COMPOSITIONS AND METHODS FOR THE TREATMENT OF HEMOGLOBINOPATHIES

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 Bcl11a coding region; (b) to disrupt a Bcl11a gene regulatory region; (c) to modify an adult human β-globin locus; (d) to disrupt a HbP silencing DNA regulatory element or pathway, such as a Bcl11a-regulated HbP 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
Non-Relapse Mortality by Donor Type

Cumulative Incidence

Years Post-Transplantation

P < 0.01

FIG. 3
Median Time to Initial ANC ≥ 500

Days Post Transplant

p < 0.0001

11 Days

25 Days

Expanded+ Unmanipulated
N=13

Conventional dcBT
N=29

FIG. 4
FIG. 5

Cord Blood Transplants
Broken Down by Disease

- Malignant
- Nonmalignant

Total # of Transplants

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of CBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>5</td>
</tr>
<tr>
<td>2006</td>
<td>18</td>
</tr>
<tr>
<td>2007</td>
<td>22</td>
</tr>
<tr>
<td>2008</td>
<td>23</td>
</tr>
<tr>
<td>2009</td>
<td>47</td>
</tr>
<tr>
<td>2010</td>
<td>41</td>
</tr>
<tr>
<td>2011</td>
<td>43</td>
</tr>
</tbody>
</table>
Nucleotide Sequence of a Human Fetal Hemoglobin (HbF) Silencing Region

(chr11:5212342-5215944 in HG18)

1  CCAGTGAGCAGGTGGTTAAGATAAGCAGGGTTTCATTAGTTTGGAGA  50
51  ATGAAAAATGAACTCCATTTCCCATATTTACCCTTAACCTCTCGCCGAGATTG  100
101  GCTGTCTGCTATGTTGTCTTGACTCAGAAACCTGTTCTTCCCTACAT  150
151  ATCTCCCACCGCATCTCTCTACGCAGTTTCTAAACTATCCTCTCTA  200
201  GCTTCTTATTTGAAAGTGTTTATAGCGTATAGTGAAATGTATTTCT  250
251  GAGTGTATACTTTATTTCTCTGTCAATTATCTAAGAAAAACTTATCTA  300
301  AGGAGAGTGTGTTAAACAAAGTTGACTCAATAAAAGTTTCTTTGCCCTC  350
351  AGGAGAGTGTGTTAAACTCTGCTTCTTCTTGGTAAATTCTT  400
401  ATGGTGGCTAAGTTTTCATTTGCTTCTCTTCTTCTTTCTGTAATTTCT  450
451  CCTATAATAAAAACTCAattttttttattataggaaatataaatattcattaa  500
501  aagaaatattttttaaaaatGAGATGTTTTTTACAGTTATGGAAATAGG  550
551  TTCTGGGAACATGAATTTTTAAGTTTACATTTTATDGACGATAAAATCA  600
601  AATTATATACAAATTTTTTGAATTTAAATATGGATACGTAATG  650
651  AAAGAATGCAAGTGAAAAGTAGATTATTTCAATATCGCCAGATTTAATT  700
701  TAACGAAGTCTCCTGGGAAATATCGTACTACAGAACATTTTTTACAGATGTG  750
751  TCTTTAATTTTTTTGTGAAATTAGACCCAGGAAATGAGATCCTCAGTT  800
801  TTTCATTCTTCTTTCTGAATTTCATTTTAAATGCCCAAATGGCAGACAAATACAC  850
851  ACCCATGACGATATCCTCAaaaggaaggattggaagaaaggaagggaggaag  900
901  tggagaaggaaggaaggaaggggaagggagagagagagatggaagggatgga  950
951  ggagaaggaaggaAATAATAATAGGagagagagagaaagggaggg  1000
1001  agagagagagagagagagggagtggagaaggaagagaggaaagggaaggaaga  1050
1051  ggagagagagagagagagagagagagaggaagaaggggaaggggaaaaagggg  1100
1101  ggagagagagagagagagagagagagagagagagagagagagagagagagagagaga  1150
1151  caaatatatgtgtgtgtgtatatatatagagaataac Tatcccatcttttctgtgga  1200
1201  cactgtacatagataaccctctttggaacccctttaatagtatcccaaa  1250
1251  atgtgagatgtgtgataagttcattacactacattacattatagatat  1300

FIG. 6
FIG. 6 (Cont.)
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   AAAGATGGATGATGTGCCTGAGATTCTGATCACAGGGAAAATGTTATAA  50
51  AATAGGGTAGAGAGGAGCCATGAATGACCTTTAAACTTTGTTACCAAGTTA  100
101  TTTTTCTGTAACCTggaagccaacgaagatattgaataattcacaagaag  150
151  tgtgtggcatggtttgttgcttttaaaagattattctcacttagtg  200
201  aagaaatgtatatttagagaaatggaagagacaatagctgggtt  250
251  ctttgccagtaGGGAAAAGTGCAATGCCATTCTATTACTGAGCTTG  300
301  GACCATGACGCTGATGTGCATCGTGAAACAAGAATAGG GCCACATTGT  350

FIG. 7
(SEQ ID NO: 13)

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

1   GCAATGAAAATAATGTTTTATTAGGCAAGATCCAGATGTCACAAGCC  50
51  CTTCTATAATATCCCAGCTTTTAGTAGGACTTGGGAACAAAAGGAACC  100
101 TTTAATAGAAAATGGGACAGGAAAGGAGCTTTAGTATCTTTTGAGGC  150
151 CAGGGGCTATTGCACCCAGCCACCCACCTTCTGTAGGCGAGCTGCTGACTG  200
201 GTGGGGTGAATCTTCGTCCAAAGTGATGGCGGCGACACAGACACAGACAG  250
251 TTGCCAGGGGGCTGTGGGAGGAAGATGACGGGATGACATGATTGCGAA  300
301 AAGGGCCTAGTTGGAGTCAGATAATCCAGCCTATCTCCACAACATAA  350
351 TAAAGCAGAATGCTAGCTGATGATCTGCTATTAGCAATATGAAACC  400
401 TTCTACTACAGTTCAAATTTATATAGCAAGAATATTTATATGAGATAT  450
451 TGCTATTTGGCCTTACCCAGAAATTTATCACGTATTCTTTGAAATGTCG  500
501 AAAGGGGCAATGATACATTGATCATATTGGCCCTGAAAGAGAGATTAG  550
551 GGAAAGTATAGAAATAAGATAAAACAAAAAGTATTATATATAGGAGAGAG  600
601 CATTTTTTAAATACAATGCAaattaccctgtttgctaatatgtg  650
651 tacacatattaaacattacatattaccataaccataaatagtaataatgatta  700
701 tgtatcatattaaataaaaagAAAATAAGGATGAGGATGATGATGATG  750
751 AAATAAGCACACATAATTCCAAATAGTAATGACTGAGCAACTGTGTA  800
801 AAGTTTTTTTAAGTATATTACATATATGACGATATTCTTTGTT  850
851 ATACACAATTTGTTAGGCCATCAAATGATAATGAAAAATGCGgagaaga  900
901 aaaaaaaaaaaagcaagaataaaacaagaaaaaaatattgcattgtaataacagc  950
951 aaataaagaaactaaacCAATCTGAGACTTCCACACTGATGCAATCAT 1000
1001 TCGTCTTTCCTCCACTCTAATAGGACTCTTTATTGCTTCTCCCTCTCT 1050
1051 GACATGAACATTAACATAGAAGAAGAGGGAAGGAAGAAGAAACTCAAGGG 1100
1101 CATAGACTCACCGTGAGTTCTCAGGATCCACGTCAGCTTCAGCTAGTG 1150
1151 CAGCTACCTAGTGCCAGATGGTGGAGGTCCGCTGCTGAGTTGCCAGAC 1200
1201 GCCCATCATAAGCCGGACACTTTCTTCTGAGCTAGCCTACCACTTA 1250
1251 GGTTGCCCCATAACAGCATCAGGAGTGAGGACAGATCCCAGGAGACTCAA 1300

FIG. 8
2550
2600
TCCTCTAGTCCACTAAGAATACTGCGTTTTAAAATCATTTCCTTGATTCA
2606
FIG. 8 (Cont.)
(SEQ ID NO: 14)

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

1  CACCCCTGAATTCCTCAGGATCCACGTGCAGCTTGTCACTAGGCAGCTCAC  50
51  TCAGTGTCAGGAGGTCAGCTGTGAGGTGTGCAGCTGCAGCTCTAGCTTTGAC  100
101  CTAAAGGCACGACACTTTCTGGTCCATGACCTTCACCTTAGGGTTGAC  150
151  CATAACAGCATCCAGAGATGCAGACATCCACAAGGAAGCTCAAGAAGCTCT  200
201  GGGTCAGAGGGTAGACCAAGGCTACAGCTTCAAGGGTGGAATAAGGACCAAT  250
251  AGGCAAGAGAGTCAGTCCTATACAGAAACCCAAAGAGCTTTCTCTGCTCTC  300
301  CACATGCCCAGGTTCTATTGCTCTCTTAAACCTGTCTGTGAACCTTGAT  350
351  ACCAACCTGCCCCGCGCTACCAAACTCCCCACCTCACTGGTTTTGACCCTGGCC  400
401  CCACAGGGGCACTACGAGCTCTTCTCTCGAGGTACGATGCAGATGACCCATGG  450
451  TGTCCTGCTGGAGTTGCTAGTGCAACACAGTTGTCAGAACGCAATGTAA  500
501  GCAATAGATGGCTCTGCCCCGTACTTTATGCCCAGCCCTGGCTCTGGCCC  550
551  TCCCTGCTCTGGAGTAGATTGCGCAACCCTAGGGTTGAACTCCACAGG  600
601  GTGAGG  606

FIG. 9
(SEQ ID NO: 24)

Nucleotide Sequence of Human Bcl11a cDNA (CCDS1862.1)

1   ATGTCTCGCCGAAGCAAGGCCAAGGCAAACCCAGCCTTAAGCAGAACGGAATT
51  CTCGCCCCAGCTCTTGAAGGCTTCACTTACAGTGAATCGAAGAAGAGGAC
101 GACCCGGATGGAATGCAAAGGGAATCTGACCTCTACCTGACCTGGGAG
151 TGCCAGATGAACCTCCCCCATTTGGGAACAGACATTTCTATTATTTATCAGACCA
201 ACGGAAACAATGTGACGCTCTGCTGAGGAAGGCTGAGTTGAACG
251 CACCTTCCTACTCAGATGAAACTCAGATCAGAATCCCGTCCTGTTATCCCGA
301 GTGGGCATCCAATGTCACCCAGAAGAGGCATGACATTTGTTTACAACGTACT
351 TAGAAGATTTGCCAACAAGCAGACATAGCAAACAAATTCTGCACT
401 GAGGGGCGCTCTTCTCCCTCTGTGCACTTAGGAGCTCTAATCCACCG
451 CCTGGATGATGTGAAGATGAAGGTCCGGCCAAGCTTTGTTATAGATGAGCC
501 CAGCACTACACACTGTACAACACATTAAGGTCAACAAGCCACCTTACAGTGATG
551 TTCTCTTGCAACACACGAAACACTGATGCTAGTAAATCATCTAGA
601 AGCAACACGGGAGTCACTCCTGCCAGCCGGTGTCTGTATCTTCACGGACT
651 AGGTGCAAGATGTCCCTCTCCAGCCACCTCTCCATAGGATCTACATGGAG
701 ACAATACCCCCTCTATTACGTGAAGTACACGGAGTACGATCAGAGAC
751 GCTTCCGGCCTGGCAAGAGGGGAGCTTTACACACCACCTCCCCCTCGTATTAG
801 TCCACACGGAGACATCATTTGACCCCCAGCCAGCATAGAGCCTGGGGG
851 CGGAAGAGATGGCCCTGCCCACCCATCACCCGGAGTGCCCTTGACAGGTTG
901 TCAGGGGTTGAATCCATGTGTAAGTGGATTCCTCCGCCGATGGTTTCTTAG
951 AGACTTACAGAGCTGGCAAGGAACAGCTAGCCACCGCTGTCGCCAG
1001 GCGGCCCCAGCCTATGCAAGGTATCTGCAACCATTTCCACACCGGTAGC
1051 AAGCCGCCCTCTTCTGGGACGCCCCCCTTCCCCTTCTGCAATCCGCC
1101 TCCTCCCTCCCCGCCCCGGTCAAGTCAAGTCACTGGAAGTTCTGCGGGG
1151 AGAGGTCTTTAACACACACACCTGTGTGTCACGGGCGCGCAGCCAGAC
1201 GGCGAGAAGGCTTACAGTGCAACCTTGCGAACACGGCTGGCACCAGC
1251 CAGCAAGATGAGGGCCACTGAGACGACATGACACAATCGTCCCCC
1301 TGACGGTGACAGCGACCTTCTTCCACCGCCGAGCTCCCGGAAACC
1350

FIG. 10
FIG. 10 (Cont.)
Directed Evolution for Creating a BCL11A Gene-targeting Endonuclease
(SEQ ID NO: 28)

Nucleotide Sequence of 1-HjeMI, Codon Optimized for Expression in E.coli

1 ATGGGATCCCCATGGGACCTGACCTACGCTTACCTGGTGTCTGTTTCGA  50
51 AGGTTGACGTTAACCCTCTTCATCAACCAAAAGGTAATATACCTGACCTACG  100
101 AACTGGGTATGCAACTGCTATCACAAGACGGTTGCTGATCTCACAATAC  150
151 AAAGACATCTCGGCTTGGGTGTAAGTTTTCTTTCCGTAAGCGTAAGGAAAT  200
201 CGAAATGGTTTCTCTGCGATACCGTACCCGTAACAGAATCACCTGAAAAACTTCA  250
251 TCCTGCCTGATCTTTGAACAAATACCCTGATGCTGTCTAACAGCAGTACGAC  300
301 TACCTCGTCTTCAAAAGACGCTCTCTGCTAATCTTACCTCTGACGA  350
351 TCTGGCGAATACCCGTGTTCTAAGCATTATCTCATCTCGACGA  400
401 TTATCAGACCTCTTACTTTCTGCTTTGCTGCTTTGTTCTGATCAGCT  450
451 GAAAGGTTGCTTCTCTACCTCAACCAAACTGAACAAAGATGAGCAGTACCCTGAT  500
501 CGCTTACCTGGGATACATCAGCAGAAAAGAGGTTGACCATCTCTGATCTCTGCTA  550
551 TCACAAATACCTGTTTCTTCAACAGAAATCTCCTGAGAAACACCAAC  600
601 TGCTTCTGCCTGAAAGTGGACGCGGTGAGGCTCTGTCTTAAAAGTGGTTA  650
651 ATTCATCCAGGGTTGCTCCGTTAAGACGGGTAACTGCTCGTAAACAAGAAACTGACGT  700
701 ACAAACTGGATACAAAGCAGCTGCCAATAATCTCTGTTACTCTGAAAAA  750
751 ATCCAGCTGCCGCTAAACTAC  771

FIG. 14
(SEQ ID NO: 29)

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

1  ATGGGCAGCCACATGGACCTGACCTACGGCTATCTGTGGTGCTGGCTGTTGA
51  GGGCGACGGCTATTATTTAGCATACCCAGGAGAGGCAATAATCTGACGTATG
101  AACTGGGCACTGAGCTCTCCATCAAGGACGTGCACCTACATCTACAAGATC
151  AAGGACATCTCGCGGTGGGAAAAGTGTTCTTTAGGAGAGGACGAGAT
201  CGAGATGTGTCGAGCGTAGCAAGGACAAAACGACCTGAGAGAACTTC
251  TCCTGCCCATCTTCGACAAGTACCCCATGCTGAGCAAAACGCAAGTACGAC
301  TATCTCCGATTCAAGGATGCGCTCCTGTGCAATCATCTATAGCGACGA
351  CTTGCCGAGTAGCGCCAGGACCAAGGCTGACATCGTAGCTGAGAGACA
401  TCAATACATCGGCTTCGACGCTACAGGACAGCTACCTGCTGAGAC
451  GAGGGCTGTGGTCACGCACTACAAGGCTCAAGGACAGCGATTATTTGAT
501  CGCGAGCTCCGATATAGCCAGGAAGGAGGCGACATTCTCAGCTTCGCCGA
551  TCCAAATAACCTGAGCTTTCAAGGACCAAAAATCTACCTGACAAGACCA
601  TGATAGCCAGCTCGGAGTACCCGCGTGAGAGGCGCTCAAGAAGCTGGTA
651  GTTCAGTCCAGGGTGCGCCGTCAGTTGGCGTATACAAAGAAGCTGCA
701  ACAAATTCGAGAAACGGCTGCGCAAGATCTCCGATACACGGAGAAA
751  ATCCAGTCGCCCAGTAACATC

FIG. 15
(SEQ ID NO: 30)

Amino Acid Sequence of the Homing Endonuclease I-HjeMI

1  MGDLYAYLVLFGDFYFSITKKGKYLTYELGIELSIKVQLIYKIKDI  50
51  LGVGKVSFRKRNEIMVSLRIRDKNHLKNFILPKIFDKYPMLSNKQYDYLRI  100
101  FKDALLSNIISDDLEPEYARSNESINSVDSINTSYFSAWLVGFEAEGC  150
151  FSTYKLKDDDYLIASFDIAQKDGDILISAIHKLGSFTTNYLDKTNCSR  200
201  LKVGLRNVKYNVVKFIQGAPVKLGNNKLLQYKLWIKLRLKISRYSEKIQL  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  ATGGGATCCACATGGACCTGACCTACGTTACCTGGTTGGTCTGTTCTGA  50
51  AGGTGACGTTTACTCAGCTACTGCTAAAGCGGTAAGTACCTGAAATTACG  100
101  AGCTGGTATCACGCTGTCTATCAAGACGCTACGCTGATCTACAATAAATC  150
151  AAAGACATCCTGGGTGTTGTAATGTTATTTCGGGAATAATAGGCAAGCA  200
201  TGAATGTTTCTCTGGATACCGGAACTACGAAATCATACCTGAAAAACTCA  250
251  TCTGGGATCTTTCGACCAAAATACCCGATGCTGTCTAAACAGCAGTAGAC  300
301  TACCTCGGTTTCAAAGACGCTCTCTGCCTGCTACATTATCTACTCTGACGA  350
351  TCTGGGGAATACGCTCTTCTTAAAGGAATCTATCAACTCTGTGGACTCTA  400
401  TTATCAACACCTCTTTACTTTTCTCGGTGGCTGGTGTGGTTTCTCGAGCT  450
451  GAAGGTTTCTTCAACGCTAACAAAGCGAGTAAGATAAGTACCTGACGCGC  500
501  TGGGTTCAGTATCGCTAGAAAGACGGTGACATCTCTGATCTCGCTATCC  550
551  ACAAAATACCCTGTCTTTCCACCAAAACGTTACAAGGCAAACAAACAACTGC  600
601  TCTCATCTGAAGGTGACCGGTTACGTCTGGTTAACAACGTTGTTAAATT  650
651  CATCCAGGTTGCTGGTTAAACTCTCGGTGTAACAAGGAAACTGCAATACA  700
701  AACTGTGATCCAAACAGCTGGTAAATCTCTGTTACTCTGAAAATCTC  750
751  CAGCTGGGCTGCTAAACTAC  768

