl. J. Med.*, 2011, vol. 365, 807-14 [0068] [0069] [0072]
- FORGET. *Ann. NY Acad. Sci.*, 1998, vol. 850, 38-44 [0070]
- CERTO et al. *Nat Methods*, 2011, vol. 8, 671-6 [0079]
- JARJOUR et al. *Nucleic Acids Res*, 2009, vol. 37, 6871-80 [0079]
- BEARD et al. *J. Clin. Invest.*, 2010, vol. 120, 2345-54 [0085]
- MUNOZ et al. *Nucleic Acids Res.*, 2011, vol. 39 (2), 729-743 [0085]
- TROBRIDGE et al. *Mol Ther*, 2008, vol. 18, 725-33 [0089]
- WANG et al. *Exp. Hematol.*, 2008, vol. 36, 823-31 [0089]
- WANG et al. *Nat. Med.*, 2011, vol. 17, 96-104 [0089]
- DAHLBERG et al. *Blood*, 2010, vol. 117, 6083-90 [0099]
- DELANEY et al. *Nat Med*, 2010, vol. 16, 232-6 [0099] [0101]
- VARNUM-FINNEY et al. *Nat Med*, 2000, vol. 6, 1278-81 [0099]
- BOITANO et al. *Science*, 2011, vol. 329, 1345-8 [0102]
- WATTS et al. *Blood*, 2010, vol. 116, 5859-66 [0102]
- ZHANG et al. *PLoSMed*, 2006, vol. 3, e173 [0102]
- BUTLER et al. *Cell Stem Cell*, 2011, vol. 6, 251-64 [0103]
- JACOBY et al. *Nucl. Acids Res.*, 2012, vol. 40 (11), 4954-4964 [0119]
- TAYLOR et al. *Nucl. Acids Res.*, 2012, vol. 40, W110-6 [0119]
- TAKEUCHI et al. *Nucl. Acids Res.*, 2009, vol. 37 (3), 877-890 [0125] [0136]
- TAKEUCHI et al. *Proc. Natl. Acad. Sci. U.S.A.*, 2011 [0125]
- CERMAK et al. *Nucl. Acids Res.*, 2011, vol. 39, e82-e82 [0133]
- GIARRATANA et al. *Nat. Biotechnol.*, 2005, vol. 23 (1), 69-74 [0160]
- PÁSZTY et al. *Science*, 1997, vol. 278 (5339), 876-878 [0164]
- PORCU et al. *Blood*, 1997, vol. 90 (11), 4602-4609 [0164]
- YANG et al. *Proc. Natl. Acad. Sci. U.S.A.*, 1995, vol. 92 (25), 11608-11612 [0165]
- BENDER et al. *Mol Cell*, 2000, vol. 5 (2), 387-393 [0165]

摘要

本发明提供治疗诸如地中海贫血和镰状细胞病的血红蛋白病的组合物和方法。组合物和方法包含/包括一种或多种内切核酸酶或内切核酸酶融合蛋白,该内切核酸酶或内切核酸酶融合蛋白包括一种或多种归巢内切核酸酶和/或归巢内切核酸酶融合蛋白和/或 CRISPR 内切核酸酶和/或 CRISPR 内切核酸酶融合蛋白: (a) 以破坏 Bcl11a 编码区; (b) 以破坏 Bcl11a 基因调节区; (c) 以修饰成人 β -球蛋白基因座; (d) 以破坏 HbF 沉默 DNA 调节元件或通路,诸如 Bcl11a-调节的 HbF 沉默区; (e) 以突变一种或多种 γ -球蛋白基因启动子来实现 γ -球蛋白基因表达增加; (f) 以突变一种或多种 δ -球蛋白基因启动子来实现 δ -球蛋白基因表达增加; 和/或 (g) 以修正一种或多种 β -球蛋白基因突变。