FIG. 17
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 ATGGGCAGCCACATGGACCTACGCTCATCTTGTCGCGCTTTCGA 50
51 GGGCGACGGCTATTTTACCATACTACGCTTATAGGCAAGTATCTGACTACG 100
101 AGCTGGGCATCACAATCTGTCAATCAGAGGCTGCTCTCTGACTACG 150
151 AAGGAGATCTCGCCCTGGGAACGTATGCTTTAGAAAGACTGGACATCA 200
201 TGGATGGTGCTAGCTGCAGAATTCACGGACACAGTCAATATAGGACGAC 250
251 TCCTGGCCATCTGGCAACCTACATCTGCTGAGCAGGACAGTACG 300
301 TACCTGCGATTTCAAGGATGCCTCTCTGTCACATCATCTATAGGACGA 350
351 CCTGCAGGATGGCCAGCCAGGACAAGGTACATGACTCAATAGGACGAC 400
401 TCAACATACCTCCTGACTGGACGGCTGCTGTTGCTCTACGAGGCA 450
451 GAGGGCGCTCTCACAAGCATTCCAGGATACGCGTCAGGACATCTACGGACTGA 500
501 GgGCTTTTGGATAGCCAGGAGCGCGCTTCCTGCTCCTGCTGATGAC 550
551 AAAGATACCTGAGCTTCAAGACCCAGACACGCTAAACGCAAGACCAACTGTG 600
601 AGCAGCACTACGAGGACCTGCGCGTGAGACGTGCAATACGATGTAAGTTT 650
651 CATCCAGGCTGCGCTCAAGGTCTGGTGAACAGAAGCCTGCGATGACA 700
701 AACTTTGGATACAGCGCTGCGCTAAGATCTCCGGATACAGCGAGAAAATC 750
751 CAGCTGCCAGTAACTAC 768
```

FIG. 18
(SEQ ID NO: 33)

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

1  MGSHMDLTYAYLVLFEQGDGYFTIAKAGKYLNYELGITLSIKDAQLIYKI  50
51  KDILGVGNVYFRKYROHEMVSLRIQDKHNLFILPIFDKYPMLSNKQYD
101  YLRFKDALLSNIYSSDLPEYARSNENINSVDSINTSYFSAWLVGFIEA  150
151  EGCFTTYKASKDLTAGFSIQAQKDGDIISAIHAKLYLSFTTKPYKDKTNC  200
201  SHLKVTGVRSVNNVKFIQGAPVKLGNKQYKLWIKQLRKISRYSEKI  250
251  QLPSNY  256

FIG. 19
Distribution of Amino-acid Residues Different between the BCL11A Gene-targeting Endonuclease and its Parental LHE-HjeMl
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  ATGTCGCCCTACATGTCCCGTCGGAGTCCATTACCGTGGATTTCTCAC  50
51  CGGTTCGCGAGCGCAGAAGGCTTCCTTTGCTGCGCATCGCGAACAACAA  100
101 ACAAGTCGACGCTGCTACTCCACTGAGCTCGCTTTCAATTACACTT  150
151 CATAAACAGGACAAGAGCATTTCTTGAGAACAATCCAGTCAACATGGAAGGT  200
201 GGGCGTGATCGCAACACAGCGTGACACACCGCGTGCTGCTAGGTCAGC  250
251 GTTTGAGGACTGAGCTTACATGACCATTCTTGGAAAATACCTCACTG  300
301 ATTACGCAAGCTCGGACTACATGCTGCTTTAAGCAGGCTGTTTCGCTG  350
351 CATGGAGAAGGAGCACCTTTATGAAGATTAATGTATGCACGGAGCTGCTG  400
401 GCATATTAGCAGCTCAGATWGAGGAAGCTGAGGCTAGGAGCTGAGAACGCC  450
451 TTTCGGAGATCACTCTCGAAGGAGCGTCCCTCATCAAAGAACAATCC  500
501 TAATTTCAAGTGCTGCGGAGTTTTACTCTCGGCAGGAGGGTTTGCTTCTTG  550
551 TTAACCGATCAAGCTCAAGTCGCTAGGCTGAGGGCGGTCCAGCTGAGAGGAGGAGGAGAGAG  600
601 TTTGCGATTTACCCAAACATCAAGGATAGGAACCTCTCATGAACTCCTCTGAT  650
651 TACCTACTTTGGCTGCGCTACATTAGGAGAGAAAAACGAGAGTGGAGTTCT  700
701 CCTGGGCTACTGCTACGCAAAATTCTCGGAATCACAAGCACAGATC  750
751 ATTCGGTCTTTCCAGAAAACAGTCATCGGCGTGAGCTCGGAGGACTTT  800
801 CGAGGATGTGTAAGCTCAGTACGAGGAGAAGAACGACACCTGA  850
851 CAAGAACTTGCGGCTGCGAGATCAAGAAGATTTAAGCTGAAACATGCACAAG  900
901 GGCAGAGCTATTC  912
```
Amino Acid Sequence of I-Onulhoming endonuclease.

(SEQ ID NO: 15)

SAYMSRESI NPWILTGFAD AEGSFLLRIR NNNKSSVGYS TELGFQITLH NKOSILENI 60
QSTWKGVGIA NSGDNAVSLK VTRFEDLKVI IDHFKEYPLI TQKLGDYMLF QAFCVMENK 120
EHLKINGIKE LVRIKAKLNW GLTDELKKAFFEIISKERSL INKNIPNKW LAGFTSCEGC 180
FFVNLIKSKS KLGVQVQVLF SITQHIDKDN LMNSLITYLG CGYIKEKNKS EFSWLDFFVT 240
KFSDINDKII PVQENTLIG VKLEDLEDWC KVAKLIIEKK HLTESGLDEI KKIKNMNKG 300
RVF 303

FIG. 22B
<table>
<thead>
<tr>
<th>Ladder</th>
<th>fGlobinOnu (L) (-) control</th>
<th>fGlobinOnu (L) (+)</th>
<th>fGlobinOnu (R) (-) control</th>
<th>fGlobinOnu (R) (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 23
(SEQ ID NO: 35)
Nucleotide Sequence of MegaTAL:5.5 RVD + Y2 I-AniI

1  gtggatctacgcacgctggctacagtacgacagacaggagatctaa  50
51  accgaaggtcggttccgacagttggccgacaccacacgacgtgtgtgggcc  100
101  atgggtttcacacgcgcacctcgctacgtgctgcctcaacctgcacgctacaga  150
151  ttaggaccgtcgtgctacgtatcgcacataatcgccggtttgccaaga  200
201  ggccacacacgaagacatcggtggggtcggcgaacacgaggtgtccgggctgac  250
251  gccgccctggaggccttgctacggtacggtggaggtgagagttccggccg  300
301  ttacagttgacacagcagccaaacttgtgaaagttgcaaaaacgtggcggctg  350
351  gaccgcaatggacgcatcccagcatcgcgcaatcgactacggtcctgcccc  400
401  ccctgaacctgacccgggaacagatgttgtggctataccagccaaatggc  450
451  gccaagcaacgcgtcaaaaccggtgcacgcgcgtgtgctgcgggtctgtgcca  500
501  ggaccatggcctcactccgaccaagtggtgtgctataccaggccacagt  550
551  gcggcaagaagcagcgtcgaacagggtgcacgcggtgtgctgcgggtctgtgctgc  600
601  caggaccatgcctgaccrccgggaacaccattggtgtggctataccagcagaac  650
651  tggccggaacagcagcgtcgaacagcgggtgcacgcggtgtgctgcgggtctgtgct  700
701  gccaacgaccatggcctgaccrccgggaacaccattggtgtggctataccagcagaac  750
751  aatgccggcaagcagcgtcgaacagcgggtgcacgcggtgtgctgcgggtctgtgct  800
801  gtccaggaccatggcctgaccrccgggaacaccattggtgtggctataccagcagaac  850
851  acgatggcggcaagcagcgtcgaacagcgggtgcacgcggtgtgctgcgggtctgtgct  900

FIG. 24
FIG. 24 (Cont.)
(SEQ ID NO: 36)
Amino Acid Sequence of MegaTAL:5.5 RVD + Y2 l-Anil

```
1  VDLRTLGYSSQEQEKIPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAA 50
51 LGTAVVYQHIITALPEATHEDIVGKGQWSGARALEALLTDAGELRGPP 100
101 LQLDGTQVIKAKRGVTVAMEAVHASRNALTGAPLNLTPDVVAIASNNG 150
151 GKQALETQRLLPVLCDQHGLTDQVVAIAASHDGKQALETVQRLLPVLC 200
201 QDHGLTDQVVAIAASNIGGKQALETQRLLPVLCDQHGLTDQVVAIASN 250
251 NGKQALETQRLLPVLCDQHGLTDQVVAIAASHDGKQALETVQRLLPV 300
301 LQDHLGLTDQVVAIAASNGGGKQALESIQAQLSRPDPALALTNHLVAL 350
351 ACLGGRPAMDKKKGLPAHELIRRVRNIRIGERTSHRVAISRVGSDLTY 400
401 AYLVGITYEGDGYFSITKKGYLYTELGIELSIKDQVQLIYKIKKLGIGIV 450
451 SFKRRNEIMVVALRIRDKHNLKSLPIFEGYPSNKQYDYLRFRNALL 500
501 SGIYLEDLPDYTRSDPLNIESIINTSYFSAWLVGFIEAGGCSVYKL 550
551 NKDLYLIAFSQLAQGRGFLISAIRKYLSSFTKLVYLDKTCNKLKVTVS 600
601 RSVENIIFLONAPVKLGINKKLQYKWLQRLKIRSYSEIKIPSNYVD 650
651 RASYRHLPYLYFCFS
```
(SEQ ID NO: 37)

Nucleotide Sequence of Cas9 Endonuclease
(from Mali et al., Science (2013))

FIG. 26
FIG. 26 (Cont.)
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 TGTACAAAAAGCAGGGTTAAAGGAACCAATTTCAGTGACTGAGATCCGG 50
51 TACCAAGGTCGGGCAGGAAAGGGCTATTCCCATGATTTCTTCTTCTATT 100
101 TGCATATACGATACAAAGGCTGTTAGAGAGATAATTAGAATTTTGACT 150
151 GAAAACAAAGATATTAGTACAAAATACGTGAAGTAAAGTAAATAATT 200
201 TCTTGATTGTTTCAGTTTTAAAATTATGTGTTTTAAATGACTATCATA 250
251 TGCTTACCCGTAATTGCCAGATTAACTTTGTCTTGTTATATATCTTG 300
301 TGGAAAGGACGAAAACACCGNNNNNNNNNNNNNNNNNNNNNNNNGGTTTTAGAGCTA 350
351 GAAATAGCAATTTAAAATAAGGCTAGTTCCGGTTATCAACTGGAAAGTGG 400
401 CACCGAGTCGCTTTTTTCTAGACCAGCTTTCTTGTACAAAGTGG 450
451 CATTA

FIG. 27
(SEQ ID NO: 62)

Nucleotide Sequence of a 1-CpaMI homing endonuclease
(ORF, codon optimized for mammalian expression)

ATGAACACCA GCTCTAGCTT CAATCCCTGG TTCTCTGACCG GCTTTAGCGA TGCAGAGTG C 60
TCTTTCAGCA TCCTGATACA GGCCAACAGC AAGTACCTCA CCGTTGGAG GATCAAGCCC 120
GTTTTCGGCA TCGGGTTCGCA CAAGAGAGAC GTAGAGCTTC TGAAGAGAAAT CAGAGCTAT 180
CTGGGCGTGG GCAAGATACA CATTCAGGCG AAAGACAGCA TTCTAGTCAG GATTGACAGC 240
CCAAAGGAGC TGGAGGTGAT CATCAGCACC TTTGAAGACT ACCCCTCTGAT AACGGCCAAG 300
TTGGCCGACT ACACCCCTTT TAAGAAGGCC CAGACGTAA TCTCTGGAAA GAGAGACGCTG 360
AGCCAGAAGG GCCTGCTTAA ACTGGTAGGC ATTAAGGCAG GCCCTGAATCT CGGGTTGAAC 420
GGCAGCCTCA AGGAGGCGGT CCCGAAACTGG GAAAGACTGC AGATCAGCAG GCCAGACTAC 480
GTGTTCAAGG GCATCCCGGA CCCCAAAGTG ATCAGGGGCT TGCCTGCAGG CGATAGCAGC 540
TTATAATGGA AAATAGCAGGA CTTCCCCCCAG TCACCTGTCA ATAAAAAGGT GCAGCTGAAG 600
TTGCCGACTG GACTGAACAT CAGAGCGAAA GCCCTTTATCC AATACTCTGT GCGCTACTTT 660
GACCTGTCAG ACAACCTGAA GAAATGTACAT TTCGACCTGA ACAGGGCACG GTGCGAGGTG 720
GTGAAGTTCA GCCACATCAG CGACGAAGTG ATCCCCCTGT CTGACAGAAT CAGCATAAAA 780
GGCAAGAAGA GCGTGACTC ATACAAGTGC AAAGGAAGTG CCGACATTAT CAGAGCAAAG 840
AACCACCTTA CTAGCGAGTG CTGCCAGGAA ATCTTGGAAC TCAAGGCAAG TATGAAA CGAAG 900

FIG. 28
Amino Acid Sequence of I-CpaMI homing endonuclease

MTSSSFNPW FLTGFSDAEC SFSILIQANS KYSTGWRIKP VFAIGHLKKD LEKKRIQSY 60
LGVKHIIHG KDSIQRIDS PKELEVIINH FENYPLVTAK WADYTLFKA LDVILLKEHL 120
SQQGLLLLQG IKASLNGLN GSLKEAFPNW EELQTIDFRSY VFKGIPDPNW ISGFAQGDSS 180
FNVKISNSPT SLLNRVQRL FGIGLNIRED ALIQYLVAYF DLSNKLNIY FDLNSARFEV 240
VKFSDITDKI IPFDKYSIQ GKSLLDYINF KEVADIKSK NHLTSEGQIE ILDIKASMNK 300

FIG. 29
In vitro digestion
BCL11A

TRE2

1st round

2nd round

FIG. 30
FIG. 31
Nucleotide sequence of BCL11A gene targeting nuclease-encoding plasmid

(Seq No: 64)

(Exon1B/C11A)
FIG. 32 (Cont.)
(SEQ ID NO: 65)

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

gacggatcgggagatctcccgatcccccctatgggtgcaacctcaagtacattgctctgatgccgcaatgtt
aaagccagatcctgctccccctgcttttgttggaggtcgtgaatgtggtgacggttgggttaaggcggttttgcgtgctctccgcg
atgtacgccccgacaaatagccgtaacacagctgctttgaccttgattacagtattataattataaatctataacgtgcgcgtc
attagcttcatagcctctatataggaatgctcgcggctacataactacgtaatggtccgctggctgagccagc
cccaacgacccccgcctggacaagctcaaatgaacctgtgtctgtaccatttcagatgtgctggtttgggc
gatcacatcaatggccggtgatagcgggtttgacactcaccgggattcccaagttcaccaccccattgcagctaa
tgagggttggttttttggcacaaccagaaatcaagggagctcttcaccctaaatgctgataaactcgcgcccccattgacgc
caaatgggcgttaggtgtacgctggagggtcttttatataacgagactctctgccagacagctgtctgtcttgagtc
ccacctgggtctggaacctttttgtattttctgagcctaaagccacatggtgcactgggctccccaaacatgagcccctaggtgggc	gcagagatbcccctttttgttgtttccgcctctcctggagaacccagaacgggtatgattctctgtggatgttgc
ctgtgcctccccgtgttcttggacaagctcactgcattgcatgctgccccggagccccttttactgccaagggc
cagttgattactgttgtgacagcgaaagctgtgacctgacgcggagaagctgtgtttcaatggcggtgtggtggatcagttcccctctaatggtgccagctgtc
acgctgaacagtctgccaggccctctctacaagggcaggagggcccccacctgccttgggccccaaatgtgggtgc
agggtgtgtgagttggctgagagcagggctagcctgccttcgccatccctggccaagcagctctccgctgtaaaccagggctgggagctgcttggcc
ccaaaggtactacagcctggccagctcctttccaccggctacctcaagggtagctctgctgctgctgttttgc
agggtgatgtgaccaacctctgctcttttgctgtgatgctcctgggctgacgctgcgtgctcgcctggcaggtg
FIG. 33 (Cont.)
FIG. 33 (Cont.)
FIG. 34
COMPOSITIONS AND METHODS FOR THE TREATMENT OF HEMOGLOBINOPATHIES

CROSS REFERENCE TO RELATED APPLICATIONS

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

SEQUENCE LISTING

The present application includes a Sequence Listing in electronic format as a text file entitled “sequence listing 54428.0006.WOUI_ST25” and crested on Feb. 22, 2013 and which has a size of 174 kilobytes (KB). The contents of the text file “sequence listing 54428.0006.WOUI_ST25” are incorporated by reference herein.

BACKGROUND OF THE DISCLOSURE

1. Technical Field of the Disclosure

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.

2. Description of the Related Art

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.

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, Aγ, 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 adult α-, β-heterotetramers are developmentally expressed; embryonic Hb Gower1 (ζ2, ε4), Hb Gower2 (α2, ε4) and Hb Portland (ζ2, γ2); fetal: HbF (Fetul) (α2, γ2) and Adult: HbA2 (α2, ζ2) and HbA (Adult) (α2, β2).

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.

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 HbS (α2, β2). Hemoglobin S (α2, β2), which carries this mutation, is referred to as HbS, as opposed to the normal adult hemoglobin (HbA). Under conditions of low oxygen concentration Hb-S 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 altertions in numerous pathways that result in tissue damage via vaso-occlusion and hemolysis.

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.

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 underprescribed, compliance is poor, and it does not adequately protect health.

Hematopoietic stem 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

[0013] 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 gene, or adding an exogenous β-like globin gene.

[0014] 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.

[0015] Unfortunately, several factors, including the low cell dose available in many cord blood units 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.

[0016] 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.

[0017] 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 (Ollner 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.

[0018] 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 (Delti1) 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.

[0019] 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.

[0020] 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.

[0021] 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

[0022] 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 promoters 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 β-globin gene; and/or (e) to correct one or more β-globin gene mutation(s).

[0023] 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 an 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.

[0024] 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.

[0025] 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: 5212342-5213944 in HG18) that contains a binding site for the regulatory protein Bcl11a.

[0026] 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.

[0027] 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.

[0028] 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.

[0029] 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.

[0030] 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 erythrocyds 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.

[0031] 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 (TALEBN) 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.

[0032] 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 viral pseudotyped lentiviral vector, a foamy virus vector, an adenoviral vector, and an adenoviral-associated viral (AAV) vector.

[0033] 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 reagents 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.

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**

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

**FIG. 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)).

**FIG. 2** depicts three exemplary rare clearing nuclease technologies.

**FIG. 3** is a graph showing that the risk of non-relapse mortality is highest among double CBT recipients. Non-relapse mortality after double CBT (DCB) matched unrelated donor (MUD), mismatched unrelated donor (MMUD), and matched related donor (SIB) transplant.

**FIG. 4** shows that a culture of B progenitors with Deltal+ results in more rapid neutrophil recovery in a myeloablative double CBT setting. The individual and median times (solid line) to ANC of ≥500/µl 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.

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

**FIG. 6** (SEQ ID NO: 1) is the sequence of the 3.6 kb region for which the HBF silencing region falls, which spans chr1:5212342-5215944 in HG18.

**FIG. 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.

**FIG. 8** (SEQ ID NO: 13) is the human beta-globin gene from 1 kb upstream of the cap through the polyA, which spans from chr1:5203272-5205877 in HG18 (reverse strand).

**FIG. 9** (SEQ ID NO: 14) is a 606 bp region of the human beta-globin spanning from the promoter into Introns 2 (chr1:5204380-5204985 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.

**FIG. 10** (SEQ ID NO: 24) is the cDNA sequence for human Bcl11a cDNA (CCDS18562.1).

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

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

**FIG. 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.

**FIG. 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.

**FIG. 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.

**FIG. 16** (SEQ ID NO: 30) is the amino acid sequence of the homing endonuclease I-HjeML.

**FIG. 17** (SEQ ID NO: 31) is the nucleotide sequence of a BCL11A gene targeting nuclease (Bcl11A HjeML), which is based on the homing endonuclease I-HjeMI (detained through directed evolution in IVC and in bacteria), which is codon optimized for expression in E. coli.

**FIG. 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.

**FIG. 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).

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

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

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

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

**FIG. 23** is a agarose gel showing the activity of an I-Onul homing endonuclease targeting the HBF silencing region.

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

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

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

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

FIG. 29 (SEQ ID NO: 63) is an amino acid sequence of L-CpRl homing endonuclease.

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


FIG. 32 (SEQ ID NO: 64) is a nucleotide sequence of a BCL11A gene targeting nuclease-encoding plasmid pExodusBCL11A/Ahje.

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

FIG. 34 is a restriction map for the plasmid pExodusBCL11A/Ahje.

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

 TABLE 1

<table>
<thead>
<tr>
<th>β-globin chain</th>
<th>Sequence</th>
<th>Sequence Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>GTGCACCTCACTCAGGAGGAGGAGG</td>
<td>SEQ ID NO: 3</td>
</tr>
<tr>
<td>Sickle</td>
<td>GTGCACCTCACTCAGGAGGAGGAGG</td>
<td>SEQ ID NO: 4</td>
</tr>
</tbody>
</table>

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.

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

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 7x10^45 bp).

As used herein, the term “Transcription Activator-Like Effector Nucleases” or “TALEn effector nucleus” or “TALEN” refers to a class of artificial restriction endonucleases that are generated by fusing a TALE effector DNA binding domain to a DNA cleavage domain.

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.

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, Jinke 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).

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,” “one or more,” and “any” terms such as “a” or “an” include both the singular and the plural.
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.

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.

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 Achieving High-Efficiency, Multiplex Gene Disruption and Gene Editing Functions

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-Onu1 homing endonuclease(s), and/or one or more Cas9 endonuclease(s) in combination with one or more RNA guide strand(s) which may be transiently or persistently expressed in targeted cells shown to have clinical benefit in human. 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.

Four protein scaffolds are known in the art for achieving targeted gene modification and disruption in eukaryotes: zinc finger nucleases (ZFNs), TALE effector nucleases (TALENs), homing endonucleases (HEs), and Cas9 endonucleases in combination with a RNA guide strand. The present disclosure employs TALE effector nucleases, homing endonucleases, and/or Cas9 endonucleases either alone or in combination. TALE 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 strand(s) are extremely efficient in mediating gene disruption, Stoddard, supra and Mali et al., Science (2013), supra.

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 660 bp region (SEQ ID NO: 14) that spans the majority of thalassemia mutations as well as the HBS mutation.

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.

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 exonucleases, such as a TREX2 exonuclease or a TALE effector exonuclease, may be delivered to CD34+ HSC for the induction of targeted genetic deletions in critical regions for HBF.

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 GAIA-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 Bc11a 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.

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 compression of a nuclease, such as a TALEN and/or a TREX exonuclease, to achieve greatly enhanced gene disruption efficiency in transduced cells.

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).


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 (Pregene, Seattle, Wash.).

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 nucleases-induced targeted gene disruption by diminishing break persistence and reducing the potential for large scale translocations mediated through alternative end joining pathways.

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 5 Jan. 2012 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 Bc11a Gene Expression

Knockout of Bc11a 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 Bc11a mediated silencing of HBF. Heterozygous and homozygous knockout of the endogenous Bc11a gene in these mice results in δ-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 Bc11a acts as rheostat, modulating the degree of HBF suppression. Consistent with this, decrease of function mutations in Bc11a 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).

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-HgeMI homing endonuclease(s), I-CpaMI homing endonuclease(s), and/or I-Ohn homing endonuclease(s) and/or one or more Cas9 endonuclease(s) to achieve the disruption of a sequence that encodes Bc11a 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 Bc11a gene-specific RNA guide strands to mediate the targeting of the Cas9 endonuclease to a Bc11a gene sequence.

The Bc11a gene has multiple exons spanning over 100 kb 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 Bc11a 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 Bc11a. These targets comprise DNA sequence modules for which pools of highly active endonuclease variants have been isolated and sequenced. The human Bc11a cDNA sequence (CCDS18621.1) is presented herein as SEQ ID NO: 24 (FIG. 10).

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 Bc11a,
and the formation of the corresponding repressive complex, and de-repressing γ-globin expression.

Genome Disruption to Block Bcl11a-Mediated Silencing of HBF

[0104] 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.

[0105] 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.

[0106] 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 a quantitatively elevated level of γ-globin gene expression and HBF protein production in adult erythroid tissues (Sankaran et al., N. Engl. J. Med. 365:507-14 (2011)).

[0107] 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.

[0108] 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 addition factors. Forget, Ann. NY Acad. Sci. 850:35-44 (1998).

[0109] 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.

[0110] 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 HFPH 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 HFPH 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 HFPH 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.

[0111] 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.

[0112] 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

[0113] 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.

[0114] 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 composi-
tions 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.

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 TALE effector modular DNA binding platform to achieve additional therapeutic advantages.

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.

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., Nucleic Acids Res 37:6871-80 (2009).

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).

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 HSC’s 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.

Because these therapeutic methods do not require the integration 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.

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

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.

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).

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 bindings and result in a therapeutically efficacious increase in δ-globin expression.

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, TALE Effector Nuclease, and TREX2 Endonucleases

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 nucleases 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.

Exemplified herein are covalently 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. 17:906-104 (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.

[0128] 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 recombinant 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 knockdown and recombination-based gene editing in HSC in cell lines and in primary CD34+ cells.

[0129] 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-F11) for efficient transduction of hematopoietic cells. Helper-dependent adenoviral vectors offer up to a 50 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 recombinant templates. Alternatively it can carry the small Cas9 polypeptide and guide RNAs. These modified chimeric adenovirus vectors may, therefore, be employed for both gene knockdown and recombination-based gene editing in HSC.

[0130] Integration-deficient lentiviral and foamy viral 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 knockdown 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, HB/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 de-repression. 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.

[0131] The efficiency of gene targeting, levels of globin gene expression in individual targeted cells as well as populations of cells and of their progeny, the extent of targeting on erythropoiesis and on stem cell function, and on hematologic parameters and organ function may be continued in model organisms.

[0132] 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.

[0133] 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 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.

[0134] 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.

[0135] 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 hemoglobinstation, 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

[0136] 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 reagents 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.

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).


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. 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).

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 (SRI) (Boitano et al., Science 329:1345-8 (2011) or HoxB4 (Watts et al., Blood 116:5859-66 (2010) and Zhang et al., PLoS Med 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.

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.

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 Deltal1 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 SRI (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).

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 permitting 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

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.

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.

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.

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.

**[0148]** 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.

**[0149]** 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.

**[0150]** 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.

**[0151]** 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.

**[0152]** 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 0.16% 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.

**[0153]** 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.

**[0154]** 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.

**[0155]** Using the defined optimal methods for generation of ex vivo expanded cord blood stem/progenitor cells, a bank of the 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.

**[0156]** The present disclosure will be best understood in view of the following non-limiting Examples.

**EXAMPLES**

**Example 1**

Selection of Bell1a Gene Targeting Homing Endonucleases Based on I-HjeMI, I-CpaMI, and I-Onul Using In Vitro Compartmentalization (IVC)

**[0157]** The open reading frame (ORF) of a parental I-AGL-1DADG 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(11)(Web Server issue); W110-6 (2012)), codon optimized for expression in E. coli, was cloned between the Ncol and Notf restriction sites of pET21a(+)(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).

**[0158]** Three rounds of IVC were conducted after each round of site-directed saturation mutagenesis in order to enrich variant nucleases 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 thor-
oughly mixed with the saturation buffer (100 mM potassium glutamate (pH 7.5), 1.0 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 16,000xg for 15 minutes at 4°C. Five hundred microfilters 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, Mass.), 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 16,000xg 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, Grand Island, N.Y.). 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.

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.

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.

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 - continued

<table>
<thead>
<tr>
<th>Target Round Site*</th>
<th>Sequence Identifier</th>
<th>Amino Acid Residues</th>
</tr>
</thead>
</table>
| TTAGAGAGAG| SEQ ID Y20, S22, T24, T31, R20, R86, R88 | Enzymes differ from those in the target site for the parental PIR-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).

### TABLE 3

<table>
<thead>
<tr>
<th>Base Homing Endonuclease 1-ClpA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>f 683/2508</td>
<td>ATGGAGATTGATC SEQ ID chr2: 60, 542, 847-60, 542, 868</td>
</tr>
<tr>
<td></td>
<td>Disrupts Bcl11a-X and XL forms</td>
</tr>
<tr>
<td>r 1589/2508</td>
<td>GCGGCTGATAC SEQ ID chr2: 60, 542, 630-60, 542, 651</td>
</tr>
<tr>
<td></td>
<td>Disrupts Bcl11a-X and XL forms</td>
</tr>
<tr>
<td>f 525/2508</td>
<td>GCGGCTGATAC SEQ ID chr2: 60, 543, 005-60, 543, 026</td>
</tr>
<tr>
<td></td>
<td>Disrupts all Bcl11a forms</td>
</tr>
</tbody>
</table>

**Example 2**

Optimization of Activity of BCL11A Gene-Targeting 1-HjeMI Variants Using Two-Plasmid Gene Elimination Cleavage Assay in Bacteria

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 Neo and NotI sites of the pENDO (FIG. 12, Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006) expression plasmid. NovaXGF
US 2015/016969 A1

Jun. 18, 2015

(EMD Millipore (Novagen)) competent cells harboring the pCdB reporter plasmid (FIG. 31, Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006); Takeuchi et al., Nucl. Acids Res. 37(3):877-8890 (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 the pEndo plasmid encoding the pEnonoM plasmid harboring the pCdB reporter plasmid containing the pEndo plasmid encoding the pEnonoM plasmid (in order to produce expression of I-HM84 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×M9 salt, 1% glycerol, 0.8% tryptone, 1 mM MgSO4, 1 mM CaCl2, 2 mM 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 plasmids were extracted from the surviving colonies on the selective plates.

[0164] The ORFs encoding active I-HM84 variants were amplified via error-prone PCR using the Gene Morph II Random Mutagenesis Kit (Agilent Technologies, Santa Clara, Calif.). After digestion with NotI, NotII, and DpnI, the resulting fragments were re-cloned 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 pEonBD reporter were grown in 2×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

[0165] Activity of an exemplary BCL11A gene-targeting endonuclease (BCL11A-Ag; FIG. 17, SEQ ID NO: 31), its catalytically inactive variant (BCL11A-Ag-ID188N), and its parental LHE-I-HM84 (FIG. 14, SEQ ID NO: 28) was measured in bacterial cells that harbor the pCdB reporter plasmid (Doyon et al., J. Am. Chem. Soc. 128(7):2477-2484 (2006)) containing 4 copies of either the target site for LHE-I-HM84 (H-I-HM84 target) or the BCL11A gene target (TCCA AGTGAATTCGGTGGTGTTG (SEQ ID NO: 39); underlined nucleotides differ from those in the target site for the parental LHE-I-HM84). The pCdB reporter plasmid encodes “control of cell death B” (“cldB”, 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 ±S.D. of 3 independent experiments.

Example 4

Detection of Targeted Mutagenesis at the Endogenous Human BCL11A Gene

[0166] HEK 293T cells (1.6x10⁴) were seeded 24 hours prior to transfection in 12-well plates, and transfected with 0.5 ug of each expression plasmids for the BCL11A gene targeting nucleases 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: Bel11A_UP1 5'-GCT GGA ATG GTT GCA GIA AC-3' (SEQ ID NO: 66); Bel11A_down1 5'-CAA ACA GCC ATT CAC CAG TG-3' (SEQ ID NO: 67). The PCR amplicon was incubated in 1×NEB buffer 4 plus 1×BSA (New England Biolabs) with or without 0.5 μM 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×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 protease 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: Bel11A_up2 5'-CGT CCA GCT CTC TAA GTT TCC-3' (SEQ ID NO: 68); Bel11A_down2 5'-TCC AAC CGC ACA GCA 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 FIG. 30).

Example 5

Selection of Fetal Hemoglobin Silencing Region
Targeting Endonucleases Based on 1-Onal Using In Vitro Compartmentalization

[0167] Exemplary homing endonuclease (HE) target sequences, which are evenly distributed throughout the 350 bp region (SEQ ID NO: 2) that includes the region of Bel11a 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>
<thead>
<tr>
<th>Position</th>
<th>Chromosomal Location</th>
<th>Sequence</th>
<th>Sequence Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>13/303</td>
<td>chr11: 5, 214, 235-5, 214, 256</td>
<td>TGGGTGCTAT</td>
<td>SEQ ID NO: 5</td>
</tr>
<tr>
<td>79/303</td>
<td>chr11: 5, 214, 165-5, 214, 190</td>
<td>CATGGGACCTT</td>
<td>SEQ ID NO: 7</td>
</tr>
<tr>
<td>143/303</td>
<td>chr11: 5, 214, 105-5, 214, 126</td>
<td>TGGGTGCTAT</td>
<td>SEQ ID NO: 8</td>
</tr>
</tbody>
</table>
TABLE 4 - continued

<table>
<thead>
<tr>
<th>Chromosomal Position</th>
<th>Chromosomal Location</th>
<th>Sequence</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>f 124/303 chr11: 5, 214, 089-5, 214, 110 ACTAAGTGAGA</td>
<td>ACTAAGTGAGA</td>
<td>SEQ ID NO: 9</td>
<td></td>
</tr>
<tr>
<td>f 200/303 chr11: 5, 214, GCCACCACTTT</td>
<td>GCCACCACTTT</td>
<td>SEQ ID NO: 10</td>
<td></td>
</tr>
<tr>
<td>f 211/303 chr11: 5, 214, 037-5, 214, 058</td>
<td>TCTTGAAATTAT</td>
<td>SEQ ID NO: 11</td>
<td></td>
</tr>
</tbody>
</table>

[0168] Table 5 presents a region from -100 bp to 210 bp upstream of globin genes, which is identical for both Ay- and Gy-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.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCTGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 16</td>
</tr>
<tr>
<td>TCTGGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 17</td>
</tr>
<tr>
<td>TCTGGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 18</td>
</tr>
<tr>
<td>TCTGGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 19</td>
</tr>
<tr>
<td>TCTGGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 20</td>
</tr>
<tr>
<td>TCTGGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 21</td>
</tr>
<tr>
<td>TCTGGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 22</td>
</tr>
<tr>
<td>TCTGGGAAACGGTCCCTGGCTAAACTCCACACCATCAGGTTGCC</td>
<td>SEQ ID NO: 23</td>
</tr>
</tbody>
</table>

[0169] Table 6 presents amino acid positions within a parental LHE L-ONuI homing endonuclease (FIG. 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>
<thead>
<tr>
<th>Amino-acid Positions Subjected to Saturation Mutagenesis in IVC to Create Targeted Homing Endonucleases against a Human Fetal Globin Silencing Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>TTTCCCAAGTATGGAGCCCTTCTTTATA</td>
</tr>
</tbody>
</table>
TABLE 6-continued

<table>
<thead>
<tr>
<th>Round</th>
<th>Target site*</th>
<th>Sequence Identifier</th>
<th>Amino-acid residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>TTTGAAATTATTTCAACCTTTTA</td>
<td>SEQ ID NO: L26, R28, R30, N32, 41</td>
<td>S40, E42, G44, K80, T82</td>
</tr>
<tr>
<td>1</td>
<td>TTTCCATTTCCATTTCTTTTA</td>
<td>SEQ ID NO: F182, N184, V199, 201, 01</td>
<td>K225, K227, D236, V238</td>
</tr>
<tr>
<td>2</td>
<td>TTTCCATTTCCATTTCTTTTA</td>
<td>SEQ ID NO: F182, N184, I196, D190, 43</td>
<td>K191, Q197, V199, V238, T240</td>
</tr>
</tbody>
</table>

FIG. 23 presents the results of a of a cleavage assay with ‘half-targt’ 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-Onul homing endonuclease, which are responsible for the lcr (L) and right (R) half sites of the gGlobin silencing region target.

Active I-Onul 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

FIG. 24 presents the results of a cleavage assay with ‘half-target’ 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-Onul homing endonuclease, which are responsible for the lcr (L) and right (R) half sites of the gGlobin silencing region target.

Active I-Onul endonucleases are selected from a pool that cleaves the full-length human fetal globin silencing region target.

Example 7

A Cas9-Based Endonuclease System for Disrupting a Bcel11a-Regulated Fetal Hemoglobin (HbF) Silencing Region

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 hemo-globinopathies that are disclosed herein: (a) to disrupt a Bcel11a coding region; (b) to disrupt a HbF silencing DNA regulatory element or pathway, such as a Bcel11a-regulated HbF silencing region; (c) to mutate one or more Y-globin gene promoter(s) to achieve increased expression of a Y-globin gene; (d) to mutate one or more D-globin gene promoters to achieve increased expression of a D-globin gene; and/or (e) to correct one or more B-globin gene mutation(s). The bacterial CRISPR system is described in Jin 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).

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 G19 -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.

Thus, within certain aspects of the present disclosure, Cas9-mediated genome editing is employed: (1) to disrupt the Bcel11a binding site within the HbF silencing region, (2) to disrupt Bcel11a 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.

[0178] 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 | Sequence Elements of a Generic Cas9 Guide RNA |
| Description | Sequence | Nucleotide |
| U6 Promoter | SEQ ID NO: 44 | GGACGAAACACC |
| Generic Target- Sequence | SEQ ID NO: 45 | GNNGGGGGGGGGGGGGGGGGGGGGGG |
| Guide RNA | SEQ ID NO: 46 | GCTGGGCTTCTGT TGCAGTAGGG |
| Scaffold | SEQ ID NO: 47 | GAAAATGGGAGAC AAATAGCTGG |
| Poly T Tail | SEQ ID NO: 48 | GCCATTTCTATTA TCAGACTTGG |

| TABLE 8 | Target-specific Sequences for Exemplary Cas9 Guide RNAs |
| Designation | Description | Use | Sequence | Nucleotide |
| GCN200G-B | Targets the GATA-1 recognition motif | Used singly, with #C to delete the putative Bcl11a and GATA-1 motif jointly, or with #K, #P or #G to delete the entire Bcl11a ChIP peak and surrounding sequences | GCCATTTCTATTA TCAGACTTGG |
| GCN200G-C | Targets the putative Bcl11a recognition motif | Used singly, with #B to delete the putative Bcl11a and GATA-1 motif jointly | GCCATTCCATTA 7C6ACTTGG |
| GCN200G-D | Targets immediately adjacent to the putative Bcl11a recognition motif | Used singly, with #B to delete the putative Bcl11a and GATA-1 motif jointly | GCCATTCCATTA 7C6ACTTGG |
| GCN200G-E | Targets downstream of the Bcl11a binding peak | Used with #B and/or #H to delete the entire Bcl11a ChIP peak and surrounding sequences | GCCATTCCATTA 7C6ACTTGG |
| GCN200G-F | Targets downstream of the Bcl11a binding peak | Used with #B and/or #H to delete the entire Bcl11a ChIP peak and surrounding sequences | GCCATTCCATTA 7C6ACTTGG |
### Example 8

**Vector Systems for Expressing Endonucleases**

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:

**[0179]** 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 vectors, and is able to mediate significant levels of transient gene expression in HSC.

**[0180]** 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 (AdS-F35) was demonstrated by Dr. Rawlings (SCH) and more recent data suggest that the serotype 11 fiber (AdS-F11) 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.

**[0181]** Integration-deficient Lentiviral and Foamyviral Vectors (IDLV and IDFV). These vectors provide 6 kb (IDLV) to 9 kb (IDFV) payloads, and have well documented capabilities to transduce human HSCs. Both IDLV and IDFV vectors can be used for gene knock out 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.

**[0182]** 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.

**[0183]** 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.

**[0185]** 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.

**[0186]** 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 assessing 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 narrow space.

### Example 9

**Characterization of Homing and Cas9 Endonucleases for Efficient Gene Targeting**

**[0187]** 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.

**[0188]** 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 erythroid production 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.

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 □-globin/□-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 □-globin mRNA as well as Hbf expression at the single cell and population level.

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.

N-MEL cells are efficiently transduced using the Amasys nucleofection system. Using 2 μg 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.

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.

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 □-globin/□-globin ratio of 0.1%. After infection with a Bcl11a knockdown vector 20-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.

Due to concerns that altering Bcl11a pathways may lead to a relative increase in □-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 CI, providing a direct measure of both the level of β- and □-expression, as well as the ratio of □-globin/□-globin.

To determine the percent of cells that show an altered □-globin/□-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 □- and β-globin RNA simultaneously for several cycles and once adequate material is present to allow for accurate splitting of the sample, □- and β-globin are assessed by qRT-PCR as above.

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 ratio of Hbf° to HbA and HbA2 are determined. The number of cells expressing Hbf° are compared pre- and post-transduction using gluteraldehyde fixing cells, permeabilizing with detergent and adding an Hbf° specific antibody followed by quantitation by flow cytometry.

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 grade CD34+ Hematopoietic Stem Cells

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 Dony. Giarratana et al., Nat. Biotechnol. 23(1): 69-74 (2005). The qRT-PCR assays above reveal a □-globin/□-globin ratio of 4%. This low ratio allows for the sensitive detection of increases that are secondary to genome editing.

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 cytoplasms 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.

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.

[0201] Clinical effectiveness is assessed in vivo in mouse models of hemoglobinopathies. Knockout of murine Bcl11a leads to a dramatic dose-dependent increase in $$beta$$-globin in mice containing a human $$beta$$-globin locus on a transgenic 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.

[0202] 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 microtiter 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. Piszty et al., Science 278(5339):876-878 (1997). In addition, the $$beta$$-globin/$$beta$$+$$gamma$$-globin ratio is assessed in mice containing the A25 and A85 human $$beta$$-globin YACs. Porcu et al., Blood 90(11):4602-4609 (1997).


---

**SEQUENCE LISTING**

<160> NUMBER OF SEQ ID NOS: 69

<210> SEQ ID NO 1
<211> LENGTH: 3603
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

ccagtgaagca ggttggttga agataaaggcag ggtttcatta gtttgagaag aatgaaatag
60
aacaccttc acactatgcc ttaaccttgc cttgagactg aagttctgtg cattgcttgc
120
ttgacctgag aacccgtgttc ttctctcatc atctctctct gcctctcttc tcagcagttg
180
tttctaaaaa tatctctcata gttttcattt tgcgaactgc ttatagcttga
240
atgtatctta gacgttaact tattgttgtc tcgaacatcta tactaagaga acctatctaa
300
aagcagatgt tttaccaagct gtgcatcata taagctcttt cttggcctct aagatctttt
360
gctctccagg gacatttttag aagttctgga ggaaccatct atctgctgcag tttaaacttg
420
cattgcttct cttcctctgg gtaaatcttc cttcataataa aaccttaatc ttttatattta
480
ttgaaatctaa tttcattaca aagatatct aaataatga tagtttttta ttacacagtt
540
tgaaataggg ttgtaggtgaa acgaatattta aggttaagct tttaatgcac gataaactca
600
aatatatatt acaaataatttt ttaaatgtttt aaataatgtg aagatctaatg aaagaattga
660
aagttgaaag tagatatttccc ataatcgcgc agatatataa taacagagtt cctgggaaata
720
tgacgaatctt caaatcttttt ccagattgtc tttaaaaaga aatggttggaa ttgaaccagg
780
gaatgaagtt cccagactttg tttatacctt ttoattgca ctatattgcc acatctggcc
840
acaataaca acccatgagc atataaaaaa gggagatttg aaggaagagag gaggagaaga
900
tggagaaaggg aagggagaag agggaggagag aagggagtgt aagggagaag
960
gagaaaaaaa taaaagggag gggagggaga aagagggag aagagggag aagagggag
1020
agggagaag gagaaggagag ggggaagagag gggagaagagag gagaagagag
1080
agggagaag gggaaaaggg gggaaaaggg gagggagagag aagagag
1140
actcagctt caaatctttgc tggagatacg gatcaataga aactcataact cttgcttgta
1200
cactctataa cattctatac ctttggtgaaa acctctctta tggcttatac cgttgtagtt
1260
-continued

cctgataagtcattaccta gcaattacat tttagatat tcgaaacac atgcactgtgta
1320

agtccaaag acatgttaca aatgtcttt gacgacaata atcaaaaaa cttaaaacgg
1380

gtacgctccct aatgtctac caacagtaga attgataa attcgtatat tcggcagaata
1440

gaatattaca cagaatagaa aagatccaca ttgctgaa ccgctagcga taccaatgca
1500

tttacagca ttggggcgtat taaagtaag ccagagctcag ttcacacat atgcacccatt
1560

tttgcataag aaagagatgat gatggtgcttg agcatcttgat ccacagggcga aatgtaataa
1620

aatagggtag agagaggagca tgaatgacct ttaaacccttg ttacaagttta ttctctctgta
1680

acctggagaac ccagcagacaa tatttgaataa ttaaagaaaaa tgggtggtcat ggtgtggtatal
1740

gttcattttt aagtagacca ttcaattgta gaaagaatgta ttattагagт agaагагаagт
1800

gggcgagcaaa acgttgccgct ccgtggtcagt aaggagaaaaa tggagaatgc acaacttataatt
1860

atcagacttg agacatcgaag tggatgtcag tggtgtaacac aagaataggcc aaccattttgt
1920

gagttaattg ctgctgataaa atcagaaaactccttctgga taactgtatt tgamgtcact
1980

cctgttaaat gggaaagat gttctttagat atgtgattct ggagctcact aagacatacagaa
2040

cccaggtcaac aatagagagac ggaacagagac atacaagaaag aatccataaag tgggtgtgtgtg
2100

aatgctacact aactcctgtt taaaatagca ggtgatgtac tctactaaata tctcttatatt
2160

atcttagaat gtgttgtgga gggagatttt aagtttcctca aactataata agaaaaaattga
2220

ttttaaaaat aatgtcattgt aaacttctaa ggccacatcgag acctgtgtatt gatotgctgtc
2280

ttgtaattcc aacagccctgg cttttctactt taattctgtt tcaacttgaat tgaatttttac
2340

catggggaga tctgggtatca ttgtttttct aatagattgtc ctattgagagg attttgtaa
2400

cagaataagg tttttttttttttttctgcc tgaattttgct ttctctcctc aatggcacaag
2460

ctagatcga aatgttcaat ctgggtctac ccgcaatctct gcttatggggt ttcgacaaad
2520

tctctgctct cagcaacctcg aatgtgcgag ctggtgacca tgcacacaca ogccggcgtga
2580

tttggatttt ttaagatagca tgggtgttca acagttttgt cagggcgttcc tggagatcct
2640

gacgctcaggt gctctggtcct ccaagtctgt ggtcttacca gcttgaacca
2700

cgcgtcagcc cccgggctaa gtttcaaaaag ggagaatca caactctcct tggctcagga
2760

aacacagtgg acctgtcaggg gcagtagata tagctattata tagataggga caaatgtgaag
2820

cagagcagca cttggtatat tagggagga aatgatttttta aacgtttgacc tggggaattt
2880

gggaaggaag tactttatga cttsaatgga aaacactgga agagaatccag ttttgatata
2940

aacaagtaaa taaccatgct ggtctgcttg tggcttatag acatcttat aataaacgccc
3000

agtagactggt tttgttagct ggtttgccac tggggtttgga gctctagggat tagaattcgc
3060

aggtgggata ctgggttcatat gttagatttt ctgatatact gttgaatccaa
3120

agtgggtgatg atacaaagtt taaaatttg aaccttttta aacacacaact tttttatgtga
3180

aactatcaag aagttgatag cccaaattatt cccctgagatatt gaaactaatc
3240

agtattgattct ttaaatgtag ctaagggctt ggggtttaga atggaaagtc aactgtgtgtg
3300

agtttgacag ctgggttagag ataagttttct atgtgtagcc aagcatggagaaa gcttagaacaag
3360

aaagggaaag ggaggtgagag atgagggagaa tgtctgtaag gagatgagc cacacttgtat
3420

gggaggtgata ctaaggactc taggggtcga gaaatagggg ttcatttctct ctaaattctaatt
3480

ccatcttgt gcctgtgtgtg gcgggtcaca cctggtgtctc cggcattttt tggaggggag
3540
gaggtggtgcatcctgagtctagtagctctactag gaccaagccc gca 3600
ta 3603

<210> SEQ ID NO 2
<211> LENGTH: 350
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

aaagatggtgagtgctgtagagtgctgcatcacaagggc aatgtttataa aatagggtag 60
agagggcacc gtaaagtaatttaaatttt ttttctgtta aacccgggagc 120
cagccaaggtagttattatagctcatgaagggcat gttggtgcat gtttgatttt gttctcttttaa 180
agagggcctgctgtgcaggtgagggaca catttcattactcagactgtggcagagagaaatg gagaacaaat 240
gaccaagccc gcaagaggtgagtgctgctactggtcacc cacaagtggg cacagttttgc 300

<210> SEQ ID NO 3
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

gtgcaacctca ctccagaggg g 21

<210> SEQ ID NO 4
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

gtgcaacctca ctccaggtgga g 21

<210> SEQ ID NO 5
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 5

tttcaacactta ttcaccttt ta 22

<210> SEQ ID NO 6
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

tgtggtgcctta ttctgtgttt gc 22

<210> SEQ ID NO 7
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 7

cattgtcact ttctccctta ct 22

<210> SEQ ID NO 8
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 8

```
taaatatcat ttcttacacta ag 22
```

<210> SEQ ID NO 9
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 9

```
actaagtggg aataatcttt ta 22
```

<210> SEQ ID NO 10
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10

```
gccacacct ttcttgaatt at 22
```

<210> SEQ ID NO 11
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11

```
tttggaatta ttcataatct tt 22
```

<210> SEQ ID NO 12
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 12

```
tttaaggtca ttcgatggtc ct 22
```

<210> SEQ ID NO 13
<211> LENGTH: 2606
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 13

```
gcaaatggaa ttaaagtttt ttataagggca gatacagagat gctcaagggc cttctataata 60
tccccccccctccggtgtagttgggctacttagggggattactacagacaaggggattccgg 120
aagaaaagca gcctagctgt accgctctgt gcagccatta gccacaccac gcaccaccttt 180
tggatatgca gcgtgtgactg gtaggtgtgaa ttctttgcca aaggtgatggg ccagccaca 240
gaccagcagc tgcccccaag gccgttgccag gaagataaga ggtatgatgc atgattgacaa 300
aagggctcatg cttggaacata gaataatccga gctttatccc aaccataaa taaagcaga 360
attgagtgcg atgatgtgct gctatagca atatgaaacc cttctacatca gttacaattt 420
atatgcagag ataatgatat gcaagatata tgcctattgcc ttaaccagaga aattatcaact 480
gttaatttctc agaatgtggtgc aagagagcat gatatatgtgt atcattatat cctggaaga 540
aagagattgg ggaagtctgta aagatasaagaa aagatattaas aagagaag 600
```
ccttttttaa aatticaaat gcaaaatattc ccctgatttgg tcaataatgty tacaatattt 660
aaaacactaa accttttaaacc ataataaatgt ataatgttaa tgtaatcactt ataaaataaaa 720
gaaatataag taggggagat atgaatagtc aataaagcag actatatattc caaaatgtaa 780
tgtcataggc agactgtgta aagttttttt ttaagttact ttaatgtact cagagatatt 840
toccttttgt atacacaagt ttaagggcat attaataata gaatatataa gcgggaagaa 900
aaaanaagaa agcagataac ataaaaaag aaccaattgtat tatgaacagc aataaaaga 960
aactaaacag atctctgagac ttccacaacttg atgcaaatcat tcgtcctgttt cccattcctaa 1020
actgtacccct gttacttacte ccotctacct gocatgaact taaccaataa aaaaaaaggg 1080
aaaaagaaaa tacccagctta cataagccaa cccgtgagtt ctcaggttcc acgtgcaacgt 1140
tgcacagcag acgccttgcct cgcgtggcga aaggttcccct gagggttggc aggttacgcca 1200
ggccatacact aagggcaaga cagactctttc tcgcacattgc ttcctcctta ggttggcaca 1260
tacagcaactt aggaagttgac agatccccttaa aagactcctaa gacacctccg ttgccaagtt 1320
agcacaaccc caggtcttaac ggctggaagc tctctctccag ggtcagcagtt cccotatgtc 1380
tccagcaaccc aagacgcttc tctgctctcct cagctgccagg tctctatggc ctcgcttaac 1440
tgtgcctttga aaccgccttacca caacggctcct aagggcctcct aacacatgctg 1500
acccctggcc aacgccgcat aacggcctagc tctctcctcg aggtcagatg tccacgttgc 1560
tggttgag ggtctgaagtt aacaccagtt ttgctgagac aataaagatgc taatagttgc 1620
tctgctcctga gttgcctgctt ggcggctcct gctgccctgg gcggtgcttt 1680
ggcgctccctt ggcgtgtgctt gtcgctttag ctgctggactt cccgttctac 1740
cctgtcctttt ccgaggtttgg cccgctcctg ggtctgtcagc tggggtcact 1800
atatgtcctt gtgccccata ccctcagctac aacctctac taaaaccactt tctcctgccttga 1860
agttctttaa cgataatattc gaaaaaaccagt ttggaagagt cttgtctcctag 1920
atatctatct tcgcttctct aaaaacttaa ctgcgctttccc atttaaatag 1980
aatgatatgta cgtatatattc gaaatccagc cttgatgaagc atatgataa aatctttccc 2040
actctcagag gatagatttc aaaactccta cctggtctact actcatctac tctccttatt 2100
tgattaaaa catttttgcgt ataagaagaa aataatatat atatatgtggt gtatatatac 2160
acaccatact atacatttct aatcggaacttg gttgcttggt ctttcttaatt tctgcctatgc 2220
atgctaatata tatttctgta aataaatatat taaaacatct atggtgtacttg tggctcattag 2280
atattttcag ggtgtttggga aacaattgga aaaaaattg aagacttttta atggatgaagg 2340
ataataacat ttatataagact cccagccctttt agcatctcactt cgccttgaataa 2400
gtgggtctca cgagaggggt ccttgggcac actcagctgt tcgtgcttgg aacactcttg 2460
aaaggtgtctct ggggggacca aagcctcctctg tcgctgctgag aatgaggggg aagggggaaa 2520
aggtctttca cttggtctcag attttttttt tctctctagtca cacaagatt gatgcttcttt 2580
aaatctttt cccttgtgccg agttcc 2640

<210> SEQ ID NO 14
<211> LENGTH: 606
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 14
caccctgag ttctcaggat ccacgtgcag ttgtagctcag ttagcagcctc tcaagtgtgc 60
aaaggtgccc ttaggtgcct ccaaggtgagc caggcactca ctaagggcag cagacacttt 120
cctgcacatca gccttcacgt ttagggtgccc ctaaacagctc tcaaggtgtag acagatcccc 180
aaagacctca aagaacccctt gggtcacaagg gtagaccccc aggcccttaa ggggaggaaa 240
ataacacctg aggccagagg agtcaggtgcc tatagagaaac ccaaggtgtc tctctgttccc 300
cacagcgcct ggctctatgg gcctctctaaa acctgtcttg taaacotggat accaaccctg 360
caccggtctcc accaccaact ccataaggctgc ccacaggggca gtaaggcag 420
actccctccc agggacagca tggcaccatgg tgtctgttgg aqgttggtag tgaacaaagct 480
tggtgcagga ccaaattgaa ccaataagatct cctgcctggt gctcttttatgc ccaagccctg 540
gctctgccc ttcctgctcc tggggagaga ttgcaacacc ctaaggtgtg gctccccacagg 600
gtcaggg
<210> SEQ ID NO 15
<211> LENGTH: 303
<212> TYPE: PRT
<213> ORGANISM: Ophiostoma novo-ulmi
<400> SEQUENCE: 15
Ser Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr 15
Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn 30
Asn Lys Ser Ser Val Gly Tyr Ser Thr Glu Leu Gly Phe Glu Ile Thr 45
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Asn Glu Ser Thr Trp 60
Lys Val Gly Val Ile Asn Ser Gly Asp Asn Ala Val Ser Leu Lys 75
Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys 90
Tyr Pro Leu Ile Thr Glu Lys Leu Gly Asp Tyr Met Leu Phe Lys Glu 105
Ala Phe Cys Val Met Glu Asn Lys Glu His Leu Lys Ile Asn Gly Ile 120
Lys Glu Leu Val Arg Ile Lys Ala Leu Asp Val Gly Leu Thr Asp 135
Glu Leu Lys Ala Phe Pro Glu Ile Ile Ser Lys Gly Arg Ser Leu 150
Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser 165
Gly Glu Gly Cys Phe Phe Val Asn Ala Lys Ser Lys Lys Lys Val 180
Gly Val Glu Val Glu Leu Val Phe Ser Ile Thr Glu His Met Ile Lys Asp 200
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile 215
Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Thr Val 230
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Glu Glu Asn 245
<210> SEQ ID NO 16
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 16

```
tgggggccct ttcctccacat tatctaatgtcaaatatcttg tctgaaacggg ttccttggtct
```
aactccaccc atggggttgccc agccttgcccttgaccaataacccggtcggc

<210> SEQ ID NO 17
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17

```
tgggggccct ttcctccacat tatctaatgtcaaatatcttg tctgaaacggg ttccttggtct
```
aactccaccc atggggttgccc agccttgcccttgaccaataacccggtcggc

<210> SEQ ID NO 18
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

```
tgggggccct ttcctccacat tatctaatgtcaaatatcttg tctgaaacggg ttccttggtct
```
aactccaccc atggggttgccc agccttgcccttgaccaataacccggtcggc

<210> SEQ ID NO 19
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

```
tgggggccct ttcctccacat tatctaatgtcaaatatcttg tctgaaacggg ttccttggtct
```
aactccaccc atggggttgccc agccttgcccttgaccaataacccggtcggc

<210> SEQ ID NO 20
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 20

```
tgggggccct ttcctccacat tatctaatgtcaaatatcttg tctgaaacggg ttccttggtct
```
aactccaccc atggggttgccc agccttgcccttgaccaataacccggtcggc

<210> SEQ ID NO 21
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
-continued

<400> SEQUENCE: 21

**tgagggcccoc** tcccccaacat tagccgtaag ccaccatgtg ccctgaagtg ccctggccta**

60

aaactcaccoc aggggtggtgag cacgcttgc ctcgcaatag ggcttgaca**

110

<210> SEQ ID NO 22
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22

**tgagggcccoc** tcccccaacat tagccgtaag ccaccatgtg ccctgaagtg ccctggccta**

60

aaactcaccoc aggggtggtgag cacgcttgc ctcgcaatag ggcttgaca**

110

<210> SEQ ID NO 23
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 23

**tgagggcccoc** tcccccaacat tagccgtaag ccaccatgtg ccctgaagtg ccctggccta**

60

aaactcaccoc aggggtggtgag cacgcttgc ctcgcaatag ggcttgaca**

110

<210> SEQ ID NO 24
<211> LENGTH: 2508
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 24

**atgctgcgc cagaccaagc cccacccccc ccaacgatag ccggagcagc ctcgccgag**

60

cctcttgag ccattcttac agatgtggca acacacccgg gcgcgcgtgg ccgctccgagaa**

120

ggggcctgag aacccctag ctcgctgagc gtcgctgtgta cccctcac ccgggcacag**

180

cattttactt tagcgcacaa acgcatacag tgcgtgctg agacagagct**

240

gtggatagcc cccccctata tccacaaac ccacatcagctgc ccctggcgacac**

300

gtccgacatg aggtagcagc gttgggggtc atggggttag ctagtggtggtt**

360

tgcgcgcacag aggccacagt acgtgatgtac ctcgctggag ccccccacgc**

420

cggtattta aagtgacgct agcagcactg cccatagcgc agggcgaga**

480

gggggacagcc gcggagggct ccctgcacct gcgcgcgtgc ccctgcggcg**

540

tagccgatcag gctgctgctg cagacacacag acacacagctg cattggtgac**

600

ggctggtgac gattcagcagcc acacacagctg cattggtgac**

660

tgcgcgctgc aggggggaac cccacccccg gcggagggct ccctgcggcg**

720

tctggtgcgt gcgcgcgtgc gcgcgcgtgc gcgcgcgtgc gcgcgcgtgc**

780

cacccacccc cccctgttag ccacccgag acagacactg cccaccceaga cgcgatagag**

840

cgcgctgggg gcggagagat gcgcgcgagcc acccgcctacc gggggtcggc ttgcaggggtg**

900

tcgcgggtgca atccacagctg ccccgccactg atctctttag caggaggtag**

960

gagcggcgg gcgacacgct cccacccgag ccctgctgctg ccctgcggcg**

1020

gggggacagcc gcggagggct ccctgcggcg gcgcgcgtgc gcgcgcgtgc**

1080

tgcgcgctgc aggggggaac cccacccccg gcgcgcgtgc gcgcgcgtgc**

1140
ttctgaggca agacgctcaaa atttcagagc aacctctgttg tgcacccgca cagccacagc 1200
ggcggaagcg cctcaagtq caacctgtgc gcacacgqct ggcacccagcg cagcagcttg 1260
aagcgc codecs tgcggagcac caactcacaac tgcctcccaac tgcaggtcaac gtcggacagc 1320
gttcttcaca cccgcccagct cccggagac ccgcaccccg acttggctgg cagccgcagc 1380
agcogctca agtcctgttg tgcgaagtct aagcagcagc aagcaccacca aatgatccgg 1440
gagaacgggg acaaacggagg aagggagagc gacaggaagc aggaagaaga ggaagaagag 1500
gagggagagg agctgagggc gacgagagag gttgcagcaag gttgagttgct gatcgttgag 1560
gcgccgacc accgcggagca ccagctccgg gcgcgctcgg tcggcgctgg gcggagagag 1620
cogcogcctgc cggagcctctc gggttcgagt gttgtcagctct ccgtgcaagc ttttcagcag 1680
gtcttcacac agttccttgag cggagagcag caccgcctcc gcctggccgc gcggagagag 1740
ccacagggcaca cttggctgca agcctcgttg gcggcagagc tcggcagcagtc agcctgagc 1800
acggtaatc gcggcctgtgt ctccgcgggc gcgtgtcct gcggcagctgc tccaaaag 1860
cctgtgctgg gcgcgcgcgc ccgcctgcgc ccccttccta agcgcacatacc gttgcagag 1920
gagtcgcaac tgcgcgcgcg gcgcgcgcgc caccgcctcc gcgcgcgcgc gcgcgcgcgc 1980
gcggctcagc gcgcgcgcgc ccccttccta aggctcctcc tttgtcctcg agtcacctag 2040
caacgccctt ttcgcgcttc gcggcagcgc ttcgcgcttc aggctcctcc gcgcgcgcgc 2100
acccgccccg ggagcgtcgg cggcgcgttc tgcggcgcgc gcggcgcgcgc gcgcgcgcgc 2160
agccgagccc aatagggctgc gccgggcgcgc gcggcgcgcgc gcgcgcgcgc gcgcgcgcgc 2220
agcgacactt gttgtcagct tggggagatt cttcggatact tgcagcactct cactgctccac 2280
aggagaaggc acagggaggg aagggcttct aatggcagcg tgcgcacata cgcgcgcgcc 2340
cagcactgca agtcgcacag gcacactgaa aagccatgcgg gggtgggggg gagaagtttac 2400
aatggtaaa ttgtaatagt gcctttttgg tgtctcagctg ccctggagag gcacatggaa 2460
aagggcaca ggtgatgagt tgtgaaaatt cgtaaaag cttgaattag 2508

---

<210> SEQ ID NO 25
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

atgggatca tattgcagac aa

---

<210> SEQ ID NO 26
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

agccatttga ttcaaccgca gc

---

<210> SEQ ID NO 27
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

cacacagcga ttcaaccagtc ga
atgggatccc acatggaccc gacctagct taccgggtg gctcggtcga agggcgacggt
tagtttctta tcaacaaaaa gggttaatac ctcagccagt aactgggtat cgaactgtc t
tatccagct cgctagtct cttcaggtct ccgcctgctg tcttctcctg gatcactctg

aaaacactca tctgccgat ccgtagcag tgcgaacaag aacagtaga cgcagacgac
tactgccgt tcaagacgcc tctctgtct ctacttactc tacatattct taacactgca
tgctgtgcct ggttggtggtt catcgagct gaga gggctgct tctttaccc ccaactgaac

aaagatgcag attacgctgt cgtctcttct gacatcgctg agaagacggc tgggtcatctg

atcttggca tacaagatt cccttgcttc cctgtctgtt accaagggac aacagcgga cccaaaccaac
tcggtcgag tcgaagtgcg cctgtctgtat cttcttacca accaggtgaa acttactacg

agtggcagcc acatggaccc gacctagcc tccacggtat gcccttgctg gggcgccggc
tatatttacg cggcagatg tccgcttgg acactgggcag cagcctctcc

atcagaccgc tcaggcatc ctacagatcc aaggacatcc ccggcgttggg caaaggtgtct

ttttagcagc ggacagcagt cgaagtggctg agcctagccg gacggcacaa aacaccaacctg

aagacaattc tccgtcatc ccagggcaag tccaggctc tggccagaaa gcgtaagcac

tgctcgagat ctcgcaggct tgtctgttcct acatacttc atagcagcag cccgcctgag
taaccgagaa gcaagcagtc aatcactagc gttcaacagc tcaacaaactctactacctc

agcagtgcag tggcagcgtt ctcgagcc gggcgctgtc tcaacacca aacagctacg aacaggtcagc

agaagacgcattttcctgc cgcggcagct ctatagcacc aagagcagcg ccacatttctc

atctcggct ccacaaatct cgagcatct tccagcacaaga ccccccaactcta cagaagcgat

tggcagcgg cccggtctgt gggtaacag aagctgcagc acaacacgtt gataaaacgag
tgctcagcaca cgcagacgaa atcccagcgc ggactacttc c
US 2015/0166969 A1

-continued

<210> SEQ ID NO 30
<211> LENGTH: 254
<212> TYPE: PRT
<213> ORGANISM: Trichoderma reesei

<400> SEQUENCE: 30

Met Gly Asp Leu Thr Tyr Ala Tyr Leu Val Gly Leu Phe Glu Gly Asp
1   5  10  15
Gly Tyr Phe Ser Ile Thr Lys Gly Lys Tyr Leu Thr Tyr Glu Leu
20  25  30
Gly Ile Glu Leu Ser Ile Lys Asp Val Gln Leu Ile Tyr Lys Ile Lys
35  40  45
Asp Ile Leu Gly Val Gly 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
65  70  75  80
Ile Leu Pro Ile Phe Asp Lys Tyr Pro Met Leu Ser Asn Lys Glu Tyr
85  90  95
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
115 120 125
Asp Ser Ile Ile Asn Thr Ser Tyr Phe Ser Ala Trp Leu Val Gly Phe
130 135 140
Ile Glu Ala Glu Gly Cys Phe Ser Thr Tyr Lys Leu Asn Lys Asp Asp
145 150 155 160
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
180 185 190
Leu Asp Lys Thr Asp Ser Arg Leu Lys Val Thr Gly Val Arg Ser
195 200 205
Val Lys Asn Val Val Lys Phe Ile Gln Gly Ala Pro Val Lys Leu Leu
210 215 220
Gly Lys Leu Lys Leu Gly Tyr 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
245 250

<210> SEQ ID NO 31
<211> LENGTH: 768
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

<222> OTHER INFORMATION: Nucleotide Sequence for a BCL11A Gene Targeting Nuclease Based on the Homing Endonuclease I-HFEMI; Codon Optimized for Expression in E. coli and Obtained through Directed Evolution in IVC and in Bacteria

<400> SEQUENCE: 31

atgggatcccc acatggacct gacctacgct taccgggtg tgctgcttcca aggggagcgt 60
tacttcacca tgctaaagc ggtgaagtac ctgaattaag aagctggtat cagctgtct 120
atcagaagct ctctgggtat ctacaaaaat aagcacatcc tggggtgttg tactggttat 180
ttccgggaat atagggagca tgaattggt ttctgctgga ttccggacaa gatacctg 240
-continued

```
aaaaacctca tctgcgcat cttgacaaa taccgagtc tgctctaasac gcagtcgac
300
```

```
tacctggttc tcgaagacgc tcctctgtct aacatattct actgtaagca tgtgcaagaa
360
```

```
tacgttcgttc ctaacagtcc tataactct cttgacctc ttacaaacac ctcttacctc
420
```

```
tctgttgggct tgtggttgct tcctgagtc agaagttctc tcctgaccta caaagcgct
480
```

```
aagaatagtg agctggacgc tggtgctcgt atcgtgtcga aagacgtgta actcgtgacc
540
```

```
tctgctatcc acaataactc gtttcctcacc aacaagcgct tacaagcacg aacaaacagc
600
```

```
tctctctagc ggtgacccc gtcagcgtct ctttccacgt ctgaaacagc tgggttaaat tccttagggt
660
```

```
gtcctgtaga aacgtgtgcc taacagccaa ctgcgtacca aacttggagc caacaagctg
720
```

```
cgttaaaacct ctctggtacc tgcagaaaa tctgtgcaagc ttaactac
786
```

<210> SEQ ID NO 32
<211> LENGTH: 786
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FRAGMENT:
<223> OTHER INFORMATION: 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> SEQUENCE: 32

```
atggtgatgcc acatgcacct gacotacgcct taactggtct gcctgttcca gggcgacggc
60
```

```
tatttttacaa tagctagggc cgacagattgt cttgacaca acgctggccc catcacttcc
120
```

```
atcaagggactc tcagaatcc tctcgaagc aacagatccc ccgctgaggg ccaaggtgac
180
```

```
ttttagaagtc acagcagaca tgaagttggt ctgccttgagc ttacggccaa aaaccctgct
240
```

```
aagaacatccatc tctgtgcttt cttgacagc taccctctgc agcagctgct cagtcagct
300
```

```
tacgtgatct ctaaaggtgcacctctct acatctctcc aagcagacg gctgctggag cagtcagct
360
```

```
tacggccct gcacagctgc aacatagcct gcaccgactc caatcagct ccctaggagc
420
```

```
agcgcctgggc tgggtgggttt cattgagggc gaggctctc tcaccacta caaggctcct
480
```

```
aagacagtt gcacaagcgc ggpctttccc atgcctcagc aaggtgacgc cattcctacc
540
```

```
tcgccagac acaatatcct gacgccgcc acacaaaacc ctaaagagac gacaaagctg
600
```

```
agccacacta agctcagagg ctgagacgg gcaccataacg tggtaaatgt ctccagcgg
660
```

```
agcgcctgcta agctgtgggt taacaaagag ctgcagtcac aacctttggg aaaccagcctg
720
```

```
cgcggtacct cccgtatacg cgagaaacct ccgctgctcc acaactac
786
```

<210> SEQ ID NO 33
<211> LENGTH: 256
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FRAGMENT:
<223> OTHER INFORMATION: Amino Acid Sequence of a BCL11A Gene Targeting Nuclease Based on the Homing Endonuclease I-HjeMI

<400> SEQUENCE: 33

```
Met Gly Ser His Met Aep Leu Thr Tyr Ala Tyr Leu Val Gly Leu Phe
1 5 10 15
```

```
Glu Gly Aep Gly Tyr Phe Thr Ile Ala Lye Ala Gly Lye Tyr Leu Ann
20 25 30
```

```
Tyr Glu Leu Gly Ile Thr Leu Ser Ile Lye Aep Ala Gln Leu Ile Tyr
35 40 45
```
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 65 70 75 80
Lys Asn Phe Ile Leu Pro Ile Phe Asp Lys Tyr Pro Met Leu Ser Asn 85 90 95
Lys Gln Tyr Asp Tyr Leu Arg Phe Lys Asp Ala Leu Leu Ser Asn Ile 100 105 110
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 Asn Thr Ser Tyr Phe Ser Ala Trp Leu 130 135 140
Val Gly Phe Ile Glu Ala Glu Gly Cys Phe Thr Thr Tyr Lys Ala Ser 145 150 155 160
Lys Asp Lys Tyr Leu Thr Ala Gly Phe Ser Ile Ala Glu Lys Asp Gly 165 170 175
Asp Ile Leu Ile Ser Ala His Lys Tyr Leu Ser Phe Thr Thr Lys 180 185 190
Pro Tyr Lys Asp Lys Thr Asn Cys Ser His Leu Lys Val Thr Gly Val 195 200 205
Arg Ser Val Asn Asn Val Val Lys Phe Ile Glu Ala Pro Val Lys 210 215 220
Leu Leu Gly Asn Lys Leu Glu Lys Leu Thr Ile Lys Glu Leu 225 230 235 240
Arg Lys Ile Ser Arg Tyr Ser Glu Lys Ile Glu Pro Ser Asn Tyr 245 250 255

<210> SEQ ID NO 34
<211> LENGTH: 912
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:  
<222> OTHER INFORMATION: Nucleotide Sequence of I-Orn, Codon Optimized for Expression in E. coli
<400> SEQUENCE: 34
atgtagcgct acatgtcccg tcgcagttcc atattaacctg gsgatctcag cggtttcgc 60
gacgoggaag gttctcttttt gttcgacatc gcgcacacaca acaagtcceag cgtcggctac 120
tccactgagc tcgcgtcat ttaactcact tataacaggg acaaggtcat ttctgtgcc 180
atccccgctcatgtaggg tggggctgat gcggacacgg ggctgacacgt cgtcggctg 240
aagttccgac gttgtgagga cctgaggtgc attgagcc atttgaaaaa atacccactg 300
attaagcaga cgctcgtgg aatcagctgcc tttaagcagg cggtttgctg cttgccgacc 360
aaggagatt gtagagattg ttttagcttg aagagcagac gcctagaagag aatgctcagat 420
tggggctgct cgagaggttat gcggagcc tgttcggagac tctctctcag ggatcgctcc 480
tctacataac gcagatcttc tatttccttg ctgggtgccg ccgttctctgc ggcggcaaggt 540
tggtttcttt ctatcgtgga caagacag tcggcgctcg gctatgtcgg ccgctgtgg 600
tttcgtctttt aacagttcag tatagagaca gctatgtcgg cggctttc 660
aaggtggtctg ctagagctg gataaagcag ttcgccggctgt ccgtttgctg cctggtctg 720
aagagatcct ogagatcctog acgagttcctg ctgttgcact gcctgtgctc 780
ggcgtgaagctcgagacctcgggaattggtgctgtaagctgtgtctaaagctgtatacggagaaaa 840
aagccacctgasaagctggctctggagcagatcaagaggtttacatgaaggtatcttgagaa 900
ggccaggtatct 912

<210> SEQ ID NO: 35
<211> LENGTH: 998
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nucleotide Sequence of MegaTAL:5.5 RVD + Y2
I-Ami

<400> SEQUENCE: 35

gttgatctacgcggcgctggctcagcagcagcaggaagatcaccgcaagctcggagttg 60
cgtgtacagctggtgagcagttctgtcgctgccattggttcagtttcggtttcctgaaac 120
atcgttgagctctgccactacgctccggcagcctctgggaccccgccgtatctacggtct 180
atatccttcgggtcgtgcagcagagaagagaggtttgtggagtgatccagagctttgg 240
tcggccagcgcgtcgtgttcaggagatctgggttgagcagcctcctgtccgcctttttg 300
atcagttggtgcacagcagatcagctctcccgcccttcagttttttctgagtttttcttgccttt 360
bagaiagtacagtgaggttgtggtctttggttttcagttttttctgagtttttcttgccttt 420
catgcagcagagattgctcttttttcttcagttttttctgagtttttcttgccttt 480
c tgtggtcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 540
tgcacagtctagccagatcagctctcccgcccttcagttttttctgagtttttcttgccttt 600
caggacatcgctcttgctggcagctctcccgcccttcagttttttctgagtttttcttgccttt 660
cagcagtctgctcagaagcagctggttcagttttttcttcagttttttctgagtttttcttgccttt 720
cagcagctgctcagaagcagctggttcagttttttcttcagttttttctgagtttttcttgccttt 780
cagcagctgctcagaagcagctggttcagttttttcttcagttttttctgagtttttcttgccttt 840
atcgtcagcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 900
cgtgtcagcagtagtctagccagatcagctctcccgcccttcagttttttctgagtttttcttgccttt 960
gggcagacaagaagcagctggttcagttttttcttcagttttttctgagtttttcttgccttt 1020
gcggcagatctcagctctcccgcccttcagttttttctgagtttttcttgccttt 1080
gcagttgatctagccagatcagctctcccgcccttcagttttttctgagtttttcttgccttt 1140
gggcagacaagtgcttctcagcgtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1200
gccgtatgtctggttctctctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1260
tacttctacagctcccatacggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1320
atgagcagctctttgctgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1380
gttggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1440
aagttctcccatactggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1500
ttcgggtcctaatacttttgttgcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1560
tccgtatagctccgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1620
gttgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1680
ttcgggtcctaatacttttgttgcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1740
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1800</td>
<td>ttctacta aggttaacct agacaagact aattgtagca aattgaagig cactagtgtt</td>
</tr>
<tr>
<td>1801</td>
<td>1860</td>
<td>aatacggtc agacaagact aattgtagca aattgaagig cactagtgtt</td>
</tr>
<tr>
<td>1861</td>
<td>1920</td>
<td>aatacggtc agacaagact aattgtagca aattgaagig cactagtgtt</td>
</tr>
<tr>
<td>1921</td>
<td>1980</td>
<td>aatacggtc agacaagact aattgtagca aattgaagig cactagtgtt</td>
</tr>
<tr>
<td>1981</td>
<td>1998</td>
<td>aatacggtc agacaagact aattgtagca aattgaagig cactagtgtt</td>
</tr>
<tr>
<td></td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>His Gly Leu Thr Pro Asp Glu Val Val Ala Ile Ala Ser Asn Gly Gly</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>Gly Lys Gin Ala Leu Glu Ser Ile Val Ala Gin Leu Ser Arg Pro Asp</td>
<td>325</td>
<td>330</td>
</tr>
<tr>
<td>Pro Ala Leu Ala Leu Thr Asp His Leu Val Ala Leu Ala Cys</td>
<td>340</td>
<td>345</td>
</tr>
<tr>
<td>Leu Gly Gly Arg Pro Ala Met Asp Ala Val Lys Lys Gly Leu Pro His</td>
<td>355</td>
<td>360</td>
</tr>
<tr>
<td>Ala Pro Glu Leu Ile Arg Arg Val Asn Arg Arg Ile Gly Gly Arg Thr</td>
<td>370</td>
<td>375</td>
</tr>
<tr>
<td>Ser His Arg Val Ala Ile Ser Arg Val Gly Gly Ser Asp Leu Thr Tyr</td>
<td>395</td>
<td>390</td>
</tr>
<tr>
<td>Ala Tyr Leu Val Gly Leu Tyr Gly Gly Asp Gly Tyr Phe Ser Ile Thr</td>
<td>405</td>
<td>410</td>
</tr>
<tr>
<td>Lys Lys Gly Lys Tyr Leu Thr Tyr Glu Leu Gly Ile Glu Leu Ser Ile</td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>Lys Asp Val Gin Leu Ile Tyr Lys Ile Lys Lys Ile Leu Gly Ile Gly</td>
<td>435</td>
<td>440</td>
</tr>
<tr>
<td>Ile Val Ser Phe Arg Lys Arg Asn Glu Ile Glu Met Val Ala Leu Arg</td>
<td>450</td>
<td>455</td>
</tr>
<tr>
<td>Ile Arg Asp Lys Asn His Leu Lys Ser Lys Ile Leu Pro Ile Phe Glu</td>
<td>465</td>
<td>470</td>
</tr>
<tr>
<td>Lys Tyr Pro Met Phe Ser Asn Gin Tyr Asp Tyr Leu Arg Phe Arg</td>
<td>485</td>
<td>490</td>
</tr>
<tr>
<td>Asn Ala Leu Leu Ser Gly Ile Ile Tyr Leu Glu Asp Leu Pro Asp Tyr</td>
<td>500</td>
<td>505</td>
</tr>
<tr>
<td>Thr Arg Ser Asp Glu Pro Leu Asn Ser Ile Glu Ser Ile Ile Asn Thr</td>
<td>515</td>
<td>520</td>
</tr>
<tr>
<td>Ser Tyr Phe Ser Ala Trp Leu Val Gly Phe Ile Glu Ala Glu Gly Cys</td>
<td>530</td>
<td>535</td>
</tr>
<tr>
<td>Phe Ser Val Tyr Lys Leu Asn Lys Asp Asp Tyr Leu Ile Ala Ser</td>
<td>545</td>
<td>550</td>
</tr>
<tr>
<td>Phe Asp Ile Ala Gin Arg Asp Gly Asp Ile Leu Ile Ser Ala Ile Arg</td>
<td>565</td>
<td>570</td>
</tr>
<tr>
<td>Lys Tyr Leu Ser Phe Thr Thr Lys Val Tyr Leu Asp Lys Thr Asn Cys</td>
<td>580</td>
<td>585</td>
</tr>
<tr>
<td>Ser Lys Leu Lys Val Thr Ser Val Arg Ser Val Glu Asn Ile Ile Lys</td>
<td>595</td>
<td>600</td>
</tr>
<tr>
<td>Phe Leu Gin Asn Ala Pro Val Lys Leu Leu Gly Asn Lys Lys Leu Gin</td>
<td>610</td>
<td>615</td>
</tr>
<tr>
<td>Tyr Lys Leu Trp Leu Lys Gin Leu Arg Lys Ile Ser Arg Tyr Ser Glu</td>
<td>625</td>
<td>630</td>
</tr>
<tr>
<td>Lys Ile Lys Ile Pro Ser Asn Tyr Val Asp Arg Ala Ser Tyr Arg His</td>
<td>645</td>
<td>650</td>
</tr>
<tr>
<td>Leu Tyr Pro Tyr Leu Phe Cys Phe Ser</td>
<td>660</td>
<td>665</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 37
<211> LENGTH: 4146
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Engineered type II CRISPR system for human cells; expression format and full sequence of cas9 gene insert of Mali et al.

SEQUENCE: 37

gccacccattg acaagaagta cttccatggg ctcgatatcg gcacaacacag cggctgcttg 60
gcggctatta cggcgatgta caagggtgcgc agcagaaaaat tcgaagttctt ggccaatacc 120
gcatgccaca gcataaagaa gaaacctatt ggccgccctc ttgtgcaacttc cggg*gagcag 180
gccgaagccca cgccgctcaaa aagaacagca cgccgcatat acaccgcaag aagaatcgg 240
gatgcatac tcgagggattc cttttgaatg cggaggtctg cagcgagcat gctcctctt 300
catgagtcgg agaggtctctt ttggtggaag gggatataaa agcgagagcg caacccacatc 360
ttggccatata cgtttggaga cggcgcttcat cattaaagat acacaacact atacacatcg 420
aggaagaagc ttgtagacag tacctgataag gctgacttcg gcgtgatctg tctgctgctg 480
ggctatcca tcaatattcgg ggcacacttc cctcatggag gggacctcgc ccacacaaccc 540
agcggatgctc acaacaactct tctcacaacctg tgtgagcctc aatcactcgt tttgaaagagc 600
aacccgacta aacgctagcg agttggccgc aacccacact cttgacgtagtt cggcgcaac 660
tcggcgccgct agggaacact ctcgctggggg aagagagaaag ggcggcttgtt 720
gccgtatttc tcggcttctc agctggggctc acacccaaact ttaaatctct cttgctgcttctg 780
gccgaagatgc ccagcctttca aatgcagcaga gaaacctcag atgatgacct gcacacttgtg 840
cggcggcagc tccgacgacg aatcttttttg ccggataagc agctgacact 900
gcattttcgc tgtgagatct tctgagctgct aacccgagga tcccaaaactc tcgcgtgcag 960
gcatagta tccacgcttgc tgtgagcagc caccaaacact ctaaatgttct gaaaagccctt 1020
gtcgcgcagc aacccgctga gaaggataag aagaaaaacttc tctagtctgtc taaatatg 1080
taccgctagc aaccgcaggg cggagaactt ctatcaaatc tataacaggccttt 1140
atggtggaaa aataggaagcg caccggagag tgtgtggttta agctttaaag aagagactctg 1200
ttcggcacac agcgcacact ttgcagaagc acatcccccc accagatctca cttgcccggac 1260
cggcgcttca tctccagggc ccagagagct ttctactcct tgtgagaga aacccagagg 1320
aagattgaga aatcccttact attccgctaa cctaatctag cggcccctct ccggccgggaa 1380
aattcgttagct tccggtgagtt gcctgcaaatacht ctatccactt cggcgcagct 1440
ggggaacttg tgtgttaagg gcctttggcc cgcttcctca tccagaggt gataaatcct 1500
gttaaaaaaa atgcctagga aagttctctt cttataacagt ctgctgatt tgcagaacatc 1560
acagtttata acggctacgt caggccctaa ttcgctcagag cggagttgag aagacccagca 1620
ctcgtcgcgt gaggagcagaa aagagctacgt tggagctcctt ttctcaagag cgcggggaaa 1680
gttacgctga aacccgctttg aatgacttata ctaaagactgt cttggaatggt gcgtcgtgtt 1740
gaatcagcg gatggagga ctcgctttcc acatccttgt gcacgtactta aagttctcttg 1800
aaatctattt aagcaactttat ctttttgac acatccgagac cggagcacat tctgagggac 1860
attctcccct gccatagctt gtttgaagag agggagattgc ttgaagacgg ctttgaaact 1920
tcgcgttact tctctgagca aagagctctg aacccgcttc gcgcgcggcg aagccagagc 1980
tggttgccggc tggggatagaga atctttttat gtggagccag cgcagagcag ccagagac 2040
atctctgtatt ccagacgctt gcatggattt gcacaacgaga actccagtgtc gtttgacccct 2100
-continued

gatgacttc tcaaccttaa ggaggcacg cagaaacac aagtttctgg ccagggggac 2160
agtcttcaag agcatacgc taaccttgc agtgaacccgc catcaaaaa gggaaatctg 2220
cagacgtta agtgtgttga tgaacctgtc aaagttaatgg gaaggcataa gcoocgaaat 2280
atcgctagct agatgtgcgc agagacccca actacgctga agggcagata gaacattga 2340
gaagaaatge agaggattga agaggtata aaagacccgg ggctcccaat ctttaaggas 2400
caccaagtgg aaaaacccca gcttcagaaat gagaagcttc aaccgtacta ctgtgcaaaac 2460
gggagggcaag tgtaagttgg tcaaggaacct gacatcaact ggtcttccag ctaagcgttg 2520
gatcatactg tgccctggag ttttttctaaa gatgttctta tgtataataa agtgttgca 2580
agatcctgaga aaaaataggg gaaaggtctg aacgttccct cgaagaagat tgtcaagaaa 2640
atgaaaaatgt attggcgccaa gctgcgtaaac cgcaaaactgta tccacccagac gaagttcagat 2700
aatctgacta aggtgtagac aggtggctgtg tctgtggttg ataaagcccg cttctccccaa 2760
agagctagtt tggagacacag cagactaccc acaagctttg cccaaatctt gccatcagc 2820
atgaaacacca aagtctgatca aaaaacccaa cttgattcag aaggtgaaagtt tattactctg 2880
aatcctgagc tggctccacag ttttaggcag aacgttccct ctgctctctc 2940
acaatatta accatgcgca tctagtgtac ccgagctgag tggtagggac tggacctttac 3000
aaaaaatata ccaagctgttt aacgcttttt gatttggcag acctctgggt ctaaaactgg 3060
agagaaagta tggcagacag tggcagcgaat tgggagaagg gcggctgaag tttgtgttttt 3120
taagcactta tttacagttaa tcggccagcc gaggacacac tggcatactg aagagaacgg 3180
agacgcacc tttgcaaac aacgggaaac acagggaaaa tcggtggtgg caaggtgaggg 3240
gatttcggga cagctgggaa ggtctgtcgc atgtggcggg tgaacatctg taaaagaccc 3300
gagatcagca cggagggcttt cctcagagga aagtctctcc cgaaaaagac cagcggccag 3360
cctggccgac gcacaacaag ttgaggcccc cgaagaacct gcgagggcttc ctctcctaca 3420
gtggctttaca ctgtctctgt tcgagcccaaa tggaggaaggg ggaagttctaa aataacccaa 3480
agcgctcagag aggctgtgcc catcacaac ccgacctggct caaggttccag aaaaaaccggg 3540
acgactctca tcggagccag aaggtatataa gggctcaaaaa aagccctcct cattgccttc 3600
ccagacttat cctctttttc gcctggaacc gcggcgggaa ggtctgtcgc tagtggcggg 3660
gatgctcaga aagggattaa gcggccactgg cctttttttc aagtttaatgt cttgtgtctg 3720
ggcgcacact atgaggcagct ccagggcatg atggcaggaa gcgctgcttc 3780
gttgaaacac acaaaactct cctgttgacg atcatttgac aaaaacgagc aatccctcaca 3840
agatgtgattc tgccagcaca taaccttgtg aaggtgtgttt tccctttacaa taagccacag 3900
gataagccca tccggagcag gggcagaaac aatccctctg ttgaccttct gcaaaacttg 3960
ggagctgctgg cagcctcaac gtttttgcag aacaacctag acaagagccg gataactctttct 4020
acaaaaaggg tctcggccag cacactgattt ctaagctcaaa ttcgacggct ctttgaacac 4080
agatctgacc tctcctcgct cgggtgagac aagccggtctt ccctccagaa gcagggagag 4140
gtggta 4146

<210> SEQ_ID NO 38
<211> LENGTH: 455
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nucleotide Sequence of RNA Guide Strand for use with Cas9 Endonuclease
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (320) .. (338)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 38

```
tgtacaanaa agcagccctt aagaagcca attcagtcga ctggactccg taccoagtc 60
ggcaaggag aaggctatt tccccatgatt cctctcatatt tgcataatc atacaagct 120
gtagagagaa taattagaaat taattagact gtaacacca aqatatattgc acaaaatcg 180
tgactcata tggctacgct aacttgaagat taattgcatg tcttgccttt atatatatctgg 240
gacatacata tggctacgct aacttgaagat taattgcatg tcttgccttt atatatatctgg 300
tgcaagac ggaaacogn nnnnnnnnnnnnnnnnnt ttagagctga aataagcattct 360
gttaaaaatatgg ggctactgct tttcaaccttt gaaacagctt caccagcttgc gtttttttttt 420
tttcaaccttt gctttcttgtt acaacaagttgg catta g 455
```

<210> SEQ ID NO 39
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: BCL11A gene target site

<400> SEQUENCE: 39

```
tttaagtctgt ttcctggttgg tg 22
```

<210> SEQ ID NO 40
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Targeted Homing Endonuclease Target Site

<400> SEQUENCE: 40

```
tttcaacatttt ttctacccct 22
```

<210> SEQ ID NO 41
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Targeted Homing Endonuclease Target Site

<400> SEQUENCE: 41

```
tttcacttatatttt ttctacccct 22
```

<210> SEQ ID NO 42
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Targeted Homing Endonuclease Target Site

<400> SEQUENCE: 42

```
tttcaacatttt ttctacccct 22
```

<210> SEQ ID NO 43
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Targeted Homing Endonuclease Target Site  
<240> SEQUENCE: 43  

```
tttcattta ttcaatatct tt
```

<210> SEQ ID NO 44  
<211> LENGTH: 12  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: U6 Promoter Sequence of a Generic Cas9 Guide RNA  
<240> SEQUENCE: 44  

```
ggacgaaca cc
```

<210> SEQ ID NO 45  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Generic Target-specific Sequence of a Generic Cas9 Guide RNA  
<220> FEATURE:  
<221> NAME/KEY: misc_feature  
<222> LOCATION: (2)...(20)  
<223> OTHER INFORMATION: n is a, c, g, or t  
<240> SEQUENCE: 45  

```
gnnnnnnnn nnnnnnnn
```

<210> SEQ ID NO 46  
<211> LENGTH: 77  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Guide RNA Scaffold Sequence of a Generic Cas9 Guide RNA  
<240> SEQUENCE: 46  

```
gtttaagc tagaataagc aagtttaaat aeggctagtc cgattcaac ttgaaaaagt
```

<210> SEQ ID NO 47  
<211> LENGTH: 10  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Poly T Tail of a Generic Cas9 Guide RNA  
<220> FEATURE:  
<221> NAME/KEY: misc_feature  
<222> LOCATION: (1)...(9)  
<223> OTHER INFORMATION: n is a, c, g, or t  
<240> SEQUENCE: 47  

```
nnnnttttttt
```

<210> SEQ ID NO 48  
<211> LENGTH: 23  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Target-specific Sequence for Cas9 Guide RNA
GGN20GG-B

<400> SEQUENCE: 48

gcatttctaa tatacgact tgg

<210> SEQ ID NO 49
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Target-specific Sequence for Cas9 Guide RNA
GGN20GG-C

<400> SEQUENCE: 49

gctgggcttc tggtagcaga tgg

<210> SEQ ID NO 50
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Target-specific Sequence for Cas9 Guide RNA
GGN20GG-D

<400> SEQUENCE: 50

gaaatggga gacaatagc tgg

<210> SEQ ID NO 51
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Target-specific Sequence for Cas9 Guide RNA
GGN20GG-E

<400> SEQUENCE: 51

gaataattca agaaggtgg tgg

<210> SEQ ID NO 52
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Target-specific Sequence for Cas9 Guide RNA
GGN20GG-F

<400> SEQUENCE: 52

gatattgaat astcagaggg agg

<210> SEQ ID NO 53
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Target-specific Sequence for Cas9 Guide RNA
GGN20GG-G

<400> SEQUENCE: 53

gcctgagatt ctgatcagac agg

<210> SEQ ID NO 54
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Target-specific Sequence for Cas9 Guide RNA
<220> FEATURE:
<223> OTHER INFORMATION: Engineered target site

<400> SEQUENCE: 54
<210> SEQ ID NO 55
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Engineered target site

gttcaatcttaaagcagtgattgtaaat

<210> SEQ ID NO 56
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Engineered target site

<400> SEQUENCE: 56
<210> SEQ ID NO 57
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Engineered target site

<400> SEQUENCE: 57
<210> SEQ ID NO 58
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Engineered target site

<400> SEQUENCE: 58
<210> SEQ ID NO 59
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Engineered target site

<400> SEQUENCE: 59
<210> SEQ ID NO 60
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
Continued

OTHER INFORMATION: Engineered target site

SEQUENCE: 60
tttaggaggt ttcctggtgta at 22

SEQ ID NO 61
LENGTH: 22
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Engineered target site

SEQUENCE: 61
tttaggaggt ttcctgggtg g 22

SEQ ID NO 62
LENGTH: 900
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Nucleotide Sequence of a I-CpaMI homing endonuclease; ORF, codon optimized for mammalian expression

SEQUENCE: 62
atgacacca gctctaggtt caatccctgg ttcctgaccc gctttagcga tgcagagttcgc 60
cttttctgaac ggcacacg ccagtcctct cgccttggag tgcacagc 120
gtctctgca acaagggagcc ctgaggtttc tcagagatg ccagagatctat 180
cctggccggg gcatctagct ctgggggagcc cacagccagc ttcagttcag gattgacagc 240
cctggccgc gtagggtgtgcat ctcacacac ttcgcagact accccttggg aaccgccaaag 300
tgggccgctt aaccctttc tggagaggtg cattggtgaa ttctctgtaa ggaagcaagtg 360
agccagaagg ggtgttttaa actgtgtaggc attaaggccga gctgatagtct cgggtggagttcc 420
ggacagctca aagggcggttg ccagacagcc ggaagactgc gacgtgactagc 480
ggttcagg gcatcggagc cccacagttg atcagcggct tgggtgcagg cagatagcaco 540
tttatgtga aaatactgacc ccctccgcaac tgcctgtgct caataaaaggt gcagcttcag 600
ttggcagcat gttgagattc cagagagaa ggtcttacct ataatccggtt ggtctttctct 660
ggctctgtgc acatcctgaa gacatctcct tgcagctgat acgcggccagcatcgaggggtg 720
gtgaagctg gggatctcag cggacaagat atccctttct cgcacacgtat cagatacaac 780
ggcacaggg tggggactca ctcacagttg cgggaaggtgc cctagcattc cagacgggaag 840
aagcaatctttactcaggg ggattcagaa tcttgacgcct taaaagcagc atagaacaag 900

SEQ ID NO 63
LENGTH: 360
TYPE: PRT
ORGANISM: Cryphonectria parasitica

SEQUENCE: 63
Met Asn Thr Ser Ser Ser Ser Phe Asn Pro Trp Phe Leu Thr Gly Phe Ser
1 5 10 15
Asp Ala Glu Oys Ser Phe Ser Ile Leu Ile Glu Ala Asn Ser Lys Tyr
20 25 30
Ser Thr Gly Trp Arg Ile Lys Pro Val Phe Ala Ile Gly Leu His Lys
35 40 45
<210> SEQ ID NO 64
<211> LENGTH: 5264
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nucleotide sequence of BCL11A gene targeting nuclease-encoding plasmid pHExodumBCL11Ahje
<400> SEQUENCE: 64

gacgagtcgg gagacagccc gatcagcatc atccatagc tgcctagtcttc 60
cgccagtt ttaggaggtg cagttgttggt gcgtgtctgct gagtagctgg 120
cgtggcacaat ttagatgtggc gatgtgctg gatgtgctgct gatgtgctg 180
ttaggttgctg ctagtggcagctgc atctctctct cttcgccccc 240
gatgtgtgctg ctagcagctgc ctagcagctgc ctagcagctgc ctagcagctgc 300
ttagacagctgc ctagcagctgc ctagcagctgc ctagcagctgc ctagcagctgc 360
cctcctgctg ctagcagctgc ctagcagctgc ctagcagctgc ctagcagctgc 420
ttagacagctgc ctagcagctgc ctagcagctgc ctagcagctgc ctagcagctgc 480
ttagacagctgc ctagcagctgc ctagcagctgc ctagcagctgc ctagcagctgc 540
atgcgccagta catgacotta tgggaactctc ctaccttgcca gtacatacct actattagc 600
tgcttatata cattgtctag gttttttggc agtacactaa tggcgttgga ttagctgttg 660
aactcaaggg attttaaat cttcaacccaa ttgaoyctcaaa tgggagtttg ttttggccac 720
aananacac gcgtcttc tataaatgtta ataatccacca ctacagtgttt gaaatattgca 780
gtanaggtgt agcttgaggg gattatataa gcagactctc ctgctgtata acagaaacac 840
ctgctctactg cttatagcaaa attataacag ctcctatata gggacaccaaa gctgggctac 900
gtttaaacctt aagcctcgta ccctactcag atcccactgt ccggtcttgt ggaattttcg 960
aggtcacttg auggcocgac cattggatat cctactatgt tcccatgata tcgcaccatc 1020
aagaagaca gcaaaaaccg cggggcgaag ccaagggcag tsgctacgct ctaactggct 1080
ggcctgtcg aggggcctcg cttatcttaca atgcttaagc cggccgagtt tctgacactc 1140
agctggggca ttcactcttc cacattgaca gctcagcttc ttcataaat caaggaactc 1200
cctggctggt gcctactgtg cttttgaggta acagctgttc acggcgacgc 1260
atcggagcactacaacacccactc gagaactacctt attcgtggcaact ttaccccaa gtagcctatg 1320
cggacccaca agccgtcagca tcaactgcgaa ttcggattgc ccctctctgct ccacactac 1380
tataacgcag acctggccaa gtaacccaggt aagcaagatg caactaattg cgtgacccagc 1440
atctcacaactctcatac tccgctcttg cttgtgcttc tcctactgac ccgaggtctgc 1500
ttcaccaacct cccagccactt ccaagacact ttcgtgcaag cggcttcttc ctatagccac 1560
aagggcgggcc acatrctctact cctggctgcac ccacaaacac ccgagctcagc gacacaaacac 1620
tacacacaca cagcccaagct tagccacactt aagttccocgc gcctgagggag cgtcaataac 1680
gtggctattaag tcccgtactgc cggcctcttg agctgcttggt gtaacacaca gtcgactatc 1740
aaactttgga taactgcaac gcggcaagctt ccctgatactt gcagagaaaa ccagctgccc 1800
agtaactct acctcctacag ggcccttttag aacctgtcgt taccgcctcag ctgggtctct 1860
agtgcgocact actcaggtgt gttccctctcc cccgctgttt cctagtacact gggagrggcoc 1920
acctcccactt ccctctcccc ataaatagcg gaaattgcct cgcattcgct gatagggtgtg 1980
cattctcctt cttggggggcg gttgggggcc gcagcggagc ccgaggtttg gcgaageaat 2040
agcaggggtat ctgctcgcgtc gttgggctctt atgctttctg gacggaasag aacccagctgg 2100
ggccttagg cttacacccgc ccgctcggatg aggcggcgat taaagcccgg ccggtggtgtg 2160
gtatacgcca cctggctggc tattcttcgg agcgccttct gcgcctgctc ttcgggtcttc 2220
ttcctctcc ttcctgcgtcct ttcgggcggtt aagcttctaa tcgggggtctc 2280
cccctttggt ttcgatttgc tgtttacagg cactctgacc ccaaaaaact tgtattagggt 2340
gatggttcaac atcggcggcc ctaactgcgag ttagctgttt ttcggtcctt gcgtggggag 2400
tccagctcttt ctaatcttgg accttcttct ccacagcggc ccaactctcata ccctactctg 2460
gttattttct tgttttata agggtatttg ccggaggctgc cttatgtttg aaaaatagcg 2520
ctgattatac aaaaaatca cggcaatttt ctggtgacat ttttttttggct 2580
sgaagctcgg aggctccca gccggcagaa gcagctccgaat cctactcttc catactttgct 2640
getacggctgt tattctggta agcactgctt ggccacaggc ctcacgcctg ctcgttgcctt 2700
taacgctctg cgcctccttc ttcggccgctt atgcctttct atgcctttct tgaagctgtcc 2760
tctgtggcgc gaccccggg gttcgaataa cccaacagac gcaccccaac ctcgcataac 2820
gagatttccga ttcaccgccg gccttctatg aaaggtttggg cttcggaaatc ttttccccggg 2880
acgcgcgctg gtcttgcttc cagcgcgggga atctcagctc ggagggcttc gcaccaccaca 2940
acttttgatt tggagtcttt aatggttca aataaaagca ataactctaca aatcccaaat 3000
ataacgcat tttttcgctc caactctgtg ggttgcttgc cactactaca aatgtttctt 3060
atcagtgcctg ttaacctcgct gaccttcgct agagtttgggg gcatactttg tagatgtgtg 3120
ttcctgtgtg aatattgttat cgcctcacaact ttcaccacacata cagagagcgc ggaaggtcat 3180
agtgtaaacg ttgggtggcgc taaatgtgtg aatcaactac caataatgtgtcttgctgctac 3240
tgcggcggctg ccgctctggg aacctgtgctg gcagctcgtga ttatgtgaact gcaccaccgcg 3300
cggggaggg ggttgggtgt atggggtggtct ctcgtcttcotcctgtgcgct gcagcggtgctg 3360
getcgctcgtg cgtcggctggcg gagccggttat cgcctcactc aaagcgtgta actagggttat 3420
cacggaggcc aggggttaac gcggagaagac acatgtgagc aaaaagccag cccaaagggca 3480
ggaagctcgcc aagcgtgctgg tgtgctgggt tttcctctagct gcctggcgcctg cctgagcgac 3540
atccacaaaa tgaactgctca agtagagaggt gcggaaaccc gcaggagacta taaggtacct 3600
agggcggtctc ccogttggagc cttcctgggtc cttctccgtc ctgctctgtct gctcgacggt 3660
atatcctcgg gctcggcctgct cttcggggga gcgggctcgt ctcctgtcag tcaogctgtga 3720
gettiactgg acctgggtgg ggtgctggctg ccgggctgtg gcctggctgac gcaccacggcg 3780
tgcgtgccga cgcgtggcgc ttcctggctct ctttccttggct gatgttctccgcc gctgggttga 3840
acagacattc actcgtgctcc gacgcagcgtc gcagagctacgt tgaactactgt aaggttgctgct 3900
gctgtgccag aaggtctcgtg aaggtgtggcg ctaactcggct ctacactaga aagacgttatt 3960	tgcgtactgc gctgtgtggcc cctggtcccgt ttctctctaaag aaggtgggtgt gcgttctgtg 4020
cggccaaaac aacccagcgtg tgtactgcgt tttttgttgg cagccagcag ataccggccac 4080
gaaaaaaag atccacagaa gactccggtta cttctcttc aggggctgtg cgcgtttggg 4140
acgaaaaactc aggttaaggg atttgggtcct gtagatttct ataagaggtc ttcaccctaa 4200
tctcttttttt tcacaaatg agttaaaatc catttttaaacttataagtga taactctggt 4260
tctgagggtgtt ccacgtttggtat gtagggtcct gcgattttct cctcgtctctt 4320
catcctaacgt tgcggtactc cccgtgcgtgt agataactc gtcaggggag ggtttacctc 4380
cgggccccg gttgctggctc ataccgcgag acacccgctc acoggtctcga gattctacgt 4440
cataacaaaa ccgcgcggagc aagggcggac gcggagtttg gcctgcctact ttcctggcct 4500
cctcactagc tatttttgtg cttgcgggag atcaggaatagg cctcgtgga caatatagtgtt 4560
tgcgtgcatgt tgtggcctgt gcgaagcgcgt cgctgctgctg tgggtggttg 4620
etctttcgtctc ctgctcgcac gcaggtcctctc atgcctcgaat atgcctgttgc 4680
aaaaaggtttg tagctccgttc ggtctccgct tcagtgtgactaatagtgct gcggcagtgt 4740
taacagctcgt tggattggca gcagctgctca atttccttac tgcattaactc ttcggtaaat 4800
tttttttgtg gaagggtgaag tatactactctc aacgattgtgt gcattgggcag 4860
cgtggttgctgt ctcacgccac acagctcagct aagattgctg atgggggcatc 4920
aaggttcgct cattgggaaa ggtttctcgg gcggaaaaact ctcagggactc ttcgctgtg 4980
tgggtcttgac gttgtgccttc gcccagcttg ccccaacctg atctccgaca ttttctactct 5040
tcccacagctgt ttcggggggc gcacaaaaag gaggccaaat gtcggccaaa aaggggaaat 5100
-continued

gggcgcacg gaaatgtgta atactcatac tcttctcttt ttcaaatatt ctaaagcatttt
5160
atcagggag tttgttcatg acgcgataca tatttaagat tatattagaa aataaaacaa
5220
taggggtccc ggcacatttt cccgaaaaag tgcacactgga cgtc
5264

<210> SEQ ID NO: 65
<211> LENGTH: 5100
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Nucleotide sequence of TREX2-encoding plasmid pNodus CMV-Tre2

<400> SEQUENCE: 65

agaagagacg gagatccccac gatcctcctat ggtgcaactc tgcctgctgtg
60
cgcctagct aagacgtact ctgtctctctt ttctgtctgatt ggaggtcgtc gatcagttgtggc 120
cgcagcaaat ttaagctcta acaagggccag gttgagcgga caatcgtcag aagatcctgc
180
tttagggttag ggtattcgtg cttcgttgatc atgtcagggc cagatataac gcgtgcattat
240
gattatagcc tattatctaa tattaattccc ttaagggtggc attatgctcat aagccatttata
300	ttaggtctcc ggtttgtatt cttcctaatcc aatgcggagct ggtgatcggcc ccaacgcacc
360
cccgcccacactt caagcgttaa aatgacagct ccccatttatt aagccgattc ggcattttcc
420
ttgaagccct aaggtgcact atttacagcc aaattgagct ctttgagact gtaaagtgtg
480
atcatactgcc aagtgcctcc caaatgggct cttgacggtt gataattgccc gcttggctg
540
taatgagga cttgcgtcgt cttgagacac tcaatgtgga gacatgctcc gttctgagcct
600
tcgttctcct catctgttag tgcagagatt attttttttcct ttcgacgcttct cagagttgata
ttcgagctcatc cttgacgcttct ttcgagctcatc cttgacgcttct ttcgagctcatc
ttcgagctcatc cttgacgcttct ttcgagctcatc cttgacgcttct ttcgagctcatc 1020
gagaaccacc caggggttta tccttgctcc tttggtctgc ccctgtttct ggcacggtctc
1080
aacattgtca tggccggtcag ggcctcccttt atcgcagcg ccagttgagat tattgtggctt
1140
agcagcggaa gctctgtgtca cttgacggag aatggctttc atggcgtcgt gttaaggaca
1200
tcgagctgt tgaatcctgc ccaggggtg ggcctctgg cttgatggc ccaatgtgtcc
ttgtagaact cccacactgc tgcagcggtat cttgagcaga cttgaggctgc cttg ccac
1260
gattatag cttgagagct tggctccggtct caataagcttc tgcagctgtgc tgcagcctgc accatgcacc
1320
gacatgtct gcctgcagct cttgccgttcg tgcagcctgc cttgagacac gacatggctgc
1380
ggcccaggtt caccggggca cagaaatcag ccggtgctgc ttcctcttcc ccgggtctct gacatgtct 1440
caggctgacg ccagttggcc atccttacc gcgtgcctgc ggccagtgat cggccctcct gacatgtct
1500
ttcctctggc gccgtggtcg tggcctgtgg tggcctgtgg tggcctgtgg cagccgctcg ccggtttggtc 1560
cattggtcg gcagttgtct gcggggtgtg ccggtggtcg gcggggtgtg ccggtggtcg gcggggtgtg 1620
catatgggt ctaataagcttt ggcacaggtg ggcagagct tctgagccct aatccttacc gcgtgcctgc
1680
cattcgcgggg cccccctccc cccgggcttt cttgtgttgtg ccccggggttt cttgtgttgtg 1740
gacccotggaa ggtgcccttc ccaactgccct ttcctaaata aatgaggaat ttgcatacgc
1800
ttgcttgaagt aggtgctaat otatttctgag gggtgaggttt ttggcagccaa gcaagggaga
1860
ggattgagga caaaatagca gcacgttggc ggtgctgtta gtgtctatag ctttggagga
1920
ggaagaaccc acgcgagtcg acctggagact ctggagcttc ggcggagcgc
1990
cgctgcgcgt ggtggtctttg acctgaggtgc agcgacgcc ccacgccggc
2040
cgctccttc gttcttcccc ctctctctct cccgcacgct ggcgcgcttt tccgctcagc
2100
tctaatcccg gggctcccc ttaggttccag atttaacctg ttacgccacc tccacccca
2160
aaaacctgta tagggctgatg ttgctaacag cctctataagc cgggtttctg
2220
cctttgagct tgggtgcaat cggctctcag tgggctcttc cctcttccaa tgggcaaccc
2280
acctcaacet atccctgctc ctctctctct cctctctctc cctctcttca
2340
ttgatattaa aatagcgtaa ttttttttttt tttaataggt tttgtgtggct
2400
tggcggtgctt cctggctgcc gcccctgggt ctggagctgc cctgtggtcag
2460
tcgggaatct gcctgctaat tctgtattgc tctctctctc tctctctctc tctctctctc
2520
cocttcctct gtctgctgctt gtacgctgct ccggctctct cctctctctc cctctctctc
2580
cctttccttc ggtctctctc cggctgctt tgggtggtgc aatagaggtc cccgctctct
2640
ccacccactct ccacccactct ccacccactct ccacccactct ccacccactct ccacccactct
2700
ggattgcttt tccgggagcc caggtgctgt gcgtctgcttg cttgatgggct
2760
tggctgccc acacccactct gggtttcgtg caacggattc cggctctctc cctctctctc
2820
atcataaaatt ttcacacttaa acaaacatatt ttcacacttaa cttgctctct tccgctcagc
2880
ctctactatg tacctttcatt tgtctgtgatt cttgctctct cctctctctc cctctctctc
2940
ttcgctctct aggtgcttctat tgtctctctat tgtctctctat tgtctctctctc cttgctctct
3000
cggtcgcgca ctgctgtgctt cttgctctct cttgctctct cttgctctct cttgctctct
3060
atggggtggcg cgttggagct ttcgctctct cttgctctct cttgctctct cttgctctct
3120
tgagcgcgca aacggctgctt cttgctctct cttgctctct cttgctctct cttgctctct
3180
cctgtctctc tccgtgcttc gcgtctgctt cttgctctct cttgctctct cttgctctct
3240
gggtaatact gcgtcttctct cgggttctcgc gcgtctctct cttgctctct cttgctctct
3300
ggctcagcatt ggcctcagcatt ggcctcagcatt ggcctcagcatt ggcctcagcatt ggcctcagcatt
3360
cgccctgcc gcggagtcgca acacggtgctt gcgtctctct cttgctctct cttgctctct
3420
ggcgggggg ggcgggggg ggcgggggg ggcgggggg ggcgggggg ggcgggggg
3480
acccctggct ttcctggggc ctggctgtgt gcgccctgct cgggtttctc cgggtttctc
3540
catagtcctc ggtctgcttc tccagctgct gcgtctgcttc gcgtctgcttc gcgtctgcttc
3600
gtggctcagc tccacccactct ccacccactct ccacccactct ccacccactct ccacccactct
3660
tcgggaatct tccagctgct gcgtctgcttc gcgtctgcttc gcgtctgcttc gcgtctgcttc
3720
agagcaagct gtttggagttt cttgctctct cttgctctct cttgctctct cttgctctct
3780
actgagctct ctggctcagc tggctgtgctc gcgtctgcttc gcgtctgcttc gcgtctgcttc
3840
gcggctgcttc gcgtctgcttc gcgtctgcttc gcgtctgcttc gcgtctgcttc gcgtctgcttc
3900
catagtcctc ggtctgcttc tccagctgct gcgtctgcttc gcgtctgcttc gcgtctgcttc
3960
tcgggaatct tccagctgct gcgtctgcttc gcgtctgcttc gcgtctgcttc gcgtctgcttc
4020
aggtctcca cctgatcct ttttaattaa aatagaaatt ttaaatcaat ctaaagtata 4080
ntagtaaa attggtgtaa caagttacaa atgatattca cttgagcacc tatotcagcc 4140
atctgttcat ttcggatcct catagttgct tgcagcgcg cgcgtgtagt aactacagata 4200
cgggaggtc ttcgatcgc cccagttgct gcaatgatat cggagagcc aggctcagcc 4260
ctgcaagatt tactacagct aaacagccca gcccggaggg ccgcgcgcag agtgctgctc 4320
gcaacccatt ccagcttccct ccagttcctt atatggctgc gggaaagctag agtaagatgt 4380
tgcaagacta tattttgtgc gcaggtgtgt gcactttgta cagggatcgc ggtgctaacg 4440
tgctggttgc gatagtgcct atctcaagttcc ggttccaaac gataaagcgcc agtcatatg 4500
tccccatgt tgcagcagaaaa aagcggtttgcc tctctctgct cctcgtatgg tgcgaagat 4560
aatgtggcgc cagtgttact acactagttg ctcggagcc tgcataatttc tcctcttgcc 4620
attgctacgc ttaagttgct tctctgtgact ggtgagtaact caacccagct atatggagaa 4680
tgtgatacgc ggccccgcag gttgcttctgc cggcgccaa taccggataa taccgcgcac 4740
catcgccagaa ctttaaaagtt gcctcatcct ggtaaacggt cttccggggcc aacaaactca 4800
agggttctac ccggagcgcata cttccagcttc cgctgctcacc caacatcactt 4860
taccgccatct ttcctcttttc cagctggttgc ggtggcagaa aacacgagag geaaaaacctg 4920
gaaaaaggg gataaggggc gcacacgaaat tggtaaataac cttactccttttctttc 4980
ttataaagaa gcactatattt gcggtagcgc cttcgtgctgc gatacatatt tgaatgattt 5040
tgaaaaagaa asccaaataggg ggttccgctgc acatctttttt gcaasagttgc acctgacgct 5100

<210> SEQ ID NO 66
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer

<400> SEQUENCE: 66
gtgatggacgt tgcccacgtaac 20

<210> SEQ ID NO 67
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR Primer

<400> SEQUENCE: 67
cacaacagcag tccacacaggt 20

<210> SEQ ID NO 68
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: PCR Primer

<400> SEQUENCE: 68
cctgccacgct tccaagtttc 21

<210> SEQ ID NO 69
<211> LENGTH: 21
<212> TYPE: DNA
1-96. (canceled)

97. A polynucleotide encoding an endonuclease selected from the group consisting of a homing endonuclease (HE) and a CRISPR endonuclease, wherein said endonuclease includes a nucleotide sequence selected from the group consisting of a BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (HbF) silencing region, a BclI1a-regulated HbF silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation.

98. The polynucleotide of claim 97, wherein the endonuclease is:

a) a HE that binds to said BclI1a coding region or said BclI1a gene regulatory region;
b) a HE selected from the group consisting of an l-Onul endonuclease, an l-HjeMII endonuclease, and an l-CpaMII homing endonuclease;
c) a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region;
d) an l-Onul homing endonuclease;
e) an l-Onul homing endonuclease comprising the amino acid sequence encoded by a variant of the nucleotide sequence of SEQ ID NO: 34 that encodes an l-Onul homing endonuclease that can specifically bind to the fetal hemoglobin (HbF) silencing region;
f) an l-Onul homing endonuclease comprising one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein each of the amino acid substitutions is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82;
g) an l-Onul homing endonuclease comprising 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, 1186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

99. The polynucleotide of claim 97 or claim 98, further comprising a polynucleotide encoding a TAL effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

100. A vector system comprising a vector and:

a) 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 BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (HbF) silencing region, a BclI1a-regulated HbF silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation; or

b) a polynucleotide encoding a Cas9 endonuclease, optionally wherein the Cas9 endonuclease comprises the deoxyribonucleic acid sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 nuclease, and an RNA guide strand that mediates the binding of the Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region, optionally wherein the 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.

101. The vector system of claim 100, wherein said endonuclease is:

a) a HE that binds to said BclI1a coding region or said BclI1a gene regulatory region;
b) a HE selected from the group consisting of an l-Onul homing endonuclease, an l-HjeMIII homing endonuclease, and an l-CpaMII homing endonuclease;
c) a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region;
d) an l-Onul homing endonuclease;
e) an l-Onul homing endonuclease comprising the amino acid sequence encoded by a variant of the nucleotide sequence of SEQ ID NO: 34 that encodes an l-Onul homing endonuclease that can specifically bind to the fetal hemoglobin (HbF) silencing region;
f) an l-Onul homing endonuclease comprising one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein each of the amino acid substitutions is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82;
g) an l-Onul homing endonuclease comprising 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, 1186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

102. The vector system of claim 100, further comprising a polynucleotide encoding a TAL effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

103. The vector system of claim 100, claim 101, or claim 102, wherein said vector is selected from the group consisting of a AAV6, a modified adenovirus vector, an integration-deficient lentiviral vector (IDLV), and an integration-deficient foamy viral vector (IDFV).

104. A polypeptide encoded by 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 BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (HbF) silencing region, a BclI1a-regulated HbF silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation.
silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation.

105. The polypeptide of claim 104, wherein said endonuclease is:
   a) a HE that binds to said BclI1a coding region or said BclI1a gene regulatory region;
   b) a HE selected from the group consisting of an I-Onul homing endonuclease, an I-HieMI homing endonuclease, and an I-CpaMI homing endonuclease;
   c) a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region;
   d) an I-Onul homing endonuclease;
   e) an I-Onul homing endonuclease comprising 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 the fetal hemoglobin (HbF) silencing region;
   f) an I-Onul homing endonuclease comprising one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein each of the amino acid substitutions is selected from the group consisting of T26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82; or
   g) an I-Onul homing endonuclease comprising 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, N186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

106. A composition for the treatment of a hemoglobinopathy, said composition comprising:
   a) 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 BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (HbF) silencing region, a BclI1a-regulated Hbf silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation;
   b) a vector system, comprising a vector and:
      i) 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 BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (HbF) silencing region, a BclI1a-regulated Hbf silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation;
      ii) a polynucleotide encoding a Cas9 endonuclease, optionally wherein the Cas9 endonuclease comprises the nucleotide sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 endonuclease, and a RNA guide strand that mediates the binding of the Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region, optionally wherein the 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; or
   c) a polypeptide encoded by 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 BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (HbF) silencing region, a BclI1a-regulated Hbf silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation.

107. The composition of claim 106 wherein said endonuclease is:
   a) a HE that binds to said BclI1a coding region or said BclI1a gene regulatory region;
   b) a HE selected from the group consisting of an I-Onul homing endonuclease, an I-HieMI homing endonuclease, and an I-CpaMI homing endonuclease;
   c) a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region;
   d) an I-Onul homing endonuclease;
   e) an I-Onul homing endonuclease comprising 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 the fetal hemoglobin (HbF) silencing region;

108. The composition of claim 106 wherein said polynucleotide further comprises a polynucleotide encoding a TAL effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

109. The composition of claim 106 wherein said vector is selected from the group consisting of a AAV6, a modified adenovirus vector, an integration-deficient lentiviral vector (IDLV), and an integration-deficient foamyviral vector (IDFV).

110. A cell comprising:
   a) 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 BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (Hbf) silencing region, a BclI1a-regulated Hbf silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation;
   b) a vector system comprising a vector and:
      i) 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 BclI1a coding region, a BclI1a gene regulatory region, an adult human β-globin locus, a fetal hemoglobin (Hbf) silencing region, a BclI1a-regulated Hbf silencing region, a γ-globin gene promoter, a δ-globin gene promoter, and a site of a β-globin gene mutation;
      ii) a polynucleotide encoding a Cas9 endonuclease, optionally wherein the Cas9 endonuclease comprises the nucleotide sequence of SEQ ID NO: 37 or a variant thereof which encodes a functional Cas9 endonuclease, and a RNA guide strand that mediates the
binding of the Cas9 endonuclease to a fetal hemoglobin (HbF) silencing region, optionally wherein the 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; or

c) a polypeptide encoded by 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.

111. The cell of claim 110 wherein said endonuclease is:
   a) a HE that binds to said Bcl11a coding region or said Bcl11a gene regulatory region;
   b) a HE selected from the group consisting of an I-Onul homing endonuclease, an I-HjeMI homing endonuclease, and an I-CpaMI homing endonuclease;
   c) a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region;
   d) an I-Onul homing endonuclease;
   e) an I-Onul homing endonuclease comprising 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 the fetal hemoglobin (HbF) silencing region;
   f) an I-Onul homing endonuclease comprising one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein each of the amino acid substitutions is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82; or
   g) an I-Onul homing endonuclease comprising 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, N186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

112. The cell of claim 110, further comprising a polynucleotide encoding a TALE effector nuclease (TALEN), a TALE-HE fusion protein, and/or a TREX2 nuclease.

113. The cell of claim 110 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).

114. The cell of claim 110 wherein said cell is:
   a) a stem cell;
   b) 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; or
   c) 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.

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

116. The genome edited stem cell of claim 115 wherein said homing endonuclease is:
   a) a HE that binds to said Bcl11a coding region or said Bcl11a gene regulatory region;
   b) a HE selected from the group consisting of an I-Onul homing endonuclease, an I-HjeMI homing endonuclease, and an I-CpaMI homing endonuclease;
   c) a HE that can specifically bind to a fetal hemoglobin (HbF) silencing region;
   d) an I-Onul homing endonuclease;
   e) an I-Onul homing endonuclease comprising 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 the fetal hemoglobin (HbF) silencing region;
   f) an I-Onul homing endonuclease comprising one or more amino acid substitutions within the amino acid sequence encoded by the nucleotide sequence of SEQ ID NO: 34, wherein each of the amino acid substitutions is selected from the group consisting of L26, R28, R30, N32, S40, E42, G44, Q46, A70, S72, S78, K80, and T82; or
   g) an I-Onul homing endonuclease comprising 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, N186, S190, K191, Q197, V199, S201, K225, K227, D236, V238, and T240.

117. The genome edited stem cell of claim 115 wherein the homing endonuclease is fused to a TAL effector (TALE) DNA binding domain and/or a TREX2 nuclease domain.

118. The genome edited stem cell of any one of claims 115 to 117 wherein:
   a) the correction template comprises a nucleotide sequence that permits the modification of key regulatory or coding sequences within a globin gene locus;
   b) the cell 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; or
   c) 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.

* * * *